Amino acid sequencing, molecular cloning and modelling of the chick liver class-Theta glutathione S-transferase CL1

Chwan-Deng HSIAO,* Elena O. MARTSEN,† Jing-Yu LEE, Shu-Ping TSAI and Ming F. TAM*

Institute of Molecular Biology, Academia Sinica, Taipei 11529, Taiwan R.O.C.

Glutathione S-transferase CL1–2 heterodimers purified from 1-day-old chick livers were digested with *Achromobacter* proteinase I. The resulting fragments were separated for amino acid sequence analysis. Oligonucleotide probes were constructed based on sequence similarity to class-Theta glutathione S-transferases for PCR using a chicken liver cDNA library as template. A full-length clone (1725 bp) encoding a polypeptide comprising 261

amino acids was isolated. Including conservative substitutions, this protein has 70–73 % sequence similarity with other mammalian class-Theta glutathione S-transferases. Based on known X-ray crystal structures of class-Alpha, -Mu and -Pi glutathione S-transferases, a model is constructed for the N-terminal 232 residues of CL1.

INTRODUCTION

Glutathione S-transferases (GSTs; EC 2.5.1.18) comprise a large group of multifunctional enzymes that catalyse the conjugation of glutathione to a wide variety of electrophiles and xenobiotics [1–3]. Besides their detoxification roles, these dimeric proteins are also involved in the reduction of organic hydroperoxides [4], the isomerization of prostaglandins [5] and the binding of non-substrate hydrophobic ligands such as bile acids, bilirubin, a number of drugs and thyroid hormones [2,6].

The rat, human and mouse GSTs have been extensively investigated at the individual enzyme level. On the basis of their physical, chemical, immunological, enzymic and structural properties, GSTs were conventionally grouped into three non-homologous multigene families, namely Alpha, Mu and Pi [7]. Evidence for the existence and the purification of a new class-Theta GST from rat and human hepatic cytosol was provided by Hiratsuka et al. [8], Meyer et al. [9] and Hussey and Hayes [10]. These class-Theta GSTs can utilize sulphate esters and dichloromethane as substrates, and are suggested to be important in the prevention of hepatocarcinogenesis [8]. Rat and human class-Theta GSTs have unique properties. They are not retainable on *S*-hexylglutathione affinity matrices and have low reactivity towards 1-chloro-2,4-dinitrobenzene (CDNB), the typical GST substrate.

Previously, we reported the characterization of cytosolic GSTs from chick livers [11]. Five groups of GST subunits with molecular masses ranging from 24 to 27 kDa were identified. These GST subunits were designated CL1–CL5 according to their electrophoretic mobility on a denaturing polyacrylamide gel. On the basis of immuno-cross-reactivity, substrate specificity and N-terminal sequencing data, CL1 and CL2 are related to the class-Theta and -Mu GSTs respectively. Unlike mammalian class-Theta GSTs, the avian CL1 isoenzyme was isolated with an *S*-hexylglutathione affinity column [11]. Results from isoelectrofocusing and denaturing gel electrophoresis indicate that CL1 can exist as CL1–2 heterodimers [11]. The sequence of a CL2 cDNA clone has been reported [12].

The three-dimensional structures of class-Alpha [13], -Mu [14]

and -Pi [15] homodimers have been elucidated. Even though these proteins have only 20–32 % sequence identity, they share a two-domain structure, a smaller α/β Domain I and a larger α Domain II. Domain I can be considered as the glutathione binding domain, while Domain II seems to be the primary binding site for the xenobiotic substrates.

In this paper, we report the cloning and the nucleotide sequence of a cDNA coding for CL1, a class-Theta GST from chicken liver. The CL1 GST, with 260 amino acids, is at least 17 residues longer than other class-Theta GSTs reported [16]. Based on sequence similarity with other classes of GSTs, a model is constructed for the N-terminal 232 residues of CL1. The structure of CL1 has overall similarity with those of other classes of GSTs.

MATERIALS AND METHODS

Materials

Male white Leghorn chicks were obtained from Taiwan University Experimental Station and killed 1 day after hatching. The chicken liver cDNA library was obtained from Stratagene Co. (La Jolla, CA, U.S.A.). Mono P columns, Ampholytes and Polybuffers were from Pharmacia (Uppsala, Sweden). *Achromobacter* proteinase I was purchased from Wako Chemicals (Osaka, Japan). The Sequenase kit was obtained from United States Biochemical Corp. (Cleveland, OH, U.S.A.). [α - 32 P]dCTP and [α - 35 S]thiol[dATP] were obtained from New England Nuclear (Wilmington, DE, U.S.A.). Taq I polymerase was purchased from Promega Corp. (Madison, WI, U.S.A.). Chemicals for protein sequencing were obtained from Applied Biosystems Inc. (Foster City, CA, U.S.A.). All other chemicals used were reagent grade or better.

Protein and peptide purification

GST CL1–2 subunits were purified by affinity chromatography and chromatofocusing according to Chang et al. [11]. The CL1–2 dimers were separated on a Brownlee Aquapore reversed-phase column (C_8 ; 0.46 cm $\times$ 22 cm) equilibrated in 0.08 % (v/v) trifluoroacetic acid/water. Elution was carried out at a flow rate of

Abbreviations used: GST, glutathione S-transferase; CDNB, 1-chloro-2,4-dinitrobenzene.

* To whom correspondence should be addressed.

† Present address: Tampa Bay Research Institute, 10900 Roosevelt Blvd., St. Petersburg, FL 33716, U.S.A.

The nucleotide sequence data reported have been submitted to the GenBank/EMBL/DBJ Nucleotide Sequence Databases under accession no. U13676.

1 ml/min with a linear gradient of 1% per min of 0.05% (v/v) trifluoroacetic acid in acetonitrile. The elution profile of the proteins was monitored by UV absorption at 220 nm. The CL1 subunits were collected and subjected to automated cycles of Edman degradation on an ABI gas/liquid-phase model 470A/900A sequencer equipped with an on-line model 120A phenylthiohydantoin analyser [17].

Alternatively, the CL1–2 dimers were digested directly with *Achromobacter* proteinase I (16 h at 25 °C) in 0.01 M Tris/HCl, pH 9.6, and 100 mM KCl. Digestion was carried out at a substrate/enzyme ratio of 100:1 (w/w). Resulting peptides for sequence analysis were separated on a Vydac reversed-phase column (C₁₈; 0.21 cm × 25 cm). Elution was carried out at a flow rate of 0.2 ml/min with a linear gradient of 0.5% per min of 0.05% (v/v) trifluoroacetic acid in 80% (v/v) acetonitrile (buffer B).

Cloning of CL1

The chicken liver Lambda Zap library was screened by the PCR method of Amaravadi and King [18] with approximately 1×10^6 clones. The sense primer, 5'-ATGGGGCTGGAGCTCTA-TCTG-3', was constructed according to the published sequence of the rat GST 12 subunit, coding for the N-terminal heptapeptide of the protein [16]. Two antisense primers were constructed according to the experimentally determined internal sequence of CL1. The C-terminal antisense primer, 5'-(A/G/T)AT(G/A)-AT(G/A)GG(C/T)TT(C/A)TC(C/T)TG(T/C)AG(A/G)AA-3', was deduced from a peptide that has high sequence similarity to residues 150–157 of rat GST 5 [19]. An additional antisense nested primer, 5'-CGG(A/G)TACCA(A/G)TG(A/G)TCAGG-3', was constructed according to a CL1 hexapeptide that has sequence identical to residues 80–85 of GST 5 [19]. Amplification was for 35 cycles under the following conditions: 94 °C for 1 min, 45 °C for 1 min and 72 °C for 1 min. The PCR products were analysed electrophoretically on a 1.5% agarose gel. Positive clones should yield DNA fragments of approx. 500 bp. Plates containing positive clones were replated at 300–500 plaque-forming units/plate for secondary screening. PCR products from positive clones of secondary screening were reamplified with the N-terminal and the nested primers to eliminate false-positives. Plates containing positive clones after the tertiary screening were plated at 50–100 plaque-forming units/plate and individual plaques were analysed by PCR.

Plasmid DNA from positive plaques was rescued from the Lambda Zap vector according to the manufacturer's instructions. The DNA sequence of the insert was determined by the dideoxy sequencing method of Sanger et al. [20] with synthetic oligonucleotides as primers.

Model construction

The amino acid sequence of CL1 was aligned with those of class Alpha, -Mu and -Pi GSTs using the multiple-sequence alignment algorithm of Feng and Doolittle [21]. Sequence-conserved regions were located under the QUANTA environment on a Silicon Graphics IRIS 4D/25 workstation. The protein model assembled for CL1 is based on the 2.2 Å resolution X-ray structure of GST 3-3 co-crystallized with glutathione [14]. The main-chain coordinates of CL1 were copied from GST 3 under the QUANTA environment, and several gap regions were determined by a fragment-search process and manually adjusted. The 'regularization modelling' technique in QUANTA was employed to build and relax the side-chain co-ordinates and the undefined terminal residues. The three-dimensional structures of GSTs from different families are highly variable in the C-terminal

region. The C-terminal region of CL1 (residues 233–260) was deleted from the final model to increase the reliability of the contrived structure. The CL1 model was obtained by 200 cycles of the 'region regularization' tool using the 'adopted basis set NR' protocol in both minimization stages. These processes removed close contacts and reduced strains among side chains. The intramolecular hydrogen bonds and hydrophobic interactions of the model were optimized by theoretically heating the resulting model to 300 K with an equilibration time of 0.6 ps (600 steps). The final model of the CL1 subunit was obtained by applying the CHARMM energy-minimization process with the CHARMM/QUANTA program package [22]. The program ran through 100 cycles of steepest descent minimization followed by 200 cycles of Powell conjugate gradient minimization and 300 cycles of conjugate gradient minimization.

RESULTS AND DISCUSSION

Amino acid sequencing of CL1

GST CL1–2 subunits were isolated by a combination of affinity and chromatofocusing chromatography [11]. The subunits were further separated on a reversed-phase column and the N-terminal sequence of the CL1 subunit was determined. A sequence of GLELYDLLSQPSRAVYIFARSNSLPFE/KHVSMLMKE was obtained for the CL1 subunit. The signal for cycle 29 was too weak for a definite identification of the residue. Including conservative substitutions, the N-terminal region of CL1 has 87% sequence similarity (33/38 residues) with that of rat GST 12.

CL1–2 subunits were digested with an endopeptidase (*Achromobacter* proteinase I). The resulting fragments were separated by HPLC on a reversed-phase column (Figure 1) and collected for N-terminal sequence analysis. The sequences of the selected peptides were: f1, IAPELREHFK; f2, TPDHWYPSDLQK; f3, LTQSNAILRYIARK; f4, VE/ATEELNVVLSQFEK; f5, LGLDFPNLPYLIDGDVK; f6, FLQDKPFIVGNDISLALV. Peptides f3 and f5 are derived from the class-Mu GSTs, covering residues 69–82 and residues 52–68 of the CL2 subunit [12] respectively. Peptides f2 and f6 are derived from the class-Theta GSTs. Peptide f2 has 75 and 66% sequence similarity, and peptide f6 has 79 and 74% sequence similarity, with GST 5 and GST 12 respectively. Peptides f1 and f4 do not have similarity with any known GST sequence.

Figure 1 HPLC profile of *Achromobacter* proteinase I digests of GST CL1–2 isolated from 1-day-old chick liver

Fractions isolated for sequence determination are indicated (f1–f6).

GAATTCGACGGAGCGTCCCCGCGGAGCCGGGCGGCGGCAGC	ATG GGG CTG GAG CTG TAC CTG GAT	66
	<u>M G L E L Y L D</u>	
CTG CTC TCG CAG CCC TGC CGC TCC ATC TAC ATC TTC GCC CGG ACC AAC AAC ATC CCG		123
<u>L L S O P C R S I Y I F A R T N N I P</u>		
TTC GAG TTC AAG CAC GTG GAG CTC TTC AAA GAC TCG GTG CTG GGG AAG AAG CCG GCG		180
<u>F E F K H V E L F K D S</u>		
GCG GCG AGC GGC GCG GAG CGG CCC CGC ACA GGG CCA TCA AAT AGT GAA GGG GAT GGC		237
<u>A A S G A E R P R T G P S N S E G D G</u>		
AAA ATC AGC CTC TTG AAG AAA GTA CCA GTA CTG AAG GAC GGA GAC TTC ACC CTA GCA		294
<u>K I S L L K K V P V L K D G D F T L A</u>		
GAA TGC ACT GCC ATT CTG CTG TAT CTG AGT CGA AAG TAC AAC ACT CCT GAC CAC TGG		351
<u>E C T A I L L Y L S R K Y N T P D H W</u>		
TAC CCA TCA GAC ATA AAA AAA CGG GCC CAG GTA GAT GAA TAC CTC TCG TGG CAT CAT		408
<u>Y P S D I K K R A Q V D E Y L S W H H</u>		
GCT AAC ATC CGG GCT AAT GCT CCT AAG ACC ATG TGG ATC AAG GTG CTG ATT CCC CTC		465
<u>A N I R A N A P K T M W I K V L I P L</u>		
TTC ACA GGG CAA CCA CAG CCA TCA GAG AAG CTG CAG GAG GTT ATG GAG GGG CTG TCC		522
<u>F T G Q P Q P S E K L Q E V M E G L S</u>		
ACT TCC CTG AAG CAA TTT GAG GAG AGA TTT CTG CAG GAC AAG GCT TTT ATC ATC GGG		579
<u>T S L K Q F E E R F L O D K A F I I G</u>		
AGC GAG ATT TCC TTG GCA GAT CTT GTG GCC ATT GTG GAG CTG ATG CAA CCT GTT GGA		636
<u>S E I S L A D L V A I V E L M Q P V G</u>		
GTT GGT TGT GAC ATC TTT GAA GAC AGG CCC AGG TTG ATG GAG TGG CGC AGG CGG GTG		693
<u>V G C D I F E D R P R L M E W R R R V</u>		
GAG GAA GCT GTG GGG AAG GAG CTT TTC TTC CAA GCC CAT GAG ATG ATC CTG AAT ATC		750
<u>E E A V G K E L F F Q A H E M I L N I</u>		
AAA GAA CTG AGC AAC ATT CAG ATT GAT CCA CAG CTG AAA GAG CAC CTG GCA CCT GTG		807
<u>K E L S N I Q I D P Q L K E H L A P V</u>		
TTG ATG AAG ATG TTG AAA TGA ATTAACCTATATTATTTTTTTCCTGCCCTATTATTATTTTTTAC		875
<u>L M K M L K *</u>		
TTCTGTGATCTCCAGCTCTACACCTAACTGGAAAAAACAACACTGCATTCTTACATAGTCTTGCAAGGTGTCC		950
TCCCCAGCTTCACTAAGCCAACCTGCATCTTTGGGGGGTGGGAGCTGGGGAAGTTTAAAGACCTGAAACATTAG		1025
ATTTGTATGAACGGGAGAGGCCCTATGGGTCCCTTAACACATGCATCTGGGTATATTTTAGTGTCGCTCAGAAA		1100
TGTTGGAAACATGAATAGTATTTCTGAAAAATTGGAGCCAGCCTGCAAACTGGAACATCAGAGATCTATCTAGG		1175
ATGAGGATCATTTGTTTCATGTGTTGTAACATTTGTAGCTCCTTCTTAAGTGGCAAGCACAGCAAAGCATACCT		1250
GATCCTCACTTCCCCAGCATGCGTCTCTGCATCAGTGAGTATGGAACAAAGCTACAGTAGTCACTAAAGTAAAAA		1325
GAAGACTTTTATTTTATTTTATTTTAAATTTAATTTCTCGTTGCTGGAGCACATTGCTTGGCCACCCATCTAC		1400
ATTCCATACCTCTGTTTGCCCTTAAAGATCCTGTGAGAAGTTTCTTTACATTCATCTATGATTGTAATCATTGCA		1475
ATTCTTTAACCACTGCTCTCCACTTTTCAGAGCTGGGTTTCTTAGGCATGTGGTCTGTGAGTACACGTGTGGTGG		1550
GGCAGTTCTTGGGAGATCTGTGT <u>AATAA</u> AGACCGAGCAACTTCAACACAGCTGAGGGTGGCCACAGCCTTCACATA		1625
ATTGCCACTTGAATATCCTCTGTTGATATTATACATGGAATACTCTGAAAAAATAAAAG (A)		1725

Figure 2 Nucleotide and deduced amino acid sequence of pGCLT-1

Regions of CL1 that have been sequenced by Edman degradation are underlined and in italics. Arrows indicate the direction and extent of the primers used in PCR screening. ‡ indicates a discrepancy between the experimentally determined amino acid sequence and that predicted from the cDNA sequence. The termination codon is denoted by an asterisk (*) and the polyadenylation signal is underlined.

	† †
--	---

Figure 3 Comparison of the deduced amino acid sequence of GST CL1 (pGCLT-1) with those of rat GST 5 [19], rat GST 12 [16], human GSTT1 [23], *M. domestica* GST1 [24], *Drosophila* GST1 [25] and *Lucilia cuprina* GST1 [26]

The amino acid sequences are given in single-letter code. Gaps were introduced for optimal alignment and are represented as dots. Dashes (—) indicate identity with amino acid residues of CL1. Chemically similar residues (I/L/M/V, K/H/R, D/E/N/Q, S/T, A/G and F/W/Y) [36] are labelled with †.

Cloning and cDNA sequence of CL1

Oligonucleotides were constructed according to the N-terminal sequence of rat GST 12 (peptides f2 and f6) for library screening.

Three separate clones were isolated from primary screening. After further purification, inserts from the lambda phages were subcloned into a pBluescript SK⁺ phagemid. Restriction mapping and sequencing analysis indicated that these three clones are

CL1	1	 <u>GLELYDL</u> <u>LLSQPCR</u> SIYIFARTNNIPFEFKHVEL.FKDSVLGKKPAAASG.....AERPRTGPSNSEGD	63
Alpha	2	AEKP_KLH.YFNA.....RG.....RMESTRWLLAAAGVEFEKFIKSAEDL.....DKLRNDG.....	48
Mu	1	... <u>PTLG</u> .YWNV.....RG.....LTHPIRLILEXTDSSYEKRYAMGDAPDYDRSOWLNEKFKLG.....	53
Pi	1	..PPYTTT.YFPV.....RG.....RCEAMRMLLADODOSWKEEVVTMET.....W...PPLKPS.....	44
		β_1 α_1 β_2 α_2	
CL1		GKISLLKKVPVLKDGDFTLAECTAIIILYLSRKYNTPDHWYPSDIKKRAOVDEYLSWHHANIRANAPKTMWIKVL	137
Alpha		..YLMFOQVPMVEIDGMKLVOTRAILNYIASKY...NLYGKDIKERALIDMYIE...GIADLGEMILLPVC	112
Mu		...LDFPNLPYLIDGSRKITOSNAIMRYLARKH...HLCGETEERIRADIVEN...OVMNRMOLIML.CY	115
Pi		...CLFRQLPKFODGDLTLYOSNAILRHLGRSF...GLYGKDKOKEAALVDMVND...GVEDLRCKYATL.IY	106
		β_3 β_4 α_3 α_4	
CL1		IPLFTGQPOP.. <u>SEKLOEV</u> .MEGLSTSLKQFEERFLQDKAFIIGSEISLAD..LVAIVELMQPVGVC.DIFED	205
Alpha		PPEEKDAKLALIKEKIKNRYFPFAFEKVLKSH.....QDYLVGNKLSRADIHLVELLYVVEFLDSSLISF..	178
Mu		NPDEEKOKPEFLKT.IPEK.MKLYSEFL...G...KRPWFAGDKVTYVDFLAYDILDYOYHIFEPKCLDAF..	177
Pi		T.NYEAGKEKYKE.LPEH.LKPFETLLSONQG...GQAFVVGSSQISFADYNLLDLRLIHOVLNPSCLDAF..	171
		α_5 α_6	
CL1		RPRMEWRRRVEEAVGKELFFQAHEMILNIKELSNIQIDPQLKEHLAPVLMKMLK	260
Alpha		.PLLKALKTRISNLPTVKKFLQPGSPR.KPPMDEKSLEEARKIFRF	222
Mu		.PNLKDFLARFEGLKKISAYMKSSRYL.STPIFSKLAQWSNK	217
Pi		.PLL SAYVARLSARPKIKAFLASPEHV.NRPIINGNGKQ	207
		α_7 α_8 β_5 α_9	

Figure 4 Comparison of the deduced amino acid sequence of GST CL1 (pGCLT-1) with those of class-Alpha [13], -Mu [14] and -Pi [15] GSTs

The amino acid residues of class-Alpha GSTs are numbered according to Sinning et al. [13]. The amino acid sequence is given in single-letter code. * denotes residues conserved among all four classes of GSTs. The secondary structures of class-Alpha, -Mu and -Pi GSTs and the predicted secondary structures of the CL1 model are underlined and specified.

identical, and they were designated pGCLT-1. A total of 11 synthetic primers were used for sequencing both strands of these inserts with overlapping regions of at least 50 nucleotides.

The nucleotide sequence of pGCLT-1 and the deduced amino acid sequence are shown in Figure 2. The direction and extent of the primers used in PCR screening are also indicated. The insert of pGCLT-1 is 1725 nucleotide residues long, containing 42 nucleotide residues in the 5' non-coding region, a 262-codon open reading frame (molecular mass 29823 Da) and a complete (897 nucleotide residues) 3' non-coding sequence including the poly(A) addition signal.

The deduced amino acid sequence of pGCLT-1 does not match exactly with the results of peptide sequencing (Figure 2). The data imply that there are multiple class-Theta GSTs in chicken livers. A similar conclusion has been reached for rat [9] and human liver [10] class-Theta GSTs.

The deduced amino acid sequence of pGCLT-1 is listed together with other members of the Theta family for comparison in Figure 3. The sequences included are rat GST 5 [19], rat GST 12 [16], human GSTT1 [23], *Musta domestica* GST1 [24], *Drosophila* GST1 [25] and *L. cuprina* GST1 [26]. The class-Theta GSTs from plants [27–29] have only 21–24% sequence identity with CL1 and are not included for comparison. CL1 encoded by pGCLT-1 has a 19-amino-acid residue insert near the N-terminus. Consequently, it is 17–21 residues longer than mammalian class-Theta GSTs. CL1 has 50 and 54% sequence identity with rat GST 5 and human GSTT1 respectively. The similarity between avian and insect class-Theta GSTs is much less. CL1 is 51–52 residues longer than, and has 24% (*L. cuprina* GST1) to 26% (*M. domestica* GST1) sequence identity with insect GSTs. With conservative substitutions, the similarities between CL1

and insect class-Theta GST sequences are 49% (*M. domestica* GST1) to 53% (*L. cuprina* GST1).

Mammalian class-Theta GSTs are not retainable on a glutathione or S-hexylglutathione column [8–10,30]. Presumably, the affinity of these isoenzymes towards GSH or S-hexylglutathione is rather low. With the exception of Yd, from dog livers [30], these enzymes have low activity towards the typical GST substrate CDNB. Conversely, class-Theta GSTs from insects [24–26] and plants [27,31] were purified with glutathione affinity columns and they possess CDNB-conjugating activity. The structural elements responsible for these differences remain to be investigated.

Molecular modelling of GST CL1

The crystal structures of class-Alpha [13], -Mu [14,32] and -Pi [15,33] homodimeric GSTs have been elucidated. Even though the sequence identities among these three classes of GSTs are rather low, the overall folding of these isoenzymes is similar. These dimeric proteins are composed of two identical subunits, and each subunit contains a two-domain structure. Domain I has an α/β structure, consisting of four β -strands and three α -helices. Domain II, a larger α domain, has five or six α -helices (Figures 4 and 5). These two domains are connected by a short peptide segment. The dimer is held together by hydrogen bonds and van der Waals interactions between Domain I of one subunit and Domain II of the adjacent subunit.

The CL1 amino acid sequence was compared with class-Alpha [13], -Mu [14] and -Pi [15] GSTs using the multiple-sequence alignment algorithm [21] and optimized using the Holley secondary structure prediction method under the QUANTA en-

Figure 5 Comparison of the CL1 model with the X-ray crystallographic structures of class-Alfa [13], -Mu [14] and -Pi [15] GSTs

The ribbon drawing showing the comparison was generated using the Molscript program of Kraulis [37].

vironment (Figure 4). We identified 13 conserved residues among the four classes of GSTs. Sequence identities among these four classes of GSTs are rather low. Taking pairwise comparisons, the sequence identities are: 20% (Alpha-Mu), 32% (Alpha-Pi), 16% (Alpha-Theta), 30% (Mu-Pi), 16% (Mu-Theta) and 20% (Pi-Theta).

Three obvious differences were detected by aligning the sequence of class-Theta, -Alpha, -Mu and -Pi GSTs. The first variance is from residues 8 to 21 of CL1 (residues 11–14, 9–12 and 10–13 of class-Alfa, -Mu and -Pi GSTs respectively). Interestingly, there are three conserved residues (Tyr-5, Arg-14 and Arg-21 of CL1) located in the middle and at each end of this insertion. Arg-21 is conserved among class-Theta, -Alpha and

-Pi GSTs. Consequently, CL1 has a long loop in this region. The other difference in Domain I is from residues 44 to 67 of CL1 (residues 39–50, 36–54 and 37–45 of class-Alfa, -Mu and -Pi GSTs respectively). There is an extension in the loop connecting the $\beta 2$ and the $\alpha 2$ of the class-Mu enzyme. This loop is not observed in class-Alfa or -Pi or in the CL1 model. Within this region there is an insertion in the CL1 sequence (residues 57–66) that is unique for Theta isoenzymes. However, this is one of the two major structural variations that occurs in all the known crystal structures of GSTs (Figure 5). The residues are not involved in subunit interactions or GSH binding. The last major difference is located at the C-terminus of Domain II (residues 233–260 of CL1). This region is structurally highly variable, as

Figure 6 Stereoviews of (a) the C α backbone of the CL1 model and (b) the least-squares fit of the C α positions of CL1 (thin line) and rat liver GST 3 (bold line)

defined by X-ray crystallographic studies of other classes of GSTs. Therefore the CL1 model was built without the C-terminal region.

We have shown previously that CL1 can be co-purified with CL2, a class-Mu GST [11]. Presumably there are structural similarities between these two proteins that lead to the formation of heterodimers. Therefore, in spite of closer sequence identity between CL1 and class-Pi GSTs, the model for CL1 was constructed with the class-Mu GST structure as the template.

The CL1 α -backbone model constructed from homology modelling, secondary structure prediction and energy optimization steps is presented in Figure 6. The overall structure of CL1 preserves the folding topology and the molecular shape of the template model. It contains an α/β structure and a helical structure for Domains I and II respectively.

Domain I (residues 1–97) of the CL1 model has four β -strands and three helices, forming a $\beta_1\alpha_1\beta_2\alpha_2\beta_3\beta_4\alpha_3$ topology. The four β -strands form a core β -sheet that is flanked by two (α_1 and α_3)

and one (α_2) α -helices on either side. In the core β -sheet, β_1 is parallel to β_2 , but antiparallel to β_3 and β_4 . Loops are formed between β_1/α_1 and α_2/β_3 . The loop between β_1 and α_1 consists of 16 amino acids and is the longest among the four classes of GSTs. Consequently, this loop protrudes into the cavity of the G-site (Figure 5). The last defined secondary structure of Domain I is α_3 . The chain then traverses from α_3 to form the entire structure of Domain II (residues 104–260), the larger domain. The two domains are connected via a short polypeptide segment (residues 97–103).

In our CL1 model (residues 1–232), Domain II contains five helices (α_4 , α_5 , α_6 , α_7 and α_8). Other classes of GSTs have only four helices in the same region. These helices are connected by a variety of loops and β -turns. The α_4 (residues 108–132) and α_6 (residues 148–173) helices are very long and are connected by a short loop and a two-turn helix (α_5 ; residues 138–145). Their axes are roughly antiparallel. The α_5 and α_6 helices together are equivalent to the α_5 helix of other classes of GSTs. The α_5 of

class-Alpha, -Mu and -Pi GSTs has a proline kink in the middle and assumes an overall appearance of a long bent helix. Sinning et al. [13] have pointed out that this helix should more accurately be described as two helices. In the CL1 model, the corresponding residues within this region exist as two helices, α_5 and α_6 . There are two proline residues (residues 145 and 147) located at the end of, and two residues beyond, α_5 . Consequently, α_5 is structurally well distinguished from α_6 .

Helix α_7 of the CL1 model is connected to α_6 by a short coil of about six residues and is packed almost parallel against α_4 and α_6 . Compared with the same region of other classes of GSTs, α_7 has a two-residue deletion in the middle of the helix. However, the overall length of α_7 is still four residues greater than the α_6 of other classes of GSTs.

The last helix in the second domain, α_8 , is almost perpendicular to α_7 . Consequently α_7 is sandwiched between α_4 , α_6 and α_8 . There is a very short (one-and-a-half turn) helix-like motif consisting of seven residues between α_7 and α_8 . Moving down to the next helix, α_9 , the structure winds back to Domain I. Therefore these two domains are connected by two short polypeptide segments, but the interface of the two domains is large.

The overall least-squares structure fit between the C α carbons of class-Mu GSTs and the CL1 model is illustrated in Figure 6(b). The overall structure of the CL1 model is similar to that of class-Mu GSTs, with the hydrophobic regions buried inside and polar regions distributed on the surface. Nonetheless, several structural variances exist among CL1 and other classes of GSTs. Instead of a β -turn structure, there is a long loop between the first β -strand and the first α -helix. This loop obviously constitutes a major difference in the overall shape of Domain I. For other classes of GSTs [13,14,34], most of the interactions between glutathione and the enzyme are from Domain I and some of the residues involved (Tyr-9 and Arg-15 in class-Alpha; Tyr-6, Trp-7 and Leu-12 in class-Mu; Tyr-7 and Arg-13 in class-Pi) are located in this region without any obstruction. The long loop in the CL1 model protrudes into the GSH binding cleft and could interfere with the enzyme/glutathione interaction. The functional effect of this protrusion remains to be investigated. Nevertheless, Tyr-5 and Arg-14 are conserved and appear to be in the active site. In addition, Arg-54 and Val-72 are conserved between class-Theta and -Alpha GSTs (Arg-45 and Val-55 in the latter). These two residues have been implicated in GSH binding for class-Alpha GSTs [13].

Compared with other class-Theta GSTs, CL1 has a 19-amino-acid insertion in Domain I (Figure 3). This region makes up a section of β_2 and the whole of α_2 in the CL1 model. Omission of this region might account for the low affinity of mammalian class-Theta GSTs towards GSH and their inability to bind on a GSH affinity column.

The residues located at the C-terminal end of α_4 form part of the electrophilic substrate binding site [13,35]. These residues are important for providing an environment in which the electrophilic substrates can react with glutathione. There is a four-residue insertion (residues 119–122) at the centre of α_4 , resulting in an extra turn in this helix for CL1. Combined with the insertion of α_5 , the structural changes around this region might affect the substrate binding and selectivity of the CL1 GST.

In summary, we have isolated a class-Theta GST cDNA clone. Sequencing results indicate that there are multiple class-Theta GSTs in chicken liver. A model is proposed for this clone based on homology modelling, secondary structure prediction and energy optimization steps.

We thank Dr. Sally E. Pemble (University College and Middlesex School of Medicine, London, U.K.) for providing the plasmids pGSTL11 and pGSTA27 for the early screening experiments. E.O.M. is the recipient of an NSC postdoctoral fellowship (NSC83-0112-C001-015). This research was supported in part by grant NSC 83-0803-B-001-110 from the National Science Council, Taiwan, Republic of China.

REFERENCES

- Ketterer, B. and Cole, B. (1991) in *Oxidative Stress: Oxidants and Antioxidants* (Sies, H., ed.), pp. 171–194, Academic Press, London
- Mannervik, B. and Danielson, U. H. (1988) *CRC Crit. Rev. Biochem.* **23**, 283–337
- Tsuchida, S. and Sato, K. (1992) *CRC Crit. Rev. Biochem. Mol. Biol.* **27**, 337–384
- Tan, K. H., Meyer, D. J., Coles, B. and Ketterer, B. (1986) *FEBS Lett.* **207**, 231–233
- Christ-Hazelhof, E., Nugteren, D. H. and Van Dorp, D. A. (1976) *Biochim. Biophys. Acta* **450**, 450–461
- Ishigaki, S., Abramovitz, M. and Listowsky, I. (1989) *Arch. Biochem. Biophys.* **273**, 265–272
- Mannervik, B., Ålin, P., Guthenberg, C., Jensson, H., Tahir, M. K., Warholm, M. and Jorvall, H. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7202–7206
- Hiratsuka, A., Sebata, N., Kawashima, K. et al. (1990) *J. Biol. Chem.* **265**, 11973–11981
- Meyer, D. J., Coles, B., Pemble, S. E., Gilmore, K. S., Fraser, G. M. and Ketterer, B. (1991) *Biochem. J.* **274**, 409–414
- Hussey, A. J. and Hayes, J. D. (1992) *Biochem. J.* **286**, 929–935
- Chang, L.-H., Chuang, L.-F., Tsai, C.-P., Tu, C.-P. D. and Tam, M. F. (1990) *Biochemistry* **29**, 744–750
- Liu, L.-F. and Tam, M. F. (1991) *Biochim. Biophys. Acta* **1090**, 343–344
- Sinning, I., Kleywegt, G. J., Cowan, S. W. et al. (1993) *J. Mol. Biol.* **232**, 192–212
- Ji, X., Zhang, P., Armstrong, R. N. and Gilliland, G. L. (1992) *Biochemistry* **31**, 10169–10184
- Reinemer, P., Dirr, H. W., Ladenstein, R., Schaffer, J., Gallay, O. and Huber, R. (1991) *EMBO J.* **10**, 1997–2005
- Ogura, K., Nishiyama, T., Okada, T. et al. (1991) *Biochem. Biophys. Res. Commun.* **181**, 1294–1300
- Hewick, R. M., Hunkapiller, M. W., Hood, L. E. and Drey, W. J. (1981) *J. Biol. Chem.* **256**, 7990–7997
- Amaravadi, L. and King, M. V. (1994) *BioTechniques* **16**, 98–103
- Pemble, S. E. and Taylor, J. B. (1992) *Biochem. J.* **287**, 957–963
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Feng, D. F. and Doolittle, R. F. (1987) *J. Mol. Evol.* **25**, 351–360
- Brooks, B., Brucoleri, R., Olafson, B., States, D. J., Swaminathan, S. and Karplus, M. (1983) *J. Comp. Chem.* **4**, 187–217
- Pemble, S., Schroeder, K. R., Spencer, S. R. et al. (1994) *Biochem. J.* **300**, 271–276
- Fournier, D., Bride, J. M., Poirie, M., Berge, J.-B. and Plapp, F. W. (1992) *J. Biol. Chem.* **267**, 1840–1845
- Toung, Y.-P. S., Hsieh, T.-S. and Tu, C.-P. D. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 31–35
- Board, P., Russell, R. J., Marano, R. J. and Oakeshott, J. G. (1994) *Biochem. J.* **299**, 425–430
- Bartling, D., Radzio, R., Steiner, U. and Weiler, E. W. (1993) *Eur. J. Biochem.* **216**, 579–586
- Meyer, R. C., Goldsbrough, P. B. and Woodson, W. R. (1991) *Plant Mol. Biol.* **17**, 277–281
- Grove, G., Zarlengo, R. P., Timmerman, K. P., Li, N.-Q., Tam, M. F. and Tu, C.-P. D. (1988) *Nucleic Acids Res.* **16**, 425–438
- Igarashi, T., Kohara, A., Shikata, Y., Sagami, F., Sonoda, J., Horiet, T. and Satoh, T. (1991) *J. Biol. Chem.* **266**, 21709–21717
- Lopez, M. F., Patton, W. F., Sawilovich, W. B. et al. (1994) *Biochim. Biophys. Acta* **1205**, 29–38
- Raghunathan, S., Chandross, R. J., Kretsinger, R. H., Allison, T. J., Penington, C. J. and Rule, G. S. (1994) *J. Mol. Biol.* **238**, 815–832
- Garcia-Saez, I., Parraga, A., Phillips, M. F., Mantle, T. J. and Coll, M. (1994) *J. Mol. Biol.* **237**, 298–314
- Reinemer, P., Dirr, H. W., Ladenstein, R. and Huber, R. (1992) *J. Mol. Biol.* **227**, 214–226
- Ji, X., Johnson, W. W., Sesay, M. A. et al. (1994) *Biochemistry* **33**, 1043–1052
- Gribskov, M. and Burgess, R. R. (1986) *Nucleic Acids Res.* **14**, 6745–6763
- Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950