Identification and bioinformatic analysis of a putative calcium-dependent protein kinase (CDPK6) from *Toxoplasma gondii*

N.Z. Zhang[1], S.Y. Huang[1], D.H. Zhou[1], Y. Xu[1,2], J.J. He[1] and X.Q. Zhu[1,2]

[1]State Key Laboratory of Veterinary Etiological Biology,
Lanzhou Veterinary Research Institute,
Chinese Academy of Agricultural Sciences, Lanzhou, Gansu Province, China
[2]College of Animal Science and Technology, Anhui Agricultural University,
Hefei, Anhui Province, China

Corresponding authors: S.Y. Huang
E-mail: siyang.huang@hotmail.com

ABSTRACT. *Toxoplasma gondii* is recognized as an opportunistic human pathogen with a worldwide distribution. Development of effective vaccines is considered the only ideal way to control *T. gondii* infection. However, only one live vaccine is commercially available for use in sheep and goats. Therefore, the identification of more effective antigenic proteins is very important. In this study, we identified a novel putative calcium-dependent protein kinase of *T. gondii*, TgCDPK6, and further analyzed its potential antigenicity using a bioinformatic approach. The physical and chemical characteristics, transmembrane domain, epitopes, advanced structure, and functional sites of TgCDPK6 were predicted by multiple bioinformatic approaches. Twenty-six post-translational modification sites were identified in the protein. The secondary structure showed that 58.35% amino acids of TgCDPK6 are exposed to the solvent interface, and the high hydrophilic domains were distributed in amino acid positions 21-59, 68-81, 156-205, 245-271,

280-294, 297-324, 334-356, 367-393, 474-498, and 543-553. The advanced structure of TgCDPK6 was developed by a homology modeling method and was validated by PROCHECK, which showed that most amino acid residues were in the most favored regions. Using these analyses, 10 potential epitopes were predicted. The results indicated that TgCDPK6 could be a vaccine candidate antigen against *T. gondii*.

INTRODUCTION

The apicomplexan parasite *Toxoplasma gondii* is an important intracellular protozoan parasite that can infect almost all warm-blooded animals and humans (Cenci-Goga et al., 2011; Robert-Gangneux and Dardé, 2012). *T. gondii* can cause severe and life-threatening diseases in developing fetuses and in immunocompromised individuals, such as microcephalus, hydrocephalus, intracerebral calcification, and blindness in infants and some symptomatic diseases in AIDS and transplanted patients (Dubey and Jones, 2008; Robert-Gangneux and Dardé, 2012). *T. gondii* is also an important pathogen of farm animals, causing considerable economic loss to the livestock industry (Innes, 2010).

T. gondii utilizes distinct protein kinase families as signaling mediators that transfer upstream signals to their target proteins to accomplish infection in hosts (Lim et al., 2012). Calcium-dependent protein kinases (CDPKs) serve as important effectors of calcium signaling, have a conserved regulatory mechanism, and are involved in regulating a number of critical biological processes, including attachment to host cells, invasion, gliding motility, and parasite egress (Billker et al., 2009; Sibley, 2011). Previous studies indicated that TgCDPK1 is an essential regulator of calcium-dependent exocytosis in *T. gondii* (Lourido et al., 2010), and the kinase provides an ideal target for new anti-*Toxoplasma* drugs (Ojo et al., 2010). Comparative genomic and phylogenetic analyses of the whole-genome sequence revealed that *T. gondii* contains 11 CDPK-like genes (Nagamune and Sibley, 2006). However, the function of most CDPKs is still not quite clear, and further studies should be performed.

Previous studies reported the functions of this kind of protein kinase in plants (Harper and Harmon, 2005). Several studies indicated that CDPKs regulated a block of essential phenotypes including parasite motility, invasion of host cells, and egress in apicomplexans (Billker et al., 2009). The objectives of this study were to identify a novel protein from *T. gondii* RH strain, TgCDPK6, and to predict its signal peptides, epitopes, topological structures, and some other physical and chemical characteristics using a bioinformatic approach.

MATERIAL AND METHODS

Mice and parasite

Specific-pathogen-free grade female Kunming mice that were 6 to 8 weeks old were purchased from the Center of Laboratory Animals, Lanzhou Institute of Biological Products (Lanzhou, China). All mice were fed with basal diet and tap water *ad libitum* and used for

serial intraperitoneal passage of *T. gondii* tachyzoites. All mice were handled in accordance with good animal practice according to the Animal Ethics Procedures and Guidelines of the People's Republic of China.

Tachyzoites of the highly virulent RH strain of *T. gondii* (Genotype I) were preserved in liquid nitrogen in our laboratory. They were thawed in a 38°C waterbath rapidly before use, and the mice were infected with the thawed tachyzoites intraperitoneally. After 3 serial passages in mice, tachyzoites were harvested from the peritoneal fluid of mice, and purified according to the method previously described (Chen et al., 2013). The tachyzoites of *T. gondii* were used for extracting total RNA.

Extraction of *T. gondii* RNA and amplification of TgCDPK6

Total RNA was extracted from *T. gondii* RH strain tachyzoites using the RNAprep Pure Tissue Kit (TIANGEN, China) according to manufacturer recommendations. The forward primer K6F (5'-CCG*GGTACC*ATGGACGATCCTGCGAACTTCCGA-3') and reverse primer K6R (5'-CGC*GCGGCCGC*TTAGTCGTGGCGCATATACGCGAC-3') were synthesized by Sangon Biotech Co. Ltd. (China). The *Kpn*I and *Not*I recognition sites were introduced in primers and are shown in italics. Then, RNA template and reagents were added to a sterile, nuclease-free polymerase chain reaction (PCR) tube following manufacturer instructions for the One Step RT-PCR Kit Ver. 2 (TaKaRa, China). Briefly, 0.5 μL 20 μM K6F, 0.5μl 20 μM K6R, 12.5 μL 2X 1-step buffer, 1 μL PrimeScript 1-step enzyme mix, 3 μL total *T. gondii* RNA (200 ng), and 7.5 μL RNase-free ddH$_2$O were combined. The following one-step reverse transcription PCR conditions were initiated: 50°C for 30 min; this was followed by 35 cycles of 94°C for 35 s (denaturation), 56°C for 35 s (annealing), and 72°C for 1 min 40 s (extension) and a final extension of 72°C for 10 min. The amplicons (4 μL) produced by PCR were separated by electrophoresis on a 1.0% (w/v) agarose gel that was stained with Goldview and photographed using a gel documentation system (UVItec, Ultra-violet Products ltd., Cambrige, UK).

Cloning, sequencing, and identification of the *TgCDPK6* gene

The gene of interest was purified by a gel extraction kit (TIANGEN), inserted into the pMD18-T vector (TaKaRa), and then transformed into DH 5α *Escherichia coli* competent cells. The Luria-Bertani (LB) plate plus ampicillin (AMP, 100 mg/L), coated with transformed competent cells, was incubated at 37°C for 12 h. Single colonies were cultured in LB liquid medium containing 100 mg/L AMP at 37°C for 16 h. All colonies were identified by PCR amplification, and the positive colonies were sent to Sangon Biotech Co. Ltd. (China) for sequencing. The sequencing results were identified by the basic local alignment search tool on the ToxoDB database (http://toxodb.org/toxo/). The correct recombinant plasmid was named pMD-CDPK6.

Bioinformatic analysis of the TgCDPK6 amino acid sequence

The gene sequence of *TgCDPK6* from the ME49 strain of *T. gondii* was obtained from ToxoDB (http://toxodb.org/toxo/) (Gene ID: TGME49_218720). The general features including molecular weight, isoelectric point, amino acid composition, molecular extinction

coefficient, and the half-life of TgCDPK6 were determined by the ProtParam program, and motifs were predicted by PROSITE. The hydrophilicity/hydrophobicity was analyzed with ProtScale. The TargetP 1.1 server (Emanuelsson et al., 2000) and PredictProtein were used to predict the subcellular location (Rost et al., 2004).

The topological structural information of TgCDPK6 was performed by the online services PredictProtein and SMART (Letunic et al., 2012), and the 3-dimensional (3D) structure of TgCDPK6 was predicted based on a homology modeling technique with the online service SWISS-MODEL (Kiefer et al., 2009). The template protein code named 3HX4 in PDB was selected by the CPH models (Nielsen et al., 2010), which recognized the template based on a profile-profile alignment guided by secondary structure and exposure predictions. The 3D modeling of TgCDPK6 was constructed by the program Swiss-PDB Viewer (Schwede et al., 2003) and was evaluated by a Ramachandran plot, which is accessible online (http://tomcat. cs.rhul.ac.uk/home/mxba001). The flexibility and accessibility of the structural proteins were analyzed by the Karplus-Schulz and Plot-Emini methods. Finally, the potential epitopes of TgCDPK6 were predicted by the DNAStar 8.0 (DNAStar, USA) software with the Jameson-Wolf index.

RESULTS

Identification of the recombinant vector pMD-CDPK6

Identification of the pMD-CDPK6 recombinant vector in colonies by PCR showed that the size of the amplified gene fragment was about 1700 bp, which was the same size as that of *TgCDPK6* gene fragments (Figure 1). The restriction product showed 2 bands, about 1700 and 2700 bp, corresponding to the expected fragment sizes (Figure 1). The sequencing result of the positive colony showed 99% identity to the corresponding sequence in the ME49 strain of *T. gondii*. These results indicated that the target gene was successfully inserted into the pMD-T vector.
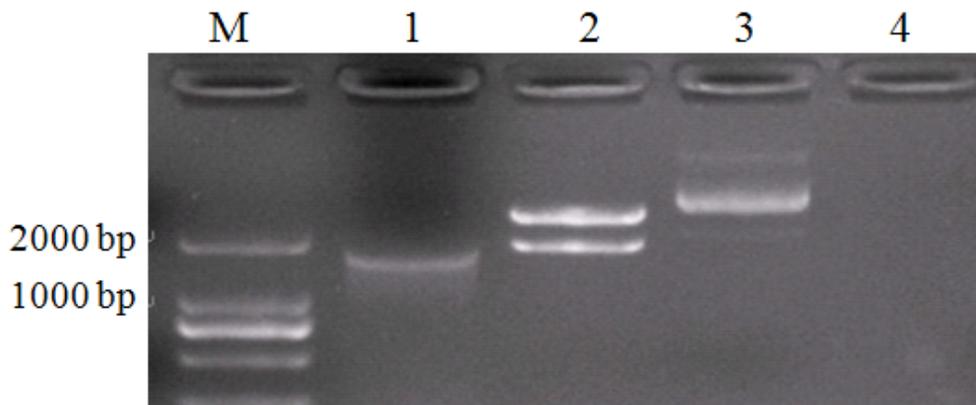
Figure 1. Identification of the recombinant vector pMD-CDPK6 by PCR amplification and restriction enzyme digestion. *Lane M* = DNA marker; *lane 1* = PCR product; *lane 2* = pMD-CDPK6 digested by *Kpn*I and *Not*I; *lane 3* = recombinant vector pMD-CDPK6; *lane 4* = blank control.

Gene information and general characteristics of TgCDPK6 protein

The *TgCDPK6* gene contained a 1674-bp open reading frame that encoded a protein of 558 amino acids. The molecular weight of TgCDPK6 protein was 63.960 kDa, and its theoretical isoelectric point was 5.57. At 280 nm in water, the molar extinction coefficient of TgCDPK6 was 59,860 $mol^{-1} \cdot cm^{-1}$, while in 0.1% acrylonitrile butadiene styrene (1 g/L), the molecular extinction coefficient was 0.936 $mol^{-1} \cdot cm^{-1}$. The half-life of the protein in yeast and *E. coli* was estimated as >20 and >10 h, respectively. TgCDPK6 protein was classified as unstable (instability index: 47.33).

Prediction of functional motifs

The "EF-hand" Ca^{2+}-binding motif plays an essential role in eukaryotic cellular signaling, and proteins containing this motif constitute a large and functionally diverse family. The CDPK family is one of them. EF-hands can effect a major conformational change of CDPKs upon calcium binding, exposing previously buried hydrophobic residues that interact with target proteins (Gifford et al., 2007). Analyzing the sequence of TgCDPK6, we found 6 EF-hand calcium-binding domains and other post-translational modification (PTM) sites, including 2 *N*-glycosylation sites, 1 cAMP- and cGMP-dependent protein kinase phosphorylation sites, 7 protein kinase C phosphorylation sites, 7 casein kinase II phosphorylation sites, 7 N-myristoylation sites, 1 tyrosine kinase phosphorylation site, and 1 protein kinase ATP-binding region signature (Table 1).

Table 1. Analysis of the post-translational modification sites of TgCDPK6.

Motif	Position (aa)	No.
N-glycosylation site	415-418, 472-475	2
cAMP- and cGMP-dependent protein kinase phosphorylation site	362-365	1
Tryosine kinase phosphorylation site	82-88	1
EF-hand calcium-binding domain	19-30, 60-71, 432-444, 466-478, 501-513, 538-550	6
N-myristoylation site	447-452	1
Protein kinase ATP-binding region signature	132-146	1
Protein kinase C phosphorylation site	10-12, 14-16, 61-63, 109-111, 296-298, 331-333, 509-511	7
Casein kinase II phosphorylation site	112-115, 139-142, 321-324, 331-334, 440-443, 474-477, 509-512	7

Prediction of hydrophilicity

Analysis of amino acid hydrophilicity/hydrophobicity can help to understand the protein folding and interaction sites and to predict the protein secondary structure and antigenic epitopes. Here, the hydrophilic/hydrophobic regions of TgCDPK6 were predicted by the parameter of Hphob using the Abraham & Leo selection in the ProtScale program. The results indicated that TgCDPK6 contained multiple hydrophilic regions and some classical high hydrophilic domains that were distributed in amino acid positions 21-59, 68-81, 156-205, 245-271, 280-294, 297-324, 334-356, 367-393, 474-498, and 543-553.

Prediction of the subcellular location and secondary structure

A signal peptide sequence and cleavage site were not identified in TgCDPK6 by Tar-

getP. The location of a protein always correlates with its function; therefore, we predicted the location of TgCDPK6 with PredictProtein. From the results, we can predict that TgCDPK6 may be located in other parts of the cell instead of the mitochondria because the location in the mitochondria score was 0.219 (<0.9), while in other parts, the score was up to 0.854. The α-helices and β-folds have a high chemical bond energy, which can maintain the structure of a protein and make interactions with antibodies difficult; they are usually found in the internal regions of the protein and generally do not act as epitopes. The random coil is looser, prone to distortion, hovers and is exposed on the surface of the protein, and may be a potential epitope. Protein Fold (PROF) indicated that the proportions of α-helixes, β-folds, and random coils in TgCDPK6 are 42.37, 7.90, and 49.73%, respectively. The solvent accessibility composition predicted that most amino acids were exposed to the solvent interface (58.35%), and the residual amino acids were embedded in the protein interior.

Homology modeling of TgCDPK6

The tertiary structure of a protein is necessarily linked with its biological function. Learning about the 3-D modeling of a protein is easier, faster, and more efficient by homology modeling method than by X-ray crystallography, nuclear magnetic resonance, and some other physical and biochemical methods. We obtained the template protein from the PDB protein database (code: 3HX4) and aligned amino acids 113-556 of TgCDPK6. The model of TgCD-PK6 was constructed using the automated modeling program and performed by the Swiss PDB-Viewer 4.0.1 (SIB, Switzerland) (Figure 2). Analysis of the quality of the TgCDPK6 model by PROCHECK showed that up to 88.3% of the phi/psi angles of residues fell in the most favored regions, 10.4% of residues fell in additional allowed regions, 0.7% of residues fell in generously allowed regions, and arginine at position 293 and glutamic acid at position 477 (0.5%) fell in disallowed conformations.
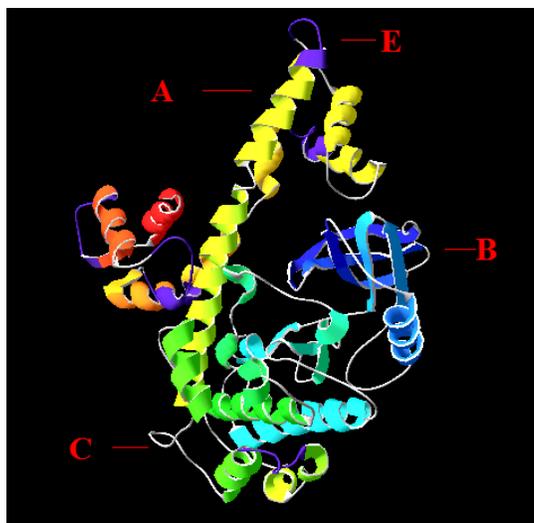
Figure 2. Ribbon diagram of the tertiary structure prediction of TgCDPK6 protein. **A.** α-helix; **B.** β-folded; **C.** random coil; **E.** potential epitopes (purple).

Prediction of linear B-cell epitopes

Each antigen has its own epitopes that determine the specific immunogenicity. In the immune response, T-cell epitopes and B-cell epitopes are recognized by the specific T-cell and B-cell antigen receptor, respectively. Predicting the epitopes is of great help for the development of diagnostic reagents and design of immunogenic peptides and new vaccines. For these reasons, the B-cell epitopes of TgCDPK6 were predicted by Jameson-Wolf index of the DNA-Star 8.0 software (Figure 3). However, potential linear B-cell epitopes are also characteristic of variability, fragment mobility, and hydrophilicity. Based on the analysis of these characters (Figure 3) and the secondary structure of TgCDPK6, 10 potential epitopes of TgCDPK6 were predicted (Table 2).
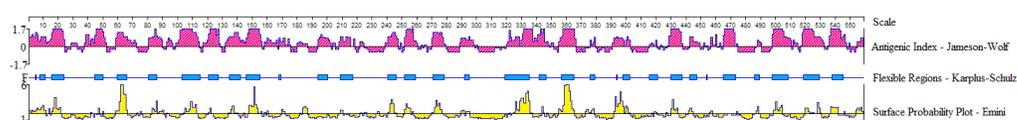
Figure 3. Analysis of the antigenic index (red), flexible regions (blue) and surface probability (yellow) of TgCDPK.

Table 2. Analysis of the antigenic epitopes of TgCDPK6.

Position	Sequence	Score interval (the lowest to the highest)
16-23	PGDSVACS	1.28-3.40
103-112	DTSTKPHFFS	1.30-3.00
121-127	GKDKSQH	1.22-3.10
147-154	SEASRWEI	1.28-3.40
358-365	AARCGKSH	1.78-3.40
430-436	PRHWRRV	1.39-3.40
498-505	RLHKLKKL	1.13-3.40
518-525	DIGMLHDA	1.28-3.40
537-543	LTVAEIQ	1.49-3.40
465-472	PWIQCYTA	1.05-3.00

DISCUSSION

Bioinformatics is becoming increasingly important in vaccine development, both to design novel simplified epitope-based vaccines and to understand the specific immune response of selected vaccine epitopes (Lundegaard et al., 2010; Zhang et al., 2013). Compared with traditional methods, bioinformatic analysis is effective, low-cost, and has been widely used to predict protein structures, functions, and other biological characteristics. Here, we conducted a comprehensive analysis of TgCDPK6, a member of the CDPK family in *T. gondii*. CDPKs, a large protein family consisting of a calcium-binding domain fused to the kinase domain, are widely distributed among plants, algae, *Paramecium tetraurelia*, and the apicomplexan parasites and are involved in a diverse array of functions (Harper and Harmon, 2005). In this study, we cloned a new gene of *T. gondii*, *TgCDPK6*, and constructed a comprehensive analysis of the protein, especially its epitopes.

The length of the cloned *TgCDPK6* gene from the RH strain was 1674 bp and showed 99% identity to the corresponding sequence in the *T. gondii* ME49 strain. The amino acid sequence of TgCDPK6 contains 558 residues and has a molecular weight of 63.960 kDa. No signal peptide sequence is predicted in TgCDPK6, and the protein is not a mitochondrial protein.

The prediction of functional domains showed that TgCDPK6 contains 1 protein kinase ATP-binding region and 7 protein kinase C phosphorylation sites, which may perform the putative function of phosphorylation of an unproven substrate. Two N-terminal EF-hand calcium-binding domains are predicted in TgCDPK6, which is different from the canonical CDPKs in plants. Other PTM sites are also predicted in TgCDPK6, which may play an important role in its cellular control mechanism; there are 2 *N*-glycosylation sites, 1 cAMP- and cGMP-dependent protein kinase phosphorylation site, 7 casein kinase II phosphorylation sites, 7 N-myristoylation sites, and 1 tyrosine kinase phosphorylation site.

Analyses of the structural contents of an unknown protein to predict its immunity by bioinformatics contributed to the development of a new vaccine candidate antigen against *T. gondii* (Rosenberg et al., 2009; Cong et al., 2010; Bai et al., 2012). The secondary structures, such as the random coils exposed on the surface of the protein, generally make a significant impact on the epitopes. In this study, the proportion of random coils in TgCDPK6 was 49.73%. Combined with the hydrophilicity/hydrophobicity and antigen index analyses, 10 potential epitopes were predicted. To visualize the results, the 3D structure of TgCDPK6 was also constructed.

In conclusion, the results of this study indicate that TgCDPK6 is a potential DNA vaccine candidate that is worth being tested experimentally against *T. gondii* infection in further studies.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

Bai Y, He S, Zhao G, Chen L, et al. (2012). *Toxoplasma gondii*: bioinformatics analysis, cloning and expression of a novel protein TgIMP1. *Exp. Parasitol.* 132: 458-464.

Billker O, Lourido S and Sibley LD (2009). Calcium-dependent signaling and kinases in apicomplexan parasites. *Cell Host Microbe* 5: 612-622.

Cenci-Goga BT, Rossitto PV, Sechi P, McCrindle CM, et al. (2011). *Toxoplasma* in animals, food, and humans: an old parasite of new concern. *Foodborne Pathog. Dis.* 8: 751-762.

Chen J, Huang SY, Li ZY, Yuan ZG, et al. (2013). Protective immunity induced by a DNA vaccine expressing eIF4A of *Toxoplasma gondii* against acute toxoplasmosis in mice. *Vaccine* 31: 1734-1739.

Cong H, Mui EJ, Witola WH, Sidney J, et al. (2010). Human immunome, bioinformatic analyses using HLA supermotifs and the parasite genome, binding assays, studies of human T cell responses, and immunization of HLA-A*1101 transgenic mice including novel adjuvants provide a foundation for HLA-A03 restricted CD8+T cell epitope based, adjuvanted vaccine protective against *Toxoplasma gondii*. *Immunome. Res.* 6: 12.

Dubey JP and Jones JL (2008). *Toxoplasma gondii* infection in humans and animals in the United States. *Int. J. Parasitol.* 38: 1257-1278.

Emanuelsson O, Nielsen H, Brunak S and von Heijne G (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300: 1005-1016.

Gifford JL, Walsh MP and Vogel HJ (2007). Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-

helix EF-hand motifs. *Biochem. J.* 405: 199-221.

Harper JF and Harmon A (2005). Plants, symbiosis and parasites: a calcium signalling connection. *Nat. Rev. Mol. Cell Biol.* 6: 555-566.

Innes EA (2010). A brief history and overview of *Toxoplasma gondii*. *Zoonoses Public Health* 57: 1-7.

Kiefer F, Arnold K, Künzli M, Bordoli L, et al. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* 37: D387-D392.

Letunic I, Doerks T and Bork P (2012). SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* 40: D302-D305.

Lim DC, Cooke BM, Doerig C and Saeij JP (2012). *Toxoplasma* and *Plasmodium* protein kinases: roles in invasion and host cell remodelling. *Int. J. Parasitol.* 42: 21-32.

Lourido S, Shuman J, Zhang C, Shokat KM, et al. (2010). Calcium-dependent protein kinase 1 is an essential regulator of exocytosis in *Toxoplasma*. *Nature* 465: 359-362.

Lundegaard C, Hoof I, Lund O and Nielsen M (2010). State of the art and challenges in sequence based T-cell epitope prediction. *Immunome. Res.* (6 Suppl 2): S3.

Nagamune K and Sibley LD (2006). Comparative genomic and phylogenetic analyses of calcium ATPases and calcium-regulated proteins in the apicomplexa. *Mol. Biol. Evol.* 23: 1613-1627.

Nielsen M, Lundegaard C, Lund O and Petersen TN (2010). CPHmodels-3.0-remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res.* 38: W576-W581.

Ojo KK, Larson ET, Keyloun KR, Castaneda LJ, et al. (2010). *Toxoplasma gondii* calcium-dependent protein kinase 1 is a target for selective kinase inhibitors. *Nat. Struct. Mol. Biol.* 17: 602-607.

Robert-Gangneux F and Dardé ML (2012). Epidemiology of and diagnostic strategies for toxoplasmosis. *Clin. Microbiol. Rev.* 25: 264-296.

Rosenberg C, De CS, Jongert E, Gargano N, et al. (2009). Induction of partial protection against infection with *Toxoplasma gondii* genotype II by DNA vaccination with recombinant chimeric tachyzoite antigens. *Vaccine* 27: 2489-2498.

Rost B, Yachdav G and Liu J (2004). The PredictProtein server. *Nucleic Acids Res.* 32: W321-W326.

Schwede T, Kopp J, Guex N and Peitsch MC (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* 31: 3381-3385.

Sibley LD (2011). Invasion and intracellular survival by protozoan parasites. *Immunol Rev.* 240: 72-91.

Zhang NZ, Chen J, Wang M, Petersen E, et al. (2013). Vaccines against *Toxoplasma gondii*: new developments and perspectives. *Expert. Rev. Vaccines* 12: 1287-1299.