J Mol Model (2005) 11: 194–199
DOI 10.1007/s00894-005-0234-0

ORIGINAL PAPER

**Sivaramakrishnan Thirumalai Nathan · Nisha Mathew
Muthuswami Kalyanasundaram
Kothandapani Balaraman**

Structure of glutathione *S*-transferase of the filarial parasite *Wuchereria bancrofti*: a target for drug development against adult worm

Received: 30 June 2004 / Accepted: 8 December 2004 / Published online: 29 April 2005

Abstract A three dimensional structural model of Glutathione-*S*-transferase (GST) of the lymphatic filarial parasite *Wuchereria bancrofti* (*wb*) was constructed by homology modeling. The three dimensional X-ray crystal structure of porcine π-class GST with PDB ID: 2gsr-A chain protein with 42% sequential and functional homology was used as the template. The model of *wb*GST built by MODELLER6v2 was analyzed by the PROCHECK programs. Ramachandran plot analysis showed that 93.5% of the residues are in the core region followed by 5.4 and 1.1% residues in the allowed and generously allowed regions, respectively. None of the non-glycine residues is in disallowed regions. The PROSA II *z*-score and the energy graph for the final model further confirmed the quality of the modeled structure. The computationally modeled three-dimensional (3D) structure of *wb*GST has been submitted to the Protein Data Bank (PDB) (PDB ID: 1SFM and RCSB ID: RCSB021668). 1SFM was used for docking with GST inhibitors by Hex4.2 macromolecular docking using spherical polar Fourier correlations.

Keywords *Wuchereria bancrofti* · Filariasis · Glutathione *S*-transferase · Parasite · Protein structure · Homology · Comparative modeling · Structure prediction

Introduction

Infection due to *Wuchereria bancrofti* accounts for more than 90% of the 120 million people afflicted by lymphatic Filariasis [1]. The drug of choice viz., diethylcarbamazine (DEC) used for treatment of this disease kills the brood of microfilariae (mf) circulating in the blood stream at the time of administration and does not kill the adults of *W. bancrofti*, living in the lymphatics. Since the pharmacokinetics of DEC is such that most of it is either metabolized or excreted within 24 h of administration, it is not expected to kill the subsequent brood of mf released into the blood stream by the adult females of *W. bancrofti*. Thus, the elimination of *W. bancrofti* in infected human beings is possible only by targeting the adult worms. This necessitates the development of molecules with potential to kill the adult parasites.

The current approach in the anti-parasitic drug discovery process involves the identification of novel targets from the databases on the parasite genome and metabolic pathways. Glutathione *S*-transferase enzymes (GST enzymes) are a family of detoxification enzymes that catalyze the conjugation of glutathione (GSH) with various endogenous and xenobiotic electrophiles. The GSTs have been considered as good targets for anti-parasitic drug development and studies have identified antischistosomal [2], antimalarial [3] and antifilarial [4–6] activity of compounds known for their GST inhibiting activity.

The GSTs appear to be the major detoxification system present in helminthes since there is no evidence that the oxygen-dependent P-450 system is expressed in adult worms [7, 8]. Nematode GSTs are topologically related to the πGSTs. They lack the extra helix of αGSTs or the μ loop characteristics of the μ-class GSTs. An interesting variation from the typical πGSTs was noted at the hydrophobic substrate-binding site of the Ov-GST2 that it possesses an open hydrophobic substrate-binding cleft [4]. However, the glutathione-binding site is closely related to that found in mammalian enzymes. The significant difference between the tertiary structure of the helminth GSTs and that of the host enzymes make the GSTs promising chemotherapeutic targets [4, 9–11]. Furthermore β-carbonyl substituted GSH conjugates

S. T. Nathan · N. Mathew (✉) · M. Kalyanasundaram
K. Balaraman
Vector Control Research Centre, Indian Council of Medical Research, Indira Nagar, Pondicherry, 605006, India
E-mail: nishamathew@yahoo.com
Tel.: +91-413-2272396
Fax: +91-413-2272041

display high activity and selectivity towards the filarial OvGST2 over human πGST and the same has been validated [4, 12].

In target-based drug development, the three-dimensional structure (3D) of the target is essential for defining the active site and designing and docking of small ligands to the target. Comparative or homology modeling uses experimentally determined protein structures (''templates'') to predict the conformation of other proteins with similar amino acid sequences (''targets'') [13–15]. This is possible since small changes in the amino acid sequence usually result in small changes in the 3D-structure of the protein [16]. To date comparative modeling remains the only method that can provide models with a root mean square error lower than 2 Å in cases of high sequence homology [17]. It has been confirmed that the application of comparative modeling in cases of high sequence identities (>35%) usually results in accurate models. The BLAST programs are widely used tools for searching protein and DNA databases for sequence similarities [18]. At the VCRC, Pondicherry, an attempt was made to generate a model 3D-structure of the enzyme of the *W. bancrofti* viz., *wb*GST by utilizing information on the amino acid sequences [11] and applying automated comparative protein modeling and bioinformatics tools and the same is reported here.

Methods

The single letter code amino acid sequence of GST of *W. bancrofti* (Accession number AY195867, Protein_id AAO45827. 1) was retrieved from NCBI database and taken as the target sequence. The modeling of 3D structure of *wb*GST followed a stepwise procedure, starting with a template structure search that related to the target sequence. This was done by protein–protein BLAST [18] (http://www.ncbi.nlm.nih.gov/BLAST) against PDB mode. From a number of hits, porcine π-class GST 2gsr-A chain [19] was taken as a template for modeling. The alignment of target sequence (*wb*GST) onto the template (2gsr-A) structure was carried out using ClustalW [20] (http://www.ebi.ac.uk/clustalw/) pairwise alignment tool for proteins by applying user-defined parameters. Thus, the 3D models generated based on the crystal structures of the porcine π-class GST with PDB ID: 2gsr-A chain was taken from the ExPDB server (www.expasy.org/swissmod/SM_Check_ExPDB.html).

After the selection of a suitable template, the automated comparative protein modeling program MODELLER6v2 [14, 21, 22] (http://salilab.org/modeller/modeller.html) was used to generate the model by an alignment of the target sequence with the selected template 2gsr-A sequence on alignment (.ali) file and the model-fast (.top) file was run. Since the length of the template was 207, an insertion was introduced at the 119th position to match with the target sequence length 208. In the first step of model building, distance and dihedral angle restraints on the target sequence were derived from its alignment with the template 3D-structures. The spatial restraints and the energy minimization steps were performed with the CHARMM22 force field for proper stereochemistry of proteins. Then, optimization of the model was carried out by the molecular dynamics simulated annealing method.

The modeled structure *wb*GST with minimum objective function was taken and evaluated by the PROCHECK [23] and WHATCHECK [24] programs to find errors in the modeled protein structure. The overall stereochemical quality of the protein was assessed by Ramachandran plot analysis [25, 26]. The modeled structure was further evaluated by PROSA II [27] by comparing the *z*-scores of the target and template. Molecular visualization of the final model was carried out with two different visualizing programs Swiss-pdb viewer [28] (http://www.expasy.org/spdbv/) and Hex4.2 [29] (http://www.biochem.abdn.ac.uk/hex/).

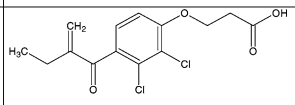
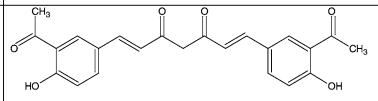
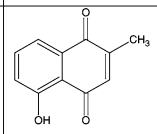
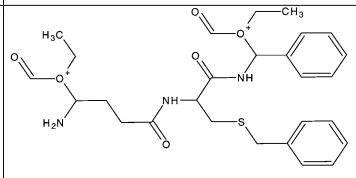
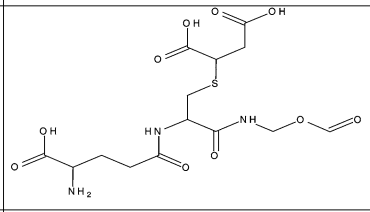
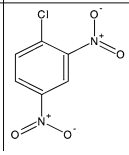
The modeled 3D structure of *wb*GST was used for docking with GST inhibitors to find the active pocket of ligand binding sites using the Hex4.2 macromolecular docking program with default parameters of the docking module. Hex employs spherical polar Fourier correlations, which removes the time consuming-requirement for explicitly generating different orientations of the mobile molecule. The compounds listed in Table 1 known for their inhibitory activity on GST were used to find the active sites on the modeled GST structure. Since Hex accepts only few file formats like PDB or MOL, the structures of the inhibitors were drawn and saved in PDB format by using Marvin sketch before docking. Then the receptor (*wb*GST) and the ligand were loaded and docking was carried out in the ''Full rotation mode''. During the docking process, initially Fourier transformation takes place followed by steric scan, final search, refinement and finally total dockings. From the total number of solutions, the ligand-binding site with minimum energy (kJ mol^{-1}) value was taken as the best solution.

Results

The results of the protein-protein BLAST search for a suitable template structure related to the target sequence (*wb*GST) showed porcine π-class GST 2gsr-A chain with 42% sequence similarity as the most suitable template for modeling. ClustalW pairwise alignment results of *wb*GST and 2gsr-A sequences showed 42% sequence similarity and 0.005% gap frequency with the porcine π-class GST (2gsr-A) sequence, as shown in Fig. 1.

In total, three models were generated by MODELLER6v2 with the following objective functions 888.5681, 971.9692 and 1040.0116. The model with minimum objective function (888.5681) was selected as the best among the three model structures and subjected to

Table 1 Hex 4.2 docking results of GST inhibitors

S.No	Inhibitors	Structure	Ligand binding site	E-min (KJ mol^{-1})
1	Ethacrynic acid		116	-161.8
2	Curcumine		116	-158.2
3	Plumbagin		116	-116.1
4	Terrapin199		116	-175.4
5	GSH analogue		116	-196.5
6	CDNB		106	-123.74

"Internal" evaluation of self-consistency checks such as a stereochemical check on the modeled structure to find the deviations from normal bond lengths, dihedrals and nonbonded atom-atom distances. The PROCHECK Ramachandran plot analysis for the final model is given in Fig. 2. The assessment of the overall stereochemical quality of the protein by Ramachandran plot to find the stable conformations of N–C_α and C_α–C main chain bonds rotations in the proteins by torsion angles ϕ and ψ indicated that there are four different regions and the two most favored ones (red and deep yellow) are the "core" and "allowed" regions in which 10°×10° pixels having more than 108 residues and no steric clashes; the "generously allowed" regions (light yellow) are extended out by 20° (two pixels); and "disallowed" regions (white) for all amino acids except glycine, which is unique in that it lacks a side chain. As shown in the Ramachandran plot, 93.5% of the residues are in the core region, 5.4% in the allowed and 1.1% in the generously allowed regions. No residue was found in the disallowed region of the Ramachandran plot showing that no residue has a disallowed conformation.

Fig. 1 ClustalW pairwise alignment of *wb*GST and 2gsrA sequences using user-defined parameters of gap open and extension penalties with BLOSUM62 matrix. The highly aligned regions (*) are showed in blue color

```
wbGST     MSYKLTYFPIRGLAEPIRLVLVDQGIKFTDDRINASDWPSMKSHFHFGQLPCLYDGDHQI
2gsrA     PPYTIITYFPVRGRCEAMRMLLADQDQSWKEEVVTMETWPPLKPSCLFRQLPKFQDGDLTL
           .*.:****:** .*.:*::*.**. .:.:: :. . **.:*.   * *** : ***  :

wbGST     VQSGAILRHLARKHNLNGGNELETTHIDMFCEGIRDLHTKYAKMIYQAYDTEKDSYIKDI
2gsrA     YQSNAILRHLGRSFGLYGKDQKEAALVDMVNDGVEDLRCKYATLIYTNYEAGKEKYVK-E
           **.******.*...* * :: *:: :**. :*:.**: ***.:**  *:: *:.*:*

wbGST     LPVELAKFEKLLATRDDGKNFILGEKISYVDFVLFEELDIHQILDPHCLDKFPLLKAYHQ
2gsrA     LPEHLKPFETLLSQNQGGQAFVVGSQISFADYNLLDLLRIHQVLNPSCLDAFPLLSAYVA
          ** .*  **.**: .:.*: *::*.:**:.*: *:: * ***:*:* *** ****.**

wbGST     RMEDRPGLKEYCKQRNRAKIPVNGNGKQ
2gsrA     RLSARPKIKAFLASPEHVNRPINGNGKQ
          *:. ** :* :  . ::.: *:******
```

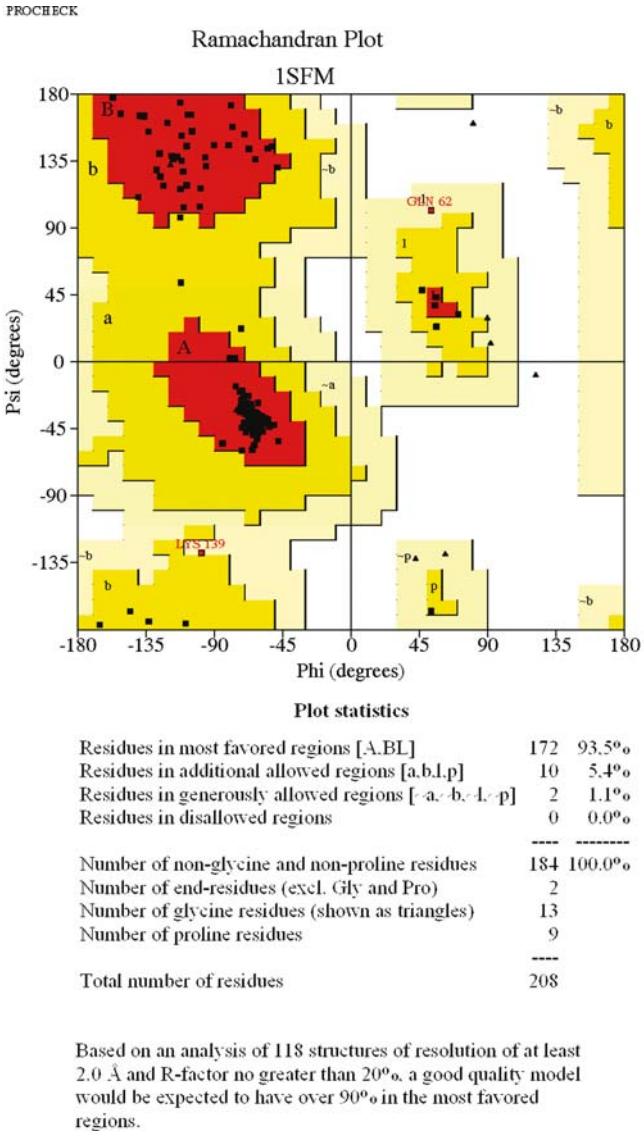


Fig. 2 Ramachandran plot of ϕ/ψ distribution of *wb*GST model 1SFM produced by PROCHECK

The PROCHECK results summary showed four labeled residues in all Ramachandrans out of 206 while the torsion angles of the side chain designated by χ_1-χ_2 plots showed only two labeled residues out of 151. All main-chain and side-chain parameters were found to be in the better region. Goodness factors (G-factors) show the quality of covalent and overall bond/angle distances. These scores should be above −0.50 for a reliable model. The observed G-factors for the present model were 0.10 for dihedrals, −0.15 for covalent and overall 0.01. The distribution of the main-chain bond lengths and bond angles were 99.3 and 99.9% within limits, respectively. RMS distances from planarity do not indicate any large deviations. All covalent bonds lie within a 6.0RMSD range about the standard dictionary value. In the case of covalent bond angles, except for residue ALA at 108, which was found to have a covalent bond angle 135.1° with a slightly higher RMSD of 13.4 against the allowed value of 13.2, all others were within the standard values. No close contacts (distances smaller than 2.2°A for heavy atoms and 1.6°A for hydrogen) were observed.

The PROSA II energy plots for the refined model 1SFM with its template are given in Fig. 3. It can be seen that in the modeled structure 1SFM except for a few residues showing slightly positive interaction energies, all other residues show negative interaction energies. The template structure, which is experimentally determined, shows negative interaction energies for all residues. The *z*-scores of pair, surface and combined energy were −7.47, −5.00 and −5.33 for *wb*GST while for the template 2gsrA the values were −10.03, −7.68 and −6.28, respectively. The comparable *z*-score values and interaction energies further confirm the quality of the modeled structure.

The evaluated final model was deposited in Protein data bank (PDB) with PDB ID: 1SFM and RCSB ID: RCSB021668 under the theoretical models category. The modeled 3D structure 1SFM of *wb*GST is given in Fig. 4 as a Hex view. The overall topology of the *wb*GST model consists of four-stranded β-sheet and eight α-helices arranged in a smaller N-terminal domain I (α_1-α_3, β_1-β_4) and a larger C-terminal domain II (α_4-α_8). The N-terminal domain I is an α/β structure built up of the four-stranded β-sheet and three α-helices showing a $\beta\alpha\beta\alpha\beta\beta\alpha$ folding pattern. Domain II is composed of five α-helices with all α-helical fold.

The results of docking of the 1SFM molecule with different GST inhibitors using Hex 4.2 are given in Table 1. The total number of solutions observed in Docking 1SFM as receptor and the known GST inhibitors ethacrynic acid, curcumine, plumbagin, GSH analogue, Terrapin-199 and the substrate CDNB as ligands were 53, 108, 25, 113, 134 and 53, respectively.

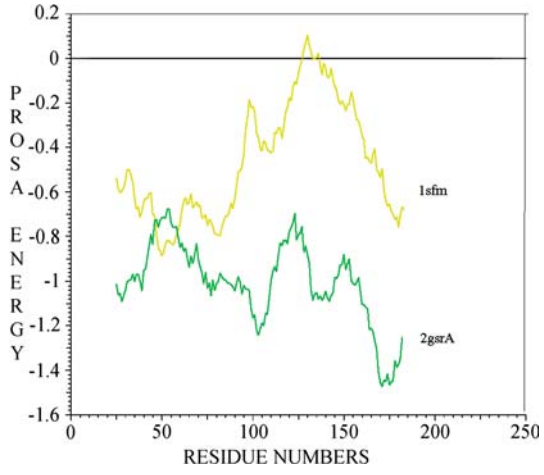


Fig. 3 PROSA energy profiles calculated for the refined model 1SFM and its template 2gsr A. In *x*-axis: Residue numbers, *y*-axis: Prosa energy. *Yellow lines*-Target 1SFM (*wb*GST); *Green lines*-Template 2gsrA

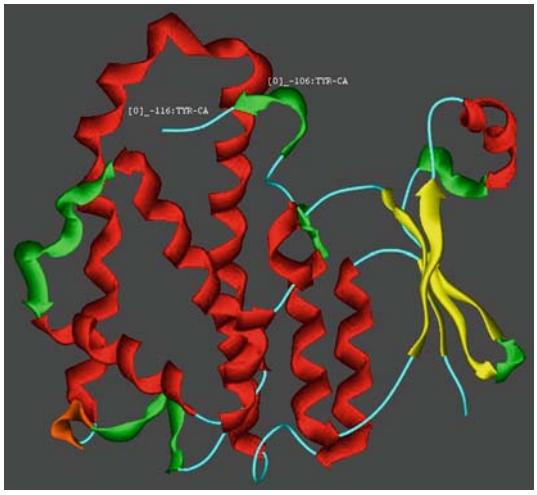


Fig. 4 Computationally modeled 3D Structure of *wb*GST submitted to Brookhaven Protein data bank (PDB ID: 1SFM) under theoretical models category (αHelices are Red, βstrands are Yellow, coils are Blue, Turns are Green and 310 Helices are Brown. The ribbon structure shown in the representation is HEX4.2 view)

From the results, the primary ligand-binding site was found to be tyrosine 116 for all five inhibitors, whereas the substrate CDNB was bound at tyrosine 106. The binding energies varied from −116.1 to −196.5 for the compounds studied, corresponding to their first docking solutions. Among the compounds studied, the GSH analogue showed maximum affinity towards *wb*GST as it exhibited the lowest binding energy. Figure 5 illustrates the key-binding site (tyrosine 116) for the GSH analogue in *wb*GST. Here, Fig. 5 is rotated through a horizontal axis by 180° relative to the view in Fig. 4 to get a clear view of the ligand binding otherwise the ligand binding would appear at the top and distant side of the Fig. 4.

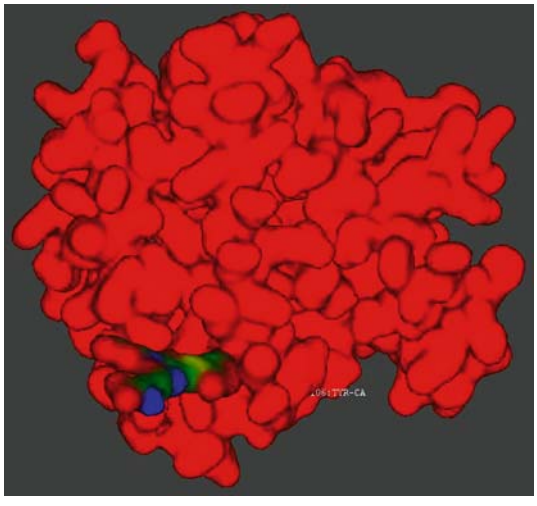


Fig. 5 Primary binding site of GSH analogue in 1SFM obtained by Hex 4.2 docking

Discussion

The lack of an experimentally determined structure of the target protein frequently limits the application of structure-based drug design methods. Homology or comparative modeling uses experimentally determined protein structures as templates to predict the conformation of other proteins with similar amino-acid sequence [14, 22]. In the absence of a crystal structure of the antifilarial target enzyme *wb*GST, an attempt has been made to model its 3D-structure, refine and validate by internal [23, 25] and external [26] evaluation methods for correctness of the model. Application of comparative modeling in cases of high sequence identities (> 35%) usually results in accurate models [17]. The quality of a template increases with its overall sequence similarity to the target and decreases with the number and length of gaps in the alignment.

Out of three models of *wb*GST built by MODELLER, the one with the minimum value for the objective function was selected as the best model for which a Ramachandran plot analysis showed 93.5% of the residues in core region followed by 5.4 and 1.1% residues in the allowed and generously allowed regions, respectively. None of the non-glycine residues are in disallowed regions. Ideally, 90% of the residues in the "core" regions of a Ramachandran plot are one of the better guides to stereochemical quality of the protein structure [26].

Besides the evaluation of the stereochemical quality for the model, PROSA II [27] was used to evaluate the model further and found that the energy profile and the *z*-scores are comparable to the template and the model is in the reasonably folded region. In the energy graph, positive values point to strained sections of the chain, whereas negative values correspond to stable parts of the molecule. In the present model, most of the values are in the negative region, indicating the folding quality of the model.

The GSTs have been established as good targets for antischistosomal [2] and antimalarial [3] drug development. They appear to be the major detoxification system in helminthes [7]. The highly conserved residues Y 7, P 51, D 57, A 65, I 66, G144, and D 151, which have been found in all known GST sequences, are also present in the filarial GSTs from *Onchocerca volvulus* (causative agent for river blindness), *W. bancrofti* and *B. malayi* [11]. These three sequences are topologically related to the π-class GST [11]. The conserved residues of the active site are responsible for common chemistry across the super family. Among the seven conserved residues all residues except G144 and D151 are in domain I. Domain I represents a typical GSH-binding domain of the $\beta\alpha\varpi$ $\beta\alpha\beta\beta\alpha$ folding pattern [30]. One important variation from the typical πGSTs was noted at the hydrophobic substrate-binding site of filarial GSTs that it possesses an open hydrophobic substrate-binding cleft [4]. However, the glutathione-binding site is closely related to

that found in mammalian enzymes. Exploitation of this significant divergence from the host GSTs by selective inhibition of parasite-derived GSTs makes this enzyme a realizable antifilarial target for drug development [4].

The docking results of the present study showed that in 1SFM the site near the tyrosine residue at 116 of *wb*GST was found to be the primary site of ligand binding for ethacrynic acid, curcumine, plumbagin and Terrapin-199, while the substrate CDNB bound near the tyrosine residue at 106. The observed differences in binding sites may be attributed to the variation in the chemical nature of the molecules/mechanism by which they undergo chemical reactions. Both these residues are located at the top of the α4-helix in domain II of *wb*GST. In the case of CDNB, the tyrosine (residue 106) assisted binding by a nucleophilic substitution reaction and for other molecules a Michael type of addition assisted by another tyrosine (residue 116) might have taken place. This is in agreement with the earlier findings [4] that tyrosine 116 is a potential target for binding a number of π-GST inhibitors. Interestingly, recent reports [4, 6, 12] showed that β-carbonyl substituted glutathione conjugates and plumbagin, which are GST inhibitors, exhibit antifilarial activity. The glutathione analogue [4], which was reported to have a 10-fold selective inhibition of the OvGST2 from *O. volvulus* compared to human π-GST, has been found to bind to the GST of *W. bancrofti*, further emphasizing the importance of this study. The results of the present work, the modeled 3D-structure of *wb*GST, 1SFM may play a useful role in the rational design and development of specific inhibitors, leading to molecules that kill the adult filarial worm.

Acknowledgments The authors are grateful to Dr. P.K. Das, Director, Vector Control Research Centre for providing the facility and his encouragement during the study.

References

1. WHO/CDS/CPE/CEE/2002.28. Global programme to eliminate lymphatic filariasis
2. Michele AM, DeWight RW, John AT (1995) J Mol Biol 246:21–27
3. Harwaldt P, Rahlfs S, Becker K (2002) Biol Chem 383:821–830
4. Brophy PM, Campbell AM, van Eldik AJ, Teesdale-Spittle PH, Liebau E, Wang MF (2000) Bioorg Med Chem Lett 10:979–981
5. Rao UR, Salinas G, Mehta K, Klei TR (2000) Parasitol Res 86:908–915
6. Mathew N, Paily KP, Vanamail P, Kalyanasundaram AM, Balaraman K (2002) Drug Dev Res 56:33–39
7. Brophy PM, Pritchard DI (1994) Exp Parasitol 79:89–96
8. Precious WY, Barrett J (1989) J Parasitol Today 5:156–160
9. Liebau E, Wildenburg G, Brophy PM, Walter RD, Henkle-Duhrsen K (1996) Mol Biochem Parasitol 80:27–39
10. Wildenburg G, Liebau E, Henkle-Duhrsen K (1998) Exp Parasitol 88:34–42
11. Rathaur S, Fischer P, Domagalsky M, Walter RD, Liebau E (2003) Exp Parasitol 103:177–181
12. Campbell AM, van Eldik AJ, Liebau E, Barrett J, Brophy PM, Teesdale-Spittle PH, Wang MF (2001) Chem-Biol Interact 133:240–243
13. Johnson MS, Srinivasan N, Sowdhamini R, Blundell TL (1994) Crit Rev Biochem Mol Biol 29:1–68
14. Sali A (1995) Curr Opin Biotechnol 6:437–451
15. Rost B, Sander C (1996) Annu Rev Biophys Biomol Struct 25:113–136
16. Chothia C, Lest AM (1986) EMBO J 5:823–826
17. Sanchez R, Sali A (1997) Curr Opin Struct Biol 7:206–214
18. Altschul SF, Thomas LM, Alejandro A, Schäffer JZ, Zheng Z, Webb M, David JL (1997) Nucleic Acids Res 25:3389–3402
19. Dirr H, Reinemer P, Huber R (1994) J Mol Biol 243:72–92
20. Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson T (1994) J Nucleic Acids Res 22:4673–4680
21. Sali A, Blundell TL (1993) J Mol Biol 234:779–815
22. Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M (1995) Proteins 23:318–326
23. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Crystallogr 26:283–291
24. Hooft RWW, Vriend G, Sander C, Abola EE (1996) Nature 381:272
25. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) J Mol Biol 7:95–99
26. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Proteins 12:345–364
27. Sippl MJ (1993) Proteins 17:355–362
28. Guex N, Peitsch MC (1997) Electrophoresis 18:2714–2723
29. Ritchie DW, Kemp GJL (2000) Proteins: Struct Funct Genet 39:178–194
30. Gilliland GL (1993) Curr Opin Struct Biol 3:875–884