

Molecular Characterization of *tg*d057, a Novel Gene from *Toxoplasma gondii*

Kiew-Lian Wan*, Ti-Ling Chang and James W. Ajioka[†]

Centre for Gene Analysis and Technology, School of BioSciences and Biotechnology,
Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor DE, Malaysia,

[†]Department of Pathology, Cambridge University, Cambridge CB2 1QP, UK

Received 19 November 2003, Accepted 13 December 2003

The expressed sequence tag (EST) effort in *Toxoplasma gondii* has generated a substantial amount of gene information. To exploit this valuable resource, we chose to study *tg*d057, a novel gene identified by a large number of ESTs that otherwise show no significant match to known sequences in the database. Northern analysis showed that *tg*d057 is transcribed in this tachyzoite. The complete cDNA sequence of *tg*d057 is 1169 bp in length. Sequence analysis revealed that *tg*d057 possibly adopts two polyadenylation sites, utilizes the fourth in-frame ATG for translation initiation, and codes for a secretory protein. The longest open reading frame for the *tg*d057 gene was cloned and expressed as a recombinant protein (rd57) in *Escherichia coli*. Western analysis revealed that serum against rd57 recognized a molecule of ~21 kDa in the tachyzoite protein extract. This suggests that the *tg*d057 gene is expressed *in vivo* in the parasite.

Keywords: EST, Multiple polyadenylation site, Recombinant protein expression, Tachyzoite

Introduction

The apicomplexan protozoan *Toxoplasma gondii* is an opportunistic pathogen and causes toxoplasmosis. Although *T. gondii* infection is typically asymptomatic in healthy adults (Frenkel, 1988), it can be fatal in immunocompromised individuals, such as in those undergoing immunosuppressive therapy after organ transplantation (Isrealski and Remington, 1993) or in AIDS immunocompromised hosts (Luft and

Remington, 1992). Toxoplasmosis is also recognized as a major problem in animals, causing abortion and neonatal mortality in sheep, goats, pigs, and cattle (Dubey and Beattie, 1988). However, despite decades of research, effective chemotherapy and vaccine prophylaxis remain elusive. This may partly be due to our lack of knowledge of the genes and molecules involved in the complex biology of this organism.

Recent efforts to identify new *T. gondii* genes have employed the expressed sequence tag (EST) approach (Wan *et al.*, 1996; Ajioka *et al.*, 1998; Manger *et al.*, 1998; Li *et al.*, 2003), and have resulted in the generation of over 72,000 ESTs from the parasite (GenBank dbEST release 111403). Analyses of a number of these ESTs have revealed that while putative functions could be assigned to more than 500 novel gene sequences, the majority of these ESTs could not be identified exclusively based on database matches (Ajioka *et al.*, 1998). Since these ESTs represent single pass sequences from the 5' end of cDNAs, they would not show significant similarity to potential orthologs if they correspond to untranslated regions, which are usually poorly conserved between organisms, or to parts of coding regions that do not contain common motifs. The probability of assigning a putative identification to these ESTs would increase with similarity searches using full-length cDNA sequences and might help define their taxonomic distribution, i.e. confined to *T. gondii*, to coccidians, to apicomplexans. In addition to the identification of these ESTs, basic molecular characterizations are essentially required initial steps towards the understanding of the roles of the genes and molecules in the parasite. Here, we chose to study the EST locus as represented by the cDNA clone *tg*d057. We provide evidence that the *tg*d057 gene is expressed in *T. gondii* tachyzoites and that it is likely to be unique to the species.

Materials and methods

Parasite culture *T. gondii* tachyzoites were cultured *in vitro* using human foreskin fibroblasts (HFF) according to the procedure

Note: Nucleotide sequence data reported in this paper is available at the GenBank database under accession number [AY313956](#).

*To whom correspondence should be addressed.
Tel: ++603-8921 5997; Fax: ++603-8925 2698
E-mail: klwan@pkrisc.cc.ukm.my

described by Roos *et al.* (1994).

Northern analysis Total RNA was prepared from HFF and *T. gondii* using Tri Reagent (Molecular Research Center Inc., Cincinnati, USA) according to the manufacturer's protocol. Approximately 30 µg of total RNA was electrophoresed in a 1.2% agarose gel under denaturing conditions, transferred to a nylon membrane (Amersham Pharmacia Biotech, Buckinghamshire, UK) and probed with the *tg*d057 DNA insert labeled using the ECLTM Direct Nucleic Acid Labeling and Detection System (Amersham Pharmacia Biotech). The membrane was washed to final stringency with 0.5X SSC (75 mM NaCl, 7.5 mM sodium citrate), and 0.4% sodium dodecyl sulfate at 55°C, and exposed for autoradiography at -70°C.

Generation and analysis of the cDNA sequence Phage stock for the cDNA clone *tg*d057, derived from a *T. gondii* tachyzoite cDNA library (Wan *et al.*, 1996), was subjected to *in vivo* excision using the ExAssistTM/SOLR system (Stratagene, La Jolla, USA). Plasmid DNA was extracted from the resulting clones using the WizardTM Minipreps DNA purification system (Promega, Madison, USA) and sequenced using the ABI PRISMTM Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, USA) on an ABI 377 DNA sequencer (Applied Biosystems). The complete sequence of both strands was determined using standard SK and T7 vector primers and gene-specific primers, *tg*d057-1f (5'-CGC AAA AGA CCA ACT TCG GCT GT-3') and *tg*d057-1r (5'-AGT CGA GTA CAG GTC AGG TGA GGA T-3'). Amplification of the 5' and 3' ends of the full-length transcript was carried out using the *T. gondii* tachyzoite cDNA library as a template. 5' RACE products were generated using the T3 and *tg*d057-1r primers, and the 3' RACE products were generated using T7 and *tg*d057-1f. Amplified products were cloned and sequenced as described above.

Nucleotide and amino acid sequence analyses were carried out using the MacDNASIS package (Hitachi Software Engineering, USA). Similarity searches were performed against the National Center for Biotechnology Information (NCBI) GenBank and the Institute for Genome Research (TIGR) Gene Indices databases utilizing the basic local alignment search tool (BLAST) (Altschul *et al.*, 1997). The signal peptide prediction was done using SignalP V2.0.b2 (www.cbs.dtu.dk/services/SignalP-2.0) (Nielsen *et al.*, 1997).

Expression and purification of the recombinant *tg*d057 protein (rd57) The predicted open reading frame (ORF) for *tg*d057 was amplified from the *T. gondii* tachyzoite cDNA library using gene-specific primers; *tg*d057ORF.F (5'-AAT AAA GAT CGC CCG TGT CGT-3') and *tg*d057ORF.R (5'-CTC GAC CTC AAT GTT GTA TTC-3'), and the amplified fragment was cloned into pBAD/Thio-TOPO expression vector (Invitrogen, San Diego, USA). Transformants with the correct insert sequence and orientation were grown in LB media with ampicillin (50 µg/ml). Cultures were then induced with 0.2%, 0.02%, or 0.002% arabinose, and samples were collected at 0, 2, 4, and 16 h after induction. Following centrifugation, protein pellets were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). For large-scale protein preparation, cultures were induced with 0.002%

arabinose and grown for 16 h. Recombinant protein was purified using a metal chelating resin column (Invitrogen).

Polyclonal antibody production Polyclonal serum against *tg*d057 was raised in rabbits hyperimmunized with purified rd57. Approximately 200 µg of rd57 was used for the primary immunization, and a boost of 100 µg was given after a 2-week interval. Serum was prepared from the first bleed and used in subsequent experiments.

Western analysis Parasite lysates were prepared by sonication, separated on 14% polyacrylamide gels, and transferred to nitrocellulose membranes. Membranes were subsequently blocked with 5% skim milk in phosphate-buffered saline, incubated with anti-rd57 polyclonal antibodies for 2 h, washed four times, incubated with horse radish peroxidase-conjugated goat anti-rabbit antibodies (Pierce Chem., Rockford, USA), and washed an additional four times (Su *et al.*, 2003). Detection was carried out using the ECLTM Direct Nucleic Acid Labeling and Detection System (Amersham Pharmacia Biotech).

Results

Determination and analysis of the complete cDNA sequence Sequencing of the cDNA clone *tg*d057 resulted in a sequence of 1041 bp, excluding the poly (A) tail. When the *tg*d057 DNA insert was used as a probe against the northern blots of the total RNAs from HFF and *T. gondii*, a clear thick band of approximately 1100 bp, was detected in the *T. gondii* RNA (Fig. 1). The size of this *T. gondii* RNA band indicated that the *tg*d057 clone contained a truncated cDNA insert. No signal was detected in the HFF RNA.

The distal 5' end of the transcript was obtained using a modified RACE method (Frohman *et al.*, 1988). Amplification with a vector and an internal primer produced a fragment of approximately 800 bp in length. Four of the eight 5' RACE subclones sequenced provide an additional 44 bp to the distal 5' end sequence. Amplification of the 3' end resulted in a band of approximately 600 bp in length. Two of the six 3' RACE subclones sequenced revealed an identical sequence and terminated at the same poly (A) site as the cDNA clone *tg*d057. Sequencing of the other four subclones generated an additional 84 bp at the 3' end, with a poly (A) tail at the end of each of the sequences. Thus, the alignment of sequences generated from both the 5' and 3' RACE products produced two transcripts of the *tg*d057 gene of 1085 and 1169 bp (Fig. 2).

Translation of this sequence in all six frames shows that the longest ORF for *tg*d057 was comprised of 630 bp encoding 210 amino acid residues with a theoretical molecular mass of 24.5 kDa (Fig. 2). This ORF is defined by an ATG codon at nucleotide position 38 and a TGA stop codon at nucleotide position 668. In addition to this first ATG codon, three other in-frame ATG codons were detected in the N terminal region of the predicted polypeptide sequence. The alignment of sequences immediately upstream of the ATG codons revealed

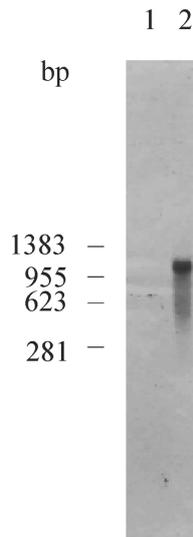


Fig. 1. Northern analysis of *tgd057*. Lane 1, total RNA from HFF; Lane 2, total RNA from *T. gondii*. The sizes indicated were determined using RNA markers (Promega, USA).

that sequences surrounding all four, with the exception of the fourth, ATG codons showed poor conservation with respect to the *T. gondii* consensus sequence (Seeber, 1997) (Fig. 3), suggesting that the parasite utilizes the fourth in-frame ATG codon as the site for translation initiation. In relation to the fourth in-frame ATG codon, a putative signal peptide (Nielsen *et al.*, 1997) and a signal peptide cleavage site between acid amino residues 21 and 22 were predicted. Thus, the mature polypeptide is 150 amino acids in length and has an expected molecular weight of 17.9 kDa.

A comparison of the complete cDNA sequence of the *tgd057* gene against the GenBank non-redundant nucleotide and protein databases produced no matches with scores of 90 or greater, other than the matches with *T. gondii* ESTs. Further analysis with the TIGR Gene Indices database (*T. gondii* Gene Index; TgGI version 3.0) showed that the most significant EST cluster recognized by *tgd057* is TC5308. This tentative consensus represents 97 ESTs, of which 94 were identified as tachyzoite ESTs; only one was found to be bradyzoite derived.

Recombinant protein expression, purification, and analysis The entire predicted ORF for *tgd057* was expressed as a recombinant protein in *E. coli*. A band showing increasing intensity was detected after 2 h induction with arabinose (Fig. 4A). The ~43.0 kDa size of this band was consistent with the expected size of the recombinant *tgd057* protein (rd57), which contains the predicted *tgd057* polypeptide, the N-terminal fusion HP-thioredoxin protein, and a C-terminal polyhistidine tag. The rd57 protein was purified using a his-tag affinity column, and subsequent SDS-PAGE gel analysis and Coomassie staining showed a single readily detected band (Fig. 4B). The expression of the recombinant protein was also confirmed by western analysis

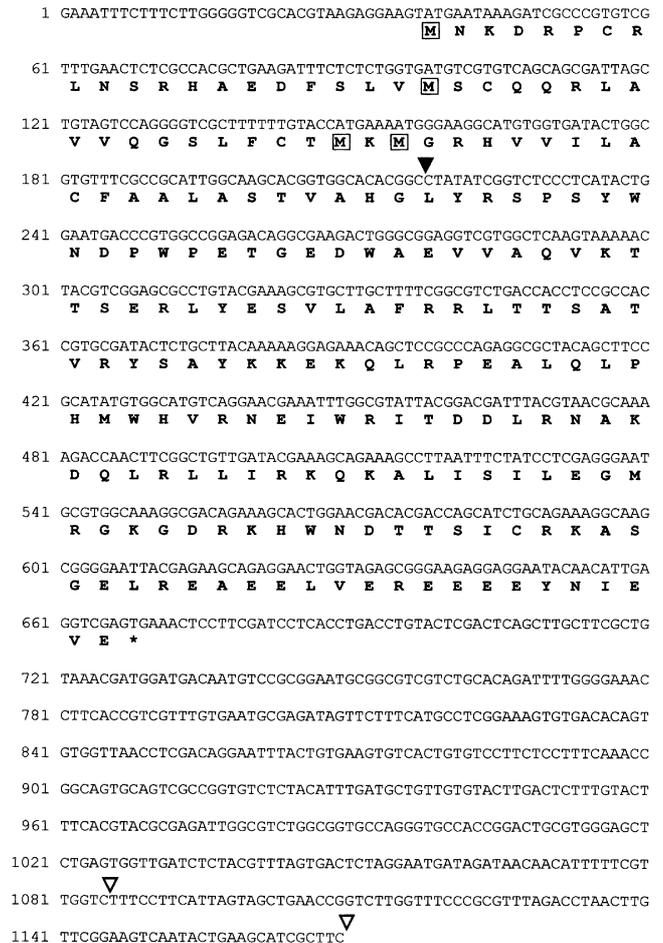


Fig. 2. The DNA and predicted amino acid sequence of *tgd057*. The predicted amino acid sequence is presented in single lettering below the corresponding DNA sequence. Potential start codons are boxed and the asterisk corresponds to the putative stop codon. The solid inverted triangle denotes a signal peptide cleavage site, whereas open inverted triangles correspond to polyadenylation sites.

carried out using horseradish peroxidase-conjugated anti-histidine monoclonal antibody and antiserum against rd57 (data not shown).

In vivo expression of *tgd057* Western analysis demonstrated that polyclonal rabbit anti-rd57 recognized a molecule of approximately 21 kDa in size in the *T. gondii* tachyzoite lysate (Fig. 5). As controls, western analyses were also carried out using anti-HP-thioredoxin/histidine and prebleed serum. Both probes were unable to recognize any protein in the parasite lysate (data not shown).

Discussion

The anonymous gene sequences revealed by this *T. gondii* EST study are interesting candidates for further

ATG#	Position	Sequence
1	1	... g g a a g t A T C a ...
2	21	... c t g g t g A T C t ...
3	28	... t g t a c c A T C a ...
4	40	... a t g a a a A T C g ...
<i>T. gondii</i> consensus		... g n c a a a A T C g ...

Fig. 3. Comparison of potential translational initiation sites in *tg*d057. Sequences surrounding the first four in-frame ATG codons in *tg*d057 were aligned against the consensus sequence of *T. gondii* (Seeber, 1997). Residues identical to the consensus sequence are highlighted in black. The positions of ATG codons from the beginning of the predicted open reading frame are indicated.

characterization, especially if they are encountered frequently within the dataset. Cluster analysis carried out on a subset of these ESTs identified four anonymous EST loci amongst the top 15 most redundant sequences (Ajioka *et al.*, 1998). The relative abundances of these unidentified ESTs imply that they code for molecules that have significant functions in the biology of the parasite. Amongst these is a cluster of nearly 100 ESTs that represent the *tg*d057 gene.

In direct studies of anonymous EST loci such as this, it is important to establish the fundamental characteristics of the derived gene. Northern analysis provided evidence that the *tg*d057 gene is transcribed in the parasite. The presence of a heavy band identified by *tg*d057 on the *T. gondii* RNA blot indicated that at least one transcript of the gene is present in the parasite mRNA population. Using specific antisera prepared from the purified recombinant protein, we further determined that *tg*d057 is expressed in at least one form of the parasite, as antiserum to rd57 readily detected the native *tg*d057 protein in *T. gondii* tachyzoite lysates. However, the apparent molecular mass of the native *tg*d057 protein was found to be slightly larger than the predicted polypeptide mass, suggesting that native *tg*d057 carries few post-translational modifications.

The *tg*d057 gene also possesses a few noteworthy features. Sequence data from 3' RACE shows that two polyadenylation sites are present in the complete cDNA sequence of *tg*d057. Though not apparent on the northern blot, two transcripts would be generated by polyadenylation at both sites. Moreover, a number of genes in *T. gondii* such as *P22* (Prince *et al.*, 1990), *P28* (Parmley *et al.*, 1993), and *tg*c002 (unpublished data) have been reported to utilise more than one polyadenylation site. Although the function of these multiple polyadenylation sites in the parasite is unclear, analyses of *tg*d057 and *tg*c002 demonstrated the value of 3' RACE for determining the distal 3' end(s) of *T. gondii* transcripts from EST data.

Sequence analysis revealed that *tg*d057 possibly utilizes the fourth in-frame ATG codon for translation initiation.

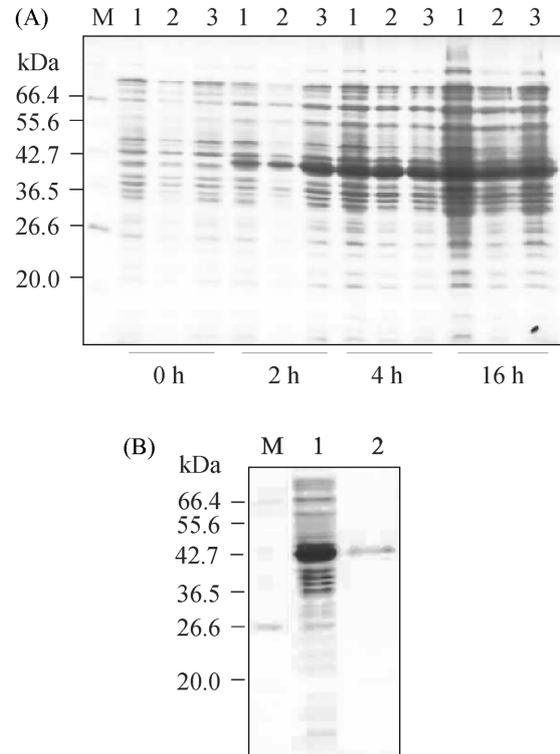


Fig. 4. SDS-PAGE of rd57. (A) Analysis of rd57 expression. Samples were collected at various time points after induction with 0.002% (lane 1), 0.02% (lane 2), or 0.2% (lane 3) arabinose. (B) Analysis of rd57 purification. Electrophoresis was carried out on rd57 in arabinose-induced *E. coli* extracts (lane 1) and purified rd57 (lane 2). The protein sizes indicated were determined using a broad range protein marker.

Translation in eukaryotes is normally initiated at the first 5' proximal in-frame ATG codon (Pain, 1996). However, for *tg*d057, the sequence around the fourth in-frame ATG codon was found to share highest similarity with a compiled *T. gondii* consensus sequence (Seeber, 1997). The usage of the fourth ATG was also consistent with the prediction of a signal peptide, and suggested that *tg*d057 may be secreted by the parasite (Joiner and Roos, 2002). During invasion, *T. gondii* releases a wide range of highly abundant proteins from specialized secretory organelles, such as micronemes, rhoptries, and dense granules (Carruthers, 2002). Thus, we speculate that further experiments may localize *tg*d057 to one of these secretory compartments.

Comparisons of the complete cDNA sequence of the *tg*d057 gene with other database sequences produced no putative identification. Further sequence similarity searching, despite the availability of a large number of ESTs for other closely related apicomplexan parasites such as *Neospora caninum* (Li *et al.*, 2003; Ellis *et al.*, 2000), *Eimeria tenella* (Li *et al.*, 2003; Wan *et al.*, 1999; Ng *et al.*, 2002), *Sarcocystis neurona* (Li *et al.*, 2003; Howe, 2001), *Plasmodium falciparum* (Li *et al.*, 2003; Chakrabarti *et al.*, 1994) and

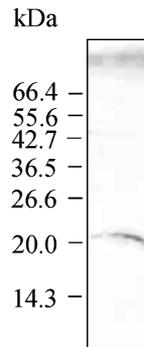


Fig. 5. Western blot analysis of *tgd057* expression in *T. gondii*. Anti-rd57 polyclonals were used as a probe against tachyzoite total protein extract. The protein sizes indicated were determined using a broad range protein marker.

Cryptosporidium parvum (Strong and Nelson, 2000), produced no significant EST match. Taken together, these results imply that *tgd057* is a novel *T. gondii* gene exclusive to the parasite. Results of matches with the *T. gondii* clustered EST database also suggest that *tgd057* is highly dominant in the tachyzoite, and that it accounts for almost 0.6% of the dataset. Moreover, its abundance in *T. gondii* suggests that it is an important molecule, and one which has potential use as a valuable diagnostic tool for detection purposes.

Acknowledgments We thank Louis Weiss, Vern Carruthers and David Sibley for technical advice and helpful discussions. This project was funded by a grant from the MPKSN (IRPA 09-02-02-0055) and from the Wellcome Trust Research Development Award for Tropical Medicine to K.L.W.

References

- Ajioka, J. W., Boothroyd, J. C., Brunk, B. P., Hehl, A., Hillier, L., Manger, I. D., Marra, M., Overton, G. C., Roos, D. S., Wan, K. -L., Waterston, R. and Sibley, L. D. (1998) Gene discovery by EST sequencing in *Toxoplasma gondii* reveals sequences restricted to the Apicomplexa. *Genome Research* **8**, 18-28.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.
- Carruthers, V. (2002) Host cell invasion by the opportunistic pathogen *Toxoplasma gondii*. *Acta Tropica* **81**, 111-122.
- Chakrabarti, D., Reddy, G. R., Dame, J. B., Almira, E. C., Lapis, P. J., Ferl, R. J., Yang, T. P., Rowe, T. C. and Schuster, S. M. (1994) Analysis of expressed sequence tags from *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **66**, 97-104.
- Dubey, J. P. and Beattie, C. P. (1988) *Toxoplasmosis of animals and man*, CRC Press, Boca Raton, USA.
- Ellis, J. T., Ryce, C., Atkinson, R., Balu, S. and Holmdahl, O. J. M. (2000) *Neospora caninum*: gene discovery through analysis of expressed sequence tags; in *A European perspective on Neospora caninum*, Hemphill, A. and Gottstein, B. (eds.), *Intl. J. Parasitol.* **30**, 877-924.
- Frenkel, J. K. (1988) Pathophysiology of toxoplasmosis. *Parasitol. Today* **4**, 273.
- Frohman, M. A., Dush, M. K. and Martin, G. R. (1998) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**, 8998-9002.
- Goldberg, A. L. and Gaff, S. A. (1986) The selective degradation of abnormal proteins in bacteria; in *maximizing Gene Expression*, reznikoff, W. and Gold, L. (eds), pp. 287-314, Butterworths, Burlington, USA.
- Howe, D. K. (2001) Initiation of a *Sarcocystis neurona* expressed sequence tag (EST) sequencing project: A preliminary report. *Vet. Parasitol.* **95**, 233-239.
- Isreali, D. M. and Remington, J. S. (1993) Toxoplasmosis in the non-AIDS immunocompromised host. *Curr. Clin. Trop. Infect. Dis.* **13**, 322.
- Joiner, K. A. and Roos, D. S. (2002) Secretory traffic in the eukaryotic parasite *Toxoplasma gondii*: less is more. *J. Cell Biol.* **157**, 557-563.
- Li, L., Brunk, B. P., Kissinger, J. C., Pape, D., Tang, K., Cole, R. H., Martin, J., Wylie, T., Dante, M., Fogarty, S. J., Howe, D. K., Liberator, P., Diaz, C., Anderson, J., White, M., Jerome, M. E., Johnson, E. A., Radke, J. A., Stoeckert Jr., C. J., Waterston, R. H., Clifton, S. W., Roos, D. S. and Sibley, L. D. (2003) Gene discovery in the Apicomplexa as revealed by EST sequencing and assembly of a comparative gene database. *Genome Res.* **13**, 443-454.
- Luft, B. J. and Remington, J. S. (1992) Toxoplasmic encephalitis in AIDS. *Clin. Infect. Dis.* **15**, 211-222.
- Manger, I. A., Hehl, A., Parmley, S., Sibley, L. D., Marra, M., Hillier, L., Waterston, R. and Boothroyd, J. C. (1998) Expressed sequence tag analysis of the bradyzoite stage of *Toxoplasma gondii*: Identification of developmentally regulated genes. *Infect. Immun.* **66**, 1632-1637.
- Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**, 1-6.
- Ng, S. -T., Jangi, M. S., Shirley, M. W., Tomley, F. M. and Wan, K. -L. (2002) Comparative EST analyses provide insights into gene expression in two asexual developmental stages of *Eimeria tenella*. *Exp. Parasitol.* **101**, 168-173.
- Pain, V. M. (1996) Initiation of protein synthesis in eukaryotic cells. *Eur. J. Biochem.* **236**, 747-771.
- Parmley, S. F., Sgarlato, G. D. and Remington, J. S. (1993) Genomic and corrected cDNA sequence of the P28 gene from *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **57**, 161-166.
- Prince, J. B., Auer, K. L., Huskinson, J., Parmley, S. F., Araujo, F. G. and Remington, J. S. (1990) Cloning, expression, and cDNA sequence of surface antigen p22 from *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **43**, 97-106.
- Roos, D. S., Donald, R. G. K., Morrisette, N. S. and Moulton, A. L. C. (1994) Molecular tools for genetic dissection of the protozoan parasite *Toxoplasma gondii*. *Methods Cell Biol.* **45**, 27-63.
- Seeber, F. (1997) Consensus sequence of translational initiation sites from *Toxoplasma gondii*. *Parasitol. Res.* **83**, 309-311.
- Strong, W. B. and Nelson, R. G. (2000) Preliminary profile of the *Cryptosporidium parvum* genome: an expressed sequence tag and genome survey sequence analysis. *Mol. Biochem. Parasitol.*

- 107**, 1-3.
- Su, Y.-C., Lim, K.-P. and Nathan, S. (2003) Bacterial expression of the scFv fragment of a recombinant antibody specific for *Burkholderia pseudomallei* exotoxin. *J. Biochem. Mol. Biol.* **36**, 493-498.
- Wan, K. -L., Blackwell, J. M. and Ajioka, J. W. (1996) *Toxoplasma gondii* expressed sequence tags (ESTs): Insight into tachyzoite gene expression. *Mol. Biochem. Parasitol.* **84**, 203-214.
- Wan, K. -L., Chong, S. -P., Ng, S. -T., Shirley, M. W., Tomley, F. M. and Jangi, M. S. (1999) A survey of genes in *Eimeria tenella* merozoites by EST sequencing. *Intl. J. Parasitol.* **29**, 1885-1892.