Chapter 6

Introduction

The Elongation Factor Tu protein is encoded by one of the best conserved genes present in the apicoplast genome as other proteins such as SufB and RpoC. As discussed earlier also, these proteins are required in the transcriptional and translational machinery and in the [Fe – S] cluster mobilization by the organelle. Based upon the differences observed in the *P. vivax* sequences, structural elucidation of these proteins was required to identify any changes produced in the active sites of the protein molecule due to amino acid mutations. Thus, a study was carried out to define the structure of these proteins and to identify their differences, if any, in comparison with the established structures of homologous proteins.

Experimental Approach to Conformation Determination

There are two prominent experimental techniques to determine molecular conformation. These are X – ray crystallography and NMR spectroscopy. Among the two, X – ray crystallography involves crystallization of the complete protein molecule at highly pure level so as to study the crystal dimensions using X – Ray diffraction technique. Due to the high specificity provided by the technique, it becomes the most useful for any structural conformation. The NMR spectroscopy is useful for only small peptides, as it provides a spectrum that gets complicated to read as the polypeptide size increases. Though very useful, both the techniques are time consuming, difficult, expensive and require high level of purity. To counteract this problem, theoretical techniques of protein structure modeling have been developed. These techniques make use of available X – Ray crystallographic structures and try to predict the structure of proteins that are closely related with these crystal structures. This is based upon the fact that approximately one third of all sequences existing in the present database are recognizably related (by homology and similarity) to at least one known crystal structure.

Homology Modeling

Homology or comparative modeling is a technique to predict the three – dimensional structure of a given protein sequence (Target) based on an alignment to one or more known protein structures (Templates). This is based on a fact that structural and functional similarity can be assumed if the sequence identity between the Target and the Template is 25 – 30% provided the evolutionary distance is not large (Abagyan and Batalov, 1997; Chothia and Lesk, 1986). This assumption is made on a fact that during evolution protein folds vary much less than the amino acid sequences (due to insertions or mutations) and may therefore be retained even when the sequence similarity is relatively low (Doolittle, 1981; Chothia and Lesk, 1982 a,b; Sander and Schneider, 1991; Pastore and Lesk, 1991). If the proteins have same folds, turns or loops they may have close biological functions. Although theoretical modeling encounters limitations and cannot substitute for a high – resolution structure determination, helpful information may be gained about the overall fold and the protein surface properties, justifying its increasing importance as a powerful tool in structural biology.

Tools for Homology Modeling

A number of softwares for basic level homology modeling are freely available online, such as online Swiss – Model (Glaxo Wellcome Research) or downloadable (Deep View – SPDBV and Modeller 8v). For advanced level modeling, that includes threading or ab – initio predictions, some commercial softwares can be used, such as INSIGHT II, Accelrys, etc. For our study we used the basic level tools only.

DeepView – the Swiss-PdbViewer (or **SPDBV**), is an interactive molecular graphics program for viewing and analyzing protein structures. In combination with Swiss-Model (a server for automated comparative protein modeling maintained at http://www.expasy. org/swissmod) new protein structures can also be modeled. **MODELLER** is also used for homology or comparative modeling of protein three-dimensional structures (Sali, 1993; Sali and Sanchez, 1997). The user provides an alignment of a sequence to be modeled with known related structures and Modeller

automatically calculates a model containing all non-hydrogen atoms. Modeller implements comparative protein structure modeling by satisfaction of spatial restraints, and can perform many additional tasks, including de novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, comparison of protein structures, etc.

Results and Discussions

To define the structures using sequences of *P. vivax* EF – Tu and RpoC proteins, the protein sequence was first submitted to search homologues having certain degree of similarity with the desired protein. Among the obtained homologues, the one having a known X- Ray crystallographic structure and showing highest similarity score was chosen. If this score matches or shows similarity from more than one structure, then all may act as the desired Templates. These homologous proteins generally fall in a group of protein family and the database of such protein families was accessed from PROSITE that is available at the SWISS PROT database (http://expasy.hcuge.ch/sprot/prosite.html). Once the homologue was determined, the sequence alignment for the two protein sequences was performed using various software tools such as Clustal W (Thompson *et. al.*, 1994), FASTA (Pearson, 2000), BLAST P (Altschul *et. al.*, 1990), etc. All these are based on certain algorithm programs that make use of matrices such as PAM – Percentage of Acceptable point Mutations (Dayhoff *et. al.*, 1968; Dayhoff *et. al.*, 1978; Dayhoff *et. al.*, 1983) and BLOSUM – Blocks substitution matrices (Henikoff and Henikoff, 1992).

Once the alignment searches give a desired sequence and structure of known proteins, first these known protein sequences and structures were aligned. This forms the base for the alignment of unknown protein. After the known structures are aligned, they are examined to identify the structurally conserved regions (SCRs) from which an average structure, or framework, can be constructed for these regions of the proteins. Variable regions (VRs), in which each of the known structures may differ in conformation, are also identified because special techniques are required to model these regions of the unknown protein.

Elongation Factor Tu

The 3 – Dimensional structure for protein EF – Tu was defined using Swiss Model. The protein sequence was first submitted to the Swiss Model database to search for a suitable template sequence. This search gave the list of all X – Ray crystallographic structures available in the database. From this list, the Template structure (Swiss Prot Id. 1b23 p chain of *Thermus aquaticus*) showing high resolution score was selected and downloaded. Following this, the sequence of the *P. vivax* EF – Tu protein (target) and template were submitted to Swiss Model First Approach Mode. The software automatically developed a model based upon the templates which was viewed in Deep View – Swiss PDB Viewer and further analysed.

A 3 – Dimensional model of *P. falciparum* EF – Tu is already reported (Sato and Wilson, 2000) in the PROSITE database (Q25820). As discussed earlier also, the *P. falciparum* EF – Tu structure has been defined into three functional Domains. Domain I carries active sites for binding of phosphoryl (G19HVDHGK25), Mg^{2+} (D83CPG86), tRNA – GTP complex (N138KED141), GDP binding pocket (G24, N138, K139, D141, S176, L178). Domain II and III have sites for antimicrobial agents like amythiamicin and kirromycin respectively (Vogeley *et. al.*, 2001). Our study for *P. vivax* EF – Tu protein model was focused on these three domains.

The obtained *P. vivax* EF – Tu protein model (Figure 6.1) was aligned to *P. falciparum* EF – Tu model in SPDB Viewer. The alignment showed certain variations (Figure 6.2) in the helices and the loops of the structures. There were increase in some α helices and shifts in coils were also visible. Further critical analysis revealed that the variations noted in the sequences of the two proteins (Table 3.3, 3.4) were responsible for these structural variations (Figure 6.3 a – d). The site for tRNA – GTP binding complex was intact. The absence of Proline at 263^{rd} position shortens a loop in the Domain II but does not affect the amythiamicin binding site. A shift in the loop and coil in region of Domain III was also visible but again this variation did not have any visible effect on the kirromycin binding site. Even the amino acid variation in the close vicinity of the active sites of protein molecule (Table 3.5) did not produce any variations at the structural level.

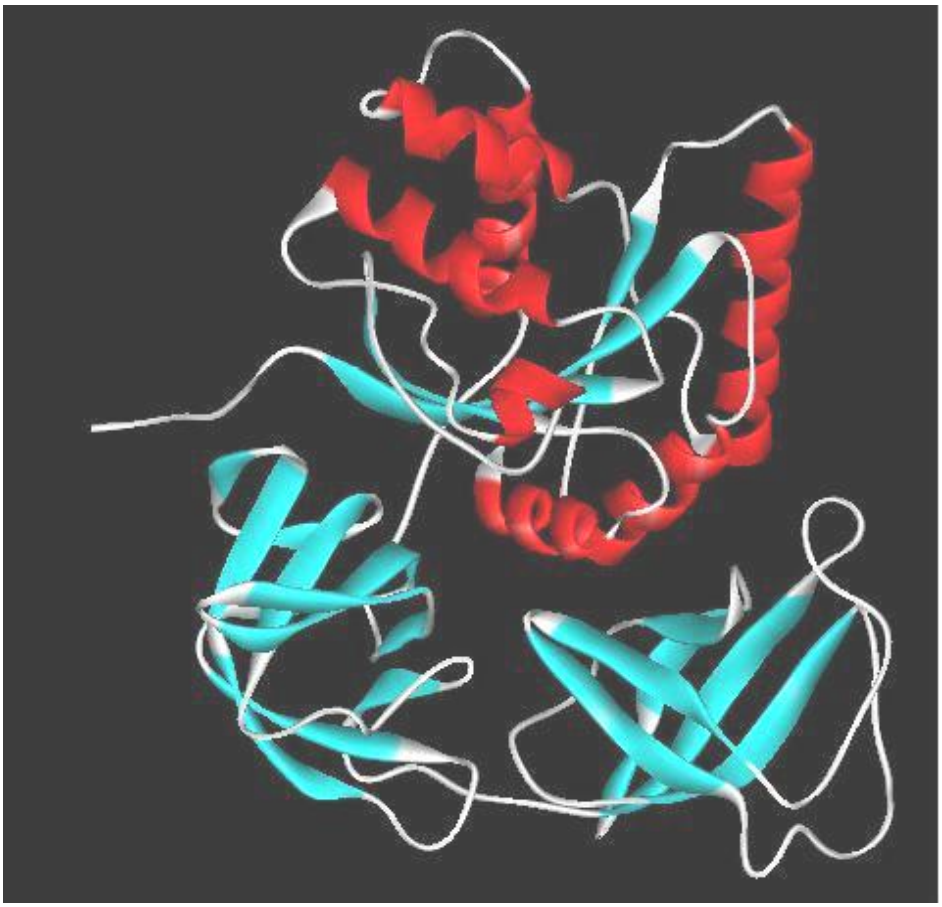

Figure 6.1: *Plasmodium vivax* EF – Tu structure as obtained from SWISS Model. The image is developed using ViewerLite (Accelrys).

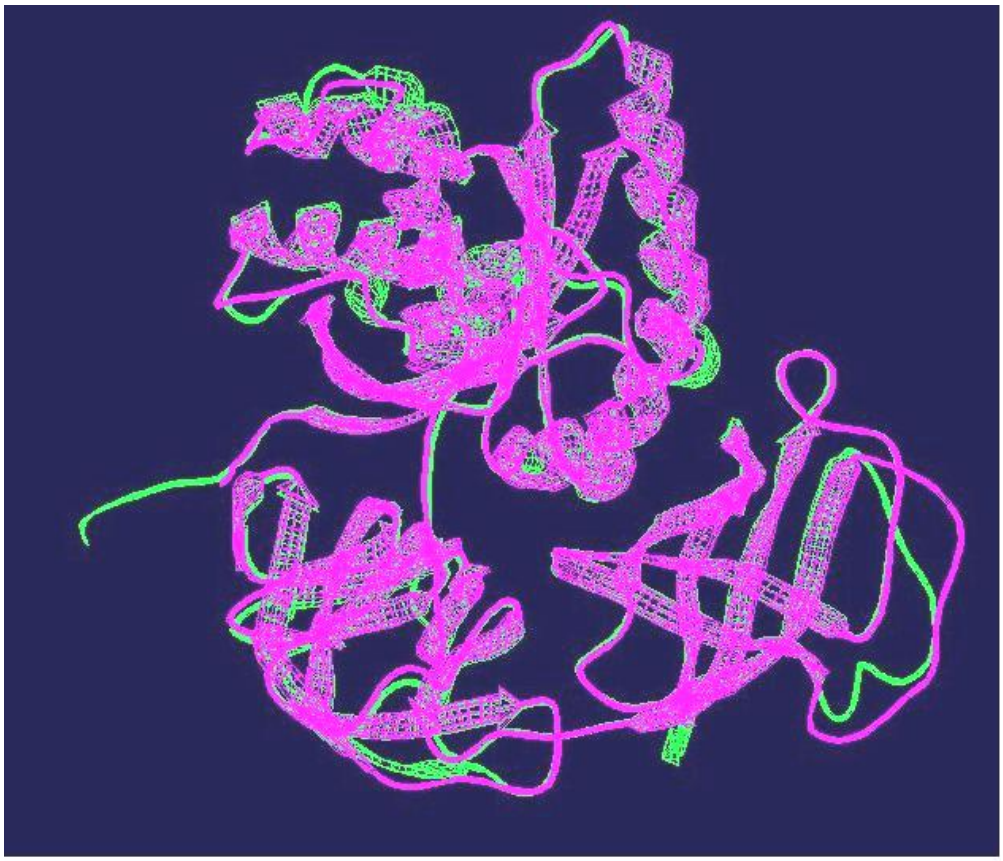

Figure 6.2: Structural Alignment of *P. vivax* (green) and *P. falciparum* (pink) EF – Tu protein models.

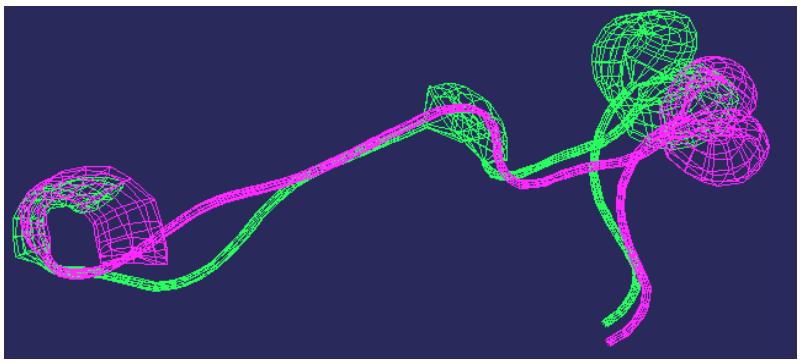

Figure 6.3a: Variation in the conformation of the structure visible due to variations in amino acid at positions L28F, T29S, T30S and A31P.

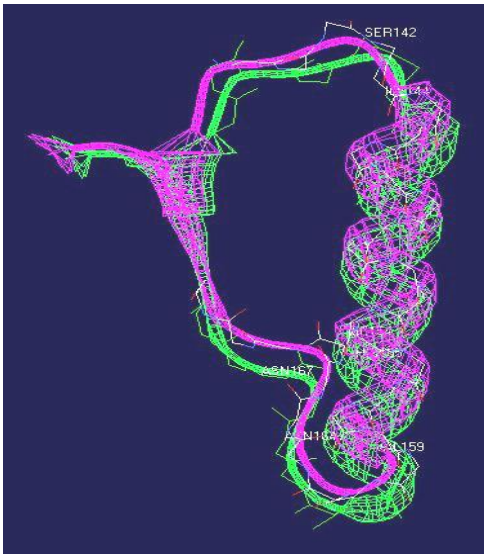

Figure 6.3b: An extended helix is visible in *P. vivax* (green) EF – Tu structure due to amino acid variations:

D142S

V144I

V159I

Y167N

Figure 6.3c: An extended helix and a broadened loop in the *P. vivax* (green) EF – Tu structure due to amino acid variations:

T172A

I179V

K186R

W196G

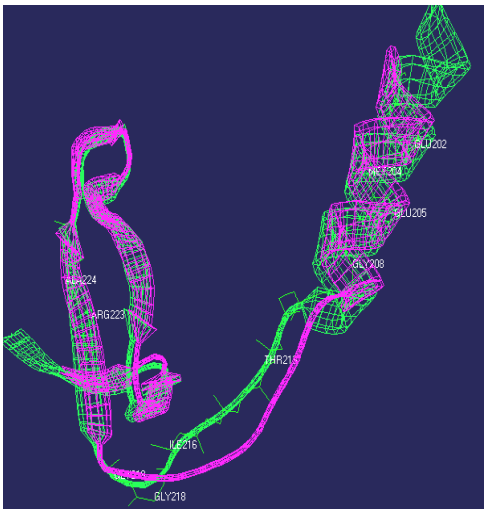

Figure 6.3d: An extended helix and a shift in the *P. vivax* (green) EF – Tu structure due to amino acid variations:

N202E

I204M

Q205E

P213T

L216I

N218G

D219G

S223A

M224R

Though the topological variations among the structures of the two protein molecules could be studied with the help of SPDBV but the variations in the functional properties, which includes an in – depth view in the functionally active / binding sites of the *P. vivax* EF – Tu molecule, due to the amino acid variations, could not be analysed. For this, the model in query was redesigned and studied with the help of Modeller.

Three-dimensional (3D) model of PvEF-Tu was built by comparative protein structure modeling with the program Modeller 6v1 (Sali.and Blundell, 1993; Sa´nchez and Sali, 1997). The closest template structure found with the BLAST search of ExPDB database (http://swissmodel.expasy.org/SM_Blast.html) was of *T. aquaticus* Elongation Factor EF-Tu: GTP Ternary Complex [Protein Data Bank (PDB), p chain of 1b23], with 2.6 A^{o} resolution crystal structure and shared 44% amino acid identity to the target *P. vivax* sequences. The input consisted of the template structure and the alignment of the target sequence with the structures. The output, obtained without any user intervention, was a 3D model of the target with all non-hydrogen atoms (Figure 6.4).

The model was derived by minimizing violations of many distance and dihedral angle restraints extracted from the template structures. The model also included GTP cysteinyl tRNA, Mg^{2+} and SO_4^{2-} ions, inherited from the 1b23 template. The quality and stereochemistry of the model were evaluated using the program PROCHECK (Laskowski *et. al.*, 1993). The main chain conformations for 99.2% of amino acid residues were within the favored or allowed regions of the Ramachandran plot (Figure 6.5) and the overall G factor was -0.14, indicating that the molecular geometry of the model is of good quality. The present model is likely to be right in the overall structure, but may contain local errors since decisions had to be made based on geometric criteria without any experimental information.

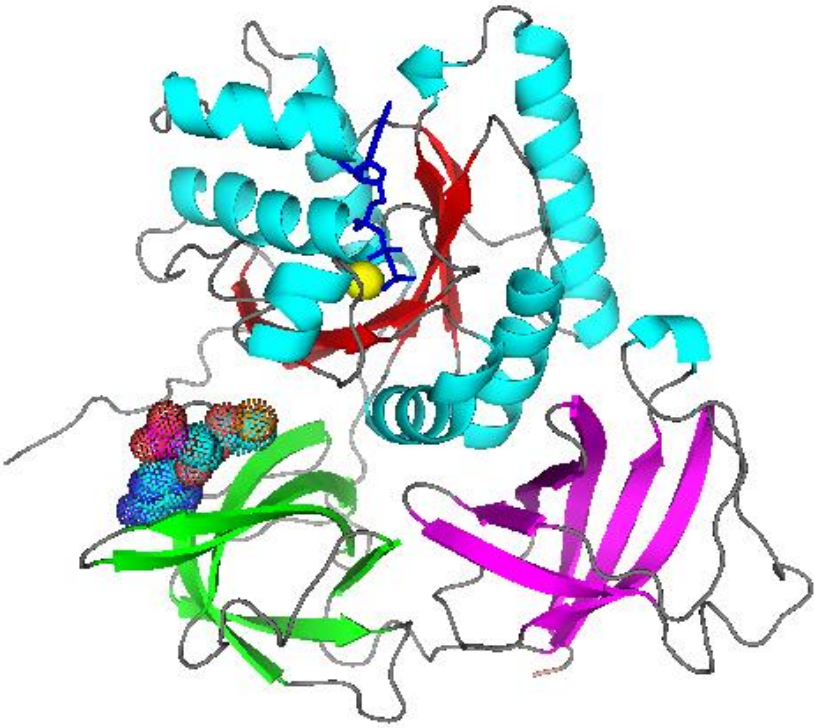

Figure 6.4: Ribbon model of the overall structure of *P. vivax* EF-Tu calculated using Modeller 6. Domain I (blue, red), domain II (lime-green) and domain III (magenta) as in the alignment in Figure 3.10

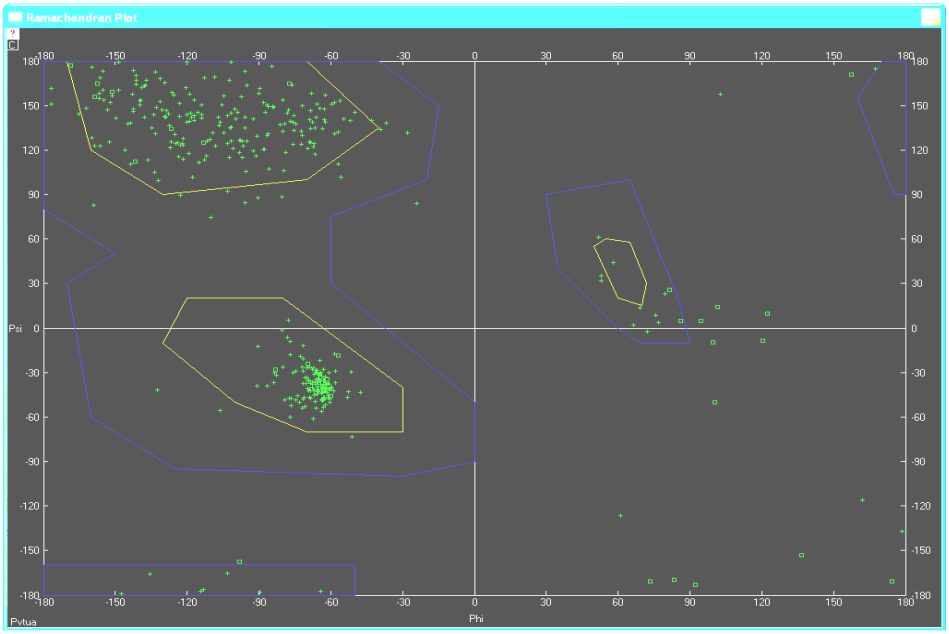

Figure 6.5: Ramachandran plot for *P. vivax* EF – Tu structure

To further analyze topologically, the conformational binding of ligands with the EF – Tu protein molecule, a surface representation view of the modeled structure (Figure 6.6) was designed using PYMOL. The view shows all the important ligand binding sites along with the ligands such as Mg^{2+} ion, GTP, end of amino acyl tRNA- Cys residues attached with terminal adenosine. This figure also shows in a color-coded fashion identical (slaty blue), conservative (light blue), non-conservative substitutions (dark blue), important sites implicated in binding to tRNA (gray), EF-Ts (white), in addition to important site implicated in kirromycin (lime green) and amythiamicin (hot pink) drug resistance. This representation also shows that mutational positions known to cause the kirromycin resistance, overlaps with tRNA and EF-Ts contact sites whereas mutational positions known to cause amythiamicin resistance overlap with sites that forms the pocket which binds the amino-acyl group of amino-acyl tRNA.

The surface representation view confirms that the variation in the amino acids should not lead to any variations in the conformation of the binding sites of the EF – Tu protein molecule. Thus, the protein should be performing similar function as an Elongation factor in *Plasmodium vivax* as in other organisms. The conclusions drawn out with this study are based upon homology modeling. The actual functional properties of this protein can only be defined by structure conformational methods such as X – Ray crystallography.

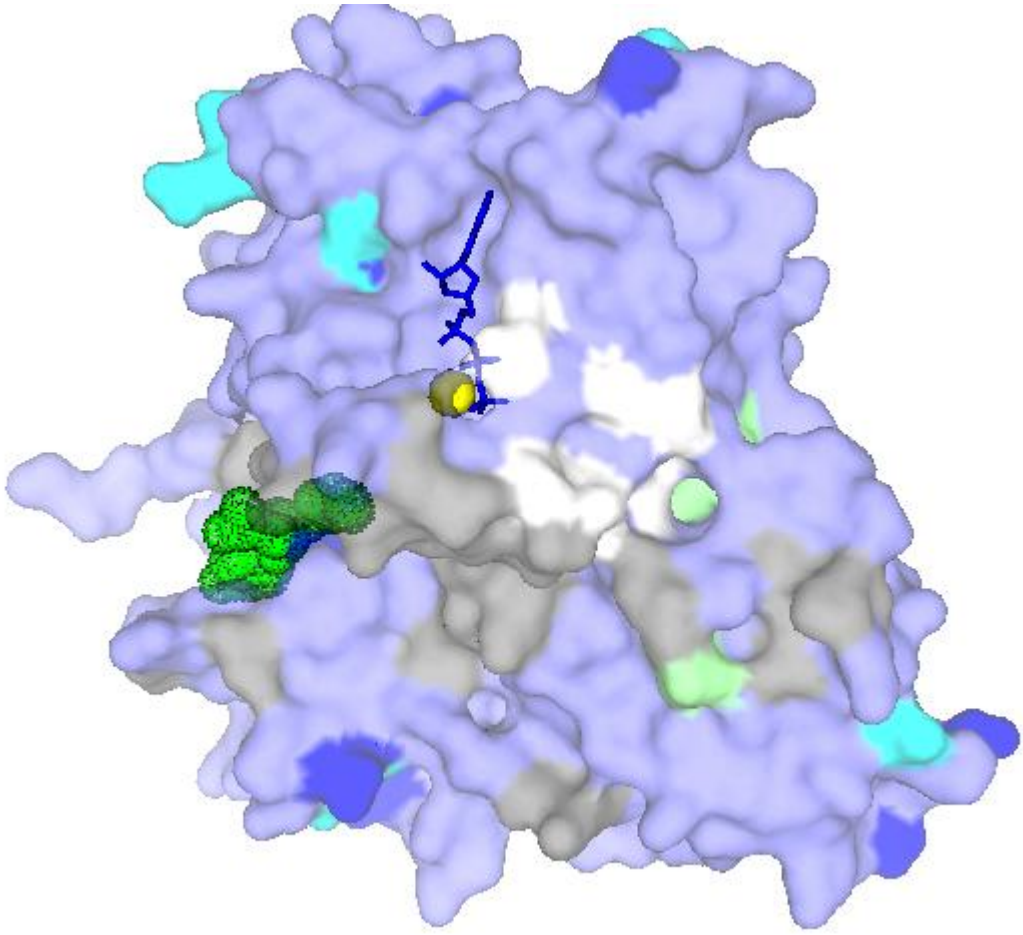

Figure 6.6: A surface representation of *P. vivax* EF-Tu prepared using **PYMOL.** Surface conservation is represented in three shades of blue, the slate blue represents the identity, cyan represents the similarity and the blue represents differences from *P.falciparum*. The binding regions for EF-Ts (white) and tRNA (grey) have been highlighted, these conserved areas would normally colour slate blue. The identical buried residues are not visible. Amino acids from drug resistant EF-Tu mutants are shown for GE2270 and amythiamicin (hot pink) and kirromycin (lime green).

RNA Polymerase C

The protein sequence of RpoC was submitted in Swiss Model database, to search for closely related Templates. From the obtained list of 3 – Dimensional X – Ray crystallographic structures, mostly from *Thermus thermophilus* (Swiss Prot Id: 1ynjD, 1ynnD, 2be5N, 2a6hN) and *E. coli* (Swiss Prot Id: 2a6eN), the structures showing best resolution of 2.9 - 3.2 were chosen and downloaded. These structures were of the complete RNA polymerase protein having regions for B, C1, C2 subunits. The complete structure had to be used as template to define the *P. vivax* RNA polymerase region C. The *P. vivax* RpoC sequence (Target) and the chosen templates were submitted to Swiss Model First Approach Mode. The structure (Figure 6.7) generated by the software was analysed in SPDBV.

The *P. vivax* RpoC protein when analysed against the *E. coli* and *T. thermophilus* RNA polymerase beta prime (β') subunit, showed about 34% match. Though this similarity is sufficient to consider this protein as a conserved protein, an in – depth analysis was done by performing homology model of the *P. vivax* protein with that of *T. thermophilus* protein.

A primary analysis of the *P. falciparum* RpoC amino acid sequence reveals the presence of active sites – a) in the form of a cluster of cysteine residues (C59, C71, C73, C89, C92) which may form a zinc binding site, and b) conserved sequences G343KRV347D and N456ADFDG461D, which are thought to be as DNA template binding sites (Gardner *et. al.*, 1991). These conserved regions matched with those in *E. coli* and are also present in the *P. vivax* protein sequence.

A view in the obtained structure of *P. vivax* RpoC protein also shows these sites to be conserved (Figure 6.8). Though there are certain variations where the loops are extended or the helices have shortened (Figure 6.9) but these variations are not affecting pockets of active sites.

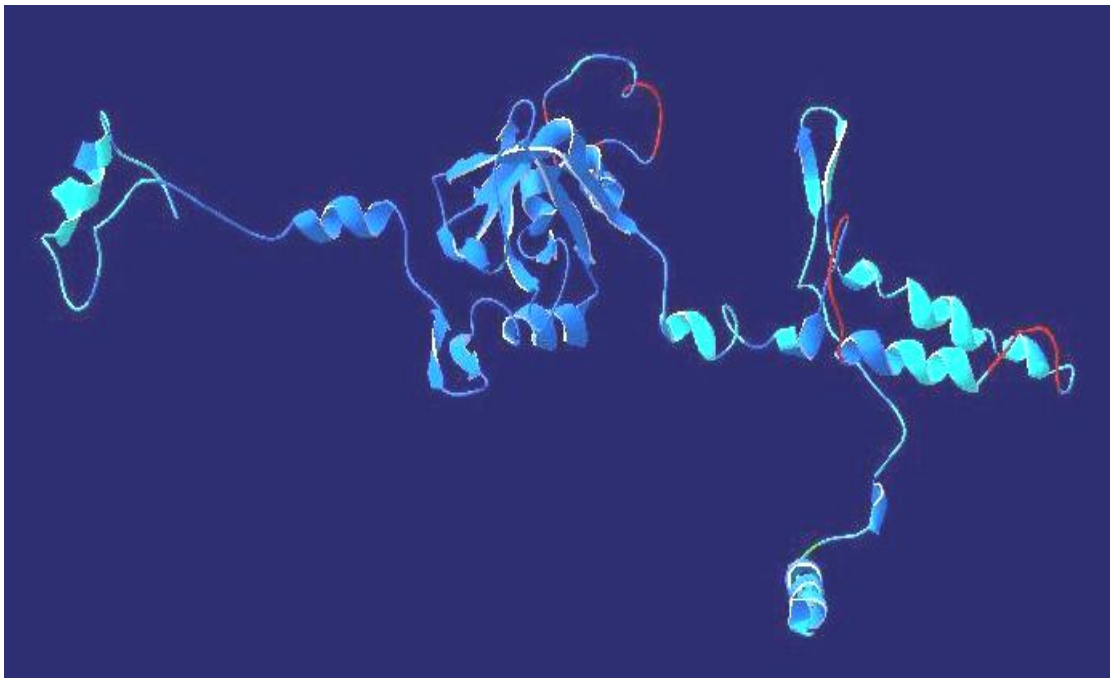

Figure 6.7: Model of RpoC as obtained from the SWISS MODEL site.

Figure 6.8: Amino acids shown in the active sites for DNA binding in *P. vivax* plastid RNA Polymerase C

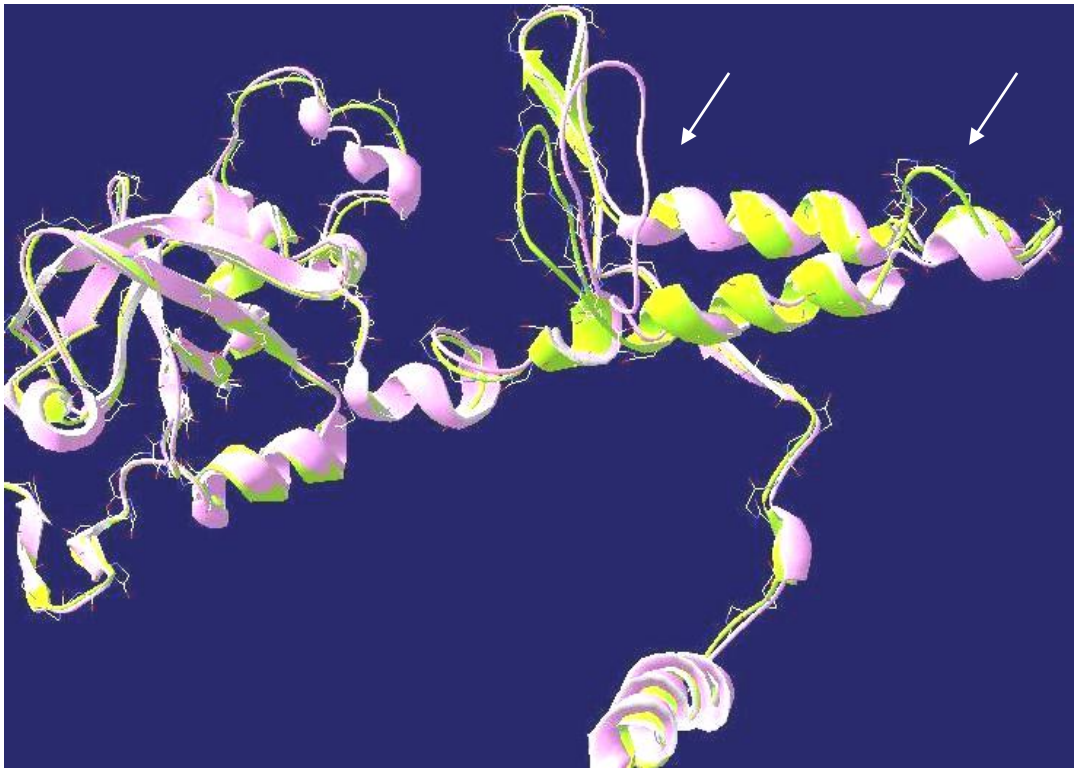

Figure 6.9: Differences in *P. vivax* (green) and *T. thermophilus* (pink) RpoC structures. Arrows indicating shift in loop and helices.

.

There are no details about *P. falciparum* RpoC structure in the protein database. But the protein sequence analysis reveals conservation of active sites in the sequence as for *P. vivax*. This would mean that the function of both *P. vivax* and *P. falciparum* RpoC proteins should be conserved with that of *E. coli* protein. A better analysis of this *P. vivax* protein structure and function can only be performed along with the complete sequences of *P. vivax* plastid RNA Polymerase B and D proteins.