**In-silico** structural, functional and immunogenic characterization of *Taenia solium* TS14 protein

Chitra Joshi and Siddharth Gautam*


Received: 26-09-2016
Accepted: 15-03-2017

**ABSTRACT**

TS14, a *Cysticercosis cellulosae* derived protein, has been exploited for immunodiagnosis of cysticercosis in humans and pigs. However, the information on structure, function, stability and immunogenicity of TS14 derived from different isolates is primarily lacking. The present study deals with *in-silico* characterization of six TS14 isolates. High thermostability and an isoelectric point of 9.41 were recorded. Based on N-terminal amino acid residues, high resistance to intracellular proteases with extended *in-vivo* and *in-vitro* half-lives was predicted. TS14 is foreseen as a secretory protein with a signal peptide and an extracellular localization. Structural analysis of TS14 exhibited the dominance of helices in the secondary structure (92% coverage) with majority of residues showing high and medium solvent accessibility. High lysine content and presence of multiple nucleotide binding sites in TS14 suggests interaction with RNA/DNA and a role in their metabolism. Immunogenic profiling predicted presence of four distinct B-cell epitopes. Mutational analysis based on the single amino acid substitutions among six TS14 isolates demonstrated minor variations in structural stability; however, all the substitutions were well tolerated. Moreover, all the isolates revealed almost identical immunogenic profile with an equivocal potential to elicit the antibody-mediated immune response.

**Key words**: B-cell epitopes, Cysticercosis, *In-silico*, Mutations, *Taenia solium*, TS14.

**INTRODUCTION**

Neurocysticercosis (NCC) is a serious parasitosis threatening human life and is a subject of major public health concern, particularly in the developing countries (Wandra et al., 2000; Deckers and Dorny, 2010). NCC has been incriminated as the primary cause of active epilepsy in the developing world, with a prevalence reaching up to 53.8% (Del Brutto et al., 2005; Montano et al., 2005; Nicoletti et al., 2005). Interestingly, NCC is now becoming an important (re-) emerging zoonosis in developed countries due to the immigration of infected individuals from endemic areas (Roman et al., 2000; White Jr and Atmar, 2002; Sorvillo et al., 2007).

Pig is the intermediate host and an important source of infection for humans, and thus, it is important to diagnose cysticercosis in both the species. Several serological techniques have been described to detect *T. solium* infection specific antibodies in humans and pigs, viz. radioimmunoassay, hemagglutination test, complement fixation test, dipstick assay, latex agglutination, ELISA and immunoblot (D’Souza and Hafeez, 1999; Rocha et al., 2002; Biswas et al., 2004). Currently, among the antibody-based diagnostic assays, enzyme-linked immunoelectrotransfer blot (EITB) is considered as the gold standard which uses lentil lectin purified glycoproteins (LLGP) as antigen with the diagnostic specificity of 100%, and sensitivity of 98%. This antigen consists of seven major diagnostic proteins in the molecular weight range of 13–50 kDa (Tsang et al., 1991; Del Brutto et al., 2001) and reaction against any one of them is regarded as confirmatory diagnosis for cysticercosis.

Among various proteins analyzed, TS14, a native 14 kDa polypeptide, has been widely exploited in immunoblot assays for detection of antibodies specific to *Cysticercus cellulosae* in humans (Greene et al., 1999). Subsequently, TS14 has also been employed in an ELISA format for serodiagnosis of *T. solium* cysticercosis in humans and pigs (da Silva et al., 2006; da Silva et al., 2012). Presently, immunodiagnosis of cysticercosis using TS14 is considered as one of the most promising methods for detection of anti-cysticercosis antibodies in serum and cerebrospinal fluid. Although TS14 has been extensively exploited for diagnostic purpose, its physicochemical, structural and functional attributes have not been characterized yet. Besides, no information on the immunogenicity of TS14 is available. Moreover, the information regarding the stability of different isolates of TS14, which could be utilized as diagnostic antigens in different regions of the world, is primarily lacking. For this
reason, the present study deals with the in-silico characterization of TS14 using bioinformatics tools in order to reveal the unknown attributes of TS14 which would facilitate more efficient utilization of TS14 as a diagnostic antigen.

MATERIALS AND METHODS

Physicochemical characterization, structural modeling, and visualization of TS14: Physicochemical properties of TS14 including approximate half-life, molecular weight, isoelectric point, amino acid composition, aliphatic index and extinction coefficient were estimated through ProtParam tool (http://web.expasy.org/protparam/) of ExPASy, a bioinformatics resource portal (Gasteiger et al., 2005). For the estimation of protein secondary structures, the amino acid sequence corresponding to the TS14 Bareilly isolate (Accession no. AIM56899.1) was submitted to the QUARK (http://zhanglab.ccmb.med.umich.edu/QUARK/) and Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) online servers. RaptorX (http://raptorx.uchicago.edu/StructurePrediction/predict/) was employed for the prediction of solvent accessibility (ACC) of amino acid residues and the results were presented as 3 states viz. buried, medium, and exposed; based on two cutoff values, 25%, and 75%. Residues less than 25% ACC were considered as buried and those above 75% ACC as exposed residues. Prediction of possible N- and O-glycosylation site was done through open access server GlyoEP (http://www.imtech.res.in/raghava/glycoep/submit.html) with a threshold kept at 0.5 to attain high confidence (Chauhan et al., 2013).

Since the tertiary (3D) structure of TS14 was not available in the PDB database, the amino acid sequence was submitted to several protein structure predicting programs including QUARK, Phyre2 and RaptorX for prediction of the 3D structure. The structural models generated by these servers were validated for their structural stability through SAVES (http://nihserver.mbi.ucla.edu/SAVES/). SAVES is a meta server integrating multiple programs and the best predicted 3D model was chosen based on the scores of ERRAT (Colovos and Yeates, 1993), VERIFY_3D (Luthy et al., 1992) and PROVE (Pontius et al., 1996). Visualization of the best TS14 model was achieved by Pymol (http://www.pymol.org/).

Assessment of biological and functional characteristics: For the prediction of TS14 function, the amino acid sequences and the best 3D model generated were submitted to the ProFunc (http://www.ebi.ac.uk/thornton-srv/databases/ProFunc) web server. Additionally, RaptorX and Phyre2 were also employed for predicting the possible functionality of TS14 based on the structural homology.

Sub-cellular localization of proteins, an important determinant of biological function, was assessed through freely accessible TargetP 1.1 server (http://www.cbs.dtu.dk/services/TargetP/) which assigns sub-cellular location based on the presence of the N-terminal pre-sequences (Emanuelsson et al., 2000). Sub-cellular localization was further confirmed by LocTree3 using Predict Protein open access server (https://www.predictprotein.org/). LocTree3 integrate de novo and homology-based predictions to achieve a prediction accuracy of over 80% for eukaryotes (Goldberg et al., 2014). Furthermore, to examine the possibility of a membrane-associated function, the presence of any glycosphatidylinositol (GPI) anchor within the amino acid chain was assessed by big-PI Predictor (http://mendel.imp.ac.at/sat/gpi/gpi_server.html) (Eisenhaber et al., 1999).

Molecular dynamic simulation of TS14: Protein functions depend largely on their interaction with other molecules, which in turn is mediated by specific binding sites on the surface of a protein molecule. Multiple tools were employed for the prediction of ligand binding sites of TS14 such as RaptorX-Binding (http://raptorx.uchicago.edu/BindingSite/), Predict Protein (Yachdav et al., 2014) and ProFunc (http://www.ebi.ac.uk/thornton-srv/databases/profunc/). RaptorX-Binding determines the ligand binding site by aligning the functional domains of query molecule against the sequences and structures available in the template library. Predict Protein, on the other hand, employs a neural network based tool ISIS2 for determination of protein-protein binding sites (Ofran and Rost, 2007) and SomeNA for protein-polynucleotide binding sites.

Identification of B-cell epitopes, the actual sites interacting with the antibodies, was achieved through open access server IEDB Analysis Resource (http://tools.iedb.org/bcel/) with Bepipred linear epitope prediction method (Larsen et al., 2006). The threshold was fixed at 0.35 with expected specificity and sensitivity of 75% and 50% respectively.

Mutational analysis: Numerous TS14 protein sequences derived from different Taenia solium isolates are available in public domain, with slight variations in their amino acid profile. The amino acid sequences of five different TS14 isolates, based on their evolutionary relationship with TS14 Bareilly isolate, were retrieved from NCBI (DQ864645 China; JP906195 China; AF356336 China; AF082829 Peru and AF257776 Mexico). Multiple sequence alignment was done to delineate the divergent amino acid residues. The impact of single amino acid substitutions on the stability of different TS14 isolates was assessed through online server 1-Mutant2.0 with default settings, except for the increment of temperature to 25 °C (http://folding.biofold.org/cgi-bin/i-mutant2.0.cgi). Considering the potential of a single amino acid substitution in altering protein functions, the effect of single site mutations among different TS14 isolates was...
assessed through SIFT (Sorting Intolerant From Tolerant), a homology-based open access server (http://sift-dna.org).

Considering the diagnostic significance of TS14, it is important to understand if there is any variability in the immunogenic potential of TS14 across different isolates. For this reason, IEDB Analysis Resource was employed to ascertain any changes in the B-cell epitopes due to amino acid substitutions at specific locations. The mutational analysis including introduction of site specific amino acid substitutions and associated changes in the polar contacts was achieved through Pymol. All the figures were generated and visualized through Pymol.

RESULTS AND DISCUSSION

Physicochemical properties: Expasy ProtParam tool was utilized for estimation of physicochemical properties of TS14 Bareilly isolate. TS14 exhibited a molecular weight of 9625.2 Daltons and an isoelectric point of 9.41. Amino acid composition reveled that TS14 is rich in lysine residues (16.5%) and contains 19 positively charged residues (Table 1). The extinction coefficient of the protein which is an indicator of light absorbed by protein was found to be 0.881 M-1cm-1 and could be utilized during the purification process. Estimation of half-life revealed a persistence of TS14 for as long as 30 hr in the mammalian cells. The instability index was found to be 14.0, way below the instability threshold of 40.0, further revealing the high stability of TS14. Additionally, high thermostability of TS14 protein, as suggested by aliphatic index of 82.59, was also observed.

Predicted biological functions and immunogenic typing of TS14: SomeNA determined presence of multiple nucleotide binding sites across the TS14 molecule. Based on the gene ontology assigned by ProFunc (GO terms: 0008152; 0005488; 0000166) and its ability to interact with nucleotide molecules, TS14 is predicted to have metabolic functions, primarily associated with the regulation of DNA/RNA metabolism and protein synthesis. Additionally, TS14 was also aligned against tryptophanyl-tRNA synthetase molecule of two different protozoan parasites (PDB id: 4J75 and 3I05) with sequence coverage ranging from 72-75% and a structural identity of 21-25%. Search against DNA-binding site by ProFunc revealed a structural similarity of 83.3%

---

**Table 1: Amino acid composition of TS14 Bareilly isolate**

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Number of residues</th>
<th>%</th>
<th>Amino acid</th>
<th>Number of residues</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A)</td>
<td>8</td>
<td>9.4</td>
<td>Leu (L)</td>
<td>5</td>
<td>5.9</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>2</td>
<td>2.4</td>
<td>Lys (K)</td>
<td>14</td>
<td>16.5</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>2</td>
<td>2.4</td>
<td>Met (M)</td>
<td>2</td>
<td>2.4</td>
</tr>
<tr>
<td>Asp (D)</td>
<td>3</td>
<td>3.5</td>
<td>Phe (F)</td>
<td>3</td>
<td>3.5</td>
</tr>
<tr>
<td>Cys (C)</td>
<td>1</td>
<td>1.2</td>
<td>Pro (P)</td>
<td>4</td>
<td>4.7</td>
</tr>
<tr>
<td>Glu (E)</td>
<td>8</td>
<td>9.4</td>
<td>Ser (S)</td>
<td>4</td>
<td>4.7</td>
</tr>
<tr>
<td>Gly (G)</td>
<td>3</td>
<td>3.5</td>
<td>Thr (T)</td>
<td>5</td>
<td>5.9</td>
</tr>
<tr>
<td>His (H)</td>
<td>3</td>
<td>3.5</td>
<td>Trp (W)</td>
<td>1</td>
<td>1.2</td>
</tr>
<tr>
<td>Ile (I)</td>
<td>5</td>
<td>5.9</td>
<td>Tyr (Y)</td>
<td>2</td>
<td>2.4</td>
</tr>
<tr>
<td>Val (V)</td>
<td>8</td>
<td>9.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Post-translational modifications and structural attributes: Consensus results from Phyre2 and QUARK revealed that TS14 secondary structure is dominated by helices with an overall coverage of 92%. Rest of the regions was occupied by coils with no sheets in the protein molecule (Fig 1). The ability of individual residues to interact with other cellular components was determined by the solvent accessibility which revealed 22% of residues as highly exposed, 40% with medium accessibility and 38% as buried residues. Prediction of glycosylation site revealed one potential N-linked glycosylation site at position 29th (Asparagine) and one potential O-linked glycosylation site at 84th (Threonine) position of the amino acid sequence. Based on the scorings of SAVES, the 3D model of TS14 generated by QUARK was found to be more stable than those predicted by Phyre2 and RaptorX. The best tertiary structure of TS14 is shown in Fig 2.

---

**Fig 1.** 3D model of TS14 revealing different secondary structures showing majority of structural coverage by helices (blue) and the left over length by loops (pink). Predicted TS14 models lacked any sheet structure.
with a human spliceosomal protein hPrp31 (PDB id: 3siv) which is known to regulate gene expression through RNA splicing. Scanning the TS14 structure for enzyme catalytic sites revealed 32.94% sequence identity and 53.4% structural identity with the catalytic site of dethiobiotin synthetase enzyme (PDB id: 1dah) which is an important enzyme in biotin metabolism.

Regarding the sub-cellular localization, TS14 is foreseen as a secretory protein with a signal peptide. This was further confirmed by gene ontology which stated extracellular localization of TS14 with a prediction confidence of 60% (GO term: 0005576). No potential GPI modification site could be determined across the chain of TS14. Except for the polynucleotide binding sites, no specific ligand binding site of major significance could be determined on TS14.

Prediction of B-cell epitopes on TS14 revealed the presence of four distinct antigenic determining sites (Fig 3). The largest epitope consisted of 16 amino acid residues. The rest of the three epitopes were smaller and one of them consisted of 4 residues whereas the rest of two were just 2 residues long (Table 2). Two epitopes were very close to each other with atoms of one only 5.7 Å apart from the other.

**Mutations and protein stability**: Considering TS14 Bareilly isolate as the prototype molecule, multiple sequence alignment revealed three amino acid substitutions; P replaced by L in all other isolates at position 10, S replaced by G in Peru and Mexican isolates at position 48, and K replaced by E in one Chinese isolate (accession no. JF906195) at position 68. Interestingly, out of the three amino acid mutations observed, two lied in the B-cell epitopic sites. Based on the free energy change, all the three substitutions were found to decrease the overall stability of the protein structure when compared with the prototype molecule, suggesting that TS14 Bareilly isolate represents the most stable form of the protein. *In-silico* mutagenic analysis also demonstrated that site specific mutations do not introduce any extra polar interactions within the molecule (Fig 4). However, a clear-cut difference in the length of H-bonds with neighboring atoms was evident due to mutations at positions 10 and 48. The length of one H-bond at Gly48 in Peru and Mexican isolates was 2.4 Å as compared to 2.5 Å in Ser48 in rest of the isolates (Fig 4B). Likewise, a difference of 0.4 and 0.9 Å was observed among two H-bonds at position 10 (Fig 4A).

The impact of amino acid substitutions on protein function was assessed by SIFT which revealed that all the substitutions were well tolerated and do not cause any phenotypic alteration in the protein. Substitution of amino acids at position 10 and 48 doesn’t seem to alter the number and length of B-cell epitopes in the TS14 molecule. However, substitution of K by E at position 68 in Chinese isolate led to an increase in the length of one B-cell epitope by two residues (Table 2).

TS14 is a 14 KD glycoprotein that has been widely exploited for detection of cysticercosis in humans as well as pigs. In practice, detection of antibodies is generally preferred over antigen detection because of a longer persistence of antibodies in the body fluids. Antigen detection, on the other hand, supplements antibody detection and confirms an active infection i.e. presence of live parasites. Both diagnostic methods, however, require a thorough understanding of the targeted protein. Although TS14 has been implicated in cysticercosis diagnosis for over two decades, its molecular and functional characterization remains largely unexplored.

The stability of the targeted antigen is a critical factor in any immunodiagnostic assay. On physicochemical characterization, we observed a reasonably high *in-vivo* and *in-vitro* stability of TS14. This is reflected by high resistance.
Table 2: Predicted B-cell epitopes in *Taenia solium* TS14 molecule

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Starting position</th>
<th>End position</th>
<th>Amino acid residues</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>34</td>
<td>AEKNKPKDVANSTKKG</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>46</td>
<td>49</td>
<td>PISK</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>63</td>
<td>64</td>
<td>LE</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>67 (66*)</td>
<td>68 (69*)</td>
<td>GK (KGEI*)</td>
<td>2 (4*)</td>
</tr>
</tbody>
</table>

*JP906195 China*

to intracellular proteases as well as high stability in the extracellular environment. Additionally, TS14 is also foreseen to withstand high temperatures without any loss of structural integrity. The high stability of TS14 may be partially explained by the presence of two potential glycosylation sites. Post-translational glycosylation has been reported to assist protein folding, stabilization and protection against proteases (Dwek, 1998; Dalziel *et al.*, 2014). In view of the diagnostic application, high stability is a desirable character and highlights the usefulness of TS14 as a targeted antigen for the detection of antibodies. This also becomes important particularly in developing countries where refrigeration of diagnostic antigens and maintenance of cold chain for long is not readily feasible.

Through *in-silico* protein function prediction, we made an effort to determine the biological function of TS14. In doing so, we observed that TS14 may be a key player in cellular metabolism, particularly nucleic acid metabolism. It closely resembled the tryptophanyl-tRNA synthetase of two protozoan parasites, an enzyme essential for protein synthesis machinery. Additionally, TS14 also showed high structural similarity with human Prp31 which is a key component of RNA splicing and provides stability to the small nuclear ribonucleoproteins (Makarova *et al.*, 2002; Schaffert *et al.*, 2004). Moreover, high lysine content (16.5%) of TS14 makes it more positive, favoring its interaction with RNA/DNA. These findings suggest that TS14 may be a key component in the regulation of pathways involved in RNA processing and protein synthesis, and further studies in *in-vivo* and *in-vitro* models are required for a deeper understanding of its biological functions.

It was evident that TS14 is an extracellular protein as it lacked any GPI modification site and possesses a signal peptide. The present result is in compliance with Hancock *et al.*, (2003) who reported that TS14 is an extracellular secreted protein which accumulates in the cyst fluid and lacks glycosylphosphatidylinositol-anchor attachment sites (Hancock *et al.*, 2003). The extracellular localization becomes important because secreted proteins are more soluble and have better access to the ligands as compared to the intracellular proteins and thus making them biologically more active (Mergulhao *et al.*, 2005). Assessment of antigenic epitopes revealed the presence of four distinct regions that could be recognized by the B-cells paratopes. TS14 isolates derived from six different geographical regions were examined and all of them revealed same epitopic sites, suggesting their relevance as a potent diagnostic antigen.

Site-directed mutagenesis was also performed to unravel the influence of single amino acid substitutions on the stability and immunogenicity of TS14 derived from
different isolates. We found that substitutions of amino acid residues alter the stability of protein modestly. Among the analyzed isolates, Bareilly isolate was found to represent the most stable form and a slight drop in the overall structural stability were observed in other isolates due to single site mutations. Also, marginal changes in the length of polar interactions were observed among different isolates. However, these changes were modest and well tolerated and do not lead to any alterations in the function and epitopic sites of the protein molecule. Amino acid substitutions were observed in two of the epitopic sites. This can greatly affect the diagnostic implications of TS14 antigen; as the antibodies generated against the epitope of one isolate may fail to recognize the same epitope of other isolates. However, considering that there is no difference in the length and number of polar interactions among native and mutated amino acid residues, the antibodies generated against one isolate may recognize the antigens of other isolates, and further studies are indicated to confirm the same. Conclusively, the results of the present study suggest that TS14 protein derived from different isolates share almost identical immunogenic profile and can be exploited for immune-diagnosis of cysticercosis around the globe. Since the present study involves in-silico analysis without generation of any empirical data, further studies to validate the results in in-vitro and in-vivo models is highly recommended.

ACKNOWLEDGEMENT

The authors are grateful to Director, Indian Veterinary Research Institute (IVRI), Izatnagar, India, and Dr. B.C. Saravanan, Senior Scientist, Division of Parasitology, IVRI, for providing laboratory facilities for molecular characterization of TS14 which paved the foundation for this work.

REFERENCES


