Amino acid sequence of glutathione S-transferase rGSTM5* from rat testis

Ming F. TAM¹, Cheng-Hsilin HSIEH, Shu-Ping TSAI and Tsuey-Chyi S. TAM

Institute of Molecular Biology, Academia Sinica, Taipei 11529, Republic of China

Glutathione S-transferase rGSTM5* was isolated from rat testis with a combination of glutathione affinity column and reverse-phase column chromatography. The protein was digested with *Achromobacter* protease I or endoproteinase Arg-C. The peptide fragments were isolated for electrospray MS and N-terminal peptide sequencing analyses. The primary amino acid sequence

of rGSTM5* comprises 217 residues and has a calculated average molecular mass of 25495.3 Da. The result is identical to that obtained for rGSTM5* with liquid chromatography–MS from a mixture of rat testicular GSTs. Therefore, rGSTM5* has not been post-translationally modified.

INTRODUCTION

Glutathione S-transferases (GSTs; EC 2.5.1.18) comprise a large family of dimeric enzymes that catalyse the nucleophilic addition of the tripeptide glutathione to a wide spectrum of compounds that have electrophilic functional groups. These isozymes can be found in most aerobic micro-organisms, plants and animals [1]. GSTs function primarily as phase II detoxification enzymes, protecting cells against both endogenous and xenobiotic alkylating agents [2]. This group of enzymes can catalyse reactions towards a large number of diverse substrates, showing the remarkable tolerance of these isozymes for both the types of electrophilic functional group and the structure of the molecules to which they are catalysed [3].

GSTs may also function as intracellular carrier proteins. The binding of bilirubin [4], haem derivatives and thyroid hormones [5] to GSTs has been well documented. It has been proposed that GSTs may have an additional protective role by binding and immobilizing certain reactive electrophilic molecules, and subsequently inactivating these compounds [1]. Besides their detoxification roles, these dimeric proteins are also involved in the reduction of organic hydroperoxides [6] and isomerization of prostaglandins [7].

Rat GSTs have been extensively studied. At least 16 cytosolic GSTs have been identified. On the bases of physical, immunological, structural and enzymic properties, these isozymes are grouped into classes Alpha (rGSTA1–5, rGSTA6*), Mu (rGSTM1–3, rGSTM4*–6*), Pi (rGSTP1), Sigma (rGSTS1) and Theta (rGSTT1, rGSTT2). By convention, subunits whose cDNAs have not been cloned are designated with an asterisk [8].

Protein N-terminal processing and modification have been documented for GSTs [9–11]. Under conditions *in vitro*, Alpha- [12] and Mu-class GSTs [13] can act as substrates for phosphorylation and methylation, respectively. The modified proteins have a lower ligandin or enzymic activity. However, the occurrence of these modifications *in vivo* has not been established.

We have analysed rat GSTs by MS, searching for post-translational modifications. The modifications we detected are limited to N-terminal processing [11] or C-terminal truncations [14]. We also observed the tissue-specific expression of rGSTM5*

in rat testis [15]. Since the cDNA for this isozyme has not been isolated, we resorted to protein sequencing to decipher whether this protein has been post-translationally modified *in vivo*.

MATERIALS AND METHODS

Materials

Sexually mature Sprague–Dawley rats were obtained from the National Animal Center (Taiwan, Republic of China). Glutathione, S-hexylglutathione, epoxy-activated Sepharose 6B and Superdex 30 were obtained from Sigma Chemical Co. (St. Louis, MO, U.S.A.). The glutathione affinity column was prepared according to the method of Simons and Vander Jagt [16]. The Vydac C₁₈ reverse-phase columns were purchased from H-P Separations Group (Hesperia, CA, U.S.A.). *Achromobacter* proteinase I and endoproteinase Arg-C were obtained from Wako (Osaka, Japan) and Boehringer Mannheim GmbH (Mannheim, Germany), respectively. Trifluoroacetic acid, Problot membranes and chemicals for peptide sequencing were obtained from PE-Applied Biosystems (Foster City, CA, U.S.A.). All other chemicals used were reagent grade or better.

Isolation of rGST5*

Testicular GSTs were isolated from adult Sprague–Dawley rats (200–250 g) as outlined by Hsieh et al. [15]. Briefly, tissues were homogenized in 4 vol. of buffer A [10 mM Tris/HCl (pH 8.0)/0.2 mM dithiothreitol/2 mM EDTA/0.5 mM PMSF] with 20% sucrose. Debris was removed by centrifugation and the supernatant was dialysed overnight against buffer A before loading on to a glutathione-Sepharose affinity column. The column was pre-equilibrated in buffer A and washed with buffer A containing 0.2 M NaCl after sample loading. GSTs were eluted from the column with buffer A containing 0.2 M NaCl, 5 mM glutathione and 5 mM S-hexylglutathione.

The GST isozymes were separated on a Vydac C₁₈ regular bore column (4.6 mm × 250 mm). Samples were introduced from a ABI model 140B HPLC system equipped with a model 1000S diode array detector. The column was equilibrated with 55%

Abbreviations used: GST, glutathione S-transferase (the GST subunits are designated according to the class-based subunit nomenclature proposed by Mannervik et al. [23], originally devised for human GST); ES/MS, electrospray MS; LC/MS, liquid chromatography–MS.

¹ To whom correspondence should be addressed (e-mail MBR405@CCVAX.SINICA.EDU.TW).

solvent 1 (0.08 % trifluoroacetic acid in water) and 45 % solvent 2 (0.06 % trifluoroacetic acid in 80 % acetonitrile). A linear gradient of 45–75 % solvent 2 was developed over 90 min at a flow rate of 600 μ l/min. Proteins eluted from the column were monitored by UV absorbance at 214 nm and compared with the published elution profile [15] for isozyme identification. In the HPLC system that we were using, the rGSTM5* peak has a retention time between those of rGSTM2 and rGSTP1 and can be identified unambiguously.

Endoprotease digestion

The organic solvent and acid were removed from the purified rGSTM5* fractions in a SpeedVac concentrator. The protein (5–10 μ g) was redissolved in 0.1 M Tris/HCl (pH 8.3)/6 M guanidinium HCl/1 mM Na₂EDTA. The protein was then reduced for 1 h in 2 mM dithiothreitol under nitrogen atmosphere at 38 °C. A 50 mM solution of iodoacetic acid (neutralized with NaOH) was added to give a final concentration of 5 mM. The mixture was kept under nitrogen for 1 h at 38 °C in the dark. Protein was recovered from the mixture by passing through a Superdex 30 gel-filtration column (5 mm $\times$ 100 mm) with 50 mM Tris/HCl (pH 9.2)/10 % acetonitrile as eluting buffer. The carboxymethylated rGSTM5* was then digested with *Achromobacter* protease I at 42 °C for 16 h. The substrate-to-enzyme ratio was maintained at 20:1.

For endoprotease Arg-C digestion, the pH of the rGSTM5* fraction from the reverse-phase column was raised to 8 with 1 M Tris/HCl. The sample volume was reduced to approx. 20 % of the original by drying in a SpeedVac concentrator. Urea and CaCl₂ were added to a final concentration of 2 and 0.1 M, respectively. Endoprotease Arg-C was then added at a substrate-to-enzyme ratio of 20:1, and incubation was carried out at 37 °C for 16 h.

Peptide mapping

The proteinase digests were subjected to liquid chromatography–MS (LC/MS) analysis. Samples were introduced from a Prodigix-4P (Microtech Scientific) HPLC system. Peptides eluted from a microbore Vydac C₁₈ reverse-phase column (1 mm $\times$ 250 mm) were detected with a model UVIS 204 detector (Linear Instruments) equipped with a 250 nl illuminated-volume high-pressure flow cell. The column was equilibrated with 95 % solvent 1 and a linear gradient of 5–70 % solvent 2 was developed over 162.5 min at 28 μ l/min. The eluant passed through a splitter (1:3) that was placed after the UV detector flow cell and approx. 6–8 μ l of sample/min were subjected to on-line electrospray MS (ES/MS) analysis. The rest of the samples were collected manually for Edman degradation.

ES/MS analyses were performed on a Quattro-Bio-Q mass spectrometer (Micromass, Altrincham, Cheshire, U.K.). The instrument was calibrated in the positive ion mode with a NaI/CsI mixture (250–2050 Da) as external standard. Data were collected in centroid mode from m/z 250 to 1850 at a scanning rate of 1 scan/4 s.

Peptide sequencing

Automated cycles of Edman degradation were performed with an ABI Procise model 491 protein sequencer. The peptide samples were directly blotted on to precut (4 mm $\times$ 4 mm) Problot membranes under light suction from a water aspirator. The samples on the membranes were subjected to sequencing analyses without further precycle treatment.

RESULTS AND DISCUSSION

rGSTM5* isolated from the reverse-phase column was reduced, carboxymethylated and subjected to LC/MS analysis. The modified protein had a molecular mass of 25673 ± 3 Da (results not shown). Compared with our previous result (25495 Da) on the native protein [15], the data clearly indicate the presence of three cysteines on rGSTM5*.

The *Achromobacter* protease I digests of rGSTM5* were separated on a microbore reverse-phase column for on-line LC/MS analysis (Figure 1), and 21 fractions were collected for Edman degradation, as indicated. Among the 21 fractions, we obtained sequences for 19 samples. Due to the high sequence homology among GSTs of the same class, the relative positions of these peptide fragments from rGSTM5* can be inferred by comparing with sequences of other class-Mu GSTs. Sequencing data from these fragments cover a total of 217 residues (Figure 2).

Fractions L11, L15 and L19 did not yield any sequence on Edman degradation. In addition, L19 did not give any meaningful ionization products on MS analysis and probably represents plastic contaminants from the Eppendorf tubes during sample handling. L21 has a molecular mass of 25677.2 Da and sequencing results revealed that it represents the undigested but carboxymethylated rGSTM5*. Theoretically, it should have a molecular mass of 25669.4 Da. The large deviation between experimental and calculated molecular mass was probably caused

Table 1 Amino acid sequencing and MS analyses of *Achromobacter* protease I and endoprotease Arg-C digests of rGSTM5*

Abbreviation used: Ct, carboxymethylcysteine.

		Molecular mass (Da)	
		Observed	Calculated
<i>Achromobacter</i> protease I digests			
L1	QATWGNK	803.6	803.9
L2	ISAYMK	711.6	711.9
L3	RKPEFLK	916.8	917.2
L4	GLPDQLK	769.6	769.9
L5	QPWFAGDK	947.5	948.1
L6	RYAMGDAPDYDRSQWLDDK	2302.0	2302.5
L7	LYSEFLGK	956.5	956.1
L8	HNLC†GETEERIRVDILEK	2340.8	2341.6
L9	TSRFLPSPVYLK	1407.4	1407.7
L10	QVMDTRIQMGTLC†YSPDFEK	2420.3	2420.8
L11	No sequence	1218.7	
L12	VTQSNAILRYLGRK	1618.5	1618.9
L13	LDLDFPNLPYLIDGSHK	1956.6	1957.2
L14	FKLDLDFPNLPYLIDGSHK	2232.2	2232.6
L15	No sequence	2600.2	
L16	ITFADFLVYDVLQHRMFEPK	2584.7	2585.0
L17	C†LDAFPNLMDFVVFEGFLKK	2381.5	2381.8
L18	C†LDAFPNLMDFVVFEGFLK	2253.3	2253.6
L19	No Sequence	—	
L20	PVTGLGYWDIRGLGHAIRLLLEYTETSYEK	3523.5	3524.0
L21	Undigested rGSTM5*	25677.2	25669.4
Endoprotease Arg-C digests			
R1	YLGR	507.3	507.6
R2	GLGHAIR	722.5	722.9
R3	YAMGDAPDYDR	1272.7	1273.3
R4	VDILEKQVMDTR	1446.5	1446.7
R5	LLLEYTETSYEK	1773.5	1774.0
R6	FLPSPVYLKQATWGNK	1848.9	1849.2
R7	PVTGLGYWDIR	1219.2	1219.4
R8	SQLWDDKFKLDLDFPNLPYLIDGSHKVTQSNAILR	4088.2	4088.6

Figure 1 HPLC profile of protease digests of rGSTM5*

Purified rGSTM5* was digested with *Achromobacter* protease I and the peptides generated were analysed with LC/MS. Fractions collected for Edman degradation are marked (L1–L21).

Figure 2 Amino acid sequence of rGSTM5* as determined by Edman degradation of *Achromobacter* protease I and endoproteinase Arg-C digests

Molecular masses of the peptides found by ES/MS are indicated in parentheses. The carboxymethyl/cysteines are represented by C.

rGSTM5*	PVT ^{††} LG [†] YWD [†] IRGLGHAIRLLLEYTETS [*] YEEKRYAMGDAPDYDRSQWLDDKF	50
rGSTM1	-MI-----NV-----T-P-----DS-----NE-----	
rGSTM2	-M-----A-----F-----D-----D-K-S-----SE-----	
rGSTM3	-M-----A-----DS-----T-----F-----NE-----	
rGSTM4*	AMI-----NV-----T-P-----DSN-----V-----NF-----SE-----	
	<u>β1</u> <u>α1</u> <u>β2</u> <u>α2</u>	
rGSTM5*	KLDLDFPNLPYLIDGSHKVTQSNAILRYLGRKHNLCGETEEERIRVDILE	100
rGSTM1	--G-----R-I-----M--A--H-----A--V--	
rGSTM2	--G-----I-----V--	
rGSTM3	--G-----I-----	
rGSTM4*	N-G--I-----T--	
	<u>β3</u> <u>β4</u> <u>α3</u> <u>α4</u>	
rGSTM5*	KQVMDTRIQMGTLCYSPDFEKRKPEFLKGLPDQLKLYSEFLGKQPWFAGD	150
rGSTM1	N---N-M-LIM---N---Q-----TI-EKM-----R-----	
rGSTM2	N-A---L-LAMV---RK---Y-E---EKM-----N	
rGSTM3	N-L---N-MVLAR---N---L---GY-EQ---GMMR-----R-----	
rGSTM4*	N-----HLMIV-C-----Q-----SI-EKM-I-----R-----	
	<u>α4</u> <u>α5</u>	
rGSTM5*	KITFADFLVYDVLQDHRMFEPKCLDAFPNLMDFVVFHFEGLKKISAYMKTS	200
rGSTM1	-V-YV---A--I---YHI-----K--LAR-----S-	
rGSTM2	---YV-----I-----K--AR-----D--SG	
rGSTM3	-----IA-----ERNQV--AT-----K--IAR-----D--S-	
rGSTM4*	-V-YV---A--I---Y---PE-----K--LAR-----S-	
	<u>α6</u> <u>α7</u> <u>α8</u>	
rGSTM5*	RFLPSPVYLKQATWGNK	217
rGSTM1	-Y-ST-IFS-L-Q-S--	
rGSTM2	---SK-IFA-M-F-NP-	
rGSTM3	----R-LFT-M-I--S-	
rGSTM4*	S---R--FT-IPQ--TD	

Figure 3 Amino acid sequence comparison of rGSTM5*, rGSTM1 [24], rGSTM2 [25], rGSTM3 [26] and rGSTM4* [27]

Broken lines indicate identity with amino acid residues of rGSTM5*. *, G-site residues [18]; †, H-site residues [21]. The secondary structures of rGSTM1 are underlined and specified.

by collecting the data in centroid mode. For the rest of the peptides, the results obtained from ES/MS and calculated from sequencing data deviated less than 1 Da from each other (Table 1).

rGSTM5* was further subjected to endoproteinase Arg-C digestion without prior carboxymethylation, and the resulting peptides were analysed by LC/MS and Edman degradation (Table 1). The recovery of the peptides was drastically less than that from the Lys-C digests, probably due to the presence of unmodified sulphhydryl groups. We sequenced 8 fragments covering 109 residues (Figure 2). One peptide, R8, encompassing residues 43–81, was too large for complete sequencing. We determined the sequence of the first 30 residues and the molecular mass of R8 matched that deduced from the sequences determined from peptides derived from the Lys-C digests. Most importantly, we determined the molecular mass and confirmed the sequence at residues 202–217. The results show indisputably that rGSTM5* has a C-terminal lysine and the data from Lys-C digests cover the whole protein.

The molecular mass deduced from the primary sequence of rGSTM5* was 25495.3 Da. The result is identical to that determined experimentally by LC/MS for the same protein isolated from rat testis. Therefore, rGSTM5* has not been post-translationally modified *in vivo*.

The amino acid sequence of rGSTM5* is listed together with other members of the rat class-Mu enzymes for comparison (Figure 3). rGSTM5* has a sequence that shares 75% (rGSTM1) to 79% (rGSTM2) identity with other class-Mu GSTs from rat.

rGSTM5* is expressed specifically in testis [15]. It has been isolated as rGSTM2–5* and rGSTM3–5* dimers by a combination of S-hexylglutathione affinity chromatography and

chromatofocusing [17]. However, the rGSTM5*–5* homodimer obtained after guanidine hydrochloride treatment of the rGSTM3–5* complex and chromatofocusing was enzymically inactive.

It is noteworthy that residues that are involved in glutathione binding, as illustrated in the crystallographic structure of rGSTM1 [18], are conserved in all five rat class-Mu GSTs (Figure 3). Out of the eight residues that are implicated in the binding of the xenobiotic substrate, 9,10-dihydrophenanthrene, rGSTM5* and rGSTM1 differ in four places [19]. These residues are Ile-9, Gly-111, Tyr-208 and Leu-209 on rGSTM5*, and the corresponding residues on rGSTM1 are Val-9, Ile-111, Phe-208 and Ser-209. The contribution of residues at positions 9 and 111 of class-Mu isozymes in the stereoselectivity and catalytic efficiency towards enones and epoxides has been documented [20]. In addition, the residue at position 209 has been shown to affect the catalytic activity of both class-Mu [21] and class-Alpha GSTs [22]. Therefore, the substrate specificities and kinetic properties of rGSTM5* probably vary significantly from those of rGSTM1. The function, the natural substrate(s) and the properties of this isozyme remain to be investigated.

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

REFERENCES

- 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
- Armstrong, R. N. (1991) Chem. Res. Toxicol. **4**, 131–140
- Bhargava, M. M., Listowsky, I. and Arias, I. M. (1978) J. Biol. Chem. **253**, 4112–4115

-
- 5 Ishigaki, S., Abramovitz, M. and Listowsky, I. (1989) *Arch. Biochem. Biophys.* **273**, 265–272
 - 6 Tan, K. H., Meyer, D. J., Coles, B. and Ketterer, B. (1986) *FEBS Lett.* **207**, 231–233
 - 7 Christ-Hazelhof, E. and Nutgeren, D. (1979) *Biochim. Biophys. Acta* **572**, 43–51
 - 8 Hayes, J. D. and Pulford, D. J. (1995) *CRC Crit. Rev. Biochem. Mol. Biol.* **30**, 445–600
 - 9 Ålin, P., Mannervik, B. and Jornvall, H. (1986) *Eur. J. Biochem.* **156**, 343–350
 - 10 Ålin, P., Jansson, H., Cederlund, E., Jornvall, H. and Mannervik, B. (1989) *Biochem. J.* **261**, 531–539
 - 11 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
 - 12 Taniguchi, H. and Pyerin, W. (1989) *Biochem. Biophys. Res. Commun.* **162**, 903–907
 - 13 Johnson, J. A., Finn, K. A. and Siegel, F. L. (1992) *Biochem. J.* **282**, 279–289
 - 14 Yeh, H.-I., Lee, J. Y., Tsai, S.-P., Hsieh, C.-H. and Tam, M. F. (1996) *Biochem. J.* **314**, 1017–1025
 - 15 Hsieh, C.-H., Tsai, S.-P., Yeh, H.-I., Sheu, T.-C. and Tam, M. F. (1997) *Biochem. J.* **323**, 503–510
 - 16 Simons, P. C. and Vander Jagt, D. L. (1981) *Methods Enzymol.* **77**, 235–237
 - 17 Ishikawa, T., Tsuchida, S., Satoh, K. and Sato, K. (1988) *Eur. J. Biochem.* **176**, 551–557
 - 18 Ji, X., Zhang, P., Armstrong, R. N. and Gilliland, G. L. (1992) *Biochemistry* **31**, 10169–10184
 - 19 Ji, X., Johnson, W. W., Sesay, M. A., Dickert, L., Prasad, S. M., Ammon, H. L., Armstrong, R. N. and Gilliland, G. L. (1994) *Biochemistry* **33**, 1043–1052
 - 20 Shan, S.-O. and Armstrong, R. N. (1994) *J. Biol. Chem.* **269**, 32373–32379
 - 21 Johnson, W. W., Liu, S., Ji, X., Gilliland, G. L. and Armstrong, R. N. (1993) *J. Biol. Chem.* **268**, 11508–11511
 - 22 Liu, L. F., Liaw, Y. C. and Tam, M. F. (1997) *Biochem. J.* **327**, 593–600
 - 23 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
 - 24 Lai, H.-C. J., Grove, G. and Tu, C.-P. D. (1986) *Nucleic Acids Res.* **14**, 6101–6114
 - 25 Lai, H.-C. J. and Tu, C.-P. D. (1986) *J. Biol. Chem.* **261**, 13793–13799
 - 26 Abramovitz, M. and Listowsky, I. (1987) *J. Biol. Chem.* **262**, 7770–7773
 - 27 Lai, H.-C. J., Qian, B., Grove, G. and Tu, C.-P. D. (1988) *J. Biol. Chem.* **263**, 11389–11395
-