Efficient expression systems for cysteine proteases of malaria parasites

Too good to be true?

Emir Salas Sanby and María de los A. Chávez Planes

Centro de Estudios de Proteínas; Facultad de Biología, Universidad de la Habana; Havana, Cuba; Red CYTED-PROMAL (210RT0398); Proteómica y Quimiogenómica de Inhibidores de Proteasas de Origen Natural con Potencial Terapéutico en Malaria

Keywords: cysteine proteases, malaria parasites, efficient expression systems, codon optimization, auto-induction, Escherichia coli, Pichia pastoris

Abbreviations: CAI, codon adaptation index; GST, glutathione S-transferase; FCU, frequency (per thousand) of codon usage; 6xHis-tag, a stretch containing six consecutive histidine residues tag; IPTG, isopropyl-1-thio-β-D-galactopyranoside; LB, Luria-Bertani medium; MBP, maltose-binding protein; OD₆₀₀nm, Optical density at 600 nm

Introduction

Malaria, the most prevalent protozoal infection worldwide, remains a major public health problem globally.¹ It is caused in humans by five species of the genus Plasmodium and causes annually the death of about 1 million people; more than 350 million new infections and 3.2 billion persons are at continuous risk of infection (http://www.dndi.org/diseases/malaria.html). Due to the lack of an effective vaccine, chemotherapy constitutes the only line of defense against malaria. However, the emergence and wide spread of drug-resistant strains reduce the effectiveness of current therapy and demand the urgent development of new drugs with novel mechanisms of action.²

Over the last ten years, the study of peptidases from malaria parasites has acquired considerable importance. Some of these enzymes have been proposed to play central roles in diverse processes such as cell invasion, differentiation, cell cycle progressions, catabolism of host proteins, parasite feeding and evasion of the host immune response; making them attractive targets for therapeutic intervention.³ Peptidases involved in hemoglobin degradation, and particularly those belonging to Clan CA family C1 (CA1), have proven very attractive as metabolic targets, considering the experimental evidence that has arisen from both the use of specific inhibitors and gene-disruption experiments.⁴ Papain-like cysteine peptidases from human malaria parasites P. falciparum (Falcipains 2, 2’ and 3)⁵⁻⁷ and P. vivax (Vinckepains 2, 3 and 4)⁸⁻¹⁰ have been extensively characterized. In addition, several cysteine peptidases from rodent malaria parasites have attracted some attention, given the relevance of murine models for the in vivo evaluation of inhibitors with therapeutic potential. These include Vinckepains (P. vinckei),¹¹ Berghelpains (P. berghei)¹² and Chabaupains (P. chabaudi).¹³ Other
closely-related isoforms of these enzymes have also been identified and characterized in different plasmodial species, but these are not involved in hemoglobin degradation and their roles in parasite physiology are now being elucidated.

A sequence analysis of the mature domain of these enzymes shows that they are fairly typical C3A cysteine peptidases (fC). Fitting all the conserved residues important for enzymatic catalysis. However, they display several unusual features for a papain family enzyme. An N-terminal extension of the catalytic domain has been identified (or computationally predicted) in several of them, displaying a moderate degree of sequence similarity. At least for Falcipain-2, this segment has proven to be functionally relevant for refolding of mature enzyme in vitro. The X-ray structure of mature Falcipain-2 confirms the presence of a short α-helix in this segment, with conserved residues making multiple side-chain interactions with the protein core. Second, they have an unusual motif near the C terminus that is absent in Papain, which in Falcipain-2 is involved in hemoglobin binding. Furthermore, plasmodial enzymes possess four disulfide bonds, one more than in the case of Papain, with a total of nine Cys residues. Like the rest of the enzymes of the family, they contain an inhibitory prodomain which is removed by auto-proteolysis. In contrast to other C3A enzymes, this prodomain seems to be unnecessary for the correct folding of plasmodial cysteine peptidases.

Since their early identification, it was evident that heterologous expression systems would be needed to produce sufficient amounts of these enzymes for their further characterization. However, many of these expression systems have displayed relatively low productivity. Under the rationale that highly-efficient expression systems will accelerate the high-throughput screening identification of novel inhibitors with therapeutic potential, we review in this paper the expression systems developed so far for plasmodial cysteine peptidases in order to identify the possible causes of this phenomenon. We found the striking differences in codon and nucleotide composition between plasmodial and host genes to be the principal cause of expression failure, and discuss several strategies to overcome these limitations. In addition, we highlight the potentialities of non-bacterial expression systems as alternatives for the efficient expression of these particular enzymes.

Analysis of Current Heterologous Expression Systems for Cysteine Peptidases from Human and Rodent Malaria Parasites

The ideal expression system for these targets should meet several requirements. Primarily, it should provide abundant amounts of soluble and active enzymes, with structural and functional characteristics resembling those of the natural molecule. In addition, the system should allow easy downstream purification of the recombinant product, being operationally simple and economically efficient. Table 1 reviews some characteristics of the expression systems described to date for expression of these enzymes.

A rapid analysis of this table highlights relevant similarities. First, E. coli has been the most popular expression host used for the recombinant production of these enzymes, accounting for the vast majority of them. This selection is based on well-known advantages it displays over other expression systems, such as (1) high productivity, (2) fast growth at a high cell density in inexpensive media, (3) well-characterized genetics and (4) the availability of a large number of cloning vectors and mutant host strains.

Second, almost all the constructs have been based on native plasmodial genes. Only one study was performed using a codon-optimized synthetic gene for the efficient expression in the selected host. Since early identification, it was evident that heterologous expression systems would be needed to produce sufficient amounts of these enzymes for their further characterization. However, many of these expression systems have displayed relatively low productivity. Under the rationale that highly-efficient expression systems will accelerate the high-throughput screening identification of novel inhibitors with therapeutic potential, we review in this paper the expression systems developed so far for plasmodial cysteine peptidases in order to identify the possible causes of this phenomenon. We found the striking differences in codon and nucleotide composition between plasmodial and host genes to be the principal cause of expression failure, and discuss several strategies to overcome these limitations. In addition, we highlight the potentialities of non-bacterial expression systems as alternatives for the efficient expression of these particular enzymes.

Analysis of Current Heterologous Expression Systems for Cysteine Peptidases from Human and Rodent Malaria Parasites

The ideal expression system for these targets should meet several requirements. Primarily, it should provide abundant amounts of soluble and active enzymes, with structural and functional characteristics resembling those of the natural molecule. In addition, the system should allow easy downstream purification of the recombinant product, being operationally simple and economically efficient. Table 1 reviews some characteristics of the expression systems described to date for expression of these enzymes.

A rapid analysis of this table highlights relevant similarities. First, E. coli has been the most popular expression host used for the recombinant production of these enzymes, accounting for the vast majority of them. This selection is based on well-known advantages it displays over other expression systems, such as (1) high productivity, (2) fast growth at a high cell density in inexpensive media, (3) well-characterized genetics and (4) the availability of a large number of cloning vectors and mutant host strains.

Second, almost all the constructs have been based on native plasmodial genes. Only one study was performed using a codon-optimized synthetic gene for the efficient expression in the selected host. Since early identification, it was evident that heterologous expression systems would be needed to produce sufficient amounts of these enzymes for their further characterization. However, many of these expression systems have displayed relatively low productivity. Under the rationale that highly-efficient expression systems will accelerate the high-throughput screening identification of novel inhibitors with therapeutic potential, we review in this paper the expression systems developed so far for plasmodial cysteine peptidases in order to identify the possible causes of this phenomenon. We found the striking differences in codon and nucleotide composition between plasmodial and host genes to be the principal cause of expression failure, and discuss several strategies to overcome these limitations. In addition, we highlight the potentialities of non-bacterial expression systems as alternatives for the efficient expression of these particular enzymes.

Analysis of Current Heterologous Expression Systems for Cysteine Peptidases from Human and Rodent Malaria Parasites

The ideal expression system for these targets should meet several requirements. Primarily, it should provide abundant amounts of soluble and active enzymes, with structural and functional characteristics resembling those of the natural molecule. In addition, the system should allow easy downstream purification of the recombinant product, being operationally simple and economically efficient. Table 1 reviews some characteristics of the expression systems described to date for expression of these enzymes.

A rapid analysis of this table highlights relevant similarities. First, E. coli has been the most popular expression host used for the recombinant production of these enzymes, accounting for the vast majority of them. This selection is based on well-known advantages it displays over other expression systems, such as (1) high productivity, (2) fast growth at a high cell density in inexpensive media, (3) well-characterized genetics and (4) the availability of a large number of cloning vectors and mutant host strains.

Second, almost all the constructs have been based on native plasmodial genes. Only one study was performed using a codon-optimized synthetic gene for the efficient expression in the selected host. Since early identification, it was evident that heterologous expression systems would be needed to produce sufficient amounts of these enzymes for their further characterization. However, many of these expression systems have displayed relatively low productivity. Under the rationale that highly-efficient expression systems will accelerate the high-throughput screening identification of novel inhibitors with therapeutic potential, we review in this paper the expression systems developed so far for plasmodial cysteine peptidases in order to identify the possible causes of this phenomenon. We found the striking differences in codon and nucleotide composition between plasmodial and host genes to be the principal cause of expression failure, and discuss several strategies to overcome these limitations. In addition, we highlight the potentialities of non-bacterial expression systems as alternatives for the efficient expression of these particular enzymes.
genes in heterologous systems. These divergences often cause problems during expression, such as decreased mRNA stability and slow translation rates; early termination of transcription and translation; occurrence of insertions, deletions and frame-shifts; and finally inhibition of protein synthesis or even arrest of cell growth. The relative abundance of codon-specific tRNAs in different host systems explains partially why some genes are better expressed than others in a selected system. The Codon Adaptation Index (CAI) parameter is commonly used to describe how well the codons of the target gene match the codon usage preference of the host organism. Consequently, the maximum value of CAI of 1.0 indicates a perfect adaptation; but in practice CAI values ≤ 0.9 are considered as very good predictors of high expression.

Figure 2 shows that plasmodial genes for cysteine peptidases are distant from highly expressed genes in *E. coli*. First, they show CAI values of ~0.6 whereas those for reference genes are around 0.75. However, the most striking difference is found in the nucleotide composition of plasmodial genes, which are among the A/T richest genomes sequenced so far. The average A+T content of reference *E. coli* genes is near 45%, while it is ~70% in plasmodial ones. Interestingly, among them, the genes for Vivipains seem to show a lower A+T content, suggesting a less problematic expression in this host. The divergences in nucleotide composition observed for the rest of the genes are mainly focused in the first (62.1% vs. 38.7%) and third (82.6% vs. 41.9%) codon positions. Consequently, it is frequent to find in these plasmoidal genes, codons that are rarely found in highly expressed *E. coli* genes with more favorable codons throughout the whole sequence. This strategy allows not only the optimization of CAI value for the selected host, but also in improved solubility of the fusion protein.

The best strategy to eliminate differences in codon usage is the recodondization of specific plasmoidal genes, by replacing codons that are rarely found in highly expressed *E. coli* genes with more favorable codons throughout the whole sequence. This strategy successfully used by us to enhance the expression of Falcipain-2. As can be observed in Figure 2, the optimized construct showed an almost perfect match to the codon usage preferences of the host (CAI = 0.96) and a base composition (percent of A + T = 51.3) very close to that of the host reference genes. Consequently, the construct was expressed to very high-level in several *E. coli* strains, with the recombinant enzyme representing 35–55% of the total proteins (depending on the host). These expression levels were 3.5–5.5 fold higher than those obtained with the native *P. falciparum* enzyme under similar conditions, exemplifying the positive effect that codon harmonization could exert in the expression of these enzymes.

Although not strictly a genetic strategy, we also introduced an auto-inducing medium to achieve high-cell-density *E. coli* cultures as an additional step to further enhance productivity. As a result, saturation OD_{600} in auto-inducing medium was considerably higher than that obtained in LB under similar culture conditions, bringing a ~1.5-fold increase in the volumetric productivity of the system. In addition to higher productivity, the system proved to be more economic and operationally simpler than classical IPTG induction, with a high degree of adaptability to other related targets.

Soluble and active expression. In addition to the above-mentioned approaches, the optimization of expression conditions can also influence the global yield of active enzymes. The rationale of this strategy is not based on the augmentation of expression levels, but on increasing the proportion of properly folded product by modifying the expression conditions. Consequently, the main goal would be the identification of induction conditions to achieve the expression of soluble and active enzymes, bypassing solubilization, renaturation and reactivation of active enzyme from insoluble/inactive aggregates. Despite sporadic and fortunate successes, many classical approaches have repeatedly failed, including changes in culture/induction temperature and time, promoter strength, inducer type and...
of plasmodial cysteine proteases: the methylotrophic yeast Pichia pastoris for Falcipain-2 and Trichoplusia ni insect cells for Falcipain-1. In both cases the results were unsatisfactory, showing no detectable concentration, growth media, E. coli host strain, cellular density at which induction began and addition of folding enhancers.

As mentioned above, the most reproducible condition to achieve soluble and active plasmodial cysteine proteases has been its fusion to MBP, with the inconvenience of low efficiency removal of the fusion partner.22

The formation of insoluble inclusion bodies has been the most common and important productive limitation of the expression systems developed so far for plasmodial cysteine proteases. Although the compositions of refolding buffers have been optimized with high-throughput technology, the current refolding methodology is so inefficient that it constitutes the limiting-step in the production of these enzymes. Any improvement in the existing refolding methodologies will constitute an actual progress, with great impact on final productivity.

Table 1. Heterologous expression systems of Papain-like cysteine proteases from human and rodent malaria parasites

| Enzyme  | System | Expression vector | E. coli host strain | Inclusion body formation | Soluble protein yield | Folding efficiency | refolding steps | reference |
|---------|--------|-----------------|-------------------|-------------------------|----------------------|------------------|---------------|----------|
| Falcipain-2 | P. pastoris | pMT238 | Ab101 | No | 1 bodies | Not achieved | ND | - | Sato et al., 1994 |
| Falcipain-2 | P. pastoris | pMT238 | Ab101 | No | 1 bodies | Not achieved | ND | - | Sato et al., 1994 |
| Falcipain-2 | P. pastoris | pMT238 | Ab101 | No | 1 bodies | Not achieved | ND | - | Sato et al., 1994 |
| Falcipain-2 | P. pastoris | pMT238 | Ab101 | No | 1 bodies | Not achieved | ND | - | Sato et al., 1994 |
| Falcipain-2 | P. pastoris | pMT238 | Ab101 | No | 1 bodies | Not achieved | ND | - | Sato et al., 1994 |
| Falcipain-2 | P. pastoris | pMT238 | Ab101 | No | 1 bodies | Not achieved | ND | - | Sato et al., 1994 |
| Falcipain-2 | P. pastoris | pMT238 | Ab101 | No | 1 bodies | Not achieved | ND | - | Sato et al., 1994 |
| Falcipain-2 | P. pastoris | pMT238 | Ab101 | No | 1 bodies | Not achieved | ND | - | Sato et al., 1994 |
| Falcipain-2 | P. pastoris | pMT238 | Ab101 | No | 1 bodies | Not achieved | ND | - | Sato et al., 1994 |
| Falcipain-2 | P. pastoris | pMT238 | Ab101 | No | 1 bodies | Not achieved | ND | - | Sato et al., 1994 |
| Falcipain-2 | P. pastoris | pMT238 | Ab101 | No | 1 bodies | Not achieved | ND | - | Sato et al., 1994 |
| Falcipain-2 | P. pastoris | pMT238 | Ab101 | No | 1 bodies | Not achieved | ND | - | Sato et al., 1994 |
| Falcipain-2 | P. pastoris | pMT238 | Ab101 | No | 1 bodies | Not achieved | ND | - | Sato et al., 1994 |

Where more than one expression vector was used, that finally used for large-scale expression is indicated in bold letter. ND, not described.

Figure 2. Comparison of plasmodial genes for cysteine proteases (●) with highly expressed E. coli genes (*) corresponding to glycolytic and tricarboxylic acid cycle enzymes. The cluster corresponding to vivipain genes is indicated with dashed lines. The optimized falcipain-2 gene is also indicated with a star. CAI values and global A + T content were calculated using the CAIcal server (http://genomes.urv.es/CAIcal/).27

Non-Bacterial Expression Systems

Non-prokaryotic expression systems have been used only twice for the expression of plasmodial cysteine proteases: the methylotrophic yeast Pichia pastoris for Falcipain-2 and Trichoplusia ni insect cells for Falcipain-1. In both cases the results were unsatisfactory, showing no detectable

©2013 Landes Bioscience. Do not distribute
expression and limited recovery of the active protein, respectively.4,5,22 However, eukaryotic systems might be predicted to be better hosts than E. coli for the expression of these enzymes, as they share more compatible translational machinery and are capable to perform many post-translational modifications (disulfide bond formation, glycosylation, post-translational processing, etc.) critical for the folding and activity of C1A cysteine proteases.

Figure 3 shows the distribution of CAI values of these plasmoidal genes for several eukaryotic systems, commonly used for the heterologous expression of Papain-like cysteine proteases. As expected, all the genes showed higher CAI values for eukaryotic hosts than for E. coli, indicating better adaptation to the codon usage preferences of these organisms, with the highest values for Saccharomyces cerevisiae. Interestingly, again vivipain genes deviate from the general behavior of the group, clearly indicating species-specific differences that must be considered for the design of efficient expression systems. Among these organisms, yeast hosts clearly constitute the best alternative to support highly-efficient expression systems for the target enzymes, considering their productivity, easy handling and cost-effective cultivation. In particular, the methylotrophic yeast Pichia pastoris is among the most used expression platforms for biomedical and biotechnological applications, mainly due to the following: (1) the presence of a strong, tightly regulated, and easily manipulated promoter derived from the P. pastoris alcohol oxidase 1 gene (AOX1); (2) the existence of an efficient secretory pathway to export recombinant products; (3) the low costs of fermentation and (4) the low risk of carrying human pathogenic agents. An additional advantage of this system is the presence of a single (non-secreted) endogenous C1A cysteine protease (in contrast with insect or mammalian cells), which facilitates the analysis and subsequent purification of the heterologous enzyme from culture supernatant. Furthermore, this organism can grow in a wide pH range, both in complex or chemically defined media and using a great variety of carbon and nitrogen sources, allowing optimization of expression conditions to expedite downstream processing of recombinant product. All these combined, can lead to the cost-effective production of correctly folded, safe and fully active recombinant proteins in milligram-to-gram quantities.4,22

Table 2 compares the nucleotide composition of E. coli and P. pastoris genes (codon usage tables) with those from plasmoidal cysteine protease genes. Interestingly, significant differences can be observed with E. coli genes in all the parameters compared, while only moderate divergences were found for P. pastoris. Apparently, both plasmoidal and P. pastoris genes show a similar A/T bias, making them more alike. In contrast, E. coli genes display a modest C/G bias, accounting for bigger differences in nucleotide composition. When compared with E. coli, P. pastoris also showed a more convergent codon usage in relation to plasmoidal genes (Fig. 4). Although in both cases a similar number of unmatched A/T-rich codons were found, the profile obtained for P. pastoris is more flat, representing lesser differences in the relative frequency of codon usage. Several C/G-rich codons, almost absent in plasmoidal genes but frequently used in host genes, were also identified; notably, seven of them in E. coli and only one in P. pastoris, emphasizing the divergences previously described. All together, these elements explain the higher CAI values obtained for plasmoidal genes in P. pastoris over the E. coli host.

In addition, yeast-based systems have been successfully used for the expression of numerous functionally active papain-like cysteine proteases from diverse organisms, including some protozoan parasites.4,22 Remarkably, all these successful attempts included the whole inhibitory prodomain of the enzymes in the expressed construct, in contrast with the failed expression of the mature domain of Falcipain-2 in P. pastoris.23 The absence of the prodomain might be one of the plausible causes of this failure, given that significant divergence in codon usage seems unlikely. Although it was postulated that the prodomain is not strictly necessary for the folding of
the enzyme, it is unclear whether its inclusion could exert a positive influence in the expression levels of the entire protein in this host. As previously described for *E. coli*, the optimization of codon composition and A + T content of the target gene would also be beneficial to improve the resultant expression level further. Other strategies to increase productivity, such as the evaluation of different secretion signals and mutated AOX1 promoter libraries, the generation/detection of multi-copy transformants, optimization of expression conditions (pH, media composition, feed regime, etc.) and the use of protease inhibitors during expression, can be used as part of the vast repertoire of possibilities offered by this versatile expression platform.

**Conclusions**

Thanks to the development of bioengineered organisms, the design of highly-efficient expression systems for cysteine proteases of malaria parasites is not only possible, but closer than ever before. The confluence of (1) rational gene design and synthesis, (2) availability of highly-expressed fusion tags to enhance expression levels, solubility and purification of problematic target genes, (3) tools for the fine genetic and metabolic regulation of protein expression and (4) a variety of bacterial and eukaryotic hosts with significant protein expression and (4) a variety of bac-

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

We would like to thank Drs. Manuel Mansur, Valentina Cartaneo, Martín Blasco and Cohn Berry for helpful discussion and critical reading of the manuscript. This material is based on work partially supported by Grants F/4081-2 and F/4081-2 from Internaciona Foundation for Science (Sweden) and fellowships from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina), Red Latinoamericana de Ciencias Biológicas (RELAB), Universidad de las Naciones Unidas/ Biotecnología para América Latina y el Caribe (UNU-BIOLAC) and Programa Pablo Neruda/Red Iberoamericana de Biotecnología Isla Negra (RIABIN).

**References**

1. Mathews CS, En诚实 M, Lopez AD. Measuring the burden of neglected tropical diseases: the global burden of disease framework. *Hum Negr Trop Dis* 2007; 1:e116, PMID:18000977, http://dx.doi.org/10.1186/1475-2875-1-116.

2. Jia J, Tesarik R, Kita K. Synthesis and evaluation of Plasmadium falciparum drug resistance. *Parasitol Int* 2009; 58:219-52, PMID:19593522, http://dx.doi.org/10.1016/j.parint.2009.04.004.

3. Wiegandt-Gerlach C, Giecher HD, Diederich WE. Proteases of Plasmadium falciparum as potential drug targets and inhibitors (in: Tanev A, Eds). *Curr Top Med Chem* 2010; 10:146-67, PMID:20666957, http://dx.doi.org/10.2174/156802610790725460.

4. Rosenthal PJ. *Falcipain* and other cysteine pro-
tases of malaria parasites. *Adv Exp Med Biol* 2004; 549:57-79, PMID:15037929; http://dx.doi.org/10.1007/1-4020-1622-5_3.

5. Shomai BR, Sijwali PS, Singh AK, Rosenthal PJ. Characterization of native and recombinant *Falcipain-2*, a principal endopeptidase cysteine protease and essential hemoglobinase of Plasmadium falciparum. *J Biol Chem* 2000; 275:29000-10, PMID:10887394; http://dx.doi.org/10.1074/jbc.M000592001.

6. Sijwali PS, Shomai BB, Gia J, Singh A, Rosenthal PJ. Expression and characterization of the Plasmadium falciparum haemoglobinase *Falcipain-3*. *Biochem J* 2001; 362:481-9, PMID:11636777; http://dx.doi.org/10.1042/BJ20010340.

7. Yuan G, Gao X, Ji L, Zhang J, Zhang G, Yao Y, et al. Identification and biochemical characterization of *Falcipain-2*, an amphotoglobin from the malarial parasite Plasmadium falciparum. *J Biol Chem* 2000; 275:25425-37, PMID:10777949; http://dx.doi.org/10.1074/jbc.M603776200.

8. Singh AK, Sijwali PS, Pandey KC, Rosenthal PJ. *Plasmadium falciparum*, biochemical characterization of the cysteine protease *Falcipain-2*. *Exp Parasitol* 2002; 101:113-20, PMID:12057629; http://dx.doi.org/10.1006/etpar.2001.5797.

9. Na HK, Shomai BR, Sijwali PS, Choe Y, Pandey KC, Singh AK, et al. Identification and biochemical characterization of *Falcipain-3*, a cysteine protease of the malaria parasite Plasmadium vivax. *Biochem J* 2004; 388:529-36, PMID:15037929; http://dx.doi.org/10.1042/BJ20031487.
16. Singh A, Thouni BR, Chau V, Goh L, Singh A, et al. Critical role of amino acids 12 and 13 in mediating activity and specificity of stockpains, a papain-family cysteine protease of vector malaria parasite. Biochim J 2002; 368:429-35; PMID:12054365; http://dx.doi.org/10.1002/bip.20048.

17. Singh PS, Thouni BR, Rosenfeld PJ. Folding of the Plasmodium falciparum cysteine proteinase falcipain-2 is mediated by a conserved-like motif, and not the prodomain. J Biol Chem 2002; 277:30938- 5; PMID:12054365; http://dx.doi.org/10.1074/jbc.M204836200.

18. Pu-Yen KC, Wang SX, Singh PS, Loo AL, McKerrow JH, Rosenfeld PJ. The Plasmodium falciparum cysteine proteinase falcipain-2 captures its substrate, hemoglobin, via a unique motif. Proc Natl Acad Sci U S A 2003; 100:5184-9; PMID:14634582; http://dx.doi.org/10.1073/pnas.0306208100.

19. Sardesai ES, Malhotra AC, Tiwé SA, de la Isla A, Chater P. High-level expression of falcipain-2 in Escherichia coli by codon optimization and auto-induction. Protein Expr Purif 2012; 83:58-69; PMID:22450163; http://dx.doi.org/10.1016/j.ypep.2012.03.008.

20. Singh PS, Bhanot LS, Rosenfeld PJ. Systematic optimization of expression and refolding of the Plasmodium falciparum cysteine proteinase falcipain-2. Protein Expr Purif 2001; 22:128-34; PMID:11388810; http://dx.doi.org/10.1006/ypep.2000.1416; e1-2001.0100.

21. Goh SL, Goh LB, Lim TS. Cysteine proteinase falcipain-1 in Plasmodium falciparum is biochemically distinct from its isozymes. Parasitol Res 2005; 97:295-301; PMID:15964982; http://dx.doi.org/10.1007/s00436-005-1430-7.

22. Panes M. High-level expression of Falcipain-2 in Escherichia coli using the Pichia pastoris expression system. Yeast 2006; 23:249-76; PMID:16564221; http://dx.doi.org/10.1002/yea.1208.

23. Nakamura Y, Gomperts T, Barrow T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res 2000; 28:293-296; PMID:10806429; http://dx.doi.org/10.1093/nar/28.1.293.

24. Goh LB, Lim TS. Characterization of aspartic acid variation at strategic positions in parasite and human proteinase inhibitors: selective inhibition of falcipains in Plasmodium falciparum. Biochem Biophys Res Commun 2005; 335:762-70; PMID:15997962; http://dx.doi.org/10.1016/j.bbrc.2005.07.147.

25. Kanju JF. Effects of rare codon clusters on high-level expression of heterologous proteins in E. coli. Curr Opin Biotechnol 1995; 6:494-500; PMID:7797366; http://dx.doi.org/10.1016/0958-1669(95)90109-6.

26. Vyas S, Vyas AK. E-CAI: effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems. Infect Immun 2003; 71:490-9; PMID:12913884; http://dx.doi.org/10.1128/IAI.71.9.4961-4969.2003.

27. Peragallo R, Benito H, García-Vidal M. E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). BMC Biotechnol 2008; 8:65; PMID:18236346; http://dx.doi.org/10.1186/1472-6750-8-65.

28. Bruneau D, Nallanthi TS, Toth B. Production and activity of recombinant papain-like cysteine protease. Methods 2004; 32:199-206; PMID:14489863; http://dx.doi.org/10.1016/j.ymeth.2004.06.003.

29. Muzafar-Patrick S, Fadul EA, McNeil B, Harvery LM. Heterologous protein production using the Pichia pastoris expression system. Yeast 2005; 22:249-76; PMID:15884221; http://dx.doi.org/10.1002/yea.1062.

30. Nakamura Y, Gomperts T, Barrow T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res 2000; 28:293-296; PMID:10806429; http://dx.doi.org/10.1093/nar/28.1.293.