Contents lists available at ScienceDirect

Asian Pacific Journal of Tropical Medicine

journal homepage:www.elsevier.com/locate/apjtm

Document heading doi: 10.1016/S1995-7645(14)60018-3

A unique insertion of low complexity amino acid sequence underlies protein-protein interaction in human malaria parasite orotate phosphoribosyltransferase and orotidine 5′-monophosphate decarboxylase

Waranya Imprasittichai[1*], Sittiruk Roytrakul[2], Sudaratana R. Krungkrai[3], Jerapan Krungkrai[1**]

[1]*Department of Biochemistry, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand*
[2]*National Center for Genetic Engineering and Biotechnology, Pathumthani 12120, Thailand*
[3]*Unit of Biochemistry, Department of Medical Science, Faculty of Science, Rangsit University, Pathumthani 12000, Thailand*

ARTICLE INFO

Article history:
Received 10 August 2013
Received in revised form 15 September 2013
Accepted 15 January 2014
Available online 20 March 2014

Keywords:
Malaria
Plasmodium falciparum
Pyrimidine biosynthesis
Orotate phosphoribosyltransferase
Orotidine 5′-monophosphate decarboxylase
Multienzyme complex
Proteomics

ABSTRACT

Objective: To investigate the multienzyme complex formation of human malaria parasite *Plasmodium falciparum* (*P. falciparum*) orotate phosphoribosyltransferase (OPRT) and orotidine 5′-monophosphate decarboxylase (OMPDC), the fifth and sixth enzyme of the *de novo* pyrimidine biosynthetic pathway. Previously, we have clearly established that the two enzymes in the malaria parasite exist physically as a heterotetrameric $(OPRT)_2(OMPDC)_2$ complex containing two subunits each of OPRT and OMPDC, and that the complex have catalytic kinetic advantages over the monofunctional enzyme. **Methods:** Both enzymes were cloned and expressed as recombinant proteins. The protein-protein interaction in the enzyme complex was identified using bifunctional chemical cross-linker, liquid chromatography-mass spectrometric analysis and homology modeling. **Results:** The unique insertions of low complexity region at the α2 and α5 helices of the parasite OMPDC, characterized by single amino acid repeat sequence which was not found in homologous proteins from other organisms, was located on the OPRT-OMPDC interface. The structural models for the protein-protein interaction of the heterotetrameric $(OPRT)_2(OMPDC)_2$ multienzyme complex were proposed. **Conclusions:** Based on the proteomic data and structural modeling, it is surmised that the human malaria parasite low complexity region is responsible for the OPRT-OMPDC interaction. The structural complex of the parasite enzymes, thus, represents an efficient functional kinetic advantage, which in line with co-localization principles of evolutional origin, and allosteric control in protein-protein-interactions.

1. Introduction

Malaria remains a major and growing threat to the public health, especially, in the endemic areas[1]. One of the five parasite species that infect humans, *Plasmodium falciparum* (*P. falciparum*) is the most lethal[2]. It is estimated that the disease afflicts 450 million and kills 1.5-2.7 million people each year[3]. Chemotherapy of malaria is available but is complicated by antimalarial drug resistance, including the first-line drug artemisinin[4], and their toxicity[5]. Present control measures highlight the priority for more effective and less toxic new drugs with different mechanism of action[6,7].

Pyrimidine nucleotides are essential metabolites. Unlike human and other mammalian cells[8], *P. falciparum* cannot salvage preformed pyrimidine bases or nucleosides from human host, but relies solely on nucleotides synthesized

*Inter-Department Program of Biomedical Science, Faculty of Graduate School, Chulalongkorn University.
**Corresponding authors: Professor Jerapan Krungkrai, Department of Biochemistry, Faculty of Medicine, Chulalongkorn University, Rama 4 Rd., Bangkok 10330, Thailand.
Tel: +662-2564482
Fax: +662-2524986
E-mail: jerapan.k@chula.ac.th; jerapan.k@gmail.com
Foundation project: The study was supported in part by Faculty of Graduate School (to W.I.) and Faculty of Medicine (contract no. RA11/54(1) to J.K.), Chulalongkorn University.

through the *de novo* pathway[9]. The pathway involves six sequential enzymes catalyzing the conversion of following precursors: HCO_3^-, ATP, Glu, Asp, and 5-phosphoribosyl-1-pyrophosphate (PRPP), to form uridine 5′-monophosphate (UMP) (Figure 1). There are some key differences in the pathway between malaria parasites and higher eukaryotes, including human host, for instance, the parasite enzyme inhibitors are found to have strong antimalarial activity for *P. falciparum in vitro* and *Plasmodium berghei* (*P. berghei*) propagated in mice[10-20]. The final two steps of UMP synthesis require the addition of ribose 5-phosphate from PRPP to orotate by orotate phosphoribosyltransferase (EC 2.4.2.10, OPRT) to form orotidine 5′-monophosphate (OMP) and the subsequent decarboxylation of OMP to form UMP by OMP decarboxylase (EC 4.1.1.23, OMPDC). In humans, both of these enzymes exist as a bifunctional UMP synthase[8].

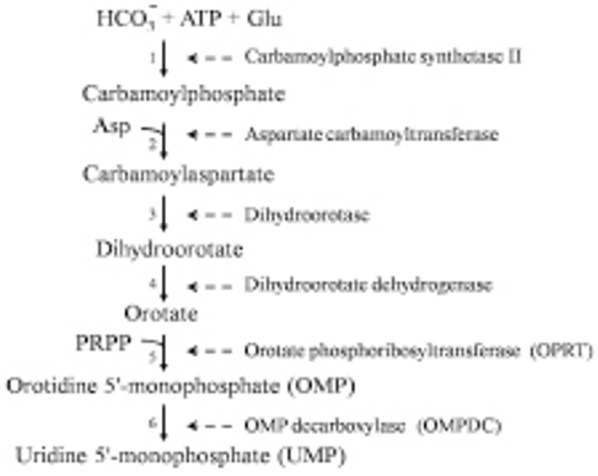

Figure 1. Six sequential steps/enzymes in pyrimidine biosynthetic pathway.

The fifth and sixth enzymes in *P. falciparum* form the heterotetrameric $(PfOPRT)_2(PfOMPDC)_2$ complex, while the bifunctional UMP synthase exert these enzymatic actions in human cell. PRPP stands for 5-phosphoribosyl-1-pyrophosphate.

Currently, we have shown that *P. falciparum* OPRT and OMPDC exist physically as a heterotetrameric $(PfOPRT)_2(PfOMPDC)_2$ enzyme complex containing two subunits each of *Pf*OPRT and *Pf*OMPDC, expressed and encoded by two separate genes[21,22]. The enzyme complex has a catalytic kinetic advantage compared to the monofunctional enzyme[23,24]. In addition, both enzymes have unique insertions of amino acid sequences[16,22], classified as low complexity region (LCR) by having single amino acid repeat sequence which was found in all *Plasmodium* species, but not in homologous proteins from other organisms[25]. The biochemical function/relevance of the LCR is, however, not understood[26].

Here, we further identified protein-protein interfaces in the heterotetrameric $(PfOPRT)_2(PfOMPDC)_2$ complex of the parasite by bifunctional chemical cross-linker, liquid chromatography-mass spectrometric analysis and homology modeling. We established that the unique insertion LCR of the *Pf*OMPDC is responsible for the OPRT-OMPDC interaction, which is the first report for the *Plasmodium* species.

2. Materials and methods

2.1. Construction of E. coli expression plasmid

*Pf*OPRT gene (846 bp) was cloned into pQE30Xa vector (Qiagen) using competent *Escherichia coli (E. coli)* M15[pREP4] strain. The *E. coli* cells were grown in LB medium containing ampicillin and kanamycin[22]. *Pf*OMPDC gene (969 bp) was cloned into pTrcHisA vector (Invitrogen) using *E. coli* TOP10 strain and propagated in LB medium containing only ampicillin[22].

2.2. Protein expression and purification of recombinant PfOPRT and PfOMPDC

The *E. coli* cells harboring the constructed plasmids described above were cultivated in their corresponding LB media at 37 ℃ until the cell density at OD_{600} nm was 0.5. They were then induced with 1 mM isopropyl β-D-thiogalactopyranoside for 18 h at 18 ℃. The recombinant *Pf*OPRT and *Pf*OMPDC proteins were purified by using two sequential steps of a nickel-nitrilotriacetic acid (Ni-NTA) metal-affinity chromatography (Qiagen) and a superose 12 gel filtration fast protein liquid chromatography (FPLC) (Amersham Biosciences) as previously described[16,22].

2.3. Protein and enzymatic assay

Protein concentrations were determined by the Bradford assay[27] with bovine serum albumin as standard. *Pf*OPRT and *Pf*OMPDC activities were measured at 37 ℃ in quartz cuvette using a UV-visible Shimadzu spectrophotometer model UV 1601 with a temperature control device[21].

2.4. Chemical cross-linking of PfOPRT and PfOMPDC

In vitro formation of heterotetrameric $(PfOPRT)_2(PfOMPDC)_2$ complex or dimeric forms of either $(PfOPRT)_2$ or $(PfOMPDC)_2$ were carried out by bifunctional chemical cross-linking using 3,3′-dimethyl suberimidate (DMS) which linked peptides having the nearest neighbors of Lys residues at $-NH_2$ groups to yield amidine derivative of the peptides with a maximum link distance of 1.0 nm[22].

2.5. SDS-Polyacrylamide gel electrophoresis

SDS-PAGE was performed on a Bio-Rad minislab gel apparatus with 5% stacking and varying acrylamide percentages of running gels[28]. Gels were stained with Coomassie Brilliant Blue R-250 dye.

2.6. In-gel digestion

Visualized protein bands in the SDS-PAGE gel were cut into small pieces and destained extensively with 50 mM NH_4HCO_3 followed by methanol prior to washing with deionized water. Chopped gels were dehydrated in acetonitrile and reduced with 10 mM dithiothreitol in 10 mM NH_4HCO_3 for 1 h at 56 ℃. Reduced gels were then alkylated by 100 mM iodoacetamide in 10 mM NH_4HCO_3 for 1 h at 25 ℃ in dark. Alkylated gels were further digested with 10 ng/μL sequencing grade trypsin in 10 mM NH_4HCO_3 for 3 h at 37 ℃. Trypsin digested peptides were eluted from the gels with 0.1% formic acid/50% acetonitrile, evaporated at 40 ℃ overnight, and then stored at -20 ℃ until use[29].

2.7. Liquid chromatography-mass spectrometry and proteomic data analysis

Nanoscale LC separation of tryptic peptides was performed with a NanoAcquity system (Waters Corp.) equipped with a symmetry C_{18} 5 μm, 180 μm × 20 mm Trap column and a BEH130 C_{18} 1.7 μm, 100 μm × 100 mm analytical reversed-phase column (Waters Corp.). The samples were initially transferred with 0.1% formic acid to the trap column with a flow rate of 15 μL/min for 1 min. Mobile phase A was 0.1% formic acid in water, and mobile phase B was 0.1% formic acid in acetonitrile. The peptides were separated with a gradient of 15%-50% mobile phase B over 15 min at a flow rate of 500 nL/min followed by a 3 min rinse with 80% of mobile phase B. The column temperature was maintained at 35 ℃. The lock mass was delivered from the auxiliary pump of the NanoAcquity pump with a constant flow rate of 500 nL/min at a concentration of 200 fmol/μL of [Glu^1]fibrinopeptide B to the reference sprayer of the NanoLockSpray source of the mass spectrometer.

Analysis of tryptic peptides was performed using a SYNAPT™ HDMS mass spectrometer (Waters Corp.). For all measurements, the mass spectrometer was operated in the V-mode of analysis with a resolution of at least 10 000 full-width half-maximum. All analyses were performed using positive nanoelectrospray ion mode. The time-of-flight (TOF) analyzer of the mass spectrometer was externally calibrated with [Glu^1]fibrinopeptide B from m/z 50 to 1 600 with acquisition lock mass corrected using the monoisotopic mass of the doubly charged precursor of [Glu^1]fibrinopeptide B. The reference sprayer was sampled with a frequency of 20 s. Accurate mass LC-MS data were acquired with data direct acquisition mode. The trap collision energy was 6 V, transfer collision energy was 4 V. The quadrupole mass analyzer was adjusted such that ions from m/z 300 to 1 800 were efficiently transmitted. The MS/MS was recorded in the range of 50 to 1 990 Da with a scan time of 0.5 s/scan. The MassLynx version 2.3 software was applied to analyze the LC-MS/MS data.

2.8. Homology modeling of *Pf*OPRT and *Pf*OMPDC in monomer, dimer and heterotetramer

The three-dimensional (3D) structures of *Pf*OPRT (residues 64-281) and *Pf*OMPDC (residues 1-323) were, respectively, determined by using crystal structures of *Saccharomyces cerevisiae* (S. *cerevisiae*) OPRT (PDB code 2PS1, chain A) and *P. falciparum* 3D7 OMPDC (PDB code 2Q8Z, chain A) as template. Homology models of the structures were generated by the Phyre program[30]. Homology search and multiple alignment of amino acid sequences were performed using the BLAST[31] and the CLUSTALW[32] programs, respectively. After the binding sequences identified by LC-MS/MS data, models for protein-protein interactions of dimeric (*Pf*OPRT)$_2$ and (*Pf*OMPDC)$_2$, and heterotetrameric (*Pf*OPRT)$_2$(*Pf*OMPDC)$_2$ were generated. The constructed dimeric models were stereo rotated with the amino acid sequences used for binding between chain A and chain B of *Pf*OPRT and *Pf*OMPDC. The dimeric models were then applied as the template to generate heterotetrameric model by further stereo rotating with the amino acid sequences identified for *Pf*OPRT-*Pf*OMPDC interactions.

3. Results

3.1. Chemical cross-linking of *Pf*OPRT and *Pf*OMPDC

The recombinant *Pf*OPRT and *Pf*OMPDC proteins expressed in *E. coli* were purified by two sequential steps of a Ni-NTA-affinity chromatography and a superose 12 gel filtration column on FPLC system. The recombinant *Pf*OPRT and *Pf*OMPDC were catalytically active in its dimeric form and had specific activities of 4.7 and 7.9 μmol/min per mg protein, respectively. Using this strategy, we obtained 3-4 mg each of nearly 95% pure and active *Pf*OPRT and *Pf*OMPDC from 1 L of cell culture. By gel filtration FPLC and SDS-PAGE, the *Pf*OPRT was a 67 kDa homodimer of two 33 kDa (*Pf*OPRT)$_2$ subunits (Figure 2A), whereas the

*Pf*OMPDC showed a monomeric subunit of 38 kDa (Figure 2B) and a 76 kDa homodimeric $(PfOMPDC)_2$. To prepare chemical cross-linked products of *Pf*OPRT and/or *Pf*OMPDC, the pure enzyme was incubated at 25 ℃ with DMS at a ratio of 1:2 protein to cross-linker[22]. By SDS-PAGE, monomers and dimers of $(PfOPRT)_2$ and $(PfOMPDC)_2$ were detected. Chemical cross-linked heterotetrameric form of *Pf*OPRT and *Pf*OMPDC was achieved by further DMS treatment at a ratio of 1:2 protein to cross-linker. After 30 min incubation, more than 50% of the proteins were cross-linked as tetrameric form (Figure 2C). The time-dependent oligomerization of each enzyme corresponded to that expected for sequential cross-linking of monomer to form dimer, and each dimer to form heterotetramer. The molecular mass of each oligomer was determined as previously described[22].

Table 1

LC-MS/MS data analysis of *P. falciparum* orotate phosphoribosyltransferase in monomeric, dimeric forms and in heterotetrameric complex with orotidine 5′-monophosphate decarboxylase.

Peptide	Peptide score [a]	MH+ (Da)	Monomeric OPRT	Dimeric OPRT	Tetrameric OPRT-OMPDC
ENEFLCDEEIYK	34.340 000	1 032.689 1	5.811 603	10.671 790	7.537 772
EYGDKNVIVGNLDDDDKDILNLKK	15.600 000	2 713.714 6	8.755 316	–	5.822 202
GIPMVSLTSHFLFESK	85.930 000	1 504.672 6	7.166 355	6.090 981	6.881 962
IYFKDIFEK	44.310 001	1 194.595 1	7.241 354	5.591 504	6.965 446
KNIIIIDDVFTCGTALTEILAK	71.029 999	2 461.818 3	7.857 266	4.154 335	–
KYSNIFYLYDR	68.019 997	1 089.601 6	10.809 390	7.890 219	7.214 506
LSFDYLLGASYK	60.650 002	1 375.949 4	7.222 945	3.474 503	7.305 166
NVIVGNLDDDDK	50.020 000	1 154.303 4	5.289 937	5.497 626	6.619 940
NVIVGNLDDDDKDILNLK	103.830 000	2 013.217 9	14.657 430	4.562 683	5.403 011
NVIVGNLDDDDKDILNLKK	35.790 001	2 122.657 7	9.319 596	–	6.336 068
SFVHLK	8.360 000	499.332 6	7.925 779	6.753 919	8.925 819
VGIPLYSILSYK	61.270 000	1 183.736 7	4.989 854	6.546 392	6.391 722
VVAFIVLLNR	47.689 999	1 144.144 1	6.420 709	9.936 290	6.922 265
VVAFIVLLNRNEYEINENNQK	33.000 000	2 520.391 1	9.863 172	4.334 970	4.282 399
YSNIFYLYDR	73.480 003	1 353.232 0	13.433 300	10.828 940	9.282 876
YSNIFYLYDRK	26.860 001	1 481.522 2	4.393 924	9.054 119	6.417 926

[a]The results were tested by the Decyder MS program with significance at $P<0.05$.

Table 2

LC-MS/MS data analysis of *P. falciparum* orotidine 5′-monophosphate decarboxylase in monomeric, dimeric forms and in heterotetrameric complex with orotate phosphoribosyltransferase.

Peptide	Peptibe score	MH+ (Da)	Monomeric OMPDC	Dimeric OMPDC	Tetrameric OPRT-OMPDC
AAQMYYDQINAILK	58.610 001	2 210.549 1	13.511 260	13.359 900	12.414 570
APDNIIR	45.570 000	1 000.651 4	8.347 193	7.843 130	7.181 398
APDNIIREEK	41.049 999	1 180.240 4	9.541 573	5.696 492	4.269 348
DICYDEEK	18.820 000	1 065.693 5	10.504 400	4.767 869	9.609 203
DICYDEEKNK	21.639 999	1 313.594 6	6.805 417	–	3.950 628
DILLKAPDNIIR	40.070 000	1 374.595 9	5.955 823	7.277 666	6.832 736
DILLKAPDNIIREEK	9.010 000	1 766.359 0	7.218 944	8.964 727	5.962 379
FIFEYLK	22.330 000	4 811.505 1	10.328 490	7.360 944	–
ILINIGR	26.360 001	800.088 8	7.875 441	5.720 076	5.588 307
INDIGNTVK	36.799 999	976.666 6	8.848 516	6.200 231	3.874 776
KFIFEYLK	25.840 000	1 089.976 7	9.965 045	5.086 093	6.271 910
MNFAFYIPYGSVGIDVLK	46.419 998	1 422.111 6	6.851 215	5.939 178	4.606 118
NAINTCLCIGLDPDEK	80.889 999	1 833.055 4	6.117 382	9.577 628	–
NAINTCLCIGLDPDEKDIENFMK	24.330 000	2 139.558 1	6.240 061	5.302 535	5.325 563
NVFDYLYELNIPTILDMK	62.799 999	2 217.121 0	14.356 230	10.730 460	–
SDSCTVNIYMGTNMLK	41.169 998	1 388.550 8	5.956 773	4.115 105	4.277 855
SEEFFYFFNHFCFYIINETNK	29.080 000	2 047.242 8	6.717 936	5.769 700	–
TTNPDSAIFQK	59.430 000	1 221.389 4	14.289 200	13.020 870	6.978 079
TTNPDSAIFQKNLSLDNK	25.290 001	1 985.607 7	8.036 843	7.232 890	6.672 452
TYFPNCYILSPGIGAQNGDLHK	41.490 002	2 463.589 0	8.362 220	11.827 000	5.371 594
YALTFK	20.879 999	3 332.587 9	12.878 050	5.566 563	8.408 722
YINNVSIK	23.230 000	4 797.700 2	8.022 080	6.701 042	–

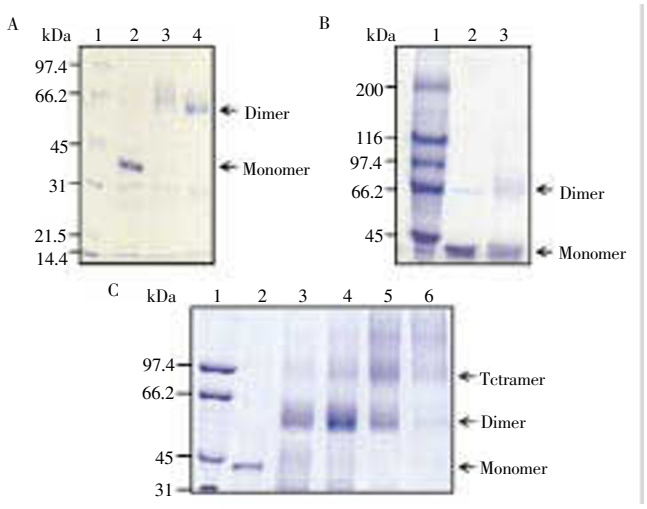

Figure 2. SDS-PAGE analysis of monomeric, dimeric and tetrameric forms in *P. falciparum* OPRT and OMPDC.

The oligomers of *Pf*OPRT and/or *Pf*OMPDC, which were cross-linked by reacting with 3,3′-dimethyl suberimidate, were identified on 7%-12% polyacrylamide gels. (A) *Pf*OPRT cross-linked dimeric form on 12% gel, lane 1: molecular mass standard, lane 2: the non-cross-linking monomeric form (10 μg protein), lanes 3, 4: cross-linking products of dimeric form after 15 and 30 min reactions. Monomer on lane 2 and dimeric product on lane 4 gels were excised for LC-MS/MS analysis. (B) *Pf*OMPDC chemically cross-linked dimeric form on 7% gel, lane 1: molecular mass standard, lane 2: the non-cross-linking monomeric form (20 μg protein), lane 3: cross-linking products of the dimeric form after 30 min reaction. Monomer on lane 2 and dimeric product on lane 3 gels were excised for LC-MS/MS analysis. (C) *Pf*OPRT and *Pf*OMPDC cross-linked tetrameric form on 10% gel, lane 1: molecular mass standard, lane 2: the non-cross-linking monomeric form (10 μg protein each), lanes 3-6: cross-linking products of tetrameric form after 10, 20, 30 and 60 min reactions. Tetrameric product on lane 4 gel was excised for LC-MS/MS analysis. The apparent monomer, dimer, and tetramer have log molecular mass of 4.58, 4.85, and 5.15, respectively.

3.2. Liquid chromatography-mass spectrometry of chemical cross-linking products

The non cross-linked monomeric forms, cross-linked dimeric and heterotetrameric forms on the SDS-PAGE gel were subjected to in-gel trypsin digestion. The entire peptide samples were injected into the HPLC Q-TOF, and the LC-MS/MS data were analyzed by the MassLynx software and summarized in Table 1 for *Pf*OPRT and Table 2 for *Pf*OMPDC. The results showed the amino acid sequences of *Pf*OPRT and *Pf*OMPDC obtained as monomer, dimer, and heterotetramer forms. Combining the chemical cross-linking and LC-MS/MS led us to identify amino acid sequences that play role in the binding between domains and/or polypeptide chain of individual and both enzymes. The blank spaces in the Tables 1 and 2 represent the cross-linked peptides/sequences. Amino acid sequences that were present in the monomeric form but was not in the dimeric or heterotetarmeric form confirms that this sequence underlies domain-domain interface and contains amino acid residues used for binding between domains.

3.3. Homology models and proteomic data analysis

As for the homology model, the Phyre program was utilized to construct models of *Pf*OPRT and *Pf*OMPDC. The model of *Pf*OPRT had no *N*-terminal domain (residues 1-63) which has the unique insertion LCR sequence (Figure 3). This *N*-terminal sequence contained one α-helix (residues 13-36). Sequence alignment between *Pf*OPRT and *S. cerevisiae* OPRT showed 30% identity. The most striking finding in all *Plasmodium* OPRTs were two large insertions; one, at the *N*-terminal region as described above and, another, at the internal region (residues 178-196) with a high hydrophobic index of +1 to +3[16].

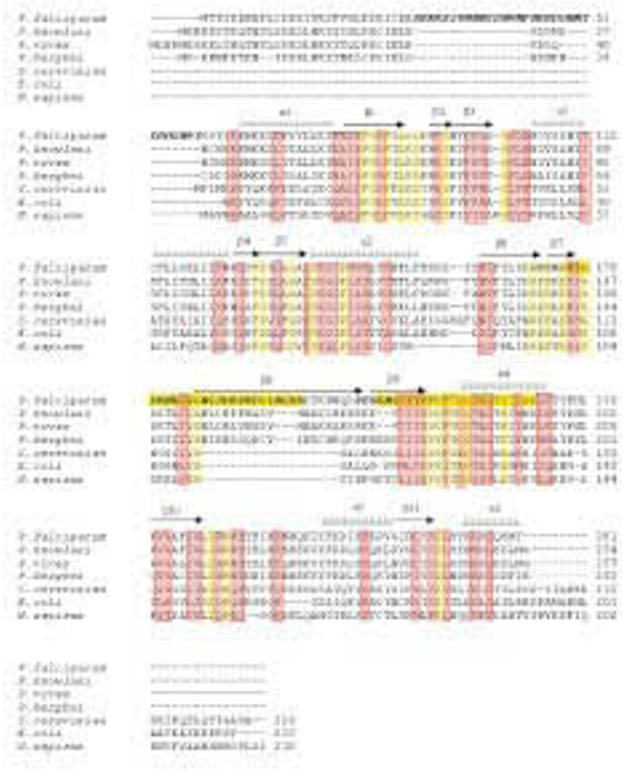

Figure 3. Sequence alignment of OPRTs.

Sequence alignment of seven different OPRTs was performed using CLUSTALW[32]. Species name and accession numbers are as follows: *P. falciparum* OPRT (*PF*E0630C), *Plasmodium knowlesi(P. knowlesi)* (XP_002259686.1), *Plasmodium vivax (P. vivax)* (XP_001613829), *P. berghei* (*PB*ANKA_111240), *S. cerevisiae* (NP_013601), *E. coli* (X00781) and *Homo sapiens (H. sapiens)* (NP_000364). Conserved identical residues are highlighted in yellow, similar residues are shown in pink. In *P. falciparum* sequence, LCRs are bold and italic letters, cross-linking peptides are shown with yellow drag, sequences 168-191 and 202-223 are responsible for homodimeric and heterotetrameric formation, respectively. The Phyre program was used to predict α-helix and β-strand in the *N*-terminal sequence (residues 1-63), having one α-helix (residues 13-36) and no β-strand[30].

Based on the sequence alignments of OMPDCs from available *Plasmodium* species, the model for *Pf*OMPDC using the full-length sequence contained a large unique insertion LCR in the α2, α3, α4 and α5 helices at the *N*-terminal domain which were found in all *Plasmodium* OMPDCs (Figure 4)[22].

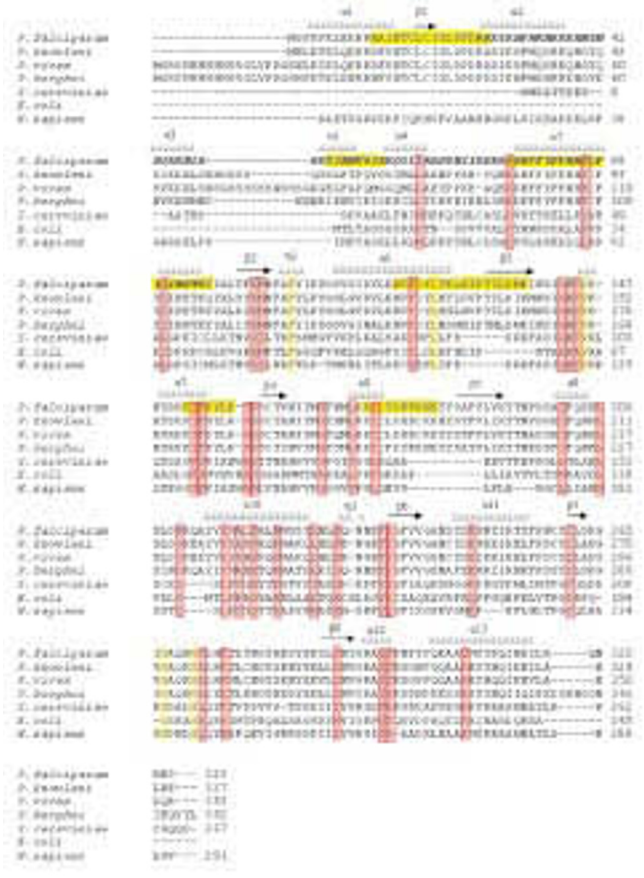

Figure 4. Sequence alignment of OMPDCs.

Sequence alignment of seven different OMPDCs was performed using CLUSTALW[32]. Species name, accession numbers and PDB ID are as follows: *P. falciparum* OMPDC (2ZA2), *P. knowlesi* (XP_002261749.1), *P. vivax* (2FFC), *P. berghei* (2FDS), *S. cerevisiae* (1DQW), *E. coli* (1L2U) and *H. sapiens* (2EAW). Conserved identical residues are highlighted in yellow, similar residues are shown in pink. In *P. falciparum* sequence, LCRs are bold and italic letters, cross-linking peptides are shown with yellow drag, sequence 175-184 is used for homodimeric formation, sequences 12-27, 52-59, 76-96, 121-138 and 152-158 are responsible for heterotetrameric formation.

The monomeric *Pf*OPRT model consisted of 6 α-helices and 11 β-strands (Figure 5A), while the monomeric *Pf*OMPDC structure model exhibited 13 α-helices and 8 β-strands (Figure 5B). Both enzymatic models highlight the LCR sequences that conserve amongst *Plasmodium* species, and participate in homodimeric and heterotetrameric formation.

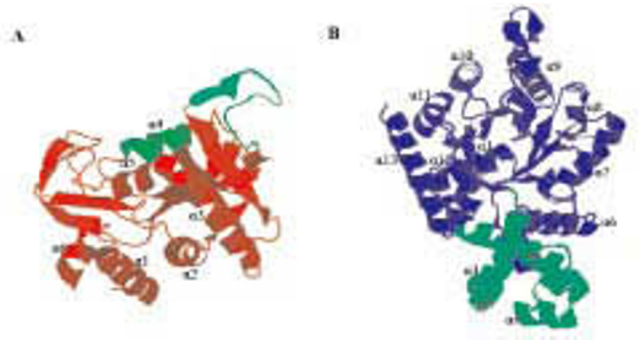

Figure 5. Phyre homology model of monomeric *P. falciparum* OPRT and OMPDC.

The Phyre program was used to predict α-helix and β-strand[30]. The α-helix numberings are shown in the built models, while the β-strands are not shown here but details can be found in the sequence alignments for *Pf*OPRT (Figure 3) and *Pf*OMPDC (Figure 4). (A) *Pf*OPRT model (64-281 residues) is depicted with the 1-63 residues omitted from the Phyre program due to its unique insertion sequence LCR with very low identity to other OPRTs. The model shows LCRs, α4 and α8 that used for dimeric and tetrameric formation (green ribbon α-helix and β-strand). The active site was identified using Swiss model as previously described[45]. (B) *Pf*OMPDC model (1-323 residues) is a completed full-length sequence in which its N-terminus harboring α2, α3, α4, and α5-helices as unique insertions (LCRs) to the enzyme, and they play role in dimeric and tetrameric formation (green ribbon α-helices). The active site was identified based on our previously described crystal structure[35].

The LC-MS/MS data of *Pf*OPRT (Table 1) identified the sequence 168-EYGDKNVIVGNLDDDDKDILNLKK-191 and 173-NVIVGNLDDDDKDILNLKK-191 containing amino acids responsible for β7 strand (residues 165-169), in β turn at β7-β8 position (residues 170-176) and β8 strand (residues 177-199), which presumably locate at domain-domain interface and bind between chain A and chain B of *Pf*OPRT (Figure 6A). Interestingly, this binding sequence is a main part of the internal unique insertion LCR (residues 178-196) of *Pf*OPRT (highlighted in Figure 3)[16]. In contrast, the LC-MS/MS data of *Pf*OMPDC (Table 2) showed the sequence 175-DICYDEEKNK-184, which was not a LCR insertion sequence and located in β8 helix (residues 173-176) (highlighted in Figure 4)[22], as the one for binding between chain A and chain B of *Pf*OMPDC (Figure 6B).

Furthermore, the LC-MS/MS data of heterotetrameric $(PfOPRT)_2$ $(PfOMPDC)_2$ complex (Tables 1 and 2) indicated that the sequence 202-KNIIIIDDVFTCGTALTEILAK-223, spanning in α4 helix (residues 214-225) of *Pf*OPRT (Figure 6A and 6C), was used to bind between the dimeric $(PfOPRT)_2$ and the dimeric $(PfOMPDC)_2$ at five positions as follows: 12-NAINTCLCIGLDPDEK-27 in α2 helix; 52-YINNVSIK-59 in β turn at α3-α4 position; 76-SEEFFYFFNHFCFYIINETNK-96 in α5 helix; 121-NVFDYLYELNIPTILDMK-138 in α6 helix; and 152-FIFEYLK-158, partially covering in α7 helix (Figure 6B and 6C) of *Pf*OMPDC. All peptides involve in the tetrameric

formation as described above are highlighted with yellow color on *P. falciparum* amino acid sequences (Figures 3 and 4). In addition, the peptide 12-NAINTCLCIGLDPDEK-27 (Table 2) was cross-linked using K27 to the peptide 202-KNIIIIDDVFTCGTALTEILAK-223 of *Pf*OPRT (Table 1) to form heterotetrameric $(PfOPRT)_2(PfOMPDC)_2$, whereas the peptide 12-NAINTCLCIGLDPDEKDIENFMK-34 of *Pf*OMPDC (Table 2) was not cross-linked with DMS and it was then digested with trypsin when the enzyme stayed in its dimeric form. These results suggest that the conformational change occurs during heterotetrameric association of dimeric $(PfOPRT)_2$ and $(PfOMPDC)_2$ (Figure 6 C).

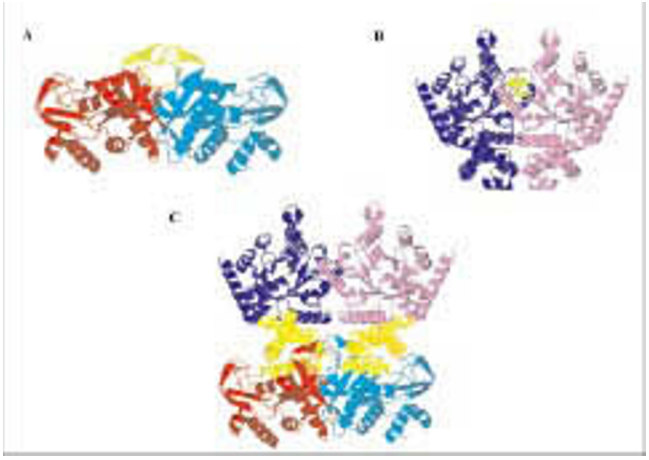

Figure 6. Proposed homology model of *P. falciparum* OPRT and OMPDC in dimeric and heterotetrameric forms.
The models were constructed by the Phyre program[30]. (A) Model of dimeric $(PfOPRT)_2$. (B) Model of dimeric $(PfOMPDC)_2$. (C) Model of heterotetrameric $(PfOPRT)_2(PfOMPDC)_2$ complex. In A and B, light green ribbons represent the cross-linking and interacting sites for dimeric $(PfOPRT)_2$ and dimeric $(PfOMPDC)_2$, respectively. In C, yellow ribbons indicate locations of protein-protein interface and cross-linking sequences for the heterotetrameric $(PfOPRT)_2(PfOMPDC)_2$ formation.

It is noted here that, the sequence 202-223 (α 4 helix) of *Pf*OPRT and 76-96 (α 5 helix) of *Pf*OMPDC were only located in the domain-domain interface of the heterotetrameric $(PfOPRT)_2(PfOMPDC)_2$. Indeed, the α 2 and α 5 helices, responsible for domain-domain interactions, are unique insertion LCR sequences located at the *N*-terminal extension of *Pf*OMPDC[22].

Based on the identified binding sequences by LC-MS/MS data, the structural models for protein-protein interactions of dimeric $(PfOPRT)_2$ and $(PfOMPDC)_2$, and heterotetrameric $(PfOPRT)_2(PfOMPDC)_2$ are then proposed and shown in Figure 6.

4. Discussion

Recent findings have highlighted the potential vulnerability of the human parasite toward agents which affect the pyrimidine biosynthetic pathway[12-20]. However, at present only three enzymes have crystal structures and crystallographic analyses: the fourth[17], fifth (*Pf*OPRT)[33], and sixth (*Pf*OMPDC)[34,35] enzymes of the *P. falciparum* pathway. Much importance has been given to these enzymes since compounds presumed to affect the enzyme activities have been shown to be selectively toxic toward the human parasite[17,18,20]. In humans and most eukaryotes, the genes encoding OPRT and OMPDC are fused into OPRT-OMPDC and expressed as bifunctional UMP synthase[36,37], although an inversely linked OMPDC-OPRT gene, where the OMPDC is located at *N*-terminus and OPRT at *C*-terminus, have been reported in the trypanosomatid *Leishmania donovani*[38]. With one exception, both genes in the human malarial parasites are acquired from eubacterial origins, are located on different chromosomes, and expressed as a heterotetrameric $(PfOPRT)_2(PfOMPDC)_2$ complex with in vivo concentration of 30-46 nM, relatively close to the UMP synthase concentration (17-32 nM) in human cells[16,21,22]. The structural complex of the parasite enzyme, thus, represents an efficient functional kinetic advantage, which is in line with co-localization principles for evolutional origin, and allosteric control in protein-protein interactions[39,40].

We have, for the first time, examined the protein-protein interactions of the heterotetrameric $(PfOPRT)_2(PfOMPDC)_2$ complex using chemical cross-linking between the nearest Lys residues, and mapping the sites of the cross-link with the LC-MS/MS analysis, as previously described for protein structural model[41-43]. Aside from establishing the domain-domain interactions between the *Pf*OPRT (α 4 helix) and *Pf*OMPDC (α 2, α 5, α 6, α 7 helices, and β turn at α 3- α 4 position) complex, we have also shown that each *Pf*OPRT (β 7 and β 8 strands) and *Pf*OMPDC (α 8 helix) form a relatively tight dimeric structure. Our Phyre structural model of *Pf*OPRT is comparable to the recent report of the Swiss model using the same template of *S. cerevisiae* OPRT[44], in which the unique insertion of 63 amino acids from *N*-terminus is omitted due to its low identity[45]. Chemical cross-linking and LC-MS/MS results indicate that: 1) the dimer $(PfOPRT)_2$ uses LCR insertion sequences for interaction between chain A and B, 2) the dimer $(PfOMPDC)_2$ is homologous to all known monofunctional OMPDC structures[35,38,46], and 3) the tetramer assembling each dimer of $(PfOPRT)_2$ and $(PfOMPDC)_2$ utilizes both the LCR sequences (α 2 and α 5 helices of *Pf*OMPDC) and the over-represented hydrophobic amino acid sequences with some acidic and basic residues (α 6 and α 7 helices, β turn at α 3- α 4 position of *Pf*OMPDC; α 4 helix of *Pf*OPRT) for the interactions. The α 2 and α 5 helices are located at a protruding domain in the crystal structure of *Pf*OMPDC[35]. In contrast, the bifunctional *Leishmania donovani* and human UMP synthase structural models do not possess any significant domain-domain contacts between OPRT and OMPDC[38]. Hence, the formation of tetrameric $(PfOPRT)_2(PfOMPDC)_2$ is unique in the case of the insert LCR sequence in the *Pf*OMPDC that found in all *Plasmodium* species plays a functional role for domain-domain

interactions. The present investigation clearly supports our recent hypothesis that the unique insertions are responsible for *Pf*OPRT–*Pf*OMPDC assembly formation *in vivo*[22,35].

The origin of LCR in the malarial parasite proteins remains unknown. The insertion LCR sequences and binding sequences identified in both *Pf*OPRT and *Pf*OMPDC contain acid, basic and hydrophobic amino acids, particular aromatic residues. Taken together, it is plausible that the interactions between *Pf*OPRT and *Pf*OMPDC domains apparently favored in the interfaces include mainly the hydrophobic pairs and some opposite charged pairs or salt bridges, which are supported by computational prediction of protein–protein interfaces[47–50]. This conclusion appears to agree well with previous results on hyperthermostability of the $(Pf\text{OPRT})_2(Pf\text{OMPDC})_2$ complex, comparing to the monofunctional form[23].

Since the actual protein interfaces are inferred from X-ray crystallographic structures[51,52], further studies on the crystal structural analyses of the enzyme complex with full length proteins will provide additional information on the binding mechanisms including solvent-inaccessible area, hydrogen bonds, salt bridges, interface residues, and solvation free-energy between the two domains. Use of the interface peptides or inhibitors to disrupt the complex formation can be exploited for a rational drug design approach for more potential antimalarials[53,54].

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgements

The study was supported in part by Faculty of Graduate School (to W.I.) and Faculty of Medicine (contract no. RA11/54(1) to J.K.), Chulalongkorn University. We would like to thank Dr. N.M.Q. Palacpac for critical review of this manuscript. We acknowledge the proteomics staffs at National Center for Genetic Engineering and Biotechnology for technical assistance.

References

[1] Guerin PJ, Olliaro P, Nosten F, Druilhe P, Laxminarayan R, Blinka F, et al. Malaria: current status of control, diagnosis, treatment, and a proposed agenda for research and development. *Lancet Infect Dis* 2002; **2**: 564–573.

[2] Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* 2005; **434**: 214–217.

[3] Hay SI, Okiro EA, Gething PW, Patil AP, Tatem AJ, Guerra CA, et al. Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007. *PloS Med* 2010; **7**: e1000290.

[4] Dondorp AM, Nosten F, Yi P, Das D, Phyo A, Tarning J, et al. Artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med* 2009; **361**: 455–467.

[5] White NJ. Antimalarial drug resistance. *J Clin Invest* 2004; **113**: 1084–1092.

[6] Ridley RG. Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature* 2002; **415**: 686–693.

[7] Krungkrai SR, Krungkrai J. Malaria parasite carbonic anhydrase: inhibition of aromatic/heterocyclic sulfonamides and its therapeutic potential. *Asian Pac J Trop Biomed* 2011; **1**: 233–242.

[8] Jones ME. Pyrimidine nucleotide biosynthesis in animals: genes, enzymes, and regulation of UMP biosynthesis. *Annu Rev Biochem* 1980; **49**: 253–279.

[9] Gero AW, O'Sullivan WJ. Purines and pyrimidines in malarial parasites. *Blood Cells* 1990; **16**: 467–484.

[10] Krungkrai J, Cerami A, Henderson GB. Pyrimidine biosynthesis in parasitic protozoa: purification of a monofunctional dihydroorotase from *Plasmodium berghei* and *Crithidia fasciculata*. *Biochemistry* 1990; **29**: 6270–6275.

[11] Krungkrai J, Cerami A, Henderson GB. Purification and characterization of dihydroorotate dehydrogenase from the rodent malaria parasite *Plasmodium berghei*. *Biochemistry* 1991; **30**: 1934–1939.

[12] Krungkrai J, Krungkrai SR, Phakanont K. Antimalarial activity of orotate analogs that inhibit dihydroorotase and dihydroorotate dehydrogenase. *Biochem Pharmacol* 1992; **43**: 1295–1301.

[13] Seymour KK, Lyons SD, Phillips L, Rieckmann KH, Christopherson RI. Cytotoxic effects of inhibitors of *de novo* pyrimidine biosynthesis upon *Plasmodium falciparum*. *Biochemistry* 1994; **33**: 5268–5274.

[14] Krungkrai J. Purification, characterization and localization of mitochondrial dihydroorotate dehydrogenase in *Plasmodium falciparum*, human malaria parasite. *Biochim Biophys Acta* 1995; **1243**: 351–360.

[15] Flores MVC, Atkins D, Wade D, O'Sullivan WJ, Stewart TS. Inhibition of *Plasmodium falciparum* proliferation *in vitro* by ribozymes. *J Biol Chem* 1997; **272**: 16940–16945.

[16] Krungkrai SR, Aoki S, Palacpac NMQ, Sato D, Mitamura T, Krungkrai J, et al. Human malaria parasite orotate phosphoribosyltransferase: functional expression, characterization of kinetic reaction mechanism and inhibition profile. *Mol Biochem Parasitol* 2004; **134**: 245–255.

[17] Hurt DE, Widom J, Clardy J. Structure of *Plasmodium falciparum* dihydroorotate dehydrogenase with a bound inhibitor. *Acta Cryst* 2006; **D62**: 312–323.

[18] Meza-Avina ME, Wei L, Buhendwa MG, Poduch E, Pai EF, Kotra LP. Inhibition of orotidine 5′-monophosphate decarboxylase and its therapeutic potential. *MiniRev Med Chem* 2008; **8**: 239–247.

[19] Krungkrai SR, Wutipraditkul N, Krungkrai J. Dihydroorotase of human malarial parasite *Plasmodium falciparum* differs from host enzyme. *Biochem Biophys Res Commun* 2008; **366**: 821–826.

[20] Abdo M, Zhang Y, Schramm VL, Knapp S. Electrophilic aromatic selenylation: new OPRT inhibitors. *Org Lett* 2010; **12**: 2982–2985.

[21] Krungkrai SR, Prapunwattana P, Horii T, Krungkrai J. Orotate phosphoribosyltransferase and orotidine 5′-monophosphate decarboxylase exist as multienzyme complex in human malaria

parasite *Plasmodium falciparum*. *Biochem Biophys Res Commun* 2004; **318**: 1012-1018.

[22]Krungkrai SR, DelFraino BJ, Smiley JA, Prapunwattana P, Mitamura M, Horii T, et al. A novel enzyme complex of orotate phosphoribosyltransferase and orotidine 5′-monophosphate decarboxylase in human malaria parasite *Plasmodium falciparum*: physical association, kinetics and inhibition characterization. *Biochemistry* 2005; **44**: 1643-1652.

[23]Kanchanaphum P, Krungkrai J. Kinetic benefits and thermal stability of orotate phosphoribosyltransferase and orotidine 5′-monophosphate decarboxylase enzyme complex in human malaria parasite *Plasmodium falciparum*. *Biochem Biophys Res Commun* 2009; **390**: 337-341.

[24]Kanchanaphum P, Krungkrai J. Co-expression of human malaria parasite *Plasmodium falciparum* orotate phosphoribosyltransferase and orotidine 5′-monophosphate decarboxylase as enzyme complex in *Escherichia coli*: a novel strategy for drug development. *Asian Biomed* 2010; **4**: 297-306.

[25]Pizzi E, Frontali C. Low-complexity regions in *Plasmodium falciparum* proteins. *Genom Res* 2001; **11**: 218-229.

[26]Frugier M, Bour T, Ayach M, Santos MAS, Rudinger-Thirion J, Theobald-Dietrich A, et al. Low complexity regions behave as tRNA sponges to help co-translational folding of plasmodial proteins. *FEBS Lett* 2011; **584**: 448-454.

[27]Bradford MM. A rapid and sensitive method for the quantitation of micrograms quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 1976; **72**: 248-254.

[28]Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970; **227**: 680-685.

[29]Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols* 2006; **1**: 2856-2860.

[30]Kelley LA, Sternberg MJE. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* 2009; **4**: 363-371.

[31]Altschul SF, Madden TL, Schaffer AA, Zang J, Zang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**: 3389-3402.

[32]Thompson JD, Higgins DG, Gibson TJ. CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**: 4673-4680.

[33]Takashima Y, Mizohata E, Tokuoka K, Krungkrai SR, Kusakari Y, Konoshi S, et al. Crystallization and preliminary X-ray diffraction analysis of orotate phosphoribosyltransferase from human malaria parasite *Plasmodium falciparum*. *Acta Cryst* 2012; **F68**: 244-246.

[34]Krungkrai SR, Kusakari Y, Tokuoka K, Inoue T, Adachi H, Matsumura H, et al. Crystallization and preliminary crystallographic analysis of orotidine 5′-monophosphate decarboxylase from human malaria parasite *Plasmodium falciparum*. *Acta Cryst* 2006; **F62**: 542-545.

[35]Tokuoka K, Kusakari Y, Krungkrai SR, Matsumura H, Krungkrai J, Horii T, et al. Structural basis for the decarboxylation of orotidine 5′-monophosphate (OMP) by *Plasmodium falciparum* OMP decarboxylase. *J Biochem* 2008; **143**: 69-78.

[36]Nara T, Hashimoto T, Aoki T. Evolutionary implications of the mosaic pyrimidine-biosynthetic pathway in eukaryotes. *Gene* 2000; **257**: 209-222.

[37]Makiuchi T, Nara T, Annoura T, Hashimoto T, Aoki T. Occurrence of multiple, independent gene fusion events for the fifth and sixth enzymes of pyrimidine biosynthesis in different eukaryotic groups. *Gene* 2007; **394**: 78-86.

[38]French JB, Soysa DR, Yates PA, Boitz JM, Carter NS, Chang B, et al. The *Leishmania donovani* UMP synthase is essential for promastigote viability and has an unusual tetrameric structure that exhibits substrate-controlled oligomerization. *J Biol Chem* 2011; **286**: 20930-20941.

[39]Marcotte ED, Pellegrini M, Ng H-L, Rice DN, Yeates TO, Eisenberg D. Detecting protein function and protein-protein interactions from genome sequences. *Nature* 1999; **285**: 751-753.

[40]Kuriyan J, Eisenberg D. The origin of protein interactions and allostery in colocalization. *Nature* 2007; **450**: 983-990.

[41]Back JW, de Jong L, Muijsers AO, de Koster CG. Chemical cross-linking and mass spectrometry for protein structural modeling. *J Mol Biol* 2003; **331**: 303-313.

[42]Lee Y J. Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes. *Mol Biosyst* 2008; **4**: 816-823.

[43]Leitner A, Walzthoeni T, Kahraman A, Herzog F, Renner O, Beck A, et al. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics* 2010; **9**: 1634-1649.

[44]Gonzalez-Segura L, Witte JF, McClard RW, Hurley TD. Ternary complex formation and induced asymmetry in orotate phosphoribosyltransfrase. *Biochemistry* 2007; **46**: 14075-14086.

[45]Zhang Y, Deng H, Schramm VL. Leaving group activation and pyrophosphate ionic state at the catalytic site of *Plasmodium falciparum* orotate phosphoribosyltransferase. *J Am Chem Soc* 2010; **132**: 17023-17031.

[46]Bello AM, Poduch E, Liu Y, Wei L, Crandall I, Wang X, et al. Structure-activity relationships of C6-uridine derivatives targeting *Plasmodia orotidine* monophosphate decarboxylase. *J Med Chem* 2008; **51**: 439-448.

[47]Karlin S, Brocchieri L, Bergman A, Mrazek J, Gentles AJ. Amino acid runs in eukaryotic proteomes and disease associations. *Proc Natl Acad Sci USA* 2002; **99**: 333-338.

[48]Ofran Y, Rost B. Analysing six types of protein-protein interfaces. *J Mol Biol* 2003; **325**: 377-387.

[49]Skrabanek L, Saini HP, Bader GD, Enright AJ. Computational prediction of protein-protein interactions. *Mol Biotechnol* 2008; **38**: 1-17.

[50]Yan C, Wu F, Jernigan RJ, Dobbs D, Honavar V. Characterization of protein-protein interfaces. *Protein J* 2008; **27**: 59-70.

[51]Dafforn TR. So how do you know you have a macromolecular complex? *Acta Cryst* 2007; **D63**: 17-25.

[52]Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 2007; **372**: 774-797.

[53]Singh SK, Maithal K, Balaram H, Balaram P. Synthetic peptides as inactivators of multimeric enzymes: inhibition of *Plasmodium falciparum* triosephosphate isomerase by interface peptides. *FEBS Lett* 2001; **501**: 19-23.

[54]Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* 2007; **450**: 1001-1009.