Chapter 4

In silico analysis of dihydrofolate synthase-folylpolyglutamate synthase (DHFS-FPGS)

4.1 Introduction

The role of folylpolyglutamate synthase (FPGS) in the cellular retention of reduced folate co-factors and inhibitors has made it an interesting target for cancer and bacterial chemotherapy (Gangjee *et al.*, 2002). Although FPGS occurs in all cells, only one crystal structure has been determined to date; namely that of the prokaryote *Lactobacillus casei* (Sun *et al.*, 1998). In contrast with FPGS, dihydrofolate synthase (DHFS) has been much less studied to date, since it only occurs in organisms capable of folate biosynthesis (section 1.5.1). In mammals (that import dietary folate) and even certain bacteria such as *L. casei* and *H. influenzeae*, DHFS activity is absent (Sun *et al.*, 1998). Eukaryotes such as *A. thaliana* and yeast have separate genes encoding DHFS and FPGS activity, but *P. falciparum* is the first eukaryote to date to contain both activities within a single gene product (Ravanel *et al.*, 2001) and (Salcedo, 2001)

.

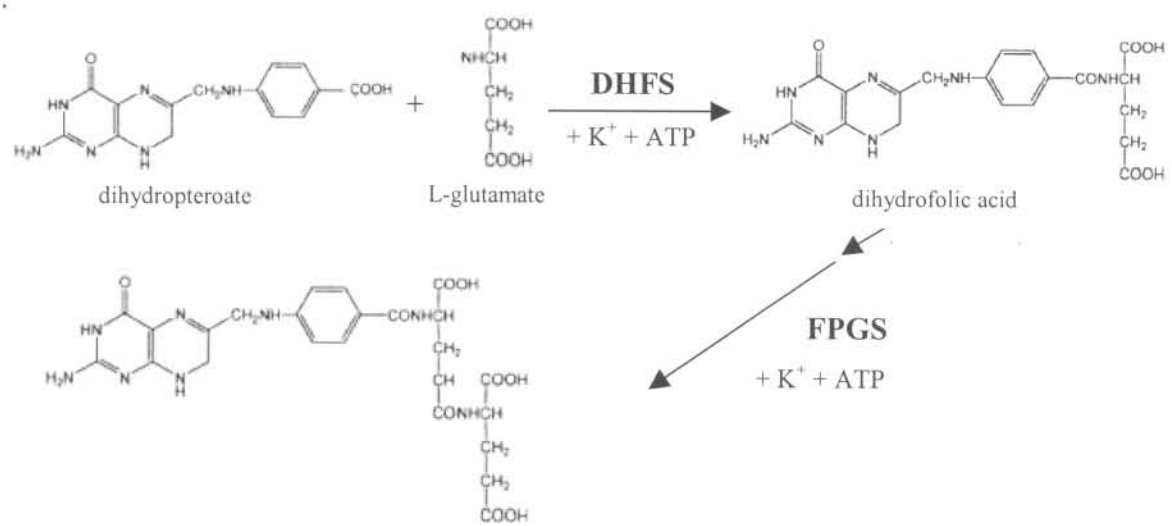


Figure 4.1: DHFS and FPGS enzyme reactions (adapted from Yuthavong, 2002).

Since DHFS and FPGS perform essentially the same enzyme reactions, namely the ATP dependant addition of a glutamate residue to a pteridine derivative, it is possible that the

bifunctional enzyme only contains one active site (Figure 4.1). This idea is supported by the fact that a single mutation abolishes both DHFS and FPGS activities in the bifunctional *E. coli* enzyme (Keshavjee *et al.*, 1991). The *L. casei* crystal structure revealed that the enzyme consisted of two domains (Figure 4.2): the N-terminal domain consisting of a mononucleotide binding fold (P-loop) similar to that found in proteins of the adenylate kinase family and a C-terminal domain, which is very similar to the folate-binding dihydrofolate reductase enzyme (Sun *et al.*, 1998). The active site is located between these two domains next to the P-loop and the folate-binding site is located in a C-terminal domain hydrophobic pocket between two alpha helices, A10 and A11 (Sun *et al.*, 1998). Despite the similarities between the FPGS C-terminal domain and folate-binding site of human DHFR, it is proposed that the orientation of the folate substrate, dihydrofolate (DHFR) and 5,10-methylene tetrahydrofolate (FPGS) might differ (Sun et al., 1998). In FPGS, the pteridine ring structure (fused white circles, Figure 4.2) will interact with the hydrophobic area between helices A10 and A11 and the growing chain of glutamate residues (dashed line, Figure 4.2) will project into the active site (dihydrofolate is in the opposite orientation). Another important feature is the 10-amino acid Ω-loop (cyan line, Figure 4.2), which has an important role in binding of the K^+ cations as well as interdomain stabilisation through hydrophobic interactions with the C-domain (Sun *et al.*, 1998). The linker region between the N- and C-terminal domains (green line, Figure 4.2) is important for domain movement and differences in the flexibility of this region might account for the different lengths of glutamate residues added in the various species (Sun *et al.*, 2001).

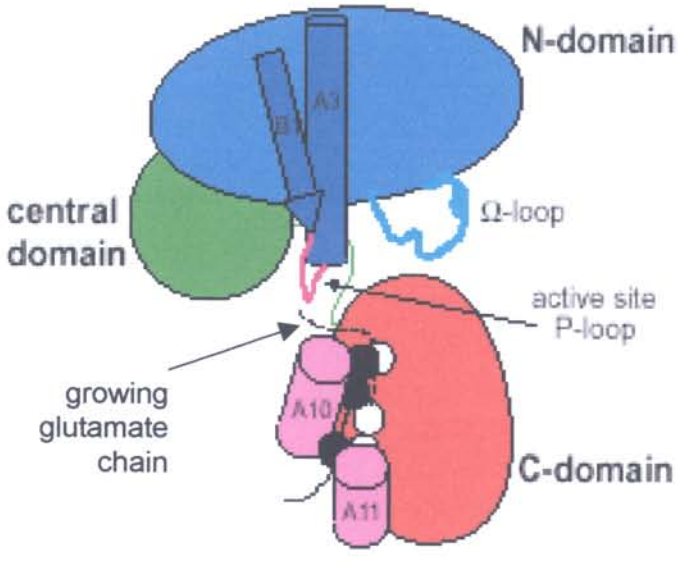


Figure 4.2: Schematic representation of *L. casei* FPGS. The two different binding modes of the folate substrates between helices A10 and A11 are compared for FPGS (white circles) and another folate-binding enzyme, DHFR (black circles) (Sun *et al.*, 1998).

To date, *P. falciparum* DHFS-FPGS has been characterized in terms of the DNA sequence, the predicted amino acid sequence and functional complementation. Comparison of these sequences with homologues indicated the presence of the ATP-binding P-loop, Ω loop as well as the FPGS signature sequence and selected residues required for catalytic activity (Salcedo *et al.*, 2001). Overall, alignments between the *P. falciparum* DHFS-FPGS and other homologues indicate low similarities (~30%), and even lower identities (~17%) (Lee *et al.*, 2001). Given this information, the possibility of designing selective drugs that are not detrimental to the human host, are increased. The predicted molecular mass of *P. falciparum* DHFS-FPGS is 56 kDa and the predicted pI 6.19 (http://us.expasy.org/cgi_bin/pi_tool). Apart from this limited information, nothing else of this enzyme is known. This chapter aims at gaining insight on certain predicted structural and genetic features of *P. falciparum* DHFS-FPGS to direct future mutagenesis experiments for determination of the enzymes' structure-function relationships.

4.2 Methods

4.2.1 Sequence alignments

DHFS and FPGS sequences were obtained for as many different organisms possible from the Swissprot and Swissall databases (http://srs.ebi.ac.uk). Monofunctional DHFS and FPGS as well as bifunctional DHFS-FPGS sequences were given in FASTA format for alignment by CLUSTAL X v 1.81 using the default gap penalty values (Thompson *et al.*, 1997). A phylogenetic tree based on this alignment was also obtained from CLUSTAL X. (ftp://ftp-igbmc.u-strasbg.fr/pub/**ClustalX**/). The sequences of various *Plasmodium* strains, the rodent parasite *Plasmodium yoelii*, as well as the less pathogenic human parasite *Plasmodium vivax*. were obtained by running a TBLASTN search (Altschul *et al.*, 1990) with the *P. falciparum* DHFS-FPGS sequence against *Plasmo*DB (http://PlasmoDB.org). The highest scoring hits ($E < 10^{-130}$) were aligned with the human host FPGS sequence and *P. falciparum* DHFS-FPGS sequences by CLUSTAL X to determine the interspecies sequence conservation as well as host-parasite differences.

4.2.2 Structure predictions

The secondary structure was predicted from the primary amino acid sequence with a range of independent computer programs: profile fed neural network systems (PHD) (Rost and Sander, 1993), GOR4 (Garnier *et al.*, 1996), hierarchical neural networks and SCRATCH Sspro (http://www.igb.uci.edu /tools/scratch/). A hydrophobicity profile for the primary amino acid sequence was obtained using Kyte and Doolittle parameters and a window size of 10 amino acids. An extensive search of the database PROSITE (Bairoch *et al.*, 1997) was run on the predicted primary amino acid sequence to identify possible amino acid sequence motifs involved in the activity or structure of the enzyme. Similarity searches also used the 3D-structure database PDB with the program SAM-T99 (Karplus *et al.*, 1998) (http://www.cse.ucsc.edu/research/compbio/HMM-apps/sam-t99) to identify proteins that could possibly share structural features with *P. falciparum* DHFS-FPGS. A Predict Protein search (http://www.embl-heidelberg.de/predictprotein/) was done, which is an extensive program searching various protein features such as the SEG low complexity sequences (Wootton and Federhen, 1996). A BLASTP search (Altschul *et al.*, 1990) of the ProDom database http://www.toulouse.inra.fr/prodom.html (Corpet *et al.*, 1999) was performed to

identify proteins with similar domains. A COILS version 2.2 search (Lupas, 1996) was performed to identify possible coiled-coil regions within the protein. To predict the solvent accessibility of amino acids and thus overall protein globularity, a GLOBE search was done (http://www.columbia.edu/~rost/Papers/98globe). Ten preliminary models of the *P. falciparum* DHFS-FPGS enzyme, based on the *L. casei* crystal structure was built with the program MODELER version 6v2 (Sali and Blundell, 1993) (Accelrys ®) using default parameters and visualised with the INSIGHT II program (Accelrys ®) (www.accelrys.com/insight/Modeler). Ramachandran plots for the 10 models were generated with PROCHECK (Morris *et al.*, 1992).

4.3 Results

4.3.1 Inter-species DHFS and FPGS alignments and phylogenetic analysis

Since no distinction between *P. falciparum* DHFS and FPGS domains can be made at the primary amino acid level, the predicted bifunctional sequence was aligned with homologous monofunctional DHFS or FPGS or bifunctional DHFS-FPGS enzymes (Figure 4.4). Sequences were obtained for prokaryotes: monofunctional *Streptococcus pneumoniae* DHFS, *Bacillus subtillus, Lactobacillus casei* and *Haemophilus influenzeae* FPGS and the bifunctional DHFS-FPGS of *Escherichia coli, Buchnera aphidicola* and *Buchnera biphidocola* as well as eukaryotes: yeast DHFS and FPGS, *Candida albicans*, *Arabidopsis thaliana*, human and mouse FPGS. A phylogenetic tree drawn from the alignment shows as expected that mammal (mouse and human) FPGS group together. Prokaryotic bifunctional DHFS-FPGS group together while prokaryotic monofunctional FPGS group together separately. *P. falciparum* DHFS-FPGS seems to be the closest related to prokaryotic bifunctional DHFS-FPGS. Yeast DHFS does not group with any other homologue and *Arabidopsis* and *Neurospora crassa* also group separately from the rest (Figure 4.3).

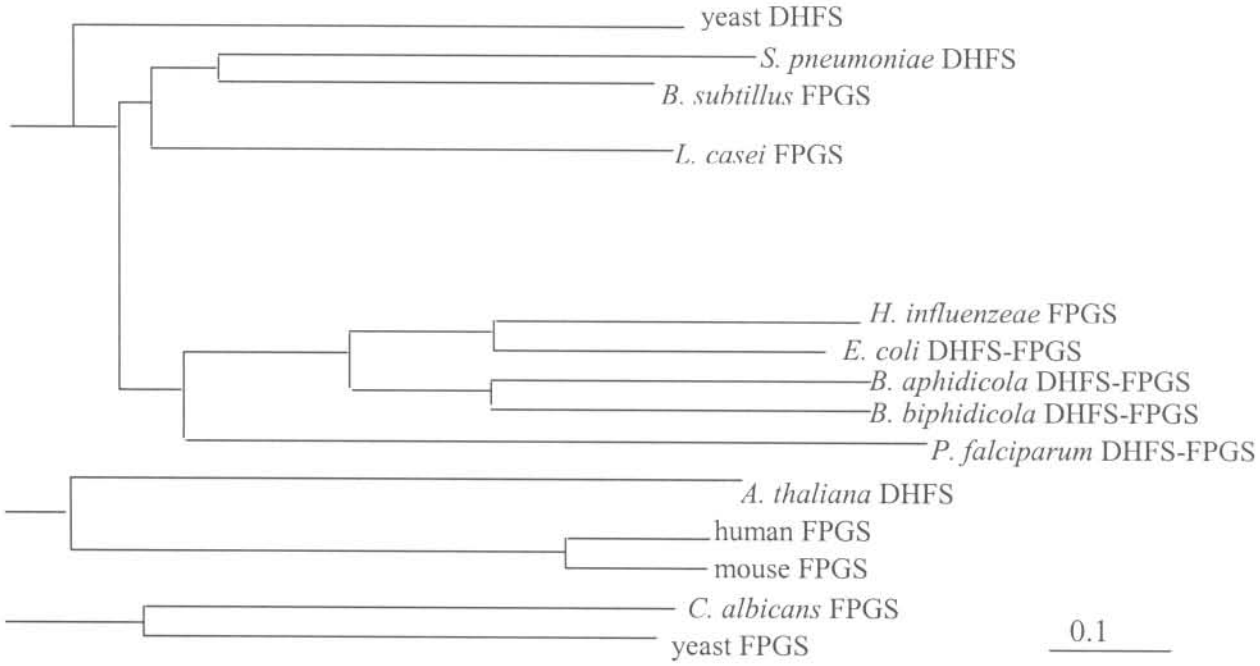
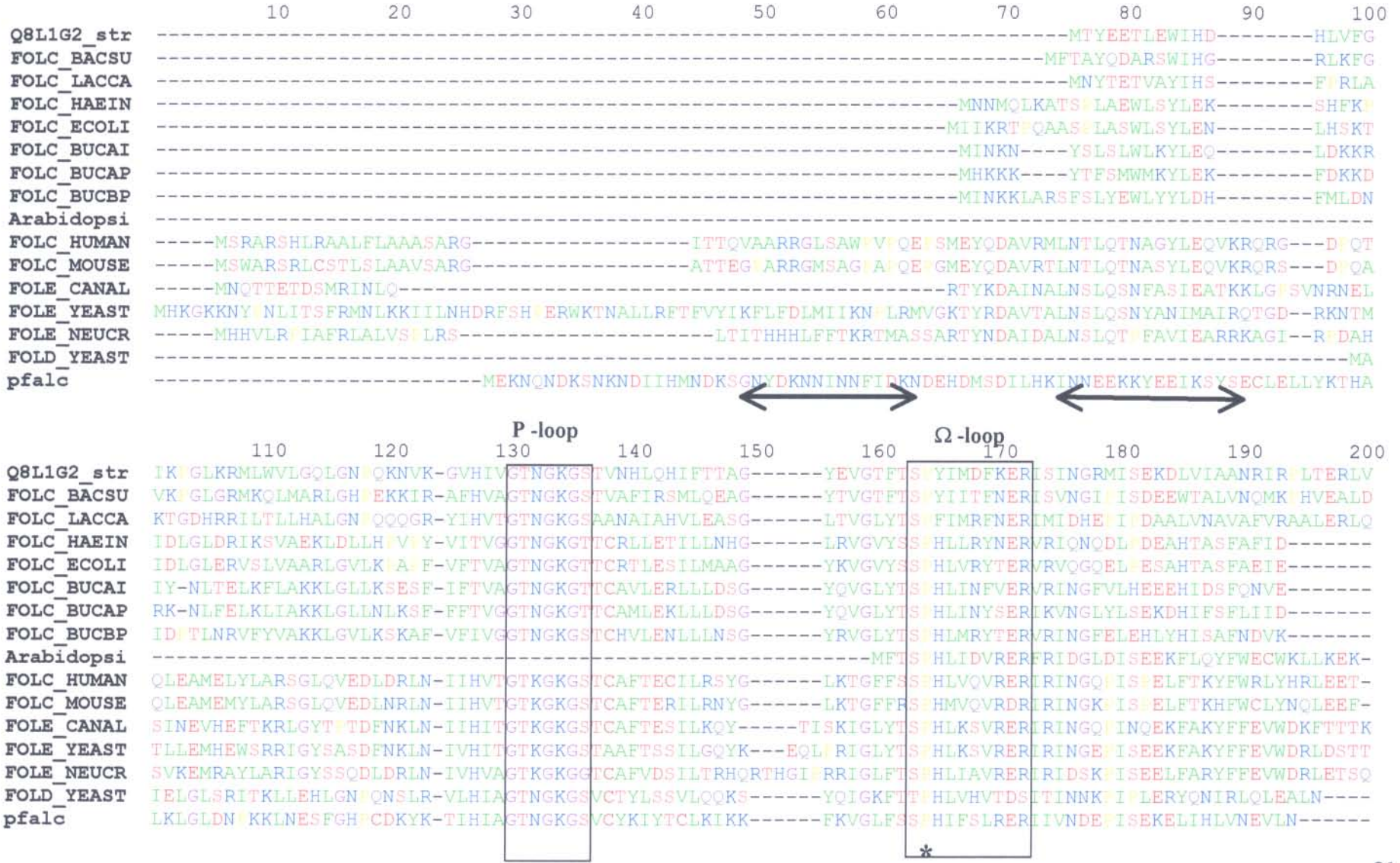


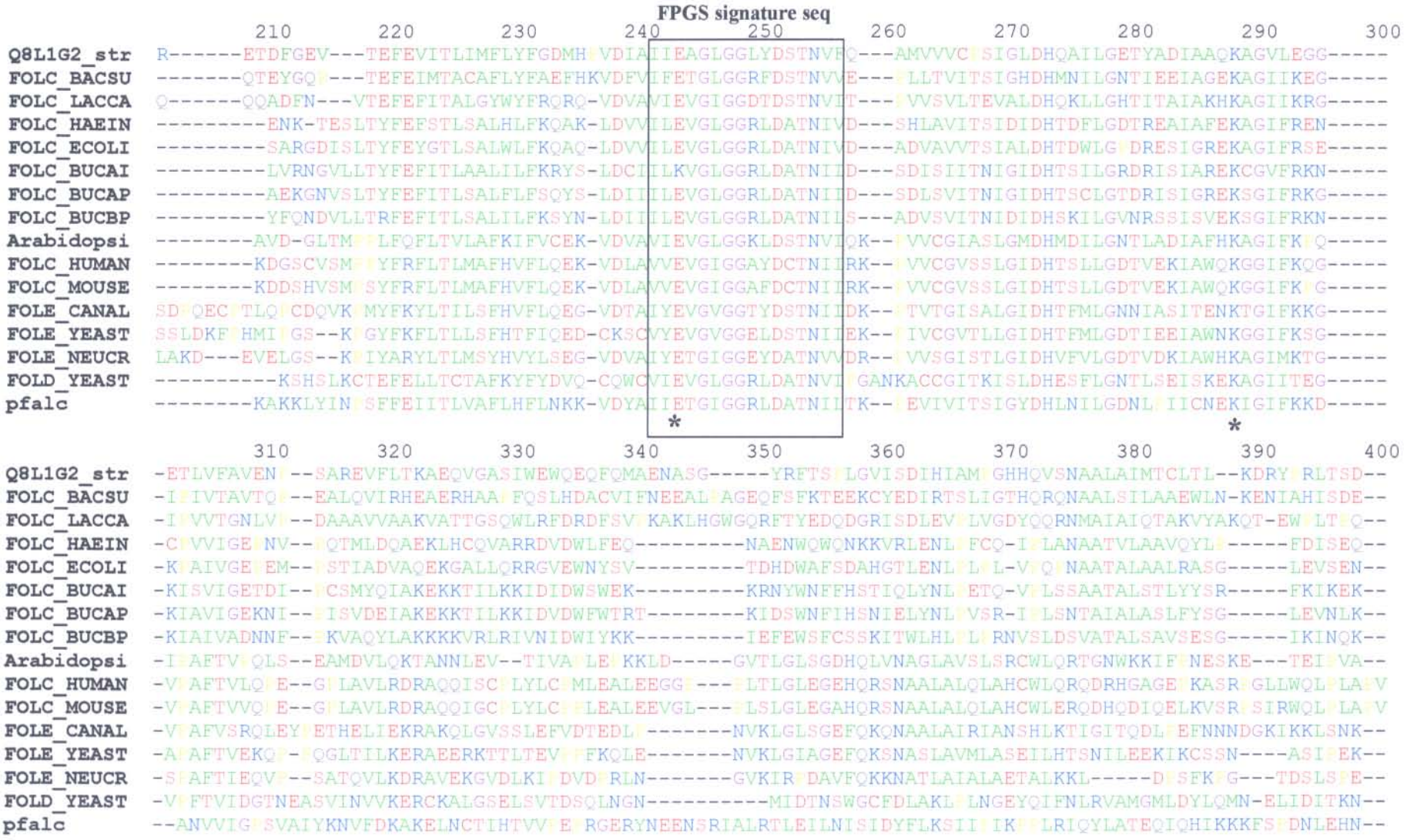
Figure 4.3: Phylogenetic analysis of the *P. falciparum* DHFS-FPGS protein based on its alignment with homologous proteins. The scale line indicates the evolutionary distance in arbitrary units to compare relative distances between proteins.

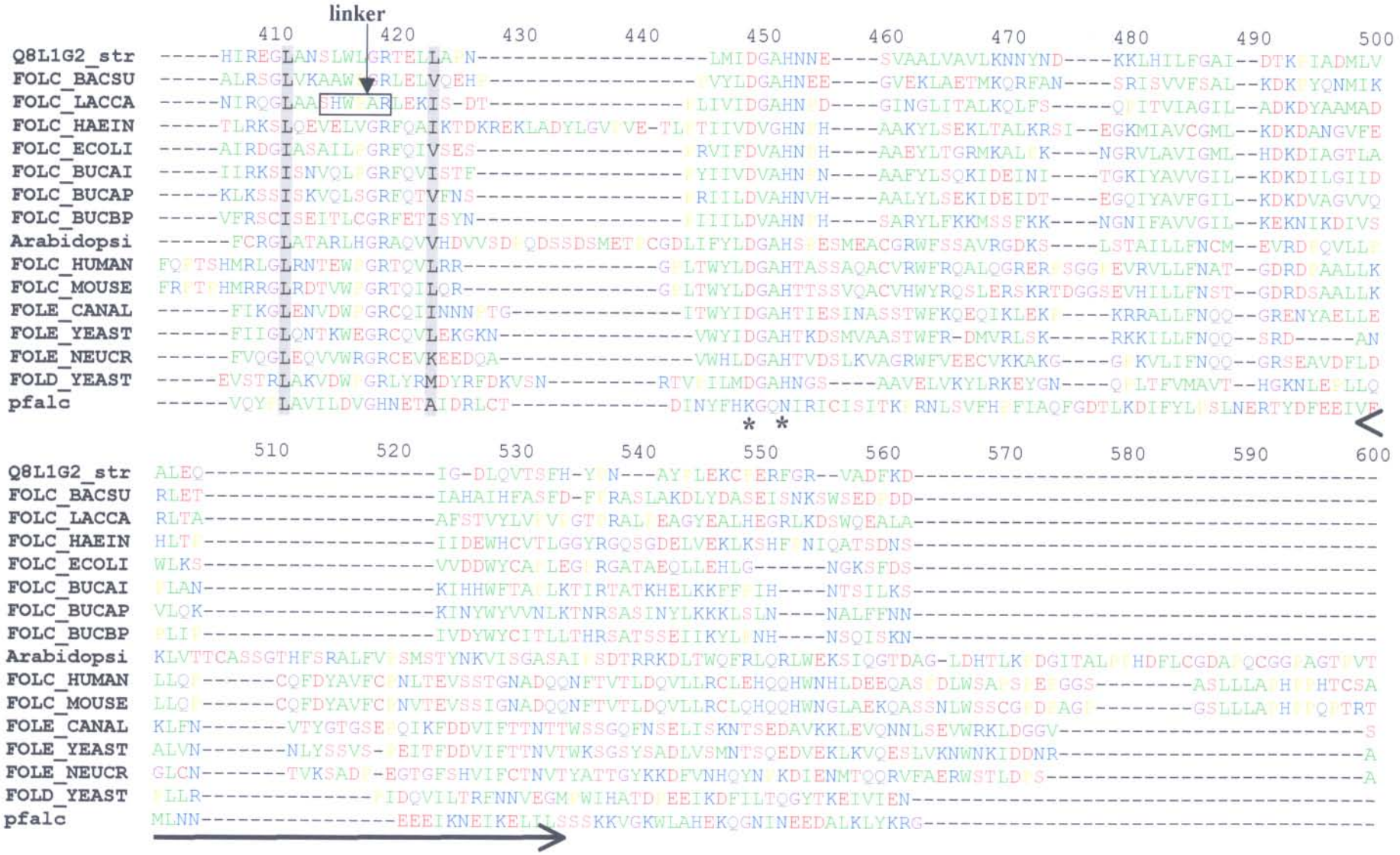
From the alignment, certain conserved features were verified within all the homologues, such as the ATP-binding P-loop "GTNGKGX" (Figure 4.4, 130-138), interdomain stabilising Ω-loop (Figure 4.4, 163-173) and FPGS signature sequence (Figure 4.4, 241-256). Interestingly, the FPGS signature sequence is highly conserved, even for monofunctional DHFS enzymes, such as that of *Streptococcus* and *A. thaliana.*

Overall, the N-terminal half of the sequence seems to be much more conserved between species than the C-terminal half as is verified by a superimposed homology model of *P. falciparum* DHFS-FPGS (red ribbons) on the *L. casei* FPGS crystal structure (green ribbons) (Figure 4.5). The yellow loop indicates the *P. falciparum* sequence that corresponds with the linker area in the alignment. N- and C-terminal halves are assigned according to the position of the *L. casei* linker hexapeptide region (Figure 4.4, 414-419) between the N-and C-terminal domains as determined by X-ray crystallography (Sun *et al.*, 1998).

The amino acid composition of the linker region itself appears not to be conserved, but two conserved hydrophobic residues are indicated by grey highlights on opposite sides of the linker region and the carbon atom backbone structure is conserved (Figure 4.5, yellow loop). The first 40 amino acids of the *P. falciparum* sequence does not align (Figure 4.4; Figure 4.5, red unstructured loop projecting from the N-terminal domain) with the homologous sequences and 20 of these amino acids also coincide with a low complexity region (Figure 4.4, black line at position 48-65). Stars throughout the alignment indicate other essential residues that cause a loss in function when mutated in *L. casei* FPGS (Figure 4.4) (Sheng *et al.*, 2000). These are well conserved between the different homologues and are also present in *P. falciparum,* except for D449 and H552 in the C-terminal region (Figure 4.4; Figure 4.5 yellow and blue ball and stick structures). The Ramachandran plot for the second homology model of ten (Figure 4.6) shows that only 4 residues (0.8% of the total), fall within disallowed areas.


10 20 30 40 50 60 70 80 90 100
Q8L1G2_str
FOLC_BACSU
FOLC_LACCA
FOLC_HAEIN
FOLC_ECOLI
FOLC_BUCAI
FOLC_BUCAP
FOLC_BUCBP
Arabidopsi
FOLC_HUMAN
FOLC_MOUSE
FOLE_CANAL
FOLE_YEAST
FOLE_NEUCR
FOLD_YEAST
pfalc
P -loop
Ω -loop
110 120 130 140 150 160 170 180 190 200
Q8L1G2_str
FOLC_BACSU
FOLC_LACCA
FOLC_HAEIN
FOLC_ECOLI
FOLC_BUCAI
FOLC_BUCAP
FOLC_BUCBP
Arabidopsi
FOLC_HUMAN
FOLC_MOUSE
FOLE_CANAL
FOLE_YEAST
FOLE_NEUCR
FOLD_YEAST
pfalc
*


FPGS signature seq
210 220 230 240 250 260 270 280 290 300
Q8L1G2_str R-------ETDFGEV---TEFEVITLIMFLYFGDMH VDIAIIEAGLGGLYDSTNVF Q---AMVVVC SIGLDHQAILGETYADIAAQKAGVLEGG-----
FOLC_BACSU -------QTEYGQ ---TEFEIMTACAFLYFAEFHKVDFV IFETGLGGRFDSTNVVE --- LLTVITSIGHDHMNILGNTIEEIAGEKAGIIKEG-----
FOLC_LACCA Q-------QQADFN---VTEFEFITALGYWYFRQRQ-VDVA VIEVGIGGDTDSTNVIT --- VVSVLTEVALDHQKLLGHTITAIAKHKAGIIKRG-----
FOLC_HAEIN ---------ENK-TESLTYFEFSTLSALHLFKQAK-LDVV ILEVGLGGRLDATNIVD ---SHLAVITSIDIDHTDFLGDTREAIAFEKAGIFREN-----
FOLC_ECOLI ---------SARGDISLTYFEYGTLSALWLFKQAQ-LDVV ILEVGLGGRLDATNIVD ---ADVAVVTSIALDHTDWLG DRESIGREKAGIFRSE-----
FOLC_BUCAI ---------LVRNGVLLTYFEFITLAALILFKRYS-LDCI ILKVGLGGRLDATNIID ---SDISIITNIGIDHTSILGRDRISIAREKCGVFRKN-----
FOLC_BUCAP ---------AEKGNVSLTYFEFITLSALFLFSQYS-LDII ILEVGLGGRLDATNIID ---SDLSVITNIGIDHTSCLGTDRISIGREKSGIFRKG-----
FOLC_BUCBP ---------YFQNDVLLTRFEFITLSALILFKSYN-LDII ILEVGLGGRLDATNILS ---ADVSVITNIDIDHSKILGVNRSSISVEKSGIFRKN-----
Arabidopsi --------AVD-GLTM LFQFLTVLAFKIFVCEK-VDVA VIEVGLGGKLDSTNVIQK-- VVCGIASLGMDHMDILGNTLADIAFHKAGIFK Q-----
FOLC_HUMAN --------KDGSCVSM YFRFLTLMAFHVFLQEK-VDLA VVEVGIGGAYDCTNIIRK-- VVCGVSSLGIDHTSLLGDTVEKIAWQKGGIFKQG-----
FOLC_MOUSE --------KDDSHVSM SYFRFLTLMAFHVFLQEK-VDLA VVEVGIGGAFDCTNIIRK-- VVCGVSSLGIDHTSLLGDTVEKIAWQKGGIFK G-----
FOLE_CANAL SD QEC TLQ CDQVK MYFKYLTILSFHVFLQEG-VDTA IYEVGVGGTYDSTNIIDK-- TVTGISALGIDHTFMLGNNIASITENKTGIFKKG-----
FOLE_YEAST SSLDKF HMI GS--K GYFKFLTLLSFHTFIQED-CKSC VYEVGVGGELDSTNIIEK-- IVCGVTLLGIDHTFMLGDTIEEIAWNKGGIFKSG-----
FOLE_NEUCR LAKD---EVELGS--K IYARYLTLMSYHVYLSEG-VDVA IYETGIGGEYDATNVVDR-- VVSGISTLGIDHVFVLGDTVDKIAWHKAGIMKTG-----
FOLD_YEAST ----------KSHSLKCTEFELLTCTAFKYFYDVQ-CQWC VIEVGLGGRLDATNVI GANKACCGITKISLDHESFLGNTLSEISKEKAGIITEG-----
pfalc --------KAKKLYIN SFFEIITLVAFLHFLNKK-VDYA IIETGIGGRLDATNIITK-- EVIVITSIGYDHLNILGDNL IICNEKIGIFKKD-----
310 320 330 340 350 360 370 380 390 400
Q8L1G2_str -ETLVFAVEN --SAREVFLTKAEQVGASIWEWQEQFQMAENASG-----YRFTS LGVISDIHIAM GHHQVSNAALAIMTCLTL--KDRY RLTSD--
FOLC_BACSU -I IVTAVTQ --EALQVIRHEAERHAA FQSLHDACVIFNEEAL AGEQFSFKTEEKCYEDIRTSLIGTHQRQNAALSILAAEWLN-KENIAHISDE--
FOLC_LACCA -I VVTGNLV --DAAAVVAAKVATTGSQWLRFDRDFSV KAKLHGWGQRFTYEDQDGRISDLEV LVGDYQQRNMAIAIQTAKVYAKQT-EW LT Q--
FOLC_HAEIN -C VVIGE NV-- QTMLDQAEKLHCQVARRDVDWLFEQ---------NAENWQWQNKKVRLENL FCQ-I LANAATVLAAVQYL -----FDISEQ--
FOLC_ECOLI -K AIVGE EM-- STIADVAQEKGALLQRRGVEWNYSV---------TDHDWAFSDAHGTLENL L L-V Q NAATALAALRASG-----LEVSEN--
FOLC_BUCAI -KISVIGETDI-- CSMYQIAKEKKTILKKIDIDWSWEK---------KRNYWNFFHSTIQLYNL ETQ-V LSSAATALSTLYYSR-----FKIKEK--
FOLC_BUCAP -KIAVIGEKNI-- ISVDEIAKEKKTILKKIDVDWFWTRT--------KIDSWNFIHSNIELYNL VSR-I LSNTAIALASLFYSG-----LEVNLK--
FOLC_BUCBP -KIAIVADNNF-- KVAQYLAKKKKVRLRIVNIDWIYKK---------IEFEWSFCSSKITWLHL L RNVSLDSVATALSAVSESG-----IKINQK--
Arabidopsi -I AFTV QLS--EAMDVLQKTANNLEV--TIVA LE KKLD-----GVTLGLSGDHQLVNAGLAVSLSRCWLQRTGNWKKIF NESKE---TEI VA--
FOLC_HUMAN -V AFTVLQ E--G LAVLRDRAQQISC LYLC MLEALEEGG --- LTLGLEGEHQRSNAALALQLAHCWLQRQDRHGAGE KASR GLLWQL LA V
FOLC_MOUSE -V AFTVVQ E--G LAVLRDRAQQIGC LYLC LEALEEVGL--- LSLGLEGAHQRSNAALALQLAHCWLERQDHQDIQELKVSR SIRWQL LA V
FOLE_CANAL -V AFVSRQLEY ETHELIEKRAKQLGVSSLEFVDTEDL -------NVKLGLSGEFQKQNAALAIRIANSHLKTIGITQDL EFNNNDGKIKKLSNK--
FOLE_YEAST -A AFTVEKQ - QGLTILKERAEERKTTLTEV FKQLE-------NVKLGIAGEFQKSNASLAVMLASEILHTSNILEEKIKCSSN----ASI EK--
FOLE_NEUCR -S AFTIEQV --SATQVLKDRAVEKGVDLKI DVD RLN-------GVKIR DAVFQKKNATLAIALAETALKKL-----D SFK G---TDSLS E--
FOLD_YEAST -V FTVIDGTNEASVINVVKERCKALGSELSVTDSQLNGN----------MIDTNSWGCFDLAKL LNGEYQIFNLRVAMGMLDYLQMN-ELIDITKN--
pfalc --ANVVIG SVAIYKNVFDKAKELNCTIHTVV E RGERYNEENSRIALRTLEILNISIDYFLKSII IK LRIQYLATEQIQHIKKKFS DNLEHN--


linker
410 420 430 440 450 460 470 480 490 500
Q8L1G2_str
FOLC_BACSU
FOLC_LACCA
FOLC_HAEIN
FOLC_ECOLI
FOLC_BUCAI
FOLC_BUCAP
FOLC_BUCBP
Arabidopsi
FOLC_HUMAN
FOLC_MOUSE
FOLE_CANAL
FOLE_YEAST
FOLE_NEUCR
FOLD_YEAST
pfalc
* *
510 520 530 540 550 560 570 580 590 600
Q8L1G2_str
FOLC_BACSU
FOLC_LACCA
FOLC_HAEIN
FOLC_ECOLI
FOLC_BUCAI
FOLC_BUCAP
FOLC_BUCBP
Arabidopsi
FOLC_HUMAN
FOLC_MOUSE
FOLE_CANAL
FOLE_YEAST
FOLE_NEUCR
FOLD_YEAST
pfalc

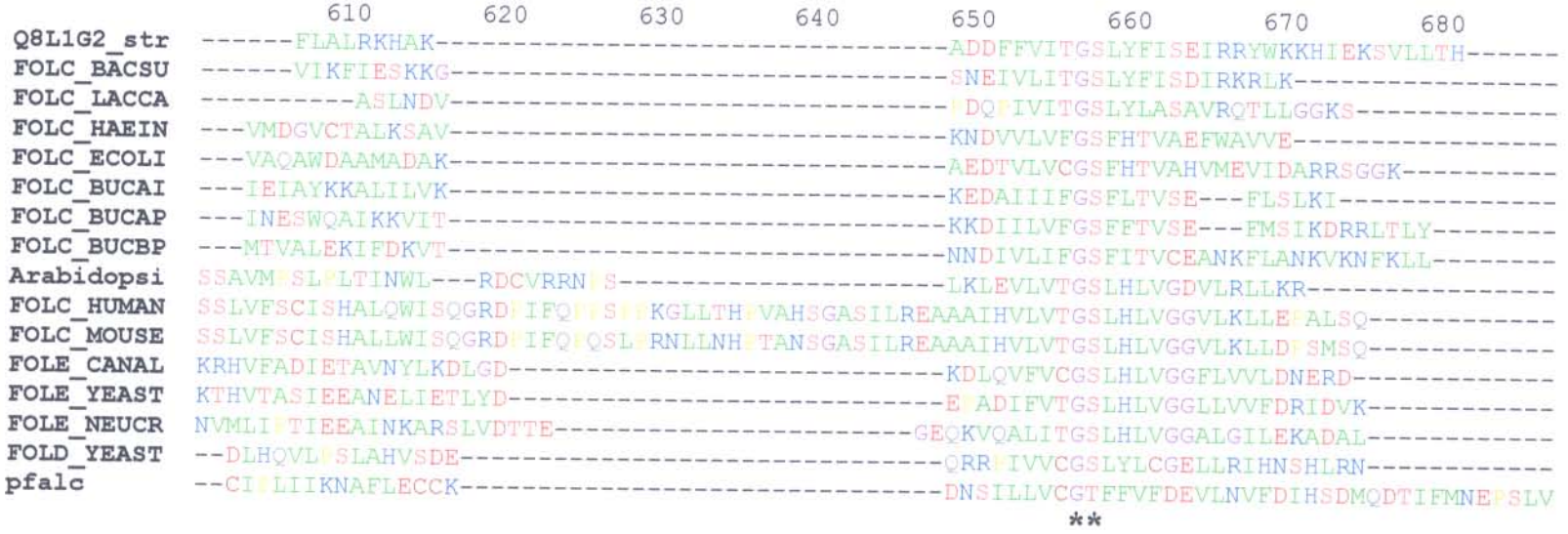


Figure 4.4: Alignment of *P. falciparum* DHFS-FPGS with homologous proteins. Q8L1G2_str: *Streptococcus* DHFS, FOLC_BACSU: *B. subtillus* DHFS-FPGS, FOLC_LACCA: *L. casei* FPGS, FOLC_HAEIN: *H. influenzeae* FPGS, FOLC_ ECOLI: *E. coli* DHFS-FPGS, FOLC_BUCAP: *B. aphidicola* DHFS-FPGS, FOLC_BUCBP: *B. biphidicola* DHFS-FPGS, Arabidopsi: *A. thaliana* FPGS, FOLC_HUMAN: human FPGS mitochondrial precursor, FOLC_MOUSE: mouse FPGS mitochondrial precursor, FOLE_CANAL: *C. albicans* FPGS, FOLE_YEAST: yeast FPGS, FOLE_NEUCR: *N. crassa* FPGS, FOLD_YEAST: yeast DHFS. p.falc: *P. falciparum* DHFS-FPGS. Thick black lines indicate areas of low complexity in the *P. falciparum* sequence. Stars indicate essential residues in the *L. casei* enzyme. Boxed areas indicate the P-loop, Ω-loop, FPGS signature sequence and *L. casei* linker features. The grey highlighted residues indicate conserved hydrophobic residues flanking the linker region of *L. casei.* Neutral amino acids are coloured green, negatively charged amino acids are coloured different shades of red according to pKa values, positively charged amino acids are coloured shades of blue according to pKa values, proline and glycine normally found in turns are coloured yellow and purple respectively.

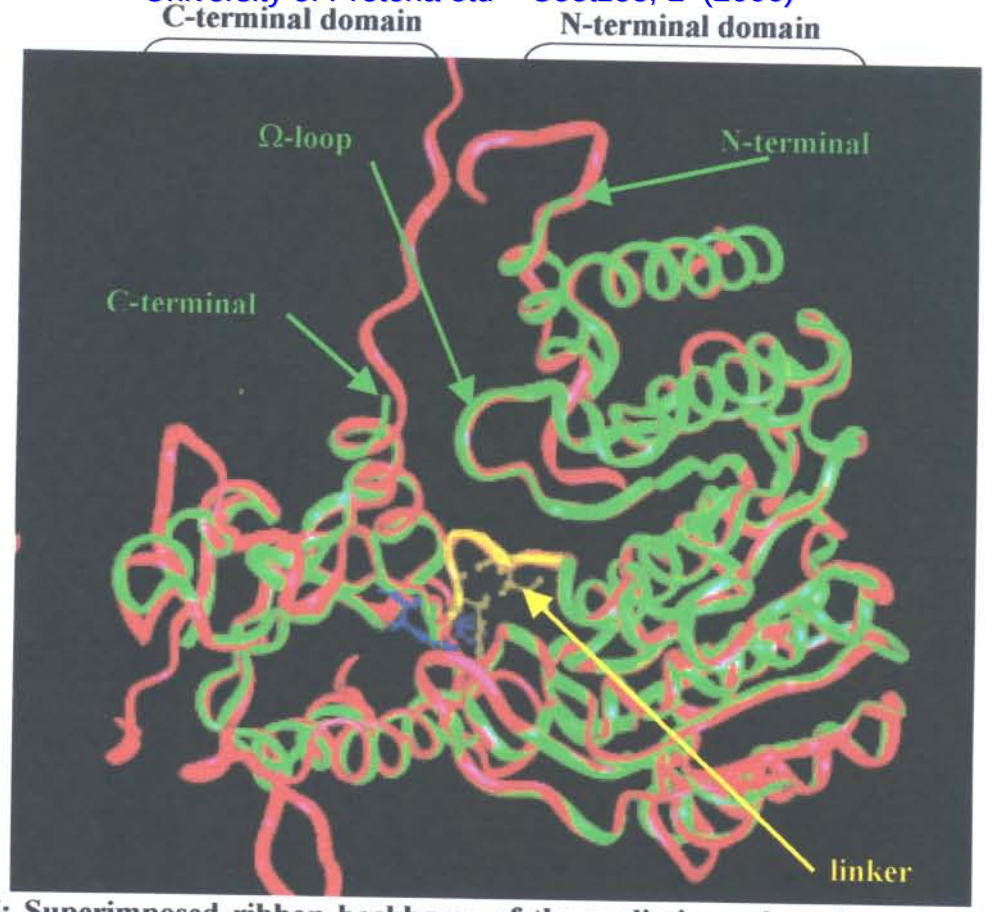


Figure 4.5: Superimposed ribbon backbones of the preliminary homology model of *P. falciparum* DHFS-FPGS (red) on the *L. casei* FPGS (green) crystal structure. The aligned Ω-loop, P-loop and linker area (yellow loop) are indicated with green arrows. Catalytic residues (D449 and H552) that do not align are indicated in blue ball and stick structures (*L. casei*) and yellow ball and stick structures (*P. falciparum*).

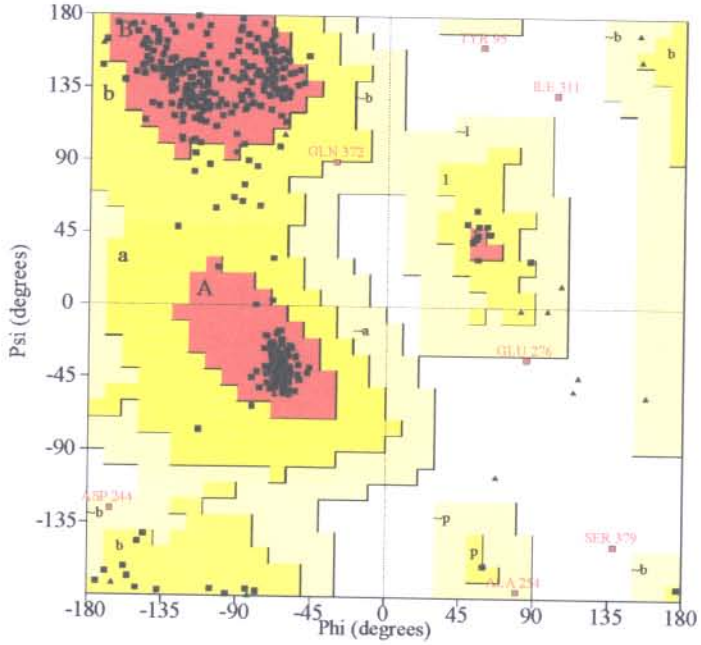


Figure 4.6: Ramachandran plot of the *P. falciparum* DHFS-FPGS homology model. Residues in most favoured regions [red: A, B, L], residues in additional allowed regions [yellow: a, b, l, p], residues in generously allowed regions [beige: ~a,~b,~l,~p] and residues in disallowed regions [white: Y95, E276, I311 and S379].

4.3.2 Conservation of DHFS-FPGS within the *Plasmodium* species and comparison with human FPGS.

The TBLASTN search of *Plasmo*DB using *P. falciparum* DHFS-FPGS as the query sequence, identified high scoring homologous gene sequences in *P. yoelii* ($E=4 \times 10^{-173}$) and *P. vivax* ($E=2.7 \times 10^{-130}$). Alignments of *P. falciparum* DHFS-FPGS with these *Plasmodium* sequences and human FPGS show ~10% identity and 24% similarity between the Plasmodium and human species (stars and :, Figure 4.7). The N-terminal extension also observed in Figure 4.5 and the alignment (Figure 4.4) seems to be specific to *P. falciparum*. An interesting feature is that the human enzyme has inserted sequences not found in the *Plasmodium* sequences (Figure 4.7, yellow highlighted sequences). Normally *Plasmodium* enzymes contain inserted sequences when compared to human homologues, which coincide with low complexity areas, (Pizzi and Frontali, 2001). This, however does not seem to be the case for DHFS-FPGS, where the human homologue, FPGS contains the majority of inserted sequences.

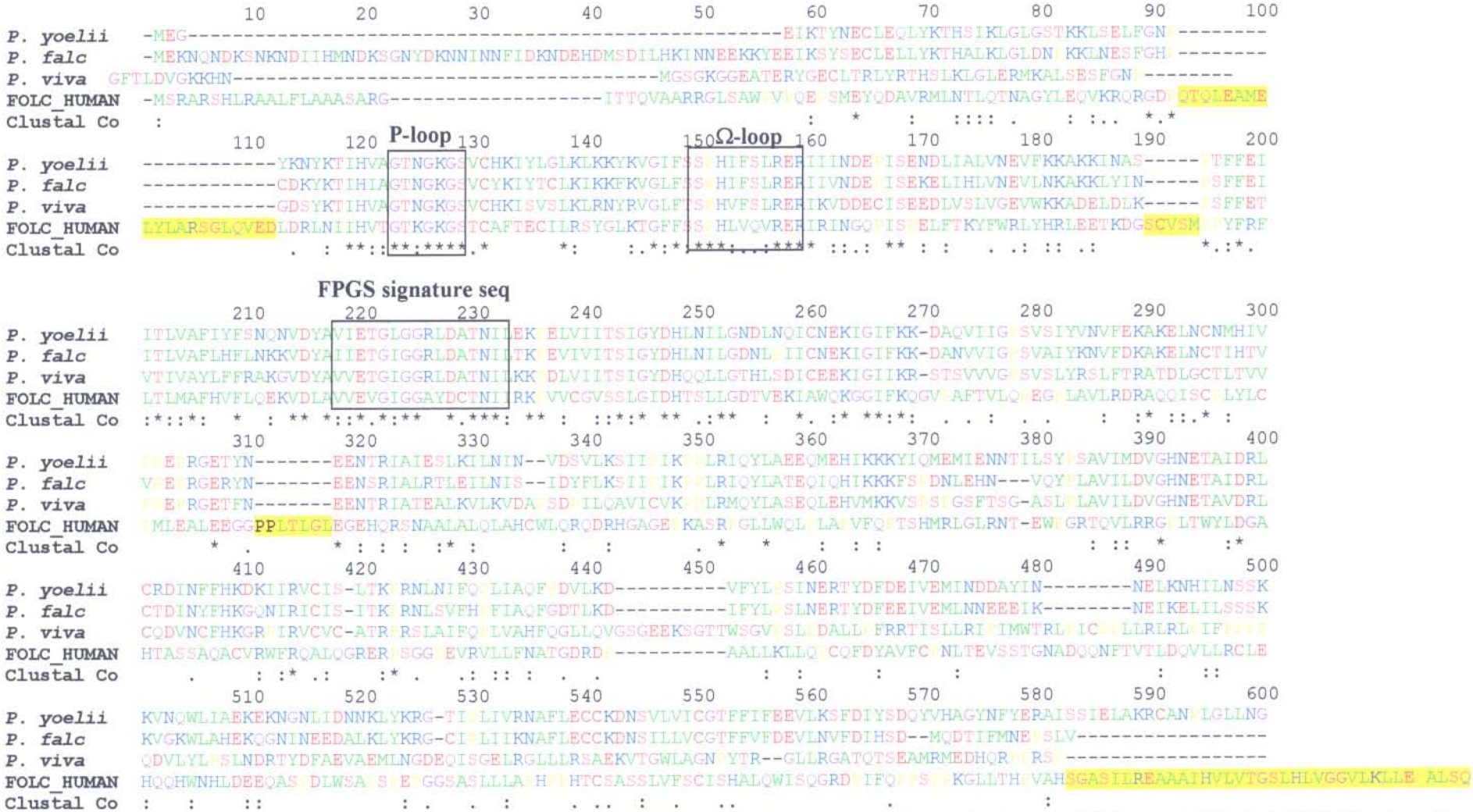


Figure 4.7: *Plasmodium* DHFS-FPGS vs Human FPGS alignment: *P. yoelii* DHFS-FPGS, *P. falc*: *P. falciparum* DHFS-FPGS, P. viva: *P. vivax* DHFS-FPGS, FOLC_HUMAN: human FPGS. Blocks indicate key enzyme features such as the P-loop, Ω-loop and FPGS signature sequence. Clustal Co= conserved sequence; stars indicate conserved residues (identities) and (:) similarities. Human inserted sequences are highlighted in yellow.

4.3.3 Secondary structure prediction of *P. falciparum* DHFS-FPGS

The primary amino acid sequence was analysed with a variety of programs for the prediction of secondary structure elements. According to the GOR4 program, most of the sequence consists of α-helices and random coils. α-Helices are predominantly found at positions 50-70, 150-180, 290-310 and 430-450 (Figure 4.8, blue lines). Random coils are predominant at positions 0-40, 280-290, 310-320, 380-390 and 410-420 (Figure 4.8, purple lines). The N-terminal extension as identified through alignments is part of a random coil as expected, but the other two low complexity areas coincide with α-helices.

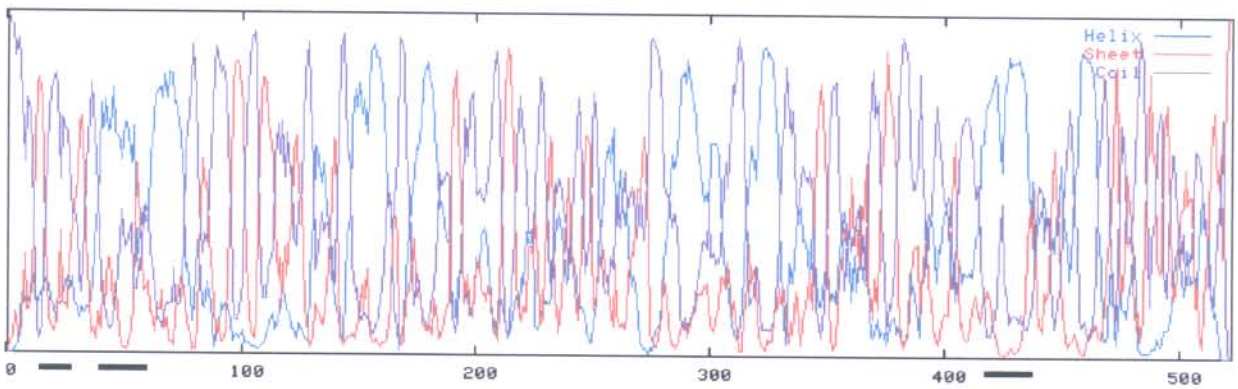


Figure 4.8: Secondary structure prediction of *P. falciparum* DHFS-FPGS with GOR4. A schematic representation indicating the possibility of each structural element (y axis) versus the position in the primary amino acid sequence (x axis). Black lines below the figure indicate the positions of low complexity sequence.

Alignments between the HNN, SCRATCH, GOR4 and PHD secondary structure predictions show mostly the same secondary structures (Figure 4.9). At a closer look it is observed that the two low complexity sequence areas consist of mixed helix-coil structures in contrast to the random coil observed for the N-terminal extension (Figure 4.9, black underlining). Taken on average of the different prediction methods the percentage of secondary structure is predicted to be 46% α-helices, 11.9% extended β sheets and 42.1 % random coils. Given that the requirements for the classification of proteins according to secondary structure are: all-alpha (%Helices>45 and % Extended beta sheets<5), all-beta (%Helices<5 and %Extended beta sheets>45) and alpha-beta (%Helices>30 and %Extended beta sheets > 20), DHFS-FPGS is predicted as a mixed class protein since the predicted secondary structures do not fall in the above categories. No coiled–coils were predicted by the program COIL and the overall topology was predicted to be compact and globular with 199 exposed residues by the program GLOBE.

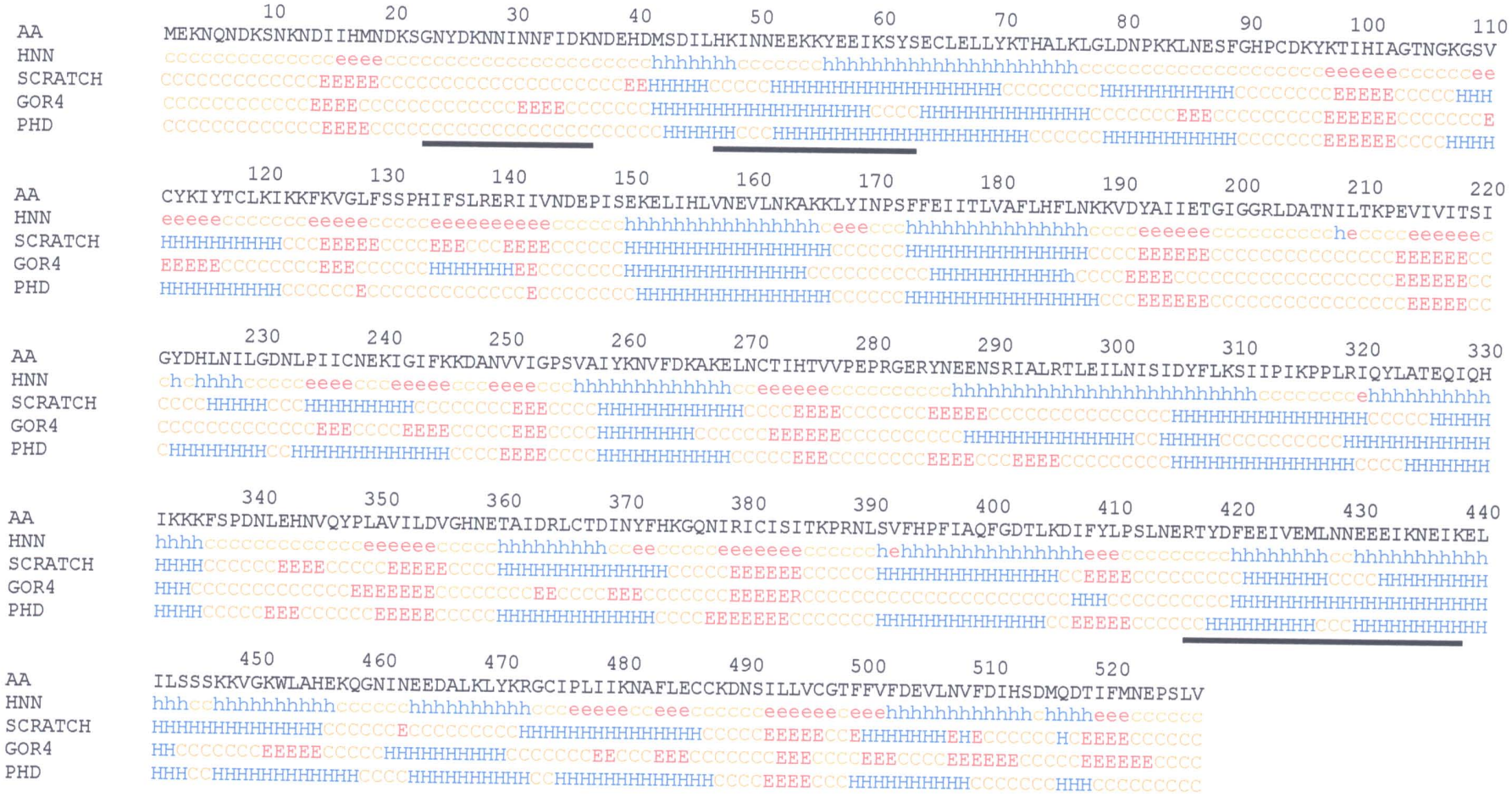


Figure 4.9: Alignment of independent secondary structure predictions of *P. falciparum* DHFS-FPGS from hierarchial neural networks (HNN), SCRATCH Sspro GOR4 and profile fed neural networks (PHD) based on the primary amino acid sequence (AA) indicated in black letters. Orange: random coil (c), red: extended β sheet (e) and blue letters: α-helix (H). The random coil observed in the first 40 amino acids corresponds with the inserted sequence that does not align with other DHFS of FPGS homologues. Black lines below the alignments indicate the positions of low complexity sequence

4.3.4 DHFS-FPGS hydrophobicity profile

The protein is predicted to be neither predominantly hydrophobic nor hydrophilic. The 40 amino acid N-terminal extension corresponds with the most hydrophilic area within the protein (Figure 4.10). Position 280-290, corresponding to a random coil is also very hydrophilic. Hydrophobic areas correspond mostly with isolated β-sheets. It is interesting that the omega loop, of which the primary function is interdomain stabilisation through hydrophobic interactions, does not occur in this plot at one of the hydrophobic maxima. In fact not one of the enzyme features seem to be hydrophobic, except for the conserved hydrophobic residues flanking the *L. casei* linker area (Figure 4.10). The other hydrophobic maxima are distributed between important catalytic sites or conserved motifs, rather than corresponding directly with these features and it may be that hydrophobic interactions between the hydrophobic maxima are required for shaping of the active site pockets or general stabilisation of enzyme structural features.

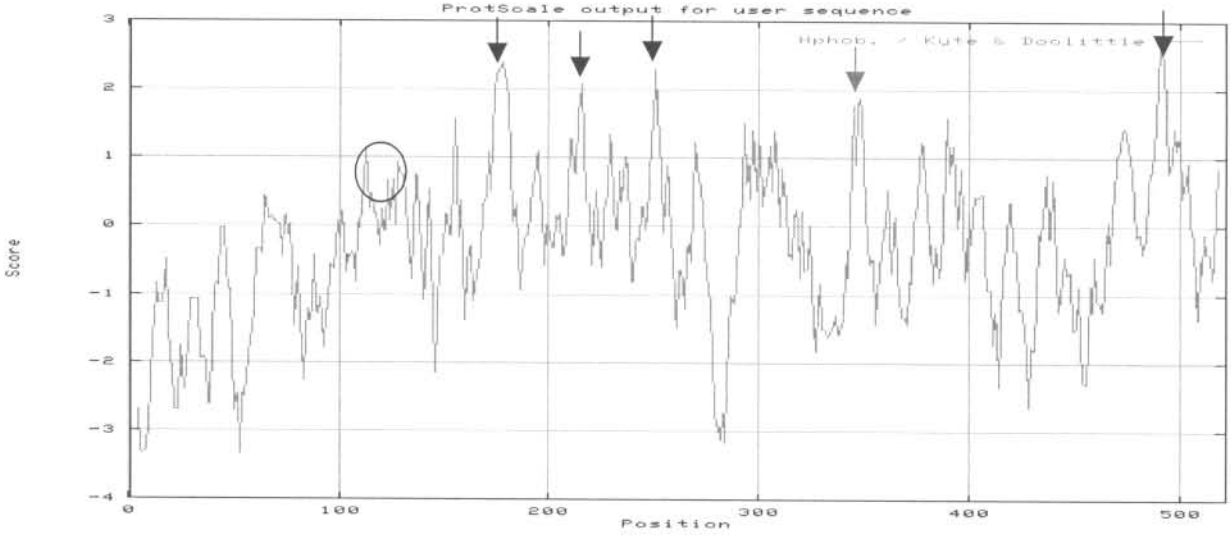


Figure 4.10: Hydrophobicity profile for the primary amino acid sequence of *P. falciparum* DHFS-FPGS using Kyte and Doolittle parameters. Hydrophobic maxima are indicated with arrows. The position of the omega loop is encircled

4.4 Discussion

4.4.1 Sequence conservation of DHFS-FPGS

The alignment of the P-loop, Ω-loop and FPGS signature sequence corresponds to that reported by (Salcedo *et al.*, 2001), yet our alignment did not show the conservation of C-terminal residues, D449 and H 452 (Figure 4.4) involved in FPGS activity as shown for *L. casei* (Sheng et al., 2000). It might be that the *P. falciparum* enzyme uses additional residues for activity, or has a slightly altered conformation of its active site due to its bifunctionality. Alignments of other *Plasmodium* strains also fail to align these two residues with the human homologue, which might indicate species-specific functional differences. Another alternative might be that the D416 and H419 residues of *P. falciparum* DHFS-FPGS should be aligned 33 residues to the right (Salcedo *et al.*, 2001) with the conserved D449 and H452 (Figure 4.4), but this would introduce a gap in the alignment around the linker area, which is an important structural feature and is shown by the homology model that the structure of the linker area remains conserved (Figure 4.5).

All the other catalytic residues align well and although the amino acid composition for the linker area is not conserved for *P. falciparum* DHFS-FPGS, it is structurally conserved in terms of its backbone structure and fold (Figure 4.5). The N-terminal is much more conserved than the C-terminal for all species, which is also shown through the superimposition of the *P. falciparum* homology model with the crystal structure. The *P. falciparum* enzyme, however, contains an N-terminal extension, which consists of a low complexity sequence and is a parasite-specific trait. Alignments with other *Plasmodium* strains indicate that this feature is unique to *P. falciparum* and might be implicated in the pathogenicity of the strain.

The phylogenetic tree deduced from the alignment shows as expected separate pro- and eukaryotic groups. This also confirms that the alignment is a good representation of the homologues. Interestingly, the *P. falciparum* sequence is closer related to the prokaryotic bifunctional enzymes than the eukaryotic enzymes. This is difficult to evaluate because no other bifunctional eukaryotic sequences are known at this stage and it may be that the bifunctionality forces the grouping of the *P. falciparum* enzyme towards other bifunctional enzymes. Alignment with the human homologue indicate low sequence similarity (24%) and

identity (10%) comparable to previous literature reports of 30% similarity and 15% identity (Lee *et al.*, 2001).

4.4.2 Predicted secondary structure

The overall topology of *P. falciparum* DHFS-FPGS is predicted to be globular, with almost half the residues exposed. The protein is classified as mixed class since there are almost equal numbers of α-helices and random coils predicted (~45%) and only a few extended β-sheets (~10%). The characteristic N-terminal extension coincides with a random coil as expected for low complexity sequences (Pizzi and Frontali, 2001), yet the other low complexity sequences correspond to mixed helix-random coil structures. The N-terminal extension furthermore coincides with the most hydrophilic portion of the enzyme as predicted by the Kyte and Doolittle hydrophobicity plot. Taken together with all the other information on this parasite-specific feature, the N-terminal extension as well as the divergent C-terminal domain might be interesting features for structural investigation.