Discussion
Discussion

Malaria remains a dominant public health issue in many developing regions of the world and it is the cause of millions of deaths annually with particularly high mortality among children (Snow et al., 2005). Despite the initiation in 1998 by World Health Organization of a campaign to 'Roll Back Malaria', the rates of disease and death caused by *Plasmodium falciparum* malaria in sub-Saharan Africa and other endemic areas are growing (Plowe, 2005). Malaria control efforts include attempts to develop an effective vaccine, eradicate mosquito vectors and develop new drugs (Doerig et al., 2005; Greenwood, 2005; Zhong et al., 2006). However the development of a vaccine has proven very difficult and a highly effective vaccine will probably not be available in the near future (Hoffman et al., 1995). Efforts to control *Anopheles* mosquitoes have had limited success, although the use of insecticide impregnated bed nets does appear to reduce malaria-related death rates (Alonso et al., 1991). In addition, methods to replace natural vector populations with mosquitoes which are unable to support parasite development, like introduction of parasite-inhibiting alleles in vector population will take decades before proving its effectiveness (Zhong et al., 2006). Currently there are only limited safe drugs for the treatment of this disease, however, the emergence of resistance to these commonly employed chemotherapeutics led to the reemergence of the disease as a major public health problem and its spread in new locations and populations (Tripathi et al., 2005). Growing drug-resistance among the *Plasmodia* motivates the development of additional anti-malarial drugs (Yeh and Altman, 2006).

Antimalarial drugs have been used for centuries. Early natural products, including the bark of cinchona tree and extracts of wormwood were used for centuries to control malaria. In the last 50 years, extensive efforts have led to the development of a number of effective synthetic antimalarial drugs. The most important of these, chloroquine has been the mainstay of antimalarial chemotherapy. However, resistance to chloroquine has been steadily increasing since the drug's initial use in late 1950s. Thus the use of chloroquine for the presumptive treatment of falciparum malaria or for chemoprophylaxis is usually no longer appropriate. Moreover, resistance to chloroquine of *P. vivax*, the second most lethal human malaria parasite, is increasing in South Asia (Murphy et al., 1993).
Newer drugs, such as mefloquine, halofantrine, atovaquone-pro-guanil and artemether-lumefantrine retain much efficacy but have limitations, not the least of which is their high cost. Novel uses of old drugs, such as chlorproguanil-dapsone, and artemisinin combination therapy offer definite possibilities for the near future, but still have regulatory, policy and implementation hurdles to jump. Consequently, new drugs that can control infection are badly needed (Beeson et al., 2001).

One strategy for developing new drugs is to identify and target parasite-specific metabolic pathways. A large number of biochemical pathways essential for the survival of parasite have been identified. Most notable discovery has been the identification of a curious feature of *Plasmodium* and other apicomplexan parasites, the presence of a plastid, an organelle found in plants and algae, which are thought to have arisen by endosymbiosis (or engulfment) of a cyanobacterial-like prokaryotic cell (McFadden and Roos, 1999). Plastid biosynthetic pathways are essential for parasite growth and are attractive therapeutic targets because of their fundamental differences from mammalian cells. A number of new drug targets like type II fatty acid synthesis pathway (Surolia and Surolia, 2001), shikimate pathway (Roberts et al., 1998), isoprenoid pathway (Jomaa et al., 1999), dihydrofolate reductase (DHFR) (Peterson et al., 1988), dihydroopteroate synthase (DHPS) (Wang et al., 1995), and prokaryote-like protein synthesis mechanism (Fichera and Roos, 1997) are being explored. A large pool of leads is important to arrive at few efficient drugs, as most of the candidates will be discarded due to problems with drug delivery, potency, cost and resistance. This emphasizes the need for the identification of more drug targets from *P. falciparum*.

The selection of parasite proteins as drug targets is usually based upon their importance for the survival of parasite, conservation across the different species of parasite and absence of their functional or sequence homologs in the host. In addition, information on the three dimensional structure of the parasite protein should be available. A wealth of information was made available for all three organisms that comprise the life cycle of the malaria parasite (i) *Plasmodium falciparum*, (ii) *Homo sapiens* and (iii) *Anopheles gambiae*. Annotation project predicted the proteomes of all these three organisms. The availability of complete proteome of these organisms allowed us to explore the proteome of these organisms in order to find out unique proteins that belong
to *P. falciparum*. Of the 5,334 proteins predicted in *P. falciparum* about 60% (3,208 hypothetical proteins) did not have sufficient similarity to proteins in other organisms. Thus, almost two-thirds of the proteins appear to be unique to malarial parasite. Unfortunately, for a majority of these genes their physiological function is not known or they do not code for essential enzymatic activities thus excluding the possibility of using them as drug targets. Assigning function(s) for these hypothetical proteins and understanding their roles in the biology of the parasite are the biggest challenges of the post genomic era.

5.1 **In silico prediction of 3D structure: a useful tool to generate structural information**

Three-dimensional structural information for *Plasmodium* proteins is critical for the identification of suitable drug targets. The *P. falciparum* genome has been published and predicted to code for 5334 genes, however, three-dimensional structures were known only for 40 proteins. Homology modeling was used to predict 3D structures for ~500 annotated *Plasmodium* proteins (Ramasamy et al., 2005). Models generated using homology modeling are more reliable than those of using fold recognition methods, due to the lack of complete understanding about the mechanism of protein folding. Nonetheless, automatically modeled proteins are error prone and require further validation. Therefore, in the present study, the modeling of *Plasmodium* proteins was semi-automated with error checking and correction being carried out manually.

An identity of 40 per cent was kept as a threshold for the selection of putative candidates for theoretical prediction of 3D structures, as the alignments sharing lesser identities might result in the generation of poor models (Guex et al., 1999). This resulted in the selection of 476 annotated *P. falciparum* proteins for comparative protein modeling. Alignment is the most crucial factor that dictates the quality of models generated. As the majority of our candidate sequences shared only less than 50% identity with their templates, utmost care was taken to avoid any incorrect alignment. The use of align2D aligner of the modeller package helped us to create biologically more meaningful alignments, thereby to predict models with high quality. Although align2D command implements a regular global dynamic programming method for the comparison of the two sequences, it also relies on the observation that evolution tends to place residue insertions and deletions in the regions that are solvent exposed, curved, outside secondary structure
segments and between two Cα positions close in space. The use of variable gap penalty for loops and core regions that are derived from template structure, guides the program to align the given sequences in its structural and evolutionary context (Fiser and Sali, 2003) and thus it favored the generation of high quality models in the present study. In the cases of alignments of sequences with very low identity or ambiguity, a profile based alignment procedure was used to achieve more reliable alignments. In the present study profile based alignment was used in 78 cases where ambiguities were found in the alignment and models generated, mainly because of low identity between target and template sequences.

The use of profile based and align2D alignment procedures resulted in the generation of good models in 85 per cent of the cases, as classified by the program Procheck (Laskowski et al., 1996), despite the fact that majority of these proteins share less than 50 per cent identity with their respective templates. Similarly, the RMS deviation calculated for the models with its respective template structures confirmed that the majority of generated models were of high quality. More than 85 per cent of the models were having RMSD of less than one, which are comparable with medium resolution NMR structures (Pieper et al., 2004). As a check for the efficiency of current pipeline of modeling, three P. falciparum proteins, for which structures were known, have been modeled based on structures from organisms other than P. falciparum and compared with their original crystal structures. RMS deviations calculated between the modeled 3D structures and their crystallized counterparts were found to be within the acceptable limit (0.5), thus proving the efficiency of the current strategy used for modeling.

5.1.1 Database of homology models of P. falciparum annotated proteins

A relational database was set up in ICGEB for easy storage and retrieval of information related to 3D predicted models (Ramasamy et al., 2005). A PHP based query page allows the users to search the database with protein id or description about the protein function as given by the PlasmoDB (Kissinger et al., 2002). Apart from coordinates for the generated models, the database provides following information: (i) name of the template used, (ii) information on the function of templates, (iii) coordinates of the template, (iv) coordinates for the overlapped structures of target and templates, (v) Ramachandaran plot for the models generated, (vi) Procheck summary output which
describes the quality of the model and (vii) protein sequences of template, target and alignment between them.

To our surprise, a quick search revealed the presence of 65 hypothetical proteins in the model database. This facilitated the assignment of putative functions for these proteins. However, only a small domain of the protein might be represented in the model database, which might not represent the real biochemical function of whole protein. Database has been integrated with PlasmoDB, the official database of *P. falciparum* genome sequencing project. The integration with PlasmoDB facilitates the user to form complex user-defined queries.

5.2 Identification of novel drug targets: using a comparative genomic approach

One of the important criteria for a protein to be a drug target is that its orthologous proteins should not be present in the host. The availability of proteomes predicted for both the parasite and host provided the impetus to find unique proteins present in the parasite. A comparative genomics strategy was employed to identify proteins unique to the parasite. Each of the 476 protein sequences, for which we generated structural information, was compared with the human proteome using BLASTP algorithm (Altschul et al., 1997) This analysis identified few parasite proteins that did not have any orthologs in the host proteome (Table 1). However, for many of these parasite proteins functional homologs were present in the host proteome. The list included enzymes like ferredoxin, acyl carrier protein, structural proteins like actin II and transporters like ABC transporters. Surprisingly some of them were annotated as hypothetical proteins. A prokaryotic homolog of proteasome protein, HsIV, was also found in the list. This parasite protein was found to have no homology with any of the known human proteins. This protein was found to be identical to a subunit of prokaryotic proteasome complex HslUV. Subsequently, the presence of other component (HsIU) of the same machinery was identified and the protein was found to be absent in human proteome. The presence of a complete prokaryotic proteasome machinery in the parasite and its absence in human proteome encouraged us to explore it further.

Proteasomes are major components of eukaryotic cellular machinery, mediating the normal turnover of proteins and the degradation of improperly folded or denatured proteins. In addition to these housekeeping functions, proteasomes play a key role in cell
cycle progression and the regulation of numerous transcription factors (Choi et al., 1996). The 26S proteasome is known to carry out protein degradation in eukaryotes and the HslUV system in prokaryotes. The presence of these proteosomes in an organism was considered mutually exclusive until the discovery of HslUV system in *Leishmania*. This is the first report of identification of HslUV system in *P. falciparum*. Even though the presence of this prokaryotic proteasome (HslUV) in a few protozoan parasites has been reported by Couvreur et al. (2002), the present study, the first of its kind, attempts to identify, annotate, clone and characterize these novel genes in any eukaryotic organism.

Rationales behind the choice of target are:

(i) The HslUV system’s prokaryotic origin and the absence of sequence homologs in the host. Genes constituting HslUV system in the parasite shared more than 60 per cent identity with its prokaryotic homologs while sharing less than ten per cent identity with the genes of eukaryotic proteasome machinery.

(ii) Lactacystin, which is an inhibitor of HslUV, was shown to be an anti-malarial, too.

(iii) The conservation of PfhslUV system in all *Plasmodium* spp further supported its candidature to be a potential drug target to control malaria.

(iv) Proteasomes are being actively developed as drug targets in many diseases like cancer (Mitchell, 2003) and HIV infection (Schmidtke et al., 1999).

(v) The successful identification of novel inhibitors specific to individual subunits/catalytic activities of 26S proteasome provided the confidence of developing specific inhibitors against PfhslUV proteasome also.

5.3 Identification and annotation of PfhslUV system: homolog of prokaryotic machinery

The prediction of the exact boundaries of the gene, exons and splice junctions in newly sequenced genomes, especially in an organism like *P. falciparum* whose overall AT content exceeds 80 per cent, is not trivial. The PfhslV gene, the proteolytic component of PfhslUV system, was predicted to encode an mRNA of more than two kb in the version 2 of PlasmoDB (Kissinger et al., 2002), even though its prokaryotic homologs are only 0.5-0.6 kb long. Most importantly, regions coding for catalytic site
residues were completely absent in the gene. Therefore, efforts were taken to reannotate the gene through computational and experimental approaches. The PfhslV gene has been reannotated to code a 0.5 kb long mRNA consisting of five exons. The PfhslV protein as described in PlasmoDB did not contain the catalytic site and its flanking residues in its sequence. Major reason for this incorrect prediction was, consideration of genomic region coding for the active site and flanking residues as introns by automatic gene prediction software (Majoros et al., 2003). An exon, earlier overlooked, coding for the catalytic site, was computationally found and experimentally verified by cloning the gene from cDNA library and sequencing.

Exceptionally long mRNA in the case of PfhslV, as predicted by PlasmoDB, might be due to the incorrect prediction of start and stop codons by prediction programs. The correct translation start site of PfhslV gene was identified as an ATG, 111 bases upstream to the active site residue, threonine. This change in the translation start site has been supported by the following four evidences, (i) the ATG satisfied the AxxATGG rule, a motif that overlaps the start sites in P. falciparum genes (Saul and Battistutta, 1990), (ii) sequences upstream to this ATG contain long stretches of low complexity regions (GC less than nine per cent; coding for runs of Asn), which suggested that this region may be a part of 5' UTR, (iii) with this start codon PfhslV contains a 37 amino acid long propeptide; leader peptides or propeptides are usually of <40 amino acids in length as in Leishmania, where as the HslV protein has a 22 amino acid long leader peptide (Couvreur et al., 2002) and (iv) sequences upstream to the catalytic site do not contribute to the activity and are to be cleaved off to make the HslV protein biochemically active. It would be uneconomical for the parasite to synthesize such a lengthy protein sequence just to cleave it off.

Errors in the prediction of 5' end of genes are more common in P. falciparum due to the lack of EST information for the N-terminal regions of genes, other inherent problems of gene prediction algorithms or both (Gardner et al., 2002). Besides, AT richness in P. falciparum poses problems to gene prediction algorithms, which are originally designed for GC rich eukaryotic organisms. Further complications in the annotation of Plasmodium genes are due to the prediction of 5' UTR regions as protein coding exons, which has resulted in the genome database containing runs of Asn in their
N-terminal region. It is commonly observed that in *P. falciparum*, prediction of 5' UTR regions as protein coding exons by the gene prediction software has resulted in annotated genes in the genome database containing runs of Asn in their N-terminal region.

The PfhsLV protein as predicted on the PlasmoDB contained a Lys repeat at its C-terminal, which was absent in other prokaryotic HslV proteins. This gave a hint that this region could originally be coded by an AT rich low complexity region and may not be a part of actual protein. During this study we determined the correct stop codon and found it to be present in a region that was considered an intron by prediction algorithms (Majoros et al., 2003). The presence of low complexity sequences after the corrected stop codon suggested that this part of exon must be coding for 3' UTR.

Further evidence for precision of our annotation of gene boundaries was provided by the fact that genomic regions flanking the corrected gene boundaries have GC content as low as nine per cent, while the coding region had a GC content of 34 per cent. This finding supported the refined PfhsLV gene structure predicted during this study. Thus entire PfhsLV gene was reconstituted computationally and confirmed experimentally. The recent release of PlasmoDB (V4), however, agreed well with our prediction of 3' end, but errors in 5' end of the gene still remain.

A single copy of the PfhsLV gene is present in chromosome 12 of *P. falciparum*, which codes for an open reading frame of 624 nucleotides and codes for a 207 amino acid long protein with a molecular weight of ~23 kDa. The PfhsLV protein contains 37 amino acids before the active site residue Thr, as a prodomain, which is cleaved off to produce a mature and active protein containing 170 amino acids with a molecular weight of ~19 kDa. The spliced gene has a GC content of 34%, which is by far one of the highest for any *P. falciparum* gene. The hslV gene in *E. coli* has a GC content of 53%. The GC content of the hslV gene in bacteria *Borrelia burgdorferi* is 35%, whose overall GC content is only 28.3%. Given the amino acid composition of HslV protein, this could be the least GC content any organism could have for hslV gene. The maintenance of 34% GC content in the coding region, in spite of *P. falciparum*’s tendency to accumulate AT rich codons signifies the functional importance of the PfhsLUV system in the parasite.

The PfhsLUV gene, regulatory and substrate recognition component of PfhsLUV system, was predicted to encode an mRNA of more than 3 kb in the version 2 of
PlasmoDB, even though their prokaryotic homologs coded only an ~1.2 kb mRNA. Therefore, through computational and experimental approaches, the PfhsIU gene was reannotated to code a 1.3 kb mRNA consisting of an exon. A single copy of the PfhsIU gene is present in the chromosome 9 of *P. falciparum*, which codes for an open reading frame of 1740 nucleotides with a GC content of 26 per cent. It codes for a 579 amino acid long protein with a molecular weight of ~66.3 kDa.

5.4 *In vivo* characterization of PfhsIUV system: expression, processing, localization and activity of PfhsIUV system in the parasite

Following the annotation and cloning of PfhsIV and PfhsIU genes, they were further characterized both *in vivo* and *in vitro*. Expression analyses indicated a differential expression of both genes during different intra-erythrocyte developmental stages of the parasite. It was observed that the transcription of RNA for PfhsIV gene was very low, medium and very high at ring, trophozoite and schizont stages, respectively. The *hsIV* and *hsIU* genes were known to express differentially in prokaryotic organisms as well. In *E. coli*, the *hsIV* gene expresses only in elevated temperatures (Yoo et al., 1996). A gradual increase in the amount of mRNA from ring to schizont stage was observed in both PfhsIV and PfhsIU genes. Surprisingly, PfhsIV protein was not detected by western blot analysis in the ring stage of the parasite while mRNA for the gene was present. The differential expression of PfhsIV and PfhsIU genes correlated well with the metabolic activities of different intra-erythrocytic stages of the malarial parasite. Metabolically less active ring stage showed no or a little quantity of PfhsIV protein while metabolically active trophozoites showed considerable amount of the same. The highest expression was seen in schizont stage of the parasite where a lot of used up proteins are to be degraded to avoid unnecessary carry over to the next cycle and the occurrence of many rounds of cell cycle needs the continuous synthesis and degradation of regulatory molecules.

The PfhsIV protein contains 37 amino acids before the active site residue Thr, as a prodomain, which is cleaved off to produce a mature and active protein containing 170 amino acids with a molecular weight of ~19 kDa. Both processed and unprocessed forms of PfhsIV protein were found to co-exist in the parasite.
A similar pattern of expression for both *PfhslV* and *PfhslU* suggested their combined activity as a proteasome complex and suggested an active role in the physiology of the parasite. HslU, a member of AAA+ ATPases works as a chaperone in prokaryotes (Neuwald et al., 1999). The biochemical function of HslU is to recognize, unfold and feed the target protein for degradation by the catalytic protease, HslV. Apart from working as a regulatory molecule in the proteasome system, HslU functions as a chaperone (Rohrwild et al., 1996) aiding in the folding of prokaryotic proteins. The involvement of PfhslU as chaperone in protein folding or in stress management in the parasite is still a speculation.

The determination of the location of the PfhslUV system in the parasite was important in order to understand its functional significance in *Plasmodium*. Till recently, *hslUV* genes were thought to be exclusively present in prokaryotic systems and the idea of its sub-cellular localization was not thought of. The PfhslV protein was computationally predicted and experimentally determined to be localized in the cytoplasm of the parasite which is in contrast to the common expectation of its organelle localization, a characteristic feature of proteins with bacterial origin. The computational analysis of the N terminal sequence of PfhslV and PfhslU using programs such as PlasmoAP, PATS, PlasMit and SignalP did not find any target signal. Instead PSORT and other programs suggested a cytoplasmic localization for PfhslV and PfhslU proteins (Emanuelsson et al., 2000; Bendtsen et al., 2004). IFA experiments proved the presence of PfhslV protein in the cytoplasm. The HslV protein encoded in *Leishmania* was predicted to have a mitochondrial-targeting signal (Couvreur et al., 2002), although experimental evidences for its localization in mitochondria and its absence in the cytoplasm are lacking. The presence of PfhslV protein, a prokaryotic homolog, in the cytoplasm suggested that it had gained some vital physiological function(s) in the cytoplasm of the parasite.

The preferential accumulation of the PfhslUV system in the parasite was ascertained. Co-localization experiments with parasites ERC (PfERC) protein suggested that the PfhslV was a soluble cytoplasmic protein and does not show any preferential accumulation in endoplasmic reticulum region. The existence of a nuclear body, called clastosome, enriched in 26S proteasomes, ubiquitin and the protein substrates of
proteasome in eukaryotes was reported earlier (Lafarga et al., 2002). Such local bodies in the nucleus of the parasite were not observed during our analysis. While both eukaryotic and prokaryotic proteasome systems are present in the cytoplasm, it is not clear which of this machinery is taking care of protein degradation in organelles. Our analysis did not exclude the possibility of either of these proteasomes getting transported into organelles. In such cases, other proteins in the parasite might be involved in the transport of proteasome proteins to the organelles. In plants, some proteins lack a transit peptide but are targeted to plastids via an unknown process. Gardner and co-workers suggested that proteins that use an alternative-targeting pathway in *P. falciparum* would have escaped detection during the annotation of the genome (Gardner et al., 2002).

After confirming the expression and localization of the PfhsIV protein it was necessary to find out if PfhsIV and PfhsIU proteins formed a complete complex in the parasite. The formation of a complete complex is very essential for the proteasome to carryout its biochemical and physiological function. In prokaryotes it was proved to form a four ringed complex; two rings of HsI hexamer and two rings of HsIU hexamer (Azim et al., 2005). If the whole complex was formed in the parasite, as observed in prokaryotes, the size of whole complex would be around 800 kDa (Bochtler et al., 2000). Western analysis on chemically cross-linked parasite lysate did not lead to any conclusion as the lysate failed to enter the gel. But, when the chemically cross-linked and non-cross linked parasite lysates were fractionated on GPC column and probed with anti-PfhsIV antibodies, bands specific to PfhsIV could be detected in fractions equivalent to >600 kDa. This supported the formation of whole PfhsIUV complex in the parasite. The high molecular size of the complex might prevent it from entering the SDS-PAGE gel, thus producing a smear near the well in western blot. The fact that both chemically cross-linked and non-cross- linked parasite lysates produced similar results in GPC fractionation suggested that the formation of PfhsIUV complex is through some very strong protein-protein interactions. But, our experiments dismissed the possibility of the involvement of a covalent bond formation between the subunits.

We also observed large complex formation in the parasite of both processed and un processed PfhsIV. Therefore, our analysis also suggested that the processing of prodomain is probably a self-cleaving mechanism and takes place once the assembly of
HslUV complex is complete. Our reconstituted quaternary structure of HslUV complex revealed that the active site and the prodomain regions were projected towards the cavity of barrel structure, thus excluding the possibility of another protease gaining access to cleave the prodomain after the assembly of PfhslUV complex or the hexamer of PfhslV.

The HslV proteases, in general, contain three protease activities; (i) chymotrypsin-like activity, (ii) trypsin-like activity and (iii) peptidyl glutamyl peptide hydrolase activity (Lowe et al., 1995) each having specificities for different substrates. The lysate of mixed stage malarial parasite showed trypsin-like proteolytic activity typical of proteasomes and was inhibited by its specific inhibitor. Owing to the fact that parasites have many proteases, this activity observed could not be attributed to proteasomes alone. To add to this complication, it is very difficult to quantitatively assess the contribution of two different proteosome systems present in *P. falciparum* towards the observed proteolytic activity.

An assay system, required to study the *in vivo* activity of PfhslUV system specifically, was established in this study. Knowledge about the substrate specificities of HslUV system is very limited. Till date, only a few proteins like Arc-repressor, SulA, RcsA and casein were known as substrates for HslUV system in prokaryotes (Kuo et al., 2004; Burton et al., 2005). Arc-repressor protein, a natural substrate for HslUV system in *E. coli*.

A small N terminal fragment of Arc was fused with a reporter gene, GFP and cloned in *P. falciparum* transfection vector. A parasite line was established with episomally transfected Arc-GFP fusion construct to assess the activity of PfhslUV system in the parasite. This construct, if expressed in the parasite, was expected to produce a fusion protein of Arc and GFP. The PfhslUV system was expected to recognize and degrade the Arc protein and thereby its fusion partner GFP also. Thus monitoring the GFP level in the parasite would indicate the level of activity of the PfhslUV system, indirectly. It was observed that the expression of GFP was very high in the ring stage where the expression of *PfhslUV* genes was minimal and the fluorescence of GFP went down in trophozoite and schizont stages where the expression of PfhslUV system went up. Our successful demonstration of inverse correlation between the expression patterns of Arc-GFP protein and PfhslUV proteins showed that PfhslUV system in the parasite is

82
indeed functional and degrades Arc-GFP fusion protein. This again confirmed the activity of the PfhsIUV system in the cytoplasm. Burton and coworkers showed that HslUV efficiently degrades Arc repressor and degradation depends on sequences near the N terminus of Arc. They have also demonstrated that 24 amino acids at the N terminus of the Arc protein is important for the interaction with HslU and this sequence is a degradation tag that binds directly to HslU very strongly (Burton et al., 2005). Protozoan parasites are the only group of organisms where both HslUV system and ubiquitin conjugation cascade co-exist. The PfhsIUV system recognizing and degrading the ubiquitinated proteins is a possibility, which still needs to be explored.

The establishment of this transfected parasite line, thus opens up a gate for the in vivo screening of a number of drug molecules that might be developed in future against PfhsIUV proteasomes.

5.5 PfhsIUV system as a potential drug target: validation by silencing

Proteasome machinery has been a major target to design drugs against cancer, HIV infection etc. (Schmidtke et al., 1999; Mitchell, 2003). The selection of HslUV proteasome machinery as a potential drug target for malaria is based upon the fact that proteasome inhibitors are known to act as antimalarials (Gantt et al., 1998). However, there are no evidences to believe that the effects observed in these studies were due to the inhibition of HslUV system. Thus the function and indispensability of the whole machinery had to be proved before considering them as drug targets. A systematic functional and genetic analysis would prove the functional role of a given gene. Knocking out the gene and complementing it would be the best way to understand the physiological function of a gene. However, it is very difficult and time consuming to knockout the genes in *P. falciparum*. An alternate way to demonstrate a gene’s indispensability is to use RNAi, which is expected to specifically silence a gene. Many earlier studies have demonstrated the use of RNAi in understanding the function of genes in *Plasmodium*. Parasites treated with falcipain's dsRNAs showed marked reduction in the levels of corresponding endogenous falcipain mRNAs (Malhotra et al., 2002). Parasites treated with dihydroorotate dehydrogenase (DHODH) dsRNAs showed marked reduction in the levels of corresponding endogenous DHODH mRNAs and reduction in parasitemia (McRobert and McConkey, 2002). The abrogation of protein phosphatase
expression by synthetic short interfering RNA (siRNA) led to the inhibition of parasite DNA synthesis (Kumar et al., 2002).

The in vitro cultures treated with the PfhsIV gene specific siRNA showed deformation in morphology and reduction in parasitemia. The shrinking of the parasite cytoplasm started to appear within the five hours of siRNA treatment followed by the death of parasites. More than 90 per cent reduction in the parasitemia was observed within the same life cycle of parasite. Even those parasites survived that did not progress to the next stage of their life cycle. The siRNA treatment on late ring and early trophozoite stages produced similar results.

Due to the lack of knowledge about the co-existence of HslUV system and 26S proteasome in the same organism, the effect of proteasome inhibitors was attributed to the inactivation of 26S proteasomes. The unavailability of inhibitors specific to either of the systems made it impossible to study the contribution of 26S proteasomes in the abnormalities observed. But in this study, we used RNAi to specifically silence PfhsUV system to show its functional significance in the parasite biology. The inhibition of translation of PfhsIV mRNA in ring or trophozoite stage of parasite using siRNA showed a dramatic effect on the survival of the parasite. The indispensability of PfhsIV gene, thereby the whole PfhsUV system, for the survival of the parasite was successfully demonstrated in the present study.

Significant reduction in parasitemia, more than 90 per cent, in a few hours after the treatment with siRNA, suggested that PfhsUV system was not only working as waste disposal machinery but also involved in the regulation of some vital pathways in the parasite. Earlier studies have also suggested the possession of essential functions by the proteasomes in the growth and development of protozoan parasites (Lindenthal et al., 2005). Proteasome inhibition by lactacystin was shown to affect the stage transformation of both forms of Trypanosomes (Gonzalez et al., 1996) and to prevent the encystations of Entamoeba invadens (Gonzalez et al., 1997). Lactacystin is also known to block the development of pre-erythrocytic and intra-erythrocytic stages of Plasmodium spp (Gantt et al., 1998). These studies have clearly demonstrated the involvement of proteasomes in the cell cycle progression of different parasites. The inhibition of progression of the parasites, which survived siRNA treatment, to further developmental stages suggested the
involvement of PfhsIUV systems in cell cycle regulation. The abundance of PfhsIUV proteins in schizont stage, where there are repeated cycles of DNA replication provided further evidence. The involvement of HslUV system in cell cycle progression was known in *E. coli*, where it regulates the timing of cell cycle by controlling the degradation of cell cycle inhibitor SulA (Kanemori et al., 1999). It had also been demonstrated in other eukaryotic systems that the inactivation of 26S proteasomes using proteasome inhibitors led to cell cycle growth arrest (Lindenthal et al., 2005).

Our results suggest that the effects observed in parasites treated with proteasome inhibitors in earlier studies may be actually due to the inhibition of PfhsIUV system or of both 26S proteasome and PfhsIUV system. This hypothesis was further supported by the fact that visual morphological changes observed in siRNA treated parasites such as, the condensation of nuclei and shrinking of whole cytoplasm of the parasite in to smaller dots were similar to those observed in lactacystin or MLN-273 treated cultures (Gantt et al., 1998; Lindenthal et al., 2005).

5.6 Cloning, expression and characterization of PfhsI and PfhsIU proteins: the components of PfhsIUV proteasome system

We established the functional significance of the PfhsIV protease in the parasite by using ARC-GFP chimera parasite lines and RNAi studies. To further characterize the protease activities of PfhsIV, we tried to express the active recombinant protein. The PfhsIV and PfhsIU genes were amplified from cDNA library and cloned in to two different cloning vectors. These constructs were then subcloned into different *E. coli* expression vectors like pET29b, pQE30 and pMAL2P. As in the case of many of the proteases from *Plasmodium*, PfhsIV is also a cysteine rich protein. This caused difficulties in the production of active recombinant proteins in *E. coli* (Sijwali et al., 2002). Although, *E. coli* produced recombinant PfhsIV in large quantities, most of the PfhsIV protein got accumulated in to insoluble inclusion bodies. In spite of various induction strategies worked out, difficulties were there in expressing their soluble and active forms. Five different expression cell lines, four different vectors were used to express the PfhsIV protein in soluble fraction. Ten different time intervals, 15 different IPTG concentrations, 8 temperature conditions and their permutation and combinations were tried, but in vein. Our efforts to refold the protein yielded a soluble protein but the
proteolytic activities were not retained by it. However, we were able to produce active recombinant protein with MBP fusion and with histidine tag by inducing E. coli culture at 1.2 OD for 30 min. The column refolded PfhsIV protein was used to rise polyclonal antibodies in mice and rabbit.

HslIV proteases in general contain three protease activities; (i) chymotrypsin-like activity, (ii) trypsin-like activity and (iii) peptidyl glutamyl peptide hydrolase activity (Lowe et al., 1995) each having specificities for different substrates. The PfhsIV was found to harbour all the three activities. However, low affinities (high $K_m$ values) to different substrates and low $V_{max}$ signified PfhsIV as a weak peptidase compared to the HslIV protease of E. coli (Seol et al., 1997). But it should be kept in mind that low activity may actually be because of (i) the determination of the activity of PfhsIV in the absence of its regulator PfhsIU; HslIV is considered to be a weak peptidase on its own and (ii) the folding of eukaryotic HslIV (PfhsIV) in E. coli might not be so good as in the parasite.

The addition of HsIU protein in the reaction is known to increase the proteolytic activities of HslIV protein in E. coli (Kwon et al., 2004). HsIU protein forms a homo hexamer and inserts its C-terminal into the hexameric rings of HslIV to activate it. To test, if a recombinant PfhsIU can enhance activities of PfhsIV, we cloned and expressed recombinant PfhsIU. The recombinant PfhsIU was found to have high ATPase activity. The addition of recombinant PfhsIU to PfhsIV had no effect on the proteolytic activities of PfhsIV. One of the reasons for this, could be the poor interaction of PfhsIV and PfhsIU-MBP fusion protein. Large fusion partner such as MBP may hinder the interaction of these proteins. In addition, PfhsIU may not be of any help to PfhsIV to cleave small peptide substrates in $in vitro$ reactions. The inhibition of proteolytic activities of PfhsIV by general inhibitors showed the possibility of developing specific inhibitors against PfhsIU proteasome system as anti-parasite drugs. The $in vitro$ protease assay using recombinant PfhsIV protein can be used for screening a variety of potential PfhsIV inhibitors to identify possible novel antimalarials.