Conformational Heterogeneity and Catalytic Promiscuity in Glutathione Transferases

Matthew T. Honaker

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

William M. Atkins, Chair

Carlos Catalano

Kelly Lee

Program Authorized to Offer Degree:

Medicinal Chemistry

University of Washington

Abstract

Conformational Heterogeneity and Catalytic Promiscuity in Glutathione Transferases

Matthew T. Honaker

Chair of the Supervisory Committee:
Professor William M. Atkins
Department of Medicinal Chemistry

Enzymological paradigms have shifted recently to acknowledge the biological importance of catalytic promiscuity. Detoxification enzymes, such as glutathione transferases, are known to be highly promiscuous. One common suggestion is that promiscuous enzymes are more conformationally heterogeneous than their substrate-specific counterparts. Here, a series of structurally similar glutathione transferase (GST) variants ranging from high specificity to high promiscuity are used as a platform for examining conformational heterogeneity as a putative mechanism for promiscuity. Quantitative promiscuity indices range from 0.72 for promiscuous GSTA1-1 to 0.05 for specific GSTA4-4, with intermediate values for each of the GST mutants. Fluorescence lifetime distribution analysis of Trp21 at the domain interface indicates a modest correlation of core protein dynamics with promiscuity. The mobility during the excited-state lifetime of Trp21 increases sharply for GSTA1-1. Differential scanning calorimetry (DSC) indicates a reversible low temperature transition for the promiscuous GSTA1-1 that is not observed with substrate specific GSTA4-4. This transition is assigned to rearrangement of the C-terminus at the active site of GSTA1-1 based on the effects of ligands and mutations. Near-UV and far-UV circular dichroism indicate that this transition is due to repacking of tertiary contacts with the remainder of the subunit, rather than 'unfolding' of the C-terminus *per se*. Analysis of the DSC data using a modified Landau Theory indicates that the local conformational landscape

of the active site of GSTA1-1 is smooth, with barrierless transitions between states. The partition function of the C-terminal states is a broad unimodal distribution at all temperatures within this DSC transition. Extraction and analysis of low temperature transitions for GSTA4-4 and the GST mutants with intermediate promiscuity using the same method indicates a correlation between the natural logarithm of partition function distribution widths with promiscuity, suggestive of an entropic basis for promiscuity. A similar trend in barrier heights differentiated barrierless GSTA1-1 from the less promiscuous GSTs, with free energy barriers increasing as promiscuity decreased. The correlation of conformational heterogeneity and level of promiscuity across the GST variants examined suggest a conformational selection mechanism for catalytic promiscuity.

Table of Contents

List of Figures

List of Tables

Page

Acknowledgements

For truly outstanding leadership and insight, I would like to acknowledge and thank my mentor, Dr. William Atkins, without whom none of this would have been possible. It has been a pleasure and an amazing opportunity to work with Dr. Atkins. I would like to thank the rest of my committee, Drs. Carlos Catalano, Kelly Lee, and Yvonne Lin for their guidance and hard work. My colleagues in the Atkins Lab, especially Laura Shireman, Kip Conner, Mauro Acchione, Caleb Woods, Larissa Balogh, Jed Lampe, Sumit Mahajan, Abinath Nath, Mike Dabrowski, Michele Scian, and Liming Hou also deserve my gratitude and thanks for their knowledge, support, and friendship. Thanks also to Marcus Collis and John Sumida for helpful discussions. In addition, I would like to thank the faculty and staff of the Department of Medicinal Chemistry.

Dedication

For their support, encouragement, and belief in me,

this work is dedicated to my parents,

Anne and Michael Honaker,

and to my wife,

Lindy Oliver Honaker.

Chapter 1: Introduction and Background

A. Introduction

Historically, enzymes are highly efficient and specific catalysts, with catalytic efficiencies typically ranging from 10^5 to 10^9 M^{-1} s^{-1} (*1*). The basis for this remarkable specificity was first described using a model where the enzyme active site was evolved to fit precisely to the appropriate cognate substrate, the 'lock and key' model (*2*). This model, first proposed by Emil Fischer, provides ideal specificity between enzyme and substrate, but does little to allow for any sort of catalytic versatility or protein dynamics as the lock and key are both fixed in structure and geometry prior to catalysis. Similarly, the induced fit model (*3-5*), where the enzyme provides transition state complementarity after substrate binding, rather than strict substrate precision, permits negligible availability for off target catalysis. An alternative perspective for single substrate enzymes in the conformational selection model, wherein there are multiple equilibrating enzyme conformations in the substrate-free form and the substrate 'selects' a conformation for which it has the highest affinity (*6*).

While the paradigm of specific, single substrate - single reaction enzyme generally dominates in most examinations of enzyme function, the catalytic versatility of some enzymes was demonstrated as early as 1965, when it was shown that carbonic anhydrase catalyzes not only the reversible hydration of carbon dioxide, but also acts as an esterase to catalyze the hydrolysis of *p*-nitrophenyl acetate (*7*). Observation of catalytic versatility, or promiscuity, implies a dynamic variability between enzyme structure and function. In addition, the well described behavior of allosterism (*8, 9*), implies some degree of plasticity and conformational heterogeneity in an enzyme's specificity and structure. Rationalization of high degrees of

transition state complementarity with clear evidence of conformational fluctuations in enzymes at multiple timescales (*10-13*), substrate promiscuity (*14-16*), and protein regulation on several levels (*8, 17-21*) demands a necessary coupling between binding and conformational change and invites comparisons between conformational heterogeneity and enzyme activity with application to enzyme promiscuity.

B. Ensemble Perspective

Conformational fluctuations of enzymes can lead to ensembles of free enzyme, E, and both the ES and EP complexes (*22-24*). Most enzyme mechanisms appear to include multiple conformational changes as a general feature (*11*). Assuming binding generally occurs within the same active site for all substrates, two simplified scenarios demonstrate limiting models of catalytic ensembles (*22, 23*). In the classic induced fit model, conformational changes occur only after binding of substrate, where substrate binding leads to an ES complex with a different conformation from enzyme E (Figure 1.1). In contrast, for the conformational selection model, an ensemble of conformers pre-exists to allow binding of substrate to a pre-organized enzyme structure suitable for that substrate; a population shift, or conformational selection mechanism (Figure 1.2). Hypothetically, the conformational selection model is exploited for catalytically promiscuous enzymes or multifunctional enzymes that metabolize multiple substrates; different conformations in the ensemble would likely be selected by different substrates and the ensemble could provide a mechanism to achieve high promiscuity.

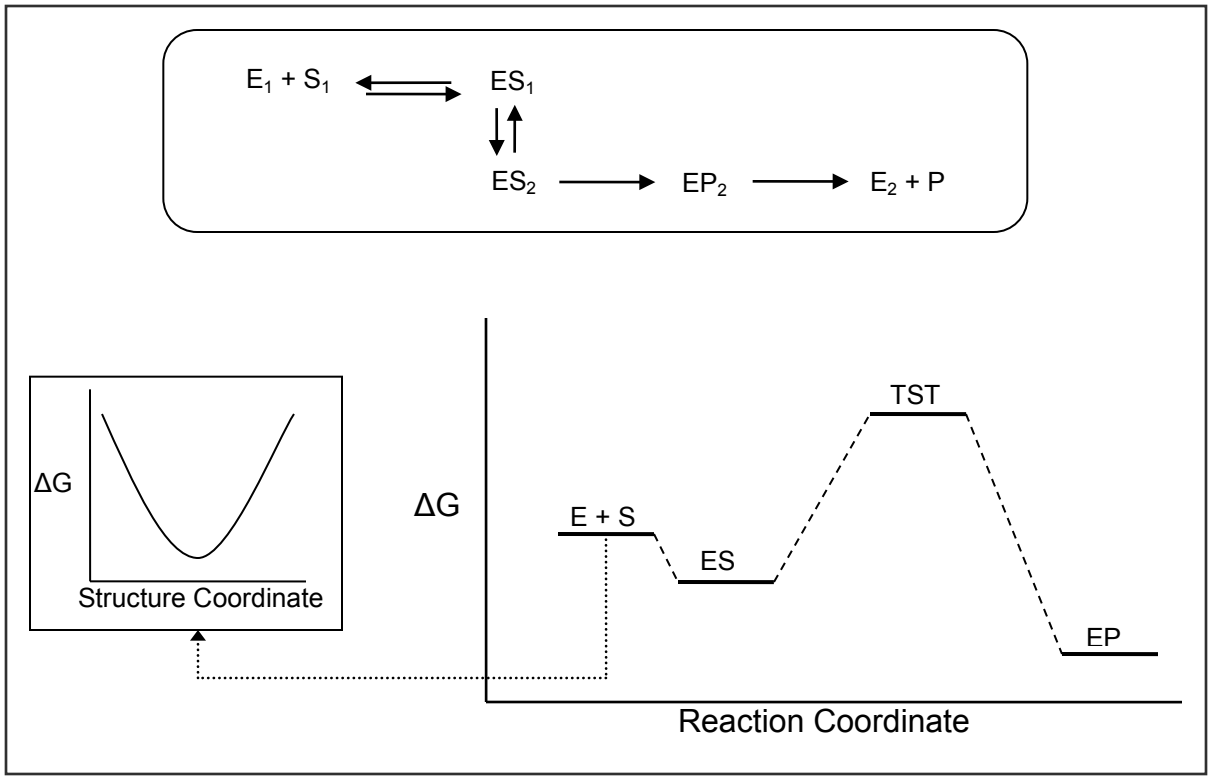


Figure 1.1 Induced fit schematic. An induced fit mechanism where substrate binding leads to an ES complex and transition state (TST) with a different conformation than the unbound enzyme. The free enzyme exists at the bottom of a smooth energy well.

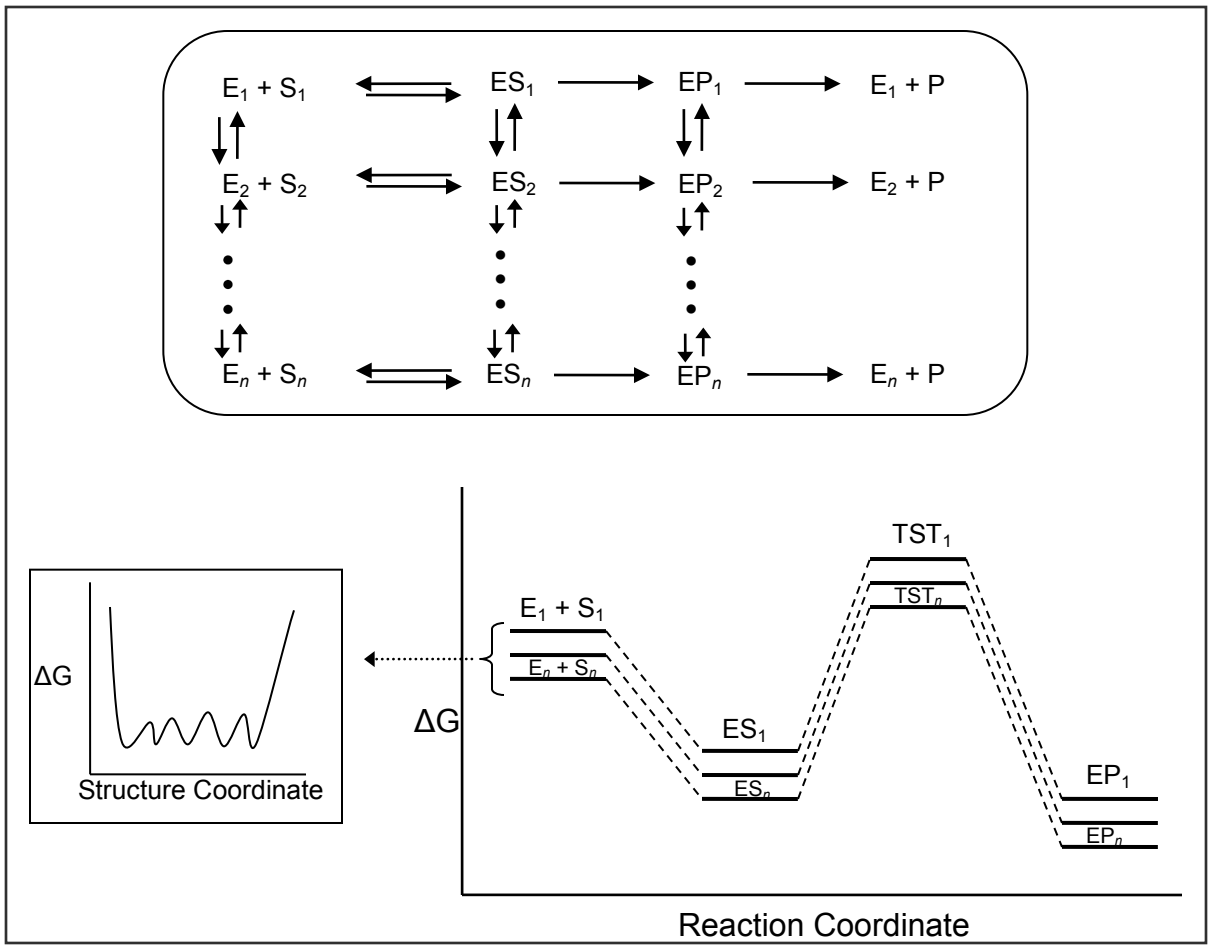


Figure 1.2 General conformational selection scheme. An ensemble of free enzyme conformations allows multiple reaction pathways with different substrates. The conformations of free enzyme exist at the bottom of a rugged energy well.

The induced fit model of enzyme activity includes the premise that binding of substrate can induce conformational changes in the enzyme for proper orientation of catalytic groups, and thus transition state complementarity, but only after binding (*25*). Implicit in this model is a singular protein conformation in the absence of effectors or substrate, with the conformation determined by the amino acid sequence and conditions of folding. This conformation can shift structure in response to binding of substrate or an allosteric ligand, as in the Koshland-Nemethy-Filmer (KNF) model of allostery (*25, 26*), and binding of a ligand can induce conformational changes similar to folding (*27*). As originally developed, the induced fit model accentuates enzyme specificity by allowing the conformational change to occur only if the correct substrate

binds, limiting catalysis with competitive species, and the active site must generally surround the transition state excluding access to other molecules (*5, 25, 26, 28*). In this view, the enzyme lies at the bottom of a smooth energy landscape with a structure evolved to complement the substrate, adjusting conformation after binding to fit the transition state (*29*). The structure of an enzyme displays enough flexibility to allow substrate dependent configurational adjustment. An induced fit mechanism is often attributed to enzymes exhibiting transitions between 'open' and 'closed' conformations in response to ligand binding, where the 'closed' conformation is not observed if ligand is not present. This has been argued in the example of phosphoenolpyruvate carboxykinase, where crystal structures have revealed a lid-gated active site, and the 'closed' conformation is only found in ligand bound states. (*30*).

The Monod-Wyman-Changeux model of allostery (*8*) allows for the existence of two (or more) pre-existing conformers of a given enzyme in the absence of ligand. Similar to this model of allostery is the notion that an enzyme can exist in multiple conformation simultaneously, even in the absence of substrate. In this case the energy landscape is rugged, with multiple wells of approximately equal energy for various conformers, one of which is active, and maximizes the catalytic efficiency for transformation of the substrate (*29*). There is no substantial difference between folding funnels of ensembles of conformers differing by loop or domain motions and quaternary allosteric shifts of subunits (*31*). Both exhibit a range of pre-existing structures, differing only by degree and type of structural change. Conformational selection through substrate matching provides a mechanism to allow access of multiple substrates to different conformers of free enzyme of the appropriate structure for each substrate. Conformational selection pathways have been indicated in studies of bacterial phosphotriesterase, as both 'open'

and ‘closed’ enzyme conformations have been observed in ligand free crystal structures and NMR experiments (*32*).

In either limiting model, the reaction coordinate is orthogonal to the conformation coordinate. Although there has been recent discussion aimed at the possibility of conformational dynamics that facilitate the transition state crossing, the models considered here do not include transition state conformational heterogeneity for a single substrate. Presumably there is no interconversion of substates at the transition state level, nor does dynamic motion somehow impart kinetic energy to speed the reaction. The relative contribution to the overall catalytic mechanism from induced fit or conformational selection can vary for each enzyme, but both are a function of enzyme dynamics (*22*).

An excellent example demonstrating elements of both induced fit and conformational selection is an engineered molten globule form of chorismate mutase. A combination of genetic selection and structure-based design was used to convert the native dimeric chorismate mutase from *Methanococcus jannaschii* into a structurally disordered monomer (mMjCM) with similar catalytic efficiency as the native enzyme toward rearrangement of chorismate into prephenate (*33*). Comprehensive examination of mMjCM demonstrated a resting state exhibiting the properties of a molten globule. This includes poor NMR signal dispersion, rapid hydrogen-deuterium exchange, and non-cooperative thermal denaturation normally associated with folding intermediates of defined secondary structure, but a significant lack of pre-organized tertiary structure (*33, 34*). Global conformational ordering was observed upon binding of a transition state analog or ligand (*33*) as in a predominantly induced fit mechanism, albeit one in which the ‘native’ macrostate is described by fluxional ensemble of isoenergetic conformational microstates. Elegant computational studies involving mMjCM demonstrated that the lowest

catalytic activation barriers are seen in the region of the folding landscape near the energy minimized native structure, and catalytic efficiencies decreased as the reorganization energy and structural distance from the native conformer increased (*35*). The area of the folding surface near the native structure, where catalysis remained efficient, was flat and without significant energy barriers between conformers. Parallel studies of the wild-type chorismate mutase dimer demonstrated well folded native structure in the presence and absence of ligand, and a much smaller region of structural divergence as a function of catalytic efficiency (*33-35*). This raises the intriguing possibility that the complexities of the folding and catalytic energy landscapes are related, and that enzymes can be considered as existing in ensembles of coupled conformational states, which can be altered as ligand binding and catalysis occurs (*36*). Dihydrofolate redutase (DHFR) is also known to go through multiple conformational changes as it catalyses the reduction of dihydrofolate by reduced nicotinamide adenine dinucleotide phosphate (NADPH) to give tetrahydrofolate and $NADP^+$ (*37, 38*). DHFR cycles through a minimum of five major intermediate states during the catalytic cycle, differing in conformation, relative energy level, and interconversion rates (*37*). Conformational changes of this enzyme are related to substrate binding in a ligand dependent manner, and often occur well removed from the active site (*36*), suggesting that even for a single substrate-single product reaction, multiple conformations and parallel reaction paths can coexist. In addition, at very low NADPH concentrations, the relative 'flux' through a conformational selection pathway predominates in NADPH binding, rapidly shifting to induced fit as the NADPH concentration increases (*22, 39*).

The observation that substrate specific enzymes such as DHFR and chorismate mutase are best described by ensemble behavior, possibly throughout their entire reaction coordinate including catalytic transition states, emphasizes that catalysis does not require conformational

homogeneity, and ensembles are likely a common feature of multiple species along the reaction coordinate. Hypothetically, this behavior could 'exploited' or optimized to afford efficient catalysis by enzymes that are required to metabolize structurally unrelated substrates. An ensemble perspective based on conformational heterogeneity produced by enzyme dynamics is perhaps an even more apt description of the catalytic landscape for a promiscuous enzyme that exhibits catalytic power towards multiple substrates. When enzymes are considered as dynamic distributions, the ensemble can present incoming ligands with several binding conformations of electrostatic surfaces of active site regions. This allows for a plastic energy surface across which catalysis can occur in a variety of modes, enabling enzyme promiscuity.

C. Promiscuity

Promiscuity is broadly used in protein science to denote many categories of multifunctional behavior (*29, 40, 41*), including recognition of multiple protein partners in regulatory networks (*42*), transport of dissimilar substrates (*40*), use of structural regions other than the enzymatic active site for some function (*29, 43*), and both metabolism and inhibition of an enzyme with diverse compounds (*14, 15, 44*). However the focus of this work is catalytic promiscuity of enzymes, or the ability of some enzymes to catalytically metabolize substrates of widely varying structure. Many, if not most, enzymes can accept a small range of structurally similar substrates (*45*), and thus exhibit some degree of promiscuity. This is an obvious consequence of using flexible macromolecules as catalysts, and provides both an efficient route for the evolution of new function, and a mechanism for advantageous plasticity inherent in the physiological role of certain enzyme classes. Several reviews (*29, 41, 43, 46-48*) illustrate the overwhelming diversity of promiscuous function of this type present in an array of proteins and enzymes; it is clear that many enzymes are promiscuous in some manner. Several enzymes

demonstrate physiological roles removed from their catalytic function, also called moonlighting, which is often regulatory in nature. For example, cellular location can determine the role of an enzyme, as is the case with the *E. coli* protein PutA, which exhibits dehydrogenase activity toward proline and pyrroline-5-carboxylate when associated with the plasma membrane in conditions of high substrate concentration, but acts as a transcriptional repressor in the cytoplasm, binding to DNA at low substrate concentrations (*48-50*). Similarly, thymidine phosphorylase displays location dependent differential functionality (*48, 51*). This enzyme catalyzes dephosphorylation of thymidine, deoxyuridine, and analogs when in cellular cytoplasm; in contrast it stimulates chemotaxis and endothelial cell growth as an extracellular protein.

Oligomeric state can also control the catalytic function of an enzyme. As a monomer, human glyceraldehyde-3-phosphate dehydrogenase is a nuclear uracil-DNA glycolase; as a tetramer it converts glyceraldehyde-3-phosphate to 1,3-diphosphatoglycerate (*52*). Enzymes can also be catalytically promiscuous; demonstrating catalytic activity with more than one substrate and/or production of more than one product.

Catalytically promiscuous enzymes are frequently separated into two broad categories, promiscuous enzymes demonstrating catalysis of reaction(s) other than those selected for by evolution, and enzymes exhibiting strong substrate or product infidelity, also termed multispecific (*43*). This distinction is somewhat artificial and depends on recognition of a 'natural' physiological reaction for any given enzyme before degree of specificity breach, or promiscuity, can be considered. One important contribution of these categories of promiscuity is in the context of evolution; catalytically promiscuous enzymes provide an evolutionary mechanism to obtain new enzyme through natural selection. Enzymes which display a

coincidental, ‘non-natural’ reaction may be intermediates on the course from an original specific catalytic function to a new specific catalytic function. As evolutionary intermediates, catalytically promiscuous enzymes exhibit some catalytic activity(s) divergent from the prime physiological role for which that enzyme was evolved (*43*). When a reaction is advantageous, it can be retained by natural selection and allow efficient emergence of new function from existing protein scaffolds through point mutation and gene duplication (*41, 47, 53*), often with little loss of original function. Enzymes with broad substrate or product specificity, multispecific enzymes, may be endpoints as the result of selection for inherently promiscuous catalysis.

Regardless of classification, catalytically promiscuous enzymes share ambiguity in form and function and are considered to be promiscuous in that they do not subscribe to the single substrate-single reaction-single enzyme paradigm. Enzyme active sites contain a milieu of reactive environments, nucleophilic and electrophilic side chains, acidic and basic groups, and reactive cofactors. Both hydrophobic and hydrophilic regions are common, as are structurally rigid portions in close proximity to more flexible and dynamic areas. Conformational adjustments of varying magnitude can mediate interactions and contacts between substrate and enzyme, alter pK_a values, create oxidizing species, and exclude water molecules. This conformational diversity can be subtle. For example, the isomerization of a tyrosine side chain allows both glycosyl hydrolase and transferase activity by trypanosomal *trans*-sialidase (*54, 55*). Alternatively, the conformational change may large as observed in the large flexible loop fluctuations that mediate detoxification of at least 10 separate aminoglycosides by an animoglycoside kinase found in Gram positive infectious bacteria (*56*).

Examples of modest catalytic promiscuity are becoming more evident in the literature, and many enzymes demonstrate a small range of catalytic promiscuity, catalyzing reactions with

a range of structurally similar, but distinct substrates (*43*). This may provide an evolutionary advantage for induction of metabolic pathways when the 'natural' substrate is in low availability, or chance formation of a new chemical moiety from preferred substrate. Alternate catalytic activities may also only be apparent *in vitro*, and may not serve any useful function. Subtle shifts in substrate positioning can result in catalytic promiscuity due to imperfect control of substrate within the active site, as seen in cases promiscuous regioselectivity. Small adjustments in the active site can also lead to divergent formation of the hydrogen bond network between D-2-keto-3-deoxy-gluconate aldolase (KDGA) and two substrates that differ only in stereochemistry (*57*). The active site of this enzyme from *Sulfolobus solfataricus* allows sufficient flexibility for both D-2-keto-3-deoxy-gluconate and its stereoisomer D-2-keto-3-deoxy-galactonate to each bind (separately), and present the 2 position carbonyl to Lys155 for Schiff base formation subsequent hydration and cleavage at similar rates for both substrates. Another example of catalytic promiscuity requiring only limited structural rearrangements and illustrating use of promiscuous enzymes in biocatalysis is the reduction of 13 benzaldehyde derivatives by 7α-hydroxysteroid dehydrogenase (*58*). A small range of benzaldehyde analogs were reduced at the aldehyde carbonyl group to the corresponding alcohol with electron-withdrawing groups demonstrating greater conversion efficiency. Alternate catalytic residues can be used to promote promiscuous activities, as noted in the case of serum paroxonase which hydrolyses lactones in a reaction mediated by a histidine dyad, but also exhibits phosphotriesterase activity induces by other nucleophilic residues (*59, 60*).

As with natural evolution, promiscuous enzymes also provide an excellent platform for *in vitro* evolution of new protein function (*61, 62*). Promiscuous activity can also be induced or modified through mutation of several residues, or even a single residue. Mutation of Met192 to

Ala increased activity toward non-native substrates by 10^5 fold, with very little reduction in the native activity (*63*). Catalytic activity toward a variety of substrates can be tuned through mutations in the hydrophobic groove of cyclooxygenase COX-2 (*64*). Location of the mutations could reposition the bis-allylic carbon in fatty acid substrates for catalysis at various points along the chain. Forced use of non-physiological substrates can also take advantage of promiscuous enzymes. A polyketide synthase from *Streptomyces griseus*, RppA, catalyses the formation of 1,3,6,8-tetrahydroxynaphthalene from five molecules of malonyl-CoA (*65*). When placed in an environment where alternate acyl CoAs with carbon chain lengths from 4 to 8 were the only available CoA source, RppA was able produce a mixture of α-pyrone and phloroglucinol products. The size and shape of the starter substrates could tune the active site for catalysis, despite the differentiation from natural substrate.

Clearly, many, if not most, enzymes exhibit some level of promiscuous activity, *in vitro* or *in vivo*. Ability to use alternate substrates may produce useful selective advantage or be used to engineer new biocatalyst function. Adjustment of the active site, whether small local shifts of side chains, or wholesale shifts in dynamic loops provide the necessary conformational diversity for off target catalysis. Fortuitous promiscuous catalytic function of enzymes may provide efficient used of limited sequence variability for expanded functional diversity for the evolution of new function, while denying loss of original function to any great degree. However, some classes of enzyme appear to be the result of natural selection for inherent promiscuity, related to their physiological role. Detoxification enzymes, whose hallmark feature is broad specificity, provide prime examples catalytic promiscuity as evolutionary endpoints, and will be the focus of the remainder of this work.

Origination of an entirely new detoxification enzyme for every molecular compound not used in a defined metabolic pathway is inefficient, and perhaps impossible. In fact, it is unreasonable to suggest that individual enzymes could be evolved for every toxin a species has been exposed to and toxins to which the species *might* be exposed. Instead a small number of detoxification enzymes are used, which can collectively accommodate a wide range of substrate structures. The levels and chemical natures of exogenous toxins are unpredictable and vary exceedingly quickly on an evolutionary timescale, so routes of elimination must be accommodating and adaptive. It is likely that the most efficient method to achieve this plasticity is through inherent catalytic promiscuity, particularly through substrate promiscuity at the single protein level. Detoxification enzymes provide the primary and most important platform for examination of the structural and thermodynamic basis for inherent catalytic and substrate promiscuity. Native detoxification enzymes, including glutathione transferases (GSTs) (*14, 15, 66*), cytochrome P450's (CYPs) (*14, 44, 67*), and uridyl diphosphate glucuronic acid transferases (UGTs) (*68, 69*), encompass both highly specific and extraordinarily promiscuous isoforms within each class. Many examples of specific isoforms in each of these enzyme classes are known, with recognized endogenous substrates, and structurally similar congeners apparently lacking endogenous substrates presumably confer a selective advantage through the biotransformation of exogenous chemical moieties. CYP3A4, for example, metabolizes more than 50% of marketed drugs, covering an extremely diverse array of chemical structures (*70*).

Relating the concepts of protein dynamics and conformational ensembles present above to the overwhelming substrate promiscuity noted in detoxifications enzymes requires a quantitative method to evaluate substrate promiscuity. An entropy-based metric developed by Nath and Atkins has proved to be useful for comparing relative levels of substrate promiscuity

between structurally similar enzymes across a substrate basis set (*14*). This method is based primarily on a measure of uncertainty about the outcome of a process, in this case substrate metabolism. The entropy H for a set of N possible results can be defined in terms of the probability p of each result such that

$$H = -\sum_{i=1}^{N} p_i \log p_i \qquad \text{(Eq. 1.1)}$$

where p_i is the probability of the ith result being true. As previously noted, the catalytic efficiency, e, for an enzyme for given substrate is a measure of enzyme specificity, and can be defined as

$$e = \frac{k_{cat}}{K_M} \qquad \text{(Eq. 1.2)}$$

Presented with equal, low concentrations of N substrates, the probability, p_i, that the ith substrate will be metabolized first is

$$p_i = \frac{e_i}{\sum_{i=1}^{N} e_i} \qquad \text{(Eq. 1.3)}$$

In addition, a metric for the dissimilarity of substrates can be defined by applying a bitwise (0 or 1) operator based on whether (1) or not (0) a given substrate has a particular chemical moiety or not. After creation of a keyset based on these operators, and Tanimoto distance matrix can be constructed where $\langle \delta \rangle_i$ is the normalized mean distance of a given substrate structure (i) from the set taken as a whole. In the manner, a promiscuity factor for enzyme catalysis across a basis set J_{cat}, can be defined as

$$J_{cat} = -\frac{N}{\left(\sum_{i=1}^{N} \langle \delta \rangle_i \log N\right)} \sum_{i=1}^{N} \langle \delta_i \rangle \frac{e_i}{\sum_{j=1}^{N} e_j} \log \frac{e_i}{\sum_{j=1}^{N} e_j} \qquad \text{(Eq. 1.4)}$$

where J_{cat} can range from 0, completely specific, to 1, completely promiscuous. This method considers only degree of difference across a substrate basis set and catalytic efficiency irrespective of the type of chemical transformation or labiality of bonds broken or formed. As such, it is suitable primarily for comparisons across ranges of structurally similar but promiscuously diverse enzymes. Human A-class glutathione transferases (GSTs), in particular GSTA1-1 as an example of a catalytically promiscuous enzyme, compared to structurally similar but specific GSTA4-4, are admirably well suited for this type of analysis.

D. Glutathione Transferases

Glutathione transferases (EC 2.5.1.18), historically glutathione *S*-transferases (GSTs), are a large family of isozymes discovered in 1961 through identification of catalytic activity for the addition of tripeptide glutathione (γ-Glu-Cys-Gly, GSH, Figure 1.3) to 1,2-dichloro-4-nitrobenzene in cytosolic rat liver extracts (*71, 72*).

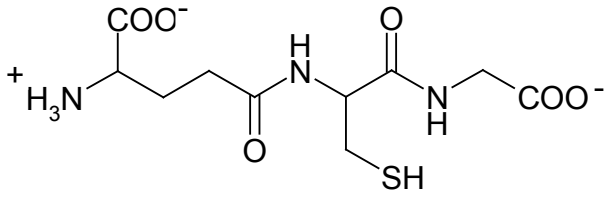


Figure 1.3 Molecular structure of glutathione.

GSTs are grouped into three main families that demonstrate glutathione transferase activity. Mitochondrial and cytosolic GSTs are soluble and the third family is microsomal. These GSTs are membrane-associated proteins in eicosanoid and glutathione (MAPEG) metabolism (*73-76*). The largest family, the cytosolic GSTs, is typically divided into eight gene classes based on substrate specificity and/or structural similarity: alpha (A, α), kappa (K, κ), mu (M, μ), omega (O, ω), pi (P, π), sigma (S, σ), theta (T, θ), and zeta (Z, ζ) (*72, 77-82*). There are one or more

members in each class, existing as homodimers or intraclass heterodimers, and each dimer is identified via class and isoform with the usual species prefix, for example hGSTA1-1 is a human alpha class homodimer of isoform 1 (*83*).

These enzymes are predominantly involved in the detoxification of a wide array of endogenous and exogenous compounds (*72, 75, 77, 84, 85*), and catalyze nucleophilic attack by reduced GSH on nonpolar substrates containing electrophilic carbon, nitrogen, or sulfur atoms, including halonitrobenzenes, quinones, arene oxides, and α,β-unsaturated carbonyls (*72, 75, 85, 86*), of the general form:

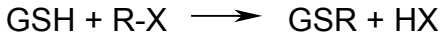


Figure 1.4 General glutathione transferase catalyzed reaction.

Glutathione transferases also efficiently catalyze essential Δ^5-Δ^4 isomerizations of steroid hormone intermediates (*87*), and are believed to influence signal transduction pathways through metabolism of cycloprostaglandin 15d-PGJ_2 (*88*) and lipid peroxidation product 4-hydroxynonenal (*89, 90*). Overexpression of GSTs in mammalian tumor cells has been implicated in resistance to chemotherapeutics (*91, 92*), and elevated levels of GSTs are associated with increased tolerance of certain insecticides (*93*) and herbicides (*94*). In addition, GSTs are also involved in the biosynthesis of useful GSH conjugates such as leukotriene A (*72*), and demonstrate ligandin and sequestration activity toward a number of hydrophobic molecules (*95*). Commensurate with their physiological role as detoxification enzymes, certain single isoforms exhibit a high degree of substrate promiscuity; coupled with a large number of isozymes, GSTs metabolize an impressive range of potential substrates.

As Phase II xenobiotic detoxification enzymes, cytosolic GSTs catalyze the first step in the transformation of toxic electrophiles to excretable mercapturic acids through conjugation to glutathione produced by glutamate cysteine ligase and glutathione synthase (*96*). Generally,

exogenous toxins are cleared through a tightly integrated three-phase pathway. Phases I and II convert lipophilic xenobiotics into more water soluble compounds, which can then be eliminated through Phase III transporters. Phase I consists primarily of the oxidative cytochrome P450 system, while GSTs comprise a major part of the Phase II pathway. GSTs catalyze Michael additions to α,β-unsaturated carbonyls, nucleophilic aromatic substitutions, and epoxide ring openings to result in GSH conjugates, as well as reducing hydroperoxides (*85*). Efflux from the cell of GSH conjugates is accomplished through several transmembrane transport proteins, including the ATP-dependent glutathione S-conjugate pump (*97*), and multidrug resistance P-glycoprotein pumps PGP, MDR1 and MDR2 (*98, 99*), a multispecific organic anion transporter (MOAT) (*100*). In addition, efflux is also promoted by RLIP76, a dinitrophenol-glutathione ATPase (*101*). The conjugates are eliminated as mercapturic acids after sequential removal of the γ-glutamyl and glycine moieties, followed by N-acetylation. Exogenous substrates removed through this pathway are numerous and structurally diverse, consisting of cancer drugs such as busulfan, chlorambusil, cyclophsphamide, and ethacrynic acid; environmental toxins including acrolein, DDT, and malathion; and carcinogens, particularly ubiquitous polycyclic aromatic hydrocarbons (PAHs) (*75*). Cancer prodrugs are also being developed to exploit the over expression of certain GSTs in tumor cells (*102-107*).

GSTs also play a large role in protection against endogenous products of oxidative stress. Reactive oxygen species, such as peroxides, hydroxyl radicals, and superoxide anions, damage membrane lipids, DNA, proteins, and carbohydrates through direct and indirect means (*108*). GSTs catalyze GSH addition to quinone species produced from oxidation of catecholamines which prevents superoxide generation via redox cycling (*75*). Furthermore, glutathione transferases conjugate GSH with 2-hydroxy-2-alkenals formed through lipid peroxidation, as

well as crotonaldehyde, cholesterol-5,6-oxide, and 9,10-epoxystearic acid (*75*). This helps prevent lipid damage caused by free radical initiated peroxidation of polyunsaturated fatty acids. Metabolism of lipid mediators by GSTs also influences several signaling pathways. Addition of GSH to 15-deoxy-$\Delta^{12,14}$-prostaglandin J_2 (15d-PGJ_2) prevents it from modifying Kelch-like ECH-associated protein 1 (Keap1), which in turn prevents Keap1 from targeting nuclear factor-erythroid 2 p45-related factor 2 (Nrf2) for degradation (*89, 109-111*). This stimulates gene expression through the antioxidant response element (ARE). 15d-PGJ_2 can also behave as an activating ligand for peroxisome proliferator-activated receptor γ (PPAR γ), a transcription factor involved in the regulation of adipocyte differentiation. Similarly, lipid peroxidation product 4-hydroxynonenal (HNE) is also believed to function as an intracellular signaling molecule, stimulating gene expression through ARE, and Nrf2 (*112*). Both 15d-PGJ_2 and HNE have been implicated in inhibition of nuclear factor κB dependent gene expression (*113, 114*). GSTs can catalyze the addition of GSH of both of these lipid mediators.

Structurally, cytosolic GSTs share in common the canonical GST fold, despite low sequence similarity across classes (>50% within class), with each subunit consisting of two distinct domains (Figure 1.5). Each subunit of these dimeric enzymes has a molecular weight of 25 kDa on average; approximately one-third of the protein constitutes the N-terminal domain, topologically similar to thioredoxin, and is considered to be in the same superfamily, along with glutaredoxin and glutathione peroxidase (*115*). This domain adopts a β1-α1-β2-α2-β3-β4-α3 motif forming a mixed four strand β-sheet structure, where the β-sheets are essentially in the same plane, with the α-helices facing the solvent above and below. The N-terminal domain, domain I, is highly conserved across classes, and provides the binding site for glutathione.

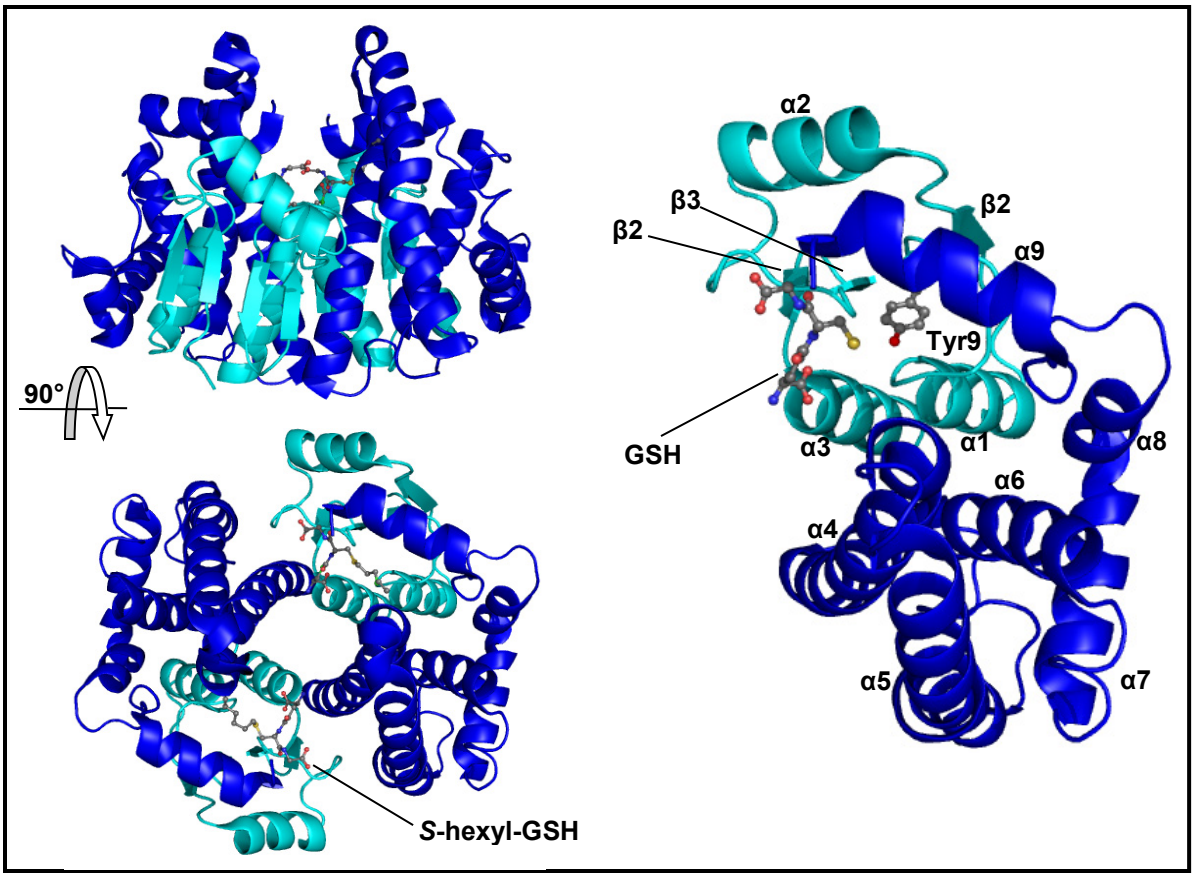


Figure 1.5 Crystal structures of GSTA1-1 dimer with *S*-hexyl GSH bound (*left*) or monomer with GSH only bound (*right*). A) GSTA1-1 homodimer is shown in ribbon representation with *S*-hexyl GSH bound depicted in ball-and-stick form. Domain I is colored *cyan*, and domain II is in *dark blue*. The top view is perpendicular to the C2 symmetry axis relating the two subunits, and the bottom view along the C2 axis. PDB:1k3y B) GSTA1-1 monomer with GSH (ball-and-stick), along the C2 symmetry axis. Also in ball-and-stick is Tyr9, the major catalytic residue. Secondary structure elements are labeled, and domain I is colored *cyan*, with domain II in *dark blue*. PDB:1pkw

The most conserved region of the GSH binding site (G-site) is the core ββα motif. It includes an unfavorable, but highly conserved *cis*-proline in the β3 – α2 connecting loop, which recognizes the γ-glutamyl portion of GSH (*72*). Glutathione binds to the G-site in an extended conformation, anchored by several electrostatic interactions and hydrogen bonds. The bound GSH molecule is generally oriented with γ-Glu toward the dimer interface, Gly near the surface of the enzyme, and the cysteinyl sulfur atom aimed toward the subunit to which it is bound. Domain I also supplies two residues providing hydrogen bonding partners for the amino and

carboxylate groups of γ-glutamine, a serine or threonine and a glutamine or glutamate, depending on the isoform. Less conserved across GST classes are residues interacting with the sulfur atom of GSH. The theta class employs the hydroxyl group of Ser9 (*116*) for activation of the GSH sulfhydryl group, while the A, M, P, and S class GSTs utilize tyrosine residues (*117-120*). The specificity of the G-site for glutathione is quite high, and presumably GSTs tend to remain saturated, as the concentration of GSH in normal cells (1-10 mM) is approximately three orders of magnitude higher than the dissociation constant between glutathione and enzyme (*84*).

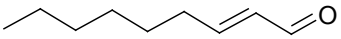
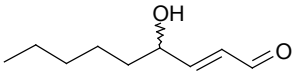
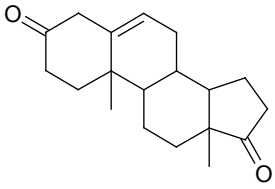
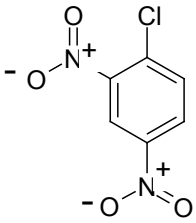
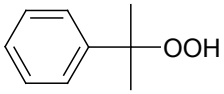
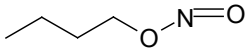
The C-terminal domain II exhibits much more variation than domain I. Connected by a short linker to domain I, the C-terminal domain consists of an all α-helical fold, with a four helix core bundle motif, unique to GSTs. The total number of helices varies across class, with the pi and mu classes each having five (*117, 121, 122*), and the alpha class with six (*122*). Although formed by residues from both domains, the hydrophobic substrate binding crevice (H-site) obtains most of the isoform based substrate specificity difference from domain II residues. Each class tends to have characteristic features in the domain, notably the mu-loop of the mu class GSTs and the α9-helix of the alpha class. Upon assembly into functional dimeric form (*123*), soluble GSTs demonstrate twofold (C2) symmetry across subunits. The subunit interface displays some variation between classes. Alpha, mu, and pi class GSTs contain a ball and socket type interaction, where an aromatic residue (Phe52 in alpha, Phe56 in mu, and either Phe47 or Tyr49 in pi) from domain I nestles into a hydrophobic pocket between the α4 and α5 helices of domain II. General hydrophobic/hydrophilic interactions characterize the subunit interface of theta and sigma classes. A crystal structure for the kappa class has not yet been solved (*72, 85*).

The catalytic mechanism for GSTs involves activation of the sulfhydryl group of bound GSH, effectively lowering the pK_a of the cysteine thiol from 9.2 in solution, to 6.7 (GSTA1-1).

This results in the formation of a nucleophilic thiolate anion at physiological pH (*72, 86, 124, 125*). The alpha class GSTs have recruited two conserved residues for stabilization of the sulfur in the GST·GSH complex through hydrogen bonding (*72*). Of these residues, Tyr9 and Arg15 in rGSTA1-1, hGSTA1-1, and hGST4-4 (*126-128*), the tyrosine appears to contribute more significantly to thiol stabilization. The arginine, also involved in a salt bridge interaction between the two GST domains, contributes to additional lowering of the tyrosyl hydroxyl group pK_a (*122*). The pK_a value of Tyr9 in hGSTA1-1 and hGSTA4-4 is 8.1 (*126*) and 6.7 (*128*) respectively, compared to 10.3 for free tyrosine. It is now generally believed that the shared proton of the reactive species complex resides on the phenolic oxygen of the tyrosyl residue in the alpha class GSTs (*72*). Additional second sphere interactions increase the strength of the first sphere hydrogen bond interactions, and generally contribute to the electropositive field near the thiolate (*72*). The precise chemical reaction mechanism is substrate dependent, but begins with nucleophilic attack by the stabilized thiolate anion.

Two isoforms of human glutathione transferase, GSTA1-1 and GSTA4-4, embody several notable structural features and catalytic properties for the examination of catalytic promiscuity and the putative relation to protein dynamics and conformational heterogeneity. GSTA1-1 represents an archetypal example of a detoxification enzyme, exhibiting catalytic activity toward a diverse array of substrates, while homologous GSTA4-4 demonstrates high substrate selectivity toward long chain alkenals derived from lipid peroxidation, particularly 4-hydroxynonenal (*129-131*). The substrate profiles of both isoforms overlap considerably, however GSTA4-4 typically displays catalytic efficiencies more than 1000-fold higher toward long chain alkenals. Table1.1 illustrates the structural diversity of substrates metabolized by GSTA1-1, and the substrate selectivity exhibited by GSTA4-4.

Table 1.1 Selected substrates of GSTs and associated catalytic efficiencies.

Substrate	Structure	k_{cat}/K_M (mM^{-1} s^{-1})		Reference
		GSTA1-1	GSTA4-4	
Nonenal		10.3	5650	142
4-hydroxynonenal		13	3100	135
Δ^5-androstene-3,17-dione		614	0.63	142
1-chloro-2,4-dinitrobenzene		130	7.9	15
Cumene hydroperoxide		80.2	0.76	142
Butylnitrite		367	2.42	142

This pair of isozymes also exhibits a striking convergence of sequence and structure. GSTA1-1 and GSTA4-4 are 54% sequence identical, and employ the same residues for activation of the thiol anion. The crystal structures also overlay almost exactly (Figure1.6).

Both of these isoforms have a class-specific amphipathic C-terminal α9 helix located at the active site of each subunit, residues 208-222 (*129, 132*), which is highly dynamic in the GSTA1-1 isoform, and localized along one side of the H-site in GSTA4-4 (*90, 133*). A major contributing factor to the differing levels of promiscuity is thought to arise from the differential local mobility of this C-terminus and the α4-α5 helix-turn-helix motif on the other side of the substrate binding H-site, which encompasses the majority of the sequence variation between these isoforms (*123, 134, 135*).

Figure 1.6 Aligned crystal structures of apo-GSTA1-1 and apo-GSTA4-4. GSTA1-1 is shown in yellow (PDB:1pkz) and GSTA4-4 is shown in blue (PDB:1gum). Structures were aligned using PyMol.

Crystallographic studies conducted with this enzyme in the apo form have shown either a lack of electron density for the C-terminal helix (*136, 137*), or evidence of 'open' and 'closed' conformations (*132*). Other investigations into GSTA1-1 using NMR have indicated a very flexible and dynamic C-terminus, with chemical exchange in the 1-10 ms time range (*83*). Similar examinations of GSTA1-1 have demonstrated that the C-terminus becomes much more localized upon binding of ligand or conjugate, and somewhat less so when GSH is bound to the G-site (*83, 132, 138*). In addition, stopped-flow kinetic experiments indicate that motions of the C-terminal helix control product release for some substrates (*139-141*). Combined, these results show strong evidence for the presence of a GSTA1-1 conformational ensemble. Efforts to modulate the dynamic properties of C-terminus and shift the substrate selectivity profiles from one isoform to the other (*15, 135, 142*) have resulted in a series of GST variants displaying intermediate catalytic profiles between naturally promiscuous GSTA1-1 and substrate selective

GSTA1-1. Two exchange mutants, GSTA1-1 V111F/R217Y and GSTA4-4 F111V/Y217R, add a stabilizing edge-on-face aromatic interaction between the C-terminal helix and the α4-α5 helix-turn-helix region of GSTA1-1, while removing it from GSTA4-4 (*15*). Another variant, termed GIMFhelix, was constructed using the core domain of GSTA1-1 and replacing the entire C-terminus, residues 208-222, with the α9 helix of GSTA4-4, along with the following substitutions: A12G, L107I, L108M, and V111F (*135*). Table 1.2 compares the catalytic efficiencies toward CDNB, the traditional substrate used for assessing glutathione transferase activity, and HNE one of the presumed physiological substrates of GSTA4-4 (*90*).

Table 1.2 Catalytic efficiencies of GSTs towards CDNB or HNE.

Enzyme	k_{cat}/K_M (mM^{-1} s^{-1})		Reference
	CDNB	HNE	
A1-1 wild-type	130	13	15
A1-1 V111F/R217Y	56	21	15
A4-4 F111V/Y217R	60	110	15
GIMFhelix	14	1010	135
A4-4 wild-type	4.5	1400	15
	7.9	3100	135

Introduced here, evidence of catalytic promiscuity in GSTA1-1 and substrate specificity in GSTA4-4 combined with increased conformational heterogeneity of GSTA1-1 over GSTA4-4 will be explored in the rest of this work.

Chapter 2: Glutathione Transferases Examined with Time-resolved Fluorescence

A. Introduction

As introduced in Section 1D, GSTA1-1 and GSTA4-4 demonstrate strong similarity in amino acid sequence, and are remarkably alike in overall structure, with an average root mean squared distance (RMSD) of 0.91 Å in the aligned peptide backbones (*15, 129*). Sequence variations are distributed throughout; however the differences are concentrated in the C-terminal helix and α4-α5 helix-turn-helix structural motif, the 'tower' region extending above the H-site and creating the intersubunit cleft at the dimer interface. These sequence variations, and resulting modifications in the tertiary interactions, are posited to provide the basis of differential conformational heterogeneity between the two isoforms (*15*). This in turn appears to result in the highly divergent catalytic profiles that distinguish promiscuous GSTA1-1 from substrate selective GSTA4-4.

Specifically, the C-terminal α9 helix of GSTA1-1 is highly disordered in the absence of ligand, and becomes localized in the presence of various ligands, with the final conformation determined by the ligand identity (*83, 131, 136, 138, 143*). In these crystallographic studies of GSTA1-1, the authors noted an almost total absence of electron density associated with the C-terminal helix, residues 208-222, in the apo-enzyme structures. Also apparent was a much more defined helix structure when an ethacrynic acid–GSH conjugate ligand was bound. The C-terminal helix closes over the active site, essentially acting as a lid to exclude solvent. Moreover, the location of the helix is able to shift to accommodate the bulkier ligand, in comparison to the [GSTA1-1·benzyl-GSH] complex. Additionally, a crystal structure for the [GSTA1-1·*S*-hexyl-GSH] complex has also been determined (*137*). Here, the C-terminal helix is in a location that is

different from the benzyl-GSH. Furthermore, even when the helix is closed over the active site, there remains enough space for two different orientations of the ligand. Even when well ordered, the helix exhibits high crystallographic B-factors compared to the rest of the structure. A more recently, published crystal structure of apo-GSTA1-1 exhibited a reasonably well resolved α9 helix, including residues up to 215 in the A subunit, and up to 219 in the B subunit (*132*). Here also, the temperature factors were high, indicating a flexible structure. In all cases, the non-conjugate ligand did not stabilize the helix to the extent the GSH-conjugates did. Furthermore, the α4-α5 helix-turn-helix that constitutes the tower region also demonstrated increased temperature factors in the crystal structures. This conformational plasticity around the active site seems to be the mechanism that enables GSTA1-1 to accommodate a wide array of substrates with similar catalytic efficiency between them.

In contrast, the C-terminal helix region of GSTA4-4, shown in Figure 2.1, is located along one side of the narrow hydrophobic cleft that accommodates the preferred long chain alkenal substrates of this isoform (*129, 144*). This helix is well resolved in the available crystal structures, even in the absence of ligand. A notable feature is the presence of an edge-on-face aromatic interaction between residues Phe111 and Tyr217 (Figure 2.2). Conspicuously absent in GSTA1-1, this interaction connects a phenylalanine residue located in the tower region at position 111 to a tyrosine in the α9 helix and it has been shown to stabilize the active site region of this enzyme (*15*). GSTA1-1 has recruited a valine (V111) and an arginine (R217) at these positions. Exchange mutants were made, where the respective residues were swapped between isoforms, creating GSTA1-1 V111F/R217Y and GSTA4-4 F111V/Y217R. Hypothetically, addition of the aromatic-aromatic interaction to GSTA1-1 would reduce protein dynamics in the C-terminal helix and tower regions, while the corresponding loss should destabilize the

corresponding areas in GSTA4-4. In fact, hydrogen/deuterium exchange mass spectrometry (HDX) experiments demonstrated an increase in protein dynamics and solvent exposure of the C-terminal helix and tower regions of GSTA4-4 F111V/Y217R, while the reverse was true of GSTA1-1.

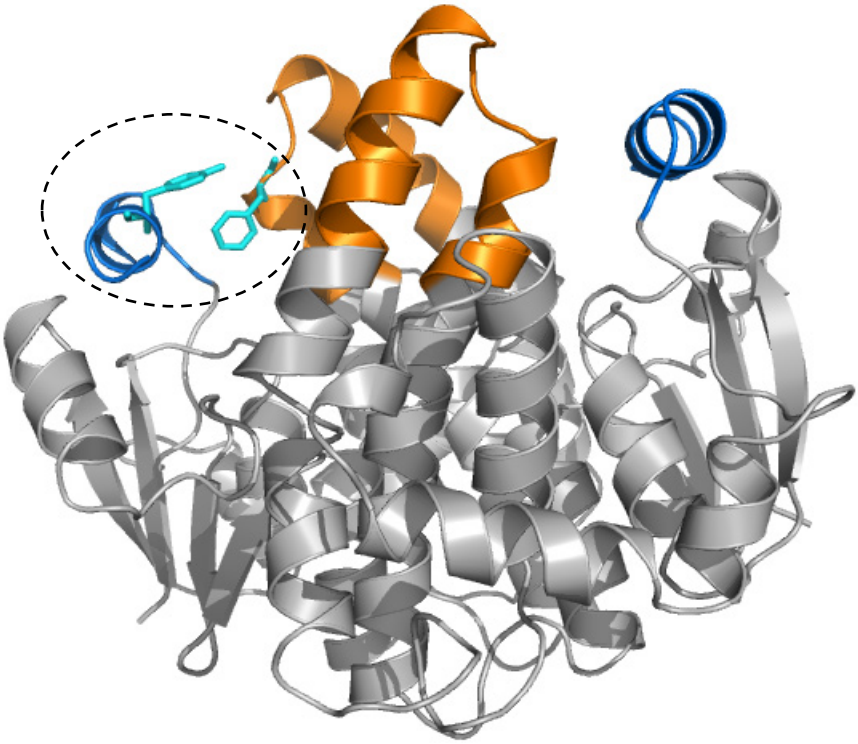


Figure 2.1 Ribbon diagram of GSTA4-4. Residues of interest, Phe111 and Tyr217 are shown in *cyan*, while the C-terminal α9 helix and α4-turn-α5 tower motif is illustrated in *orange*. The area indicated by the *dashed circle* is expanded in Figure 2.2. PDB: 1GUM

To gain additional insight into the mechanism through which GSTA1-1 gains catalytic promiscuity, time-resolved fluorescence was used to probe the conformational heterogeneity of these enzymes, and lifetime distribution analysis was used to obtain a relative measure for the breadth of conformational ensembles extant on the nanosecond timescale.

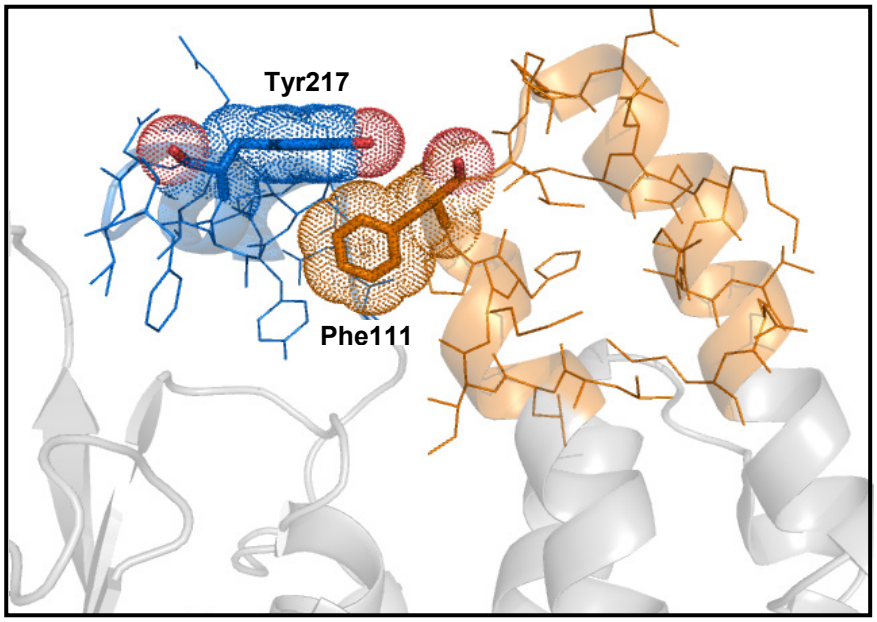


Figure 2.2 Interdomain interactions of GSTA4-4. The aromatic edge-on-face interaction is displayed in a close view. PDB:1GUM

B. Theory of Time-resolved Fluorescence Distribution Analysis

Time-resolved fluorescence spectroscopy is an excellent method to study conformational heterogeneity and dynamics of protein structure. This method monitors events occurring during the excited singlet state lifetime of the fluorophore under consideration, which can range from picoseconds to hundreds of nanoseconds, making it possible to examine biologically relevant events such as whole protein rotation, local conformational shifts, and amino acid side chain fluctuations or molecular interactions with solvent, ligand, or neighboring residues (*145, 146*). The excited state of a fluorophore, particularly tryptophan, is highly sensitive to the local physiochemical environment, and investigation of intrinsic protein fluorescence, particularly fluorescence lifetime based methods, is a widely used and effective method for study of protein dynamics and conformations. Tryptophan, is a complex fluorophore with two nearly isoenergetic transitions, and it has several advantages for probing protein conformational shifts and local

microenvironment. Tryptophan, with a quantum yield of 0.13 in neutral aqueous solution of at 23 °C, can be electively excited at 295 - 305 nm, which avoids complicating fluorescence emission from tyrosine or phenylalanine (*146*). Additionally, Trp is present in protein at a level of approximately 1 mole % (*146*), so proteins often contain very few or single tryptophan residues, and exhibit strong environment dependent variation in emission characteristics. Azurins, small (~15 kDA) copper containing, proteins display the highly blue-shifted fluorescence emission spectra, often as low as 308 nm (*147*). In this case, the tryptophan residue is buried in a hydrophobic environment, with very little exposure to water, while in contrast the tryptophan present in glucagon is entirely solvent exposed, and demonstrates an emission maximum of 352 nm. Decay rates are also variable, with fluorescence lifetimes in proteins ranging by more than a factor of 100 (*145*).

The indole photophysics and fluorescence decay rates of protein tryptophan residues are complex, and often multiexponential; free tryptophan in solution has biexponential decay components of 3.1 ns and 0.5 ns, resulting from rotational isomers (*148*), and most proteins, even single tryptophan structures, exhibit at least biexponential intensity decays. Another complicating factor is the presence of two electronic transitions, to the ^1L_a and ^1L_b states in the last absorption band, oriented nearly perpendicular to each other (*149*). While emission from the ^1L_b state is infrequent in proteins, complete absence of hydrogen bonding to the imino group will result in structured ^1L_b emission and occasional introduction of additional fluorescence lifetime complexities. The dominant ^1L_a transition shifts to lower energies in more polar environments and greater strength of hydrogen bonding. Moreover, the measured fluorescence lifetime, τ_m, depends on the natural lifetime, τ_n, and sum of nonradiative (k_{nr}) decay rate constants (Equation 2.1).

$$\tau_m = 1/\left(\tau_n^{-1} + \sum k_{nr}\right) \quad \text{(Eq. 2.1)}$$

In addition to solvent effects, nonradiative decay processes resulting from quenching of tryptophan fluorescence in proteins comes from numerous sources: proton transfer from charged amino groups, quenching from electron acceptors, electron transfer by disulfides, amides, peptide bonds, and resonance energy transfer (*150-153*). These quenching effects, which essentially reduce the excited state lifetime, and hence the measured quantum yield, are strongly distance dependent. In particular, quenching via electron transfer, decreases exponentially with distance. Nearby charged groups can also stabilize or destabilize the charge transfer state (*154*). Dynamic quenching by groups intrinsic to the protein matrix, generally ionic species, will be determined by structural fluctuations occurring on the same nanosecond timescale as the emission process, and the quenching moieties must be sufficiently close to allow for collisions between the spheres of charge interaction. In summary, fluorescence decay rates in proteins, dominated by tryptophan residues, are strongly dependent on the local microenvironment resulting from variations in protein conformation and dynamics.

Time-resolved fluorescence data is typically examined through a multi-exponential decay model, where the intensity as a function of time, *I(t)* is described by,

$$I(t) = \sum_i \alpha_i e^{(-t/\tau_i)} \quad \text{(Eq. 2.2)}$$

and τ_i is the lifetime of the *i*th component, with its associated pre-exponential factor, α_i, which is a measure of its fractional contribution (or amplitude) to the total intensity of emission. The advantages to this model are simplicity and utility. The multi-exponential model can be used to fit almost any intensity decay, and is a simple sum of the average lifetime of each resolved component, modulated by that component's relative contribution to the total intensity decay. This method does not, however, provide any information on the functional form of the intensity

decay, and it is often quite difficult to obtain an intuitive grasp of the decay pattern from a simple numerical lifetimes. In addition, tryptophan fluorescence decay in proteins nearly always consists of at least a double-exponential decay pattern, and resolvability of fluorescence lifetimes is limited using this model (*146*). As the population of fluorophores becomes more conformationally heterogeneous, intensity decays also become more complex, and it is often difficult to assign any quantitative physical meaning to exponential lifetime components obtained when using such a simple model. Replacement of the multi-exponential model with different functional forms to describe the intensity decay can often result in a much better understanding of the underlying decay processes at work. In particular, the forms of complex intensity decay can often be visualized quite well from the distributions of fluorescence lifetimes.

From an ensemble perspective, a set of quasi-continuous fluorescence lifetime distributions can arise from the strong dependence of tryptophan fluorescence on local conformational changes and dynamics. Pioneered by Alcala, Gratton, and Prendergast (*155-159*), this model assumes that for a given fluorophore, the conformation of the protein determines the various quenching mechanisms that affect its excited state lifetime, and that protein dynamics will moderate the number of conformations, and thus the number of environments, that fluorophore will experience during its excited state lifetime, consequently resulting in a lifetime change. The width of the distribution of lifetimes reports on the conformational heterogeneity of substates available for sampling, during the excited state lifetime of the fluorophore. Using this model, the fluorescence intensity, I_F, as a function of time (t) can be described by

$$I_F(t) = \sum_{i=1}^{n} f_i \tau_i^{-1} e^{-t/\tau_i} \qquad \text{(Eq. 2.3)}$$

where n is the number of independent decay components in the system and f_i is the ith fraction of light contributing to the total intensity, proportional to the product of that fraction and its

associated lifetime τ_i. When the total number of components is large, the sum can be expressed as a limit through

$$I_F(t) = \int_0^\infty f(\tau)\tau^{-1}e^{-t/\tau}d\tau \quad \text{(Eq. 2.4)}$$

In the frequency domain, if sinusoidally modulated light at an angular frequency ω is used for excitation as a function of time such that

$$E(t) = E_0\left[1 + M_e \sin\omega t\right] \quad \text{(Eq. 2.5)}$$

and E_0 is the average intensity and M_e is the modulation of excitation, then the fluorescence response as a function of time is

$$F(t) = F_0\left[1 + M_f \sin(\omega t - \varphi)\right] \quad \text{(Eq. 2.6)}$$

where F_0 is the average fluorescence. The phase angle shift (φ) and modulation (m) are the observed quantities and are related to the fluorescent intensity via the sine, $S(\omega)$, and cosine, $G(\omega)$, Fourier transforms of the fluorescence intensity shown by

$$\varphi = \tan^{-1} S(\omega)/G(\omega) \quad \text{(Eq. 2.7)}$$

The distribution functional used for fitting is described in terms of fractional contribution to the lifetime and parameters related to the characteristic behavior of the distribution. The distribution can take a variety of forms. For example uniform, Gaussian, and Lorentzian distributions have all been used to describe fluorescence decay, and can be unimodal or multimodal depending on the system under study. To study the dynamics of the C-terminus of GSTA1-1, previous members of the Atkins lab made the double mutant W21F/F222W, which replaces the single Trp of the wild-type with Phe and incorporates a Trp at the C-terminus. This mutant is well described and has catalytic properties similar to wild-type GSTA1-1 (*134*). Using this model, the fluorescence lifetime distributions of GSTA1-1 and GSTA4-4 wild-types, as well as several

mutants were examined. Conveniently, the GST isoforms include only single tryptophan residues at position 21 in the native structures enabling examination without complicating factors of multiple tryptophan reporting and resonance energy transfer between Trp residues.

C. Materials and Methods

All chemicals and reagents were purchased from Sigma-Aldrich or Fischer Scientific unless otherwise noted.

Protein Expression and Purification

The gene encoding hGSTA1-1 wild-type, was previously subcloned into the pKK233-3 vector, and the gene encoding hGSTA4-4 wild type with ubiquitin tag (pRB173 vector) was previously subcloned in to the pRB269 vector. The ubiquitinase gene was subcloned into the pRB173 vector. GSTA1-1 W21F/F222W was previously constructed and subcloned into the pKKGTB34. The hGSTA1-1 V111F/R217Y, and hGSTA4-4 F111V/Y217R mutants were previously constructed using overlap extension PCR and subcloned into pCWOri+ vectors (*15*). All vectors were used to transform *Escherichia coli*, BL21 strain (Agilent Technologies – Stratagene Products, La Jolla, CA). For GSTA1-1 and mutants, bacterial cultures were grown for 12 h in lysogeny broth (LB). Glycerol stock cultures were made through the addition of 15% (v/v) glycerol, frozen in dry ice/acetone and stored at -80°C. Glycerol stocks were used to streak 100 μg/mL ampicillin LB agar plates and cultures were grown overnight at 37°C. Isolated colonies were used to inoculate 5 mL LB broth starter cultures with 50 μg/mL ampicillin added, and grown at 37°C with 225 rpm shaking for 5 – 8 hours. These starter cultures were then used to inoculate 500 mL terrific broth (TB) cultures containing 100 μg/mL ampicillin and grown at 37 °C with 225 rpm shaking until an optical density (OD) of 0.5 (approx. 3 h) was reached.

Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce protein expression, followed by 18 h continued incubation, after which the cells were harvested via centrifugation. For GSTA4-4, LB agar plates contained 50 μg/mL ampicillin and 34 μg/mL chloramphenicol, and other media contained 100 μg/mL ampicillin and 34 μg/mL chloramphenicol. Growth conditions were same, but incubation time after induction was shortened to 6 hours.

Cell pellets were resuspended in PBS, pH 7.4, and cells were lysed using a French press at 10,000 psi. Cellular debris was removed through centrifugation at 125,000 G for 1 h, and the supernatant containing protein was purified via affinity chromatography using *S*-hexyl-GSH-agarose (Pierce Biotechnology, Rockford, IL). The column was pre-equilibrated with 50 mM Tris-HCl, pH 7.5, supernatant loaded and washed with 5 column volumes 50 mM Tris-HCl, followed by elution with 10 mM glutathione, 50 mM Tris-HCl, pH 9.4, all at 4 °C. Activity was checked by assaying conjugation of reduced glutathione to 1-chloro-2,4-dinitrobenzene, monitored though change in absorbance at 340 nm. Active fractions were pooled and exchanged into 100 potassium phosphate buffer, pH 6.5 (GSTA1-1 wild-type and mutants) or 100 mM sodium phosphate buffer, pH 6.5 (GSTA4-4 wild-type and mutants). In some cases, additional purification was necessary, in which case a diethylaminoethyl column was equilibrated using 10 mM Tris-HCl containing 1 mM ethylaminediaminetetraacetic acid, pH8.1. Pooled fractions were loaded, and the same buffer used to elute. Purity was checked using SDS-PAGE, and found to be > 95% pure. Protein concentrations were determined via absorbance, with the following molar extinction coefficients at 280 nm for folded dimer: GSTA1-1 wild-type, 4.27 X 10^4 $M^{-1}cm^{-1}$; GSTA1-1 V111F/R217Y, 3.65 X 10^4 $M^{-1}cm^{-1}$; GSTA1-1 W21F/F222W, 4.54 X 10^4 $M^{-1}cm^{-1}$;

GSTA4-4 3.18 X 10^4 $M^{-1}cm^{-1}$; GSTA4-4 F111V/Y217R 3.26 X 10^4 $M^{-1}cm^{-1}$. All enzymes were stored at -80°C until used.

Fluorescence Measurement and Analysis

Steady-state fluorescence spectra were collected using a SLM-Aminco spectrofluorometer (SLM Instruments Inc., Urbana, IL) with slit-widths set to 4 mm, and excitation monochromometer set to 295 nm. Emission spectra were collected from 300 to 450 nm, and are averages of three scans over this wavelength range. Integration time for each point was set to 0.1 seconds, and protein concentrations used were 1 μM dimer in 100 mM potassium phosphate buffer, pH 6.5. Standard 10 mm quartz fluorescence cuvettes were used. Steady state fluorescence quenching studies were carried out using the same instrument, protein concentrations and conditions. 10 M ultrapure acrylamide (Invitrogen Life Technologies, Grand Island, NY) was used for the quenching studies, titrated in 5 μL aliquots, with a final volume change of <5 %. Construction of Stern-Volmer plots used fluorescence intensity at peak maximum. Fluorescence lifetime measurements were collected using a Spex Fluorolog Tau3 frequency domain spectrofluorometer with a 450-watt xenon arc lamp (Horiba Jobin Yvon, Edison, NJ). For each experiment, 20-25 frequencies were chosen, ranging from 1 to 260 MHz. Ten replicate measurements were obtained at each frequency, with an integration time of 60 s for each measurement. The excitation monochromator was set at 295 nm, and emission signals was collected through a Schott WG 305 nm high pass filter at each frequency. The iris opening and slit widths were adjusted to maximize signal, while retaining appropriate modulation values. The lifetime reference standard used was *p*-terphenyl dissolved in ethanol (1.05 ns), and its concentration was adjusted to achieve comparable signal strength between reference and sample. Data were collected at a final protein concentration of 50 μM GST dimer. Final *S*-hexyl-GSH

concentration was 2 mM. GSTA1-1 samples were in 100 mM potassium phosphate buffer, pH 6.5, GSTA4-4 samples were in 100 mM sodium phosphate buffer, pH 6.5. Samples with *S*-hexyl-GSH included 1% ethanol for solubility. All experiments were performed at 12°C. The model detailed in the previous section was used for fluorescence lifetime distribution analysis, with the data presented using a Gaussian function of the form

$$f(\tau) = Ae^{-\frac{(\tau-b)^2}{2c^2}} \quad \text{(Eq. 2.8)}$$

where the fitting parameters are the amplitude, A, peak center, b, and width, c, is related to the full-width at half-maximum (FWHM) by

$$FWHM = 2\sqrt{2\ln 2}c \quad \text{(Eq. 2.9)}$$

and the decay data fit using the Globals software package, developed at the Laboratory for Fluorescence Dynamics at the University of Illinois at Urbana-Champaign.

D. Results and Discussion

Steady-state Fluorescence

Steady-state fluorescence spectra were collected for GSTA1-1 and GSTA4-4 wild-type enzymes, as well as GSTA1-1 W21F/F222W, GSTA1-1 V111F/R217Y, and GSTA4-4 F111V/Y217R. As illustrated in Figure 2.3 A, the fluorescence spectra for GSTA1-1 wild-type and GSTA1-1 V111F/R217Y overlay almost exactly, with peak maxima at 325 and 327 nm, respectively. A slight red-shift was noted for GSTA4-4 wild-type, moving the peak maximum to 329 nm, and the peak maximum of GSTA4-4 F111V/Y217R was 328 nm. The fluorescence spectra for these GST variants are indicative of the hydrophobic environment of Trp21, the native tryptophan at the domain interface. GSTA1-1 W21F/F222W demonstrated a large red-shift to 346 nm at peak maximum, due to the much greater solvent exposure of Trp222 at the end of the C-terminus.

Fluorescence quenching experiments using non-penetrative acrylamide revealed similar results Figure 2.3 B. Solvent-accessible Trp222 fluorescence was readily quenched, with a quenching K_D of 4.68 M^{-1}. The native Trp21 residues in GSTA1-1 wild-type and GSTA1-1 V111F/R217Y were nearly identically quenched, with K_D's of 1.50 M^{-1} and 1.56 M^{-1}. GSTA4-4 F111V/Y217R exhibited a slightly higher quenching constant of 1.95 M^{-1}. Interestingly, GSTA4-4 wild-type demonstrated more accessibility to acrylamide with a quenching constant of 2.24 M^{-1}. In all cases, the fluorescence quenching as a function of acrylamide concentration was linear, indicating a single tryptophan residue was responsible for the fluorescence observed in these enzymes.

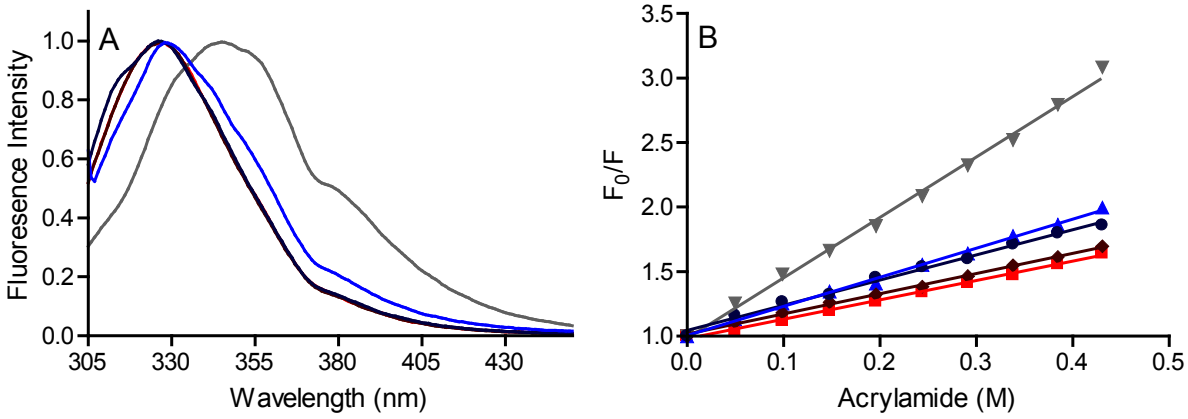


Figure 2.3 Normalized steady-state fluorescence spectra and Stern-Volmer fluorescence quenching plots of GSTs. A) Steady-state fluorescence spectra of GSTA1-1 wild-type (*red*), GSTA1-1 W21F/F222W (*grey*), GSTA1-1 V111F/R217Y (*maroon*), GSTA4-4 F111V/Y217R (*navy*), and GSTA4-4 (*blue*) are shown. B) Plots of the ratio of fluorescence in the absence (F_0) and presence (F) of quencher are shown in the same colors. Linear regression lines are also shown, where the slope is the quenching constant (K_D) described in the text. GSTA1-1 wild-type (■), R^2 = .9977, GSTA1-1 W21F/F222W (▼), R^2 = .9960, GSTA1-1 V111F/R217Y (♦), R^2 = .9977, GSTA4-4 F111V/Y217R (●), R^2 = .9930, GSTA4-4 wild-type (▲), R^2 = .9965.

Time-resolved Fluorescence

To further investigate the C-terminal α9 helix of GSTA1-1, time-resolved fluorescence parameters were obtained for the GSTA1-1 W21F/F222W mutant, with and without *S*-hexyl-GSH. The retrieved data were analyzed using several forms of distribution, including unimodal and bimodal uniform, Gaussian, Lorentzian, and Cauchy distributions. In all cases, the best fit

was obtained using a unimodal Gaussian distribution in combination with a short lifetime discreet component (analogous to a distribution of 0 width), as is common in analyses of protein fluorescence (*160*).

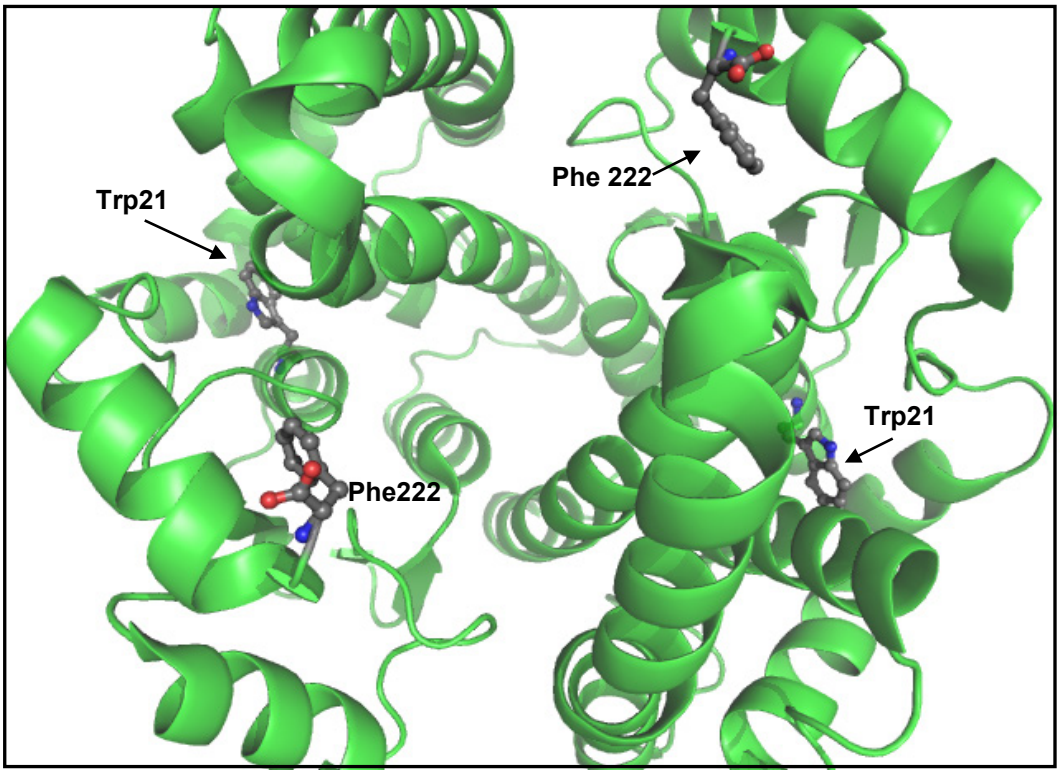


Figure 2.4 GSTA1-1 Ribbon diagram. The view is along the two fold axis that relates the subunits. The native GSTA1-1 residues mutated in the W21F/F222W are labeled.

As expected, this mutant, where the phenylalanine residue at the end of the helix was replaced with a tryptophan (and the native Trp replaced with Phe), exhibited a very broad distribution centered at 6.5 ns, with a minor short-lived component at 0.9 ns. The distributions are illustrated in Figure 2.5, and data in Figure 2.6. Generally, in distributive analysis of fluorescence lifetime data, the width of the Gaussian reflects the width of the potential energy well of conformational substates. The conformational heterogeneity of the structural ensemble at the time of excitation is correlated to the width of the distribution, provided the majority of substates do not equilibrate

faster than the excited state lifetime. Thus, the full-width at half-maximum (FWHM) provides a convenient measure of the conformational heterogeneity of the ensemble. The recovered FWHM of 4.75 ns for GSTA1-1 W222F/F222W is indicative of a wide potential energy well of many conformational substates. The C-terminal tryptophan is highly non-localized, and experiences a wide variety of microenvironments, with differential distances to quenching moieties and variable degrees of solvent exposure. This agrees well with HDX experiments, which indicated a high level of deuterium exchange in the C-terminal helix region in both the fast (10 s) and slow (>10 s) time regimes (*15*). In contrast, upon the addition of saturating concentrations of a product-like ligand, *S*-hexyl-GSH, the broad distribution is eliminated in favor of a longer lifetime component, with a negligible width. Moreover, fits of these data to a sum of discreet exponential terms yielded a fit nearly as good, with no significant change in average lifetime values. The C-terminal helix in this case is presumably closed down over the ligand and locked into place.

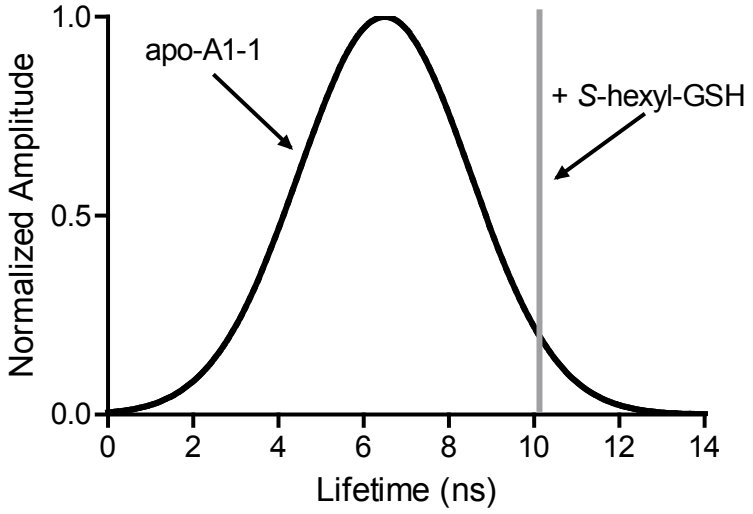


Figure 2.5 Lifetime distributions of GSTA1-1 W21F/F222W. The normalized Gaussian distribution of fluorescence lifetimes for the apo-A1-1 C-terminal Trp222 residue is displayed in *black*. The *grey* line indicates the mean lifetime after addition of saturating concentrations of *S*-hexyl-GSH, expanded in width for clarity. Data and fits are in Figure 2.6.

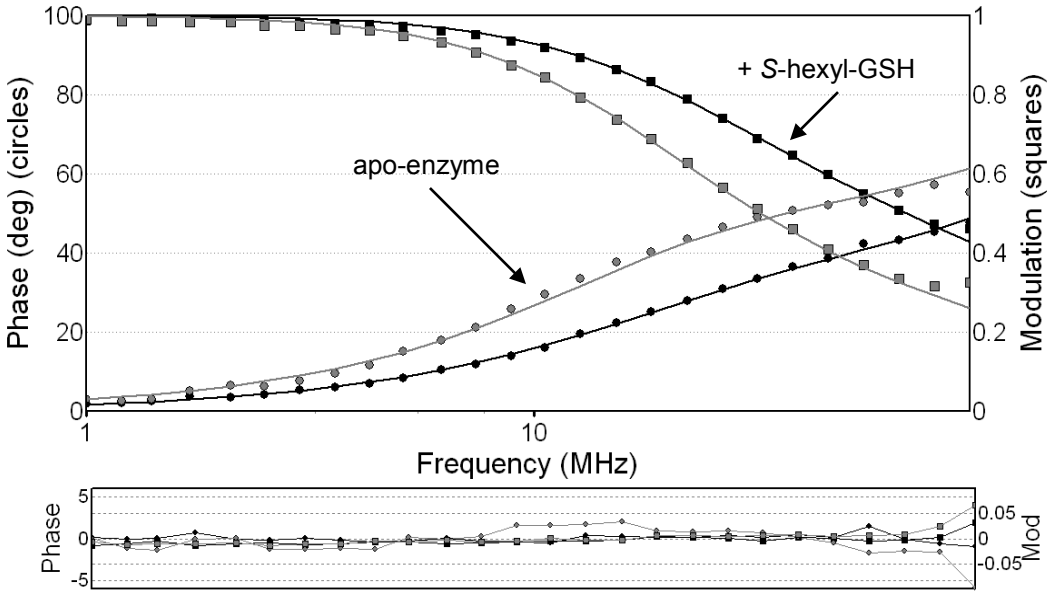


Figure 2.6 Time-resolved fluorescence data for GSTA1-1 W21F/F222W. Top panel: Phase (*circles*) and modulation (*squares*) data are shown, with fits indicated by lines. Apo-enzyme is *grey* and +*S*-hexyl-GSH in *black*. Bottom panel: Residuals obtained after best fits. Data also appears in reference 15.

The native tryptophan residue at position 21 is located at the intrasubunit domain interface, and previous time-resolved fluorescence and fluorescence quenching studies probing the analogous residue in rat glutathione transferase A1-1 (rGSTA1-1) have shown that the enzyme undergoes ligand dependent microconformational reorganization, despite its distance from the H-site (*161*). Additionally, while G-site functionality does not appear to be affected by this mutation to this residue, at the H-site, located 15 Å away from the native tryptophan, ligand binding becomes less favorable upon replacement with an alanine at position 21 (*162*). Trp21 extends from the α1 helix of domain I into a hydrophobic pocket in domain II, and appears to be located very near a non-substrate binding site for anionic dyes 8-anilino-1-napthalene sufonate (ANS) and bromosulfophthalene (BSP) (*162, 163*), and may impact the ligandin activity of GSTs for similar compounds. While substitution of the native tryptophan with a phenylalanine did not adversely affect the catalytic function of GSTA1-1, and only a slight decrease in thermal stability was noted for this mutant, replacement with alanine created a pocket at the domain interface

(*162, 163*), increasing the susceptibility of the enzyme to thermal denaturation and reduced the transition state stabilization energy for the Meisenheimer complex formed during the conjugation of GSH to CDNB.

Trp21 was exploited in order to obtain information about the solution dynamics of the domain interface region of GSTA1-1 as compared to GSTA4-4 from an ensemble based perspective. As above, the data were fit to a combination of a Gaussian distribution and a short-lived discreet component, with distributions illustrated in Figure 2.7 and data in Figure 2.8. The FWHM for GSTA1-1 was found to be 2.99 ns, while the recovered FWHM for GSTA4-4 was 0.63 ns. These results suggest a significant increase in the conformational heterogeneity even at sites removed from the hydrophobic substrate binding pocket.

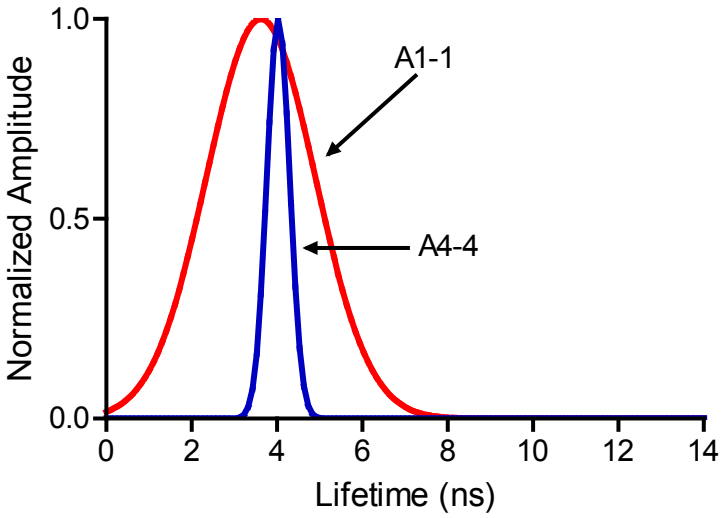


Figure 2.7 Fluorescence lifetime distributions of GSTA1-1 and GSTA4-4 wild-type. Results from distribution analysis using data obtained from the fluorescence of the native Trp21. GSTA1-1 shown in *red*, and GSTA4-4 in *blue*. Data used for these analyses are in Figure 2.8.

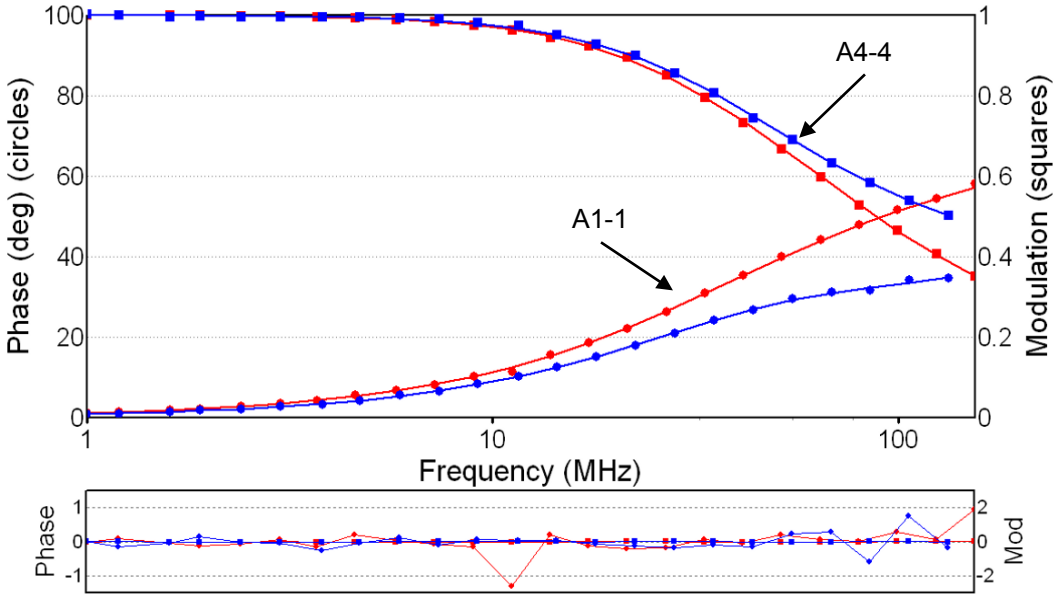


Figure 2.8 Time-resolved fluorescence data for GSTA1-1 and GSTA4-4 wild-type. Top panel: Phase (*circles*) and modulation (*squares*) data are shown, with fits indicated by lines. GSTA1-1 is shown in *red* and GSTA4-4 in *blue*. Bottom panel: Residuals obtained after best fits. Data also appears in reference 15.

Similarly, fluorescence lifetime data were obtained for the exchange mutants, GSTA1-1 V111F/Y217R and GSTA4-4 F111V/R217Y (Figures 2.9 and 2.10). Interestingly, the widths, 1.08 ns for GSTA1-1 V111F/Y217R and 0.81 ns for GSTA4-4 F111V/R217Y, were precisely between those recovered for the wild-type enzymes of both isoforms, suggesting that the Phe111/Tyr217 interaction of GSTA4-4 contributes to the overall C-terminal and tower region packing. A summary of these results in presented in Table 2.1. Collectively, it appears that the conformational heterogeneity of GSTA1-1 describes a greater breadth of structural substates, particularly in the H-site region, dominated by C-terminal helix dynamics, and that the edge-on-face aromatic interaction between the phenylalanine and tyrosine residues, at position 111 and 217 respectively, influence global dynamics in addition to local motions involving the C-terminal helix.

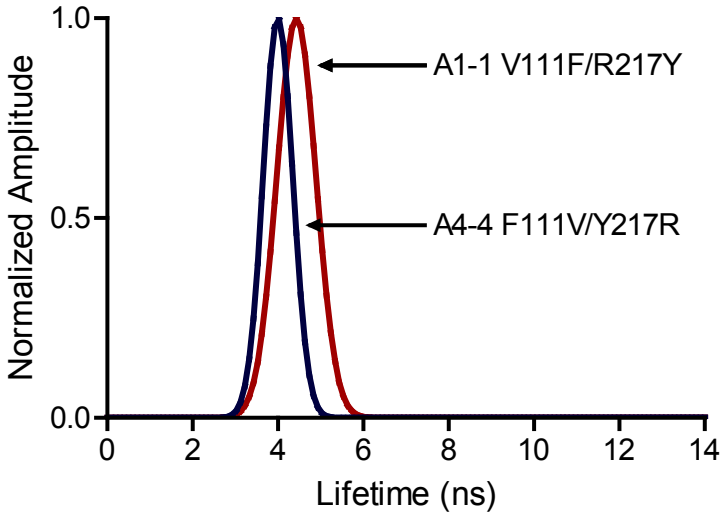


Figure 2.9 Fluorescence lifetime distributions of GSTA1-1 and GSTA4-4 exchange mutants. Results from distribution analysis using data obtained from the fluorescence of the native Trp21. GSTA1-1 V111F/R217Y shown in *maroon*, and GSTA4-4 F111V/Y217R in *navy*. Data used for these analyses are in Figure 2.10.

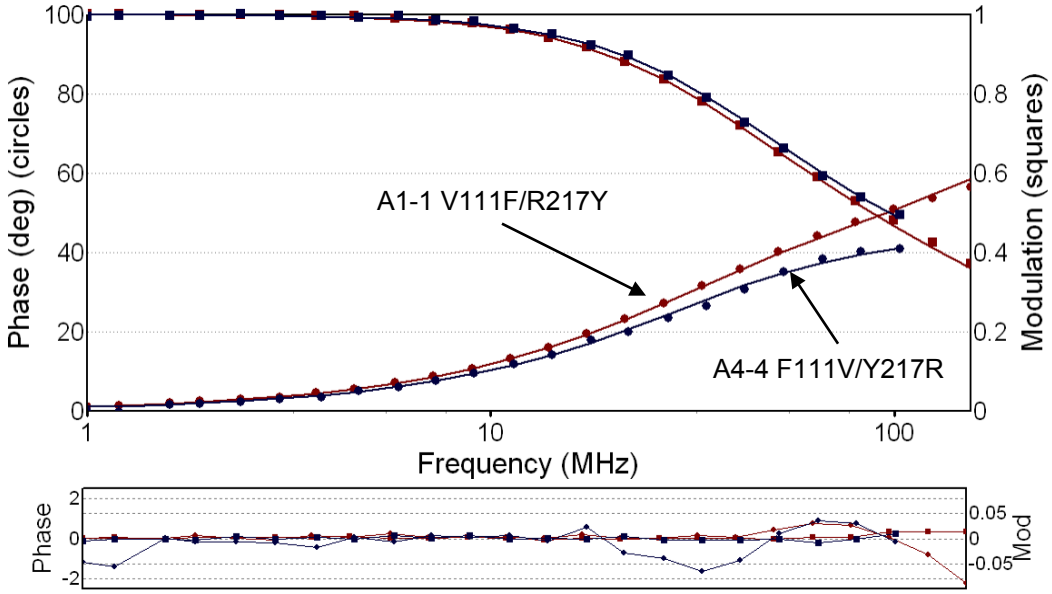


Figure 2.10 Time-resolved fluorescence data for GSTA1-1 and GSTA4-4 exchange mutants. Top panel: Phase (*circles*) and modulation (*squares*) data are shown, with fits indicated by lines. GSTA1-1 V111F/R217Y is shown in *maroon* and GSTA4-4 F111V/Y217R in *navy*. Bottom panel: Residuals obtained after best fits. Data also appears in reference 15.

Table 2.1 Excited state fluorescence decay parameters. Upper limits for the S.E. values are ±0.003 for SAS with wild-type (WT) proteins and ±0.01 for W21F/F222W; for center parameters, the upper limit of S.E. is ±0.025 for WT, and ±0.176 for W21F/F222W. Data also appears in reference 15.

Parameter	GSTA1-1 WT	GSTA4-4 WT	GSTA1-1 V111F/R217Y	GSTA4-4 F111V/Y217R	GSTA1-1 W21F/F222W (apo)	GSTA1-1 W21F/F222W +*S*-hexyl-GSH
SAS1[a]	0.881	0.593	0.694	0.700	0.69	0.81
Center[b] (ns)	3.621	4.032	4.431	4.00	6.49	10.09
Width[c] (ns)	1.271	0.267	0.457	0.346	2.016	0.001
SAS2[d]	0.119	0.407	0.306	0.300	0.31	0.19
Lifetime[e] (ns)	0.539	0.410	1.045	0.471	0.945	1.662
χ^2	1.880	0.272	0.733	3.407	1.679	7.769

[a]The fractional intensity of the distributed component of the fluorescent lifetime
[b]The median value of the distributions
[c]The variance
[d]The fractional intensity of the discreet component
[e]For the discreet component

As illustrated in Figure 2.11, the C-terminal helix portion of GSTA1-1 represents the most mobile portion of this enzyme; however it appears that even core residues retain significant motion when compared to the more specific GSTA4-4. The pair of isozymes, GSTA1-1 and GSTA4-4, comprise an exemplary and tunable platform for the investigation of the basis of high catalytic promiscuity versus specificity within a common fold.

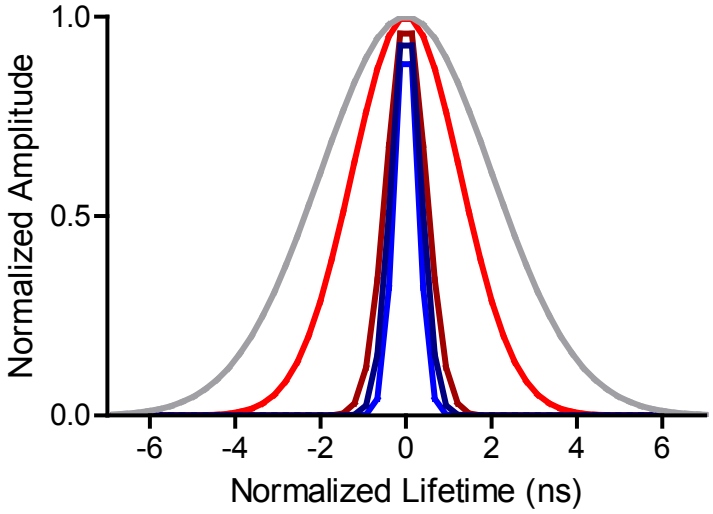


Figure 2.11 Fluorescence lifetime distribution summary. Lifetime distributions are displayed normalized about the zero point. GSTA1-1 W21F/F222W, *grey*; GSTA1-1 WT, *red*; GSTA1-1 V111F/R217Y, *maroon*; GSTA4-4 F111V/Y217R, *navy*; GSTA4-4 WT, *blue*.

These fluorescence lifetime data indicate a progression in protein dynamics and conformational heterogeneity at the local residue level, and on the nanosecond timescale. The very broad distribution noted for the C-terminal helix of GSTA1-1 is an additional indicator of the broad ensemble of substates already established through alternate methods, and serves as a validation for this type of analysis. Surprisingly, the data collected from the time-resolved fluorescence emission characteristics of the core Trp21 also revealed an increase in the breadth of accessible microconformational states which mirrors the general progression in catalytic character from promiscuous GSTA1-1 wild-type, through the exchange mutants to specific GSTA4-4 wild-type.

Chapter 3: Calorimetric Analysis of Glutathione Transferases

A. Introduction

In general, functional promiscuity has been correlated with increased structural flexibility of enzymes (*15, 164-166*). There is also an increasing trend to consider enzymes as conformational ensembles with complex free energy landscapes analogous to protein folding landscapes (*33, 35, 36, 167*). However, the thermodynamic basis for the conformational landscapes that foster promiscuity remains unexplored. Hypothetically, there are significant differences in the conformational dynamics and folding ensembles of substrate specific enzymes compared to catalytically selective enzymes. We return again to catalytically promiscuous GSTA1-1 and substrate selective GSTA4-4 for examination of the folding landscape as it relates to promiscuity. GSTA1-1 is considered to be much more dynamic than GSTA4-4, principally around the active site. (*15, 132, 134, 168*). As illustrated in Figure 3.1, the crystallographic B-factors for the C-terminus and α4-α5 helix-turn-helix tower region are higher than the rest of the enzyme, and they are higher than the B-factors for the corresponding regions in GSTA4-4. Other experimental methods, detailed above, have provided further evidence of this difference in local dynamic character.

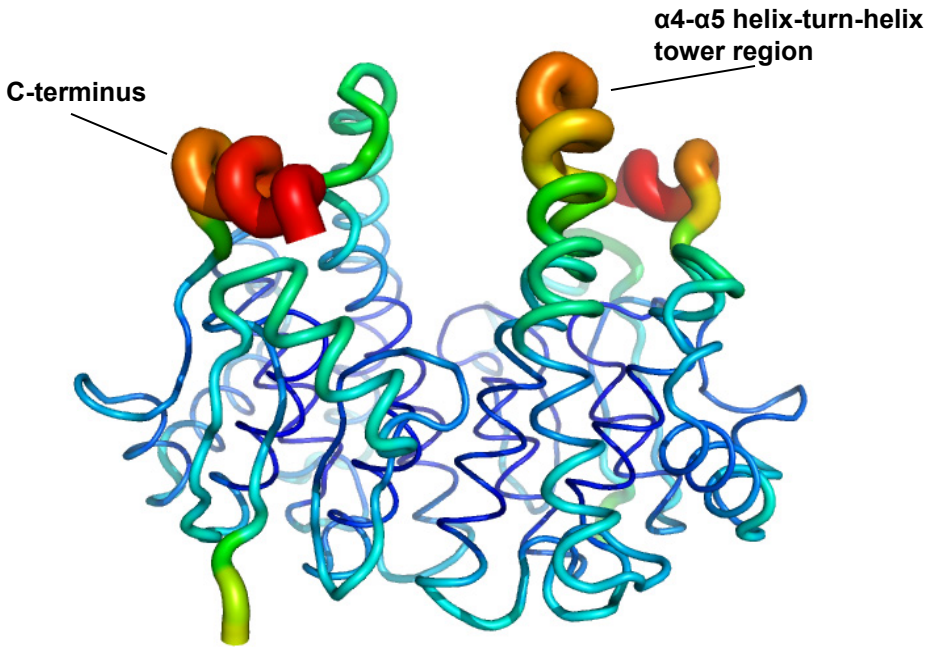


Figure 3.1 Apo-GSTA1-1 demonstrating C-terminus and tower region crystallographic B-factors. Spectrum shift to red and increasing radius of representation indicate higher crystallographic temperature factors. PDB:1PKY

Differential scanning calorimetry, commonly used to examine protein folding landscapes, was used to provide a thermodynamic characterization of the conformational landscapes of GSTA1-1 and GSTA4-4. In addition, the GSTA1-1 V111F/Y217R and GIMFhelix mutants, as catalytic intermediates between the wild-type enzymes, were also investigated.

B. Differential Scanning Calorimetry and the Variable-barrier Model

Differential scanning calorimetry (DSC) measures the apparent molar heat capacity of a macromolecule as a function of temperature. The standard definition of heat capacity is the temperature derivative of the enthalpy (or other energy) function, or simply, a measure of a system's capacity to store energy. Generally, the information obtained using this technique includes the absolute partial heat capacity of the system under investigation and the thermodynamic parameters associated with a temperature induced transition (*169, 170*). Thus,

the enthalpy change (ΔH), entropy change (ΔS), and change in heat capacity (ΔC_p), as well as the transition midpoint, T_m are all accessible in a typical DSC experiment. Furthermore, once these basic parameters are evaluated, other thermodynamic quantities, such as the change in Gibbs free energy (ΔG), and the equilibrium constant, K_{eq}, for the process under study, can be determined. Given a simple two-state system at equilibrium, these quantities can be evaluated from the excess heat capacity as a function of temperature (T) at constant pressure:

$$\Delta H = \int \Delta C_p dT \qquad \text{(Eq. 3.1)}$$

and

$$\Delta S = \int \frac{\Delta C_p}{T} dT \qquad \text{(Eq. 3.2)}$$

The equilibrium constant for the process is related to the van't Hoff enthalpy (ΔH_{vH}) through

$$\frac{\delta \ln K_{eq}}{\delta T} = \frac{\Delta H_{vH}}{RT^2} \qquad \text{(Eq. 3.3)}$$

where R is the gas constant and the van't Hoff enthalpy is defined by the excess heat capacity at the transition maximum ($C_{p(m)}$) such that

$$\Delta H_{vH} = 4RT_m^2 \frac{C_{p(m)}}{\Delta H} \qquad \text{(Eq. 3.4)}$$

additionally, at the T_m

$$\Delta G = \Delta H - T\Delta S = 0 \qquad \text{(Eq. 3.5)}$$

Conventionally used in protein science to examine folding/unfolding behavior, the main observable, excess heat capacity, is a direct summation of the changes in the heat capacity of protein conformations and related solvent effects as a function of temperature and includes both enthalpic and entropic terms (*171-173*). Contributions to the heat capacity changes observed over the course of a DSC scan vary in magnitude and sign depending on the system (protein +

solvent) under study, as can their relative contributions to the heat capacity of the protein itself versus the entropy of the system (*173*). Among the possible sources of heat capacity changes are: hydrophobic effects, hydrogen bonding, electrostatics, vibrational terms, changes in the equilibrium of states, polar hydration effects, and protein conformational entropy (*172, 174*), and all of these terms contribute to the apparent excess heat capacity observed over the course of an unfolding transition in varying degrees. In practical terms, DSC is often used to evaluate the thermal stability of a protein, usually in conjunction with changes in the solvent environment (pH, salt concentration, denaturant concentration). Often, only the change in T_m is of interest, although the changes in enthalpy and entropy are frequently evaluated. Protein-protein interactions and ligand binding thermodynamics are also possible to investigate using this technique (*171, 172*). More recent advances have also allowed use of DSC as a tool for exploring kinetic stability in the folding/unfolding regimes of many proteins (*175*). Extensive investigations into the unfolding of lysozyme using DSC and other methods, have resulted in a through thermodynamic characterization of the unfolding process for this enzyme (*176-178*). At pH 7 and 25°C, the ΔH for unfolding of lysozyme was found to be 56.4 kcal mol^{-1} and ΔC_p between folded and unfolded states was 1560 cal K^{-1} mol^{-1}, while ΔS and ΔG were determined at 140 cal K^{-1} mol^{-1} and 14.5 cal K^{-1} mol^{-1} respectively. DSC has also been used to study intermediate states along the reaction coordinate from native to fully unfolded (*179-181*). Apo-α-lactalbumin presents a stable intermediate molten globule state, exhibiting a high amount of native-like secondary structure with few consistent tertiary interactions (*179*). A multidimensional analysis evaluating the temperature induced unfolding of this protein at varying denaturant conditions demonstrated that the enthalpy change at 25°C between the native and molten globule forms was 7.7 kcal mol^{-1}, while a value of 31.8 kcal mol^{-1} was obtained for

the transition between native and fully unfolded. A similarly large divergence between the native to molten globule and native to unfolded states was noted in the entropy and heat capacity changes, resulting in a calculated loss of ~19% of hydrogen bonds between native and molten globule. Theoretically, a large gain in conformational entropy is responsible for the stability of the intermediate state, despite loss of tertiary structure, which has implications for the evaluation of conformational landscapes using DSC.

Differential scanning calorimetry also allows direct access to the canonical thermodynamic partition function (*169, 170, 182, 183*). If a thermodynamic system is described by a set of discreet macrostates, then the partition function Q can be expressed as the summation over the entropic (ΔS_i) and enthalpic (ΔH_i) terms from each state by Equation 3.6, where R and T are the gas constant and temperature, respectively.

$$Q = \sum_{i=0}^{n} e^{\left(\frac{\Delta S_i}{R}\right)} \cdot e^{\left(-\frac{\Delta H_i}{RT}\right)} \qquad \text{(Eq. 3.6)}$$

Evaluation of the partition function allows projection of the high dimension protein energy landscape onto a physically relevant free energy surface (*175, 184*). In principle, if enthalpy is used as a measure of the folding progress, the free energy surface of protein folding can be derived immediately from the DSC data using an inverse Laplace transform, however this is notoriously difficult to achieve using actual experimental data obtained from protein unfolding studies (*185, 186*). Muñoz and Sanchez-Ruiz have developed a phenomenological model based on a variable free energy barrier to unfolding (*186*). This method is based on the classical Landau model of critical transitions (*187*). Employing this variable-barrier analysis allows a calorimetric transition to be described in terms of a continuous distribution of enthalpic microstates, where Q is the continuous distribution of microstates, H is the enthalpy scale, $\rho(H)$ is the density of enthalpic microstates, such that

$$Q = \int \rho(H) \cdot e^{\left(-\frac{H}{RT}\right)} \cdot dH \qquad \text{(Eq. 3.7)}$$

Here, these microstates are assigned a constant energy value, and the other thermodynamic parameters including entropy and heat capacity arise from the probability distribution of enthalpic ensemble of microstates. The probability of finding a protein in enthalpic microstate *H* at a given temperature *T* is determined by the probability density at that temperature, shown by

$$P(H \mid T) = \frac{1}{Q} \rho(H) \cdot e^{\left(-\frac{H}{RT}\right)} \qquad \text{(Eq. 3.8)}$$

The relationship between the probability at temperature *T* and the 'characteristic' temperature, T_0, is found through

$$P(H \mid T) = C \cdot P(H \mid T_0) \cdot e^{-\lambda H} \qquad \text{(Eq. 3.9)}$$

where C is a normalization constant found by setting the integral of P(H|T) equal to 1 and λ is the reduced temperature, Equation 3.10.

$$\lambda = \frac{1}{R}\left\{\frac{1}{T} - \frac{1}{T_0}\right\} \qquad \text{(Eq. 3.10)}$$

$P(H|T_0)$ is the probability density at the T_0, determined by

$$P(H \mid T_0) = C' \cdot e^{\left(-\frac{G_0(H)}{RT_0}\right)} \qquad \text{(Eq. 3.11)}$$

where *C'* is a normalization constant similar to *C* and the free energy function, $G_0(H)$, is based on the Landau theory of critical transitions, and expanded as a power series of *H*, truncated at the quartic term, as shown by

$$G_0(H) = -2\beta \cdot \left(\frac{H}{\alpha}\right)^2 + |\beta| \cdot \left(\frac{H}{\alpha}\right)^4 \qquad \text{(Eq. 3.12)}$$

This yields a description of the distribution of states and hence the degree of 'roughness' of the landscape, where parameters α and β are the critical coefficients and enthalpy *(H)* is the order parameter. When β>0, it corresponds to the energy barrier separating two enthalpic macrostates, and 2α approximates the enthalpy difference between macrostates. When β is negative, there is a single macrostate in a smooth energy well, and α and β simply describe the shape of the free energy function. At this point the first and second enthalpy moments, H^1 and H^2, can be calculated by

$$\left\langle H^n \right\rangle = \int H^n P\left(H \mid T\right) dH \qquad \text{(Eq. 3.13)}$$

and compared to DSC data in terms of the excess heat capacity with respect to the native state (C_p^{EX}) through Equation 3.14 in a least-squares fitting routine.

$$C_p^{EX} = \frac{d\left\langle H \right\rangle}{dT} = \frac{\left\langle H^2 \right\rangle - \left\langle H \right\rangle^2}{RT^2} \qquad \text{(Eq. 3.14)}$$

In order to account for the asymmetry of unfolding, where the native and the unfolded have differing heat capacities, one value of α is used for positive enthalpy values (α_N) and another for negative enthalpy values (α_P) when integrating, given by Equations 3.15-3.17

$$\alpha_N + \alpha_P = \sum \alpha \qquad \text{(Eq. 3.15)}$$

$$\alpha_N = \sum \alpha \cdot f/2 \qquad \text{(Eq. 3.16)}$$

$$\alpha_P = \sum \alpha \cdot \left(2 - f\right)/2 \qquad \text{(Eq. 3.17)}$$

where f is the asymmetry factor ($0 > f > 1$). The fitting parameters used in this model are the barrier height, β, the enthalpy order parameter expressed as the sum over positive and negative enthalpy terms, $\Sigma\alpha$, the characteristic temperature, T_0, and the asymmetry factor, f.

This model is purely phenomenological, and is valid for first-order or continuous transitions. The transition order is determined by the magnitude of the barrier height (β), such

that a negative or small (~2.5*RT*) barrier indicates a continuous, or downhill folding mechanism, while the presence of an energetic barrier to folding describes general two-state-like behavior, which does not necessarily have to conform precisely to two-state ideality. This method has been used to examine unfolding barrier heights (or lack thereof) for several proteins (*186, 188, 189*). One well-studied example of a downhill folder is the peripheral subunit binding domain (PSBD) of dihydrolipoamide succinyl transferase from the 2-oxo-glutarate dehydrogenase complex of *E. coli* (*186, 190, 191*). This domain was produced via peptide synthesis to generate a small (~40 residues) protein comprising two parallel helices connected by a long partially structured loop (*190, 192*), termed BBL. The equilibrium unfolding transition of BBL was previously examined through numerous techniques (*190*) indicating a downhill folding regime for this protein. Application of the variable-barrier model to this protein agreed with previous results, with fits to DSC data proving to be barrierless, and as a positive control, thioredoxin exhibited a significant barrier height of ~40 kJ mol^{-1} (*186*). Based on this analysis established by others, DSC was used to examine thermally-induced conformational changes in GSTA1-1 and GSTA4-4, and well as the GST mutants exhibiting intermediate catalytic character between the two wild-type enzymes.

C. Materials and Methods

Protein Expression and Purification

hGSTA1-1, hGSTA4-4, hGSTA1-1 V111F/R217Y, hGSTA1-1 W21F/F222W, and hGSTA4-4 F111V/Y217R were expressed and purified as described previously (Section 2C). Additionally, the hGST GIMFhelix mutant enzyme was generated from a previously subcloned pET-21a(+) (Novagen) vector (*135*), and expressed in *E. coli* BL-21 cell cultures, and purified in the same manner as GSTA4-4 wild-type. Protein concentration was determined using a molar

extinction coefficient of 22,800 M^{-1} cm^{-1} on a monomer basis at 276 nm. The rat GSTA1-1 (rGSTA1-1) Δ209-222 C-terminal truncation mutant was previously subcloned into the pKKGTB plasmid, and expressed using the same protocol as GSTA1-1 wild-type (*139*). Protein concentrations for this enzyme were determined using the Thermo Scientific Pierce BCA Protein Assay Kit (Thermo Fisher Scientific Inc., Rockford, IL).

Differential Scanning Calorimetry (DSC)

Thermal denaturation experiments were conducted on VP-CAP DSC microcalorimeter (MicroCal, Northamption, MA). In all cases, protein samples were dialyzed against 200x v/v 100 mM potassium phosphate buffer, pH 6.5 at 4°C and the dialysate used for buffer reference scans, and loaded into the reference cell for protein scans. The instrument was thermally pre-equilibrated with 12 buffer scans, or until no change was noted in instrument response, before protein scans were obtained. Thermograms were collected at scan rates of 0.5 °C/min, 1°C/min, and 2 °C/min; all data shown were collected at 2 °C/min, as no dependence on scan rate was noted. Unless otherwise noted, DSC scans were run from 10 °C to 80 °C, with a 10 minute equilibration at 10 °C. Reversibility of the transitions was evaluated by scanning to just above the highest thermal transition midpoint, cooling to 10 °C and rescanning. The data were initially analyzed using the software supplied with the instrument, Origin 7.0 (OriginLab, Northampton, MA). Buffer reference scans were subtracted, and the data normalized to dimer concentration before further analysis. Baselines for traditional analysis were calculated though the cubic option in Origin 7.0, while baselines were extrapolated from the slope of the pre-transitional baseline for variable-barrier analysis. Origin 7.0 was also used to deconvolute the complex DSC transitions prior to variable-barrier analysis, as noted in the text. The variable-barrier model presented above was used to fit DSC data using a custom Octave (www.octave.org) script

(Appendix A) employing a Levenberg-Marquardt minimization algorithm. The least-squares fitting parameters used were T_0, β, $\Sigma\alpha$, and f. Goodness of fit was determined by R^2 values and confidence interval analysis.

Circular Dichroism (CD)

Spectra were obtained using a Jasco 720 CD spectropolrimeter (Easton, MD). The samples used for data collection were dialyzed against 200x v/v 20 mM potassium phosphate, pH 6.5 at 4 °C, and the dialysate used to obtain solvent reference scans. Far-UV CD spectra were collected using 10 μM protein, and 50 μM protein used for near-UV CD spectra. 10 mm and 1 mm cuvettes were used for far-UV and near-UV data collection, respectively. Averaged CD signals, corrected for solvent and FFT smoothed, were converted to mean residue molar ellipticity [Θ].

D. Results and Discussion

Differential Scanning Calorimetry

Extending the investigation describe in Chapter 2, where it is hypothesized that conformational heterogeneity contributes to catalytic promiscuity, the thermodynamic characteristics of GSTs were examined. Differential scanning calorimetry (DSC) data were obtained for the GSTA1-1 and GSTA4-4 wild-type enzymes, in addition to the GSTA1-1 V111F/R217Y, GSTA1-1 W21F/F222W, and GIMFhelix variants. The DSC data collected for GSTs demonstrated high reversibility to nearly 60 °C in repetitive scans, although the complete scans to 80 °C were irreversible, precluding a specific interpretation of enthalpy values in the traditional manner due to distortion of the post-transitional baseline. However, no dependence on scan rate was noted, indicating equilibrium conditions at each incremental temperature step in the process (*193*). However, some general characteristics and comparisons of the DSC scans can be explored, prior

to a more involved investigation and analysis using the variable-barrier analysis model detailed above. The DSC thermogram of apo-GSTA1-1 exhibited a significant change in excess C_p at temperatures as low as 30 °C, and much lower than the main unfolding transition, as seen in Figure 3.2 A.

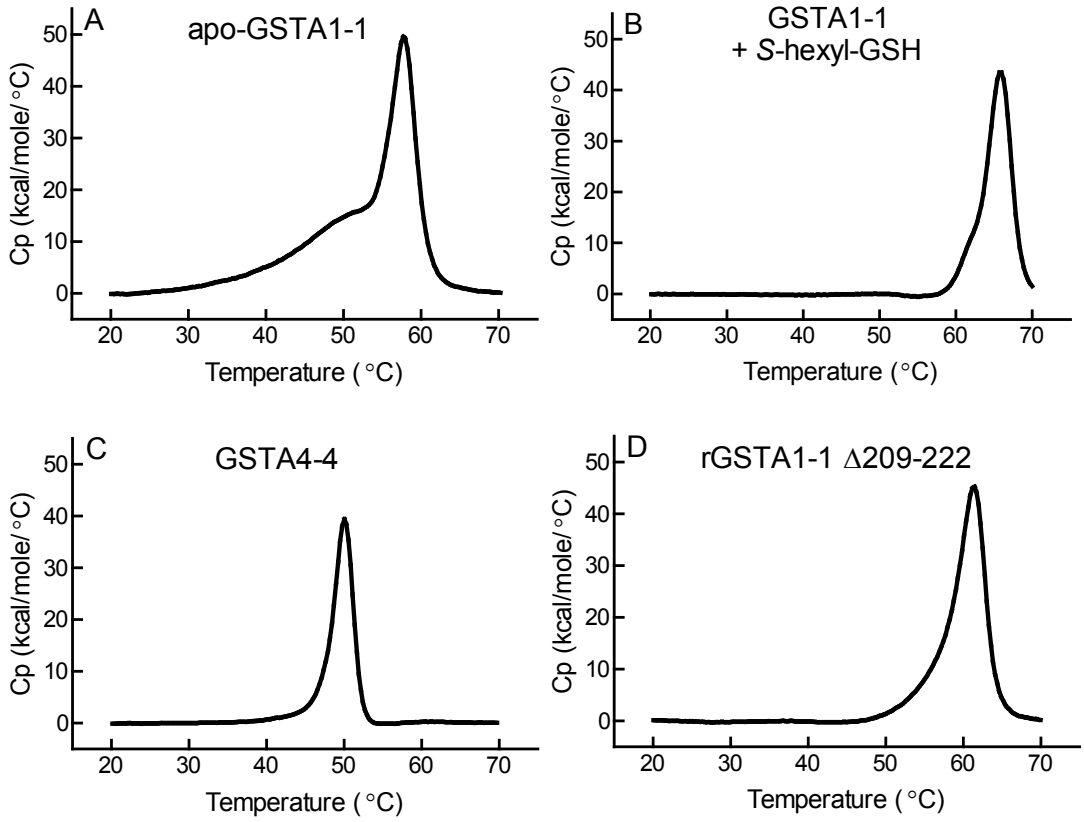


Figure 3.2 DSC thermograms of GSTs. Shown are A) GSTA1-1 wt. B) GSTA1-1 wt with 2mM *S*-hexyl-GSH. C) GSTA4-4 wt. D) rGSTA1-1 Δ209222, with C-terminus truncated. Data are represented after protein dimer concentration normalization, buffer scan subtraction, and baseline subtraction using the cubic option in Origin 7.0. Data also appears in reference 66.

The peak maximum at 58 °C, observed for the main unfolding transition, lies between the T_m determined from the temperature dependence of catalytic function of 56 °C and the structural unfolding T_m based on circular dichroism data (*163*). Deconvolution of complex DSC traces displaying similar low-temperature characteristics have been used to resolve transitions for

separate structural domains, as noted for the thermal unfolding of phenylalanine hydroxylase, where the low temperature transition was attributed to the unfolding of the N-terminal regulatory domain, followed by the remainder of the protein (*194*). Additionally, the DSC curve for the λ repressor of *E. coli* exhibits two separated peaks, with the low temperature transition due to denaturation of the N-terminal portion (*195*). The C-terminal helix of GSTA1-1 is likewise known to behave in a manner independent from the remainder of the protein, and it is likely that the low temperature transition observed for this enzyme corresponds to changes in this structural element. However, previous investigations into the thermal melting behavior of GSTA1-1 have shown a notable retention of helical structure in this same temperature regime (*162, 163*). These results suggest that the lower temperature transition of GSTA1-1 not related to unfolding *per se*, but may arise from temperature dependent sampling of different locations by the C-terminal helix or restructuring of tertiary contacts, further indicating a remarkable conformational heterogeneity and a large conformational entropy contribution to the excess heat capacity. Moreover, the addition of saturating concentrations of *S*-hexyl-GSH, known to cause the carboxy terminus to adopt a well-defined conformation (*83, 136, 138*), eliminated the low temperature asymmetry features from the thermogram (Figure 3.2 B). The DSC trace retained only the main unfolding transition, shifting the main peak maximum to 66 °C. Additionally, data were collected using a GSTA1-1 variant from rat for which the α9 helix has been truncated at residue 209 (rGSTA1-1 Δ209-222), as demonstrated in Figure 3.2 D. The prominent low temperature transition noted for wild-type hGSTA1-1 was conspicuously absent in the DSC thermogram obtained for this variant, with a slight increase in little impact on thermal stability, shifting the peak maximum to 61 °C. GSTA4-4, which has a static C-terminus localized along the H-site portion of the active site, was also examined via DSC. In marked contrast to GSTA1-1, GSTA4-

4 did not display a low temperature transition, but yielded a sharp unfolding peak, indicating a much more cooperative unfolding scheme than GSTA1-1 (Figure 3.2 C). Surprisingly, this isoform is significantly less thermally stable than the GSTA1-1 isoform, with a peak maximum at 50 °C. The striking absence of a low temperature asymmetry in the GSTA4-4 trace further suggests that the low temperature transition of GSTA1-1 is related to that enzyme's conformational heterogeneity around the active site. Thermograms were similarly collected for GSTA1-1 V111F/R217Y (Figure 3.3 B) and GIMFhelix variants (Figure 3.3 C). In these traces, the observed excess heat capacity prior to the main unfolding transition decreased following the same general trend as the catalytic character progression from highly promiscuous GSTA1-1 to substrate specific GSTA4-4. Interestingly, all the enzymes that retained the majority of the GSTA1-1 core structure were significantly more stable than GSTA4-4, with main transition peak maxima of 58 °C and 57 °C for GSTA1-1 V111F/R217Y and GIMFhelix respectively, perhaps providing a stable platform for conformational heterogeneity around the active site. DSC data were also collected for GSTA1-1 W21F/F222W (Figure 3.3 E). The thermogram for this enzyme displayed a very similar profile to wild-type GSTA1-1, with the same low temperature asymmetry feature, however the observed excess heat capacity was demonstrably lower throughout the temperature range. Previous results have indicated that replacement of the native Trp21 with phenylalanine results in a slightly less thermally stable protein with enhanced solvent accessibility at the intradomain interface, and the possible presence of a fast unfolding intermediate (*163*). In addition, the bulkier indole side chain may reduce the conformational entropy at the C-terminus (*196*).

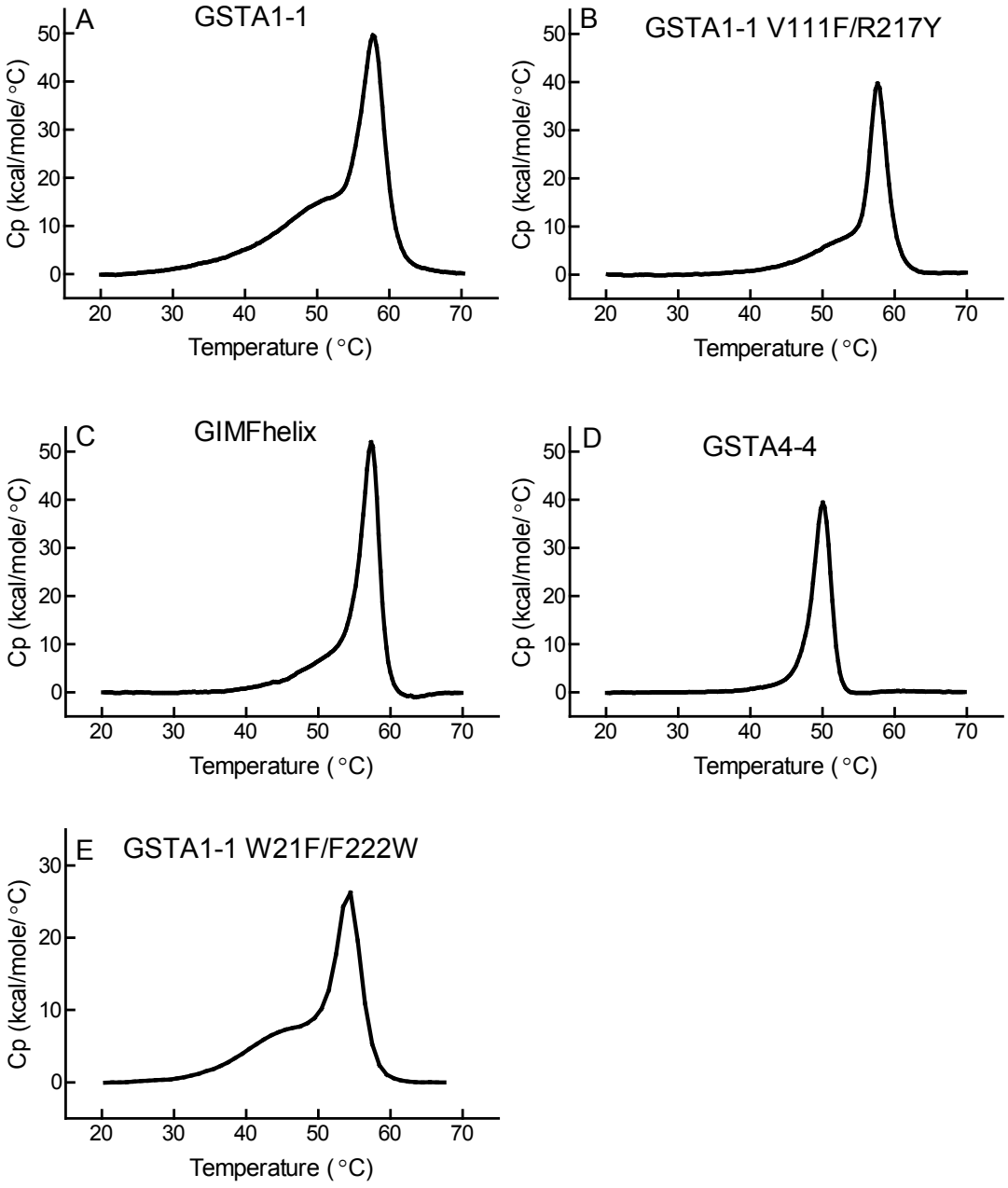


Figure 3.3 DSC thermograms of GSTs. Shown are A) GSTA1-1. B) GSTA1-1 V111F/Y217R. C) GIMFhelix. D) GSTA4-4. E) GSTA1-1 W21F/F222W. Please note the difference in y-axis scale. Data are represented as curves after protein dimer concentration normalization, buffer scan subtraction, and baseline subtraction using the cubic option in Origin 7.0.

Circular Dichroism

In order to further explore the possibility that the low temperature DSC transition of GSTA1-1 does not result from unwinding of the C-terminal helix, circular dichroism (CD) studies were also completed. Spectra collected for GSTA1-1 in the far-UV region revealed ellipticity minima at 208 and 222 nm, indicative of a high degree of α-helix character (Figure 3.4 A). When examined as a function of temperature, there is no apparent change in the ellipticity before 50 °C for GSTA1-1, providing a strong indication that the C-terminal helix does not unwind in the temperature region covered by the first DSC transition. More interesting are the CD data collected for GSTA1-1 in the near-UV region (Figure 3.5). The peak at 292 nm in these spectra is characteristic of the presence of a tryptophan exciton couplet. Several aromatic residues, Phe197, Tyr165, and Tyr166, are within 6 Å of Trp21 and are candidate partners for the exciton pair. Spectra from scans taken from 10 °C to 44 °C exhibit insignificant change in the tertiary packing environment around Trp21, and the fine structure displays a similar lack of change at temperatures prior to 50 °C.

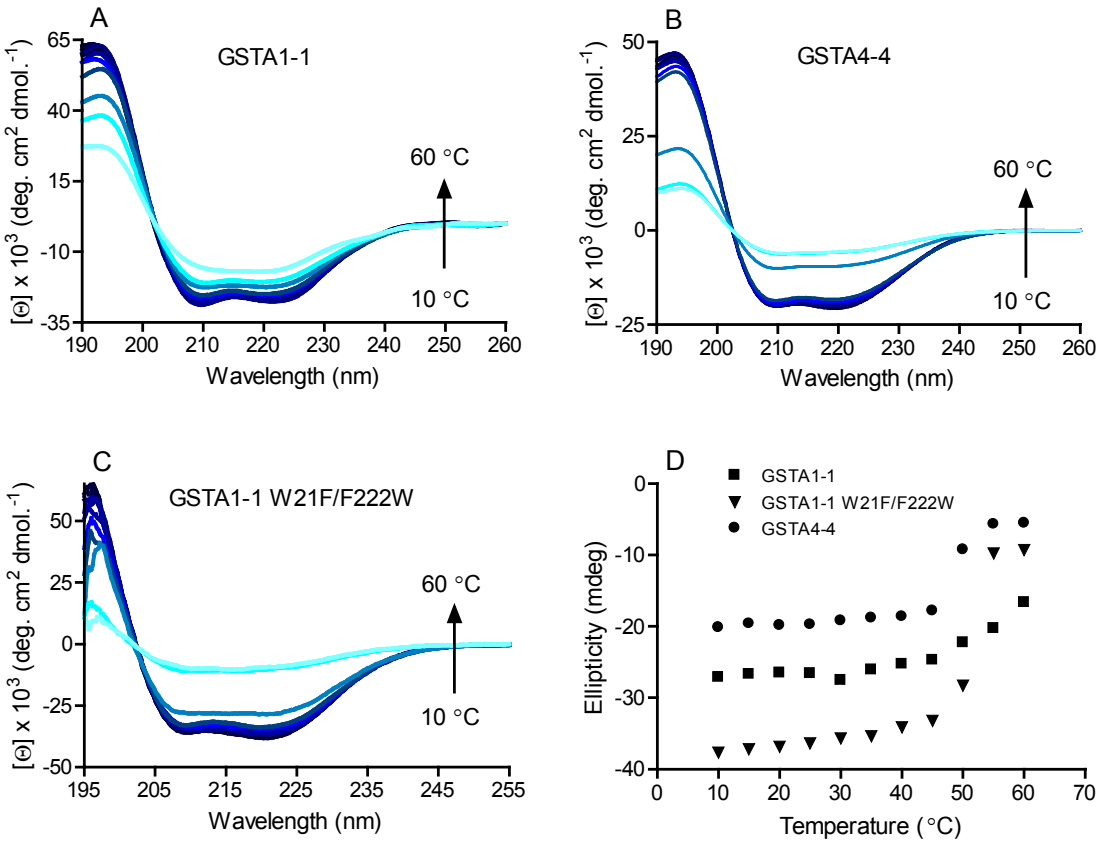


Figure 3.4 CD spectra of GSTs with increasing temperature. A) GSTA1-1 wt. B) GSTA4-4 wt. C) GSTA1-1 W21F/F222W. Panels A-C are in order of increasing temperature, from 10 °C (*navy*) to 60 °C (*cyan*), in 5 °C increments. D) Ellipticity at 222 nm as a function of temperature, GSTA1-1 wt (■), GSTA1-1 W21F/F222W (▼), and GSTA4-4 (●). Data also appears in reference 66.

In further pursuit of information about the nature of the low temperature excess heat capacity observed for GSTA1-1, CD data were collected using the GSTA1-1 W21F/F222W mutant. Spectra collected in the far-UV region were similar to those obtained for wild-type GSTA1-1, although they indicated a reduced resistance to thermal unfolding (Figure 3.4 C). This mutant, where the core Trp21 was replaced with a phenylalanine residue, and Phe222, located at the end of the C terminus, was replaced with a tryptophan, allows more direct evaluation of the C-terminal helix through near-UV CD spectroscopy. The near-UV CD spectra of this mutant exhibit significant divergence from those of the wild-type enzyme (Figure 3.5). The peak

maximum in this region was shifted to 285 nm, clearly indicating a more solvent exposure to the tryptophan in the location as opposed to Trp21 in the core of the protein. Additionally, the signal intensity for this peak decreases drastically starting at 30 °C, in the same temperature range as the early DSC transitions stemming from changes in the tertiary packing environment around Trp222.

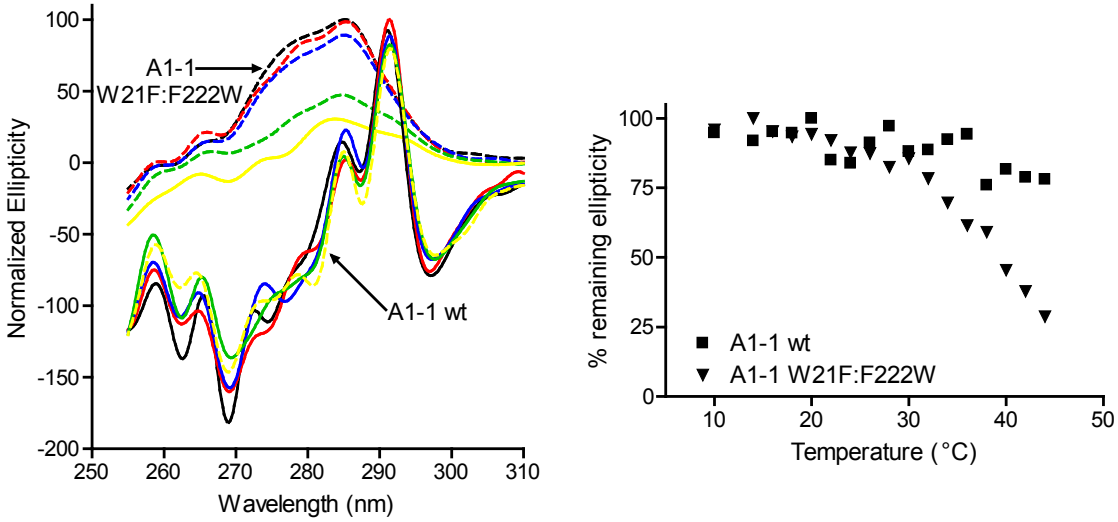


Figure 3.5 Near-UV CD spectra of GSTA1-1 wild-type and GSTA1-1 W21F/F222W. Left panel: GSTA1-1 wt spectra are shown in *solid lines*: 10 °C, *black*; 20 °C, *red*; 30 °C, *blue*; 40 °C, *green*; 44 °C, *yellow*. GSTA1-1 W21F/F222W spectra are represented as *dashed lines*: : 10 °C, *black*; 20 °C, *red*; 30 °C, *blue*; 40 °C, *green*; 44 °C, *yellow*. Right panel: the % remaining ellipticity at peak maximum is shown as a function of temperature. (■) GSTA1-1 wt, (▼) GSTA1-1 W21F/F222W. Data also appears in reference 66.

These results suggest that the low temperature DSC transition asymmetry observed for GSTA1-1 (and the W21F/F222W mutant) are due to repacking of the C-terminal α9 helix, rather than unfolding in the temperature rage below 50 °C. Far-UV CD spectra for GSTA4-4 also exhibited strong α-helix character, with ellipticity minima at 208 and 222 nm, similar to the spectra of GSTA1-1, however GSTA4-4 proved to be more sensitive to temperature induced unfolding than GSTA1-1; the spectra indicated loss of helical structure at temperatures much lower than was noted for GSTA1-1 (Figure 3.4 B).

Variable-barrier Free Energy Analysis

To more thoroughly understand the apparent local molten globule-like behavior around the active site attributed predominantly to the C-terminus, the DSC data were also examined in terms of the variable-barrier model developed by Muñoz and Sanchez-Ruiz (*186, 188*). This model provides a direct analysis of the free energy barrier in the folding landscape, or lack thereof, as detailed in Section 3B. In this analysis, plots of the free energy function versus enthalpy ($G_0(H)$ vs. H) provide one-dimensional representations of the complex high dimension conformational ensemble surface. Additionally, a view of the breadth of accessible microstates arises from the distribution of enthalpic states at a given temperature. In preparation for the application of this analytic method for initial comparison between GSTA1-1 and GSTA4-4, the thermogram obtained for aop-GSTA1-1 wild-type was deconvoluted in to two separate transitions in order to separate the low temperature region from the main unfolding transition (Figure 3.6 A&D). The variable-barrier analysis was performed on this two component system, as opposed to a one component model for the singular transition noted for GSTA4-4. The low temperature transition of GSTA1-1 proved to be barrierless, as indicated by a single smooth well, and the distribution of enthalpic states at each temperature for this transition was unimodal and extremely broad (Figure 3.6 B&C). This agrees with the lack of a temperature-dependent change in NMR spectra indicating a very small enthalpy difference between states (*83*). In contrast, the higher temperature transition for global unfolding demonstrated a significant free energy barrier of 67.8 kJ/mol (Figure 3.6 E). The probability density shifts from a single sharp unimodal distribution centered at low enthalpy at 325 K where the protein is completely folded, to a bimodal distribution at 335 K, and then to a single broad distribution of enthalpic states at 345 K where the population is entirely unfolded. Comparison between the probability distributions and free

energy contours provide further evidence suggesting that the low temperature transition is due to local conformation heterogeneity around the active site, with barrierless transitions between states, while the high temperature transition is associated with unfolding of the core of GSTA1-1.

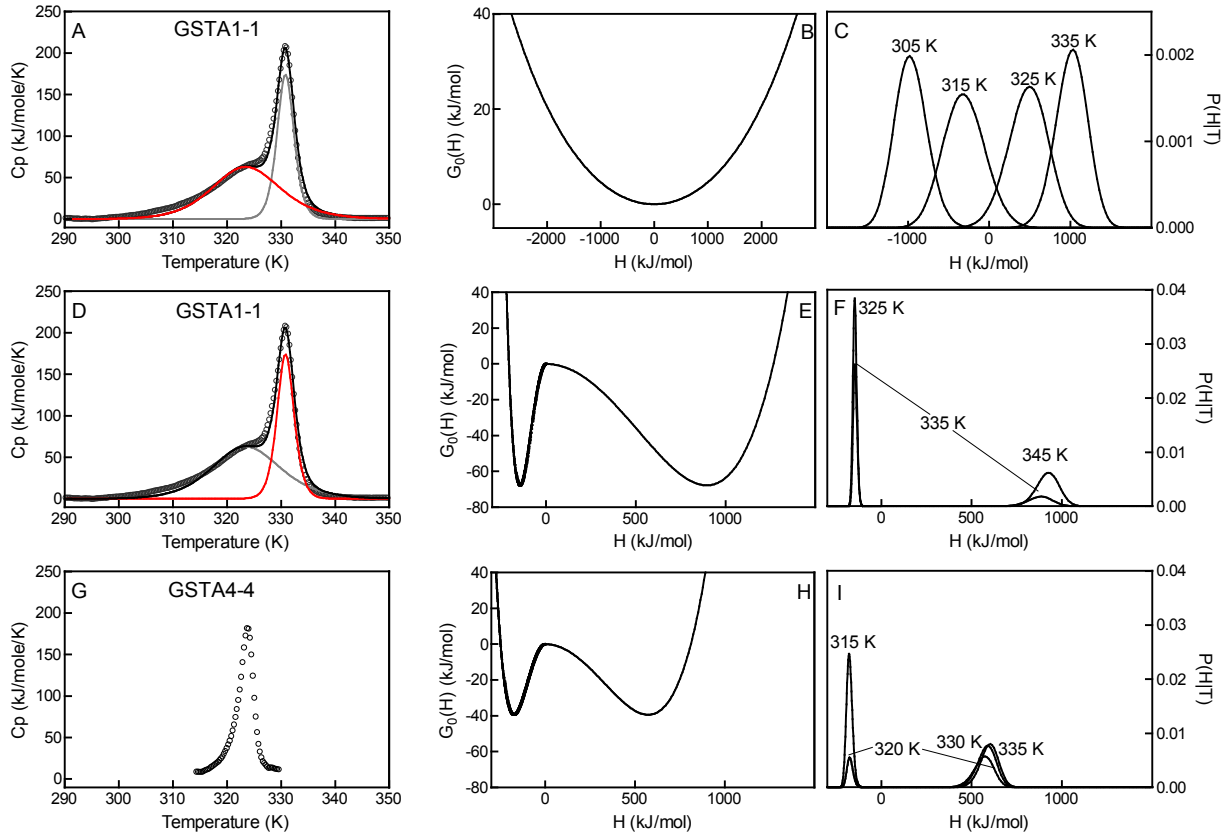


Figure 3.6 DSC deconvolutions (left), free energy profiles (center), and probability distributions of enthalpic states (right) for GSTA1-1 and GSTA4-4. *Top row*: GSTA1-1 low temperature transition DSC data (A), free energy profile (B), and probability distributions (C). *Middle row*: GSTA1-1 high temperature transition data (D), free energy profile (E), and probability distributions (F). *Bottom row*: GSTA4-4 DSC data used for variable-barrier analysis (G), free energy profile (H), and probability distributions (I). The GSTA1-1 thermogram was deconvoluted into a low temperature transition (A, *red line*), and a high temperature transition (D, *red line*). Data points are shown as *open circles*, and the overall fit as *black lines*. The deconvolutions are shown as *red* and *grey lines*. GSTA4-4 DSC data (G) was used without deconvolution for this analysis. Thermograms are shown after buffer scan subtraction, protein dimer concentration normalization, and native-state baseline analysis. The GSTA1-1 high temperature transition parameters recovered are as follows: $\Sigma\alpha$=1042.1 kJ/mol, T_0=337.4 K, β=67.8 kJ/mol, and f=0.28 (R^2=0.9287). The entire data set for GSTA4-4 gave the following: $\Sigma\alpha$=749.6 kJ/mol, T_0=324.9 K, β=39.4 kJ/mol, and f=0.47 (R^2=0.9287). The low temperature transition fit parameters are shown in Table 3.1. Data also appears in reference 66.

When applied to GSTA4-4, the free energy contour exhibits similar characteristics to the one observed for GSTA1-1, except the depth of the free energy wells is lower, reflecting the reduced stability of this isoform, with an energy barrier of 39.4 kJ/mol (Figure 3.6 H). Interestingly, a similar value of 40 kJ/mol was previously recovered for the unfolding of thiorexodin (*186*), which has the same fold as the conserved core domain of GSTs. The distributions recovered for GSTA4-4 were also significantly narrower at all temperatures.

With these results in hand, GSTA1-1 V111F/R217Y, GSTA1-1 W21F/F222W, and GIMFhelix were also investigated. Ideally, we aimed to compare the active site conformational landscape exhibited in the low temperature transition of GSTA1-1 among the mutants and isoforms. For this effort, the DSC thermograms of the mutant enzymes and GSTA4-4 wild-type were deconvoluted using the same protocol that was used for GSTA1-1 wild-type. To be clear, GSTA4-4 and GIMFhelix did not have separable low temperature transitions, and the deconvolution providing the low temperature transitions for the following analysis is for comparative purposes. In effect, the barrier heights and probability distributions recovered are the energy barriers and distributions for these enzymes if the thermodynamic state functions followed the same path of GSTA1-1. With this in mind, the variable-barrier analysis was then used to fit the low temperature transitions recovered. The results acquired from this analysis generally followed the expected trend in barrier height for the deconvoluted low temperature transitions for each enzyme, placing GSTA1-1 V111F/R217Y and GIMFhelix between GSTA1-1 and GSTA4-4 wild-type enzymes with free energy barriers of 7.2 kJ/mol and 101.6 kJ/mol, respectively. The barrier height determined for GSTA4-4, as the least promiscuous enzyme, was 1676.1 kJ/mol. The free energy as a function of enthalpy for the deconvoluted low temperature transitions are shown in Figure 3.7. Surprisingly, the barrier heights for GSTA1-1 W21F/F222W

and GSTA1-1 V111F/R217Y were similar, with the recovered barrier height for GSTA1-1 W21F/F222W at 9.5 kJ/mol.

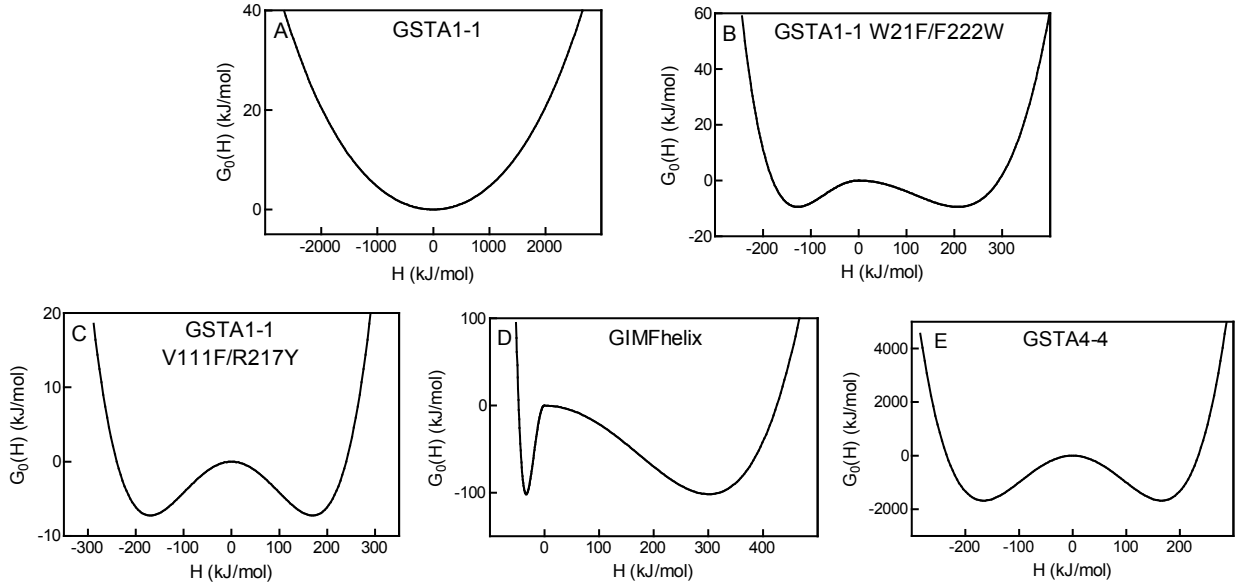


Figure 3.7 Free energy profiles of GSTs. The free energy profiles recovered from variable-barrier analysis of low temperature transitions are shown. A) GSTA1-1 wild-type, B)GSTA1-1 W21F/F222W, C) GSTA1-1 V111F/R217Y, D) GIMFhelix, E) GSTA4-4. Please note the difference in scales on both axes.

This suggests that the wild-type GSTA1-1 enzyme is exquisitely poised for its detoxification function, and even subtle alterations in the enzyme structure can have effects not readily apparent using standard structural analyses. Research by others in the Atkins lab involving pre-steady state kinetics of GSTA1-1 W21F/F222W has indicated that this marginal energy barrier may result in a delay for accessing certain conformational macrostates not evident in the wild-type enzyme (*197*). That is, this mutant exhibits pre-steady state lags in stopped-flow catalytic experiments are due to conformational rearrangement required for catalysis (unpublished data). Some caution should be used in evaluation of these results. The recovered β values are interpreted only as relative barriers for access to the same conformational landscape as observed for the low temperature transition of GSTA1-1. The probability distributions for all of the low

temperature transitions for the GSTs, excepting barrierless GSTA1-1 wild-type were unimodal at temperatures where the proteins were all folded. As the temperature is increased the distributions shifted to bimodal coverage, followed by a unimodal distribution after complete unfolding. The distributions of enthalpic states at 298 K are shown in Figure3.8. GSTA1-1 wild-type exhibited the broadest distribution, followed by GSTA1-1 W21F/F222W and GSTA1-1 V111F/R217Y, which were similar. Both the GIMFhelix and GSTA4-4 demonstrated very narrow distributions at all temperatures.

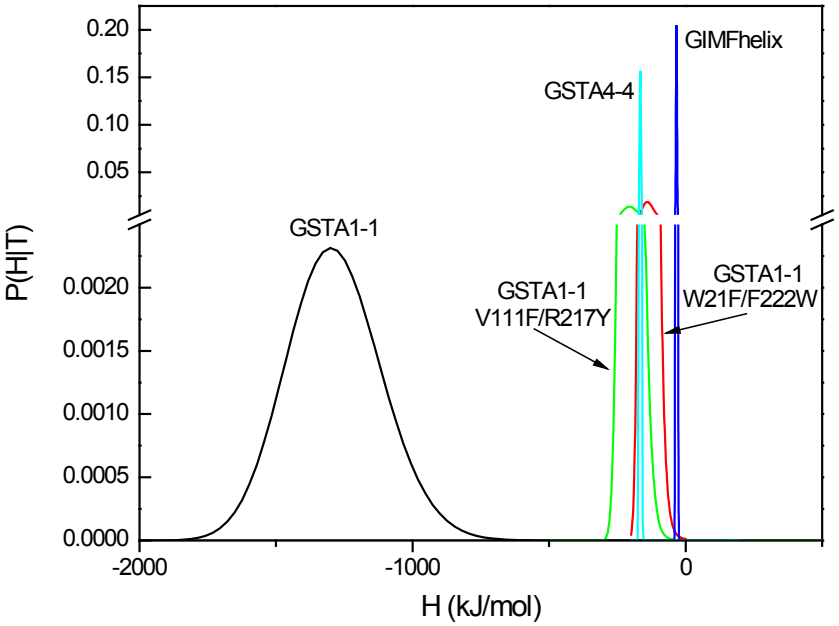


Figure 3.8 Distributions of enthalpic states at 298 K. Probability distributions of enthalpic states at 298 K for the low temperature transitions after deconvolution are shown. GSTA1-1 is in *black*, GSTA1-1 W21F/F222W is in *red*, GSTA1-1 V111F/R217Y is in *green*, GIMFhelix is in *blue*, and GSTA4-4 is in *cyan*.

Table 3.1 Low temperature transition variable-barrier analysis parameters.

Protein	Barrier height (β) (kJ/mol)	Σα (kJ/mol)	T_0 (K)	f	R^2
GSTA1-1	-30	3664.7	318.8	1.0	0.9933
GSTA1-1 W21F/F222W	9.5	333.77	320.5	0.76	0.9925
GSTA1-1 V111F/R217Y	7.2	338.9	327.70	1.0	0.9931
GIMFhelix	101.6	334.3	332.7	0.20	0.9992
GSTA4-4	1676.1	331.8	322.0	1.0	0.7800

The large difference in barrier height between GSTA1-1 and GSTA4-4 has implications for their respective physiological roles. As a detoxification enzyme, GSTA1-1 is able to bind and metabolize a diverse array of substrates. The lack of free energy barriers for remodeling of the molten globule-like active site domain enables significant conformational rearrangements in an efficient manner. In contrast, substrate specific GSTA4-4 must overcome a substantial energetic penalty to achieve the same conformational freedom. An examination of the unfolding landscape using the same variable-barrier method has been previously applied to two protein of similar size and topology, bovine α-lactalbumin (BLA) and hen-egg-white lysozyme (HEWL) (*189*). BLA and HEWL are nearly identical in size and posses the same overall tertiary structure, however BLA readily forms a molten globule under mild denaturing conditions (*198, 199*). Similar to the results obtained for A1-1 and A4-4 above, a high thermodynamic barrier was recovered for HEWL, while a marginal barrier was noted for BLA (*189*). BLA is generally considered to retain a high degree of secondary structure in the unfolded state, and to easily adopt molten globule properties when it binds to membranes (*200*). HEWL functions as a bactericide that needs to work in harsh extracellular conditions, and a high thermodynamic unfolding barrier will increase the half-life of HEWL in its active form (*201*). In addition to being a secreted milk protein, BLA functions inside the cell as a conformationally dynamic substrate-specifier for galactosyltransferase and binds reversibly to membranes by adopting a molten globule-like state (*202*). These results, in addition to the results obtained for A1-1 and A4-4, point to a systematic selection of marginal free energy barriers to conformational dynamism in proteins required by their physiological role to exhibit a high degree of conformational freedom.

Chapter 4: Catalytic Promiscuity of GSTs

A. Introduction

Recently there have been numerous published examples of functional promiscuity among enzymes and proteins (*16, 41, 43, 56, 61, 203*). The functional importance of promiscuity has been outlined above and includes the utility of catalytic diversity in evolution of new function and in enzymatic detoxification. In addition, functional promiscuity may be important in the immune system (*204-206*) and signal transduction (*207-210*). In 1976, Jensen hypothesized that primordial enzymes possessed very broad substrate selectivity profiles to afford a wide range of metabolic capabilities (*211*). Certainly ancient organisms adhering to this hypothesis would find it much easier to diversify into new ecological niches in a gradual manner, able to compete at a low level in a variety of environments. Mutations to the genes encoding versatile enzymatic function can shift the enzyme catalytic profiles to more specific and efficient utilization of available resources, without total loss of original function. Metabolic pathways and catalytic networks can then be integrated with a great deal of redundancy. It appears also, that some classes of enzymes have either kept or reacquired a high level of catalytic promiscuity as necessary to their physiological role. Take, for example, two sequiterpene cyclases from grand fir (*Abies grandis*). Two enzymes, δ-selinene and γ-humulene, synthesis 34 and 52 different sequiterpenes from (*E*,*E*)-farnesyl diphosphate (*212*). This makes efficient use of a single substrate and only two enzymes to provide a broad spectrum defense again insects and herbivores (*41, 213*). Serum albumins, while not primarily enzymes, also catalyze the conversion of several harmful molecules into less toxic metabolites (*41*). Generally, detoxification enzymes are considered to be highly promiscuous regarding both the variety of chemical transformations they can perform and diverse array of substrates (*14, 15, 43, 47, 69, 75, 214, 215*). Glutathione

transferases, particularly GSTA1-1, are among this group of enzymes characterized by their ability to catalyze reactions towards a large number of structurally diverse compounds (*14, 15, 75, 84, 85*). Drug metabolizing enzymes have existed for at least 2.5 billion years (*216*), and GSTs constitute an ancient super family of enzymes thought to have evolved from a thioredoxin-like progenitor in response to oxidative stress (*85, 217, 218*). The cytosolic superfamily of GSTs appear to have diversified though divergent evolution, with the A-class presumed to be the most modern of the GST classes (*72, 85, 219*) arising from duplication of the theta-class gene and incorporation of the α-helix, along with recruitment of an arginine into the inner coordination sphere of the catalytic sulfur. GSTA1-1 would seem to be a clear case of an enzyme whose catalytic promiscuity is a result of natural selection, either through retention of primordial catalytic versatility or through reacquiring a useful trait for detoxification. GSTA4-4, though quite similar in both sequence and structure, retains only a fraction of the promiscuity that GSTA1-1 does. The substrate promiscuity of both of these isoforms, as well as the GST mutants discussed in the previous chapters, will be quantified below.

B. Materials and Methods

Enzyme Activity Assays

Protein expression and purification was completed as described in Section 2B. Enzyme activities were determined by monitoring changes in absorbance on a dual-beam Olis Modernized Aminco DW-2 absorbance spectrophotometer (Olis Inc., Bogart, GA). Assays were conducted in a thermostated cell compartment at 30 °C with 5 mm standard quartz cuvettes in 100 mM sodium phosphate buffer. The pH for each substrate was adjusted to minimize the nonenzymatic reaction. The reaction mixture contained enzyme and freshly prepared reduced

GSH, and was allowed to temperature equilibrate for 5 minutes before the reaction was initiated with the addition of substrate. Absorbance change for each substrate concentration was monitored for 60 seconds, and the initial rates were determined from the slope of this line and the molar extinction coefficients at the appropriate wavelengths as follows: reduced glutathione (GSH) conjugation with 1-chloro-2,4-dinitrobenzene was monitored at 340 nm (ε_{340} = 9.6 mM^{-1} cm^{-1}), a final concentration of 5 mM GSH, 5% ethanol (v/v), and a 0.025-1.6 mM substrate concentration range, at pH 6.5; addition of GSH to ethacrynic acid was monitored at 270 nm (ε_{270} = 5.0 mM^{-1} cm^{-1}), a final concentration of 1 mM GSH, 2% ethanol (v/v), and a 0.025-0.8 mM substrate concentration range at pH 6.5; conjugation of GSH to phenethyl isothiocyanate was monitored at 274 nm (ε_{274} = 8.89 mM^{-1} cm^{-1}), a final GSH concentration of 1 mM, 2% acetonitrile (v/v), and a 0.0125-0.4 mM substrate concentration range, at pH 6.5; *p*-nitrophenolate release was monitored at 400 nm (ε_{400} = 8.79 mM^{-1} cm^{-1}) for the conjugation of GSH to *p*-nitrophenyl acetate, with a final GSH concentration of 1 mM, 2% ethanol (v/v), and a 0.05-2.4 mM concentration range, at pH 7.0; GSH addition to *trans*-2-nonenal was followed by depletion of *trans*-2-nonenal monitored at 225 nm (ε_{225} = -19.22 mM^{-1} cm^{-1}), with a final GSH concentration of 0.5 mM, 2 % ethanol (v/v), and a 0.00312-0.2 mM substrate concentration range, at pH 6.5; and peroxidase activity towards cumene hydroperoxide was followed by detection of NADPH oxidation monitored at 340 nm (ε_{340} = -6.5 mM^{-1} cm^{-1}), with a final GSH concentration of 1 mM, 5% ethanol (v/v), and a 0.025-2.5 mM substrate concentration range, at pH 7.0. Fresh NADPH to a final concentration of 0.1 mM and 0.6 units/mL glutathione reductase was also added for this assay. A minimum of seven substrate concentrations were used within each concentration range, and triplicate measurements were acquired for each substrate concentration. The nonenzymatic reactions were completed in the same manner, without addition

of enzyme, and the resulting standard curves used to subtract the background reaction from the enzyme catalyzed one at each concentration point. The initial rates were subsequently fit to the Michaelis-Menten equation to determine k_{cat} and K_M, and the catalytic efficiencies determined from the ratio k_{cat}/K_M.

Promiscuity Quantification

The method of Nath and Atkins detailed in Section 1C was used to determine J_{cat} indices for a measure of substrate promiscuity. The catalytic efficiencies used for these calculations were determined as described in the previous section, and a custom Python (www.python.org) script was used for the calculations (Appendix A). The PubChem Substance and Compound Database (http://pubchem.ncbi.nlm.nih.gov) was used to obtain an 881 bit descriptor keyset (Appendix B) for the Tanimoto dissimilarity matrix. The PubChem identification numbers are as follows: 1-chloro-2,4-dinitrobenzene, CID = 6; ethacrynic acid, CID = 3278; phenethyl isothiocyanate, CID = 16741; *p*-nitrophenylacetate, CID = 13243; *trans*-2-nonenal, CID = 5283335; cumene hydroperoxide, CID = 6629. The promiscuity indices were resampled by leaving out each substrate one at a time and recalculating J_{cat} (jackknife analysis).

C. Results and Discussion

Glutathione Steady State Kinetic Parameters

Evaluation of catalytic promiscuity necessarily requires a measure of the catalytic properties of the enzymes under question, as described in Section 1C. Thus, the steady state kinetic parameters for the GSTs previously discussed were acquired. Table 4.1 lists the k_{cat} and K_m values recovered for a basis set of six substrates, shown in Figure 4.2. GSTA4-4 wild-type and GIMFhelix mutant demonstrated a strong discrimination among the substrates examined,

with high catalytic activity toward nonenal, and much lower activity towards the other substrates presented. GSTA4-4 F111V/Y217R was nearly as selective, and GSTA1-1 V111F/R217Y exhibited surprisingly low activity towards nonenal. GSTA1-1 wild-type and GSTA1-1 W21F/F222W were much more promiscuous, although GSTA1-1 wild-type was more efficient at metabolizing CDNB than the rest of the substrates examined, and GSTA1-1 W21F/F222W proved to be more catalytically efficient with *p*NPA and ethacrynic acid as substrates. The bulkier indole group at the end of the C-terminus may provide more productive positioning for aromatic substrates with longer substituents. Previous results have shown a similar shift in the substrate selectivity profiles moving from GSTA1-1 to GSTA4-4 and GIMFhelix using a slightly larger array of substrates (*15, 142*). The stacked bar plot (Figure 4.1) displays the substrate selectivity profiles for each of the GSTs as a ratio of each substrates contribution to the total catalytic efficiencies for all of the substrates.

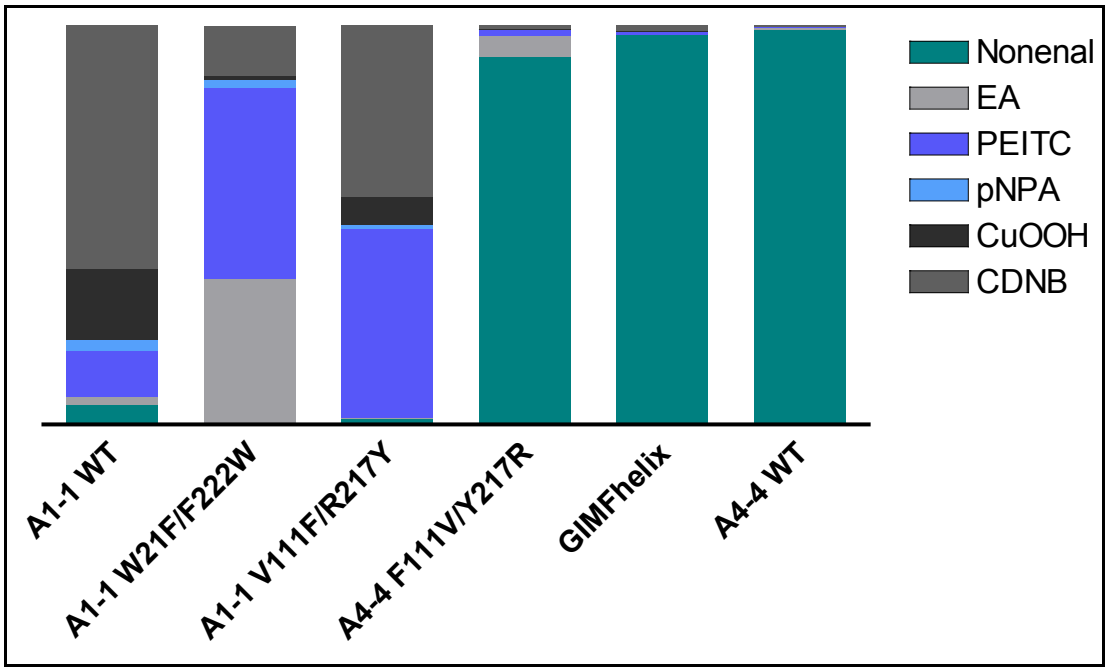


Figure 4.1 Stacked bar plot expressing substrate-selectivity profiles for GSTs. The profiles show the fractions of the total sum of catalytic efficiencies with the six substrates (Figure 4.2) used in the characterization of the enzymes (Table 4.2). The contributions to the bar heights represent the percentage contribution of each respective substrate.

Table 4.1 Catalytic parameters recovered for GSTs. The k_{cat} and K_M values are averages of triplicate evaluations for each substrate-enzyme pair, along with the SD of triplicate assays.

	CDNB		EA		PEITC		*p*NPA		Nonenal		CuOOH	
Enzyme	k_{cat} (s^{-1})	k_M (mM)	k_{cat} (s^{-1})	k_M (mM)	k_{cat} (s^{-1})	k_M (mM)	k_{cat} (s^{-1})	k_M (mM)	k_{cat} (s^{-1})	k_M (mM)	k_{cat} (s^{-1})	k_M (mM)
A1-1 WT*	64.70 ± 1.21	0.45 ± 0.02	0.51 ± 0.06	0.11 ± 0.04	1.90 ± 0.08	0.068 ± 0.008	7.522 ± 0.240	1.25 ± 0.08	0.31 ± 0.02	0.027 ± 0.005	8.21 ± 0.17	0.19 ± 0.01
A1-1 W21F/F222W	179.36 ± 20.06	1.64 ± 0.25	4.705 ± 0.205	0.012 ± 0.003	5.86 ± 0.31	0.014 ± 0.002	4.97 ± 0.37	0.239 ± 0.034	1.39 ± 0.09	1.31 ± 0.12	13.77 ± 0.98	1.35 ± 0.13
A1-1 V111F/R217Y	100.75 ± 13.27	1.58 ± 0.32	0.601 ± 0.002	0.172 ± 0.004	6.167 ± 0.183	0.079 ± 0.004	1.738 ± 0.241	1.04 ± 0.19	0.385 ± 0.050	0.151 ± 0.031	13.77 ± 0.98	1.386 ± 0.134
A4-4 F111R/Y217R*	15.63 ± 1.86	4.20 ± 0.75	1.68 ± 0.07	0.085 ± 0.014	1.73 ± 0.28	0.31 ± 0.09	0.17 ± 0.03	3.15 ± 0.78	130.16 ± 17.34	0.38 ± 0.06	0.88 ± 0.09	2.40 ± 0.38
GIMFhelix*	40.00 ± 1.59	2.01 ± 0.12	0.26 ± 0.08	0.29 ± 0.19	1.97 ± 0.14	0.20 ± 0.03	1.57 ± 0.44	3.51 ± 1.44	26.42 ± 0.50	0.020 ± 0.001	3.19 ± 0.24	1.83 ± 0.23
A4-4 WT*	9.67 ± 0.65	2.45 ± 0.24	1.06 ± 0.06	0.12 ± 0.02	0.67 ± 0.11	0.38 ± 0.11	1.92 ± 0.69	6.11 ± 2.81	97.97 ± 2.51	0.078 ± 0.004	0.57 ± 0.11	3.28 ± 0.96

*The results for these enzymes were kindly provided by Dr. Bengt Mannervik, Dept. of Neurochemistry, Stockholm University, Stockholm, Sweden

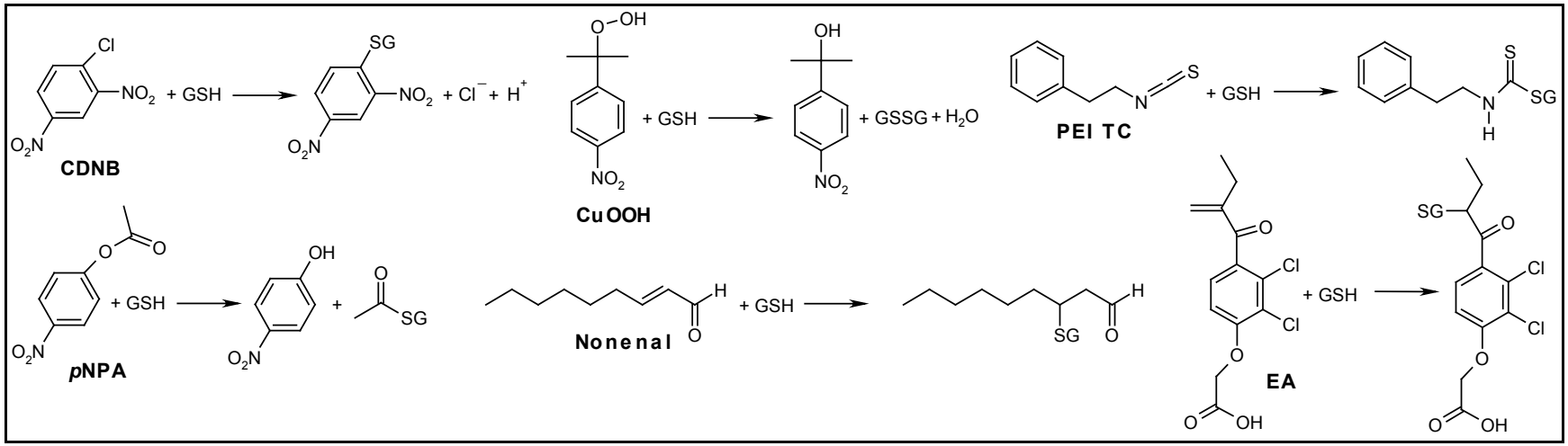


Figure 4.2 Substrate basis set used for promiscuity analysis. Abbreviations: CDNB, 1-chloro-2,4-dinitrobenzene; CuOOH, cumene hydroperoxide; PEITC, phenethyl isothiocyanate; *p*NPA, *para*-nitrophenyl acetate; nonenal, *trans*-2-nonenal; EA, ethacrynic acid.

Substrate Promiscuity Quantification

The method of Nath and Atkins has been successfully used to describe the catalytic promiscuity of several groups of enzymes, including proteases (*14*), cytochrome P450s (*14, 44*) and GSTs (*14*) using a variety of basis sets. This method, described in Section 1C, is based on information entropy, and uses catalytic efficiency as a measure of the likelihood that an enzyme will metabolize a given substrate when presented with equal low concentrations of all the substrates in the basis set. A higher catalytic efficiency increases the probability of that substrate being preferentially metabolized. Included in the determinations of promiscuity indices is a weighting factor based on the structural dissimilarity across the basis set. In this manner, the impact of adding a very similar substrate to a substrate set is less than the impact of adding a structurally distinct substrate to the set. Substrates that are structurally more diverse than others in the same set have a larger impact on the promiscuity score. For this work, an 881 bit descriptor keyset from the PubChem Chemical Structure Database (Appendix B) was used. Substrate promiscuity, represented as J_{cat} values, can range between 0, which is completely specific, and 1, completely promiscuous. The overall dissimilarity of the basis set chosen is reflected in δ_{set}, where $\delta_{set} = k/(k+l)$. The number of structural features present in at least one, but not all members in the set is represented by k, and l is the number of features in all the members of the set. Despite there being only six members in the basis set chosen for this analysis, the compounds chosen demonstrated a high degree of dissimilarity, with $\delta_{set} = 0.9$. The catalytic efficiencies in Table 4.2 were we used to determine promiscuity indices for all of the enzymes presented earlier in this work, also presented in Table 4.2.

Table 4.2 Promiscuity indices (J_{cat}) and catalytic efficiencies of GSTs.

Protein	J_{cat}	CDNB k_{cat}/K_M (mM^{-1} s^{-1})	EA k_{cat}/K_M (mM^{-1} s^{-1})	PEITC k_{cat}/K_M (mM^{-1} s^{-1})	*p*NPA k_{cat}/K_M (mM^{-1} s^{-1})	Nonenal k_{cat}/K_M (mM^{-1} s^{-1})	CuOOH k_{cat}/K_M (mM^{-1} s^{-1})
GSTA1-1 WT	0.72	143.3	4.8	27.7	6.0	11.5	42.2
GSTA1-1 W21F/F222W	0.68	102.4	296.3	392.6	62.9	1.06	8.4
GSTA1-1 V111F/R217Y	0.57	57.1	0.5	62.9	1.3	1.8	9.1
GSTA4-4 F111V/Y217R	0.22	3.7	19.7	5.6	0.05	345.2	0.4
GIMFhelix	0.09	19.9	0.9	10.0	0.5	1344.7	1.7
GSTA4-4 WT	0.05	4.0	9.1	1.8	0.3	1248.5	0.2

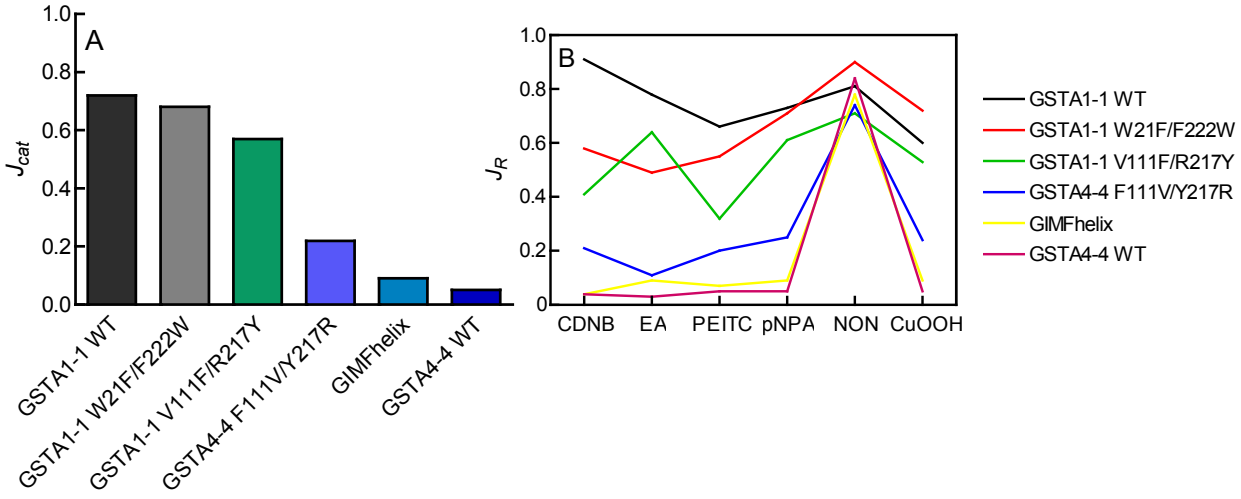


Figure 4.3 Promiscuity indices (J_{cat}) of GSTs and resampling analysis of J_{cat} on a per substrate basis. The entire basis set (Figure 4.1) was used to calculate J_{cat} values (A) and each substrate was removed individually and J_{cat} recalculated (B).

As illustrated in Figure 4.3 A, GSTA1-1 wild-type was the most promiscuous of the enzymes in this analysis, followed closely by GSTA1-1 W21F/F222W, with J_{cat} values of 0.72 and 0.68, respectively. GSTA4-4 wild-type (J_{cat} = 0.05) and GIMFhelix (J_{cat} = 0.09) were much more substrate specific. GSTA1-1 V111F/R217Y and GSTA4-4 F111V/Y217R exhibited intermediate promiscuity, falling between GSTA1-1 and GSTA4-4. Resampling of the promiscuity indices utilizing jackknife analysis (Figure 4.3 B) further illustrates the high specificity of GSTA4-4. Removal of nonenal, which is similar to the presumed cognate substrate 4-hydroxynonenal for

this isoform, and recalculation of the promiscuity indices increases J_{cat} to a level resembling GSTA1-1, while removal of any other substrate did not affect the calculated promiscuity scores of GSTA4-4. Similar resampling analyses performed regarding GSTA1-1 and GSTA1-1 W21F/F222W demonstrated only moderate change in J_{cat} as each substrate was removed, while the exchange mutants, GSTA1-1 V111F/R217Y and GSTA4-4 F111V/Y217Y exhibited intermediate effects. This further validates the utility of this substrate basis set for comparing promiscuity among the GSTs used in this work.

Chapter 5: Linking Catalytic Promiscuity and Conformational Heterogeneity

A. Introduction

The mechanisms by which enzymes achieve catalytic promiscuity are not well understood. The structure and sequence similarity across the GSTs discussed in the preceding chapters makes it difficult to attribute the difference in substrate specificity to any one structural characteristic. One common suggestion is that promiscuous enzymes are more 'flexible' than their substrate-specific counterparts, as suggested in the previous chapters. Intuitively, an enzyme cannot exhibit catalytic activity toward many dissimilar substrates without a concomitant versatility of structure. However, due to the difficulties in quantifying both 'promiscuity' and 'flexibility' no systematic evaluation of this putative relationship has been performed. The increased conformational heterogeneity and dynamics of the C-terminus and tower region of catalytically promiscuous GSTA1-1 as compared to substrate specific GSTA4-4 described above provide an excellent platform with which to investigate this link between plasticity of form and function. Using this series of GSTs as a platform, it becomes possible to seek quantitative trends in the relationship between catalytic promiscuity as summarized in Table 4.2 and active site 'flexibility' as measured by dynamics of the protein core from the fluorescence lifetime distributions in Chapter 2 or as measured by the DSC experiments in Chapter 3. The relationship between fluorescence lifetime distributions and quantitative promiscuity of a subset of four GSTs, GSTA1-1 wild-type, GSTA1-1 V111F/R217Y, GSTA4-4 F111V/Y217R, and GSTA4-4 wild-type will be discussed here. In addition, the DSC data discussed in Chapter 3 for a slightly different subset of five enzymes, GSTA1-1 wild-type,

GSTA1-1 W21F/F222W, GSTA1-1 V111F/R217Y, GIMFhelix, and GSTA4-4 wild-type, will be compared to the promiscuity values determined in the previous chapter (Chapter 4).

B. Results and Discussion

The local environment at the domain interface, where Trp21 is located (Figure 2.4), was examined using FWHM of the fluorescence lifetime distribution of this residue for a subset of these proteins, as in Chapter 2. For this examination, the GSTA1-1 W21F/F222W mutant was not used because the fluorescent reporter was been removed from the protein core. When examined in relation to J_{cat}, (Figure 5.1) each increase in the conformational heterogeneity in the nanosecond range at Trp21 displayed an increase in the catalytic promiscuity of the enzyme. The increase in the fluorescence lifetime distribution FWHM from GSTA4-4 wild-type through the exchange mutants GSTA4-4 F111V/Y217R and GSTA1-1 V111F/R217Y was modest and almost linear. This was followed by a sharp increase at GSTA1-1 wild-type. This suggests that there is a threshold in dynamic mobility in the domain interface region that needs to be exceeded in order to achieve the full catalytic promiscuity observed in GSTA1-1 wild-type. Substitution of the native Trp21 with alanine has been shown to impact the catalytic prowess of this enzyme, and to reduce CDNB binding affinity (*162*), demonstrating a linkage between the H-site portion of the active site domain interface. The correlation between promiscuity and protein dynamics in this region show a similar effect. Retention of the core domain of GSTA4-4 exerts a stabilizing influence in this domain, which in turn prevents the active site of GSTA4-4 F111V/Y217R from full access to conformational space even with the edge-on-face aromatic interaction between the C-terminus and the tower region removed. This is evident in the modest increase in promiscuity and fluorescence lifetime distribution in this mutant compared to GSTA4-4 wild-type. However, in GSTA1-1 V111F/R217Y restriction of active site mobility though inclusion of the aromatic

interaction inhibits both promiscuity and conformational heterogeneity at the domain interface, as seen through the reduction in J_{cat} and fluorescence lifetime distribution FWHM from wild-type GSTA1-1. These data suggest that dynamics of buried regions of the protein, outside the immediate vicinity of the active site, are functionally linked to catalytic promiscuity, although the correlation is not linear and high levels of promiscuity are only achieved with broad conformational heterogeneity. The wild-type enzyme has significantly more core 'flexibility' than other isoforms or variants.

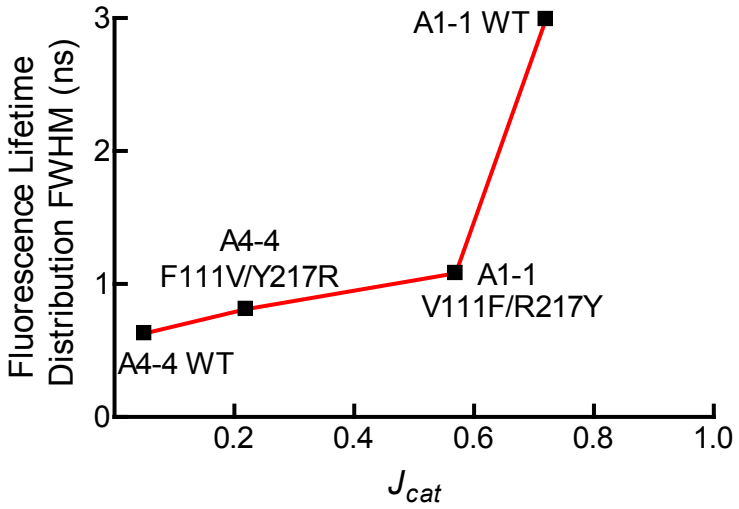


Figure 5.1 Substrate promiscuity relationship to fluorescence lifetime distribution. The fluorescence lifetime distribution full-width at half-maximum is compared to promiscuity (J_{cat}). The *red line* simply connects the points for illustration. A correlation with a Pearson *r* value of 0.8210 was determined with an R^2 of 0.6740.

The conformational distribution of the active site and core are available from the DSC experiments described in Chapter 3. The distributions of enthalpic states at 298 K recovered from DSC were also examined, shown in Figure 5.2 A. Taking the FWHM of the probability distribution of the deconvoluted low temperature transitions as a measure of active site conformational heterogeneity, GSTA4-4 and GIMFhelix both displayed a very narrow range of enthalpic states present at 298 K, and both exhibited high substrate selectivity. The FWHM of

this probability distribution for GSTA1-1 V111F/R217Y increased ~10-fold over GSTA4-4, from 5.8 kJ/mol to 67.5 kJ/mol, as did J_{cat}, from 0.05 to 0.57. Surprisingly, the probability distribution of enthalpic states for GSTA1-1 W21F/F222W was slightly narrower than for GSTA1-1 V111F/R217Y at 49 kJ/mol, even with a slightly higher J_{cat} (0.68). Both were well below wild-type GSTA1-1, with a FWHM of 405 kJ/mol, even with J_{cat} values only slightly lower than the value of 0.72 determined for GSTA1-1. A similar trend was observed in the free energy barrier heights revealed for the deconvoluted low transitions of the enzymes (Figure 5.2 B). The barrier height for GSTA4-4, at 1676.1 kJ/mol, likely prevents all but the most limited access to active site conformational space, centered around the best structure for catalysis of long chain alkenals. The GIMFhelix barrier height was considerably lower (101.6 kJ/mol) for the same transition, but still enough to considerably restrict the conformational heterogeneity of this GST variant. GSTA1-1 V111F/R217Y and GSTA1-1 W21F/F222W demonstrated comparable energetic barrier heights, 7.2 and 9.5 kJ/mol, respectively. These barriers restrict transitioning of the active site structure between macrostates only slightly, enough to reduce the promiscuity compared to GSTA1-1 wild-type with the basis set examined.GSTA1-1 wild-type exhibits a barrierless transition of the active site region across an impressively wide range of enthalpic states, enabling effective promiscuity for this protein's primary role as a detoxification enzyme. With increasing substrate promiscuity, proceeding from GSTA4-4 to GSTA1-1, the free energy barrier heights decreased and enthalpic distribution widths increased. Notably, when the FWHM values for the distributions of enthalpic states were natural logarithm (ln) transformed, a reasonably linear correlation was discovered between these values and the promiscuity indices recovered in Chapter 4 (Figure 5.A). Through Boltzmann's equation ($S=k_B \ln W$), entropy S is equal to the natural log transform of the number of microstates W multiplied by Boltzmann's

constant (k_B). This correlation suggests a strong relationship between increasing active site entropy and a higher level of catalytic promiscuity. In addition, the natural logarithms of recovered barrier heights decreased in a linear fashion with decreasing promiscuity, providing a link between catalytic rates and substrate specificity (Figure 5.2 B). Low, or nonexistent, free energy barriers between state widely separated in conformational space results in similar catalytic ability across several substrates, but high barriers limit a specific enzyme to its cognate substrate.

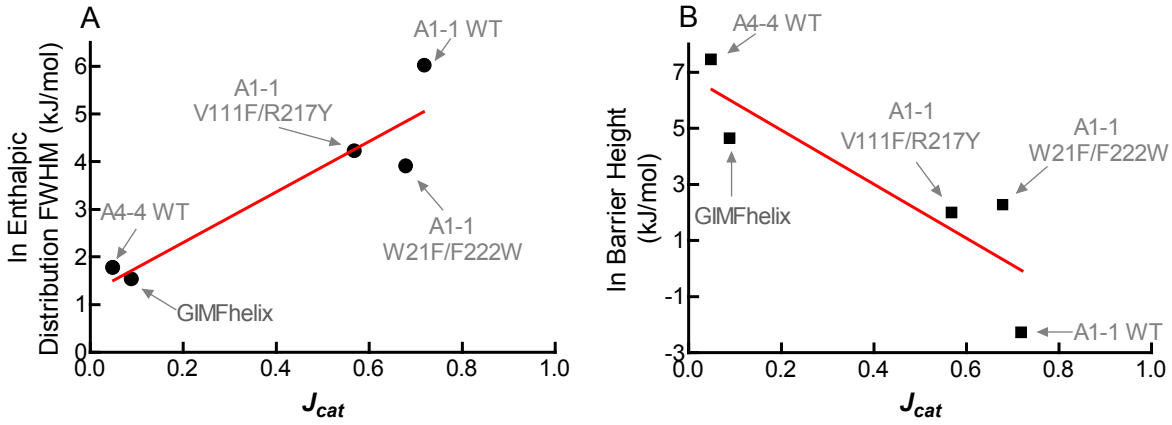


Figure 5.2 Plots of distributions of enthalpic states or free energy barrier heights vs. catalytic promiscuity for GSTs. The GST variants are labeled in *grey*. A) The ln transforms of the full-width at half-maximum values for the probability distributions of enthalpic states at 298 K from the deconvoluted low temperature transitions are shown (●). The linear best fit is indicated with the *red line*, with a slope of 5.3, and R^2 = 0.8619. A correlation with a Pearson *r* value of 0.9284 was determined with the same R^2. B) The ln transforms of the free energy barrier heights from the deconvoluted low temperature transitions are shown (■). The linear best fit is indicated by the *red line*, with a slope of -9.65, and R^2 = 0.7666. Here the Pearson *r* value was -0.8756. The barrier height of -30 kJ/mol for GSTA1-1 wild-type was substituted with 0.1 kJ/mol to allow the ln transformation.

C. Conclusions

Based on the data, a substantial degree of catalytic promiscuity can be conferred at a much lower level of conformational heterogeneity than is exhibited in GSTA1-1 wild-type, and the additional increase in structural dynamics observed for this enzyme results in only modest increases in promiscuity. This is consistent with accumulating evidence that moderate promiscuity is a common feature among enzymes. The data may effectively define an upper limit

to substrate promiscuity, and the corresponding level of conformational heterogeneity required, beyond which an enzyme would exhibit no substrate preference whatsoever. This enzyme would then begin interfering with normal metabolic processes, or retain so little defined structure that the enzyme's catalytic power would be nearly abolished. From a protein folding perspective, GSTA1-1 seems to poised through entropy – energy compensation at just the appropriate level of the characteristic folding funnel to have a molten globule-like active site to enable promiscuity, while the rest of the core structure remains both flexible enough to allow for significant conformational changes accommodating a diverse array of substrates, and restrictive enough to provide overall stability and some preference in substrates.

Glossary

15d-PGJ2	15-deoxy-$\Delta^{12,14}$-prostiglandin J_2
A	alanine
Ala	alanine
ARE	antioxidant response element
Arg	arginine
ATP	adenosine triphosphate
BLA	bovine α-lactalbumin
C	cysteine
CD	circular dichroism
CDNB	1-chloro-2,4-dinitrobenzene
cm	centimeter
CoA	coenzyme A
C_p	heat capacity
CuOOH	cumene hydroperoxide
CYP	cytochrome P450
Cys	cysteine
Da	Dalton
DDT	dichlorodiphenyltrichloroethane
DHFR	dihydrofolate redutase
DNA	deoxyribonucleic acid
DSC	differential scanning calorimetry
E	glutamic acid

EA	ethacrynic acid
F	phenylalanine
FFT	fast Fourier transform
FWHM	full width at half maximum
G	glycine
Glu	glutamic acid
Gly	glycine
GSH	glutathione
GST	glutathione transferase
H	enthalpy
HDX	hydrogen-deuterium exchange mass spectrometry
HEWL	hen-egg-white lysozyme
HNE	4-hydroxynonenal
I	isoleucine
Ile	isoleucine
IPTG	isopropyl-β-D-thiogalactopyranoside
K	Kelvin
K	lysine
k_B	Boltzmann's constant
kcal	kilocalorie
KDGA	D-2-keto-3-deoxy-gluconate aldolase
Keap 1	Kelch-like ECH-associated protein 1
kJ	kiloJoule

KNF	Koshland-Nemethy-Filmer
L	leucine
LB	lysogeny broth
Leu	leucine
ln	natural logarithm
M	methionine
MAPEG	membrane-associate proteins in eicosanoid and glutathine metabolism
Met	methionine
MDR	multidrug resistance protein
mM	millimolar
mm	millimeter
mMjCM	*Methanococcus jannaschi* chorismutase monomer
MOAT	multiple organic anion transporter
mol	mole
NADPH	nicotinamide adenine dinucleotide phosphate
nm	nanometer
NMR	nuclear magnetic resonance
NON	*trans*-2-nonenal
Nrf2	nuclear factor-erythroid 2 p45-related factor 2
ns	nanosecond
OD	optical density
PAH	polyaromatic hrdrocarbon
PBS	phospahte buffered saline

PEITC	phenethyl isothiocyanate
PGP	P-glycoprotein
Phe	phenyalanine
*p*NPA	*para*-nitrophenyl acetate
PPARγ	peroxisome proliferator-activated receptor γ
PSBD	peripheral subunit binding domain
R	gas constant, 8.318 J K^{-1} mol^{-1}
R	arginine
RMSD	root mean squared distance
S	entropy
S	serine
s	second
Ser	serine
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TB	terrific broth
T_m	transition midpoint
Tris-HCl	tris(hydroxymethyl)aminomethane hydrochloric acid
Trp	tryptophan
TST	transition state
Tyr	tyrosine
UGT	uridyl diphosphate glucuronic acid transferase
UV	ultraviolet
V	valine

Val	valine
v/v	volume by volume
W	tryptophan
wt	wild-type
Y	tyrosine
Θ	molar ellipticity
μM	micromolar

References

1. Miller, B. G., and Wolfenden, R. (2002) Catalytic proficiency: the unusual case of OMP decarboxylase, *Annual review of biochemistry 71*, 847-885.
2. Fischer, E. (1894) Einfluss der Configuration auf die Wirkung der Enzyme, *Bererichte der Deutschen Chemishen Gesellschaft 27*, 2985-2991.
3. Haldane, J. B. S. (1930) *Enzymes*, Green anc Co., London, UK.
4. Koshland, D. E. (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis, *Proc Natl Acad Sci U S A 44*, 98-104.
5. Koshland, D. E., Jr., Ray, W. J., Jr., and Erwin, M. J. (1958) Protein structure and enzyme action, *Federation proceedings 17*, 1145-1150.
6. Ma, B., Shatsky, M., Wolfson, H. J., and Nussinov, R. (2002) Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations, *Protein Sci 11*, 184-197.
7. Pocker, Y., and Stone, J. T. (1965) The catalytic versatility of erythrocyte carbonic anhydrase. The enzyme-catalyzed hydrolysis of rho-nitrophenyl acetate, *J Am Chem Soc 87*, 5497-5498.
8. Monod, J., Wyman, J., and Changeux, J. P. (1965) On the Nature of Allosteric Transitions: a Plausible Model, *J Mol Biol 12*, 88-118.
9. Jackson, M. B. (2006) *Molecular and cellular biophysics*, Cambridge University Press, Cambridge, UK ; New York.
10. Frauenfelder, H., Sligar, S. G., and Wolynes, P. G. (1991) The energy landscapes and motions of proteins, *Science 254*, 1598-1603.
11. Hammes, G. G. (2002) Multiple conformational changes in enzyme catalysis, *Biochemistry 41*, 8221-8228.
12. Kasimova, M. R., Milstein, S. J., and Freire, E. (1998) The conformational equilibrium of human growth hormone, *J Mol Biol 277*, 409-418.
13. Nashine, V. C., Hammes-Schiffer, S., and Benkovic, S. J. (2010) Coupled motions in enzyme catalysis, *Curr Opin Chem Biol 14*, 644-651.
14. Nath, A., and Atkins, W. M. (2008) A quantitative index of substrate promiscuity, *Biochemistry 47*, 157-166.
15. Hou, L., Honaker, M. T., Shireman, L. M., Balogh, L. M., Roberts, A. G., Ng, K. C., Nath, A., and Atkins, W. M. (2007) Functional promiscuity correlates with conformational heterogeneity in A-class glutathione S-transferases, *J Biol Chem 282*, 23264-23274.
16. Griswold, K. E., Aiyappan, N. S., Iverson, B. L., and Georgiou, G. (2006) The evolution of catalytic efficiency and substrate promiscuity in human theta class 1-1 glutathione transferase, *J Mol Biol 364*, 400-410.
17. Wellen, K. E., and Thompson, C. B. (2012) A two-way street: reciprocal regulation of metabolism and signalling, *Nature reviews 13*, 270-276.
18. Atkins, W. M. (2005) Non-Michaelis-Menten kinetics in cytochrome P450-catalyzed reactions, *Annu Rev Pharmacol Toxicol 45*, 291-310.
19. Rist, W., Graf, C., Bukau, B., and Mayer, M. P. (2006) Amide hydrogen exchange reveals conformational changes in hsp70 chaperones important for allosteric regulation, *J Biol Chem 281*, 16493-16501.

20. Zalatan, J. G., and Herschlag, D. (2009) The far reaches of enzymology, *Nature chemical biology 5*, 516-520.
21. Hanover, J. A., Krause, M. W., and Love, D. C. (2012) Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation, *Nature reviews 13*, 312-321.
22. Ma, B., and Nussinov, R. Enzyme dynamics point to stepwise conformational selection in catalysis, *Curr Opin Chem Biol 14*, 652-659.
23. Ma, B., Kumar, S., Tsai, C. J., Hu, Z., and Nussinov, R. (2000) Transition-state ensemble in enzyme catalysis: possibility, reality, or necessity?, *Journal of theoretical biology 203*, 383-397.
24. Benkovic, S. J., and Hammes-Schiffer, S. (2003) A perspective on enzyme catalysis, *Science 301*, 1196-1202.
25. Koshland, D. E., Jr. (1959) Enzyme flexibility and enzyme action, *Journal of cellular and comparative physiology 54*, 245-258.
26. Koshland, D. E., Jr., Nemethy, G., and Filmer, D. (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits, *Biochemistry 5*, 365-385.
27. Sugase, K., Dyson, H. J., and Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein, *Nature 447*, 1021-1025.
28. Herschlag, D. (1988) The role of induced fit and conformational changes of enzymes in specificity and catalysis, *Bioorganic Chemistry 16*, 62-96.
29. James, L. C., and Tawfik, D. S. (2003) Conformational diversity and protein evolution--a 60-year-old hypothesis revisited, *Trends Biochem Sci 28*, 361-368.
30. Sullivan, S. M., and Holyoak, T. (2008) Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection, *Proc Natl Acad Sci U S A 105*, 13829-13834.
31. Kumar, S., Ma, B., Tsai, C. J., Sinha, N., and Nussinov, R. (2000) Folding and binding cascades: dynamic landscapes and population shifts, *Protein Sci 9*, 10-19.
32. Jackson, C. J., Foo, J. L., Tokuriki, N., Afriat, L., Carr, P. D., Kim, H. K., Schenk, G., Tawfik, D. S., and Ollis, D. L. (2009) Conformational sampling, catalysis, and evolution of the bacterial phosphotriesterase, *Proc Natl Acad Sci U S A 106*, 21631-21636.
33. Vamvaca, K., Vogeli, B., Kast, P., Pervushin, K., and Hilvert, D. (2004) An enzymatic molten globule: efficient coupling of folding and catalysis, *Proc Natl Acad Sci U S A 101*, 12860-12864.
34. Pervushin, K., Vamvaca, K., Vogeli, B., and Hilvert, D. (2007) Structure and dynamics of a molten globular enzyme, *Nature structural & molecular biology 14*, 1202-1206.
35. Roca, M., Messer, B., Hilvert, D., and Warshel, A. (2008) On the relationship between folding and chemical landscapes in enzyme catalysis, *Proc Natl Acad Sci U S A 105*, 13877-13882.
36. Benkovic, S. J., Hammes, G. G., and Hammes-Schiffer, S. (2008) Free-energy landscape of enzyme catalysis, *Biochemistry 47*, 3317-3321.
37. Boehr, D. D., McElheny, D., Dyson, H. J., and Wright, P. E. (2006) The dynamic energy landscape of dihydrofolate reductase catalysis, *Science 313*, 1638-1642.
38. Boehr, D. D., McElheny, D., Dyson, H. J., and Wright, P. E. (2010) Millisecond timescale fluctuations in dihydrofolate reductase are exquisitely sensitive to the bound ligands, *Proc Natl Acad Sci U S A 107*, 1373-1378.

39. Hammes, G. G., Chang, Y. C., and Oas, T. G. (2009) Conformational selection or induced fit: a flux description of reaction mechanism, *Proceedings of the National Academy of Sciences of the United States of America 106*, 13737-13741.
40. Atkins, W. M., and Qian, H. (2011) Stochastic ensembles, conformationally adaptive teamwork, and enzymatic detoxification, *Biochemistry 50*, 3866-3872.
41. Copley, S. D. (2003) Enzymes with extra talents: moonlighting functions and catalytic promiscuity, *Curr Opin Chem Biol 7*, 265-272.
42. Patil, A., Kinoshita, K., and Nakamura, H. (2010) Hub promiscuity in protein-protein interaction networks, *International journal of molecular sciences 11*, 1930-1943.
43. Khersonsky, O., and Tawfik, D. S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective, *Annual review of biochemistry 79*, 471-505.
44. Foti, R. S., Honaker, M., Nath, A., Pearson, J. T., Buttrick, B., Isoherranen, N., and Atkins, W. M. (2011) Catalytic versus inhibitory promiscuity in cytochrome P450s: implications for evolution of new function, *Biochemistry 50*, 2387-2393.
45. O'Brien, P. J., and Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities, *Chem Biol 6*, R91-R105.
46. Huberts, D. H., and van der Klei, I. J. (2010) Moonlighting proteins: an intriguing mode of multitasking, *Biochim Biophys Acta 1803*, 520-525.
47. Khersonsky, O., Roodveldt, C., and Tawfik, D. S. (2006) Enzyme promiscuity: evolutionary and mechanistic aspects, *Curr Opin Chem Biol 10*, 498-508.
48. Jeffery, C. J. (1999) Moonlighting proteins, *Trends Biochem Sci 24*, 8-11.
49. Muro-Pastor, A. M., Ostrovsky, P., and Maloy, S. (1997) Regulation of gene expression by repressor localization: biochemical evidence that membrane and DNA binding by the PutA protein are mutually exclusive, *Journal of bacteriology 179*, 2788-2791.
50. Ostrovsky de Spicer, P., and Maloy, S. (1993) PutA protein, a membrane-associated flavin dehydrogenase, acts as a redox-dependent transcriptional regulator, *Proc Natl Acad Sci U S A 90*, 4295-4298.
51. Furukawa, T., Yoshimura, A., Sumizawa, T., Haraguchi, M., Akiyama, S., Fukui, K., Ishizawa, M., and Yamada, Y. (1992) Angiogenic factor, *Nature 356*, 668.
52. Meyer-Siegler, K., Mauro, D. J., Seal, G., Wurzer, J., deRiel, J. K., and Sirover, M. A. (1991) A human nuclear uracil DNA glycosylase is the 37-kDa subunit of glyceraldehyde-3-phosphate dehydrogenase, *Proc Natl Acad Sci U S A 88*, 8460-8464.
53. O'Loughlin, T. L., Patrick, W. M., and Matsumura, I. (2006) Natural history as a predictor of protein evolvability, *Protein Eng Des Sel 19*, 439-442.
54. Buschiazzo, A., Amaya, M. F., Cremona, M. L., Frasch, A. C., and Alzari, P. M. (2002) The crystal structure and mode of action of trans-sialidase, a key enzyme in Trypanosoma cruzi pathogenesis, *Molecular cell 10*, 757-768.
55. Colman, P. M., and Smith, B. J. (2002) The trypanosomal trans-sialidase: two catalytic functions associated with one catalytic site, *Structure 10*, 1466-1468.
56. Fong, D. H., and Berghuis, A. M. (2002) Substrate promiscuity of an aminoglycoside antibiotic resistance enzyme via target mimicry, *Embo J 21*, 2323-2331.
57. Theodossis, A., Walden, H., Westwick, E. J., Connaris, H., Lamble, H. J., Hough, D. W., Danson, M. J., and Taylor, G. L. (2004) The structural basis for substrate promiscuity in 2-keto-3-deoxygluconate aldolase from the Entner-Doudoroff pathway in Sulfolobus solfataricus, *J Biol Chem 279*, 43886-43892.

58. Liu, Y., Lv, T., Ren, J., Wang, M., Wu, Q., and Zhu, D. (2011) The catalytic promiscuity of a microbial 7alpha-hydroxysteroid dehydrogenase. Reduction of non-steroidal carbonyl compounds, *Steroids 76*, 1136-1140.
59. Khersonsky, O., and Tawfik, D. S. (2006) The histidine 115-histidine 134 dyad mediates the lactonase activity of mammalian serum paraoxonases, *J Biol Chem 281*, 7649-7656.
60. Yeung, D. T., Lenz, D. E., and Cerasoli, D. M. (2005) Analysis of active-site amino-acid residues of human serum paraoxonase using competitive substrates, *The FEBS journal 272*, 2225-2230.
61. Bornscheuer, U. T., and Kazlauskas, R. J. (2004) Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways, *Angewandte Chemie (International ed) 43*, 6032-6040.
62. Gillam, E. M. (2005) Exploring the potential of xenobiotic-metabolising enzymes as biocatalysts: evolving designer catalysts from polyfunctional cytochrome P450 enzymes, *Clinical and experimental pharmacology & physiology 32*, 147-152.
63. Bone, R., Silen, J. L., and Agard, D. A. (1989) Structural plasticity broadens the specificity of an engineered protease, *Nature 339*, 191-195.
64. Vecchio, A. J., Orlando, B. J., Nandagiri, R., and Malkowski, M. G. (2012) Investigating substrate promiscuity in cyclooxygenase-2: the role of Arg-120 and residues lining the hydrophobic groove, *J Biol Chem 287*, 24619-24630.
65. Funa, N., Ohnishi, Y., Ebizuka, Y., and Horinouchi, S. (2002) Properties and substrate specificity of RppA, a chalcone synthase-related polyketide synthase in Streptomyces griseus, *J Biol Chem 277*, 4628-4635.
66. Honaker, M. T., Acchione, M., Sumida, J. P., and Atkins, W. M. (2011) Ensemble perspective for catalytic promiscuity: calorimetric analysis of the active site conformational landscape of a detoxification enzyme, *J Biol Chem 286*, 42770-42776.
67. Ekroos, M., and Sjogren, T. (2006) Structural basis for ligand promiscuity in cytochrome P450 3A4, *Proc Natl Acad Sci U S A 103*, 13682-13687.
68. Meech, R., Rogers, A., Zhuang, L., Lewis, B. C., Miners, J. O., and Mackenzie, P. I. (2012) Identification of residues that confer sugar selectivity to UDP glycosyltransferase 3A (UGT3A) enzymes, *J Biol Chem 287*, 24122-24130.
69. Dong, D., Ako, R., Hu, M., and Wu, B. (2012) Understanding substrate selectivity of human UDP-glucuronosyltransferases through QSAR modeling and analysis of homologous enzymes, *Xenobiotica 42*, 808-820.
70. Thummel, K. E., and Wilkinson, G. R. (1998) In vitro and in vivo drug interactions involving human CYP3A, *Annu Rev Pharmacol Toxicol 38*, 389-430.
71. Booth, J., Boyland, E., and Sims, P. (1961) An enzyme from rat liver catalysing conjugations with glutathione, *Biochem J 79*, 516-524.
72. Armstrong, R. N. (1997) Structure, catalytic mechanism, and evolution of the glutathione transferases, *Chem Res Toxicol 10*, 2-18.
73. Ladner, J. E., Parsons, J. F., Rife, C. L., Gilliland, G. L., and Armstrong, R. N. (2004) Parallel evolutionary pathways for glutathione transferases: structure and mechanism of the mitochondrial class kappa enzyme rGSTK1-1, *Biochemistry 43*, 352-361.
74. Robinson, A., Huttley, G. A., Booth, H. S., and Board, P. G. (2004) Modelling and bioinformatics studies of the human Kappa-class glutathione transferase predict a novel third glutathione transferase family with similarity to prokaryotic 2-hydroxychromene-2-carboxylate isomerases, *Biochem J 379*, 541-552.

75. Hayes, J. D., Flanagan, J. U., and Jowsey, I. R. (2005) Glutathione transferases, *Annu Rev Pharmacol Toxicol 45*, 51-88.
76. Jakobsson, P. J., Morgenstern, R., Mancini, J., Ford-Hutchinson, A., and Persson, B. (1999) Common structural features of MAPEG -- a widespread superfamily of membrane associated proteins with highly divergent functions in eicosanoid and glutathione metabolism, *Protein Sci 8*, 689-692.
77. Mannervik, B., Alin, P., Guthenberg, C., Jensson, H., Tahir, M. K., Warholm, M., and Jornvall, H. (1985) Identification of three classes of cytosolic glutathione transferase common to several mammalian species: correlation between structural data and enzymatic properties, *Proc Natl Acad Sci U S A 82*, 7202-7206.
78. Meyer, D. J., Coles, B., Pemble, S. E., Gilmore, K. S., Fraser, G. M., and Ketterer, B. (1991) Theta, a new class of glutathione transferases purified from rat and man, *Biochem J 274 (Pt 2)*, 409-414.
79. Meyer, D. J., and Thomas, M. (1995) Characterization of rat spleen prostaglandin H D-isomerase as a sigma-class GSH transferase, *Biochem J 311 (Pt 3)*, 739-742.
80. Pemble, S. E., Wardle, A. F., and Taylor, J. B. (1996) Glutathione S-transferase class Kappa: characterization by the cloning of rat mitochondrial GST and identification of a human homologue, *Biochem J 319 (Pt 3)*, 749-754.
81. Board, P. G., Baker, R. T., Chelvanayagam, G., and Jermiin, L. S. (1997) Zeta, a novel class of glutathione transferases in a range of species from plants to humans, *Biochem J 328 (Pt 3)*, 929-935.
82. Board, P. G., Coggan, M., Chelvanayagam, G., Easteal, S., Jermiin, L. S., Schulte, G. K., Danley, D. E., Hoth, L. R., Griffor, M. C., Kamath, A. V., Rosner, M. H., Chrunyk, B. A., Perregaux, D. E., Gabel, C. A., Geoghegan, K. F., and Pandit, J. (2000) Identification, characterization, and crystal structure of the Omega class glutathione transferases, *J Biol Chem 275*, 24798-24806.
83. Zhan, Y., and Rule, G. S. (2004) Glutathione induces helical formation in the carboxy terminus of human glutathione transferase A1-1, *Biochemistry 43*, 7244-7254.
84. Wilce, M. C., and Parker, M. W. (1994) Structure and function of glutathione S-transferases, *Biochim Biophys Acta 1205*, 1-18.
85. Sheehan, D., Meade, G., Foley, V. M., and Dowd, C. A. (2001) Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily, *Biochem J 360*, 1-16.
86. Armstrong, R. N. (1991) Glutathione S-transferases: reaction mechanism, structure, and function, *Chem Res Toxicol 4*, 131-140.
87. Johansson, A. S., and Mannervik, B. (2001) Human glutathione transferase A3-3, a highly efficient catalyst of double-bond isomerization in the biosynthetic pathway of steroid hormones, *J Biol Chem 276*, 33061-33065.
88. Paumi, C. M., Smitherman, P. K., Townsend, A. J., and Morrow, C. S. (2004) Glutathione S-transferases (GSTs) inhibit transcriptional activation by the peroxisomal proliferator-activated receptor gamma (PPAR gamma) ligand, 15-deoxy-delta 12,14prostaglandin J2 (15-d-PGJ2), *Biochemistry 43*, 2345-2352.
89. Wakabayashi, N., Dinkova-Kostova, A. T., Holtzclaw, W. D., Kang, M. I., Kobayashi, A., Yamamoto, M., Kensler, T. W., and Talalay, P. (2004) Protection against electrophile and oxidant stress by induction of the phase 2 response: fate of cysteines of the Keap1 sensor modified by inducers, *Proc Natl Acad Sci U S A 101*, 2040-2045.

90. Hubatsch, I., Ridderstrom, M., and Mannervik, B. (1998) Human glutathione transferase A4-4: an alpha class enzyme with high catalytic efficiency in the conjugation of 4-hydroxynonenal and other genotoxic products of lipid peroxidation, *Biochem J 330 (Pt 1)*, 175-179.
91. Coles, B., and Ketterer, B. (1990) The role of glutathione and glutathione transferases in chemical carcinogenesis, *Critical reviews in biochemistry and molecular biology 25*, 47-70.
92. Tew, K. D. (1994) Glutathione-associated enzymes in anticancer drug resistance, *Cancer research 54*, 4313-4320.
93. Ranson, H., Rossiter, L., Ortelli, F., Jensen, B., Wang, X., Roth, C. W., Collins, F. H., and Hemingway, J. (2001) Identification of a novel class of insect glutathione S-transferases involved in resistance to DDT in the malaria vector Anopheles gambiae, *Biochem J 359*, 295-304.
94. Cummins, I., Dixon, D. P., Freitag-Pohl, S., Skipsey, M., and Edwards, R. (2011) Multiple roles for plant glutathione transferases in xenobiotic detoxification, *Drug Metabolism Reviews 43*, 266-280.
95. Litwack, G., Ketterer, B., and Arias, I. M. (1971) Ligandin: a hepatic protein which binds steroids, bilirubin, carcinogens and a number of exogenous organic anions, *Nature 234*, 466-467.
96. Keen, J. H., and Jakoby, W. B. (1978) Glutathione transferases. Catalysis of nucleophilic reactions of glutathione, *J Biol Chem 253*, 5654-5657.
97. Ishikawa, T. (1992) The ATP-dependent glutathione S-conjugate export pump, *Trends Biochem Sci 17*, 463-468.
98. Gottesman, M. M., and Pastan, I. (1993) Biochemistry of multidrug resistance mediated by the multidrug transporter, *Annual review of biochemistry 62*, 385-427.
99. Leier, I., Jedlitschky, G., Buchholz, U., Cole, S. P., Deeley, R. G., and Keppler, D. (1994) The MRP gene encodes an ATP-dependent export pump for leukotriene C4 and structurally related conjugates, *J Biol Chem 269*, 27807-27810.
100. Eaton, D. L., and Bammler, T. K. (1999) Concise review of the glutathione S-transferases and their significance to toxicology, *Toxicol Sci 49*, 156-164.
101. Awasthi, S., Sharma, R., Singhal, S. S., Zimniak, P., and Awasthi, Y. C. (2002) RLIP76, a novel transporter catalyzing ATP-dependent efflux of xenobiotics, *Drug Metab Dispos 30*, 1300-1310.
102. Morgan, A. S., Sanderson, P. E., Borch, R. F., Tew, K. D., Niitsu, Y., Takayama, T., Von Hoff, D. D., Izbicka, E., Mangold, G., Paul, C., Broberg, U., Mannervik, B., Henner, W. D., and Kauvar, L. M. (1998) Tumor efficacy and bone marrow-sparing properties of TER286, a cytotoxin activated by glutathione S-transferase, *Cancer research 58*, 2568-2575.
103. Rosen, L. S., Brown, J., Laxa, B., Boulos, L., Reiswig, L., Henner, W. D., Lum, R. T., Schow, S. R., Maack, C. A., Keck, J. G., Mascavage, J. C., Dombroski, J. A., Gomez, R. F., and Brown, G. L. (2003) Phase I study of TLK286 (glutathione S-transferase P1-1 activated glutathione analogue) in advanced refractory solid malignancies, *Clin Cancer Res 9*, 1628-1638.
104. Rosen, L. S., Laxa, B., Boulos, L., Wiggins, L., Keck, J. G., Jameson, A. J., Parra, R., Patel, K., and Brown, G. L. (2004) Phase 1 study of TLK286 (Telcyta) administered weekly in advanced malignancies, *Clin Cancer Res 10*, 3689-3698.

105. Findlay, V. J., Townsend, D. M., Saavedra, J. E., Buzard, G. S., Citro, M. L., Keefer, L. K., Ji, X., and Tew, K. D. (2004) Tumor cell responses to a novel glutathione S-transferase-activated nitric oxide-releasing prodrug, *Mol Pharmacol 65*, 1070-1079.
106. Saavedra, J. E., Srinivasan, A., Buzard, G. S., Davies, K. M., Waterhouse, D. J., Inami, K., Wilde, T. C., Citro, M. L., Cuellar, M., Deschamps, J. R., Parrish, D., Shami, P. J., Findlay, V. J., Townsend, D. M., Tew, K. D., Singh, S., Jia, L., Ji, X., and Keefer, L. K. (2006) PABA/NO as an anticancer lead: analogue synthesis, structure revision, solution chemistry, reactivity toward glutathione, and in vitro activity, *J Med Chem 49*, 1157-1164.
107. Townsend, D. M., Findlay, V. J., Fazilev, F., Ogle, M., Fraser, J., Saavedra, J. E., Ji, X., Keefer, L. K., and Tew, K. D. (2006) A glutathione S-transferase pi-activated prodrug causes kinase activation concurrent with S-glutathionylation of proteins, *Mol Pharmacol 69*, 501-508.
108. Marnett, L. J., Riggins, J. N., and West, J. D. (2003) Endogenous generation of reactive oxidants and electrophiles and their reactions with DNA and protein, *The Journal of clinical investigation 111*, 583-593.
109. Jowsey, I. R., Smith, S. A., and Hayes, J. D. (2003) Expression of the murine glutathione S-transferase alpha3 (GSTA3) subunit is markedly induced during adipocyte differentiation: activation of the GSTA3 gene promoter by the pro-adipogenic eicosanoid 15-deoxy-Delta12,14-prostaglandin J2, *Biochem Biophys Res Commun 312*, 1226-1235.
110. Itoh, K., Mochizuki, M., Ishii, Y., Ishii, T., Shibata, T., Kawamoto, Y., Kelly, V., Sekizawa, K., Uchida, K., and Yamamoto, M. (2004) Transcription factor Nrf2 regulates inflammation by mediating the effect of 15-deoxy-Delta(12,14)-prostaglandin j(2), *Molecular and cellular biology 24*, 36-45.
111. McMahon, M., Itoh, K., Yamamoto, M., and Hayes, J. D. (2003) Keap1-dependent proteasomal degradation of transcription factor Nrf2 contributes to the negative regulation of antioxidant response element-driven gene expression, *J Biol Chem 278*, 21592-21600.
112. Ishii, T., Itoh, K., Ruiz, E., Leake, D. S., Unoki, H., Yamamoto, M., and Mann, G. E. (2004) Role of Nrf2 in the regulation of CD36 and stress protein expression in murine macrophages: activation by oxidatively modified LDL and 4-hydroxynonenal, *Circulation research 94*, 609-616.
113. Ceaser, E. K., Moellering, D. R., Shiva, S., Ramachandran, A., Landar, A., Venkartraman, A., Crawford, J., Patel, R., Dickinson, D. A., Ulasova, E., Ji, S., and Darley-Usmar, V. M. (2004) Mechanisms of signal transduction mediated by oxidized lipids: the role of the electrophile-responsive proteome, *Biochem Soc Trans 32*, 151-155.
114. Levonen, A. L., Landar, A., Ramachandran, A., Ceaser, E. K., Dickinson, D. A., Zanoni, G., Morrow, J. D., and Darley-Usmar, V. M. (2004) Cellular mechanisms of redox cell signalling: role of cysteine modification in controlling antioxidant defences in response to electrophilic lipid oxidation products, *Biochem J 378*, 373-382.
115. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures, *J Mol Biol 247*, 536-540.
116. Board, P. G., Coggan, M., Wilce, M. C., and Parker, M. W. (1995) Evidence for an essential serine residue in the active site of the Theta class glutathione transferases, *Biochem J 311 (Pt 1)*, 247-250.

117. Ji, X., von Rosenvinge, E. C., Johnson, W. W., Tomarev, S. I., Piatigorsky, J., Armstrong, R. N., and Gilliland, G. L. (1995) Three-dimensional structure, catalytic properties, and evolution of a sigma class glutathione transferase from squid, a progenitor of the lens S-crystallins of cephalopods, *Biochemistry 34*, 5317-5328.
118. Liu, S., Zhang, P., Ji, X., Johnson, W. W., Gilliland, G. L., and Armstrong, R. N. (1992) Contribution of tyrosine 6 to the catalytic mechanism of isoenzyme 3-3 of glutathione S-transferase, *J Biol Chem 267*, 4296-4299.
119. Kolm, R. H., Sroga, G. E., and Mannervik, B. (1992) Participation of the phenolic hydroxyl group of Tyr-8 in the catalytic mechanism of human glutathione transferase P1-1, *Biochem J 285 (Pt 2)*, 537-540.
120. Wang, R. W., Newton, D. J., Huskey, S. E., McKeever, B. M., Pickett, C. B., and Lu, A. Y. (1992) Site-directed mutagenesis of glutathione S-transferase YaYa. Important roles of tyrosine 9 and aspartic acid 101 in catalysis, *J Biol Chem 267*, 19866-19871.
121. Reinemer, P., Dirr, H. W., Ladenstein, R., Schaffer, J., Gallay, O., and Huber, R. (1991) The three-dimensional structure of class pi glutathione S-transferase in complex with glutathione sulfonate at 2.3 A resolution, *Embo J 10*, 1997-2005.
122. 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., and et al. (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.
123. Dirr, H., Reinemer, P., and Huber, R. (1994) X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function, *Eur J Biochem 220*, 645-661.
124. Graminski, G. F., Kubo, Y., and Armstrong, R. N. (1989) Spectroscopic and kinetic evidence for the thiolate anion of glutathione at the active site of glutathione S-transferase, *Biochemistry 28*, 3562-3568.
125. Gustafsson, A., and Mannervik, B. (1999) Benzoic acid derivatives induce recovery of catalytic activity in the partially inactive Met208Lys mutant of human glutathione transferase A1-1, *J Mol Biol 288*, 787-800.
126. Bjornestedt, R., Stenberg, G., Widersten, M., Board, P. G., Sinning, I., Jones, T. A., and Mannervik, B. (1995) Functional significance of arginine 15 in the active site of human class alpha glutathione transferase A1-1, *J Mol Biol 247*, 765-773.
127. Atkins, W. M., Wang, R. W., Bird, A. W., Newton, D. J., and Lu, A. Y. (1993) The catalytic mechanism of glutathione S-transferase (GST). Spectroscopic determination of the pKa of Tyr-9 in rat alpha 1-1 GST, *J Biol Chem 268*, 19188-19191.
128. Hubatsch, I., and Mannervik, B. (2001) A highly acidic tyrosine 9 and a normally titrating tyrosine 212 contribute to the catalytic mechanism of human glutathione transferase A4-4, *Biochem Biophys Res Commun 280*, 878-882.
129. Bruns, C. M., Hubatsch, I., Ridderstrom, M., Mannervik, B., and Tainer, J. A. (1999) Human glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products, *J Mol Biol 288*, 427-439.
130. Mannervik, B., and Danielson, U. H. (1988) Glutathione transferases--structure and catalytic activity, *CRC critical reviews in biochemistry 23*, 283-337.
131. 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., and

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, *Journal of Molecular Biology 232*, 192-212.

132. Grahn, E., Novotny, M., Jakobsson, E., Gustafsson, A., Grehn, L., Olin, B., Madsen, D., Wahlberg, M., Mannervik, B., and Kleywegt, G. J. (2006) New crystal structures of human glutathione transferase A1-1 shed light on glutathione binding and the conformation of the C-terminal helix, *Acta Crystallogr D Biol Crystallogr 62*, 197-207.
133. Gustafsson, A., Etahadieh, M., Jemth, P., and Mannervik, B. (1999) The C-terminal region of human glutathione transferase A1-1 affects the rate of glutathione binding and the ionization of the active-site Tyr9, *Biochemistry 38*, 16268-16275.
134. Nilsson, L. O., Edalat, M., Pettersson, P. L., and Mannervik, B. (2002) Aromatic residues in the C-terminal region of glutathione transferase A1-1 influence rate-determining steps in the catalytic mechanism, *Biochim Biophys Acta 1597*, 157-163.
135. Nilsson, L. O., Gustafsson, A., and Mannervik, B. (2000) Redesign of substrate-selectivity determining modules of glutathione transferase A1-1 installs high catalytic efficiency with toxic alkenal products of lipid peroxidation, *Proc Natl Acad Sci U S A 97*, 9408-9412.
136. Cameron, A. D., Sinning, I., L'Hermite, G., Olin, B., Board, P. G., Mannervik, B., and 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.
137. Le Trong, I., Stenkamp, R. E., Ibarra, C., Atkins, W. M., and Adman, E. T. (2002) 1.3-A resolution structure of human glutathione S-transferase with S-hexyl glutathione bound reveals possible extended ligandin binding site, *Proteins 48*, 618-627.
138. Adman, E. T., Le Trong, I., Stenkamp, R. E., Nieslanik, B. S., Dietze, E. C., Tai, G., Ibarra, C., and Atkins, W. M. (2001) Localization of the C-terminus of rat glutathione S-transferase A1-1: crystal structure of mutants W21F and W21F/F220Y, *Proteins 42*, 192-200.
139. Nieslanik, B. S., Dabrowski, M. J., Lyon, R. P., and Atkins, W. M. (1999) Stopped-flow kinetic analysis of the ligand-induced coil-helix transition in glutathione S-transferase A1-1: evidence for a persistent denatured state, *Biochemistry 38*, 6971-6980.
140. Nieslanik, B. S., Dietze, E. C., Atkins, W. M., Trong, I. L., and Adman, E. (1999) The locally denatured state of glutathione S-transferase A1-1: transition state analysis of ligand-dependent formation of the C-terminal helix, *Pacific Symposium on Biocomputing*, 554-565.
141. Nieslanik, B. S., Ibarra, C., and Atkins, W. M. (2001) The C-terminus of glutathione S-transferase A1-1 is required for entropically-driven ligand binding, *Biochemistry 40*, 3536-3543.
142. Blikstad, C., Shokeer, A., Kurtovic, S., and Mannervik, B. (2008) Emergence of a novel highly specific and catalytically efficient enzyme from a naturally promiscuous glutathione transferase, *Biochim Biophys Acta 1780*, 1458-1463.
143. Kuhnert, D. C., Sayed, Y., Mosebi, S., Sayed, M., Sewell, T., and Dirr, H. W. (2005) Tertiary interactions stabilise the C-terminal region of human glutathione transferase A1-1: a crystallographic and calorimetric study, *J Mol Biol 349*, 825-838.

144. Balogh, L. M., Le Trong, I., Kripps, K. A., Tars, K., Stenkamp, R. E., Mannervik, B., and Atkins, W. M. (2009) Structural analysis of a glutathione transferase A1-1 mutant tailored for high catalytic efficiency with toxic alkenals, *Biochemistry 48*, 7698-7704.
145. Beechem, J. M., and Brand, L. (1985) Time-resolved fluorescence of proteins, *Annual review of biochemistry 54*, 43-71.
146. Lakowicz, J. R. (2006) *Principles of fluorescence spectroscopy*, 3rd ed., Springer, New York.
147. Vivian, J. T., and Callis, P. R. (2001) Mechanisms of tryptophan fluorescence shifts in proteins, *Biophys J 80*, 2093-2109.
148. Szabo, A. G., and Rayner, D. M. (1980) Fluorescence decay of tryptophan conformers in aqueous solution, *Journal of the American Chemical Society 102*, 554-563.
149. Callis, P. R. (1997) 1La and 1Lb transitions of tryptophan: applications of theory and experimental observations to fluorescence of proteins, *Methods in enzymology 278*, 113-150.
150. Nishimoto, E., Yamashita, S., Szabo, A. G., and Imoto, T. (1998) Internal motion of lysozyme studied by time-resolved fluorescence depolarization of tryptophan residues, *Biochemistry 37*, 5599-5607.
151. Poveda, J. A., Prieto, M., Encinar, J. A., Gonzalez-Ros, J. M., and Mateo, C. R. (2003) Intrinsic tyrosine fluorescence as a tool to study the interaction of the shaker B "ball" peptide with anionic membranes, *Biochemistry 42*, 7124-7132.
152. Kroes, S. J., Canters, G. W., Gilardi, G., van Hoek, A., and Visser, A. J. (1998) Time-resolved fluorescence study of azurin variants: conformational heterogeneity and tryptophan mobility, *Biophys J 75*, 2441-2450.
153. Tcherkasskaya, O., Ptitsyn, O. B., and Knutson, J. R. (2000) Nanosecond dynamics of tryptophans in different conformational states of apomyoglobin proteins, *Biochemistry 39*, 1879-1889.
154. Ababou, A., and Bombarda, E. (2001) On the involvement of electron transfer reactions in the fluorescence decay kinetics heterogeneity of proteins, *Protein Sci 10*, 2102-2113.
155. Alcala, J. R., Gratton, E., and Prendergast, F. G. (1987) Fluorescence lifetime distributions in proteins, *Biophys J 51*, 597-604.
156. Alcala, J. R., Gratton, E., and Prendergast, F. G. (1987) Interpretation of fluorescence decays in proteins using continuous lifetime distributions, *Biophys J 51*, 925-936.
157. Gratton, E., Jameson, D. M., Weber, G., and Alpert, B. (1984) A model of dynamic quenching of fluorescence in globular proteins, *Biophys J 45*, 789-794.
158. Gratton, E., Limkeman, M., Lakowicz, J. R., Maliwal, B. P., Cherek, H., and Laczko, G. (1984) Resolution of mixtures of fluorophores using variable-frequency phase and modulation data, *Biophys J 46*, 479-486.
159. Alcala, J. R., Gratton, E., and Prendergast, F. G. (1987) Resolvability of fluorescence lifetime distributions using phase fluorometry, *Biophys J 51*, 587-596.
160. Lasagna, M., Gratton, E., Jameson, D. M., and Brunet, J. E. (1999) Apohorseradish peroxidase unfolding and refolding: intrinsic tryptophan fluorescence studies, *Biophys J 76*, 443-450.
161. Wang, R. W., Bird, A. W., Newton, D. J., Lu, A. Y., and Atkins, W. M. (1993) Fluorescence characterization of Trp 21 in rat glutathione S-transferase 1-1: microconformational changes induced by S-hexyl glutathione, *Protein Sci 2*, 2085-2094.

162. Balchin, D., Fanucchi, S., Achilonu, I., Adamson, R. J., Burke, J., Fernandes, M., Gildenhuys, S., and Dirr, H. W. Stability of the domain interface contributes towards the catalytic function at the H-site of class alpha glutathione transferase A1-1, *Biochim Biophys Acta 1804*, 2228-2233.
163. Wallace, L. A., Burke, J., and Dirr, H. W. (2000) Domain-domain interface packing at conserved Trp-20 in class alpha glutathione transferase impacts on protein stability, *Biochim Biophys Acta 1478*, 325-332.
164. Schreiber, G., and Keating, A. E. (2011) Protein binding specificity versus promiscuity, *Curr Opin Struct Biol 21*, 50-61.
165. Meier, S., and Ozbek, S. (2007) A biological cosmos of parallel universes: does protein structural plasticity facilitate evolution?, *Bioessays 29*, 1095-1104.
166. Tokuriki, N., and Tawfik, D. S. (2009) Protein dynamism and evolvability, *Science 324*, 203-207.
167. Hammes, G. G., Benkovic, S. J., and Hammes-Schiffer, S. (2011) Flexibility, diversity, and cooperativity: pillars of enzyme catalysis, *Biochemistry 50*, 10422-10430.
168. Dirr, H. W., Little, T., Kuhnert, D. C., and Sayed, Y. (2005) A conserved N-capping motif contributes significantly to the stabilization and dynamics of the C-terminal region of class Alpha glutathione S-transferases, *J Biol Chem 280*, 19480-19487.
169. Freire, E., and Michael L. Johnson, L. B. (1994) [23] Statistical thermodynamic analysis of differential scanning calorimetry data: Structural deconvolution of heat capacity function of proteins, in *Methods in enzymology*, pp 502-530, Academic Press.
170. Freire, E. (1995) Differential scanning calorimetry, *Methods in molecular biology (Clifton, N.J 40*, 191-218.
171. Sturtevant, J. M. (1987) Biochemical applications of differential scanning calorimetry, *Annual review of physical chemistry*, 463-488.
172. Sturtevant, J. M. (1977) Heat capacity and entropy changes in processes involving proteins, *Proc Natl Acad Sci U S A 74*, 2236-2240.
173. Prabhu, N. V., and Sharp, K. A. (2005) Heat capacity in proteins, *Annual review of physical chemistry 56*, 521-548.
174. Privalov, P. L., and Dragan, A. I. (2007) Microcalorimetry of biological macromolecules, *Biophys Chem 126*, 16-24.
175. Freire, E., van Osdol, W. W., Mayorga, O. L., and Sanchez-Ruiz, J. M. (1990) Calorimetrically determined dynamics of complex unfolding transitions in proteins, *Annual review of biophysics and biophysical chemistry 19*, 159-188.
176. Pfeil, W., and Privalov, P. L. (1976) Thermodynamic investigations of proteins. I. Standard functions for proteins with lysozyme as an example, *Biophys Chem 4*, 23-32.
177. Pfeil, W., and Privalov, P. L. (1976) Thermodynamic investigations of proteins. II. Calorimetric study of lysozyme denaturation by guanidine hydrochloride, *Biophys Chem 4*, 33-40.
178. Pfeil, W., and Privalov, P. L. (1976) Thermodynamic investigations of proteins. III. Thermodynamic description of lysozyme, *Biophys Chem 4*, 41-50.
179. Xie, D., Bhakuni, V., and Freire, E. (1991) Calorimetric determination of the energetics of the molten globule intermediate in protein folding: apo-alpha-lactalbumin, *Biochemistry 30*, 10673-10678.
180. Chen, B. L., Baase, W. A., and Schellman, J. A. (1989) Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations, *Biochemistry 28*, 691-699.

181. Chen, B. L., and Schellman, J. A. (1989) Low-temperature unfolding of a mutant of phage T4 lysozyme. 1. Equilibrium studies, *Biochemistry 28*, 685-691.
182. Biltonen, R. L., and Freire, E. (1978) Thermodynamic characterization of conformational states of biological macromolecules using differential scanning calorimetry, *CRC critical reviews in biochemistry 5*, 85-124.
183. Freire, E., and Biltonen, R. L. (1978) Thermodynamics of transfer ribonucleic acids: the effect of sodium on the thermal unfolding of yeast tRNAPhe, *Biopolymers 17*, 1257-1272.
184. Munoz, V. (2007) Conformational dynamics and ensembles in protein folding, *Annu Rev Biophys Biomol Struct 36*, 395-412.
185. Kaya, H., and Chan, H. S. (2000) Polymer principles of protein calorimetric two-state cooperativity, *Proteins 40*, 637-661.
186. Munoz, V., and Sanchez-Ruiz, J. M. (2004) Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments, *Proc Natl Acad Sci U S A 101*, 17646-17651.
187. Callen, H. B., and Callen, H. B. (1985) *Thermodynamics and an introduction to thermostatistics*, 2nd ed., Wiley, New York.
188. Naganathan, A. N., Sanchez-Ruiz, J. M., and Munoz, V. (2005) Direct measurement of barrier heights in protein folding, *J Am Chem Soc 127*, 17970-17971.
189. Halskau, O., Jr., Perez-Jimenez, R., Ibarra-Molero, B., Underhaug, J., Munoz, V., Martinez, A., and Sanchez-Ruiz, J. M. (2008) Large-scale modulation of thermodynamic protein folding barriers linked to electrostatics, *Proc Natl Acad Sci U S A 105*, 8625-8630.
190. Garcia-Mira, M. M., Sadqi, M., Fischer, N., Sanchez-Ruiz, J. M., and Munoz, V. (2002) Experimental identification of downhill protein folding, *Science 298*, 2191-2195.
191. Packman, L. C., and Perham, R. N. (1986) Chain folding in the dihydrolipoyl acyltransferase components of the 2-oxo-acid dehydrogenase complexes from Escherichia coli. Identification of a segment involved in binding the E3 subunit, *FEBS Lett 206*, 193-198.
192. Robien, M. A., Clore, G. M., Omichinski, J. G., Perham, R. N., Appella, E., Sakaguchi, K., and Gronenborn, A. M. (1992) Three-dimensional solution structure of the E3-binding domain of the dihydrolipoamide succinyltransferase core from the 2-oxoglutarate dehydrogenase multienzyme complex of Escherichia coli, *Biochemistry 31*, 3463-3471.
193. Ibarra-Molero, B., Sanchez-Ruiz, J. M., Arrondo, J. L. R., and Alonso, A. (2006) Differential Scanning Calorimetry of Proteins: an Overview and Some Recent Developments in *Advanced Techniques in Biophysics*, pp 27-48, Springer Berlin Heidelberg.
194. Thorolfsson, M., Ibarra-Molero, B., Fojan, P., Petersen, S. B., Sanchez-Ruiz, J. M., and Martinez, A. (2002) L-phenylalanine binding and domain organization in human phenylalanine hydroxylase: a differential scanning calorimetry study, *Biochemistry 41*, 7573-7585.
195. Pabo, C. O., Sauer, R. T., Sturtevant, J. M., and Ptashne, M. (1979) The lambda repressor contains two domains, *Proc Natl Acad Sci U S A 76*, 1608-1612.
196. Dirr, H. W., and Wallace, L. A. (1999) Role of the C-terminal helix 9 in the stability and ligandin function of class alpha glutathione transferase A1-1, *Biochemistry 38*, 15631-15640.

197. Acchione, M. (2012) Unpublished data.
198. Arai, M., and Kuwajima, K. (1996) Rapid formation of a molten globule intermediate in refolding of alpha-lactalbumin, *Folding & design 1*, 275-287.
199. Kuwajima, K. (1996) The molten globule state of alpha-lactalbumin, *Faseb J 10*, 102-109.
200. Griko, Y. V., Freire, E., and Privalov, P. L. (1994) Energetics of the alpha-lactalbumin states: a calorimetric and statistical thermodynamic study, *Biochemistry 33*, 1889-1899.
201. Plaza del Pino, I. M., Ibarra-Molero, B., and Sanchez-Ruiz, J. M. (2000) Lower kinetic limit to protein thermal stability: a proposal regarding protein stability in vivo and its relation with misfolding diseases, *Proteins 40*, 58-70.
202. Halskau, O., Underhaug, J., Froystein, N. A., and Martinez, A. (2005) Conformational flexibility of alpha-lactalbumin related to its membrane binding capacity, *J Mol Biol 349*, 1072-1086.
203. Huberts, D. H., Venselaar, H., Vriend, G., Veenhuis, M., and van der Klei, I. J. (2010) The moonlighting function of pyruvate carboxylase resides in the non-catalytic end of the TIM barrel, *Biochim Biophys Acta 1803*, 1038-1042.
204. Zimmermann, J., Oakman, E. L., Thorpe, I. F., Shi, X., Abbyad, P., Brooks, C. L., 3rd, Boxer, S. G., and Romesberg, F. E. (2006) Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics, *Proc Natl Acad Sci U S A 103*, 13722-13727.
205. Sethi, D. K., Agarwal, A., Manivel, V., Rao, K. V., and Salunke, D. M. (2006) Differential epitope positioning within the germline antibody paratope enhances promiscuity in the primary immune response, *Immunity 24*, 429-438.
206. James, L. C., Roversi, P., and Tawfik, D. S. (2003) Antibody multispecificity mediated by conformational diversity, *Science 299*, 1362-1367.
207. Nussinov, R., and Ma, B. Protein dynamics and conformational selection in bidirectional signal transduction, *BMC biology 10*, 2.
208. Anastassiadis, T., Deacon, S. W., Devarajan, K., Ma, H., and Peterson, J. R. (2011) Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity, *Nat Biotechnol 29*, 1039-1045.
209. Denison, M. S., Soshilov, A. A., He, G., DeGroot, D. E., and Zhao, B. (2011) Exactly the same but different: promiscuity and diversity in the molecular mechanisms of action of the aryl hydrocarbon (dioxin) receptor, *Toxicol Sci 124*, 1-22.
210. Mikhaylova, M., Hradsky, J., and Kreutz, M. R. (2011) Between promiscuity and specificity: novel roles of EF-hand calcium sensors in neuronal Ca2+ signalling, *Journal of neurochemistry 118*, 695-713.
211. Jensen, R. A. (1976) Enzyme recruitment in evolution of new function, *Annual review of microbiology 30*, 409-425.
212. Steele, C. L., Crock, J., Bohlmann, J., and Croteau, R. (1998) Sesquiterpene synthases from grand fir (Abies grandis). Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of delta-selinene synthase and gamma-humulene synthase, *J Biol Chem 273*, 2078-2089.
213. Langenheim, J. H. (1994) Higher plant terpenoids: A phytocentric overview of their ecological roles, *Journal of Chemical Ecology 20*, 1223-1280.
214. Guengerich, F. P. (2001) Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity, *Chem Res Toxicol 14*, 611-650.

215. Park, H., Lee, S., and Suh, J. (2005) Structural and dynamical basis of broad substrate specificity, catalytic mechanism, and inhibition of cytochrome P450 3A4, *J Am Chem Soc 127*, 13634-13642.
216. Nebert, D. W., and Dieter, M. Z. (2000) The evolution of drug metabolism, *Pharmacology 61*, 124-135.
217. Koonin, E. V., Mushegian, A. R., Tatusov, R. L., Altschul, S. F., Bryant, S. H., Bork, P., and Valencia, A. (1994) Eukaryotic translation elongation factor 1 gamma contains a glutathione transferase domain--study of a diverse, ancient protein superfamily using motif search and structural modeling, *Protein Sci 3*, 2045-2054.
218. Martin, J. L. (1995) Thioredoxin--a fold for all reasons, *Structure 3*, 245-250.
219. Pemble, S. E., and Taylor, J. B. (1992) An evolutionary perspective on glutathione transferases inferred from class-theta glutathione transferase cDNA sequences, *Biochem J 287 (Pt 3)*, 957-963.

Appendix A: Data Analysis Code

Octave Script for Variable-barrier Analysis

```
%       Variable-barrier model for analysis of DSC data
%       Matthew Honaker 2012
%       The first function calculates the free energy functional and
%       probability density functions at T0 and as a function of T,
%       followed by calculation of the first and second (n=1, n=2) enthalpy
%       moments, followed by d<H> for each temperature.
%       See Munoz and Sanchez-Ruiz, PNAS, 101(51), 17646-17651
%

clear
%       This function is for fitting only

function [k]=moment(p)

R=0.00831447;                   %R, universal gas constant in kJ/mol/K
h1=-10000;                      %lower bound of enthalpy range
h2=10000;                       %upper bound of enthalpy range

z=[];
for T=[];

%calculate alpha for positive and negative enthalpy values
an=(p(1)*p(4))/2;
ap=(p(1)*(2-p(4)))/2;

%Integrate probability density at T0 with Gauss-Konrod quadrature - P(H|T0)
[P1_p]=quadgk(@(H) exp(-(-2.*p(3).*(H./ap).^2+abs(p(3)).*(H./ap).^4)./(R.*p(2))),0,h2);
[P1_n]=quadgk(@(H) exp(-(-2.*p(3).*(H./an).^2+abs(p(3)).*(H./an).^4)./(R.*p(2))),h1,0);
y1=[P1_p]+[P1_n];               %sum of integrations
c2=1./y1;                       %normalize

%Integrate probability density at T (with respect to T0) with Gauss-Konrod quadrature - P(H|T)
[P2_p]=quadgk(@(H) c2.*exp(-(-2.*p(3).*(H./ap).^2+abs(p(3)).*(H./ap).^4)./(R.*p(2))).*exp(-
(1./R).*(1./T.-1./p(2)).*H),0,h2);
[P2_n]=quadgk(@(H) c2.*exp(-(-2.*p(3).*(H./an).^2+abs(p(3)).*(H./an).^4)./(R.*p(2))).*exp(-
(1./R).*(1./T.-1./p(2)).*H),h1,0);
y2=[P2_p]+[P2_n];               %sum of integrations
c1=1./y2;                       %normalize

%Integrate 1st enthalpy moment at T with Gauss-Konrod quadrature - <H1>
[P3_p]=quadgk(@(H) H.*c2.*exp(-(-
```

```
2.*p(3).*(H./ap).^2+abs(p(3)).*(H./ap).^4)./(R.*p(2))).*exp(-(1./R).*(1./T.-1./p(2)).*H),0,h2);
[P3_n]=quadgk(@(H) H.*c2.*exp(-(-
2.*p(3).*(H./an).^2+abs(p(3)).*(H./an).^4)./(R.*p(2))).*exp(-(1./R).*(1./T.-1./p(2)).*H),h1,0);

%integrate 2nd enthalpy moment at T with Gauss-Konrod quadrature - <H2>
[P4_p]=quadgk(@(H) H.^2.*c2.*exp(-(-
2.*p(3).*(H./ap).^2+abs(p(3)).*(H./ap).^4)./(R.*p(2))).*exp(-(1./R).*(1./T.-1./p(2)).*H),0,h2);
[P4_n]=quadgk(@(H) H.^2.*c2.*exp(-(-
2.*p(3).*(H./an).^2+abs(p(3)).*(H./an).^4)./(R.*p(2))).*exp(-(1./R).*(1./T.-1./p(2)).*H),h1,0);

H1=(P3_p+P3_n).*c1;          %sum of integrations - returns 1st enthalpy moment
H2=(P4_p+P4_n).*c1;          %sum of integrations - returns 2nd enthalpy moment
z(end+1)=H2-H1.^2;           %create vector of d<H> values

k=[z]';                      %transpose vector (if needed)

endfor
endfunction

%
%       This function is called by the fitting routine
%       It takes the d<H> values calculate by the moment function and returns Cp_ex values
%       T are the temperatures in Kelvin for which there are Cp data points, and p is
%       a vector of fitting parameters (sum alpha, T0, beta)
%

function [Cp]=Cp_fit(t,p)
R=0.00831447;                %R, universal gas constant in kJ/mol/K
[Cp]=moment(p)./(R.*t.^2);   %calculate Cp_ex
endfunction

t=[];
data=[];

%Add the desired least-squares fitting routine here:
global verbose
verbose=1;
pin=[];                      %initial guesses
stol=0.001;
niter=50;
wt=ones(length(data),1);
dp=.001*(pin*0+1);
dFdp='dfdp';
options.bounds=[]
[f,p,cvg,iter,corp,covp,covr]=leasqr(t,data,pin,"Cp_fit",stol,niter,wt,dp,dFdp,options)
```

Python Script for Promiscuity Calculations

```
## Jvalue.py
## Calculate J-values from a set of inhibitor potencies and bit keys
## Abhinav Nath, Feb 2010
## Edits by Matthew Honaker, Apr 2011
import math
from scipy.io import read_array, write_array

# File contains substrate names and descriptor bitstrings - goes between quotes
inf = file('descriptor file name goes here.txt').readlines()

#Columns in enzyme data file, tab separated - goes between quotes
cyps = 'enzyme          names  go       here     tab       separated'.split()

#Number of compounds
print len(inf)

#Read descriptor bitstrings
descriptors = []
for line in inf:
        if not line[0] == '#':
                descriptors.append([line.split()[0], line.split()[1]])

#Return Tanimoto coeff of two bitstrings
def tanimoto(string1, string2):
        if not len(string1) == len(string2):
                return -1
        else:
                a = float(0)
                b = float(0)
                c = float(0)

        for i in range(len(string1)):
                if string1[i] == '1' and string2[i] == '1':
                        c += 1
                elif string1[i] == '1':
                        a += 1
                elif string2[i] == '1':
                        b += 1
        return c/(a+b+c)

#Returns I
def I(e):
        tot = sum(e)
        return -sum([(e[i]/tot)*math.log(e[i]/tot) for i in range(len(e))])/math.log(len(e))
```

```
#Returns J weighted by Tanimoto coeff normalized by d_set
def J(e, desc):
	ave_tans = []
	for i, d1 in enumerate(desc): ## Arithmetic mean
		s = 0
		for j, d2 in enumerate(desc):
			if not i == j:
				s +=  1 - tanimoto(d1[1], d2[1])
		ave_tans.append(s/(len(desc)-1))
	tot = sum(e)

#	for i in range(len(desc)):
#		print desc[i][0], ave_tans[i]/d_set(desc)
	return -len(e)*sum([(ave_tans[i]/d_set(desc))*(e[i]/tot)*math.log(e[i]/tot) for i in
range(len(e))])/(sum(ave_tans)*math.log(len(e)))

#Returns distance between the two most distant members of a set
def d_set(desc):
	k = 0.
	l = 0.
	for i in range(len(desc[0][1])):
		ones = [desc[j][1][i] for j in range(len(desc))].count('1')
		if ones > 0:
			if ones == len(desc):
				l += 1
			else:
				k += 1
	return k/(k+l)

#pots = [l.split() for l in file(' data file name ').readlines()]
pots = read_array('data_file_name_goes_here.txt')
for i in range(len(cyps)):
	print cyps[i], I(pots[:,i]), J(pots[:,i],descriptors), d_set(descriptors)

##Data columns are returned in columns enzyme names - Ivalues - Jvalues - d_set values
```

Appendix B: PubChem Substructure Fingerprints

Description from PubChem:
V1.3 http://pubchem.ncbi.nlm.nih.gov

The PubChem System generates a binary substructure fingerprint for chemical structures. These fingerprints are used by PubChem for similarity neighboring and similarity searching. A substructure is a fragment of a chemical structure. A fingerprint is an ordered list of binary (1/0) bits. Each bit represents a Boolean determination of, or test for, the presence of, for example, an element count, a type of ring system, atom pairing, atom environment (nearest neighbors), etc., in a chemical structure. The native format of the PubChem Substructure Fingerprint property is binary data with a four byte integer prefix, where this integer prefix indicates the length of the bit list.

Section 1: Hierarchic Element Counts - These bits test for the presence or count of individual chemical atoms represented by their atomic symbol.
Bit Position Bit Substructure

```
0 >= 4 H
1 >= 8 H
2 >= 16 H
3 >= 32 H
4 >= 1 Li
5 >= 2 Li
6 >= 1 B
7 >= 2 B
8 >= 4 B
9 >= 2 C
10 >= 4 C
11 >= 8 C
12 >= 16 C
13 >= 32 C
14 >= 1 N
15 >= 2 N
16 >= 4 N
17 >= 8 N
18 >= 1 O
19 >= 2 O
20 >= 4 O
21 >= 8 O
22 >= 16 O
23 >= 1 F
24 >= 2 F
25 >= 4 F
26 >= 1 Na
27 >= 2 Na
28 >= 1 Si
29 >= 2 Si
30 >= 1 P
31 >= 2 P
32 >= 4 P
33 >= 1 S
```

```
34 >= 2 S
35 >= 4 S
36 >= 8 S
37 >= 1 Cl
38 >= 2 Cl
39 >= 4 Cl
40 >= 8 Cl
41 >= 1 K
42 >= 2 K
43 >= 1 Br
44 >= 2 Br
45 >= 4 Br
46 >= 1 I
47 >= 2 I
48 >= 4 I
49 >= 1 Be
50 >= 1 Mg
51 >= 1 Al
52 >= 1 Ca
53 >= 1 Sc
54 >= 1 Ti
55 >= 1 V
56 >= 1 Cr
57 >= 1 Mn
58 >= 1 Fe
59 >= 1 Co
60 >= 1 Ni
61 >= 1 Cu
62 >= 1 Zn
63 >= 1 Ga
64 >= 1 Ge
65 >= 1 As
66 >= 1 Se
67 >= 1 Kr
68 >= 1 Rb
69 >= 1 Sr
70 >= 1 Y
71 >= 1 Zr
72 >= 1 Nb
73 >= 1 Mo
74 >= 1 Ru
75 >= 1 Rh
76 >= 1 Pd
77 >= 1 Ag
78 >= 1 Cd
79 >= 1 In
80 >= 1 Sn
81 >= 1 Sb
82 >= 1 Te
83 >= 1 Xe
84 >= 1 Cs
85 >= 1 Ba
86 >= 1 Lu
87 >= 1 Hf
88 >= 1 Ta
89 >= 1 W
90 >= 1 Re
```

```
91 >= 1 Os
92 >= 1 Ir
93 >= 1 Pt
94 >= 1 Au
95 >= 1 Hg
96 >= 1 Tl
97 >= 1 Pb
98 >= 1 Bi
99 >= 1 La
100 >= 1 Ce
101 >= 1 Pr
102 >= 1 Nd
103 >= 1 Pm
104 >= 1 Sm
105 >= 1 Eu
106 >= 1 Gd
107 >= 1 Tb
108 >= 1 Dy
109 >= 1 Ho
110 >= 1 Er
111 >= 1 Tm
112 >= 1 Yb
113 >= 1 Tc
114 >= 1 U
```

Section 2: Rings in a canonic Extended Smallest Set of Smallest Rings (ESSSR) ring set - These bits test for the presence or count of the described chemical ring system. An ESSSR ring is any ring which does not share three consecutive atoms with any other ring in the chemical structure.
Bit Position Bit Substructure

```
115 >= 1 any ring size 3
116 >= 1 saturated or aromatic carbon-only ring size 3
117 >= 1 saturated or aromatic nitrogen-containing ring size 3
118 >= 1 saturated or aromatic heteroatom-containing ring size 3
119 >= 1 unsaturated non-aromatic carbon-only ring size 3
120 >= 1 unsaturated non-aromatic nitrogen-containing ring size 3
121 >= 1 unsaturated non-aromatic heteroatom-containing ring size 3
122 >= 2 any ring size 3
123 >= 2 saturated or aromatic carbon-only ring size 3
124 >= 2 saturated or aromatic nitrogen-containing ring size 3
125 >= 2 saturated or aromatic heteroatom-containing ring size 3
126 >= 2 unsaturated non-aromatic carbon-only ring size 3
127 >= 2 unsaturated non-aromatic nitrogen-containing ring size 3
128 >= 2 unsaturated non-aromatic heteroatom-containing ring size 3
129 >= 1 any ring size 4
130 >= 1 saturated or aromatic carbon-only ring size 4
131 >= 1 saturated or aromatic nitrogen-containing ring size 4
132 >= 1 saturated or aromatic heteroatom-containing ring size 4
133 >= 1 unsaturated non-aromatic carbon-only ring size 4
134 >= 1 unsaturated non-aromatic nitrogen-containing ring size 4
135 >= 1 unsaturated non-aromatic heteroatom-containing ring size 4
136 >= 2 any ring size 4
137 >= 2 saturated or aromatic carbon-only ring size 4
138 >= 2 saturated or aromatic nitrogen-containing ring size 4
139 >= 2 saturated or aromatic heteroatom-containing ring size 4
140 >= 2 unsaturated non-aromatic carbon-only ring size 4
141 >= 2 unsaturated non-aromatic nitrogen-containing ring size 4
```

```
142 >= 2 unsaturated non-aromatic heteroatom-containing ring size 4
143 >= 1 any ring size 5
144 >= 1 saturated or aromatic carbon-only ring size 5
145 >= 1 saturated or aromatic nitrogen-containing ring size 5
146 >= 1 saturated or aromatic heteroatom-containing ring size 5
147 >= 1 unsaturated non-aromatic carbon-only ring size 5
148 >= 1 unsaturated non-aromatic nitrogen-containing ring size 5
149 >= 1 unsaturated non-aromatic heteroatom-containing ring size 5
150 >= 2 any ring size 5
151 >= 2 saturated or aromatic carbon-only ring size 5
152 >= 2 saturated or aromatic nitrogen-containing ring size 5
153 >= 2 saturated or aromatic heteroatom-containing ring size 5
154 >= 2 unsaturated non-aromatic carbon-only ring size 5
155 >= 2 unsaturated non-aromatic nitrogen-containing ring size 5
156 >= 2 unsaturated non-aromatic heteroatom-containing ring size 5
157 >= 3 any ring size 5
158 >= 3 saturated or aromatic carbon-only ring size 5
159 >= 3 saturated or aromatic nitrogen-containing ring size 5
160 >= 3 saturated or aromatic heteroatom-containing ring size 5
161 >= 3 unsaturated non-aromatic carbon-only ring size 5
162 >= 3 unsaturated non-aromatic nitrogen-containing ring size 5
163 >= 3 unsaturated non-aromatic heteroatom-containing ring size 5
164 >= 4 any ring size 5
165 >= 4 saturated or aromatic carbon-only ring size 5
166 >= 4 saturated or aromatic nitrogen-containing ring size 5
167 >= 4 saturated or aromatic heteroatom-containing ring size 5
168 >= 4 unsaturated non-aromatic carbon-only ring size 5
169 >= 4 unsaturated non-aromatic nitrogen-containing ring size 5
170 >= 4 unsaturated non-aromatic heteroatom-containing ring size 5
171 >= 5 any ring size 5
172 >= 5 saturated or aromatic carbon-only ring size 5
173 >= 5 saturated or aromatic nitrogen-containing ring size 5
174 >= 5 saturated or aromatic heteroatom-containing ring size 5
175 >= 5 unsaturated non-aromatic carbon-only ring size 5
176 >= 5 unsaturated non-aromatic nitrogen-containing ring size 5
177 >= 5 unsaturated non-aromatic heteroatom-containing ring size 5
178 >= 1 any ring size 6
179 >= 1 saturated or aromatic carbon-only ring size 6
180 >= 1 saturated or aromatic nitrogen-containing ring size 6
181 >= 1 saturated or aromatic heteroatom-containing ring size 6
182 >= 1 unsaturated non-aromatic carbon-only ring size 6
183 >= 1 unsaturated non-aromatic nitrogen-containing ring size 6
184 >= 1 unsaturated non-aromatic heteroatom-containing ring size 6
185 >= 2 any ring size 6
186 >= 2 saturated or aromatic carbon-only ring size 6
187 >= 2 saturated or aromatic nitrogen-containing ring size 6
188 >= 2 saturated or aromatic heteroatom-containing ring size 6
189 >= 2 unsaturated non-aromatic carbon-only ring size 6
190 >= 2 unsaturated non-aromatic nitrogen-containing ring size 6
191 >= 2 unsaturated non-aromatic heteroatom-containing ring size 6
192 >= 3 any ring size 6
193 >= 3 saturated or aromatic carbon-only ring size 6
194 >= 3 saturated or aromatic nitrogen-containing ring size 6
195 >= 3 saturated or aromatic heteroatom-containing ring size 6
196 >= 3 unsaturated non-aromatic carbon-only ring size 6
197 >= 3 unsaturated non-aromatic nitrogen-containing ring size 6
198 >= 3 unsaturated non-aromatic heteroatom-containing ring size 6
```

```
199 >= 4 any ring size 6
200 >= 4 saturated or aromatic carbon-only ring size 6
201 >= 4 saturated or aromatic nitrogen-containing ring size 6
202 >= 4 saturated or aromatic heteroatom-containing ring size 6
203 >= 4 unsaturated non-aromatic carbon-only ring size 6
204 >= 4 unsaturated non-aromatic nitrogen-containing ring size 6
205 >= 4 unsaturated non-aromatic heteroatom-containing ring size 6
206 >= 5 any ring size 6
207 >= 5 saturated or aromatic carbon-only ring size 6
208 >= 5 saturated or aromatic nitrogen-containing ring size 6
209 >= 5 saturated or aromatic heteroatom-containing ring size 6
210 >= 5 unsaturated non-aromatic carbon-only ring size 6
211 >= 5 unsaturated non-aromatic nitrogen-containing ring size 6
212 >= 5 unsaturated non-aromatic heteroatom-containing ring size 6
213 >= 1 any ring size 7
214 >= 1 saturated or aromatic carbon-only ring size 7
215 >= 1 saturated or aromatic nitrogen-containing ring size 7
216 >= 1 saturated or aromatic heteroatom-containing ring size 7
217 >= 1 unsaturated non-aromatic carbon-only ring size 7
218 >= 1 unsaturated non-aromatic nitrogen-containing ring size 7
219 >= 1 unsaturated non-aromatic heteroatom-containing ring size 7
220 >= 2 any ring size 7
221 >= 2 saturated or aromatic carbon-only ring size 7
222 >= 2 saturated or aromatic nitrogen-containing ring size 7
223 >= 2 saturated or aromatic heteroatom-containing ring size 7
224 >= 2 unsaturated non-aromatic carbon-only ring size 7
225 >= 2 unsaturated non-aromatic nitrogen-containing ring size 7
226 >= 2 unsaturated non-aromatic heteroatom-containing ring size 7
227 >= 1 any ring size 8
228 >= 1 saturated or aromatic carbon-only ring size 8
229 >= 1 saturated or aromatic nitrogen-containing ring size 8
230 >= 1 saturated or aromatic heteroatom-containing ring size 8
231 >= 1 unsaturated non-aromatic carbon-only ring size 8
232 >= 1 unsaturated non-aromatic nitrogen-containing ring size 8
233 >= 1 unsaturated non-aromatic heteroatom-containing ring size 8
234 >= 2 any ring size 8
235 >= 2 saturated or aromatic carbon-only ring size 8
236 >= 2 saturated or aromatic nitrogen-containing ring size 8
237 >= 2 saturated or aromatic heteroatom-containing ring size 8
238 >= 2 unsaturated non-aromatic carbon-only ring size 8
239 >= 2 unsaturated non-aromatic nitrogen-containing ring size 8
240 >= 2 unsaturated non-aromatic heteroatom-containing ring size 8
241 >= 1 any ring size 9
242 >= 1 saturated or aromatic carbon-only ring size 9
243 >= 1 saturated or aromatic nitrogen-containing ring size 9
244 >= 1 saturated or aromatic heteroatom-containing ring size 9
245 >= 1 unsaturated non-aromatic carbon-only ring size 9
246 >= 1 unsaturated non-aromatic nitrogen-containing ring size 9
247 >= 1 unsaturated non-aromatic heteroatom-containing ring size 9
248 >= 1 any ring size 10
249 >= 1 saturated or aromatic carbon-only ring size 10
250 >= 1 saturated or aromatic nitrogen-containing ring size 10
251 >= 1 saturated or aromatic heteroatom-containing ring size 10
252 >= 1 unsaturated non-aromatic carbon-only ring size 10
253 >= 1 unsaturated non-aromatic nitrogen-containing ring size 10
254 >= 1 unsaturated non-aromatic heteroatom-containing ring size 10
255 >= 1 aromatic ring
```

256 >= 1 hetero-aromatic ring
257 >= 2 aromatic rings
258 >= 2 hetero-aromatic rings
259 >= 3 aromatic rings
260 >= 3 hetero-aromatic rings
261 >= 4 aromatic rings
262 >= 4 hetero-aromatic rings

Section 3: Simple atom pairs – These bits test for the presence of patterns of bonded atom pairs, regardless of bond order or count.

Bit Position Bit Substructure

263 Li-H
264 Li-Li
265 Li-B
266 Li-C
267 Li-O
268 Li-F
269 Li-P
270 Li-S
271 Li-Cl
272 B-H
273 B-B
274 B-C
275 B-N
276 B-O
277 B-F
278 B-Si
279 B-P
280 B-S
281 B-Cl
282 B-Br
283 C-H
284 C-C
285 C-N
286 C-O
287 C-F
288 C-Na
289 C-Mg
290 C-Al
291 C-Si
292 C-P
293 C-S
294 C-Cl
295 C-As
296 C-Se
297 C-Br
298 C-I
299 N-H
300 N-N
301 N-O
302 N-F
303 N-Si
304 N-P
305 N-S
306 N-Cl
307 N-Br

```
308 O-H
309 O-O
310 O-Mg
311 O-Na
312 O-Al
313 O-Si
314 O-P
315 O-K
316 F-P
317 F-S
318 Al-H
319 Al-Cl
320 Si-H
321 Si-Si
322 Si-Cl
323 P-H
324 P-P
325 As-H
326 As-As
```

Section 4: Simple atom nearest neighbors – These bits test for the presence of atom nearest neighbor patterns, regardless of bond order (denoted by "~") or count, but where bond aromaticity (denoted by ":") is significant.

Bit Position Bit Substructure

```
327 C(~Br)(~C)
328 C(~Br)(~C)(~C)
329 C(~Br)(~H)
330 C(~Br)(:C)
331 C(~Br)(:N)
332 C(~C)(~C)
333 C(~C)(~C)(~C)
334 C(~C)(~C)(~C)(~C)
335 C(~C)(~C)(~C)(~H)
336 C(~C)(~C)(~C)(~N)
337 C(~C)(~C)(~C)(~O)
338 C(~C)(~C)(~H)(~N)
339 C(~C)(~C)(~H)(~O)
340 C(~C)(~C)(~N)
341 C(~C)(~C)(~O)
342 C(~C)(~Cl)
343 C(~C)(~Cl)(~H)
344 C(~C)(~H)
345 C(~C)(~H)(~N)
346 C(~C)(~H)(~O)
347 C(~C)(~H)(~O)(~O)
348 C(~C)(~H)(~P)
349 C(~C)(~H)(~S)
350 C(~C)(~I)
351 C(~C)(~N)
352 C(~C)(~O)
353 C(~C)(~S)
354 C(~C)(~Si)
355 C(~C)(:C)
356 C(~C)(:C)(:C)
357 C(~C)(:C)(:N)
358 C(~C)(:N)
```

```
359 C(~C)(:N)(:N)
360 C(~Cl)(~Cl)
361 C(~Cl)(~H)
362 C(~Cl)(:C)
363 C(~F)(~F)
364 C(~F)(:C)
365 C(~H)(~N)
366 C(~H)(~O)
367 C(~H)(~O)(~O)
368 C(~H)(~S)
369 C(~H)(~Si)
370 C(~H)(:C)
371 C(~H)(:C)(:C)
372 C(~H)(:C)(:N)
373 C(~H)(:N)
374 C(~H)(~H)(~H)
375 C(~N)(~N)
376 C(~N)(:C)
377 C(~N)(:C)(:C)
378 C(~N)(:C)(:N)
379 C(~N)(:N)
380 C(~O)(~O)
381 C(~O)(:C)
382 C(~O)(:C)(:C)
383 C(~S)(:C)
384 C(:C)(:C)
385 C(:C)(:C)(:C)
386 C(:C)(:C)(:N)
387 C(:C)(:N)
388 C(:C)(:N)(:N)
389 C(:N)(:N)
390 N(~C)(~C)
391 N(~C)(~C)(~C)
392 N(~C)(~C)(~H)
393 N(~C)(~H)
394 N(~C)(~H)(~N)
395 N(~C)(~O)
396 N(~C)(:C)
397 N(~C)(:C)(:C)
398 N(~H)(~N)
399 N(~H)(:C)
400 N(~H)(:C)(:C)
401 N(~O)(~O)
402 N(~O)(:O)
403 N(:C)(:C)
404 N(:C)(:C)(:C)
405 O(~C)(~C)
406 O(~C)(~H)
407 O(~C)(~P)
408 O(~H)(~S)
409 O(:C)(:C)
410 P(~C)(~C)
411 P(~O)(~O)
412 S(~C)(~C)
413 S(~C)(~H)
414 S(~C)(~O)
415 Si(~C)(~C)
```

Section 5: Detailed atom neighborhoods – These bits test for the presence of detailed atom neighborhood patterns, regardless of count, but where bond orders are specific, bond aromaticity matches both single and double bonds, and where "-", "=", and "#" matches a single bond, double bond, and triple bond order, respectively.

Bit Position Bit Substructure

```
416 C=C
417 C#C
418 C=N
419 C#N
420 C=O
421 C=S
422 N=N
423 N=O
424 N=P
425 P=O
426 P=P
427 C(#C)(-C)
428 C(#C)(-H)
429 C(#N)(-C)
430 C(-C)(-C)(=C)
431 C(-C)(-C)(=N)
432 C(-C)(-C)(=O)
433 C(-C)(-Cl)(=O)
434 C(-C)(-H)(=C)
435 C(-C)(-H)(=N)
436 C(-C)(-H)(=O)
437 C(-C)(-N)(=C)
438 C(-C)(-N)(=N)
439 C(-C)(-N)(=O)
440 C(-C)(-O)(=O)
441 C(-C)(=C)
442 C(-C)(=N)
443 C(-C)(=O)
444 C(-Cl)(=O)
445 C(-H)(-N)(=C)
446 C(-H)(=C)
447 C(-H)(=N)
448 C(-H)(=O)
449 C(-N)(=C)
450 C(-N)(=N)
451 C(-N)(=O)
452 C(-O)(=O)
453 N(-C)(=C)
454 N(-C)(=O)
455 N(-O)(=O)
456 P(-O)(=O)
457 S(-C)(=O)
458 S(-O)(=O)
459 S(=O)(=O)
```

Section 6: Simple SMARTS patterns – These bits test for the presence of simple SMARTS patterns, regardless of count, but where bond orders are specific and bond aromaticity matches both single and double bonds.

Bit Position Bit Substructure

460 C-C-C#C
461 O-C-C=N
462 O-C-C=O
463 N:C-S-[#1]
464 N-C-C=C
465 O=S-C-C
466 N#C-C=C
467 C=N-N-C
468 O=S-C-N
469 S-S-C:C
470 C:C-C=C
471 S:C:C:C
472 C:N:C-C
473 S-C:N:C
474 S:C:C:N
475 S-C=N-C
476 C-O-C=C
477 N-N-C:C
478 S-C=N-[#1]
479 S-C-S-C
480 C:S:C-C
481 O-S-C:C
482 C:N-C:C
483 N-S-C:C
484 N-C:N:C
485 N:C:C:N
486 N-C:N:N
487 N-C=N-C
488 N-C=N-[#1]
489 N-C-S-C
490 C-C-C=C
491 C-N:C-[#1]
492 N-C:O:C
493 O=C-C:C
494 O=C-C:N
495 C-N-C:C
496 N:N-C-[#1]
497 O-C:C:N
498 O-C=C-C
499 N-C:C:N
500 C-S-C:C
501 Cl-C:C-C
502 N-C=C-[#1]
503 Cl-C:C-[#1]
504 N:C:N-C
505 Cl-C:C-O
506 C-C:N:C
507 C-C-S-C
508 S=C-N-C
509 Br-C:C-C
510 [#1]-N-N-[#1]

511 S=C-N-[#1]
512 C-[As]-O-[#1]
513 S:C:C-[#1]
514 O-N-C-C
515 N-N-C-C
516 [#1]-C=C-[#1]
517 N-N-C-N
518 O=C-N-N
519 N=C-N-C
520 C=C-C:C
521 C:N-C-[#1]
522 C-N-N-[#1]
523 N:C:C-C
524 C-C=C-C
525 [As]-C:C-[#1]
526 Cl-C:C-Cl
527 C:C:N-[#1]
528 [#1]-N-C-[#1]
529 Cl-C-C-Cl
530 N:C-C:C
531 S-C:C-C
532 S-C:C-[#1]
533 S-C:C-N
534 S-C:C-O
535 O=C-C-C
536 O=C-C-N
537 O=C-C-O
538 N=C-C-C
539 N=C-C-[#1]
540 C-N-C-[#1]
541 O-C:C-C
542 O-C:C-[#1]
543 O-C:C-N
544 O-C:C-O
545 N-C:C-C
546 N-C:C-[#1]
547 N-C:C-N
548 O-C-C:C
549 N-C-C:C
550 Cl-C-C-C
551 Cl-C-C-O
552 C:C-C:C
553 O=C-C=C
554 Br-C-C-C
555 N=C-C=C
556 C=C-C-C
557 N:C-O-[#1]
558 O=N-C:C
559 O-C-N-[#1]
560 N-C-N-C
561 Cl-C-C=O
562 Br-C-C=O
563 O-C-O-C
564 C=C-C=C
565 C:C-O-C
566 O-C-C-N
567 O-C-C-O

568 N#C-C-C
569 N-C-C-N
570 C:C-C-C
571 [#1]-C-O-[#1]
572 N:C:N:C
573 O-C-C=C
574 O-C-C:C-C
575 O-C-C:C-O
576 N=C-C:C-[#1]
577 C:C-N-C:C
578 C-C:C-C:C
579 O=C-C-C-C
580 O=C-C-C-N
581 O=C-C-C-O
582 C-C-C-C-C
583 Cl-C:C-O-C
584 C:C-C=C-C
585 C-C:C-N-C
586 C-S-C-C-C
587 N-C:C-O-[#1]
588 O=C-C-C=O
589 C-C:C-O-C
590 C-C:C-O-[#1]
591 Cl-C-C-C-C
592 N-C-C-C-C
593 N-C-C-C-N
594 C-O-C-C=C
595 C:C-C-C-C
596 N=C-N-C-C
597 O=C-C-C:C
598 Cl-C:C:C-C
599 [#1]-C-C=C-[#1]
600 N-C:C:C-C
601 N-C:C:C-N
602 O=C-C-N-C
603 C-C:C:C-C
604 C-O-C-C:C
605 O=C-C-O-C
606 O-C:C-C-C
607 N-C-C-C:C
608 C-C-C-C:C
609 Cl-C-C-N-C
610 C-O-C-O-C
611 N-C-C-N-C
612 N-C-O-C-C
613 C-N-C-C-C
614 C-C-O-C-C
615 N-C-C-O-C
616 C:C:N:N:C
617 C-C-C-O-[#1]
618 C:C-C-C:C
619 O-C-C=C-C
620 C:C-O-C-C
621 N-C:C:C:N
622 O=C-O-C:C
623 O=C-C:C-C
624 O=C-C:C-N

625 O=C-C:C-O
626 C-O-C:C-C
627 O=[As]-C:C:C
628 C-N-C-C:C
629 S-C:C:C-N
630 O-C:C-O-C
631 O-C:C-O-[#1]
632 C-C-O-C:C
633 N-C-C:C-C
634 C-C-C:C-C
635 N-N-C-N-[#1]
636 C-N-C-N-C
637 O-C-C-C-C
638 O-C-C-C-N
639 O-C-C-C-O
640 C=C-C-C-C
641 O-C-C-C=C
642 O-C-C-C=O
643 [#1]-C-C-N-[#1]
644 C-C=N-N-C
645 O=C-N-C-C
646 O=C-N-C-[#1]
647 O=C-N-C-N
648 O=N-C:C-N
649 O=N-C:C-O
650 O=C-N-C=O
651 O-C:C:C-C
652 O-C:C:C-N
653 O-C:C:C-O
654 N-C-N-C-C
655 O-C-C-C:C
656 C-C-N-C-C
657 C-N-C:C-C
658 C-C-S-C-C
659 O-C-C-N-C
660 C-C=C-C-C
661 O-C-O-C-C
662 O-C-C-O-C
663 O-C-C-O-[#1]
664 C-C=C-C=C
665 N-C:C-C-C
666 C=C-C-O-C
667 C=C-C-O-[#1]
668 C-C:C-C-C
669 Cl-C:C-C=O
670 Br-C:C:C-C
671 O=C-C=C-C
672 O=C-C=C-[#1]
673 O=C-C=C-N
674 N-C-N-C:C
675 Br-C-C-C:C
676 N#C-C-C-C
677 C-C=C-C:C
678 C-C-C=C-C
679 C-C-C-C-C-C
680 O-C-C-C-C-C
681 O-C-C-C-C-O

682 O-C-C-C-C-N
683 N-C-C-C-C-C
684 O=C-C-C-C-C
685 O=C-C-C-C-N
686 O=C-C-C-C-O
687 O=C-C-C-C=O
688 C-C-C-C-C-C-C
689 O-C-C-C-C-C-C
690 O-C-C-C-C-C-O
691 O-C-C-C-C-C-N
692 O=C-C-C-C-C-C
693 O=C-C-C-C-C-O
694 O=C-C-C-C-C=O
695 O=C-C-C-C-C-N
696 C-C-C-C-C-C-C-C
697 C-C-C-C-C-C(C)-C
698 O-C-C-C-C-C-C-C
699 O-C-C-C-C-C(C)-C
700 O-C-C-C-C-C-O-C
701 O-C-C-C-C-C(O)-C
702 O-C-C-C-C-C-N-C
703 O-C-C-C-C-C(N)-C
704 O=C-C-C-C-C-C-C
705 O=C-C-C-C-C(O)-C
706 O=C-C-C-C-C(=O)-C
707 O=C-C-C-C-C(N)-C
708 C-C(C)-C-C
709 C-C(C)-C-C-C
710 C-C-C(C)-C-C
711 C-C(C)(C)-C-C
712 C-C(C)-C(C)-C

Section 7: Complex SMARTS patterns – These bits test for the presence of complex SMARTS patterns, regardless of count, but where bond orders and bond aromaticity are specific.
Bit Position Bit Substructure
713 Cc1ccc(C)cc1
714 Cc1ccc(O)cc1
715 Cc1ccc(S)cc1
716 Cc1ccc(N)cc1
717 Cc1ccc(Cl)cc1
718 Cc1ccc(Br)cc1
719 Oc1ccc(O)cc1
720 Oc1ccc(S)cc1
721 Oc1ccc(N)cc1
722 Oc1ccc(Cl)cc1
723 Oc1ccc(Br)cc1
724 Sc1ccc(S)cc1
725 Sc1ccc(N)cc1
726 Sc1ccc(Cl)cc1
727 Sc1ccc(Br)cc1
728 Nc1ccc(N)cc1
729 Nc1ccc(Cl)cc1
730 Nc1ccc(Br)cc1
731 Clc1ccc(Cl)cc1
732 Clc1ccc(Br)cc1
733 Brc1ccc(Br)cc1

734 Cc1cc(C)ccc1
735 Cc1cc(O)ccc1
736 Cc1cc(S)ccc1
737 Cc1cc(N)ccc1
738 Cc1cc(Cl)ccc1
739 Cc1cc(Br)ccc1
740 Oc1cc(O)ccc1
741 Oc1cc(S)ccc1
742 Oc1cc(N)ccc1
743 Oc1cc(Cl)ccc1
744 Oc1cc(Br)ccc1
745 Sc1cc(S)ccc1
746 Sc1cc(N)ccc1
747 Sc1cc(Cl)ccc1
748 Sc1cc(Br)ccc1
749 Nc1cc(N)ccc1
750 Nc1cc(Cl)ccc1
751 Nc1cc(Br)ccc1
752 Clc1cc(Cl)ccc1
753 Clc1cc(Br)ccc1
754 Brc1cc(Br)ccc1
755 Cc1c(C)cccc1
756 Cc1c(O)cccc1
757 Cc1c(S)cccc1
758 Cc1c(N)cccc1
759 Cc1c(Cl)cccc1
760 Cc1c(Br)cccc1
761 Oc1c(O)cccc1
762 Oc1c(S)cccc1
763 Oc1c(N)cccc1
764 Oc1c(Cl)cccc1
765 Oc1c(Br)cccc1
766 Sc1c(S)cccc1
767 Sc1c(N)cccc1
768 Sc1c(Cl)cccc1
769 Sc1c(Br)cccc1
770 Nc1c(N)cccc1
771 Nc1c(Cl)cccc1
772 Nc1c(Br)cccc1
773 Clc1c(Cl)cccc1
774 Clc1c(Br)cccc1
775 Brc1c(Br)cccc1
776 CC1CCC(C)CC1
777 CC1CCC(O)CC1
778 CC1CCC(S)CC1
779 CC1CCC(N)CC1
780 CC1CCC(Cl)CC1
781 CC1CCC(Br)CC1
782 OC1CCC(O)CC1
783 OC1CCC(S)CC1
784 OC1CCC(N)CC1
785 OC1CCC(Cl)CC1
786 OC1CCC(Br)CC1
787 SC1CCC(S)CC1
788 SC1CCC(N)CC1
789 SC1CCC(Cl)CC1
790 SC1CCC(Br)CC1

791 NC1CCC(N)CC1
792 NC1CCC(Cl)CC1
793 NC1CCC(Br)CC1
794 ClC1CCC(Cl)CC1
795 ClC1CCC(Br)CC1
796 BrC1CCC(Br)CC1
797 CC1CC(C)CCC1
798 CC1CC(O)CCC1
799 CC1CC(S)CCC1
800 CC1CC(N)CCC1
801 CC1CC(Cl)CCC1
802 CC1CC(Br)CCC1
803 OC1CC(O)CCC1
804 OC1CC(S)CCC1
805 OC1CC(N)CCC1
806 OC1CC(Cl)CCC1
807 OC1CC(Br)CCC1
808 SC1CC(S)CCC1
809 SC1CC(N)CCC1
810 SC1CC(Cl)CCC1
811 SC1CC(Br)CCC1
812 NC1CC(N)CCC1
813 NC1CC(Cl)CCC1
814 NC1CC(Br)CCC1
815 ClC1CC(Cl)CCC1
816 ClC1CC(Br)CCC1
817 BrC1CC(Br)CCC1
818 CC1C(C)CCCC1
819 CC1C(O)CCCC1
820 CC1C(S)CCCC1
821 CC1C(N)CCCC1
822 CC1C(Cl)CCCC1
823 CC1C(Br)CCCC1
824 OC1C(O)CCCC1
825 OC1C(S)CCCC1
826 OC1C(N)CCCC1
827 OC1C(Cl)CCCC1
828 OC1C(Br)CCCC1
829 SC1C(S)CCCC1
830 SC1C(N)CCCC1
831 SC1C(Cl)CCCC1
832 SC1C(Br)CCCC1
833 NC1C(N)CCCC1
834 NC1C(Cl)CCCC1
835 NC1C(Br)CCCC1
836 ClC1C(Cl)CCCC1
837 ClC1C(Br)CCCC1
838 BrC1C(Br)CCCC1
839 CC1CC(C)CC1
840 CC1CC(O)CC1
841 CC1CC(S)CC1
842 CC1CC(N)CC1
843 CC1CC(Cl)CC1
844 CC1CC(Br)CC1
845 OC1CC(O)CC1
846 OC1CC(S)CC1
847 OC1CC(N)CC1

848 OC1CC(Cl)CC1
849 OC1CC(Br)CC1
850 SC1CC(S)CC1
851 SC1CC(N)CC1
852 SC1CC(Cl)CC1
853 SC1CC(Br)CC1
854 NC1CC(N)CC1
855 NC1CC(Cl)CC1
856 NC1CC(Br)CC1
857 ClC1CC(Cl)CC1
858 ClC1CC(Br)CC1
859 BrC1CC(Br)CC1
860 CC1C(C)CCC1
861 CC1C(O)CCC1
862 CC1C(S)CCC1
863 CC1C(N)CCC1
864 CC1C(Cl)CCC1
865 CC1C(Br)CCC1
866 OC1C(O)CCC1
867 OC1C(S)CCC1
868 OC1C(N)CCC1
869 OC1C(Cl)CCC1
870 OC1C(Br)CCC1
871 SC1C(S)CCC1
872 SC1C(N)CCC1
873 SC1C(Cl)CCC1
874 SC1C(Br)CCC1
875 NC1C(N)CCC1
876 NC1C(Cl)CC1
877 NC1C(Br)CCC1
878 ClC1C(Cl)CCC1
879 ClC1C(Br)CCC1
880 BrC1C(Br)CCC1

Curriculum Vita

Matthew T. Honaker

EDUCATION

University of Washington, School of Pharmacy

- Ph.D. in Medicinal Chemistry, 2012
- ARCS Foundation Grant 2004-2007

Western Kentucky University

- M.S. in Chemistry 2004
- American Institute of Chemists Graduate Student Award 2004
- President's Scholar 2001-2004

Western Kentucky University

- B.S. in Recombinant Genetic Technology 2002

RESEARCH EXPERIENCE

Graduate Research Assistant, University of Washington, 2004-present

- Revealed links between the conformational heterogeneity and catalytic promiscuity using multiple glutathione transferase mutants.
- Characterized the active site conformational landscapes and stabilities of glutathione transferases through free energy barrier and partition function analyses of differential scanning calorimetry data.
- Examined protein dynamics using time-resolved fluorescence distributions and *in silico* methods.
- Performed kinetic assays with multiple substrates and isoforms of glutathione transferases.

Summer Internship, Amgen, 2007

- Conducted stability studies of biotheraputics through matrix-assisted laser desorption/ionization mass spectrometry analysis of H/D exchange rates.

Graduate Research Assistant, Western Kentucky University, 2002-2004

- Developed novel synthetic methods for the synthesis of tertiary and ditertiary phosphines.
- Identified and characterized compounds using hydrogen, carbon, and phosphorus NMR.
- Used inert atmosphere techniques for manipulation of air-sensitive compounds.

Undergraduate Research Assistant, Western Kentucky University, 2000-2001

- Quantified metal ion uptake by *Sesbania drummondii* with inductively coupled plasma mass spectrometry.
- Surveyed and identified new species of speleological microorganism populations via terminal restriction fragment length polymorphism analysis.

TEACHING EXPERIENCE

University of Washington
Teaching Assistant, Medicinal Chemistry 2006-2007

- Prepared and taught class review sections for pharmacy students

Western Kentucky University
Teaching Assistant, Organic Chemistry 2004

- Assisted in preparation and execution of laboratory section

Teaching Assistant, Chemistry 2002-2003

- Prepared and taught laboratory section

OTHER EXPERIENCE

- Completed NIH workshop: Global and Multi-Method Analysis of Macromolecular Interactions with SEDPHAT, 2011
- Representative, Graduate and Professional Student Senate University of Washington 2011-2012

PUBLICATIONS

Honaker, M.T.; Acchione, M.; Zhang, W.; Mannervik, B.; Atkins, W.M. Enzymatic detoxification and drug metabolism: the role of molten globule active sites. In preparation.

Honaker, M.T.; Acchione, M.; Sumida, J.P., Atkins, W.M. (2011) Ensemble perspective for catalytic promiscuity: calorimetric analysis of the active site conformational landscape of a detoxification enzyme. *Journal of Biological Chemistry*, *286*, 42770-42776.

Foti, R.S.; **Honaker, M**.; Nath, A.; Pearson, J.T.; Buttrick, B.; Isoherranen, N.; Atkins, W.M. (2011) Catalytic versus inhibitory promiscuity in cytochrome P450s: implications for evolution of new function. *Biochemistry 50*, 2387-2393.

Hou, L.; **Honaker, M.T.**; Shireman, L.M.; Balogh, L.M; Roberts, A.G.; Ng, K.C.; Nath, A.; Atkins, W.M. (2007) Functional promiscuity correlates with conformational heterogeneity in A-class glutathione S-transferases. *Journal of Biological Chemistry, 282*, 23264-23274.

Honaker, M.T.; Hovland, J.M.; Salvatore, R.N. (2007) The synthesis of tertiary and secondary phosphines and their applications in organic synthesis. *Current Organic Synthesis, 4*, 31-45.

Honaker, M.T.; Salvatore, R.N. (2005) A mild and efficient synthesis of disecondary phosphines using CsOH: efforts toward phosphine macrocycles. *Letters in Organic Chemistry*, *2*, 198-200.

Honaker, M.T.; Salvatore, R.N. (2004) A mild and efficient CsOH-promoted synthesis of ditertiary phosphines. *Phosphorous, Sulfur and Silicon and the Related Elements, 179*, 277-283.

Honaker, M.T.; Sandefur, B.J. Hargett, J.L.; McDaniel, A.L.; Salvatore, R.N. (2003) CsOH-promoted P-alkylation: A convenient and highly efficient synthesis of tertiary phosphines. *Tetrahedron Letters, 44*, 8373-8377.

PUBLISHED ABSTRACTS

Salvatore, R.N.; **Honaker, M.T.**; Sandefur, B.J.; Hargett, B. (2005) CsOH-promoted direct mono-p-alkylation of primary phosphines: A mild and highly chemoselective synthetic route toward secondary and disecondary phosphines. *Abstracts of Papers of the American Chemical Society, 230*, U3093-U3094.

McDaniel, A.L.; **Honaker, M.T.**; Davidson, K.; Pesterfield, L.L.; Salvatore, R.N. (2005) Synthesis and characterization of bis-diphosphine complexes with transition metals. *Abstracts of Papers of the American Chemical Society 229*, U1010.

Honaker, M.T.; Sandefur, B.J.; McDaniel, A.L.; Pesterfield, L.L.; Salvatore, R.N. (2003) CsOH-promoted P-alkylation: A convenient and highly efficient synthesis of tertiary phosphines and their metal complexes. *Abstracts of Papers of the American Chemical Society, 225*, U342.