# Structural analysis of the *Babesia microti* thioredoxin reductase: a potential drug target for babesiosis treatment.

## Neelima Arora<sup>1\*</sup>, Amit Kumar Banerjee<sup>2</sup>, Mangamoori Lakshmi Narasu<sup>1</sup>

<sup>1</sup>Centre for Biotechnology (CBT), Institute of Science and Technology (Autonomous), Jawaharlal Nehru Technological University Hyderabad, Kukatpally, Hyderabad, Telangana, India

<sup>2</sup>Biology Division, CSIR-Indian Institute of Chemical Technology, Uppal Road, Tarnaka, Hyderabad, Telangana, India

#### Abstract

Babesia microti, the causative agent of human babesiosis which was endemic in United States and parts of Europe with sporadic cases in other regions is expanding its geographical range. Being a common transfusion threat, *Babesia* has become a cause of concern in researchers and epidemiologists. Human babesiosis is now recognized as an emerging disease and gains more attention now than ever before. Thioredoxin reductase, a promising drug target in apicomplexan parasite has been validated in several species. This study focused on *in-silico* analysis of physicochemical properties, functional and structural aspects of thioredoxin reductase of *B. microti*. Comparative modeling approach was adopted for developing the three dimensional structural model of Thioredoxin reductase of *Babesia microti*. The model developed was found to be of reasonably good stereo-chemical quality by structure validation servers. This study will provide valuable insights about the function and structure of the enzyme thioredoxin reductase of *B. microti* and aid in developing effective chemotherapeutic agents for control and treatment of *Babesia*.

Keywords: Babesia microti, Babesiosis, Thioredoxin reductase.

Accepted on April 28, 2018

## Introduction

Babesiosis in humans and animals is caused by apicomplexan zoonotic hematotropic parasites of the genus *Babesia* transmitted by ticks. *Babesia* is quite ubiquitous in nature with a wide range. *Babesia* is one of the most common blood parasites of mammals. Babesiosis is one of the most common infections of animals and causes profound economic, medical and veterinary impact globally. After years of being recognized as major pathogen in cattle causing huge economic losses, babesiosis has recently gained recognition as an important pathogen of human [1]. First case of babesiosis was reported in Croatia in 1957 [2]. Babesiosis caused by the *B. microti* in human in United States was first documented in 1969 [3].

Out of more than 100 known species of *Babesia*, only a fraction cause human babesiosis. Among these, genetically distinct *Babesia microti* is the predominant species in US and major culprit responsible for causing human babesiosis. The spectrum of Babesiosis ranges from asymptomatic to severe malaria-like symptoms including chills, sweats, headache, arthralgia, myalgia, anorexia, cough and occasionally causing deaths in humans. *Babesia* disease cycle progresses in two hosts involving black legged deer ticks of belonging to *Ixodes* genus as definite host and a vertebrate intermediate host primarily. *Peromyscus leucopus* while humans are accidental

hosts or dead-end hosts who acquire the infection with the bite of infected ticks. The disease is tightening its grip in many states of United States as well as other parts of the world including Europe and a few Asian countries. Co-morbidity of babesiosis with Lyme disease and possibility of underreporting in asymptomatic patients often compounds the situation. The fact that Babesiosis can be transmitted congenitally as well as through blood transfusion and is difficult to diagnose makes the situation worse [4,5]. Babesia can even survive for prolonged period in blood storage conditions. Babesiosis has emerged as a major transfusion transmitted threat of the recent times [5]. Babesia remains understudied unlike another apicomplexan parasite Plasmodium and Toxoplasma. Owing to increased number of incidences over past few years, human babesiosis is now classified as emerging disease [6]. CDC has included babesiosis as nationally notifiable diseases in 2011. Immunocompromised individuals like those suffering with hemoglobinopathy, HIV infection. cancer. organ transplantation are at high risk of contacting babesiosis and demonstrate more severe symptoms and persistent infection. Other factors that make one vulnerable to contacting babesiosis are age, drug regime of immunosuppressive drugs or recent splenectomy. The disease shows a seasonal trend coinciding with the seasonal activity of ticks. Like any other protozoan parasites, Babesia inhabits an oxygen rich environment inside

blood cells of mammalian host and faces a challenge of shielding its system against oxidative stress which may wreak havoc by damaging its membrane lipids, nucleic acids, and proteins. To counteract Reactive Oxygen Species (ROS), Babesia employs anti-oxidant systems including Thioredoxinthioredoxin reductase (Trx/TrxR) and Glutathione-Glutathione Reductase (GSH/GR). Thioredoxin system comprises of Thioredoxin reductase (TrxR), various thioredoxins and Thioredoxin-dependent peroxidases (TPx). Thioredoxin reductase has been biochemically characterized in Babesia [7]. Thioredoxin reductase belongs to a family of dimeric flavoenzymes. Thioredoxin reductase is an essential enzyme required for counteracting the oxidative stress and hence, in survival of these parasites. It is considered an attractive drug target as it plays key role in many cellular processes and its inhibition can affect many vulnerable stages in apicomplexan parasites. Not much is known about the functional and structural aspects of thioredoxin reductase of Babesia. Comparative modeling or homology modeling is an alternate method that aid in deriving structural insight about protein in absence of experimentally derived structure. Homology modeling has been used successfully for providing structural information of key enzymes in various instances in important human pathogens in past [8-17]. The enzyme of Babesia has pronounced and significant differences with the human counterpart. Hence, we undertook the exercise of developing a 3-dimensional model of Thioredoxin reductase enzyme of Babesia microti for obtaining insights about its structure and function

## **Materials and Methods**

## Sequence retrieval and physicochemical characterization

Uniprot database (http://www.uniprot.org/) was searched using keyword "*Babesia microti*" and thioredoxin reductase". The query yielded 7 entries. Primary sequence of the enzyme thioredoxin reductase of *Babesia microti* (Accession: A0A1R4AAX8) was collected in Fasta format from the database. Physicochemical properties of the enzymes were determined using Expasy-ProtParam and Protscale tool [18]. Charge and Pepstats of Emboss was used to determine position-wise distribution of charge mole percentage of various classes of amino acids in Thioredoxin reductase respectively.

## Propensity of crystallization

CRYSTALP2 web server available at [19] and Xtalpred [20] were used to predict the propensity of enzyme to crystallize.

## Functional characterization of thioredoxin reductase of Babesia microti

CYS REC server [21] that identifies the positions of cysteines, total number of cysteines and computes the most probable SS bond pattern of pairs in the protein sequence was used to analyse SS bonds in primary sequence.

Motifs were predicted using default parameters in MEME suite (Multiple Em for Motif elicitation) [22].

Consurf server was employed for identification of biologically important protein residues in the protein sequence [23,24]. As three-dimensional structure of the protein is not available in Protein Data Bank (PDB) (http://www.rcsb.org/pdb), this study aimed at development of 3D model of thioredoxin reductase of *Babesia microti* and conducting in-depth sequence and structural analysis.

#### Secondary structure prediction from target sequence

Secondary structure of a protein denotes the arrangement of residues in alpha helix, beta sheets, extended strands or turns. Secondary structure of the selected protein was predicted employing NPS server [25] using various methods viz. DPM [26], DSC [27], GOR4 [28], HNNC [29], PHD [30], Predator [31], SOPMA [32], SIMPA96 [33] and secondary consensus [34] keeping default parameters for 4 state predictions keeping output width=70.

## Prediction of protein disorder

Often post translational modifications, attachment of signal peptides, evolutionary path of a protein molecule are dictated by the disordered regions present in a protein molecule. Thus, identification of such important disordered regions is of immense importance to understand the structure and function of a protein. Disordered regions of the thioredoxin reductase were identified using DisEMBL [35], GLOBPLOT [36], RONN [37] and Protein Disorder Prediction System (PRDOS) [38] server.

## Generation of three dimensional models

Comparative modeling approach was adopted to derive a 3D model of thioredoxin reductase of *Babesia microti*. The spatial restraint based Modeller 9.19 version was used for this purpose. The complete modeling exercise included target based template searching and selection, target-template alignment, model development employing spatial restraint, model evaluation, loop optimization and final structural quality evaluation. For further assurance and better template consideration, we have also attempted parallel template selection through Swiss model [39].

## Selection of template

For template selection, the first approach model generation method of Swiss model [39] was used. On the basis of obtained result, high-resolution X-ray crystallography structure of thioredoxin glutathione reductase at a resolution of 2.30 Å (PDB ID: 2x99, A chain) [40] having 48.22% identity with the query sequence was selected as template for homology modeling exercise. In parallel, template searches were performed through Modeller 9.19 version [41] *via* profile development and searching. Reasonable number of hits (20; 1aogA, 1ojt, 3grs, 1trb, 1dx1A, 1ebdA, 1nhp, 1fecA, 1gesA, 1h6vA, 1jehA, 3ladA, 1lpfA, 1lvl, 1mo9A, 1onfA, 1vdc,

1xdiA, 1q1rA, 1xhcA) was obtained through this process and all the obtained templates were compared on the basis of their sequence length coverage (sequence identity) and crystal structure resolution.

### Target-template alignment

Out of the selected total number of templates (20 from Modeller analysis and 2x99A from SwissModel), further reduction in number of templates was done following the sequence identity and structural resolution based criterion. The following templates were selected: 1mo9A, 1xhcA, 1nhpA, 1xdiA, 11vIA, 1h6vA, 1q1rA. A template-template comparison was done to reach such conclusion. Target-template alignment was done using the Modeller based scripts (Modeller version 9.19). The template (2x99A) selected through the Swiss Model tool was also included in the final template list.

#### Model development and evaluation

The target sequence was carefully aligned with the generated multi-template alignment. Total 10 models were generated based upon the multi-template-target alignment. All the models were subjected to PDF, DOPE and GA341 evaluation. The best model was selected after carefully analyzing the obtained values, especially the DOPE scores. GA341 output suggested that the structures were similar to the original crystal structures in quality.

### Loop optimization and further structural evaluation

The best model was selected and visually inspected using VMD [42] and Chimera [43]. The observation suggested that the model contained multiple sheets and helices along with several loops. Considering the generation of multiple loops, multi-template based approach was adopted so as to generate a structure with comparatively shorter loops. Loop optimization was mandatory considering the presence of multiple loops in the structure. Based on the visual inspection and other external systematic evaluation (RAMPAGE [44], ERRAT [45] etc.), regions with loops having comparatively poor energy were subjected to refinement. Total 10 structures with loop refinement were generated and further evaluated. Obtained results pertaining to the template selection through structure development and evaluation are described in the result section.

## Structure validation

Followed by the structure validation through DOPE score, GA341 and modeler objective function, all the structures were further subjected to validation through RAMPAGE, PROSA (Protein Structure Analysis) and ANOLEA (Atomic Non-Local Environment Assessment). Analysis detail is provided in the result section.

## Secondary structural analysis from final protein structure

To understand the three-dimensional structure with details of secondary, tertiary and quaternary features, further structural analysis with PDBSUM and PROMOTIF [46] was performed. Detailed insight on all the sheets, helices, loops, hairpins, disulphide linkages along with specific motifs and orientation was obtained for the generated structure.

### Active site prediction

Identification of active site is important to understand the plausible pockets present in a protein which can be considered for safe binding sites for potential inhibitors which may aid in hindering the structural and functional mechanism of an important protein molecule belonging to a pathogen. With such objective, potential active site mapping was done using CASTp tool [47] to retrieve information regarding the microenvironment of the pockets of the target protein.

## Ligand selection and docking analysis

Antibiotic resistance has become a cause of concern globally. The drug pipeline should be replenished from time to time with novel natural or synthetic molecules to provide remedy to the patients from various infections. Similar situation demands for novel active molecules for *Babesia* too. Recent report by Harbut et al. [48] suggested that Auranofin is having excellent potential to exhibit bactericidal role affecting the thiol-redox homeostasis. Interestingly, it was found in the same study that Auranofin can be used for other gram positive bacteria too such as *Bacillus subtilis, Enterococcus faecalis, Enterococcus faecium* and *Staphylococcus aureus*. Auranofin functions through creating oxidative stress for the bacteria through diminishing their reducing capacity. Thus, this potential molecule may become a common purpose inhibitor for several gram positive pathogens.

Following the vital information, similar compounds from the PubChem databases were utilized which showed maximum structural similarity with the auranofin molecule. Total 33 such molecules were selected including different conformations that were used for the docking purpose against the developed structure. Structures having 3D information were only considered. SwissDock server [49] was utilized for this purpose.

#### Protein dynamics simulation for fluctuations

Fluctuation of protein provides insight in relation to the stability and flexibility of a protein. Simulated molecular motion analysis provides ample information in this regard. Different types of molecular motion analysis were performed for the target protein to understand the flexibility and plausible direction of the motion in the target protein. CABSflex [50] was used for this purpose.

## **Results and Discussion**

The goal of this computational analysis was to retrieve indepth molecular information with relation to the target protein considered for this study. Analyzing the protein in a comparative manner through integrated sequential, structural and functional analyses yielded novel insights for *Babesia*  *microti* thioredoxin reductase protein. The following result section describes such information in detail.

## Sequence retrieval and physicochemical characterization

Initial investigation of the protein sequence suggested some important information that is summarized in the Table 1. The length of the protein was found to be 553 amino acids having overall molecular weight of 60329.49. The higher molecular weight suggests that out of various types of thioredoxin reductase, this protein might have similarity with glutathione reductase, trypanothine reductase and such other enzymes [51].

The enzyme was predicted to be non-crystallizable with 0.463 confidence by CRYSTALP and least crystallizable both in EP crystallization (Class: 3) and RF crystallization classes (Class: 11). SPpred (Soluble Protein prediction) results showed a score indicating that thioredoxin reductase of *Babesia microti* is a soluble protein.

Theoretical pI of the protein was estimated to be 7.56, suggesting the point where the net charge of the protein could be nil. The theoretical pI value helps in different experimental set up pertaining to the protein for isolation and purification. The extinction coefficient value is another important parameter in relation to experimental calculations providing the absorption value of the protein. The obtained instability index value (Table 1) 31.12 represents the theoretical stability of the protein even after being a quite large molecule. Similarly, the aliphatic index value of 93.58 (Table 1) suggests the major relative volume occupancy of the aliphatic side chains for this Babesia thioredoxin reductase. Grand Average of Hydropathicity (GRAVY) value of -0.016 demonstrates the hydrophilic nature of the protein.

Table	1.	Physicochemical	properties	of	thioredoxin	reductase	of	В.
micro	ti d	etermined using p	rotparam.					

S. no.	Property	Value
1.	Number of amino acids	553
2.	Molecular weight	60329.49
3.	Theoretical pl	7.56
4.	Total number of negatively charged residues	56
5.	Total number of positively charged residues	57
6.	Formula	C <sub>2687</sub> H <sub>4276</sub> N <sub>722</sub> O <sub>798</sub> S <sub>2</sub> 7
7.	Total number of atoms	8510
8.	Ext. coefficient	43945
9.	Ext. coefficient*	43320
10.	Instability index	31.12
11.	Aliphatic index	93.58
12.	Grand average of hydropathicity (GRAVY)	-0.016

13.	Estimated half-life (mammalian reticulocytes, in vitro)	>30
14.	Estimated half-life (yeast, in vivo)	>20
15.	Estimated half-life (Escherichia coli, in vivo)	>10

## Amino acid composition

Prediction results showed that leucine (Table 2) was the most abundant amino acid followed by glycine and valine. Secondary structure of a protein is representation of repetitive geometrical conformations formed as a result of intermolecular and intramolecular hydrogen bonding. Most of the servers predicted that random coils were predominant structures in the protein followed by alpha helices and extended strands.

Eleven cysteines at position 17, 105, 110, 200, 251, 278, 306, 419, 471, 547 and 552 were predicted using CYSRec. Out of these, 7 are not S-S bounded, 1 is probably S-S bounded. Only 2 cysteines (536) showed high score and are probably S-S bounded. Most probable pattern is 105-110.

**Table 2.** Amino acid composition of thioredoxin reductase of B.microti predicted using protscale.

S. no.	Amino acid	Number	Percentage
1	Ala	37	6.70%
2	Arg	21	3.80%
3	Asn	29	5.20%
4	Asp	34	6.10%
5	Cys	11	2.00%
6	Gln	12	2.20%
7	Glu	22	4.00%
8	Gly	48	8.70%
9	His	13	2.40%
10	lle	38	6.90%
11	Leu	51	9.20%
12	Lys	36	6.50%
13	Met	16	2.90%
14	Phe	22	4.00%
15	Pro	23	4.20%
16	Ser	39	7.10%
17	Thr	34	6.10%
18	Тгр	3	0.50%
19	Tyr	18	3.30%
20	Val	46	8.30%

Proportion of different classes of amino acids in thioredoxin reductase of *B. microti* is represented in Table 3 suggesting dominance of non-polar amino acids followed by small and

polar amino acids. Periodicity of polar and non-polar amino acids along with the existence of disordered regions determines the secondary structure of a protein molecule, especially its tendency of developing  $\alpha$ -helices and  $\beta$ -sheets [52].

*Table 3.* Mole percentage of different classes of amino acids present in thioredoxin reductase of B. microti using emboss.

S. no.	Property	Residues	Number	Mole (%)
1	Tiny	(A+C+G+S+T)	169	30.561
2	Small	(A+B+C+D+G+N+P+S+T+V)	301	54.43
3	Aliphatic	(A+I+L+V)	172	31.103
4	Aromatic	(F+H+W+Y)	56	10.127
5	Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	313	56.6
6	Polar	(D+E+H+K+N+Q+R+S+T+Z)	240	43.4
7	Charged	(B+D+E+H+K+R+Z)	126	22.785
8	Basic	(H+K+R)	70	12.658
9	Acidic	(B+D+E+Z)	56	10.127

Improbability of expression in inclusion bodies for the considered protein was found to be 0.792 suggesting the probable tendency of the protein with reference to its presence in the inclusion bodies and thus hinting the possible way of purifying the same without losing the enzymatic activity. The other important physicochemical properties along with their respective value ranges are provided in Table 4.

**Table 4.** Minimum and maximum values of physicochemical properties of thioredoxin reductase of B. microti predicted using protscale.

S. no.	Property	Minimum value	Maximum value
1.	Bulkiness	0.197	0.828
2.	Polarity	0.001	0.556
3.	Refractivity	0.108	0.515
4.	Recognition factor	0.065	0.632
5.	Hydrophobicity (Kyte and Doolittle)	0.246	0.819
6.	Transmembrane tendency	0.329	0.839
7.	Buried Residues	0.218	0.846
8.	Accessible residues	0.323	0.701
9.	Ratio hetero end/side	0.030	0.442
10.	Average area buried	0.115	0.542
11.	Average flexibility	0.329	0.852
12.	Relative mutability	0.218	0.767

#### Sequential motif analysis

Sequential and structural motifs are majorly conserved in nature and serve as determinant for important structural component or vital supporting functional elements. The protein sequence was subjected to motif analysis and 3 motifs were found to be conserved in the protein sequences which are represented in Table 5.

*Table 5.* Motifs discovered in thioredoxin reductase of B. microti using MEME.

Logo	p-value	Start	E-value	Site count	Width
VSSCVCVNVKC	1.27E-15	98	1.90E+01	2	13
	1.28E-14	540			
RIKISNIPITI	2.30E-23	451	1.60E+01	2	21
MMAIOUIIİĞÊÎĞÊĞÊ	1.18E-22	399			
	1.82E-08	211	2.50E+01	3	6
	1.50E-07	375			
	2.17E-07	14			

#### Secondary structure analyses from target sequence

A multi-server based analysis was performed to gain insight into the secondary structure of the *B. microti* thioredoxin reductase protein. Table 6 depicts the obtained result from various tools along with the percentage of secondary structural elements predicted.

**Table 6.** Secondary structure of thioredoxin reductase of Babesiamicroti predicted using NPS server.

Secondary structure	DSC	HNNC	MLRC	PHD	Predator	Sec. cons.
Alpha helix	25.50%	30.56%	29.66%	22.24%	20.43%	25.50%
310 helix	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Pi helix	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Beta bridge	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Extended strand	16.64%	23.69%	22.06%	24.41%	20.25%	20.25%
Beta turn	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bend region	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Random coil	57.87%	45.75%	48.28%	53.35%	59.31%	50.27%

The obtained results hinted towards dominance of coiled structural component followed by helices and extended strands. No further hint was obtained for the other delicate structural components such as 310 helix, Pi helix, beta turns, bend regions or hinge regions. Thus, it is necessary to have structure based secondary structure analysis for minute information missing from the sequence based analysis. As thioredoxin reductase are of different types [51] with difference in

structure, mechanism of action, variation in coding genes, therefore, each thioredoxin structure should be analysed with individual attention.

 Table 7. Disordered regions by definition using different servers.

S. no	Server	Definition by type	Region
1	Pondr	Disordered regions	172-177, 187-188, 213-218, 336-357, 367-367, 377-382
2	DisEMBL	Disordered by loops/coils definition	4-71, 62-70, 86-112, 133-144, 180-195, 208-240, 289-322, 327-380, 394-423, 456-481, 488-495, 509-553
		Disordered by hot-loops definition	93-103, 134-145, 167-196, 208-217, 460-469, 536-553
		Disordered by Remark-465 definition	538-553
3	GLOBPLOT	Disordered by Russell/Linding definition	61-71, 92-106, 210-219, 230-237, 463-468, 536-553
		Potential globular domains (GlobDoms) by Russell/Linding definition	1-91, 107-551

#### Transmembrane region prediction

As the transmembrane helix regions scoring above 500 only are considered significant, only 2 regions (2-21, 239-266, inside to outside helices) and 1 (96-114, Outside to inside helices) predicted were considered (Figure 1).



Figure 1. Transmembrane region prediction using TMPRED.

#### Disordered regions in the target protein

Identification of disordered regions in a protein molecule is of vital importance to understand the structural, functional and evolutionary aspect of a protein. It is imperative to understand the disordered regions of this protein considering its structural and functional diversity. The predicted disordered regions are provided in Table 7 and Figure 2.

#### Sequence-structure conservation analysis

Consurf server [23,24] was used for finding the evolutionary significance of amino acids in thioredoxin reductase protein sequence of *B. microti* with the default options of BLAST E-value threshold: 0.001, maximum number of homologs: 50, iteration=1. Multiple sequence alignment was built using MAFT and UNIREF90 was used for collection of homologues

using HMMER homology search algorithm. Out of 1366 HMMER hits, 1242 were found to be unique and the calculation was performed on the 50 sequences closest to the query. The conservation scores versus residue number and the unrooted phylogenetic tree constructed using the tree building facility of CLUSTAL-W employing the multiple sequence alignment obtained from MUSCLE are shown in Figure 3.



*Figure 2.* Protein disorder of thioredoxin reductase of Babesia microti predicted using (a) PONDR, (b) GLOBPLOT (c) IUPRED (d) DisEMBL and (e) PRDOS.

#### Structure generation through homology modeling

As mentioned in the methodology section, homology modeling for the target sequence was attempted using Modeller 9.19 version. Following the protocol discussed, profile was built and relevant structures were searched through Modbase. Rigorous experiment was conducted to develop a model from a single template where PDB ID 2x99 chain "A" (best template found through SwissModel server) and 1mo9 chain "A" (best template found in Modbase template search) were used as templates. But due to template-target length difference, single template based approach was not adequate to provide expected result. Therefore, a multi-template based approach was adopted. Figure 4A shows the potential important templates obtained during template search along with their respective sequence identity, structural resolution and sequence coverage. Altogether, twenty one (21) templates were obtained and the best identity and templates with good structural resolution containing (1mo9A, 1xhcA, 1nhpA, 1xdiA, 1lvlA, 1h6vA, 1q1rA and 2x99A) were considered for multi-template based modeling (Figure 4B). Generated target-template alignment is shown in Figure 5.



Figure 3. Consurf results showing conserved amino acid with scores.





**Figure 4.** Template selection for three dimensional structural development through modeller. (A) The obtained initial hits as templates. (B) Comparative identity of the selected templates for multi-template based modelling.

_aln.pos 1mo9A	10 20 30 40 50 60 KVWNARNDHLTINOWATRIDEILEADDCCEVIYNVDEN
1xhcA	
InhpA	
11vlA	
1h6vA	
lq1rA 2x99A	CTSOWLRKTVDSAAVILFSKTTCPYCK
target	
_consrvd	
_aln.p	70 80 90 100 110 120 130
1mo9A	
InhpA	
IxdiA	
11VIA 1h6vA	
lq1rA	
2x99A	KVKDVLAEAKIKHATIELDQLSNGSAIQKCLASFSKIETVPQMFVRGKFIGDSQTVLKYYSNDELAGI MTLDAGLLITAAVDDLCDOTYHAAHHSDKLSNLLDKNFLFTADKTNFVSDK
_consrvd	
	140 150 160 170 180 190 200
1mo9A	DPREYDAIFIGGGAAGRFGSAYLRAMGGRQLIVDRWPFLGGSCPHNACVPHHLF
1xhcA	SKVVIVGNGPGGFELAKQLSQT-YEVIVIDKEPVPYYSKPMLS-HY
1xdiA	VTRIVILGGGPAGYEAALVAATSHPETTQVTVIDCDGIGGAAVLDDCVPSKTF
11vlA	IQTTLLIIGGGPGGYVAAIRAGQLGIPTVLVEGQALGGTCLNIGCIPSKAL
lqlrA	NDNVVIVGTGLAGVEVAFGLRASGNE-GNIRLVGDATVIPHHLPPLSKAY
2x99A	VNESKYDYDLIVIGGGSGGLAAGKEAAKYGAKTAVLDYVEPTPIGTTWGLGGTCVNVGCIPKKLM
consrvd	MSTDAVKYDFAVIGGGPGGHAAAKEAARLGAKVIIFDHVKPSHRGISWGFGGVCVNVGCVPKKLM
aln.pos	210 220 230 240 250 260 270
1mo9A	SDCAAELMLARTFSGQYWFPDMTEKVVGIKEVVDLFRAGRNGPHGIMNFQSKEQLNLEYILM-CP
1xhcA 1nhnA	LAFPYSLDWYRKRGIEIRLAEE LEGKVKDVNSVDVMTGEKMESDGVAUTSNTF
1xdiA	IASTGLRTELRRAPHLGFHKISLPQIHARVKTLAAAQSADITAQLLSMGVQVIA-GR
11v1A	IHVAEQFHQASRFTEPSPLGISVASPRLDIGQSVAWKDGIVDRLTTGVAALLKKHGVKVVH-GW
lqlrA	LAGKA-TAESLYLRTPDAYAAQNIQLLGGTQ
2x99A	HQAGLLSHALEDAEHFGWSLDRSKISHNWSTWVEGVQSHIGSLNWGYKVALRDNQVTYLN-AK
_consrvd	hindulgvakguikkisigirdvidiinniiikulvoniikulapbikooliv-kovkivakkv
-1-	780 290 900 910 000
1mo9A	200 290 300 310 320 330 340 AKVIDNHTVEAACKVFKAKNLILAVGAGPGTLDVPGVNAKGVFDHATLVEE
1xhcA	AKLIDRGRKVVITEK-GEVPYDTLVLATGARAREPQIKGKEYLLTLRTIFDADR
innpA 1xdiA	GELIDSTPGLARHRIKATAA-DGSTSEHEADVVLVATGASPRILPSAODDGERILTWPOLYD-
11vlA	AKVLDGKQVEVDGQRIQCEHLLLATGSSSVELPMLPLG-GPVISSTEALA-
1g1rA	GKFIGPHKIMATNN-KGKEKVYSAERFLIATGERPRYLGIPGDK-EYCISSD-DLFS- VTAINRDROOVILSDGRALDYDRLVLATGGRDRDLDVASGAVGKANNFDVLDTLEDAFC
2x99A	GRLISPHEVQITDK-NQKVSTITGNKIILATGERPKYPEIPGAV-EYGITSDDLFS-
consrvd	SLKSANEITYTDT-DGREIVVCAKYILLAIGDRPHTPPEIDKSLVITSDDLFY
_aln.pos	350 360 370 380 390 400
1xhcA	IKESIENSGEAIIIGGGFIGLELAGNLAEAGYHVKLIHRGAMFLGLDEELSNMIKDMLEETGVK
1xhcA 1nhpA 1xdiA	IKESIENSGEAIIIGGGFIGLELAGNLAEAGYHVKLIHRGAMFLGLDEELSNMIKDMLEETGVK LKOKTVDPEVNNVVIGSGYIGIEAAEAFAKAGKKVTVIDILDRPLGVYLDKEFTDVLTEEMEANNIT LAND-DD-UTUNGGGYIGIEYDAEPGNDVTFIGUDUTUN SONDUL D-VEDANAU VEPGAPBOUD
1xhcA 1nhpA 1xdiA 11vlA	IRESTEN-SCHAITIGGGFIGLELAGNIARACYUVKLINRGANFL-GLDEELSNNIKUMLEETOV LKQKTVDBFUNVVIGSGTIGLELAAFAVAGKKVIVTUILDEDLGVULUEFTDVLTEDGENNIT LDALDD-HLIVVGSGVTGAEFVDAYTELGVVVVVASQDRVLD-YEDADALVLEESFAEROVF PKALDQ-HLIVVGSGVTGLEJGJAYRFLGQVSVVARETILD-TYDSEIXHGGIA
lxhcA lnhpA lxdiA llvlA lh6vA	IRESTEN-SGRATIIGGGTIGLELANIARACYUVULINGANFI-GLDELSINTIKULLETUV LUKYUVDEVINNVVIGGYIGLELANIARACYUVULINGANFI-GLDELSINTIKULLETUV LUKYUVDEVINNVVIGGYIGLELANIT LUKLDD-HLIVVGGGVIGLENARTELAVVIVVISODEVID-YERANALIVLESFAREVY PRULDC-HLIVVGGGVIGLEJGAYENAATIGAVIVSEDILID-TIDSLITAVVASISKAGIA LUKUS
1xhcA 1nhpA 1xdiA 11v1A 1h6vA 1q1rA 2x99A	IRESTRNSCRATTIGGEFTGLEIAANIARACYUVLIHREAMTLGLDRELSIMIKIMLEETGVK LEGGTVDBEVINNVUTGSCYTGLEIAARIAKAKAVKIVIDILDEDKOVJUKRETDVIDTEREMINIT LOALD
lixhcA lixhcA linhpA lixdiA livlA lifevA lq1rA 2x99A target construd	IRESTRNSCRAIT LOGGFTGLEIACHLARACYWLI HRCAMFLGLDERLSNMIKHMLEITGVK LEGKTVDDEVNNVVIGSGYTGTEAEIVDAYTELGUVUTVVIDILDRUGVYLIKETTDVLITEREANNI LEGKTVDDEVNNVVIGSGYTGTEAEIVDAYTELGUVUTVVASQUMVLD-YTDADAINULEESTAERGVR DRALDQ-HLUVVOGGYTGLEICIATRULGAGVSVVEAEREILD-TYDEEITAPVAESIKULGIA LEYFCG-KLUVVASGYLICECGTIATUKAINUVTLLDTARVLERVTAPVSATYEHLERKACVD LEYFCG-KLUVVASGYLICECGTIATUKAINUVTLLDTARVLERVTAPVSATYEHLERKACVD LEYFFG-KLUVASSYLAECGFIAEGCGYTGWSI-LLG-GFDQRACKGYMENHGUT LINDDGKTLUVASSYLAECGFIAEGCGFIAEGCOVTWVNSSI-LLG-GFDQRACKGYMENHGUT LINDDGKTLUVASSYLAECGFIAEGCGFIAEGCGVTWVISSI-LEG-GFDQRACKGYMENHGUT
lixhcA lixhcA lixdiA lixdiA livlA lh6vA lq1rA 2x99A target _consrvd	IRESTENSCRATT GOOGFTGLETACHLARACTWYKLINRCMMTGLORELSSMITKOMLETTCWK LEGKTVD5FVMWVUGSGYGTGLETALSATARAKAKGWVVTDILDADLGVVJDKET FDVUTTENEMNIT LDALDD-HLIVUSGYTGLETALSATARAKAKGWVVTDILDADLAUVLESSTARBGVR PKLIDQ-HLIVUSGYTGLEIGIATRELGUGVVWARSI-LLD-TTDSELTAPPASSINICIGIA LEPYCDG-KTLWUSGSYGLEGIATRELGUGVVWARSI-LLD-GTDGUDAMKIGENEHHEIK INRDLGHTLWUSGSYGLEVALATIAKIGUGVVWARSI-LLD-GFDGUGAEKVGDYMENHGVK LEWTDGHTLWUSGSYGLECASTILASIGGUVVWARSI-LLD-GFDGUGAEKVGDYMENHGVK LEWTDGHTLWUSGSYGLECASTILASIGGUVVWAVSI-LLD-GFDGUGAEKVGDYMENHGVK LEWTDGHTLWUSGSYGLECASTILASIGGUVVWAVSI-LULA-GFDGUGAEKVGDYMENHGVK
laboa laboa laboa laboa laboa laboa laboa consrvd aln.p 4 lmo9A	IRESTENSCRAITICCCFTGLEIACHLARACTIVKLIMRCAMTGLORELSMHIKHMLEITCVK LUKKYTVDFUNWVIGSCYTGLEIALACHARACHARACHVITULIDABLAVVIINETTOVITUTEHEANNII LDALDD-HLIVUSCOTGLEIGLEIXTRILGVVIVVIDASCHWILD-TEDLALIVUSESFARBOUR PKLIDC-HLIVUSCOTGLEIGLEIXTRILGVUSVUSES-LLD-TTBEILANVIDASIXHLEIL LUYCD
labola lixhcA linhpA lixdiA livlA livlA lifvA lifvA lifvA lifvA 	IRESTRMSCRATTIGGEFTGLEIAGNIARACYWKLINRCAMTGLDRELSSMITKMLEITUWK LEKKYTNDEYNWVIGSEYTGLEAALSTAARACKAUKYVIDILOBENGVYIDKETBUNDITEEHEANNIT LDALDD-HLIVUSGUTGLEIGIATNEILGUVVIVIDSLOBEVID-YEDADAIVLEESTANRCAU PRALDG-HLIVUSGEVTGLEIGIATNEILGUVVIVISSI-LLA-GFDDDKANKIGEHEEHEIK IRBQLIANUEUVVIGGEYTGLEIGIATNEILGUVVIVISI-LLA-GFDDDKANKIGEHEEHEIK IRBQLIANUEUVVIGGEYTGLEIGIATNEILGUVVIVISI-LLA-GFDDDKANKIGEHEEHEIK IRBQLIANUEUVVIGGEYTGLEVANTIEKSUUVVIVISI-LLA-GFDDDKANKIGEHEEHEIK IRBQLIANUEUVVIGGEYTGLEVANTIEKSUUVVIVISI-LLA-GFDDDKANKIGEHEEHEIK IRBQLIANUEUVVIGGEYTGLEVANTIEKSUUVVIVISI-LLA-GFDDDKANKIGEHEEHEIKE IRBQLIANUEUVVIGEYTGUVINISI-VANTIEKSUUVINISI-LLA-GFDDDKANKIGEHEEHEIKE IRBQLIANUEUVVIGEYTGUVINISI-VANTIEKSUUVINISI-LLA-GFDDDKANKIGEHEEHEIKE IRBQLIANUEUVVIGEYTGUVINISI-VANTIEKSUUVINISI-VILA-GFDDDKANKIGEHEEHEIKE IRBQLIANUEUVVIGEYTGUVINISI-VANTIEKSUUVINISI-VILA-GFDDDKANKIGEHEEHEIKE IRBQLIANUEUVVIGEYTGUVINISI-VANTIEKSUUVINISI-VILA-GFDDDKANKIGEHEEHEIKE IRBQLIANUEUVVIGEYTGUVINISI-VANTIEKSUUVINISI-VILA-GFDDDKANKIGEHEEHEIKE IRBQLIANUEUVVIGEVIGUVINISI-VANTIEKSUUVINISI-VILA-GFDDDKANKIGEHEEHEIKE IRBQLIANUEUVVIGUVINISI-VANTIEKSUUVINISI-VILA-GFDDDKANKIGEHEEHEIKE IRBQLIANUEUVVIGUVINISI-VANTIEKSUUVINISI-VILA-GFDDDKANKIGEHEEHEIKE IRBQLIANUEUVVIGUVINISI-VANTIEKSUUVINISI-VILA-GFDDDKSEVIEHEIKEKSUU IRBQLIANUEUVVIGUVINISI-VANTIEKSUUVINISI-VILA-GFDDDKSEVIENEISI 
labok lakcA lnhpA lixdiA llvlA lqlrA 2x99A target consrvd aln.p lmo9A lxhcA lnhpA lxdiA	IRESTEN-SCRATT COCFT GLETACHLARACTIVKLI HRGANTL-GLORELSINI KIMLETCVK LIKKTVDFVNWVVGSCYGLETALARATAKAKKIVIVIDLI DABLOVINCH TOVLITEHEANNII LDALDD-HLIVVGSCYGLETALARATAKAKKIVIVDI LDABLOVINCH TOVLITEHEANNII LDALDD-HLIVVGSCYGLEGIATRICLAGVSVVARST-LD-YTSDLADVLESSTARERGV PKLDQ-HLIVVGSCYGLEGIATRICLAGVSVVARST-LLD-TTDSELTAPVASSKIVCID LBYCDG-KTLVVGSYVALCCAFTLAGVSVVVARST-LLD-GTDQUARKGUPYEHIERAK LBYDDGKTLVVGSYVALCCAFTLAGUSVVVANST-LLD-GTDQUARKGUPYEHIERAK LBYDD
laboa lacha lacha lacha lacha lacha 	IRESTENSCRAITICGCFTGLEIACHLARACTIVKLIHRCAMTGLORELSMHIKHMLETCVK LKQKTVDPSVMVVVTGSCYGLEIAACHARACKUVVVTVLDIADEVAVVTNETESTENENNII LDALDD-HLIVVUSGYUGLEIGAINFILGUVVVVVDSQDHVLD-YEDDALIVLESFARBOUR PKLDG-HLIVVUSGYUGLEIGAINFILGUVVVVVBSI-LLR-GFDQDHANKIGEHREHRDIK IBRQLIANDHLVVIGGYUGLEGAINFILGUVIAVVBSI-LLR-GFDQDHANKIGEHREHRDIK IBRQLIANDHLVVIGGYUGLEGAINFILGUNIAVVBSI-LLR-GFDQDHANKIGEHREHRDIK IBRQLIANDHLVVIGGYUGLEGAINFILGUNIAVVBSI-LLR-GFDQDHANKIGEHREHRDIK IBRQLIANDHLVVIGGYUGLEGAINFILGUNIAVVBSI-LLR-GFDQDHANKIGEHREHRDIK IBRQLIANDHLVVIGGYUGLEGAINFILGUNIAVVBSI-LLR-GFDQDHANKIGEHREHRDIK IBRQLIANDHLVVIGGYUGLEGAINFILGUNIAVVBSI-LLR-GFDQDHANKIGEHREHRDIK IBRQLIANDHLVVIGGYUGLEGAINFILGUNIAVVBJI-LRGFTGH 
Imboa Ishca Inhpà Iskdi Iskdi Iskdi Ishca Imboa Ishca Inhpà Iskdi Ishca Ishca Ishca Ishca	IRESTEN—SCEATTICCCFTGLEIACHLARACTIVIKLIHRGANTL—CLORELISMITKIMLETTCVK LEKKTVDFVMWVIGSCYTGLETALASTAARACHVIVTUILDBALVAVIIKETTOVITTEHEMMIT LDALP——D-HLIVVUSCYTGLEIGLETARTIAARACHVIVTUILDBALVAVIIKETTOVITTEHEMMIT LDALP——D-HLIVVUSCYTGLEIGLETARTIAACHVIVTUUDBALVAVIIKETTOVIVASISKITGLEI LPYCP——G-HLIVVUSCYTGLEIGLETARTIAACHOUVTUVASI-LLA-GFDQUARKVGPHEHBERKU LBYCP——G-HTIVVISASTVALECAFTLAGICLOVTWIVSI-LLA-GFDQUARKVGPHEHBERKU LBYCP——GKTLIVVGASTVALECAFTLAGICLOVTWIVSI-LLA-GFDQUARKVGPHEHBERKU LBYCP——GKTLIVVGASTVALECAFTLAGICLOVTWIVSI-LILA-GFDQUARKVGPHEHBERKU LBYCP——GKTLIVVGASTVALECAFTLAGICLOVTWIVSI-LILA-GFDQUARKVGPHEHBERKU LBYCP——GKTLIVUSASTVALECAFTLAGICLOVTWIVSI-GFDFQCSEKVANLAENTGTH — * * * * * * * * * * * * * * * * * *
Imboa Ishca Inhpà Iskdi Ilvlà Ihfva Iqlrà 2x99a aln.p Imo9a Iskca Inhpa Iskca Inhpa Iskdi Inhva Iglrà 2x99a	IRESTENSCRAITICOCFTICLELACHLARACTIVICLIRRCANTCLORELISMITKIMLETCVK LEXXTVDFVNWVVVVVVVXVVVVVVVVVVVVVVVVVVVVVVVVVVVV
lmbyA lxhcA lnhpA lxdiA llvlA lh6vA lq1rA 2x99A target 	IRESTRISCRAITICOCFTGLEIACHLARACTIVKLIRRCMFLGLORELSMITKMELETCVK LKGKTVD5VMWVUTGSCYGLEIAACHLARACHLARUVVUTULIDABUGVUTUKETSUTUTEHEANNIT LDALDD-HLIVUSGUTGLEIGIATREIAGUVUTUUTABUGVITUKETSUTUTEHEANNIT LDALDD-HLIVUSGUTGLEIGIATREIAGUSVUTUUTABUT HIPKLDG-HLIVUSGUTGLEIGIATREIAGUSVUTUUTABUT LBYCDG-HTIVUGASVALECASTLAGIGLUVUTWISSI-LLA-GFDQUBANKIGHMEHHGIK LMNDDGHTIVUGASVALECASTLAGIGLUVUTWISSI-LLA-GFDQUBANKIGHMEHHGIK LMNDDGHTIVUGASVALECASTLAGIGLUVUTWISSI-LLA-GFDQUBANKIGHMEHHGIK LMNDDGHTIVUGASVALECASTLAGIGLUVUTWISSI-LLA-GFDQUBANKIGHMEHHGIK LMNDDGHTIVUGASVALECASTLAGIGUVUTWISSI-LLA-GFDQUBARKIGHMEHHGIK LMNDDGHTUVUGASVALECASTLAGIGUVUTWISSI-LLA-GFDQUBARKIGHMEHHGIK LMNDD
labba labcA labcA labcA labcA latrA latrA latrA latrA latrA labcA	IRESTENSCRAITICGCFTGLEIACHLARACTIVKLINRCAMTGLORELSMITKMLETCVK LEXXTVDEVNWVVVVSCSYGTGLENAARAKAKAKVVVVDILDAPKVVVVIDLIACHLAVKVINETTBUEANNI LDALDD-HLIVVUSGYTGLEIGLATRIKLAKVVVDILDIADKUV-YEDDAALVVESSTREDGVA PKLDG-HLIVVUSGYTGLEIGLATRIKLAKVVVVIDULDAPKUV-YEDDAALVVESSTREDGVA INVODGKTLVVCASYVALECASTLAKIGLUVVVVVSI-LLA-GFDQDAALVVESSTRELEI LBYGDGKTLVVCASYVALECASTLAKIGLUVVVVSI-LLA-GFDQDAALVCESHEHEIKI INNDDGKTLVVCASYVALECASTLAKIGLUVVVVSI-LLA-GFDQDAALVCESHEHEIKI INNDDGKTLVVCASYVALECASTLAKIGUVVAVSSI-LLA-GFDQDAALVCESHEHEIKI INNDDGKTLVVCASYVALECASTLAKIGUVVAVSSI-LLA-GFDQDAALVCESHEHEIKI INNDDGKTLVVCASYVALECASTLAKIGUVVAVSSI-LLA-GFDQDAALVCESHEHEIKI INNDDGKTLVVCASYVALECASTLAKIGUVVAVSSI-LLA-GFDQDAALVCESHEHEIKI INNDDGKTLVVCASYVALECASTLAKIGUVVAVSSI-LLA-GFDQDAALVCESHEWI INNDDGKTLVVCASYVALECASTLAKIGUVVAVSSI-LLA-GFDQDAALVCESHEWI INNDD
lauya lakaca lakaca lakaca lakaca lakaca lainy construt lakaca la	IRESTENSCHAITIGGEFIGLELAGNIARACIWKLINRGAMTCLORELISMIKKMLETEVK LEKKTVDEYNWWVIGSGYGERLARAFARAKKIVVIDLIDABVGWVIDKETENVITENEMMINI LDALDD-HLIVVGSVTGEFIGLETARFARAKKIVVIDLIDABVGWVIDKETENVITENEMMINI LDALDD-HLIVVGSVTGEFIGLETARFARAKKIVVIDLIDABVGHVID-YEDDALVLEESFAREVK FKALDQ-HLIVVGSVTGELGIATKRIGAGVSVEARERLIDITISELTAVDFYHEIRERAVU LBVFDG-HTIVVGASYVALECAFIJAGIGLVVIWVSILDIAAVLEVTAVFBVYBAVKGUPKEHHEIKI LBVFDG-HTIVVGASYVALECAFIJAGIGLVVIWVSILDIAAVLEVTAVFBVYBAVKGUPKEHHEIKI LBVFDG-HTIVVGASYVALECAFIJAGIGLVVIWVSI-LLA-GFDQGASKVGDYKEHHEIKI LBVFDG-HTIVVGASYVALECAFIJAGIGLVVIWVSI-LLA-GFDQGASKVGDYKEHHEIKI LBVFDG-HTIVVGASYVALECAFIJAGIGLVVIWVSI-LLA-GFDQGASKVGDYKEHHEIKI LBVFDG-HTIVVGASYVALECAFIJAGIGLVVIWVSI-LUA-GFDQGASKVGDYKEHHEIKI LINDDG-HTIVVGASYVALECAFIJAGIGLVVIWVSI-VIAUKSTULG- FIJASILLAB
laboa laboa	IRESTENSCRAITICOCFTGLEIACHARACTIVKLIMRCMPTCLORISMITKMMLETCVK LKKYTVDFUNWVVIGSOTGLEIAACHARAKAKIVVIDLIDABU/WVINKTENELMINIT LDALDD-HLIVUSGOTGLEIAACHARAKAKIVVIDLIDABU/WVINKTENELMINIT LDALDD-HLIVUSGOTGLEIGIATRILAGUGVIVVIDUIDABU/WINKTENUSHESTARBOW BWLDQ-HLIVUSGOTGLEIGIATRILAGUGVIVVIDUIDABU/WINKTJUPUSESTARDI LPYCDC-HTIVUGASYVALECASTIAGICLVVIWVISJ-LLA-GFDQUBANKICHHEHIST INRQLIAONDIVVIGCUSTERVANTIFM-GUNUTLDIAAUVENTJDYSAITUPUSESTARDI LPYCDG-HTIVUGASYVALECASTIAGICLVVIWVISJ-LLA-GFDQUBANKICHHEHIST INRQLIAONDIVVIGUENTATATARABUGVILDIAAUVENTJDYSAITUPUSESTARDI LPYCDG-HTIVIGASYVALECASTIAGUGVIWVINJSJ-LLA-GFDQUBAKKICHHEHIST INNDGHTUVIGASYVALECASTIAGUGVIWVINJSJ-LLA-GFDQUBAKKICHHEHIST INNDGHTUVIGASYVALECASTIAGUGVIWVINJSJ-LLA-GFDQUBAKKICHHEHIST IISSSMUTEISEN ISSSMUTEISEN ISSSMUTEISEN INTOTENCO
labora lahca	IRESTRISCRAITICGCFTGLEIAANIARACYIVKLINRCAMTGLORELSIMIKKMLETCVK LIKKYTVDSTAWIVGCSTGLEIAANIARACKIVVKLINGLAUGVYINKETSUVKITEHEKNINTI LOALDD-HLIVUSGUTGLEIAANIARACKIVVTUDILOBUSVAVINKETSUVKITEHEKNINTI LOALDO-HLIVUSGUTGLEIGAANIARACKIVVUTUDIASUDHUD-YEDDAALVLEESTREGVR PKLIDO-HLIVUSGUTGLEIGAANIARACKIVVUTULOBANIAVID-YEDDAALVLEESTREGVR INNDDGHIVUSGUTGLEIGAANIARACKIVUTUDIASUDHUD-TIDSELTAVIAVISSIKULUSI LIVIGGHIVUSGUTGLEIGAANIARACKIVULUSIASUVUTUVASI LIVIGGHIVUSGUTGLEIGAANIARACKIVULUSIASUVULUSI LIVIGGHIVUSGUTGLEIGAANIARACKIVULUSIASUVULUSI LIVIG
labia ladia	IRESTENSCRAITICOCFTGLEIAONIARACTIVKLIMEGANTCLORELSINIKKIMLETEVK LIKKYTUPSVANVVUGSOTGLEIAAONIARAKKIVIVUTLIDAPKVINKETTUVITENEMMINI LIAALSD-HLIVVOSOTGLEIGAINRILAGUVIVVINGUSUPATIAEIN-ITIBELIANNI LIAALSO-HLIVVOSOTGLEIGAINRILAGUVIVVINGUVIUDIASUKUUSI LIYKALSO-HLIVVOSOTGLEIGAINRILAGUVIVVINGUVIUDIASUKUUSI LIYKALSO-HLIVVOSOTGLEIGAINRILAGUVIULDIAARUKUTAPVINGUTASIKKUUSI LIYKAO-HLIVVOSOTGLEIGAINRILAGUVIULDIAARUKUTAPVINGUTASIKKUUSI LIYKAO-HLIVVOSOTGLEIGAINRILAGUVIULDIAARUKUTAPVINGUVIULISESTARENT LIYKA
labca labca	IRESTENSCHAITIGGEFIGLELAGNIARAGYIVKLIMEGAMTGLOBELISMIKKMLETEVK LKQKTVDSFUNWVVIGGSTGLENAARAKAKKIVVIDLIDABVGVVIDKITSUTUTEHEMINI LDALDD-HLIVVGSGVIGLENAARAKAKKIVVIDLIDABVGVVIDKITSUTUTEHEMINI LDALDD-HLIVVGSGVIGLEGATRELAGVIGVVIVASJOHVL-YEDDALVLESFARBOVR BYKLDQ-HLIVVGSGVIGLEGATRELAGVIGVVIVASJ-LLA-GTDQUBANKIGHMEHHGI LBYCDGKTLVIGASVALECAGTLAGIGLOVIVAVSJ-LLA-GTDQUBANKIGHMEHHGI LBYDDGKTLVIGASVALECAGTLAGIGLOVIVAVSJ-LLA-GTDQUBANKIGHMEHHGI LBYDDGKTLVIGASVALECAGTLAGIGLOVIVAVSJ-LLA-GTDQUBANKIGHMEHHGI LBYDDGKTLVIGASVALECAGTLAGIGLOVIVAVSJ-LLA-GTDQUBANKIGHMEHHGI LBYDDGKTLVIGASVALECAGTLAGIGLOVIVAVSJ-LLA-GTDQUBANKIGHMEHHGI LBYDDGKTLVIGASVALECAGTLAGIGLOVIVAVSJ-LGTDQUSSKALIAKKILLEDIL FYNSSTLLAN
lahch la	IRESTENSCRAITICOCFTGLEIACHARACTIVIKLINRCMETGLORELISMIKKMLETCVK LEKKTVDEYNWWVVGSCHTGLEAACHARACHARACHVIVUTUDIGDEVGVVINKETENUTEHEMINIT LOALDD-HLIVUSGUTGLEIAARHARACHVIVUTUDIGDEVGVVINKETENUTEHEMINIT LOALDD-HLIVUSGUTGLEIGATREIGANGVIVUTUDIGDEVIN-TTSDEILAPHARASEKKUCEIA HUVVGGYTGLEIGATREIGANGVIVUTUDIGANGVINTUTUDIAANUTABUTABUTABUTABUTABUTABUTABUTABUTABUTAB
lahca lahca	IRESTENSCRAITIGGEFIGELIAANIAARGYIWILINGAMITGLOELISMIKKMILETTÜVK LIKKYTVDEVMWVVIGSUTGAEANAANARGVIVVIDILOBAUGVIVAETSUTAENEANIIT LOALDD-HLIVUSGUTGAEIVAANTELGVVVIVUSGURVLADILAU-TSUDAALVLEESTARGVA PKLDCHILVUSGUTGAEIVAANTELGVVVIVUSGURVLADILAU-TSUDAALVLEESTARGVA INNDCHILVUSGUTGAEIVAANTELGVVIVUSGI-LLA-GFDQUGAEKVGENKEHSIK INNDGKILVUSGUTGAEVAANTELGVVIVUSGI-LLA-GFDQUGAEKVGENKEHSIK INNDGKILVUSGUTGAEVAANTELGVIVUSUSGI-LLA-GFDQUGAEKVGENKEHSIK INNDGKILVUSGUTGAEVAANTELGVIVUSUSGI-LLA-GFDQUGAEKVGENKEHSIK INNDGKILVUSGUTGAUSTAANTAKANGVILLANAAUVAUVAUVASUVILA-GFDQUGAEKVGENKEHSIK INNDGKILVUSGUTGAUSTAANTAKANGVILLANAAUVAUVAUVAUVAUVAUVAUVAUVAUVAUVAUVAUVAUV
inhpä inhpä inhpä inhpä inhpä inhpä inhpä inhpä inhpä inhpä inhpä inhpä inhpä inhpä inhpä inhoä inhpä inhoä in	IRESTEN-SCRAITICOCFTGLETACHARACTWYLINKLINGCMTL-GLORELISMIKKMLETTUW LUKKYTDSYNWYUGSYTGLETAACHARACKYNYUTULDSAUWYINKETSYNTEHEMINIT LOALSD-HLIVWSGYTGLETAATATAKAKKYWYUTULDSAUWYINKETSYNTEHEMINIT LOALSD-HLIVWSGYTGLEIGATARILAGUGWYUTWASICHU-TTBOELIAPKYNESSKUTCH LWYFG-KTLWUGSYTGLEGATTRIAGUGWYUTULDTAAVUKTJAPWSTYAPKILISTAAVU LWYFG-KTLWUGSYTGLEGATTRIAGUGWYUTULDTAAVUKTJAPWSTYAPUSHUKU LWYFG-KTLWUGSYTGLEGATTRIAGUGWYUTWASICLUG-STUPUS LWYFG-KTLWUGSYTGLEGATTRIAGUGWYUTWASICLUG-STUPUS LWYFG-KTLWUGSYTGLEGATTRIAGUGWYUTWASICLUG-STUPUS LWYFG-KTLWUGSYTGLEGATTLAGUGWYUTWASICLUG-STUPUS LWYFG-KTLWUGSYTGLEGATTLAGUGWYUTWASICLUG-STUPUS LWYFG-KTLWUGSYTGLEGATTLAGUGWYUTWASICLUG-STUPUS LWYFG-KTLWUGSYTGLEGATTLAGUGWYUTWASICLUG-STUPUS LWYFG-KTLWUGSYTGLEGATTLAGUGWYUTWASICLUG-STUPUS LYYFG-KTLWUGSYTGLEGATTLAGUGWYUTWASICLUG-STUPUS LYYFG-KTLWUGSYTGLEGATTLAGUGWYUTWASICLUG-STUPUS LYYFATANIANAN ISSNTTI KEEDANGRUGWYUTSTOFTGEUKEQPESAEL-AKILGLUM FINSSELLAN
labia labia labia labia labia libi	INFSTRM-SCRAITICOCFTGLEIAANIAARACTWYKLINRCMATGLORELISMIKKMMLETTÜVK         LOKKTVDSTVMWVVIGSOTGLEIAAANIAARACTWYVTUNIDSDLAUGVIJKETSUNITEHEMINITI         LOKKTVDSTVMWVVIGSOTGLEIAANIAARACTWYVTUNIDSDLAUGVIJKETSUNITEHEMINITI         LOKADD-HLIVUSGOTGLEIGIATRILGAUGVIVUNIDSIADLAUVLESSTAREOVR         JEKLJO-HLIVUSGOTGLEIGIATRILGAUGVIVUNIDSIADLAUVLESSTAREOVR         JEKLJO-HLIVUSGOTGLEIGIATRILGAUGVIVUNIDSIALUA-TTIDEILANDYTASSIXICIAI         LUPYCPC-HTIVIGASYALECASTIAGICLOVIWINSI-LLA-GFDQUGARKVGDYMENHEIKI         JAND
lahca la	IRESTENSCRAITICOCFTICALIANIARACYUVILIARACATUVALIMERANTIGLORLISSMIKEMALETTÜVK           LEKKYTVDESUNAVUSICASTICALARATAKARKAUVUVILIARACVIJALESTARLISSMIKEMALETTÜVK           LEKKYTVDESUNAVUSICASTICALARATAKARKAUVUVILIARACVIJALESTARLISSMIKEMALETTÜVK           LANDD-HLIVUSGOVICALEVIANTELGUVUVVASORVLD-YEDADALUVEESTARLIGAN           LEVICC-HLIVUSGOVICALEJATATICALGUVUVVASORVLD-YEDADALUVEESTARLIGAN           LEVICC-HLIVUSGOVICALEJATATICALGUVUVAVSSI-LLA-GFDQUARKUCEHMEHSIK           INDELO-HLIVUSGOVICALEGATILAGIALUVVAVSSI-LLA-GFDQUARKUCEHMEHSIK           LINDDGHTLVICASYVALECASTILASIGGUVUVAVSSI-LLA-GFDQUARKUCEHMEHSIK           LINDDGHTLVICASYVALECASTILASIGGUVUVAVSSI-LLA-GFDQUARKUCEHMEHSIK           ILISSENUTI IELONAC         -RVQAVVATUVAVSSI-LLA-GFDQUARKUCEHMEHSIK           ILISSENUTI IELONAC         -RVQAVVATUVAVSSI-LLA-GFDQUARKUCEHMEHSIK           ILISSENUTI IELONAC         -RVQAVVATUVAVSSI-LLA-GFDQUARKUCEHMEHSIK           ILISSENUTI IELONAC         -RVQAVVATUVAVSSI-LLA-GFDQUARKUCEHMEHSIK           ILISSENUTI IELONAC         -RVQAVVATUVAVSSI-LLA-GFDQUARKUCEHMENT           ILISSENUTI IELONAC         -RVQAVVATUVAVSSI-LLA-GFDQUARKUCEHMENT           ILISSENUTI IELONAC         -RVQAVVATUVAVSSI-LLA-GFDQUARKUCEHMENT           ILISSENUTI IELONACH         -REGUINIVAUVAUVAUVAUVAUVAUVAUVAUVAUVAUVAUULARABISIELI           ILISSENUTI IELONACH         -REGUINIVAUVAUVAUVAUVAUVAUVAUVAUVAUVAUVAUULKEHSKUSUL           ILISSENUTI IELONACH
labiba labiba ladia likyla lik	TRESTENSCRATT COCOFT GLETAGNIARAGYNYLLINRGAMT,-GLORELISMIKRMLETTCVF LOKKYTDSYNWYUGSOFT GLETAAGNIARAGYNYLLINGACHYNYLDIAGD LWYDYNYL RETRENN TENER HENDYLDO-HLIVVGSOT GLETAARAFAAGNG WYTVTU DAG DLEYDYN DYN BLADWYL BY BARBOW BYRLDO-HLIVVGSOT GLETAARAFAAGNG WYTVARS I-LLA-GYDQIMANK GEHMEHEI F TREGTAT-TNEHUVU GGYTGLEGAATRIAANGWYLLDIARACHWYLD YN BUDYNS SINGL LYFDG-KTLVVGASYNALCAGTLAGIGUVYWNS I-LLA-GYDQGARKYGDYREHHEI F HUNDGHUVVGGYTGLEGAATRIAANGWYLLDIARACHWYLD YN BYNYLWYLS YN LLA-GYDQGARKYGDYREHHEW LYFDG-KTLVVGASYNALCAGTLAGIGUVYWNS I-LLA-GYDQGARKYGDYREHHEW LWNDGHUVVGASYNALCAGTLAGIGUVYWNS I-LLA-GYDQGARKYGDYREHHEW LWNDGHUVVGASYNALCAGTLAGIGUVYWNS I-LLA-GYDQGARKYGDYREHHEW TI ISSGNTT I KEDANG
labha ladia ladia linya	INFSTRM-SCRAITICOCFTGLEIAANIAARACTIVUKLIMRCMMTGLORLISMIKTMMLETTUVK LOKUTVDSTVMVVVIGSUTGLEAAANIAARACHVIVUTULIDAAVUTUSTEHEMINITI LOALSD-HLIVUSGUTGLEIGAANRIAARACHVIVUTULIDAAVUTSTHEMENNITI LOALSD-HLIVUSGUTGLEIGAANRIAANGVIVUTULIDAAVUTSTASTAUGIA HUVVGGUTGLEIGAANRIAANVILIDAANVANTUVIDUTAVUTAVUTSTULAVUTSTASTAUGIA LUVCSC-HTLVUGASTVALECASTLAGIGLUVTWINSI-LLA-GTDQUGASKUGDYMENHEIKI LUVSTC-HTLVUGASTVALECASTLAGIGLUVTWINSI-LLA-GTDQUGASKUGDYMENHEVK LUNDSGHTLVUGASTVALECASTLAGIGLUVTWINSI-LLA-GTDQUGASKUGDYMENHEVK LUNDSGHTLVUGASTVALECASTLAGIGLUVTWINSI-LLA-GTDQUGASKUGDYMENHEVK LUNDSGHTLVUGASTVALECASTLAGIGLUVTWINSI-LLA-GTDQUGASKUGDYMENHEVK LUNDSGHTLVUGASTVALECASTLAGIGLUVTWINSI-LLA-GTDQUGASKUGDYMENHEVK LUNDSGHTLVUGASTVALECASTLAGIGLUVTWINSI-LLA-GTDQUGASKUGDYMENHEVK LUNDS
labid la	IRESTRM-SCRAITICOCFTGLEIAANIAARACTWYKLINRCAMTGLORELISSMIKKMALETUV LEKKYTDSYNWYVIGCHTGLAAARAARACHUVYUVIDIDSKUVIDKITSMIKTENELSINTI LONLDD-HLIVUSGUTGLEIGAARRIAAKUVIDIDLOBAUGVINKET SUUTENELSINTI LONLDD-HLIVUSGUTGLEIGAARRIAAKUVIDUIDASULT-YEDDALAVUEESTARBOW PKLLDO-HLIVUSGUTGLEIGAARRIAAKUVIDUIDASULT-YEDDALAVUEESTARBOW INSULTAGRUVVIGGUTGLEIGAARRIAAKUVIDUIDASULT-YEDDALAVUEESTARBOW INSULTAGRUVVIGGUTGLEIGAARRIAAKUVIDUIDASUNTABVISI-LLA-GFORGUARKUGINEHHEIK INSULTO-HTIVIGASYVALECASTLASIGGUVIWWISSI-LLA-GFORGUARKUGINEHHEIK LINDDGRUVVIGGUTGLEAARRIAAKUSAUSULLA-GASTLASUGUVIWWISSI-LLA-GFORGUARKUGINEHHEIK INSULTO-HTIVIGASYVALECASTLASIGGUVIWWISSI-LLA-GFORGUARKUGINEHHEIK INSULTGRUVVIGGUTGLEAARRIAAKUSAUSULLASIAUUNIGENDI-STUDYFIGLIGEQUSEKI-AKILGUL IISGSMUTHIERDING
labiba lacita lacita likyla likyla likyla likyla likyla likyla lacita likyla lakyla likyla lakyla lakyla likyla lakyla likyla lakyla likyla la	IRESTENSCRAITICOCFTGLEIAANIAARACYUVILIBACMATCLORELISMIKKMILETEVE LAKYTVDSYNWYUGSOTGLEIAARIAARAKAKUVIVILIDJAPUVAVINKETTOVIEHEMINII LANLSD-HILIVUSGOTGLEIAARIAARAKUVIVIIDIJABUAVUINKETTOVIEHEMINII LANLSD-HILIVUSGOTGLEIGIATRILAAVIVIUVIANGUVIL-YEDDALUVESTARKUTA BYKLDQ-HILIVUSGOTGLEIGIATRILAAVIAANUVILIDIJAAVUSTI-LIDA-STRUCIA LBYKDQ-HILIVUSGOTGLEIGIATRILAAVIAAUUTLIDIJAAVUSTI-JUB-STRUCIA LBYKDQ-HILIVUSGOTGLEIGIATRILAAVIAAUUTLIDIJAAVUSTI-JUB-STRUCIA LBYKDQ-HILIVUSGOTGLEIGIATRILAAVIAAUUTLIDIJAAVUSTI-JUB-STRUCIA LBYKDQ-HILIVUSGOTGLEIGIATRILAAVIAAUUTLIDIJAAVUSTI-JUB-STRUCIA LBYKDQ-HILIVUSGOTGLEIGAATRILAAVIAAUUTLIDIJAAVUSTI-JUB-STRUCIASI LBYKD
labha ladia ladia ladia livyla likyla likyla livyla li	INFESTEN-SCRAITICOCOFTICETIAGNIARACTIVULI IRRCAMTCLORELISSMI KURMETTEVEN LOKUTYDOFUNDIVUSCASUTCETAARATAKAKUVIVUTU LIDADAVUTESTARUSUTENEN MITTU LOKUTYDOFUNDIVUSCASUTCETAARATAKAKUVIVUTU LIDADAVUTESTARUSUTENEN MITTU LOKUTYDOFUNDI CONTIGUESTARTAKAKUVIVUTU LIDADAVUTESTARUSUTENEN MITTU LOKUTYDOFUNDI CONTIGUESTARTAKAKUVIVUTU LIDADAVUTESTARUSUTENEN MITTU LUVICOTICUESTICETAARTAKAKUVIVUTU LIDADAVUTESTARUSUTENEN MITTU LUVICOTICUESTICATARTAKANUVILUDAAVUSTUTADAVUTYDUSTUKUTASUVAU LUVICOTICUESTICATARTAKANVILUDAAVUSTUKUTADAVUSTUKUTADIVUSATURVITURISTARUSUTENEN LUVICOTICUESTICAARTAKAKUVIVUTUSI LILA-GTDQÜGAKUVGIYMENINGEN LUVICOTICUESCASITICAARTAKANVILUTAARAVUSTUKUTADIVUSATUKUTASUVAU LUVICOTICUESCASITICAARTAKANAN LUNDOCHTLVICASIVALECASITICASUVIVUVAUSISTUL-GTDRÜGSKUVAIMEENTETTE INTERTIVUKASIVALECASITICASUVIVUVAUSISTUL-GTDRÜGSKUVAIMEENTETTE INTERTIVUKASIVALECASITICASUVIVUVAUSISTUL-GTDRÜGSKUVAIMEENTETTE INTERTIVUKASIVALECASITICASUVIVUVAUSISTUL-GTDRÜGSKUVILAAVUSENTANLUKUT-LILLI INTERTIVUKASIVALECASITICASUVIVAUSISTULUA-GEDRÜGSKUVAIMEENTET INTERTIVERENAM
labia labia labia labia libia libia libia labia	IRESTENSCRAITICOCFTGLETAGNARACYUVULIBACMOTGLORELISMIKKMLETUV LOKUTVDSUNWUVUSGOYGAETAAATAARAKUVUVULIDABUVUVUVUSUISMUSTASUKUESHA LOKUTVDSUNWUVISGOYGAETVAATIKAUVUVUVUSUISMUSTASUKUES LOKUTVDSUNWUVISGOYGAETVAATIKAUVUVUVAUSUUT-YEDDAAUVLESSARBOVA BYKLDCHILVUGASYALECASTLAGIGLUVUVUVASI-LLA-GTDQUMANKUGHMEHHEI LUVUSGOYGAETVAATIKAUVUILUDAAVUSUVIA-YEDDAAUVLESSARBOVA LUVUS
labiba lacita labiba lacita libyA li	IRESTEN-SCRAITICOCFTGLETAGNARACTWYLINKALMECHT-CLORLISMIKKEMLETTWY LEXKYTVDSYNWYUGSYTGLETAARATAKAKKYWYUTUIDSAVYINKETSYNWITEHEMINIT LEXKYTVDSYNWYUGSYTGLETAARATAKAKKYWYUTUIDSAVINKETSYNWITEHEMINIT LEXKYTVDSYNWYUGSYTGLETAARATAKAKKYWYUTUIDSAVINUS TILLA-STDQUAAKYGENSERVA FXLD
labiba labiba labiba labiba labiba libya libya libya libya labiba	INFESTEN-SCRAITICOCFTCLETLACHARACTURAL HERCMETCLORELISENTIKEMENTETURY         LORATUDETVANVUGSOTGAETVAATTELAATAKAKAKUVIUTULDADAVAUTUKET FUNCTENELANNII         LORATUDETVANVUGSOTGAETVAATTELAVAVUTUUTUADAVAUTUKETSATAKUSTENELANNII         LORADD-HLIVVOSGUTGAETVAATTELAGUGVIVANSI-LLA-GTDQUBANKICEHEEHEIT         ISBQLIAOLAUVVOGGUTGAETVAATATEMBUVIUTUDAAVUSTATAVAUTUKETSATSUKUESIA         LEYEDC-KTLVVGASVALECAETLAGICLUVIVANSI-LLA-GTDQUBANKICEHEEHEITE         ISBQLIANEUVVOGGUTGAETVAATATEMBUVIUTUDAASVENTAVIVANSI-VLA-GETDQUBANKICEHEEHEITE         LANDDC-KTLVVGASVALECAETLAGICLUVIVANSI-LLA-GFDQUBANKICEHEEHEITE         ISUSSIVTEIKEENAK       - EVOUVVIVANTEN-GENEL-STIDFVFLGLESQUBASKICEHEEHEITE         ISUSSIVTEIKEENAK       - EVOUVVITO
labiba ladia labiba ladia livla livva labiba	INFSTEN-SCRAITICOCFTGLEIAANAARACYUVULIDACUMCUMULISHUKUMLETTÖVEL         LOKUTVDSTANNUVIGSOTGLEIAAANAARACHUVUTULIDADAUVULESTARISUKULEIAUKUTULIDAUVUTUSIDAUVULESTARISUKU         LOKUTVDSTANNUVIGSOTGLEIAANAARACHUVUTULIDADAUVULESTARISUKU         LOKUTVSTANNUVIGSOTGLEIAANAALAANACHUVUTULIDAAUVULESTARISUKUS         LEYCS
labid labid labid labid livlA livvA labid	IRESTEN-SCRAITICOCFTCLEIAANIAARACTUVALIMECANTCLORLISMIKKMLETEV LAQKTUPSEVANVVIGSOFTGLENAARIAARAKKUVIVIDIOBAUVAUTUKETSULISMIKKMLETEVE LAQKTUPSEVANVVIGSOFTGLENAARIAARAKKUVIVIDIOBAUVAINIKETSULISMIKKMLETEVE HALVVIGSOFTGLEIAARIAARAKAKUVIVIDIOBAUVAINIKETSULESTAREVA PKALD
labiba labiba labiba labiba labiba libya libya libya libya labiba	INFESTEN-SCRAITICOCOFTICLELAGNARACYUVUL IARCAMELCLUELISMILKEMLETEVEK LEXKYTVDEVANWVIGSOYTGELAARATAKAKUVUVUL IARCAMUVUKUTENELAMINI LANLSD-HILVUVGSOYTGELGAATRILAGUVUVUVAGUVUKUTENTADUVALUVESTAREVKI IKALSO-HILVUVGSOYTGELGAATRILAGUVUVUVASI LULA-STADUALUVESTAREVKI IKALSO-HILVUVGSUVIGELGAATRILAGUVUVULDIAARUVUTUDIAAUKUVIAUVASIKUUESTAREVKI IKALSO-HILVUVGSUVIGELGAATRILAGUVULDIAAUKUVIDUVASUVUTUSTASUVUKUSUKUVUULESTAREVKI IKALSO-HILVUVGSUVIGELGAATRILAGUVUVUVASI-LLA-GFDQUARKVODYKEHHEIKI LUVTS
labiba la	INFERTM-SCRATTICOGEFICALTAANAACTIVULI MECMATCLORELISMITKUMLETTEVEN         LOKUTUDETVANVUGGUTGETVAATATAANAACTIVUTUDI DADLAVUTUESTARBUUR         HELLYUSGUTGETVAATATAANAACTIVUTUDI DADLAVUTUESTARBUUR         HELLYUSGUTGETVAATATAANAACTIVUTUDI DADLAVUTUESTARBUUR         HELLYUSGUTGETVAATATAANAACTIVUTUDI DADLAVUTUSESTARBUUR         HELLYUSGUTGETVAATATAANATAANATUVUTUDI DADLAVUTUSTASTATUGI DUTUNGESTARBUUR         LEYED
lababa labababab	IRESTEN-SCRATT COCOFT GLETAGNIARACTIVIL IRRGAMTGLORELISMIKEMENTETV LORUTDESTANNIVGSOTGAETVAATRILAVKIVIUT LIDAGVAVITALESTAREVA REALT
labaha labahabaha labahabaha labahabaha labahabahabaha labahabahabahabahabahabahabahabahabahaba	IRESTEN-SCRAITICOCFTGLEIAANAARACYUVULIDACUMCUTUSULSUNUKUMULESTARTOVA LOKUTVDEYUNUVUGSUTGLEIAANAARACKUVUVULIDASUVUKUSUTUSUESTARTOVA BYKLDQ-HLIVUSGUTGLEIAANAARACKUVUVULIDASUVIKUTDISUDALUVLESTARTOVA BYKLDQ-HLIVUSGUTGLEIAANAARACKUVUVULIDASUKUTJAD-YEDDALUVLESTARTOVA BYKLDQ-HLIVUSGUTGLEIAANAARACKUVULIDASAAVAUVTJAD-YEDDALUVLESTARTOVA BYKLDQ-HLIVUSGUTGLEIAANAARACKUVULIDASAAVAUVTJAD-YEDDALUVLESTARTOVA BYKLDQ-HLIVUSGUTGLEIAANAARACKUVULIDASAAVAUVTJAD-YEDDALUVLESTARTOVASISUKUELIA LUYDCHIVUTGASVALECASTIAASIGUVUVANSI-LLA-GFDQQGARKVEDYKENHEVK LUNDGHUVUSGUTGLECASTIAASIGUVUVANSI-LLA-GFDQQGARKVEDYKENHEVK LUND
labiba la	INFSTRM-SCRAITICOCFTGLEIAAMARACKIWIKLINGAMETGLORELISMIKKMLETEVT       LOKKYTDSTANNVUSGOSTGLEAAAAAAAAKUVIVUSUUGADAVAVILAKETSANTIKEMELMINI LOKKYTDSTANNVUSGOSTGLETAAAAAAAAKUVIVUSUUGADAVAVILAKETSANTIKEMELMINI LOKKYTDSTANNVUSGOSTGLETAAATAIAAANAKUVIVUSUUGADAVATASTANGUSASIKKUGI LUYG
labid la	TRESTEN-SCRATT GOOGFTGLETAGNARACTWYLLINGKAMT-C-GLORELSSMITKEMLETEVE LOKYTVDSTWAWVIGSGYGAETAAATAAAKACKWYVTUNID BOLYGWYLAKET SUVITENEMANII LOKYTVDSTWAWVIGSGYGAETAAATAAAKACKWYVTUNID BOLYGWYLAKETSTUNETENEMNI FUKLS
labaha labaha ladia labaha ladia labahabaha labahabaha labahabaha labahabahabaha labahabahabahabahabahabahabahabahabahaba	TRESTENSCRATT COCFT CLETACHARACTIVAL IMECANT CLORELISMITKEMLETEVE LOGYTVDEYNWVVGSCYGGET LAAATAKAAKAYWVVUTDLODA DVWVIACH STUTTENELMINI LOALDD-HLIVVGSCYGGET LAAATAKAAKAYWVUTDLODA DVWIACH STUTTENELMINI LOALDD-HLIVVGSCYGGET LAAATAKAAKAYWVUTDLODA DVWIACH STUTTENELMINI LOYD
labiba labibaba labiba labiba	TRESTEN-SCRATT GOOGFTGLETAGNARAGYUVUL TARCMET, -GLORELSSMITKENLETTEVE LOKUTYDEVUNUVUGSUTGLETAAATAAAAAKUVIVUL TOLDALVAVINGET SUUTENELSSMIT LOKUTYDEVUNUVUGSUTGLETAAATAAAAKUVIVUL TOLDALVAVINGET SUUTENELSSMIT LOKUTYDEVUNUVUGSUTGLEGATATIALAANAKUVIVUL TANGALVAVINGE TOLVASSIKKUGI LUVIG
labida labida	TRESTEN-SCRATT GOOGFTGLETAGNARACYWYL INGALWYL CHURCH SWITTENER MINT LOALDD-HLIVUSGOTGLETAARATAAAKKWYUTU IDD ADVAWINGET SWITTENER MINT LOALDD-HLIVUSGOTGLETAARATAAKKWYUTU IDD ADVAWINGET SWITTENER MINT LOALDD-HLIVUSGOTGLEGTATRELAGVYUVANGUNUL-YEDDALVLEESTARROVA BYRLDQ-HLIVUSGOTGLEGTATRELAGVYUVANGUNUL-YEDDALVLEESTARROVA INGLNEUVVGCOTGLEGTATRELAGVYUVANGUNUL-YEDDALVLEESTARROVA BYRLDQ-HLIVUSGOTGLEGTATRELAGVYUVANGUNUL-YEDDALVLEESTARROVA BYRLDQ-HLIVUSGOTGLEGTATRELAGVYUVANGUNUL-YEDDALVLEESTARROVA BYRLDQ-HLIVUSGOTGLEGTATRELAGVYUVANGUNUL-STRATENTA PVGAVVENIERINGV LWNDGETLIVUSGOTGLEGAATRICAGVYUVANGUNUL-GETRGLEGTARD VENIERIA BYRDG-HLIVUSGOTGLEGAATRICAGVYUVANGUNUL-GETRGLEGTARD VENIERIA BYRDGETLIVUSGOTGLEGAATRICAGVYUVANGUNUGETRGLEGAVROUMENTIGT INFORMATING
labaha labahabaha labahabaha labahabahabahabahabahabahabahabahabahaba	TRESTEN-SCRATTICCGTICLIAAGNARACYUVULIREANTICLEELSNITKULLETOV LEQUTVDSEVINVUIGGSITCLIAAGNARACYUVUILIREANTICLEELSNITKULLETOVI LEQUTVDSEVINVUIGGSITCLAEFUDATTELEVUVUVASQDEVUL-YEDADALVLESFSFAROV BRALBD-HLIVVOGSUTCLISLIGATRICLOVITWASQDEVUL-YEDADALVLESGSFAROVA INFORMATION CONTRACTION OF A CONTRACT AND A
labiba la	INSTEN-SCRATTICCGTICLIAARAACCUVUTINECAMTL-CLIERCAMTL-CLIERCENNITHCLIETOW LEQUIVDESVENUWUGSGUTCALFUDATTELCVUVUVASQDEVLP-YEDADAALVLEESFAEROW BRLDD-ELIVUVGSGUTCALFUDATTELCVUVUVASQDEVLP-YEDADAALVLEESFAEROW BRLDD-ELIVUVGSGUTCALFUDATTELCVUVUVASQDEVLP-YEDADAALVLESFAEROW INSTALSO-ELIVUVGSGUTCALFUDATTELCVUVVASQDEVLP-YEDADAALVLESFAEROW INSTALSO-ELIVUVGSGUTCALFUDATATELCVUVVASQDEVLP-TEDADAALVLESFAEROW INSTALSO-ELIVUVGSGUTCALFUDATATELCVUVVASQDEVLEVTAPVASESKUCGIA LYPEDO-ETIVUGSTVALECASTICAMENTIVUVASDI-LLR-CSTDQCAERVOLTMENTOK LWND

*Figure 5.* Generated target sequence and multiple template alignment.

Once the target-template alignment was generated, ten protein structures were developed for the target protein using the multi-template and target sequence alignment. All the proteins were further subjected to evaluation through molpdf, DOPE and GA341 scoring. The obtained results are displayed in Table 8 along with their respective model numbers. The 4<sup>th</sup> model was found to be comparatively good based on the above mentioned evaluation scoring and was considered for the further analysis. The obtained model is represented in new cartoon format and surface view in Figure 6. The developed structure showed the alpha helices and beta sheets almost similar to other thioredoxin reductase proteins along with some long loops.

**Table 8.** Molpdf, DOPE score and GA341 score of best ten threedimensional structural models developed with multiple templates.

Model number	molpdf	DOPE	GA341
Model 4	27163.44336	-55457.5	1
Model 1	26907.65625	-55397.7	1
Model 10	26742.1875	-55374.7	1
Model 7	26880.80469	-55139.8	1
Model 8	26782.80469	-54850.4	1
Model 6	27027.62305	-54804.3	1
Model 5	27123.99609	-54774.9	1
Model 2	27446.76563	-54381.3	1
Model 3	27149.9707	-54333.4	1
Model 9	27038.10547	-54202.5	1

#### Loop optimization of the best structure

The existence of multiple loops suggested the further need for loop optimization. Therefore, all loop regions with poor energy profile and structural orientation were identified using energy profiling and structural evaluation and visualization and the structure was subjected to loop optimization.

**Table 9.** molpdf and DOPE score of best ten loops optimized three
 dimensional structural models developed with multiple templates.

Structure model number	molpdf	DOPE
Model 8	1184.14331	-10252.12012
Model 2	1651.75635	-9835.49707
Model 1	903.81464	-9727.25488
Model 4	2020.79944	-9114.76074
Model 9	5959.34863	-8566.29785
Model 6	3826.18115	-8330.34473
Model 5	4257.93311	-8059.69629
Model 3	5646.4541	-8047.19873
Model 7	4218.79346	-7961.8457



**Figure 6.** Three dimensional structure of thioredoxin reductase of B. microti visualized using VMD. (A) Representation of the  $\alpha$  helices (purple),  $\beta$  sheets (yellow) and loops (cyan) of the final protein generated after loop optimization through multiple-template based modeling approach. (B) Surface view of the protein where purple colored surface refers to the  $\alpha$ -helix region and cyan loops.



**Figure 7.** Three dimensional structure of thioredoxin reductase of *B.* microti with the loops optimized for 10 structures visualized using chimera.



Figure 8. Superimposed target-template structures. The target is represented in light grey and all other templates are depicted in various other colors superimposed with the target protein molecule.

The evaluation scores for the 10 models developed during loop optimization considering the previous multi-template based best structure as starting structure are provided in Table 9. The model 8 was found to be the best loop optimized model and was considered for all relevant analysis. Figure 7 represents the structures with loops after optimization.

Structural analysis of the Babesia microti thioredoxin reductase: a potential drug target for babesiosis treatment

#### Structure validation

The final structure was identified after the loop modeling and subjected to superimposition with the considered multiple templates. The obtained RMSD was within the allowed region for most of the templates. The superimposition of the final modeled structure with multiple templates is shown in Figure 8. Respective RMSD value ranges and distance observed between the target and the templates are represented in Figure 9. The RMSD values ranged from 0.506 Å to 2.457 Å.



**Figure 9.** Target-templates superimposition distance matrix with distance values and standard deviations. This is a 2D matrix representation of the target and the template proteins where ranges of deviations are represented in angstrom.



Figure 10. Stereochemical quality check using RAMPAGE.

The final structure was further evaluated for stereo-chemical structural quality through various standard tools such as RAMPAGE [44], PROSA [53] and ANOLEA.

#### **RAMPAGE** analysis

Stereochemical quality of model was checked using RAMPAGE server (Figure 10). Ramachandran plot obtained using RAMPAGE revealed that 90.2% of residues were within the most favored regions. Residues falling in additionally allowed regions and outlier residues in outlier region were 7.1% and 2.7% respectively.

#### Validation through PROSA

PROSA (Protein Structure Analysis) was used to evaluate the quality of 3D models of protein structures. Z-score is a measure of overall model quality and denotes the deviation of the total energy of the structure compared to energy distribution derived from random conformations. The PROSA

score was -7.71 for the modeled protein, which indicates its correctness.



**Figure 11.** PROSA analysis of the modeled structure for structural evaluation (A) The black dot shows the position of the modeled structure with comparison to the existing X-ray crystallography generated and NMR based structures. (B) Comparative estimation of the energy profile with relation to the knowledge-based energy profiles. (C) Depiction of the structure based on the energy values where "blue" represents lowest energy and "red" represents highest energy in the protein.

PROSA profile for the target protein model was found better than all other template structures (PDB ID: 1mo9, chain "A"-11.77; PDB ID: 2x99, chain "A"-11.87; PDB ID: 1xhc, chain "A"-9.98; PDB ID: 1nhp, chain "A"-12.77; PDB ID: 1xdi, chain "A"-11.07; PDB ID: 1lvl, chain "A"-11.0; PDB ID: 1h6v, chain "A"-10.93 and PDB ID: 1q1r chain "A"-11.43). Zscore computed by PROSA for thioredoxin reductase of *B. microti* was found to be better than the Z-score of all the templates (Figure 11A). Negative values in PROSA plot with comparison to the knowledge based energy values indicated the stable regions of the protein and authentic model development (Figure 11B).



*Figure 12.* Complex protein topology displaying the direction of multiple helices and parallel and anti-parallel beta sheets generated for the target protein.



Figure 13. Depiction of top 10 pockets in target protein molecule.

## Analysis by ANOLEA

Analysis of the modeled protein structure with ANOLEA (Atomic Non-Local Environment Assessment) [53] also provided expected outcome. The analysis was conducted with a window size of 9 amino acids. The results suggested that out of 553 amino acids, only 114 (20.61%) showed comparatively higher energy. The Figure 11C also provided the similar information with a structural view. Computed non-local normalized energy Z-score of 2.80 was considerable for this large protein. The amino acids with high energy values were residue number 42; 52-56; 99-101; 132-135; 137-139; 151;

212; 228-233; 256; 310-318; 332-337; 351-354; 440-460; 494-498; 500-501; 507-539; 541-549.



**Figure 14.** (A): Representation of the identified pores; (B): major residues participating in the interaction with ligand molecules and (C-L): docking modes of the top 10 molecules along with respective hydrogen bond interactions. Information related to detailed interactions is provided in Table 19.

## Structural topological analysis

Structural topological analysis is important for this protein considering its complex structural and functional diversity. The

outcome of the PROMOTIF topological analysis is provided in Figure 12.



**Figure 15.** Flexibility analysis of the target protein. (A) Cartoon representation of top 10 models with fluctuations depicted in different color code. (B) Contact map generated for all 10 proteins. (C) Representation of obtained residue wise RMSF values for the structural models depicting the fluctuation for each residue in the protein structure.

The figure depicts the alpha helices, parallel and anti-parallel beta sheets and loop regions in detail along with the residue numbers. It is evident that the overall structure contains 22  $\beta$ -sheet regions, 19 helices and several loop regions.

#### Secondary structural analysis from structure

Detail secondary structural analysis was performed for the final protein structure developed. The post loop optimized best structure was utilized for this purpose and PROMOTIF was considered for the analysis.

As expected, the alpha helix portion was found to be maximum in the structure with 179 amino acids out of 553 being part of it which was 32.4% followed by the strand portion (108 amino acids (19.5%)). 3-10 helix structures were found to be comprising of 5 amino acids (0.9%) and other important structural component were constituted by 261 (47.2%) amino acids. The overall structure contains several structural components such as beta sheets, beta-alpha-beta structural unit, beta-hairpins, beta-bulges, strands, helices, helix-helix interactions, beta turns, gamma turns and disulphide linkage. The detail description is provided in the following section.

**Beta sheets:** Total 6 beta sheets were found in the *Babesia* thioredoxin reductase structure. The number of strands varied from 2 to 5 as shown in Table 10. No barrel like structures were found for these predicted sheets. Presence of parallel and anti-parallel sheets were observed for this structure as represented with their respective topology in Table 10.

**Table 10.** Beta sheets present in the modeled structure along with number of strands, type and topology.

Sheet	No. of strands	Туре	Barrel	Тороlоду
А	4	Parallel	No	-1X 2X 1X
В	3	Antiparallel	No	1 1
С	2	Antiparallel	No	1
D	5	Parallel	No	-2X -1X -1X 3X
E	3	Antiparallel	No	1 1
F	5	Antiparallel	No	-1 -3 1 1

**Beta-alpha-beta units:** Beta-alpha-beta units are important structural components of thioredoxin reductase structure. Table 11 represents the strands along with the start, end residues and residues distributed in the helix and loop regions. This topology suggests the presence of alpha helical region sandwiched by beta sheets and most of the time such structural arrangement serves as a conserved topology in a protein structure.

Table 11. Distribution of the amino acid residues in the beta-alpha-beta units.

Strand 1			Strand 2	Strand 2			No. of residues	
Start	End	Length	Start	End	Length	No. of residues in loop	No. of residues in helices	
Tyr 59	lle 64	6	Val 84	Phe 87	4	1	19	14
Lys 238	Val 242	5	His 261	Val 266	6	1	18	14
His 261	Val 266	6	His 292	Tyr 295	4	1	25	15

#### **Beta-hairpins units**

The beta-hairpin structural components suggest the presence of the anti-parallel beta sheets separated by a short loop region. Obtained structural regions are listed in Table 12 along with their respective hairpin class. Many hairpin classes were observed as provided in Table 12. Beta hairpin structural elements often indicate the folding pattern of a protein molecule and aid in unraveling the protein folding.

**Table 12.** Beta-hairpin regions in the modeled structure along with their classes.

Strand 1			Strand 2			Hairpin class
Start	End	Length	Start	End	Length	
Arg178	Lys182	5	lle187	Asp191	5	2:4
lle187	Asp191	5	Glu196	Val199	4	3:5 IG
Pro299	Cys306	8	Lys309	Phe314	6	2:2 II
Lys309	Phe314	6	Ser318	Phe322	5	2:4
Thr408	lle410	3	Phe416	Gly420	5	5:5
Leu435	Phe442	8	Cys471	Glu478	8	28:28
Cys471	Glu478	8	Val484	Val490	7	4:6

**Beta-bulges:** Beta bulges are important structural regions which are responsible for slight structural deviation in the beta sheets through introducing additional H-bonds in a beta sheet. They are categorized based on the length of disruption in the beta strand. Apart from simple distortion in the native structures, a beta bulge may have far reaching functional impact as well. Often, beta bulged regions may indicate an important structural region which may have been mutated yet maintained structural conservation considering the importance of the structural conservation in that particular region. The obtained beta bulges along with their respective types and impacted residue as res X are presented in Table 13.

 Table 13. Beta-bulges in the modeled structure along with classes.

Bulge type	Res X	Res 1	Res 2
Parallel bent		Lys202	Asp60
Antiparallel classic	His213	Thr329	Gly330
Antiparallel G1	Cys306	Gly308	Lys309

6	1	18		14	
4	1	25		15	
Antiparallel clas	sic	Lys309	Glu305	Cys306	
Antiparallel wid	e	Lys311	Asp320	Thr321	
Antiparallel clas	sic	Val313	lle300	Asn301	
Antiparallel clas	sic	lle476	Val485	Gly486	
Antiparallel G1		Glu478	Asp482	Leu483	

**Strands:** The combination of helices and strands in thioredoxin reductase structure makes it unique and complex for its important function. The strand regions are presented in Table 14 along with their respective length.

**Helices:** Apart from strand regions, multiple helix regions also develop the backbone of the large thioredoxin protein. Obtained helices are presented in Table 15 along with their respective length.

Table 14. Components of the strand for the generated protein model.

Start	End	Sheet	No.of residues
Tyr59	lle64	A	6
Val84	Phe87	А	4
Arg178	Lys182	В	5
lle187	Asp191	В	5
Glu196	Val199	В	4
Ala201	Leu206	А	6
Arg211	Pro212	С	2
Leu222	Val223	D	2
Lys238	Val242	D	5
His261	Val266	D	6
His292	Tyr295	D	4
Pro299	Cys306	E	8
Lys309	Phe314	E	6
Ser318	Phe322	E	5
Thr324	Tyr327	D	4
Arg331	Ser332	С	2
Val365	Ala367	А	3
Thr408	lle410	F	3
Phe416	Gly420	F	5

Leu435	Phe442	F	8
Cys471	Glu478	F	8
Val484	Val490	F	7

 Table 15. Helix regions found in the modeled structure of thioredoxin reductase of Babesia microti.

Start	End	Туре	No.of residues
Leu7	Val13	н	7
Arg14	Leu16	G	3
Lys36	Thr45	Н	10
Val48	Ser53	Н	6
Pro68	Arg79	Н	12
Gly103	Val108	Н	6
Cys110	Leu132	Н	23
Trp145	Lys170	Н	26
Thr225	Asp228	Н	4
Tyr246	Gly257	н	12
Arg276	Asn288	Н	13
Leu340	Val343	Н	4
Gly369	lle371	G	3
Ala379	Phe394	Н	16
Glu423	Tyr430	Н	8
Glu432	Asn434	G	3
Leu458	Glu462	н	5
Ala494	Leu507	н	14
Lys511	Leu516	Н	6
Ser527	Val533	Н	7

**Helix-helix interactions:** Helix-helix interaction is crucial in understanding the folding pattern of a protein structure especially in membrane proteins. Detail information about the obtained helix-helix interaction is provided in Table 16.

 Table 16. Depiction of helix-helix interaction of the modeled protein.

Helices		Helix types		Inte	raction type	No. interacting residues		
						Helix 1	Helix 2	
3	4	н	Н	С	N	2	1	
5	6	н	Н	n	n	1	2	
5	8	н	Н	С	С	4	3	
5	14	н	Н	I	I	7	8	
6	7	н	Н	С	n	4	3	
6	8	н	Н	С	I	2	4	

6	9	н	Н	С	С	1	2	
6	14	н	Н	Ν	I	1	1	
7	8	н	Н	I	I	4	5	
7	9	н	Н	n	с	1	1	
7	10	н	Н	I	I	4	4	
7	14	н	Н	n	n	1	1	
7	17	н	Н	С	С	2	1	
8	9	н	Н	I	С	2	1	
8	10	Н	Н	n	с	1	1	
9	10	н	Н	Ν	I	2	3	
10	11	н	Н	С	С	3	2	
11	15	н	Н	n	n	2	2	
18	19	н	Н	I	Ν	1	1	
18	20	Н	Н	Ν	Ν	2	1	
19	20	н	Н	I	I	3	5	

Beta turns: Information on beta turns provides important hint about the reverse turn in a large protein molecule. Table 17 represents the observed  $\beta$ -turns found in the modeled protein. Abundance of proline and glycine is common in  $\beta$ -turns. The obtained result suggests that turn types VIII and IV are dominant for this protein. In some of cases, H-bonds were also found. The type IV turn suggests they are not perfectly defined turns whereas type VIII suggest that these turns are having-60 degree Phi (i+1), -30 degree Psi (i+1), and -120 degree Phi and Psi angle for i+2 residues for developing the dihedral angle.

 Table 17. Beta turns observed in the modeled protein.
 Particular
Turn	Sequence	Turn type	H-bond
Met1-Arg4	MILR	IV	
Arg4-Leu7	RAGL	IV	Yes
Val57-Asp60	VKYD	VIII	
Lys91-His94	KPSH	IV	
His94-Thr97	HRGT	IV	
Arg95-Ser98	RGTS	IV	
Lys182-Asn185	KSAN	IV	
Ser183-Glu186	SANE	IV	
Ala184-lle187	ANEI	IV	
Asp191-Gly194	DTDG	I	
Asp219-Leu222	DKSL	IV	
Asp228-Tyr231	DLFY	IV	
Leu229-Leu232	LFYL	IV	
Phe230-Asn233	FYLN	IV	
Leu232-Asp235	LNND	IV	

Pro236-Thr239	PGKT	IV	
Leu271-Phe274	LRGF	IV	
Lys294-Ile297	KYGI	IV	
Lys304-Asp307	KECD	IV	
Cys306-Lys309	CDGK	II	Yes
Phe314-Asn317	FSDN	IV	
Phe322-Val325	FDTV	VIII	
Leu335-Met338	LADM	IV	Yes
Leu347-Asn350	LSPN	I	Yes
Thr360-Pro363	TSVP	VIII	
Val362-Val365	VPNV	I	Yes
Val372-Arg375	VENR	II	Yes
Met401-Ser404	MDYS	IV	
Thr412-Glu415	TPYE	VIII	
Ser444-Gln447	SLEQ	IV	
Arg452-Thr455	RMKT	IV	
Met453-Arg456	MKTR	IV	Yes
Thr455-Leu458	TRHL	IV	
Asp464-Leu467	DTDL	VIII	
Glu478-Thr481	EKGT	IV	
Lys479-Asp482	KGTD	11'	
Gly491-Ala494	GPNA	IV	Yes
lle520-Thr523	IHPT	VIb	
Thr534-Ser537	TKDS	I	
Asp536-Glu539	DSGE	IV	Yes
Glu539-Val542	ESWV	IV	
Ser540-Ser543	SWVS	I	
Gly545-Gly548	GGCG	IV	
Gly548-Lys551	GGGK	IV	

**Gamma turns:** Followed by beta turns, gamma turns are the predominant tight turns present in the protein structures. Table 18 represents the observed gamma turns in the modeled structure. Abundance of inverse gamma turns is common as found in this case also. Rare classic gamma turns are also found in this structure (Table 18). The gamma turns may have impact on structural folding of the protein and may help in developing the required quaternary structural configuration.

Table 18. Observed	gamma i	turns in th	e structure.
--------------------	---------	-------------	--------------

Start	End	Sequence	Turn type
Met338	Leu340	MNL	INVERSE
Pro376	Leu378	PQL	INVERSE

lle410	Thr412	IFT	INVERSE
Met453	Thr455	MKT	INVERSE
Lys479	Thr481	KGT	CLASSIC
Val533	Lys535	VTK	INVERSE

**Disulphide linkage:** Existence of disulphide linkage is important for proteins involved in redox associated reaction. The analysis suggested that the modeled structure is having disulphide linkage between the cysteine 105 and cysteine 110 residue number.

## Active site identification

Castp server was used to predict possible binding sites of modeled structure. Out of all the sites predicted, best 10 sites were selected as shown in Figure 13. Binding sites having highest surface area and volume were selected as the most probable active sites for the modeled protein (Table 19).

Table 19. The top 10 pocket	s with area and	l volume information.
-----------------------------	-----------------	-----------------------

Pocket ID	Area (SA)	Volume (SA)
1	1437.150	1334.336
2	517.250	693.072
3	266.266	132.424
4	137.962	81.917
5	201.313	77.787
6	127.716	77.102
7	135.291	48.060
8	88.826	42.928
9	65.263	33.284
10	82.267	29.309
11	88.254	25.771

## Molecular docking analysis

As mentioned in the method section, molecular docking analysis was carried out employing SwissDock docking facility. All the molecules were minimized employing CHARMM force field. The PSF, PAR files and the docking clusters were generated using the protocol provided. The parameters used are passive flexibility distance=0.0, wanted confs=5000, nbfact seval=5000, nbseeds=250, sdsteps=100, abnr steps=250, clustering radius=2.0 Å and maximum cluster size=8.

The target protein and the selected ligands were prepared through molecular mechanical calculations before docking exercises as per the tool protocol. The Supplementary Table 1 contains the detailed docking results of all the molecules considered. The docking analysis revealed the interactions between the ligand molecules and the target protein considered. The ligands are mostly auranofin derivatives as mentioned in the method section. The obtained ten best molecules along with their specific interactions are represented in Table 20. The energy values of the ligand-protein binding affinity were found between -16.51 and -20.88 kcal/mol. Some of the residues in the target proteins showed extensive affinity towards forming hydrogen bonds with the ligand molecules which are presented in "H-H bond receptor atom" column of Table 20. Interestingly, LYS 29, SER 245, CYS 110, LYS 210, SER 27 and LYS 202 showed more affinity towards the ligand molecules than other residues in the target protein. The poses of the docking are presented in Figure 14.

#### Target protein flexibility analysis

Molecular motion analysis provided information related to the most flexible as well as the rigid part of the target protein considered for this study. Estimated output of the flexibility of the modeled targeted protein computed through CABSflex server is provided in Figure 15. The outcome shows the deviation of the structural components of native protein as depicted in the Figure 15A. Fluctuation pattern is depicted through different color codes where maximum fluctuations are depicted towards the "C" and "N" terminal of the protein and medium range of RMSF was observed for the rest of the structural part. The contact map of ten different modes of the protein shows residue-residue fluctuations for all the amino acid residues. The RMSF fluctuation plot (Figure 15B) of all the modes shows that RMSF (root mean square fluctuation) value ranged approximately between 0 Å and 8 Å. The analysis suggests that the movement of the protein is higher around amino acid residue number 100, 245, 450 (Figure 15C) (region represented in green and dark brown in Figure15A). The fluctuations obtained around the C-terminal and the N-terminal of the protein may be ignored considering them as free ends.

**Table 20.** Docking result of the top 10 molecules with specific interactions and hydrogen bonds. The "Compound ID" column contains the conformations after the ID represented with ("\_") followed by the number.

Compound ID	Clust er	Clust er rank	Energy (kcal/mo I)	H- bond numbe r	First H-H bond receptor atom	First H-H bond ligand atom	Bond length (Å)	Second H- H bond receptor atom	Second H- H bond ligand atom	Bond length (Å)	Third H-H bond receptor atom	Third H-H bond ligand atom	Bond length (Å)
CID7067547 2	7	1	-20.88	2	LYS 29 HZ3	1.5 LIG 1 O5	2.157	SER 27 OG	1.52 LIG 1 H9	2.353			
CID1106849 9	0	2	-20.02	1	CYS 110 SG	1.3 LIG 1 O8	3.689						
CID5741707 2_1	2	0	-18.61	2	LYS 29	1.15 LIG 1 07	2.498	LYS 202 HZ1	1.15 LIG 1 O8	2.36			
CID88293	0	0	-18.39	1	SER 245 HG1	1.1 LIG 1 O9	2.064						
CID5694524 4	1	5	-18.37	1	CYS 110 SG	1.14 LIG 1 O6	3.846						
CID1090457 8_1	4	0	-17.54	1	CYS 110 SG	1.30 LIG 1 O8	3.294						
CID153711_ 3	2	5	-17.44	2	CYS 110 SG	1.22 LIG 1 O8	3.298	CYS 110 SG	1.22 LIG 1 07	3.711			
CID1110057 7_1	1	0	-17.40										
CID153711_ 12	8	0	-17.00	1	LYS 202 HZ3	1.50 LIG 1 O7	2.159						
CID153711	14	0	-16.51	3	LYS 29	1.92 LIG 1 07	2.378	LYS 202 HZ1	1.92 LIG 1 O9	2.001	LYS 202 HZ3	1.92 LIG 1 O1	2.17

## Conclusion

Lack of vaccine for combating human Babesiosis also implores us to search for novel drug targets and development of new, safe and effective chemotherapeutic agents. The infection is caused by the intra-erythrocytic parasitic genus *Babesia*. The transmission cycle of this parasite is maintained in both vertebrate and non-vertebrate hosts and the parasite infects a wide range of vertebrates. Across various geographical regions, human Babesiosis is caused by different parasitic species such as *B. microti, WA1 piroplasm, B. divergens* etc. Thus, it is cumbersome to develop a single potent and effective vaccine for effective prevention or control of human babesiosis. Several research efforts have been made to develop and crosscheck the possibility of a recombinant vaccine against different species of *Babesia*, such as *Babesia bovis*, *B. microti* etc. None of these efforts have been successful in providing complete solution in this regard. The complex process of host cell invasion of this parasite should be further explored to target several molecular events and multiple stages for vaccine design and development.

An effective inhibitor which may act against almost all the species would be a great support to the existing treatment as we cannot ignore the lethal consequences of the infection in several cases. Keeping in view the possibility of relapse and the emergence of atovaquone resistant strains after treatment, there is a dire need for exploring new drug targets and drugs.

In the current study, we developed a reasonably good high quality three dimensional structure of thioredoxin reductase of *B. microti* using homology modeling approach. The good quality of refined model was obtained using the state-of-the-art computational methodology and the outcome of the analysis was confirmed by using standard structure validation tools. As no experimentally determined structure of the considered enzyme is available, we believe that this model will aid in initiating drug discovery exercises and studies based on structural and functional insights obtained in this study. The docking analysis with derivatives of inhibitor auranofin which has already shown huge potential against an array of gram positive bacteria may pave a way towards procuring a series of excellent and effective medication against the pathogen.

## Acknowledgement

Neelima Arora thanks University grants Commission for Postdoctoral fellowship.

## References

- 1. Lobo CA, Cursino-Santos JR, Alhassan A, Rodrigues M. Babesia: an emerging infectious threat in transfusion medicine. Knoll LJ PLoS Pathogens 2013; 9: 1003387.
- 2. Skrabalo Z, Deanovic Z. Piroplasmosis in man; report of a case. Doc Med Geogr Trop 1957; 9: 11-16.
- Western KA, Benson GD, Gleason NN, Healy GR, Schultz MG. Babesiosis in a Massachusetts resident. N Engl J Med 1970; 283: 854-856.
- Esernio-Jenssen D, Scimeca PG, Benach JL, Tenenbaum MJ. Transplacental/perinatal babesiosis. J Pediatr 1987; 110: 570-572.
- 5. Leiby DA1. Babesiosis and blood transfusion: flying under the radar. Vox Sang 2006; 90: 157-165.
- Kjemtrup AM, Conrad PA. Human babesiosis: an emerging tick-borne disease. Int J Parasitol 2000; 30: 1323-1337.
- Regner EL, Thompson CS, Iglesias AA, Guerrero SA, Arias DG.Biochemical characterization of thioredoxin reductase from Babesia microti. Biochimie 2014, 99: 44-53.
- Banerjee AK, Arora N, Murty US. Structural model of the Plasmodium falciparum thioredoxin reductase: a novel target for antimalarial drugs. J Vector Borne Dis 2009; 46: 171.
- 9. Banerjee AK, Arora N, Murty US. Analyzing a potential drug target N-Myristoyltransferase of Plasmodium falciparum through in silico approaches. J Glob Infect Dis 2012; 4: 43.

- Arora N, Banerjee AK, Murty US. In silico characterization of Shikimate Kinase of Shigella flexneri: a potential drug target. Interdiscipl Sci Comp Life Sci 2010; 2: 280-290.
- Arora N, Banerjee AK, Murty US. Homology model of 2C-methyl-d-erythritol 2, 4-cyclodiphosphate (MECP) synthase of Plasmodium falciparum 3D7. Electron J Biol 2010; 6: 52-57.
- 12. Banerjee AK, Arora N, Murty US. Aspartate carbamoyltransferase of Plasmodium falciparum as a potential drug target for designing anti-malarial chemotherapeutic agents. Med Chem Res 2012; 21: 2480-2493.
- 13. Duddela S, Sekhar PN, Padmavati GV, Banerjee AK, Murty US. Probing the structure of human glucose transporter 2 and analysis of protein ligand interactions. Med Chem Res 2010; 19: 836-853.
- Arora N, Chari UV, Banerjee AK, Murty US. A computational approach to explore Plasmodium falciparum 3D7 chorismate synthase. Intern J Genom Proteom 2007; 3.
- 15. Banerjee AK, Murty US. Extracting the significant descriptors by 2D QSAR and docking efficiency of NRTI drugs: a molecular modeling approach. Intern J Genom Proteom 2007; 2.
- 16. Arora N, Narasu ML, Banerjee AK. Shikimate kinase of Yersinia pestis: a sequence, structural and functional analysis. Int J Biomed Data Min 2016; 5: 2.
- 17. Shinde SP, Banerjee AK, Arora N, Murty US, Sripathi VR, Pal-Bhadra M, Bhadra U. Computational approach for elucidating interactions of cross-species miRNAs and their targets in Flaviviruses. J Vector Borne Dis 2015; 52: 11.
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel, RD. Protein identification and analysis tools on the ExPASy server. Proteomics Protocols Handbook Humana Press 2005.
- Kurgan LA, Razib AA, Aghakhani S, Dick S, Mizianty MJ, Jahandideh S. CRYSTALP2: sequence-based protein crystallization propensity prediction. BMC Struct Biol 2009; 9: 50.
- Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA, Godzik A. XtalPred: a web server for prediction of protein crystallizability. Bioinformatics 2007; 23: 3403-3405.
- 21. Timothy LB, Charles E. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Sec Int Conf Intel Sys Mol Biol AAAI Press Menlo Park California 1994; 28-36.
- 22. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, Casadio R, Ben-Tal N. ConSeq: the identification of functionally and structurally important residues in protein sequences. Bioinformatics 2004; 20: 1322-1324
- 23. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. ConSurf 2016: an improved

methodology to estimate and visualize evolutionary conservation in macromolecules. Nucl Acids Res 2016.

- 24. Combet C, Blanchet C, Geourjon C, Deleage G. NPS@: network protein sequence analysis. Trends Biochem Sci 2000; 25: 147-150.
- 25. Deleage G, Roux B. An algorithm for protein secondary structure prediction based on class prediction. Protein Eng 1987; 1: 289-294.
- King RD, Sternberg MJ. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. Protein Sci 1996; 5: 2298-2310.
- Garnier J, Gibrat JF, Robson B. GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol 1996; 266: 540-553.
- Guermeur Y, Geourjon C, Gallinari P, Deleage G. Improved performance in protein secondary structure prediction by inhomogeneous score combination. Bioinformatics 1999; 15: 413-421.
- 29. Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 1993; 232: 584-599.
- 30. Frishman D, Argos P. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. Protein Eng 1996; 9: 133-142.
- 31. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 1995; 11: 681-684.
- 32. Levin JM, Robson B, Garnier J. An algorithm for secondary structure determination in proteins based on sequence similarity. FEBS Lett 1986; 205: 303-308.
- Deleage G, Blanchet C, Geourjon C. Protein structure prediction: implications for the biologist. Biochimie 1997; 79: 681-686.
- Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. Structure 2003; 11: 1453-1459.
- Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: exploring protein sequences for globularity and disorder. Nucleic Acids Res 2003; 31: 3701-3708.
- Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 2005; 21: 3369-3376.
- Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res 2007; 35: 460-464.
- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucl Acids Res 2014; 42: 252-258.
- 39. Angelucci F, Dimastrogiovanni D, Boumis G, Brunori M,

Miele AE, Saccoccia F, Bellelli A. Mapping the catalytic cycle of Schistosoma mansoni thioredoxin glutathione reductase by X-ray crystallography. J Biol Chem 2010; 285: 32557-32567.

- Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993; 234: 779-815.
- 41. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996; 14: 27-28, 33-38.
- 42. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera-a visualization system for exploratory research and analysis. J Comp Chem 2004; 25: 1605-1612.
- Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 2003; 50: 437-450.
- 44. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 1993; 2: 1511-1519.
- 45. Laskowski RA. PDBsum: summaries and analyses of PDB structures. Nucleic Acids Res 2001; 29: 221-222.
- Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res 2006; 34: 116-118.
- 47. Harbut MB, Vilcheze C, Luo X, Hensler ME, Guo H, Yang B, Chatterjee AK, Nizet V, Jacobs WR, Schultz PG, Wang F. Auranofin exerts broad-spectrum bactericidal activities by targeting thiol-redox homeostasis. Proc Nat Acad Sci 2015; 112: 4453-4458.
- 48. Grosdidier A, Zoete V, Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Res 2011; 39: W270-277.
- 49. Kuriata A, Gierut AM, Oleniecki T, Ciemny MP, Kolinski A, Kurcinski M, Kmiecik S. CABS-flex 2.0: a web server for fast simulations of flexibility of protein structures. Nucleic Acids Res 2018.
- 50. Hirt RP, Muller S, Embley TM, Coombs GH. The diversity and evolution of thioredoxin reductase: new perspectives. Trends Parasitol 2002; 18: 302-308.
- Xiong H, Buckwalter BL, Shieh HM, Hecht MH. Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. Proc Nat Acad Sci 1995 3; 92: 6349-6353.
- 52. Wiederstein M, Sippl M. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007; 35: 407-410.
- 53. Melo F, Feytmans E. Assessing protein structures with a non-local atomic interaction energy. J Mol Biol 1998; 277: 1141-1152.

## \*Correspondence to

Neelima Arora Centre for Biotechnology (CBT) Institute of Science and Technology (Autonomous) Jawaharlal Nehru Technological University Hyderabad Telangana India