J Mol Model (2013) 19:5213–5223
DOI 10.1007/s00894-013-2009-3

ORIGINAL PAPER

Highlighting a π–π interaction: a protein modeling and molecular dynamics simulation study on *Anopheles gambiae* glutathione S-transferase 1-2

Yan Wang · Qing-Chuan Zheng · Ji-Long Zhang · Ying-Lu Cui · Qiao Xue · Hong-Xing Zhang

Received: 12 July 2013 / Accepted: 8 September 2013 / Published online: 12 October 2013

Abstract Cytosolic insect theta class glutathione S-transferases (GSTs) have not been studied completely and their physiological roles are unknown. A detailed understanding of *Anopheles gambiae* GST (Aggst1-2) requires an accurate structure, which has not yet been determined. A high quality model structure of Aggst1-2 was constructed using homology modeling and the ligand–protein complex was obtained by the docking method. Molecular dynamics (MD) simulations were carried out to study conformational changes and to calculate binding free energy. The results of MD simulation indicate that Aggst1-2 undergoes small conformational changes after ligands dock to the protein, which facilitate the catalytic reaction. An essential hydrogen bond was found between the sulfur atom of glutathione (GSH) and the hydrogen atom of hydroxyl group in Ser9, which was in good agreement with experimental data. A π–π interaction between Phe204 and CDNB ligand was also found. This interaction seems to be important in stabilization of the ligand. Further study of binding free energy decomposition revealed a van der Waals interaction between two ligands that may play a key role in nucleophilic addition reaction. This work will be a good starting point for further determination of the biological role of cytosolic insect theta class GSTs and will aid the design of structure-based inhibitors.

Keywords Glutathione S-transferase · Homology modeling · Molecular docking · Molecular dynamics simulation · MM-GB/SA calculation · π–π interaction

Y. Wang · Q.-C. Zheng (✉) · J.-L. Zhang · Y.-L. Cui · Q. Xue · H.-X. Zhang (✉)
State Key Laboratory of Theoretical and Computational Chemistry, Institute of Theoretical Chemistry, Jilin University, Changchun 130023, Jilin, People's Republic of China
e-mail: zhengqc@jlu.edu.cn
e-mail: zhanghx@jlu.edu.cn

Introduction

Glutathione S-transferases (E.C.2.5.1.18; GSTs), which catalyze the nucleophilic addition of the tripeptide glutathione (γ-L-glutamyl-L-cysteinyl-glycine, GSH) to substrates with electrophilic groups, are found in various species [1]. Living organisms come into contact with a variety of foreign chemicals or xenobiotics through exposure to environmental contaminants. These xenobiotics can interact deleteriously with living organisms, causing toxic and sometimes carcinogenic effects [2]. Thus, all living organisms must have the ability to rapidly eliminate xenobiotics that would otherwise cause them harm. The detoxification of xenobiotics by enzymes has been classified into three distinct phases. Phases I and II involve the conversion of a lipophilic, non-polar xenobiotic into a more water-soluble and therefore less toxic metabolite. These metabolites can be eliminated more easily from the cell through Phase III detoxification. GSTs are major Phase II detoxification enzymes found mainly in the cytosol [3]. These enzymes catalyze the conjugation of active xenobiotics to an endogenous water-soluble substrate, such as GSH, UDP-glucuronic acid or glycine. Conjugation of electrophilic compounds with the thiol group of GSH is catalyzed mainly by GSTs; products are more excretable than non-GSH conjugated substrates following this process [4]. Detailed study of a protein's structure is of great importance in understanding its actual function. The crystal structure of GSTs was solved for the first time in 1993 by Xinhua Ji and coworkers [5, 6]. To date, crystal structures of more than 300 GSTs from different species have been determined (http://www.rcsb.org/pdb/home/home.do). Of these, about 50 structures belong to insect GSTs. Based on their biochemical, immunological, and structural properties, cytosolic insect GSTs have been grouped into six major classes: Delta, Epsilon, Omega, Sigma, Theta and Zeta.

The detoxification properties of GSTs mean that they can metabolize several major classes of insecticides. Elevated GST activity is thus an important mechanism by which insects develop resistance to insecticides. So far only a few genes of theta GSTs have been identified in mosquitoes (two in *Anopheles gambiae*, and five putative in *Aedes aegypti*). These putative insect theta GSTs have not been characterized biochemically and their physiological role is unknown [7]. *Anopheles gambiae* GST 1-2 (Aggst1-2), which consists of 209 residues, belongs to cytosolic insect GST class theta. Although the physiological role of insect theta class GST remains uncertain, Aggst1-2 may participate in the detoxification of insecticides, and insecticide resistance may also be related to this enzyme. The three-dimensional (3D) structure of Aggst1-2 has not yet been determined. In order to gain a more detailed understanding of Aggst1-2, it is necessary to build an accurate 3D structure of the protein to explore its structure–function relationship.

The chemical structures of GSH and 1-chloro-2,4-dinitrobenzene (CDNB) are shown in Fig. 1. In the present study, a high quality model of Aggst1-2 was constructed by homology modeling and molecular dynamics (MD) simulations. The model structure was then used to search the active site and carry out binding studies with GSH and CDNB. MD and molecular mechanics generalized born surface area (MM-GB/SA) theory, which have proved to be useful and valuable tools [8–11], were employed to discover the interaction between protein and ligands. The modeling and simulation results illustrated close agreement between theoretical and experimental results, and may be helpful for further investigation of the mechanism of action of Aggst1-2.

Theoretical methods

Homology modeling and molecular docking were performed on a HP Z600 workstation using the Discovery Studio 2.5 software package [12]. Molecular dynamic (MD) simulations and MM-GB/SA calculations were carried out on the Inspur TS10000 serve with MD program AMBER11 [13]. The ff03ua force field [14] in AMBER11 was chosen for MD simulations. The MM-GB/SA method [15, 16] was used to calculate the binding free energies of the CDNB-GSH-Aggst1-2 complex.

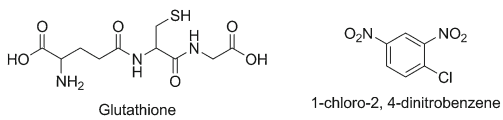

Fig. 1 The chemical structure of glutathione (GSH) and 1-chloro-2,4-dinitrobenzene (CDNB). The two atoms used in this study are defined thus: S1 refers to the sulfur atom in GSH, and C1 refers to the carbon atom in CDNB that is connected to the chlorine atom

Homology search

The primary sequence of Aggst1-2 (accession no.Q94999) was obtained from NCBI database (http://www.ncbi.nlm.nih.gov). This Aggst1-2 query sequence was then used to search against Protein Data Bank database for the related protein structure using the BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

3D model building

Initially, sequence alignment was carried out on Discovery Studio 2.5 using the Align Multiple Sequence protocol. The 3D-structure of Aggst1-2 was then constructed based on the crystal structure of *Anopheles dirus* species B GSTs 1–3, chain A (PDB code: 1JLV) using the Building Homology Models module.

Energy minimization (EM) was performed using the Minimization module in Discovery Studio 2.5 in order to improve the initial model; 400 steps of steepest descent (SD) minimization and 600 steps conjugate gradient (CG) minimization were carried out, which further improved the quality of the initial model.

After optimization calculations, the structure was checked using Profile-3D [17] and Ramachandran plot analysis [18].

Binding site analysis

The Define and Edit Binding Site module is a suite of programs in Discovery Studio 2.5 for calculating, editing, partition and displaying the binding sites of a receptor. There are two site-finding routines that can be used to automatically locate binding sites. One identifies cavities within the receptor, while the other builds a binding site based on a ligand molecule in a known location. When the search was completed, the largest site is displayed on the structure. These results can be used to guide protein–ligand docking.

Docking ligands to Aggst1-2

Docking GSH to Aggst1-2

Amino acid residues that form G-sites (glutathione-binding sites) appear to be highly conserved across different GST classes [19–22]. A superimposition method was used here for docking GSH to Aggst1-2. Based on the structure of the GST complex with GSH—the template structure mentioned above (PDB code: 1JLV, *Anopheles dirus* species B GSTs 1–3)—a structural alignment was used to hard dock GSH to the G-site of Aggst1-2. Thus, the GSH-Aggst1-2 complex was obtained and then used as the starting structure for the subsequent docking study.

Docking of CDNB to the GSH-Aggst1-2 complex

The initial structure of CDNB was generated by Discovery Studio 2.5 software. By means of the B3LYP density functional

method using the Gaussian09 [23] package of programs, geometry optimizations for CDNB were performed using the 6-31G(d) basis set. CDOCKER [24] was then used to dock CDNB into the H-site of the GSH-Aggst1-2 complex. CDOCKER is a grid-based molecular docking method that employs CHARMm. The receptor, i.e., the GSH-Aggst1-2 complex, was held rigid, while the CDNB ligand was allowed to flex during the refinement. As mentioned above, the ligand binding site was located automatically using the Find Sites from Receptor Cavities module of Discovery Studio 2.5. Thus, knowledge of the binding site was acquired. It was possible to specify the ligand placement in the active site using a binding site sphere with a radius of 9.0 Å. Random conformations of the CDNB ligand were generated from the initial ligand structure through high temperature MD simulation at 1,000 K, followed by random rotations, respectively. The random conformations were refined by grid-based simulated annealing and a final full force field minimization. The top ten poses were saved for comparison and analysis. Finally, the pose with the lowest CDOCKER interaction energy was used for the subsequent MD simulation experiments.

Molecular dynamics simulations

The initial structures of the Aggst1-2 and CDNB-GSH-Aggst1-2 complexes in this study were obtained in a previous homology modeling and docking study. The MD simulations, including EM, were performed using the AMBER11 software package [13]. To keep the whole system neutral, a certain number of sodium ions (Na^+) were added using the tleap module in AMBER based on a coulomb potential grid. The initial model of GSH was extracted from the 1JLV_A crystal structure and the CDNB model was constructed by Discovery Studio2.5. The PRODRG program [25] was used to convert these two ligands to standard PDB format. Structural optimization of CDNB and GSH were conducted at the B3LYP/6-31G(d) level using Gaussian09 [23] and the restrained electrostatic potential (RESP) fitting procedure was used for charge derivation based on the optimal conformation. The force field parameters of CDNB and GSH were supplied by general AMBER force field (GAFF) [26] in the Antechamber module [27] of AMBER11. The two systems, i.e., the Aggst1-2 and the CDNB-GSH-Aggst1-2 complex system, were solvated with TIP3P water model [28] in a truncated octahedron box with 9.0 Å distance around the solute. During the minimizations and MD simulations, the particle mesh Ewald summation method [29] was applied to treat long-range electrostatic interactions with a periodic boundary condition. All bonds involving hydrogen atoms were restricted by the SHAKE algorithm. The integration step-time of MD simulation was 2 fs. The protein and small molecules were held fixed with a 500 kcal mol^{-1} $Å^{-2}$ constraint, and both solvent and ions were minimized for 5,000 steps of SD method followed by a further 5,000 steps of CG method for each system. After that, each system was totally minimized for another 20,000 steps with no restraint—first 8,000 steps using the SD method, and then 12,000 steps using the CG method. After minimization, the two systems were heated gradually from 0 to 300 K for 500 ps in the canonical ensemble (NVT ensemble), applying harmonic restraints with a force constant of 10.0 kcal mol^{-1} $Å^{-2}$ on the protein and small molecules. A Langevin thermostat was adopted. Subsequently, the two systems were equilibrated in a NPT ensemble under constant pressure (1.0 bar) for 1 ns. The relaxation time for barostat bath was set to 2.0 ps. Then, 20 ns MD simulations were performed for both the Aggst1-2 system and the CDNB-GSH-Aggst1-2 complex system. The 20 ns simulations' trajectory was analyzed using AmberTools1.5 [30] in the Amber11 software package. Hydrogen bonds were determined via the distance between the heavy atoms using a cutoff of 3.0 Å and the angle between the acceptor, and donor atoms using a cutoff of 135°. Pi-stacking interactions were determined following the methodology of McGaughey et al. [31]. This method finds stacked and staggered π–π interactions by performing the following tests:

(1) The distance between the centroid of each pair of aromatic rings was determined to find those that fall within the center distance cutoff (default 8.0 Å);
(2) For these, an atom from each ring should be within the closest atom distance cutoff (4.5 Å);
(3) The angle θ between the normal of one ring and the centroid-centroid vector must fall between 0° and ± the theta angle cutoff (default 60°), and the angle gamma between the normal to each ring must fall between 0° and ± the gamma angle cutoff (default 30°).

VMD [32] Chimera [33] and PyMOL [34] software were used to visualize the trajectories and to depict structural representations.

MM-GB/SA calculations

The binding free energy of the docking complex was calculated using the MM-GB/SA method [8, 35]. The MM-GB/SA method calculates binding free energy between substrates and enzymes. The binding free energy (ΔG_{bind}) in MM-GB/SA between a ligand (L) and a receptor (R) to form a complex RL was calculated as:

$$\Delta G_{bind} = G_{complex} - \left(G_{receptor} + G_{ligand}\right) \quad (1)$$

$$G = E_{MM} + G_{sol} - TS \quad (2)$$

$$E_{MM} = E_{int} + E_{ele} + E_{vdw} \quad (3)$$

$$E_{MM} = E_{int} + E_{ele} + E_{vdw} \quad (4)$$

In Eq. (2), E_{MM}, G_{sol} and TS represent molecular mechanics components in gas phase, the stabilization energy due to solvation, and a vibrational entropy term, respectively. E_{MM} is

given as the sum of E_{int}, E_{ele} and E_{vdw}, which are internal, Coulomb and van der Waals interaction terms, respectively. Solvation energy G_{sol} is separated into an electrostatic solvation free energy (G_{GB}) and a nonpolar solvation free energy (G_{SA}). The former solvation free energy was calculated using the generalized Born model [36]. The latter solvation free energy (G_{SA}) was calculated by:

$$G_{SA} = \gamma SASA + \beta \quad (5)$$

The solvent accessible surface area (SASA) in Eq. (5) was evaluated using the ICOSA model in AMBER11. In this study, the values for γ and β were set to 0.0072 kcal mol^{-1} Å^{-2} and 0 kcal mol^{-1}, respectively.

The normal mode analysis was performed to estimate the change of conformational entropy upon ligand binding (−TS) using the nmode module of AMBER11. For the docking complex system, 500 snapshots were extracted from the last 10 ns MD trajectory at intervals of 10 ps. Because normal mode analysis is computationally expensive, only 100 snapshots of the 500 snapshots were applied for the entropy calculation.

Energy decompositions were performed. Here, only per-residue decomposition was included. The per-residue decomposition was to separate the energy contribution of each residue from the association of receptor with the ligand into three terms: van der Waals contribution (ΔE_{vdw}), electrostatic contribution (ΔE_{ele}), and solvation contribution (ΔG_{gb}+ΔG_{surf}). By analyzing the free energy decompositions, the important interaction within the docking complex can be found.

Results and discussion

Homology modeling of Aggst1-2

According to the BLAST search, the sequences of the three proteins share relative high identity (Table 1). 1JLV_A was chosen as the final template structure for several reasons. When a homology modeling method is used, a high resolution of the template structure is necessary in order to get a high quality model. 1JLV_A has the highest resolution (1.75 Å) among these three proteins. As mentioned before, GSH will hard dock into the G-site using the superimposing method due to the conserved G-site in all GSTs. Thus, a crystal structure of the GST-GSH complex is crucial for further docking study. 1JLV_A is the only structure bound to GSH. Although the sequence identity between Aggst1-2 and 1JLV_A was not the highest compared to the other two proteins (55 % for 1JLW_A, 53 % for 1PN9_A), an identity of 51 % is quite enough for further straightforward sequence alignment, as shown in Fig. 2. Therefore 1JLV_A was chosen as the template structure. Automated homology model building was performed using the build homology model module in the Discovery Studio 2.5 software package with its default parameters. The best-ranked model was chosen based on probability density functions. This model, made up of 208 residues, was refined by EM. The final structure of Aggst1-2 is shown in Fig. 3a.

Table 1 Numbers of residues, sequence identity and resolution of three proteins intended to be used as template structures

	Numbers of residues	Sequence identity (%)	Resolution (Å)
1JLV_A	207	51	1.75
1JLW_A	217	55	2.45
1PN9_A	209	53	2.00

As can be seen in Fig. 3a, this enzyme contains seven helices and four sheets. The N-terminal domain consists of four beta sheets and three flanking alpha helices. This domain adopts a conformation similar to the thioredoxin domain that is found in many proteins [3]. The larger C-terminal is composed of seven alpha helices. There are two distinct binding sites in each subunit: one is the G-site, the other is a variable hydrophobic substrate binding site (H-site). The G-site is the binding site of endogenous GSH. This binding site is composed mainly of N-terminal residues, including active residue that activate the sulfydryl group of GSH to generate catalytically active thiolate anion. The H-site, which varies between different classes and subtypes, is found in the C-terminal domain.

The lowest energy conformation after EM optimization was superimposed with 1JLV_A. Their root-mean-square deviation (RMSD) value is 1.29 Å (shown in Fig. 3b), indicating a good overall structure alignment with 1JLV_A. The structure was checked by Profile-3D (Fig. 4) and the verify score for this protein is 82.43, which is higher than the lowest score 42.48 and close to the top score 94.37. Note that verify scores above zero correspond to an acceptable side chain environment. It shows that all residues are reasonable as shown in Fig. 4. A Ramachandran plot (Fig. 5) was used to evaluate the protein structure. The statistical score of the Ramachandran plot shows that 96.64 % are in the most favored region (cyan area), 0.96 % in the semi-favoured region (magenta area) and 2.40 % in the disfavoured region. All these above assessments indicate that the homology model is reasonable and can be used for the subsequent docking study.

Docking ligands to Aggst1-2

In Aggst1-2, the G-site is conserved across all GSTs [19–22] while the H-site contributes to the substrate diversity of the GST enzyme. As shown in Fig. 2, the sequence of G-site residues in both template and target are nearly the same, which

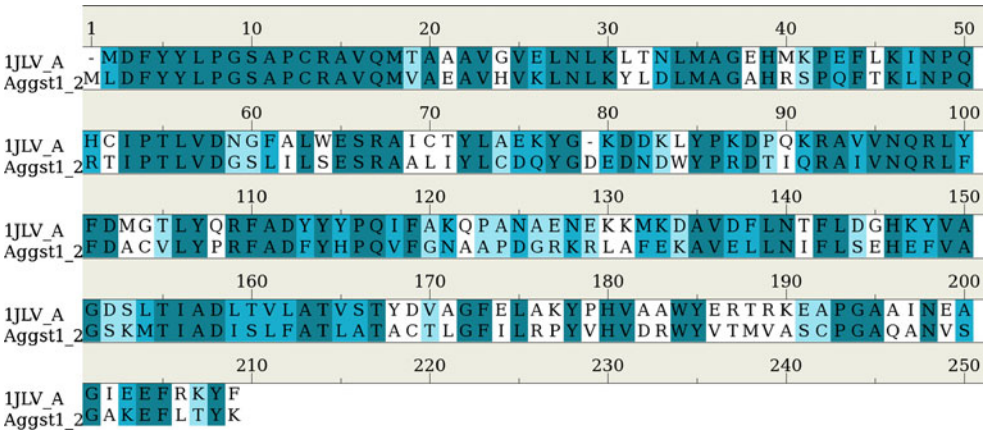

Fig. 2 Sequence alignment between 1JLV_A and Aggst1-2. G-site residues (Leu6, Pro7, Gly8, Ser9, Ala10, Pro11, Glu49, Arg50, Thr51, Ile52, Pro53, Ser65, Arg66, Tyr106 and Tyr207) have a high sequence identity, aligning to the template structure sequence

is one of the reasons we choose 1JLV_A as the template. Leu6, Pro7, Gly8, Ser9, Ala10, Pro11, Glu49, Arg50, Thr51, Ile52, Pro53, Ser65, Arg66, Tyr106, Tyr207 are G-site residues in Aggst1-2. Only Thr53 is not the same as template 1JLV_A (Cys52 in the template). The GSH-Aggst1-2 complex was obtained by the hard docking method. Among the 15G-site residues, positively charged residue Arg50 and Arg66 are essential for stabilizing the GSH binding mode. Another important G-site residue is Ser9. The CDNB ligand was docked into the H-site using the CDOCKER protocol of Discovery Studio2.5. The H-site for docking CDNB is composed of six residues (Tyr106, Phe109, Ala110, Phe118, Phe204 and Tyr207). Hydrophobic residues such as Phe118 and Phe204 are indispensible for the stabilization of CDNB. There may be other interactions between these aromatic residues and CDNB. Ten conformations were obtained for CDNB ligand and the final docking mode of the CDNB-GSH-Aggst1-2 complex with the lowest CDOCKER interaction energy (-21.2961 kcal mol^{-1}) was shown in Fig. 6. Ser9 is reported to be conserved between various species of the Theta class GST [22, 37]. In the crystal structure (1JLV_A), there was a distance of 3.6 Å between the sulfur atom of GSH and the hydrogen atom of the hydroxyl group of Ser9. A hydrogen bond occurred between these two atoms. Ser9 may contribute to the stabilization of the GSH thiolate. As shown in Fig. 6, there is a hydrogen bond (represented by green dashed line) between the two atoms mentioned above. This indicates that the CDNB-GSH-Aggst1-2 complex obtained from docking study is reasonable.

MD simulations

20 ns MD simulations were carried out on the Aggst1-2 and CDNB-GSH-Aggst1-2 complex systems to study conformational change and to calculate binding free energy. The Aggst1-2 system reached equilibrium in the last 5 ns and the CDNB-GSH-Aggst1-2 complex system equilibrized in the last 10 ns, shown in Fig. 7. The average RMSD value for the Aggst1-2 system is 1.75 Å. As shown in Fig. 7a, RMSD values deviate quite largely from the average value. This indicates that the Aggst1-2 structure without ligands (CDNB and GSH) undergoes some conformational changes in certain areas. The average RMSD value of CDNB-GSH-Aggst1-2 complex system is 2.11 Å. Contrasting with the Aggst1-2 system, RMSD values have a small deviation from the average value, which implies that the ligand-Aggst1-2 complex structure is rather stable. The RMSD plots for these two systems are quite different. The structural difference between these two systems may explain the reason for Aggst1-2 system's large deviation from average RMSD values. The main

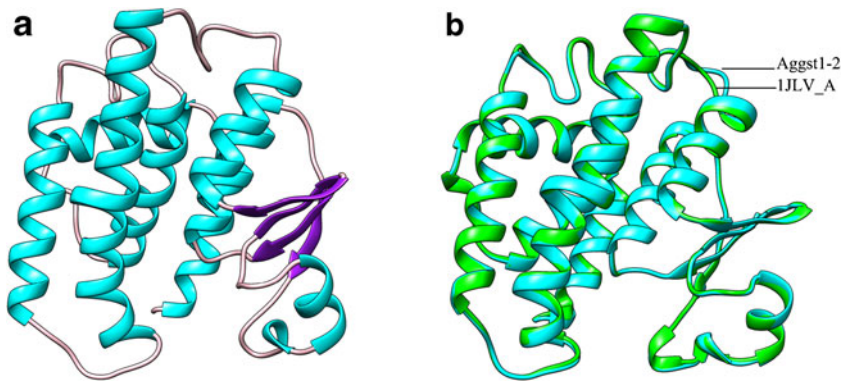

Fig. 3 **a** Final 3D-structure of Aggst1-2. The α-helix, β-sheets and loops are colored *cyan*, *purple* and *pink*, respectively. **b** Structural alignment between the refined Aggst1-2 model and the 1JLV_A template; *cyan* Aggst1-2, *green* template

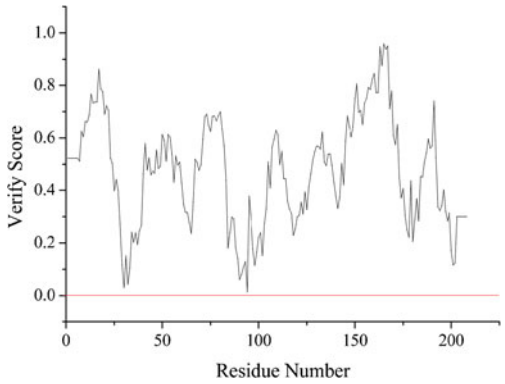

Fig. 4 3D-profile-verified results of the Aggst1-2 model; residues with positive verify score are folded reasonably

distinction between the two systems lies in the ligand docking to the protein structure.

To investigate the effect of ligand docking to the protein structure, the root mean square fluctuation (RMSF) was measured. As shown in Fig. 8, certain residues have higher RMSF values. Here, Arg39 to Arg50 are defined as domain I residues. Domain I residues in the Aggst1-2 system underwent larger conformational changes during simulation, while domain I residues in ligand-Aggst1-2 complex systems showed less flexibility. Since the protein structure is more flexible without ligands than with ligands-Aggst1-2 complex structure, this gives good explanation to large deviation of RMSD values in Aggst1-2 system.

Domain I residues locate on a small α-helix in the N-terminal domain. This small α-helix is adjacent to the G-site and is at the interface of Aggst1-2. When GSH bound to Aggst1-2, Arg39 and Arg50 form hydrogen bonds with GSH, this small α-helix is 'held fixed' through such hydrogen bond interactions. Thus, domain I residues become less flexible.

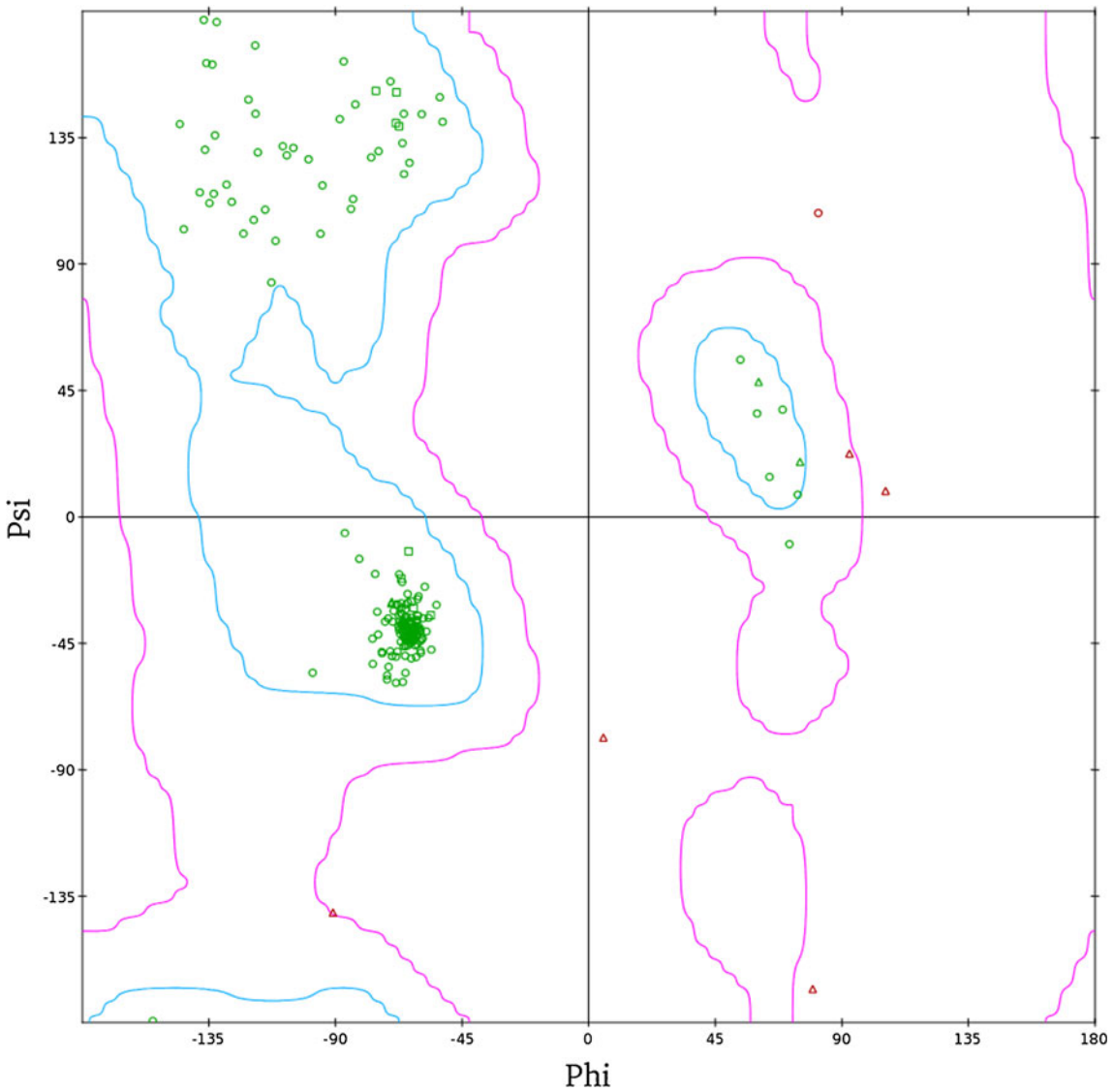

Fig. 5 Ramachandran plot of the Aggst1-2 model from Discovery Studio2.5

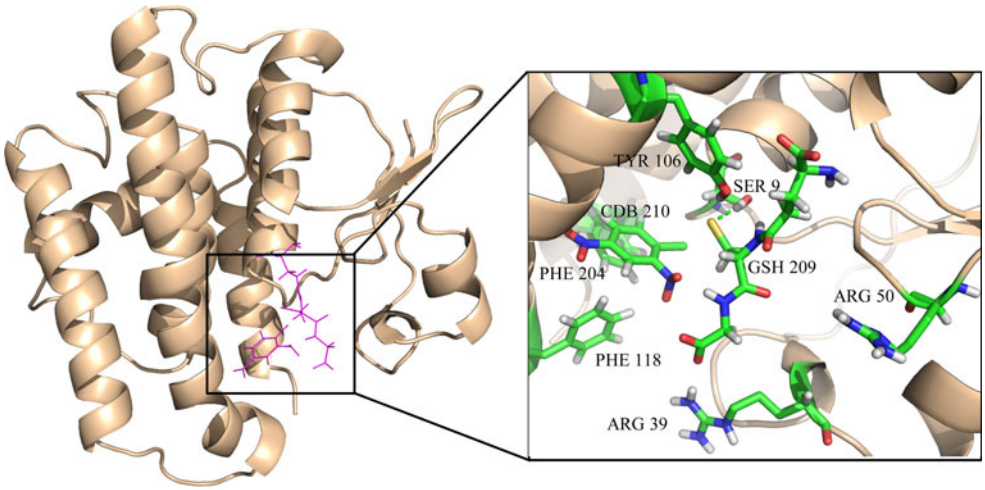

Fig. 6 Final docking pose of the CDNB-GSH-Aggst1-2 complex with the lowest CDOCKER interaction energy and a close-up view of the active site (G-site and H-site) residues. In the overview of the docking pose, the protein structure is represented in wheat cartoon style and the ligands (GSH and CDNB) in *line representation* and in *magenta*. In the close-up view of the active site, the two ligands and key residues are in *stick representation*; *green dashed line* the important hydrogen bond

Asn120 reveals another higher RMSF value as shown in Fig. 8. Its adjacent residue, Phe118, interacts strongly with CDNB. Because of CDNB docking into Aggst1-2 and the interaction between Phe118 and CDNB, Asn120 is less flexible in ligand-Aggst1-2 complex systems. After ligands dock into the active sites, the structure of certain area become less flexible.

Hydrogen bonds are well known to play an indispensable role in the structure and function of biological molecules. The hydrogen bonds between two ligands and their adjacent residues were analyzed and the results are shown in Fig. 9 and Table 2. Board [37] pointed out that a serine residue hydroxyl group was within hydrogen-bonding distance of the glutathione sulfur atom. In the CDNB-GSH-Aggst1-2 complex there is a strong H-bond with 99.95 % occupancy between the hydrogen atom of the hydroxyl group in Ser9 and the sulfur atom from GSH. These results are in good agreement with experimental data [37]. It is worth noting that hydrogen bonds form within GSH and domain I residues (Arg39 and Arg50). These two residues are stabilized by hydrogen bonds and became stable. Arg39 and Arg50 interact with the two carboxylate group of GSH in the form of hydrogen bonds. Thus, these two residues in domain I and GSH stabilize each other. Some of the flexible domain I residues are part of the G-site residues that interact directly with GSH. In the presence of GSH, domain I residues became much less flexible. Thus, we suggest that GSH plays an important role in stabilizing G-site residues.

The result of hydrogen bond analysis showed that CDNB did not form hydrogen bonds with its adjacent residues. As can be seen from Fig. 6, there is cleft between the two domains of Aggst1-2. This area lacks certain interactions such as hydrogen bonds and van der Waals interactions. CDNB is located in this area. But the binding mode of the ligand seemed to be rather stable throughout the entire MD simulation. There may be other interactions between CDNB and its surrounding residues. The centroid distance of two aromatic

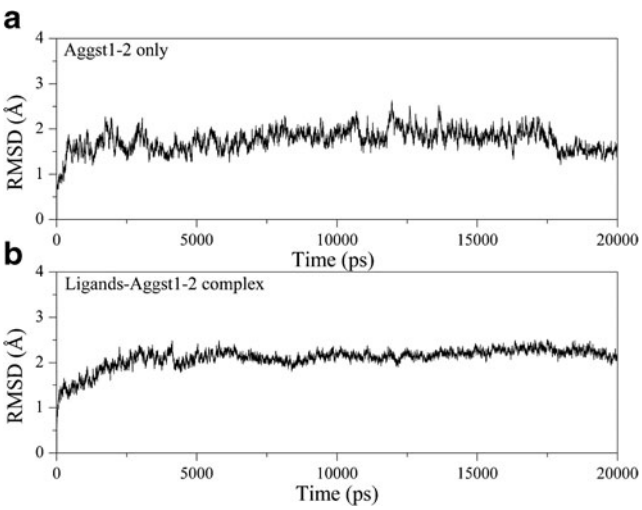

Fig. 7 Calculated root-mean-square deviations (RMSD) of the backbone atoms referenced to the corresponding starting structure. **a** Apo protein Aggst1-2 system. **b** CDNB-GSH-Aggst1-2 complex system

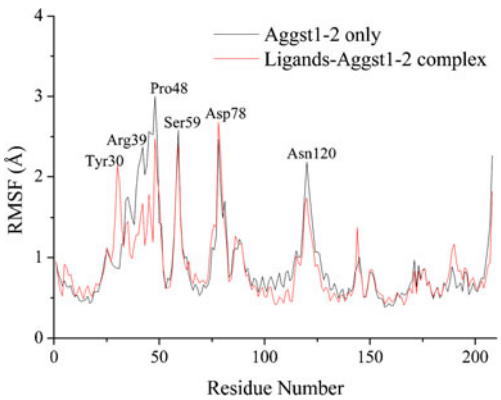

Fig. 8 Root-mean-square fluctuations (RMSF) of the Aggst1-2 backbone observed for the apo protein (*black line*) and the CDNB-GSH-Aggst1-2 complex (*red line*) during the course of 20 ns MD simulations. Flexible residues are labeled: Tyr30, Arg39, Pro48, Ser59, Asp78 and Asn120

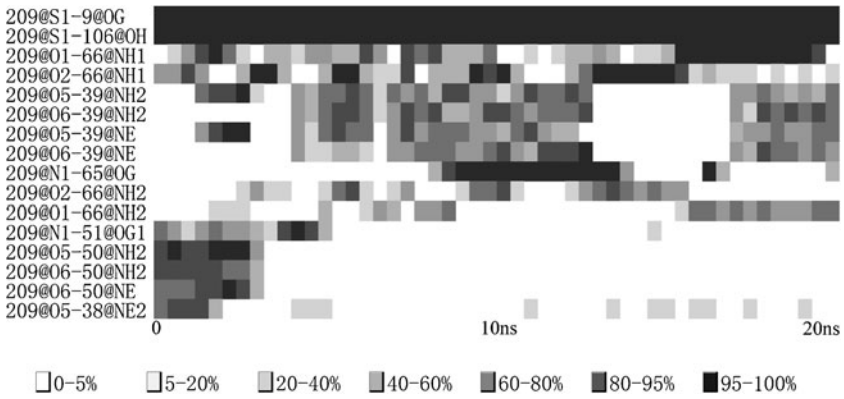

Fig. 9 Percentage occupancy of GSH hydrogen bonds varying with simulation time. Hydrogen bonds shown in Table 2

rings (one aromatic ring is from Phe204 and the other is from CDNB) was monitored and the results are shown in Fig. 10a. Considering an average distance of 4.49 Å, a π–π interaction may occur between these two aromatic rings [31]. In order to detect π–π interaction between these two aromatic rings, theta and gamma angles of each frame during 20 ns simulation time period 10,000 snapshots in total were monitored and the results are shown in Fig. 11. Most theta angle values fell between 0 and 60° and the majority of gamma angle values are in the 0–30° region. Because the average centroid distance is much lower than 8.0 Å, and is a bit lower than 4.5 Å, it is reasonable to use just two of the three conditions, which are the average centroid distance and theta-gamma angles, to determine π–π interactions. Of the 10,000 snapshots, 7,745 are in π–π interaction. Therefore, there is indeed a π–π interaction between these two aromatic rings. When this interaction was absent, the distance between the sulfur atom (S1 atom) of GSH and the carbon atom (C1 atom connected to chlorine atom) of CDNB was quite large. Due to the large distance, the catalytic reaction (nucleophilic addition reaction between these two ligands) cannot take place. In this case, the π–π interaction plays an important role in ligand stabilization. The stabilized ligand conformation may lead to further catalytic reaction.

The nucleophilic addition reaction between the S1 atom of GSH and C1 atom (connected to the chlorine atom) of CDNB would probably occur under if there is a correct orientation and proper distance between GSH and CDNB [38]. This reaction would lead to conjugate formation. The distance between C1 and S1 is defined as $d_{C1–S1}$. This value was monitored over the simulation time and the result is shown in Fig. 10b. A DFT calculation carried out by Zheng and coworkers [39] pointed out the distance between C1 and S1 atom was 2.679 Å in transition state 1. The average value of $d_{C1–S1}$ is 4.63 Å. The

Table 2 Properties of hydrogen bonds between glutathione (GSH) and its adjacent residues, including occupied, distance, angle and lifetime

H-bond		Occupied(%)	Distance(Å)	Angle(°)	Lifetime
1	209@S1-9@OG	99.95	2.924	14.22	1,665.8
2	209@S1-106@OH	99.55	2.992	14.48	221.2
3	209@O1-66@NH1	46.69	2.879	26.62	8.2
4	209@O2-66@NH1	43.40	2.863	26.71	12.7
5	209@O5-39@NH2	41.51	2.986	32.73	6.6
6	209@O6-39@NH2	39.69	2.974	30.79	7.7
7	209@O5-39@NE	38.11	2.942	27.07	7.8
8	209@O6-39@NE	31.71	2.963	29.07	6.6
9	209@N1-65@OG	30.27	2.815	14.16	137.6
10	209@O2-66@NH2	22.60	3.011	37.05	4.8
11	209@O1-66@NH2	19.30	3.062	39.18	3.2
12	209@N1-51@OG	14.48	2.772	21.98	8.4
13	209@O5-50@NH2	14.44	2.917	25.12	21.9
14	209@O6-50@NH2	12.54	3.016	38.96	6.7
15	209@O6-50@NE	12.26	2.973	33.84	7.3
16	209@O5-38@NE2	10.81	2.949	31.03	3.9

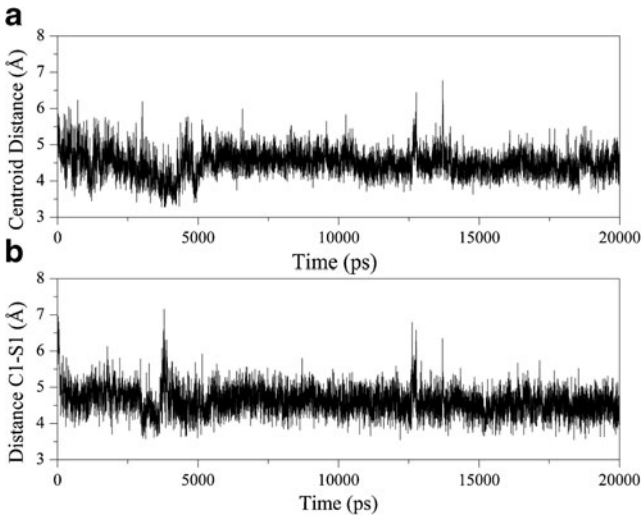

Fig. 10 Two kinds of distance data observed for certain atoms or groups during the course of 20 ns MD simulation. **a** The centroid distance varying with simulation time between two aromatic rings (one aromatic ring from Phe118 and the other from CDNB). **b** The distance between S1 atom of GSH and C1 atom of CDNB

strong interaction between GSH and CDNB, discussed below, indicates that d_{C1-S1} may be in the transition state 1 range to facilitate the nucleophilic addition reaction.

MM-GB/SA calculation and energy decomposition

To explore the interaction between CDNB and protein, MM-GB/SA calculations and energy decomposition were performed using the MMPBSA module of the AMBER11 software package. A total of 500 snapshots were extracted from the last 10 ns trajectory for the binding free energy analyses. The predicted binding affinities of the docking complex system are listed in Table 3.

Table 3 Molecular mechanics generalized born surface area (MM-GB/SA) binding free energy results for 1-chloro-2,4-dinitrobenzene (CDNB)-GSH-Aggst1-2 complex (kcal mol^{-1}). ΔE_{ele} Electrostatic contribution, ΔE_{vdw} van der Waals contribution, ΔG_{np} nonpolar contribution of solvation energy, ΔG_{gb} polar contribution of solvation energy, $-T\Delta S$ the conformational of entropic contribution at temperature of 298.15 K[a]

System	ΔE_{ele}	ΔE_{vdw}	ΔG_{np}	ΔG_{gb}	ΔG_{tol}[a]	$-T\Delta S$	ΔG_{bind}[b]
CDNB	33.33	−29.22	−3.35	−21.25	−20.49	14.23	−6.26

[a] $\Delta G_{tol}=\Delta E_{ele}+\Delta E_{vdw}+\Delta G_{np}+\Delta G_{gb}$

[b] $\Delta G_{bind}=\Delta E_{ele}+\Delta E_{vdw}+\Delta G_{np}+\Delta G_{gb}-T\Delta S$

Overall, as shown in Table 3, the binding free energy of the CDNB to GSH-Aggst1-2 complex system is −6.26 kcal mol^{-1}. It should be noted that van der Waals (ΔE_{vdw}) interactions play a major role in the binding free energy contributions in CDNB-GSH-Aggst1-2 complex system (−29.22 kcal mol^{-1}). The electrostatic energy (ΔE_{ele}) shows the unfavorable contributions (33.33 kcal mol^{-1}). The polar solvation energy (ΔG_{gb}) and the nonpolar solvation energy are contributed to the ligand binding affinity.

The residues that contribute to ligand binding are explored by per-residue binding free energy decomposition analyses. The results are shown in Fig. 12 and Table 4. As illustrated in Fig. 12, residues labeled in the CDNB-GSH-Aggst1-2 complex system have strong interactions with the CDNB ligand, especially the residue Phe204. As mentioned above, the aromatic ring of Phe204 formed a π–π interaction with CDNB. There is indeed a strong interaction (−1.788 kcal mol^{-1}) between them. There is also the interaction energy of −1.038 kcal mol^{-1} between Phe118 and CDNB. As mentioned above, residues Phe118 and Phe204 are critical to making the protein-CDNB complex conformation more stable. It is also noteworthy that GSH has an interaction energy of

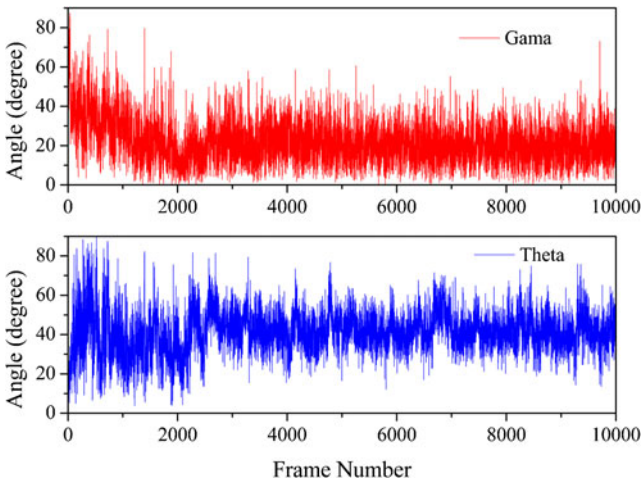

Fig. 11 Theta and gamma angle values observed for CDNB and Phe204 during the course of 20 ns MD simulations. *Blue* Theta angle values, *red line* gamma angle values

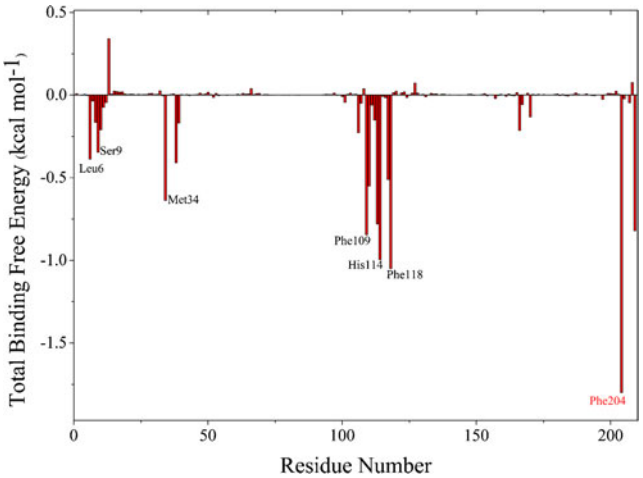

Fig. 12 Binding free energy decomposition between the ligand CDNB and residues of Aggst1-2. Residues important for ligand binding are labeled: Leu6, Ser9, Met34, Phe109, His114, Phe118 and Phe204

Table 4 Interaction energies (kcal mol^{-1}) of the CDNB ligand with each key residue (lower than −0.1 kcal mol^{-1}) in the active site for Aggst1-2

Residue	Van der Waals	Electrostatic	Polar solvation	Non-polar solvation	Total
Leu6	−0.456	−0.124	0.177	−0.019	−0.422
Gly8	−0.514	−0.964	1.372	−0.058	−0.164
Ser9	−0.635	0.369	−0.081	−0.026	−0.373
Ala10	−0.116	0.258	−0.346	−0.020	−0.205
Met34	−0.774	0.293	−0.119	−0.075	−0.675
His38	−0.375	−0.710	0.675	−0.019	−0.430
Arg39	−0.143	−1.740	1.708	−0.010	−0.185
Tyr106	−0.167	0.188	−0.244	−0.001	−0.225
Phe109	−0.832	0.717	−0.609	−0.094	−0.818
Ala110	−0.979	0.849	−0.332	−0.137	−0.599
Phe112	−0.068	−0.010	−0.059	0.000	−0.137
Tyr113	−0.500	−0.465	0.250	−0.014	−0.729
His114	−0.893	−0.400	0.370	−0.073	−0.996
Val117	−0.509	−0.186	0.218	−0.019	−0.496
Phe118	−1.083	0.030	0.171	−0.156	−1.038
Leu170	−0.135	0.092	−0.084	−0.001	−0.128
Phe204	−1.594	−0.340	0.277	−0.131	−1.788
GSH209	−2.068	−0.290	1.911	−0.267	−0.714

−0.714 kcal mol^{-1} with CDNB, which indicates that these two ligands interact with each other. The interaction between GSH and CDNB was composed mainly of van der Waals interactions (−2.068 kcal mol^{-1}). The strong van der Waals interaction of GSH and CDNB may ensure that the nucleophilic addition reaction happens.

Conclusions

To date, the crystal structure of Aggst1-2 has not been identified. A 3D model of Aggst1-2 was constructed based on the crystal structure of *Anopheles dirus* species B GSTs 1–3, chain A (PDB code: 1JLV) and was refined by EM and MD simulations. The model was assessed by means of Profile-3D and Ramachandran plot. The results indicated that the structure is reliable. The complex CDNB-GSH-Aggst1-2 was obtained through docking studies. By means of MD simulations, the binding mode of the complex is in good agreement with the known experimental data. After MD simulations, our results arrived at some significant conclusions. The CDNB-GSH-Aggst1-2 complex is more stable than Aggst1-2, which indicates that these two ligands are essential for the stabilization of protein structure. Among the two ligands, GSH is especially indispensible for the stabilization of G-site residues. Through hydrogen analysis, a hydrogen bond between hydrogen atom of hydroxyl group in Ser9 and the sulfur atom from GSH was detected during MD simulation. This hydrogen bond fits the experimental data well. The absence of hydrogen bonds for CDNB ligand led us to identify an important π–π interaction. This interaction between CDNB and Phe204 plays a major role in the stabilization of the CDNB ligand, thus leading to subsequent reaction. Using per-residue binding free energy decomposition, several residues critical for ligand binding were identified. A strong van der Waals interaction between GSH and CDNB was identified using the per-residue binding free energy decomposition method. This interaction may be crucial for nucleophilic addition reaction. These results will be helpful in further determining the biological roles of Aggst1-2 and will provide insights to help design inhibitors of this protein.

Acknowledgments This work is supported by Natural Science Foundation of China (Grant Nos. 21273095, 20903045 and 21203072).

Reference

1. Armstrong RN (1991) Glutathione S-transferases: reaction mechanism, structure, and function. Chem Res Toxicol 4(2):131–140
2. Ames BN, Profet M, Gold LS (1990) Nature's chemicals and synthetic chemicals: comparative toxicology. Proc Natl Acad Sci USA 87(19): 7782–7786
3. Sheehan D, Meade G, Foley VM, Dowd CA (2001) Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. Biochem J 360(Pt 1):1–16
4. Habig WH, Pabst MJ, Jakoby WB (1974) Glutathione S-transferases the first enzymatic step in mercapturic acid formation. J Biol Chem 249(22):7130–7139
5. Xiao G, Liu S, Ji X, Johnson WW, Chen J, Parsons JF, Stevens WJ, Gilliland GL, Armstrong RN (1996) First-sphere and second-sphere

electrostatic effects in the active site of a class mu glutathione transferase. Biochemistry 35(15):4753–4765
6. Ji X, Armstrong RN, Gilliland GL (1993) Snapshots along the reaction coordinate of an SNAr reaction catalyzed by glutathione transferase. Biochemistry 32(48):12949–12954
7. Ranson H, Hemingway J (2005) Mosquito glutathione transferases. Methods Enzymol 401:226–241
8. Gohlke H, Kiel C, Case DA (2003) Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. J Mol Biol 330(4):891–914
9. Hou T, Wang J, Li Y, Wang W (2010) Assessing the performance of the MM/PBSA and MM/GBSA methods. I. The accuracy of binding free energy calculations based on molecular dynamics simulations. J Chem Inform Model 51(1):69–82
10. Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. J Comput Chem 32(5):866–877
11. Chu W-T, Zhang J-L, Zheng Q-C, Chen L, Xue Q, Zhang H-X (2013) Insights into the drug resistance induced by the BaDHPS mutations: molecular dynamic simulations and MM/GBSA studies. J Biomol Struct Dyn 31:1127–1136
12. Studio D (2009) version 2.5. Accelrys, San Diego, CA
13. Case D, Darden T, Cheatham III T, Simmerling C, Wang J, Duke R, Luo R, Walker R, Zhang W, Merz K (2010) AMBER 11. University of California, San Francisco
14. Yang L, Tan C, Hsieh MJ, Wang J, Duan Y, Cieplak P, Caldwell J, Kollman PA, Luo R (2006) New-generation amber united-atom force field. J Phys Chem B 110(26):13166–13176
15. Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J Chem Inf Model 51(1):69–82. doi:10.1021/ci100275a
16. Swanson JM, Henchman RH, McCammon JA (2004) Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. Biophys J 86(1):67–74
17. USA AISD (2009) Profile-3D user guide.
18. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystall 26(2):283–291
19. Widersten M, Kolm RH, Björnestedt R, Mannervik B (1992) Contribution of five amino acid residues in the glutathione-binding site to the function of human glutathione transferase P1-1. Biochem J 285(Pt 2):377–381
20. Casalone E, Allocati N, Ceccarelli I, Masulli M, Rossjohn J, Parker MW, Di Ilio C (1998) Site-directed mutagenesis of the Proteus mirabilis glutathione transferase B1-1 G-site. FEBS Lett 423(2): 122–124
21. Andújar-Sánchez M, Clemente-Jiménez JM, Las Heras-Vázquez FJ, Rodríguez-Vico F, Cámara-Artigas A, Jara-Pérez V (2003) Thermodynamics of glutathione binding to the tyrosine 7 to phenylalanine mutant of glutathione S-transferase from *Schistosoma japonicum*. Int J Biol Macromol 32(3):77–82
22. Winayanuwattikun P, Ketterman AJ (2004) Catalytic and structural contributions for glutathione-binding residues in a Delta class glutathione S-transferase. Biochem J 382(Pt 2):751–757
23. Frisch M, Trucks G, Schlegel HB, Scuseria G, Robb M, Cheeseman J, Scalmani G, Barone V, Mennucci B, Petersson G (2009) Gaussian 09, revision A. 1. Gaussian Inc, Wallingford
24. Wu G, Robertson DH, Brooks CL, Vieth M (2003) Detailed analysis of grid–based molecular docking: a case study of CDOCKER—A CHARMm–based MD docking algorithm. J Comput Chem 24(13): 1549–1562
25. Schuttelkopf AW, van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr 60(8):1355–1363
26. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25(9):1157–1174
27. Wang J, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model 25(2):247–260
28. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935
29. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N· log (N) method for Ewald sums in large systems. J Chem Phys 98(12): 10089–10092
30. Case D, Darden T, Cheatham Iii T, Simmerling C, Wang J, Duke R, Luo R, Walker R, Zhang W, Merz K (2010) Amber Tools 1.5. Amber
31. McGaughey GB, Gagné M, Rappé AK (1998) π-Stacking interactions alive and well in proteins. J Biol Chem 273(25):15458–15463
32. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14(1):33–38
33. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612
34. DeLano WL (2002) The PyMOL molecular graphics system. Version 1.1, 2002, Schroinger LLC, 2002, http://www.pymol.org
35. Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 112(16):6127–6129
36. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large–scale conformational changes with a modified generalized born model. Proteins: Structure, Function, and Bioinformatics 55(2):383–394
37. Board P, Coggan M, Wilce M, Parker MW (1995) Evidence for an essential serine residue in the active site of the Theta class glutathione transferases. Biochem J 311(Pt 1):247–250
38. Velazquez F, Peak-Chew SY, Fernández IS, Neumann CS, Kay RR (2011) Identification of a eukaryotic reductive dechlorinase and characterization of its mechanism of action on its natural substrate. Chem Biol 18(10):1252–1260
39. Zheng Y-J, Ornstein RL (1997) Mechanism of nucleophilic aromatic substitution of 1-chloro-2, 4-dinitrobenzene by glutathione in the gas phase and in solution. Implications for the mode of action of glutathione S-transferases. J Am Chem Soc 119(4):648–655