Glubodies: randomized libraries of glutathione transferase enzymes

Eugene W Napolitano, Hugo O Villar, Lawrence M Kauvar, Deborah L Higgins, Doug Roberts, Janis Mandac, Sandra K Lee, Åsa Engqvist-Goldstein, Robert Bukar, Brenda L Calio, Hans Martin Jäck[†] and John A Tainer[‡]

Background: The immunoglobulin framework has been mutagenized to engineer recombinant libraries of proteins as potential diagnostics and novel catalysts, although the often shallow binding cleft may limit the utility of this framework for binding diverse small organic molecules. By contrast, the glutathione S-transferase (GST) family of enzymes contains a deep binding cleft, which has evolved to accommodate a broad range of hydrophobic xenobiotics. We set out to determine whether GST molecules with novel ligand-binding characteristics could be produced by random mutagenesis of segments of the binding cleft.

Results: We have identified two ligand-recognition segments (LRSs) in human GST P1, which are near the active site in the folded protein, but have characteristics indicating that the integrity of their sequence is not essential for the overall structure or activity of the protein. Libraries of GST P1-derived proteins were produced by substituting randomized sequences for an LRS or inserting random sequences into an LRS. The recombinant proteins in the libraries, collectively designated as 'glubodies,' generally retain enzymatic activity but differ markedly both from each other and from the parent enzyme in sensitivity to inhibition by diverse small organic compounds. In some instances, a glubody is inhibited by completely novel structures.

Conclusions: We have shown that a non-antibody framework can be used to create large libraries of proteins with a wide range of binding specificities for small organic molecules. The glubodies provide a rich source of data for correlating the structural and functional features of proteins relevant to ligand binding. The criteria applied for identifying an LRS in GST P1 are generally applicable to other protein frameworks.

Address: Terrapin Technologies, Incorporated, 750 Gateway Boulevard, South San Francisco, CA 94080, USA.

Present address: [†]Loyola University, Chicago, IL 60626, USA and [‡]Scripps Clinic and Research Foundation, La Jolla, CA 92037, USA.

Correspondence: Eugene W. Napolitano
e-mail: gnap@trpntech.com

Key words: ligand recognition segment (LRS), glutathione S-transferase (GST), randomized libraries, substrate specificity

Received: **25 May 1995**
Revisions requested: **16 Jun 1995**
Revisions received: **22 Mar 1996**
Accepted: **16 Apr 1996**

Chemistry & Biology May 1996, **3**:359–367

Introduction

Although immunoglobulins function primarily by recognizing foreign protein and peptide structures [1], they can also bind to carbohydrates and other small molecules [2]. Antisera, hybridomas, and a variety of recombinant antibodies [3] have thus been exploited to produce diagnostic reagents [4] and novel catalysts [5]. There are other families of protective proteins, which have evolved in parallel with the immunoglobulins and function primarily in elimination of both exogenous poisons and toxic metabolic byproducts of oxidative metabolism. Several enzymes can be included in this category; each enzyme manifests a broad spectrum of affinities for chemically diverse compounds, in contrast to mature antibodies, which bind a particular target with high affinity [6]. The glutathione S-transferase (GST) family of enzymes are among the proteins that can recognize the widest range of chemicals [7]. In humans, the GST family comprises at least a dozen cytosolic enzymes. They catalyze the conjugation of the tripeptide glutathione (GSH) to a broad range of hydrophobic electrophiles of both endogenous and exogenous origin, thus eliminating the potentially destructive electrophilic moieties. GSTs seem to have evolved a protein framework optimal for binding small organic molecules of the type often used as drugs, and indeed several drugs are known to be metabolized by GSTs.

Mutagenesis has been extensively used to probe protein structure and function; in general, these studies have focused on particular residues or short segments within the amino-acid sequence. A recent study with GST homodimers has focused on Cys47; mutation of this residue can induce positive cooperativity in ligand binding by the GST subunits [8]. New techniques have been developed for making and screening large numbers of mutants containing short gene segments that have been nearly exhaustively replaced with random sequences.

These studies often concentrate on establishing a correlation between protein structure and the catalytic activity using a particular substrate. For example, libraries of mutagenized β-lactamase [9] and herpes thymidine kinase [10,11] have been selected in *Escherichia coli* for the ability to mediate growth in ampicillin or suboptimal concentrations of thymidine, respectively. Mutant GST libraries have been screened for the ability to metabolize particular selective agents. In one case, mutants were isolated with increased catalytic activity toward merchlorethamine, an electrophilic nitrogen mustard [12]; in another case, phage display was used to generate mutant GSTs with altered specificities for specific glutathione (GSH) derivatives [13]. Although such studies provide information about binding of the enzyme to a particular agent, the selection techniques provide few insights into the range of structures to which the protein framework can be adapted by focused mutagenesis. Since natural GSTs display broad specificity in ligand binding, and the framework is amenable to mutagenesis, this protein provides an appropriate system for examining issues of plasticity in ligand binding.

The GST binding pocket has been subdivided phenomenologically into a GSH-binding domain called the G-site, and an H-site where the hydrophobic substrates are bound prior to being conjugated to GSH [14]. The dimeric human P1-1 isoform (a member of the Pi (P) class of GSTs) is one of the best characterized enzymes of the GSTs isolated from a wide range of plants and animals. GST P1-1 has been cloned [15], and the crystal structure of the homodimer has been determined with a bound ligand [16]. The structure revealed that some portions of the H-site in this enzyme are not critical to the overall domain fold. The architecture of this enzyme is thus more suitable for the directed mutagenesis studies reported here than the Alpha (A) class of GST, which has been used previously to study altered substrate specificity [12,13]. The amino-acid segments that constitute the binding site in the A class are integral to the protein structure.

Our analysis of published GST structures identified two regions near the active site that appear to be suitable for generating novel binding proteins. These ligand recognition segments (LRSs) are largely solvent exposed, show high mobility as measured by crystallographic temperature factors or molecular dynamics simulations, lack obvious importance for the structural integrity of the enzyme or the proposed catalytic mechanism, and participate in substrate recognition. Thus, an LRS is analogous to the somewhat larger CDR3 hypervariable loop in antibodies, which is thought not to contribute to the structural integrity of the antibody, but to be the most important structural segment contributing to antigen specificity [17].

Two general classes of randomized libraries can in principle be generated: one in which the variants have an existing region replaced with a randomized sequence, or one with variants containing an insert of randomized sequence. Here we generate both types of libraries. Three LRS variant libraries of homodimeric GST P1-1 were constructed by PCR mutagenesis. The resulting proteins, designated glubodies, are largely still catalytically active as glutathione S-transferases, confirming that the segments are dispensable for catalytic activity. Significant alterations in inhibitor potencies and selectivity are found among active variants, however, indicating the potential of LRS mutagenesis within the GST framework for generating proteins with varied ligand binding properties.

Results and discussion

Glubody design

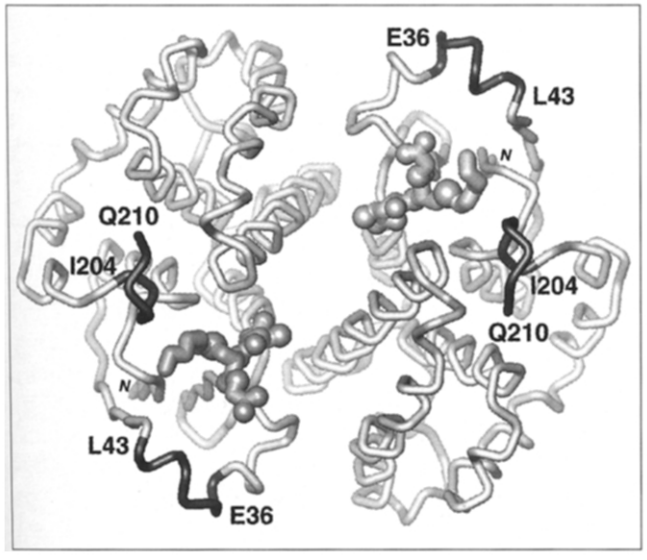
Structural analysis of GST isozymes using Insight II (Biosym Technologies, San Diego, CA) provided the rationale for the development of a diversified set of glubodies. Two crystal structures were available: GST P1-1 from human placenta in a complex with S-hexyl glutathione refined at 2.8 Å resolution [16] and rat Mu (M) class GST 3-3 determined at 2.2 Å resolution in a complex with glutathione (Brookhaven Data Bank code: 1GST) [18]. In addition, the structures of three human M-class isozymes, generated by homology modeling [19], were analyzed.

Because the structure of the enzyme has now been solved in the presence of a bound GSH conjugate, the topological elements involved in GSH and substrate recognition can be identified. The G-site contains the residues directly involved in GSH binding and activation. Tyr7 at this site has been implicated in formation of the thiolate species, which is thought to initiate formation of the glutathione conjugate. In classical protein structure–activity studies, mutations in Tyr7 significantly reduce the catalytic efficiency [20]. Three non-adjacent structural elements with significant flexibility relative to each other form the H-site: the β1–α1 loop (residues 8–15); the carboxyl terminus (residues 201–210); and the α2 helix (residues 36–44), which also forms part of the G-site [16]. Equivalent H-site elements occur in both P- and M-class isozymes.

We aimed to expand the repertoire of ligands recognized by GST recognition, while minimizing structural changes (as monitored by preservation of enzymatic activity). In principle, variation of any H-site residues could alter ligand binding. The goal of maintaining the enzyme's structural and functional integrity, however, merits a more thorough analysis of these three segments.

The first potential LRS (residues 8–15) borders Tyr7, a residue known to be required for catalytic activity. Moreover, the direct interactions of residues 8–15 with neighboring residues that in turn interact with numerous

Figure 1



Glubody replacement loops. The backbone of dimeric human GST P1-1 is oriented to display the LRSs randomized by PCR mutagenesis (shown in black); the ball and stick figure molecule (in dark gray) that is docked into the active site is S-hexyl glutathione. The view shown here is down the molecular two-fold axis and into the ligand-binding site.

non-contiguous residues indicates that the β1–α1 segment is important for maintaining the integrity of domain folding. Randomization of residues 8–15 could thus cause a cascade of structural changes ultimately affecting the catalytic function. In contrast, the α2 helix and the carboxyl terminus, which also form part of the H-site (highlighted in Fig. 1), seem to be dispensable for protein folding and catalysis.

Although residues 36–44 of the α2 helix in the GST P1-1–ligand complex contact the hexyl moiety of the bound ligand S-hexyl glutathione, there are pronounced differences in the sequence of this potential LRS between GST isozymes. For instance, it contains a two residue insertion for the human isozyme compared to the porcine enzyme, with no significant effect on activity or stability [16]. Moreover, the M-class GST isozymes resemble the P-class isozymes in subunit fold and association, but this region is markedly divergent, with no helical character in the reported structure of the M-class isozyme [18]. Thus, modifications in this region may alter substrate recognition without disruption of folding or catalysis.

Two residues in the α2 segment are also involved in GSH recognition. Trp38 acts as a hydrogen bond donor to the Gly residue of GSH and Lys44 forms an important salt bridge with the terminal carboxylate of this substrate. Residues downstream of the α2 segment neighbor the GSH binding area. From all the structural data available, we concluded that alterations to the α2 segment from Glu36 to Leu43 should affect ligand binding, with minimal disruption of enzymatic activity, as desired for an LRS.

The H-site carboxyl terminus (residues 204–210) in the M and P classes is also a potential LRS. It is a solvent-exposed loop, and residues 203–204 neighbor the hexyl moiety of the bound ligand in the crystal structure [16], although they are not involved in GSH recognition. This segment does not interact with other protein subdomains, and thus its integrity should not be important to the overall protein fold. Both the α2 LRS candidate and the carboxy-terminal segment have above average crystallographic temperature factors, suggesting that these regions are somewhat flexible. This may allow inducible formation of complementarity with diverse substrates. The flexibility further suggests that these segments do not make critical contributions to the overall stability of the protein.

Properties of the novel proteins

Using PCR mutagenesis, we created three glubody libraries by randomizing amino acids within regions 36–43 and/or 204–210, or by inserting five random amino acids followed by Pro between Asn206 and Gly207. These LRS-mutagenized libraries contained between 1 x 10^6 and 5 x 10^6 recombinants. Sequencing 20 random clones from each library demonstrated the fidelity of the mutagenesis procedure; greater than 90 % of the clones had randomized sequence in the correct region without changing the register of the reading frame. A minimum of 20 randomly picked clones from each of the three glubody libraries were screened for immunoreactivity by Western blot analysis to verify that the plasmid constructions could yield adequate quantities of protein for analysis. Over 75 % showed significant amounts of recombinant protein from bacterial extracts as visualized with antibodies against both GST P1-1 and a peptide fragment of c-*myc*, which served as a carboxy-terminal immunological tag [21]. This appended peptide tag had a measurable effect on compound inhibition of the recombinant parental enzyme compared to its native form; accordingly, the glubody properties described below are always compared to those of the recombinant parental protein rather than the native GST P1-1.

Randomly picked clones producing substantial amounts of glubody protein were tested individually for GST activity as measured by their ability to conjugate the standard small chromogenic substrate, 1-chloro-2,4 dinitrobenzene (CDNB) to GSH, which causes a change in absorbance at 340 nm. The mutated proteins varied considerably in their enzymatic properties. Of 40 randomly picked glubodies from the three libraries, 21 retained measurable enzyme activity (i.e. at least 3 to 5-fold over background). Several additional clones had low activity, while the activities of

the rest were undetectable over background. Sequencing of all 40 clones showed that they contained randomized sequence in the expected region. Thus, many modifications of the chosen LRS segments of GST P1-1 do not abolish binding of either CDNB or glutathione in the active site.

Of the 21 clones that retained CDNB conjugation activity, 17 express a sufficient amount of recombinant protein to be detected by immunoblotting using anti-GST P1-1 or anti-c-*myc* monoclonal antibodies. The 17 expressors were characterized by a preliminary screen for inhibition of catalytic activity by a panel of 20 compounds, providing an initial inhibition profile. Five of the glubody clones were subsequently examined in much greater detail, beginning with the determination of specific activities relative to the parental enzymes by using CDNB assays. Aliquots from clone extracts were evaluated for protein content by screening dot blots with labeled antibodies to GST. Whereas the recombinant enzyme had activity equivalent to the native enzyme, the glubodies exhibited an ≈17-fold range in specific activity (Table 1). To test for activity from potentially contaminating bacterial proteins in the cell extracts, one clone was grown in larger quantities, purified by S-hexyl glutathione affinity chromatography [22], and retested, with no significant change in specific activity or in the other properties examined below.

To further characterize alterations in binding specificity, the extracts were assayed in the presence of potential inhibitors from a variety of chemical classes. The potency of some of these inhibitors had previously been found to vary significantly among the major natural GST isoforms [23]. Most inhibitors have moderate affinity for the natural P1-1 enzyme, and only 1 out of the 18 compounds selected for screening did not bind significantly to P1-1. The recombinant parental P1-1 exhibits lower affinity for the

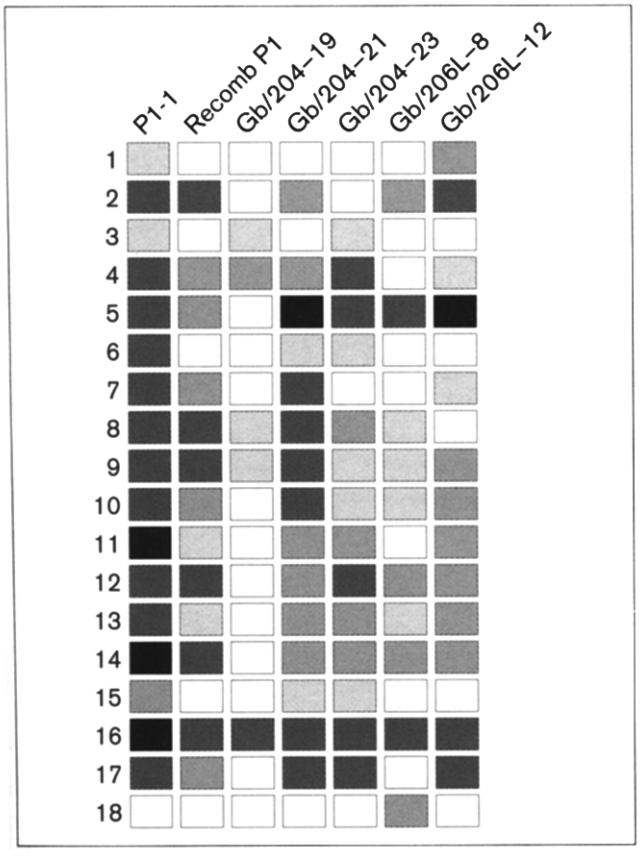
Table 1

Specificity of GST P1-1 and glubodies.

		Enzyme inhibition (−log IC_{50})						
#	Compound name	P1-1	recomb P1	Gb/204-19	Gb/204-21	Gb/204-23	Gb/206L-8	Gb/206L-12
1	Fluorescein isothiocyanate, isomer 1	400 ± 200	N.I.	N.I.	N.I.	N.I.	N.I.	210 ± 30
2	9-Phenyl-2,3,7-trihydroxy-6-fluorone	11.0 ± 3.	80 ± 30.0	N.I.	140 ± 30	N.I.	166 ± 56	90 ± 40
3	4-(6-Hydroxy-3-oxo-3H-xanthen-9-yl)benzoic acid	400 ± 150	N.I.	350 ± 200	N.I.	310 ± 120	N.I.	N.I.
4	6-Hydroxy-3-oxo-3H-xanthene-9-propionic acid	45 ± 5.0	120 ± 50	260 ± 140	140 ± 50	45.8 ± 5.0	N.I.	410 ± 230
5	1,2,3,4-Tetrafluoro-5,8-dihydroxy-anthraquinone, 99 %	16.3 ± 0.7	300 ± 170	N.I.	7.9 ± 2.2	29.0 ± 18.0	77.9 ± 4.4	8.0 ± 1.0
6	Cephaloglycin	20 ± 10	N.I.	N.I.	400 ± 70	300 ± 35	N.I.	N.I.
7	Cephapirin sodium salt	42 ± 9	120 ± 30	N.I.	87 ± 6	N.I.	N.I.	300 ± 100
8	Cephalothin sodium salt	20.0 ± 6	70 ± 20	470 ± 130	44 ± 8	210 ± 50	490 ± 250	N.I.
9	Cephaloridine	30 ± 10	80.0 ± 10	480 ± 140	42.4 ± 5.8	340 ± 130	390 ± 170	180 ± 60
10	Cefoperazone sodium salt, 90 %	39 ± 1	150 ± 50	N.I.	73.3 ± 14.7	330 ± 73	328 ± 12	260 ± 100
11	γ-Glutamyl-S-hexyl cysteinyl-glycine	9.6 ± 0.6	300 ± 150	N.I.	190 ± 40	115 ± 54	N.I.	240 ± 35
12	γ-Glutamyl-S-2-methylnapthalene cysteinyl-glycine	11.0 ± 2.0	50 ± 10	N.I.	135 ± 30	67.1 ± 8.9	220 ± 100	130 ± 50
13	γ-Glutamyl-S-butyl cysteinyl-glycine	80 ± 10	300 ± 100	N.I.	290 ± 75	200 ± 100	360 ± 60	110 ± 8
14	γ-Glutamyl-S-benzyl cysteinyl-glycine	3.0 ± 3.0	23.0 ± 8.	N.I.	300 ± 200	150 ± 74	140 ± 30	140 ± 50
15	2-Mandelamido-4-nitrophenol	200 ± 100	N.I.	N.I.	400 ± 100	310 ± 110	N.I.	N.I.
16	3-Nitro-5,6-benzo-2H-chromene	< 1	12.0 ± 7.0	60 ± 20	10.9 ± 2.1	75.1 ± 3.6	18.0 ± 3.0	20.0 ± 3.0
17	1-Benzene sulfonamideoxime-naphthaquim-4-one	19 ± 2.0	100 ± 65	N.I.	79 ± 22.0	84.1 ± 8.8	N.I.	30.0 ± 20.0
18	3-Cyano-5-pentyl-2,3-dihydro-pyridine-2-thione	N.I.	N.I.	N.I.	N.I.	N.I.	140 ± 50	N.I.
	Specific activity (μM min^{-1} mg^{-1})	57	52	21	3	8	10	16

The inhibitory potency of 18 compounds for native GST P1-1, dimerized recombinant enzyme (recomb P1) and five catalytically active glubodies (Gb) are quantified as −log IC_{50} (μM) calculated from a series of five-fold dilutions. Standard deviations (SD) are based on at least three determinations. Specific activities of glubodies were calculated based on normalization of protein levels assayed by immunological methods. N.I. = no measurable inhibition.

Figure 2



Glubodies have novel profiles of inhibition by a panel of small molecules. Black boxes represent the most potent compounds and white boxes represent no detectable inhibition (numerical data in Table 1). P1-1 is the parental native enzyme; recomb P1 is the recombinant enzyme expressed with a carboxy-terminal c-*myc* tag. The gray scale thresholds of IC_{50} are, from black to white: < 10 μM; 10–100 μM; 100–300 μM; 300–500 μM; no measurable inhibition.

compounds compared to the wild type, apparently due to the c-*myc* peptide tail, but parallels its inhibition patterns overall. The H-site is in part formed by the carboxyl terminus and proximal alterations in amino acid sequence are thus likely to influence inhibitor binding patterns of the tagged protein. The variation of the inhibition profiles of the recombinant P1-1 compared to the natural form is, however, no more extensive than that between the natural variants, even using a set of inhibitors known to differ in potency for the various natural GSTs.

The IC_{50} data for the five glubodies examined in detail (Table 1) show that some of the novel proteins have markedly different inhibitor profiles using this panel of compounds. The qualitative differences are easily seen by plotting the data with a gray scale representing the inhibition potencies (Fig. 2). In general, the glubodies have a wide range of reactivity with the panel of inhibitor compounds. Some glubodies, such as Gb/204-21 and Gb/204-23, recognize more compounds than the recombinant parent, demonstrating that this approach can generate enzymes with broader specificities. By contrast, Gb/206L-8 has a more restricted range of inhibitor susceptibility; this change is accompanied, however, by novel recognition of compound #18, which does not bind to any of the other proteins. An even more extreme change in specificity is seen in Gb/204-19, which is not strongly inhibited by any of the GSH conjugates tested, despite the fact that this enzyme has high specific activity using the standard CDNB substrate. This glubody is particularly interesting as it included a shift in reading frame, resulting in expression of a protein with an extreme carboxy-terminal alteration (Table 2). This frame shift extends the carboxyl terminus and clearly affects the ability of the protein to recognize ligands, although it exerts a much less pronounced effect on the catalytic activity (Table 1). Unlike Gb/204-19, gluthathione S-transferases of the A and M classes have extended carboxy termini that do not appear to increase their substrate specificities.

The correlations of –log IC_{50} values determined for each glubody against all compounds were calculated in a pairwise fashion. The degree of correlation was assessed by computing correlation coefficients (S-Plus advanced data analysis software, StatSci, Seattle, WA). For all pairs of the natural human isozymes examined, as well as with the recombinant P1-1, the natural isozymes appear highly correlated (not shown). The observed variability in the glubodies is summarized in Table 3 by calculating the pairwise correlation coefficients of the affinities (expressed as –log IC_{50}) for the compounds shown in Table 1. The wild type and recombinant parental enzyme show

Table 2

Sequences of randomized regions of glubodies.

Clone	Amino acid 204-210
Parental	INGNGKQ
Gb/204-21	NCARCNT
Gb/204-23	TGYGCVL
Gb/204-19	QVWGSGGRPQNKNSSQKRL
Clone	**N206 – loop insertion – G207**
Gb/206L-8	VDLPCP
Gb/206L-12	SNVCLP
Clone	**Amino acid 36-43**
Parental	ETWQEGSL
Gb/36-1	FRRLLPSL
Gb/36-2	WRVVEILV

Sequence position refers to the wild type P1 monomer. The underlined P refers to an added proline residue after the insertion of five random amino acids. This proline residue is not present in the native enzyme. Sequences from the Gb/36 library represent clones that underwent some preliminary testing in enzyme-inhibition assays.

Table 3

Correlations in enzyme-inhibition properties.

	P1-1	rP1-1	Gb/204-19	Gb/204-21	Gb/204-23	Gb/206L-8	Gb/206L-12
P1-1	1.000						
rP1-1	0.836	1.000					
Gb/204-19	0.291	0.356	1.000				
Gb/204-21	0.749	0.720	0.307	1.000			
Gb/204-23	0.503	0.390	0.314	0.547	1.000		
Gb/206L-8	0.470	0.541	0.254	0.478	0.175	1.000	
Gb/206L-12	0.592	0.591	0.025	0.647	0.412	0.502	1.000

Glubody variants of GST P1-1 with high sequence identity range in character from similar to widely divergent in binding properties compared to the recombinant parental protein or to each other. Pairwise correlation coefficients calculated from the data in Table 1 are shown.

the highest correlation, indicating the highest degree of similarity in affinity patterns among all the proteins compared. Among the novel glubodies, pairs of effectively uncorrelated patterns of affinity can be seen, such as for Gb/206L-8 and Gb/204-23, or Gb/206L-12 and Gb/204-19. Evidently, LRS mutagenesis generates glubodies that vary widely in sensitivity to inhibitors, used as a measure of binding affinities for individual compounds. No correlation is apparent between binding specificity and specific activity against CDNB, or between binding properties and mutation type, at least within an LRS.

Enzyme-inhibition data were also collected for individual members of the Gb/36 library, in which the other candidate LRS, from Glu36 to Leu43, was mutagenized. Two clones (sequences in Table 2) seem to be as different from each other, and from the other glubodies, as any of the other pairs. Similarly, spot checking of several additional clones from the carboxy-terminal libraries suggests that the clones examined in detail and documented in Table 1 do not exhaust the diversity of this library; rather, every clone examined seems to be significantly different from most of the others.

Implications

The defining features of an LRS found in GST are discernable in many proteins in the crystallographic database. The salient elements are a solvent-exposed linear segment of ~8 residues that is adjacent to the substrate binding site, and low relevance of the segment to the overall structural integrity of the protein as indicated by few non-local contacts and higher than average flexibility. Flexibility can be assessed by crystallographic temperature factors or mobility in molecular dynamics simulations. In both the α2 and carboxy-terminal LRS, some residues are in direct contact with the bound ligand on one side and are solvent exposed on the other side. In neither case do the mutated regions have direct interactions with other regions of the protein, and the mutations are thus not likely to affect the stability of the protein. Mutagenizing regions with significant flexibility within a protein does not severely disrupt its folding, as we have shown here using GST and the sensitive assay of catalytic activity.

Glubody variants of GST P1-1 range in character from similar to widely divergent in ligand binding properties compared to the parental protein with which they share nearly 97% sequence identity (Table 2). It is interesting to note that glubody variants with totally different variable regions can have either similar or divergent binding properties. Thus, glubodies provide a rich source of data for exploring the relationship between protein structure and function, and the techniques described here provide tools for engineering proteins with particular desirable properties [24].

Other mutagenesis studies have been carried out to alter the ligand-binding properties of GST. Gulick and Fahl [12] mutated several regions of a rat 2-2 GST A-class enzyme to search for proteins better able to detoxify an electrophilic toxin, mechlorethamine. The successful clones were mutated primarily in residues 9–11, supporting our identification of residues 8–15 as a likely target for altering substrate specificity. A bacteriophage library has recently been prepared [13] that displays an A-class GST that had previously had selected hydrophobic amino acids mutagenized. Using a panning technique, phage were selected for the ability to bind a set of glutathione derivatives, and thus it was not possible to determine how many of the unselected clones were stably folded but had altered substrate specificity. The selected GST mutants had drastic reductions in specific activity for CDNB, generally one to three orders of magnitude lower than the least active glubody described here.

Nonetheless, the success of the mutagenized GST panning experiments supports the possibility that the LRS mutagenesis strategy outlined here could be applied to assay properties of the novel binding sites by non-enzymatic means, such as competitive ligand binding. Further support is provided by studies on phage displayed cytochrome b-562, containing mutations in the solvent exposed loops that are free of non-local

interactions and would not be expected to affect folding [25]. Panning against a specific antigen yielded clones with consensus-sequence motifs within the randomized regions. As noted above, however, the breadth of novelty in binding properties is not addressed by these methods.

A different approach to improving the utility of glubody libraries is to replace random mutagenesis with more directed methods, for example using pre-made nucleotide triplets to create the randomized regions [26]. In this way, libraries can be created to explore specific structure–activity hypotheses (e.g. perhaps replacing all the residues in an LRS with hydrophobic amino acids). Such an approach allows smaller libraries to be synthesized that effectively sample the diversity of the full combinatorial library [27]. Aside from practical utility in creating new diagnostics and catalysts, the ability to create homologous proteins that differ widely in ligand recognition should be useful in uncovering the cryptic similarities among binding sites of unrelated proteins. These similarities have recently been documented by examining correlations in cross-reactivity of various proteins to drugs and other small molecules [24].

Significance

We have identified structurally permissive segments of GST and subjected them to mutagenesis, generating enzymes with altered ligand-binding properties. These results demonstrate that a non-antibody framework can be used to create large libraries of proteins with a wide range of binding specificities for small organic molecules. Modifying the deep binding cleft of GST, which has evolved to accommodate small molecules, may facilitate the development of diagnostic reagents for such molecules as well as the discovery of novel catalysts, perhaps including new enzymes capable of metabolizing environmental toxins.

The manner in which the glubodies were engineered, using intensive randomization of an LRS, opens the possibility that similar libraries can be generated from many other proteins and enzymes that contain amino-acid segments with the characteristics of an LRS. These varied panels of binding proteins should prove useful for correlating structural and functional features relevant to ligand binding in general. In contrast, most previous studies of mutant enzymes have emphasized structural features related to catalytic function [28]. Generalities such as preferential use of particular amino acids in binding sites, which have been noted in existing proteins [29], can now be tested on a broader scale, thus contributing to a deeper understanding of protein–ligand recognition processes and to more efficient drug design.

Materials and methods

Genetic engineering

Three glubody libraries (Gb/P36, Gb/P204 and Gb/206L) were prepared by similar procedures, using the primers defined in Table 4 to target PCR mutagenesis to particular regions. The GST P1 expression vector (pKXHP1) [30] has had an internal *Sfi*I site at nucleotide position 573–585 removed by overlap PCR mutagenesis, using primers L569 and R594. A G-to-A substitution at position 582 removed the *Sfi*I site while leaving the amino acid sequence unchanged.

Randomized glubody libraries were constructed essentially following the method for antibodies of Barbas *et al.* [31]. Gb/P36 contains an eight amino acid randomized region beginning at position 36 in the parental amino acid sequence. Gb/P204 has seven randomized residues beginning at position 204, very near the carboxyl terminus. Gb/206L contains an insertion of five randomized amino acids followed by a constant Pro residue between N206 and G207. All were made analogously using the appropriate primers. For the carboxy-terminal mutants, only one PCR reaction was used; for the internal fragment, two sequential reactions were used to simplify primer synthesis.

For randomization of amino acids corresponding to position 36–43 (ETWQEGSL), two primary amplifications were performed. Reaction 1 contained 10 pmol each of primers Lext and P36Rint, Perkin-Elmer Taq polymerase buffer with $MgCl_2$ added to 2 mM, 10 ng of template pKXHP1, all four dNTP's at 250 μM each and 2.5 units of Taq polymerase in a final volume of 50 μl. Reaction 2 was identical to reaction 1 except that it contained the primers P36LintR and R210. Using an Omnigene thermal cycler, the reaction mixes were put

Table 4

PCR primers for glubody construction.

Primer	Sequence (5' → 3')
L569	TCAAGGCCTTCCTAGCCTCCCCTGA
R594	TCAGGGGAGGCTAGGAGGCCTTGA
Lext	CATGCCATGACTCGCGGCCCAGCCGGCCATGGCATGCCTCCATACACAGTTGTTTA
P36Rint	CACGGTCACCACCTCCTCCTTCCA
P36LintR	AAGGAGGAGGTGGTGACCGTGNNSNNSNNSNNSNNSNNSNNSNNSAAAGCCTCCTGCCTATACGGG
R210	CCAGCATTCTGCGGCCGCCTGTTTCCCGTTGCCATTGATGG
R210	CCAGCATTCTGCGGCCGCSNNSNNSNNSNNSNNSNNSNNGGGGAGGTTCACGTACTCAGG
R206Loop	CCAGCATTCTGCGGCCGCCTGTTTCCCGTTGCCCGGGSNNSNNSNNSNNSNNSNNATTGATGGGGAGGTTCAC

The number in the primer name refers to the nucleotide position of 5' nucleotide of the primer within the sequence of human GST P1-1 (where appropriate). N = A, C, G, or T (equimolar); S = G or C (equimolar).

through 25 cycles of denaturation (94 °C, 1 min), annealing (65 °C, 1 min), and extension (72 °C, 1 min) followed by a final cycle of extension (72 °C, 10 min). The reaction products were gel purified, subjected to overlap extension and assembled as follows: 100 ng of purified product from reaction 1 was combined with 100 ng of purified product from reaction 2 and added to a PCR reaction mix containing Taq polymerase buffer with $MgCl_2$ added to 2 mM, dNTP's at 250 μM and 2.5 units of Taq polymerase in a final volume of 50 μl. This assembly mix was then taken through seven rounds of denaturation (94 °C, 1 min) and annealing (65 °C, 2.5 min) after which 10 pmols each of primers Lext and R210 were added and the PCR amplification continued for 25 cycles as above. The resulting product is a GST P1 mutant with a randomized loop domain and is designated Gb/36 cDNA. This fragment was gel purified, digested with *Sfi*I and *Not*I, and gel purified once again.

For randomization of amino acids 204–210 (INGNGKQ), a primary PCR reaction was performed as described above except that 10 pmol of the primer R210R was used to replace R210 in the amplification. PCR products were purified and digested as above. The mutant cDNA generated from this reaction was designated Gb/204 cDNA. The loop insertion library was made in a similar manner using 10 pmol of the primer R206Loop and the cDNA from this library designated Gb/206L.

Digested cDNA (1 μg) from each library was ligated to 1 μg *Sfi*I/*Not*I-restricted pHEN-1 phagemid vector [21] in a standard ligation reaction, and electrotransformed into the non-suppressor *E. coli* strain HB2151 by established procedures [3]. Individual clones were induced to produce protein as described below.

Preparation of bacterial extracts

Overnight cultures from individual clones grown in LB medium with 100 μg ml^{-1} of ampicillin and 1 % glucose were diluted into LB supplemented with antibiotic and 0.1 % glucose and incubated at 37 °C with shaking until the culture reached an OD_{600} of 0.8–0.9. IPTG was added to a final concentration of 1 mM and the culture grown at 30 °C overnight with shaking. Bacterial cultures were harvested and centrifuged at 7000 x g in a Sorvall SS-34 rotor for 5 min at 4 °C. Cell pellets were frozen and thawed twice using dry ice/ethanol and a water bath at 37 °C and resuspended in one tenth the original culture volume of lysis buffer (10 mM Tris-HCl, pH 7.8, 50 mM EDTA, 15 % glucose and lysozyme (Sigma) at 1 mg ml^{-1}). PMSF was added to a final concentration of 250 μM and the solution allowed to sit on ice for 1 h. The suspension was sonicated on ice with a Branson Sonifier 450 at 50 % duty cycle and 6 output setting for a total of 60 pulses. Samples were centrifuged for 30 min. at 25 000 x g in a Sorvall SS-34 rotor at 4 °C. The supernatant was collected and stored at 4 °C.

Enzymatic activity determination

GST activity was measured by the ability to conjugate CDNB to GSH (reduced glutathione). IC_{50} values were measured in a standard conjugation assay which contains 1 mM GSH and 1 mM CDNB in 200 mM sodium phosphate, pH 6.8 [15]. Compounds were assayed for their inhibitory activity at 250, 50, 10, 2 and 0.4 μM. Conjugation of GSH and CDNB was followed for 5 min at 30 °C by measuring the absorbance increase at 340 nm in a thermostated microplate reader (Molecular Devices, Menlo Park, CA). For IC_{50} determinations, individual glubody extracts were normalized for constant activity i.e. to produce absorbance readings at 340 nm of 40–50 mOD by addition of *E. coli* extract from a pHEN-1 transformant to equalize the amount of bacterial protein present in each assay.

Specific activity determination

Quantitation of glubody protein was accomplished by dot-blotting and subsequent scanning using a UMAX scanner (Fremont, CA) of serial dilutions of glubody extracts onto nitrocellulose strips where detection was either with polyclonal anti-GST P1-1 antisera or anti-c-*myc* monoclonal antibody. Specific activities were calculated based on amount of enzyme activity per mg of protein using the purified native GST P1-1 enzyme as the standard.

Reagents

Compounds 11–14 were synthesized at Terrapin Technologies, Incorporated (South San Francisco, CA). All other chemicals were reagent grade and purchased from either Aldrich Chemical Company (Milwaukee, WI) or Sigma Chemical Company (St. Louis, MO). Taq polymerase was from Perkin-Elmer Cetus (Norwalk, CT). Restriction enzymes and T4 DNA Ligase were from New England Biolabs (Beverly, MA). dNTP's were from Pharmacia Biotech (Alameda, CA). Anti-c-*myc* monoclonal antibody 9E10 was purchased from Santa Cruz Biotechnology (Santa Cruz, CA). The PCR primers were prepared on a Biosearch 8750 DNA synthesizer using reagents from Prime Synthesis (Astin, PA). pHEN-1 phagemid vector was generously provided by Dr Jorg Berg of University of California at San Francisco. The GST P1-1 vector, pKXHP1, was the generous gift of Dr Bengt Mannervik, University of Uppsala, Sweden.

References

1. Tainer, J.A., Deal, C.D., Geysen, H.M., Roberts, V.A. & Getzoff, E.D. (1991). Defining antibody–antigen recognition: towards engineered antibodies and epitopes. *Int. Rev. Immunol.* **7**, 165–188.
2. Edmundson, A.B., Guggat, L.W., Shan, L., Fan, Z.C. & Hanson, B.L. (1994). Structural aspects of conformational changes in ligand binding by antibody fragments. *Res. Immunol.* **145**, 56–61.
3. Marks, J.D., Hoogenboom, H.R., Bonnert, T.P., McCafferty, J., Griffiths, A.D. & Winter, G. (1991). By-passing immunization: human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* **222**, 581–597.
4. Vanderlaan, M., Stanker, L.H., Watkins, B.E. & Roberts, D.W., Eds. (1991). *Immunoanalysis for Trace Chemical Analysis.* Washington, D.C., ACS Symposium Series #451.
5. Lerner, R.A., Benkovic, S.J. & Schultz, P.G. (1991). At the crossroads of chemistry and immunology: catalytic antibodies. *Science* **252**, 659–667.
6. Jakoby, W.B. & Ziegler, D.M. (1990). The enzymes of detoxification. *J. Biol. Chem.* **265**, 20175–20178.
7. Mannervik, B. & Danielson, U.H. (1988). Glutathione transferases: structure and catalytic activity. *CRC Crit. Rev. Biochem.* **23**, 283–337.
8. Ricci, G., *et al.*, & Federici, G. (1995). Site-directed mutagenesis of human glutathione transferase P1-1. *J. Biol. Chem.* **270**, 1243–1248.
9. Palzkill, T. & Botstein, D. (1992). Probing β-lactamase structure and function using random replacement mutagenesis. *Proteins* **14**, 29–44.
10. Munir, K.M., French, D.C., Dube, D.K. & Loeb, L.A. (1994). Herpes thymidine kinase mutants with altered catalytic efficiencies obtained by random sequence selection. *Protein Eng.* **7**, 83–89.
11. Black, M.E. & Loeb, L.A. (1993) Identification of important residues within the putative nucleoside binding site of HSV-1 thymidine kinase by random sequence selection: analysis of selected mutants *in vitro*. *Biochemistry* **32**, 11618–11626.
12. Gulick. A.M. & Fahl, W.E. (1995). Forced evolution of glutathione S-transferase to create a more efficient drug detoxification enzyme. *Proc Natl Acad Sci USA* **92**, 8140–8144.
13. Widersten, M. & Mannervik, B. (1995). Glutathione transferases with novel active sites isolated by phage display from a library of random mutants. *J. Mol. Biol.* **250**, 115–122.
14. Mannervik, B., Guthenberg, C. Jakobsen, I. & Warholm, M. (1978). Glutathione conjugation: reaction mechanism of glutathione S-transferase. In *Conjugation Reactions in Drug Biotransformation.* (Aito, A., ed.), pp. 101–122, Elsevier/North Holland Biomedical Press, Amsterdam.
15. Kano, T., Sakai, N. & Muramatsu, M. (1987). Structure and expression of a human class Pi glutathione S-transferase messenger RNA. *Cancer Res.* **47**, 5626–5630.
16. Reinemer, P., *et. al.*, & Parker, M.W. (1992). Three dimensional structure of class Pi glutathione-S-transferase from human placenta in complex with S-hexylglutathione at 2.8 Å resolution. *J. Mol. Biol.* **227**, 214–226.
17. Chothia, C. & Lesk, A.M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196**, 901–917.
18. 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.
19. Cachau, R.E., Erickson, J.W. & Villar, H.O. (1994). Novel procedure for structure refinement in homology modeling and its application to the human class Mu glutathione S-transferases. *Protein Eng.* **7**, 831–839.
20. Kolm, R.H., Sroga, G.E. & Mannervik, B. (1992). Participation of the phenolic hydroxyl group of Try-8 in the catalytic mechanism of human glutathione transferase P1-1. *Biochem. J.* **285**, 537–540.
21. Hoogenboom, H.R., Griffiths, A.D., Johnson, K.S., Chiswell, D.J., Hudson, P. & Winter, G. (1991). Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res.* **19**, 4133–4137.
22. Wheatley, J.B., Hughes, B., Bauer, K. & Schmidt, D.E. (1994). Study of chromatographic parameters for glutathione S-transferases on high-performance liquid chromatography affinity stationary phase. *J. Chromatography* **676**, 81–90.
23. Flatgaard, J.E., Bauer, K.E. & Kauvar, L.M. (1993). Isozyme specificity of novel glutathione S-transferase inhibitors. *Cancer Chemother. Pharmacol.* **33**, 63–70.
24. Kauvar, L.M., *et. al.*, & Rocke, D.M. (1995). Predicting ligand binding to proteins by affinity fingerprinting. *Chemistry & Biology* **2**, 107–118.
25. Ku, J. & Schultz, P.G. (1995). Alternate protein frameworks for molecular recognition. *Proc. Natl. Acad. Sci. USA* **92**, 6552–6556.
26. Lyttle, M.H., Napolitano, E.W., Calio, B.L. & Kauvar, L.M. (1995). Mutagenensis using trinucleotide β-cyanoethyl phosphoramidites. *Biotechniques* **19**, 274–281.
27. Kauvar, L.M. (1994). Panels of analyte binding ligands. U.S. Patent #5,340,474.
28. Roberts, V.A., *et. al.*, & Tainer, J.A. (1990). Antibody remodeling: a general solution to the design of a metal-coordination site in an antibody binding pocket. *Proc Natl Acad Sci USA* **87**, 6654–6658.
29. Kolm, R.H., Stenberg, G., Widersten, M. & Mannervik, B. (1995). High-level bacterial expression of human glutathione transferase P1-1 encoded by semisynthetic DNA. *Protein Expr. Purif.* **6**, 265–271.
30. Stenberg, G. (1991). Glutathione transferases: molecular cloning, site-directed mutagenesis and structure–function studies. Ph.D. thesis, University of Stockholm.
31. Barbas, C.F., Bain, J.D., Hoekstra, D.M. & Lerner, R.A. (1992). Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. *Proc. Natl. Acad. Sci. USA* **89**, 4457–4461.