

# Solution Structure of the Carboxyl Terminus of a Human Class Mu Glutathione S-transferase: NMR Assignment Strategies in Large Proteins

Scott A. McCallum<sup>1</sup>, T. Kevin Hitchens<sup>2</sup> and Gordon S. Rule<sup>2\*</sup>

<sup>1</sup>Department of Biochemistry  
University of Virginia School of  
Medicine, Charlottesville  
VA 22908, USA

<sup>2</sup>Department of Biological  
Sciences, Carnegie Mellon  
University, 4400 Fifth Avenue  
Pittsburgh, PA 15213, USA

Strategies to obtain the NMR assignments for the H<sub>N</sub>, N, CO, C<sup>α</sup> and C<sup>β</sup> resonance frequencies for the human class mu glutathione-S-transferase GSTM2-2 are reported. These assignments were obtained with deuterated protein using a combination of scalar and dipolar connectivities and various specific labeling schemes. The large size of this protein (55 kDa, homodimer) necessitated the development of a novel pulse sequence and specific labeling strategies. These aided in the identification of residue type and were essential components in determining sequence specific assignments. These assignments were utilized in this study to characterize the structure and dynamics of the carboxy-terminal residues in the unliganded protein. Previous crystallographic studies of this enzyme in complex with glutathione suggested that this region may be disordered, and that this disorder may be essential for catalysis. Furthermore, in the related class alpha protein extensive changes in conformation of the C terminus are observed upon ligand binding. On the basis of the results presented here, the time-averaged conformation of the carboxyl terminus of unliganded GSTM2-2 is similar to that seen in the crystal structure. NOE patterns and <sup>1</sup>H-<sup>15</sup>N heteronuclear nuclear Overhauser enhancements suggest that this region of the enzyme does not undergo motion on a rapid time scale.

© 1999 Academic Press

**Keywords:** glutathione S-transferase; NMR; perdeuterated protein; resonance assignment; omega loop

\*Corresponding author

## Introduction

Glutathione S-transferases (GSTs) play an important role in cellular detoxification processes for plants and animals by catalyzing the conjugation of glutathione (*N*-(*N*-L-γ-glutamyl-L-cysteinyl)glycine, GSH) to any one of a large array of hydrophobic xenobiotic and endobiotic compounds (Mannervik & Danielson, 1988; Armstrong,

1997). The benefits of GST catalyzed reactions are twofold. First, the products are generally more soluble and less harmful than the original hydrophobic substrates. Second, the GSH conjugated products are targeted for further degradation and/or clearance by secondary enzymes (Ishikawa, 1992; Armstrong, 1987). Since GSTs recognize many toxic compounds, including carcinogens, GST expression and activity is generally advantageous for the organism. However, deactivation of many anticancer chemotherapeutic agents is proposed to be the result of GST-catalyzed conjugation of these compounds to GSH, thereby lowering their efficacy (Jakoby & Habig, 1980). Isozyme specific deactivation of therapeutics and the increase in expression in drug resistant cancerous cell lines implicate these enzymes in anticancer drug resistance (Baptist *et al.*, 1986; Hayes *et al.*, 1990; Tsuchida & Sato, 1992). In an attempt to obtain a more complete understanding of protein structure-function relationships, including the relationship between dynamics and catalytic func-

Abbreviations used: GST, glutathione S-transferase; GSTM2-2, glutathione S-transferase human class mu isozyme 2-2; rGSTM3-3, glutathione S-transferase rat class mu isozyme 3-3; GSH, *N*-(*N*-L-γ-glutamyl-L-cysteinyl)glycine; NMR, nuclear magnetic resonance spectroscopy; NOE, nuclear Overhauser enhancement; HSQC, heteronuclear single quantum coherence; RMSD, root-mean-square-deviation; 3D, three-dimensional; 4D, four-dimensional; NOESY, NOE spectroscopy; 2D, two-dimensional; hnNOE, heteronuclear NOE.

E-mail address of the corresponding author:  
[rule@andrew.cmu.edu](mailto:rule@andrew.cmu.edu)

tion, significant resources have been directed toward developing structural models that elucidate the mechanism of substrate specificity. A particular interest is the rational design of GST inhibitors and/or chemotherapeutic agents not recognized by GSTs (Rushmore & Pickett, 1993).

In the past few years, numerous crystal structures have supplied a wealth of information about four cytosolic classes of the GST family: mu, pi, alpha, and theta (Sinning *et al.*, 1993; Cameron *et al.*, 1995; Raghunathan *et al.*, 1994; Ji *et al.*, 1992, 1997; Reinemer *et al.*, 1991; Wilce *et al.*, 1995). All GSTs are dimeric and share similar folding topologies. Although the enzymes are dimeric, the monomeric units are functionally independent. Each monomer is structurally organized into two domains that form a cleft that contains the GSH-binding site (G-site) and the xenobiotic substrate-binding site (H-site). Domain I is an  $\alpha$ - $\beta$  structure which is smaller than the  $\alpha$ -helical domain II. The location of GSH-binding site is well defined in most crystal structures and is highly conserved in all classes of GSTs. Although the bound GSH sits at the interface of the domains, it contacts residues mainly from domain I.

In contrast to the highly conserved residues in the G-site, residues that form the H-site are highly variable. The H-site residues in class mu, pi, and alpha enzymes are concentrated to four regions in the primary sequence; the beta-strand containing residues 5-13, the carboxy-terminal end of helix 4, the peptide segment connecting beta-strand 2 ( $\beta$ 2) to helix 2 ( $\alpha$ 2), and the last 12 residues of the carboxyl terminus (C terminus; see Figure 4 for secondary structure and Figure 6a for tertiary structure). Crystal structures of pi-class enzymes suggest that the H-site is formed by residues F8, V10, R13, V35, Y108, N204, and G205 (Reinemer *et al.*, 1991; Oakley *et al.*, 1997; Ji *et al.*, 1997). The crystal structure of the rat mu-class isozyme 3-3 (rGSTM3-3) indicates that residues W7, V9, L12, I111, Y115, F208, and S209 line the H-site (Ji *et al.*, 1994). NMR (Penington & Rule, 1992) and crystallographic studies (Raghunathan *et al.*, 1994) suggest that similar residues in the human mu class 2-2 isozyme (GSTM2-2) participate in binding of hydrophobic substrates. In the class alpha enzyme, residues F10, R13, G14, L107, L108, P110, V111, M208 and F222 appear to be involved with binding of hydrophobic substrates (Sinning *et al.*, 1993; Cameron *et al.*, 1995). Primary sequence alignment of members of a single class of GSTs reveals clusters of sequence variability in the regions that form the H-site. This diversity in the primary sequence likely provides an important means to collectively broaden the substrate specificity within each class of GSTs.

Variation in the primary sequence for the different GST classes also results in class specific structural features. In particular, both the strand connecting  $\beta$ 2 to  $\alpha$ 2 and the C terminus form structural elements that are unique for alpha, mu, and pi-class enzymes. Crystal structures show that the

turn connecting  $\beta$ 2 to  $\alpha$ 2 is extended in the class mu enzymes and occludes the active site region from bulk solvent. For the pi-class enzyme, the same segment is  $\alpha$ -helical and may undergo significant changes in position during the catalytic cycle (Ricci *et al.*, 1996). Crystal structures indicate that the C terminus of the alpha and mu-class enzymes occludes the active site from the bulk solvent to a greater extent than seen in the class pi enzyme. The class alpha enzyme has an even more sterically hindered binding site than the class mu enzyme due to differences in the C terminus. Despite diverse topologies of the  $\beta$ 2- $\alpha$ 2 strand and the C terminus, mutagenesis studies suggest that both regions play an important role in the binding and catalysis of different hydrophobic substrates (Zhang & Armstrong, 1990; Zhang *et al.*, 1992; Board & Mannervik, 1991; Johnson *et al.*, 1993).

Crystallographic studies suggest that the C terminus of the class alpha and mu enzymes is flexible. Structures of the human class alpha GST show a disordered C terminus in the ligand free form of the protein. This region becomes an ordered  $\alpha$ -helix with bound ligand (Cameron *et al.*, 1995). However, the thermal factors remain high for this region in the liganded complex, suggesting some disorder or flexibility for the C-terminal helix. The C terminus of GSTM2-2 appears ordered in the presence of GSH; however, minor conformational differences were observed among seven distinct crystal morphologies (Raghunathan *et al.*, 1994). These conformational differences suggest a degree of flexibility for this region in the liganded form. Currently, there is no X-ray structure of a ligand free class mu GST, therefore, the question remains as to whether the class mu C terminus is disordered in the absence of ligand.

The conformational flexibility of the C terminus may enhance the breadth of substrate specificity by providing a mechanism of generating different H-sites from the same primary sequence. The degree of flexibility of the C terminus may also be relevant to the catalytic cycle for the class mu and possibly the class alpha enzymes. The C terminus in mu class isozymes is found to be over the active site, and sterically hinders ligand binding and product release. A detailed enzymatic analyses of the rGSTM3-3 enzyme indicates that motion of the C terminus may be the rate-limiting step in the release of product in the catalysis of GSH and 1-chloro-2,4-dinitro benzene (Zhang *et al.*, 1992). In the rat 3-3 enzyme, Y115 forms main-chain and side-chain hydrogen bonds with S209. This interaction may stabilize the position of the C terminus and Y115. Replacement of Y115 by Phe disrupts this interaction leading to faster product release rates (Johnson *et al.*, 1993).

The mobility and conformation of the C terminus are likely to be of functional significance in some GSTs. Therefore, a detailed characterization of them is an essential part of understanding the reaction mechanism. NMR spectroscopy is well suited to probe protein structure in solution and to

characterize motions over a wide range of time scales. However, the assignment of resonance frequencies is a prerequisite for the interpretation of NMR studies. While this task is somewhat routine for proteins with a mass less than 25–30 kDa using standard triple resonance techniques, the >50 kDa mass of the GST dimer compromises the ability to obtain resonance assignments because of the short spin-spin relaxation time ( $T_2$ ) associated with larger proteins.

Protein perdeuteration increases the  $T_2$  relaxation times for amide protons and aliphatic carbon nuclei. This significantly improves the sensitivity and resolution of many triple-resonance experiments commonly used for resonance assignments (Torchia *et al.*, 1988; Grzesiek *et al.*, 1993; Yamazaki *et al.*, 1994a,b). Triple-resonance NMR experiments that greatly benefit from protein perdeuteration include the HNCA, HN(CO)CA, HN(CA)CB, HN(COCA)CB, HN(CA)CO, C(CC)(CO)NH, and HN(COCA)NH experiments (Yamazaki *et al.*, 1994a,b; Matsuo *et al.*, 1996; Farmer & Venters, 1995; Grzesiek *et al.*, 1993). A disadvantage of protein perdeuteration is that many experiments that are integral parts of assigning backbone atoms of protonated proteins are not applicable due to the necessity for protons at aliphatic sites. As a result, sequence specific assignments for perdeuterated proteins rely heavily on detecting correlations between amide and intra or inter-residue  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}^\beta$  resonance resonances. Unfortunately, in large proteins,  $^{13}\text{C}$  chemical shift degeneracy and the lack of chemical shifts which are diagnostic of residue type make the linkage of adjacent spin systems and the alignment of linked spin systems within the primary sequence of the protein very challenging. In the case of the 64 kDa tandem dimer trpR complex the CT-HN(CO)CA, CT-HNCA, CT-HN(COCA)CB and CT-HN(CA)CB experiments provided enough chemical shift information to determine assignments of this rather large complex (Shan *et al.*, 1996). Unfortunately,  $\text{C}^\alpha$  and  $\text{C}^\beta$  correlations alone did not yield a large number of unambiguous assignments in our study of GSTM2-2. Similar work in 30–45 kDa proteins suggests that the  $\text{C}^\alpha$  and  $\text{C}^\beta$  connectivities need to be supplemented by additional chemical shift connectivities (Garrett *et al.*, 1997; Caffrey *et al.*, 1997; Constantine *et al.*, 1997).

Here, we report the assignment strategy used for GSTM2-2. The additional inter-residue connectivities were obtained with the HN(CA)CO, HNCO and (H)N(CACO)NH experiments. Specific labeling schemes are described that aided in the alignment of linked spin systems to specific sequences. Finally, the crystal structure of the protein was used to guide the assignment of  $\text{H}_\text{N}$ - $\text{H}_\text{N}$  NOE (nuclear Overhauser enhancement) cross-peaks. An automated Monte Carlo approach was used to combine information from specific labeling, triple-resonance experiments, and NOE experiments to obtain a unique solution for the assignments.

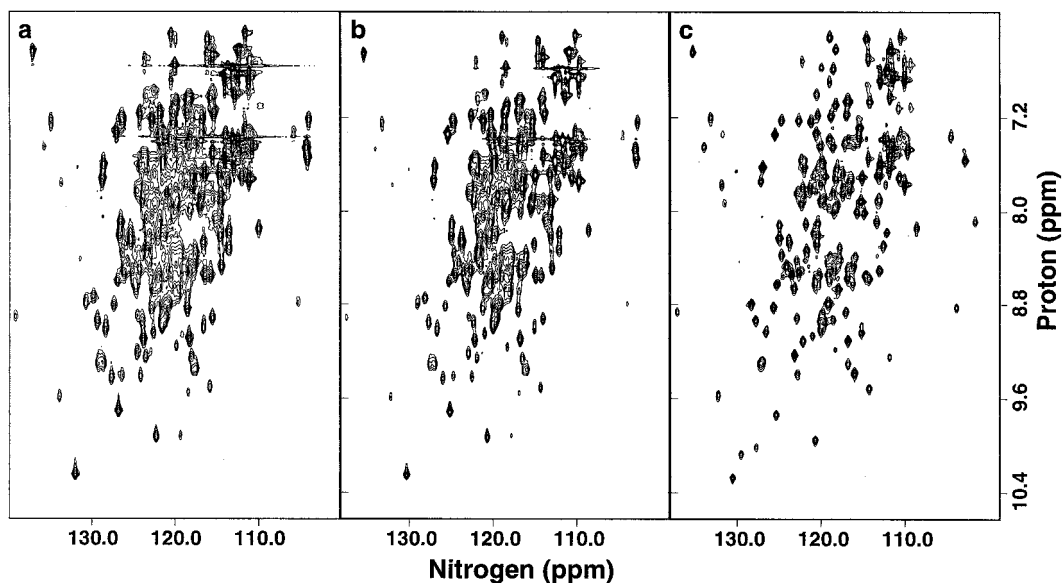
We utilized these assignments to determine the solution structure and to characterize the dynamics of the C-terminal tail for the ligand-free state of the protein. This structural study is a clear demonstration of the applicability of NMR techniques to large proteins. The diverse labeling methods and heteronuclear NMR techniques used here reflect the complexity of the assignment process for proteins as large as GST. The methods presented in this work can be used to obtain resonance assignments in other large proteins and presents opportunities to investigate the structure and dynamics of proteins not presently considered for NMR studies.

## Results and Discussion

### GST deuteration and NMR experiments

In practice, the level of GST deuteration required for adequate signal to noise depended on the NMR experiment. Figure 1 shows a comparison of 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC (heteronuclear single quantum coherence) spectra of GSTM2-2 at increasing levels of protein deuteration. Few resonances are resolved in the spectrum of the 45%  $^2\text{H}$ -enriched GST due to severe spectral overlap (Figure 1(a)). Thus, this level of deuterium incorporation is only useful for observing resonances from slow exchanging amide in spectra acquired in  $^2\text{H}_2\text{O}$ . The quality of the HSQC spectrum improved with increasing deuteration level. Figure 1(b) and (c) shows progressively enhanced spectral resolution as a result of significantly narrower amide proton linewidths. This sharpening of amide proton resonance signals is due to the reduction in the number of coupled protons to the amide group (Torchia *et al.*, 1988; Grzesiek *et al.*, 1995). Although further improvements in resolution are still obtained with perdeuterated protein (Figure 1(c)), relative to that seen for 85% deuteration (Figure 1(b)), the majority of spectral improvements are realized at 85% deuteration. For example, this level of deuterium incorporation provided sufficient spectral resolution and sensitivity in the various 3D (three-dimensional) and 4D (four-dimensional) amide-amide NOESY spectra (see Figure 2) and the  $^{15}\text{N}$  heteronuclear NOE measurement used in this study.

A deuteration level of 85% also provides adequate resolution and sensitivity of triple-resonance experiments that restrict magnetization pathways to the protein backbone (i.e. HNCO, HNCA, and HN(CO)CA experiments, see Figure 3(a) and (c)). For these experiments, a high level of deuteration at the  $\text{C}^\alpha$  position is required to increase the  $T_2$  relaxation time of this carbon. Fortunately, these sites are the most readily deuterated when *Escherichia coli* is grown in minimal media with protonated carbon sources; the ratio of deuterium to proton incorporation at this position is approximately equivalent to that of the  $^2\text{H}_2\text{O}$  content of the growth media (Rosen *et al.*, 1996). This labeling method also leaves many of the methyl groups



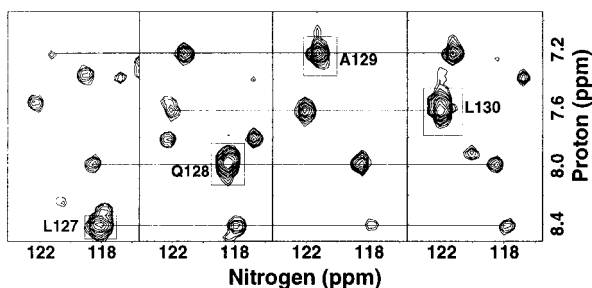
**Figure 1.** Effect of deuteration on the resolution in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of GSTM2-2. Two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra recorded at 298 K at 600 MHz are shown of 45% (a), 85% (b), and perdeuterated (c) GSTM2-2 samples.

partially protonated, which allows the measurement of amide-methyl distance constraints (Metzler *et al.*, 1996; Rosen *et al.*, 1996).

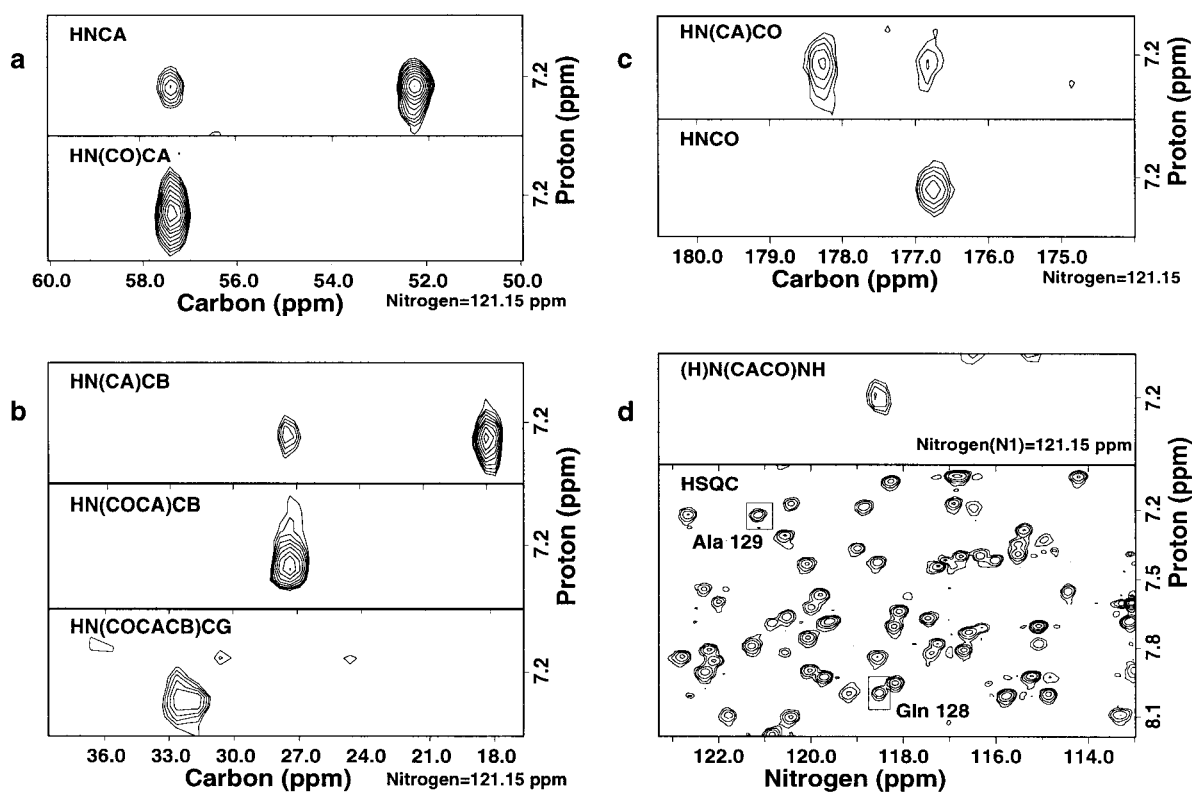
The sensitivity of experiments that measure correlations between carbon side-chain and amide proton resonance frequencies was compromised by residual protons in the 85% deuterated protein. Approximately one-third of the amide resonances had missing cross-peaks in the HN(CA)CB and HN(COCA)CB spectra of the 85% deuterated GSTM2-2. A significant reduction in the sensitivity was also seen by Kay and co-workers in studies of the trp repressor (Shan *et al.*, 1996). As with the trp repressor, the sensitivity of these experiments vastly improved when applied to perdeuterated

GSTM2-2. For virtually all of the amide peaks found in the 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC, both the (*i*) and (*i* - 1)  $\text{C}^\beta$  frequency shifts were also observed (example shown in Figure 3(b)). The sensitivity of these  $\text{C}^\beta$  experiments was sufficiently high that we extended the transfer of magnetization to obtain inter-residue aliphatic  $\text{C}^\gamma$  shifts using the HN(CO-CACB)CG experiment (McCallum *et al.*, 1998). The sensitivity of this latter experiment is an excellent demonstration of the benefits of complete deuteration (see Figure 3(b)).

For all of the experiments mentioned above, it is necessary to exchange amide deuterons for protons after protein biosynthesis. For many proteins, complete amide proton exchange is not a problem. In the case of GSTM2-2, 12-15% of the amide sites do not significantly exchange after four months of incubation in  $\text{H}_2\text{O}$ . Unfortunately, we have not identified conditions that reversibly denature the protein or significantly increase the amide proton exchange rates at these sites. Slowly exchanging amide sites can be problematic for backbone assignments. For example, in GSTM2-2 we find that some amide groups along the buried face of a helical segment exchange too slowly to be observed. These interleaved deuterated and protonated amide groups lead to breaks in sequential connectivities detected in *J*-coupled experiments. However, in cases where segments of slow exchange are sufficiently short (one to two residues), dipolar coupling can be used to potentially resolve these assignment problems. For example, such connectivities helped assign neighboring residues of deuterated amide protons in the helical segment near residue 183 (see Figure 4). Our approach to overcoming connectivity breaks greater than several residues in length was to acquire spectra using a  $^{15}\text{N}$  and  $^{13}\text{C}$  enriched GSTM2-2



**Figure 2.** Representative sections of NOESY spectra of GSTM2-2. The four panels are  $^1\text{H}$ - $^{15}\text{N}$  ( $F_4$ ,  $F_3$ ) cross-sections through the 4D  $^1\text{H}$ - $^{15}\text{N}$ -separated NOESY spectrum of GSTM2-2, and are taken at  $^1\text{H}$  and  $^{15}\text{N}$  frequencies of the self-peaks of the amide protons for residues A127, Q128, A129, and L130, respectively. The self-peak for these residues are boxed. GSTM2-2 was at a monomer concentration of 1.2 mM and was  $^{15}\text{N}$ -enriched and 85% deuterated. The spectrum was recorded at 600 MHz and 298 K. The horizontal lines indicate the NOE cross-peaks between residues.



**Figure 3.** Representative examples of inter-residue scalar connectivities observed in 3D spectra are shown for Ala129. (a) Connectivities between  $C^\alpha$  carbon atoms as observed in the HNCA and HN(CO)CA experiment. (b) Connectivities between  $C^\beta$  carbon atoms as observed in the HN(CA)CB and HN(COCA)CB experiment. Also shown here is the HN(COCACB)CG spectrum (lower spectrum) which provides the resonance frequency for the  $C^\gamma$  of Gln128. (c) Inter-residue CO connectivities. (d) Inter-residue N connectivities. The top spectrum shows the region of the (H)N(CA-CO)NH spectra corresponding with the NH frequencies of Ala129. The lower spectrum shows a portion of the HSQC spectrum; peaks from Ala129 and Gln128 are boxed. The single peak in the upper spectrum represents magnetization transferred from the amide of Gln128 to the amide of Ala129.

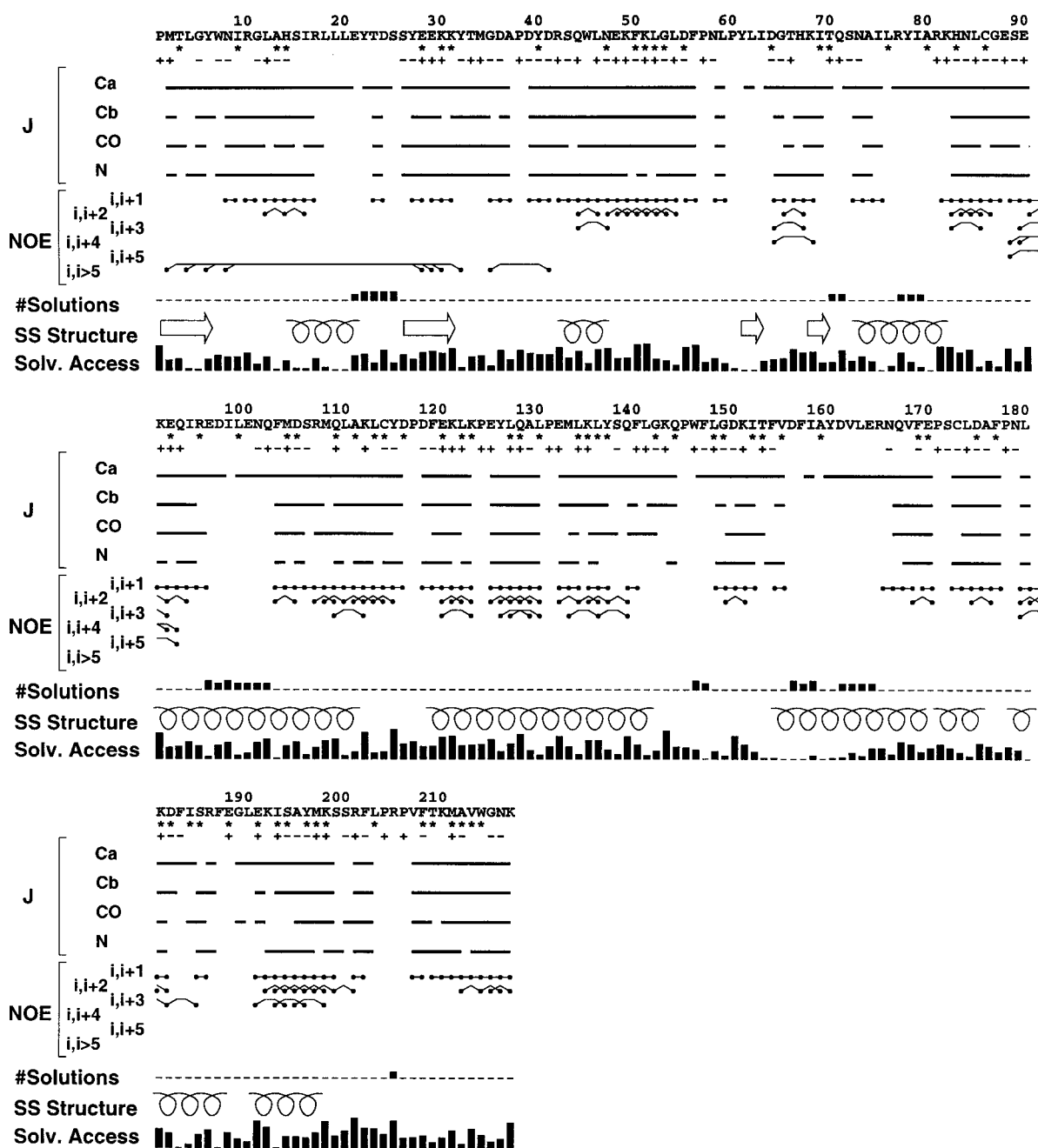
sample that was 45% deuterated. In this sample, all amide sites are approximately half protonated. In practice, it is necessary to simplify the spectra of this sample by acquisition in NMR buffer containing  $\sim 100\%$   $^2\text{H}_2\text{O}$ . Scalar-coupled connectivities between slowly exchanging amide groups and the ( $i$ ) and ( $i-1$ )  $C^\alpha$  resonance shifts were easily measured with the HNCA, and HN(CO)CA experiments. In addition, the connectivities between amide groups and ( $i-1$ ) CO shifts were recorded with the HN(CA)CO experiment. The complementary HN(COCA)CB experiment was not acquired because of poor sensitivity from this sample.

### Assignments

A summary of the nuclear spin connectivities is shown in Figure 4. The automated assignment procedure merged the connectivity and residue type information by calculating the lowest energy for the tentative assignments (see Materials and Methods). Excerpts from the lowest energy solution are shown in Table 1A. The middle section of Table 1A includes residues whose connectivities are displayed in Figures 2 and 3. The reliability of an assignment was assessed by counting how

many times the assignment of a particular residue differed after four independent Monte Carlo runs. An assignment was accepted as correct, if this number is zero (#Solutions, Figure 4). A total of 178 residues of a possible 205 were uniquely assigned. Note that the regions where the assignments are not well defined have minimal connectivities and have relatively low levels of solvent accessibility (see Figure 4). The tentative solutions for these sites are generally alternative arrangements of spin systems observed with the 45% deuterated GSTM2-2 sample. Since, only the  $\text{CO}_{(i-1)}$ ,  $C^\alpha_{(i)}$  and  $C^\alpha_{(i-1)}$  shifts were available for linking these spin systems, there is a high degree of uncertainty for assignments at these sites.

Although, we found that the inter and intra-residue  $C^\alpha$  and  $C^\beta$  shifts were readily detected (see Figure 3(a) and (b)), the large number of peaks, broad linewidths, and poor dispersion, limited the effectiveness of these correlations in the assignment process. For example, a total of less than 90 assignments were consistent with our final assignments when utilizing only  $C^\alpha$  and  $C^\beta$  shifts for residue type identification and inter-residue linkages in our Monte Carlo program. More disturbingly,



**Figure 4.** Summary of information used to assign GSTM2-2. The top line shows the primary sequence of GSTM2-2. (\*) distinguishes residues whose sequential residue type was determined by ( $^{13}\text{C}$ )-specific labeling; residues that had measured aliphatic  $\text{C}^\gamma$  or  $\text{C}^\beta$  shifts in the HN(COCCB)CG spectrum are indicated by a + or -, respectively. The inter-residue  $\text{C}^\alpha$ ,  $\text{C}^\beta$ , CO, and  $\text{N}_\text{H}$  scalar connectivities are shown as horizontal bars. Observed NOE connectivities between amide groups are categorized and displayed as  $(i+1)$ ,  $(i+2)$ ,  $(i+3)$ ,  $(i+4)$ ,  $(i+5)$ , and long range ( $i > 5$ ) transfers. The frequency of alternative assignments for a particular residue is displayed as a vertical bar (#Solutions), a smaller bar indicates fewer solutions. The solvent-exposed surface area of a residue (solvent access) is represented as vertical bars, a higher bar indicates greater solvent exposure. This was calculated from the crystal structure of GSTM2-2 using the surface program MS with a 1.4 Å probe (Connolly, 1983). The secondary structure of the protein (SS structure) is also shown. The coils represent helices and the arrows represent  $\beta$ -strands.

some sequential segments would have been misassigned using this limited information (see Table 1B).

Predicting residue type based on carbon chemical shifts was enhanced with aliphatic  $\text{C}^\gamma_{(i-1)}$  chemical shifts measured from the HN(CO-

CACB)CG spectrum. For a residue lacking an aliphatic  $\text{C}^\gamma$ , this experiment gives an inverted peak at the frequency of the last aliphatic side-chain position, thus restricted that residue to an amino acid lacking an aliphatic  $\text{C}^\gamma$  carbon. However, utilizing this information in conjunction with the  $\text{C}^\alpha$

**Table 1.** Monte Carlo assignments of GST

Residue	MC	CO_L	HN	N	N - 1	C $^{\alpha}$	C $^{\alpha}$ - 1	C $^{\beta}$	C $^{\beta}$ - 1	C $^{\gamma}$ - 1	CO	CO - 1	N1	N2	N3	N4	N5
A.																	
7Trp	[7]	...	8.37	120.4	0.0	56.3	57.7	0.0	38.4	0.00	0.0	0.0	0	0	0	0	0
8Asn	[8]	...	8.66	123.2	120.9	52.3	56.3	36.3	32.8	-1.00	174.6	175.5	9	32	0	0	0
9Ile	[9]	...	8.17	112.1	123.1	58.6	52.3	41.1	36.4	-1.00	174.5	174.6	8	0	0	0	0
10Arg	[10]	Ile	8.17	120.6	112.0	58.8	58.5	30.4	41.2	0.00	173.8	174.5	11	0	0	0	0
11Gly	[11]	...	8.06	101.5	120.8	47.6	58.8	-1.0	30.2	0.00	175.1	173.7	10	0	0	0	0
12Leu	[12]	Gly	8.42	122.9	101.9	55.0	47.6	43.4	-1.0	-1.00	175.2	175.1	13	0	0	0	0
125Glu	[125]	...	7.39	116.6	0.0	58.8	66.1	29.0	30.6	27.72	179.3	179.1	126	127	0	0	0
126Tyr	[126]	...	7.56	122.3	116.6	61.6	58.9	39.4	29.0	35.75	178.5	179.4	125	0	0	0	0
127Leu	[127]	Tyr	8.41	117.8	122.3	57.3	61.6	40.8	39.2	0.00	179.7	178.3	125	126	128	129	130
128Gln	[128]	Leu	7.99	118.5	117.7	57.9	57.3	28.1	40.7	26.54	177.3	179.7	126	127	129	130	0
129Ala	[129]	...	7.17	121.2	118.4	52.8	57.8	18.8	28.3	33.57	178.9	177.3	126	127	128	130	0
130Leu	[130]	Ala	7.60	122.0	121.2	59.1	52.8	39.5	18.8	-1.00	174.4	178.9	128	129	0	0	0
207Val	[207]	...	8.38	124.6	125.5	67.3	60.9	32.1	34.2	25.47	174.6	175.7	208	0	0	0	0
208Phe	[208]	Val	8.39	115.9	124.8	55.4	67.3	42.5	32.1	0.00	174.7	174.6	207	0	0	0	0
209Thr	[209]	Phe	7.36	104.5	116.0	61.1	55.2	71.8	42.4	-1.00	0.0	174.7	210	0	0	0	0
210Lys	[210]	...	7.39	115.5	104.5	58.9	61.0	33.8	71.9	0.00	176.2	175.6	211	0	0	0	0
211Met	[211]	Lys	6.80	111.9	115.4	56.2	58.8	33.0	33.8	0.00	175.2	176.2	210	212	0	0	0
212Ala	[212]	Met	7.22	122.6	112.2	52.6	56.4	18.8	33.0	31.11	175.9	175.3	211	213	214	0	0
213Val	[213]	Ala	8.49	118.9	0.0	63.1	52.6	31.9	18.6	-1.00	176.0	176.0	212	214	0	0	0
214Trp	[214]	Val	7.18	120.4	118.9	57.2	63.0	32.8	31.7	0.00	175.5	176.1	213	215	216	0	0
215Gly	[215]	...	7.38	112.4	120.3	46.2	57.0	-1.0	32.8	0.00	173.3	175.5	216	217	0	0	0
216Asn	[216]	Gly	7.89	113.1	112.6	52.8	46.3	38.7	-1.0	-1.00	174.6	173.3	215	217	0	0	0
217Lys	[217]	...	7.36	125.5	113.0	57.2	52.7	34.4	38.6	-1.00	179.9	174.5	215	216	0	0	0
B.																	
7Trp	[165]	...	7.87	113.8	0.0	58.8	61.6	0.0	0.0	0.00	0.0	179.5	0	0	0	0	0
8Asn	[50]	Lys	8.29	112.6	0.0	58.1	58.9	37.7	32.2	23.03	177.0	175.4	0	0	0	0	0
9Ile	[51]	Phe	8.10	120.4	112.5	55.8	58.1	32.9	37.7	-1.00	176.8	176.9	0	0	0	0	0
10Arg	[52]	Lys	7.64	120.4	0.0	54.8	55.9	44.0	32.9	24.14	176.9	176.8	0	0	0	0	0
11Gly	[53]	Leu	8.14	108.6	120.4	45.5	54.8	-1.0	44.3	25.05	174.3	176.9	0	0	0	0	0
12Leu	[12]	Gly	8.42	122.9	101.9	55.0	47.6	43.4	-1.0	-1.00	175.2	175.1	0	0	0	0	0

Output from the Monte Carlo program is shown for representative segments of the primary sequence. Column labels are: Residue, residue number and residue type; MC, spin-system identification number; CO\_L, peak was present in spectra of ( $1\text{-}^{13}\text{C}$ )-labeled sample of the indicated amino acid; HN, amide proton frequency of the indicated residue; N, amide nitrogen frequency of the indicated residue; N - 1, amide nitrogen frequency of the previous residue; C $^{\alpha}$ , C $^{\alpha}$  shift; C $^{\alpha}$  - 1, C $^{\alpha}$  shift of the previous residue; C $^{\beta}$ , C $^{\beta}$  shift; C $^{\beta}$  - 1, C $^{\beta}$  shift of the previous residue; C $^{\gamma}$  - 1, C $^{\gamma}$  shift of the previous residue; CO, carbonyl shift; CO - 1, carbonyl shift of the previous residue; N1-N5, observed NOEs to indicated residue. A chemical shift of 0.0 indicates the peak was not observed in the spectrum. A chemical shift of -1 encodes residue type information. In case of C $^{\beta}$  shifts, a -1 indicates a Gly residue. In the case of C $^{\gamma}$  shifts, a -1 indicates that the residue lacks an aliphatic C $^{\gamma}$  carbon. A, The first group of residues (7-12) is in a  $\beta$ -strand, and a long-range NOE (column N2) is observed across the sheet. The second group of residues (125-130) starts at the beginning of an  $\alpha$ -helix, thus a large number of local amide-amide NOEs are observed. The last group of residues (207-217) shows output typical for a region that lacks regular secondary structure. The spin system identification number has been set to the residue number in the final assignment solution. B, Output from the Monte Carlo program for residues 7-12 using only C $^{\alpha}$  and C $^{\beta}$  information. Note that residues 7 through 11 are assigned incorrectly, even though the tentative assignments are consistent with the C $^{\alpha}$  and C $^{\beta}$  connectivities.

and C $^{\beta}$  chemical-shifts, only about 124 assignments were identical to those found in the final assignments.

Additional inter-residue chemical shifts were necessary to constrain the assignments. The HN(CA)CO experiment provides the intra-residue CO shift, thus complimenting the inter-residue CO shift obtained from the HNCO spectrum (see Figure 3(c)). Inter-residue scalar coupled correlations of amide nitrogen shifts were obtained from the (H)N(CACO)NH experiment (see Figure 3(d)). In contrast with the HNCA experiment, only 75% of the amide peaks had observable crosspeaks in the HN(CA)CO and HN(CACO)NH spectra of GSTM2-2 (see Figure 4).

Amide-amide dipolar couplings were also an integral part of the assignment process for GSTM2-2. The dipolar transfer of magnetization is very efficient due to the slow rotational correlation time

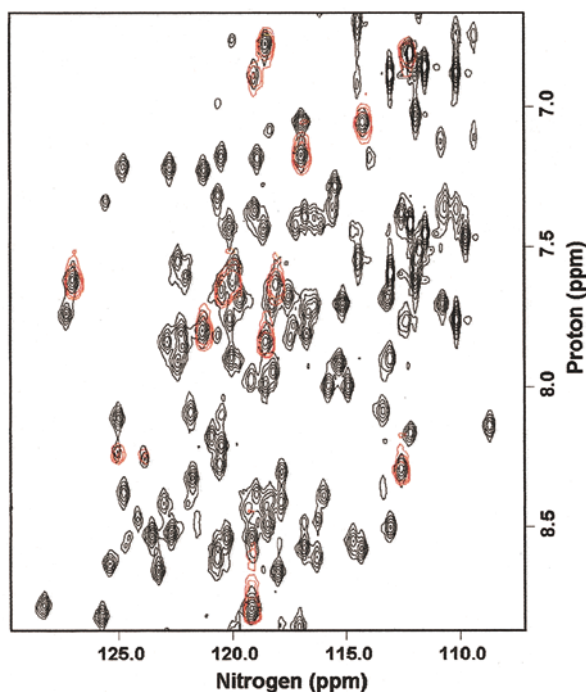
of GSTM2-2 and the reduction in competing amide-aliphatic dipolar coupling in the deuterated protein. As observed in other deuterated proteins (Torchia *et al.*, 1988; Garrett *et al.*, 1997), NOEs were detected between amide sites shown to be at distances of >6 Å in the crystal structure. Figure 4 shows that the vast majority of observed NOE connectivities were local, which is consistent with the patterns expected for a highly alpha-helical protein. However, several long-range NOEs were observed, for example between amides on separate  $\beta$ -strands in the  $\beta$ -sheet.

The scalar intra and inter-residue N, CO, C $^{\alpha}$  and C $^{\beta}$  shift connectivities and amide-amide NOESY connectivities enabled the linkage of short segments of spin systems free of degeneracies. Unfortunately, the large number of breaks in sequential connectivities due to the 12 proline residues and the 25 amide groups that did not exchange, made

unique sequence-specific assignments of these segments difficult.

The correct assignments could be obtained with sufficient residue type information obtained from carbon shifts and specific labeling schemes. One method of determining residue type was to match peaks in HSQC spectra acquired with  $^{15}\text{N}$  specifically labeled GST to peaks in the HSQC spectrum of perdeuterated,  $^{15}\text{N}$  enriched GSTM2-2. Despite the broad peaks in the HSQC spectra of protonated GSTM2-2, specific  $^{15}\text{N}$ -labeling of amino acids generally resolved all the peaks (data not shown). While  $^{15}\text{N}$ -labeling by amino acid type can, in principle, provide residue-specific assignments, a certain degree of uncertainty is associated with the merging of observed shifts in  $^1\text{H}$  and  $^{15}\text{N}$  resonance frequencies from specifically labeled samples relative to those found for the perdeuterated sample. The sources of these uncertainties are differences in sample preparation and unpredictable isotope shifts of the amide frequencies that occur as a result of perdeuteration.

Additional residue specific assignments were obtained by combining  $^{15}\text{N}/^2\text{H}$ -labeling with the residue-specific labeling of a second amino acid type at the carbonyl position ( $1\text{-}^{13}\text{C}$ ; Rule *et al.*, 1993; Lukin *et al.* 1997). A two-dimensional (2D) CO half-filtered  $^1\text{H}\text{-}^{15}\text{N}$  HSQC (HN(CO)) experiment acquired on a protein labeled in this fashion produces a spectrum with a resonance signal for every occurrence in the primary sequence where the ( $1\text{-}^{13}\text{C}$ )-labeled amino acid precedes an  $^{15}\text{N}$ -labeled residue, thereby unambiguously identifying the ( $i - 1$ ) residue type for the detected resonances. The resonance frequencies from the  $^1\text{H}\text{-}^{15}\text{N}$  HSQC, HN(CO), and HNCO spectra of samples labeled with this approach were found to be invaluable in the assignment process of the protein. The distribution of residues that had the ( $i - 1$ ) residue type determined in this manner is indicated in Figure 4. Examples of spectra taken with the  $^1\text{H}\text{-}^{15}\text{N}$  HSQC and HN(CO) experiments are shown in Figure 5. The benefits of this technique were numerous. First, minimal isotope shifts were observed. Second, perdeuteration produced sharp  $^1\text{H}$  lines. Third, the HSQC spectrum usually provided amide signals from the entire protein, facilitating the referencing of spectra and identification of different sample conditions. Fourth, the carbonyl shifts detected with the 3D HNCO experiment were helpful in matching peaks in crowded regions of the 2D spectrum. The selection of particular 1-C labels is specific to the protein being assigned. Our approach was to target residues for labeling on the basis of cost/benefit analysis with consideration to residue frequency and distribution in the primary sequence. Additional residues were chosen as needed to target problem areas. The critical importance of the 1-C-labeling for the assignment of this protein is evident in Table 1. The misassigned segment in Table 1B was corrected by the addition of information from the various 1-C-labelings. The high degree of sensitivity of the HNCO experiment

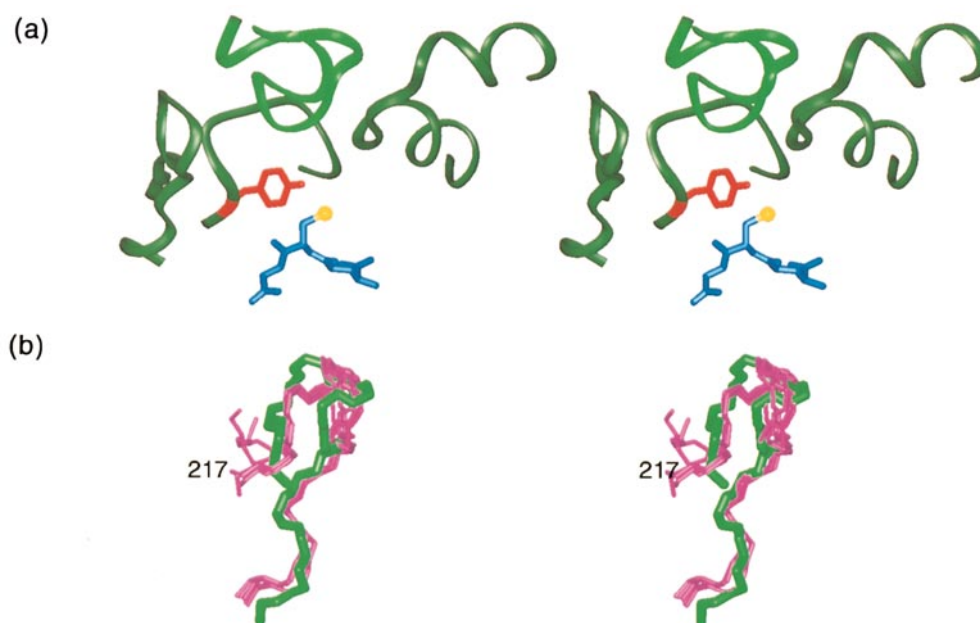


**Figure 5.** An example of residue type determination utilizing carbonyl-labeling. The Figure is composed of a portion of a  $^1\text{H}\text{-}^{15}\text{N}$  HSQC spectrum (black) superimposed on a 2D  $^1\text{H}\text{-}^{15}\text{N}$  HN(CO) spectrum (red) of  $^2\text{H}$ ,  $^{15}\text{N}$ , [ $1\text{-}^{13}\text{C}$ ]Leu-enriched GSTM2-2 sample. The red peaks indicate that these residues follow Leu in the primary sequence.

will make this labeling approach applicable to the assignment of larger proteins.

### Structure and dynamics of the C terminus

The solution structure for the segment of residues between 205-217 of the C terminus was calculated for the ligand-free form using NOESY-derived distance constraints. Of the 20 structures built, the ten lowest energy structures showed an RMSD (backbone atoms, residues 205-217) to the lowest energy structure of 1.4 Å. These structures appear to represent the most populated conformer because the NOE intensities for the C-terminal residues were similar to those measured in other, well defined, regions of GSTM2-2. The solution structure of the C terminus forms an omega loop (Fetrow, 1995) and extends vertically in the same direction as helix  $\alpha 4$  (see Figure 6). The NMR-derived structures are similar to the crystal structure of the protein in complex with GSH, showing an RMSD of 2.1 Å to the highest resolution crystal structure (1hna.pdb; Raghunathan, 1994). A similar topology of this region is also found in the rat 3-3 enzyme complexed with various ligands. Therefore, the conformation of C terminus for ligand-free GSTM2-2 appears to be ordered and does not undergo a large conformational change upon ligand binding. This observation is in contrast to



**Figure 6.** Active site and C terminus structure of GSTM2-2. (a) The green ribbon represent the crystal structure obtained of the GSTM2-2 and GSH complex (1hna.pdb). Only segments for residues 5-13, 30-45, 107-126, and 203-217 are shown. The C terminus is shown in light green, and the other segments are shown in dark green. The active site tyrosine (Y6) is colored red. The bound glutathione is shown in blue, with the sulfur colored yellow. (a) View of the C termini of the crystal structure (green) and the five lowest energy NMR structures (magenta). The orientation is rotated by  $\sim 90^\circ$  relative to the top Figure. The NMR structures were superimposed on the X-ray structure using Insight II (BIOSYM/Molecular Simulations, San Diego, CA).

the crystallographic study of an alpha class enzyme that suggests a large conformational transition of the C terminus from a disordered state to an  $\alpha$ -helix upon ligand binding.

The  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear NOE for residues in the C terminus provide further evidence of an ordered structure. Heteronuclear NOEs (hnNOE) were obtained at two different magnetic field strengths (see Table 2). The hnNOE for some residues exceeds the theoretical maximum due to the effects of chemical exchange with solvent. This effect is exacerbated at the higher magnetic field

due to the longer relaxation time of the water. Therefore, only a qualitative analysis of these data is possible. Since the hnNOE was found to be close to that expected for an immobile residue (theoretical maximum), these values also indicate that this region is not undergoing rapid motions. The lack of rapid motions reported here suggests that the high thermal factors observed in the crystallographic studies are due to static disorder or motion on a time scale of less than  $\omega_{\text{N}}$ .

The egress of the product from the active site of the class mu enzymes likely requires the transient disassociation of the C terminus from the H-site. The results presented above do not exclude this event. Slow, "hinge-type" motions of the C-terminal residues would not affect the heteronuclear NOE, nor would these greatly alter the predominate conformation of the omega loop structure. In the case of GSTM2-2, the hinge region may include residues 200-206. These residues show few  $\text{H}_{\text{N}}\text{-H}_{\text{N}}$  NOESY cross-peaks, and their associated resonance lines are broad, indicative of slow chemical exchange. A detailed analysis of the structure and dynamics of GSTM2-2 in the presence and absence of ligand are currently underway to better ascertain the mechanism of ligand accessibility of the active site.

**Table 2.** Heteronuclear NOEs for residues 205-217 of GSTM2-2

Residue	hnNOE (600 MHz)	hnNOE (750 MHz)
R205	0.86	0.95
P206	—	—
V207	0.82	0.92
F208	0.90	1.00
T209	0.84	0.81
K210	0.79	0.85
M211	—	—
A212	0.87	0.96
V213	0.91	0.91
W214	0.86	0.87
G215	0.85	0.89
N216	0.88	0.81
K217	0.79	0.82
Theoretical maximum	0.84	0.86

The signal-to-noise in the spectra used to obtain these values was approximately 10:1 for the data acquired at 600 MHz and approximately 20:1 for the data acquired at 750 MHz.

## Materials and Methods

### GST expression and isotopic enrichment

The expression system and method of purification for GSTM2-2 have been described (Vorachek *et al.*, 1991).

Labeled GSTM2-2 was expressed by growing the *E. coli* host in a synthetic media (Penington & Rule, 1992) that differed only by the supplemented labeled material, as described below. All isotopes were purchased from Cambridge Isotopes Laboratories, Inc. (Andover, MA). For  $^{15}\text{N}$ -directed labeling of Tyr, Phe, Ala, Val, Leu, Met, and Trp residues in GST, the media was supplemented with the particular  $^{15}\text{N}$ -labeled amino acid. The transaminase deficient cell line, DL39, was used to minimize label dilution (McIntosh & Dahlquist, 1990). The  $^{15}\text{N}$ -directed labeling was also combined with deuteration and individual ( $1\text{-}^{13}\text{C}$ )-specific labeling ( $^{13}\text{C}$ -labeling of the carbonyl position) of Tyr, Phe, Ala, Val, Ile, Leu, Met, Ser, and Lys residues. In these cases, a prototropic strain (AR-58) was used, and the concentration of ( $1\text{-}^{13}\text{C}$ )-labeled amino acid was doubled at the time of induction. The final concentrations of  $^{15}\text{N}$  and ( $1\text{-}^{13}\text{C}$ )-labeled amino acids were 25 mg/l Trp, 40 mg/l Tyr 40, 40 mg/l Phe, and 150 mg/l for the non-aromatic amino acids.

Deuterated GSTM2-2 was prepared using freshly transformed cells plated on  $\text{H}_2\text{O}$  containing minimal plates with the appropriate antibiotics. A single colony was used to inoculate an overnight culture containing 2-3 ml of culture media in 75% deuterium oxide supplemented with 0.3% casamino acids (Difco Laboratories, Detroit, MI). The entire content of this culture was transferred to 25-50 ml of minimal media at 80-85%  $^2\text{H}_2\text{O}$  without Casamino acids. When an  $A_{580}$  of 0.6-1.0 was reached, the culture was scaled up about five- to tenfold in volume using media containing 95-99%  $^2\text{H}_2\text{O}$ . The dilution of the culture was repeated after periods of growth until the desired volume was reached. Protein production was induced at an  $A_{580}$  of 0.8-1.0, and allowed to continue for a minimum of 12 hours. Deuterated and uniformly  $^{15}\text{N}$ -labeled protein was prepared with media containing protonated 4 g/l glucose, 1 g/l  $^{15}\text{N}$  ammonium sulfate, and 98%  $^2\text{H}_2\text{O}$ . Deuterated,  $^{13}\text{C}$ -

labeled samples of GST were obtained using protonated [ $^{13}\text{C}$ ]glucose as the carbon source. Perdeuterated and  $^{13}\text{C}/^{15}\text{N}$ -labeled protein was prepared using Bio-Express cell growth media (Cambridge Isotope Laboratories). Alternatives to glucose as a carbon source for perdeuteration, such as acetate and/or succinate, were not found to be suitable for protein production when used with our cell lines. A fractionally deuterated,  $^{15}\text{N}$  and  $^{13}\text{C}$ -labeled GST sample was prepared by using the same minimal media containing a 50%  $^2\text{H}_2\text{O}$  and 50%  $\text{H}_2\text{O}$  mixture. From the above procedures, proteins with three levels of deuterium incorporation were obtained: 85%, 100%, and 45%. Average levels of deuterium incorporation were estimated using mass spectroscopy (Finnigan-MAT TSQ7000 electrospray-tandem mass spectrometer) in addition to one-dimensional  $^1\text{H}$  NMR spectroscopy.

NMR spectra were recorded at 750, 600, or 500 MHz using Bruker DMX spectrometers equipped with deuterium decoupling and Bruker triple-resonance probes with triple axis pulsed-field gradient coils. All spectra of GST were obtained at pH 7.0 and 298 K in buffer containing 10 mM potassium phosphate, 50 mM sodium chloride, 1 mM dithiothreitol, and 0.02% sodium azide in 5%  $^2\text{H}_2\text{O}$ . The only exception was that 100%  $^2\text{H}_2\text{O}$  solvent used for the 45% deuterated sample. Table 3 provides a summary of NMR experiments used to obtain the resonance and sequence-specific assignments for GSTM2-2. The recycle delay for most NMR experiments was typically 1.3 seconds. In the case of the  $^1\text{H}\text{-}^{15}\text{N}$  heteronuclear NOE experiment, this delay was set to 5.6 seconds which included the 3.4 second presaturation period. Homonuclear (proton-proton) NOE mixing times were 100 ms. Quadrature detection was obtained using the States-TPPI method (Marion *et al.*, 1989b). Pulsed-field gradients were used to suppress undesired coherences and water magnetization (Bax & Pochapsky, 1992). Water suppression schemes varied depending upon the experiment, but generally included gradients applied as ZZ-filters

**Table 3.** Summary of NMR experiments recorded on GSTM2-2

Experiment	Acquired data matrix size (nucleus)				Spectral widths (ppm)				Number of transients
	$t_1$	$t_2$	$t_3$	$t_4$	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	
$^{15}\text{N}$ HSQC <sup>a,j</sup>	128(N)	1024(H)			32	15			4-32
3D NOESY-HSQC <sup>b</sup>	112(H)	38(N)	1024(H)		9.5	26	15		16
3D $^{15}\text{N}\text{-}^{15}\text{N}$ NOESY <sup>c,j</sup>	50(N)	50(N)	1024(H)		32	32	15		32
4D $^{15}\text{N}\text{-}^{15}\text{N}$ NOESY <sup>d</sup>	29(N)	24(H)	27(N)	1024(H)	26	6	26	15	2
HNCA <sup>e,j</sup>	60(C)	40(N)	1024(H)		28	32	15		16
HN(CO)CA <sup>e</sup>	44(C)	34(N)	1024(H)		26	30	15		48
HN(CA)CB <sup>e,j</sup>	45(C)	30(N)	1024(H)		58	32	15		64
HN(COCA)CB <sup>e</sup>	40(C)	30(N)	1024(H)		58	32	15		16
HNCO <sup>f,j</sup>	44(C)	35(N)	1024(H)		15	32	15		4-16
HN(CO) <sup>f,k</sup>	128(N)	1024(H)			32	15			8-128
HN(CA)CO <sup>g</sup>	36(C)	35(N)	1024(H)		15	30	15		64
(H)N(CACO)NH <sup>g</sup>	32(N)	32(N)	1024(H)		30	30	15		96
HN(COCACB)CG <sup>h</sup>	40(C)	30(N)	1024(H)		64	30	15		80
$^1\text{H}\text{-}^{15}\text{N}$ heteronuclear NOE <sup>i,j</sup>	120(N)	1024(H)			27	15			32

<sup>a</sup> Bodenhausen & Rubin (1980); Bax *et al.* (1990).

<sup>b</sup> Talluri & Wagner (1996).

<sup>c</sup> Frenkiel *et al.* (1990); Ikura *et al.* (1990).

<sup>d</sup> Grzesiek *et al.* (1995).

<sup>e</sup> Yamazaki *et al.* 1994a,b).

<sup>f</sup> Muhandiram & Kay (1994).

<sup>g</sup> Matsuo *et al.* (1996).

<sup>h</sup> McCallum *et al.* (1998).

<sup>i</sup> Farrow *et al.* (1994).

<sup>j</sup> Data acquired at both 600 and 750 MHz.

<sup>k</sup> Some data acquired at 800 MHz.

and a “watergate” selection of desired off-resonance signals (Piotto *et al.*, 1992; Talluri & Wagner, 1996). Spectral deconvolution in the time domain (Marion *et al.*, 1989a) and spin-lock purge pulses (Messerle *et al.*, 1989) were used to further suppress residual water signals in some experiments. NMR spectra were processed with FELIX 95 software (BIOSYM Technologies, San Diego, CA) using linear prediction, and zero-filling in the indirectly detected dimensions for 3D and 4D experiments.

## Assignments

Chemical-shift assignments were accomplished utilizing an in-house Monte Carlo algorithm based on that described by Ho and co-workers (Lukin *et al.*, 1997). The advantages of using an automated method are that unbiased solutions result, and that it is easy to determine if assignments are unique. The Monte Carlo algorithm generates trial solutions in an attempt to find the solution with the lowest energy. The random swapping of one or more spin systems generates different trial solutions with the probability of accepting any given solution depending on  $kT$ . At an infinite temperature, all swaps are accepted, even if the swap results in a higher energy solution. As the temperature of the system decreases, the probability that an unfavorable swap is accepted decreases. A final solution is reached after several heating and cooling cycles. The number of residues that were swapped was initially set to one but was allowed to increase to four to five sequential residues as the temperature of the system decreased. For our application, the software uses four  $J$ -coupled chemical shifts for connecting sequential amide resonances:  $C_{(i)}^{\alpha}$  and  $C_{(i-1)}^{\alpha}$  chemical-shifts derived from the HNCA and HN(CO)CA experiments,  $C_{(i)}^{\beta}$  and  $C_{(i-1)}^{\beta}$  chemical-shifts from the HNCB, and HN(COCA)CB experiments,  $CO_{(i-1)}$  and  $CO_{(i)}$  chemical shifts from the HNCO and HN(CA)CO experiments, and amide nitrogen chemical shift of the preceding residue,  $N_{H(i-1)}$ , from the (H)N(CACO)NH experiment. The contribution of each of these individual correlations to the total energy of the solution is determined by using the chemical shift difference between the  $X_{(i-1)}$  (i.e.  $C_{(i-1)}^{\alpha}$ ) shift of one residue and the  $X_{(i)}$  (i.e.  $C_{(i)}^{\alpha}$ ) shift of the preceding residue as the independent variable in a Gaussian energy function. The width and height of this Gaussian function are used to control the tolerance of frequency mismatches and the maximum contribution of each type of chemical shift connectivity to the total energy of the system. Frequency mismatches at the extreme of the Gaussian distribution are assigned a repulsive (positive) energy. The relative contribution of any one type of connectivity to the overall energy is adjustable. This control allows information from spectra with limited sensitivity or resolution to be weighted less significantly. For example, the connectivities between the amide and CO and/or the scalar coupled  $N_H$  resonance frequencies were generally scaled less than the  $C^{\alpha}$  and  $C^{\beta}$  connectivities.

In a manner similar to that described by Lukin *et al.* (1997), we utilized the  $C^{\alpha}$ ,  $C^{\beta}$ , and CO chemical shifts to calculate a residue type probability for each spin system. The expected frequency ranges for carbon shifts were adjusted using the secondary structure as determined from the crystal structure. In addition, we also utilized the chemical shifts derived from the  $(i-1)$  residue to determine a probability for a residue type of the preceding amino acid. For this probability calculation, the  $C_{(i-1)}^{\alpha}$  chemical shift obtained from the HN(CO-CACB)CG experiment was also incorporated.

Additional residue type information was derived from two types of specific labeling schemes:  $^{15}\text{N}$  directed amino acid labeling, and  $(1-^{13}\text{C})$ -directed amino acid labeling with  $^{15}\text{N}$  and  $^2\text{H}$ -labeled GST. These two labeling schemes identify the amino acid type for the  $(i)$  and  $(i-1)$  residue, respectively. Because amide frequencies showed non-uniform isotope frequency shifts in the perdeuterated protein, it is difficult to unambiguously match these with equivalent peaks in the HSQC spectrum of the protonated protein, particularly in crowded regions of the spectrum. Therefore, chemical shifts obtained from the  $^{15}\text{N}$ -directed labeling were input as a table of nitrogen and proton frequencies. A 0.4 and 0.1 ppm chemical shift tolerance was allowed for nitrogen and proton frequencies, respectively, and the program determined a best-fit for this specific assignment data. Each peak in the spectra of the  $[1-^{13}\text{C}]$ amino acid labeled samples was matched to a particular peak in the spectrum of the perdeuterated protein. Unlike the program written by Lukin *et al.* (1997), these  $(1-^{13}\text{C})$ -determined spin systems are not defined as a “constrained set”, but are allowed to swap into all positions in the primary sequence. If the residue type of a peak matches its assigned residue, the total energy of the solution is reduced.

Amide-amide NOE data was also utilized by the Monte Carlo program in conjunction with the X-ray crystal structure. For each backbone amide site, expected NOEs were calculated using the crystal structure with a 5.5 Å cut off. The presence of a predicted cross-peak in the NOESY spectra contributed to the total energy of the tentative assignment. The absence of a predicted peak or the presence of an experimental peak that was not predicted from the structure does not effect the energy.

In determining the assignments of GSTM2-2 information from the known crystal structure was used in two instances; the secondary structure was used to modify chemical shift-based residue type predictions and predicted NOEs were compared to experimental NOEs. To assess the contribution of this information we assigned GSTM2-2 using two different approaches. In the first, no NOE information was utilized. In the second, NOEs were predicted from an extended chain structure. In both of these cases the observed chemical shifts were used to predict the secondary structure of a residue. In the absence of NOEs, only 170 residues were correctly assigned. Addition of the sequential NOEs raised this number to 175 residues. Although the constraints from NOEs did not markedly affect the number of correct assignments, they increased the rate of convergence of the correct solution and increased the confidence level of the resultant assignments.

## C-terminal structure calculations

NMR structures of the C-terminal residues were calculated using X-PLOR 3.81. Forty-two constraints were obtained from NOESY spectra. Thirty-one of these constraints were amide-amide NOEs between residues 207 to 217. Six were NOEs between the amides of residues 209-217 and the methyl group of A212. Two NOEs were from the methyl group and amide of I9 to the amide of R205. Three additional constraints, from a methyl group of V213 to the amide groups of residues C114, Y115, and D116 were also utilized. Methyl protons are not highly deuterated in the 85% deuterated samples; methyl groups of residues A212, V213 and I9 were assigned utilizing intra and inter-residues NOEs. All NOE intensities were converted to distance constraints as described

(Briercheck *et al.*, 1998). During the calculation the positions of residues 1-204 were highly constrained with a total of 3000 HN-HN and HN-HB distances measured from the crystal structure. These distances were constrained to  $\pm 0.1$  Å using a square well potential. Structures were calculated by subjecting the crystal structure to 40 ps of molecular dynamics as the temperature of the system was raised to 2000 K. This was followed by an additional 200 ps of molecular dynamics at 2000 K. The structures were then cooled to 200 K over 40 ps and subjected to 400 cycles of energy minimization. During the initial heating cycle the scale factor for the experimental NOEs was set to zero, and thus they did not contribute to the target function. During the 200 ps time period, the scale factor for these NOEs was gradually raised to 100 and maintained at that level for the rest of the calculation. The scale factor for the constraints derived from the crystal structure was set at 50 for the entire calculation. This protocol was sufficient to randomize the conformation of carboxy-terminal residues prior to adding experimental constraints.

## Acknowledgments

We thank Virgil Simplaceanu for his technical support. The Bruker 600 MHz spectrometer was obtained through an equipment grant from NIH (S10 RR11248-01). This work was supported by a grant from NIH (GM46722) and the Eberly Family Professorship in Structural Biology to G.S.R. Some NMR studies were carried out at the National Magnetic Resonance Facility at Madison (operation subsidized by the NIH Biomedical Research Technology Program under grant RR02301; equipment funded by the University of Wisconsin, NSF Academic Infrastructure Program under grant BIR-9214394, the NIH Shared Instrumentation Program under grants RR2781 and RR08438, the NIH Biomedical Research Technology Program under NIH grant RR02301, the NSF Biological Instrumentation Program under grant DMB-8415048, and the US Department of Agriculture).

## References

- Armstrong, R. N. (1987). Enzyme-catalyzed detoxification reactions: mechanism and stereochemistry. *CRC Crit. Rev. Biochem.* **22**, 39-88.
- Armstrong, R. N. (1997). Structure, catalytic mechanism, and evolution of the glutathione transferases. *Chem. Res. Toxicol.* **10**, 2-18.
- Baptist, G., Tulpurle, A., Sinha, B., Katki, A. G., Meyers, C. E. & Cowan, K. H. (1986). Overexpression of a novel anionic glutathione S-transferase in multi-drug-resistant human breast cancer cells. *J. Biol. Chem.* **261**, 15544-15549.
- Bax, A. & Pochapsky, S. S. (1992). Optimized recording of heteronuclear multidimensional NMR spectra using pulsed field gradients. *J. Magn. Reson.* **99**, 638-643.
- Bax, A., Ikura, M., Kay, L. E., Torchia, D. A. & Tschudin, R. (1990). Comparison of different modes of two-dimensional reverse-correlated NMR for the study of proteins. *J. Magn. Reson.* **86**, 304-318.
- Board, P. G. & Mannervik, B. (1991). The contribution of the C-terminal sequence to the catalytic activity of GST2, a human alpha-class glutathione transferase. *Biochem. J.* **275**, 171-174.
- Bodenhausen, G. & Rubin, D. J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Letters*, **69**, 185-189.
- Briercheck, D. M., Wood, T. C., Allison, T. J., Richardson, J. P. & Rule, G. S. (1998). The NMR structure of the RNA binding domain of *E. coli* rho factor suggests possible RNA-protein interactions. *Nature Struct. Biol.* **5**, 393-399.
- Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Gronenborn, A. M. & Clore, G. M. (1997). Determination of the secondary structure and global topology of the 44 kDa ectodomain of gp41 of the simian immunodeficiency virus by multidimensional nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **271**, 819-826.
- Cameron, A. D., Sinning, I., L'Hermite, G., Olin, B., Board, P. G., Mannervik, B. & Jones, T. A. (1995). Structural analysis of human alpha-class glutathione transferase A1-1 in the apo-form and in complexes with ethacrynic acid and its glutathione conjugate. *Structure*, **3**, 717-727.
- Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science*, **221**, 709-713.
- Constantine, K. L., Mueller, L., Goldfarb, V., Wittekind, M., Metzler, W. J., Yanchunas, J., Jr, Robertson, J. G., Malley, M. F., Friedrichs, M. S. & Farmer, B. T., II (1998). Backbone and side-chain dynamics of uncomplexed human adipocyte and muscle fatty acid-binding proteins. *J. Mol. Biol.* **267**, 1223-1246.
- Farmer, B. T., II & Venters, R. A. (1995). Assignment of side-chain resonances in perdeuterated proteins. *J. Am. Chem. Soc.* **117**, 4187-4188.
- Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. & Kay, L. E. (1994). Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by  $^{15}\text{N}$  NMR relaxation. *Biochemistry*, **33**, 5984-6003.
- Fetrow, J. S. (1995). Omega loops: nonregular secondary structures significant in protein function and stability. *FASEB J.* **9**, 708-717.
- Frenkiel, T., Baur, C., Carr, M. D., Birdsall, B. & Feeney, J. (1990). HMQC-NOESY-HMQC a three-dimensional NMR experiment which allows detection of nuclear Overhauser effects between protons with overlapping signals. *J. Magn. Reson.* **90**, 420-425.
- Garrett, D. S., Seok, Y.-J., Liao, D.-I., Peterkofsky, A., Gronenborn, A. M. & Clore, G. M. (1997). Solution structure of the 30 kDa N-terminal domain of enzyme I of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR. *Biochemistry*, **36**, 2517-2530.
- Grzesiek, S., Hao, R. & Bax, A. (1993).  $^{13}\text{C}$  line narrowing by  $^2\text{H}$  decoupling in  $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -enriched proteins. Application to triple resonance 4D J connectivity of sequential amides. *J. Am. Chem. Soc.* **115**, 4369-4370.
- Grzesiek, S., Wingfield, P., Stahl, S., Kaufman, J. D. & Bax, A. (1995). Four dimensional  $^{15}\text{N}$ -separated NOESY of slowly tumbling perdeuterated  $^{15}\text{N}$ -enriched proteins. Application to HIV-1 Nef. *J. Am. Chem. Soc.* **117**, 9594-9595.
- Hayes, J. D., Pickett, C. B. & Mantle, T. J. (1990). The glutathione S-transferases and their contributions to drug resistance in nature. In *Glutathione S-transferases and Drug Resistance* (Hayes, J. D., Pickett, C. B. & Mantle, T. J., eds), pp. 3-16, Taylor and Francis, London.

- Ikura, M., Bax, A., Clore, G. M. & Gronenborn, A. M. (1990). Detection of nuclear Overhauser effects between degenerate amide proton resonances by heteronuclear three-dimensional nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* **112**, 9020-9022.
- Ishikawa, T. (1992). The ATP-dependent glutathione S-conjugate export pump. *Trends Biochem. Sci.* **17**, 463-468.
- Jakoby, W. B. & Habig, W. H. (1980). *Enzymatic Basis of Detoxification* (Jakoby, W. B., ed.), vol. 2, pp. 63-94. Academic Press, New York.
- Ji, X., Zhang, P., Armstrong, R. N. & Gilliland, G. L. (1992). The three dimensional structure of a glutathione S-transferase from the mu gene class. Structural analysis of the binary complex of isozyme 3-3 and glutathione at 2.2-Å resolution. *Biochemistry*, **31**, 10169-10184.
- Ji, X., Johnson, W. W., Sesay, M. A., Dickert, L., Prasad, S. M., Ammon, H. L., Armstrong, R. N. & Gilliland, G. L. (1994). Structure and function of the xenobiotic substrate binding site of a glutathione S-transferase as revealed by X-ray crystallographic analysis of product complexes with the diastereomers of 9-(S-glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene. *Biochemistry*, **33**, 1043-1052.
- Ji, X., Tordova, M., O'Donnell, R., Parsons, J. F., Hayden, J. B., Gilliland, G. L. & Zimniak, P. (1997). Structure and function of the xenobiotic substrate-binding site and location of a potential non-substrate-binding site in a class pi glutathione S-transferase. *Biochemistry*, **36**, 9690-9702.
- Johnson, W. W., Liu, S., Ji, X., Gilliland, G. L. & Armstrong, R. N. (1993). Tyrosine 115 participates both in chemical and physical steps of the catalytic mechanism of a glutathione S-transferase. *J. Biol. Chem.* **268**, 11508-11511.
- Lukin, J. A., Gove, A. P., Talukdar, S. N. & Ho, C. (1997). Automated probabilistic method for assigning backbone resonances of ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) labeled proteins. *J. Biomol. NMR*, **9**, 151-166.
- Mannervik, B. & Danielson, U. H. (1988). Glutathione transferases-structure and catalytic activity. *CRC Crit. Rev. Biochem.* **23**, 283-337.
- Marion, D., Ikura, M. & Bax, A. (1989a). Improved solvent suppression in one and two-dimensional NMR spectra by deconvolution of time domain data. *J. Magn. Reson.* **84**, 425-430.
- Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989b). Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J. Magn. Reson.* **85**, 393-399.
- Matsuo, H., Kupce, E., Li, H. & Wagner, G. (1996). Use of selective  $\text{C}^{\alpha}$  pulses for improvement of HN(CA)CO-D and HN(COCA)NH-D experiments. *J. Magn. Reson. ser. B*, **111**, 194-198.
- McCallum, S. A., Hitchens, T. K. & Rule, G. S. (1998). Unambiguous correlations between backbone amide and aliphatic gamma resonances in deuterated proteins. *J. Magn. Reson.* **134**, 350-354.
- McIntosh, L. P. & Dahlquist, F. W. (1990). Biosynthetic incorporation of  $^{15}\text{N}$  and  $^{13}\text{C}$  for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Quart. Rev. Biophys.* **23**, 1-38.
- Messerle, B. A., Wider, G., Otting, G., Weber, C. & Wuthrich, K. (1989). Solvent suppression using a spin lock in 2D and 3D NMR spectroscopy with  $\text{H}_2\text{O}$  solutions. *J. Magn. Reson.* **85**, 608-613.
- Metzler, W. J., Wittekind, M., Goldfarb, V., Mueller, L. & Farmer, B. T., II (1996). Incorporation of  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ -[Ile, Leu, Val] into a perdeuterated,  $^{15}\text{N}$ -labeled protein: potential in structure determination of large proteins by NMR. *J. Am. Chem. Soc.* **118**, 6800-6801.
- Muhandiram, D. R. & Kay, L. E. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. ser. B*, **103**, 203-216.
- Oakley, A. J., Rossjohn, J., Bello, M. L., Caccuri, A. M., Federici, G. & Parker, M. W. (1997). The three-dimensional structure of the human pi class glutathione transferase P1-1 in complex with the inhibitor ethacrynic acid and its glutathione conjugate. *Biochemistry*, **36**, 576-585.
- Penington, C. J. & Rule, G. S. (1992). Mapping the substrate binding site of a human class mu glutathione transferase using nuclear magnetic resonance spectroscopy. *Biochemistry*, **31**, 2912-2920.
- Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661-665.
- Raghunathan, S., Chandross, R. J., Kretzinger, R. H., Allison, T. J., Penington, C. J. & Rule, G. S. (1994). Crystal structure of human class mu glutathione transferase GSTM2-2. Effects of lattice packing on conformational heterogeneity. *J. Mol. Biol.* **238**, 815-832.
- Reinemer, P., Dirr, H. W., Ladenstein, R., Schaffer, J., Gally, O. & Huber, R. (1991). The three-dimensional structure of class pi glutathione S-transferase in complex with glutathione sulfonate at 2.3 Å resolution. *EMBO J.* **10**, 1997-2005.
- Ricci, G., Caccuri, A. M., Bello, M. L., Rosato, N., Mei, G., Nicotra, M., Chiessi, E., Mazzetti, A. P. & Federici, G. (1996). Structural flexibility modulates the activity of human glutathione transferase P1-1. Role of helix 2 flexibility in the catalytic mechanism. *J. Biol. Chem.* **271**, 16187-16192.
- Rosen, M. K., Gardner, K. H., Willis, R. C., Parris, W. E., Pawson, T. & Kay, L. E. (1996). Selective methyl group protonation of perdeuterated proteins. *J. Mol. Biol.* **263**, 627-636.
- Rule, G. S., Tjandra, N., Simplaceanu, V. & Ho, C. (1993). Assignment strategies for  $^{15}\text{N}$ - $^1\text{H}$  correlated spectra of large proteins in solution. *J. Magn. Reson. ser. B*, **102**, 126-128.
- Rushmore, T. H. & Pickett, C. B. (1993). Glutathione S-transferases, structure, regulation, and therapeutic implications. *J. Biol. Chem.* **268**, 11475-11478.
- Shan, X., Gardner, K. H., Muhandiram, D. R., Rao, N. S., Arrowsmith, C. H. & Kay, L. E. (1996). Assignment of  $^{15}\text{N}$ ,  $^{13}\text{C}_{\alpha}$ ,  $^{13}\text{C}_{\beta}$ , and HN resonances in an  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  labeled 64 kDa Trp repressor-operator complex using triple-resonance NMR spectroscopy and  $^2\text{H}$  decoupling. *J. Am. Chem. Soc.* **118**, 6570-6579.
- 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., Olin, B., Mannervik, B. & Jones, T. A. (1993). Structure determination and refinement of human alpha class glutathione transferase A1-1, and a comparison with the mu and pi class enzymes. *J. Mol. Biol.* **232**, 192-212.
- Talluri, S. & Wagner, G. (1996). An optimized 3D NOESY-HSQC. *J. Magn. Reson. ser. B*, **112**, 200-205.
- Torchia, D. A., Sparks, S. W. & Bax, A. (1988). Delineation of  $\alpha$ -helical domains in deuterated staphylococcal nuclease by 2D NOE NMR spectroscopy. *J. Am. Chem. Soc.* **110**, 2320-2321.

- Tsuchida, S. & Sato, K. (1992). Glutathione S-transferases and cancer. *CRC Crit. Rev. Biochem. Mol. Biol.* **27**, 337-384.
- Vorachek, W. R., Pearson, W. R. & Rule, G. S. (1991). Cloning, expression, and characterization of a class-mu glutathione transferase from human muscle, the product of the GST4 locus. *Proc. Natl Acad. Sci. USA*, **88**, 4443-4447.
- Wilce, M. C. J., Board, P. G., Feil, S. C. & Parker, M. W. (1995). Crystal structure of a theta-class glutathione transferase. *EMBO J.* **14**, 2133-2143.
- Yamazaki, T., Lee, W., Revington, M., Mattiello, D. L., Dahlquist, F. W., Arrowsmith, C. H. & Kay, L. E. (1994a). An HNCA pulse scheme for the backbone assignment of  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$ -labeled proteins: application to a 37-kDa Trp repressor-DNA complex. *J. Am. Chem. Soc.* **116**, 6464-6465.
- Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. & Kay, L. E. (1994b). A suite of triple resonance NMR experiments for the backbone assignment of  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  labeled proteins with high sensitivity. *J. Am. Chem. Soc.* **116**, 11655-11666.
- Zhang, P. & Armstrong, R. N. (1990). Construction, expression, and preliminary characterization of chimeric class mu glutathione S-transferases with catalytic properties. *Biopolymers*, **29**, 159-169.
- Zhang, P., Liu, S., Shan, S., Ji, X., Gilliland, G. L. & Armstrong, R. N. (1992). Modular mutagenesis of exons 1, 2, and 8 of a glutathione S-transferase from the mu class. Mechanistic and structural consequences for chimeras of isozyme 3-3. *Biochemistry*, **31**, 10185-10193.

*Edited by P. E. Wright*

*(Received 22 July 1998; received in revised form 9 November 1998; accepted 17 November 1998)*