Homology Modeling of Cephalopod Lens *S*-Crystallin: A Natural Mutant of Sigma-Class Glutathione Transferase with Diminished Endogenous Activity

Chyh-Chong Chuang,* Shih-Hsiung Wu,*# Shyh-Horng Chiou,*# and Gu-Gang Chang§
*Institute of Biological Chemistry, Academia Sinica, #Institute of Biochemical Sciences, National Taiwan University, and §Department of Biochemistry, National Defense Medical Center, Taipei, Taiwan, Republic of China

ABSTRACT The soluble *S*-crystallin constitutes the major lens protein in cephalopods. The primary amino acid sequence of *S*-crystallin shows an overall 41% identity with the digestive gland sigma-class glutathione transferase (GST) of cephalopod. However, the lens *S*-crystallin fails to bind to the *S*-hexylglutathione affinity column and shows very little GST activity in the nucleophilic aromatic substitution reaction between GSH and 1-chloro-2,4-dinitrobenzene. When compared with other classes of GST, the *S*-crystallin has an 11-amino acid residues insertion between the conserved α4 and α5 helices. Based on the crystal structure of squid sigma-class GST, a tertiary structure model for the octopus lens *S*-crystallin is constructed. The modeled *S*-crystallin structure has an overall topology similar to the squid sigma-class GST, albeit with longer α4 and α5 helical chains, corresponding to the long insertion. This insertion, however, makes the active center region of *S*-crystallin to be in a more closed conformation than the sigma-class GST. The active center region of *S*-crystallin is even more shielded and buried after dimerization, which may explain for the failure of *S*-crystallin to bind to the immobilized-glutathione in affinity chromatography. In the active site region, the electrostatic potential surface calculated from the modeled structure is quite different from that of squid GST. The positively charged environment, which contributes to stabilize the negatively charged Meisenheimer complex, is altered in *S*-crystallin probably because of mutation of Asn99 in GST to Asp101 in *S*-crystallin. Furthermore, the important Phe106 in authentic GST is changed to His108 in *S*-crystallin. Combining the topological differences as revealed by computer graphics and sequence variation at these structurally relevant residues provide strong structural evidences to account for the much decreased GST activity of *S*-crystallin as compared with the authentic GST of the digestive gland.

INTRODUCTION

Crystallins are soluble proteins in eye lenses, which play an important role in the maintenance of lens transparency and optical clarity (De Jong et al., 1989). In the past decade, some taxon-specific crystallins were found to have structures related to cytosolic enzymes (Piatigorsky and Wistow, 1991; Piatigorsky, 1992). Among them the *S*-crystallin in cephalopods lenses has been shown to have an amino acid sequence similar to glutathione transferase (GST; enzyme nomenclature number EC 2.5.1, 18) (for review, see Tomarev and Piatigorsky, 1996).

We have isolated and characterized the octopus hepatopancreatic GST and lens *S*-crystallin (Lin and Chiou, 1992; Tang et al., 1994; Chiou et al., 1995; Tang and Chang, 1996). *S*-Crystallin constitutes the major cephalopod lens protein, however, it possesses very little GST activity as compared with the digestive gland GST (Tang et al., 1994; Ji et al., 1995). Thus, *S*-Crystallin is a natural mutant of GST and serves as an excellent model to investigate the structure-function relationships of GST. More recently, we have delineated the complete amino acid sequence of octopus *S*-crystallin by cDNA cloning technique (Lin and Chiou, 1992; Chiou et al., 1995). From the steady-state kinetic studies, we have deduced a plausible kinetic mechanism for the octopus digestive gland GST-σ (Tang and Chang, 1995) and lens *S*-crystallin (Tang and Chang, 1996).

Glutathione transferase catalyzes the conjugation of glutathione with various endogenous and xenobiotic electrophilic compounds that constitute a major detoxification mechanism of biological systems (Mannervik and Danielson, 1988; Rushmore and Pickett, 1993; Hayes and Pulford, 1995; Armstrong, 1997). On the basis of amino acid sequence homologies, substrate specificities, and immunocross-reactivity with human GSTs, the supergene family of mammalian cytosolic GST isoenzymes is grouped into four gene classes: GST-alpha (GST-α), GST-mu (GST-μ), GST-pi (GST-π), and GST-theta (GST-θ). The three-dimensional structures of these GST classes have been resolved by x-ray crystallographic analysis (Sinning et al., 1993; Ji et al., 1992; Dirr et al., 1994a; Wilce et al., 1995). The cephalopod GST was found to belong to a separate class, i.e., sigma-class GST (GST-σ) (Ji et al., 1995).

Homology protein modeling, which uses the coordinates of known structures as templates to predict the conformation of another protein that has a similar amino acid sequence, has been proven to be useful in establishing the tertiary structures of homologous proteins (Sternberg,

Received for publication 6 July 1998 and in final form 20 October 1998.
Address reprint requests to Professor Gu-Gang Chang, Department of Biochemistry, National Defense Medical Center, P. O. Box 90048, Taipei 100, Taiwan, Republic of China. Tel.: +886-2-2364-1207; Fax: +886-2-2365-5746; E-mail: ggchang@ndmc1.ndmctsgh.edu.tw or to Prof. Shih-Hsiung Wu, Institute of Biological Chemistry, Academia Sinica, Nankang 115, Taipei, Taiwan, Republic of China. Tel.: +886-2-2785-5696; Fax: +886-2-7883743; E-mail: shwu@gate.sinica.edu.tw.

0006-3495/99/02/679/12 $2.00

1996). All GSTs share a common topology in tertiary structure (Armstrong, 1994; Wilce and Parker, 1994; Dirr et al., 1994b). The essential amino acid residues at the active site region are highly conserved. The GSTs are thus proposed to have a common ligand docking mode and the same chemical mechanism in the detoxification reaction. Based on these observations, the tertiary structures of several GST or related proteins have been theoretically built recently by homology modeling to explore the structural basis for substrate binding and catalysis (Chelvanayagam et al., 1997; Koehler et al., 1997; Marsh and Ferguson, 1997; Orozco et al., 1997).

The primary amino acid sequences of *S*-crystallin from octopus and the GST-σ from squid digestive gland have an overall 41% identity. From the evolutionary point of view, both proteins should share a closely related ancestral relationship (Ji et al., 1995). In this report, by using the coordinates of GST-σ, we describe the construction of a possible tertiary structure of *S*-crystallin. This model provides a plausible explanation for the diminished GST activity and low affinity of the cephalopod *S*-crystallin to the *S*-hexylglutathione affinity column.

METHODS

The modeling processes involved four steps: 1) sequence alignment between the target and the template; 2) building an initial model; 3) refining the model; and 4) evaluating the model. Modeling was performed using commercial software from MSI/Biosym Technologies (San Diego, CA) and Tripos Inc. (St. Louis, MO) and public domain software from many research groups.

Sequence alignment

Pairwise alignment of *S*-crystallin with the squid GST-σ was performed using the BESTFIT algorithm of the GCG program package (Genetics Computer Group, 1994). The pairwise alignment was compared with a multisequence alignment of all five crystal structures of four main classes of GSTs (pig π, human π, human α, human μ, rat μ, and squid σ) and the *S*-crystallin sequence, which was created using the PILEUP option of the GCG program, in order to identify conserved and variable regions of the sequences and aid in determining the most robust gap arrangement.

In a separate attempt to optimally align the target sequence onto the template structure, a threading procedure was used. The *S*-crystallin sequence was threaded onto the squid GST structure with the fold-recognition method developed by Fischer and Eisenberg (1996). The final alignment used for model building was achieved by manual judgment of the above three kinds of sequence alignment.

Initial model

With sequence alignment in hand, the first stage of model building was started by first stenciling coordinates from the template onto the target sequence to create a scaffolding for the model using the Tripos COMPOSER package (Sybyl, Tripos user manual, version 6.3, Tripos, 1996; for review, see Johnson et al., 1994). This was performed for all residues where the sequences are in the defined structure conserved regions (SCRs), which are identified from the structure-based alignment. The coordinates in the SCRs are copied directly to the target. Side-chain geometry is built using the automated procedure incorporated into the program COMPOSER. Where ever the two aligned residues type are the same, the template side-chain geometry is directly adopted by the target, otherwise the side chain is built by mutating the template residue into the target residue. First the template residue is transformed onto the model backbone by least-squares fitting of the N-, C_α-, and C-backbone atoms. Then the coordinates for compatible side-chain atoms in the target residue are copied directly from the corresponding atoms in the template residue. The remaining target atoms get their coordinates from the canonical side-chain conformation for that residue type in the appropriate secondary structure. The left intervening stretches between the SCRs were considered structure variable regions (SVRs), these are defined as the loop regions between the conserved secondary structures of GST family in the structure-based alignment as shown in Fig. 1.

The variable regions are modeled by either directly copying from template when the SVRs in question have the same number of residues in the template and target or identifying a suitable segment from a known structure in a compiled structure data bank (Hobohm and Sander, 1994). A loop search is made for a segment having the desired number of residues and the proper end-to-end distances across the three anchor residues at either side of the putative loop such that the loop can be fitted joining the contiguous conserved regions (Summers and Karplus, 1990). The best fragment is melded with the SCRs after making small alterations of the torsion angles at the melded regions. Side chains in the SVRs are modeled by the same way used in the SCRs modeling.

After the initial model is completely built, the side chains are modeled further by using the SCWRL (side-chains with a rotamer library) protocol incorporated in the program SCWRL (Bower et al., 1997). Side-chain geometry is kept fixed for all the conserved residues in the SCRs in this procedure. This method takes a structure with residues in their favorable backbone-dependent rotamers (Dunbrack and Karplus, 1994) and systematically resolves the conflicts that arise from that structure. An energy function for simple repulsive steric clash check is defined in SCWRL. Each residue begins in its most favorable backbone-dependent rotamers. Side-chain main-chain (fixed) steric clash is relieved by changing to progressively less favorable rotamers until one is found that does not conflict with the main chain. Then side-chain to side-chain clashes are relieved through a combinatorial search to find its minimal steric clash score by allowing the clashing residues to explore all their respective rotamers.

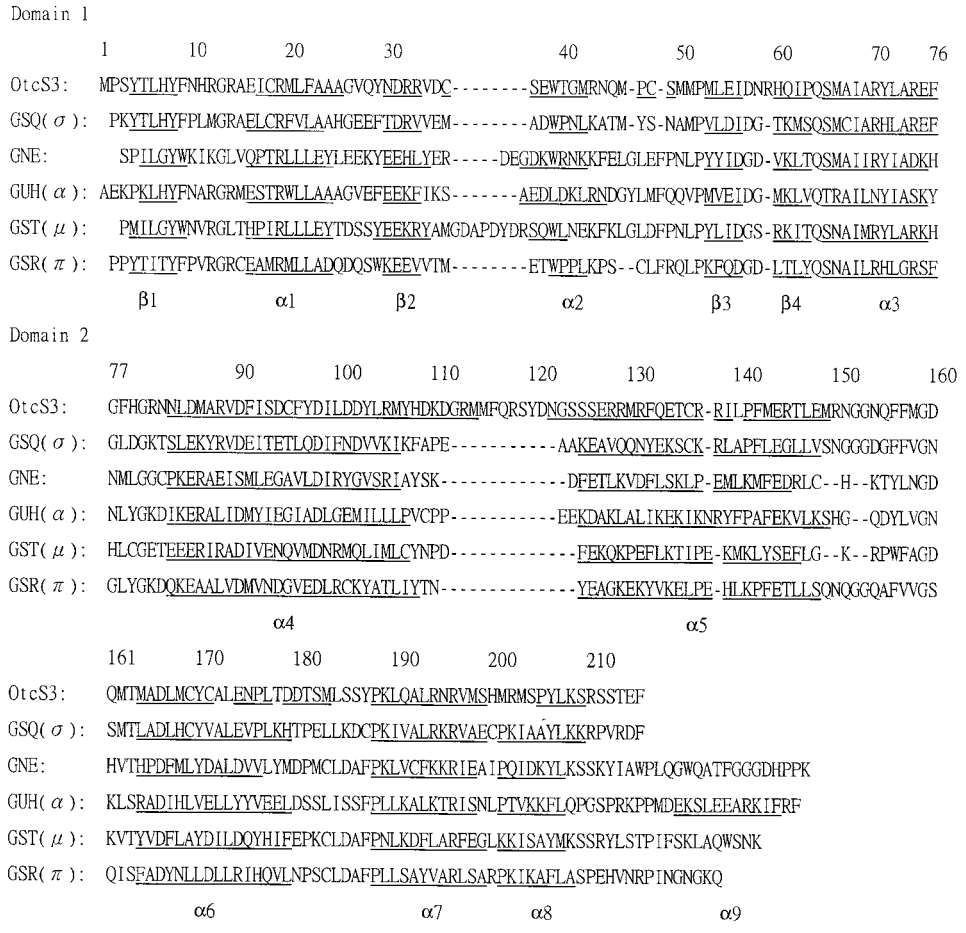

FIGURE 1 Pairwise alignment of *S*-crystallin with GST-σ and other classes of GST. The α-helices and β-strands in *S*-crystallin are underlined and labeled. The identity between *S*-crystallin and GST-σ is 41% with a *z*-score of 17.35.

Model refinement

The starting structure is refined by successive iterations of molecular dynamics (MD) followed by energy minimization (EM) using DISCOVER program (MSI/Biosym Technologies) and AMBER force field (Pearlman et al., 1995). Each SVR is first submitted to molecular dynamics followed by energy minimization. The number of constraints applied to the protein are progressively reduced during subsequent cycles (Du et al., 1992; Vinals et al., 1995).

The nonbonded interaction cutoff distance was left at the default value of 100 Å during the whole refinement procedure. No water molecules were introduced into the system and the solvent effect was approximated by a distance-dependent dielectric constant ($4 \times r$). The integration time step for the molecular dynamics calculation is 1 fs, and snapshots of the protein structure are taken every picosecond. Before the final round of optimization, *S*-crystallin is placed in its dimeric form by applying the symmetry operations to the target monomer as the same way operated to the squid GST structure (Ji et al., 1995).

Model evaluation

The deviation from the standard geometry and atomic overlap were determined more rigorously on a residue-by-residue basis using the PROCHECK program (Laskowski et al., 1993). PROSAII (Sippl, 1993) and Profile-3D (Bowie et al., 1991; Luthy et al., 1992) were used to assess the compatibility between modeled conformation and the sequence of *S*-crystallin and developed an energetic profile of the modeled structure. Pairwise residue interaction energies were calculated in PROSAII using β-carbon interactions, and a 50-residue window was used for energy averaging at each residue position. Positive energy indicates putatively misfolded region. In model evaluation by Profile-3D, a self-compatibility score (*S*) is determined for each residue in the sequence. A smoothing window size of 21 residues is used. Negative (*S*) score indicates misfolded region. Finally, the model was examined for correlation with the experimental results.

Calculation of electrostatic surface potential map

The electrostatic potentials of the modeled *S*-crystallin and the x-ray structure of squid GST-σ were calculated by the program GRASP (Nicholls et al., 1991). The linearized Poisson-Boltzmann equation was solved by the finite difference method. Protein surfaces on which electrostatic potential is calculated are solvent-accessible surface, assuming the probe radius of 1.4 Å to represent a water

molecule. The dielectric constants for the exterior and interior of the protein were set at 80 and 2, respectively, with an ionic strength of zero. All basic and acidic side chains were fully charged except for histidines, which were treated as neutral. Contributions from the product conjugate *S*-(2,4-dinitrophenyl) glutathione (*S*-(DNP)GS) in the squid GST structure were neglected during the calculation. The dimeric structure of the squid GST was generated by applying the symmetry operation to the target monomer (PDB code 1GSQ). In the calculation of the electrostatic surface potential map of monomer, the contributions from the other monomer were also neglected.

RESULTS AND DISCUSSION

Homologous modeling

S-Crystallin has an overall 41–42% sequence identical to the squid GST with 50% sequence identity for the N-terminal domain (residue 1–76) and 36% for the C-terminal domain (residue 83–215). For all the sequence alignment methods we tested, the alignment in the N-terminal domain was agreeable. But, with the lower sequence identity and an 11-residue insertion, the alignment in the C-terminal domain appears quite diverse particularly in the α5 helix and the loop between α4 and α5. Our first pairwise alignment of *S*-crystallin and squid GST sequences by the BESTFIT method has an 11-residue insertion in the middle of the α5 fragment of squid GST structure. A further multiple sequence alignment (PILEUP) of *S*-crystallin with GST sequences, the three-dimensional structures of which have been solved by x-ray crystallography (pig π, human π, human α, human μ, rat μ, and squid σ-GST), also showed the same result.

In the fold-recognition method (Fischer and Eisenberg, 1996), a new compatibility function and a dynamic programming algorithm were used to optimize alignment between the probe sequence and the target (Needleman and Wunsch, 1970; Smith and Waterman, 1981). The pairwise alignment obtained from this method showed the same alignment in the N-terminal domain (and up to the end of α4 helix) when compared with the BESTFIT pairwise alignment and PILEUP multiple sequence alignment. But in the C-terminal domain alignment, this method has different result. The 11-residue was inserted in the loop between α4 and α5 helices. Wilce and Parker (1994) analyzed a large amount of sequences of GST family from several sources, including the main classes of GSTs and lens *S*-crystallins. When compared with the multiple sequence alignment based on the solved crystal structures (Ji et al., 1995), the alignment of consensus GST sequences in their result also showed a long insertion in the loop between α4 and α5 helices of GST structures. This central insert, which is also a characteristic feature of several *S*-crystallins of the squid, had been reported to be encoded by a separate exon in the lens genes of squid (Tomarev et al., 1992, 1995). These strongly support the alignment result of the threading method even though it is well known that, in fold recognition, identifying the correct fold in a set of structures is a much easier problem than providing the correct alignment for the probe and the fold (Lemer et al., 1995; Wilmanns and Eisenberg, 1995).

The final alignment used for model building was shown in Fig. 1. By the method of gonnet+predss (Fischer and Eisenberg, 1996), the highest scoring fold of *S*-crystallin was obtained, as expected, with squid GST-σ (*z*-score 17.35), which is above the threshold of the confidence threshold set by a *z*-score of 4.8 $\pm$ 1.0. An 11-residue long insertion was located between α4 and α5 helices of GST structure. Comparing with previous works, the seven conserved residues observed in the alignment of consensus sequences of GST family (Wilce and Parker, 1994) and the 16 conserved residues in the structure-based alignment (Ji et al., 1995) are all correctly aligned. All critical amino acids involved in GSH binding and catalysis are conserved (see below).

The overall topology of the model is similar to the x-ray structure of GST which consists of four-stranded β-sheet and eight α-helices arranged in a smaller N-terminal domain I (α1-α3, β1-β4) and a larger C-terminal domain II (α4-α8) (Fig. 2). The N-terminal domain I is an α/β structure built up of the four-stranded β-sheet and three α-helices that represents a typical GSH-binding domain of the $\beta\alpha\beta\alpha(\alpha)\beta\beta\alpha$ folding pattern (Gilliland, 1993). This GSH binding domain, consisting of a motif 1 of $\beta\alpha\beta$ followed by a motif 2 of $\beta\beta\alpha$, is found in many GSH-binding proteins including glutathione reductase, glutathione peroxidase, glutaredoxin, glutathione synthetase, as well as glutathione transferases (Gilliland, 1993). However, the C-terminal domain II has a novel topology and is more diverse between different classes of GSTs or other GSH-binding proteins, corresponding to various substrate specificities among these proteins. Domain II of *S*-crystallin, like other GSTs, is composed of five α-helices folded in a similar pattern. The two domains are linked with a six-residue linker and interact with each other through α1-α3 and α4-α6 mostly by hydrophobic interactions.

The most distinct difference between the template and *S*-crystallin model is the 11-residue insertion (number 111–121) between α4 and α5 helices (Fig. 2 *B*). Based on the results of alignment by the threading method, the 23-residue segment including the 11-residue insertion, the 6-residue loop between α4 and α5 helices, and the three anchor residues at either side of α4 and α5 was modeled by the following procedure. First, suitable segments with known structures and the same number of residues were identified from a compiled structure data bank. Next, the best segment used to fit into the three anchor residues at either side of α4 and α5 was selected with the lowest root mean square deviation when melded with the SCRs. The final modeled structure was achieved by energy minimization and molecular dynamics as described in the Method section. The center of this 11-residue insertion forms a loop to connect both α-chains and both sides of the insertion extend α4 and

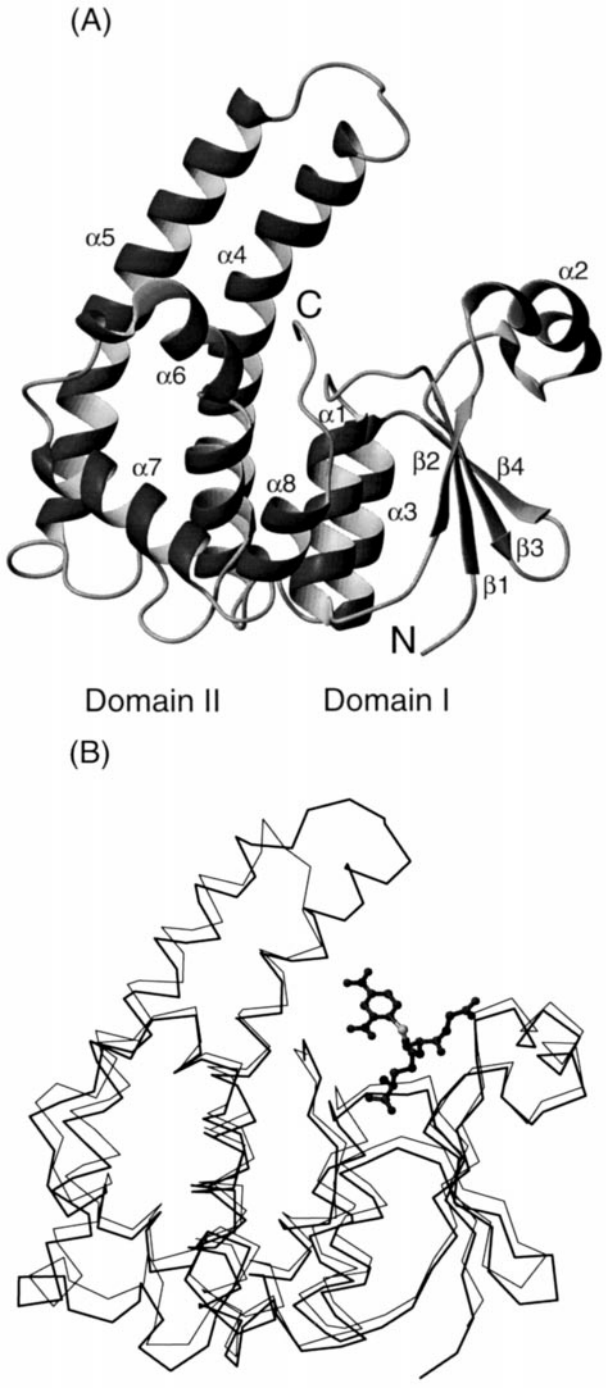

FIGURE 2 Overall topology of *S*-crystallin monomer. The overall topologies of the modeled *S*-crystallin monomer (*A*) and its superposition (*heavy line*) with the template GST-σ (*light line*) (*B*) were shown. Like other GSTs, the protein contains two domains. The N-terminal domain I consists of $\beta\alpha\beta\alpha\beta\beta\alpha$ folding motif in which the β-strand is in a $\uparrow\beta2\uparrow\beta1\downarrow\beta3\uparrow\beta4$ arrangement. The C-terminal domain II consists of five α-helices in which the $\alpha4$ and $\alpha5$ are longer than those of the squid GST-σ. The diagram was generated by the program MOLMOL (Karadi et al., 1996). The disposition of the longer helices relative to the active site of the enzyme was shown in *B*, which also indicates the binding of *S*-(DNP)GS (in ball-and-stick model) at the active center.

$\alpha5$ longer than those of the template. The longer $\alpha4$ and $\alpha5$ chains are slightly bent toward the binding cleft (G-site) of GSH and the connected loop just resides in front of the binding cleft.

Like all GST, *S*-crystallin is in a dimeric form in nature (Chiou, 1984) (Fig. 3). The interaction of two subunits in the model is similar to that of the template, squid digestive gland GST-σ. The two monomers interact between domain I ($\beta4$, $\alpha3$) of one monomer and domain II ($\alpha4$, $\alpha5$) of the adjacent monomer. Several key residues involved the molecular recognition for dimerization in the template are conserved in *S*-crystallin such as Met47 (Met45 in GST-σ),

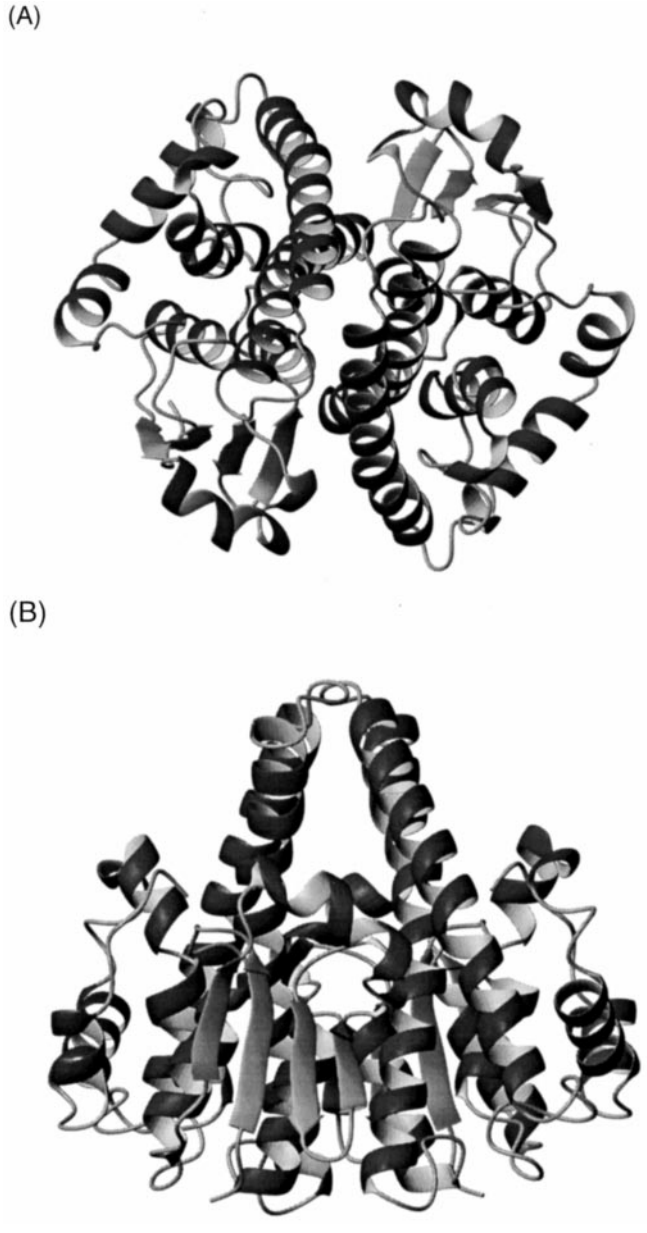

FIGURE 3 Overall *S*-crystallin dimer model along the vertical (*A*) and central (*B*) axes. The loop region between $\alpha4$ and $\alpha5$ is shown on the top in *B*. The diagram was generated by the program MOLMOL (Karadi et al., 1996).

Arg70 (Arg68 in GST-σ), and Phe142 (Phe129 in GST-σ), and some are conservative mutations such as Ser93 (Thr93 in GST-σ) and Met143 (Leu130 in GST-σ).

The model was evaluated by PROCHECK, a program to check the stereochemical quality of protein structures. The ψ and φ angle values for all nonglycyl, nonprolyl but one residue (Ser185) are in the most favored regions (83.6%) and the additional allowed regions (14.4%). The only one out-lying residue, Ser185, is found in the loop between α6-α7, located far from the putative active site region of the protein.

Besides the evaluation of stereochemical quality for the model, two programs (ProsaII and Profile 3D) based on the energy evaluation of sequence arrangement of the protein in three-dimensional structure were used to further evaluate the model for complementary structural analyses. The energy profile over the whole sequence of the model is in the reasonable folded region, indicating that the *S*-crystallin model is native-like.

A good model of *S*-crystallin must be able to explain the characteristic properties of this lens protein, i.e., low binding affinity with GSH-agarose column, diminished but not completely vanished GST activity in the nucleophilic aromatic substitution reaction (S_NAr) between GSH and 1-chloro-2,4-dinitrobenzene (CDNB). The endogenous GST activity (k_{cat}) of *S*-crystallin in the GSH/CDNB or the GSH/ethacrynic acid system is 0.06 s^{-1} and 0.3 s^{-1}, respectively, which are only 1/1,912 or 1/787, respectively, as compared with those of the digestive gland GST-σ (Tang et al., 1994).

The overall topology of the modeled structure of *S*-crystallin is similar to that of the GST-σ except for the 11-residue insertion between α4 and α5 helices. Tomarev et al. (1995) have demonstrated that addition of this insertion peptide into the GST-σ reduced the specific activity of the enzyme by 30-fold. However, the reverse experiment, i.e., deletion of this insertion peptide from *S*-crystallin, did not result in increased GST activity. Thus, other factors must be also involved in the loss of GST activity of *S*-crystallin. We then proceeded to compare the active site region of *S*-crystallin and GST-σ.

Active site region of *S*-crystallin and explanation for the diminished endogenous GST activity

Based on the x-ray crystallographic data of GST-σ (Ji et al., 1995), the putative important amino acid residues at the G-site (glutathione binding site) of domain I are all conserved in *S*-crystallin (Fig. 4), e.g., Asn64 (Asn62 in GST-σ), Ser65 (Ser63 in GST-σ), Met51 (Met50 in GST-σ), Trp39 (Trp38 in GST-σ). The Asp98B (Asp96B in GST-σ) from the other subunit also appears in the active center. Only two residues are different, i.e., Arg43 in *S*-crystallin

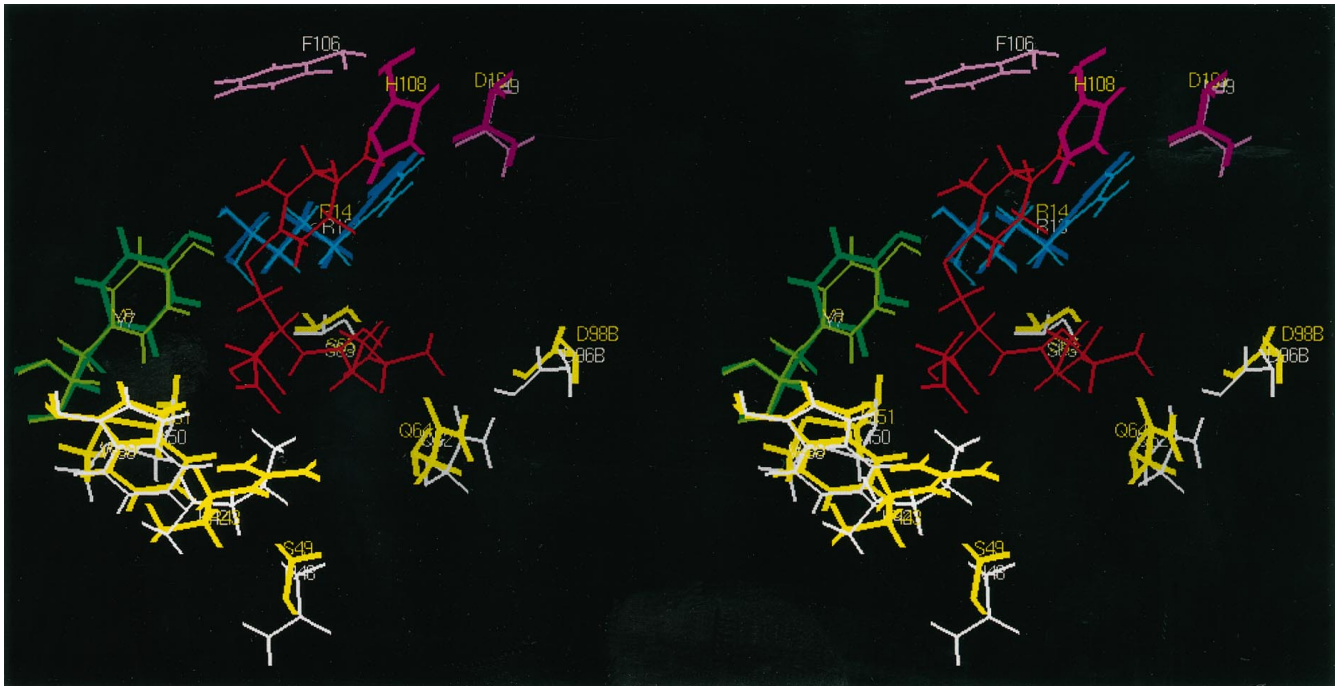

FIGURE 4 Stereoview of the alignments of *S*-crystallin and GST-σ at the active site region. GST-σ (*light color*) (Ji et al., 1995; PDB code 1GSQ) and *S*-crystallin (*dark color*; this work) are shown in complexed with *S*-(2, 4-dinitrophenyl) glutathione (*red*). The putative essential amino acid residues involved in the electrostatic interactions between glutathione and the proteins were shown in white (GST-σ) and yellow (*S*-crystallin). The important residues Tyr8 (*green*) and Arg14 (*blue*) of *S*-crystallin are almost completely superimposible with those of GST-σ. The major differences between GST-σ and *S*-crystallin are shown in pink color for GST-σ and magenta color for *S*-crystallin, which include change of Asn99 in GST-σ to Asp101 in *S*-crystallin and Phe106 in GST-σ to His108 in *S*-crystallin. D98B and D96B indicate that these residues are from other subunit of the dimeric proteins.

but Lys42 in GST-σ and Ser49 in *S*-crystallin but Asn48 in GST-σ. These substitutions, however, do not alter the overall structure and might not affect the GSH binding property.

For the GST-catalyzed nucleophilic aromatic substitution of GSH and electrophilic substrate CDNB, the enzymatic reaction involves an initial ionization of glutathione's sulfhydryl group to yield a better nucleophilic glutathiolate anion (GS^-) (step I in Fig. 5). The latter then attacks the *ipso* carbon atom in CDNB, which bears the leaving chloride atom, forming a negatively charged Meisenheimer complex (step IIa in Fig. 5). Finally, the chloride ion leaves and forms a water-soluble conjugate *S*-(DNP)GS (step IIb in Fig. 5) for excretion from the cells. The rate-limiting step for the cephalopods GST-catalyzed S_NAr reaction has been proposed at the σ-bond formation in the Meisenheimer complex (Tang and Chang, 1995). Direct evidence for enzymatic stabilization of the Meisenheimer complex in the S_NAr reaction catalyzed by GST-μ and GST-π have been provided by using a transition state analog 1-(*S*-glutathionyl)-2,4,6-trinitrocyclohexadienate in complex with the enzyme (Ji et al., 1993; Prade et al., 1997).

To assist the process of the above reaction sequence, a tyrosyl residue has been demonstrated in most classes of GST to act as a general base to abstract the hydron from GSH, lowering the pK_a of the thiol of bound GSH and results in the formation of a better nucleophilic glutathiolate that facilitates the subsequent reactions (Liu et al., 1992; Atkins et al., 1993; Meyer et al., 1993; Orozco et al., 1997). The only exception is GST-θ, which uses a serine residue for the same purpose through a hydrogen bonding (Wilce et al., 1995; Marsh and Ferguson, 1997; Chelvanayagam et al., 1997). The corresponding position in GST-σ is a tyrosine residue (Tyr7), which is conserved in *S*-crystallin (Tyr8).

Another highly conserved positively charged arginine residue (Arg14 in *S*-crystallin) may be critical in stabilizing the negatively charged Meisenheimer complex. The functional importance of the corresponding Arg15 in GST-α has been demonstrated by site-specific mutagenesis (Björnestedt et al., 1995). The location and orientation of Tyr8 and Arg14 in *S*-crystallin are similar to those of GST-σ (Fig. 4). In summary, all those amino acid residues involved in GSH binding and the catalytically essential groups are conserved in *S*-crystallin, yet the endogenous GST activity of *S*-crystallin in the S_NAr reaction is three orders of magnitude smaller than that of the hepatopancreatic GST-σ.

The major differences between *S*-crystallin and GST-σ are found in the hydrophobic site, located in domain II, for the xenobiotic substrate binding. The hydrophobic site is found within the same cavity of G-site immediately adjacent to the sulfur atom of GSH. A tyrosine residue at the hydrophobic site has been demonstrated to play various roles within and between the GST isoenzyme family (Lo Bello et al., 1997). For GST-μ, GST-π, and GST-θ, a tyrosine residue (Tyr115 in GST-μ, Tyr108 in GST-π, Tyr115 in GST-θ), located near the top of α4 helix, assists in stabilizing the enol or enolate intermediate in the Michael addition reaction or plays a role in opening the epoxide ring (Johnson et al., 1993; Barycki and Colman, 1993). In GST-μ, a hydrogen bond between Tyr115 and Ser209 blocks the channel to the active site and freezes the segmental motion of the protein. This tyrosine was changed to phenylalanine in GST-σ (Phe106) (Ji et al., 1995) that makes the active site of GST-σ relatively open compared with the partially occluded active sites of GST-α and GST-μ. The contribution of this aromatic residue in the S_NAr reaction of the CDNB/GSH system is presumed to solely provide hydrophobic interactions through the aromatic ring (Lo Bello et al., 1997). For GST-α, a valine residue (Val104) is at the corresponding position. His108 in *S*-crystallin replaced the Phe106 in GST-σ. As shown in Fig. 4, the aromatic ring of phenylalanine 106 in GST-σ and the imidazole ring of His108 in *S*-crystallin are almost 90° perpendicular to each other. Changing of the hydrophobic phenylalanine to a hydrophilic histidine residue may hamper the binding of hydrophobic substrate that might explain, at least in part, the diminished S_NAr reactivity of *S*-crystallin. However, the following finding provides a more compelling explanation for the diminished enzyme activity of *S*-crystallin in the CDNB/GSH system.

Besides the 11-residue insertion between α-4 and α-5 helices, the most prominent difference between *S*-crystallin and GST-σ is that an Asn99 in GST-σ was replaced by Asp101 in *S*-crystallin. Although this aspartate residue is almost superimposible with Asn99 in GST-σ (Fig. 4), the Asp101 in *S*-crystallin is located in the vicinity of Arg14. The distance between the geometric center of the resonance

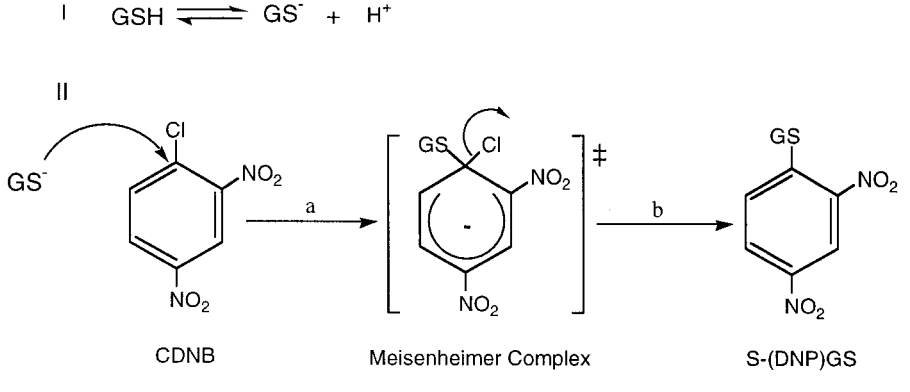

FIGURE 5 Proposed chemical mechanism for the base-catalyzed nucleophilic aromatic substitution of GSH and CDNB.

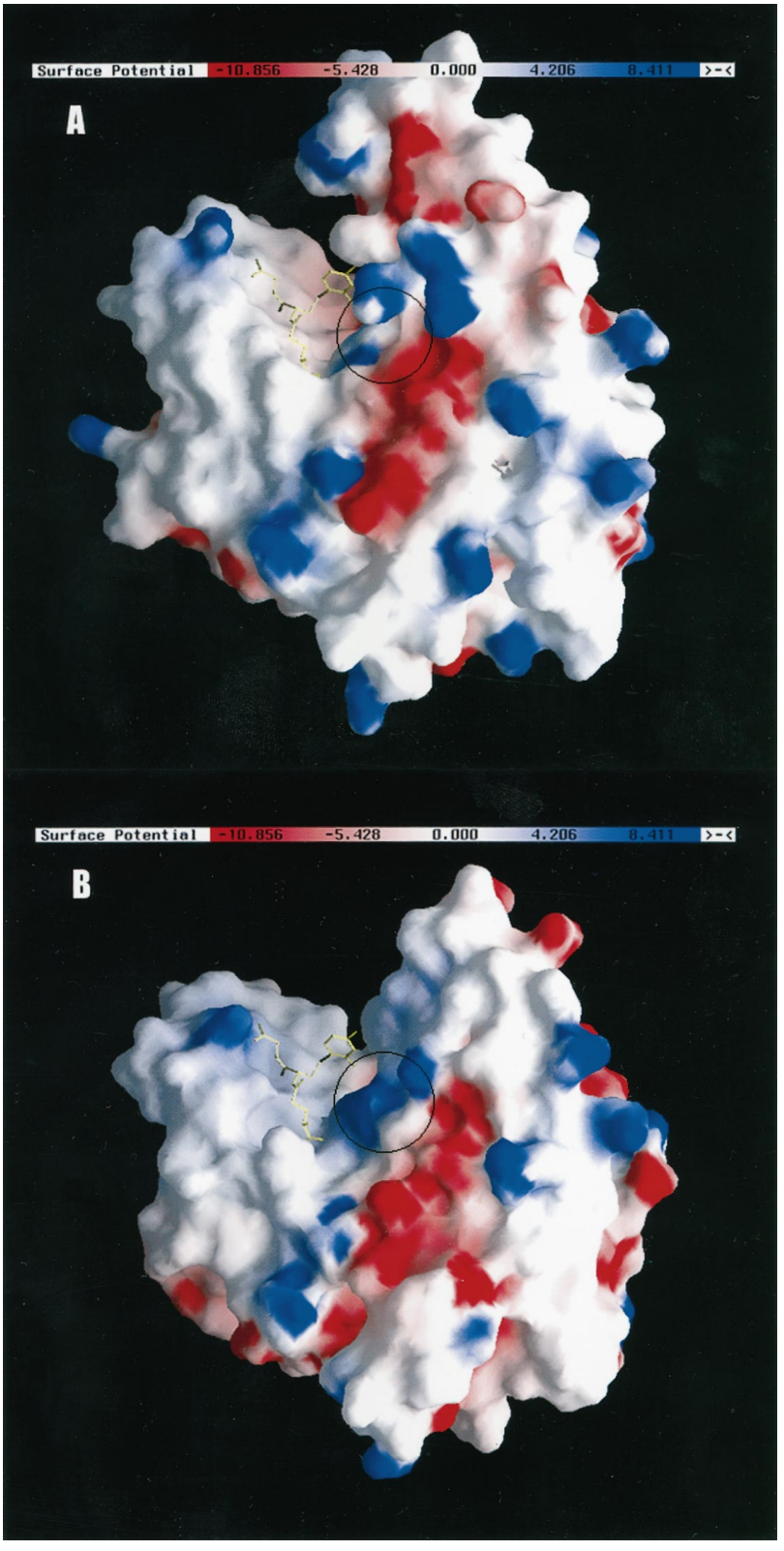

FIGURE 6 Comparison of the surface electric potential of *S*-crystallin and GST-σ monomers. Both *S*-crystallin (*A*) and GST-σ (*B*) are in the same orientation and complexed with *S*-(2,4-dinitrophenyl) glutathione (*yellow*). Areas of positive potential are blue and those with negative potential are red. The salient electrostatic differences between GST-σ and *S*-crystallin in the active site region around Arg14 are shown in the area highlighted with a circle. The diagram was generated by the program GRASP (Nicholls et al., 1991).

structure of carboxyl and guanidino groups is 3.71 Å, which is close enough to have charge-charge interactions. These charge interactions might diminish the important function of arginine residue in stabilizing the negatively charged Meisenheimer complex. This finding strongly suggests that the much less GST-catalyzed S_NAr reactivity of *S*-crystallin is caused by a decrease in the positive charge environment at the active center.

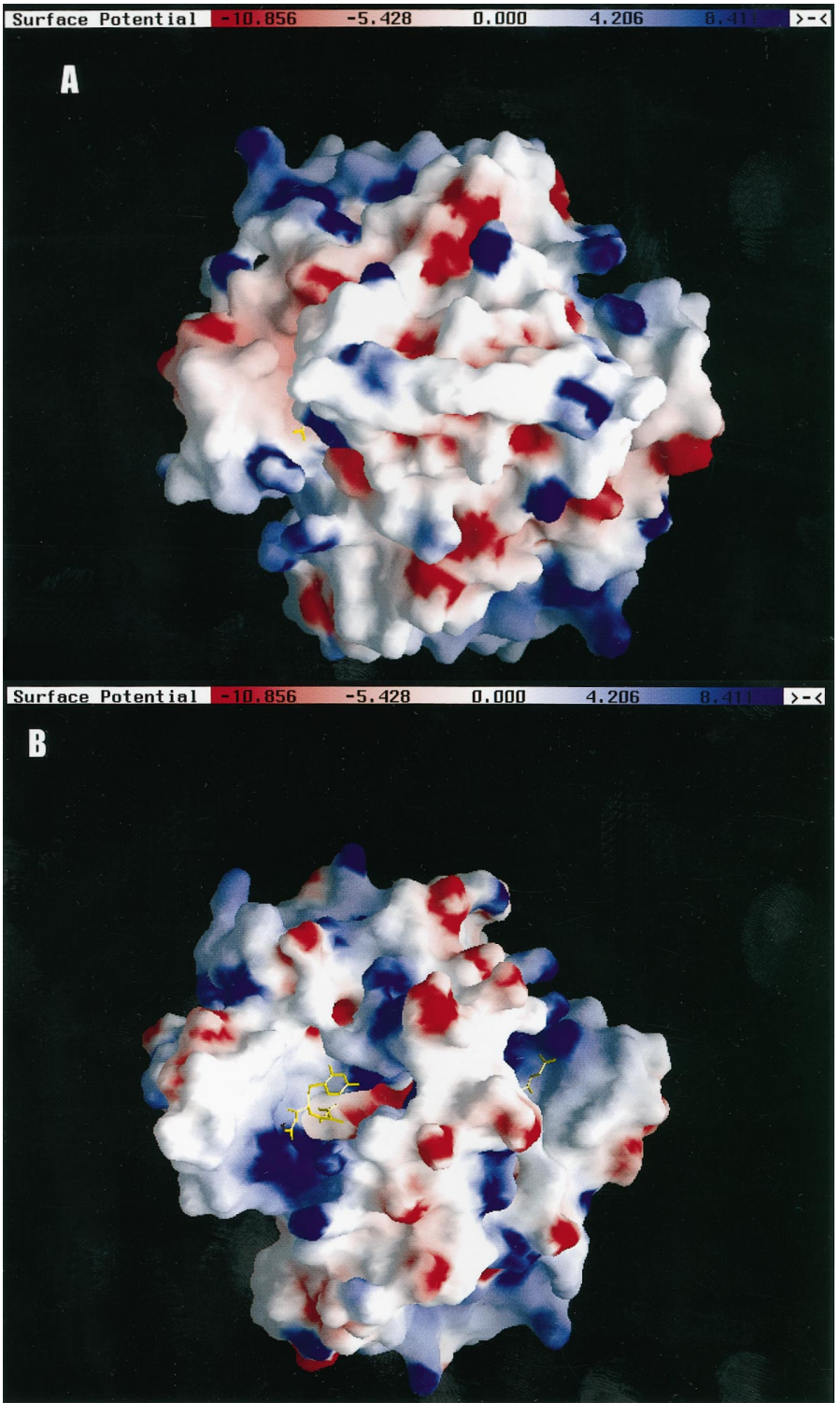

FIGURE 7 Comparison of surface electric potential of the dimeric *S*-crystallin and GST-σ. Both *S*-crystallin (*A*) and GST-σ (*B*) are in the same orientation and complexed with *S*-(2,4-dinitrophenyl) glutathione (*yellow*). Areas of positive potential are blue and those with negative potential are red. The shielded active site region in *S*-crystallin is clearly shown by the buried *S*-(2,4-dinitrophenyl) glutathione in *A*. The diagram was generated by the program GRASP (Nicholls et al., 1991).

Overall electrostatic surface potential map of *S*-crystallin and explanation for low affinity of *S*-crystallin to the GSH-Sepharose affinity column

The above discussion suggests that the diminished GST activity of cephalopods *S*-crystallin is caused by the loss of charge stabilization of the Meisenheimer complex. This argument is further supported by the comparison of surface electrical potential near the active-site region between *S*-crystallin and GST-σ (Fig. 6). Although the structure of *S*-crystallin shows a very similar overall topology to that of GST-σ (Fig. 2 *B*), the two molecules have a quite different map of the electrostatic surface potential, particularly at the active center. Two marked features are noticed in Fig. 6. First, in the active center region, the overall electrostatic surface potential is more positive (blue color) in GST-σ but

slightly negative (slightly pink color) in *S*-crystallin. Quite different electrostatic surface potential are observed around Arg14, the major residue responsible for the positively charged environment in the active site region (circled in Fig. 6). Second, because of the extra loop inserted between α4 and α5 in *S*-crystallin, the α4-turn-α5 motif appears to become an α4-loop-α5 element. We noticed an additional feature of this loop, which shields part of the active center, making the active site region in a narrow cleft and less accessible to substrate. This situation is even more prominent after dimerization (Fig. 7). As shown in Fig. 7 *A*, the bound *S*-(DNP)GS is almost invisible in the dimeric form of *S*-crystallin. The lack of binding affinity of *S*-crystallin to the immobilized *S*-hexylglutathione therefore could be simply owing to steric hindrance caused by an inaccessible G-binding site. Because *S*-crystallin still possesses residual GST activity, it suggests that free GSH must be able to diffuse into the active center. It should be noted that modeling of large insertions is usually not very precise. Confirmation of the above conjecture must await the experimental evidence to show that *S*-crystallin is able to bind with the GSH-affinity column if longer spacer is introduced between GSH and the matrix.

The above modeling and comparison of various GSTs indicate that the cephalopod *S*-crystallin is a natural mutant of the important detoxification enzyme, glutathione transferase. Shoichet et al. (1995) have suggested that amino acid residues involved in function are not optimized for conformational stability, and there is an inverse relationship between the effect of mutations on stability and catalysis. Our present work seems to support this prediction. Lens is an organ of closed, limited hydrated cells. In vertebrates, the crystallins inside the lens do not turnover during the life span (de Jong et al., 1989). Therefore crystallins must have a relatively stable structure against oxidation or other stresses and can survive at high concentration. The imaging system of cephalopods has some similarities to those of vertebrates (Tomarev and Piatigorsky, 1996). Although many lens proteins are related to stress and may have other functions besides their structural roles in reflecting and focusing light to the retina, conformational stability of crystallins is still a major concern in lens research. Insertion of a surface loop is one of the most efficient means to modify a protein without introducing instability to the protein (El Hawrani et al., 1994). Furthermore, a mutation of Asn to Asp at position 101 in *S*-crystallin creates an extra ion pair Asp 101-Arg14 that links α1 of domain I and α4 of domain II, which might also result in a more stable protein than the original GST enzyme. Our work seems to suggest that when recruiting cytosolic GST as one of the lens structural proteins during evolution, some mutations have occurred that may endow the protein with better conformational stability at the expense of enzymatic activity. However, this speculation is based upon modeling results and additional evidence is needed to test the hypothesis.

Conclusion and future perspectives

The discovery of enzyme crystallins in various taxons of disparate species is part of the allure of crystallin research. Of vital importance and no less interesting in a structural analysis of unique crystallins with some conventional or exotic enzymatic functions is knowledge of their primary structures achieved by the modern molecular cloning and sequencing. Only after such information is obtained would it be possible to assign functional roles for the individual crystallin.

In this study, we constructed a possible three-dimensional structure for the *S*-crystallin. Our model provides a plausible explanation for the less GST activity of *S*-crystallin and its lack of binding capacity to the GSH affinity column. The salient feature of our work includes proposing several experimentally testable ideas for further delineating the structure/function relationships of GST. For example, the functional roles of Arg14, Asp101, His108, and/or other critical amino acid residues in the GST-catalyzed reaction can be further elucidated by site-specific mutagenesis and crystallographic analyses. The biological significance of these dual-function crystallins may eventually be understood through the structural comparison of crystallins and their corresponding enzymes based on both the available x-ray structures of enzymes and the homology modeling of various crystallins with known sequences.

This work was supported by the National Science Council, Taipei, Taiwan, Republic of China (Grants NSC 86–2311-B002–032-B15 to S.-H. Wu; NSC 84–2311-B001–050-BA and 86–2311-B002–031-B15 to S.-H. Chiou; and NSC 86–2311-B016–001-B15 to G.-G. Chang).

REFERENCES

Armstrong, R. N. 1994. Glutathione *S*-transferases: structure and mechanism of an archetypical detoxication enzyme. *Adv. Enzymol. Relat. Areas Mol. Biol.* 69:1–44.

Armstrong, R. N. 1997. Structure, catalytic mechanism, and evolution of the glutathione transferase. *Chem. Res. Toxicol.* 10:2–18.

Atkins, W. M., R. W. Wang, A. W. Bird, D. J. Newton, and A. Y. Lu. 1993. The catalytic mechanism of glutathione *S*-transferase (GST): spectroscopic determination of the pK_a of Tyr-9 in rat alpha 1–1 GST. *J. Biol. Chem.* 268:19188–19191.

Barycki, J. J., and R. F. Colman. 1993. Affinity labeling of glutathione *S*-transferase, isozyme 4–4, by 4-(fluorosulfonyl)benzoic acid reveals Tyr115 to be an important determinant of xenobiotic substrate specificity. *Biochemistry.* 32:13002–13011.

Björnestedt, R., G. Stenberg, M. Widersten, P. G. Board, I. Sinning, T. A. Jones, and B. Mannervik. 1995. Functional significance of arginine 15 in the active site of human class alpha glutathione transferase A1–1. *J. Mol. Biol.* 247:765–773.

Bower, M. J., F. E. Cohen, and R. L. Dunbrack, Jr. 1997. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modeling tool. *J. Mol. Biol.* 267:1268–1282.

Bowie, J. U., R. Luthy, and D. Eisenberg. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science.* 253:164–170.

Chelvanayagam, G., M. C. J. Wilce, M. W. Parker, K. L. Tan, and P. G. Board. 1997. Homology model for the human GSTT2 theta class glutathione transferase. *Proteins.* 27:118–130.

Chiou, S. H. 1984. Physicochemical characterization of a crystallin from the squid lens and its comparison with vertebrate lens crystallins. *J. Biochem. (Tokyo)* 95:75–82.

Chiou, S. H., C. W. Yu, C. W. Lin, F. M. Pan, S. F. Lu, H. J. Lee, and G. G. Chang. 1995. Octopus *S*-crystallins with endogenous glutathione *S*-transferase: cDNA sequence determination, expression and comparison of structure and kinetic mechanism with authentic enzymes. *Biochem. J.* 309:793–800.

De Jong, W. W., W. Hendriks, J. W. M. Mulders, and H. Bloemendal. 1989. Evolution of eye lens crystallins: the stress connection. *Trends Biochem. Sci.* 14:365–368.

Dirr, H., P. Reinemer, and R. Huber. 1994a. Refined crystal structure of porcine class Pi glutathione *S*-transferase (pGST P1–1) at 2.1 Å resolution. *J. Mol. Biol.* 243:72–92.

Dirr, H., P. Reinemer, and R. Huber. 1994b. X-ray crystal structures of cytosolic glutathione *S*-transferases: implications for protein architecture, substrate recognition and catalytic function. *Eur. J. Biochem.* 220:645–661.

Du, P., J. R. Collins, and G. H. Loew. 1992. Homology modeling of a heme protein, ligin peroxidase, from the crystal structure of cytochrome-c peroxidase. *Protein Eng.* 5:679–691.

Dunbrack, R. L., Jr., and M. L. Karplus. 1994. Conformational analysis of the backbone-dependent rotamer preferences of protein side-chains. *Nat. Struct. Biol.* 1:334–340.

El Hawrani, A. S., K. M. Moreton, R. B. Sessions, A. R. Clarke, and J. J. Holbrook. 1994. Engineering surface loops of proteins: a preferred strategy for obtaining new enzyme function. *Trends Biotechnol.* 12: 207–211.

Fischer, D., and D. Eisenberg. 1996. Protein fold recognition using sequence-derived predictions. *Protein Sci.* 5:947–955.

Genetics Computer Group. 1994. Program Manual for the Wisconsin Package, Version 8. Madison, Wisconsin.

Gilliland, G. L. 1993. Glutathione proteins. *Curr. Opin. Struct. Biol.* 3:875–884.

Hayes, J. D., and D. J. Pulford. 1995. The glutathione *S*-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit. Rev. Biochem. Mol. Biol.* 30:445–600.

Hobohm, U., and C. Sander. 1994. Enlarged representative set of protein structures. *Protein Sci.* 3:522–524.

Ji, X., R. N. Armstrong, and G. L. Gilliland. 1993. Snapshots along the reaction coordinate of an S_NAr reaction catalyzed by glutathione transferase. *Biochemistry.* 32:12949–12954.

Ji, X., E. C. von Rosenvinge, W. W. Johnson, S. I. Tomarev, J. Piatigorsky, R. N. Armstrong, and G. L. Gilliland. 1995. Three-dimensional structure, catalytic properties, and evolution of a sigma class glutathione transferase from squid, a progenitor of the lens *S*-crystallins of cephalopods. *Biochemistry.* 34:5317–5328.

Ji, X., P. Zhang, R. N. Armstrong, and G. L. Gilliland. 1992. The three-dimensional structure of a glutathione *S*-transferase from the mu gene class: structural analysis of the binary complex of isoenzyme 3–3 and glutathione at 2.2 Å resolution. *Biochemistry.* 31:10169–10184.

Johnson, M. S., N. Srinivasan, R. Sowdhamini, and T. L. Blundell. 1994. Knowledge-based protein modeling. *Crit. Rev. Biochem. Mol. Biol.* 29:1–68.

Johnson, W. W., S. Liu, X. Ji, G. L. Gilliland, and R. N. Armstrong. 1993. Tyrosine 115 participates both in chemical and physical steps of the catalytic mechanism of a glutathione *S*-transferase. *J. Biol. Chem.* 268: 11508–11511.

Karadi, R., M. Billiter, and K. Wüthrich. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics.* 14:51–55.

Koehler, R. T., H. O. Villar, K. E. Bauer, and D. L. Higgins. 1997. Ligand-based protein alignment and isozyme specificity of glutathione *S*-transferase inhibitors. *Proteins.* 28:202–216.

Laskowski, R. A., M. W. MacArthur, D. J. Moss, and J. M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26:283–291.

Lemer, M. C. R., M. J. Rooman, and S. J. Wodak. 1995. Protein structure prediction by threading methods: evaluation of current techniques. *Proteins.* 23:337–355.

Lin, C. W., and S. H. Chiou. 1992. Facile cloning and sequencing of *S*-crystallin genes from octopus lenses based on polymerase chain reaction. *Biochem. Int.* 27:173–178.

Liu, S., P. Zhang, X. Ji, W. W. Johnson, G. L. Gilliland, and R. N. Armstrong. 1992. Contribution of tyrosine 6 to the catalytic mechanism of isoenzyme 3–3 of glutathione *S*-transferase. *J. Biol. Chem.* 267: 4296–4299.

Lo Bello, M., A. J. Oakley, A. Battistoni, A. P. Mazzetti, M. Nuccetelli, G. Mazzarese, J. Rossjohn, M. W. Parker, and G. Ricci. 1997. Multifunctional role of Tyr 108 in the catalytic mechanism of human glutathione transferase P1–1: crystallographic and kinetic studies on the Y108F mutant enzyme. *Biochemistry.* 36:6207–6217.

Luthy, R., J. U. Bowie, and D. Eisenberg. 1992. Assessment of protein models with three-dimensional profiles. *Nature.* 356:83–85.

Mannervik, B., and U. H. Danielson. 1988. Glutathione transferases: structure and catalytic activity. *Crit. Rev. Biochem.* 23:283–337.

Marsh, A., and D. M. Ferguson. 1997. Knowledge-based modeling of a bacterial dichloromethane dehalogenase. *Proteins.* 28:217–226.

Meyer, D. J., C. Xia, B. Coles, H. Chen, P. Reinemer, R. Huber, and B. Ketterer. 1993. Unusual reactivity of Tyr-7 of GSH transferase P1–1. *Biochem. J.* 293:351–356.

Needleman, S. B., and C. D. Wunsch. 1970. A general method applicable to search for similarities in amino acid sequence of two proteins. *J. Mol. Biol.* 48:443–453.

Nicholls, A., K. Sharp, and B. Honing. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins.* 11:281–286.

Orozco, M., C. Vega, A. Parraga, I. Garcia-Saez, M. Coll, S. Walsh, T. J. Mantle, and F. J. Luque. 1997. On the reaction mechanism of class Pi glutathione *S*-transferase. *Proteins.* 28:530–542.

Pearlman, D. A., D. A. Case, J. W. Caldwell, W. S. Ross, T. Cheatham, D. M. Ferguson, G. Seibel, U. C. Singh, P. Weiner, and P. A. Kollman. 1995. AMBER 4.1. University of California, San Francisco.

Piatigorsky, J. 1992. Lens crystallins: innovation associated with changes in gene regulation. *J. Biol. Chem.* 267:4277–4280.

Piatigorsky, J., and G. Wistow. 1991. The recruitment of crystallins: new functions precede gene duplication. *Science.* 252:1078–1079.

Prade, L., R. Huber, T. H. Manoharan, W. E. Fahl, and W. Reuter. 1997. Structures of class pi glutathione *S*-transferase from human placenta in complex with substrate, transition-state analogue and inhibitor. *Structure.* 5:1287–1295.

Rushmore, T. H., and C. B. Pickett. 1993. Glutathione *S*-transferases, structure, regulation, and therapeutic implications. *J. Biol. Chem.* 268: 11475–11478.

Shoichet, B. K., W. A. Baase, R. Kuroki, and B. W. Matthews. 1995. A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. USA.* 92:452–456.

Sinning, I., G. J. Kleywegt, S. W. Cowan, P. Reinemer, H. W. Dirr, R. Huber, G. L. Gilliland, R. N. Armstrong, X. Ji, P. G. Board, B. Olin, B. Mannervik, and T. A. Jones. 1993. Structure determination and refinement of human alpha class glutathione transferase A1–1, and a comparison with the Mu and Pi class enzymes. *J. Mol. Biol.* 232:192–212.

Sippl, M. J. 1993. Recognition of errors in three-dimensional structures of proteins. *Proteins.* 17:355–362.

Smith, T. F., and M. S. Waterman. 1981. Identification of common molecular subsequences. *J. Mol. Biol.* 147:195–197.

Sternberg, M. J. E., editor. 1996. Protein Structure Prediction: A Practical Approach. Oxford University Press, Oxford.

Summers, N. L., and M. L. Karplus. 1990. Modeling of globular proteins. *J. Mol. Biol.* 216:991–1016.

Tang, S. S., and G. G. Chang. 1995. Steady-state kinetics and chemical mechanism of octopus hepatopancreatic glutathione transferase. *Biochem. J.* 309:347–353.

Tang, S. S., and G. G. Chang. 1996. Kinetic characterization of the endogenous glutathione transferase activity of octopus lens *S*-crystallin. *J. Biochem. (Tokyo).* 119:1182–1188.

Tang, S. S., C. C. Lin, and G. G. Chang. 1994. Isolation and characterization of octopus pancreatic glutathione *S*-transferase: comparison with lens *S*-crystallin. *J. Protein Chem.* 13:609–618.

Tomarev, S. I., S. Chung, and J. Piatigorsky. 1995. Glutathione *S*-transferase and *S*-crystallin of cephalopods: evolution from active enzyme to lens refractive protein. *J. Mol. Evol.* 41:1048–1056.

Tomarev, S. I., and J. Piatigorsky. 1996. Lens crystallins of invertebrates: diversity and recruitment from detoxification enzymes and novel proteins. *Eur. J. Biochem.* 235:449–465.

Tomarev, S. I., R. D. Zinovieva, and J. Piatigorsky. 1992. Characterization of squid crystallin genes: comparison with mammalian glutathione *S*-transferase genes. *J. Biol. Chem.* 267:8604–8612.

Vinals, C., X. De Bolle, E. Depiereux, and E. Feytmans. 1995. Knowledge-based modeling of the C-lactate dehydrogenase three-dimensional structure. *Proteins.* 21:307–318.

Wilce, M. C., P. G. Board, S. C. Feil, and M. W. Parker. 1995. Crystal structure of a theta-class glutathione transferase. *EMBO J.* 14:2133–2143.

Wilce, M. C., and M. W. Parker. 1994. Structure and function of glutathione *S*-transferases. *Biochim. Biophys. Acta.* 1205:1–18.

Wilmanns, M., and D. Eisenberg. 1995. Inverted protein folding by the residue pair preference profile method. *Protein Eng.* 8:627–639.