Homology modeling of putative thioredoxin from *Helicobacter pylori*

D Premalatha, P Ravindra\(^1\) and L Venkateshwar Rao*\(^*\)

Department of Microbiology, University College of Science, Osmania University, Hyderabad 500 007, India

\(^1\)School of Engineering and Information Technology, University Malaysia Sabah, 88999 Kota Kinabalu, Sabah, Malaysia

Received 12 June 2006; revised 27 November 2006; accepted 10 January 2007

The tertiary structure of putative thioredoxin (trx) of *Helicobacter pylori* was generated based on structural homology of the X-ray crystallographic structure of thioredoxin from *Escherichia coli*. Inspection of theoretically predicted structure indicates that the thioredoxin of *H. pylori* is similar to that of *E. coli*. Analysis of the structure revealed that thioredoxins have a common fold, characterized by a core of twisted \(\beta\)-pleated sheet flanked either side by helices. The amino terminal end of the molecule is occupied by \(\beta\)-\(\alpha\)-\(\beta\) motif and carboxy terminal end by \(\beta\)-\(\beta\)-\(\alpha\) motif. This molecule is characterized by five strands and four helices. Among the four helices, \(\alpha_2\) is the longest helix which was disrupted near proline. Proline72 is identified as \textit{cis}-proline. This structure retained overall trx-fold with the conservation of global shape and the secondary structures. This work determines the structure of thioredoxin and is found to be unique for further insight into molecular characterization.

**Keywords**: BLAST, PDB, thioredoxin, RMSD

**IPC Code**: Int. Cl. G01N33/00

**Introduction**

Thioredoxins are small, heat stable proteins with molecular mass ranging from 10 to 12 kDa. They are present in a variety of living organisms from prokaryotes to human beings and characterized by an active site at the amino terminus containing two cysteine residues separated by two amino acids, i.e. Cys-X-X-Cys where X can be any amino acid\(^1\).

Generally, thioredoxin functions as a general protein-disulfide reductase and its redox activity resides in the sequence of its conserved active site, Cys-X-X-Cys, which undergoes reversible oxidation of the cysteine residues from a dithiol to a disulfide form\(^1\). The reversible oxidation of the cysteine residues to the disulphide form serves as a redox couple for a vast spectrum of biological processes from cell division to enzyme regulation. The mechanism of action involves initial nucleophilic attack by thiol group (Cys32) on one of the sulfur atoms in the substrate, yielding an unstable transient mixed disulphide. The second thiol group (Cys35) displaces the sulphur atom of the substrate to generate oxidized thioredoxin and reduces the substrate\(^5\).

The active-site-residues of all known thioredoxin molecules are highly conserved but the remainder portion of it may vary. Several studies have indicated the multifunctional role of thioredoxin. The role of thioredoxin in the reduction of methionine sulfoxide, for the first time, was described in yeast\(^3\). Thioredoxin was found to participate in the synthesis of deoxyribonucleotides by the reduction of ribonucleotides\(^4\). It was also demonstrated that *Escherichia coli* thioredoxin was essential for bacteriophage T7 replication. It served as a sub unit of phage DNA polymerase\(^5\) and for the assembly of filamentous phage\(^6\). In plants, it controlled the activity of different photosynthetic enzymes\(^7\).

Thioredoxin from different species has shown 27-69% sequence identity to that of *E. coli* thioredoxin. The structural studies of thioredoxin from different organisms represented a thioredoxin fold\(^8\). Thioredoxin from phage T\(_4\) has also similar fold but with certain variations in the primary structure\(^3\). Thioredoxin of *E. coli* has so far been studied most extensively\(^9,10\) and represented a reference model for other studies. The 3D-structure of oxidized form of *E. coli* thioredoxin was studied by X-ray crystallography at 2.8Å resolution\(^10\); later the resolution was extended to 1.68Å (Ref. 11). The protein contained a polypeptide chain with a core of twisted \(\beta\)-sheet flanked on either side by helices\(^10,11\). However, the 3D-model of thioredoxin from *Bacillus acidocaldarius* (Bac trx), constructed by a molecular...
modeling technique, indicated larger first helix region. A study on H. pylori thioredoxin system has suggested it a stress response protein, which helped the organism to inhabit impermeable mucus gel. As the function of a protein is mostly related with its 3D-structure, the present study is the motivation for the above observation. The present communication summarizes the structural features of the putative thioredoxin of H. pylori, a gastric pathogen, employing in silico methods.

Materials and Methods

Computational Methods
The computational methods of 3D-model building involved template selection, alignment of template with the target, building of the model and evaluation of the structure. Putative thioredoxin sequence of H. pylori was mined from Swissprot database.

Selection of Template and Sequence Alignment
Different on-line tools were used to select the template. The PDB (Brookhaven Protein Data Bank) database was extensively screened using BLAST (Basic Local Alignment Search Tool) server, developed and maintained at Adam Godzik's laboratory at the Burnham Institute, to find out the distantly related homologues of query sequence. PDB database was searched by PSI-BLAST using a profile generated from NR database for proteins exhibiting similarity to the unknown structure. Putative sequence was submitted to FUGUE server (www.cryst.bioc.cam.ac.uk) situated at University of Cambridge, United Kingdom, for recognizing distant homologues by sequence-structure comparison. Putative sequence was submitted to FUGUE server (www.cryst.bioc.cam.ac.uk) situated at University of Cambridge, United Kingdom, for recognizing distant homologues by sequence-structure comparison. The sequence was also submitted to fold recognition server, 3d-pssm, (http://www.sbg.bio.ic.ac.uk/~3dpssm) situated at Imperial College of Science, Technology and Medicine, United Kingdom. The 3d-pssm results contain the scores and alignments for the 20 most probable matches. Template was selected with the help of search algorithms. The alignment of protein sequence was constructed by using the option CLUSTAL-X.

Model Building, Evaluation and Validation of Model
The atomic co-ordinates of the crystallographic structure of the oxidized form of E. coli thioredoxin (chain-A) at a resolution of 1.68Å was obtained from Protein Data Bank files. Structure was constructed using MODELLER software, which builds the model based on satisfaction of spatial restraints. Of 20 models generated, the representative model has the lowest objective function. This model was refined by loop building using SWISS PDB VIEWER software and energy minimization by sibyl (Tripos Inc., 1699, South Hanley Rd., St. Louis, Missouri, 63144, USA) installed on Silicon Graphics Workstation. Stereochemical quality of the representative model was evaluated by means of PROCHECK and VERIFY3D. Ribbon diagram of 3D-model was generated by MOLMOL.

Results
The 3D-structure of putative thioredoxin of H. pylori (Fig. 1) was built by homology modeling based on E. coli thioredoxin. The closest homologue with highest sequence identity of 35% and highest score of 101 was selected as representative model (Table 1). This model possessed the lowest objective function of 383.3299. The model discloses a single polypeptide chain with major secondary structural elements, α-helices and β-sheets (Fig. 1; Table 2). Structure reveals β-α-β motif at amino terminal end and β-β-α motif at carboxy terminal end. Central core of the structure is occupied by β-sheet of five strands flanked by four helices. Among five strands, four are parallel and one is antiparallel; while among four helices, second is the largest helix. Cis-proline,
Table 1 — Data showing closest thioredoxin homologue of H. pylori putative thioredoxin from PDB-BLAST

<table>
<thead>
<tr>
<th>PDBCODE</th>
<th>Protein</th>
<th>Chain</th>
<th>Identity to trx of H. pylori</th>
</tr>
</thead>
<tbody>
<tr>
<td>2TRX</td>
<td>Crystal structure of E. coli thioredoxin</td>
<td>Chain A</td>
<td>35%</td>
</tr>
<tr>
<td>1XOA</td>
<td>NMR-structure of thioredoxin from E. coli</td>
<td>Chain A</td>
<td>35%</td>
</tr>
<tr>
<td>1DBY</td>
<td>NMR-structure of thioredoxin from Chlamydomonas Reinhardtii</td>
<td>Chain A</td>
<td>32%</td>
</tr>
<tr>
<td>2TIR</td>
<td>Crystal structure analysis of thioredoxin from E. coli</td>
<td>Null</td>
<td>35%</td>
</tr>
</tbody>
</table>

Table 2 — Contents of α-helices and β-strands of H. pylori putative thioredoxin

<table>
<thead>
<tr>
<th>β-strand</th>
<th>Amino acids</th>
<th>α-helix</th>
<th>Amino acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>β₁</td>
<td>Leu2, Glu3, Val4</td>
<td>α₁</td>
<td>Tyr10, Ala11, Glu12, Lys13</td>
</tr>
<tr>
<td>β₂</td>
<td>Ala18, Val19, 20, 21, Asn22, Val23, Gly24</td>
<td>α₂</td>
<td>Pro29, Asp30, Cys31, Arg32, Lys33, Ile34, Ile37, Met38, Glu59, Asn60, Leu41, Ala42, Lys43, Thr44</td>
</tr>
<tr>
<td>β₃</td>
<td>Glu50, Phe51, 52, Lys53, Val54, Ser55</td>
<td>α₃</td>
<td>Leu62, Lys63, Glu64, Ser65, Leu66</td>
</tr>
<tr>
<td>β₄</td>
<td>Thr73, Leu74, Ile75, Phe76, Tyr77, Lys78</td>
<td>α₄</td>
<td>Gln93, Lys94, Pro95, Ile96, Glu97, Asp98, Ala99, Leu100, Lys101, Ala102, Leu103</td>
</tr>
<tr>
<td>β₅</td>
<td>Lys81, Glu82, Val83, Gly84, Glu85, Arg86</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

important for the stability of the molecule, is the proline72. Two cysteins are separated by residues Cys-Pro-Asp-Cys and located at the down stream of β₂ strand.

Validation of Structure

The φ and ψ distribution values of Ramachandran plot of non-glycine and non-proline residues after refinement of the model with loop building, using SWISS PDB VIEWER, and subjecting to energy minimization, using Tripos force field, are presented in Table 3. Altogether 100.0% of the residues were in favoured and allowed regions (Fig. 2).

Discussion

Thioredoxins (trx) are conserved proteins and distributed in the living organisms from prokaryotes to humans. Thioredoxin-sequences from different species have been reported to have 27-69% sequence-identity with that of E. coli thioredoxin with trx-fold8. In the present study, putative thioredoxin sequence of H. pylori exhibited 35% sequence-identity with that of E.coli thioredoxin hinting the possibility of possessing trx-fold. The protein model built in the present study retains a general trx-fold, i.e. β-pleated sheet flanked either side by helices. Further inspection of the structure indicates that larger part of the model,
including the largest second helix region with a break near the helix breaker—proline, is similar to that of E. coli thioredoxin derived by X-ray crystallography. However, the model differs from the 3D-model of B. acidocaldarius, which had the largest first helix region. Another confirmatory evidence obtained by the overall superposition of backbone atoms of H. pylori thioredoxin to the structure of E. coli thioredoxin showing an average RMSD (root mean square deviation) of 0.52Å. This indicates that the molecular topology is globally similar. The contents of α-helices and β-strands constitute 32.6% and 26.6% of total amino acids, respectively.

Analysis of PROCHECK reveals that all residues are within the limits of Ramachanran plot. So, it can be considered a good model. Loop building with SWISS PROTEIN DATABANK VIEWER and energy minimization with SYBYL (Tripos Inc., 1699, South Hanley Rd., St. Louis, Missouri 63144, USA) carried out for the model resulted in the model with reduced energy. According to thermodynamic hypothesis, native conformation of the proteins corresponds to global minima of their free energy. This fold might be the native conformation of the putative thioredoxin, which is trying to follow thermodynamic hypothesis. Proteins, described as the servants of life, can perform their function efficiently in native state only.

**References**


