Characterization and cloning of avian-hepatic glutathione S-transferases

Cheng-Hsilin HSIEH[1], Li-Fan LIU[1], Shu-Ping TSAI and Ming F. TAM[2]
Institute of Molecular Biology, Academia Sinica, Taipei 11529, Taiwan, Republic of China

Cytosolic glutathione S-transferases (GSTs) were isolated from 1-day-old Leghorn chick livers by glutathione (GSH)-affinity chromatography. After sample loading and extensive washing with 0.2 M NaCl, the column was sequentially eluted with 5 mM GSH and 1 mM S-hexylglutathione. The isolated GSTs were subjected to reverse-phase HPLC, electrospray ionization–MS, N-terminal and internal peptide sequencing analyses. The proteins recovered from the 5 mM GSH eluant were predominantly cGSTM1. A protein (cGSTM1′) with an N-terminal amino acid sequence identical to that of cGSTM1 but with the initiator methionine retained and a novel class-mu isozyme (cGSTM2*) were also recovered from this fraction. Nine class-alpha isozymes with distinctive molecular masses were identified from the 1 mM S-hexylglutathione eluant. Three of these proteins are probably variants with minor amino acid substitutions of other isozymes. Of the six remaining class-alpha isozymes, three of them have had their complete (cGSTA1 and cGSTA2) or partial (cGSTA3) cDNA sequences reported previously in the literature. A chicken liver cDNA library was screened with oligonucleotides generated from the cGSTA2 sequence as probes. Clones that encompass the complete coding regions of cGSTA3 and cGSTA4 were obtained. A clone encoding the C-terminal 187 residues of cGSTA5 was also isolated.

Key words: amino acid sequence, cDNA cloning, chromatography, mass spectrometry.

INTRODUCTION

Glutathione S-transferases (EC 2.5.1.18; GSTs) are a family of multifunctional enzymes that catalyse the conjugation of glutathione with electrophilic compounds, resulting in the protection of cellular macromolecules from xenobiotics [1]. These isozymes have also glutathione peroxidase activity and possibly protect tissues from endogenous electrophiles generated during oxidative stress [2]. In addition, GSTs can isomerize prostaglandins [3] and retinoic acids [4]. GSTs may also function as intracellular carrier proteins. The binding of non-substrate compounds such as bilirubin [5] and thyroid hormones [6] to GSTs have been well documented.

Mammalian GSTs have been intensively studied. Even though these isozymes have overlapping substrate activities, they can be distinguished according to their physical, chemical, immunological and structural properties. The cytosolic GSTs are grouped into six classes: alpha, mu, pi [7], theta [8], sigma [9] and zeta [10]. In addition, a class-kappa GST from mitochondria [11] and a membrane-bound form of the isozyme [12] have also been reported. The GST subunits are designated according to the class-based subunit nomenclature proposed by Mannervik et al. [13]

Comparatively, very sparse data are available on avian GSTs. Avian species are exposed to a wide spectrum of xenobiotics. The use of feed additives represents a necessary requirement in poultry farming. Surprisingly, GSTs of the phase-II biotransformation system have received attention only recently. At the protein level, GSTs were isolated from chick livers and designated as CL1-CL5 according to their electrophoretic mobility on denaturing gels [14]. Subsequent cloning experiments and an Expressed Sequence Tag database search have isolated class-alpha [15,16], -mu [17], -theta [18] and -sigma [19] cDNAs from domestic chicken. Characterization of the heterologously expressed proteins has shown unique properties for avian GSTs. For example, cGSTS1 is a glutathione-dependent prostaglandin D_2 synthase that expresses in liver and kidney [19]; cGSTM1 has the highest epoxidase activity reported for class-mu GSTs [20]; and cGSTA1 has high ethacrynic-acid-conjugating activity [21].

In the publication by Chang et al. [14], the purified chick-liver GSTs were characterized by N-terminal sequencing, gel electrophoresis and immunoreactivity only. With the advance in MS [22–24], we decided to re-examine the expression of GSTs in chick livers with liquid chromatography (LC)/MS and internal peptide sequencing. The results are then correlated with the cDNA clones isolated.

MATERIALS AND METHODS

Materials

Male white Leghorn chicks were obtained from a local farm and sacrificed 1 day after hatching. Glutathione (GSH), S-hexylglutathione and epoxy-activated Sepharose 6B were purchased from Sigma Chemical Co. (St. Louis, MO, U.S.A.). The GSH-affinity column was coupled as described previously [25]. *Achromobacter* proteinase I was from Wako (Osaka, Japan). A chicken Lambda ZAP cDNA library derived from a male broiler-breeders' liver was purchased from Stratagene (La Jolla, CA, U.S.A.). The digoxigenin (DIG) DNA labelling and detection kit was obtained from Boehringer Mannheim (Mannheim, Germany). Vydac C_{18} reverse-phase columns were obtained from H-P Separations Group (Hesperia, CA, U.S.A.). Trifluoroacetic acid and heptafluorobutyric acid for HPLC analysis were obtained from Aldrich (Milwaukee, WI, U.S.A.). Problot membranes and chemicals for peptide sequencing were purchased from PE Applied Biosystems (Foster City, CA, U.S.A.). All other chemicals used were reagent grade or better.

Abbreviations used: GSH, reduced glutathione; GST, glutathione S-transferase; LC, liquid chromatography.

[1] These authors contributed equally to this work and should be considered as equal first authors.
[2] To whom correspondence should be addressed (e-mail mbr405@ccvax.sinica.edu.tw).

The nucleotide sequence data reported will appear in the GenBank Nucleotide Sequence Database under the accession numbers L15387, AF133251 and AF133252 for pGCLA3, pGCLA4 and pGCLA5 respectively.

Purification of GSTs from chick livers

Unless specified otherwise, GSTs were purified from 100 1-day-old chick livers in a typical enzyme preparation. Tissues (approx. 90 g) were homogenized in 300 ml of buffer A (10 mM Tris/HCl, pH 8.0, 2 mM ethylenedinitrilotetraacetic acid, 0.2 mM dithiothreitol and 0.5 mM PMSF) containing 20 % (w/v) sucrose with a Polytron (Kinematica, Littau, Switzerland). The homogenate was centrifuged at 300000 ***g*** for 90 min. The supernatant fraction was then passed through a Sephadex G25 column to remove the lipid materials.

The crude liver extracts were loaded on to a GSH-affinity column (2.5 cm × 4.5 cm) pre-equilibrated with buffer A. After loading, the column was washed extensively (60 column volumes) with buffer B (20 mM ethanolamine, pH 9.6, 0.2 M NaCl and 2 mM dithiothreitol) until the absorption of the effluent at 280 nm was less than 0.003. The adsorbed proteins were then eluted from the column with buffer B containing 5 mM GSH. The elution profile was monitored by UV absorption at 280 nm. UV-absorbing fractions were collected and combined. After the UV absorption profile returned to baseline, the column was washed again with buffer B (3 column volumes). The remaining GSTs on the column were then eluted with buffer B containing 1 mM S-hexylglutathione.

LC–MS analysis of GSTs

The GST isozymes were analysed by LC–MS as outlined by Yeh et al. [26]. In brief, the isozymes were resolved on a Vydac (C_{18}; 1 mm × 250 mm) reverse-phase column pre-equilibrated with 55 % solvent 1 (0.08 % trifluoroacetic acid in water) and 45 % solvent 2 (0.07 % trifluoroacetic acid in 80 % acetonitrile). A linear gradient of 45– 80 % solvent 2 was developed over 105 min at a flow rate of 35 μl/min. A splitter was employed after the UV detector flow-cell. Approx. 6 μl/min of samples were introduced on-line into a Quattro quadrupole instrument (Micromass, Altrincham, U.K.) for electrospray ionization–MS analysis. The remaining samples were collected for N-terminal sequence determination. The mass spectrometer was set in the positive-ion mode. Scanning was in the multichannel analyser mode from m/z 600–1600 at 6 s/scan. Data were summed according to the total-ion current profile and processed by the MaxEnt program provided by the manufacturer. For the molecular mass determination of GSTs, the multiply-charged ion peaks from myoglobin (M_r 16951) were used as an external reference for mass scale calibration.

HPLC purification of cGSTs

GST fractions from the GSH-affinity column were separated on a Vydac C_{18} regular bore column (4.6 mm × 250 mm). The buffer system and the gradient employed were identical to those mentioned above. The column was eluted at a flow rate of 600 μl/min and the UV-absorbing peaks were collected manually.

cGSTA2 and A6* co-migrate in this buffer system. The fraction containing these two isozymes was collected and further purified on a C_{18} narrow-bore column (2.1 mm × 250 mm). The column was equilibrated with 55 % solvent 3 (0.08 % heptafluorobutyric acid in water) and 45 % solvent 4 (0.06 % heptafluorobutyric acid in 80 % acetonitrile) at a flow rate of 128 μl/min. After sample injection, the concentration of solvent 4 was increased from 45 to 70 % at a linear rate of 1 %/3 min. The elution profile was monitored at 214 nm as outlined above and the separated GST isozymes were collected manually.

Peptide mapping and sequencing

The organic solvent and acids in the purified cGST fractions were removed by vacuum in a SpeedVac concentrator. The dried pellets were resuspended in 100 mM NH_4HCO_3, pH 8.8, and 2 M urea. Digestion with *Achromobacter* protease I was carried out at 42 °C for 18 h. The substrate to enzyme ratio was estimated at 20:1.

The resulting peptides were separated on the LC–MS system outlined above. The peptides were resolved on a microbore C_{18} reverse-phase column with a linear gradient of 5–70 % solvent 2 developed over 160 min. One-sixth of the samples were subjected to on-line electrospray ionization–MS analysis. The rest of the samples were reserved for Edman degradation. For peptide mapping analysis, the mass spectrometer was calibrated in the positive-ion mode with a NaI and CsI mixture (250–2050 Da) as an external standard. Data were collected in centroid mode from m/z 250–1850 at a scanning rate of 4 s/scan [27].

Selected peptides were subjected to Edman degradation on an Applied Biosystems Procise Model 491 protein sequencer. Samples were applied to Problot membranes for sequencing analysis without prior precycle treatments [28].

Isolation of class-alpha GST cDNAs from a Lambda ZAP cDNA library

A Bluescript phagemid, pGCLa-1, containing the coding region of cGSTA2 is available from the laboratory [16]. The region encoding cGSTA2 was amplified by PCR, then used as a template for generating digoxigenin-labelled DNA probes according to the manufacturer's instructions. Screening of a Lambda ZAP cDNA library made from a single chicken liver was carried out with approx. 500000 plaques. The screening conditions suggested by the manufacturer were followed closely. After the third round of screening, the pBluescript phagemids were excised from the Lambda ZAP vectors of isolated plaques. The inserts of the pBluescript phagemids were then subjected to sequencing analysis on an ABI Prism Model 377 DNA sequencer.

RESULTS AND DISCUSSION

Identification of cGSTs eluted off the affinity column with 5 mM GSH

GSTs were purified with a GSH-affinity column. The column was first eluted with 5 mM GSH. We obtained routinely 20–25 mg of proteins, representing 0.6 % of the proteins in the crude liver extract. This fraction was analysed by LC–MS and the chromatogram is shown in Figure 1(A).

The major component (> 96 %) from this fraction has a molecular mass of 25763 ± 1 Da. The amino acid sequence for the N-terminal 45 residues of this protein is identical to that of cGSTM1 [17], a protein on which we have performed structural and biochemical studies [20]. cGSTM1 has a theoretical molecular mass of 25762 Da, therefore, we conclude that this protein is cGSTM1.

A minor protein component with a molecular mass of 25319 ± 2 Da was eluted in front of cGSTM1. This protein has an N-terminal sequence of AVLGYWDIRGLAHAIRLLLEH-TGTPYEDKLYS/GEAPDYDKSQW. The sequence indicates that this is a class-mu isozyme and differs from that of cGSTM1 in 14 out of the 45 residues. We have also cleaved this protein with *Achromobacter* protease I and determined the amino acid sequence of three internal fragments from the digests. These fragments cover residues 52–68, 122–133 and 134–145 of the corresponding cGSTM1 sequence. The results summarized in

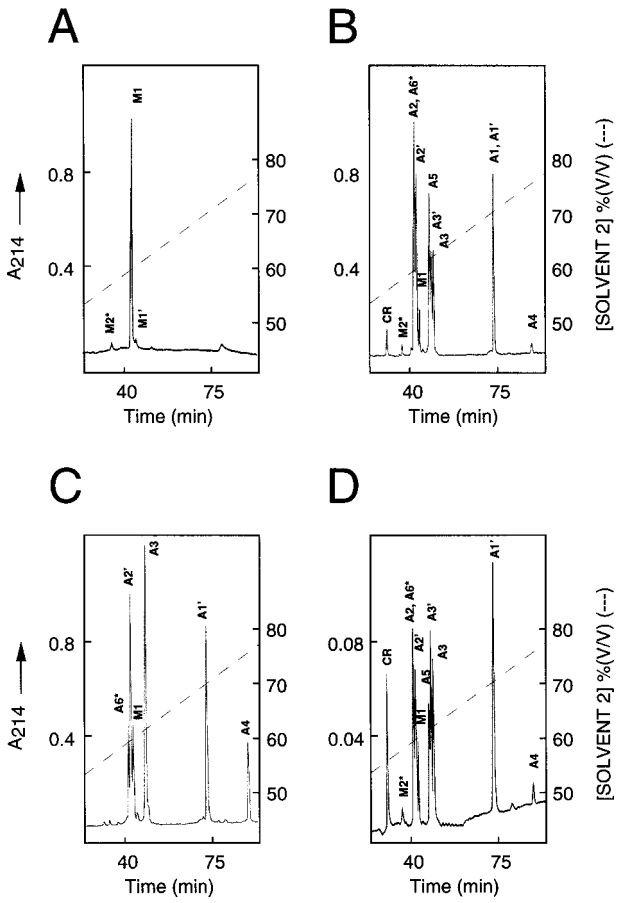

Figure 1 HPLC profile of chick-liver cytosolic GSTs

GSTs from a crude extract of 100 1-day-old chick livers were purified with a GSH-affinity column. Fractions eluted from the column with 5 mM GSH (**A**) and 1 mM S-hexylglutathione (**B**) were analysed on a C_{18} reverse-phase column. The chromatography profiles of the GSTs in the 1 mM S-hexylglutathione eluant prepared from a single adult liver and a 1-day-old chick liver are presented in (**C**) and (**D**) respectively. The GST subunits and the carbonyl reductase (CR) are indicated.

```
hGSTM2     PMTLGYWNIR GLAHSIRLLL EYTDSSYEEK KYTMGDAPDY DRSQWLNEKF   50
rGSTM1     PMILGYWNVR GLTHPIRLLL EYTDSSYEEK RYAMGDAPDY DRSQWLNEKF
cGSTM1     VVTLGYWDIR GLAHAIRLLL EYTETPYQER RYKAGPAPDF DPSDWTNEKE
cGSTM1'  MVVTLGYWDIR GLAHA
cGSTM2*    AVLGYWDIR GLAHAIRLLL EHTGTPYEDK LYS?GEAPDY DKSQW

hGSTM2     KLGLDFPNLP YLIDGTHKIT QSNAILRYIA RKHNLCGESE KEQIREDILE  100
rGSTM1     KLGLDFPNLP YLIDGSRKIT QSNAIMRYLA RKHHLCGETE EERIRADIVE
cGSTM1     KLGLDFPNLP YLIDGDVKLT QSNAILRYIA RKHNMCGETE VEKQRVDVLE
cGSTM1'
cGSTM2*    LGLDFPNLP YFIDGDVK

hGSTM2     NQFMDSRMQL AKLCYDPDFE KLKPEYLQAL PEMLKLYSQF LGKQPWFLGD  150
rGSTM1     NQVMDNRMQL IMLCYNPDFE KQKPEFLKTI PEKMKLYSEF LGKRPWFAGD
cGSTM1     NHLMDLRMAF ARLCYSPDFE KLKPAYLEQL PGKLRQLSRF LGSRSWFVGD
cGSTM1'
cGSTM2*                        LKPGYLEQL PGKLRLFSAF LGDRK

hGSTM2     KITFVDFIAY DVLERNQVFE PSCLDAFPNL KDFISRFEGL EKISAYMKSS  200
rGSTM1     KVTYVDFLAY DILDQYHIFE PKCLDAFPNL KDFLARFEGL KKISAYMKSS
cGSTM1     KLTFVDFLAY DVLDQQRMFV PDCPELQGNL SQFLQRFEAL EKISAYMRSG
cGSTM1'
cGSTM2*

hGSTM2     RFLPRPVFTK MAVWGNK                                      217
rGSTM1     RYLSTPIFSK LAQWSNK                                      217
cGSTM1     RFMKAPIFWY TALWNNKKE                                    219
cGSTM1'
cGSTM2*
```

Figure 2 Comparison of the primary sequence of chick class-mu GSTs with those of hGSTM2 [38] and rGSTM1 [39]

The amino acid sequences are given in single-letter code. Sequences that have been subjected to Edman degradation are underlined.

Table 1 Molecular mass of chick-liver cytosolic GSTs determined by electrospray ionization–MS

The errors in the molecular mass values are S.D.

	Molecular mass (Da)			
cGST	MS	Predicted	Modification	Reference
M1	25763 ± 1	25762	pGCL201 [16]	
M1′	25895 ± 2			
M2*	25319 ± 2			
A1	26238 ± 2	26237	*N*-Acetylation	pGCL301 [14]
A1′	26253 ± 2	26251	*N*-Acetylation	This work.
A2	25211 ± 2	25209	*N*-Acetylation	CLα-1 [15]
A2′*	25221 ± 2			
A3	25326 ± 1	25325	*N*-Acetylation	CLα-2 [15] and this work.
A3′*	25342 ± 2			
A4	25352 ± 1	25351	*N*-Acetylation	This work.
A5	25258 ± 2			
A6*	25134 ± 2			

Figure 2 demonstrate clearly that these sequences are similar but not identical to that of cGSTM1. We designate this novel protein as cGSTM2*. We adopt the nomenclature, proposed by Hayes and Pulford [29], that subunits whose cDNAs have not been cloned are designated by an asterisk (*). With the limited sequence information available (85 residues), cGSTM2* has 73 % identity with cGSTM1. The sequence identity between cGSTM2* and hGSTM2 (66 %) or rGSTM1 (62 %) is much less.

A protein with a longer retention time on the reverse-phase column and a molecular mass of 25895 ± 2 Da was also detected on the chromatogram. This protein was subjected to Edman degradation and an N-terminal sequence of MVVTLGYWDI-RGLAHA was obtained. This sequence is identical to that of cGSTM1 with the retention of an initiator methionine. Moreover, the molecular mass experimentally determined is 132 Da higher than that of cGSTM1 (Table 1). The result matches with that of cGSTM1 with an additional methionine. Apparently, this protein represents a form of cGSTM1 that has not been N-terminally processed. This protein was not subjected to internal sequencing analysis due to quantitative limitation. We tentatively designate this protein as cGSTM1′.

Analysis of proteins in the 1 mM S-hexylglutathione eluant

The GSH-affinity column was further eluted with 1 mM S-hexylglutathione. We routinely recovered 11–12 mg of proteins, representing approx. 0.3 % of the proteins in the crude liver extract. This fraction was analysed by LC–MS and 12 major components were identified by MS, even though only 10 peaks with significant UV absorbance were observed on the chromatogram (Figure 1B). We have performed LC–MS analysis on five separate preparations. The chromatograms presented as Figures 1(A) and 1(B) are highly reproducible with less than 5 % deviation in relative peak intensity.

The first protein eluted off the reverse-phase column had a molecular mass 30272 ± 4 Da. We have identified this protein as carbonyl reductase [30], which co-purifies with hepatic GSTs on either GSH- or S-hexylglutathione-affinity columns [28]. Subunits M1 and M2* could still be observed in this fraction but in a

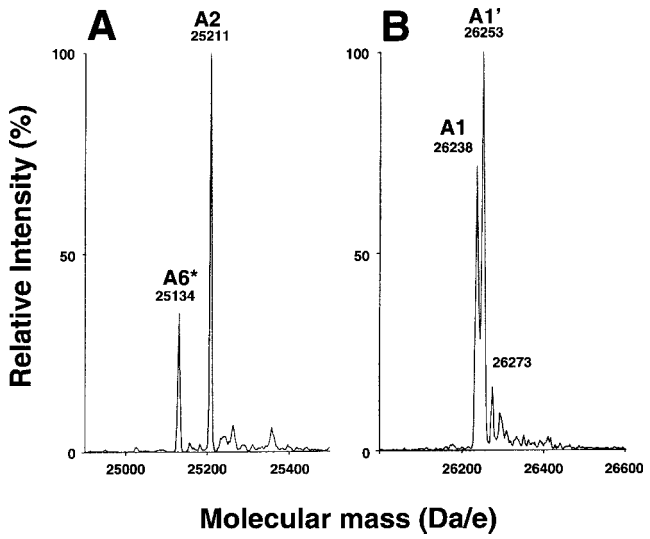

Figure 3 The deconvoluted mass spectrum corresponding to the (A2, A6*) peaks (panel A) and the (A1, A1′) peaks (panel B) of Figure 1(B).

diminished amount (Figure 1B). The remaining proteins from this fraction were N-terminally blocked class-alpha GSTs.

cGSTA1 was previously designated as CL3 [15]. The cDNA clone has been isolated and the deduced amino acid sequence has a molecular mass of 26326 Da, including the initiator methionine. A peak (Figure 1B) that eluted off the column at approx. 70% solvent 2 contains a mixture of proteins with a molecular mass of 26238 ± 2 and 26253 ± 2 Da (Figure 3B). The 26238 Da component probably represents the *N*-acetylated form of cGSTA1 without the initiator methionine. To affirm our conclusion and elucidate the identity of the 26253 Da component, we digested this mixture with *Achromobacter* protease I and subjected the resulting peptides to LC–MS and sequencing analysis. With this approach, we identified 12 peptides covering 86% (196 residues) of the A1 subunit (Figure 4) and isolated a peptide encompassing residues 141–169 that carries a Val^{166} to Leu substitution (results not shown). This substitution adds 14 Da to the molecular mass of the protein and accounts for the difference we observed between these two components. We designated the protein with the Val^{166} to Leu substitution as cGSTA1′. Assuming the ionization efficiency for both proteins is similar, subunits A1 and A1′ exist in the ratio 3:5 (Figure 3B).

We further screened a commercial cDNA library prepared from the liver of an adult male chicken, using probes generated from the cGSTA2 sequence. We isolated a partial cGSTA1′ cDNA clone that lacked the region encoding the N-terminal 45 amino acids. Except for a single G to C change at nucleotide +499, this clone has a DNA sequence identical to that of cGSTA1 [15], including the 3′ noncoding region. This single nucleotide change accounts for the observed Val^{166} to Leu substitution at the protein level.

The observation of both subunits A1 and A1′ on the chromatogram probably reflects a polymorphic effect. Using the same procedure, we analysed liver samples prepared from an adult male broiler (Figure 1C) and a single 1-day-old chick (Figure 1D). We could only detect subunit A1′ in these two preparations.

The peak that eluted off the column at approx. 60% solvent B (Figure 1B) also contains two components (Figure 3A). They have a molecular mass of 25211 ± 2 and 25134 ± 2 Da. These two components can be separated on a C_{18} reverse-phase column with heptafluorobutyric acid as the ion-pairing reagent (results not shown). These proteins were then digested separately with endoproteinase for internal peptide sequencing analysis. We determined the sequence for 89 residues from the 25211 Da protein (Figure 4) and the results are identical to those of cGSTA2 [21], a protein that we initially cloned and designated as CLα-1 [16]. The deduced cGSTA2 sequence has a molecular mass of 25299 Da, including the initiator methionine. Therefore, the mature cGSTA2 from chick liver has an *N*-acetyl group that has replaced the initiator methionine.

We have sequenced six peptides from the 25134 Da protein, covering 81 residues. The results are summarized in Figure 4 and show clearly that the sequence represents a protein that has not been reported. This protein is designated as cGSTA6*.

The protein that eluted immediately after subunits A2 and

```
cGSTA1    MAAKPVLYYFNGRGKMESIRWLLAAAGVEFEEVFLETREQYEKLLQSGILMFQQVPMVEIDGMKLVQTRAILNYIAGKYN   80
cGSTA2    MSGKPVLHYANTRGRMESVRWLLAAAGVEFEEKFLEKKEDLQKLKSDGSLLFQQVPMVEIDGMKMVQTRAILNYIAGKYN
cGSTA2'     SGKPVLHYANTRGRMESVRWLLAAAGVEFEEKFLEK       SDGSLLFQQVPMVEIDGMKMVQTRAILNYIAGKYN
cGSTA3    MAGKPKLHYTRGRGKMESIRWLLAAAGVEFEEEFIEKKEDLEKLRNDGSLLFQQVPMVEIDGMKMVQSRAILCYIAGKYN
cGSTA3'         LHYTRGRGKMESIRWLLAAAGVEFEEEFIEK      LRNDGSLLFQQVPMVEIDGMKMVQSRAILCYIAGKYN
cGSTA4    MSGKPRLTYVNGRGRMESIRWLLSAAGVEFEEIFLETREQLLKLCQDGSLLFHQLPLVEIDGMKLVQCRAILSYIAGKYN
cGSTA5                                          LKTKDDLQKLRTDGFLLFQQVPMVEIDGMKLVQTRAILNYIAGKYN
cGSTA6*                                              LRASGSLLFQQVPMVEIDGMKLVQTRAILNYIAGK

cGSTA1    LYGKDLKERALIDMYVGGTDDLMGFLLSFPFLSAEDKVKQCAFVVEKATSRYFPAYEKVLKDHGQDFLVGNRLSWADIHL  160
cGSTA2    LYGKDLKERALIDMYVEGLADLYELIMMNVVQPADKKEEHLANALDKAANRYFPVFEKVLKDHGHDFLVGNKLSRADVHL
cGSTA2'   LYGKDLK                              KEEHLANALDKAANRYFPVFEK   DHGHDFLVGNKLSRADVHL
cGSTA3    LYGKDLKERAWIDMYVEGTTDLMGMIMALPFQAADVKEKNIALITERATTRYFPVYEKALKDHGQDYLVGNKLSWADIHL
cGSTA3'   LYGK                                   NIALITERSTTRYFPVYEK   DHGQDYLVGNK
cGSTA4    LYGKDLKERALIDMYVEGISDLMQLILVFPFSPPEAKEKNLATIAEKATERYFPVFEKVLKQHGQDFLVGNRFSWADVQL
cGSTA5    LYGKDLKERALIDMYVEGLADLNELILHHEFKPANEMEKDLANILDKATNRYLPVFEKVLKDHGHDFLVGNKLSKADVHL
cGSTA6*                                          AANRYLPVFEK              LSRADVHL

cGSTA1    LEAILMVEEKKSDALSGFPLLQAFKKRISSIPTIKKFLAPGSKRKPISDDKYVETVRRVLRMYYDVKPH   229
cGSTA2    LETILAVEESKPDALAKFPLLQSFKARTSNIPNIKKFLQPGSQRKPRLEEKDIPRLMAIFH            221
cGSTA2'   LETILAVEESKPDALAKFPLLQSFKARISNIPNIK FLQPGSQRKPRLEEKDIPRLMAIFH
cGSTA3    LEAILMTEELKSDILSAFPLLQAFKGRMSNVPTIKKFLQPGSQRKPPLDEKSIANVRKIFSF            222
cGSTA3'         SDILSAFPLLQAFKGRMSNVPTIK FLQPGSQRKPPLDEK
cGSTA4    MEAILAVEEKVPSVLSGFPQLQAFKTKMSNMPTIKKFLQPGSPRKPPPDEHYVATVKKIFKLN            223
cGSTA5    LENILWLEELKPDVLAKFPLLQSFKARMSNMPNIKKFLQPGSPKKPIVQEKDVPAVLAIFS              221
cGSTA6*   LETILVVEEFKPDALAK        ARMSNMPNIK
```

Figure 4 Comparison of the primary sequence of chick class-alpha GSTs

The amino acid sequences are given in single-letter code. Sequences that have been subjected to Edman degradation are underlined.

A6* has a molecular mass of 25221 ± 2 Da. This protein has an acetylated serine at the N-terminus, that can be deblocked by mild acid treatment (results not shown). After protease digestion, we isolated peptide fragments and determined over 81 % (179 residues) of the amino acid sequence of this protein. The sequences obtained are identical to that deduced from the cGSTA2 cDNA clone except for a Thr^{188} to Ile substitution (Figure 4). This change carries an increase of 12 Da, and could account for the observed difference in molecular mass between cGSTA2 and this protein. Furthermore, this protein, but not cGSTA2, was detected in adult chicken (Figure 1C). We tentatively designate this protein as cGSTA2′. Interestingly, both subunits A2 and A2′ are expressed in an 1-day-old chick liver (Figure 1D). Whether the expression of cGSTA2 is ontogenetically controlled or the presence of both A2 and A2′ in our preparations is an allelic effect remains to be investigated.

We have reported the nucleotide sequence of pGCLα-2 that encodes the C-terminal 193 amino acids of an avian class-alpha GST previously [16]. In compliance with the class-based subunit nomenclature, as proposed by Mannervik et al. [13], the gene product of pGCLα-2 is redesignated as cGSTA3 from here on. cGSTA3 (25326 ± 1 Da) eluted off the reverse-phase column at approx. 63 % solvent 2 (Figure 1B). We have done peptide mapping on this protein and selected two fragments covering residues 186–195 and 197–211 for sequence confirmation. The sequences of these two fragments are unique for cGSTA3 (Figure 4).

We have also isolated a full-length cDNA clone for cGSTA3 (Figure 5) and the deduced amino acid sequence is summarized in Figure 4. This clone encodes a protein of 222 amino acids and has a molecular mass of 25414 Da. Therefore, the mature cGSTA3 has an acetylated alanine at its N-terminus. At the protein level, cGSTA3 has 71 % sequence identity with cGSTA2. Cross species, cGSTA3 has 68 % sequence identity with either hGSTA1 [31] or hGSTA2 [32].

The protein that eluted prior to cGSTA3 from the reverse-phase column (Figure 1B) is 16 Da higher in molecular mass (25342 ± 2 Da). The peptide map of this protein is very similar to cGSTA3 (results not shown). We have isolated and sequenced 10 peptides covering 141 residues. The results are summarized in Figure 4 and they are identical to that of cGSTA3 except for an Ala^{127} to Ser substitution. Serine is 17 Da higher in molecular mass than alanine. Even though this protein was not detected in the liver preparation of an adult chicken (Figure 1C), it existed in approx. equal proportion to cGSTA3 in the 1-day-old chick-liver sample (Figure 1D). Whether this protein is an allelic variant of the cGSTA3 remains to be investigated. We tentatively assign this protein as cGSTA3′.

In our cloning experiments, we have obtained an isolate encoding a protein of 223 amino acids (Figure 5) with a molecular mass of 25440 Da. This protein has 69, 68 and 68 % sequence identity with cGSTA1 [15], rGSTA4 [33] and mGSTA4 [34] respectively. These three isozymes have been proposed as members of a subgroup within the class-alpha GSTs that participate in the conjugation of lipid peroxidation products [34]. Site-directed mutagenesis and kinetic studies suggest that a Ala^{107} to Leu substitution can decrease the catalytic activity of mGSTA4 towards 4-hydroxynon-2-enal and 1-chloro-2,4-dinitrobenzene conjugation. Interestingly, cGSTA4 has a leucine at position 107. Experiments are underway to heterologously express this protein for enzymic characterization.

Subunit cGSTA4 is eluted off the reverse-phase column at approx. 76 % solvent 2. The identity of the protein was confirmed by peptide mapping and sequencing. The results are summarized in Figure 4. The molecular mass of the mature cGSTA4 was determined to be 25352 Da, inferring that the protein was *N*-acetylated and without the initiator methionine. This is different from its rat orthologue that has an acetylated methionine at its N-terminus [26].

cGSTA5 is eluted preceding cGSTA3′ on the reverse-phase column (Figure 1B). This protein has a molecular mass of 25258 ± 2 Da. We have sequenced seven peptides covering 91 residues from *Achromobacter* protease digests. The results are summarized in Figure 4 and show clearly that it is distinct from other chick-liver GSTs. A partial cDNA clone, encoding the C-terminal 187 amino acids of cGSTA5 has also been isolated (Figure 5). The deduced amino acid sequence of this clone has 78 and 66 % identity with that of cGSTA2 [16] and a class-alpha GST from a Syrian hamster [35] respectively.

Chick-liver GSTs have been isolated by a combination of S-hexylglutathione and glutathione affinity columns previously. The GST subunits were recovered from the liver supernatant in approx. 1.4 % yield. The proteins were analysed by gel electrophoresis and N-terminal sequencing analysis. According to their electrophoretic mobility on denaturing gels, the GSTs were designated CL1 to CL5 [14]. Cloning experiments have shown that CL1, CL2, CL3 and CL5 are class-theta (cGSTT1) [18], -mu (cGSTM1) [17], -alpha (cGSTA1) [15] and -sigma (cGSTS1) [19] subunits respectively. Immunological data suggest that CL4 belongs to class-alpha GST(s), but the corresponding clone(s) has not been identified.

In the present study, we purified chick-liver GSTs with a single GSH-affinity column. The column was washed extensively before eluting with buffers containing GSH and S-hexylglutathione sequentially. Combining both fractions, GST subunits were recovered from crude liver extracts in approx. 0.9 % yield, significantly less than that reported by Chang et al. [14]. Therefore, we cannot exclude the possibility that certain GSTs are omitted from our analysis. Indeed, we have not detected isozymes designated previously as CL1 (class-theta) or CL5 (class-sigma) in our fractions.

We have determined the electrophoretic mobility of the class-alpha GSTs in the 1 mM S-hexylglutathione eluant. Subunits A1, A1′, A3, A4, A5 and A6 have the same, while A2 and A2′ have faster mobility on a denaturing gel (results not shown). Therefore, A2 and A2′ correspond to CL4 in Chang's study [14], while CL3 is a mixture of at least six polypeptides.

Human [23], rat [22] and porcine [24] hepatic GSTs have been analysed with LC–MS. A complex array of class-mu GSTs were detected. In contrast, we observed only M1, M1′ and M2* subunits in our preparations. Quantitatively, over 96 % of the class-mu subunits are cGSTM1. The amount of cGSTM1 is more than all the chick-liver class-alpha subunits combined. In earlier experiments, the GSTs were eluted off the affinity column in a single step with S-hexylglutathione. The presence of cGSTM1 masked the signals from A2, A2′ and A6 in LC–MS analysis. Fortunately, the class-mu GSTs bind less tightly to the affinity column and can be selectively eluted with GSH.

We have identified at least six class-alpha GSTs in chick-liver cytosols. We have proved at both cDNA and protein level that cGSTA1 has polymorphic variations. A single Val^{166} to Leu substitution occurs on cGSTA1′. According to the cGSTA1 structural model [21], this residue is located on helix 6 of the C-terminal domain and away from the substrate binding site. This commutation should have minimal effect on the activity of the enzyme.

Chromatographic and protein sequencing data suggest that cGSTA2 and cGSTA3 probably have allelic variants with amino acid substitutions at residues 188 and 127 respectively. Assuming that cGSTA2 and cGSTA3 have similar overall folding topology

```
pGCLA3                                                  -262      ccgctgcggcgcggagctccgggatc     -237

pGCLA3   tcccacagccgacacgggactaggaggcggacctggagcagcgcggcggcgctgcagtcaagctcgcttcagctgccag     -160

pGCLA3   gtttgtactgtggtgtgcagctgcacctggtggtagtgaaactttgaaaggaccgagttcaatctccgagtgattatct      -81

pGCLA3   ggtggcctgacttcagtccttggtgctgggagcattgtccaggccttatatatccaacggaaaggagagctctgaaatc       -1
pGCLA4                                                                          -13  cctgctcagaagc       -1

pGCLA3   ATG GCT GGG AAA CCG AAG CTT CAC TAT ACC AGA GGA AGG GGG AAG ATG GAG TCA ATT CGG       60
pGCLA4   --- T-G --- --G --- -G- --C AC- --- GT- -AT --G --A --- CGA --- --- --C --C --A      60

pGCLA3   TGG CTG TTA GCA GCA GCT GGG GTT GAA TTC GAG GAA GAA TTT ATA GAA AAA AAG GAA GAC      120
pGCLA4   --- --- C-G T-T --- --- --A --G --G --T --A --- ATT --- C-G --- -C- -GA --G C-G      120
pGCLA5                                                          C-G A-- -C- --- --T --T       18

pGCLA3   CTA GAA AAA TTA CGC AAT GAT GGA TCC CTG CTG TTC CAG CAA GTG CCC ATG GTG GAA ATC      180
pGCLA4   T-- TTG --G --- T-- C-A --- --- --- --- --- --- --T --- C-- --A C-- --T --G ---      180
pGCLA5   --C C-- --G --- A-G -CA --- --- -T- --- --- --- --- --- --- --- --- --- --G ---       78

pGCLA3   GAT GGG ATG AAG ATG GTG CAG AGC AGA GCC ATC CTT TGC TAC ATA GCA GGG AAA TAC AAT      240
pGCLA4   --- --- --- --- T-- --- --- T-- --- --- --- --C A-- --- --C --- --- --- --- ---      240
pGCLA5   --- --- --- --- T-- --- --- -C- --- --- --- --C AA- --- --- --- --- --- --- ---      138

pGCLA3   CTC TAC GGG AAA GAC CTG AAG GAG AGA GCC TGG ATT GAT ATG TAT GTG GAG GGA ACA ACA      300
pGCLA4   --- --T --- --- --- --- --- --- --- --- --- CT- --C --C --- --- --- --A --- -T- T--      300
pGCLA5   --- --- --- --G --T --- --- --- --- --- CTA --- --C --- --- --- --A --- CTG G--      198

pGCLA3   GAC CTG ATG GGA ATG ATC ATG GCT CTC CCT TTT CAA GCA GCC GAT GTG AAA GAA AAG AAT      360
pGCLA4   --- --- --- CA- T-- --T T-- -TG T-T --- --C TCT C-- C-T --G -CA --G --G --A ---      360
pGCLA5   --T --- -AT -AG T-A --- T-- CAC -AT GAG --C A-- C-- --T A-- -AA -TG --G --A G-C      258

pGCLA3   ATT GCC TTA ATC ACT GAA CGA GCT ACA ACC AGA TAC TTT CCT GTT TAT GAA AAG GCC TTA      420
pGCLA4   C-- --- AC- --T G-A --G AAG --A --- GAG --G --- --C --- --C -T- --- --- -TT --G      420
pGCLA5   C-- --T AAT --T TTG --C AAG --C --- -A- --- --- --G --A --C -T- --G --- -TT --G      318

pGCLA3   AAA GAC CAT GGT CAG GAT TAT CTT GTT GGC AAC AAG CTA AGC TGG GCA GAC ATC CAC CTG      480
pGCLA4   --- C-G --- --C --A --C -T- --- --G --A --- CGA T-C --- --- --- --T G-T --G --C      480
pGCLA5   --- --- --- --G --T --C -T- --- --- --C --- --- C-G --- AAA --T --T G-G --T T-A      378

pGCLA3   CTG GAA GCC ATT TTA ATG ACA GAA GAA CTT AAG TCT GAT ATA CTG TCT GCA TTC CCT CTG      540
pGCLA4   A-- --- --- --- --- GCA GTG --G --G AAA GT- C-- TC- G-G --T --- -GG --T --- -A-      540
pGCLA5   --- --- AA- --- --- TGG TTG --- --G T-G --- C-C --- G-- --T G-A AA- --T --C --C      438

pGCLA3   CTA CAG GCT TTT AAA GGA AGA ATG AGC AAT GTC CCA ACA ATC AAA AAA TTC TTG CAG CCT      600
pGCLA4   --G --- --- --- --- ACC -A- --- --- --C A-G --T --- --T --G --G --- C-- --- ---      600
pGCLA5   T-G --- AG- --- --- -C- --- --- --- --- A-G --C -AC --- --G --- --- C-- --- ---      498

pGCLA3   GGC AGC CAG AGG AAG CCT CCA TTA GAT GAG AAA TCA ATT GCT AAT GTG AGG AAA ATA TTC      660
pGCLA4   --- --- -CA --- --- --C --- CC- --- --A C-T -AT G-A --A -C- --- -AA --- --T --T      660
pGCLA5   --- --- -CA -A- --A --A ATC G-T C-G --A A-G GAT G-A C-A GC- --- TTA GCC --T ---      558

pGCLA3   AGT TTC taa tcatgtggctgctacagataacgtctatagtatattgcctaagctgtgcctcttaagtataagctgc      736
pGCLA4   -AG C-A AAC tga gtgcagcgttaacttcactaagtcccaaagtgctgggaagaaagacatcaaaacccgaaacct      726
pGCLA5   --- tga gcagcagcctaactgtgctgtaactccattcaattgtcttttgtggttcctgagaagctaagtgaaatgaa      635

pGCLA3   gtagatctatcctgtggtttgtgccttataatctttgagatggaagacaatttagattaaaggctgttggacctgctga      815
pGCLA4   taaaggaagtaggaaactcagtaaagttttattgtacttaaaagcaatggtagaaataatagaagagtgaaataaagca      805
pGCLA5   ttaaattagtttaatttcactaggtttcatttccacattggcaaatgaagttcttaactcagtagtctcaaaccttaac      714

pGCLA3   ttacagaaggtagtgttttattaccaagagttcttgctaccatcagttacatgaaataaaacatactaatgtttgctgt      897
pGCLA4   ttctgtttggcaatgtgaaaaaaaaaa                                                        832
pGCLA5   atactgtcttaaatacttcatgacctttacagaataaagttgacgttaaaatgtaaaaaaaaa                    778

pGCLA3   ttggttagccattaattgcagtaaagtatttcagctgtagtttccaaagcactgcaaacaaagcctcttccctcccttt      973

pGCLA3   ttattagttgcagcccagaaaagctactctgtaacatttgg                                         1014
```

Figure 5 Nucleotide sequences of cDNA inserts of pGCLA3, pGCLA4 and pGCLA5

The coding regions are represented with capital letters. The polyadenylation signals are underlined.

as hGSTA1 [36], then these two residues are located on helix 7 and helix 5 respectively. They are on the outside surfaces of the macromolecules and should not contribute to substrate binding or enzymic activity.

GSTs with high activities toward α,β-unsaturated carbonyl compounds have been identified in rat [33], mouse [34] and human [37]. Based on sequence homology, cGSTA1 has been suggested to be a member of this subgroup of class-alpha GSTs [34]. However, cGSTA1 has very low 4-hydroxynon-2-enal-conjugating activity (M. F. Tam, unpublished results). We have now isolated cGSTA4 that has 68% sequence identity with rGSTA4. This isozyme is possibly a member of this subgroup.

In summary, we have characterized chick-liver cytosolic GSTs by LC–MS. We have identified two class-mu and six class-alpha GSTs. We have isolated two complete (cGSTA3 and cGSTA4) and one partial (cGSTA5) cDNA clones by library screening. Experiments are underway to express these clones for enzymic characterization.

This research was supported in part by Grants from the National Science Council and Academia Sinica, Taiwan, Republic of China.

REFERENCES

1 Armstrong, R. N. (1997) Chem. Res. Toxicol. **10**, 2–18
2 Beckett, G. J. and Hayes, J. D. (1993) Adv. Clin. Chem. **30**, 281–380
3 Christ-Hazelhof, E. and Nutgeren, D. (1979) Biochim. Biophys. Acta **572**, 43–51

4 Chen, H. and Juchau, M. R. (1998) Biochem. J. **336**, 223–226
5 Bhargava, M. M., Listowsky, I. and Arias, I. M. (1978) J. Biol. Chem. **253**, 4112–4115
6 Ishigaki, S., Abramovitz, M. and Listowsky, I. (1989) Arch. Biochem. Biophys. **273**, 265–272
7 Mannervik, B., Ålin, P., Guthenberg, C., Jensson, H., Tahir, M. K., Warholm, M. and Jornvall, H. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 7202–7206
8 Meyer, D. J., Coles, B., Pemble, S. E., Gilmore, K. S., Fraser, G. M. and Ketterer, B. (1991) Biochem. J. **274**, 409–414
9 Meyer, D. J. and Thomas, M. (1995) Biochem. J. **311**, 739–742
10 Board, P. G., Baker, R. T., Chelvanayagam, G. and Jermiin, L. S. (1997) Biochem. J. **328**, 929–935
11 Pemble, S. E., Wardle, A. F. and Taylor, J. B. (1996) Biochem. J. **319**, 749–754
12 Morgenstern, R., DePierre, J. W. and Jornvall, H. (1985) J. Biol. Chem. **260**, 13976–13983
13 Mannervik, B., Awasthi, Y. C., Board, P. G., Hayes, J. D., Di Ilio, C., Ketterer, B., Listowsky, I., Morgenstern, R., Muramatsu, M., Pearson, W. R. et al. (1992) Biochem. J. **282**, 305–306
14 Chang, L.-H., Chuang, L.-F., Tsai, C.-P., Tu, C.-P. D. and Tam, M. F. (1990) Biochemistry **29**, 744–750
15 Chang, L.-H., Fan, J.-Y., Liu, L.-F., Tsai, S.-P. and Tam, M. F. (1992) Biochem. J. **281**, 545–551
16 Liu, L.-F., Wu, S.-H. and Tam, M. F. (1993) Biochim. Biophys. Acta **1216**, 332–334
17 Liu, L.-F. and Tam, M. F. (1991) Biochim. Biophys. Acta **1090**, 343–344
18 Hsiao, C.-D., Martsen, E. O., Lee, J.-Y., Tsai, S.-P. and Tam, M. F. (1995) Biochem. J. **312**, 91–98
19 Thomson, A. M., Meyer, D. J. and Hayes, J. D. (1998) Biochem. J. **333**, 317–325
20 Sun, Y. J., Kuan, I. C., Tam, M. F. and Hsiao, C. D. (1998) J. Mol. Biol. **278**, 239–252
21 Liu, L. F., Liaw, Y. C. and Tam, M. F. (1997) Biochem. J. **327**, 593–600
22 Yeh, H.-I., Hsieh, C.-H., Wang, L.-Y., Tsai, S.-P., Hsu, H.-Y. and Tam, M. F. (1995) Biochem. J. **308**, 69–75
23 Rowe, J. D., Nieves, E. and Listowsky, I. (1997) Biochem. J. **325**, 481–486
24 Rouimi, P., Anglade, P., Debrauwer, L. and Tulliez, J. (1996) Biochem. J. **317**, 879–884
25 Simons, P. C. and Vander Jagt, D. L. (1981) Methods Enzymol. **77**, 235–237
26 Yeh, H. I., Lee, J. Y., Tsai, S. P., Hsieh, C. H. and Tam, M. F. (1996) Biochem. J. **314**, 1017–1025
27 Klarskov, K., Roecklin, D., Bouchon, B., Sabatie, B., Dorsselaer, A. V. and Bischoff, R. (1994) Anal. Biochem. **216**, 127–134
28 Hsieh, C.-H., Tsai, S.-P., Yeh, H.-I., Sheu, T.-C. and Tam, M. F. (1997) Biochem. J. **323**, 503–510
29 Hayes, J. D. and Pulford, D. J. (1995) CRC Crit. Rev. Biochem. Mol. Biol. **30**, 445–600
30 Tsai, S. P., Wang, L. Y., Yeh, H. I. and Tam, M. F. (1996) Biochim. Biophys. Acta **1292**, 289–292
31 Tu, C.-P. D. and Qian, B. (1986) Biochem. Biophys. Res. Commun. **141**, 229–237
32 Rhoads, D. M., Zarlengo, R. P. and Tu, C.-P. D. (1987) Biochem. Biophys. Res. Commun. **145**, 474–481
33 Stenberg, G., Ridderstrom, M., Engstrom, Å., Pemble, S. E. and Mannervik, B. (1992) Biochem. J. **284**, 313–319
34 Zimniak, P., Eckles, M. A., Saxena, M. and Awasthi, Y. C. (1992) FEBS. Lett. **313**, 173–176
35 Maggouta, F., Li, S. A., Li, J. J. and Norris, J. S. (1997) Biochem. J. **323**, 147–149
36 Sinning, I., Kleywegt, G. J., Cowan, S. W., Reinemer, P., Dirr, H. W., Huber, R., Gilliland, G. L., Armstrong, R. N., Ji, X., Board, P. G. et al. (1993) J. Mol. Biol. **232**, 192–212
37 Hubatsch, I., Ridderstrom, M. and Mannervik, B. (1998) Biochem. J. **330**, 175–179
38 Vorachek, W. R., Pearson, W. R. and Rule, G. S. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 4443–4447
39 Lai, H.-C. J., Grove, G. and Tu, C.-P. D. (1986) Nucleic Acids Res. **14**, 6101–6114

Received 5 May 1999/23 June 1999; accepted 14 July 1999