# Identification of an N-capping box that affects the α6-helix propensity in glutathione S-transferase superfamily proteins: a role for an invariant aspartic residue

Antonio ACETO*, Beatrice DRAGANI*, Sonia MELINO†, Nerino ALLOCATI‡, Michele MASULLI‡, Carmine Di ILIO* and Raffaele PETRUZZELLI*§

*Istituti di Scienze Biochimiche e ‡Medicina Sperimentale, Università degli Studi 'G. D'Annunzio', Via dei Vestini, 31, I-66013 Chieti, Italy, and †Dipartimento di Biologia, Università di Tor Vergata, Roma, Italy

We have identified an N-capping box motif (Ser/Thr-Xaa-Xaa-Asp) that is strictly conserved, at the beginning of α6 helix, in all glutathione S-transferases (GSTs) and most of the related superfamily proteins. By using CD and peptide modelling we have demonstrated that the capping box residues have an important role in determining the helical conformation adopted by this fragment in the hydrophobic environment of the protein. This is an example in which a local motif, contributing to nucleation of a structural element essential to the global folding of the protein, is strictly conserved in a superfamily of homologous proteins.

## INTRODUCTION

Glutathione S-transferases (GSTs; EC 2.5.1.18) are a family of detoxification enzymes that catalyse the nucleophilic attack of reduced glutathione (GSH) on the electrophilic groups of a wide range of hydrophobic compounds [1–4], including herbicides, insecticides, carcinogens and other xenobiotic substances [5–7]. The cytosolic GSTs have been grouped into five evolutionary classes, Alpha, Pi, Mu, Theta and Sigma, on the basis of N-terminal sequence, substrate specificity and immunological properties [8,9]. A sixth class contains membrane-bound enzymes [10] unrelated in primary structure to the soluble forms. The GSTs are dimeric proteins (molecular mass 46.6 kDa) assembled from identical or non-identical subunits from the same gene class.

Although the aligned amino acid sequences show little identity between the classes, the representative crystal structure of each cytosolic class shows that the overall polypeptide folding is very similar [11–14]. Each subunit is characterized by two distinct domains and possesses an active site that acts independently of the other subunit. The smaller N-terminal domain (domain I) adopts an $\alpha/\beta$ topology and contributes most of the contacts to GSH. The C-terminal domain (domain II) is all α-helical and provides some of the contacts to the hydrophobic binding site, which lies adjacent to the GSH-binding site. A recent study [15] has shown that all GSTs and GST-related proteins, including eukaryotic translation elongation factors [15], bacterial reductive dehalogenases [16,17], β-etherases [18], plant stress-induced proteins [19–21], yeast nitrogen metabolism regulator URE2 [22], S-crystallins from cephalopod eye lens [23–25] and several uncharacterized proteins from different sources with significant sequence similarity to GSTs, are characterized by the presence of two conserved sequence motifs. Motif I is found in domain I and consists of a sequence encompassing β-sheets β4, β5 and the α3 helix. Motif II is in the C-terminal domain and includes a long conserved loop and a subsequent α-helix. Only two residues are strictly conserved in the latter motif, a glycine residue and an aspartic residue corresponding to positions 143 and 150 respectively of pig GSTP1. Some authors have attempted to determine the role of this aspartic residue through site-directed mutagenesis, but the results are somewhat contradictory [26,27].

We here show that Asp-150 has a structural role, being an essential component of a motif named 'capping box' [28–30], present at the beginning of the α6 helix in all known GSTs and GST-related proteins. To examine the structural role of the residues that make up the capping box, we have synthesized the pig α6 peptide encompassing the GST α6 helix. By CD spectroscopy we demonstrate the propensity of this fragment to fold into a helical structure in water/2,2,2-trifluoroethanol (TFE) and water/ethanol solutions; the stabilizing effect of the residues constituting the capping box was established through measurements of the pattern of folding of the pig α6 peptide in which the residues implicated in the capping box had been replaced.

Finally, the possible significance in global folding as well as in the refolding of GST is discussed in detail.

## MATERIALS AND METHODS

### Peptide synthesis and purification

Peptides were synthesized on a Synergy Peptide Synthesizer model 432A (Applied Biosystems), which performs solid-phase synthesis with fluoren-9-ylmethoxycarbonyl derivatives. Cleavage from the resin was achieved with a trifluoroacetic acid/ethanediol mixture. The peptides were purified on a Beckman HPLC apparatus with an Aquapore RP300 reversed-phase analytical and preparative column and a gradient of 0–60% (v/v) acetonitrile containing 0.2% trifluoroacetic acid. The fractions were monitored by absorbance at 220 nm. Purified peptides were freeze-dried and stored at −80 °C.

Automated Edman degradation was performed with an Applied Biosystems model 473A pulsed-liquid sequencer with on-line detection of phenylthiohydantoin amino acids.

Abbreviations used: GSH, reduced glutathione; GST, glutathione S-transferase; TFE, 2,2,2-trifluoroethanol.

§ To whom correspondence should be addressed.

### Protein database and molecular graphics analysis

The Entrez network service (National Center for Biotechnology Information) was used to screen the amino acid sequence database. The GST X-ray structure was taken from the Brookhaven Protein Data Bank via anonymous file-transfer protocol [31]. The crystal structures were rearranged (Hyperchem program; Autodesk, Sausalito, CA, U.S.A.) by energy-gradient minimization with an Amber3-based Polak–Ribiere conjugate gradient minimization algorithm. Rearrangements were stopped when a gradient of less than 0.20 kcal/mol (1 kcal ≡ 4.184 kJ) per Å (1Å ≡ 0.1 $\mu$m) was reached. Local minima in the neighbourhood of the X-ray structure were obtained and considered for further analysis. The structure of the mutated pGSTP1 was obtained by replacing Asp-150 with Glu and rearranging the molecule as described above.

### CD analysis

CD spectra were measured on a Jasco J-700 spectropolarimeter calibrated with camphor sulphonic acid. Spectra were recorded between 190 and 250 nm with a 1.0 nm spectral step size, a 1.0 nm bandwidth and a scan rate of 50 nm/min. A 1 mm sealed and thermostatically controlled quartz cell was used for all CD spectra, except during the determination of concentration dependence experiments in which cuvettes of different pathlengths (1–10 mm) were used. The spectra were the average of five scans and were corrected by five scans of the solvent alone. For every sample the final spectrum was the average of three experiments repeated on different days. Peptide thermal denaturation was monitored by changes in CD spectra in 5 °C steps. The samples (0.3 mg/ml of a particular peptide dissolved in 10 mM citrate buffer, pH 4.8, containing 30% TFE) were equilibrated for 10 min at each temperature. This time of heating was found to be adequate to achieve equilibrium at each temperature. The changes were found to be almost completely reversible, because the spectrum returned to 85–90% of its starting value by lowering the temperature to 0 °C.

Ellipticity is reported as mean residue molar ellipticity ([$\theta$], in degrees·cm²·dmol⁻¹). The peptide helical content was calculated by using the mean residue ellipticity at 222 nm and the equation for the chain-length dependence of helices [32]:

$$[\theta]_\lambda = (f_H - ik/N)[\theta]_{H\lambda\infty}$$

where $[\theta]_\lambda$ is the observed mean residue ellipticity at wavelength $\lambda$, $[\theta]_{H\lambda\infty}$ is the maximum mean residue ellipticity of an $\alpha$-helix of infinite length [33], $f_H$ is the fraction of helix in the molecule, $i$ is the number of helical segments, $N$ is the total number of residues and $k$ is a wavelength-dependent constant (2.57 at 222 nm). Contributions from $\beta$-sheet or random coil were omitted, because no indication of the existence of $\beta$-sheet structure was found and the contribution of random-coil structure to the $[\theta]_{222}$ value was negligible, as empirically determined in 8 M urea. Thus in the peptide used in our study, of chain length 20 residues and with an $f_H$ of 0.8 (obtained by the analysis of the pGSTP1-1 structure), the expected value of the mean residue ellipticity for 100% helicity was 26524 degrees·cm²·dmol⁻¹.

The stock solutions were made by dissolving 1 mg of peptides in 1 ml of 10 mM citrate buffer, pH 4.8, containing 70% (v/v) TFE. The concentration of the stock solution of a particular peptide was monitored by absorption at 280 nm, because all peptides studied contained aromatic residues. During titrations with TFE and ethanol the peptide concentration was 0.3 mg/ml, obtained by appropriate dilutions of the stock solution. At 30–70% (v/v) TFE the CD signals were independent of the peptide concentration in the range 13–300 $\mu$M, indicating that the peptides were essentially monomeric.

## RESULTS

### Identification of structural motifs

As shown in Figure 1a, at the beginning of domain II $\alpha$6 helix, the sequence Ser/Thr-Xaa-Xaa-Asp is conserved in Alpha, Mu, Pi, Theta and Sigma class GSTs. This motif seems to correspond to the 'capping box' signal [29,30] that has been suggested to exert a strong influence on the nucleation and folding of the helix N-terminus [34]. Previous statistical analysis [35] revealed that the residues observed most frequently in a capping box are Ser/Thr as the first helical residue (Ncap residue) and Glu, Gln or Asp located three residues downstream (N3 residue). In GSTs the capping motif (Ser-147 at the Ncap and Asp-150 at the N3 residue in pig lung GSTP1-1) seems to be conserved in a structurally homologous region of the molecule, in the core of the protein and far from the active site. In addition, this sequence is present in all GST and GST-related protein sequences so far analysed (more than 50). In Figure 1(b) are shown, for instance, some representative sequences from distant members of GST superfamily proteins such as bacterial, fish and plant GSTs, and Figure 1(c) shows some sequences from GST-related proteins such as translation elongation factors 1$\gamma$, bacterial dichloromethane dehalogenases, and lignin-degrading $\beta$-etherase.

Thus, despite the low overall sequence identity (in some cases less than 20%), the capping box motif seems well conserved across the GST system and GST-related sequences. Only in some GST-related proteins [15] (as identified by their SWISS-PROT or Genbank entry names: SCII-OMMSL, ARP-TOBAC, PRP1-SOLTU, HS26-SOYBN, CELC29E4-4 and CELC02D5-3) and in GST1 from Carnation (GTT1-DIACA, P28342), this motif seems to be absent from the primary structure. Most of these sequences belong to uncharacterized proteins, and only on the basis of their sequence similarities have they been grouped as 'GST Theta-like' proteins [15]. Their sequence alignment shows [15] that while the Asp residue is still maintained at the N3 position, Ser/Thr at Ncap is often replaced by a glycine residue. Lacking three-dimensional structures of these proteins it is not possible at this stage to explain these differences from the other members of the GST superfamily. In contrast, GST Alpha, GST Mu and GST Pi are more compact groups and the capping box is clearly conserved in their sequences.

The sequence analysis also reveals the presence of a second structural motif at the N-terminus of the $\alpha$6 helix of Alpha, Mu, Pi, Theta and Sigma class GSTs (Figure 1a). This motif consists of a specific hydrophobic interaction, named the hydrophobic staple-motif [30] between the residue located before the Ncap and the residue located within the helix after N3 (I, I+5 interaction). It has been suggested [36] that this accompanying capping box motif enhances the definition of the N-terminus limit and the stabilization of the helices (the $\alpha$6 helix in GSTs). In addition, it should be noted (Figure 1) that four residues before the Ncap a glycine residue is conserved in all GST sequences (Gly-143 in pGSTP1-1), probably indicating a concerted role with the above-described structural motifs. The residues participating in these motifs, together with Gly-143, are the only conserved residues in domain II in all GST and GST-related proteins.

### Molecular graphics analysis

The analysis of the three-dimensional structures of different GST

| NCBI SeqID | Protein | Sec. Struct. α 5 > | | < Ncap N3 | α 6 > | |
|---|---|---|---|---|---|---|
| 121746 | hGSTP1 | 134 | L S Q N Q G G K T F I V G D Q I S F A D Y N L L D L L L I H E V L A P G | | | |
| 544445 | pGSTP1 | 131 | L S Q N Q G G Q A F V V G S Q I S F A D Y N L L D L L R I H Q V L N P S | | | |
| 121747 | mGSTP1 | 134 | L S Q N Q G G K A F I V G D Q I S F A D Y N L L D L L L I H Q V L A P G | | | |
| 232204 | hGSTM2 | 138 | Y S Q F L G K Q P W F L G D K I T F V D F I A Y D V L E R N Q V F E P S | | | |
| 442976 | rGSTM3 | 137 | Y S E F L G K R P W F A G D K V T Y V D F L A Y D I L D Q Y H I F E P K | | | **a** |
| 121730 | hGSTA1 | 138 | K V L K S H G Q D Y L V G N K L S R A D I H L V E L L Y Y V E E L D S S | | | |
| 409146 | Blowfly | 137 | L N T F L E G H Q Y V A G D S L T V A D L A L L A S V S T F E V A G F D | | | |
| 121697 | *Sc. japonicum* | 133 | F E D R L C H K T Y L N G D H V T H P D F M L Y D A L D V V L Y M D P M | | | |
| 476830 | Squid | 135 | L V S N G G G D G F F V G N S M T L A D L H C Y V A L E V P L K H T P E | | | |
| 232216 | Plaice | 149 | E Y L Q K T S G S F V A G K S F S L A D V S V F P G V A Y L F R F G L T | | | |
| 121701 | Chicken | 138 | L S R F L G S R S W F V G D K L T F V D F L A Y D V L D Q Q R M F V P D | | | |
| 121694 | *Dr.melanogaster* | 138 | L N T F L E G Q D Y A A G D S L T V A D I A L V A T V S T F E V A K F E | | | |
| 121695 | Maize | 146 | Y E A R L T K C K Y L A G D F L S L A D L N H V S V T L C L F A T P Y A | | | |
| 232196 | Wheat | 155 | Y E A R L E K S R Y L A G D S I T F A D L N H I P F T F Y F M T T P Y A | | | **b** |
| 642324 | *S. cerevisiae* | 166 | F D T V L R E R P Y V A G D S F S M A D I T V I A G L I F A A I V K L Q | | | |
| 232217 | *P.mirabilis* | 136 | I N D V L S K Q K C V C G D H F T V A D A Y L F T L S Q W A P H V A L D | | | |
| 520921 | *Pseudomonas* sp. | 136 | V A R Q L E H A P Y L L G D Q L S V A D I Y L F V V L G W S A Y V N I D | | | |
| 729643 | *E.coli* | 136 | V N E A L K D E H W I C G Q R F T I A D A Y L F T V L R W A Y A V K L N | | | |
| 433061 | *Methylophilus* sp. | 151 | E K Y W L K D G D Y L C G N T L S Y P D L A T C Q D L V S H D A G R I I | | | |
| 148690 | *Flavobacterium* sp. | 178 | L D V H L A D K P F I A G S N Y S I A D I M W T V L L A R I E M L N M T | | | |
| 119165 | EF1G human | 142 | L D A Y L K T R T F L V G E R V T L A D I T V V C T L L W L Y K Q V L E | | | **c** |
| 242503 | *Dr.melanogaster* GST27 | 137 | L D T F L E D Q E Y V A G D C L T I A D I A L L A S V S T F E V V D F D | | | |
| 134271 | Octopus crystallin | 141 | P F M E R T L E M R N G G N Q F S M G D Q M T M A D L M C Y C A L E N P | | | |

**Figure 1 Multiple alignment of the amino acid sequences of GST and GST-related superfamily proteins**

The 36 amino acid residues belonging to secondary structure elements of helices 5 and 6 of Alpha, Mu, Pi, Theta and Sigma GSTs and the corresponding sequences from distant members of GST and GST-related proteins were aligned manually. The position of the first residue shown relative to the full-length protein is indicated at the left of each sequence. The code numbers of the Entrez network service from which these sequences were obtained are shown in the first column. In the second column are reported the enzyme names following the GST nomenclature scheme or alternatively the protein name given by the authors. (**a**) Sequences of GST belonging to Alpha, Mu, Pi, Theta and Sigma classes with known three-dimensional structures; (**b**) corresponding sequences of GST purified from species distant from mammals; (**c**) corresponding sequences from some representative GST-related proteins. The highly conserved capping box residues are shown in a black box. The Ncap and N3 residues are also indicated at the top of the corresponding column. The highly conserved amino acids residues constituting the hydrophobic staple-motif are shown in a grey box. The highly conserved glycine residue, always located four residues before the Ncap, is shown in a black box. Abbreviation: EF1G, eukaryotic translation elongation factor.

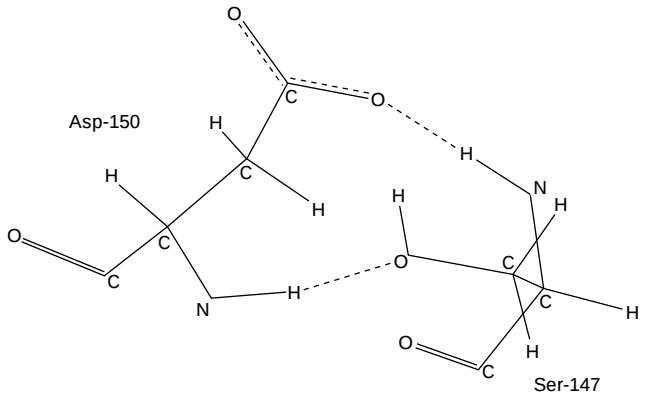


**Figure 2 The capping box hydrogen-bonded circle localized at the N-terminus of the GST α6 helix**

The spatial localization of the residues of interest (shown in atomic detail) is obtained from the published co-ordinates for pGSTP1-1. The side chain of Ncap (OH of Ser-147) forms a hydrogen bond principally with the backbone NH of N3 (Asp 150) and, reciprocally, the side chain of N3 forms a hydrogen bond with the backbone NH of Ncap. The corresponding distances, indicated by broken lines, are given in Table 1.

**Table 1 Capping box hydrogen bonds in the α6 helix of different GSTs**

The Protein Data Bank entry numbers are shown in parentheses after the GST names.

| Hydrogen bond | Distance (Å) Monomer A | Monomer B |
|---|---|---|
| pGSTP1 (PDB 1GSR) | | |
| OD2 (D150)···NH (S147) | 1.67 | 1.72 |
| NH (D150)···OG (S147) | 2.08 | 2.47 |
| hGSTP1 (PDB 1GSS) | | |
| OD2 (D152)···NH (S149) | 1.71 | 1.69 |
| NH (D152)···OG (S149) | 2.18 | 1.99 |
| hGSTA1 (PDB 1QUH) | | |
| OD2 (D156)···NH (S153) | 2.06 | 2.06 |
| NH (D156)···OG (S153) | 3.21 | 3.21 |
| rGSTM3 (PDB 1GST) | | |
| OD2 (D156)···NH (T153) | 1.88 | 1.82 |
| NH (D156)···OG (T153) | 3.12 | 3.59 |

classes [11–14] indicated that the conserved sequence S/TXXD possesses the typical structural properties of the capping box motif. Figure 2 shows the reciprocal side chain–main chain capping interactions as observed in the pGSTP1-1 crystal structure. The side chain O atom (OG) of Ser-147 is within hydrogen-bonding distance from the backbone NH of Asp-150, just as the side chain O atom OD2 Asp-150 forms a hydrogen bond with the main chain NH of Ser-147. The distances between the corresponding residues of other GST structures are also reported in Table 1. In Alpha, Pi and Mu class GSTs the capping box interactions are found within hydrogen-bonding distance. It is worth noting that whereas the Ncap residue can be either serine or threonine, no mutations have been found for the N3 position; the aspartic residue is strictly conserved in all the sequences. To

**Table 2 Peptide sequence and helical content**

$[\theta]_{222}$ is the mean residue ellipticity (degrees · $cm^2 dmol^{-1}$) at 222 nm, 0 °C and pH 4.8; *f* is the percentage helix content calculated as described in the Materials and methods section. Residues shown in bold indicate the substitutions present in the mutant synthetic peptides; residues in italics are the capping residues.

| Peptide | Sequence | $-[\theta]_{222}$ | *f* (%) |
|---|---|---|---|
| Pig$\alpha$6 | Q*I*S*F*A*D*YNLLDLLRIHQVLN | 10760 | 41 |
| Pig$\alpha$6I | Q*I*S*F*A**G**YNLLDLLRIHQVLN | 6660 | 25 |
| Pig6$\alpha$II | Q*I*S*F*A**A**YNLLDLLRIHQVLN | 7100 | 27 |
| Pig6$\alpha$III | Q**I****A**F*A*D*Y*NLLDLLRIHQVLN | 4470 | 17 |

investigate this aspect a mutation Asp-150 → Glu was simulated and the corresponding distances of the capping box interactions were again calculated. Whereas the (Ser-147)OG–NH(Glu-150) interaction remained substantially unaltered, the distance (Glu-150)OE2–NH(Ser-147) increased significantly (7.2 Å), thus becoming less compatible with a hydrogen bond.

## Peptide design

The peptide matching the $\alpha$6 helix of mammalian pGSTP1-1, named Pig$\alpha$6, was synthesized and analysed by CD spectroscopy. Pig$\alpha$6 contained the Ncap residue (Ser/Thr) at the third position to include the possible hydrophobic staple-motif interaction [36]. Moreover, it had free ends to prevent undesired effects of blocking groups on the helical content. To study the role of the conserved capping box sequence in the helical conformation, and to separate its effect from that of a helix macrodipole, we synthesized three other mutant peptides and analysed them by CD spectroscopy. By analogy with a previous study [37], mutant peptides were obtained by substituting Gly or Ala for Asp at the N3 residue (Pig$\alpha$6I and Pig$\alpha$6II respectively) and Ala for Ser at the Ncap site (Pig$\alpha$6III) of the native peptide (Table 2).

## CD analysis of the designed peptides: significance and estimation of helical tendencies

CD spectra of Pig$\alpha$6 peptide in aqueous solution were sufficiently stable only at concentrations as low as 30 $\mu$M as a consequence of its tendency to aggregate, and were therefore no longer considered in this study. In contrast, no aggregation was observed in TFE/water or ethanol/water solutions. TFE titrations were performed on 130 $\mu$M solution of each peptide at low temperature (0 °C) in 10 mM citrate buffer. As shown in Figure 3, the addition of TFE caused a substantial increase in negative ellipticity at 207 nm ($\pi$–$\pi$* transition) and 222 nm ($n$–$\pi$* transition) indicating the formation of an $\alpha$-helix.

To investigate the role of TFE in helix stabilization, comparable titrations were conducted in ethanol/water. The ethanol/water titration of Pig$\alpha$6 at 222 nm was significantly different from that obtained in TFE/water. However, the CD spectra indicated (results not shown) that also in ethanol/water the peptide showed a clear tendency to form an $\alpha$-helix structure,

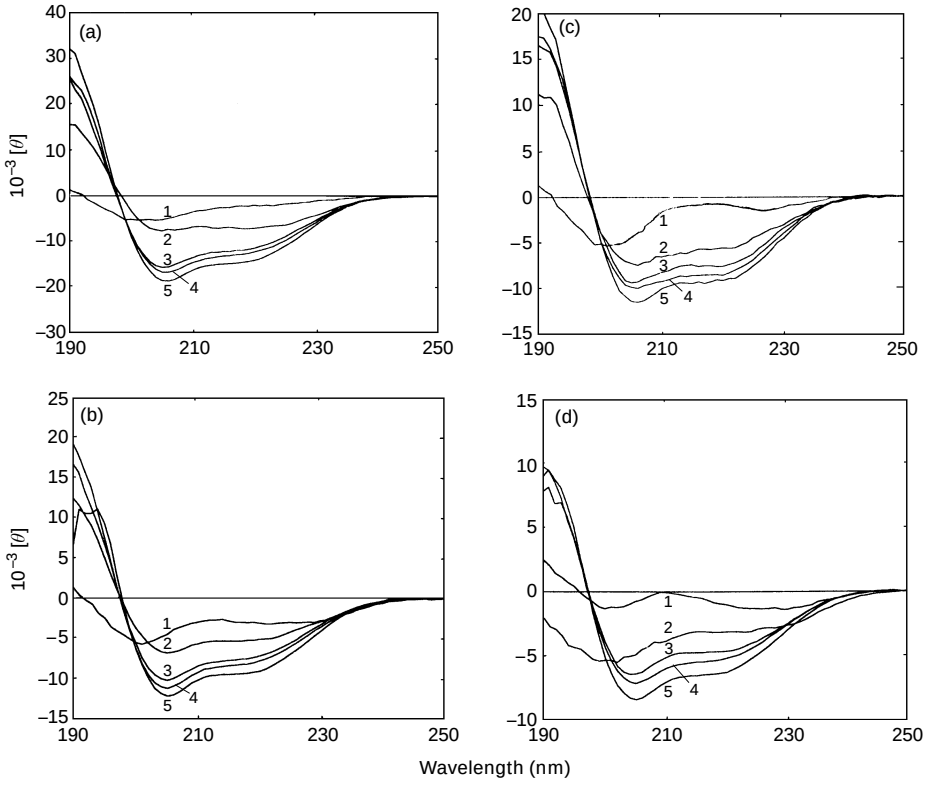


**Figure 3 TFE-dependent conformational transition of the peptides Pig$\alpha$6 (a) and mutant peptides Pig$\alpha$6I (Asp → Gly) (b), Piga6II (Asp → Ala) (c) and Pig$\alpha$6III (Ser → Ala) (d)**

Each CD spectrum sample contained 130 $\mu$M of one peptide dissolved in 10 mM citrate buffer, pH 4.8, at 0 °C and various TFE concentrations. Curves 1–5 represent 10%, 20%, 30%, 50% and 70% TFE (v/v).

with a helical content of 35 % in the presence of 70 % ethanol. This suggested that the stabilizing effect of TFE is predominantly caused by the increased hydrophobicity of the solvent relative to water, which can disrupt intermolecular hydrophobic interactions and at the same time strengthen intramolecular hydrogen bonds.

To investigate the role of the conserved Ncap and N3 residues on the helical tendency of the $\alpha$6 peptide, CD TFE titrations of mutant peptides Pig$\alpha$6I, Pig$\alpha$6II and Pig$\alpha$6III were performed (Figure 3). At 222 nm the mutant peptides showed the same titration behaviour as the parent Pig$\alpha$6 peptide, but the helical tendency of Pig$\alpha$6I, Pig$\alpha$6II and Pig$\alpha$6III peptides at each TFE concentration was much lower (approx. 50 %) than that of the native peptide (results not shown). At 30 % TFE the helical contents of Pig$\alpha$6I, Pig$\alpha$6II and Pig$\alpha$6III were 25 %, 27 % and 17 % respectively (Table 2).

The CD spectra plotted as a function of the temperature in the range 0–70 °C showed that for all peptides the mean residue ellipticity at 222 nm and at 70 °C was about half of the ellipticity at 0 °C (results not shown). These results indicated that the thermal stability of the secondary structure of the Pig$\alpha$6 peptide was not substantially affected by the substitution of the capping residues.

## DISCUSSION

A previous alignment study [38] showed that only two residues, Gly-143 and Asp-150, are conserved in GST domain II. The Asp-150 residue is buried in the hydrophobic core of the protein, close to the N-terminus of the $\alpha$6 helix. Attempts to explain its function by site-directed mutagenesis have given apparently contradictory results. In fact, a lower thermostability was found in the D152A mutant of the human GSTP1-1 than in the wild-type [26], whereas the D157N mutant of the GST YaYa resulted in a complete loss of activity [27]. However, the origin of these marked effects of single mutations of the conserved aspartic residue on the protein has not been investigated until now.

The results obtained in the present work clearly indicate that Asp-150 has an important structural role as a component of the local sequence Ser/Thr-Xaa-Xaa-Asp. This sequence, identified as a capping box motif, is conserved in the GST system. This strict conservation is a novel observation for a helix capping motif.

Considering the function attributed to the capping box residues in helix formation [30], we have also investigated, by using the peptide model and CD measurements, the specific role that this motif could have in the GST superfamily.

CD spectra of the synthesized Pig$\alpha$6 peptide, performed in water and at very low concentration, did not correspond to the typical random-coil observed only at extreme pH (results not shown). The results clearly indicate that the sequence of Pig$\alpha$6 does not possess sufficient information to decide on one 'unique' canonical conformation in water. In contrast, the native $\alpha$-helical structure is strongly stabilized in water/TFE (Figure 3, Table 2) or water/ethanol solutions. Thus, in the presence of solvents less polar than water, the isolated Pig$\alpha$6 peptide showed the same helical conformation as that adopted by the corresponding $\alpha$6 helix in the hydrophobic core of the protein.

In contrast, the mutant peptides Pig$\alpha$6I or Pig$\alpha$6II, in which the N3 residue has been replaced by Gly or Ala respectively, and the mutant peptide Pig$\alpha$6III, in which the Ncap residue has been replaced by Ala, show a much smaller helical tendency than that of the native peptide. This indicates that single substitutions of the residues constituting the capping box have a profound effect on the helical preference of the Pig$\alpha$6. Moreover, the same residue substitutions do not seem relevant for the thermal stability of this peptide. Thus our results clearly show that the conserved local sequence Ser/Thr-Xaa-Xaa-Asp corresponds to a capping box motif and has an important influence on the helical tendency, in a hydrophobic environment, of the isolated Pig$\alpha$6 peptide.

It should be noted that the present results might provide an explanation of previous reports [26,27]. In fact, the $\alpha$6 helix, although located in the core of the protein, makes hydrophobic contacts with an important structural element of the active site ($\alpha$1 helix) of GST. It is conceivable that mutations that decrease the helical tendency of the isolated Pig$\alpha$6 peptide also affect the structural preference of the corresponding $\alpha$6 helix in the protein. Thus protein mutations such as D152A or D157N at the N3 position of the $\alpha$6 helix, which are well known to be incompatible with the capping box motif [37], might also affect in a direct fashion the structure or the stability of the active site.

However, at this stage we cannot exclude the possibility that the conserved capping box residues, involved in $\alpha$6 helix formation, also have a role in the refolding of GST. The present sequence alignment (Figure 1) shows that a hydrophobic-staple motif, flanking the N-capping box, is conserved in GST and GST-related proteins. It has recently been suggested that these local sequences, when simultaneously present at the N-terminus of a helix, have an important role in protein folding by determining the direction of the forming helix with respect to the preceding structural element [36]. Taken together, this observation and the results reported here permit the speculation that the above conserved motifs could be involved in the formation of a partly structured intermediate essential to drive GST folding.

The relation between local sequence motifs and secondary-structure formation is an important and incompletely investigated aspect of the protein-folding problem. In this context the present work constitutes an example in which a local sequence, which enhances the helical propensity of an essential structural element, is strictly conserved in a superfamily of homologous proteins.

We thank Professor Bengt Mannervik (Department of Biochemistry, Uppsala, Sweden) for helpful discussions and for a critical reading of the manuscript, and Dr. F. De Pisa for her expertise in peptide synthesis. This work was supported by grant of the Ministero dell'Università e della Ricerca Scientifica e Tecnologica, and by contributo C.N.R. no. 3603712.CT14.

## REFERENCES

1 Hayes, J. D. and Pulford, D. J. (1995) Crit. Rev. Biochem. Mol. Biol. **30**, 445–600
2 Awasthi, Y. C. and Singh, S. V. (1985) Comp. Biochem. Physiol. **82B**, 17–23
3 Mannervik, B. (1985) Adv. Enzymol. Relat. Areas Mol. Biol. **57**, 357–417
4 Ketterer, B., Meyer, D. J. and Clark, A. G. (1988) in Glutathione Conjugation: Mechanisms and Biological Significance (Sies, H. and Ketterer, B., eds.), pp. 73–135, Academic Press, London
5 Chasseaud, L. F. (1979) Adv. Cancer Res. **29**, 175–274
6 Hayes, J. D. and Wolf, C. R. (1988) in Glutathione Conjugation: Mechanisms and Biological Significance (Sies, H. and Ketterer, B., eds.), pp. 316–356, Academic Press, London
7 Lamoreaux, G. L. and Rusness, L. (1989) in Glutathione – Chemical, Biochemical and Medical Aspects (Dolphin, D., Poulser, R. and Avramovic, O., eds.), part B, pp. 153–196, John Wiley, New York
8 Mannervik, B., Ålin, P., Guthenberg, C., Jensson, H., Thair, M. H., Warholm, M. and Jornvall, H. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 7202–7206
9 Mannervik, B. and Danielson, U. K. (1988) CRC Crit. Rev. Biochem. **23**, 283–337
10 Morgenstern, R., Lundqvist, G., Andersson, G., Balk, L. and DePierre, J. W. (1984) Biochem. Pharmacol. **33**, 3609–3614
11 Reinemer, P., Dirr, H. W., Ladenstein, R., Schaffer, J., Gallay, O. and Huber, R. (1991) EMBO J. **10**, 1997–2005
12 Ji, X., Zhang, P., Armstrong, R. N. and Gilliland, G. L. (1992) Biochemistry **31**, 10169–10184

13 Sinning, I., Kleywegt, G. J., Cowan, S. W., Reinemer, P., Dirr, H. W., Huber, R., Gilliland, G. L., Armstrong, R. N., Ji, X. H., Board, P. G. et al. (1993) J. Mol. Biol. **232**, 192–212
14 Wilce, M. C. J., Board, P. G., Feil, S. C. and Parker, M. W. (1995) EMBO J. **14**, 2133–2143
15 Koonin, E. V., Mushegian, A. R., Tatusov, R. L., Altschul, S. F., Bryant, S. H., Bork, P. and Valencia, A. (1994) Protein Sci. **3**, 2045–2054
16 La Roche, S. D. and Leisinger, T. (1990) J. Bacteriol. **172**, 164–171
17 Orser, C. S., Dutton, J. D., Lange, C., Jablonski, P., Xun, L. and Hargis, M. (1993) J. Bacteriol. **175**, 2640–2644
18 Masai, E., Katayama, Y., Kubota, S., Kawai, S., Yamasaki, S. and Morohoshi, N. (1993) FEBS Lett. **323**, 135–140
19 Czarnecka, E., Nagao, R. T., Key, J. L. and Gupley, W. B. (1988) Mol. Cell. Biol. **8**, 1113–1122
20 Takahashi, Y., Kuroda, H., Tanaka, T., Machida, Y., Takebe, I. and Nagata, T. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 9279–9283
21 Dominov, J. A., Stenzler, L., Lee, S., Schwartz, J. J., Leisner, S. and Howell, S. H. (1992) Plant Cell **4**, 451–461
22 Coshigano, P. W. and Magasanik, B. (1991) Mol. Cell Biol. **11**, 822–832
23 Doolittle, R. F. (1988) Nature (London) **336**, 18
24 Tomarev, S. I. and Zinovieva, R. D. (1988) Nature (London) **336**, 86–88
25 Tomarev, S. I., Zinovieva, R. D. and Piatigorsky, J. (1992) J. Biol. Chem. **267**, 8604–8612
26 Wang, R. W., Newton, D. J., Huskey, S. E., McKever, B. M., Pickett, C. B. and Lu, A. Y. H. (1992) J. Biol. Chem. **267**, 19866–19871
27 Kong, K. H., Inoue, H. and Takahashi, K. (1993) Protein Eng. **6**, 93–99
28 Baker, E. N. and Hubbard, R. E. (1984) Prog. Biophys. Mol. Biol. **44**, 97–179
29 Dasgupta, S. and Bell, J. A. (1993) Int. J. Pept. Protein Res. **41**, 499–511
30 Harper, E. T. and Rose, G. D. (1993) Biochemistry **32**, 7605–7609
31 Bernstein, F., Koetzle, T. F., Williams, B. J. B., Meyer, E. F., Bruce, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasmui, M. (1977) J. Mol. Biol. **112**, 535–542
32 Chen, Y. H., Yang, J. T. and Chau, K. H. (1974) Biochemistry **13**, 3350–3359
33 Chang, T. C., Wu, C. S. and Yang, J. T. (1978) Anal. Biochem. **91**, 13–31
34 Jimenez, M. A., Munoz, V., Rico, M. and Serrano, L. (1994) J. Mol. Biol. **242**, 487–496
35 Richardson, J. S. and Richardson, D. C. (1988) Science **240**, 1632–1652
36 Munoz, V., Blanco, F. J. and Serrano, L. (1995) Nature (London) **2**, 380–385
37 Lyu, P. C., Wemmer, D. E., Zhou, H. X., Pinker, R. J. and Kallenbach, N. R. (1993) Biochemistry **32**, 421–425
38 Wilce, M. C. J. and Parker, M. W. (1994) Biochim. Biophys. Acta **1205**, 1–18

Received 15 July 1996/26 September 1996; accepted 3 October 1996