Chapter 2

PCR-mediated synthesis of the *P. falciparum dhfs-fpgs* gene

2.1 Introduction

The genetic organisation of *dhfs* and *fpgs* varies greatly between organisms. In yeast three genes, each on a separate chromosome, encodes the DHFS and FPGS functions (Cherest *et al.*, 2000). *Arabidopsis thaliana* (Ravanel *et al.*, 2001) also have separate genes encoding for DHFS and the various isoforms of FPGS. In *E. coli* (Bognar *et al.*, 1985) and *P. falciparum* (Lee *et al.*, 2001) the DHFS and FPGS activities are encoded by a single gene. The *P. falciparum* gene consists of four exons (70% of the mRNA transcript) and three introns (Figure 2.1). Three ATG codons, from which differential translation could take place, are found at amino acid positions 1, 17 and 41 (Figure 2.1). It was shown that the whole length of the sequence (from the first ATG codon) was necessary to complement yeast DHFS-FPGS mutants (Salcedo *et al.*, 2001). It is not known whether transcription of the truncated gene (from codons 17 or 41) takes place *in vivo*.

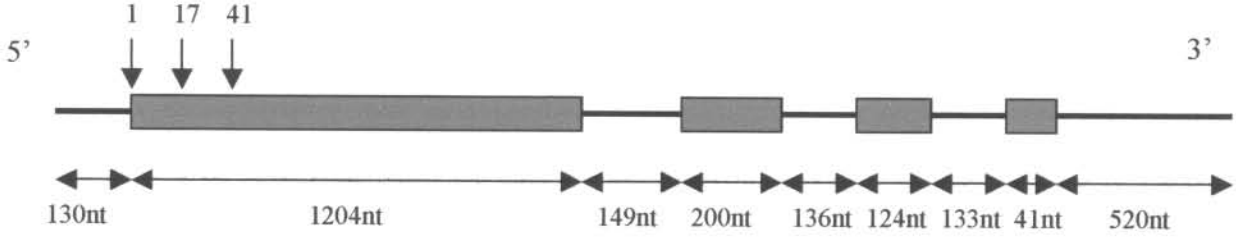

Figure 2.1: Schematic representation of the genetic organisation of *P. falciparum dhfs-fpgs*. Coloured boxes indicate the exons or coding regions. Black lines between the boxes indicate the intron positions and horizontal lines at the 5' and 3' ends indicate the 5' and 3' untranslated regions (UTR's), respectively. Position 1, 17 and 41 indicate ATG codons.

In most organisms DHFS and FPGS are natively expressed only at very low levels, making characterisation of the enzymes very difficult. Isolation of native *E. coli* DHFS-FPGS

encoded by the *folC* gene has such low yields that the gene had to be expressed in a high expression plasmid to obtain sufficient protein for catalytic studies (Bognar *et al.*, 1985). The native *P. falciparum* DHFS-FPGS is only expressed at very low levels in heterologous systems such as *E. coli* or yeast (personal communication, J. Hyde). Heterologous expression of malaria proteins is hampered by codon bias (lysine and arginine preferences) and the fact that the parasite genome is extraordinary A+T rich. The overall percentage of A+T content is 82% (including the plastid genome) and 76% for nuclear genes alone (Gardner *et al.*, 2002). The abundance of poly-A sites can furthermore act as termination signals during transcription, resulting in truncated mRNA and incomplete proteins after translation (Romanos *et al.*, 1991). In order to circumvent codon bias and high A + T content a separate plasmid encoding "rare" tRNAs for the amino acids arginine (AGA, AGG), isoleucine (AUA) and glutamine (GGA) can be introduced together with the plasmid containing the gene of interest (RIG plasmid). This plasmid enhances the expression ability of the host cells (Baca and Hol, 2000).

An alternative to the above method is the synthesis of a modified version of the gene from oligonucleotides, with a lower A+T content and preferred use of *E. coli* codons (Sugiyama *et al.*, 1996). The use of such a strategy has the further advantage that cassette mutagenesis of gene segments is made possible through the use of specific restriction enzyme sites in the gene. A chemically synthesised *P. falciparum* DHFR-TS gene was shown to express 10 times more efficiently from *E. coli* than the native DNA sequence (Prapunwattana *et al*, 1996). The gene was synthesised by cloning duplexes of complementary oligonucleotides ranging from 69 to 116 bp in length into 10 different vectors and restriction-ligation of the cloned fragments with the use of 31 different unique restriction enzyme sites. The oligonucleotides were designed to accommodate the differences in codon preferences in order to decrease the overall A + T content (Prapunwattana *et al.*, 1996).

Other methods describe the application of the polymerase chain reaction (PCR) for the synthesis of genes from overlapping oligonucleotides (Carpenter *et al.*, 1999). Completely overlapping oligonucleotides were assembled in a single PCR reaction, termed assembly PCR, through 'priming' of the complementary oligonucleotide overlaps (Figure 2.2). An additional PCR reaction with primers corresponding to the 5' and 3' outer ends was then used to obtain a sufficient amount of the full-length synthetic gene. This is, however, an

expensive method since the oligonucleotides overlap completely and requires a large amount of oligonucleotides (~125 pmoles of each) for template generation.

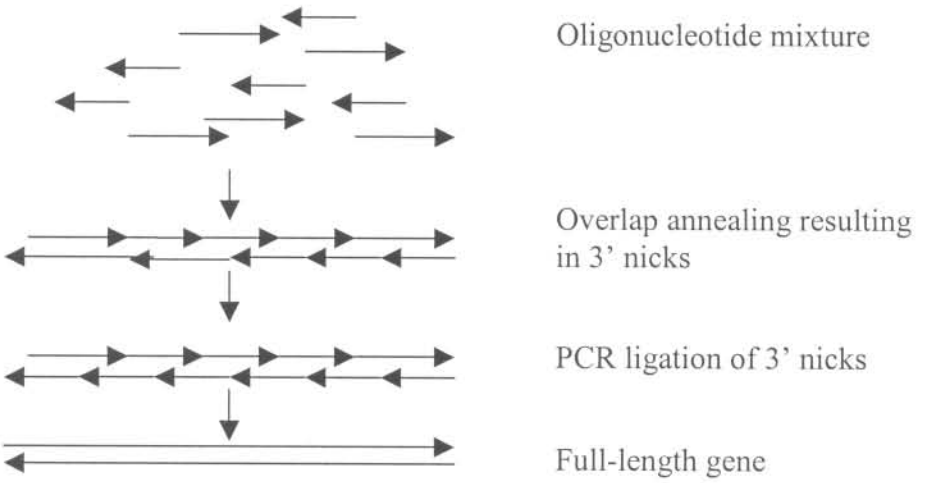

Figure 2.2: Overlap PCR used during gene synthesis (Carpenter *et al.*, 1999).

The strategy used here was to divide the full-length gene into halves, each constructed by means of PCR assembly from partially overlapping (about 20nt) oligonucleotides. The oligonucleotides prime initially only in the overlapping regions (Figure 2.4), leaving single stranded pieces of template which are extended during reiterative PCR cycles to generate the template (as will be described in the following sections). After a limited number of assembly steps and a 10 to 20-fold dilution of the mix, the PCR generated fragment is amplified with gene specific outer primers. Modifications introduced a 3' end reverse oligonucleotide for silencing of the stop codon to enable C-terminal His_6 expression from the pET22b vector. The full-length gene was constructed by overlap extension PCR of the halves and amplification with the designated outer primers.

2.2 Materials and methods

2.2.1 Oligonucleotides and primers

A total of 32 partially overlapping oligonucleotides, sixteen forward (f1-f16) and sixteen reverse (r1-r16) were designed for gene synthesis (Table 2.1). The first eight pairs of forward and reverse oligonucleotides were used for synthesis of the first half-gene and the last eight pairs for the second half-gene. Primers outF (5') and part1R (3') were used to amplify the first half fragment and primers part2F (5') and outR (3') were used to amplify the second half fragment. Oligonucleotide and primer sequences are given in appendix D.

Table 2.1: Forward (f) and reverse (r) oligonucleotides and forward (F) and reverse (R) primers used for the assembly of different gene fragments. The oligonucleotides involved in a specific overlap are indicated in brackets.

First half (nt 1-812)			Second half (nt 789-1586)		
Oligonucleotide name	**Length (nt)**	**Overlap T_m (°C) ***	**Oligonucleotide name**	**Length (nt)**	**Overlap T_m (°C)***
Forward 1 (f1)	72	53 (f1/r1)	Forward 9 (f9)	76	52 (f9/r9)
Reverse 1 (r1)	76	58 (r1/f2)	Reverse 9 (r9)	62	57 (r9/f10)
Forward 2 (f2)	76	53 (f2/r2)	Forward 10 (f10)	69	57 (f10/r10)
Reverse 2 (r2)	76	59 (r2/f3)	Reverse 10 (r10)	72	56 (r10/f11)
Forward 3 (f3)	71	60 (f3/r3)	Forward 11 (f11)	74	61 (f11/r11)
Reverse 3 (r3)	72	57 (r3/f4)	Reverse 11 (r11)	68	59 (r11/f12)
Forward 4 (f4)	69	56 (f4/r4)	Forward 12 (f12)	69	58 (f12/r12)
Reverse 4 (r4)	71	58 (r4/f5)	Reverse 12 (r12)	71	58 (r12/f13)
Forward 5 (f5)	73	58 (f5/r5)	Forward 13 (f13)	74	59 (f13/r13)
Reverse 5 (r5)	71	58 (r5/f6)	Reverse 13 (r13)	73	57 (r13/f14)
Forward 6 (f6)	76	58 (f6/r6)	Forward 14 (f14)	70	58 (f14/r14)
Reverse 6 (r6)	62	58 (r6/f7)	Reverse 14 (r14)	72	58 (r14/f15)
Forward 7 (f7)	76	49 (f7/r7)	Forward 15 (f15)	72	57 (f15/r15)
Reverse 7 (r7)	73	58 (r7/f8)	Reverse 15 (r15)	71	55 (r15/f16)
Forward 8 (f8)	68	57 (f8/r8)	Forward 16 (f16)	72	58 (f16/r16a)
Reverse 8 (r8)	75		Reverse 16a (r16a)	66	
			Reverse 16b (r16b)	69	
Primer name	**Sequence (5' to 3')**			**Length (nt)**	**T_m (°C) ***
outF	CGCGGACATATGGAAAAAAAC			21	56
part1R	CAGTTCAGTTCTTTCGCTTTATC			23	57
part2F	TGATAAAGCGAAAGAACTGA			20	51
outR	CGGATCC**TTA**CACCAGGCTC			20	61
alt4R	CTCGGATCCTG**TTT**CACCAGGCTCGGTTCGTTCAT			35	73

* T_ms were calculated with the formula: T_m=69.3 +0.41(%GC) - 650/ primer or overlap length in nt. Primer alt4R and oligonucleotide r16b were used to silence the stop codon (indicated in bold letters) at the 3' end of the gene.

2.2.1.1 Oligonucleotide design

Comparison of codon frequencies between *P. falciparum* and *E. coli* identified points of codon bias, which were eliminated in the design of the synthetic gene. Codon preference tables (www.kazusa.or.jp/codon), provided in Appendix A, were used to convert the native *P. falciparum* codons to the preferred *E. coli* codons (Figure 2.3). Factors taken into account in the design of the oligonucleotides were similar melting temperatures (<5°C deviation) of overlapping regions and primers to eliminate non-specific priming as well as minimal palindromes that could form secondary structures. The Oligo 4 Primer Analysis Software program (Molecular Biology Insights, MBI) was used to check internal stability and mispriming of the designed oligonucleotides and primers.

2.2.1.2 Oligonucleotide- and primer stock solutions and nucleic acid concentration determination

Oligonucleotides were obtained from MWG (UK) and were dissolved in 20% acetonitrile at 37°C at a stock solution concentration of 100 pmoles/µl. Oligonucleotide solutions were made up containing 1 pmole/µl of each of the oligonucleotides in TE buffer (0,1 mM Tris, 0,1 mM EDTA pH 7,6) and 20 µl aliquots were frozen away to minimise nucleotide damage through repeated freeze-thaw cycles. Primers were dissolved in 10 mM Tris (pH 8,6) at 37°C to give a final stock solution concentration of approximately 100 pmoles/µl each. Silanized tubes were used for primer and oligonucleotide storage at -20°C. The DNA concentration of each oligonucleotide and primer was determined spectrophotometrically at 260 nm. The manufacturer values for oligonucleotide mass per OD was used to determine the concentration according to the following formula:

Concentration = A_{260nm} × weight per OD (ng/µl) ×dilution factor / molecular mass (ng/nmole).

The DNA concentration of plasmids and gel-purified PCR products was determined either by ultraviolet absorbance (260nm) with 1 A_{260nm} unit equivalent to 50 ng/µl double stranded and 33 ng/µl single stranded DNA or by comparison with a range of calf thymus DNA standards spotted on glad wrap and stained with 10 µg/ml ethidium bromide (Sambrook *et al.*, 1989). Alternatively, concentrations of PCR products or isolated plasmids were also estimated by comparison of the fluorescence intensity of the sample

bands to markers of known concentrations (*EcoR*I and *Hind*III digested λ DNA, Promega, Wisconsin, USA). Sample band intensities were compared to marker band intensities of the same size to estimate the concentration.

2.2.2 PCR gene synthesis

The synthetic 1586 bp *dhfs-fpgs* gene was subdivided into ~750 bp halves. Each half fragment was constructed from 16 overlapping oligonucleotides and then amplified with the corresponding primers (Table 2.1). The outer primers were designed to incorporate an N-terminal *Nde*I site and a C-terminal *BamH*I site for the in-frame cloning of the gene into the pET15b and pET22b expression vectors (see Chapter 3). Two alternative C-terminal reverse primers were also designed: outR containing the *P. falciparum* stop codon for N-terminal His tagged expression from pET15b and alt4R for silencing of the stop codon to enable C-terminal His tagged expression from pET22b.

Gene synthesis consisted of a two-step procedure in which each fragment was assembled from 8 pairs of oligonucleotides by overlap extension cycles (refer to figure 2.4) followed by PCR amplification of a 50-fold diluted assembly mixture with the appropriate primers and *Pfu* DNA polymerase (Promega, Wisconsin, USA). The 3' to 5' proofreading ability of the *Pfu* DNA polymerase increased the fidelity of the PCR process.

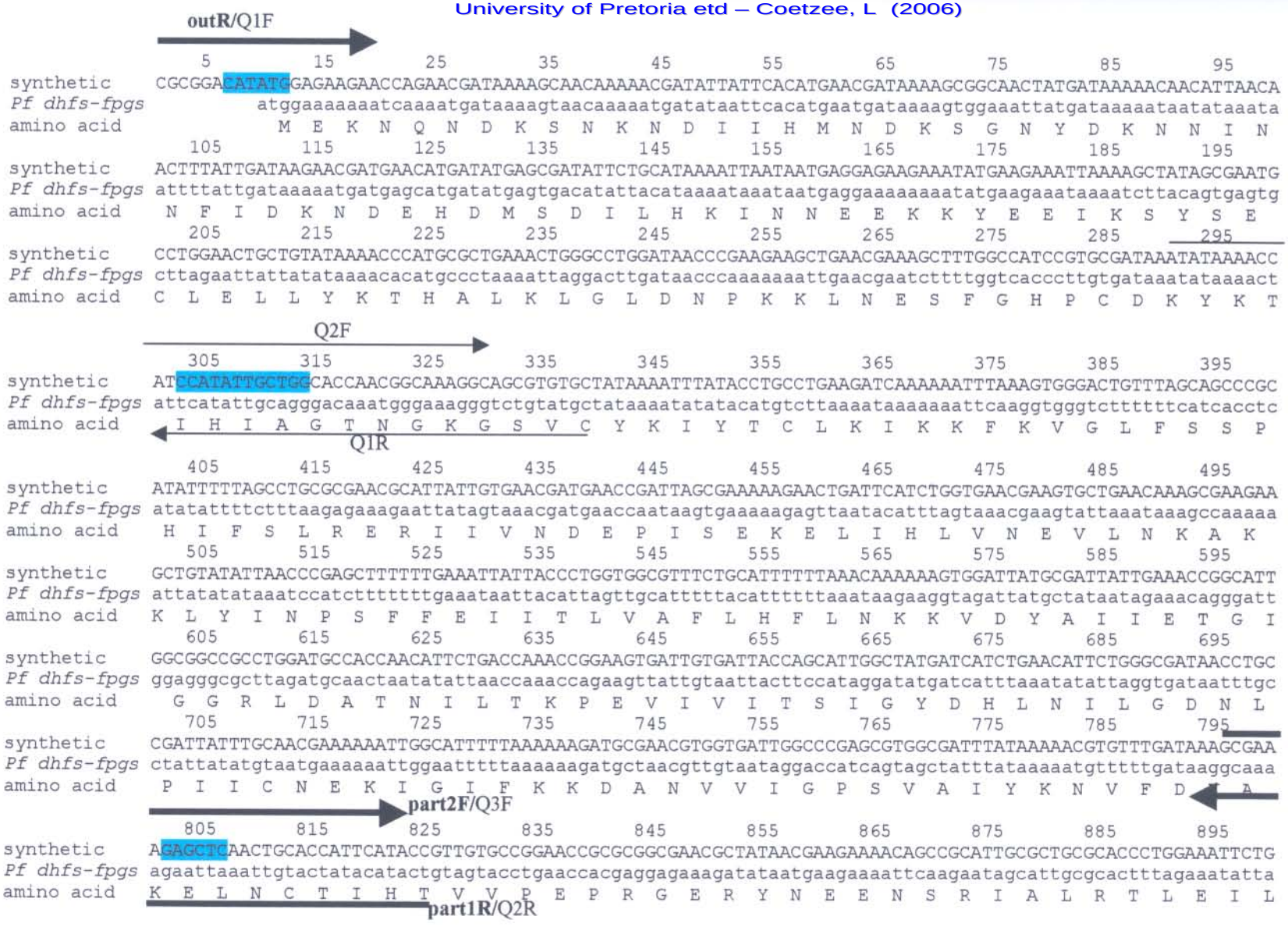
outR/Q1F
synthetic CGCGGACATATGGAGAAGAACCAGAACGATAAAAGCAACAAAAACGATATTATTCACATGAACGATAAAAGCGGCAACTATGATAAAAACAACATTAACA
Pf dhfs-fpgs atggaaaaaaatcaaaatgataaaagtaacaaaaatgatataattcacatgaatgataaaagtggaaattatgataaaaataatataaata
amino acid M E K N Q N D K S N K N D I I H M N D K S G N Y D K N N I N
synthetic ACTTTATTGATAAGAACGATGAACATGATATGAGCGATATTCTGCATAAAATTAATAATGAGGAGAAGAAATATGAAGAAATTAAAAGCTATAGCGAATG
Pf dhfs-fpgs attttattgataaaaatgatgagcatgatatgagtgacatattacataaaataaataatgaggaaaaaaaaatatgaagaaataaaatcttacagtgagtg
amino acid N F I D K N D E H D M S D I L H K I N N E E K K Y E E I K S Y S E
synthetic CCTGGAACTGCTGTATAAAACCCATGCGCTGAAACTGGGCCTGGATAACCCGAAGAAGCTGAACGAAAGCTTTGGCCATCCGTGCGATAAATATAAAACC
Pf dhfs-fpgs cttagaattattatataaaacacatgccctaaaattaggacttgataacccaaaaaaattgaacgaatcttttggtcacccttgtgataaatataaaact
amino acid C L E L L Y K T H A L K L G L D N P K K L N E S F G H P C D K Y K T
Q2F
synthetic ATCCATATTGCTGGCACCAACGGCAAAGGCAGCGTGTGCTATAAAATTTATACCTGCCTGAAGATCAAAAAATTTAAAGTGGGACTGTTTAGCAGCCCGC
Pf dhfs-fpgs attcatattgcagggacaaatgggaaagggtctgtatgctataaaatatatacatgtcttaaaataaaaaaattcaaggtgggtctttttttcatcacctc
amino acid I H I A G T N G K G S V C Y K I Y T C L K I K K F K V G L F S S P
Q1R
synthetic ATATTTTTAGCCTGCGCGAACGCATTATTGTGAACGATGAACCGATTAGCGAAAAAGAACTGATTCATCTGGTGAACGAAGTGCTGAACAAAGCGAAGAA
Pf dhfs-fpgs atatattttctttaagagaaagaattatagtaaacgatgaaccaataagtgaaaaagagttaatacatttagtaaacgaagtattaaataaagccaaaaa
amino acid H I F S L R E R I I V N D E P I S E K E L I H L V N E V L N K A K
synthetic GCTGTATATTAACCCGAGCTTTTTTGAAATTATTACCCTGGTGGCGTTTCTGCATTTTTTAAACAAAAAAGTGGATTATGCGATTATTGAAACCGGCATT
Pf dhfs-fpgs attatatataaatccatctttttttgaaataattacattagttgcatttttacatttttttaaataagaaggtagattatgctataatagaaacagggatt
amino acid K L Y I N P S F F E I I T L V A F L H F L N K K V D Y A I I E T G I
synthetic GGCGGCCGCCTGGATGCCACCAACATTCTGACCAAACCGGAAGTGATTGTGATTACCAGCATTGGCTATGATCATCTGAACATTCTGGGCGATAACCTGC
Pf dhfs-fpgs ggagggcgcttagatgcaactaatatattaaccaaaccagaagttattgtaattacttccataggatatgatcatttaaatatattaggtgataatttgc
amino acid G G R L D A T N I L T K P E V I V I T S I G Y D H L N I L G D N L
synthetic CGATTATTTGCAACGAAAAAATTGGCATTTTTAAAAAAGATGCGAACGTGGTGATTGGCCCGAGCGTGGCGATTTATAAAAACGTGTTTGATAAAGCGAA
Pf dhfs-fpgs ctattatatgtaatgaaaaaattggaatttttaaaaaaagatgctaacgttgtaataggaccatcagtagctatttataaaaatgtttttgataaggcaaa
amino acid P I I C N E K I G I F K K D A N V V I G P S V A I Y K N V F D A
part2F/Q3F
synthetic AGAGCTCAACTGCACCATTCATACCGTTGTGCCGGAACCGCGCGGCGAACGCTATAACGAAGAAAACAGCCGCATTGCGCTGCGCACCCTGGAAATTCTG
Pf dhfs-fpgs agaattaaattgtactatacatactgtagtacctgaaccacgaggagaaagatataatgaagaaaattcaagaatagcattgcgcactttagaaatatta
amino acid K E L N C T I H T V V P E P R G E R Y N E E N S R I A L R T L E I L
part1R/Q2R

Figure 2.3: Alignment of the synthetic and native (*Pf dhfs-fpgs*) *dhfs-fpgs* genes. The amino acid sequence is indicated below these sequences. Bold arrows and bold primer labels indicate the primers used for the generation of the half fragments (Figure 2.4). Thin arrows and primer labels indicate primers used for the generation of the four-quarter segments (Figure 2.5). Restriction enzyme sites incorporated by the primers are highlighted in blue. The stop codon (TAA) is underlined. A portion of the alternative reverse sequence to remove the stop codon is indicated by the sequence annotation '-stop'. Here AAA replaces the stop codon TAA, for readthrough and expression of a C-terminal tag.

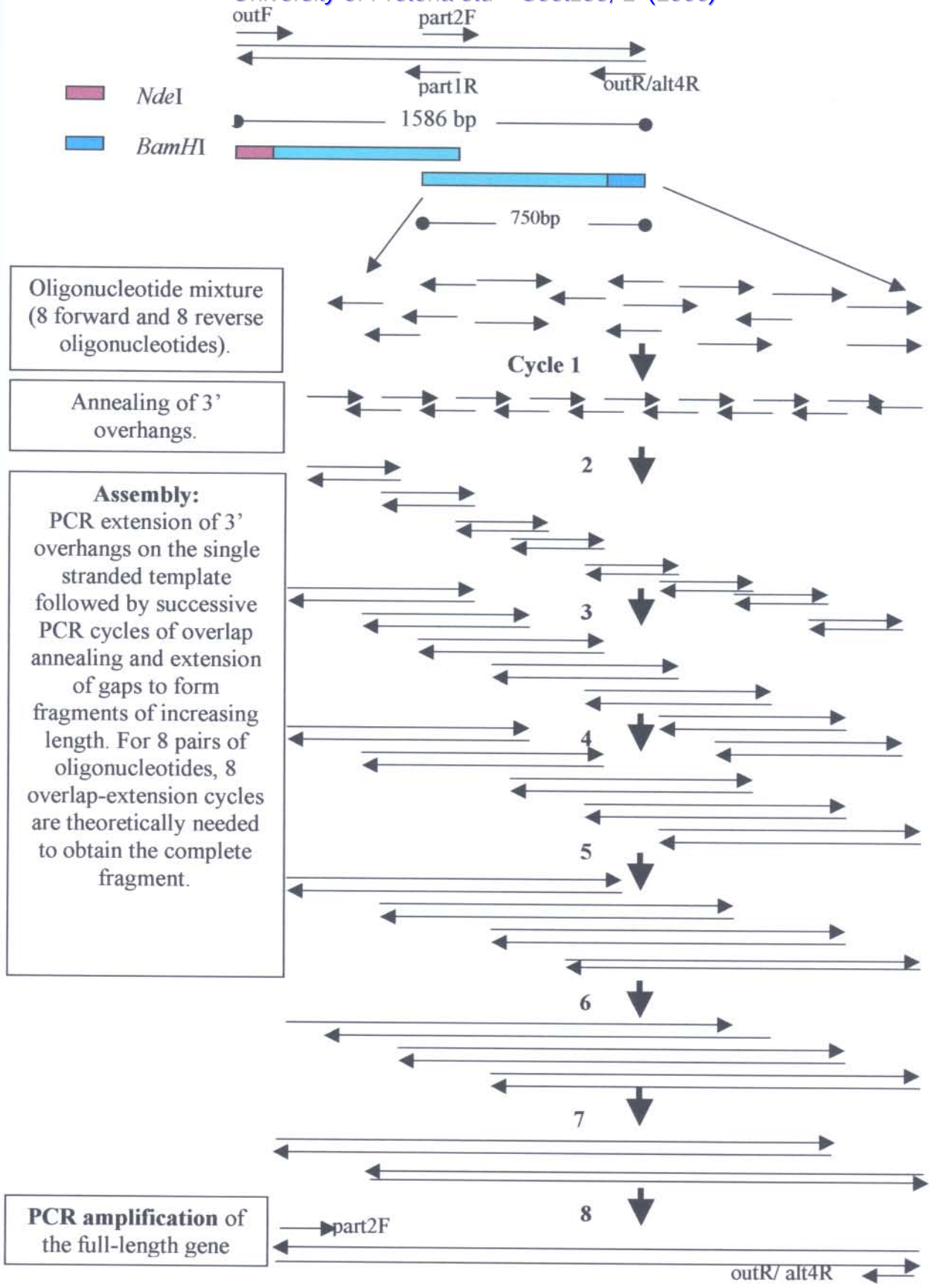

Figure 2.4: Overlap-extension PCR. The gene is divided into halves, each made of 8 forward and 8 reverse partially overlapping oligonucleotides. PCR assembly is used for the full-length construction by extension of single stranded gaps between overlaps, followed by amplification with the designated 5' and 3' end primers (part2F and outR).

In later studies the gene was divided into quarter segments by amplification of half gene segments with internal primers containing mutations to create unique restriction sites at these locations and to generate quarter fragments (Figure 2.5).

The PCR gene synthesis was optimised as follows: The Taguchi method was used to determine broad parameters for the Mg^{2+}, dNTP and *Pfu* DNA polymerase concentrations in the assembly step (Cobb and Clarkson, 1994). These parameters were then further refined with additional optimisation experiments. The final assembly PCR reactions consisted of 1pmole of each oligonucleotide, 4 mM $MgSO_4$ (Promega, Wisconsin, USA), 250 µM of each dNTP (Promega, Wisconsin, USA), 5 µl of an enzyme master mix and 1x *Pfu* DNA Polymerase buffer (Promega, Wisconsin, USA) in a final volume of 50 µl. The final amplification reactions consisted of 10 pmoles of each primer, a 50x dilution of the assembly PCR reaction, 2 mM $MgSO_4$, 250 µM of each dNTP, 1x *Pfu* DNA Polymerase buffer (Promega, Wisconsin, USA) and 5 µl of an enzyme master mix (2.5 U *Pfu* DNA polymerase for assembly or 1.25 U for amplification and 1x *Pfu* DNA Polymerase buffer) in a final volume of 50 µl. Master mixes consisting of all the reaction components except the enzyme were used throughout and added to each reaction in a hot start protocol. Assembly cycles were varied according to the number of oligonucleotides (see results section 2.3.3) and the number of amplification cycles did not exceed 30. The optimised cycling profile was as follows: initial denaturation at 94 °C for 30 sec, hot start addition of *Pfu* DNA polymerase at 80 °C for 20 sec, cycling at 94 °C for 30 sec, lowest overlap T_m (refer to table 2.1) for 30 sec and 72 °C for 2 min and a final extension step at 72 °C for 5 min. PCR products were stored at 4°C. All PCR's were performed in a Gene-Amp 9700 thermocycler (Perkin Elmer, Wellesley, USA) in 0.2 ml thin walled PCR tubes.

2.2.3 Agarose gel electrophoresis and purification

Digested plasmids and PCR products were separated on 1,5% (w/v) or 2% (w/v) agarose (Promega, Wisconsin, USA) gels, respectively. TAE buffer (0.04 M Tris-acetate, 1 mM EDTA) was used for electrophoresis. A 10 µg/ml ethidium bromide (EtBr) stock solution was included in the gel to a final concentration of 1.25 ng/ml. Samples were electrophoresed at 8 V/cm in a Minnie Submarine HE33 Agarose Gel unit (Hoeffer Scientific systems, San Francisco, USA). The bands were visualized on a Spectroline TC-312A UV transilluminator (Spectronics Corporation, Connecticut, USA) at 312 nm. Visualisation of PCR products used for cloning was done with crystal violet to prevent thymine dimers in the A+T rich areas of

the gene by ultraviolet illumination (Clark and Webb, 1955). A final concentration of 10 µg/ml crystal violet was used in TAE for the gels and running buffer. Bands were excised from the agarose gel and purified with the High Pure PCR product purification kit (Roche, Basel, Switzerland) according to the manufacturer's protocol. Gel slices were dissolved at 65°C in binding buffer (4 M guanidinium-HCl, 0.5 M potassium acetate, pH 4.2) and loaded onto a High Pure filter tube. DNA binds to the filter in the presence of chaotropic salts. The filter was washed two times with 500 µl and 200 µl wash buffer II (20 mM NaCl, 2 mM Tris HCl, pH 7.5), respectively, by means of centrifugation 15 000xg for 1 min. Each time the filtrate was discarded. The filter tube was transferred to a clean 1,5 ml tube and the plasmid DNA eluted in 50 µl elution buffer (1 mM Tris-HCl, pH 8.5) and collected by centrifugation at 16000xg for 1min at 22 °C. All buffers mentioned here, were supplied in the kit.

2.2.4. Cloning

2.2.4.1 The pGEM T Easy vector system

An A-tailing reaction was performed to add A overhangs to the blunt-ended PCR products generated by *Pfu* DNA polymerase to enable A/T cloning into the pGEM-T Easy vector according to the manufacturers protocol (Promega, Wisconsin, USA). For the vector map refer to Appendix B. Approximately 190 ng purified DNA was incubated for 30 minutes at 70°C in the presence of 0.2 mM dATPs and 5 U rTaq DNA polymerase (Takara, Shuzo, Japan) with 1x Taq buffer and 2 mM $MgCl_2$. The A-tailed product and pGEM T Easy vector were added together in a 3:1 molar ratio. 3 Weiss units T4 DNA ligase and 1x T4 DNA ligase buffer (provided in the pGEM T Easy vector kit) were added to the vector (50 ng) and the A-tailed product to a final volume of 10 µl. This reaction was incubated on ice overnight. The ligation mixture was cloned into electrocompetent or calcium chloride competent DH5α [genotype F^-Φ80 d*lacz* ΔM15 Δ(*lac*ZYA –argF) U169 *deo*R *rec*A1 *end*A1 *hsd*R17 (r_K^+, m_K^+) *pho*A *sup*E44 λ *thi*-1 *gyr*A96 *rel*A1] *E. coli* cells (see below).

Electrocompetent cells were prepared by the following method (Sambrook *et al.*, 1989): A single DH5α *E. coli* colony was inoculated into 15ml Luria-Bertani liquid medium (1% tryptone, 0,5% yeast extract, 1% NaCl, pH 7) and grown overnight with shaking at 300 rpm at 37°C. The overnight culture was diluted 1/100 into 2 flasks with 500ml LB liquid medium and grown at 30°C with shaking (300rpm) to an OD_{600nm} of 0.5. The cultures were poured into 2 pre-chilled 250 ml centrifuge bottles and incubated on ice for 20 min. All the subsequent steps

were done at 4°C. Cells were pelleted at 5000xg in a Beckman Avanti J-25 centrifuge with a fixed angle rotor at 4 °C for 10min. After removal of the supernatant, the cells were dissolved in 10 ml ice-cold water and washed with 250 ml ice-cold water. The suspension was centrifuged at 5000xg for 10 min at 4°C. This washing step was repeated twice. After the final centrifugation step the supernatant was immediately removed from the loose pellets. The pellets were resuspended in 10 ml ice-cold 10% glycerol each and incubated on ice for 30 minutes. Cells were subsequently pelleted (5000xg for 10 min at 4°C), the supernatant removed with vacuum suction and the pellets resuspended in 800 µl 10% ice cold glycerol. This was divided in 90 µl aliquots and frozen at -70°C.

Plasmids were co-precipitated with 10% tRNA (10mg/ml) for electroporation by the addition of 10% (v/v) sodium acetate of a 3M stock solution, pH 5 and 3 times the sample volume of absolute ethanol. This was incubated at -70°C for 1 hour and then centrifuged at 16 000xg for 20 min at 22°C. The plasmid pellet was washed with 3 volumes 70% ethanol, dried *in vacuo* and dissolved in 10 µl water, to which electrocompetent cells (90 µl) were added. This was then transferred to pre-chilled electroporation cuvettes and a pulse of 2000 V applied for 5 ms in the Multiporator system (Eppendorf, Germany). LB liquid medium (1 ml) was added directly after electroporation and the cells were grown for 1 hour at 30°C with shaking and plated out on LB solid media (1% noble agar, 1% tryptone, 0,5% yeast extract, 1% NaCl, pH 7) containing 0.2 mg/ml ampicillin.

Calcium chloride competent cells (J. Hyde, UMIST personal communication) were prepared as follows: 5ml of an overnight DH5α cell culture was diluted 10 times and grown with shaking to an OD_{600nm} of 0.4 at 30°C. The cells were pelleted by centrifugation for 5 min at 5000xg at 4°C and pellets were resuspended in 30 ml ice cold 0.1 M $CaCl_2$ at 4°C. This was then centrifuged at 5000xg for 5 min at 4°C. The supernatant was removed and the pellet resuspended in 25 ml ice cold 0.1 M $CaCl_2$. The centrifugation was repeated as before and the pellet then resuspended in 2.5 ml ice cold 0.1 M $CaCl_2$ .The competent cells were then incubated on ice for an hour followed by the addition of 15% ice cold glycerol and snap-freezing of 90 µl aliquots in liquid nitrogen.

$CaCl_2$ competent cells were thawed on ice, added to 50 ng of plasmid and incubated on ice for 30 min. A 42°C heat shock was applied for 1 min in a water bath, followed by incubation on

ice for 2 min. 500 µl of LB broth was added to the cells. The cells were then grown with shaking for 1 hour at 30°C and plated out on LB agar plates containing 0.2 mg/ml ampicillin.

2.2.4.2 The pMOS*Blue* vector system

Blunt ended cloning was also conducted with the pMOS*Blue* vector from Amersham Biosciences, Buckinghamshire, UK (refer to Appendix C). An optimal molar vector: insert ratio of 1:2.5 was used as recommended by the manufacturer. Before ligation, a phosphorylation reaction was performed on the purified inserts. The specified amount of insert was added to a reaction tube containing 5 mM dithiotreitol (DTT), a 1x phosphorylase kinase (pk) buffer and 1 µl phosphorylase kinase enzyme mix (provided in the pMOS*Blue* kit) in a final volume of 10µl. The reaction was incubated at 22°C for 5 min, followed by heat inactivation of the phosphorylase at 75°C for 10 min. The reaction was cooled on ice to prepare the mixture for ligation. Ligation mixtures consisted of 10 µl of the phosphorylation reaction, 50 ng of pMOS*Blue* vector and 4 Weiss units of T4 DNA ligase in a final volume of 12µl (one Weiss unit of ligase is equivalent to 200 cohesive end ligation units; Sambrook *et al.* 1989). The reaction was incubated at 22°C for 5 hours.

Competent MOS*Blue* cells (genotype: *end*A1 *hsd*R17 (r_{k12}-m_{k12}^{+}) *sup*E44 thi-1 *rec*A1 *gyr*A96 *rel*A1 lac [F' *pro*$A^{+}B^{+}lac^{q}Z$ ΔM15:Tn10(T_c^{R})]) supplied by the pMOS*Blue* blunt ended cloning kit were used for transformation of inserts ligated into the pMOS*Blue* vector (Amersham Biosciences, Buckinghamshire, UK). The transformation of MOS*Blue* competent cells was done by the addition of 5 ng of ligated plasmid (~2µl of the ligation reaction) to 20 µl competent cells. This was followed by incubation on ice for 30 min, heat shock at 42°C for 40 sec and an ice incubation for 2 min. 80 µl room temperature SOC (2% tryptone, 0.5% yeast extract, 10m M NaCl, 2.5 mM KCl, 1% glucose, pH 7) was added to this and the cells grown with shaking at 37°C for an hour, after which the cells were plated out on LB solid media containing a final concentration of 0.2 mg/ml ampicillin (Roche, Basel, Switzerland), 12.5 µg/ml tetracycline (Roche, Basel, Switzerland), 40 µg/ml 5-Bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) (Promega, Wisconsin, USA) and 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Ampicillin and tetracycline was used for primary selection and X-gal and IPTG for blue-white screening to identify the recombinant clones.

2.2.5 Plasmid isolation and identification of recombinant clones

2.2.5.1 STET–prep plasmid isolation

An adapted STET (Sucrose, TritonX-100, EDTA, Tris)-prep protocol was used as a quick and convenient method for plasmid isolation and screening of inserts (Quigley and Holmes, 1981). Clones were picked from solid media and grown overnight in liquid culture with shaking at 30°C. 1.5 ml of the liquid cell culture was centrifuged for 10 min, 3500xg at 22°C in an Eppendorf 5415R centrifuge with a 45° fixed angle rotor. The supernatant was partially removed and the pellet was resuspended in the remainder of supernatant. To this 300 µl of STET buffer (8% (w/v) sucrose, 5% (v/v) Triton X100, 50 mM EDTA, 50 mM Tris, pH 8) was added. Lysozyme, 25µl of a freshly prepared 4 mg/ml stock solution, (Roche, Basel, Switzerland) was added to digest the cell walls. The samples were placed in boiling water for 1 min and then centrifuged for 15 min at 16000xg (in an Eppendorf 5415R centrifuge) at 22 °C to pellet cell walls and debris. The genomic DNA is denatured by the high pH of the buffer, while the plasmid DNA reanneals fully. Pellets were discarded and 300 µl isopropanol added to the supernatant for precipitation of the DNA at 22°C. The tubes were centrifuged as above and the supernatant removed. The pellets containing plasmid DNA were air-dried and resuspended in 20 µl TE buffer (10 mM Tris, 1 mM EDTA, pH 8) containing a final concentration of 0.5 mg/ml RNAse (Roche, Basel, Switzerland).

Screening for recombinant clones was done by means of restriction enzyme digestion of STET-prep isolated plasmids. The *EcoR*I restriction enzyme site in the multiple cloning site of the pGEM-T Easy vector was used to screen for recombinant colonies, since it digests the plasmid at both sides of the insert (Appendix B). Double digestion was performed on recombinant MOS*Blue* cells, with *EcoR*I and *Hind*III on opposite sides of the multiple cloning site. Digestion reactions were set up each in a total volume of 20µl, consisting of >50 ng plasmid DNA, 10 U of the specific restriction enzyme and 1x of the appropriate restriction enzyme buffer (Promega, Wisconsin, USA). The tubes were incubated at 37°C for a minimum of 3 hours and the reaction stopped by loading the sample on an agarose gel.

PCR screening was used as an alternative method to identify positive clones (Amersham pMOS*Blue* instruction leaflet). Of the overnight DH5α or MOS*Blue* liquid cultures 100µl cells were centrifuged for 1 min at 16000xg at 4 °C and the pellet resuspended in 100 µl water. This was boiled for 5 min and centrifuged again (1 min at 16000xg 4 °C). The PCR reaction consisted of 10 µl of the supernatant as template, 5 pmoles of insert-specific primers (Tables 2.1 and 2.2), 1.25 U rTaq DNA polymerase (Promega, Wisconsin, USA), 1x Taq

PCR buffer, 2 mM $MgCl_2$ and 0,2 mM of each dNTP in a final volume of 20µl. The PCR profile consisted of 30 cycles at 94°C for 30 sec, 50°C for 30 sec and 72°C for 5 min. PCR products were analysed with agarose gel electrophoresis as described in section 2.2.3.

2.2.5.2. High Pure Plasmid kit plasmid isolation

The High Pure Plasmid isolation kit (Roche, Basel, Switzerland) was used to obtain pure plasmid samples. For plasmid isolation, 7 ml liquid culture was centrifuged for 30 sec, 5000xg at 22°C to pellet the cells. Cell pellets were resuspended in 250 µl suspension buffer (50 mM Tris-HCl, 10 mM EDTA, 10% w/v RNAse, pH 8) and then 250 µl lysis buffer (0.2 mM NaOH, 1% SDS) was added followed by a 5 min incubation at 25°C. After lysis, 350 µl chilled binding buffer (4 M guanidinium-HCl, 0.5M potassium acetate, pH 4.2) was added. The sample was mixed by inversion and incubated on ice for 5 min. Cell walls and large contaminants were pelleted by centrifugation, 16000xg for 10 min at 22°C. The supernatant containing plasmid DNA was passed through a High Pure Filter tube for binding of the plasmid DNA to the filter and purified further as described in section 2.2.3.

2.2.6 Sequencing

Sequencing PCR reactions were done on recombinant plasmids isolated with the High Pure Plasmid kit with T7 promoter and U19 primers (Promega, Wisconsin, USA) specific to the multiple cloning site of the pMOS*Blue* vector or T7 promoter and SP6 primers (Promega, Wisconsin, USA) complementary to the T7 and SP6 promoters flanking the multiple cloning site of the pGEM-T Easy vector (refer to appendix D). Gene-specific primers were also used to obtain internal sequences. Single-primer reactions were set up containing: 200-500 ng pure plasmid DNA as template, 3µl 5x sequencing buffer, 2µl Terminator Big-Dye Ready reaction mix version 3 (Perkin Elmer, Wellesley, USA) and double distilled deionised H_2O to a final volume of 20µl. The PCR was performed in a GeneAmp 9700 thermocycler (Perkin Elmer, Wellesley, USA) and the profile consisted of 25 cycles of: denaturation at 96 °C for 10 sec, annealing at 50 °C for 5 sec and extension at 60 °C for 4 min.

The DNA generated by PCR was isolated from all contaminants such as free dNTPs, protein and truncated DNA fragments by size exclusion filtration of the PCR mixture. Gel filtration media was prepared by addition of 600 µl G50 Superfine Sephadex (66 mg/ml) hydrated overnight in sterile water to the filter tube, followed by centrifugation at 750xg for 2 min. The PCR mixture was loaded onto the column, centrifuged again, collected in a clean tube, dried *in vacuo* and stored at -20 °C in the dark to protect the light sensitive dyes. Sequencing was performed on an ABI PRISM 377 Automatic Sequencer (Perkin Elmer, Wellesley, USA) or the ABI PRISM 3100 capillary sequencer. Another method of PCR product purification was the precipitation of DNA by the addition 3 volumes of ice-cold absolute ethanol followed by centrifugation at 16 000xg for 25-30 min at 4°C. After removal of the supernatant, the pellet was washed with 5 volumes of 70% ethanol followed by centrifugation at 16 000xg for 10 min at 4°C. The wash step was repeated with 4 volumes of 70% ethanol. After centrifugation at 16000xg for 10 min at 4°C, the DNA pellet was dried in the dark *in vacuo*.

The sequences were obtained as ABI files, which were either imported into the BioEdit version 5.0.6 software program (Hall 2001, North Carolina State University, Department of Microbiology, USA) for analysis and alignment with the required sequence, or into the Staden Package (Staden *et al.*, 2000; Medical Research Council Laboratory of Molecular Biology, Cambridge, UK) which was used for alignment of the different clone sequences with the required sequence. The electropherograms were compared with the sequence designated by the software.

2.2.7 Construction of the full-length gene

The full-length gene was assembled from the purified half-gene fragments at different concentrations (1 pmole, 0.1 pmole and 0.01 pmole). A single overlap-extension PCR procedure was used for assembly and amplification. PCR mixtures consisted of 250 mM of each dNTP, 1x of *Pfu* DNA polymerase buffer containing 2 mM $MgSO_4$ and 1.25 U *Pfu* DNA polymerase (Promega, Wisconsin, USA) in a final volume of 50µl. This PCR mixture was assembled for 5 cycles consisting of 94°C for 1 min, 50°C for 30 sec, and 68°C for 3 min. After these cycles, the temperature was lowered to 70°C for the addition of 5 pmoles of each outer primer, followed by 20 amplification cycles consisting of 94°C for 30 sec, 50°C for 30 sec, and 68°C for 2 min. A final extension step of 72°C for 5 min was used.

2.2.8 Gene repair strategies

2.2.8.1 Quarter gene segments

For ease of cloning and sequencing of the PCR generated fragments, the gene was subdivided in later experiments into two 300bp and two 500bp quarters, constructed from six and ten oligonucleotides, respectively. New primers were designed for the generation of the quarter segments (Table 2.2). These primers also incorporated unique restriction enzyme sites for restriction-ligation of the quarter fragments to assemble the full-length gene (Figure 2.5).

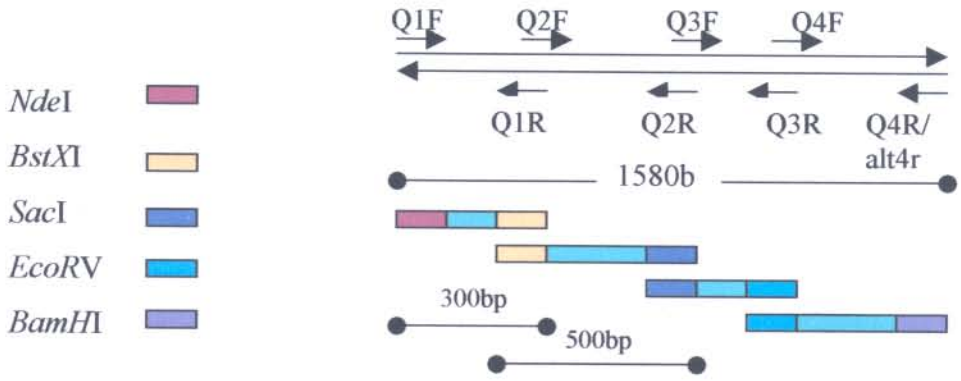

Figure 2.5: Subdivision of *dhfs-fpgs* into quarter segments with unique restriction enzyme sites in the overlaps. Primer positions are indicated by the arrows and primer names are directly above or below the arrows. Specific restriction enzyme sites used for digestion and ligation of the segments are indicated by the corresponding blocks.

Table 2.2: Forward (F) and reverse (R) primers used for the generation of the quarter fragments. T_ms were calculated with the formula: T_m=69.3 +0.41(%GC)-650/(primer length in nt). Restriction enzyme sites are indicated above the underlined sequences and cut sites are indicated with arrows.

Quarter fragment	Primer sequence (5' to 3')	length (nt)	T_m (°C)
Quarter 1 (300bp)	*NdeI* Q1F: CGCGGACATATGGAGAAGAACCAGA Q1R: *BstXI* GTTGGTGCCAGCAATATGGATGGTTTTATATTTATC	25 36	65 65
Quarter 2 (500bp)	*BstXI* Q2F: AAACCATCCATATTGCTGGCACCAAC *SacI* Q2R: CAGTTGAGCTCTTTCGCTTTATCAAACACG	26 30	62 61
Quarter 3 (300bp)	*SacI* Q3F: TGATAAAGCGAAAGAGCTCAACTGCA *EcoRV* Q3R: ATAGTTGATATCGGTGCACAGGCGATCAATC	26 31	62 62
Quarter 4 (500bp)	*EcoRV* Q4F:TGATCGCCTGTGCACCGATATCAACTATTTTCATA *BamHI* Q4R: CCGGATCCTTACACCAGGCTCGGTTCGT	35 28	67 71

2.2.8.2 Site directed mutagenesis of point mutated PCR products

Minor errors (single or triple nucleotide errors) in quarter 4 were corrected by means of site directed mutagenesis (protocol adapted from the Stratagene Quick change mutagenesis protocol). For each mutation a pair of oligonucleotides (>35bases) or megaprimers were designed to incorporate the correct sequence on both DNA strands (Table 2.3). The mutation was designed to be in the middle of each oligonucleotide (with at least 10 bases on each side). Less than 50 ng of plasmid was used as template, while the primer concentration was kept above 125 ng. The rest of the PCR mixture contained 2.5 U *Pfu* DNA polymerase (Promega, Wisconsin, USA), added in a hot start protocol, a 1x *Pfu* DNA polymerase buffer, 2 mM $MgSO_4$ and 250 M dNTPs (Promega, Wisconsin, USA) in a final volume of 50 µl. The PCR profile consisted of an initial denaturation of 95°C for 30 sec, followed by 12 to 18 cycles of 95°C for 30 sec, the primer T_m for 30 sec and 68°C for 6 min (2 minutes of extension time for each kilobase of plasmid). The number of cycles was determined by the size of the mutation, i.e. 12 cycles for a single base change and 18 for a triple base change. After PCR, the tubes were cooled on ice for 2 min and then 10 U *Dpn*I restriction enzyme was added to the PCR mixture to digest the wild type plasmid for 3 hours at 37°C. *Dpn*I only recognizes its site after bacterial methylation, leaving the mutated PCR product intact. The digested mixture was purified with the High Pure PCR purification kit (Roche, Basel, Switzerland) and the PCR product cloned into the pMOS*Blue* vector as described in sections 2.2.3 and 2.2.4.2 respectively. Colonies were screened from agar plates and sequenced to determine whether mutagenesis occurred as described in sections 2.2.6.

Table 2.3: Site-directed mutagenesis primers designed for error corrections in quarter 4.

Primer name	T_m (°C)	Sequence (Positions where errors occur are highlighted)
errors SDM4.1F SDM4.1R	 68 68	nt1279-1308:CTGAAC**C**AACGAAGAA~~~ATTAAAAA... 5′ GTGGAAATGCTGAAC~AACGAAGAA**GAA**ATTAAAAACGAAATTAAAG3′ 3′ CACCTTTACGACTTG~TTGCTTCTT**CTT**TAATTTTTGCTTTAATTTC5′
error SDM 4.2F SDM 4.2R	 65 65	nt1386-1410:GAAGAAGATGCGCTG**G**AAACTGTATA 5′ CGAAGAAGATGCGCTG~AAACTGTATAAACG 3′ 3′ GCTTCTTCTACGCGAC~TTTGACATATTTGC 5′

2.2.8.3 Restriction-ligation of correct fragments

The restriction enzyme *Hha*I (Promega, Wisconsin, USA) was used for the ligation of two correct fragments from different quarter 1 clones as part of the gene repair strategy. The fragments were cloned into the *EcoR*V site of the pMOS*Blue* vector. The restriction enzyme *Hha*I was chosen since it was the only unique restriction site between the correct parts of the two quarter 1 clones. The correct fragments were isolated from the pMOS*Blue* vector by means of *EcoR*I/*Hha*I double digestion for the 5' end correct fragment and *Sal*I/*Hha*I double digestion for the 3' end correct fragment (refer to appendix C). The vectors were first cut with *Sal*I or *EcoR*I restriction enzymes (Promega, Wisconsin, USA), respectively, and the pMOS*Blue* vector was also double digested with *Sal*I and *EcoR*I in a single step. Digestion mixtures contained ~500 ng plasmid, 3 U of each of the restriction enzymes and 1x buffer B (Promega, Wisconsin, USA) in a total volume of 20 µl. The digested bands were purified from agarose gels with the High Pure PCR product purification kit (Roche, Basel, Switzerland) and were then ligated in a 1:1:1 molar ratio as described in section 2.2.4.1.

2.2.8.4. Cassette mutagenesis PCR

Stretches of internal sequence errors were also corrected by means of cassette mutagenesis with short overlapping resynthesised gene fragments (~150 bp). The resynthesised oligonucleotides were also designed to enable future cassette mutagenesis through the use of unique restriction enzyme sites on opposite sides of important features of the DHFS-FPGS protein, such as the ATP-binding P-loop (GTNGKGS; nt 313-333) and an interdomain-stabilising Ω loop (VGLFSSPHIFSLRERI; nt 379-426). Clones containing the correct 5' and 3' areas flanking the errors were identified and areas containing the errors were resynthesised from shorter oligonucleotides (~50 bases each) using the optimised gene assembly method as described in section 2.2.2. Primers were designed for the isolation of the 5' and 3' flanking sequences of each quarter as well as the amplification of the internal assembled parts (Table 2.4 and Figure 2.6). The internal sequences were assembled from resynthesised oligonucleotides using the optimised PCR parameters (Figure 2.6). All the subsequent PCR products were purified and added in equimolar amounts to assemble the correct full-length quarter in a PCR reaction consisting of 250 µM of each dNTP, 4 mM $MgSO_4$, 1x *Pfu* DNA polymerase buffer and 2.5 U *Pfu* DNA polymerase in a final volume of 50µl. PCR cycling consisted of 5 cycles of 94°C for 30 sec, 65°C for 30 sec, 68°C for 2 min, addition of 5

pmoles of each primer at 70°C and 20 amplification cycles of 94°C for 30 sec, 65°C for 30 sec, 68°C for 2 min (Figure 2.6).

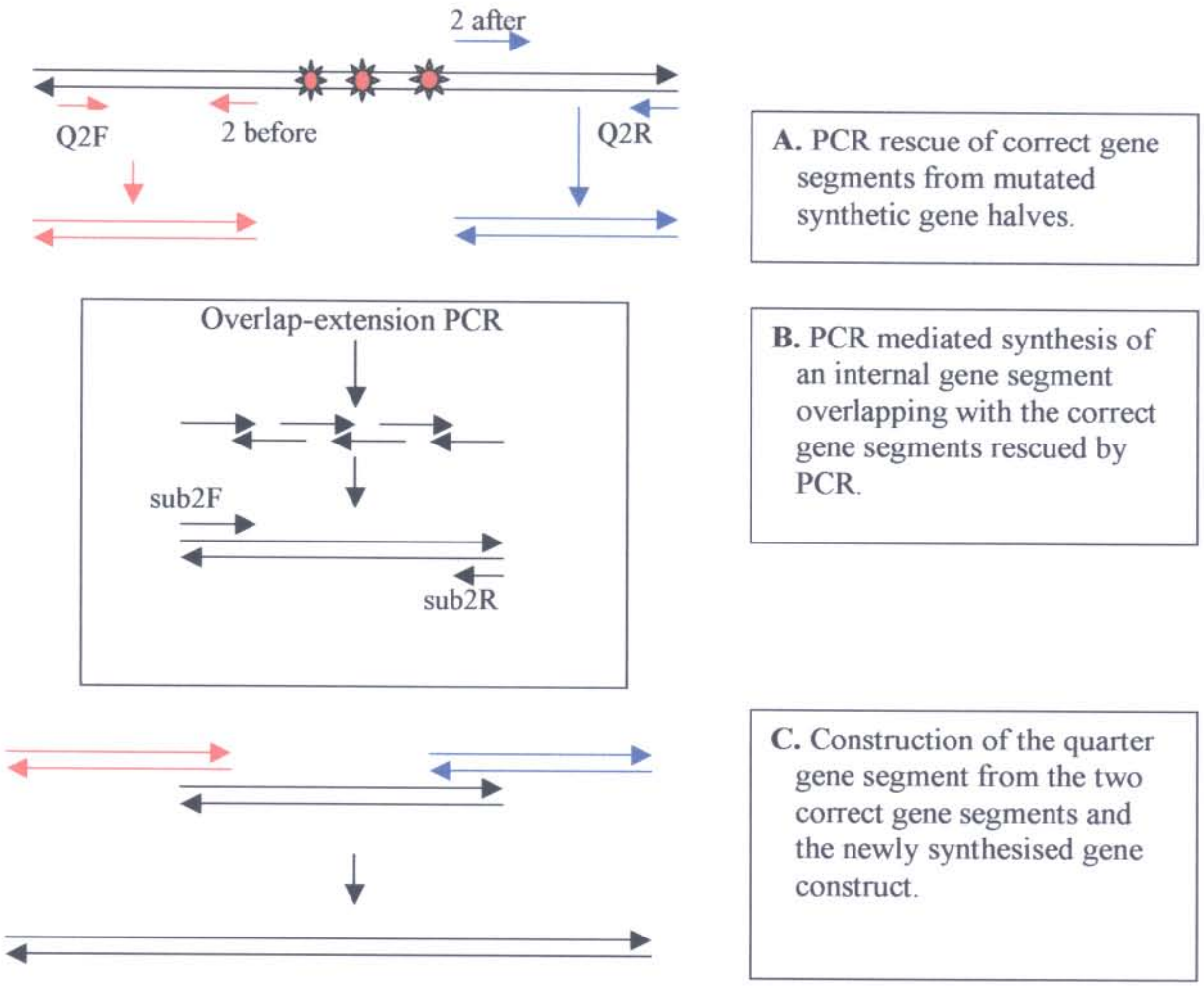

Figure 2.6: Cassette replacement of a mutated gene segment with a newly synthesised internal area of the quarter fragment. The 5' and 3' ends and the primers used for their isolation are indicated in red and blue respectively. Red stars indicate the errors.

Table 2.4: Forward (2f/3f) and reverse (2r/3r) oligonucleotides used for the resynthesis of the quarter 2 and 3 internal fragments. Before (5') and after (3') primers were used for isolation of the correct fragments with existing quarter primers (Table 2.2). Internal forward (sub) and reverse (sub) primers were used for amplification of the resynthesised PCR fragment. The oligonucleotide sequences are given in Appendix D.

Internal fragment 2			Internal fragment 3		
Oligonucleotide name	**Length (nt)**	**Overlap T_m (°C)**	**Oligonucleotide name**	**Length (nt)**	**Overlap T_m (°C)**
2f1	49	58 (2f1/2r1)	3f1	58	50 (3f1/3r1)
2r1	54	55 (2r1/2f2)	3r1	58	64 (3r1/3f2)
2f2	54	57 (2f2/2r2)	3f2	58	53
2r2	54	55 (2r2/2f3)	3r2	59	
2f3	54	61(2f3/2r3)			
2r3	54	55 (2r3/2f4)			
2f4	56	62 (2f4/2r4)			
2r4	54	55 (2r4/2f5)			
2f5	54	55 (2f5/2r5)			
2r5	44				
Primer name	**Sequence (5' to 3')**			**Length (nt)**	**T_m (°C)**
2before	AAACAGTCCCACTTTAAATTTTTTGA			26	55
sub2F	AATTTAAAGTGGGACTGTTT			20	49
sub2R	GTTGCAAATAATCGGCAGGT			20	55
2after	ACCTGCCGATTATTTGCAACGA			22	58
3before	AGAATTTCCAGGGTGCGCAGCGCAAT			26	66
sub3F	TGCGCTGCGCACCCTGGAAATTCT			24	66
sub3R	CACCGCCAGCGGATACTGCACGTT			24	68
3after	ATAACGTGCAGTATCCGCTGGCGGTG			26	68

2.2.9.5 Construction of the full-length gene from correct quarter fragments

The full-length gene was constructed by means of restriction digestion and ligation of the resultant overhangs of the first three correct quarter clones (refer to figure 2.5) followed by PCR assembly with the two alternative fourth quarters and PCR amplification. The 1100 bp fragment and the two alternate versions of the 500 bp fourth segments (Figure 2.5) with and without the stop codon which were cloned into the pMOS*Blue* vector were isolated and purified from their respective vectors by means of *Nde*I/*Eco*RV and *Eco*RV/*Bam*HI restriction digestion, respectively. The full-length gene was constructed by a single overlap-extension assembly of equimolar amounts of the purified 1100 bp fragment and each of the 500 bp fragments. The PCR reaction contained 0.2 pmole of each fragment, 1x *Pfu* DNA polymerase buffer (Promega, Wisconsin, USA) containing 2 mM $MgSO_4$, 1.25 U *Pfu* DNA polymerase and 250 µM of each dNTP in a final volume of 50µl. The PCR profile consisted of 10 cycles

of 94 °C for 30 sec, 50 °C for 30 sec and 72 °C for 2 min. After these cycles, the temperature was taken to 70°C for the addition of 5 pmoles each of the outer primers Q1F and Q4R/alt4R, followed by 10 cycles of 94 °C for 30 sec, 50 °C for 30 sec and 68°C for 2 min. The resulting isolated PCR product was then restriction digested, cloned into the pET15b and pET22b vectors, respectively, using the *Nde*I and *BamH*I sites, and sequenced as described in section 2.2.6.

2.3. Results

2.3.1 Oligonucleotide design

Comparison of the gene sequences for *P. falciparum dhfs-fpgs* and synthetic *dhfs-fpgs* indicated that the synthetic gene has a more equal distribution of the four nucleotides (42.5% G+C and 57.5% A+T) whereas the native gene has a codon bias to A+T (25.49% G+C and 74.51% A+T). This however does not change the amino acid composition, and *P. falciparum* DHFS-FPGS remains skewed towards an abundance of isoleucine (12.3%), lysine (10.2%) and leucine (9.6%). Refer to Appendix D for a list of the oligonucleotides and their complete sequences.

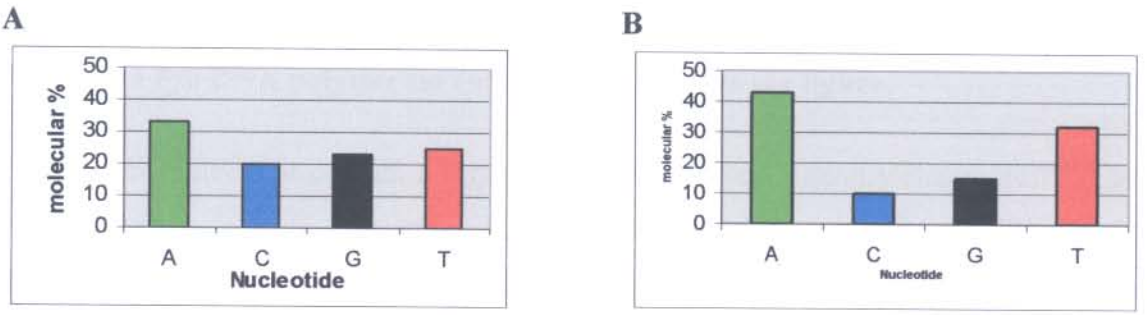

Figure 2.7: Graphs indicating the nucleotide composition of the synthetic *P. falciparum dhfs-fpgs* (A) gene compared to native *P. falciparum dhfs-fpgs* (B).

2.3.2 Optimisation of the overlap extension assembly step

The Taguchi method was used to determine the influence of the Mg^{2+}, dNTP and *Pfu* DNA polymerase concentrations on the efficiency of the PCR assembly step (Cobb and Clarkson, 1994). The overlap-extension PCR protocol was optimised for six oligonucleotides in terms of the Mg^{2+} and dNTP concentrations. The concentrations tested ranged from 150 to 250 µM dNTPs, 2 to 4 mM Mg^{2+} and 1.25 to 2 U *Pfu* DNA polymerase. For this optimisation, the assembly PCR was kept constant at 10 cycles. The amplification was conducted with the standard protocol (section 2.2.2). The results after 20 cycles of amplification are shown in Figure 2.8.

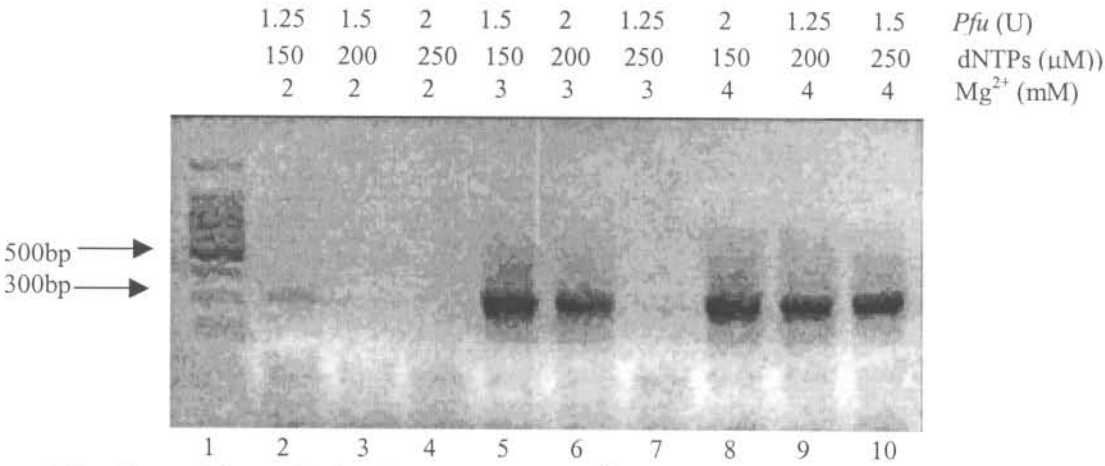

Figure 2.8: Taguchi optimisation of the Mg^{2+}, dNTP and *Pfu* DNA polymerase concentrations for the assembly of six oligonucleotides. Lane 1: 100 bp molecular marker. The arrows indicate the 500bp position and the expected band size (300bp). Mg^{2+} concentrations of each reaction are indicated in the lanes below the figure and dNTP and *Pfu* DNA polymerase concentrations above the figure.

It should be noted that overall, a higher magnesium concentration yields more PCR product. Where the magnesium concentration stays the same, an increased dNTP concentration decreases the PCR efficiency. An increase in the enzyme concentration is not directly linked to an increase in PCR product generation. At a higher magnesium concentration however, a higher enzyme concentration yields more PCR product (Figure 2.8, lane 8)

Using the information gained from the Taguchi experiment, refining of each of the above PCR parameters (the amount of dNTPs, Mg^{2+} and *Pfu* DNA polymerase) was optimised separately in the following three experiments: Firstly, the dNTP concentration (250 μM) was kept constant while using different Mg^{2+} concentrations (4 mM, 4.5 mM and 5 mM). This range was chosen since a high magnesium concentration was observed to give better yields of PCR products (Figure 2.8). 2.5 U *Pfu* DNA polymerase was used since the Taguchi experiment showed that a higher amount of enzyme at 4 mM Mg^{2+} yielded more intense PCR products (Figure 2.8, compare lanes 8, 9 and 10). Secondly, the dNTP concentration was varied (150 μM, 200 μM and 250 μM) at a constant amount of 4 mM Mg^{2+} and 2.5 U *Pfu* DNA polymerase. In the third experiment 1.5 U, 2 U and 2.5 U *Pfu* DNA polymerase was tested in PCR reactions containing 4 mM Mg^{2+} and 250 μM dNTPs. The assembly PCR cycles were kept constant at 10 cycles and amplification at 20 cycles. Refer to section 2.2.2 for the cycling profile.

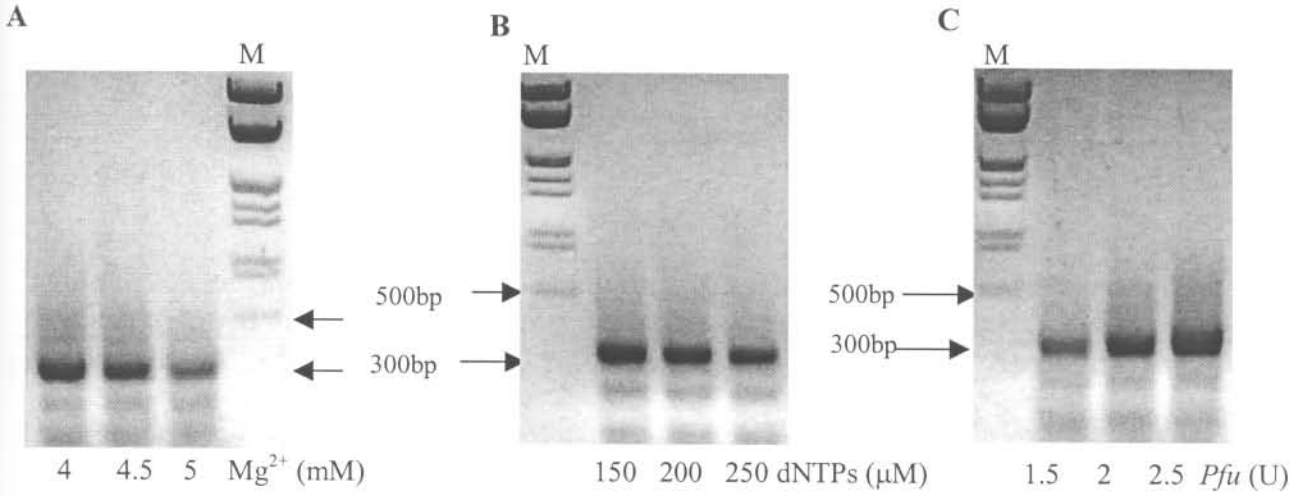

Figure 2.9: Refinement of the PCR parameters for the assembly PCR step. Optimization of the Mg^{2+} concentration (A), dNTP concentration (B) and the amount of *Pfu* DNA polymerase (C). M indicates the molecular marker, *EcoR*I and *Hind*III digested λ DNA. The arrows indicate the 500bp position and the expected band size (300bp).

From the DNA gels above it is evident that the efficiency decreases at 5mM Mg^{2+} (Figure 2.9 A). As the dNTP concentration increases, the PCR efficiency also decreases significantly (Figure 2.9 B). Increasing the amount of *Pfu* DNA polymerase however increases the PCR efficiency (Figure 2.9 C). From the above results the optimal conditions for PCR assembly was thus chosen as 4 mM Mg^{2+}, 250 mM of each dNTP, 1x *Pfu* DNA polymerase buffer and 2.5 U *Pfu* DNA polymerase.

2.3.3 Optimisation of the number of assembly cycles

The parameters in 2.3.2 were optimised for 1 pmole each of six 3' ends at a constant number of 10 assembly cycles. For the synthesis of fragments made up from different numbers of oligonucleotides, the optimal Mg^{2+} and dNTP concentrations as determined in section 2.3.2 were used and the molar amounts of the fragments adjusted to achieve the molar equivalent of six 3' ends. Thus 0.6 pmole of ten 3' ends and 1.6 pmoles of four 3' ends were taken to be equivalent to 1 pmole of six 3'ends (Table 2.5). This ensured a constant ratio of the concentration of 3' ends to the Mg^{2+} concentration under the optimised conditions. The number of assembly cycles required for fragment generation was also extrapolated from the optimised conditions. Fragments assembled from more oligonucleotides than used in the optimised conditions required proportionally more assembly cycles, which was adjusted as shown in Table 2.5.

Assembly of these fragments was thus performed with the optimised conditions of 4 mM Mg^{2+}, 250 mM dNTPs, 1x *Pfu* DNA polymerase buffer and 2.5 U *Pfu* DNA polymerase, as well as the specific amount of oligonucleotides as described above for 7 to 17 assembly cycles. A 50x, 10x and 5x dilution of the assembly mixtures was amplified for 20 cycles (section 2.2.2).

Table 2.5: The theoretical and actual number of cycles needed for complete template generation based on the number of oligonucleotides involved.

Amount of 3' ends	Theoretical amount of cycles	Number of cycles used	Amount of each oligonucleotide required (pmole)
Six 3'ends	3	10 (optimised in 2.3.2)	1
Ten 3'ends	5	17	0.6
Four 3'ends	2	7	1.6

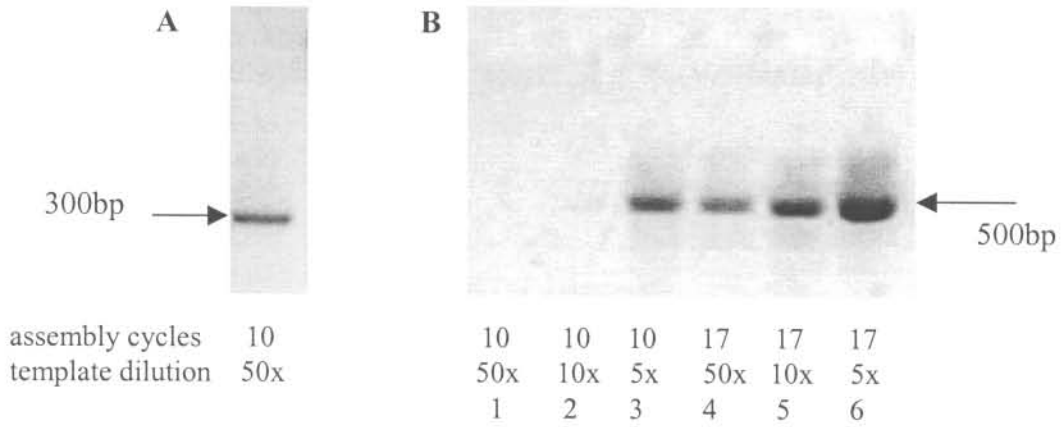

Figure 2.10: PCR products obtained for fragments assembled from 6 oligonucleotides (A) or 10 oligonucleotides (B). B lanes 1 to 6: PCR products obtained from 10 and 17 assembly cycles of 10 oligonucleotides at different dilutions of the assembly mixture (indicated below each lane). Arrows indicate the expected fragment sizes for 6 oligonucleotides (300bp) and ten oligonucleotides (500bp) after 20 amplification cycles.

It was observed that with more oligonucleotides, more assembly cycles were required to obtain PCR products at the same template dilution (compare lanes 1 and 4, lanes 2 and 5 and lanes 3 and 6, Figure 2.10 B). An assembly of 17 cycles for 10 oligonucleotides followed by a 50x dilution of the assembly reaction for amplification was taken to be preferable to an assembly of 10 cycles followed by a 5x dilution of the assembly reaction for amplification (compare lanes 3 and 4, Figure 2.10 B).

2.3.4 Construction of the full-length gene from half fragments

The full-length gene obtained from a single overlap-extension PCR of the gene halves was cloned and sequenced. In this instance 0.01 pmole of the halves gave the most specific PCR product after a total of 25 PCR cycles, when comparing the signal obtained to background 'noise' (Figure 2.11).

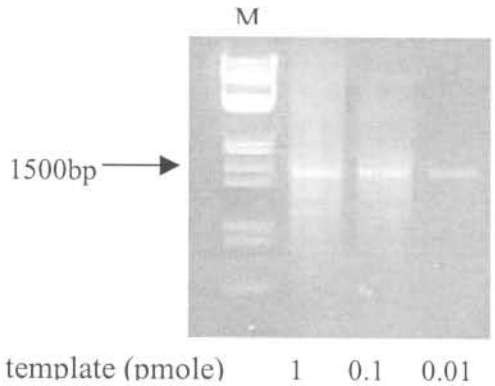

Figure 2.11: The full-length synthetic *dhfs-fpgs* gene obtained from overlap extension PCR of the two 750 bp half fragments. The dilution of each of the half fragments (template dilution) is indicated below the figure; M is the molecular marker (*EcoR*I and *Hind*III digested λ DNA).

2.3.5 Gene repair

Sequencing of gene halves and the full-length gene initially revealed errors in the form of single base insertions and deletions. The collective sequences of 65 clones were aligned in the 5' to 3' direction and the amount of errors per nucleotide position determined as a percentage of the number of oligonucleotides. This showed that the errors were mostly concentrated at the 5'-end overlaps between oligonucleotides. Scoring of the average errors of all the oligonucleotides as a percentage of the total number of errors at specific positions within the oligonucleotide, indicated an increase of the number of errors from the 3'- to the 5'-end (Figure 2.12). Insertions were localised to the 5' end whereas deletions were distributed in a more random fashion. The 3' end contained more deletions than insertions and the 5' end contained more insertions than deletions. Gene synthesis from newly synthesized shorter oligonucleotides (~50nt) revealed no errors.

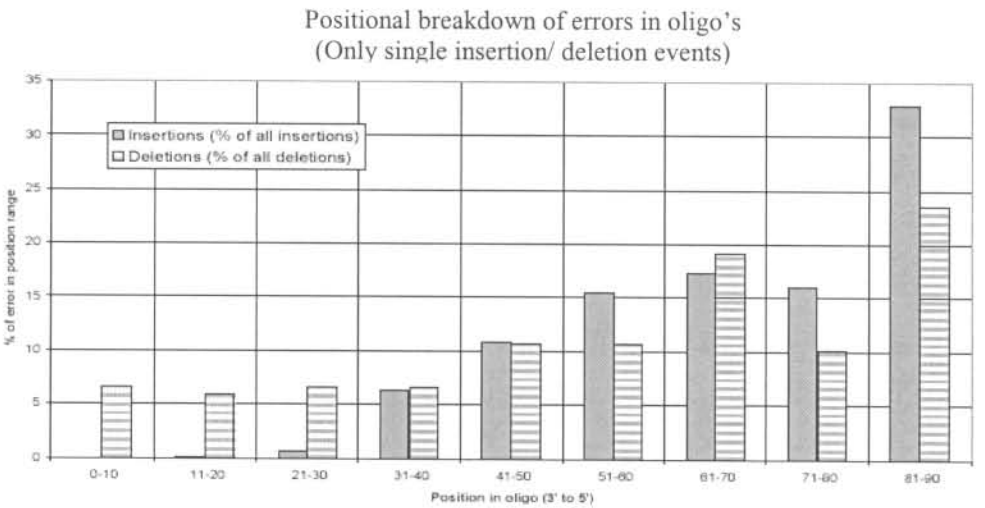

Figure 2.12: The direct relationship between the error rate and the proximity to the 5' end of each oligonucleotide. This graph indicates the relationship between the number of errors (expressed as a percentage of the total number of errors) to its 3' to 5' position in the oligonucleotide. The position ranges are divided in window sizes of 10 nucleotides. Solid bars indicate insertions and striped bars indicate deletions.

Quarters 2 and 3 were corrected by means of cassette mutagenesis of areas containing errors with short overlapping resynthesised gene fragments. The newly synthesised internal fragments used for cassette mutagenesis (Figure 2.13, lane 3) and the correct 5'-end and 3'-end sequences (Figure 2.13, lanes 2 and 4) were combined in an overlap extension PCR assembly of the quarter 2 and 3 fragments.

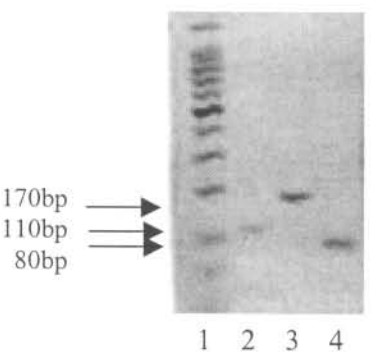

Figure 2.13: Construction of the correct quarter 2 by means of cassette mutagenesis. Lane 1: 100bp molecular marker, lanes 2 and 4: correct 5' and 3' fragments isolated by means of PCR, lane 3: fragment resynthesised from shorter oligonucleotides

After cloning and sequencing at least one correct quarter 2 and 3 clone of three clones sequenced was obtained (Figure 2.14, lanes 3 and 4 respectively).

Quarter 1 was corrected by means of restriction digestion and ligation of two different quarter 1 clones, each with the correct sequence on opposite sides of a unique restriction enzyme site, *Hha*I. This ligation product was cloned and sequenced and one of two clones sequenced contained the correct quarter 1 sequence (Figure 2.14, lane 2)

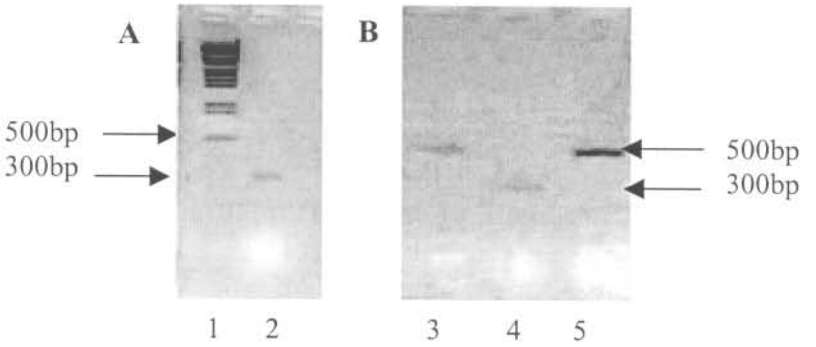

Figure 2.14: Corrected quarter fragments. Lane 1: molecular marker (*EcoR*I and *Hind*III digested λ DNA), lane 2A: quarter 1(300bp) isolated after *Nde*I/*BstX*I digestion, lane3B: quarter 2 (500bp) isolated after *BstX*I/*Sac*I digestion, lane 4: quarter 3 (300bp) isolated after *Sac*I/*EcoR*V digestion and lane 5: quarter 4 (500bp) isolated after *EcoR*V/*BamH*I digestion of the respective plasmids. The arrows indicate the expected fragment sizes.

Quarter 4 was repaired by means of two site-directed mutagenesis steps. Since two of the errors, a single base insertion and triple base deletion were within 10 bases of each other, a single primer pair was sufficient to correct both errors (Table 2.3, SMD4.1F and –R). After cloning and verification of the correct sequence, the second site-directed mutagenesis step was performed on a single base insertion, ~100 bases upstream of the other errors. The corrected quarter 4 was cloned and sequenced (Figure 2.14, lane 5). At least one of every two clones sequenced revealed the correct sequence after site directed mutagenesis.

2.3.6 Construction of the full-length gene from correct quarter fragments.

Once the clones containing the correct quarter sequences were identified, the quarter fragments were isolated from the respective plasmids and the three-quarter gene assembled by means of restriction-ligation, using the engineered restriction sites. An 1100bp fragment was obtained after restriction-ligation and cloning of the first three quarters (Figure 2.15, lane 2).

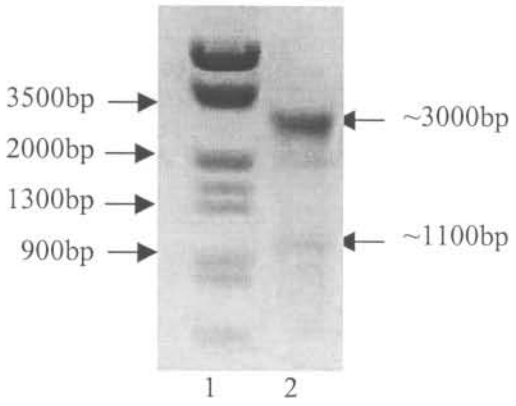

Figure 2.15: Isolation of the 1100bp fragment consisting of the 1st three fragments by means of restriction digestion. Lane 1: molecular marker (*EcoR*I and *Hind*III digested λ DNA), lane 2: *Nde*I/*EcoR*V digested pMOS*Blue* plasmid containing the 1100bp fragment.

The 1100bp fragment was used in a single overlap-extension PCR with the two alternate fourth quarters (with and without the stop codon) for the generation of two versions of the gene to enable N- and C-terminal His_6 tagged expression, respectively (Figure 2.16 lanes 1 and 2).

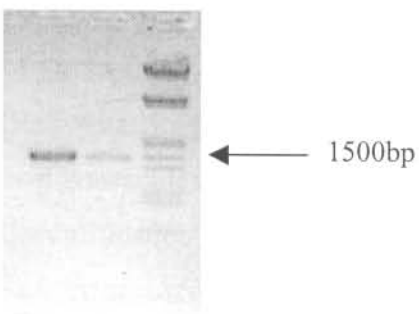

Figure 2.16: Full-length *dhfs-fpgs* constructs obtained after overlap-extension of the 1100bp fragment and alternative 500bp fragments. Lane 1: full-length gene with the stop codon, lane 2: full-length gene without the stop codon, lane 3: molecular marker. The arrow indicates the size of the full-length gene.

2.4 Discussion

2.4.1 Overlap-extension PCR as a gene synthesis method

The value of synthetic genes for the successful expression of proteins, especially those of complicated systems such as the A+T rich malaria genome, is clear through past successes (Prapunwattana *et al.*, 1996; Zhang *et al.*, 2002). Gene synthesis can be achieved through different methods such as the restriction-ligation of cloned double-stranded oligonucleotides or the PCR-based assembly of overlapping oligonucleotides. The first approach involves a large number of complicated cloning and sub-cloning steps and is limited by the length of the gene (Prapunwattana *et al.*, 1996). The longer the gene, the more unique restriction sites, cloning vectors and cloning steps are needed, since extremely long oligonucleotides (>100nt) can not be chemically synthesised without compromising the fidelity of the sequence. The second approach is a simpler method and not restricted by the length of the gene. The assembly of a gene from 125 pmoles of completely overlapping oligonucleotides using a total of 50 PCR cycles was described previously (Carpenter *et al.*, 1999). To cut the costs of this approach, we decided to use only partially overlapping oligonucleotides and to apply the PCR process more efficiently for extension of the gaps between overlapping oligonucleotides. For a 1500bp gene, this reduced the number of nucleotides required from 3000nt (for complete overlapping oligonucleotides) to ~2500nt (for partially overlapping nucleotides), reducing the cost with 17%. A further reduction in the cost was achieved by using less oligonucleotides, 1 pmole of each for only 10 assembly and 20 amplification cycles. The reduction in the number of PCR cycles reduced the possibility of error incorporation during PCR. *Pfu* DNA polymerase with 3' to 5' proofreading activity was also used to increase the fidelity of the PCR protocol.

The 1586 bp *dhfs-fpgs* gene was subdivided into halves, which were constructed by means of a two-step PCR protocol. In the first step each half was assembled from 16 oligonucleotides by means of overlap-extension PCR. In the second step the assembly mixtures were diluted and the halves amplified with specific primers. The full-length gene was then obtained by a single overlap-extension PCR step of the two halves. The 5' and 3' outer primers (Table 2.1) introduced unique restriction enzyme sites into the gene for directional cloning into expression vectors and to enable cassette mutagenesis. The 3' end primers also incorporated alternative 3' gene ends for expression from N- or C-terminal His_6 tagged vectors.

2.4.2 Optimisation of the assembly PCR parameters

The first step of the PCR protocol (assembly of oligonucleotides) was optimised in terms of the Mg^{2+}, dNTP and *Pfu* DNA polymerase concentrations. The Taguchi method indicated that an increased Mg^{2+} concentration promoted PCR efficiency (Figure 2.8). In multiplex PCR reactions, where a large number of different primers anneal to a single template, an increased Mg^{2+} concentration, sometimes up to 8mM, is required for efficient primer annealing and extension (Zangenberg *et al.*, 1999). The PCR assembly is similar to a multiplex reaction due to the large number of oligonucleotide overlaps that have to anneal to each other. Variations in the Mg^{2+} concentration, however, have an effect on other PCR components, such as the dNTP concentration and enzyme activity. Mg^{2+} chelates dNTPs, thus the amount of dNTPs have to be increased as the Mg^{2+} concentration increases. Too high Mg^{2+} concentrations could lead to mispriming due to increased non-specific annealing and also lower the available dNTP concentration. Low dNTP concentrations could result in the misincorporation of bases due to a decreased availability of dNTPs whereas high dNTP concentrations could activate the 3' to 5' exonuclease activity of the *Pfu* DNA polymerase resulting in the degradation of the free 3' ends of the oligonucleotides (*Pfu* DNA polymerase Technical data, Promega). To increase overall PCR efficiency a higher amount of enzyme (2.5 U instead of the recommended amount of 1.25U as indicated by the manufacturer) was used. The refined optimisation indicated the optimal concentrations as 4 mM Mg^{2+}, 250 µM dNTPs and 2.5U *Pfu* DNA polymerase. For all optimisation experiments, the same amplification protocol (section 2.2.2) was used. The PCR products obtained were only visible after the amplification step, since the concentration of assembly products was too low.

2.4.3 Optimisation of the number of assembly cycles

The above parameters (section 2.4.2) were optimised for the assembly of 1 pmole each of six oligonucleotides for ten assembly cycles. These conditions were applied to other fragments consisting of different numbers of oligonucleotides by varying the number of assembly cycles. Based on the number of 3' ends involved in the annealing process the number oligonucleotides used for assembly was thus adjusted in such a manner that the absolute concentration of 3' ends to the Mg^{2+} and dNTP concentration remained the same. For a larger number of oligonucleotides, a proportionally smaller pmole amount was used and vice versa, i.e. if six oligonucleotides were used at a concentration of 1 pmole each, then the mixture consisted of 6 pmoles 3' ends. Similarly, for four oligonucleotides a concentration of 1.6

pmoles of each would give a total of 6 pmoles of 3' ends and for ten oligonucleotides a concentration of 0.6 pmole each would give 6 pmoles of 3'ends. The number of assembly cycles was also varied in proportion to the optimised number of 10 cycles. For ten oligonucleotides the number of assembly cycles was increased from 10 to 17 (Table 2.5). PCR products were obtained after only 20 amplification cycles of a 50x dilution of the assembly mixture. A 50x dilution of the assembly mixture was preferable to a 10x or 5x dilution since the dilution was used to reduce the interference of incompletely assembled products in the assembly mixture during the amplification cycles. The dilution of the assembly mixture together with the addition of 5' and 3'end- specific primers thus drive the generation of the complete assembled product, underlining the importance of the use of a two-step PCR protocol.

2.4.4 Construction of the full-length gene

The full-length gene (~1580 bp) was successfully constructed from the single overlap extension PCR of purified gene halves by means of a one-step PCR protocol. Different concentrations of the halves (1 pmole, 0.1 pmole and 0.01 pmole) were used to determine the optimal template to primer ratio (theoretically the primers have to be in a 10^4 molar excess; Sambrook *et al.*, 1989). 0.01 pmole of each half fragment was sufficient to obtain PCR products within 20 amplification cycles. The lower template concentration was thus optimal to obtain the most specific PCR products. Since there was only a single overlap between the halves, there was thus no interference from incompletely assembled PCR products and no dilution of the assembly mixture was needed before amplification to increase the specificity of the reaction, in contrast to the two-step procedure used for the assembly of >3 oligonucleotides. Furthermore, since this protocol involved the annealing of one overlap, higher Mg^{2+} and enzyme concentrations (used for the two-step protocol) were unnecessary and the manufacturer's recommended amount of 2mM Mg^{2+} and 1.25 U *Pfu* DNA polymerase was used. In theory the single overlap extension occurs during one PCR cycle, but an initial 5 assembly cycles was used for the extension of the single overlap to generate a sufficient amount of the full-length gene required for specific amplification. This was followed by the addition of the outer primers and amplification of the full-length template.

2.4.5 Gene repair

Initial sequencing revealed mostly single base insertions or deletions over the whole length of the gene sequence. PCR generated errors are normally due to misincorporations caused by the decreased fidelity of the polymerase or multiple-base deletions due to secondary structures of the template. Single base insertions and -deletions observed by the sequence alignment of 65 clones revealed that these errors were concentrated at the 5' overlap areas between oligonucleotides. There was also an observed shift from mainly deletions at the 3' end of each oligonucleotide to mainly insertions at the 5' end (Figure 2.12). Taken together all these results indicated that the PCR process was not responsible for the errors. Since the chemical synthesis of oligonucleotides occur in the 3' to 5' direction and on average, the highest number of errors occurred at the 5' end of each oligonucleotide, the errors were attributed to poor quality control by the manufacturer during the chemical synthesis of the oligonucleotides. During phosphoramidite oligonucleotide synthesis the growing oligonucleotide chain is linked through the 3' OH group of the deoxyribose sugar of the first deoxyribonucleotide to a solid matrix, controlled pore glass beads (CPG). The elongation of the oligonucleotide chain by a single oligonucleotide, involves intricate chemical steps. The 4,4'- dimethoxytrityl (DMT) 5' OH protecting group of the deoxyribose sugar of the previous dexoyribonucleotide has to be removed just prior to coupling of the next deoxyribonucleotide. The deprotected 5' end is then coupled to the 3' phosphoramidite derivative of the next deoxynucleoside with the coupling agent, tetrazole. Under prolonged coupling conditions (eg the synthesis of long oligonucleotides) the acidic Tetrazol, used for deprotection during synthesis could remove the protective trityl group from the deoxyribonucleotide, resulting in double-coupling events, observed here as insertions (Fu *et al.*, 2002). Oligonucleotides are normally only guaranteed up to 60 bases in length due to this reason. After coupling, unreacted 5' ends are blocked by acetylation and the phosphite triester resulting from the coupling step is oxidised to a phosphotriester. When synthesis is complete, blocking groups on the bases are removed and the oligonucleotide chain cleaved from the solid support. This can then be purified by HPLC to remove truncated products, but the HPLC is not sensitive enough to detect single base insertions or deletions (Voet and Voet, 1995).

Error-free quarter 2 and 3 sequences could be obtained from the cassette mutagenesis of internal fragments with newly synthesised, shorter (~50nt long) oligonucleotides from as little as three clones. PCR generated errors were few (less than one in 500bp) and were directly proportional to the number of oligonucleotides involved in the assembly. The resynthesised

fragments furthermore include important structural features of the enzyme, such as the P-loop (GTNGKGS; nt 313-333), which is involved in the binding of ATP and the Ω loop (VGLFSSPHIFSLRERI; nt.379-426) that is responsible for stabilisation interactions between domains (Sun *et al.*, 1998). Future structure-function studies are thus possible by cassette mutagenesis of these internal fragments.

Other minor errors were corrected by restriction-ligation or site-directed mutagenesis. The correct first quarter was obtained from the restriction-ligation of two clones containing correct sequences on opposite sides of the unique *Hha*I restriction enzyme site. The correct fourth quarter was obtained after two successive site-directed mutagenesis experiments.

2.4.6 Construction of the full-length gene from quarter fragments

For ease of cloning and sequencing, quarter fragments were used during the repair of the gene. Correct quarter clones were identified by sequencing after the corrections were made as stated above in section 2.4.5. These first three quarter fragments were ligated, using the engineered unique restriction enzyme sites into an 1100bp fragment. The use of unique restriction enzyme sites prevented digestion of the vector or gene at any area besides the overlaps. The two different fourth quarters (with and without the stop codon) were then assembled with the 1100bp fragment into the full-length gene, which was cloned and sequenced. The correct sequences for both *dhfs-fpgs* constructs were obtained.

This chapter described the simple, quick, efficient, low cost synthesis of a codon-adapted gene by means of a two-step PCR method. Costs were cut through the use of partially overlapping oligonucleotides instead of complete overlapping oligonucleotides as well as >100 times less starting material, when compared to previous methods. Provided high quality oligonucleotides are used, correct synthetic genes of virtually any length can be obtained by this method. The oligonucleotides can also be used in future cassette mutagenesis experiments for structure-function studies. The purpose of the synthetic gene described here, *P. falciparum dhfs-fpgs* is to increase recombinant expression in *E. coli*. The synthetic gene was therefore successfully adapted to *E. coli* codon preferences and has a 17% reduction in A+T content. Synthesis of two alternate versions of the gene also enables N- or C-terminal His_6-tagged expression of the protein. The next chapter will focus on the recombinant expression of various tagged versions of this synthetic gene in *E. coli*.