Molecular Modelling of Calcium Dependent Protein Kinase 4 (CDPK4) from *Plasmodium falciparum*

Tsepo L. Tsekoa
Systems Biology Group
CSIR Biosciences
Pretoria 0001
+27 12 841 3274
ttsekoa@csir.co.za

ABSTRACT
Malaria continues to be one of the most serious global health challenges. The increasing incidence of drug resistant *Plasmodium* strains has emphasised the need for urgent action in the development of new therapeutic strategies against this disease. Development of new drug targets is of vital importance in this regard. The recent availability of genomic information and the resultant observation that in many instances, protein kinases from parasitic protozoa are phylogenetically distant from those in humans has established this group of enzymes as potential drug targets in the Malaria parasite. In order to rationally design novel inhibitors and chemical tools exclusively targeting CDPKs, reliable molecular structures are needed. Structural Bioinformatics, specifically molecular modelling, can contribute immensely to improving access to structural information for these challenging targets. Here, a three dimensional structure of *Pf*CDPK4 created by homology modelling is reported. Further, a model structure with computationally docked ATP is created. These structures will be used to facilitate the discovery and development of novel inhibitors and chemical proteomics tools for the study of this sub-family of proteins.

Categories and Subject Descriptors
J3 [**Computer Applications**]: Life and Medical Sciences – *biology and genetics.*

General Terms
Algorithms, Experimentation.

Keywords
Structural bioinformatics, homology modelling, protein kinase, *Plasmodium falciparum*, Malaria.

1. INTRODUCTION
Malaria continues to be one of the most serious global health challenges.

The increasing incidence of drug resistant *Plasmodium* strains has emphasized the need for urgent action in the development of new therapeutic strategies against this disease. Development of new drug targets is of paramount importance in this regard.

Until recently, protein kinases were not viewed as viable drug targets in parasitic infectious diseases; not least due to the fact that the human host for these parasites expresses more than five hundred different protein kinases. Availability of genomic information and the resultant observation that in many instances, protein kinases from parasitic protozoa are phylogenetically distant from those in humans has resulted in this view being reversed [1].

The protein kinase complement of *Plasmodium falciparum*, the main infectious agent of lethal malaria in humans, has been analysed in detail [2, 3].

These analyses revealed that the *P. falciparum* kinome comprises as many as 65 sequences related to typical eukaryotic protein kinases (ePKs) as defined in model organisms. A novel family of phylogenetically distinct ePK-related genes in *P. falciparum* has been identified. These kinases (up to 20 in number [2], designated the FIKK family due to a conserved amino-acid motif in their catalytic domain, are only found in the Apicomplexa and are particularly highly-represented in *P. falciparum* [4]. A further difference from the host kinome was the presence of calcium dependent protein kinases (CDPKs), normally only found in plants. The presence of these kinases and kinase families that are only remotely related to human ePKs might facilitate the discovery of novel inhibitors and drugs with exclusive specificity for *P. falciparum.* In addition to their absence in the human host, CDPKs demonstrate several characteristics of a good drug target. These include the essential involvement of CDPK4 in sexual development of the parasite [1].

In order to develop novel inhibitors and chemical tools targeting CDPKs, reliable molecular structures are needed. Currently, no reports of experimentally solved structures of the FIKK kinases or *P. falciparum* CDPKs exist. This can be attributed to a number of factors including the continued limited interest from industrial research groups in the biology of the Malaria parasite. In addition, major challenges are often experienced in attempts to express protein targets from *P. falciparum* in heterologous hosts for

subsequent crystallisation. The AT rich genome of *P. falciparum* can result in difficulties in heterologous expression due to the resultant high prevalence of rare codons relative to the host machinery [5]. The presence of low complexity insertions in *Plasmodium* proteins also contributes to the inability of these proteins to express in soluble form in heterologous hosts [6].

Structural Bioinformatics, specifically molecular modelling, can contribute immensely to improving access to structural information for these challenging targets. Here, a three dimensional structure of *Pf*CDPK4 created by homology modelling is reported. Further, a model structure with computationally docked ATP is created. These structures will be used to facilitate the discovery and design of novel inhibitors and chemical proteomics tools for the study of this sub-family of proteins.

2. METHODS

2.1 Molecular Graphics Software

All molecular modelling operations were carried out using Discovery Studio 2.1 (Accelrys Inc., San Diego, CA). Structure images were generated using Pymol 0.99.

2.2 Template Selection and Alignment

The amino acid sequence of *P. falciparum* CDPK4 (*Pf*CDPK4) was retrieved from the UniProt Protein Knowledgebase and used to search non redundant protein chains from the Protein Data Bank (PDB) *via* PSI-BLAST using the BLOSUM62 scoring matrix. The crystal structure of the kinase domain of CDPK1 from *Cryptosporidium parvum* (PDB Id: 2qg5) was ultimately selected as a template due to its relatively close sequence identity to *Pf*CDPK4 and our speculation that since the two parasitic organisms are phylogenetically related, their CDPKs may have functional similarity. The structure coordinates of *Cp*CDPK were obtained from the RCSB Protein Data Bank and the amino acid sequence of chain A extracted. A sequence alignment between *Pf*CDPK4 and *Cp*CDPK was generated using the Align123 module in Discovery Studio 2.1 (Accelrys Inc., San Diego, CA). Align123 is based on ClustalW and uses a progressive pairwise alignment algorithm. A gap open penalty of 10 and a gap extension penalty of 0.05 were used.

2.3 Homology Modelling of *Pf*CDPK4 Structure

*Pf*CDPK4 was modelled against *Cp*CDPK (PDB Id: 2qg5) using the Modeller [7] engine operated *via* Discovery Studio 2.1 (Accelrys Inc., San Diego, CA). Five models were created with the alignment produced as described above used as input. The one insertion-loop (Arg38 – Asp41; this region resulted in a gap in the alignment) from the best model (selected on the basis of PDF Total Energy score) was refined using the Modeller energy function without homology constraints. Other regions of the protein were held rigid. A subset of amino-acid side-chains from the best model after loop refinement was subjected to side chain refinement using ChiRotor [8]. This program optimises protein side-chain conformations based on systematic searching of side-chain conformation followed by CHARMm energy minimisation. Only amino acids which fell on low homology (i.e., no sequence identity or similarity) regions of the alignment were selected for ChiRotor refinement.

The final refined, minimised model was validated using Profiles 3D verify and Modeller verify in Discovery Studio 2.1 and using the WHAT-IF server (http://swift.cmbi.ru.nl/servers/html/index.html).

2.4 Computational docking of ATP

The refined, minimised model created above was used for small-molecule docking studies. First, a 12 Å ligand sphere centred on a major cavity (around the typical ATP- binding site of protein kinases) was defined. ATP was then docked using CDOCKER [9]. CDOCKER uses a CHARMm-based molecular dynamics (MD) method for ligand-docking. In the current study, 10 random conformations of each ligand were generated using high-temperature MD. After these conformations were translated into the binding sphere, candidate poses were created using random rigid-body rotations followed by simulated annealing and a final energy minimisation. The final ligand poses were scored using a variety of scoring functions including LigScore1, LigScore2, PLP1, PLP2, Jain, PMF and PMF04.

3. RESULTS

3.1 Homology Model of *Pf*CDPK4

A search of non-redundant protein structures in the current release of the Protein Data Bank (PDB) *via* PSI-BLAST yielded several homologues of *Pf*CDPK4. The crystal structure of the kinase domain of CDPK1 from *Cryptosporidium parvum* (PDB Id: 2qg5) was ultimately selected as a template. This structure was selected on the basis of its close sequence and putative functional homology to *Pf*CDPK4. On alignment, the amino acid sequences of the two proteins' kinase domains showed 39% sequence identity and 63% sequence similarity (Figure 1).

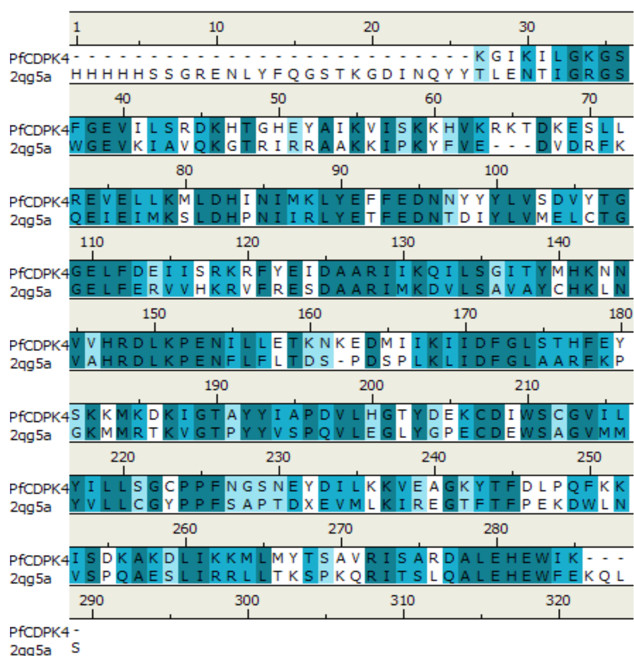

Figure 1 Alignment of the amino acid sequences of *Pf*CDPK4 and *Cp*CDPK1.

The alignment between the two protein sequences displayed long stretches of matched amino acid residues, with only one gap due to a short insertion in the *Pf*CDPK4 sequence.

One model was selected from the five structures modelled on the basis of this alignment. This initial model was further refined using several tools (Section 2.2). Validation results demonstrated that these refinement procedures resulted in incremental improvement of the model at each stage of refinement (Table 1).

Table 1 Validation statistics for homology model of *Pf*CDPK4. At each stage of refinement, the quality of model improved.

	Initial Model	**Loop Refined Model**	**Loop/Sidechain Refined and minimized (Final) Model**
DOPE Score	-28511	-28642	-28692
Verify Score	108.55	108.75	108.82
Ramachandran Z-score	1.256	1.103	1.103

The overall fold of the final model of *Pf*CDPK4 was typical of protein kinase catalytic domains. The structure was bilobal comprising an N-terminal domain of β-sheets and a larger, mainly α-helical C-terminal domain. Superimposition of *Pf*CDPK4 to the template structure showed close similarity between the two structures, with a Root Mean Square Deviation (RMSD) of only 0.153 Å.

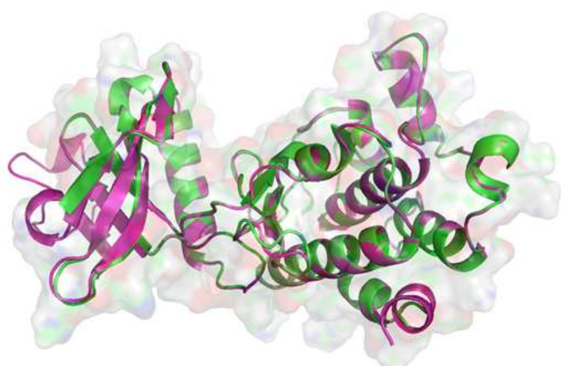

Figure 2 Cartoon representation of superimposition of *Pf*CDPK4 (Green) and *Cp*CDPK1 (Magenta). The typical bilobal kinase catalytic domain structure can be seen.

3.2 ATP-binding to *Pf*CDPK4

Protein kinases that bind ATP have the classic mononucleotide binding fold. This fold comprises a β-strand-loop-α-helix secondary structure [10] and is important for binding of ATP, particularly the phosphate tail.

This fold was identified in the structure of *Pf*CDPK4 and used to determine the positioning of a binding sphere for *in silico* docking. The sphere was centred on a major biding pocket and spanned 12 Å. The program CDOCKER [9] was used for docking. Poses were prioritized by scoring using a variety of scoring functions. Based on this ranking procedure, Pose number 6 was selected as the most likely orientation for binding as it returned consistently high scores from the various scoring functions.

In this pose, ATP nestles into the catalytic cavity between the N-terminal and C-terminal lobes. The molecule forms hydrogen bonds with several amino acid residues (Figure 3).

The adenine hydrogen bonds to a sidechain oxygen of ASP145 and to GLU48. One of the ribose hydroxyl groups is hydrogen bonded to the side chain of GLU48. Since the modelled structure does not contain water molecules or magnesium, not all analogous interactions of the adenine and ribose groups in other kinase structures could be identified (e.g., those that require bridging water).

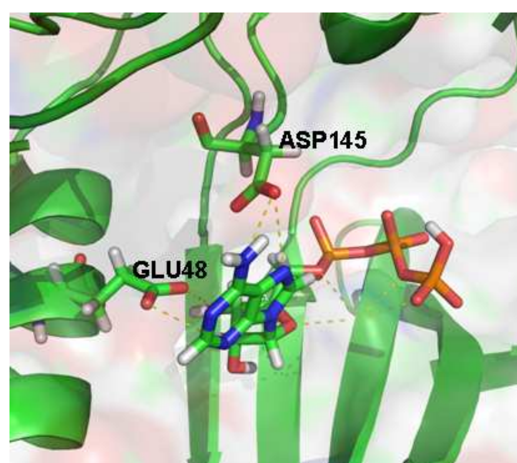

Figure 3 Binding of ATP to *Pf*CDPK4. ATP forms several hydrogen bonds (yellow dashed lines) with amino acid residues ASP145 and GLU48.

4. CONCLUSION

Since the beginning of this millennium, the technologies required to progress from cloning of a gene of interest to experimental determination of the high resolution structure have been accessible even to non-specialist crystallography laboratories, including research facilities on the African continent [11; 12]. However, the pace of determining structures of protein from the malaria parasite *P. falciparum* has been disappointingly slow. This low rate of success can be attributed to a number of factors including the continued limited interest from industrial research groups in the biology of the Malaria parasite. In addition, major challenges are often experienced in attempts to express protein targets from *P. falciparum* in heterologous hosts for subsequent crystallisation.

These limitations in obtaining structural information for challenging targets can be circumvented using structural bioinformatics tools, as demonstrated in this study. Here, the structure of Calcium Dependent Protein Kinase 4 from *P. falciparum* (*Pf*CDPK4) was modelled using computational homology modelling. The structure obtained was subsequently used for *in silico* docking of a key interacting small molecule, ATP. The structures obtained show remarkable similarity to experimentally determined structures of related protein kinases.

The ultimate goal of this study is to rationally design or identify new small molecules for application in the study of *Pf*CDPK4 as a potential drug target in the treatment of Malaria. The initial approach will be to apply Structure-based Virtual Screening of a subset the ZINC database [13, 14] and the National Cancer Institute Diversity Set II database. Promising compounds will be selected for further *in vitro* assays.

5. ACKNOWLEDGMENTS

The author would like to thank CSIR YREF for funding.

6. REFERENCES

[1] Doerig, C., Billker, O., Pratt, D., Endicott, J. 2005. Protein kinases as targets for antimalarial intervention: Kinomics, structure-based design, transmission-blockade, and targeting host cell enzymes. Biochim. Biophys. Acta. 1754(1-2), (Dec. 2005) 132-150.

[2] Ward, P., Equinet, L., Packer, J. and Doerig, C. 2004. Protein kinases of the human malaria parasite Plasmodium falciparum: the kinome of a divergent eukaryote. BMC Genomics. 5(1), (Oct. 2004), 79.

[3] Anamika, Srinivasan, N. and Krupa, A. 2005. A genomic perspective of protein kinases in *Plasmodium falciparum*. Proteins. 58(1), (Jan 2005), 180-189.

[4] Schneider, A.G. and Mercereau-Puijalon, O. 2005. A new Apicomplexa-specific protein kinase family: multiple members in *Plasmodium falciparum*, all with an export signature. BMC. Genomics. 6(1), (Mar. 2005), 30.

[5] Yadava, A., Ockenhouse, C.F. 2003. Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems. Infect. Immun. 71(9), (Sept. 2003), 4961-4969.

[6] Mehlin, C., Boni, E., Buckner, F.S., Engel, L., Feist, T., Gelb, M.H., Haji, L., Kim, D., Liu, C., Mueller, N., Myler, P.J., Reddy, J.T., Sampson, J.N., Subramanian, E., Van Voorhis, W.C., Worthey, E., Zucker, F. and Hol, W.G. 2006. Heterologous expression of proteins from *Plasmodium falciparum*: results from 1000 genes. Mol. Biochem. Parasitol. 148(2), (Aug. 2006), 144-60.

[7] Sali, A.; Blundell, T. L. 2003. Comparative protein modeling by satisfaction of spatial restraints. Mol. Biol. 234, (Dec. 1993), 779-815.

[8] Spassov, V.Z.; Yan, L.; Flook, P.K. 2007. The Dominant Role of Side-Chain Backbone Interactions in Structural Realization of Amino Acid-Code. ChiRotor: A Side-Chain Prediction Algorithm Based on Side-Chain Backbone Interactions. Prot. Sci. 16, (Mar. 2007), 494-506.

[9] Wu, G.; Robertson, D. H.; Brooks, C. L. III; Vieth, M. 2003. Detailed Analysis of Grid-Based Molecular Docking: A Case Study of CDOCKER - A CHARMm-Based MD Docking Algorithm. J. Comp. Chem. (Mar. 2003), 24, 1549.

[10] Schulz, G.E., Schiltz, E., Tomasselli, A.G., Frank, R., Brune, M., Wittinghofer, A., Schirmer, R.H.. 1986. Structural relationships in the adenylate kinase family. Eur. J. Biochem. 161(1), (Nov 1986) 127-132

[11] Tsekoa, T.L., Sayed, M.F., Cameron, R.A., Sewell B.T. and Cowan D.A. (2004) Purification, crystallization and preliminary X-ray diffraction analysis of thermostable nitrile hydratase. S. Afr. J. Sci. 100, (Dec 2004), 488-490.

[12] Kuhnert D.C., Sayed Y., Mosebi S., Sayed M., Sewell T., and Dirr H.W. 2005. Tertiary interactions stabilise the C-terminal region of human glutathione transferase A1-1: a crystallographic and calorimetric study. J. Mol. Biol. 349, (May 2005), 825-38.

[13] Irwin J.J., and Shoichet B.K. 2005. ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. J. Chem. Inf. Model. 45, (Jan 2005), 177-182.

[14] Irwin J.J. 2008. Using ZINC to Acquire a Virtual Screening Library. Curr. Protoc. Bioinform. 22, (June 2008), 14.6.1–14.6.23.