The conformational stability of a detoxification enzyme widely used as a ion-protein affinity tag

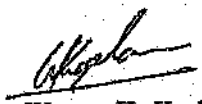
Warren H. Kaplan

A thesis submitted to the Faculty of Science, University of the Witwatersrand, in fulfilment of the requirements for the degree of Doctor of Philosophy.

Johannesburg, 1997

DECLARATION

I declare that this thesis is my own work. It is being submitted for the Degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.



Warren H. Kaplan

21st day of August 1997

ABSTRACT

A glutathione S-transferase (Sj26GST) from *Schistosoma japonicum*, which functions in the parasite's Phase II detoxification pathway, is expressed by the Pharmacia pGEX-2T plasmid and is widely used as a fusion-protein affinity tag. It contains all 217 residues of Sj26GST and an additional 9-residue peptide linker with a thrombin cleavage site at its C-terminus. Size-exclusion HPLC (SEC-HPLC) and SDS-PAGE studies indicate that purification of the homodimeric protein under nonreducing conditions results in the reversible formation of significant amounts of 160 -kDa and larger aggregates without a loss in catalytic activity. The basis for oxidative aggregation can be ascribed to the high degree of exposure of the four cysteine residues per subunit. The conformational stability of the dimeric protein was studied by urea- and temperature-induced unfolding techniques. Fluorescence-spectroscopy, SEC-HPLC, urea- and temperature-gradient gel electrophoresis, ultraviolet melting, differential scanning microcalorimetry, and enzyme activity were employed to monitor structural and functional changes. The unfolding data indicate the absence of thermodynamically stable intermediates and that the unfolding/refolding transition is a two-state process involving folded native dimer and unfolded monomer. The stability of the protein was found to be dependent on its concentration with a $\Delta G°(H_2O)$ = 26 $\pm$1.7 kcal/mol. The conformational stability was unchanged in the presence of the leading antischistosomal drug Praziquantel, which bound the protein with a Kd = 9 $\pm$1.8 μM. The strong relationship observed between the *m*-value and the size of the protein indicates that the amount of protein surface exposed to solvent upon unfolding is the major structural determinant for the dependence of the protein's free energy of unfolding on urea concentration. Thermograms obtained by differential scanning calorimetry also fitted to a two-state irreversible unfolding transition, both in the presence and absence of Praziquantel, with values of ΔC_p = 1779 cal $mol^{-1}K^{-1}$, ΔH_{cal} = 227 kcal/mol, ΔH_{VH} = 233 kcal/mol ($r = \Delta H_{VH} / \Delta H_{cal}$ = 1.02) and ΔS = 354 cal $mol^{-1}K^{-1}$. The low ΔC_p and ΔS, when compared with the theoretically determined values, implied that the thermal denaturation of Sj26GST did not result in complete unfolding of the protein.

A paper based on the results of this thesis has been published under the following reference:

Kaplan, W., Hüsler, P., Klump, H., Erhardt, J., Sluis-Cremer, N. and Dirr, H. (1997) Conformational stability of pGEX-expressed *Schistosoma japonicum* glutathione S-transferase: A detoxification enzyme and fusion-protein affinity tag. *Protein Sci.* 6, 399-406.

Dedicated to my parents and sisters

ACKNOWLEDGEMENTS

I would like to thank the following persons and institutions:

My supervisor, Professor Heini Dirr, whose contagious passion for this field triggered my interest in it. For his guidance, encouragement and enthusiasm, and for the privilege of working in his lab, I shall always be grateful.

The members of the Protein Structure-Function Research Programme, as well as the members of staff in the Department of Biochemistry at Wits.

Professor Horst Klump and Dr. Paul Hüsler of the Department of Biochemistry at the University of Cape Town for the use of their Differential Scanning Calorimeter (DSC), and for the deconvolution of the DSC curves.

Mr. David Amm for his expert construction of the electrophoresis equipment.

The University of the Witwatersrand and the South African Foundation for Research and Development for their financial support.

TABLE OF CONTENTS

ABBREVIATIONS

A	absorbance
A_{280}	absorbance at 280 nm
ΔASA	change in the solvent accessible surface area
C_m	chemically-induced midpoint of unfolding
CDNB	1-chloro-2,4-dinitrobenzene
ΔC_p	change in the heat capacity
DMSO	dimethyl sulfoxide
DSC	differential scanning calorimetry
DTNB	5,5'-dithiobis(2-nitrobenzoic acid)
DTT	dithiothreitol
EDTA	ethylenediaminetetra-acetic acid
FTP	file transfer protocol
ΔG	Gibbs free energy change
GSH	reduced glutathione
GST	glutathione S-transferase
Gdn.Cl	guanidinium chloride
G-site	glutathione-binding site
H-site	hydrophobic electrophile binding site
ΔH_{cal}	calorimetric enthalpy change
ΔH_{VH}	van't Hoff enthalpy change
IPTG	isopropylthiogalactoside
K_d	dissociation constant
LB	Luria-Bertani
LEM	linear extrapolation method
L-site	non-substrate ligand binding site
NTB	2-nitro-5-thiobenzoate
PBS	phosphate-buffered saline
pI	isoelectric point

PDB	protein databank
RMS	root mean square
rpm	revolutions per minute
ΔS	entropy change
ΔS_{conf}	conformational entropy change
SD	standard deviation
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SEC-HPLC	size-exclusion high performance liquid chromatography
Sj26GST	26 kDa GST from *Schistosoma japonicum*
TEMED	N,N,N',N'-tetramethylethylenediamine
TGGE	thermal-gradient gel electrophoresis
T_m	temperature-induced midpoint of unfolding
TSS	transformation and storeage solution
UGGE	urea-gradient gel electrophoresis
vs	versus
x g	times gravity

LIST OF FIGURES

LIST OF TABLES

FOREWORD

"structure, definite organised structure, becomes the essential characteristic of the physical universe, and this structural character accounts for many hitherto inexplicable phenomena" (Smuts, 1926).

In vivo proteins are manufactured on the ribosome, but shortly after Sanger had determined the amino acid sequence of the first protein (Sanger, 1952), Anfinsen (1956, see Anfinsen, 1973) found that it was possible for a protein to spontaneously refold *in vitro*. This implied that the final three-dimensional structure of the protein was determined by its amino acid sequence. What though did this three-dimensional structure look like? At that stage no structure of a protein was known. Proteins however, had been crystallised (Bernal and Crowfoot, 1934 in Perutz, 1992). The very fact that they could be crystallised, implied that each individual protein molecule within the crystal had assumed the same unique three-dimensional conformation as all its neighbours.

On 8 March 1958, the British journal Nature published the first three-dimensional structure of a protein (Kendrew, *et al.*, 1958). The protein was myoglobin, and in the article the authors noted: *"perhaps the most remarkable feature of the molecule are its complexity and its lack of symmetry. The arrangement seems to be almost totally lacking in the kind of regularities which one instinctively anticipates, and it is more complicated than has been predicted by any theory of protein structure"* (Kendrew, *et al.*, 1958).

The absence of this instinctively anticipated structure was in contrast to the proposed model of DNA by Watson and Crick (1953) which, being highly symmetrical and periodic made its method of replication self evident, leading the authors to comment: *"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material"* (Watson and Crick, 1953).

Close inspection of myoglobin provided no such structure-function detail. The elucidation of the second protein structure, that of haemoglobin (Perutz, *et al.*, 1960) was once again to be disappointing in terms of explaining the protein's function. At the time

Perutz commented that "*little can be said as yet about the relation between structure and function*" (Perutz, *et al.*, 1960).

From these exciting, yet humble beginnings the field of structural biology would develop to the stage where protein structures could be exploited for medical purposes (Riechmann, *et al.*, 1988) and ligand binding sites on proteins would become predictable (Laskowski *et al.*, 1996).

Having access to detailed protein structures (Kendrew *et al.*, 1958; Perutz *et al.*, 1960), now shone light on the end result of a protein's folding pathway. What though were the rules which had dictated the assumption of this unique conformation?

The challenge of learning these rules was then taken up by researchers in a diverse range of fields, in their quest to solve what was to become known as the "protein folding problem".

To date (mid 1997), the protein folding problem has not yielded. Great strides have however been made, and it is my hope that this study of a protein's structure-function relationship, and its dynamics of folding and unfolding, will in some small way contribute to a greater understanding of the protein folding problem. The protein under investigation is a glutathione S-transferase (GST), specifically the 26 kDa GST from *Schistosoma japonicum*.

CHAPTER 1 PROTEIN FOLDING AND STABILITY

1.1 INTRODUCTION

The protein folding problem has attracted researchers from all fields of science to its ranks (Karplus and Shakhnovich, 1992). As in many scientific fields the motivation for wanting to solve the problem varies from a rather simple desire to know how it is that a protein folds into its unique three-dimensional form, to those who desire to apply the new technology. What though are the possible applications?

In the last few years the human genome project has received enormous backing and the project is advancing rapidly. Once the human genome is mapped, i.e. every gene on all the chromosomes in a human cell identified, a diverse range of genetic applications are expected to follow. Importantly though, it is the protein which is coded for by each gene which ultimately results in either the correct or deficient manifestation of the genetic code. In order to carry out structure-function studies, it is therefore of far greater importance to have the three-dimensional structure of the protein rather than just the sequence of nucleotides which code for it. At present (mid 1997) just over 4000 structures have been deposited in the Protein Data Bank (PDB) (Bernstein, 1977). Many of these structures are variants of a similar structure, either falling within the same family of proteins or alternatively having merely had specific residues mutated. In excess of 50 000 proteins are thought to be coded for by the human genome, and consequently no structures will be known for the vast majority of them.

Furthermore, the conversion of substrates into products by enzymes is due to the specific addition or elimination of individual atoms to the substrate. Enzymes have therefore been recognised as being the smallest machines in existence. It is therefore hoped that by exploiting the rules used to solve the protein folding problem the field of nanotechnology will come into its own (Drexler, 1990). It is proposed that this technology will be able to facilitate the manipulation of individual atoms, applied from the rules governing protein folding, for the building of nano-sized robots. In turn these

nano-robots could then find potential uses in environments as diverse as a living cell to outer space (Stix, 1996).

The purpose of this chapter is firstly, to present some of the ideas about protein folding; from the very first suggestion that protein denaturation was due to the unfolding of the protein (Wu, 1931, the paper is reprinted in Edsall, 1995), to the development of both the kinetic and thermodynamic approaches of folding. The thermodynamic approach is then pursued and reasons discussed on why a protein's conformational stability is of interest to solving the folding problem. The nature of denaturation, and the states which may occur along the unfolding pathway are described. Since the differences between these states are due to the existence or absence of non-covalent interactions, the next section discusses the forces which result from these interactions. Finally some of the current experimental and theoretical methods used either for determining the protein's unfolding/refolding pathway, or its conformational stability are discussed.

1.1.1 Historical Development

In vivo proteins are manufactured on the ribosome where the nascent polypeptide chain folds into its final three-dimensional structure. In the 1950s it was assumed that accessory factors as well as the unfolded polypeptide chain were required for the correct folding of a protein (Baldwin and Eisenberg, 1987). In the case of a disulphide-containing protein, these factors were assumed to be essential for the correct formation of disulphide bonds between the specific cysteine residues. The reasoning was that most disulphide-containing proteins are found to unfold completely when their S-S bonds are reduced by a thiol. A given sulfhydryl (-SH) group in an unfolded polypeptide chain comprising 6-SH groups would, upon oxidation, be able to react with any of the other 5-SH groups. Any of the remaining 4-SH groups could then react with any of the last 3-SH groups etc. resulting in the following combination: 5 x 3 x 1, i.e. 15 possible S-S bonds. In order to ensure that only the correct disulphide bonds were formed, would require the positioning of the -SH groups through folding. The problem though is that the folded structure is not stable unless the S-S bonds are in-tact. In order to solve this

dilemma, one needs to turn to the experiments of Anfinsen (Anfinsen *et al.*, 1961).

1.1.2 Anfinsen's Experiment and the Thermodynamic Hypothesis

Until Anfinsen's experiments it was known that ribonuclease A (RNase A), contains 4 S-S bonds and undergoes a reversible unfolding/refolding reaction when the S-S bonds are left intact. Upon reduction of these disulphide bonds, the unfolding of RNase A is found to be complete, with reoxidation of the protein resulting in the formation of incorrectly formed S-S bonds (Baldwin and Eisenberg, 1987).
Anfinsen was able to show that it was possible to correctly pair S-S bonds when both a small amount of reducing and oxidising agent were simultaneously added to the protein. The reducing agent caused cleavage of the S-S bonds and the oxidising agent randomly reformed them. This random reformation of S-S bonds resulted in a reshuffling of the disulphides. Upon the formation of the correct disulphide, the protein was able to fold thus burying the S-S bonds. Since these disulphides were no longer accessible to the reducing agent any further reshuffling was prevented from occurring.

Because these experiments were carried out *in vitro*, Anfinsen concluded that no accessory factors were necessary for correct folding of the protein. The reshuffling reaction may even be carried out by the protein itself, where the -SH groups of cysteine residues attack S-S bonds in the same protein molecule via a disulphide interchange reaction. Protein disulphide isomerase was found however to increase the rate of correct folding, and more recently other chaperons have been identified which aid only in the rate of refolding (provided the chaperon is a true catalyst), (see Dill, 1990; Freedman, 1992 and the references within). It was therefore suggested that no accessory factors are involved in folding, rather the information needed for the correct folding of a protein is contained entirely in its amino acid sequence (for a review see Baldwin and Eisenberg, 1987). The S-S bonds themselves are regarded only as fasteners for reinforcing the correctly folded structure.

The fact that ribonuclease and indeed many other proteins can be denatured and then renatured to full activity (see Dill, 1990 and the references therein) would imply the

following: denaturation of the large number of native protein molecules in solution results in unfolding to a very large number of different conformations of nearly equal free energy (Dill, 1987). Thus refolding must begin from this enormous number of initial conformations of the different molecules in solution, along very many folding pathways, which are thought to more closely resemble funnels than tunnels in configurational space (see e.g. Mirny *et al.*, 1996). Yet the end result of refolding is the convergence to the same native structure. This represents the state of the lowest free energy (Dill, 1987), implying therefore that the native structure is thermodynamically stable. Proof of thermodynamic stability would therefore only require that the native structure is only a function of state and does not depend on the initial conditions leading to that state. By definition, such a state would be at the global minimum of free energy relative to all other states accessible on that time scale (Dill, 1990). Hence the interest in attempting to predict global energy minima for proteins (see e.g. Scheraga, 1996), and the need to determine the conformational stability of a protein.

1.1.3 Two-State Folding and the Levinthal Paradox

Early studies suggested that the denaturation of a globular proteins is an all-or-nothing process. Just as in a phase transition which a liquid undergoes to form a gas, and the transition comprises only molecules in either the liquid or gas phase, so too in the transition region of a protein's denaturation there were thought to be only two states i.e., folded and unfolded. This implied that any intermediate states, which if present along the protein's unfolding pathway would be so sparsely populated that they are not measurable. The first evidence to support this view was the coincidence of denaturation curves measured using different techniques (Lumry *et al.*, 1966; Tanford, 1968). The definitive evidence was however obtained only with the introduction of the scanning microcalorimeter (Privalov and Khechinashvili, 1974; Privalov, 1979). The calorimetric measurements showed that a co-operative folding unit coinciding with a single domain protein would denature without any visible intermediate states (Privalov, 1982).

Levinthal (1968), took the 2-state model to its logical conclusion by asking how slowly folding would occur if there were no folding intermediates, and folding took place via a stochastic search. By considering only the polypeptide backbone of a 100 amino acid residue protein, with only two rotable bonds (Φ and Ψ), and three possible conformations about each rotable bond, there would be 3^{200} possible conformations of the backbone. By assuming that bond rotation occurs at a rate of 10^{13} sec^{-1}, then the time t required to sample each polypeptide conformation is $t = 3^{2n}/(2n \times 10^{13}) = 1.3 \times 10^{80}$ sec, or 4×10^{72} years (Wetlaufer, 1973).

This calculation has been criticised (Dill, 1987) as it ignores the excluded volume of the chain and also assumes the conformations will be explored randomly as the molecule tends towards its native structure (Dill, 1985; 1987). The search would only be random if all the conformations were of equal energy, as would only occur in the case of a homopolomer (Dill, 1985; 1987). Despite this argument reducing the folding time by many orders of magnitude, it still does not explain the millisecond time scale of folding for many small proteins.

Levinthal's calculation therefore prompted the argument that folding was in fact kinetically controlled because the time scale of folding does not allow for the search through all possible conformations. As a result there would be no guarantee of reaching the thermodynamically most stable state. The high rate of folding was thought to be due to the proposed existence of folding intermediates (Privalov, 1996 and the references therein), which would direct the folding process along a defined pathway.

1.1.4 The Thermodynamic Versus the Kinetic Hypotheses

The two hypotheses, i.e. the thermodynamic hypothesis which stemmed from the work of Anfinsen (1973) and the kinetic hypothesis which arose from the calculations of Levinthal (1968) have led to two schools of thought as to what dictates the folding process. The first view attributes thermodynamic stability as determining the conformation of the folded protein. A consequence of this assumption would then be that the kinetic pathway of folding would result in the most stable product. Using a process

of energy minimisation it should in principle be possible to predict the structure of the folded protein from its amino acid sequence (see e.g. Poland and Scheraga, 1970). The second school of thought holds that the conformation of the final product is dictated by the kinetic pathway of folding. Folding is assumed to occur in the fastest route available, and the result would be the assumption of a three-dimensional structure dictated by the rate at which it was achieved, without necessarily being the thermodynamically most stable structure (see Fersht, 1995). A result of the kinetic approach to folding could then result in the protein being trapped in a local energy minimum, as opposed to the global minimum suggested by the kinetic approach.

Different approaches have been used in support of both of these schools of thought. The mutation of residues which provide no stability to the folded protein, yet affect the folding pathway, provide support for the kinetic approach. On the other hand if the thermodynamic view is correct then these mutations should have no effect.

At present, no evidence exists to reject the thermodynamic hypothesis (e.g. Dill, 1987; Scheraga, 1996), and that the native structure of a protein can be solely explained by means of this hypothesis.

1.2 MEASUREMENT OF PROTEIN STABILITY

Interest in protein stability stems from two main reasons; firstly, the need for the biotechnological industry to engineer proteins which are stable at above average temperatures (see e.g. Klibanov, 1983) or in environments traditionally regarded as being hostile to native protein structure (Zaks and Klibanov, 1984). The second reason stems from the thermodynamic approach to protein folding, where in order to understand how a protein acquires its global energy minimum it is essential to understand how stability is determined.

From the initial studies on protein unfolding, it was determined that proteins unfold from their native three-dimensional conformation to that of the denatured state. By assuming that this "denatured" state has no structure, it then becomes possible to measure the free energy change for the entire folding reaction by determining the

equilibrium constant K for the reaction, i.e. the ratio of the fraction of molecules which are unfolded f_U to the fraction of molecules which are in the native state, $1\text{-}f_U$, by measuring K as a function of denaturant e.g. temperature, pH, or urea.

Proteins exhibiting a two-state unfolding pattern appear to unfold in a highly co-operative sigmoidal curve. Initial work conducted on the energetics of unfolding implied that the much less structured state, usually designated the "unfolded state" or "denatured state," is an ensemble of many poorly structured conformations that behave as a single continuous distribution of microstates as described in Figure 1.1. From this observation emerged the two-state approximation (Lumry *et al.*, 1966), which allowed the denaturation reaction to be formally described by a single equilibrium constant K. Hence the stability of a protein to reversible denaturation may be assigned a unique value, the free energy difference between the two states, denatured (D) state and the native (N) state, i.e. $\Delta G_D = G_D - G_N$.

In order to carry out a conformational stability study of a protein, the first step is the choice of denaturant and the conditions. The pH may be kept low in order to prevent oxidation of the sulfhydryl groups, and to prevent any ionic effects, urea is used as a denaturant in place of Gdn.Cl. The next step is determining the reversibility of unfolding. This may be done by comparing the transition curves when one begins with either a folded or unfolded protein. Alternatively the denaturant may be diluted out to a fixed concentration, and the reversibility of the denaturation process compared to that of a reference. In addition the time to reach equilibrium should also be ascertained. Homogeneity, without the formation of any precipitate through aggregation should also be ascertained. This is easily done by Raleigh scatter.

For most proteins however the denatured state is insoluble, and most of the physical techniques available for characterising it (when in solution) are relatively insensitive for detecting structure with a highly flexible dynamic character (Shortle, 1996). As a result of no contradictory evidence protein chemists accepted the assumption that the denatured state is a featureless random coil-like structure (Shortle, 1996). The advantage of this assumption is that it simplifies the interpretation of experimental data.

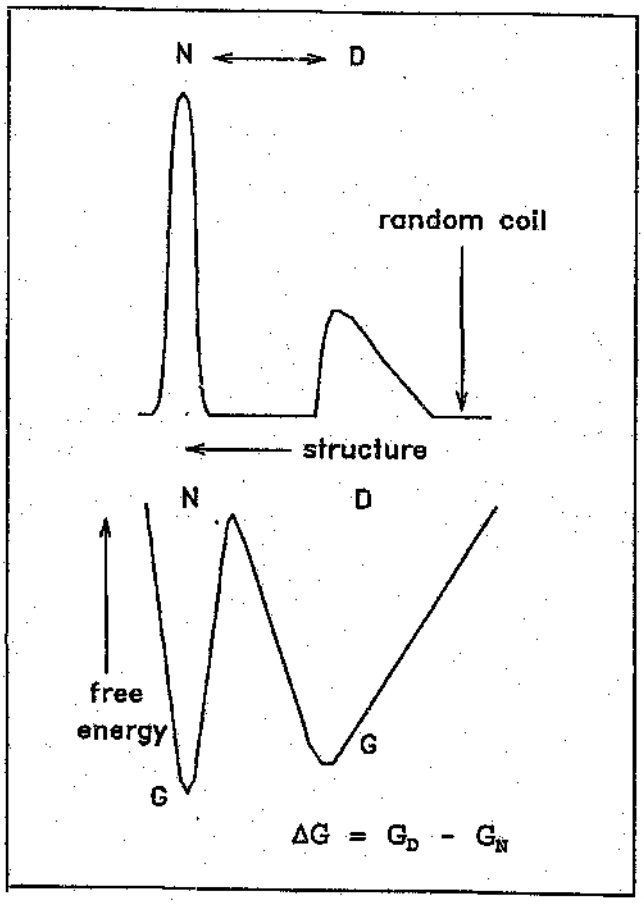


Figure 1.1. The reversible denaturation reaction: an equilibrium between the native state (N) and the denatured state (D), i.e. two macrostates or the distribution of microstates. The stability of a protein is determined by the free energy difference between these two states $\Delta G = G_D - G_N$. When the stability of a protein declines, the change in stability can result from and increase in G_N ($\Delta G_N > 0$) or from a decrease in G_D ($\Delta G_D < 0$), or a combination of such effects (adapted from Shortle, 1996).

By definition, a random coil is an energetically well-defined reference state in which no side chain-side chain interactions occur (Shortle, 1996). Hence, the closer the denatured state is to the random coil the larger is the role played by the energetics of the native state. Effectively, the experimentally measured thermodynamic parameters reflect the entire process of protein folding, with the sum total of interactions in the native state supplying all the free energy needed to drive the formation of structure (Shortle, 1996).

1.3 DENATURATION AND THE DIFFERENT PROTEIN STATES

The two-state approximation implies the presence of only a native and unfolded state separated by the denaturation process. In this section the process of denaturation and the above mentioned two states are discussed.

1.3.1 Denaturation

Numerous denaturants are used to study the unfolding of proteins. These include heat, cold, pressure, acid, alkali, or organic denaturants such as urea or Gdn.Cl. The unfolding of a protein has been investigated by both computational and experimental studies. Recently, the data has begun to converge (Schiffer and Dötsch, 1996) and while numerous suggestions have been proposed as to the nature of denaturation it has been suggested that the primary reason for protein unfolding is due to a disruption of the balance of forces between a protein's interaction with itself and its environment. (Schiffer and Dötsch, 1996). This disruption in the balance of forces is thought to be due to the perturberance in the normal water structure around a protein, with the decrease in the normal water-water interactions, resulting in an increase in the relative interaction of water with protein, which in turn may initiate a protein's unfolding.

In concentrated Gdn.Cl and urea solutions, normal water structure no longer exists (Collins and Washbaugh, 1985; Pace *et al.*, 1992). The urea-water interaction is weaker than that of the water-water interaction. This reduction in normal water-water interactions disrupts the dynamic network of hydrogen bonds that exist in pure water.

Since water molecules are not maximally involved in hydrogen bonding, they are now able to interact more strongly with solutes. If the solute is a protein, the water molecules would now be able to compete for its internal hydrogen bonds, and by so doing initiate protein unfolding.

Urea is known via NMR to interact directly with the aliphatic side chains of a denatured protein (Dötsch *et al.*, 1995 and the references therein), and has also been deduced via computational means to interact with aromatic side chains (Duffy *et al.*, 1993). It is however thought that this direct interaction with the protein does not initiate unfolding. Rather, urea and Gdn.Cl are thought to have a secondary role, i.e. by binding to the hydrophobic residues of the protein they aid in maintaining the solubility of the unfolded protein (Schiffer and Dötsch, 1996). This is in contrast to other chaotropes such as sodium thiocyanate, where no direct protein-denaturant interaction occurs which results in the precipitation of the protein (Schiffer and Dötsch, 1996).

Temperature is thought to promote denaturation of the protein by destabilising the native state via its action on that state (Shortle, 1996). The thermal denaturation at higher temperatures is attributable to the gain in conformational freedom of the chain being more advantageous than the gain in interactions of the nonpolar contacts (Dill, 1990). Because of its large entropy, the denatured state decreases rapidly in free energy as the temperature is raised ($-T\Delta S$). Some proteins begin to denature when the temperature is lowered to near 0°C (Schellman, 1987). In this situation, it is the lower enthalpy of the denatured state (formed by clathrates around the exposed hydrophobic chains) which shifts the equilibrium away from the native state (Shortle, 1996).

1.3.2 Native State

The primary structure of a protein is defined as being the actual sequence of amino acids which comprise its polypeptide chain. With few exceptions peptide groups are constrained by resonance effects to assume a planar, trans conformation. Steric interactions further limit the conformations of the polypeptide backbone by restricting the angles of Φ (Cα-N bond) and Ψ (Cα - C bond), to three small regions in the

Ramachandran diagram. The diagram provides a graphic representation of the sterically permitted and forbidden conformations, where the non-bonding interatomic distance is either within (permitted) or less than (forbidden) van der Waals distances. Glycine residues, not having Cβ atoms are not constrained to these regions.

The protein's secondary structure is defined as the local conformation of its backbone, the specific folding patterns which are identified by secondary structure include: helices, pleated sheets and turns.

The α-helix whose conformation angles fall within the allowed regions of the Ramachandran diagram, is held together by hydrogen bonds. The 3_{10} helix, which is more tightly coiled than that of the α-helix, lies in a mildly forbidden region of the Ramachandran diagram. These helices occur less frequently, usually as a single turn termination of α-helices.

β-sheets are found to occur both in parallel and in an antiparallel fashion. In both these instances, the two or more fully extended polypeptide chains associate, forming hydrogen bonds between the neighbouring chains. These β-sheets have a right-handed curl when viewed along their polypeptide chains. Frequently the chain reverses its direction through a β-bend.

Other arrangements of the polypeptide chain are more difficult to describe but are no less ordered than the α or β structures found in the rest of the protein.

The major secondary structure elements of α-helices and β-sheets occur in most globular proteins. The side chain location varies with polarity, the nonpolar residues occur predominantly in the interior of a protein, out of contact with the aqueous solvent. The charged polar residues are located on the protein surface, with the uncharged polar groups located both in the interior of the protein as well as on its surface. The globular proteins are compact, with crystal structures indicating the presence of little space inside, so that water is largely excluded from the interior. The micelle like arrangement of the side chains (polar groups outside, nonpolar groups inside) has led to the analogy been drawn with oil droplets. The packing density of the internal regions of globular proteins is however in the same range as molecular crystals of small organic molecules (Privalov,

1995). Implying therefore that the interior of a protein is very efficiently packed. The ability of hydrogen bond donors to find acceptors, is due to the fact that most hydrogen bonding partners are on residues in very close sequence to each other. Thus explaining why the majority of hydrogen bonds are formed by backbone N-H groups. The majority of bonds in protein side chains, including those occupying protein cores, almost invariably have low-energy staggered torsion angles. Therefore, despite the profusion of intramolecular interactions, the interior side chains adopt relaxed conformations (Voet and Voet, 1995).

1.3.2.1 Subunit interactions

The multiple polar and nonpolar groups in proteins make them notorious for sticking to virtually anything. One exception though is that they do not stick to other proteins except under very special circumstances. Proteins of molecular masses in excess of 100 kDa usually comprise more than one subunit, where these subunits are found to associate in a very specific manner. Several reasons have been proposed why multisubunit proteins are so common. In structural proteins such as collagen fibers the advantage of having the large protein comprising an assembly of identical subunits rather than one large molecule is analogous to the use of prefabricated components in building construction, where defective components may be replaced without the need for replacing the entire structure (Voet and Voet, 1990). From a genetic point of view the edifice is merely translated from a single gene, or in the case of non-identical subunits, from a few small genes of average size. In the case of enzymes, increasing the protein's size tends to better fix the positions of the atoms forming the enzyme's active site (Voet and Voet, 1990). Assuming each subunit has an active site, then increasing the size of the enzyme via the association of the two subunits results in greater efficiency, assuming both active sites are accessible to the substrates simultaneously. A multisubunit protein may consist of identical or non-identical polypeptide chains which associate via an interface which closely resembles the interior of a single protein subunit. The contact region contains closely packed nonpolar side chains, hydrogen bonds involving the

polypeptide backbones and their side chains. In the majority of oligomeric proteins (i.e. proteins containing identical subunits) the subunits are symmetrically arranged, whereby the subunits occupy geometrically equivalent positions in the oligomer. This therefore implies that each monomer has exhausted its capacity to bind to other monomers; otherwise higher oligomers would form. As a result of this limited binding capacity, the individual monomers pack about a single point to form a closed shell. No mirror symmetry can exist in proteins since this would require the naturally occurring L-amino acids to be converted into D-amino acids. Since D-amino acids do not occur naturally the only type of symmetry occurring in proteins is that of rotational symmetry. Numerous types of rotational symmetry occur in proteins, the most simple being that of cyclic symmetry. Here the subunits are related by a single axis of rotation. Proteins with a 2, 3, ...n-fold rotational axes are said to have C_2, C_3,...C_n symmetry respectively. An oligomer with C_2 symmetry would therefore consist of 2 subunits that are related by 360° /2 = 180°.

1.3.3 The Denatured State

From the original work of Tanford (1968;1970) evidence began to mount that the denatured state was not a random coil. In 6M Gdn.Cl the hydrodynamic radii of a number of small proteins agreed well with the theoretical estimates of the values of random coils. In no way however can a random coil found in e.g. 6M urea emulate conditions *in vivo*. Polypeptide chains swell and become less structured as solvent-side chain interactions become more favourable (Uversky and Ptitsyn, 1994; Dill and Shortle, 1991). More recent experimentation using small-angle X-ray scattering has shown that even in high concentrations of denaturant, significant amounts of residual structure may still be present (Kataoka *et al.*, 1993).

1.3.4 Intermediate States in Protein Folding

Since the rate of protein folding and the yield of a correctly folded conformation

rules out the possibility that folding is a simple stochastic process (Levinthal, 1968;1969), the process has to be directed. Thereby implying that it moves through a series of intermediate states with decreasing Gibbs free energies (Privalov, 1996). Due to these states being largely unstable and short lived, they are very difficult to characterise. Under certain conditions however, some proteins have been observed to contain states which are neither native nor completely unfolded (Tanford, 1968; Ptitsyn, 1992). What however are these intermediates? Are they merely incorrectly folded proteins trapped in a potential well or artifacts of folding? Alternatively, they may in fact be true intermediates in the folding pathway comprising a global macroscopic state of the whole protein. This would exclude a state in which only part of the molecule is folded, which would rather be representative of an independently folding domain.

It has been suggested (e.g. Dill and Stigter, 1995; Privalov, 1996) that the folding of a polypeptide chain starts by the collapse of the polymer into a compact liquid-like globule from which it later acquires the native conformation via the adjustment of its groups. While the collapse of the polypeptide into the compact globule is expected to be very fast, the adjustment of the side chains in the very viscous condensed phase would be expected to be much slower (Privalov, 1996). The collapse of the chain, would result in a significant reduction of the configurational phase space, dramatically limiting the number of folding options, and therefore was thought to be able to solve Levinthal's paradox (Ptitsyn, 1973; 1992; 1995a;b). As a result the hypothesis of a "molten globule" state gained support from the very first observations which supported this idea (Kuwajima, 1977).

Initially the molten globule was assumed to be a collapsed state of the polypeptide chain without fixed, specific long range interactions (Privalov, 1996). The collapsed state was thought to be maintained by hydrophobic interactions, i.e. the external pressure of water that does not penetrate inside the globule (Ptitsyn, 1992; Dill, 1990; Dill and Stigter, 1995). Furthermore, it was suggested that this compaction would promote a helical conformation of the chain (Dill, 1990; Chan and Dill, 1991). Thus the molten globule appeared to be a highly dynamic compact state of the polypeptide chain, without the tertiary structure but with some helicity, but not necessarily related to the secondary

structure of the native protein.

Later, water was "introduced" to this "dry molten globule," resulting in the formation of the "wet molten globule" and then the "swollen molten globule" (Finkelstein and Shaknovich, 1989). When it was suggested that a polypeptide chain in the molten globule had a "native-like fold" the "highly ordered" molten globule was introduced (Ptitsyn, 1995b; Peng and Kim, 1994; Redfield *et al.*, 1994).

A thermodynamic state of a macroscopic system is understood to be a state that differs significantly from all other states in its extensive thermodynamic parameters; such as enthalpy, entropy, volume, etc. (Privalov, 1996). Transition from one thermodynamic state to another therefore proceeds with a drastic change of these parameters. Such a change of state is termed a first-order phase transition (Privalov, 1996). If one were then to consider a protein macromolecule as a macroscopic system, then the separation of the native state from the denatured state (in a single domain protein) by a highly cooperative, two-state transition, could also be regarded as a first-order phase transition, since it proceeds with drastic changes in enthalpy, entropy and volume (Privalov, 1979, 1989, 1996). While the molten globule is known to be separated from the native state of the protein by a first-order phase transition, it is unclear as to whether it is separated from the unfolded state by a first-order phase transition (Privalov, 1996). According to statistical mechanical studies, the collapse of a polymer chain into a molten globule should result in a second order phase transition, accompanied by changes in the heat capacity and compressibility (Lifshitz *et al.*, 1978; Shakhnovich and Finkelstein, 1989; Finkelstein and Shankhnovich, 1989; Bryngleson and Wolynes, 1987).

Recently, the idea of a molten globule has been criticised (Privalov, 1996). The fact that the molten globule and unfolded protein move in different peaks in a size-exclusion chromatographic experiment (Ptitsyn and Uverski, 1994), would imply that their rates of interconversion are very slow (Privalov, 1996). Hence, the molten globule state could not be regarded as dynamic with a highly mobile conformation. Finally the native-like fold of a polypeptide chain assumes the existence of many fixed long-range interactions, the concept of describing a state with such attributes as molten i.e. liquefied state seems wrong (Privalov, 1996).

1.4 NON-COVALENT INTERACTIONS

The search for a global energy minimum, is dictated by the interactions which individual atoms in the protein have with others atoms. All these interactions are non-covalent, and are discussed in this section.

Since proteins are only marginally stable at room temperature, no type of molecular interaction is unimportant and even small interactions can contribute significantly (positively or negatively) to a protein's stability (Dill, 1990). Over the years different non-covalent interactions have been regarded as providing the dominant forces in stabilising proteins. The denaturation of proteins upon heating was thought to be due to the breaking of the hydrogen bonds with a large unfavourable enthalpy change. However, the original proposal by Kauzmann (1959), that non polar interactions imparted most of the stability to the native protein is still regarded as valid (Dill, 1990).

1.4.1 Hydrogen Bonding

The relative contributions of any forces in a protein under physiological conditions are difficult to evaluate. Firstly, because binding forces under these conditions represent competition between binding to water and binding to another group, and not the net formation of a new bond in a vacuum. Secondly, the results of studies on small model compounds in aqueous solution cannot be applied directly to proteins because the different parts of a protein are held together by a covalent peptide backbone, whereas small molecules are not (Abeles *et al.*, 1992).

Hydrogen bonds (D-H---A) are predominantly electrostatic interactions between a weakly acid donor group (D-H) and an acceptor atom (A) that bears a lone pair of electrons. In biochemical systems the D and A can be both the highly electronegative N and O atoms and occasionally S atoms as well. The hydrogen bonds are much more directional than van der Waals forces, although less so than in the case of covalent interactions. Their association energies are usually in the range of -2.8 to -7 kcal/mol, with the D---A distance in the range of 2.7 to 3.1 Å. The strength of the interaction is

dependent on the geometry, with the D-H bond being at its strongest when pointing along the acceptors lone pair orbital. Deviations from this ideal geometry are however not unusual, with the hydrogen bonds in alpha helices and in antiparallel beta sheets, the N-H bonds point along the C=O bonds rather than along the O lone pair orbital (see e.g. Voet and Voet, 1990).

The internal hydrogen bonding atoms in a protein are arranged to facilitate the maximum number of hydrogen bonds. Breakage of the internal hydrogen bonds upon unfolding of the protein results in no net loss of hydrogen bonds as they are now formed between the hydrogen bonding atoms and the surrounding water molecules. The free energy of stabilisation derived from the internal hydrogen bonds are therefore equal to the difference in the free energy of hydrogen bonding between the native and the unfolded protein. Since most of the hydrogen bonds have similar free energies, internal hydrogen bonding was, until recently, not thought to significantly stabilise the structure of the native protein relative to its unfolded state (Abeles et al., 1992). This idea is supported by the fact that in the presence of organic solvents which lower the dielectric constant of the solution, proteins unfold. Since hydrogen bonds are slightly ionic, the lowering of the dielectric constant should increase their strength, effectively stabilising the protein. More recent evidence (Myers and Pace, 1996) suggests that hydrogen bonds significantly stabilise native protein structures and provide elevated stability to thermostable proteins (Vogt and Argos, 1997).

1.4.2 Nonpolar Interactions

Two kinds of explanations have been used to describe the tendency of nonpolar molecules to clump together in an aqueous solution. The first explanation, the hydrophobic interaction (Kauzmann, 1959), describes the hydrogen bonding, present between water molecules as effectively squeezing out the nonpolar groups, and by so doing, forces these groups to aggregate. The second explanation deals with the attraction of nonpolar groups for each other. If the attraction between the groups is larger than that between the groups and water, then they will tend to come together.

1.4.2.1 Hydrophobic interactions

The hydrophobic interaction is a consequence of the high polarity and strong hydrogen bonding ability of water. Due to the strong interaction of the water molecules with each other, and their stronger interaction with each other than with a nonpolar molecule, a resistance is created to the insertion of nonpolar molecules. As a result, nonpolar groups tend to interact with each other and in so doing they reduce their exposed surface area to the surrounding water, allowing the water to interact more strongly with itself.

Contrary to what may be expected the insertion of a nonpolar molecule into water does not result in the breaking of hydrogen bonds between water molecules. The water molecules at a nonpolar surface are not able to form hydrogen bonds or dipolar interactions with the surface. However, they can form hydrogen bonds with other water molecules either at or below the surface (Collins and Washbaugh, 1985). This results in an increase in the ordering of water molecules at the surface, i.e. the system becomes entropically opposed with a reduction in the $-\Delta S$, and hence an increase in the $+\Delta G$ ($\Delta G = \Delta H - T\Delta S$). The formation of these hydrogen bonds corresponds to the reaction being enthalpically driven ($-\Delta H$). However, the net change in G remains positive.

Results of Monte Carlo calculations of the lowest energy states for a large number of different arrangements of water molecules surrounding a hydrocarbon (Jorgensen *et al.*, 1985), have led to the conclusion that the negative changes in both ΔS and ΔH were not due to a net increase in the number of hydrogen bonds of the water molecules around the nonpolar molecule. Rather the number of hydrogen bonds is approximately the same for the water molecules in the bulk solution as those around the hydrocarbon. Hence the enthalpically driven process is attributable to the interaction of the water molecules with the dissolved hydrocarbon by diopole-induced dipole and dispersion forces. The entropy decrease is due to the water molecules existing in a smaller number of different positions in the region occupied by the nonpolar molecule.

The effect of two nonpolar molecules coming together in water would then be a reduction in the number of unfavourable interactions with water because the surface area

hydrophobic molecules in contact with water is reduced. As a result the number of water molecules surrounding the nonpolar dimer is now reduced, causing an increase in the entropy of the system, with a favourable change in the Gibbs energy, -ΔG. The increase in entropy is accompanied by an increase in the enthalpy, ΔH, because there are fewer molecules in contact with the nonpolar surface.

The folding up of a protein to form hydrophobic interactions in the interior of the folded protein is one example of this process, since it decreases the exposure of the nonpolar surface to water. The formation of most chemical bonds gives off heat. While in this case one is dealing with an interaction rather than a bond, there should still be a negative change of enthalpy, implying that the reaction is more favourable at low temperatures. However, reduction in the number of nonpolar-water interactions which occur upon aggregation of nonpolar molecules (the association of nonpolar residues during protein folding) results in a positive enthalpy change. As a result of the reduced ordering of the water molecules with increasing temperature, the hydrophobic entropy is highly temperature dependent (-TΔS) (Dill, 1987), i.e. decreasing in magnitude while increasing in temperature. The consequent rise in ΔH would thus imply that the hydrophobic interaction becomes stronger with increasing temperature (Privalov and Gill, 1988). This is true up to about 60°C, above this temperature, the enthalpy change for interaction of nonpolar groups in water becomes zero and then negative. At higher temperatures therefore, the strength of the interactions between nonpolar groups in water decreases with increasing temperature and most proteins denature.

Consequently, many proteins have maximum stability at temperatures in the physiological range. Above this range, they undergo denaturation because of the positive enthalpy of denaturation. Below this range, with the exclusion of damage by ice formation, they undergo cold denaturation due to the weakening of the hydrophobic interactions, where the decrease in enthalpy from the breaking of hydrophobic interactions (and hence increase in the number of nonpolar-water interactions) is larger than the increase in the enthalpy from the breaking of other interactions (such as hydrogen bonds). So that the net enthalpy of denaturation is negative and denaturation becomes more favourable with decreasing temperature (see e.g. Franks, 1995).

1.4.2.2 Dispersion interactions

The second explanation for nonpolar interactions in water is based on the attraction of nonpolar groups for each other. Nonpolar groups are also held together by weak attractive van der Waals, or London dispersion interactions between any two molecules in contact. These forces are stronger between organic molecules than between water molecules. The strong hydrogen bonds and dipole interactions of water lead to a degree of structuring and resist close packing of the water molecules, resulting in the water being an unusually open liquid, with large amounts of empty spaces between the water molecules. The strength of the van der Waals interactions depends largely on the amount of matter that is in close contact, so that these interactions are stronger between two organic molecules, or two parts of a protein, than between the water and organic molecule or protein. As a result the organic molecules will tend to interact more favourably with each other.

As in most chemical bonds, these van der Waals interactions have a negative enthalpy change. The breaking of these interactions as well as hydrogen bonds is responsible for the positive change in enthalpy that is observed for the denaturation of most proteins at high temperatures.

It appears that neither the "squeezing out" hydrophobic interaction or the dispersion force are mutually exclusive. The water is a strongly dipolar, hydrogen bonded liquid that resists the insertion of anything that cannot interact strongly with it. On the other hand, proteins are large, dense molecules with atoms which may be highly polarised. The peptide chain is therefore more likely to fold up and interact with itself more strongly than it interacts with small water molecules.

1.4.2.3 Electrostatic interactions

The association of two ionic protein groups of opposite charge, within 4Å of each other is known as an ion pair or salt bridge (Barlow and Thornton, 1983). Repulsive forces between charged groups can cause denaturation of a protein, particularly at low

ionic strength. If the salt concentration is low, the charges on proteins are not shielded by a cloud of ions of opposite charge, so that the interactions between charges become much more important than at high ionic strength. Free ions in solution are highly solvated however, so that the free energy of solvation of two separated ions is about equal to the free energy of formation of their unsolvated ion pair. Most ion pairs are found to occur between residues that are distant in sequence along the protein polypeptide chain (Baldwin and Eisenberg, 1987). Thus they are thought to stabilise tertiary structure, rather than secondary structure, although their contribution is thought to be small. This partially accounts for the fact that buried (unsolvated) ion pairs rarely occur in proteins and that the ion pairs that are exposed to the aqueous solvent are poorly conserved amongst homologous proteins.

1.5 METHODS USED TO MONITOR THE UNFOLDING/REFOLDING PATHWAY AND STABILITY OF A PROTEIN

1.5.1 Physico-Chemical Means of Determining a Protein's Unfolding/Refolding Pathway

The most ideal method for monitoring the unfolding/refolding of a protein would be a technique whereby the three-dimensional position of each atom in a protein could be assigned a specific position in three-dimensional space as the protein unfolded/refolded from its native state to its denatured state or vice versa. Currently, only two procedures are available for characterising the necessary structures at the individual residue or even atom resolution, i.e. Nuclear Magnetic Resonance (NMR) and through the analysis of transition states. NMR is unparalleled as a spectroscopic technique in the level of detail which may be obtained for structure solution. The method may be applied to analysing stable states, including the unfolded state, and also for studying transient species by Hydrogen-Deuterium (H-D) exchange. For highly defined structures, information is given about backbone and side chains. In the case of less defined structures as well as H-D exchange, information tends to be more about the backbone and secondary structure.

The only way of analysing transition states is by kinetics, and the only way of analysing detail at individual residues is by kinetic measurements on mutant enzymes, as laid out by Matouschek *et al.* (1989, 1990) and Arcus *et al.*(1994). This "protein engineering approach" may also be used for the characterisation of intermediates. The procedure has been described in detail (Fersht *et al.*, 1992) and reviewed (Fersht, 1993). The principle applied is very simple: a suitable residue in the protein is mutated, generally to a smaller one, or to that of an alanine, and the change in stability of the protein is measured. The effects of the mutation on the stability of intermediates and transition states are also measured. In order to circumvent the need to elucidate the structure of the mutant protein, if the transition state were to be destabilised by exactly the same amount as the folded state, then the structure at the site of the mutation is most likely the full native structure. If, on the other hand, the mutation does not alter the stability of the transition state at all, then the protein must be fully unfolded at the site of the mutation. These principles may in turn be converted to quantitative terms by defining a parameter Φ which is the ratio of change in energy of a particular state to the change in energy of the native state on mutation. $\Phi = 0$ means the structure is not formed; $\Phi = 1$ means the structure is fully formed.

Exploiting both these techniques, a method has been proposed for the folding pathway of Chymotrypsin Inhibitor 2 (CI2). The method of folding proposed is that of nucleation condensation (Fersht, 1995). The "nucleus" is formed by those residues involved in the initial site around which the rest of the molecule condenses. The concept of a "collapse" of the protein around its hydrophobic core, has also been suggested by others (Dill, 1995). This concept of nucleation condensation has also been successfully carried out using computer simulation (Daggett *et al.*, 1996).

Other structural probes which have been exploited are fluorescence for monitoring changes in the environment of specific residues inside the protein. Polarisation is another technique which monitors the tumbling of specific residues within the protein. Additional spectroscopic probes include second derivative spectroscopy, and circular dichroism for monitoring α-helix formation or disappearance. These structural changes, induce changes in the hydrodynamic volume of the protein which may be monitored by size exclusion

chromatography, or gel electrophoresis in the presence or absence of denaturants such as sodium dodecylsulfate (SDS), urea or heat.

Conventional enzymology may be used as a functional probe in cases where the protein displays enzyme activity. An alteration in the orientation of atoms which constitute the active site results in a discernable increase/decrease in the catalytic activity of the enzyme. While only limited information is available by following the enzyme catalysed reaction under steady state conditions, the rate of these reactions is such that use of a stopped flow apparatus is essential for monitoring the pre-steady-state reaction. The stopped flow in turn has been exploited for monitoring the protein's folding/unfolding under non-equilibrium conditions, as opposed to that occurring under equilibrium conditions (see e.g. Service, 1996). In the case of the latter, only thermodynamically stable species are discernible. In the case of the former, transiently formed, i.e. non thermodynamically stable species or kinetic intermediates may be identified.

1.5.2 Theoretical Studies

Levinthal (1968) was the first to suggest that the way in which a protein folded was not dictated by a random stochastic search, whereby the protein adopts all possible conformations before finally assuming its energetically most stable conformation, the native state. Since the native state is unique, confirmed by the fact that proteins do crystallise, what would dictate the assumption of this energetically most stable state reached by obtaining a global energy minimum? Since it is not possible to monitor the movement of each individual atom of a protein as it moves from its position in a random coil towards its final position in the native protein, the equilibrium properties of matter may be considered from two points of view: the macroscopic and the microscopic. Thermodynamics is a macroscopic view which describes the behaviour of large numbers of molecules in terms of pressure, volume, temperature, composition, and exchanges of heat and work. The quantitative relationships between various measured properties are not based on any model of the microscopic structure of matter. For example the

experimental quantity $\Delta G = G_D - G_N$, where G_D and G_N represent the free energy of the denatured and native states respectively . The differences between the two values are determined by the laws of thermodynamics; though the absolute values of G_D and G_N are indeterminable within the framework of thermodynamics. The macroscopic view of thermodynamics starts off with experiment and is built up with the development of a scheme to explain the experimental observables. On the other hand, classical and quantum mechanics provide a microscopic description for the structure and interactions of molecules. Ideally one would like to be able to predict the thermodynamic behaviour of substances using our knowledge about individual molecules. If the mass, position, velocity, momentum, kinetic energy and potential energy of each and every atom of the protein and those of the solvent were known as a function of time, it would then in principle be possible to calculate all the properties of the system. Unfortunately the complexity of dealing with so many equations precludes this approach; therefore one is compelled to deal with the system on a statistical basis. Statistical mechanics provides the needed bridge between microscopic mechanics (classical and quantum) and macroscopic thermodynamics. (see e.g. Alberty, 1987). With the advent of powerful and affordable computers molecular dynamics methods have become a wide-spread means for simulating molecular-scale models of matter. Since the time of Newton, the N-body problem of classical mechanics was viewed as important. The reasons for the importance of this problem have evolved, and at present its importance stems from attempts to relate collective dynamics to single-particle dynamics. This would then explain how the behaviour of large collections of particles can be explained by examining the motions of individual particles.

A molecular scale simulation consists of three principle steps: (1) construction of a model, (2) calculation of molecular trajectories, and (3) analysis of those trajectories to obtain property values. The second step constitutes the simulation proper, providing the molecular positions. In molecular dynamics the positions are obtained by solving differential equations of motion and hence, the positions are connected in time, where the positions reveal dynamics of individual molecules as in a motion picture. Despite the availability of supercomputers, computing limitations require that the time involved for

the simulation needs to be dramatically reduced, in the order of nanoseconds, whereas the timescale for a protein to unfold is in the order of milliseconds. In order to do this, the simulated temperature of the system needs to be increased dramatically, often well above the boiling point of water, where the protein would find itself in a solvent no longer in the liquid phase. Other computing limitations involve the actual number of atoms whose trajectories may be simulated. In order to overcome this effect, two-dimensional or cubic lattices (see e.g. Dinner *et al.*, 1996) representative of the whole protein are used to overcome this deficiency. In still other simulation methods, the positions are computed from hybrid schemes that involve some deterministic features, as in molecular dynamics, and some stochastic features, such as Monte Carlo simulations. In Monte Carlo simulations the molecular positions are generated stochastically such that a molecular configuration depends only on the previous configuration. The results of the combination of the deterministic (molecular dynamic) with the stochastic (Monte Carlo) are then compared to the macroscopic model, with a reduction in the number of Monte Carlo "moves" indicating an improvement in the simulation. Simulations however are also prone to error, but with a constantly expanding databank of physico-chemically derived results with which to compare the simulations the principles of protein folding are gradually being learned.

CHAPTER 2 GLUTATHIONE S-TRANSFERASES

2.1 INTRODUCTION

Glutathione S-transferases (GSTs, EC 2.5.1.18) the multigene family of isoenzymes were first discovered in the early 1960's (Booth *et al.*, 1961). They are widely distributed amongst aerobic organisms (Pemble and Taylor, 1992), and localised in mammalian tissue where they are most able to carry out their function of detoxification. Together with other detoxifying enzymes, they are collectively known as enzymes of detoxication (Jakoby and Ziegler, 1990).

2.1.1 Glutathione S-transferases as Enzymes of Detoxication

The GSTs, together with flavin-containing monoxygenases, amine N-methyltransferases, sulfotransferases, and others, form the enzymes of detoxication (Jakoby and Ziegler, 1990). In all living organisms a broad system exists for the recycling of either naturally occurring or synthetic compounds. Micro-organisms in particular are capable of very effective catabolic activity, and their enzymes are generally specific to individual compounds. The problem however in dealing with nutritionally worthless compounds present in tissue of larger organisms is dealt with differently. These animals have evolved systems adapted for the elimination, rather than the utilisation of these toxic compounds. Effectively these enzymes alter xenobiotic (foreign) and endobiotic (host derived) compounds by making them either more readily excretible or less pharmacologically active.

All these enzymes have the following similarities in common: They all have a preference for hydrophobic compounds, although polar compound are not entirely excluded. Each has a very broad substrate range, and they are usually more concentrated at points of entry to the body (liver, lung, intestinal mucosa) or organ (choroid plexus of the brain). Many appear to be inducible as well, whereby the organism will respond

to the toxin/drug by producing more enzyme (Jakoby and Ziegler, 1990).

The detoxication enzymes have, in turn been divided into three main groups: Phase I enzymes, activate chemicals by forming reactive functional groups (e.g. epoxides) in them. Phase II enzymes deactivate these reactive chemicals by attaching a hydrophillic moiety (e.g. glutathionyl, glucuronyl, or sulphuryl) to them. Finally the phase III enzymes mediate the elimination of inactive and water-soluble products (Dirr *et al.*, 1994).

The GSTs form part of the phase II detoxication mechanism, by catalysing the S-conjugation between the thiol group of glutathione to numerous chemically diverse compounds including alkyl-and arylhalides, lactones, epoxides, quinones, esters and activated alkenes (Mannervik and Danielson, 1988). The consequence of this reaction, is a less reactive and more polar glutathionyl S-conjugate which in animals is further catabolized and excreted from the cell (Boyland and Chasseaud, 1969). The release from the cell also marks the first stage of glutathione (GSH) turnover. The released GSH-conjugate is transported from the blood to the kidney, where the degradation of the GSH begins by the removal of the γ-glutamyl moiety. This moiety is in turn transferred to an acceptor amino acid and relocated into the cell where it is involved in the resynthesis of GSH. Upon acetylation with acetylcoenzyme-A the remaining cysteinyl-glycine is hydrolysed by a dipeptidase catalysed reaction, where the cysteine conjugate is converted to a mercapturic acid, and excreted (Siegers and Younes, 1983; Mannervik, 1985).

In addition to cytosolic GSTs displaying much catalytic diversity, they are also able to function as very versatile ligand-binding proteins, known as ligandins (Litwack *et al.*, 1971; Listowsky, 1993). In the light of their abundance in cells (up to 10% of the cytosolic protein in some animal cells), as well as their binding properties, they are thought to mediate in intracellular storage and transport of hormones, metabolites, drugs, and a great variety of other hydrophobic non-substrate compounds (for reviews see eg.Dirr *et al.*, 1994; Wilce and Parker, 1995).

2.1.2. The 26 kDa GST from *Schistosoma japonicum*

Interest in the 26 kDa GST of *Schistosoma japonicum* (Sj26GST) first came to light when mice which were relatively resistant to chronic infection by the parasitic helminth *Schistosoma japonicum* responded highly to a 26 kDa antigen from these adult worms. This antigen was later identified as being a glutathione S-transferase (Smith *et al.*, 1986). In order to carry out several lines of investigation, such as affinity purification of antibodies, raising of specific antisera or vaccinating animals, large amounts of uncontaminated Sj26GST were needed. As a result, a bacterial plasmid (pSj5) which directed the synthesis of the entire Sj26GST molecule in *E.coli* was constructed (Smith *et al.*, 1988). Motivated by the fact that Sj26GST can be easily purified by affinity chromatography, an expression vector which would direct the synthesis of foreign polypeptides as fusions with Sj26GST was developed (Smith and Johnson, 1988). These fusions could then be purified under non-denaturing conditions by affinity chromatography. Thus was introduced the plasmid expression vector pGEX.

In order to facilitate cleavage of the fusion protein from the Sj26GST affinity tag, linkers susceptible to cleavage by site-specific protease, i.e. thrombin (pGEX-2T) and blood coagulation factor (pGEX-3X) were engineered (Smith and Johnson, 1988). The pGEX range of fusion protein affinity tags has proven so successful that it has been identified as one of the four main systems in use in *E. coli* (La Valli and Mc Coy, 1995).

Furthermore, Sj26GST with C-terminus fusion proteins has also led to some interesting structure function studies, such as the mimicry of receptor dimerisation (Yan *et al.*, 1995), and the identification of oligomeric structures of proteins (Nemoto *et al.*, 1995). An Sj26GST fusion has also facilitated the crystallisation of a peptide fragment (Lim *et al.*, 1994).

2.1.3 Construction of the pGEX Plasmid

By means of a series of manipulations, the pSj5 plasmid was modified to facilitate the expression of foreign polypeptides with the C-terminus of Sj26GST. The resulting

plasmid map is shown in Figure 2.1 and contains firstly, the *tac* promoter, essentially a hybrid of both the tryptophan and lactose promoters (De Boer *et al.*, 1983), followed by the complete coding sequence of the Sj26GST (Smith *et al.*, 1986; 1987). The normal termination codon has been replaced by a polylinker containing unique recognition sites for the restriction nucleases *Bam* HI, *Sma* I, and *Eco* RI, followed by TGA translation termination codons in all three reading frames. Secondly the β-lactamase-coding gene for inferring Ampicillin resistance on the host. Thirdly, an origin of replication or *ori* site. Fourthly, a fragment of the *lac* operon containing the over-expressed *lac* I allele of the *lac* repressor and part of the *lac* Z gene. Finally, oligodeoxynucleotides encoding the cleaveage-recognition sequence of thrombin was introduced (Smith and Johnson, 1988).

2.1.4 Overall Structure of GSTs

Since the elucidation of the first GST structure (Reinemer *et al.*, 1991), the co-ordinates of an additional 21 GST structures have been deposited in the Brookhaven Protein Data Bank (PDB) (Bernstein *et al.*, 1977). All fully functional GSTs are dimers, having molecular masses in the region of 50 kDa. Both subunits have an active site along the domain interface. The position and architecture of the active sites, which are situated in spatially equivalent positions in all GST classes, have been clearly defined in the GST crystal structures due to their crystallisation in the presence of substrate ligands. Between the two active sites in the dimer and extending along the subunit interface is a hydrophillic and solvent accessible furrow with a prominent cavity at the centre. Each site can be separated into two distinct functional regions: a hydrophillic G-site for binding the physiological substrate glutathione (GSH) in the extended form, and an adjacent hydrophobic H-site for binding structurally diverse electrophilic substrates. Each subunit in turn is divided into two domains, both of the active sites occur in the N-terminus domain 1 of their respective subunits. A third kind of site is created by the dimerisation. This hydrophobic binding site is different to the H-site in that it is much larger, and capable of binding compounds which are not potential substrates.

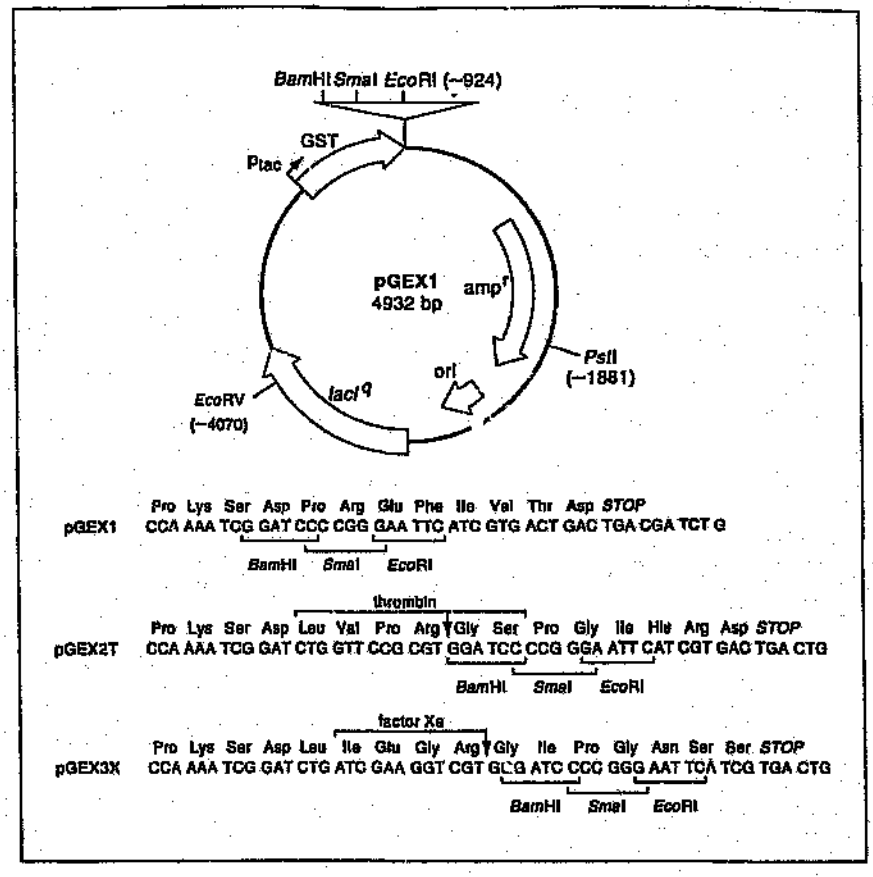


Figure 2.1. Map of the pGEX plasmid: a range of fusion-protein expression vectors that bind a cloned gene as a fusion to Sj26GST. The *lac* repressor (product of the *lacI* gene) binds the p_{tac} promoter, repressing the expression of the GST fusion protein. Upon induction with IPTG, derepression occurs and the GST fusion protein is expressed. The gene of interest can be inserted into the polylink... located at the end of the GST gene. The polylinker sequences are shown below the map of pGEX 1, where the restriction endonuclease sites are bracketed. The polylinker of pGEX-2T and pGEX-3X contain protease cleaveage sites so the cloned protein can be released from the GST moiety. The recognition sequences for thrombin and factor Xa are bracketed above the polylinker sequences, with the actual cleaveage site between Arg and Gly (adapted from Current Protocols in Molecular Biology)

2.1.5 GST Classes

GSTs have been classified, together with the microsomal enzyme into 6 gene classes, i.e. alpha, mu, pi, sigma and theta and kappa (Mannervik et al, 1985; Meyer *et al.*, 1991; Ji *et al.*, 1995; Pemble *et al.*, 1996). The classification is based on both primary and tertiary sequences. However, considerable discrepancy exists as to the classification of Sj26GST (McTigue *et al.*, 1995; Meyer 1995). Based on the fact that a class mu has successfully been used as a search model for solving the schistosomal GST structure, and a 42% sequence identity between the two enzymes, has led to Sj26GST being classified as a class mu (Hughes, 1992; Lim *et al.*, 1994) and not a sigma (Meyer, 1995).

2.1.6 Nomenclature

The nomenclature used in describing various GST classes differs widely. In the case of the 26-kDa glutathione S-transferase from *Schistosoma japonicum* abbreviations vary from sj26 (Hughes, 1992), to sjGST (McTigue, *et al.*, 1995). In this thesis the protein is abbreviated as Sj26GST. The nomenclature regarding other GST classes has been standardised (Mannervik *et al.*, 1992), with eg. rGSTM1-1 being an acronym for a glutathione S-transferases, where the prefix r, indicates the species i.e. rat, M, indicates the gene class mu, and 1-1 indicates a dimer of two type-1 subunits.

2.2 STRUCTURE OF THE 26 KDA GST FROM *Schistosoma japonicum*

Three structures of Sj26GST have been elucidated to date. Table 2.1 gives a brief summary of the Sj26GST structures currently available, together with some crystallographic data.

The co-ordinates of the structures were downloaded via a file transfer protocol (ftp) from the PDB (Bernstein *et al.*, 1977). The procedure used is summarised in Appendix A. Of the elucidated GST structures, only the class sigma GST (pdb accession codes, 1GSQ and 2GSQ) and Sj26GST have a monomer (designated subunit A) in their

Table 2.1. Sj26GST crystallographic data

pdb accession code	ligand	resolution (Å)	R-factor*	space group
1GNE	glutathione	2.5	0.219	$P4_32_12$
1GTA	none	2.4	0.197	$P6_522$
1GTB	praziquantel	2.6	0.212	$P6_522$

*R factor = $\Sigma \big| |F_{obs}| - |F_{calc}| \big| / \Sigma |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors respectively.

crystal asymmetric units. As a result the second monomer (subunit B) had to be generated in each of these cases. The procedure carried out in generating the dimeric structures is summarised in Appendix B.

2.2.1 Primary Structure

Table 2.2 shows the results of a sequence alignment of representative enzymes from different GST classes against Sj26GST. The sequences were aligned against the 218 residues of Sj26GST as specified in the 1GTA structure, according to the method of Needleman and Wunsch (1970).

The discrepancy between the 1GNE (Lim, *et al.*, 1994) sequence and those of 1GTA and 1GTB (McTigue, *et al.*, 1995), are due to the methionyl residue present at position 1 in the latter two structures, but absent in that of 1GNE, as well as the attached fusion to 1GNE. All numbering in this work is carried out according to that of Lim (1994).

2.2.2 Secondary Structure

The 217 residues of each Sj26GST subunit fold to form an N-terminal domain (domain 1) comprising a $\beta\alpha\beta\alpha\beta\beta\alpha$ folding pattern (residues 1-77), where the β-sheets are predominantly antiparallel. Connecting domain 1 to domain 2 is a short linker region (residues 78-84). The C-terminal domain 2 is larger comprising 5 α-helices (residues 85-193), and a long coil (residues 194-217) after the final helix (α-8) (Lim *et al.*, 1994; McTigue *et al.*, 1995; Reinemer *et al.*, 1996). A comparison of the secondary structural elements of representative classes of GSTs is shown in Table 2.3 (Reinemer, 1996).

2.2.3 Tertiary Structure

The overall fold of Sj26GST resembles that seen in mammalian GSTs. A least squares superposition of the 117 C positions from α-helices and β-strands onto those of

Table 2.2. Sequence alignment of various GST classes to 1GTA

PDB accession code	GST class	Similarity (%)	Root Mean Square(RMS)
1GTA	Sj26GST	100	0
1GNE	Sj26GST	99	51.7
1GST	rGSTM3-3	42	24.9
1GUH	hGSTA1-1	26	85.4
1GSR	pGSTP1-1	27	68.4
1GSQ	Squid	23	38.3

rat mu, pig pi, and human alpha GSTs, yield respective root mean square (r.m.s.) deviations of 1.35, 1.35, and 1.75 Å (McTigue *et al.*, 1995). The regions of greatest structural differences between Sj26GST and the mammalian enzymes occur at residues 32-40, and 199-217, which form part of the hydrophobic binding site discussed below.

The C-terminus of Sj26GST (residues 213-217) essentially sit on the surface of domain 2. It is also close to the 33-35 loop, a shortened version of the 33-42 loop of mu GST, that links β-strand 2 and α-helix 2 in domain 1.

2.2.3.1 Domain-domain interactions

The interactions between the two domains are made up of a single salt bridge which occurs between Lys10 (occurring between β-strand 1 and α-helix 2) and Asp159 in α-helix 6. The rest of the interactions are made up predominantly by nonpolar interactions between the residues Trp7, Pro15 and Leu18 from α-helix 1 in domain 1 and residues in α-helix 6 and α-helix 8 of domain 2. The side chain of Glu33 forms a hydrogen bond with the ring atom ε-N1 of Trp205. Residue 34 is an arginine, and its side chain interacts with the side chain of Asp213, and the carbonyl atom of Gly212 (Figure 2.2).

2.2.4 Quartenary Structure

The two subunits Sj26GST assemble to form a dimer of dimensions 57 Å x 47 Å x 44 Å. The 2-fold rotation axis relating the subunits within the dimer lies approximately 11 degrees from the helical axis of α-4, so that the symmetry-related α-4 helices cross at an angle of approximately 22 degrees (McTigue *et al.*, 1995). The dimer assembly creates a long (40 Å), narrow (6-10 Å) cleft lined primarily by polar side chains as described in Figures 2.3 and 2.4.

2.2.4.1 Dimer interface

The most prominent hydrophobic contact between the two monomers is the penetration of Phe51 between residues 90-93 (α-4), and 128-132 (α-5) of the other subunit. This "lock-and-key" type contact (Ji et al, 1994), is absent in the sigma and theta GST classes (Figure 2.3).

The following residues face the residues in subunit B of the dimer pair: Ala89, Met93, Leu94, and Phe132. The sequence number and identities of residues that participate in this hydrophobic interaction differ slightly in the other GSTs, but the above mentioned residues make up the core structure for the hydrophobic interaction between the GST dimer pairs. The region along the dimer 2-fold axis is a relatively open hydrophillic channel, and flanking this channel on both sides of the dimer is this hydrophobic interface, acting as a stabilising force for the GST dimer.

The hydrophillic interactions between the dimer pair are more variable than the hydrophobic interactions. Certain key sequence positions may be distinguished: Gln66, Arg72, Asp76, Arg88, Ser92, and Asp100. These residues are located in α-helices 3 and 4, and together with the corresponding α-helices of subunit B, are positioned nearly parallel along the 2-fold axis and form a hydrophillic core. Gln66 participates in glutathione binding, and interacts with Asp100 of subunit B. Arg72 interacts with its own symmetry-related residue in subunit B, and Asp76 interacts with Arg88 in the second subunit. In class mu GST, the polarity is changed so that Arg81 interacts with Glu90 and Asp97 of subunit B. In class pi GST, Arg77 interacts with Asp88 of subunit B. The residue Ser92 shows no direct interaction with subunit B in Sj26GST, however in the other alpha/mu/pi GST classes, this sequence position is occupied by an aspartate, and interacts with a positively charged residue in subunit B.

Table 2.3. Secondary structure comparison of GSTs: adapted from Reinemer *et al.*, (1996).

Domain	Structure element	Corresponding element in GST class					
		Sj26GST[a]	Mu[b]	Alpha[c]	Pi[d]	Sigma[e]	Theta[f]
I	β1	β1 (2-6)	β1 (2-7)	β1 (5-8)	β1 (3-7)	β1 (3-7)	β1 (2-5)
	α1	α1 (14-22)	α1 (17-26)	α1 (17-26)	α1 (15-23)	α1 (15-22)	α1 (10-21)
	β2	β2 (27-32)	β2 (28-35)	β2 (28-35)	β2 (29-32)	β2 (29-32)	β2 (27-30)
	3_{10}'						
	3_{10}''	α2 (38-43)	α2/3_{10}(43-46, 49-51)	α2 (38-47)	α2 3_{10} (38-44)	α2 (38-41)	α2 (41-44)
	3_{10}'''						
	β3	β3 (56-59)	β3 (61-64)	β3 (57-60)	β3 (52-55)	β3 (52-55)	β3 (54-57)
	β4	β4 (62-65)	β (67-70)	β4 (63-66)	β4 (58-61)	β4 (58-61)	β4 (60-63)
	α2	α3 (67-77)	α3 (72-82)	α3 (68-79)	α3 (63-74)	α3 (63-74)	α (65-76)
II	α3'						
	α3''	α4 (85-110)	α4 (90-114)	α4 (86-111)	α4 (81-104)	α4 (81-104)	α4 (86-111)
	α3'''						
	α4	α5 (113-136)	α5a (119-127)	α5' (114-129)	α5 (109-132)	α5 (112-135)	α5 (124-141)
			α5b (130-141)	α5'' (132-143)			
	α5'						
	α5''	α6(149-164)	α6 (154-169)	α6 (155-160)	α6 (148-163)	α6 (151-166)	α6 (154-169)
	α6	-	-	-	3_{10} (167-170)	-	-
	α7	α7 (173-184)	α7 (177-189)	α7 (179-190)	α7 (172-182)	α7 (175-186)	α7 (177-189)
	α8	α8 (186-192)	α8 (191-197)	α8 (192-198)	α8 (185-192)	α8 (188-196)	-

All crystal structure analyses were calculated according to the method of Kabsch and Sander (1983), and taken from the following GST structures: (a) Sj26GST (Lim *et al.*, 1994; McTigue *et al.*, 1995), (b) rGSTM3-3 (Ji *et al.*,1992), (c) hGSTA1-1 (Sinning *et al.*, 1993), (d) pGSTP1-1 (Reinemer *et al.*,1991; Dirr *et al.*,1994), (e) squid (Ji *et al.*, 1995), (f) Australian sheep blowfly *Lucila cuprina* (Wilce *et al.*, 1995).

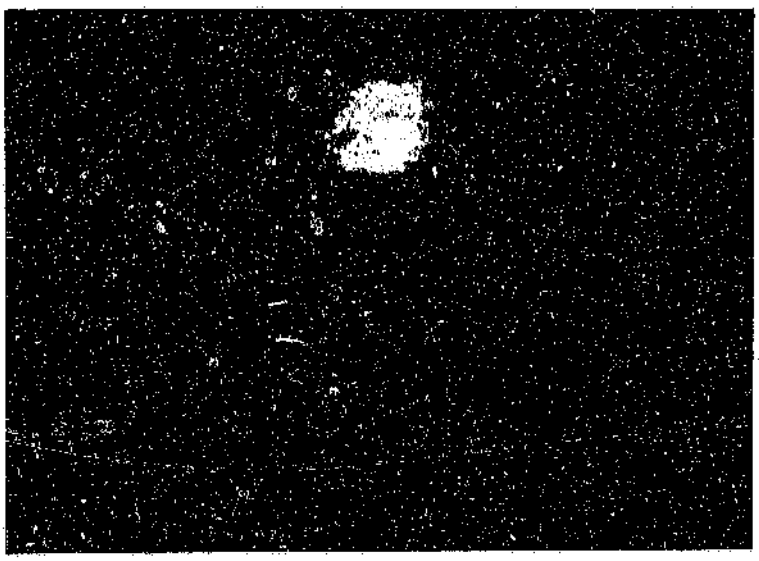
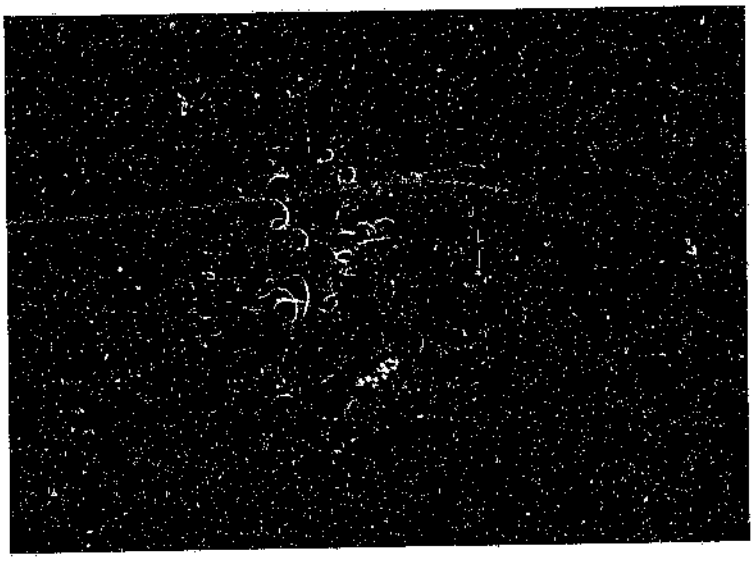


Figure 2.2. The overall tertiary structure of Sj26GST: domain 1 is indicated in green and domain 2 is in blue.

A



B

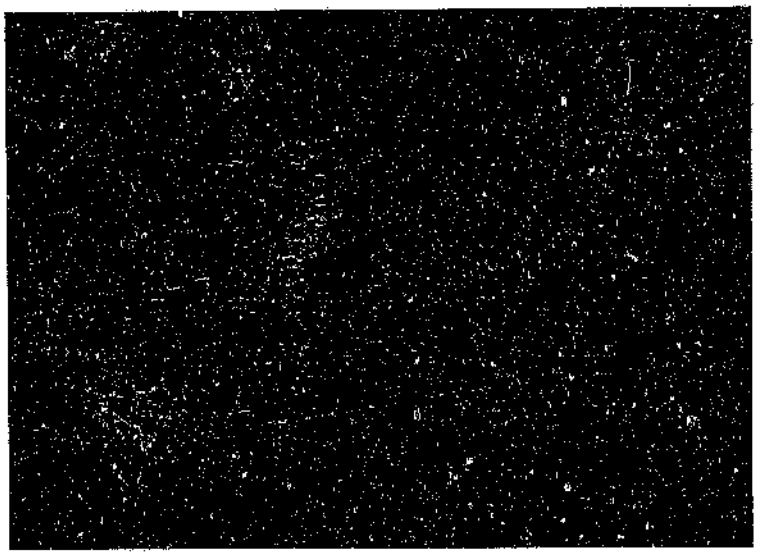
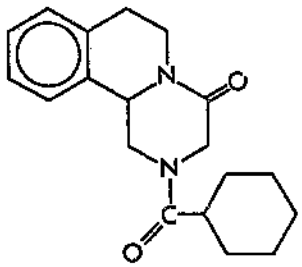


Figure 2.3. Dimeric structure of Sj26GST and squid GST: the view is down the two-fold axis with the two subunits indicated in green and blue. The most prominent hydrophobic interaction between the two subunits in Sj26GST, that of Phe51 (space filled) lock-and-key contact in Figure 2.3 A, is contrasted with the flatter subunit interaction of the class sigma GST in Figure 2.3B. The red molecule is that of the antischistosomal drug PZQ bound to the non-substrate ligand-binding site.

A



Praziquantel

B

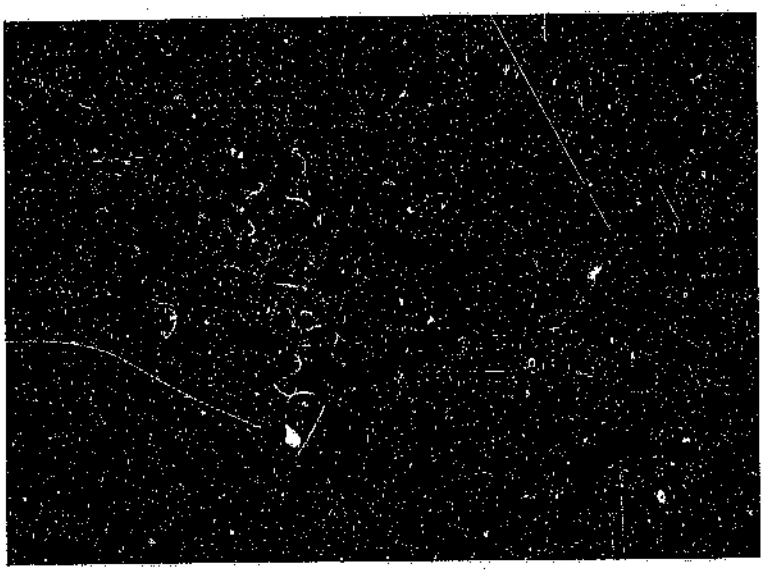


Figure 2.4. The structure of Praziquantel (A) and its interaction with Sj26GST (B). The view of the protein is perpendicular to the two-fold axis with the subunits indicated in blue and green and PZQ in red.

2.3 SUBSTRATE BINDING SITES

2.3.1 G-Site

In all GSTs, only an extended form of GSH is bound to domain 1 (Lim *et al.*, 1994). Four types of interactions with GSH occur in the binding site: (1) Stabilisation and orientation of the γ-glutamate of GSH, (2) alignment of the glutathione peptide backbone, (3) stabilisation of the terminal carboxylate of glycine, and (4) interaction with sulfhydryl of cysteine for enzymatic activation.

By co-crystallisation of Sj26GST with GSH (Lim *et al.*, 1994), the position of the G-sites can be seen in Figure 2.5. A comparison of those residues involved in binding GSH in classes alpha, mu, pi, and Sj26GST are summarised in Table 2.4 (Reinemer, *et al.*, 1996).

Site-directed mutagenesis studies have revealed the importance of specific residues in the GSH binding site. In class pi, substitution of Arg13, Gln62, and Asp96, resulted in a 20-50 fold decrease of both the catalysis and GSH binding, when compared to the wild type (Manoharan, *et al.*, 1992). Substitution of Tyr7 to phenylalanine in class pi, resulted in only a 27% GSH binding ability when compared to that of the wild type. The enzymatic catalysis though, was reduced to less that 1%. This substantial reduction in catalytic activity has also been shown in the case of class mu (Liu, *et al.*, 1992). Remarkably though, in the case of class theta (Wilce *et al.*, 1995, and Reinemer *et al.*, 1996) the equivalent tyrosine is absent and appears to have its role fulfilled by that of a serine.

2.3.2 H-Site

GSH binds to the domain 1 side, and there is more space in the domain 2 side for additional binding by an electrophilic substrate. This region has been termed the H-site, since the residues in the region are generally hydrophobic. No Sj26GST structure has been elucidated in the presence of any H-site ligands. However when compared to class

mu GST which has (Ji *et al.*, 1993,1994), the equivalent residues in Sj26GST are: ILe9, Leu12, Ser106, Tyr110, Gln203, and Gly204 as indicated in Figure 2.5.

2.4 CATALYTIC FUNCTION

The kinetic mechanism for the GST catalysed addition of the electrophile to the nucleophilic species in the active site (the thiolate anion of GSH), occurs sequentially in the ternary complex of the enzyme, GSH, and electrophile (Armstrong, 1991; see also Cameron *et al.*, 1995 for structure of a ternary complex with a class alpha GST).

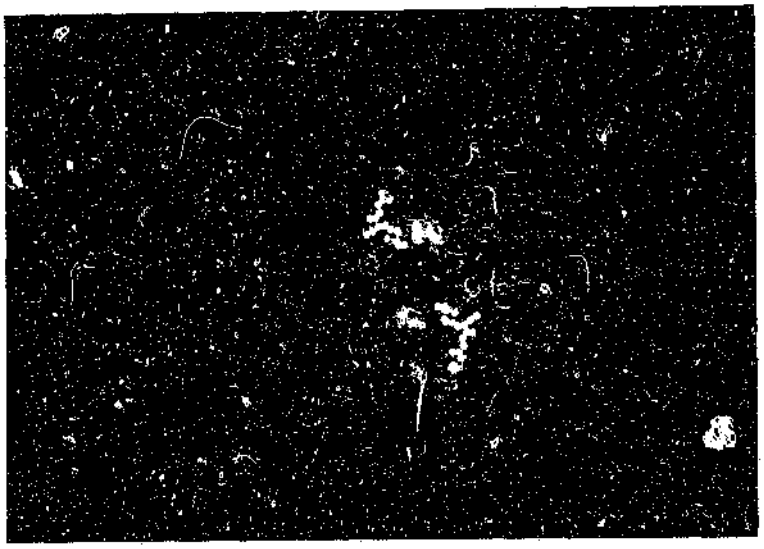
A schematic diagram of the sequential kinetic mechanism for GST can be represented as follows (Armstrong, 1991):

$$E + GSH \Leftrightarrow E{\bullet}GS^- + H^+ \overset{R\text{-}X}{\Leftrightarrow} E{\bullet}GS^-{\bullet}R\text{-}X \Leftrightarrow E{\bullet}GSR{\bullet}X^- \Leftrightarrow E + GSR + X^-$$

Where E is the GST,GSH is the physiological substrate glutathione and R-X is the hydrophobic substrate.

The G-sites of the alpha, mu, Sj26GST, and pi GST classes all have a conserved tyrosine residue present (Tyr6 in the case of Sj26GST). A hydrogen bond is present between the thiol group of GSH and the hydroxyl group of Tyr6. Replacement of Tyr7 in class pi GST with phenylalanine by site-directed mutagenesis, reduces enzyme activity to less than 1% of the wild type, although little change is observed in the affinity of the enzyme for GSH (Lim *et al.*, 1994). pKa values of the bound GSH in class pi GST is also lowered in the wild type (pKa = 6.3), Phe7 mutant (pKa = 8.7). The latter value being very similar to GSH in solution (pKa = 9) (Kong *et al.*, 1992). Other studies have confirmed the lowering of the pKa values of GSH by mutating equivalent tyrosine residue in class mu (Liu, *et al.*, 1992), and class alpha (Wang *et al.*, 1992). Thereby implying that Tyr7, or its equivalent in other classes, is essential for enhancing the nucleophilicity of the thiol group of GSH. In this instance it appears to act as a hydrogen bond donor, which promotes the formation of the thiolate by decreasing the pKa of GSH in the enzyme complex, and in so doing stabilises the thiolate species (Kolm *et al.*, 1992; Liu

A



B

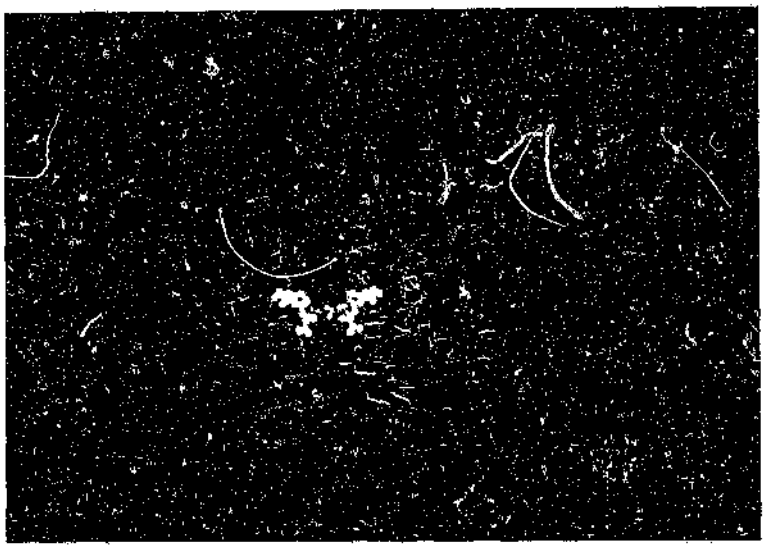


Figure 2.5. Glutathione and hydrophobic substrate binding sites: looking down the two-fold axis (A) and perpendicular to it (B). GSH is shown in white. Interactions with GSH are from residues of the same subunit, except for Asp100 which is contributed by the second subunit (space filled). The residues which comprise the H-site are in red.

Table 2.4. Interactions of the various GST classes with GSH: adapted from Reinemer *et al.*, (1996).

Glutathione moities		Glutathione S-transferase from class					
		Sj26GST[a]	Mu[b]	Alpha[c]	Pi[d]	Sigma[e]	Theta[f]
γ-Glu	NH_3^+	Gln66 Asp100*	Gln71 Asp105*	Gln66 Asp100*	Gln62 Asp96*	Gln62 Asp96*	Glu64
	COO^-	Gln66 Ser67	Ser72	Thr67	Arg13 Ser63	Gln62 Ser63	Ser65 Arg66
	C=O	-	-	-	Gln49	-	-
Cys	NH	Leu54	Leu59	Val55	Leu50	Met50	Ile52
	C=O	Trp7	Trp7	Val55	Leu50	Met50	Ile52
Gly	NH	Asn53	Asn58	-	-	-	-
	COO^-	Lys44	Arg42 Trp45 Lys49	Arg44 Arg130*	Trp38 Lys42 Gln49	Trp38 Lys42 Asn48	His38 His50

* Interaction contributed by a residue of the second subunit
(a) Sj26GST (Lim *et al.*, 1994; McTigue *et al.*, 1995)
(b) rGSTM3-3 (Ji *et al.*,1992; Raghunathan *et al.*, 1994)
(c) hGSTA1-1 (Sinning *et al.*, 1993)
(d) pGSTP1-1 (Reinemer *et al.*,1991; 1992; Dirr *et al.*,1994; García-Sáez *et al.*, 1994)
(e) squid (Ji *et al.*, 1995)
(f) Australian sheep blowfly *Lucila cuprina* (Wilce *et al.*, 1995)

et al., 1992).

Another role of the Tyr7 residue, or its equivalents which has been proposed, is that the tyrosine may function as a base, and therefore proton donor. Based on chemical modification studies with diethylpyrocarbonate (Meyer *et al.*, 1993), and electrostatic potentials calculated for the active site (Karshikoff *et al.*,1993), it has been suggested that Tyr7 in the active site is deprotonated, thereby promoting proton removal from GSH, and creating a reactive thiolate anion.

The remarkable finding that Tyr7, or its equivalents are not present in the G-sites of either class sigma or alpha classes (Ji *et al.*, 1995; Wilce *et al.*, 1995; Reinemer *et al.*, 1996) has resulted in a serine residue being proposed to take over the role of the Tyr7. While it has been suggested that the deprotonation of GSH by GST is due to the presence of histidine residues involved in base catalysis (Walker, *et al.*, 1995), no histidine residues have been found to be present in any of the mammalian class G-sites. The theta GSTs from both *Arabidopsis thaliana* (Reinemer *et al.*, 1996) and *Lucila cuprina* (Wilce *et al.*, 1995) do however have histidine residues present in their G-sites, although no catalytic role has been attributed to them. See Table 2.4.

2.5 LIGAND BINDING SITE (L-SITE)

While the ligandin function (Litwack *et al.*, 1971; Listowsky, 1993) of GSTs was well known, and it had been suggested that the possible site of non-substrate ligand binding was at the interface of the two subunits, it was not until the crystallisation of Sj26GST in the presence of the antichistosomal drug Praziquantel, (PZQ) 2-(cyclohexylcarbonyl)-1,2,3,6,7-11b-hexahydro-4H-pyrazino[2,1a]isoquinolin-4-one) (McTigue, *et al.*, 1995), or Biltricide as the drug is commercially known, (Goodman *et al.*, 1993), that the site was confirmed. The structure of PZQ is shown in Figure 2.4A and its binding to Sj26GST in Figure 2.4B. Subsequently another GST structure with an L-site ligand has also been solved (Ji *et al.*, 1996).

2.5.1 Praziquantel Binding Site in Sj26GST

The structure of the PZQ bound Sj26GST indicates the binding of one molecule of the drug per enzyme dimer, in the dimer interface groove, adjoining the two catalytic sites (Figure 2.4B). Therefore, PZQ may occupy either, but not both, of the two overlapping symmetry-related binding sites. PZQ makes contact with Gln66, Gly96, Leu99, Asp100, Tyr103 and Arg107 side chains from α-4 from one subunit, and also Tyr103 from subunit B. Interestingly, Tyr103, or the equivalent residue is missing in the case of the mammalian GSTs, where the equivalent positions are occupied by Cys99 in human alpha, Glu103 in pig pi, and Met108 in rat mu GSTs.

2.5.1.2 Praziquantel mode of action

It has been suggested (Smith *et al.*, 1986), that the binding of highly insoluble molecules such as haem and bilirubin by GST would increase their solubility. The suggested mode of action is that, haematein, the reduced form of the haem prosthetic group, produced during the metabolism of haemoglobin from the host erythrocyte, accumulates as an insoluble pigment in the lumen of the parasitic gut. It is then excreted by the parasite and lodges in the liver of the host. Sj26GST has been suggested to bind to the haematein and prevent the formation of large haematein crystals, which might block the evacuation of the parasitic gut. Thus PZQ, is thought to interfere with haem binding, by blocking the binding site (McTigue, *et al.*, 1995).

Others (Harder, *et al.*, 1987) have suggested that PZQ interacts with tegumental phospholipids, by causing spastic paralysis of the parasite musculature, leading to loss of attachment to the host, and rapid damage of the syncytical tegument. It is thought, that PZQ enhances the cell membrane's Ca^{2+} permeability, by membrane perturbation. The resulting influx of Ca^{2+} causes spastic paralysis of the parasite. When the PZQ-induced damage of the parasite's tegument is sufficiently severe, the resulting Ca^{2+} influx will disrupt various regulation processes triggered by Ca^{2+} and this then leads to the death of the parasite.

In rats up to 74% of the drug (20 mg/kg) is metabolised within 5 minutes after administration. The remaining 26% is reported to rapidly disappear, with 70-90% of the drug having been cleared from the body within 32 hours.

GSTs are almost certainly involved in the elimination of the PZQ from the system. Whether in fact the organism dies due to accumulation of haematein crystals or the altered permeability of the cell's Ca^{2+} transport mechanism, or an accumulation of the two effects, is unknown at this stage.

2.6 EVOLUTIONARY PERSPECTIVE

Evolutionary changes in proteins, which stem from random mutational events, often alter a protein's primary structure. A mutational change in a protein, if it is to be propagated, must somehow increase, or at least not decrease, the probability that the host will survive to reproduce. Many mutations are deleterious and often lethal in their effects and therefore rapidly die out. A comparison of the genes, the primary sequences and overall three-dimensional structures of related proteins may therefore shed light on the evolutionary development of the organism and provide insights into those residues which by virtue of their conservation are important for folding, stability or function of the protein.

A comparison of the cDNA and gene sequences of the class alpha, mu, pi and theta GSTs from both pro- and eukaryotes have led to the suggestion that the alpha/mu/pi precursor gene arose through the duplication of the theta gene (Pemble and Taylor, 1992; Armstrong, 1997). The theta gene is thought to stem from the kappa GST class (Pemble et al., 1996), and duplication of this gene would have afforded the protein the advantage of maintaining its original function while the duplicated copy evolved a new function.

Not only are the topologies of all GST classes similar, but the elucidation of the first GST structure (Reinemer, et al., 1991) showed that the topology of domain 1 of GST resembled that of thioredoxin from *E.coli* (Holmgren et al., 1975; Reinemer et al., 1991). Two structural features which have led to the clear evolutionary distinctions between various GST classes is the presence or absence of either the catalytic tyrosine

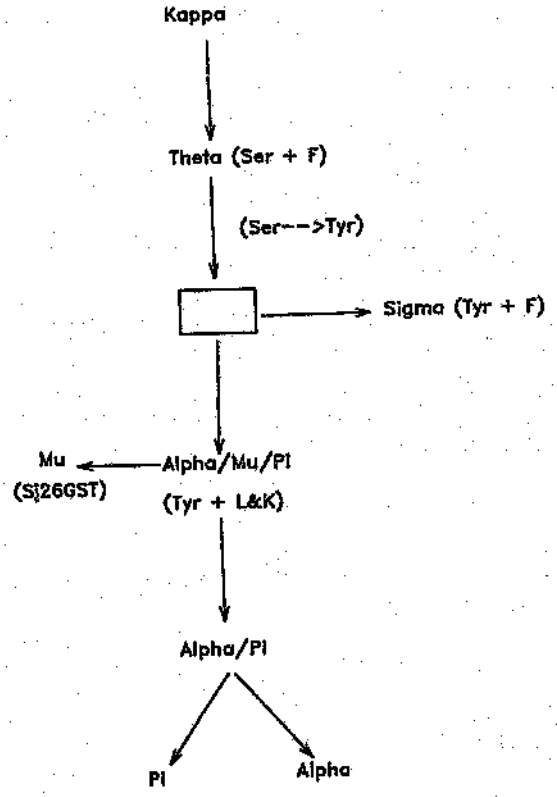


Figure 2. 6. Possible evolutionary scheme of the cytosolic GSTs. The precursor class theta enzymes have a serine in the active site together with a flat (F) subunit interface. In the class alpha/mu/pi enzymes the serine has been replaced by a tyrosine and the subunit interface is of the lock-and-key (L&K) type. The box represents the suggested sigma/alpha/mu/pi precursor (adapted from Armstrong, 1997).

or the "lock-and-key" interaction of phenylalanine from one subunit with nonpolar residues of the other subunit (Armstrong, 1997). In both the theta and sigma classes the "lock-and-key" interaction, present in the alpha, mu and pi classes, is absent. Unlike the theta class which has a catalytic serine, the sigma/alpha/mu/pi classes all have a tyrosine residue in its place. Based on the assumption that the theta gene is the ancestral precursor to the all classes other than kappa, the switch from the serine residue to that of tyrosine would have preceded the evolution of the change at the dimer interface. This would have yielded a sigma/alpha/mu/pi precursor which with the development of the "lock-and-key" interaction gave rise to the alpha/mu/pi classes (Armstrong, 1997). The highly conserved 3'-noncoding sequences of the mu and theta genes also suggest that the mu gene diverged from the theta precursor before the pi or alpha gene (Armstrong, 1997) and a possible evolutionary scheme of the different GST classes is described in Figure 2.6.

2.7 OBJECTIVES

In light of the physiological importance of Sj26GST as a detoxication protein and vaccine and drug target, and its wide application in biotechnology this work attempted to initially understand the protein's propensity to aggregate when purified in the absence of a reducing agent. In addition, it sought a clearer understanding of the conformational dynamics and subunit folding and assembly of Sj26GST in both the presence and absence of PZQ.

CHAPTER 3 EXPERIMENTAL PROCEDURES

3.1 MATERIALS

E. coli JM109 cells were obtained from Promega, and the pGEX-2T plasmid was from Pharmacia. The affinity column used was an S-hexyl GSH affinity matrix, in which the S-hexyl GSH ligand was prepared according to the method of Vince *et al.* (1971) and ligated to epoxy activated Sepharose 6B from Sigma.

GSH, BSA (Fraction V), and Dithiothreitol (DTT) were from Boehringer Mannheim. The CDNB was obtained from Sigma and sodium dodecylsulfate (SDS) calibration markers with a molecular mass range between 14.3 and 170 kDa were from Boehringer Mannheim Combitek.

The urea used in the gradient gels was from Riedel DeHaën, all other urea was Aristar Ultra Pure from BDH. Other chemicals were of an analytical grade quality.

3.2 METHODS

3.3 EXPRESSION AND PURIFICATION OF Sj26GST

3.3.1 Transformation of *E.coli* JM109 Cells

E. coli JM109 cells were transformed with the pGEX-2T plasmid using the one-step preparation method of Chung *et al.* (1989). An overnight culture was diluted 100x with Luria Bertani (LB) medium, until an $A_{600} \approx 0.3$. One volume of ice cold 2x Transformation and Storeage Solution (TSS) (10% w/v polyethylene glycol 4000, 5% v/v DMSO, 50 mM $MgCl_2$ in LB, pH6.5) was added and the suspension mixed on ice. 100 μl of the competent cells were mixed with 1-5 μl of plasmid (100 ng) in an ice cold Eppendorf tube. The solution was incubated at 4°C for 30 minutes, to which 0.9 ml of LB medium with 20 mM glucose was added. The cells were then incubated for 60 min

at 37°C with mild shaking, by placing them in an Erlenmeyer flask in an incubator. The cells were then immediately plated on a solid medium (10 g tryptone, 5 g yeast extract, 5 g NaCl, 15 g agar and 1 ml 1 M NaOH, made upto 1 liter), containing 100 μg/ml Ampicillin.

3.3.2 Growth and Induction of JM109 Cells Transformed with pGEX-2T Plasmid

An overnight culture of *E.coli* JM109 cells were diluted 10x into a fresh LB medium containing 50 μg/ml Ampicillin. The cells were grown at 37°C to a mid-log phase where the $A_{600} \approx 0.6\text{-}1.0$. The cells were induced by adding Isopropylthiogalactoside (IPTG) to a final concentration of 1mM. The cells were then grown for 4 hours at 37°C. All incubations at 37°C were accompanied by shaking at 180 rpm.

3.3.3 Preparation of Cell Extracts

The cells were pelleted by centrifuging for 10 minutes at 5000 x g. The supernatant was poured off, and the pellets resuspended in a minimal volume of Phosphate Buffered Saline (PBS) (150 mM NaCl, 16 mM NaH_2PO_4, 4 mM Na_2HPO_4, 2 mM DTT, pH 7.3) + 1% Triton X-100 . Volumes of ± 0.5 ml of the resuspended cells were added to Eppendorf tubes and placed on ice. The cells in the individual tubes were then lysed by sonicating the individual tubes using a Heat Systems (Ultrasonics Inc.) sonicator, with the following settings: output control was on the micro tip, and 50% duty cycle was selected. Each tube's contents was then sonicated for 30 seconds using the pulse mode. The colour of the suspended cells appeared to change from cream to a light beige after successful sonication. The crude extract was then centrifuged in a microfuge at 10 000 x g for 5 minutes, at 4°C. The supernatant was drawn off and purified by affinity chromatography on a regenerated S-hexyl GSH affinity column.

3.3.4 Application and Desorption of the Bound Protein to and from the Affinity Column

The cytosolic extract, containing the overexpressed protein did not usually exceed 10 ml from a 900 ml cell culture, and was applied directly to the head of the affinity column at 30 ml/h. The column was washed with the running buffer (20 mM Tris, 100 mM NaCl, 1 mM EDTA, 0.02% sodium azide, 2 mM DTT, pH 7.8) until the eluent had an A_{280} <0.05. The protein was then desorbed from the column using 5 mM S-hexyl GSH as competing ligand. The ligand was first solubilised in a minimal volume of 1 M NaOH before being diluted to volume with running buffer. 3 ml fractions of eluent were collected and their absorbances checked at 280 nm after having blanked the spectrophotometer on S-hexyl GSH. The fractions containing the protein were pooled and buffer exchanged (20 mM sodium phosphate, 2mM DTT, pH6.5) on a Sephadex G-25 column in order to remove the S-hexyl GSH. The eluted protein was identified spectrophotometrically as before. The fractions were pooled and the protein concentrated using an Amicon ultrafiltration cell with a PM10 membrane.

3.3.5 Regeneration of the S-Hexyl Glutathione Affinity Column

The column containing 30 ml of S-hexyl GSH affinity resin was re-generated by washing with 300 ml of sodium acetate buffer, pH 4. The column was then equilibrated with the running buffer. In all cases the column was assumed to have been equilibrated once the pH of the eluent was equal to that of the applied buffer.

3.3.6 Protein Determination

All protein concentrations were determined according to Bradford, (1976) and modified using the microassay technique as per instructions of BioRad.

3.3.7 Homogeneity of the Protein

The purity of the protein was confirmed via sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) according to the method of Laemmli (1970) as modified by Robyt and White (1990). The protein was run through a 4% stacking, and 15% running gel for about 3 hours at 180V. The gels were stained for about 1 hour (0.25% Coomassie brilliant blue R 250 in 45% methanol and 10 % acetic acid) and destained in 7% methanol and acetic acid (7%).

3.4 ENZYME ACTIVITY ASSAYS

The catalytic activity of Sj26GST was determined according to the method of Habig and Jakoby (1981). The enzyme ($\pm$ 0.02μM) was added to the reaction mixture containing 1 mM CDNB and 1 mM reduced GSH in 0.1 M sodium phosphate buffer, pH 6.5, containing 1 mM EDTA and 0.02% NaN_3. The formation of S-2,4-dinitrophenyl glutathione (ϵ = 9600 M^{-1} cm^{-1}) was monitored spectrophotometrically with a Hewlett Packard Vectra CS model 8452A diode array specrophotometer at a wavelength of 340 nm, over a period of 60 seconds.

Assays to determine the effects of denaturants and PZQ were monitored over 120 seconds by first incubating the protein until equilibrium (either at a specific urea concentration or in 100 μM PZQ), or for 10 minutes (at a specific temperature) and then following the above mentioned procedure. The PZQ had been solubilised in a minimal volume dimethylsufoxide (DMSO) and the incubated protein added to the standard reaction mixture which contained 100 μM PZQ. No reactivation of the enzyme was detected. All measurements were at room temperature and the linear progress curves were corrected for the corresponding non-enzymatic controls.

3.5 REACTION WITH ELLMAN'S REAGENT

DTNB (5,5-Dithiobis(2-nitrobenzoic acid), is a sufhydryl reagent able to react with aliphatic thiols by an exchange reaction to form a mixed disulphide, and a molecule of the yellow aromatic thiol, 2-nitro-5-thiobenzoate (NTB) per sulfhydryl group (Bico,

1994; Creighton, 1993; Habeeb, 1972). Just prior to experimentation, Sj26GST was buffer exchanged on a PM-10 column (Pharmacia) (0.1 M Tris, 1mM EDTA, pH 7.5) in order to remove any DTT. The number of free sulfhydryl groups in Sj26GST were determined at room temperature by titrating a solution of 5 μM Sj26GST (in the absence or presence of 8M urea) with DTNB to a final concentration of 250 μM. The release of NTB was monitored at 412 nm (ϵ = 13600 $M^{-1}cm^{-1}$).

3.6 FLUORESCENCE

The absorption of light at certain wavelengths may involve the excitation of a molecule from its ground state to that of a higher energy level. In most cases this excitation energy is lost as heat to the surroundings as the molecule relaxes to its ground state. In some cases however reradiation may occur. This process is called fluorescence see e.g. (Lakowicz, 1983; Cambell and Dwek, 1984).

Fluorescence therefore, involves two processes: absorption and the subsequent emission. Each process occurs in the time scale given by the inverse of the transition frequency (10^{-15} seconds). There is however a time lag of about 10^{-9} seconds when the molecule exists in the excited state. The lifetime of the molecule in the excited state depends on the competition between the radiative emission and any radiationless process, e.g. the transfer of excitation energy to the surrounding medium. These non radiative processes provide an alternative mechanism for the excited molecules to relax back to the ground state, and their presence will result in a reduction or quenching of the fluorescence intensity.

Fluorescence occurs at a lower frequency than that of the incident light. Since the detection frequency is different from the incident frequency, sensitivity is high because of absence of any background noise from the excitation source. Fluorescence may even be measured in concentrations as low as 10^{-8} M, about two orders of magnitude lower than absorption. Fluorescence is therefore very useful as a structural probe in monitoring binding, or conformational changes. Furthermore, there are many reactions which take place on the time scale of the excited state. The resulting sensitivity of fluorescence to

this time scale and to the environment is the basis of many of the applications to biochemistry.

The principle of fluorescence may then be thought of in the following physical terms: generally at room temperature a molecule would be in its ground state and its lowest vibrational level. Absorption of the appropriate energy results in excitation into the upper vibrational levels of the singlet excited state. The spectral transitions between the vibrational levels of different electronic states is governed by the Franck-Codon principle i.e. that an electronic transition occurs so rapidly that during the transitions the nuclei are static. Different vibration levels in the excited state will receive the transition depending on the relative positions of the ground- and -excited-state curves. In solutions, once molecules have reached the Franck-Condon excited state, several events can occur as the molecules relax back to their equilibrium populations, usually to that of lowest vibrational energy level in the ground electronic state. The first event involves the molecules in vibrational levels above the lowest one in the excited state losing their excess energy as heat. They then return to their lowest level and therefore establish a Boltzmann population distribution in the excited state. The process takes about 10^{-12} seconds to occur From this excited state, spontaneous emission can occur, and the emitted light is fluorescence.

The dependence of the fluorescence intensity on the wavelength of the exciting light is referred to as the excitation spectrum. Conversely, the emission spectrum describes the variation of the fluorescence intensity with the wavelength of the emitted light. The position of the maximum in the emission spectrum is sensitive to the polarity of the environment and mobility of the fluorophore.

3.6.1 Fluorescence Measurements

All fluorescence measurements were made at room temperature using a Hitachi model 850 fluorescence spectrophotometer. The excitation band was set to 5 nm and the emission band width to 10 nm. Samples were excited at 295 nm for measuring tryptophan fluorescence.

3.6.2 Binding Affinity Calculations

In order to determine the affinity of Sj26GST for PZQ, the binding affinity of the protein for the ligand was determined. Binding affinities are usually measured in terms of the association constant K_a. The value of K_a has units expressed as 1/concentration therefore, it is conceptually easier to consider the dissociation constant K_d which is the reciprocal of K_a and has units in concentration. Concentrations of free ligand below the K_d value result in little binding to the protein, and when the K_d equals that of the free ligand half of the protein molecules are bound to a ligand. In order to achieve an occupancy of 90% bound ligand would require a nine times greater concentration of free ligand, with 99% occupancy requiring 99 times the K_d (Creighton, 1993). The need for using very high occupancy levels of ligand is two-fold; firstly, it is necessary that the majority of protein molecules under investigation are of a uniform composition (i.e. bound or ligand free), secondly, the uptake of ligand by the protein should not significantly alter the concentration of the free ligand and thus disturb the equilibrium between the protein and ligand.

The affinity of Sj26GST for PZQ was determined by measuring the quenching of the intrinsic tryptophan fluorescence of the protein. The excitation wavelength was set at 295 nm, and the emission wavelength at 335 nm. A solution of 1 μM Sj26GST in sodium phosphate buffer, pH 6.5 was titrated with PZQ to a final concentration of 50 μM of the ligand. Controls were similarly prepared in the absence of enzyme. The final dilution factor did not exceed 10% of the initial volume.

During the titration of the protein with PZQ the fluorescence intensity was decreased. If the increasing concentrations of PZQ absorbed the excitation beam then this would alter the concentration dependence of the fluorescence, a phenomenon referred to as the primary absorption effect (Birdsall *et al.*, 1983). A consequence of the primary absorption effect would be the derivation of incorrect dissociation constants. As a result corrections to the fluorescence measurements were made according to Birdsall *et al.* (1983) as follows:

Corrections for protein controls: $F_t = F_{obs} - F_{con}$ (3.1)

Corrections for dilutions: $F_2 = F_1 \times V_f/V_i$ (3.2)

Correction for the primary absorption effect: $F_{act} = F_2 \times 10^{(Aex + Aem)/2}$ (3.3)

where F_1 and F_2 represent fluorescence values after corrections to equations 3.1 and 3.2 respectively. F_{act} is the actual fluorescence, F_{obs} is the observed fluorescence, and F_{con} is the fluorescence of the control. V_f and V_i are the final and initial volumes respectively, and A_{ex} and A_{em} are the absorbance readings at the respective excitation and emission wavelengths. The correction for the primary absorption effect can only be used for sufficiently low absorbances of the ligand at the excitation and emission wavelengths ($A < 0.2$).

The dissociation constant was then determined via the generation of a double reciprocal plot similar to that used in a Lineweaver-Burk graph:

$1/\Delta F = K_d/ [\Delta F_{max} (L)] + 1/ \Delta F_{max}$ (3.4)

where ΔF is the difference in fluorescence intensity between the ligand-bound protein and that in the uncomplexed form. ΔF_{max} represents the maximum ligand-induced change in fluorescence of the protein, and K_d is the dissociation constant.

By plotting $1/\Delta F$ against $1/(L)$ the K_d is determined directly via extrapolation to zero $1/\Delta F$, where the intercept on the $1/(L)$ axis is equal to $-1/K_d$.

The concentration of the PZQ bound to the Sj26GST was then calculated according to Lee (1982):

$L_b = \{(K_d + nP_t + L_t) - [(K_d + nP_t + L_t)^2 - 4nP_tL_t]^{1/2}\} / 2$ (3.5)

where L_b is the bound ligand, L_t is the total ligand, P_t is the total concentration of protein, and n is the number of sites on the enzyme to which the ligand binds. One PZQ molecule is bound per dimer of Sj26GST (McTigue *et al.*, 1995). The occupancy of the protein by the ligand is then calculated by plotting L_b/nP_t versus ligand concentration.

3.7 UNFOLDING/REFOLDING AND CONFORMATIONAL STABILITY CALCULATIONS

All unfolding/refolding experiments were performed at room temperature in 20 mM sodium phosphate buffer, pH 6.5, containing 0.1 M NaCl, 1 mM EDTA, 2mM

DTT, 0.02% NaN_3. Final concentrations of dimeric Sj26GST ranged from 0.2 to 20 μM. Prior to commencing the experiment, Sj26GST was confirmed to be dimeric and unaggregated by size exclusion high performance liquid chromatography (SEC-HPLC). Due to the propensity of urea to form cyanates and ammonium ions over time (Hagel *et al.*, 1971), stock solutions of urea were prepared fresh according to Pace *et al.* (1989), and the pH adjusted (pH = 6.5) using a minimal volume of concentrated NaOH. In the unfolding experiments, the protein was added to well-vortexed solutions of buffer with urea (0-8 M), to final volumes of 1 ml. The solutions were incubated for 1 hour, which was well past the previously determined equilibrium time of 10 minutes. Fluorescence intensities were measured by specifically exciting the tryptophan residues at 295 nm and measuring the emission wavelengths at both 335 and 355 nm. In the refolding studies, unfolded protein (20 μM in 6 M urea) was diluted 10-fold with buffer and no urea.

Both tryptophan fluorescence spectroscopy and enzyme activity were used to monitor the unfolding of the protein.

The absence of any intermediates in the unfolding pathways of many proteins, as shown in Figure 3.1, facilitates the assumption of a two-state unfolding mechanism:

$F \Leftrightarrow U$

where only the folded state F, and the unfolded state U are assumed to exist in significant concentrations in the transition region, and $f_F + f_U = 1$, where f_F and f_U represent the fraction of protein present in the folded and unfolded conformations respectively. Therefore, the observed value of y_{obs} (e.g. enzyme activity, tryptophan fluorescence) at any point will be $y_{obs} = y_F f_F + y_U f_U$ (3.6)

where y_F and y_U represent the values characteristic of the folded or unfolded states respectively. Both y_F and y_U are respectively calculated from the pre- and post-transition according to the following straight line equations:

$$y_F = mx + c \quad (3.7)$$

and $$y_U = mx + c \quad (3.8)$$

Combining these equations yields:

$$f_U = (y_F - y_{obs})/(y_F - y_U) \quad (3.9)$$

where f_U is the fraction of unfolded protein, and y_{obs} is the observed signal, i.e.

fluorescence intensity ratio F355/F335, or the percentage residual enzyme activity. The equilibrium constant of denaturation K_D was calculated for a dimeric protein according to the method of Bowie and Sauer (1989) as follows:

$$K_D = 2P_t[f_U^2/(1 - f_U)] \quad (3.10)$$

P_t is the protein concentration

and $$\Delta G_D = -RT\ln K_D = -RT\ln [(y_F - y_{obs})/(y_{obs} - y_U)] \quad (3.12)$$

here R is the gas constant (1.987 cal K^{-1} mol^{-1}) and T is the temperature in degrees Kelvin. The equilibrium unfolding curve was then iteratively fitted via a least-squares algorithm (Sigma Plot version 5.0, Jandel Scientific) to the following equation:

$$f(x) = a / [1 + e^{(b(x - c))}] + d \quad (3.13)$$

where a is the range of the mean f_U, b is the slope coefficient, c is the urea concentration at maximum rate of change, d is the minimum f_U, e is the exponential function, and x is the denaturant concentration.

By assuming a linear dependence of Gibbs free energy of unfolding (Schellman, 1978) the fitted values were used to compute the free energy of unfolding as follows:

$$\Delta G_D = \Delta G_D^{H2O} - m\,[\text{denaturant}] \quad (3.14)$$

where ΔG_D^{H2O} represents the difference in Gibbs free energy between the unfolded and folded protein in the absence of denaturant, and m the slope of the line is the change in accessible area of the protein upon unfolding.

3.8 UREA GRADIENT GEL ELECTROPHORESIS (UGGE)

In UGGE a uniform sample of either native or denatured protein is electrophoresed through a slab polyacrylamide gel, which is cast with a linear urea gradient, oriented perpendicular to the direction of electrophoresis. Due to the electrophoretic immobility of urea, different parts of the applied protein band will migrate in different concentrations of urea, thus altering the rate of interconversion between the native and denatured states. The sieving effect of the gel causes the unfolded protein conformations to migrate more slowly than do the more compact folded

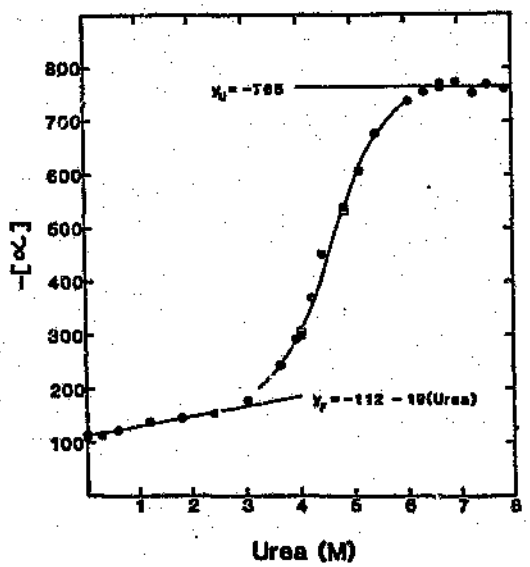


Figure 3.1. Solvent-induced unfolding transitions: where α is the specific optical rotation at 295 nm. The equations for y_F and y_U give the value of α for the folded and unfolded states respectively, where urea is the molar urea concentration (Pace *et al.*, 1992).

conformations. This technique therefore provides a continuous two-dimensional pattern of the effect of urea on the shape of the protein, and is especially sensitive to microheterogeneity of the protein. The validity of this method has been confirmed for proteins known to unfold/refold in a two-state manner (Privalov, 1974, Creighton, 1978), as well as for non two-state unfolding/refolding mechanisms (Creighton, 1986, Yu *et al.*, 1995).

Solutions and recipes for urea gradient gels:

The following solutions, with volumes and amounts having been adjusted to suit the apparatus used, were made up according to the method of Goldenberg (1990)

a. Stock solutio.

A. Acrylamide-bisacrylamide stock:

30 g of acrylamide

0.8 g of bisacrylamide

Water to 100 ml

B. 100 ml 0.5 M Tris-bicine (both components at 0.5 M, pH8.4)

C. 4 mg of riboflavin dissolved in 100 ml of water.

D. Ethanol solution for overlaying gels:

20 ml of ethanol

80 ml of water

2 mg of bromophenol blue

b. Recipes for gel solutions

0 M Urea, 15% acrylamide

10 ml of 30% acrylamide stock (solution A)

2 ml of 0.5 M Tris-bicine buffer (solution B)

2.5 ml of riboflavin stock (solution C)

5.5 ml of water

24 μl of N,N,N',N-Tetramethylethylenediamine (TEMED) (added after de-gassing, and just before casting of gels)

8 M Urea

9.6 g of solid urea

7.3 ml of 30% acrylamide (solution A)

2 ml of 0.5 M Tris-bicine buffer (solution B)

2.5 ml of riboflavin stock (solution C)

0.8 ml of water

24 μl of TEMED (added after degassing, and just before casting gels)

3.8.1 Casting Transverse Gradient Gels

The linear urea gradient gels (80 x 80 mm) were prepared as previously described (Creighton, 1986; Goldenberg, 1990) and as shown in Figure 3.2. An increasing urea gradient, in a gel of uniform acrylamide composition, is known to cause a continuous decrease in the electrophoretic mobility of a protein in the absence of any unfolding transitions (Creighton, 1979). This effect is attributed to either the increased viscosity of the gel or a change in its sieving properties (Goldenberg, 1990), and was compensated for by using an inverse acrylamide gradient.

By exploiting density differences between different concentrations of urea, the gel was cast at a constant pump rate (0.8 ml/min) by first pumping solution D through the bottom of the casting box (to a height of 2 cm), followed by 3 ml of the 0 M urea solution. The linearly increasing urea gradient was then formed by mixing identical volumes (9 ml) of the 0 and 8 M urea solutions in a well stirred beaker, together with the concomitant removal of the mixture by the pump. The 8 M urea solution was then pumped into the casting box until the boundary between solution D and the 0 M solution had reached the top of the glass plates (equivalent to a 1 cm band of both 0 and 8 M urea solutions separated by the gradient). The pump was stopped, the inlet to the casting box

disconnected from the pump and sealed, and the top of the casting box was covered with "cling wrap". All tubing was well flushed with water. The fluorescent lights in the room were turned on, and the gel in the casting apparatus was left to photo-polymerise overnight with riboflavin as catalyst.

3.8.2 Sample Preparation, Application and Electrophoresis

After polymerisation and the removal of the gel cassettes from the casting box, and the tape and spacers from the cassettes, the gels were turned 90° relative to their casting orientation. The new vertical edges were taped and the cassettes installed into an electrophoresis apparatus which facilitated the running of two gels simultaneously. The gels were pre-electrophoresed for 1 hour at 20 mA. Samples of native protein (50 μg of protein, 10% glycerol and 0.002% bromophenol blue) and of denatured protein (50 μg of protein, 36 mg of urea) were prepared. Both samples were made up to final volumes of 75 μl with 0.05 M Tris-bicine running buffer, pH ~ 8.4, which ensured good mobility of the Sj26GST in the electric field (pI = 5 (Walker *et al.*, 1993)). The samples were added across the top of their respective gels and then electrophoresed for 3 hours at 10 mA. Migration of the protein was in the direction of cathode to anode. The gels were stained in Coomassie Brilliant Blue.

3.9 SIZE-EXCLUSION HIGH PERFORMANCE LIQUID CHROMATOGRAPHY (SEC-HPLC)

Exploiting the molecular sieving properties of a variety of porous materials, molecules may be separated on the basis of their molecular size and shape (Corbett and Roche, 1984). Combining this technique with a high performance liquid chromatography system reduces the time taken for an elution profile from hours to minutes, without compromising resolution. The hydrodynamic volume of Sj26GST, at different urea concentrations (0-6 M), was measured at room temperature by analytical SEC-HPLC

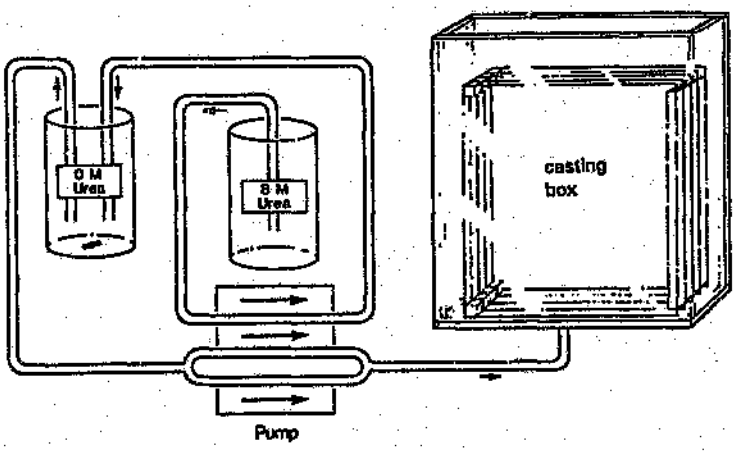


Figure 3.2. Preparation of urea gradient gels using a three-channel peristaltic pump. The glass plates are held in the perspex casting box. The gradient is formed by pumping from the low urea concentration solution in the mixing vessel into the casting box, using two channels of the pump, while simultaneously mixing the high urea concentration solution, pumped through a single channel of the pump, with the low urea concentration solution. The solution in the mixing vessel is vigorously stirred with a magnetic stirrer. The volume of the gradient is chosen to allow 1 cm regions at the top and bottom of the gels (as the gels are oriented for casting) which contain 0 and 8 M urea respectively (Goldenberg, 1992).

using an isocratic LKB model 2150 HPLC system, coupled to either a Tosohaas G2000SW or BioRad Bio-Sil SEC 250-5 column at a flow rate of 0.5 ml/min. All buffer solutions were degassed and filtered through a 0.45-micron nylon filter. Samples of 20 μM Sj26GST in different concentrations of urea were prepared 1 hour before injecting onto the column, which was pre-equilibrated with the same concentration of denaturant. Elution profiles were recorded at 280 nm using a Spectra series UV 100 detector.

3.10 THERMAL GRADIENT GEL ELECTROPHORESIS (TGGE)

TGGE can be applied to analyze conformational transitions and sequence variations of proteins, nucleic acids, and protein-nucleic acid interactions (Birmes *et al.*, 1990). A linear and highly reproducible temperature gradient is established perpendicular to the directions of electrophoresis of the macromolecule. The more compact folded conformations migrate more rapidly than do the unfolded expanded conformations, due to the sieving effect of the polyacrylamide gel. This provides a continuous two-dimensional pattern of the effect of heat on the conformation of a protein. To monitor structural changes of biomolecules spectroscopic methods, hydrodynamic techniques and observation of changes in chemical reactivity, have been applied. TGGE has the advantage in that it can fractionate, in the same experiment according to size and shape, as well as monitor structural changes. Therefore in principle, studies on the structural transitions of macromolecules may be carried out with high resolution, without purifying the biopolymer of interest (Riesner *et al.*, 1989).

3.10.1 Gel Casting

The gels were cast by assembling the two glass plates (300 x 300 mm) separated by spacers 1 mm thick and 10 mm wide. The spacers were recessed from the edge of the plates [illegible] the plates were then clamped to maintain their alignment. A molten solution of 2% agar was added to the outside of the spacers. The gel was then positioned at an angle of [illegible], and the gel solution (comprising 12.5% polyacrylamide, 0.29%

bisacrylamide, 2 mM DTT, 100 mM NaCl, 1 mM EDTA, and 0.075 M boric acid/0.03 M borax, pH 8.6) was poured between the plates. The choice of the boric acid/borax buffer selected from a table of buffers (Birmes *et al.*, 1990) was due firstly, to the pH being well above the pI of Sj26GST (see section 3.8.2) and secondly, in most buffers an increase in temperature and hence reaction enthalpy results in an increase in proton dissociation and a decreased pH. At 50°C the drop in the chosen buffer was only 0.33 pH units as opposed to other buffers in a similar pH range where the decrease ranged from 0.68 (Tris/boric acid, pH 8.3) to 1.66 (glycine/NaOH, pH 9.4) pH units.

The TGGE was performed as previously described by Birmes *et al.* (1990), but with the gel placed vertically against the aluminium gradient forming block, electrically insulated by a thin layer of teflon sheeting. Protein (250 μg in a final volume of 1 ml, containing 10% glycerol and 0.002% bromophenol blue) was added across the top of the gel and electrophoresed at 50 V for 30 minutes in order to allow for its smooth application onto the gel. The voltage was then increased to 300 V for 2 hours to allow for the uniform migration of the protein into the gel. Electrophoresis was halted and a linear gradient of 20°C to 80°C was allowed to form for 30 minutes, by passing water at 5°C through one extreme of the block, and at 85°C through the other extreme. Electrophoresis was then continued at 350 V for 5 hours while maintaining the temperature gradient. Gels were stained with Coomassie Brilliant Blue.

3.11 DIFFERENTIAL SCANNING CALORIMETRY (DSC)

Following the introduction of the first scanning calorimeter (Privalov and Khechinashvili, 1974), dramatic advances in the understanding of the unfolding of macromolecules resulted. The Differential Adiabatic Scanning Microcalorimeter Model-4 (DASM-4) comprises two near-perfectly matched platinum capillaries, each with a volume of $\pm$ 0.5 ml. Each of these cells have their own dedicated heating unit, and the cells are in contact with each other via approximately 200 thermocouples. These thermocouples ensure that as the temperature is increased/decreased at a constant rate, the two cells maintain thermal equilibrium with each other. Exploiting this aspect, a

typical scan is carried out by injecting the sample into one cell, and the reference buffer into the other. The run is then commenced, with the two cells "striving" to maintain thermal equilibrium with each other. At a specific temperature, the macromolecules in the sample cell begin to unfold in an exothermic reaction due to the breaking of non-covalent bonds. As a result of this increase in temperature in the sample cell, the heater dedicated to the reference cell attempts to compensate for this effect by additional heating to the reference cell. This energy input is recorded in micro-watts (W = cal/s), and it is in fact this energy input of the reference heater which gives the classic DSC thermograms.

Calorimetric measurements were conducted on a DASM-4 (Mashpriborintork, Moscow). See Privalov and Potekhin (1986) for additional background to its construction. Just prior to DSC experiments, all S.26GST samples were buffer exchanged on a PM-10 column (Pharmacia) with 20 mM sodium phosphate buffer, pH 6.5, to remove DTT which interfered with the baseline. Final protein concentrations ranged from 7.7 to 50 μM. Both the sample and buffer were degassed under vacuum for 10 minutes, and then injected into their respective cells over a period of about 1 minute. The chamber above the cells was then sealed and pressurised with nitrogen. In order to check the T_m for scan rate dependence, the heating rate was varied at 0.5, 1, and 2°C/min, and corrected for instrument dynamic response according to Freire *et al.* (1990). Data were collected every 0.1 min and stored to disk for further processing. Instrument baselines were determined by scanning the heat capacity versus temperature, with buffer in both reference and sample cells. Baselines were subtracted from the raw data before further analysis. Attempts at fitting the DSC curves to a multi-state reaction mechanism were not possible as the data fitted well to the following two-state reaction scheme:

$$F_{initial} \overset{K}{\Leftrightarrow} 2F_{final}$$

Where K is the equilibrium constant which is defined as a function of ΔH, ΔS and the partial specific heat capacity ΔC_p. The equilibrium constant can be expressed in terms

of the Gibbs Helmholtz free energy as follows:

$$K = \exp(-\Delta G/RT) \quad (3.15)$$

$$\Delta G = \Delta H + \Delta C_p(T - T_R) - T\{\Delta S + \Delta C_p \ln(T/T_R) - R \ln(2C_t)\} \quad (3.16)$$

Here ΔH and ΔS are the enthalpy and entropy changes at reference temperatures T_R. The enthalpy changes (ΔH_{cal}) were calculated by integrating ΔC_p vs. temperature over the transition interval, taking the protein concentration into account . The protein concentration is considered because ΔH is expressed in cal/mol; the calories are derived from the ΔC_p which is the apparent heat capacity change from the initial state $F_{initial}$ to the final state F_{final}. The number of moles are expressed as:

$n = C_t V$, where C_t is the protein concentration in molarity (M) and $V = 0.4623$ ml, which is the actual volume of both the sample and reference cells (within 1 μl of each other).

The equilibrium constant K therefore differs from a conventional equilibrium constant as the molar concentration term C_t has been included.

The molar enthalpy and entropy changes for the unfolding process were obtained by fitting the equations (3.17 & 3.18) to the experimental data.

$$\alpha_2 = (-K + (K^2 + 4K)^{1/2}) / 2 \quad (3.17)$$

Here α_2 is the fractional amount of species $F_{final} = \alpha_1 2C_t$

C_t is the total concentration of the dimers which dissociate to 2 unfolded monomers in F_{final}, and in terms of the law of conservation of mass $\alpha_1 + \alpha_2 = 1$.

The fitted molar heat capacity function (ΔC_p) which represents the DSC data can be expressed in terms of α_2, ΔH and ΔC_p as follows:

$$\Delta C_p = \Delta H (d\alpha_2/dT) + \alpha_2 \Delta C_p \quad (3.18)$$

T_m values were defined as :

$$T_m = \Delta H / [\Delta S - R \ln(2C_t)] \quad (3.19)$$

where T_m is the midpoint of the transition, and R the gas constant (1.987 cal/mol/K)

ΔH_{VH} values at individual protein concentrations were calculated from the integrated ΔH_{cal} curves, where:

$$\Delta H_{VH} = 2(m + 1)RT_m^2(d\alpha_2/dT) \quad (3.20)$$

where m is the number of monomers in the native protein, therefore for a dimeric

protein:

$$\Delta H_{VH} = 6RT_m^2(d\alpha_2/dT) \quad (3.21)$$

Here $d\alpha_2/dT$ can be expressed in terms of the change in temperature (ΔT) between the pre- and post-transition of the integrated ΔH_{cal}, as indicated in Figure 3.3.

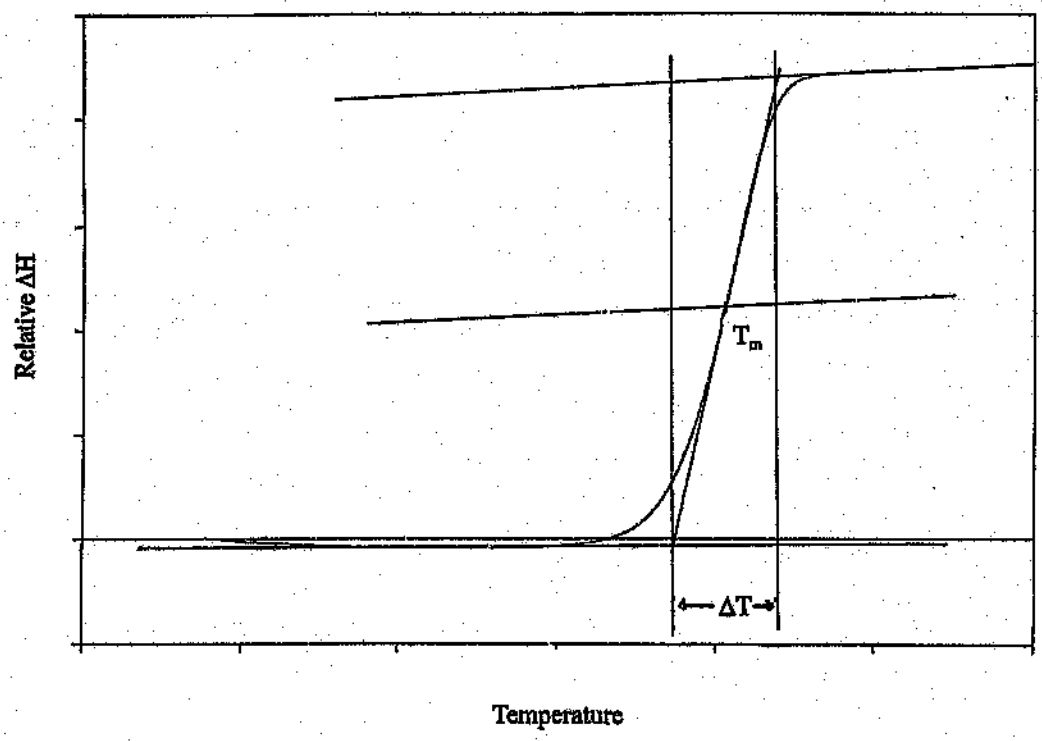


Figure 3.3. Calculation of the ΔH_{VH} from the integrated DSC curve. The temperature difference (ΔT) is calculated between the pre- and post-transition and then applied to equation 3.21.

CHAPTER 4 RESULTS AND DISCUSSION

4.1 AGGREGATION OF Sj26GST UNDER NON-REDUCING CONDITIONS

Sj26GST, as expressed by the pGEX-2T plasmid, has a tendency to undergo oxidative aggregation when extracted and purified under non-reducing conditions. SEC-HPLC elution profiles result in only the native dimeric Sj26GST when the enzyme is purified and analyzed in the presence of DTT. The absence of DTT results in a mixture of native protein together with 160 kDa and larger aggregates (60% native & 40% aggregates), as indicated in Figure 4.1A. Furthermore, SDS-PAGE in the absence of β-mercaptoethanol indicated the presence of a high molecular mass aggregate (73 kDa) in addition to the monomer (26 kDa) for protein not purified under reducing conditions (Figure 4.1B). All aggregates disappeared when DTT was present indicating a reversible process involving thiol groups. Comparable specific activities were obtained for both native dimeric (10.8 μmol/min/mg) and aggregated enzyme (9.4 μmol/min/mg). Sj26GST has 4 cysteine residues per subunit none of which is involved in intra- or inter-subunit disulphide bonds (Lim *et al.*, 1994; McTigue *et al.*, 1995). Titration of the reduced recombinant protein with DTNB yielded values of 3.5 and 3.7 free cysteines per subunit for folded protein and for unfolded protein, respectively.

Therefore, both the chemical modification data with DTNB and the crystallographic data (Lim *et al.*, 1994; McTigue *et al.*, 1995) indicate the four cysteine residues per Sj26GST subunit to be highly exposed to solvent (Figure 4.2). The formation of intermolecular disulphide bonds involving some of these cysteines could therefore explain the oxidative aggregation observed for the protein, as well as the reversal of aggregation by treatment with reducing agents. The small amount of the 73 kDa aggregate in the SDS-PAGE relative to that of the SEC-HPLC is perhaps attributable to the aggregation of the protein according to the scheme as laid out in Figure 4.3. Only the monomers of the inner circle are able to form disulphide bonds between the exposed cysteines, while those of the outer circle are spaced too far apart (> 2Å) to facilitate the

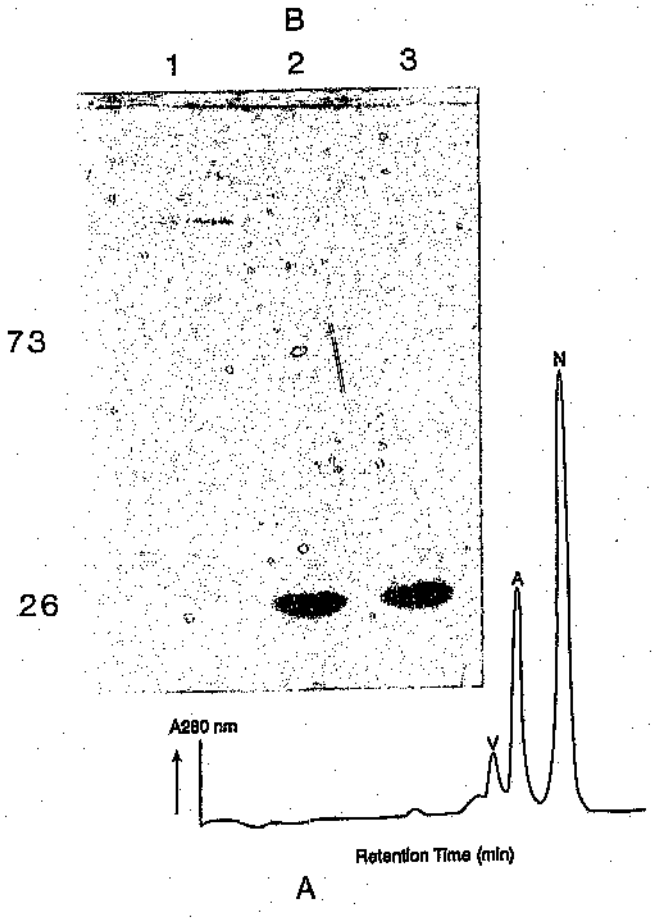


Figure 4.1. SEC-HPLC profile and SDS-PAGE of aggregated Sj26GST. A 20 μM sample of Sj26GST, which had bee purified in the absence of DTT, was applied to an SEC-HPLC column. The resolved species comprise 60% native and 40% aggregated protein (Figure 4.1A), where *N* is the native protein (52 kDa), *A* is the aggregated protein (73 kDa) and *V* is the void volume (160 kDa). Figure 4.1B represents an SDS-PAGE of aggregated Sj26GST which was loaded onto the gel in the absence (lane 2) and presence (lane 3) of β-mercaptoethanol. Both lanes show predominant species of the 26 kDa protein with a 73 kDA aggregate visible in lane 2.

Figure 4.2. Profile of the environmental exposure of the cysteine residues in Sj26GST. The degree of exposure is calculated according to the method of Lee and Richards (1971) and the cysteines are identified as the green vertical lines and from left to right are: Cys84; Cys137; Cys 168; Cys 177.

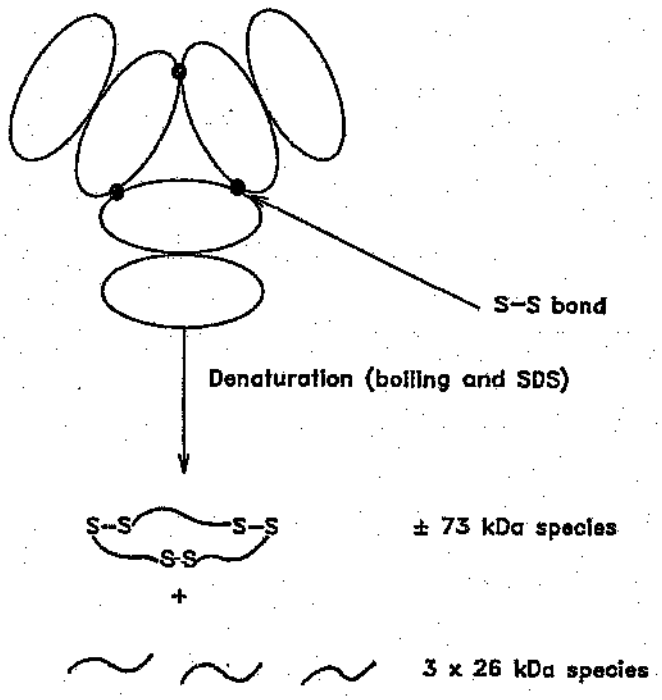


Figure 4.3. Schematic representation of the aggregation of Sj26GST. Only 3 disulphide bonds (•) are able to form in the inner subunit circle of the aggregate, with subunits on the outer side positioned too far apart to facilitate disulphide bond formation. Denaturation of the aggregate by heat and SDS results in the formation of the disulphide-linked 73 kDa species together with 3 individual 26 kDa species. The lower amount of aggregated protein in the SDS-PAGE relative to that of the SEC-HPLC is attributable to the added contribution of the 3 x 26 kDa species which migrate together with the unaggregated protein.

formation of inter-subunit disulphide bonds. The combined effects of the SDS without the reducing agent causes the native dimers to denature into their 26 kDa monomers. Each 160 kDa aggregated molecule would then result in three 26 kDa molecules and one 73 kDa aggregated molecule. Thus explaining the higher proportions of native protein to aggregated protein in the SDS-PAGE (40% aggregated/2 = 20% + 60% native = 80% native and 20% aggregated in the SDS-PAGE). The four cysteines appear in domain II; Cys84 is located near the subunit interface while the other three residues are clustered together on the surface of domain II on the opposite side of the active site. Because none of the cysteines is situated near the active site, access to these sites is not affected by the oxidative association of dimers. Aggregated protein retains its ability to bind the affinity matrix and displays a specific activity comparable to that of the native dimeric protein. The close proximity of the clustered cysteines to the peptide linker (with thrombin cleavage site) fused to the C-terminus of Sj26GST, provides an explanation for the reported obstruction of specific cleavage of GST-fusion proteins that have undergone oxidative aggregation (Abeliovich and Shlomai, 1995).

4.2 UREA-INDUCED UNFOLDING/REFOLDING REACTION

4.2.1 Reversibility of Unfolding (Structure and Function)

To study unfolding under equilibrium conditions it is necessary to show that the process is reversible (Pace *et al.*, 1989). Tryptophan fluorescence and enzyme activity were used as structural and functional probes to monitor changes in the Sj26GST molecule during unfolding/refolding. The enzyme contains four tryptophan residues per subunit (Trp7, Trp40, Trp200 and Trp205). The tryptophan emission spectrum for the folded protein (excitation at 295 nm) has an emission maximum at 335 nm which shifts to 355 nm when the protein unfolds. This red shift is accompanied by a large increase in quantum yield (Figure 4.4) suggesting that the tryptophan residues are located in a hydrophobic and quenching environment in the folded protein (Teal, 1960). Refolding of the denatured protein by a tenfold dilution in non-denaturing buffer resulted in a blue

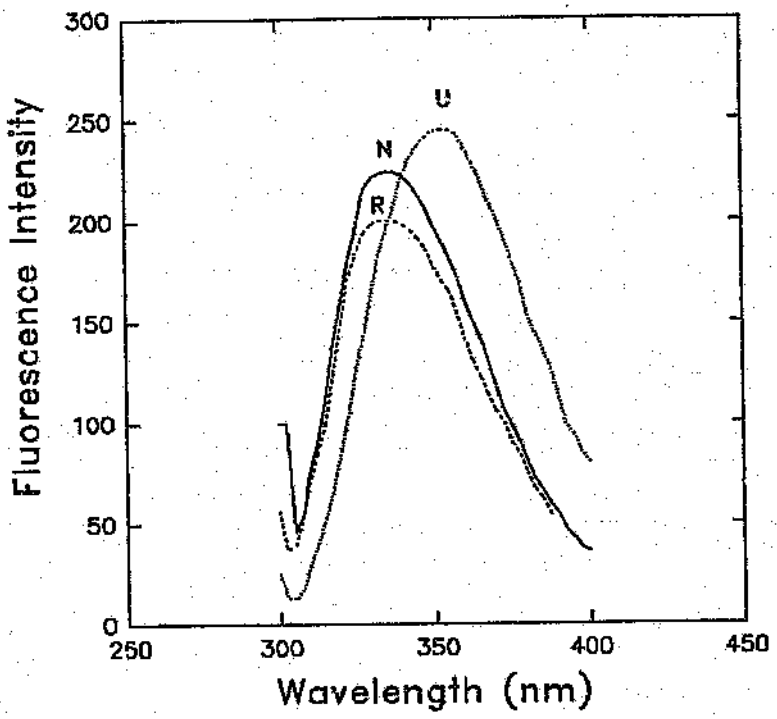


Figure 4.4. Reversibility of the unfolding of Sj26GST. Excitation was at 295 nm and the emission wavelength scan ranged from 300 to 400 nm. In all instances the protein was in 20 mM sodium phosphate buffer, pH 6.5, containing 0.1 M NaCl, 1 mM EDTA, and 0.02% NaN_3. The unfolded protein, *U* (20 μM in 6 M urea) was diluted 10-fold (refolded protein *R*), thereby equalling the concentration of the native protein, *N* (2 μM in 0.6 M urea). All measurements were performed after equilibrium was attained.

shift from 355 nm to 335 nm (Figure 4.4). About 90% of the emission intensity at 335 nm was recovered when compared with untreated folded protein. Furthermore, the recovery of enzyme activity following refolding was in excess of 80%. The data indicates that refolding proceeds mainly to the catalytically functional conformation.

4.2.2 Unfolding Transitions and Protein Concentration Dependence

Enzyme activity and tryptophan fluorescence unfolding data resulted in coincident and monophasic unfolding curves, once the raw data in Figure 4.5 A had been refined (Figure 4.5B). The midpoint of the unfolding transition increased to higher urea concentration with increasing protein concentration (Figure 4.5B insert). The monophasic and coincident transition curves for fluorescence and activity data, and the protein concentration dependence of this data, suggest a two-state unfolding/refolding model (Bowie & Sauer, 1989). This facilitated the estimation of the conformational stability of the Sj26GST in the absence of denaturant as shown in Figure 4.6. Table 4.1 gives some of the conformational stability properties derived from the unfolding of Sj26GST together with similar data available for other GSTs.

In order to conceptualise the meaning of the data in Table 4.1 a clear understanding of both the unfolding curves and the determination of the free energy, via the linear extrapolation method (LEM), must be attained.

The unfolding curves in Figure 4.5B describe an increase in the fraction of unfolded species as a function of the urea concentration. In the transition region of the unfolding curve one sees an increase in the unfolded species per unit change of urea concentration. The smaller the change in urea concentration through which the protein unfolds, i.e. moves from 0/1 fraction of unfolded species to 1/1 unfolded species, the higher is the cooperativity; the Cm value therefore implying that 1/2 of the population of protein molecules are unfolded.

The LEM of Figure 4.6 describes a decrease in the ΔG as a function of an increasing urea concentration. This decrease is due to the substitution of protein-protein non-covalent interactions for that of protein-solvent interactions. The breaking of these

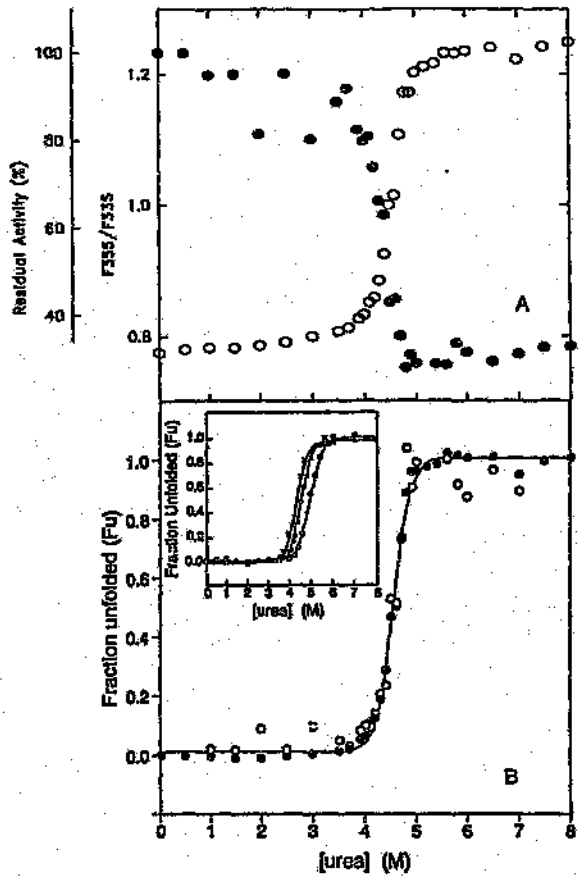


Figure 4.5. Urea-unfolding curves for Sj26GST. Tryptophan fluorescence (○) and enzyme activity (•) were used to monitor the unfolding of 2 μM Sj26GST. The unrefined data in Figure 4.5A shows the percentage residual enzyme activity and the fluorescence ratio of unfolded to folded (F355/F335) protein. The C_m value of the fraction of unfolded protein (F_u) in Figure 4.5B occurs at 4.5 M urea. The insert shows the protein-concentration dependence of Sj26GST stability against urea at 0.2μM (x), 2μM (•) and 20μM (○)

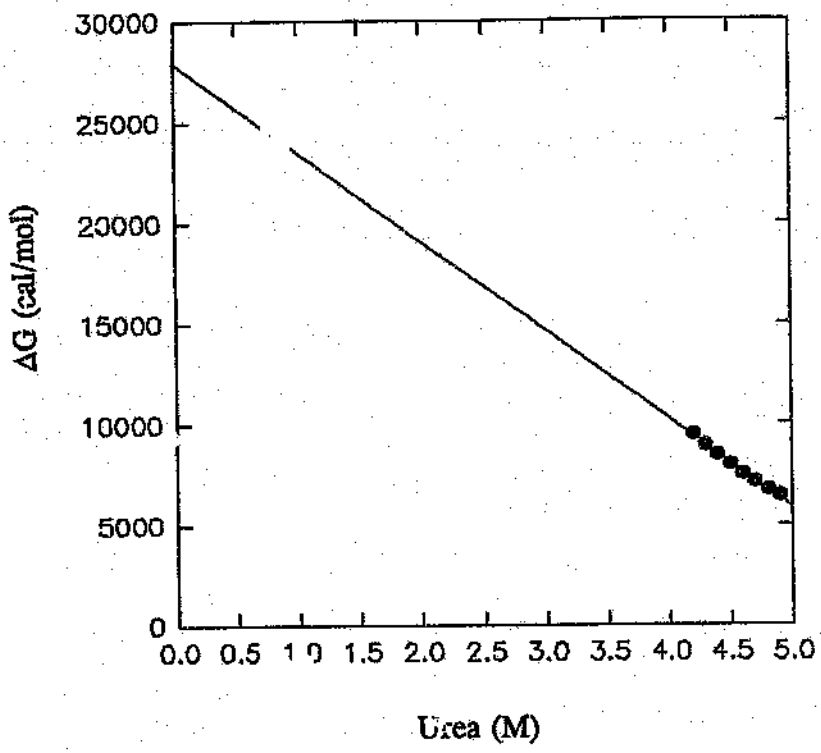


Figure 4.6. $\Delta G^{o}(H_2O)$ as a function of urea concentration. The experimental values of ΔG (•) were extrapolated to 0 M urea using equation 3.14, where $\Delta G^{o}(H_2O) = 26 \pm 1.7$ kcal/ mol and m = 4.5 kcal mol^{-1} M^{-1}.

Table 4.1. Conformational stability parameters of various GST classes. at different protein concentrations.

GST Protein	Concentration* (μM)	$\Delta G°(H_2O)$ (kcal/mol)	m-value (kcal/mol/M)	Cm-value (M urea)
Sj26GST	20	26.0 ± 1.7	4.5	5
pGSTP1-1†	1	21.6 ± 2.5	3.1	4.45
hGSTA1-1‡	1	27.5 ± 3.8	4.2	4.6

*By compensating for protein concentration effects (equation 3.10) similar ΔG values ($\pm$1.7 kcal/mol) were obtained at all three Sj26GST concentrations of Figure 4.5 (insert).

†(Dirr and Reinemer, 1991)

‡(L. Wallace, unpublished results)

bonds leads to the unfolding of the protein. It then becomes possible to quantify the unfolding of the protein in terms of an increased change in the accessible surface area (ΔASA) (Schellman, 1978), which has been found to increase linearly with an increase in amino acid number (Myers *et al.*, 1995). This dependence of the ΔG as a function of urea concentration is then specified by the slope of the graph, and is termed the parameter *m*. The larger the change in ΔG per unit denaturant concentration, the larger would be the number of broken interactions. Hence the more advanced the unfolding of the protein would be, the greater would be the cooperativity of unfolding. Therefore, as a function of urea concentration, a large change in Fu (i.e., fraction of unfolded species) over a small urea concentration range would then be comparable to a large change in ΔG (ΔASA) over the same range of urea.

The experimentally determined *m*-value of 4.5 kcal $mol^{-1}M^{-1}$ at 20 μM protein concentration compares fairly well with a value of 4.92 kcal/mol per M urea calculated from (Myers *et al.*, 1995):

$$m = 374 + 0.11(\Delta ASA) \qquad (4.1)$$

where ΔASA is the change in solvent accessible surface area upon unfolding (ΔASA = -907 + 93(number of amino acid residues) = 41129Å²).

4.2.3 Unfolding/Refolding Model

From the above data, it is reasonable to assume two basic models for the equilibrium unfolding/refolding pathway of Sj26GST. In model 1, active folded dimer (A_2), inactive folded monomer (A), and unfolded monomer (U) are significantly populated at equilibrium:

$$A_2 \overset{K_1}{\Leftrightarrow} 2A \overset{K_2}{\Leftrightarrow} 2U$$

where $K_1 = [A]^2/[A_2]$ is the equilibrium constant for the bimolecular dissociation reaction, and $K_2 = [U]/[A]$ is the equilibrium constant for the unimolecular unfolding reaction. The folded monomer is assumed to have a tertiary structure similar to that of the dimer subunit, but is catalytically inactive because the dimeric structure is required

to form a fully functional active site, due to the salt bridge formed between Asp100 of the second subunit and the N-atom of the γ-glutamate of GSH bound to the other subunit, as shown in Figure 2.5.

In model 2, only active folded dimer (A) and unfolded monomer (U) are significantly populated:

$$A_2 \overset{K_u}{\Leftrightarrow} 2U$$

where $K_u = [U]^2/[A_2]$ is the equilibrium constant for the bimolecular concerted dissociation and unfolding reaction.

The experimental data presented here (i.e., the coincident structural and enzyme activity data, protein-concentration dependence, monophasic transitions) support Model 2 which describes a highly co-operative and concerted two-state pathway for the equilibrium unfolding/refolding of Sj26GST. The size of the protein or the amount of its surface area exposed to solvent upon unfolding is the major structural determinant for the *m*-value (Schellman, 1978; Alonso & Dill, 1991; Myers *et al.*, 1995). The magnitude of the *m*-value is also indicative of a highly co-operative two-state unfolding/refolding process (Pace, 1986). Conformational stabilities for dimeric proteins generally range between 10 to 27 kcal mol^{-1} M^{-1} which is significantly higher than those for monomeric proteins (Neet and Timm, 1994). The large stabilisation energy of the dimer is usually attributable to the interactions at its subunit interface and gives a rationale for the formation of oligomers (Neet and Timm, 1994). Molecular recognition for dimerisation of the different types of GST monomers is strictly gene class, specific with no known interclass dimers, (Dirr *et al.*, 1994; Jones and Thornton, 1995; J.Hornby unpublished results). As in all other GST structures, intersubunit contacts in the Sj26GST dimer are formed primarily between domain I of one subunit and domain II of the other. The interface has extensive hydrophobic patches at either end and is hydrophillic near the dimer 2-fold axis. The most prominent hydrophobic interaction at the interface is the anchoring of the phenyl ring of Phe51 located in a loop after alpha-helix 2 in domain I of one subunit into a hydrophobic pocket between alpha-helix 4 and alpha-helix 5 in domain II of the adjacent subunit. This lock-and-key feature is also conserved at the

dimer interface of class alpha/mu/pi enzymes but is absent at the more hydrophillic interface in sigma and theta class dimers.

Recently, a new method for identifying and analysing statistically significant residue clusters (basic, acidic or of mixed charge) that occur in three-dimensional protein structures has been described (Karlin and Zhu, 1996; Zhu and Karlin, 1996). Approximately 10% of protein structures contain charge clusters. Positive charge clusters are very rare, and negative charge clusters are usually involved in coordinating metal ions. Mixed charge clusters being prominent at interchain contacts where they stabilise quartenary protein formation. The advantage of this method is in the identification of the mixed charged clusters which would appear neutral when charge potentials are used. This method has been applied to rGSTM3-3 (PDB accession code 2GST) which, while not only belonging to the class mu, has in addition a 42% sequence identity with Sj26GST. The analysis shows that no significant charge cluster exists in either of its subunits, but involves a mixed charged cluster at the interface between the two chains. Consequently, the mixed charged cluster is thought to be essential in contributing to dimer formation and would appear to substantiate the less important role played by the "lock-and-key" interaction in stabilising the alpha/mu/pi classes (J. Stevens, unpublished results). Of the 15 residues in rGST3-3 which form an intersubunit charge cluster, only 3 of the topologically equivalent residues in Sj26GST are conserved (Arg72, Arg88 and Glu95), with no conservation of charge occurring in the other residues. At this stage, however, the contribution of the individual interactions towards the conformational stability of GSTs is unknown.

4.2.4 Effect of Binding PZQ

Dimerisation of Sj26GST creates a solvent accessible cleft along the 2-fold axis which has been shown to function as a binding site for the non-substrate ligand, and antischistosomal drug, PZQ (McTigue *et al.*, 1995). PZQ makes contact with Gln66, Gly96, Leu99, Asp100, Tyr103 and Arg107 of subunit A and also Tyr103 of subunit B. In the light of the contacts made by PZQ to Gln66 and Asp100, both of which are

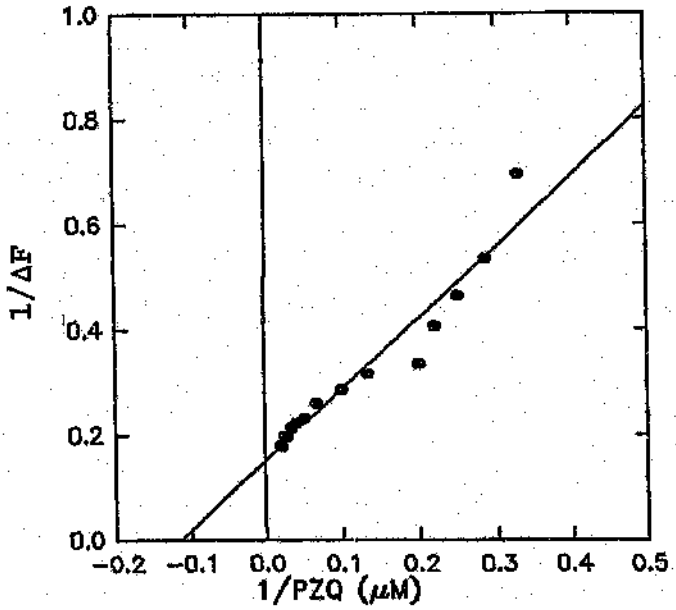


Figure 4.7. Determination of the dissociation constant, K_d for PZQ. The fluorescence quenching technique described in section 3.6.2 was used to determine the K_d.

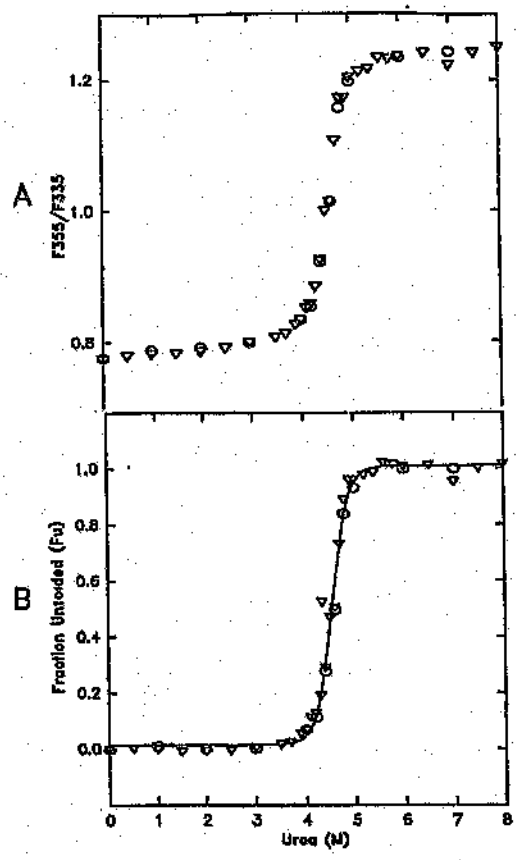


Figure 4.8. Urea unfolding curves for Sj26GST in the presence and absence of PZQ. Tryptophan fluorescence was used to monitor the unfolding of 2 μM Sj26GST in the presence (circles) and absence (triangles) of PZQ. The ratio of unfolded to folded (F355/F335) protein is shown in (A) and the fraction of unfolded protein (Fu) in (B).

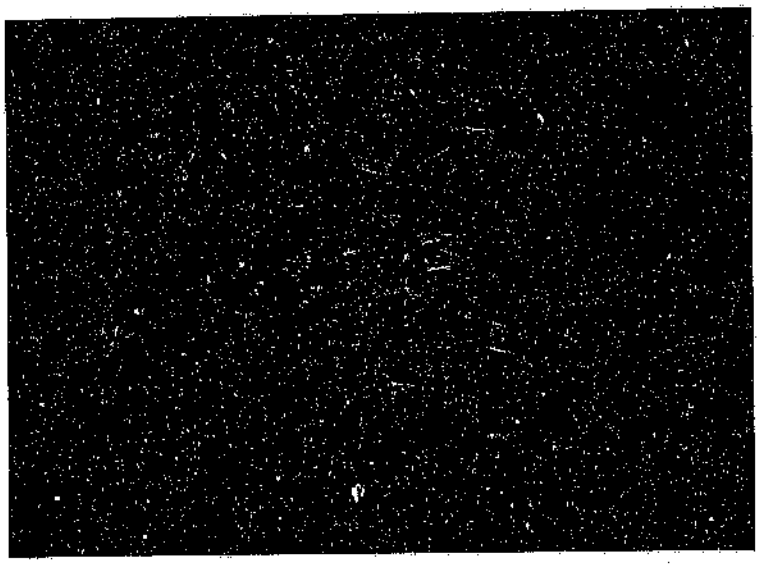


Figure 4.9. Overlay of the uncomplexed and PZQ-bound Sj26GST. The coordinates of the PZQ-bound (PDB accession code 1GTB) were overlaid with the unbound (1GTA) Sj26GST according to the least-squares superposition method of Hendrickson (1979). The view is down the two-fold axes of equivalent monomers with PZQ in red. No gross conformational differences are visible between the two structures.

involved in GSH binding, it has been inferred by others that PZQ inhibits enzyme activity (McTigue *et al.*, 1995; Meyer *et al.*, 1995; Cameron *et al.*, 1995). Despite tightly binding PZQ ($K_d = 9 \pm 1.8$ μM PZQ, as calculated from Figure 4.7), no reduction in enzyme activity was found to occur in the presence of 100 μM PZQ. A higher concentration (500μM PZQ) has also been reported to not reduce enzyme activity (Walker *et al.*, 1993).

In order to discern the effect of a saturating concentration of PZQ (450 μM) on the stability of Sj26GST, a conformational stability study was carried out (Figure 4.8). No difference in the stability of the protein in either the presence (26 kcal mol^{-1}) or absence of PZQ (26 kcal mol^{-1}) was discernible. A least squares superimposition (Hendrickson, 1979) of the ligand free and PZQ bound Sj26GST indicates no gross conformational differences in structures (Figure 4.9). The number of contacts between the two subunits is however reduced from 23 to 19 upon binding the drug. Elevated conformational stabilities of dimeric proteins relative to monomers are well established as being due to intersubunit interactions (Neet and Timm, 1994). The most dominant hydrophobic interaction between the two subunits, that of the Phe51 lock-and-key interaction, is not affected by the presence of PZQ (Figure 2.3). The absence of any changes in stabilities in either the ligand-bound or uncomplexed Sj26GST highlights the difficulties involved in attributing protein stabilisation to the effect of specific bonds. (An *addendum* dealing with reservations, expressed by one of the reviewers, regarding this section and section 4.3.5 have been included on page 158)

4.2.5 Urea Gradient Gel Electrophoresis (UGGE)

The presence of intermediate states in the unfolding/refolding pathway of some proteins which have gone undetected using spectroscopic probes, can be found using urea gradient gel electrophoresis (Creighton, 1986). UGGE was therefore used in order to ascertain whether any intermediate conformational states were present along the unfolding/refolding pathway of Sj26GST. Figure 4.10 illustrates the electrophoretic behaviour of the protein when applied across the gradient of urea in the folded state

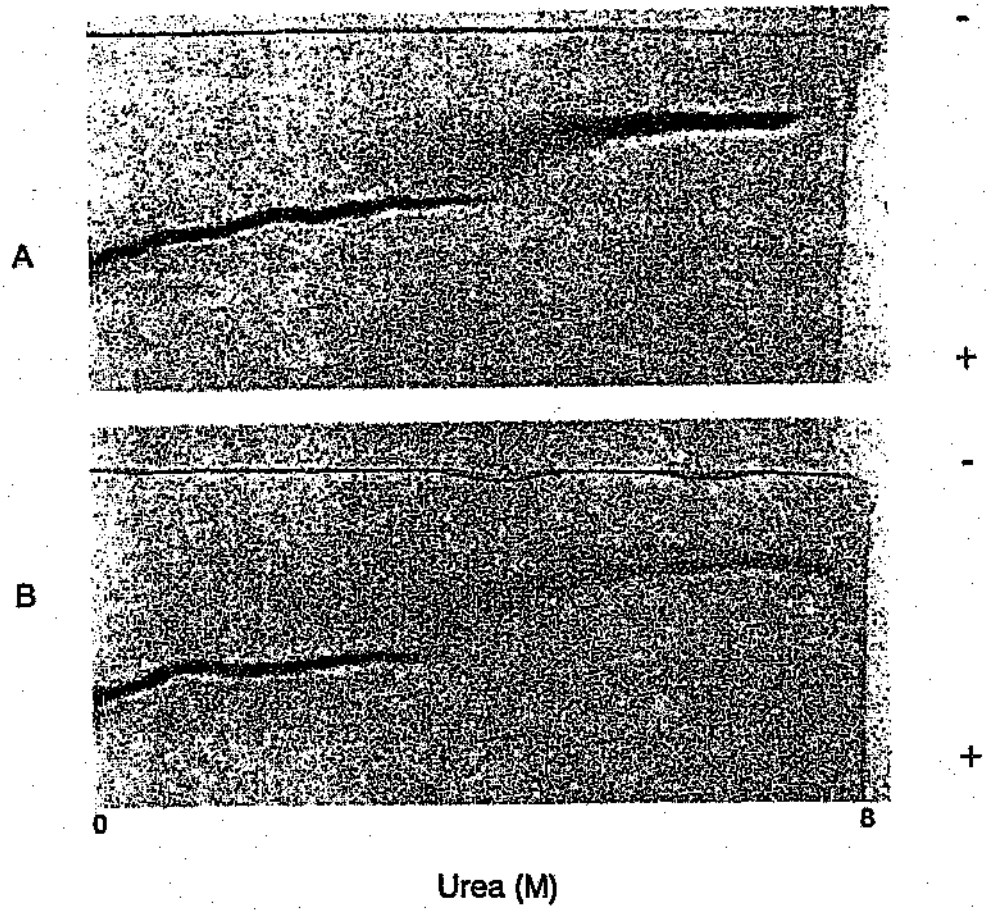


Figure 4.10. Urea-gradient gel electrophoretograms for native Sj26GST (A) and unfolded Sj26GST (B). The urea gradient (0 - 8 M) is perpendicular to the direction of electrophoresis from cathode to anode. Transition occurs between 4 M and 5 M urea.

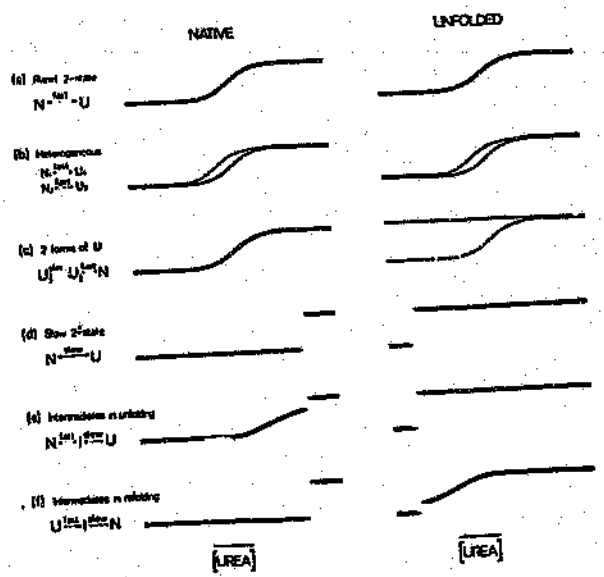


Figure 4.11. Urea gradient electrophoresis patterns expected for some folding transitions. The protein applied to the gel is either fully folded (left) or unfolded (right). The apparent rate constants for the transitions are described as "fast" or "slow" relative to the duration of the electrophoretic separation. The equilibrium midpoint is near the middle of the urea gradient in each case (Creighton, 1986).

(Figure 4.10A) and in the unfolded state (Figure 4.10B). The protein bands in the two electrophoretograms are very similar, showing a single transition between 4 and 5 M urea. Utilising both folding and unfolding electrophoretograms, a scheme has been developed for describing the rate of unfolding/refolding, and the presence of any intermediate species (Creighton, 1986, Mitchell, 1976). Figure 4.11 is a representation of the scheme taken from Creighton (1986). The discontinuity between the bands of the native and unfolded Sj26GST indicates that unfolding and refolding are relatively slow on the time scale of the electrophoresis near the transition midpoint (i.e. the two states are in slow equilibrium, allowing for their accumulation). The similarity of the two electrophoretograms would however imply that the unfolding/refolding process is not quite as slow as represented in Figure 4.11d, but the lack of continuity in the transition region would preclude it from a Figure 4.11a style of unfolding/refolding.

4.2.6 Size-Exclusion HPLC (SEC-HPLC)

SEC-HPLC is useful for studying protein unfolding/refolding due to its ability to resolve changes in the hydrodynamic volumes of macromolecules and therefore, to detect the presence of kinetically stable intermediate states provided that they are stable within the time scale of the chromatographic run (Corbett and Roche, 1984). Figure 4.12 shows elution profiles of Sj26GST in the presence of increasing concentrations of urea. Only two molecular species are observed; one species at the retention time for the native dimeric protein, and the other at the retention time for the unfolded monomer. The midpoint of the unfolding transition appears at 4.85 M urea and is similar to the midpoint value obtained by tryptophan fluorescence (Figure 4.5). Folded intermediates (such as a folded monomer) were not observed.

4.3 THERMAL UNFOLDING

The unfolding/refolding of a protein is dependent on the energetics of protein structure, and hence the thermodynamics thereof. The unfolding of small single-domain

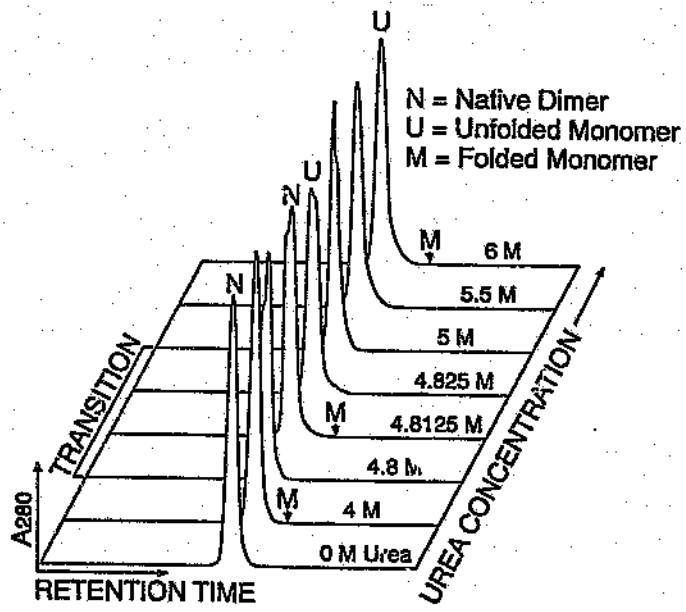


Figure 4.12. SEC-HPLC elution profiles of Sj26GST with increasing concentrations of urea. 20 μM Sj26GST was incubated for 1 hour at different concentrations of urea and applied to a SEC-HPLC column equilibrated with the same concentration of urea. The expected position of folded monomer if it had been observed is indicated by M.

proteins is well approximated by a two-state transition (Privalov, 1979). Their unfolding is however not exactly two-state, in that some intermediate states are present, but their populations are usually less than 5% (Freire and Biltonen, 1978, Privalov, 1979). In order to determine the two-state behaviour, a test of cooperativity needs to be applied. Only in the temperature-induced unfolding of proteins is this test possible (Makhatadze and Privalov, 1992). The test is dependent on the ratio of the calorimetric enthalpy to that of the van't Hoff enthalpy being equal to unity. No rigorous test for the cooperativity of denaturant-induced unfolding of proteins exists, since this would require the measurement of the number of bound ligands (denaturant molecules) to the protein upon unfolding (see eg. Ptitsyn, 1995b). Therefore the application of a two-state transition model for the solvent induced denaturation is based on indirect evidence which may lead to error (Pfeil and Privalov, 1976, Makhatadze and Privalov 1992). Accordingly, all thermodynamic characteristics of protein unfolding, obtained by solvent-induced denaturation, should be considered with caution (Makhatadze and Privalov, 1992). Due to the energetics of the protein's structure being understood as a difference between the folded and unfolded state, it is vital that the denaturation results in complete unfolding. As previously pointed out (Shortle, 1996), complete unfolding of the protein is only expected to occur in very high concentrations of chemical denaturant. The problem though is the large interaction of denaturant with protein, which is not well understood (see Makhatadze and Privalov, 1995 and the references therein). As a result, numerous empirical methods have been developed which exclude the effect of the interaction of the protein with denaturant molecules, from which certain thermodynamic information may be deduced (see the method of Pace, 1986 by which the conformational stability in this thesis was derived). This method makes use of the assumption that the process under investigation occurs in a two-state manner. Furthermore, the methods of linear extrapolation to zero denaturant concentration are not always well-substantiated (Pace, 1986). The only parameters which may be obtained by testing protein stability against chemical denaturant are the concentration of denaturant corresponding to the midpoint of the transition (Cm), the effective value of the Gibbs energy difference between the folded and unfolded states, and the parameter *m* specifying change in accessible surface

area. This information is however insufficient for the quantitative analysis of the protein structure.

Due to all the reservations regarding solvent-induced denaturation, heat was chosen as a denaturant as it provides more complete thermodynamic information (Makhatadze and Privalov, 1992).

4.3.1 Thermal Inactivation of Sj26GST

The loss of enzyme activity of Sj26GST with increasing temperature is shown in Figure 4.13A The midpoint of the transition (45-55°C) occurs at about 51°C. These data are similar to that reported for hGSTP1-1 (T_m ~ 52°C) (Kong *et al.*, 1993).

4.3.2 Thermal Gradient Gel Electrophoresis (TGGE)

TGGE was performed in order to trap any intermediates that may have occurred along the thermally induced unfolding pathway. The behaviour of Sj26GST when electrophoresed across a temperature gradient (20-80°C) is shown in Figure 4.13B. and indicates a single transition between 45 and 55°C. This corresponds with the enzyme inactivation data. The discontinuity between the bands of native and unfolded Sj26GST indicates that unfolding is slow on the time scale of the electrophoresis near the transition midpoint, thus allowing for their accumulation, and is similar to the UGGE results obtained in section 4.2.5 and Figure 4.10. The protein band in the pre-transitional region is non-parallel to the surface of application (top of the gel) due to the decreasing gel viscosity with increasing temperature (Birmes *et al.*, 1990).

4.3.3 Ultraviolet (UV)-Melting and Differential Scanning Calorimetry (DSC)

4.3.3.1 UV-Melting

UV-melting curves are extensively used in deriving thermodynamic parameters

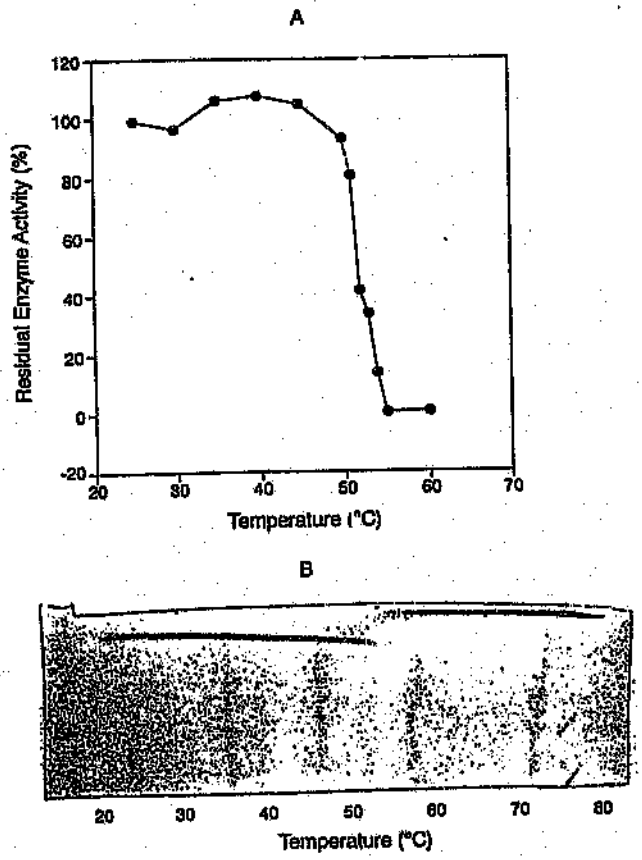
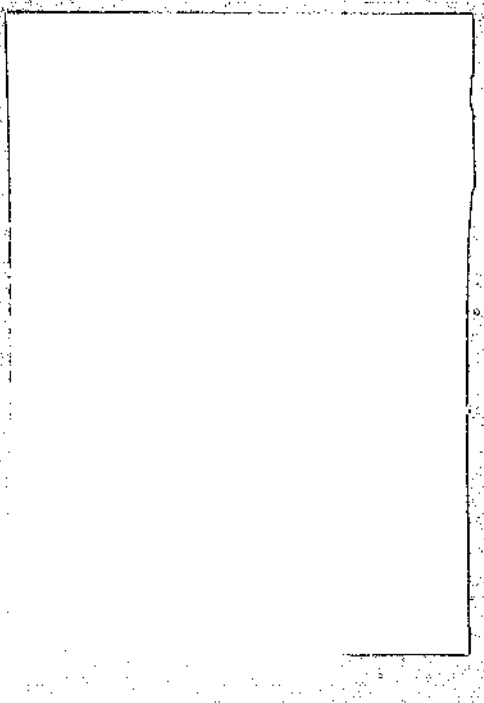


Figure 4.13. Profile of decrease in enzyme activity (A) and thermal-gradient gel electrophoretogram (B) of Sj26GST. In (A), 50μg of the enzyme was incubated at each temperature for 10 minutes, and the remaining activity assayed in 0.1 M potassium phosphate buffer, pH 6.5, 1mM GSH, 1mM CDNB. The transition occurs between 45°C and 55°C. In B, native Sj26GST was electrophoresed through the gel with a linear thermal gradient from 20°C to 80°C perpendicular to the direction of electrophoresis, from cathode to anode. The transition occurs between 45°C and 55°C. The transitions of both (A) and (B) coincide.



from DNA (Marky and Breslauer, 1987). In order to monitor the melting (unfolding) of a protein, heated at a constant rate, the increase in absorbance of buried chromophores (phenylalanine, tyrosine and tryptophan) upon unfolding was monitored at 280 nm. The UV-melting curve for Sj26GST and its first derivative are shown in Figure 4.14, and have transition midpoints (T_m) of about 58°C. The highly symmetrical derivatised melting curve indicates a monophasic transition from the folded to unfolded protein. After the scan the protein was found to have aggregated. Because urea-induced unfolding of Sj26GST did not result in any absorbance increase at 280 nm, the UV-melting curve could be attributed to aggregation rather than the increased exposure of the 3 types of chromophores. Upon unfolding only a small absorbance increase is expected at 292 nm (Schmid, 1990). In addition, numerous techniques are used to determine the concentration of a protein by making use of its absorbance at 280 nm (Layne, 1957; Perkins, 1986). None of these techniques requires the denaturation of the protein.

4.3.3.2 Differential Scanning Calorimetry (DSC)

Reliable thermodynamic information which describes the protein's unfolding can only be obtained by direct calorimetric measurement of the heat capacities of the native and denatured states of the protein. Since enthalpy and temperature are conjugate thermodynamic extensive and intensive variables, calorimetry is especially important for studying temperature induced processes. The functional dependence between such variables includes all the information on the macroscopic states which are realised in the process (Freire and Biltonen, 1978; Privalov and Potekhin, 1986; Privalov, 1989).This is why calorimetry provides such a rigorous test for the presence of intermediate states during the temperature-induced unfolding, i.e. on its correspondence to a two-state transition. A small deviation of the observed process from a two-state transition does not affect the precision of the calorimetrically derived enthalpy and entropy differences between the states. Applying the indirect equilibrium analysis in solvent-induced denaturation would render the results useless if there were any deviation from an assumed two-state reaction scheme.

Since the development of the first Differential Scanning Calorimeter (Privalov and Kechinashvili, 1974), DSC has been extensively used as the most definitive technique for establishing the presence or absence of intermediate species along the unfolding pathway of macromolecules. For reviews see Privalov & Potekhin, 1986; Sturtevant, 1987; Freire, 1994. The DSC curve for Sj26GST, indicates that the protein unfolds in a single transition where T_m-values are dependent upon protein concentration (Figure 4.15).

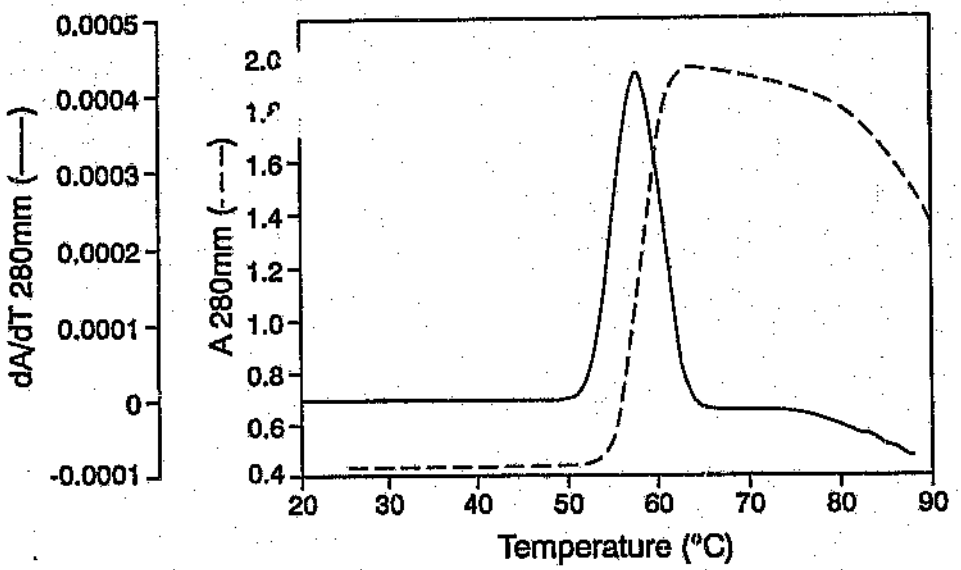


Figure 4.14. UV-melting curve of Sj26GST. A solution of Sj26GST was heated at a constant rate of 1°C/min, and the absorbance at 280 nm (- - -) measured continuously. (-), first-derivatised curve of the A_{280} data.

4.3.3.2.1 Scan rate dependence

Rescanning the same Sj26GST sample after completing the run, indicated the base line to be flat with no thermal effect visible on the DSC curve. Consequently, thermal denaturation of Sj26GST can be regarded as being irreversible (Freire *et al.*, 1990). Although the interpretation of DSC data are based on reversible processes, the validity of applying equilibrium thermodynamics to irreversible processes has been indicated (Privalov, 1982; Sturtevant, 1987), provided that the T_m values of the excess heat capacity curves, and their shape are independent of scan rate, (Freire *et al.*, 1990; Sanchez-Ruiz, 1992) i.e. that the denaturation process is not kinetically controlled.

In order to check for scan rate dependence, identical new samples of 15 μM Sj26GST were scanned at three different rates, and the data summarised in Table 4.2

The existence of scan rate effects can be attributed to the instrument dynamic response (Freire *et al.*, 1990). This slow time response of differential scanning calorimeters results in the distortion of the shape of the heat capacity function, and is more pronounced for sharp transition peaks and for instruments performed at fast scanning rates (Freire *et al.*, 1990).

In the case of a strongly rate-limited, irreversible DSC transition (see for example Galisteo *et al.*, 1991), protein concentration effects may still be expected to occur if the dissociation of the native multimeric protein into monomers takes place before the rate-determining step, or if the kinetics of the irreversible process is not first-order (Sanchez-Ruiz, 1992). This would imply that the existence of protein concentration dependence, (in the absence of scan rate independence) does not provide evidence that the denaturation process is not kinetically controlled (Sanchez-Ruiz, 1992). Ideally, the scan rate should be infinitely slow in order to combat this effect (Freire *et al.*, 1990). However, proteins are susceptible to being further destabilised when maintained at constant elevated temperatures (Freire *et al.*, 1990). Refer also to this study where thermal inactivation of enzyme activity (10 minute incubation at each temperature) results in a midpoint of denaturation approximately 10°C lower (T_m ~ 50°C) than that found for the DSC or UV-melting (T_m ~ 60°C), where the scan rate was 1°C/min.

Table 4.2 Effect of scan rate on the T_m value of Sj26GST. The T_m values were corrected for the dynamic response of the DSC according to the method of Freire *et al.* (1990).

Scan Rate (°C/min)	T_m (°C)	Corrected T_m (±0.2 °C)
0.5	58.9	60.0
1	59.6	60.3
2	61.7	60.5

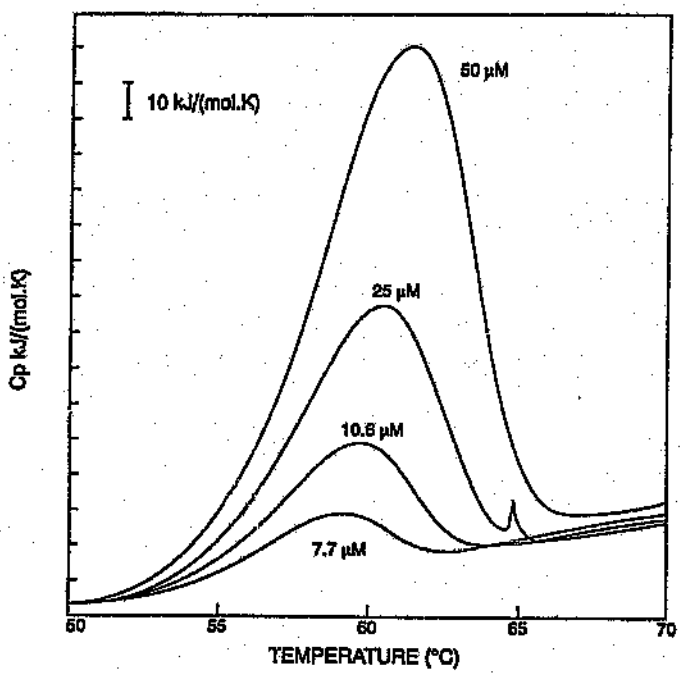


Figure 4.15. DSC curves for Sj26GST at different protein concentrations (7.7 μM to 50 μM)

4.3.4 The Unfolding Pathway of Sj26GST and the Implications of the Thermodynamic Parameters

Deconvolution of the curves in Figure 4.16 indicate that the melting of Sj26GST is a two state process (i.e., intermediates are present in negligible amounts). Table 4.3 lists the thermodynamic parameters derived from the fitted DSC scans. The correlation of $r = \Delta H_{VH} / \Delta H_{cal} \approx 1$, provides very strong evidence that the unfolding process is two-state (Sturtevant, 1987). Furthermore, the slope of a linear van't Hoff plot as shown in Figure 4.17 (Sturtevant, 1987), yields a $\Delta H_{VH} = 192$ kcal/mol. This is in reasonable agreement with $\Delta H_{cal} = 227$ kcal/mol (i.e., $r = \Delta H_{VH} / \Delta H_{cal} = 0.85$), supporting the assumption of a two-state process (Sturtevant, 1987). The average ΔC_p(1779 cal $mol^{-1}K^{-1}$) is due to the hydration of non-polar residues upon unfolding (Privalov and Gill, 1988; Myers, *et al.*, 1995). This value has been found to increase linearly with the molecular mass of the protein, and consequently it is possible to predict the ΔC_p (Myers, *et al.*, 1995) by using the following algorithm: $\Delta C_p = -251 + 0.19(\Delta ASA)$. (4.2)

For Sj26GST the theoretical $\Delta C_p = 7563$ cal $mol^{-1}K^{-1}$, which is approximately four-fold that of the value which was experimentally derived (average $\Delta C_p = 1779$ cal $mol^{-1}K^{-1}$). Since Sj26GST does not have an inordinately low number of buried nonpolar residues (81 of the 94 nonpolar residues having a 50% or less exposure to solvent), it would appear that the denatured state of the protein is not fully unfolded. Incomplete unfolding has previously been reported for other proteins when heat is used as a denaturant (Ptitsyn, 1995) with the non-unique structure of the denatured protein explaining the large standard deviation (SD) found in the ΔC_p values (SD = 739 cal $mol^{-1}K^{-1}$).

The major destabilising component of a protein's stability is the conformational entropy(ΔS_{conf}). Because there are many more arrangements of the flexible, thermally agitated unfolded chain than for the compact folded molecule, ΔS_{conf} is always positive (Baldwin and Eisenberg, 1987), where,

$$\Delta S_{conf} = k \ln w \text{ [amino acid number]} \quad (4.3)$$

where k is the Boltzmann constant = 1.98 cal $mol^{-1}K^{-1}$ and w is the number of accessible

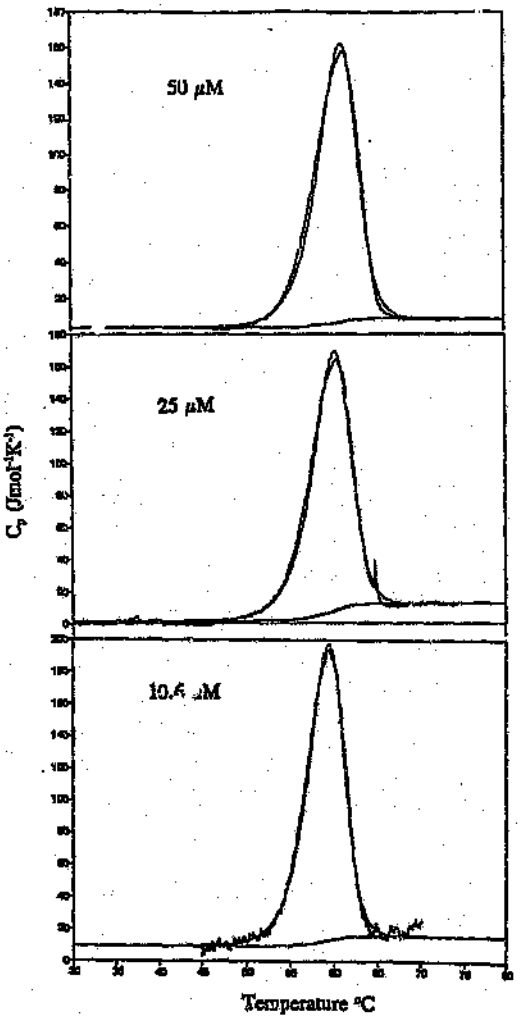


Figure 4.16. Deconvolution of the DSC unfolding curve. DSC curves were deconvoluted according to the method outlined in section 3.11. The raw data together with the fitted curve are shown for each protein concentration (A, B and C). The noise level in the 7.7 μM Sj26GST sample was too high to facilitate successful curve fitting.

Table 4.3 Table of thermodynamic parameters calculated from the fitted DSC data.

Concentration of S|26GST (μM)	ΔH_{cal} (kcal/mol)	ΔH_{vH} (kcal/mol)	T_m (K)	ΔS cal $mol^{-1}K^{-1}$	ΔCp cal $mol^{-1}K^{-1}$
50	220.3	229	334.4	344	1236
25	223.1	237	333.4	348	2621
10.6	238.0	231	332.6	372	1481
Average	227	233	-	354	1779
Standard Deviation	±9.5	±3.8	-	±14.7	±739.1

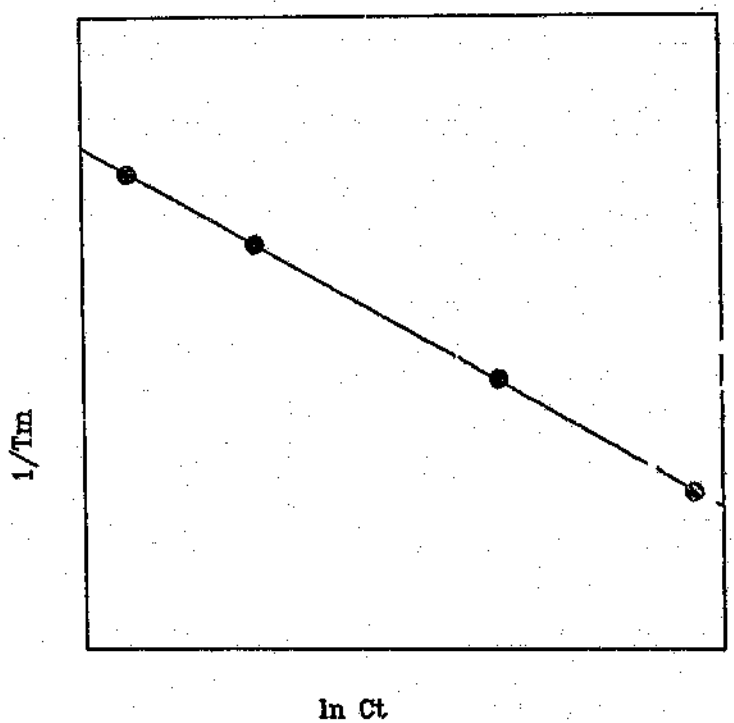


Figure 4.17. Arrhenius plot for the calculation of the ΔH_{VH}. The slope of the ln C_t vs $1/T_m$ (where C_t is the total protein concentration; Sturtevant, 1987) yields a ΔH_{VH} = 192 kcal/mol, i.e. r = $\Delta H_{VH}/\Delta H_{cal}$ = 0.85, supporting the two-state unfolding assumption.

states in which the system can exist. The value of w for a peptide bond has been taken as 4 (Dill, 1985; 1987). Therefore, from equation number $\Delta S_{conf} = 1191$ cal $mol^{-1}K^{-1}$, i.e. approximately 3 times the experimentally determined ΔS.

Further evidence of incomplete unfolding due to aggregation lies in the results of the UV-melting experiment. The increase in absorbance was attributed to the exposure of buried tyrosine, tryptophan and phenylalanine residues. However, unfolding the protein in the presence of urea resulted in no discernible increase in absorbance. Neither has it been reported to occur with only a small increase in absorbance expected upon unfolding at 292 nm (Schmid, 1990). In addition, numerous techniques which are used to determine the concentration of a protein make use of its absorbance at 280 nm (Layne, 1957; Perkins, 1986). None of these techniques requires the denaturation of the protein. Therefore, it would seem that the massive increase in absorbance at 280 nm is attributable to aggregation of the protein. If this is so, then interestingly enough, the aggregation begins with the onset of the transition in Figure 4.14. Thereby implying that the protein is already aggregating at that early stage of the transition, thus explaining the low ΔC_p values derived from the DSC where conditions are identical. The results of the UV-melting, nevertheless appear to provide excellent evidence for two-state unfolding using aggregation as a probe.

Several mechanisms have been suggested for the irreversible denaturation of proteins (Klibanov and Ahern, 1987). They may be classified into: firstly, covalent processes, such as peptide bond hydrolysis adjacent to aspartic acid residues, destruction of disulphide bonds, or deamination of asparagine or glutamine residues, and secondly, conformational processes, such as protein aggregation or the formation of incorrectly folded structures. Because these incorrectly folded structures may sometimes be recovered (e.g. by incubation at room temperature or through denaturation (eg. urea) and then rediluting out the denaturant), they are sometimes referred to as pseudo-irreversible (Freire, 1990), but for the analysis of the DSC data, or other thermal studies, these conformational transitions are regarded as being irreversible.

4.3.5 DSC in the presence of PZQ

A comparison of both the native and PZQ bound Sj26GST (15 μM Sj26GST incubated overnight with 150 μM PZQ) thermograms are shown in Figure 4.18. No changes in either the T_m or shape of the curves was discernible in the ligand-bound protein. This data is in agreement with the urea-induced denaturation of Sj26GST in the presence of PZQ, where no change in the stability was detected when compared to that of the ligand-free protein. (Further discussion of this section is included in the *addendum* on page 158)

4.4 CRITERIA FOR TWO-STATE UNFOLDING

Both solvent and thermal-induced unfolding of Sj26GST has been found to occur in a two-state manner. Various models have been reviewed which address the basis and criteria for the two-state cooperativity of protein folding (Chan *et al.*, 1995). I , with reference to this review that some of the experiments in this thesis are addressed.

Since the 1950's, three main models have been proposed to account for the two-state cooperativity of protein folding.

Model 1. "Helix-coil theory", (Schellman, 1958; Zimm and Bragg 1959; Poland and Scheraga, 1970) which is due to the local interactions among near neighbours in the sequence. The helix-coil cooperativity has been dismissed as being the principal basis for the unfolding/refolding of globular proteins because it is not two-state (Chan *et al.*, 1995). The forces involved are weak, and it is also not able to account for the proteins with β-sheets. In addition, no evidence exists that helix formation precedes the formation of a hydrophobic core in the folding pathway of proteins.

Model 2. The "sidechain packing model" of Shakhnovich and Finkelstein (1989) has attributed cooperativity to the jigsaw-puzzle-like complementary fits of side chains. This model has also been criticised (Chan *et al.*, 1995), as not being the basis of folding cooperativity since exact models (Dill, 1995) give no evidence that sidechain freezing is two-state. The side chain complementarities in proteins are also generally weak. Furthermore, the molten globule model predicted by this model is more native-like than

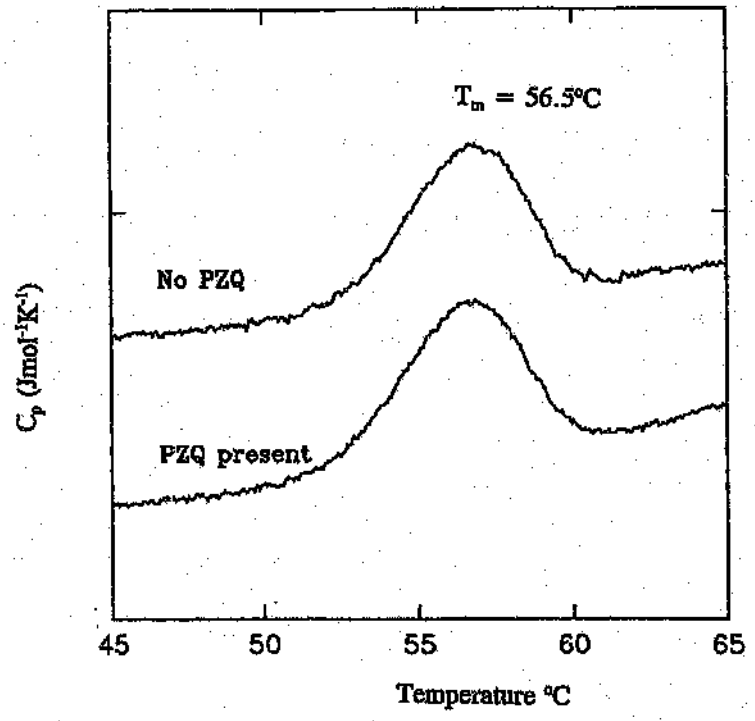


Figure 4.18. DSC curves for Sj26GST in the presence and absence of PZQ. Identical concentrations of protein (15 μM) were scanned in the presence or absence of 150 μM PZQ. No changes in the shape of the curves was discernible. The T_m values, while identical, were lowered by 3.5 °C when both samples contained DMSO which had been used to solubilise the PZQ.

that indicated by experiments (Chan *et al.*, 1995).

Model 3. In the "hydrophobic core collapse" model, cooperativity is due to the assembly of non-polar residues into a core. Exact model studies have indicated that this model provides two-state behaviour for some sequences of hydrophobic and polar monomers (Dill, 1995). The model is based on strong forces, and considerable experimental evidence exists for the kinetics predicted by the model. Included in the model is the formation of hydrophobic clusters and cores, which are concurrent with secondary structure formation. Finally, it predicts compact denatured states with sizes and degrees of disorder which are in agreement with experiments.

Protein unfolding/refolding for small single-domain proteins is known to involve two-state transitions (Lumry *et al.*, 1966, Privalov, 1979). These two-state transitions are dependent on cooperativity. The cooperativity is best described by Figure 4.19 which shows a cooperative process in which a system gradually changes from state A to B as a function of temperature or denaturant. Figure 4.19 b, c shows the experimental signature of a cooperative process, ie. a sigmoidal curve. Cooperativity may however be of two different types: two-state (first order, ie. there are two minima in free energy that are separated by a free energy barrier) or one-state (higher order, ie. when the free energy surface has only a single minimum and no barrier). Neither the observation of a sigmoidal curve (Figures 4.5 & 4.8), nor the observation of a peak of heat absorption by DSC (Figure 4.15; 4.16; 4.18), can distinguish a two-state from a one-state cooperative transition (Chan *et al.*, 1995). The definitive way of experimentally differentiating between these two types of transitions is by an analysis of the populations. At the midpoint of a two-state transition, two identifiable states will be populated. In contrast, one-state behaviour is defined by a single broad population peak at the midpoint. Whether the cooperativity of a process is one-state or two-state does not necessarily depend on the steepness of its sigmoidal curve (Figures 4.5 and 4.8). The *m*-value also serves as no criterion since the curves from exact model studies indicate that the steepness in a one-state processes can be as large as those in two-state processes. Importantly therefore, cooperativity depends on a molecular property which is much harder to measure: i.e. the distribution of the underlying populations.

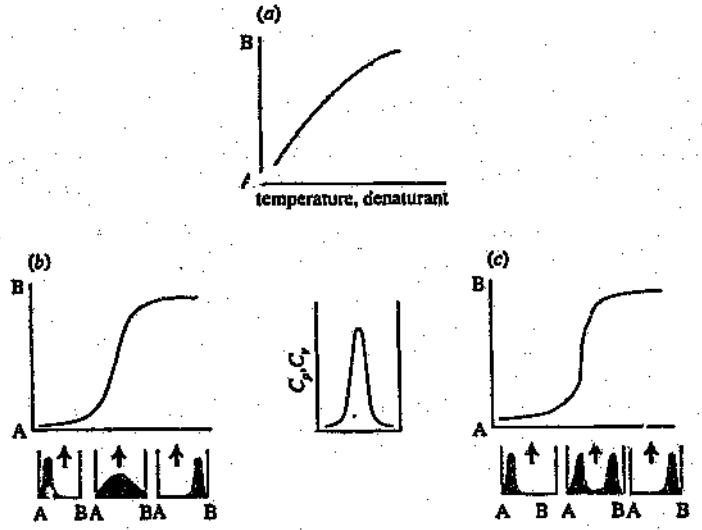


Figure 4.19. Models of cooperativity. Two states, A and B, such as native and denatured states, can change populations with an increasing denaturant concentration. (a) Gradual change, no cooperativity. (b) Cooperative transition of the one-state type. (c) Cooperative transition of the two-state type. Both one-state and two-state can have sigmoidal behaviour and absorption (a peak in the C_p plot); they cannot be distinguished on these bases, or from the steepness of the sigmoidal curve. The main distinction is whether there is a broad peak involving high populations of "intermediates" near the denaturant midpoint (one-state) or whether there are two populated states and less intermediate population (small plots at the bottom of (b) and (c) (from Chan *et al.*,1995).

Discriminating between the two forms of cooperativity from each other would therefore provide fundamental information regarding the free energy surface, and the underlying molecular mechanism of the cooperative process.

Various methods may be used to experimentally determine two-state behaviour:

1. The most definitive method is the determination of two distinct species or populations near the midpoint of the transition. In the case of slowly changing systems, this may be done by transport methods (Cann, 1970) such as SEC-HPLC (Uversky, 1993, Figure 4.12), or gel electrophoresis (4.10 and 4.13B).

2. Spectroscopic signal changes through the denaturation midpoint, which can be modelled to be linear combinations of the signals from the two limiting states and indicating two-state behaviour. If the spectroscopic signal near the denaturation midpoint cannot be fitted as a simple linear combination of the two limiting state signals, it provides evidence of one-state behaviour. In this case a significant population of "intermediate" molecules may have different spectral signatures and may contribute to the total signal.

3. The ratio of the calorimetrically determined enthalpy (determined directly via experimentation and independent of a pre-supposed model) to that of the van't Hoff enthalpy (determined by a pre-supposed model), provides additional evidence for two-state behaviour (Privalov, 1979, Privalov and Gill, 1988), see Table 4.3.

4. A less definitive measure of two-state behaviour is the coincidence of sigmoidal curves when different probes are used, see Figure 4.5 (Dill and Shortle, 1991).

Both points 3 and 4 are based on the assumption that both the native and denatured states are specific. As pointed out earlier (section 1.3), the denatured state is considerably more complex. With due consideration of the factors referred to it would appear as though the unfolding/refolding pathway of Sj26GST is indeed two-state.

4.5 CONCLUSIONS

The pGEX gene fusion system which expresses a fully functional Sj26GST is regarded as one of the four most popular systems currently in use (La Vallie and McCoy, 1995). Much of the dynamic behaviour and widespread use of this protein can be explained in terms of the structure-function studies of this thesis.

The tendency of Sj26GST to reversibly aggregate (Abeliovich and Shlomai, 1995) has been attributed to the exposed cysteine residues. The localisation of the cysteines, away from the active site, does not inhibit enzyme function and consequently a high specific activity is not indicative of unaggregated enzyme. When dealing with fusion proteins attached to Sj26GST, purification and storage under reducing conditions is imperative as this impedes aggregate formation which has been shown to interfere with the proteolytic cleaveage of some fusions from the Sj26GST tag (Abeliovich and Shlomai, 1995).

As in the case of all dimeric proteins and other GSTs (Dirr and Reinemer, 1991; Erhardt and Dirr, 1995; Sluis-Cremer and Dirr, 1995), Sj26GST has an unusually high conformational stability. Folded Sj26GST monomers are thermodynamically unstable and consequently dimerisation is critical to the protein's stability. However, binding of the antischistosomal drug and non-substrate ligand, PZQ at the subunit interface does not affect the enzyme's activity, stability, or its unfolding pathway. Neither does the binding of the physiological substrate GSH or PZQ induce any conformational changes in the protein's structure (Lim *et al.*, 1994; McTigue *et al.*, 1995). These small ligand-induced changes are in support of microconformational changes being introduced by ligand binding (Bico, *et al.*, 1995), and in contrast to the major structural changes as suggested by others (Li and Ishibashi, 1990; Nishihira, *et al.*, 1992; Nishihira *et al.*, 1993). While the binding of G-site ligands has been reported to destabilise some GSTs (Erhardt and Dirr, 1996; W. Parsons, unpublished results), the stability of binding other non-substrate ligands does not seem to have been affected in other GST classes (R.Ohlmeyer, unpublished results). Thus implying that the non-substrate binding site, whether occupied or vacant, of Sj26GST as well as some other GST classes (alpha, pi) does not impact on

the stability of the protein. Neither does the occupation of the site result in a different unfolding pathway for the protein. Thereby facilitating the uninterrupted catalytic function of the enzyme by its non-substrate ligand binding region.

The instability of the folded monomers causes the protein to unfold/refold in a highly cooperative two-state manner. The advantage of this high cooperativity is that effectively every residue "feels" the stability of the entire protein (Karplus and Shakhnovich, 1992). This cooperativity has also been regarded as affording protection to the protein (Ptitsyn, 1995b), as it would unfold before the rapid motions of the atoms at high temperatures, and in close proximity with each other, could induce any permanent structural damage. Despite this high cooperativity thermal denaturation of Sj26GST was irreversible with the low ΔC_p and ΔS values implying that thermal denaturation is incomplete. The early onset of aggregation as seen in the UV-melting experiment provides an answer to the incomplete unfolding of the protein.

In protein folding studies a domain is defined as being an independent folding unit (Privalov, 1982), and therefore, multi-domain proteins are expected to unfold/refold in a non two-state pathway. All GSTs are described as comprising two domains, however in light of the two-state unfolding/refolding pathway of this study this nomenclature, while assisting in the topological description of the protein, is not true. Furthermore, an algorithm designed to identify domains via a number of different criteria (Siddiqui and Barton, 1995) identifies the mu GST (PDB accession code 1GST) as comprising only one domain per subunit. Neither folded domains of Sj26GST (W. Parsons, unpublished results), or that of folded monomers (Erhardt and Dirr, 1995 and this study) have been found to exist. The large number of both inter-domain and subunit-subunit interactions appears to be the underlying cause for this phenomenon.

APPENDICES

APPENDIX A: FILE TRANSFER PROTOCOL (FTP)

FTP is a method of accessing remote computer systems and downloading their files. The advent of modern internet browsers such as the Netscape Navigator or Windows Explorer enable one to visit numerous *home pages* which provide an FTP facility, making the procedure easier than ever before. Occasionally, FTP sites are not accessible via internet browsers or because of congestion they may take too long to download files. In these instances one may use *anonymous ftp* described in this appendix.

To use FTP, both ends of a connection must be running a program that does FTP services running off a UNIX platform. In order to download a file from a remote system, you must start your FTP software and instruct it to connect to the FTP software running the remote machine. Usually when you connect to a remote system you must log in. This means that you must be a valid user, with a username and password for that remote machine. Because it is impossible to provide logins for everyone who wants access to a public archive, such as the Protein Data Bank, many systems use *anonymous ftp*.

The following procedure may then be followed for downloading the file **pdb1gne.ent** from the PDB. All words in bold are to be entered by the user who should then press the *return* key.

ftp> **open**
To: pdb.pdb.bnl.gov
Username: **anonymous**
Guest login ok, type your name as password.
Password: (enter your e-mail address which will not appear on the screen)
ftp> **cd fullrelease**

250 CWD command successful

ftp> cd uncompressed_files

250 CWD command successful

ftp> cd gn (the two letters preceding the .ent denote the directory that the file is in)

ftp> dir

-rw-r- -r- - 3root sys 182817 Dec 7 1994 pdb1gne.ent

226 Transfer complete

ftp> bin (changes the format of the file to a binary code)

200 Type set to I (conformation of binary code selection)

ftp> hash

Hash mark printing on (1024 bytes/ hash mark)

ftp> lcd a: (local current directory set to a; i.e. where the file will be saved)

ftp> get pdb1gne.ent

As the information on the file is now transferred the number of hash markings (#) increase until the message:

ftp>transfer complete and the number of bytes transferred is displayed. You should establish whether the number of bytes transferred is exactly the same as that in the file (182817 bytes in this instance).

ftp> quit (terminates contact with the remote server)

271 Goodbye

The downloaded file is now available in the directory you selected.

Some commonly used UNIX commands which may be beneficial are as follows:

pwd print working directory will display the directory you are currently in.

dir displays the directories available, the first column may have the following appearance:

drxwxrwxrwx

-rw-rw-rw-

Only the upper string of letters is a directory and is identified as such since it begins with

the letter *d*.

cd change directory, include a space and type the new directory name.

cd .. in the event of you wanting move upwards in the directory tree

Note: UNIX is case sensitive so that e.g. cd is not the same as CD.

APPENDIX B: GENERATION OF A SECOND MONOMER TO FORM A DIMERIC PROTEIN

Most of the GST coordinates which have been released to date have been supplied in the dimeric form. The reason for this is that the asymmetric units in the unit cells of the crystals were dimeric. The elucidation of the Sj26GST structures as well as that of the squid enzymes (1GSQ and 2GSQ) have resulted in only a single monomer being present in the respective asymmetric units. As a result the second monomer needed to be generated using the matrix supplied in the heading of the coordinate file.

The procedure involved in generating the second subunit is essentially very simple; firstly, it involves a rotation i.e. the positioning of the new monomer into its correct orientation, and secondly, the translation of the correctly oriented monomer into its new position.

Two types of coordinates exist:

1 Fractional coordinates are allocated positions which are expressed as fractions of the unit cell with the minimum value being zero and the maximum value unity.

2 Orthogonal coordinates designate actual physical positions of the atoms. Accordingly they can be used to measure distances between individual atoms which may be expressed in a unit of length (usually Å). As a result all coordinate files are given in orthogonal coordinates.

Carrying out the matrix transformation currently involves one of either two methods:

Method 1 This method is easier in that the matrix facilitates the generation of the new coordinates directly from the orthogonal coordinates without having to use any

fractional coordinates at all.

Method 2 The orthogonal coordinates need to be converted into fractional coordinates before the transformation is carried out. Having carried out the transformation, the inverse matrix of the transformed fractional coordinates is calculated which then results in the desired orthogonal coordinates.

Both these methods require that the coordinate files be opened in a word processing file by converting the format from an ASCII text. Once the matrix has been identified in the REMARKS section of the file (the matrix of pdb1gne.ent is shown below),

SCALE1	0.010555	0.000000	0.000000	0.00000	1GNE 182
SCALE2	0.000000	0.010555	0.000000	0.00000	1GNE 183
SCALE3	0.000000	0.000000	0.017203	0.00000	1GNE 184

the coordinates should be inspected in order to ascertain that the columns are continuous. In the event of there being a break in the columns due to an overflow of the line, selecting a landscape format for the page should rectify the situation. The entire remarks section should now be deleted so that atom 1 appears at the top of the file. With the exception of the x, y and z coordinates, all other columns as well as the connectivity instructions at the bottom of the file should be deleted as well. In Word Perfect 5.1 column deletion is easily accomplished by selecting the *Edit - Select - Rectangle* command. The file is now saved in an ASCII text format ready for importation into a spread-sheet programme such as Sigma Plot.

The default setting for the importation of the file into Sigma Plot (version 5.0, Jandel Scientific) is correct, and once the file is opened it should be inspected for continuity of the columns. Any discontinuity in the columns can almost certainly be attributed to not having saved or opened the file in an ASCII format and the procedure should be repeated.

When orthogonal coordinates are used to directly generate the new orthogonal coordinates (orthogonal - orthogonal) then one may proceed according to the method laid down below:

Orthogonal-Orthogonal

$$r = (a, b, c, o)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x\,a + y\,b + z\,c$$

In the case of 2GSQ, the supplied matrix is as follows:

-0.499960	0.86605	-0.000020	-0.00072
0.866050	0.499960	-0.000040	-0.00040
-0.000020	-0.000030	-1.0000	0.00168

From this matrix the new coordinates x_2, y_2, and z_2 are generated from the original coordinates x_1, y_1, and z_1 as follows:

$x_2 = (x_1\ (0.499960) + y_1\ (0.866050) + z_1\ (-0.000020)) + (-0.00072))$

$y_2 = (x_1\ (0.866050) + y_1\ (0.499960) + z_1\ (-0.000040)) + (-0.00040))$

$z_2 = (x_1\ (-0.000020) + y_1(-0.00003) + z_1\ (-1) + (0.00168))$

The new coordinates are then exported in an ASCII format for importation by the word processor.

Open the original coordinates of the file in the word processor. Move the CONNECT section at the bottom of the file a few spaces down, and then copy and paste the entire set of coordinates (atom 1 to the last hetero atom of subunit A) in the space created by the displaced CONNECT section. Cut and delete the original set of coordinates in what is now subunit B. Place the cursor below the word "END" i.e. the very last word of the file and import the new coordinates from the spread-sheet. Cut the new coordinates and paste them into their coinciding positions in subunit B. Ensure that the coordinate and atom numbers coincide by moving to the bottom of the file and then

shifting the CONNECT section upto the last HETERO atom. The file should now be saved and the result viewed.

Orthogonal-Fractional-Orthogonal

The second method is carried out using the same principle, only that once the fractional coordinates are generated by the matrix function, they are then adjusted according to the details supplied with the matrix. In order to return to fractional coordinates the inverse matrix is used. The programme Matlab (Matrix Laboratories) provides a very easy means of carrying out this procedure. Importantly, Matlab requires a semicolon (;) after each z coordinate. Upon exporting from Sigma Plot the facility exists to insert the semicolon. It therefore is easier to go the word processor-Sigma Plot-Matlab route in order to carry out the transformation.

Differentiating between the two types of transformations is easily carried out by inspection of the matrix itself; the end result of the conversion of orthogonal to fractional coordinates will have values ranging between 0 and 1. On the other hand, the conversion of orthogonal to orthogonal coordinates involve matrices where the numbers are considerably larger, as can be seen by comparing the matrices of 2GSQ (method 1) and 1GNE (method 2).

REFERENCES

Abeles, R.H., Frey, P. A. and Jencks, W.P. (1992) Forces and interactions in aqueous solution. In: *Biochemistry* Jones and Bartlett Publishers, Boston. pp.212-234.

Abeliovich, H. and Shlomai, J. (1995) Reversible oxidative aggregation obstructs specific proteolytic cleavage of glutathione S-transferase fusion proteins. *Anal. Biochem.* 228, 351-354.

Alberty, R.A. (1987) *Physical Chemistry*. Seventh edition. John Wiley & Sons. Singapore.

Alonso, D.O.V. and Dill, K.A. (1991) Solvent denaturation and stabilization of globular proteins. *Biochemistry* 30, 5974-5985.

Anfinsen, C.B. (1973) Principles that govern the folding of protein chains. *Science* 181, 223-230.

Anfinsen, C.B., Haber, E., Sela, M. and White Jr, F.H. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA.* 47, 1309-1314.

Arcus, V.L., Vuilleumier, S., Freund, S.V., Bycroft, M. and Fersht, A.R. (1994) Towards solving the folding pathway of barnase: the complete backbone ^{13}C, ^{15}N and ^{1}H NMR assignments of its pH-denatured state. *Proc. Natl. Acad. Sci. USA.* 91, 9412-9416.

Armstrong, R.N. (1991) Glutathione S-transferases: Reaction mechanism, structure, and function. *Chem. Res. Toxicol.* 4: 131-140.

Armstrong, R.N. (1997) Structure, catalytic mechanism, and evolution of the glutathione S-transferases. *Chem. Res. Toxicol.* 10, 2-18.

Bader, R. and Leisinger, T. (1994) Isolation and characterization of the *Methylophilus sp.* strain DM 11 gene encoding dichloromethane dehalogenase/glutathione S-transferase. *J. Bacteriol.* 176, 3466-3473.

Baldwin, R.L. and Eisenberg, D. (1987) Protein stability, in *Protein Engineering* (Alan R. Liss ed. Inc) pp. 127-148.

Barlow, D.J. and Thornton, J.M. (1983) Ion-pairs in proteins. *J. Mol. Biol.* 168, 867-885.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F.Jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) The protein data bank: a computer - based archival file for macromolecular structures. *J. Mol. Biol.* 112, 535-542.

Bico, P.C.G. (1994) *The ligand-binding function of the porcine class Pi glutathione S-transferase.*MSc dissertation. The University of the Witwatersrand.

Bico, P., Erhardt, J., Kaplan, W. and Dirr, H. (1995) Porcine class pi glutathione S-transferase: anionic binding and conformational analysis, *Biochim. Biophys. Acta* 1247, 225-230.

Birdsall, B., King, R.W., Wheeler, M.R., Lewis, C.A., Goode, S.R., Dunlap, R.B. and Robers, G.C.K (1983) Correction for light absorption in fluorescence studies of protein-ligand interactions. *Anal. Biochem.* 132, 353-361.

Birmes, A., Sättler, A., Maurer, K. and Riesner D. (1990) Analysis of the

conformational transitions of proteins by temperature-gradient gel electrophoresis. *Electrophoresis* 11, 795-801.

Booth, J., Boyland, E. and Sims, P. (1961) An enzyme from rat liver catalysing conjugates with glutathione, *Biochem. J.* 79, 516-523.

Borman, S. (1995) Researchers advance ability to predict structures of folded proteins. *Chem. Engineer.* 6, 20-26.

Bowie, J.U. and Sauer R.T. (1989) Equilibrium dissociation and unfolding of the arc repressor dimer. *Biochemistry* 28, 7139-7143.

Boyland, E. and Chasseaud, L.F. (1969) The role of glutathione and glutathione S-transferases in mercapturic acid biosynthesis. *Adv. Enzymol.* 32, 173-219.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248-254.

Bryngleson, J.D. and Wolynes, P.G. (1987) Spin glasses and the statistical mechanics of protein folding. *Proc. Nat'l. Acad. Sci. USA.* 84, 7524-7528.

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.

Campbell, I.D. and Dwek, R.A. (1984) *Biological Spectroscopy*. Benjamin Cummings Publishing. pp. 91-125.

Cann, J.R. (1970) *Interacting macromolecules: the theory and practice of electrophoresis, ultracentrifugation, and chromatography*, New York: Academic Press

Chan, H.S. and Dill, K.A. (1991) Polymer principles in protein structure and stability. *Annu. Rev. Biophys. Biophys-Chem.* 20,447-490.

Chan, H.S., Bromberg, S. and Dill, K.A. (1995) Models of cooperativity in protein folding. *Phil. Trans. R. Soc. Lond.* B. 348, 61-70.

Carter, D.C., Rüker, F., Ho, J.X., Lim, K., Keeling, K., Gilliland, G.L. and Ji, X. (1994) Fusion proteins as alternate crystallization paths to difficult structure problems. *Protein Pept. Lett.* 1, 175-178.

Chung, C.T., Niemela, S.L. and Miller, R.H. (1989) One-step preparation of competent *Escherichia coli* : transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA.* 86, 2172-2175.

Collins, K.D. and Washabaugh, M.W. (1985) The hofmeister effect and the behaviour of water at interfaces. *Q. Rev. Biophys.* 18, 323-422.

Corbett, R.J.T. and Roche, R.S. (1984) Use of high-speed size-exclusion chromatography for the study of protein folding and stability. *Biochemistry* 23, 1888-1894.

Creighton, T.E. (1979) Electrophoretic analysis of the unfolding of proteins by urea. *J. Mol. Biol.* 129, 235-264.

Creighton, T.E. (1986) Detection of folding intermediates using urea-gradient electrophoresis. *Methods Enzymol.* 131, 156-172.

Creighton, T.E. (1993) *Proteins: structures and molecular properties*, 1st ed. USA: W.H. Freeman and Company.

Daggett, V., and Levitt, M. (1993) Protein unfolding pathways explored through molecular dynamics simulations. *J. Mol. Biol.* 232, 600-619.

Daggett, V., Li, A., Itzhaki, L.S., Otzen, D.E. and Fersht, A.R. (1996) Structure of the transition state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* 257,430-440.

DeBoer, H.A., Comstock, L.J. and Vasser, M. (1983) The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. *Proc. Natl. Acad. Sci. USA.* 80, 21-25.

Dill, K.A. (1985) Theory for the folding and stability of globular proteins. *Biochemistry* 24, 1501-1509.

Dill, K.A. (1987) The stability of globular proteins, in *Protein Engineering* (Alan R. Liss, Inc.) pp. 187-192.

Dill, K.A. (1990) Dominant forces in protein folding. *Biochemistry* 29, 7133-7155.

Dill, K.A. and Shortle, D. (1991) Denatured states of proteins. *Ann. Rev. Biochem.* 60, 795-825.

Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D. and Chan, H.S. (1995) Principles of protein folding-a perspective from simple exact models. *Protein Sci.* 4, 561-602.

Dill, K.A. and Stigter, D. (1995) Modeling protein stability as heteropolymer collapse. *Adv. Prot. Chem.* 46, 59-104.

Dinner, A.R., Šali, A. and Karplus, M. (1996) The folding mechanism of larger model proteins: role of native structure. *Proc. Natl. Acad. Sci. USA.* 93, 8356-8361.

Dirr, H.W., Mann, K., Huber, R., Ladenstein, R. and Reinemer, P. (1991) Class π glutathione S-transferase from pig lung. Purification, biochemical characterization, primary structure and crystallization. *Eur. J. Biochem.* 196, 693-698.

Dirr, H.W. and Reinemer, P. (1991) Equilibrium unfolding of class π glutathione S-transferase. *Biochem. Biophys. Res. Comm.* 180, 294-300.

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.

Drexler, K.E. (1996) *Engines of creation, the coming era of nanotechnology*. Fourth Estate. London.

Duff, E.M. Kowalczyk, P.J. and Jorgensen, W.L. (1993) Do denaturants interact with aromatic hydrocarbons in water? *J. Am. Chem. Soc.* 115, 9271-9275.

Edsall, J.T. (1995) Hsien Wu and the first theory of protein denaturation (1931). *Adv. Prot. Chem.* 1-26.

Erhardt, J. and Dirr, H. (1995) Native dimer stabilizes the subunit tertiary structure of porcine class pi glutathione S-transferase. *Eur. J. Biochem.* 230, 614-620.

Fersht, A.R., Matouschek, A. and Serrano, L. (1992) The folding of an enzyme I. Theory of protein engineering analysis of stability and pathways of protein folding. *J. Mol. Biol.* 224, 771-782.

Fersht, A.R (1993) Protein folding and stability - the pathway of folding of barnase. *FEBS Lett.* 325, 5-16.

Fersht, A. (1995) Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. USA.* 92, 10869-10873.

Finkelstein, A.V. and Shakhnovich, E.I. (1989) Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers* 28, 1681-1689.

Frangioni, J.V. and Neel, B.G. (1993) Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. *Anal. Biochem.* 210, 179-187.

Franks, F. (1995) Protein destabilization at low temperatures. *Adv. Prot. Chem.* 46, 105-139.

Freedman, R.B. (1992) Protein folding in the cell. In *Protein Folding* (Creighton, T.E., ed.), pp. 455-539, Feeman and Company, New York.

Freire, E. (1994) Statistical thermodynamic analysis of differential scanning calorimetry data: Structural deconvolution of heat capacity function of proteins. *Methods Enzymol.* 240, 502-530.

Freire, E. and Biltonen, R.L. (1978) Statistical mechanical deconvolution of thermal transitions in macromolecules I theory and application to homogenous systems. *Biopolymers* 17, 463-479.

Freire, E., van Osdol, W.W., Mayorga, O.L. and Sanchez-Ruiz, J.M. (1990) Calorimetrically determined dynamics of complex unfolding transitions in proteins. *Annu. Rev. Biophys. Biophys. Chem.* 19, 159-188.

Galisteo, M.L., Mateo, P.L. and Sanchezruiz, J.M. (1991) Kinetic-study on the irreversible thermal-denaturation of yeast phosphoglycerate kinase. *Biochemistry* 30,

2061-2066.

García-Sáez, I., Párraga, A., Phillips, M.F., Mantle, T. J. and Coll, M. (1994) Molecular structure at 1.8 Å of mouse liver class pi glutathione S-transferase complexed with S-(*p*-nitrobenzyl)glutathione and other inhibitors. *J. Mol. Biol.* 237, 298-314.

Garel, J-R. (1992) Folding of large proteins: Multidomain and multisubunit proteins in *Protein Folding* (Creighton, T.E. ed.) pp. 455-539, W.H. Freeman and Company, New York.

Georgiou, G. and Valax, P. (1996) Expression of correctly folded proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* 7,190-197.

Gilman, A.G., Rall, T.W., Nies, A.S. and Taylor, P. (1993) *Goodman and Gilman's, the pharmacological basis of therapeutics.* Eighth editon. Macmillan Publishing Company. USA. pp. 967-969.

Goldenberg, D.P. (1990) Analysis of protein conformation by gel electrophoresis. in: Creighton, T.E. ed. *Protein structure: A practical approach, 2nd edition*. Oxford, UK: IRL press at Oxford University Press. pp. 225-250.

Habeeb, A.F.S.A. (1972) Reaction of protein sulfhydryl groups with Ellman's reagent, *Methods Enzymol.* 25, 457-464.

Habig, W.H. and Jakoby, W.B. (1981) Assays for differentiation of glutathione S-transferases. *Methods Enzymol.* 77, 398-405.

Hagel, P., Gerding, J.J.T., Fieggen, W. and Bloemendal, H. (1971) Cyanate formation in solutions of urea I. Calculation of cyanate concentrations at different temperatures and pH. *Biochem. Biophys. Acta* 243, 366-373.

Harder, A., Andrews, P. and Thomas, H. (1987) Praziquantel: mode of action. *Biochem. Soc. Trans.* 15, 68-70.

Hartman, J., Daram, P., Frizzell, R.A., Rado, T., Benos, D.J. and Sorscher, E.J. (1992) Affinity purification of insoluble recombinant fusion proteins containing glutathione S-transferase. *Biotechnol. Bioeng.* 39, 828-832.

Hendrickson, W.A. (1979) Transformation to optimise the superposition of similar structures. *Acta Crystallogr. Sect. A.* 35, 158-163.

Holmgren, A., Söderberg, B-O., Eklund, H. and Bränden, C-I. (1975) Three-dimensional structure of *Escherichia coli* Thioredoxin-S_2 to 2.8 Å resolution. *Proc. Natl. Acad. Sci. USA.*72, 2305-2309.

Hughes, A. (1993) Rates of amino acid evolution in the 26- and 28- kDa glutathione S-transferases of *Schistosoma. Mol. Biochem. Parasitol.* 58, 43-52.

Jakoby, W.B. and Ziegler, D.M. (1990) The enzymes of detoxication. *J. Biol. Chem.* 265, 20715-20718.

Ji, X., Zhang, P., Armstrong, R.N. and Gilliland, G.L. (1992) The three dimensional structure of a glutathione S-transferase from mu gene class. Structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2 Å resolution. *Biochemistry* 31, 10169-10184.

Ji, X., von Rosenvinge, E.C., Johnson, W.W., Tomarev, S.I., Piatigorsky, J., Armstrong, R.N. and Gilliland, G.J. (1995) Three dimensional structure, catalytic properties, and evolution of a sigma class glutathione S-transferase from squid. *Biochemistry* 34, 5317-5328.

Ji, X., von Rosenvinge, E.C., Johnson, W.W., Armstrong, R.N. and Gilliland, G.L. (1996) Location of a potential transport binding site in a sigma class glutathione

transferase by x-ray crystallography. *Proc. Natl. Acad. Sci. USA.* 93, 8208-8213.

Jones, S. and Thornton, J.M. (1995) Protein-protein interactions: a review of protein dimer structures. *Prog. Biophys. Molec. Biol.* 63, 31-65.

Jorgensen, W.L., Gao, J. and Ravimohan, C. (1985) Monte Carlo simulations of alkanes in water: hydration numbers and the hydrophobic effect. *J. Phys. Chem.* 89, 3470-3473.

Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* 22, 2577-2637.

Karlin, S. and Zhu, Z-Y. (1996) Characterisation of diverse residue clusters in protein three-dimensional structures. *Proc. Natl. Acad. Sci. USA.* 93, 8344-8349.

Karplus, M. and Shakhnovich, E. (1992) Protein folding: Theoretical studies of thermodynamics and dynamics. In *Protein Folding* (Creighton, T.E. ed.), pp. 129-195, Freeman and Company, New York.

Karshikoff, A., Reinemer, P., Huber, R. and Ladenstein, R. (1993) Electrostatic evidence for the activation of the glutathione thiol by Tyr7 in π-class glutathione transferases. *Eur. J. Biochem.* 215, 663-670.

Katanka, M., Hagihara, Y., Miharo, K. and Goto, Y. (1993) Molten globule of cytochrome c studied by small angle X-ray scattering. *J. Mol. Biol.* 229, 591-596.

Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation. *Adv. Prot. Chem.* 14, 1-63.

Kendrew, J.C., Bodo, G., Kintzis, H.M., Parrish, R.G., Wyckoff, H. and Phillips, D.C. (1958) A three-dimensional model of the myoglobin molecule obtained by X-ray

analysis. *Nature* 181, 662-666.

Klibanov, A.M (1983) Stabilisation of enzymes against thermal inactivation. *Adv. Appl. Microbiol.* 29, 1-28.

Klibanov, A.M. and Ahern, T.J (1987) Thermal stability of proteins, in *Protein Engineering* (Alan R. Liss, Inc.) pp. 213-218.

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, 537-540.

Kong, K.H. Inoue, H. and Takahashi, K. (1993) Site-directed mutagenesis study on the roles of evolutionally conserved aspartic acid residues in human glutathione S-transferase P1-1. *Protein Eng.* 6, 93-99.

Kuwajima, K. (1977) A folding model of alpha-lactalbumin deduced from three-state denaturation mechanism. *J. Mol. Biol.* 114, 241-258.

Laemmli, U.K. (1970) Cleaveage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.

Lakowicz, J.R. (1983) Principles of Fluorescence Spectroscopy. pp. 112-151, Plenum Press, New York.

Laskowski, R.A., Luscombe, N.M, Swindells, M.B. and Thornton, J.M. (1996) Protein clefts in molecular recognition and function. *Prot.ein Sci.* 5, 2438-2452.

LaVallie, E.R. and McCoy, J.M. (1995) Gene fusion expression systems in *Escherichia coli. Curr. Opin. Biotechnol.* 6, 501-506.

Layne, E. (1957) Spectrophotometric and turbidimetric methods for measuring proteins. *Methods Enzymol.* 3, 447-454.

Lee, A.G. (1982) Membrane studies using fluorescence spectroscopy. In: *Techniques in the life sciences*. Elsevier Scientific Publishers, Ireland.

Lee, B. and Richards, F.M. (1971) Interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55, 379-400.

Li, M. and Ishibashi, T. (1990) Evidence for identifying fatty-acids in rat liver glutathione S-transferase and its possible involvement in the secondary structure. *J. Biochem.* 108, 462-465.

Levinthal, C. (1968) Are there pathways for protein folding? *J. Chem. Phys.* 65, 44-45.

Levinthal, C. (1969) How to fold graciously. In *Mossbauer Spectroscopy in Biological Systems* (DeBrunner, P., Tsibris, J. and Munck, E., eds), pp. 22-24. University of Illinois Press, Urbana.

Lifshitz, I.M., Grossberg, A.Y. and Khokhlov, A.R. (1978) Some problems of statistical physics of polymer chains with volume interaction. *Rev. Mod. Phys.* 60, 683-713.

Lim, K., Ho, J.X., Keeling, K., Gilliland, G.L., Ji, X., Rüker, F. and Carter, D.C. (1994) Three-dimensional structure of *Schistosoma japonicum* glutathione S-transferase fused with a six-amino acid conserved neutralizing epitope of gp41 from HIV. *Protein Sci.* 3, 2233-2244.

Listowsky, I. (1993) High capacity binding by glutathione S-transferases and glucocorticoid resistance. In: *Structure and function of glutathione transferases* (Tew, K.D., Pickett, C.B., Mantle, T.J., Mannervik, B., Hayes, J.D. eds) pp 199-210,

CRC Press, Boca Rotan.

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.

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.

Lumry, R., Biltonen, R. and Brandts, J.F. (1966) Validity of the 'two - state' hypothesis for conformational transitions of proteins. *Biopolymers* 4, 917-944.

Makhatadze, G.I. and Privalov, P.L. (1992) Protein interactions with urea and guanidinium chloride-a calorimetric study. *J.Mol. Biol.* 226, 491-505.

Makhatadze, G.I. and Privalov, P.L. (1995) Energetics of protein structure. *Adv. Prot. Chem.* 307-425.

Mannervik, B. (1985) The isoenzymes of glutathione transferase. *Adv. Enzymol.* 57, 357-417.

Mannervik, B. and Danielson, U.H. (1988) Glutathione transferases - structure and catalytic activity, *CRC Crit. Rev. Biochem.* 23, 283-337.

Mannervik, B., Awasthi, Y.C., Board, P.G., Hayes, J.D., Di Ilio, C., Ketterer, B., Listowsky, I., Morgenstern, R., Muramatsu, M., Pearson, W.R., Pickett, C.B., Sato, K., Widersten, M. and Wolf, C.R. (1992) Nomenclature for human glutathione transferases, *Biochem. J.* 282, 305-308.

Manoharan, T.H., Gulick, A.M., Puchalski, R.B., Servais, A.L. and Fahl, W.E. (1992) Structural studies on human glutathione S-transferase π. *J. Biol. Chem.* 267,

18940-18945.

Marky, L.A. and Breslauer, K.J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* 26, 1601-1620.

Matouschek, A., Kellis, J.T. Jr, Serrano, L. and Fersht, A.R. (1989) Mapping the transition state pathway of protein folding by protein engineering. *Nature* 342, 122-126.

Matouschek, A., Kellis, J.T. Jr, Serrano, L., Bycroft, M. and Fersht, A.R. (1990) Transient folding intermediates characterised by protein engineering. *Nature* 346, 440-445.

McTigue, M.A., Williams, D.R. and Tainer, J.A. (1995) Crystal structure of a schistosomal drug and vaccine target: Glutathione S-transferase from *Schistosoma japonica* and its complex with the leading antischistosomal drug praziquantel. *J. Mol. Biol.* 246, 21-27.

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, 409-414.

Meyer, D.J., Xia, C., Coles, B., Chen, H., Reinemer, P., Huber, R. and Ketterer, B. (1993) Unusual reactivity of Tyr-7 of GSH transferase P1-1. *Biochem. J.* 293, 351-356.

Meyer, D.J. and Thomas, M. (1995) Characterisation of rat spleen prostaglandin H D-isomerase as a sigma class GSH transferase. *Biochem. J.* 311, 739-742.

Mitchell, R.M. (1976) The spatial distribution of moving macromolecules undergoing isomerization. *Biopolymers* 15, 1717.

Mirny, L.A, Abkevich, V. and Shakhnovich, E.I. (1996) Universality and diversity of the protein folding scenarios: a comprehensive analysis with the aid of a lattice model. *Folding Design* 1, 103-116.

Morjana, N.A., McKeone, B.J. and Gilbert, H.F. (1993) Guanidine hydrochloride stabilization of a partially unfolded intermediate during the reversible denaturation of protein disulfide isomerase. *Proc. Natl. Acad. Sci. USA*. 90, 2107-2111.

Myers, J.K., Pace, C.N. and Scholtz, J.M. (1995 Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein folding. *Protein Sci*. 4, 2138-2148.

Needleman, S.B. and Wunsch, C.D. (1970) A general method applicable to the search for similarities in the amino acid sequences of two proteins. *J. Mol. Biol.* 48, 433.

Neet, K.E. and Timm, D.E. (1994) Conformational stability of dimeric proteins: quantitative studies by equilibrium denaturation. *Protein Sci*. 3, 2167-2174.

Nemoto, T., Ota, M., Ohara-Nemoto, Y. and Kaneko, M. (1995) Identification of dimeric structure of proteins by use of glutathione S-transferase-fusion expression system. *Anal. Biochem.* 227, 396-399.

Nishihira, J., Ishibashi, T., Sakai, M., Nishi, S. Kumazaki, T. and Hatanaka, Y. (1992) Circular dichroic evidence for regulation of enzymatic activity by nonsubstrate hydrophobic ligand on glutathione S-transferase, *Biochem. Biophys. Res. Commun.* 189, 1243-1251.

Nishihira, J., Ishibashi, T., Sakai, M., Tsuda, S. and Hikichi, K. (1993) Identification of the hydrophobic ligand-binding region in recombinant glutathione S-transferase P and its binding effect on the conformational state of the enzyme. *Arch. Biochem. Biophys.* 302, 128-133.

Pace, C.N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* 131, 266-280.

Pace, C.N., Shirley, B.A. and Thomson, J.A. (1989) Measuring conformational stability of a protein. In: Creighton TE, ed. *Protein structure: A practical approach.* Second edition, Oxford UK: IRL press at Oxford University Press. pp 311-330.

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, 957-963.

Pemble, S.E., Wardle, A.F. and Taylor, J.B. (1996) Glutathione S-transferase (GST) class kappa: characterization by the cloning of rat mitochondrial GST and identification of a human homologue. *Biochem. J.* 319, 754-759.

Peng, Z.Y. and Kim, P.S. (1994) A protein dissection study of a molten globule. *Biochemistry*, 33, 2136-2141.

Perkins, S.J. (1986) Protein volumes and hydration effects. The calculation of partial specific volumes, neutron scattering matchpoints and 280-nm absorption coefficient for proteins and glycoproteins from amino acid sequences. *Eur. J. Biochem.* 157, 169-180.

Perutz, M.F., Rossman, M.G., Cullis, A.F., Muirhead, H., Will, G. and North, A.C.T. (1960) Structure of haemoglobin. *Nature* 185, 416-427.

Perutz, M.F. (1992) *Protein structure: new approaches to disease and therapy.* W.H. Freeman and Company. USA.

Pfeil, W. and Privalov, P.L. (1976) Thermodynamic investigations of proteins with lysozyme as an example. *Biophys. Chem.* 4, 23-32.

Poland, D.C. and Scheraga, H.A. (1970) *Theory of helix-coil transition.*New York: Academic Press

Privalov, P.L. (1979) Stability of proteins: Small globular proteins. *Adv. Prot. Chem.* 33, 167-241.

Privalov, P.L. (1982) Stability of proteins. Proteins which do not present a single cooperative system. *Adv. Prot. Chem.* 35,1-104.

Privalov, P.L. (1989) Thermodynamic problems of protein structure. *Annu. Rev. Biophys. Biophys. Chem.* 18, 47-69.

Privalov, P.L. (1996) Intermediate states in protein folding. *J. Mol. Biol.* 258, 707-725.

Privalov, P.L. and Khechinashvili, N.N. (1974) Thermodynamic approach to the problem of stabilisation of globular protein structure: a calorimetric study. *J. Mol. Biol.* 86, 665-684.

Privalov, P.L. and Potekhin, S.A. (1986) Scanning microcalorimetry in studying temperature-induced changes in proteins. *Methods Enzymol.* 131, 4-51.

Privalov, P.L. and Gill, S.J. (1988) Stability of protein structure and hydrophobic interaction. *Adv. Prot. Chem.* 39, 191-234.

Ptitsyn, O.B. (1992) The molten globule state. In *Protein Folding.* (Creighton, T.E., ed.), pp. 243-300, Freeman and Company, New York.

Ptitsyn, O.B. (1995a) Structure of folding intermediates. *Curr. Opin. Struct. Biol.* 5, 74-78.

Ptitsyn, O.B. (1995b) Molten globule and protein folding. *Adv. Prot. Chem.* 47,

83-229.

Raghunathan, S., Chandross, R.J., Kretsinger, R.H., Allison, T.J., Pennington, C.J. and Rule, G.S. (1994) Crystal structure of human class mu glutathione transferase GSTM2-2. Effects of lattice packing on conformational heterogeneity. *J. Mol. Biol.* 238, 815-832.

Redfield, C., Smith, R.A.G. and Dobson, C.M. (1994) Structural characterization of a highly-ordered molten globule at low pH. *Nature Struct. Biol.* 1, 23-29.

Reinemer, P., Dirr, H.W., Ladenstein, R., Schäffer, J., Gallay, O. and Huber, R. (1991) The three-dimensional structure of class π glutathione S-transferase in complex with glutathione sulfonate at 2.3 Å resolution. *EMBO J.* 10, 1997-2005.

Reinemer, P., Dirr, H.W., Ladenstein, R., Huber, R., Lo Bello, M., Federici, G. and Parker, M. (1992) Three-dimensional structure of class π glutathione S-transferase from human placenta in complex with S-hexylglutathione at 2.8 Å resolution. *J. Mol. Biol.* 227, 214-226.

Reinemer, P., Prade, L., Hof, P., Neuefeind, T., Huber, R., Zettl, R., Palme, K., Schell, J., Koelln, I., Bartunik, H.D. and Bieseler, B. (1996) Three-dimensional structure of glutathione S-transferase from *Arabidopsis thaliana* at 2.2Å resolution: Structural characterization of herbicide-conjugating plant glutathione S-transferases and a novel active site architecture. *J. Mol. Biol.* 255, 289-309.

Riechmann, L., Clark, M., Winter, G. and Waldman, H. (1988) Reshaping human-antibodies for therapy. *Nature* 332, 323-327.

Robyt, J.F. and White, B.J. (1990) *Biochemical techniques: theory and practice.* Waveland Press, Inc. USA.

Sanchez-Ruiz, J.M. (1992) Theoretical analysis of Lumry-Eyring models in

differential scanning calorimetry. *Biophys. J.* 61, 921-935.

Sanger, F. (1952) The arrangement of amino acids in proteins. *Adv. Prot. Chem.* 7, 1-67.

Schellman, J.A. (1958) The factors affecting behaviour of the isolated polymer chain. *Biopolymers* 31, 1463-1470.

Schellman, J.A. (1978) Solvent denaturation. *Biopolymers* 17, 1305-1322.

Schellman, J.A. (1987) The thermodynamic stability of proteins. *Ann. Rev. Biophys. Biophys. Chem.* 16, 115-137.

Scheraga, H.A. (1996) Recent developments in the theory of protein-folding-searching for the global energy minimum. *Biophys. Chem.* 59, 329-339.

Schiffer, C.A. and Dötsch, V. (1996) The role of protein-protein interactions in protein unfolding. *Curr. Opin. Biotech.* 7, 428-432.

Schmid, F.X. (1990) Spectral methods of characterising protein conformation and conformational changes. In: Creighton, T.E. ed. *Protein structure: A practical approach, 2nd edition*. Oxford, UK: IRL press at Oxford University Press. pp 251-285.

Service, R.F. (1996) Folding proteins caught in the act. *Science* 273, 29-30.

Shakhnovich, E.I. and Finkelstein, A.V. (1989) Theory of cooperative transitions in protein molecules. I. why denaturation of globular proteins is a first - order phase transition. *Biopolymers* 28, 1667-1680.

Shortle, D. (1996) The dentatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* 10, 27-34.

Siddiqui, A.S. and Barton, G.J. (1995) Continuous and discontinuous domains: an algorithm for the automatic generation of reliable domain defenitions. *Protein Sci.* 4, 872-884.

Siegers, C.P. and Younes, M. (1983) Clinical significance of the glutathione conjugating system. *Pharm. Res. Commun.* 15, 1-13.

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 mu and pi class enzymes. *J. Mol. Biol.* 232, 192-212.

Sluis-Cremer, N. and Dirr, H. (1995) Conformational stability of Cys45 -alkylated and hydrogen peroxide-oxidised glutathione S-transferase. *FEBS Lett.* 371, 94-98.

Smith, D.B, Davern, K.M., Board, P.G., Tiu, W.U., Garcia, E.G. and Mitchell, G.F. (1986) Mr 26,000 antigen of Schistosoma japonicum recognized by resistant WEHI 129/J mice is a parasite glutathione S-transferase. *Proc. Natl. Acad. Sci. USA.* 83, 8703-8707.

Smith, D.B and Johnson, K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* 67: 31-40.

Smuts, J.C. (1926) *Holism and evolution.* Reprinted by N&S Press (1987). Cape Town.

Stix, G. (1996) Trends in nanotechnology, waiting for breakthroughs. *Sci. American* 274, 94-99.

Sturtevant, J.M. (1987) Biochemical applications of differential scanning calorimetry. *Ann. Rev. Phys. Chem.*38, 463-488.

Tanford, C. (1968) Protein denaturation. Part A and B. *Adv. Prot. Chem.* 23, 121-275.

Tanford, C. (1970) Protein denaturation, part c, theoretical models for the mechanism of denaturation. *Adv. Prot. Chem.* 24, 1-95.

Teal, F.W.J. (1960) The ultraviolet fluorescence of proteins in neutral solution. *Biochem. J.* 76, 381-388.

Timm, D.E. and Neet, K.E. (1992) Equilibrium denaturation studies of murine beta-nerve growth factor. *Protein Sci.* 1, 236-244.

Uversky, V.N. (1993) Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature due through the molten globule. *Biochemistry* 32, 13288-13298.

Uversky, V.N. and Ptitsyn, O.B. (1994) Partly folded state, a new equilibrium state of protein molecules: four-state guanidinium chloride of β-lactamase at low temperature, *Biochemistry* 33, 2782-2791.

Vince, R., Daluge, S. and Wadd, W.B. (1971) Studies on the inhibition of glyoxalase I by S-substituted glutathione. *J. Med. Chem.* 14, 402-404.

Voet, D. and Voet, J.G. (1990) *Biochemistry* John Wiley and Sons, USA.

Walker, J., Crowley, P., Moreman, A.D. and Barrett, J. (1993) Biochemical properties of cloned glutathione S-transferases from *Schistosoma mansoni* and *Schistosoma japonicum*. *Mol. Biochem. Parasitol.* 61, 255-264.

Walker, J., Crowley, P. and Barrett, J. (1995) Chemical modification of a cloned glutathione S-transferase from *Schistosoma japonicum*: Evidence for an essential histidine residue. *Exp. Parasitol.* 80, 616-623.

Wang, R.W., Newton, D.J., Huskey, S.W., McKeever, B.M., Pickett, C.B. and Lu, A.Y.H. (1992) Site-directed mutagenesis of glutathione S-transferase YaYa. *J. Biol. Chem.* 267, 19866-19871.

Ward, L.D., Matthews, J.M., Zhang, J.G. and Simpson, R.J. (1995) Equilibrium denaturation of recombinant murine interleukin-6: Effect of pH, denaturants, and salt on formation of folding intermediates. *Biochemistry* 34, 11652-11659.

Watson, J.D. and Crick, F.H.C. (1953) A structure for deoxyribose nucleic acid. *Nature* 171, 737-738.

Wetlaufer, D.B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA.* 70, 697-701.

Wilce, M.C.J., Board, P.G., Feil, S.C. and Parker, M.W. (1995) Crystal structure of a theta class glutathione transferase. *EMBO J.* 14, 2133-2143.

Yan, H., Lim, J.T.E., Contillo, L.G. and Krolewski, J.J. (1995) Glutathione S-transferase fusion proteins mimic receptor dimerization in permeabilized cells. *Anal. Biochem.* 231, 455-458.

Yeh, H.I., Hsieh, C.H., Wang, L.Y., Tsai, S.P., Hsu, H.Y. and Tam, M.F. (1995) Mass spectrometric analysis of rat liver cytosolic glutathione S-transferases: Modifications are limited to N-terminal processing. *Biochem. J.* 308, 69-75.

Yu, M.H., Lee, K.N. and Kim, J. (1995) The Z type variation of human α_1 - antitrypsin causes a protein folding defect. *Nature Struct. Biol.* 2, 363-367.

Zaks, A. and Klibanov, A.M. (1984) Enzymatic catalysis in organic media at 100 °C. *Science* 224, 1249-1251.

Zhu, Z-Y. and Karlin, S. (1996) Clusters of charged residues in protein three-dimensional structures. *Proc. Natl. Acad. Sci. USA.* 93, 8350-8355.

Zimm, B.H. and Bragg, J.K. (1959) Theory of the phase transition between helix and random coil in polypeptides. *J. Chem. Phys.* 31, 526-535.

ADDENDUM

One of the reviewers (Ken Neet) was "greatly surprised" that the binding of PZQ did not influence the stability of the GST (sections 4.2.4 and 4.3.5). The fact that there are essentially the same number of interactions in both the uncomplexed and PZQ-bound dimer would seem to imply that this situation would lower the affinity of PZQ for the GST, or might possibly mean that the PZQ would bind equally to dimer and "native" monomer. Since PZQ cannot bind the unfolded monomer, the ligand would be expected to effect the apparent stability of the GST by simply shifting the equilibrium towards the native dimer. Alternatively, destabilisation of the protein could occur by shifting to an intermediate which is more easily destabilised. Finally, precedents of bound ligands not stabilising a protein towards denaturation were requested.

With respect to these comments two possibilities exist: firstly, the unfolding pathway would be altered as has been shown for other GST classes in the presence of G-site ligands (see eg. Erhardt and Dirr, 1996), and secondly, the unfolding pathway would remain unaltered. From Figure 4.8 and 4.18 the first possibility may be ruled out. The second possibility should then result in a shift in the C_m value (with constant *m*-value) of Figure 4.8 and be similar to that observed for protein concentration dependence of Figure 4.5 (insert), or alternatively result in a shift in the T_m value of Figure 4.18 similar to that found in Figure 4.15. The identical PZQ-bound and unbound Sj26GST unfolding curves of Figures 4.8 and 4.18 eliminates this option as well. This absence of change in either the unfolding pathway (*m*-value for Figure 4.8 and ΔH_{cal} for Figure 4.18) or midpoint of unfolding (C_m or T_m) would seem to imply that, upon ligand binding, no change in the protein-solvent interactions are occurring which affect either the protein's stability or its unfolding pathway. An overlay of the PZQ-bound and unbound Sj26GST, as shown in Figure 4.9, indicates no conformational changes of the protein upon ligand binding. While the number of contacts between the two subunits of Sj26GST are reduced from 23 to 19 upon binding PZQ, it would consequently appear that the disruption of these contacts are not critical to the protein's stability or unfolding pathway.

Regarding other examples of bound ligands not stabilising a protein towards

denaturation, PZQ and ANS are known to bind both class alpha and pi GSTs, however, DSC studies of these proteins in the presence and absence of both these ligands have not affected the DSC thermograms. Solvent-induced denaturation studies of class pi GST in the presence of other non-substrate ligands have also had no real affect upon the unfolding pathway of this enzyme (R. Ohlmeyer, unpublished results). Therefore, irrespective of whether the non-substrate ligand binding site is occupied or vacant, neither the unfolding/refolding pathway or the stability of Sj26GST and many other GST classes is affected by the presence or absence of non-substrate ligands.