PROBING THE STABILITY OF A MODEL JELLYROLL β - SANDWICH PROTEIN

by

SARAH EVERTS

B.Sc., The University of Guelph, 1998

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Department of Chemistry)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

March 2001

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Chemistry

The University of British Columbia
Vancouver, Canada

Date March 19 2001

DE-6 (2/88)

ABSTRACT

A major challenge in the field of protein folding involves dissecting the thermodynamic and kinetic pathways by which a polypeptide acquires its unique, minimal free-energy structure. While much research has focussed on α-helical containing proteins, the folding pathways of their all-β-sheet counterparts remains less well defined. To this end, thermodynamic investigations of the second N-terminal cellulose binding domain of *Cellulomonas fimi* endoglucanase C (CBD_{N2}) were carried out. This 153 residue domain is composed of 11 β-strands that assemble into two anti-parallel β-sheets with the topology of a jellyroll β-sandwich. Using circular dichroism (CD) spectroscopy, CBD_{N2} is shown to undergo fully reversible unfolding induced by both heating and the addition of denaturants. Unfolding of CBD_{N2} occurs in a largely pH independent manner.

This thesis describes the results of "native state hydrogen exchange (HX)" experiments in non-denaturing concentrations of GuanidineDCl (GuDCl). Amide hydrogen to deuteron exchange provides a means of obtaining residue specific stability, since these rates reflect the structural fluctuations and hydrogen bond disruptions required for solvent exposure. Many of these rates increase when measured in the presence of low GuDCl concentration, since the free-energy barrier against structural fluctuations is decreased, although the protein remains globally folded.

The core residues of CBD_{N2} exhibit a high degree of protection, exchanging only as a result of global unfolding fluctuations. The stability measured for these unfolding fluctuations corresponds to the global stability measured by CD denaturation. Amide groups in loop regions and near the termini exhibited much less protection, exchanging via low energy fluctuations. There appeared to be no difference in exchange patterns or residue specific stability between residues in sheet-A and sheet-B, suggesting a two-state folding mechanism and the absence of

any partially unfolded intermediates. Residues most protected from HX were highly conserved among the 8 most sequence homologous proteins.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CD	circular dichroism
CBD_{N2}	second N-terminal cellulose-binding domain from *Cellulomonas fimi*
CBD_{N1}	first N-terminal cellulose-binding domain from *Cellulomonas fimi*
ΔG_U	the apparent difference in free energy of a native protein (N) relative to it's unfolded state (U)
ΔG_{HX}	the apparent difference in free energy for an amide group between states closed and open to exchange
$\Delta G_{HX}(D_2O)$	the apparent difference in free energy for an amide group between states open and closed to exchange, extrapolated to zero concentration of chemical denaturant in D_2O solution
DSC	differential scanning calorimetry
GuDCl	Guanidine deuterochloride
GuHCl	Guanidine hydrochloride
H-bond	hydrogen bond
HSQC	heteronuclear single quantum correlation
HX	hydrogen exchange
IPTG	isopropyl β-D-thioglucopyranoside
n	the index of refraction
n_{ref}	the index of refraction of a reference solution
NMR	nuclear magnetic resonance
pD	pH of samples dissolved in D_2O, corrected for the isotope effect
PUF	partially unfolded form
wt	wild type

ACKNOWLEDGEMENTS

I wish to thank my supervisor Lawrence McIntosh for his support and direction during my time in his lab. His enthusiasm and commitment to excellent science is inspiring. I appreciated the opportunity to work on a project that deviated from the usual focus of the laboratory.

I would like to acknowledge Dr. Louise Creagh for making the D.S.C. measurements.

I wish to thank all the members of the McIntosh laboratory for their help and their humor over the past years. It was a real pleasure to work with people of excellent minds, lively hearts, and musically accommodating ears.

To all the dear hearts who came to my rescue when the tunnel seemed too long and too dark, I love and appreciate you deeply. I hope that I may give as much as I have received. Finally, I thank my mother Nikki for being my most resilient supporter, and for encouraging me amidst all my endeavors.

Chapter 1
Introduction

1.0 General Considerations

With the maturation of the genome sequencing projects, and parallel burst of activity in the biotechnology industry, the field of protein science is experiencing not simply an expansion of knowledge, but an ability to expeditiously exploit this knowledge. Anfinsen's observation, that primary sequence dictates protein structure, is routinely exploited by protein engineering research (Creighton, 1993). Yet prediction of the processes by which a protein acquires and maintains this structure is still comparatively nascent and successful manipulation of the folding trajectory is virtually non-existent. In recent years, however, a theoretical framework for the interpretation of protein folding data has been carefully refined, and a host of experiments designed to probe both the kinetics and thermodynamics of the acquisition of protein structure.

Protein folding is a kinetic process driven to the preeminent goal of minimizing free energy. While the final structure is generally determined by the lowest energy conformation, the pathway to this fold is not unique, and the rate of folding not necessarily indicative of the protein's global stability. The relationship between global stability and folding kinetics lies in the conformational or "free-energy" landscape of the protein.

The conformational landscape is an abstraction illustrating the ensemble of available conformations of a protein as a function of free energy (Dill and Chan, 1997). A "rough" landscape describes a protein that has backbone arrangements wherein slight deviations in conformation lead to radically different free energy (Figure 1.1a). A "smooth" landscape is one wherein backbone conformations grade uniformly with energy (Fig 1.1b). Both types of landscapes may have relatively deeper energy wells leading to intermediates with observable lifetimes. Folding rate is slowed both by the number of local energy minima in the folding landscape, as well as the depth of these minima.

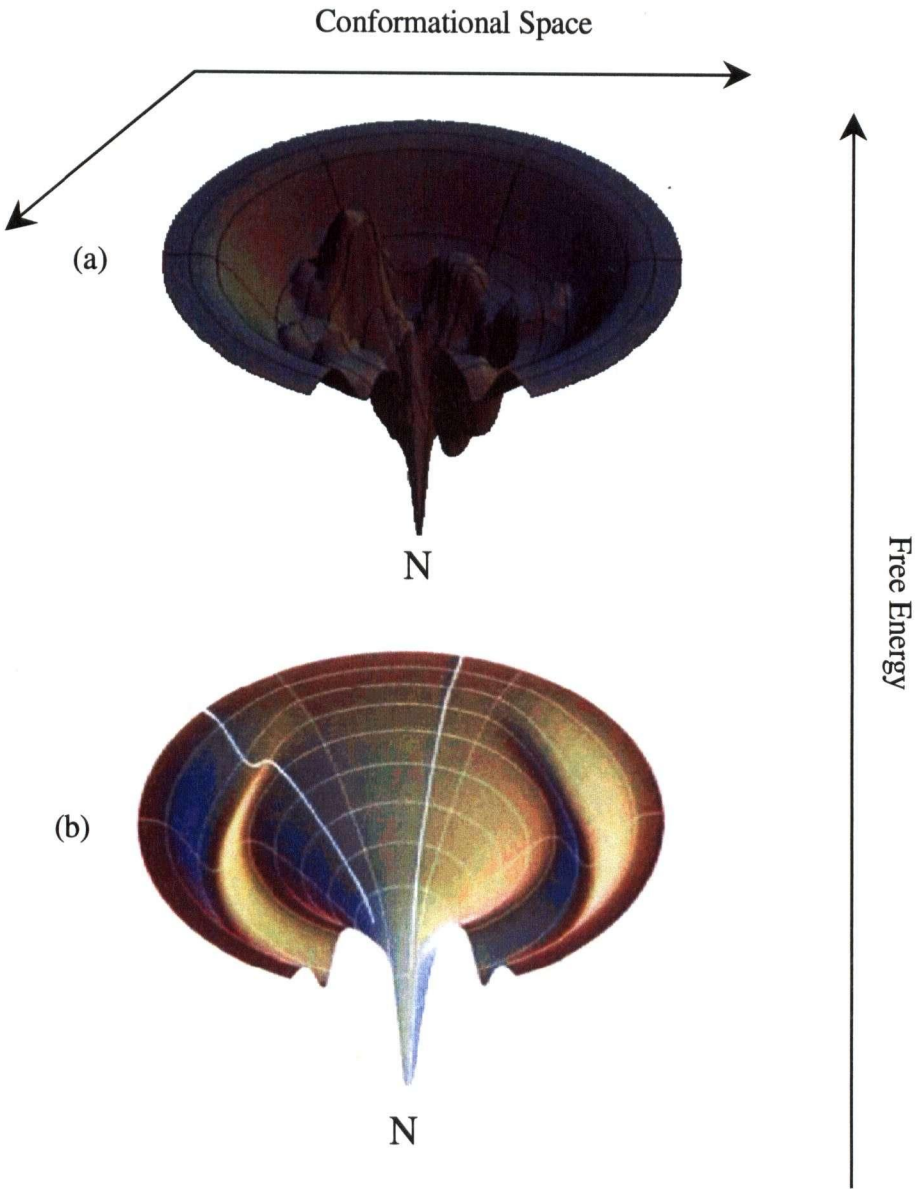


***Figure 1.1**: Examples of (a) rough and (b) smooth conformational or free-energy landscapes (Dill and Chan, 1997).*

Only recently has there been success in extending the abstraction of a conformational landscape to generalized rate-determining parameters. Baker (2000) has shown that overall protein topology seemingly determines the important features of a protein's free-energy landscape. In particular, contact order, defined as the average sequence separation of residues that are in physical contact, correlates inversely with folding rates of proteins. The salient feature in the relationship between contact and folding rate is that it appears general for proteins from both α-helical and β-sheet structural subgroups (Baker, 2000).

While Levinthal's paradox[1] dictates that proteins do not randomly sample all possible conformations along the path(s) to tertiary structure, they may begin their collapse in any conformation. Furthermore, folding proceeds as an iterative sampling of conformational "ensembles", which are groups of partially folded structures with similar energy, en route to the lowest energy conformation. Given that conformational free-energy propels the acquisition of structure, there is much interest in measuring the thermodynamic stability of structural elements within a protein. By acquiring information about the substructure stability of a protein, insights on the contours of and navigation within its conformational landscape can be obtained.

An experimental technique that probes structure at a residue specific level and relinquishes information about the structural contacts significant in the stability of the final structure is native-state hydrogen exchange.

1.1 Native State Hydrogen Exchange

Over 40 years ago Linderström-Lang observed that the hydrogen of amide groups in native proteins exchanged with those of the solvent at rates that varied by orders of magnitude (Englander *et al*, 1997). While surface amide residues are expected to exchange readily with the

[1] Levinthal argued that if a protein sampled all possible structural conformations along the folding pathway then folding rates of an average sized protein would be ~10^{30} years (Creighton, 1993).

solvent, it is remarkable that amides completely buried in the protein core display similar, albeit slower, behavior. Buried amide residues require disruption of H-bonds, solvent penetration and/or global fluctuations of structure before exchange can occur. The rate of hydrogen exchange (HX) is often a reflection of the local stability of a H-bond, since H-bonds must be disrupted prior to exchange. Interpretation of this stability in terms of the flexibility/rigidity in the region of an amide group can help provide detailed structural information at the residue specific level.

HX in a folded protein is typically described by the general two-step model:

$$closed \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} open \xrightarrow{k_{int}} exchanged \qquad (1.1)$$

In this model an equilibrium exists between the *open* and *closed* protein conformations. Exchange with the solvent can only occur from the open form, at the intrinsic rate of exchange of an unstructured peptide, k_{int}. A general expression for the observed exchange rate k_{ex} can be derived using a standard eigenvalue method for solving a series of differential equations obtained from (1.1), and is done in Appendix 1. Under conditions where $k_{cl} >> k_{op}$, corresponding to the predominance of the native state, the observed exchange rate can be written:

$$k_{ex} = \frac{k_{op} k_{\text{int}}}{k_{cl} + k_{\text{int}}} \qquad (1.2)$$

Two limiting cases emerge from this model (Englander and Kallenbach, 1983). When the intrinsic exchange rate is much faster than the closing rate, that is $k_{int} >> k_{cl}$, (1.2) reduces to:

$$k_{ex} = k_{op} \qquad (1.3)$$

In this so-called EX1 regime, the observed exchange rate is equal to the rate of opening. Those residues that exchange as a result of the same structural fluctuations will have measured exchange rates that are all equal to k_{op} of that fluctuation (and independent of k_{int}). The more advantageous EX2 regime occurs when the refolding rate is much greater than the intrinsic rate

of exchange, that is, $k_{cl} >> k_{int}$. In the EX2 regime, the observed rate of exchange can be written as:

$$k_{ex} = \frac{k_{op}}{k_{cl}} k_{\text{int}} = K_{op} k_{\text{int}} \tag{1.4}$$

K_{op} is interpreted as an equilibrium constant for the structural opening during which HX occurs. It's inverse, $1/K_{op}$, is a residue specific protection factor, $P_{op}=k_{ex}/k_{int}$. A protection factor of 100 implies that the structure of the protein slows exchange by 100-fold relative to a random-coil state. With an equilibrium constant K_{op}, the free-energy for structural opening leading to the HX of a given residue can be determined as follows:

$$\Delta G_{HX} = -RT \ln K_{op} = -RT \ln \frac{k_{ex}}{k_{\text{int}}} \tag{1.5}$$

In order to obtain ΔG_{HX}, measurements of the observed exchange rate in a native protein (k_{ex}, section 1.3) and the intrinsic rate of exchange of the unstructured peptide (k_{int}, section 1.2) are required.

1.2 The Intrinsic Hydrogen Exchange Rate for Unstructured Peptides

Over the past 40 years extensive descriptions of the factors influencing hydrogen exchange in unstructured peptides have emerged. (Molday *et al*, 1972; Englander and Mayne, 1983; Bai *et al*, 1993)

1.2.1 Theoretical Framework

In aqueous solutions, peptide HX is catalyzed by hydroxyl and hydronium ions, as well as water.

$$k_{\text{int}} = k_A[H^+] + k_B[OH^-] + k_W \tag{1.6}$$

k_w is the water catalyzed exchange rate constant, while k_A and k_B are the acid and base catalyzed exchange rate constants. k_{int} has a first order pH dependence for specific OH^- and H_3^+O catalysis. Due to the relative magnitude of these effects, acid catalysis dominates for pH < 3, while base catalysis dominates at pH > 3. The constant rate of water catalysis is so small that it is generally negligible except near pH 3.

Exchange requires the formation of a hydrogen-bonded encounter complex between catalyst and amide group followed by the partitioning of the hydrogen between the donor and acceptor according to their relative pKa values. This complex forms according to a second order diffusion limited rate constant. One can readily visualize the formation of an encounter complex with OH^- ions H-bonded to an amide hydrogen (Figure 1.2). Conversely, the mechanism for acid catalyzed hydrogen exchange remains controversial given the possibility of protonating the oxygen or nitrogen in the amide. Most native state hydrogen exchange experiments are measured near physiological pH, and thus, base catalysis takes precedence.

1.2.2 Determination of Unstructured (Intrinsic) Peptide Exchange Rate, k_{int}

HX rates of unstructured peptides are significantly influenced by the nature of the side chains flanking the amide residue in question. Molday *et al* first showed that charged groups and

polar side chains have additive inductive effects on acid and base catalyzed HX (1972). Later work revealed that the extent to which a side-chain exhibited β-sheet propensity scaled with its

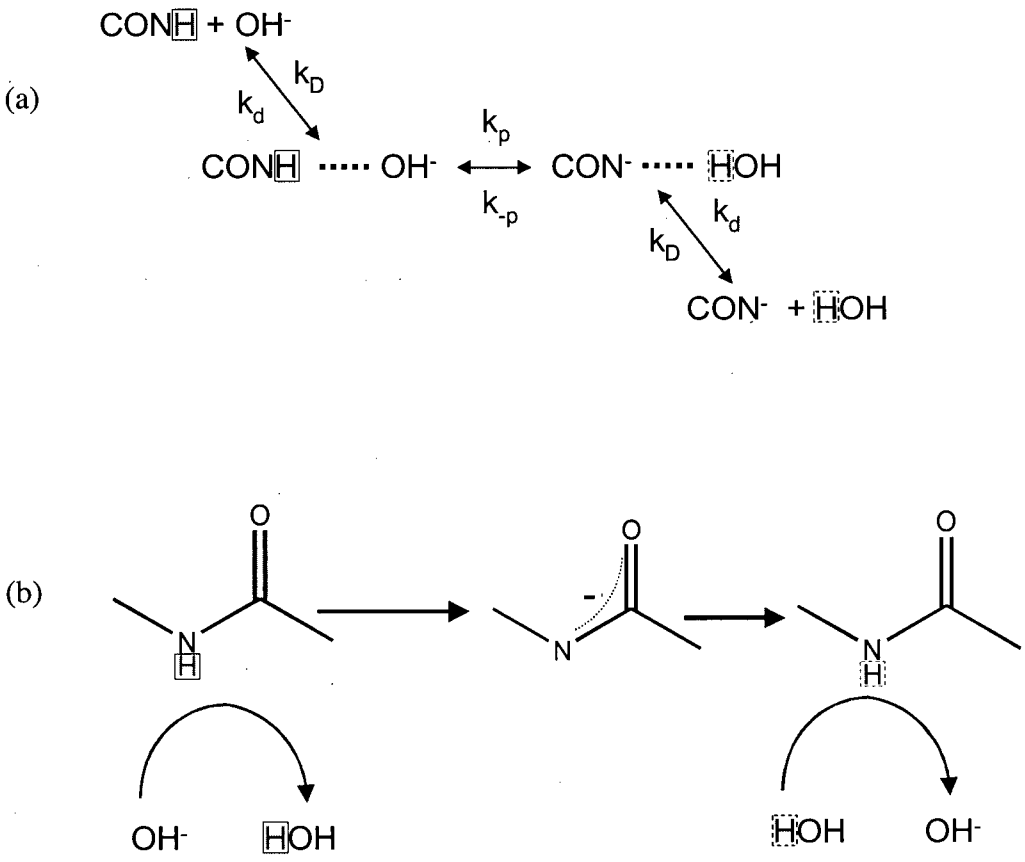


***Figure 1.2**: (a) Base Catalyzed Reaction Equation with (b) mechanism (Englander and Kallenbach, 1983).*

peptide amide HX rate (with a correlation coefficient of 0.96) (Bai and Englander, 1994). This is a reflection of the side-chain blocking effect. Namely, in shielding the amide group from making H-bonds with solvent HX catalysts, the side-chains also shield the amide group from solvent H-bonds that would compete with secondary structure formation. HX blocking energy increases 5-fold more rapidly than the β-sheet H-bond blocking energy. This is because the solvent catalyst in the HX encounter complex (an ion, for all intents and purposes) has hydration shells that constitute a volume larger than the single water molecule involved in secondary structure competing H-bonds.

Comprehensive studies of specific neighboring side-chain effects on amide HX residues have resulted in tabulation of acid and base catalysis, in H_2O and D_2O, for all amino acid pairs. Thus a modified version of (1.6) that includes the additive inductive electrostatic and steric effects of the flanking side chains, allows calculation of the intrinsic amide HX rates of an unstructured peptide, as a function of pH (or pD) and temperature (Bai *et al*, 1993). Experimental agreement is good under conditions of low ionic strength.

1.3 Measurement of HX in Native Proteins

Amide hydrogen exchange is most commonly and sensitively measured using Nuclear Magnetic Resonance (NMR) spectroscopy. In addition NH → ND exchange has been detected using infrared mass spectroscopy and NH→NT using gel filtration and scintillation counting (Tito *et al*, 2000; Loftus *et al*, 1986; Connelly *et al*, 1993; Arrington and Robertson, 2000a,b).

^{15}N-^{1}H Heteronuclear Single-Quantum Correlation (HSQC) NMR experiments produce 2-dimesional spectra with a single peak for every ^{15}N atom directly bonded to a ^{1}H atom. The ω1 axis (direct detection) displays the $^{1}H^{N}$ chemical shift of an amide while the ω2 axis (indirect detection) displays the ^{15}N chemical shift (Figure 1.3). Since HSQC spectra exhibit resonances

for every amide group in a protein, it is a powerful residue specific detection tool for measuring HX.

H $\rightarrow$ D buffer exchange can be initiated by transferring a protonated protein into a deuterated solution. As exchange occurs, HSQC peak intensity decreases since ^{1}H NMR probes are insensitive to deuterons. By fitting the peak intensity of HSQC spectra taken at incremental time points to a single exponential decay function, HX rates of residues in the protein can be measured (Figure 1.3 insert). Practically speaking, not all rates can be measured because many surface amides exchange too rapidly for accurate analysis, while others have overlapping resonances. The time between transfer into D_2O and acquisition the first spectra is approximately 15 minutes while the time to record an HSQC spectrum of satisfying signal to noise on a 500 MHz spectrometer is also 15 minutes. Acquiring at least 5 spectra to produce enough data points to precisely define a decay curve requires that the amide groups do not completely exchange before the "dead-time" of 1.5 hours. For most globular proteins, only the surface amides exchange within the experiments dead-time.

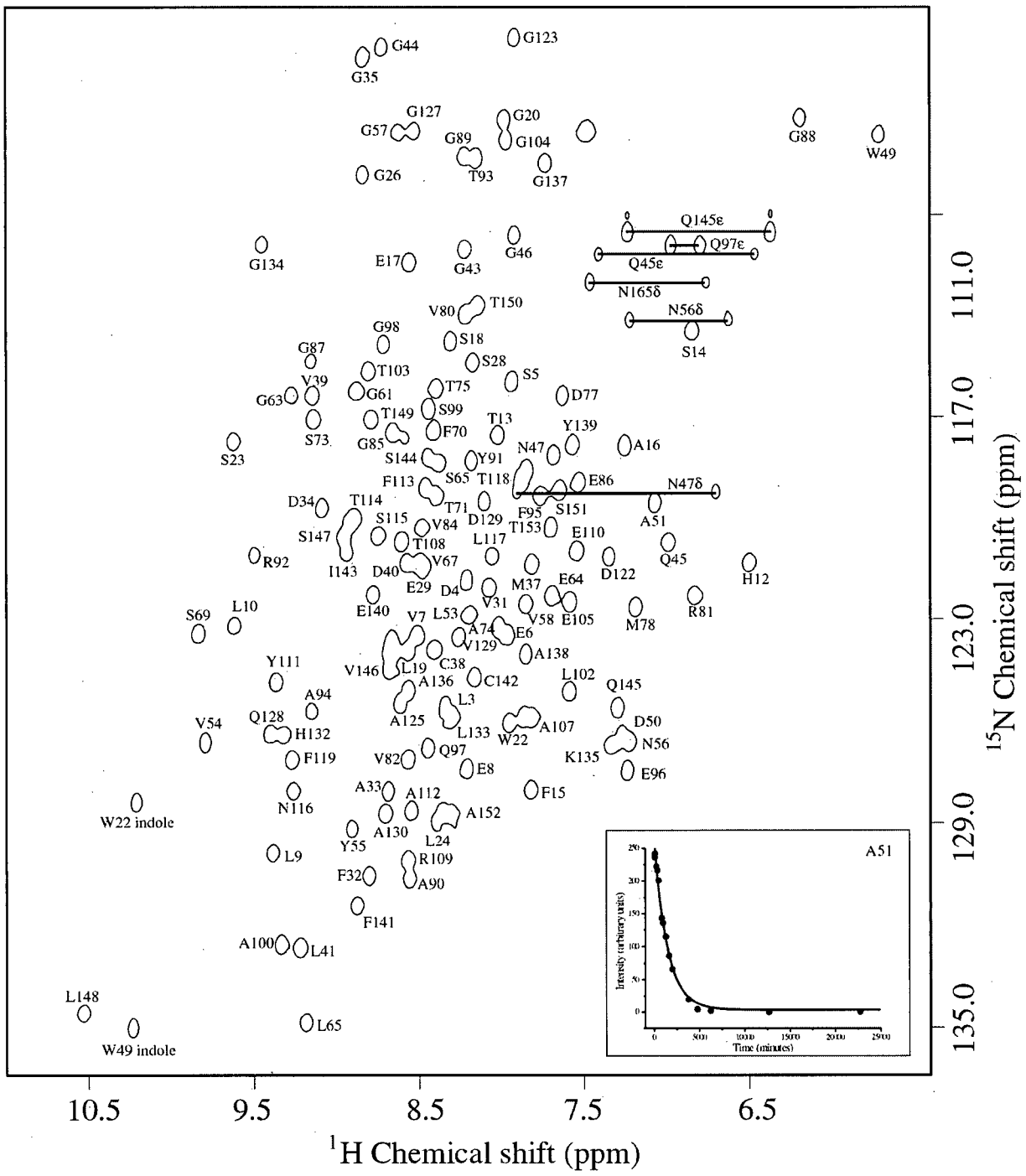


***Figure 1.3**: ^{15}N-^{1}H HSQC spectrum of CBD_{N2}. As exchange proceeds, peak intensities decrease and are fit to a single exponential decay function (insert).*

1.4 Distinguishing EX1 and EX2

Protection factors, and hence K_{op} and ΔG_{HX} can be extracted from protein HX experiments provided that the exchange occurs in the EX2 regime. This requires that the refolding rate of the structural opening leading to exchange is much faster than the intrinsic rate of exchange for all the residues exposed in the opening. That is, $k_{cl} >> k_{int}$. Comparison of global refolding rates of proteins to intrinsic rates calculated as in section 1.2.2 can confirm EX2 exchange. More direct proof of EX2 exchange is obtained by measuring the pH dependence of native protein exchange rates, k_{ex}.

In the EX2 regime $k_{ex} \propto k_{int}$, and $k_{int} \propto [OH^-]$. Therefore an increase of pH by one unit should lead to an increase in k_{ex} by ten-fold. Conversely in the EX1 regime $k_{ex}=k_{open}$ and pH-independent HX is expected. A transition from EX2 exchange to EX1 exchange is seen when the pH is raised sufficiently high to cause k_{int} to switch from being much less than the refolding rate k_{cl}, to much greater. Determination of the pH dependence of amide exchange (k_{ex}) provides a residue level probe of the exchange regime.

Since establishing the exchange regime is so integral to the interpretation of native state HX results, considerable effort is often extended to studying the pH dependence of exchange. Arrington and Robertson measured the pH dependence of native state exchange of turkey ovoid mucoid third domain (OMTKY3) to observe the transition from EX2 exchange to that of EX1 near pH ~ 8.5 (1997). It was possible to obtain k_{op} in the EX1 regime and k_{ex} in the EX2 regime. They used this information to observe the residue specific refolding rate and found that this was highly correlated to the residue specific stability, while k_{op} had no correlation.

1.5 Native State HX in Low Concentrations of Denaturant

The success of determining residue specific stabilities using (1.6) under native conditions has lead to an extension of this experiment wherein HX rates are measured in solutions containing low concentrations of denaturant. Slowing protein refolding by making more favorable contacts with the hydrophobic core of a protein than would water, denaturants lower the unfolding free-energy, ΔG_U. Present in a high enough concentration, denaturants reduce the free-energy of unfolding such that a protein is completely denatured. Present in low concentrations, below the midpoint of unfolding, denaturants spread the Boltzman distribution of protein conformations so that partially unfolded forms are transiently populated (Figure 1.4). In effect this lowers the energy barrier to structural fluctuations within a protein, without denaturing it.

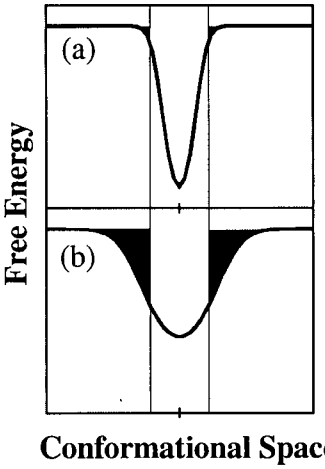


***Figure 1.4**: Boltzman distribution of a folded protein with (a) no denaturant present, and (b) in a solution of denaturant of sufficiently low concentration that the protein remains folded. While all structural conformations are possible, at any given time the native state is the most populated state. When less than half of the protein molecules are in the native state, the protein is considered denatured.*

The increased structural fluctuations occurring in the presence of low concentrations of denaturant causes overall protein HX rates to augment. Paralleling equilibrium studies of denaturant-induced protein unfolding, phenomenological estimates of the relationship between the free-energy associated with unfolding fluctuations leading to HX and denaturant concentration reveal a linear relationship (Bai *et al*, 1995).

$$\Delta G_{HX} = \Delta G_{HX}(D_2O) + m_{HX}[denaturant] \quad (1.7)$$

where $\Delta G_{HX}(D_2O)$ is the stability of a particular residue extrapolated to zero concentration of denaturant, and m is the denaturant dependence of the free energy of exchange.

Since denaturant promotes the exposure of non-polar protein surface area, residues with a large denaturant dependence (large m_{HX}) exchange during large ("global") unfolding fluctuations. Under conditions that favor the native state, $\Delta G_{HX}(D_2O)$ and m_{HX} for those residues which have the highest protection factors should correspond to the global unfolding free energy, ΔG_U, and m_U measured by CD or fluorescence spectroscopy. When the extrapolated $\Delta G_{HX}(D_2O)$ and m_{HX} values are less than the global unfolding ΔG_U and m_U, then exchange occurs as a result of so-called "sub-global" structural fluctuations.

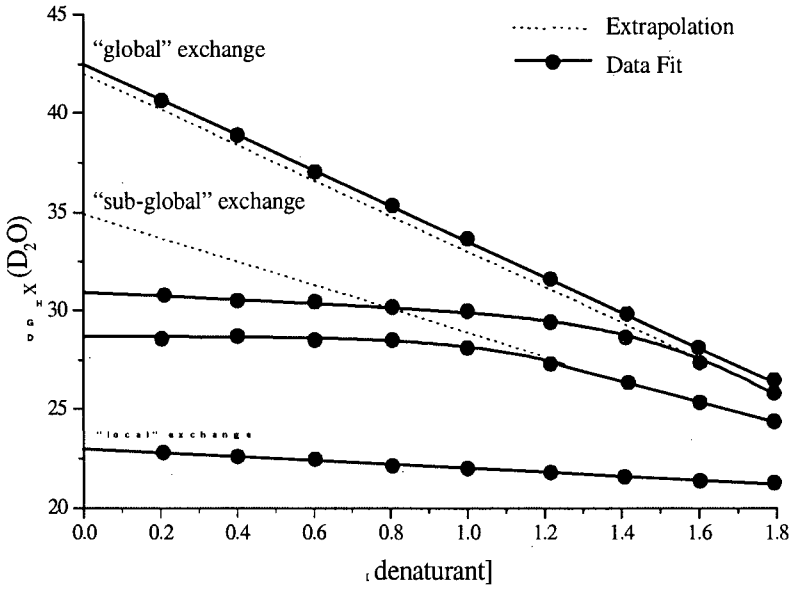


***Figure 1.5**: Typical profiles of HX in the presence of denaturant (GuDCl).*

Many residues in native state HX experiments exchange in with extremely low denaturant dependence ($m_{HX} \sim 0$), throughout a large range of denaturant concentrations. Often referred to as "local" exchange, denaturant independent exchange is attributed to local structural fluctuations that do not expose new non-polar surface area and correspond to extrapolated $\Delta G_{HX}(D_2O)$ values which are far below the global unfolding free energy (Llinás and Marqusee, 1999). Since fluctuations leading to denaturant independent exchange are low in energy, they do not require stabilization of non-polar residues with denaturant molecules.

Usage of the word "local" to describe denaturant independent exchange may be misleading because it implies little overall structural fluctuation is involved in this sort of exchange. Recent simulations on *Staphylococcal* nuclease suggest that denaturant independent exchange of many residues is consistent with significant surface area exposure of this protein (Wooll *et al*, 2000). The salient feature is that while Wooll's work indicates that fluctuations leading to denaturant independent exchange *can* expose a lot of surface area, this does not mean the surface area exposed is non-polar. If the surface area exposed were non-polar, a favorable interaction would lead to a significant HX dependence on denaturant.

A given amide residue can exchange "locally" at low concentrations of denaturant, exchange with some denaturant dependence at intermediate concentrations of denaturant ("subglobal" exchange), and exchange "globally" near the unfolding mid point concentration of denaturant. Exchange profiles of residues within the same element of secondary structure, or structural domain, will occasionally superimpose at the "subglobal" and "global" level, leading to similar extrapolated $\Delta G_{HX}(D_2O)$ and m_{HX}-values. These regions of similar structural stability, called "partially unfolded forms" (PUF's) and "stability domains," have been interpreted as domains that fold synchronously. The formation of these putative structures, identified from their corresponding stabilities, is sometimes interpreted to represent the dominant kinetic folding

pathway (Xu *et al*, 1998). Several issues immediately arise as a result of extracting kinetic conclusions from an equilibrium study.

Since native state HX studies probe the increasing population of *unfolded* forms, if any interpretations of kinetic pathways are to be made, they should be of both the unfolding pathway and the folding pathway. Or more generally, is the folding pathway the same as the unfolding pathway? In principle, if unfolding is reversible, then the protein must traverse the same conformational/energetic landscape regardless of it's initial state (native or denatured) (Creighton, 1992; Dill and Chan, 1997).

To fully understand how a protein folds, a complete construction of the conformation landscape is required. The identification of stability domains of native state HX generates a coarse grain contour of this landscape. However, interpretation of the trajectory through the conformational landscape from equilibrium studies is a delicate extrapolation. Specifically, while suggesting that the most stable domains of a protein fold first, followed by the less stable domains, is intuitive, it is also an extrapolation. One route to validation[2] of the structural folding pathway is achieved through "stability mutants."

1.6 Native State HX of Stability Mutants

Consider a hypothetical protein (Figure 1.5) with 3 domains, A, B, and C, where A is the most stable domain, unfolding only when the protein globally denatures. B and C display subglobal stability dependence, where $\Delta G_B(D_2O) > \Delta G_C(D_2O)$. One could suggest therefore that this proteins unfolds in the order of the C domain first (since it is least stable), then B, and finally A. One way to confirm step-wise unfolding is to introduce a mutation in B that increases the stability of this domain by 10 kJ/mol. If native state HX experiments are performed on the mutant, and folding proceeds in the order predicted, then the stability of the C domain should be

the same in both the wild type and mutant. If the C domain unfolds first, independently of B, then any stability increase in B will be inconsequential to C. Conversely, the stability of both A and B should increase by 10 kJ/mol, because B unfolds before A.

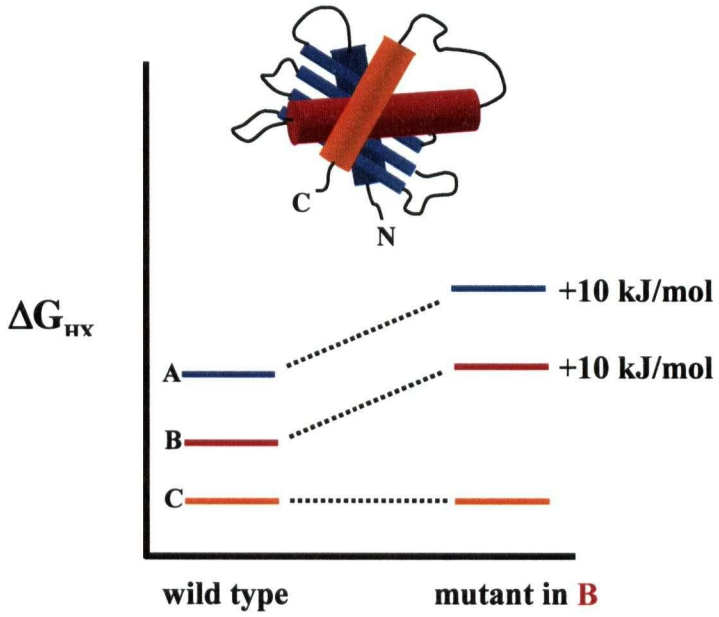


Figure 1.6: *Hypothetical protein with three "stability domains" that sequentially acquire structure with a representation of the effect of mutation in one domain on all domain stabilities as measured by native state HX.*

[2] Other techniques to establish the order of folding events are kinetics of folding/unfolding and pulsed label HX

The "stability labeling" method was used to establish that the three partially unfolded forms observed in the native state HX of cytochrome C, and the 2 partially unfolded forms in Rnase T 1 represented sequential intermediates in the folding pathway. (Xu *et al*, 1998; Chamberlin and Marqusee, 2000)

The free-energy impact of a mutation is often evaluated as the difference in stability between mutant and wild type[3], or the $\Delta\Delta G_{HX}$.

$$\Delta\Delta G_{HX} = \Delta G_{HX}(mut) - \Delta G_{HX}(wt) = -RT\ln(\frac{k_{ex}(mut)}{k_{\text{int}}}\frac{k_{\text{int}}}{k_{ex}(wt)}) = -RT\ln(\frac{k_{ex}(mut)}{k_{ex}(wt)} \qquad (1.8)$$

Direct comparison of HX rates eliminates the need to calculate the intrinsic rate of exchange of the unstructured peptide, which is an empirical approximation.

1.7 Survey of the Results of Native State HX Experiments

1.7.1 Identification of "Stability Domains"/ Partially Unfolded Forms from Denaturant Native State HX Experiments

Native state HX experiments as a function of denaturant have been performed on 11 proteins with a broad distribution of α-helical and β-sheet secondary structure. Subglobal domains have been identified in six proteins: Cytochrome C, T4 lysozyme (Llinás and Marqusee, 1999), LysN (Alexandrescu *et al*, 1999), Ribonuclease H1 (Chamberlin *et al*, 1996), Ribonuclease A (Mayo *et al*, 1993), and Cytochrome b562 (Fuentes and Wand, 1998b) (Figure 1.7). In general, the stability domains constitute an α-helix, or group of α-helices, or a β-sheet.

Proteins to display no "stability domains" are: peptostreptococcal protein L (Yi *et al*, 1997), Chymotrypsin Inhibitor A (Neira *et al*, 1997), Barnase (Chu *et al*, 1999), Barstar[4] (Bhuyan and Udgaonkar, 1998), β-Lactoglobulin (Ragona *et al*, 1999), CD2, Chymotrypsin Inhibitor 2 (Itzhaki *et al*, 1997). In general, these proteins have only global denaturant dependent

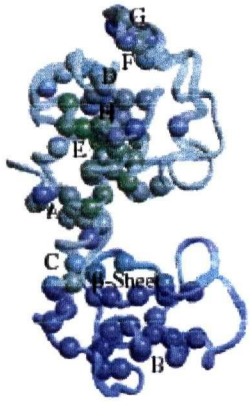
exchange or denaturant independent exchange at low denaturant that switches to a global dependence at high denaturant.

Table 1.1: A Survey of Protein Native State Experiments HX using NMR

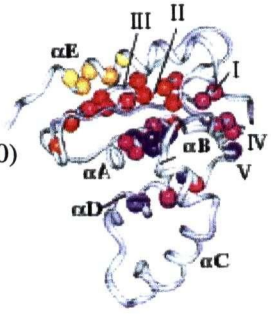
Protein	Variable	Partially Unfolded Forms	2° Structure	First Author, Publication Date
T4 Lysozyme	GuDCl	yes	7α,3β	Llinas, 1999
LysN	GuDCl	yes	3α,5β	Alexandrescu, 1999
Rnase H	GuDCl	yes	5α,5β	Chamberlin, 1996
Rnase H Thermo	GuDCl	yes	5α,5β	Hollien, 1999
Cytochrome c	GuDCl	yes	3α	Xu, 1998; Bai, 1995
Cytochrome b562	GuDCl	yes	4α	Fuentes, 1998b
Barstar	GuDCl	yes[4]	4α,3β	Bhuyan, 1998
CD2-molten globule	GuDCl	yes		Parker, 2001
RNaseA	GuDCl	no		Mayo, 1993
peptostreptococcal protein L	GuDCl	no	4β,1α	Yi,1997
beta-Lactoglobulin	Urea	no	9β,1α	Ragona, 1999
Cold Shock Protein A	Urea	no	5β	Jaravine, 2000
Barnase	Urea, GuDCl	no	2α,4β	Chu (1999)
Chymotrypsin Inhibitor 2	GuDCl	no		Itzhaki, 1997
Chymotrypsin Inhibitor 2	Temperature	no	1α,6β	Itzhaki, 1997
Rnase T1	Temperature	no	1α,7β	Mullins, 1997
Cytochrome c	Temperature	yes	3α	Xu, 1998; Bai, 1995
Cytochrome b562	Pressure	yes	4α	Fuentes, 1998a
Ovomucoid third domain	pH		3β,1α	Arrington, 1997
Alamethicin	in micelle		α	Yee, 1999
Rnase T1	Mutation		1α,7β	Huyghues-Despointes, 1999
CspA	TMAO		5β	Jaravine, 2000
Prion Protein			α,β	Hosszu, 1999
Stefin A	Dimer		5β,1α	Jerala, 1999
Carbonic anhydrase	Trp, GuDCl		4α,7β	Jonasson,1999

[3] It is generally assumed that mutations affect the folded state of a protein.

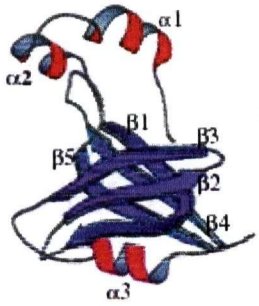
(a) Cytochrome C
(Xu *et al*, 1998)



(b) T4 Lysozyme
(Llinas and Marqusee, 1999)



(c) Ribonuclease H
(Hollien and Marqusee, 2000)



(d) LysN
(Alexandrescu *et al*, 1999)

***Figure 1.7**: Proteins that exhibit "stability domains" as determined from native state HX in low concentrations of denaturant. These regions of similar HX stability are identifies by a uniform color of the backbone, or of probes located on the backbone.*

[4] Bhuyan *et al* claim PUF's exist in Barstar but the data is inconclusive on this point.

1.7.2 Native State HX as a Function of Pressure

Hydrostatic pressure was used as a variable for native state HX experiments on apocytochrome b_{562}. Application of hydrostatic pressure initiates the reversible unfolding of proteins in water because the volume of the protein-solvent system is smaller for the unfolded state (Fuentes, 1998a). Native state HX results between 1 bar and 1.1 kbar showed 3 regions of unique stability that corresponded to the stability domains observed in experiments as a function of denaturant.

1.7.3 Native State HX as a Function of Temperature

Temperature can be used as a variable with which to study native state HX, since the Boltzmann distribution of partially unfolded states spreads out at temperatures close to, but not beyond the unfolding temperature midpoint, T_m. When temperature is used as the perturbant, values for ΔH and ΔS can be obtained at a residue specific level.

Cytochrome c temperature dependent experiments revealed the same "stability domains" as the native state HX experiments as a function of denaturant (Bai *et al*, 1995). Stability measurements from temperature native state HX of Chymotrypsin Inhibitor 2 (CI2) revealed only global and denaturant independent exchange, also consistent with earlier denaturant studies (Itzhaki *et al*, 1997). That the general results of native state HX experiments from diverse denaturants (urea, GuDCl, pressure and temperature) agree ascertains that native state HX results are not an artifact of the experimental conditions.

1.7.4 Other Mutation Work

Aside from inferring the order in which domains of similar stability fold, native state HX on mutant proteins has been used to study the effects of a single mutation on the properties of the entire protein.

NSHX experiments on stability (reduction) mutants of CI2 showed that the HX rates of all residues exchanging through "local" fluctuations were unaffected. However, for a small number of residues in the mutant protein that exchanged through "global" fluctuations, the effect of mutation on the HX rate constant was the same as for the equilibrium constant for unfolding. These residues corresponded to the most slowly exchanging residues in the wild-type protein (Neira *et al*, 1997).

Conversely, mutant studies (one located in the core, one on the surface) of Rnase T1 show that while the residues in the wild type which exchange "globally" have a $\Delta\Delta G_{HX} = \Delta\Delta G_U$ and a $\Delta G_{HX}=\Delta G_U$, this is also true of many residues that exchange through "local" fluctuations (Huyghues-Despointes *et al*, 1999). Stability mutants of barnase show the same behaviour as Rnase T1 (Perrett *et al*, 1995).

The common result among the mutant studies of CI2, barnase and Rnase T1 is that the stability of the global exchangers is affected by the mutation in the same way as the global stability (that is, $\Delta G_{HX}=\Delta G_U$ and $\Delta\Delta G_{HX}=\Delta\Delta G_U$). One factor likely affecting the results of "local" exchange behavior among these proteins is in the placement of the mutations. While the CI2 mutations[5] were embedded in elements of secondary structure, the Rnase T1 mutations were placed on the surface an in only one small part of the interior.

1.7.5 Native State HX on a Thermophilic Homologue of RNase H

Thermophilic protein homologues are nature's "stability mutants". Comparison of thermophilic and mesophilic homologues of RNase H have been made using native state HX. RNase H has stability domains that are also detected in the thermophilic protein. For all residues,

[5] Mutations in CI2: one I→V in the hydrophobic core, all others in either a β-strand or an α helix.

a proportional increase in HX stability ($\Delta G_{HX}(D_2O)$) is observed. In fact when the global stability of the thermophilic protein is decreased to that of the mesophilic protein in the presence of 0.88 M GuDCl, the residue specific stabilities of the homologues are virtually the same. Evolution has selected multiple site-mutations in the mesophilic protein to stabilize the thermophilic protein in a delocalized fashion (Hollien and Marqusee, 1999).

1.7.6 Tryptophan Indole Hydrogen Exchange

Native state H $\rightarrow$ D exchange experiments at 7 tryptophan indole side-chains was performed in human carbonic anhydrase II (HCAII) as a function of low concentrations of denaturant. Indole HX shows similar dependence on denaturant as amide HX. Several of the Trp residues showed little protection to exchange, while others exchanged in a denaturant independent way before shifting to global exchange. Only Trp97, a member of the hydrophobic core, exchanged "globally" for all denaturant concentrations. HCAII native state HX protection factors were on par with those measured on a previously identified "molten globule[6]" of this protein (Jonasson *et al*, 1999). This work highlights the lack of clear distinction between a molten globule and the native state.

1.7.7 Transition Between Conformational Isomers

Human prion protein (PrP) has a propensity to adopt alternative folds, suggesting that several alternative conformations exist in equilibrium with the native state. Native state HX experiments identified a hyper-stable protected core of approximately 10 residues in the region of a disulphide bond in PrP^C (Hosszu *et al*, 1999). The results suggest that the transition route

[6] A molten globule refers to a population of partially folded protein, at pH extremes, that maintains indeterminate tertiary contacts.

from the PrP^{C} conformer to the detrimental PrP^{Sc} conformer occurs through an unfolded state that retains only this small region of structure, instead of a highly organized intermediate.

1.7.8 Dimer Formation

The protection factors of stefin A as a monomer were compared to those under conditions where stefin A forms a dimer. Residues with increased protection in the dimer corresponded to those residues whose chemical shifts were affected by the monomer to dimer transition. As protein concentration was incrementally decreased to initiate dimer to monomer transitions rapid exchange of all protected residues in the dimer occurred, suggesting that the transition state is unfolded to a large extent (Jerala and Zerovnik, 1999).

1.7.9 Exchange of Peptide/Micelles Mixtures

Native state HX was performed on Alamethicin, a 20 residue α-helical antibiotic peptide that inserts into bacterial membranes and self-associates to form ion channels, in the presence of micelles. Protection factor data supports the theory that one face of the helix interacts most strongly with the micelle (Yee *et al*, 1999).

1.7.10 Native State HX in the Presence of Low Concentrations of TMAO

Trimethylamine N-oxide (TMAO) interactions with exposed peptide backbone are unfavorable, thereby stabilizing native proteins by predominantly increasing the free-energy of the denatured state, while only marginally increasing the free-energy of the native state. Thus increased protection as a result of TMAO reflects changes in the properties of the denatured state as opposed to the native state. Unlike denaturant, which stabilizes *elements* of a protein

(hydrophobic residues), TMAO uniformly affects the *entire* protein (since it interacts with the backbone), generating results that reflect free-energy properties of the entire protein. Native State HX on cold shock protein A (CspA) in the presence of TMAO lead to increased protection throughout the entire protein. These results suggest that the free-energy of partially and globally unfolded species are increased, which makes exchange competent conformations less accessible from the native state. No stability subdomains were observed as a function of TMAO for CspA (Jaravine *et al*, 2000).

1.8 Native State HX of β-sheets in Proteins

Many of the proteins studied using the native state HX were comprised of both α-helices and β-sheets. In all proteins the β-sheets had uniform stabilities or protection factors. That is, no "sub-global" unfolding was detected among strands of the same sheet. β-strands are often among the most stable elements of secondary structure. As a counterpoint, the strands in these proteins often lacked enough probes in the β-strands to justify conclusive statements about specific stability trends of β-sheets in proteins.

There are four proteins of majority β-sheet composition that have been studied using native state HX. The Alexandrescu group has focussed on the OB-fold superfamily, studying both LysN[7] and Cold Shock Protein A to determine whether conserved native state structures reflect conserved mechanisms of structure formation. Their results suggest that the most conserved elements of structure are the most stable to unfolding, and that this stability correlates with sequence hydrophobicity (Jaravine *et al*, 2000; Alexandrescu *et al*, 1999).

β-Lactoglobulin is a 9-stranded β-barrel protein with a terminal α-helix. Native state HX experiments indicate that it folds in a 2-state process. All the residues in this protein, with the

[7] Anticodon binding domain of the LysS isoform of *E.coli* lysyl-tRNA synthetase

exception of the last β-strand in the sequence and the helix, exchange solely as a result of global unfolding fluctuations, while all other probes exchange exclusively in a denaturant independent way (Ragona *et al*, 1999).

Stopped-flow experiments on CD2.d1[8], a 9-stranded β-barrel with one α-helix revealed folding of this protein included the transient accumulation of an intermediate. Native state HX in denaturant concentrations which traversed the native-intermediate transition revealed sub-global exchange profiles whose $\Delta G_{HX}(D_2O)$ and m_{HX} values agreed with those obtained from a kinetic burst phase transition (Parker and Marqusee, 2001).

1.9 The Second N-Terminal Cellulose-Binding Domain from Cellulomonas fimi

Enzymes that degrade allomorphs of cellulose are comprised of one or more cellulose binding domains (CBDs) as well as a catalytic domain. Classified by sequence similarity into at least 13 families, over 200 CBDs have been identified in cellulolytic bacteria and fungi. CBDs act to associate cellulases with their substrate, and increase catalysis either passively by increasing the local concentration of substrate, or more actively by aiding in the disruption of the structure of cellulose.

The *Cellulomonas fimi* 1,4-β-glucanase CenC has 2 homologous tandem N-terminal binding domains, CBD_{N1} and CBD_{N2}. With 10 and 11 β-strands, CBD_{N1} and CBD_{N2} have a similar jellyroll β-sandwich topology as elucidated from NMR structure determination (Johnson *et al*, 1996, Brun *et al*, 2000). The jellyroll motif consists of four Greek key motifs formed across an imaginary barrel, which flattens to form a sandwich of 2 anti-parallel β-sheets (Figure 1.8).

[8] CD2.d1: The N-terminal domain of rat CD2

(a)

(b)

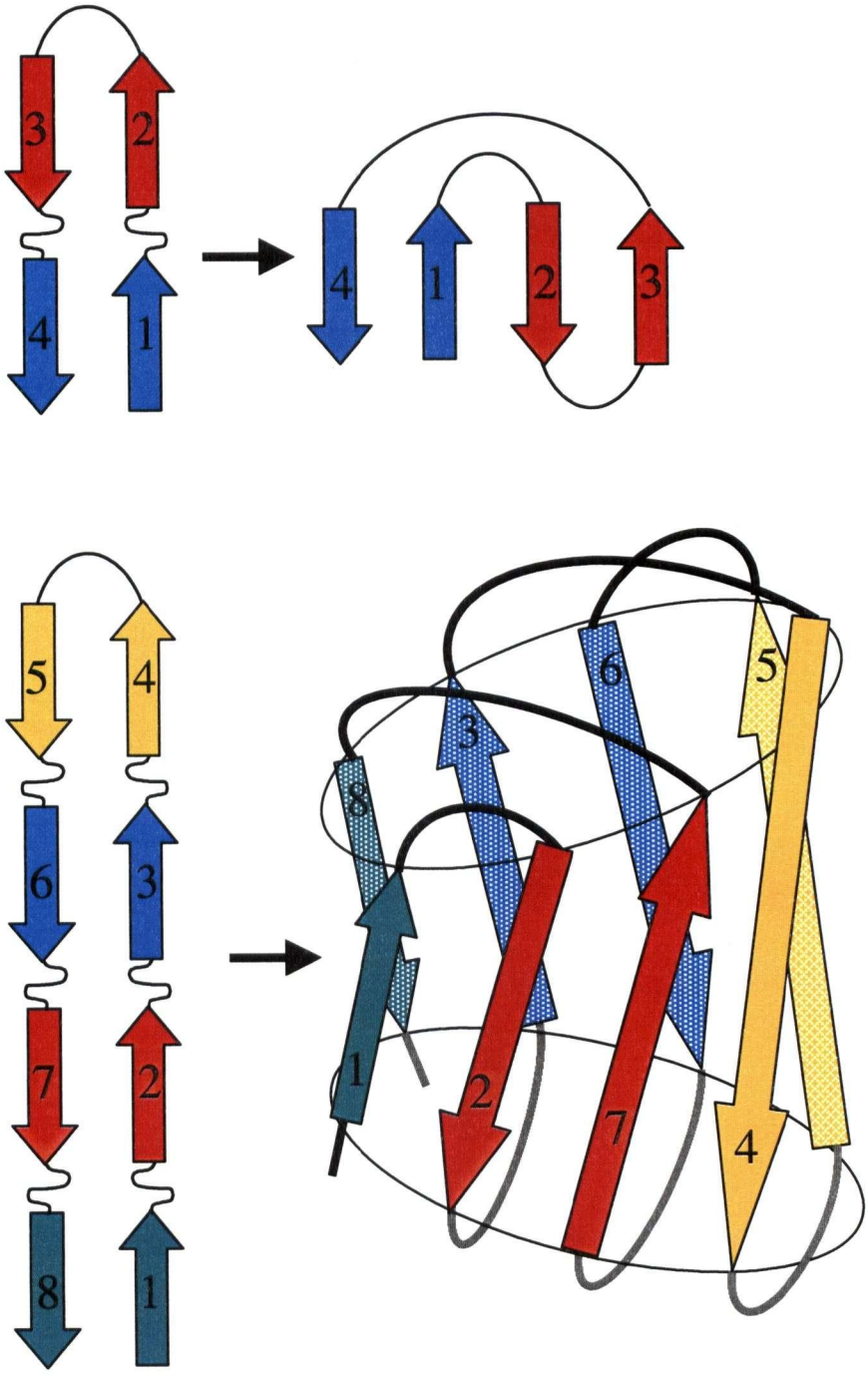


***Figure 1.8**: (a) The Greek key motif occurs when four adjacent anti-parallel β-strands are arranged in a pattern similar to the repeating unit of one of the ornamental patterns used in ancient Greece. (b) The jellyroll motif (Branden and Tooze, 1991).*

(c)

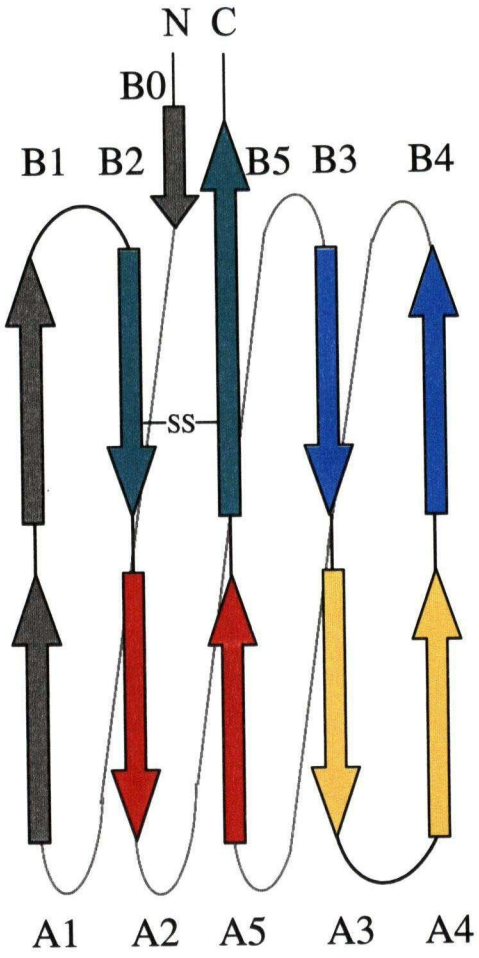


***Figure 1.8**: (c) Schematic diagram showing the jellyroll β-sandwich topology of CBD_{N2}. Strands A2-A5 and B2-B5 comprise the jellyroll motif and are colored accordingly. The position of the disulphide between Cys38 and Cys142 is indicated. The tertiary structure can be envisioned by folding the figure such that sheet B lies below sheet A. The lengths of the strands and loops are not drawn to scale.*

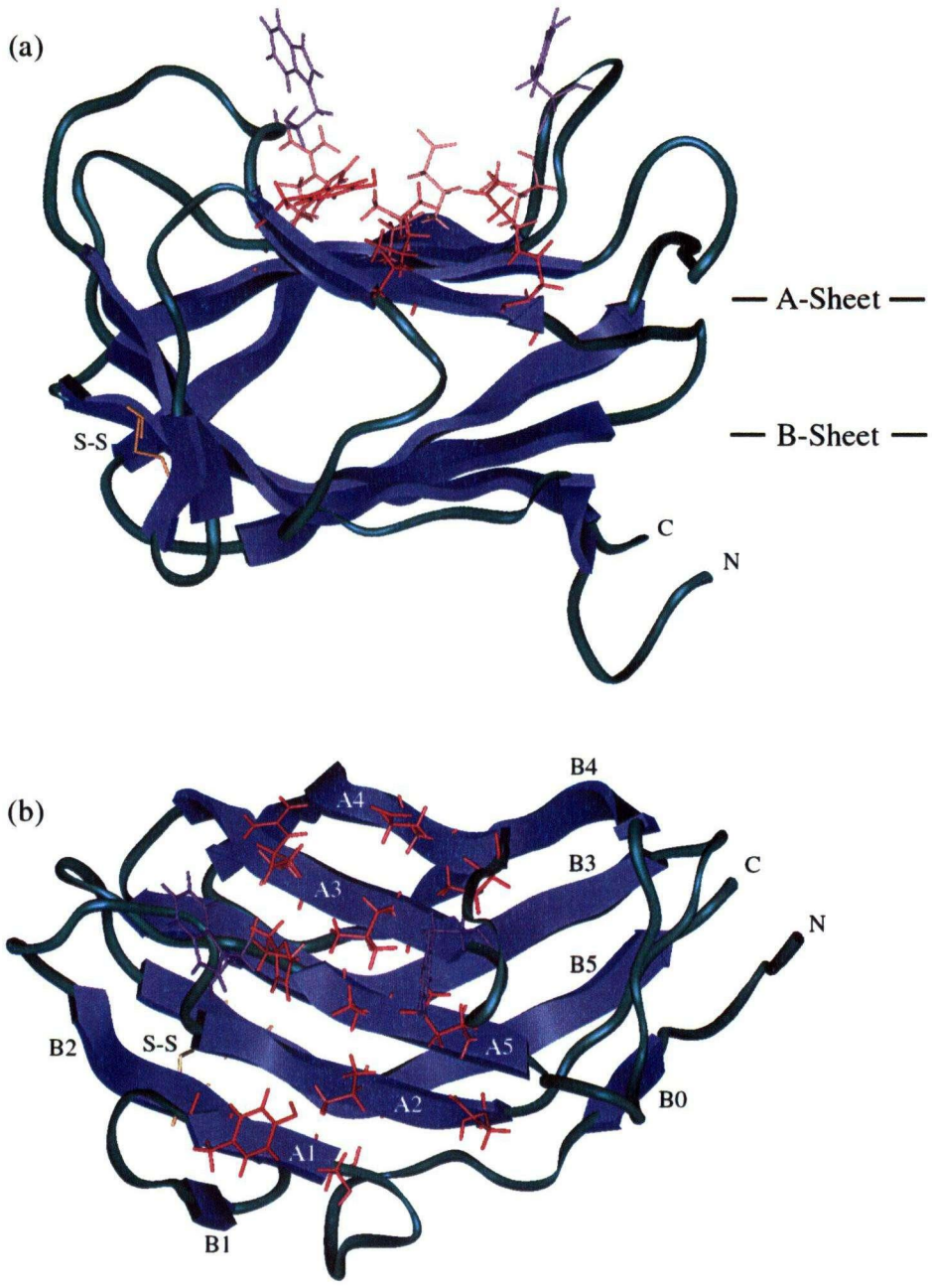


***Figure 1.9**: (a) Side and (b) Top views of the tertiary structure of CBD_{N2}. The disulphide bond comprised of Cys38 and Cys142 is colored in yellow and those residues involved in the binding are colored in pink and purple (Tyr91 and Trp49).*

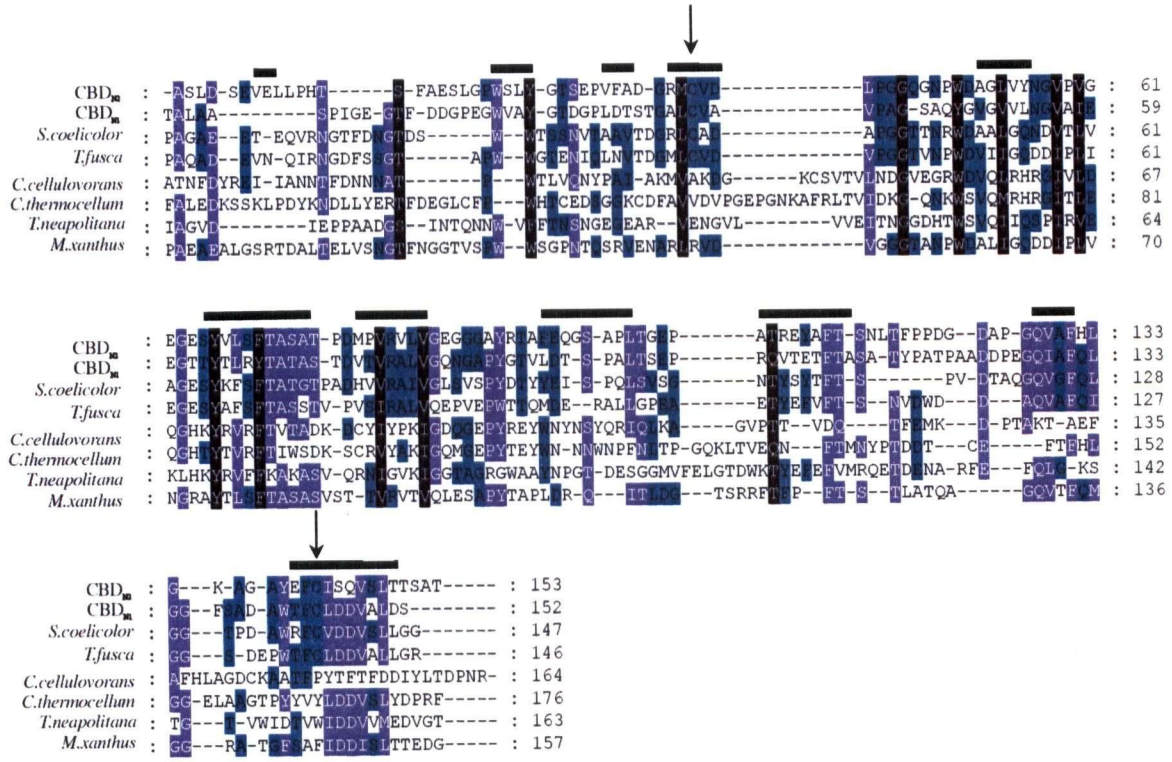


***Figure 1.10**: Sequence alignment among the 8 most sequence-similar proteins to CBD$_{N2}$. (all are endoglucanses, with the exception of one laminase) Completely conserved residues are shaded in black, residues with >75% identity are in purple, and those with > 50% identity in turquoise. Residues involved in β-strands are indicated by a thick line above the sequence. Arrows indicate the Cysteine residues involved in the disulphide bond required for the native structure. The sequence alignment was done using GeneDoc from sequences obtained from the SwissProt database.*

CBD_{N2} has 8 strands forming the jelly-roll motif with three extra strands that lengthen the A and B sheets by 1 and 2 respectively.

The cellulose binding site of CBD_{N2}, identified by cellooligosaccharide induced chemical shift deviations in HSQC spectra, is a groove across the side of one β-sheet which contains 13 residues directly involved in binding. (Figure 1.9) The 8 most similar sequences are from other endoglucanases, and identity among the sugar binding residues and within the β-strands is common. (Figure 1.10).

The β-sandwich topology is common in sugar binding domains, as evidenced by prevalence of the OB[9]-folding motif, the FN3[10] module, and the IG[11] module (Murkin, 1993; Bork 1996). These latter domains are however much smaller than CBD_{N2} having on average only 5 β-strands, although a few have similar stabilities (Plaxco *et al*, 1997). While many kinetic folding studies had focussed on the β-sandwich topology, at the beginning of this study there was correspondingly little known about the residue specific stability, and conformational landscape of this folding topology.

CBD_{N2} is an ideal representative for the β-sandwich topology, and is a good candidate for the native state HX experiment. The large size of CBD_{N2} ensured that multiple residues per secondary element could be probed for greater confidence on results. As well, the excellent stability (and reversibility) to chemical denaturant indicates that HX of CBD_{N2} could be measured throughout a large range of non-denaturing [GuDCl]'s.

The motivation for this work was to explore the thermodynamic stability of CBD_{N2}, primarily as a representative of the jellyroll class of β-sheet proteins, but also as a useful protein for industrial cellulose degradation. By identifying the factors that affect global stability of the

---

[9] The (oligonucleotide/oligosaccharide binding)-fold is a 5 stranded β-sheet coiled to form a closed β-barrel.
[10] Fibronectin type III module
[11] Immunoglobulin "superfamily"

protein, as well as the distribution of stability within the structure, we hope to gain insight on the free-energy landscape of CBD_{N2}, and identify the common characteristics (conservation, function) of stable residues.

Chapter 2
Thermodynamic and Kinetic Parameters for the Unfolding Transition of CBD_{N2}

2.0 Overview

Measurements of thermodynamic parameters for the global denaturation of CBD_{N2} were performed using circular dichroism (CD) spectroscopy. Reversible temperature and [GuDCl]-dependent unfolding transitions of CBD_{N2} were measured as a function of pD. Results from parallel Differential Scanning Calorimetry (DSC) unfolding experiments are also included. Analysis of thermodynamic parameters requires that the folding/unfolding reaction is reversible, thus the reversibility of CBD_{N2} from GuDCl and temperature denaturation is also shown.

Stopped flow fluorescence experiments to obtain CBD_{N2} refolding rates in non-denaturing concentrations of GuDCl were performed. These refolding rates are compared to intrinsic unstructured peptide amide hydrogen exchange rates in chapter 3 to confirm EX2 exchange in the native state HX experiments.

2.1 Materials and Methods

2.1.1 Sample Conditions

CBD_{N2} was produced and purified as described in Appendix B. D_2O was used instead of H_2O in all experiments to be consistent with the NMR HX data. All cited pD are those corrected for the isotope effect ($pD = pH_{measured} + 0.4$). Potassium phosphate buffer at 5mM (CD and Fluorescence) or 50mM (DSC) was dissolved in D_2O and adjusted to the correct pD using DCl or NaOD. 1-2 mg/ml of CBD_{N2} were used for the CD and DSC experiments while 0.1 mg/ml was used for the fluorescence samples.

For the GuDCl denaturation experiments, GuHCl (Sigma) was dissolved in 5 mM potassium phosphate-D_2O prior to pD adjustment. GuDCl concentration was determined from the difference between the index of refraction of a reference solution and the solution containing GuDCl, according to (2.1). Measurements to determine [GuDCl] within ±0.05 M were made at room temperature on a Spectronic Instruments refractometer (model number 334610) from the difference between the index of refraction of a reference solution, n_{ref}, and a solution containing GuHCl, n (Pace *et al*, 1992).

$$[\text{GuHCl}] = 57.147\,(n - n_{ref}) + 38.68\,(n - n_{ref})^2 - 91.6\,(n - n_{ref})^3 \qquad (2.1)$$

Thermodynamic parameters are extracted from systems in equilibrium. Thus all samples made for GuDCl denaturation were left to equilibrate overnight, and samples in thermal denaturation experiments were allowed to equilibrate 5 minutes per 1 °C temperature increment.

2.1.2 CD Spectroscopy

All CD spectroscopy experiments were performed on a JASCO-720 spectrophotometer. The scan rate for all experiments was 50 nm/min with a response time of 2 seconds to yield a

response wavelength of 1.71 nm, in keeping with manufacturer's suggestions. The bandwidth was chosen to be 2.0 nm with a resolution of 0.5 nm. Six spectra were acquired for every condition variable and averaged by the acquisition software.

2.1.3 Stopped-Flow Fluorescence Spectroscopy

Refolding kinetics of CBD_{N2} were monitored using a stopped-flow apparatus from Bio-Logic (France). The fluorescence detector was a SFM-400/S and the four syringe mixer was a MOS-450/AF-CD. The refolding buffer was located in 3 syringes, while the denatured protein was placed in the fourth. The cuvette was TC-100/10 with a dead volume of 30.2 μL. The maximum dead-time for the volumes used in these experiments was 2.3 ms (corresponding to 3 refolding buffer samples of 150 μL, a denatured protein sample of 70 μL, and a mixing time of 40 ms). Two scans were averaged for every refolding concentration.

2.1.4 Differential Scanning Calorimetry

DSC measurements were performed with a Calorimetry Sciences Corp. (Spanish Fork, UT) model 4207. Background excess thermal power scans were obtained with buffer in the sample cell and were subtracted from the scans for each 0.5 mL sample solution containing 1.5 to 2.0 mg/mL CBD_{N2}. For each pD, 4-5 thermograms were obtained. The scan rate used was 1 °C/min Reversibility of unfolding was determined by reheating the sample after cooling in the calorimeter.

2.1.5 Denaturant Induced Protein Unfolding

Many proteins exhibit two-state unfolding when monitored by conventional spectroscopic techniques (Pace *et al*, 1992). Indeed the extraction of thermodynamic unfolding parameters is

most often performed using equations that assume two-state behavior, with validation of this assumption based on the efficacy of the fit.

In a two-state analysis model, the sum of the fraction of folded (f_F) and unfolded (f_U) protein is normalized to 1:

$$f_U + f_F = 1 \tag{2.2}$$

Thus any observed signal y during an unfolding reaction can be written as:

$$y = y_F f_F + y_U f_U \tag{2.3}$$

where y_F and y_U represent the signal characteristic to the folded and unfolded state, respectively. In general there is a linear dependence on denaturant concentration of the folded and unfolded signal and therefore:

$$y_U = b_U + a_U[denaturant]$$

$$y_F = b_F + a_F[denaturant] \tag{2.4}$$

where $b_{U/F}$ is the extrapolated unfolded/folded signal at 0 M denaturant and $a_{U/F}$ is the dependence of the unfolded/folded signal on denaturant.

Combining (2.2) and (2.3) to solve for f_U:

$$f_U = \frac{y_F - y}{y_F - y_U} \tag{2.5}$$

An equilibrium constant K for the unfolding transition is the ratio of unfolded versus folded protein and can be written:

$$K = \frac{f_U}{1 - f_U} \tag{2.6}$$

Since this expression neglects the presence of denaturant in the solution, K is not a full, but an apparent, equilibrium constant. The free energy of unfolding can be elicited from the thermodynamic expression:

$$\Delta G(denaturant) = -RT\ln(K)$$
$$\Delta G(denaturant) = -RT\ln(\frac{f_U}{1-f_U}) \quad (2.7)$$

For denaturant unfolding reactions, the free energy of unfolding is observed to vary with denaturant concentration in a linear fashion (Pace *et al*, 1992):

$$\Delta G(denaturant) = \Delta G_U(H_2O) + m_U[denaturant] \quad (2.8)$$

ΔG_U (H_2O) or ΔG_U (D_2O) is the free energy of unfolding in the absence of denaturant, and m_U is the dependence of ΔG (denaturant) on [denaturant]. m_U is often interpreted as a measure of the protein surface area exposed to the denaturant upon unfolding. Setting (2.7) equal to (2.8), and substituting in (2.4) and (2.5), one can derive that the observed signal for a denaturant induced unfolding of a protein is:

$$y = \frac{b_F + a_F[den] + (b_U + a_U[den])e^{(-\Delta G_U(H_2O) - \frac{m[den]}{RT})}}{1 + e^{(-\Delta G_U(H_2O) - \frac{m[den]}{RT})}} \quad (2.9)$$

Non-linear least squares fitting of y versus [denaturant] allows for the extraction of $\Delta G_U(H_2O)$ (or in this case, $\Delta G_U(D_2O)$) and m_U.

2.1.5 Temperature Induced Protein Unfolding

Two state analysis of thermal denaturation profiles proceeds inititally in the same way as denaturant unfolding analysis. A full equilibrium constant is extracted from the spectroscopic signal changes (2.6), and is inserted into the equation for free-energy:

$$\Delta G(T) = -RT \ln K(T) = -RT \ln(\frac{f_U}{1 - f_U}) \qquad (2.10)$$

In denaturant unfolding reactions, extraction of the unfolding free-energy in native conditions, ΔG_U (H_2O), is made simple by the empirical observation that ΔG is linearly dependent on denaturant concentration (2.8). In thermal unfolding reactions, however, the free energy varies non-linearly with temperature because the heat capacity of a folded and unfolded protein is significantly different. To obtain the free energy of unfolding from a thermal denaturation profile, it is necessary to independently measure the change in heat capacity ΔC_p. The free-energy of the temperature unfolding transition (derived in Appendix C) is then written:

$$\Delta G_U(T) = \Delta H_m(1 - \frac{T}{T_m}) - \Delta C_p(T_m - T + T \ln(\frac{T}{T_m})) \qquad (2.11)$$

where T_m is the temperature at the midpoint of the unfolding transition, ΔH_m is the enthalpy change at this midpoint, and ΔC_p is the change in heat capacity of the folded and unfolded protein.

The assumption in (2.11) is that the unfolding process is two-state. Unlike GuDCl denaturation experiments, however, the equilibrium constant is not an apparent, but a full one.

Without foreknowledge of ΔC_p, ΔG_U cannot be extracted directly from a temperature denaturation experiment; however, the midpoint unfolding temperature T_m, and enthalpy change, ΔH_m, can in the following manner. Free-energy is related to enthalpy according to (2.12):

$$\Delta G(T) = \Delta H(T) - T\Delta S \tag{2.12}$$

At the midpoint temperature, $\Delta G(T_m)=0$. Therefore from (2.12):

$$\Delta S = \frac{\Delta H}{T_m} \tag{2.13}$$

(2.12) can be re-written as:

$$\Delta G(T) = \Delta H(1 - \frac{T}{T_m}) \tag{2.14}$$

Setting (2.14) equal to (2.10), and including the linear dependence of the folded and unfolded signal with temperature, the signal observed during a temperature unfolding reaction can be written as:

$$y = \frac{b_F + a_F T + (b_U + a_U T) e^{(-\frac{\Delta H}{R}(\frac{1}{T} - \frac{1}{T_m}))}}{1 + e^{(-\frac{\Delta H}{R}(\frac{1}{T} - \frac{1}{T_m}))}} \tag{2.15}$$

where b_F=folded signal at T=0
a_F=slope of folded signal on T
b_U=unfolded signal at T=0
a_U= slope of unfolded signal on T
T=temperature of the sample
T_m=melting temperature
R=Gas constant

By performing a non-linear least squares fit to (2.15), ΔH and T_m can be extracted from spectroscopic data. Since a measure of the free-energy of unfolding at 25 °C is desired, determination of the change in heat capacity due to temperature denaturation is necessary for substitution into (2.11).

ΔC_p, defined as the variation of enthalpy change with temperature, or:

$$\Delta C_p = \frac{d(\Delta H)}{dT} \tag{2.16}$$

and can be measured from the slope of ΔH versus T_m graph for temperature denaturation data taken at different pH values, or in the presence of GuDCl. The assumption is that while the different solution conditions affect T_m, ΔC_p is constant. ΔC_p is extracted from the strong temperature dependence of the enthalpy (2.16).

2.1.6 Analysis of GuDCl and Temperature Denaturation Profiles

An entire CD spectrum was obtained at every temperature or [GuDCl] increment. As a result, the unfolding transition of CBD_{N2} could be monitored at multiple wavelengths. The greatest difference in CD signal during temperature denaturation occurred between 235-240 nm. In the GuDCl denaturation experiments, the greatest difference in CD signal of denatured and native protein occurred between 222-228 nm. Since GuDCl has significant CD signal, a corresponding buffer blank was subtracted from every GuDCl denaturation spectra.

Non-linear least squares fitting of the unfolding traces to equations (2.9) or (2.15) was performed in Plotdata to extract ΔG_U (D_2O) and m_U for GuDCl unfolding experiments or ΔH_m and T_m for the temperature denaturation experiments. Thermodynamic parameters from each wavelength fit were then averaged. The standard deviation of the set of standard errors from the fit parameters was taken to be the error in the averaged parameter (Taylor, 1982).

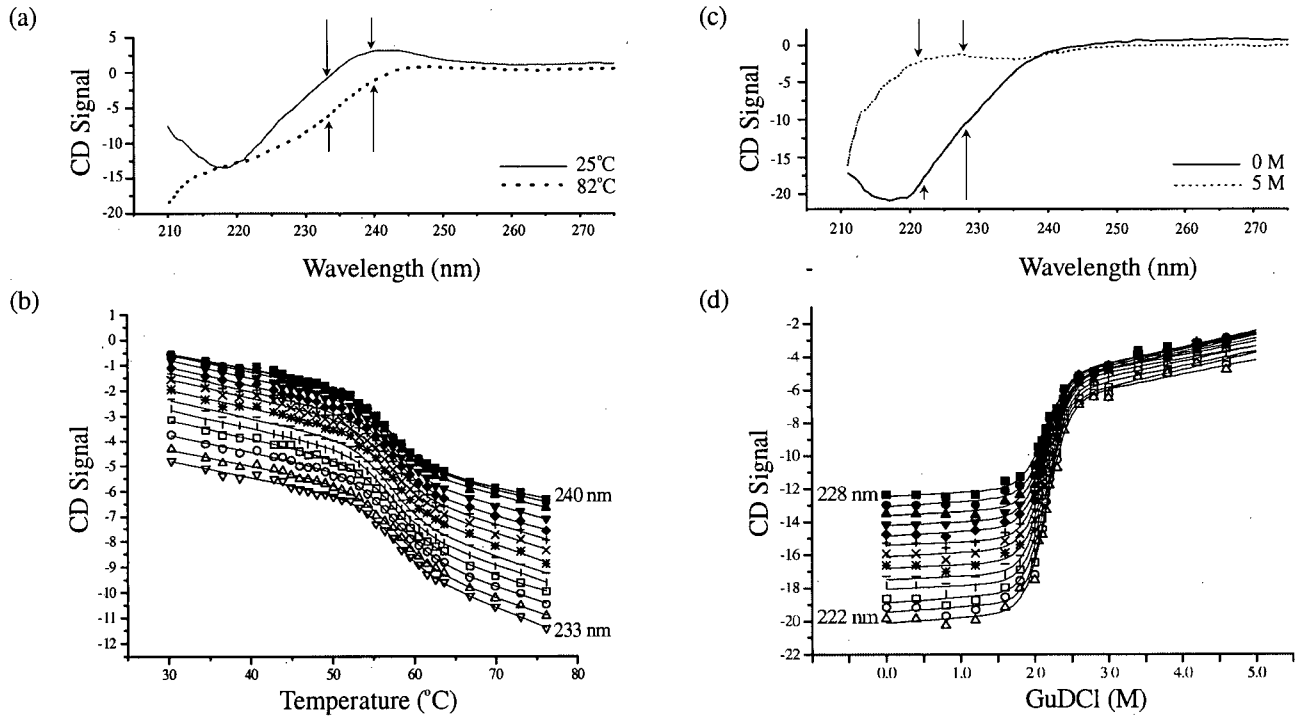


***Figure 2.1**: (a) Native and GuDCl unfolded CD spectra of CBD_{N2}. (b) Plotdata fits at multiple wavelengths to the GuDCl denaturation profiles. (c) Native and temperature unfolded spectra of CBD_{N2}. (d) Plotdata fits at multiple wavelengths to the temperature denaturation profiles.*

2.2 Thermodynamic Results

2.2.1 Unfolding Reversibility of CBD_{N2} after Thermal and Chemical Denaturation

CBD_{N2} unfolds reversibly upon thermal denaturation. Refolding from 80 °C at pD=6.0 is shown in Figure 2.2. The folded and refolded traces do not perfectly overlap, presumably because some protein aggregation occurs at high temperature. Nevertheless, the CD spectral profile of both folded and refolded CBD_{N2} are the same and distinct from that of the unfolded spectra.

Refolding from GuDCl denaturation was established by taking a spectra of CBD_{N2} in 5 M GuDCl, then diluting/concentrating it multiple times until the GuDCl concentration was below 0.1 M, where CBD_{N2} is folded. A new spectra of the refolded sample was then compared to native protein spectra. Since it is very difficult to concentrate the refolded protein to the exact concentration of the initial sample, the two spectra do not superimpose. However their respective morphology is the same, establishing that CBD_{N2} refolds reversibly from GuDCl denaturation.

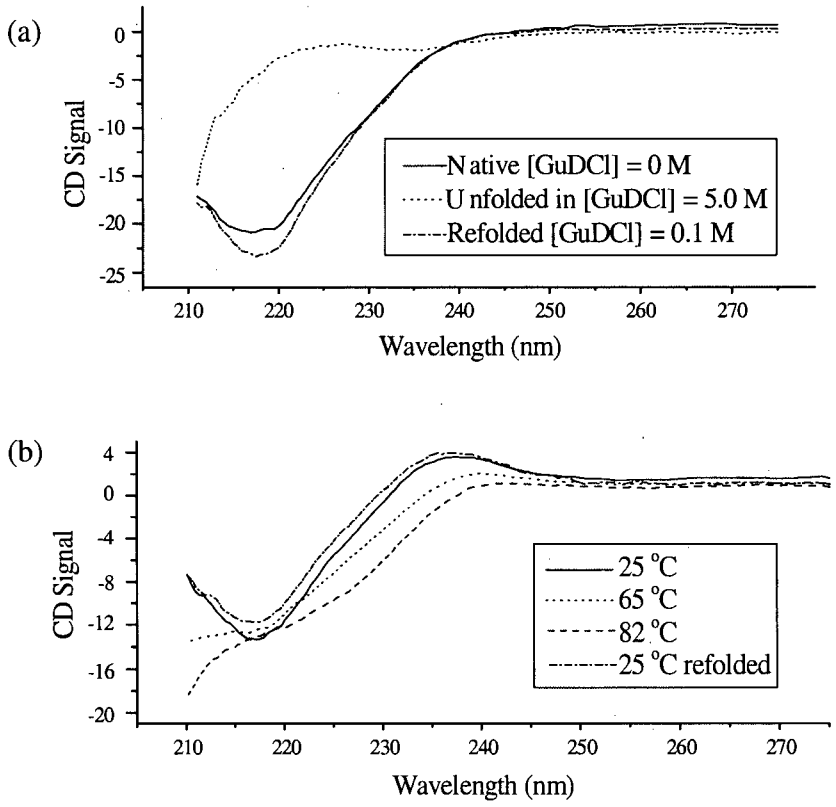


***Figure 2.2**: Refolding of CBD_{N2} after (a) GuDCl denaturation at pD=6.0 and (b) temperature.*

2.2.2 GuDCl Denaturation as a Function of pD

GuDCl denaturation of CBD_{N2} was monitored at 3 pD values (5.0, 6.0, 7.0) using CD spectroscopy. This was done to examine any stability trends as a function of pD, and to obtain a ΔG_U at pD=6.0, and T=25 °C, for comparison with NMR experiments (Figure 2.3).

Table 2.1: Thermodynamic Results for GuDCl denaturation experiments at 25 °C

pD	$[GuDCl]^{1/2}$ (M)	±	m (kJ/mol•M)	±	$\Delta G_U(D_2O)$ (kJ/mol)	±
5.0	2.05	0.02	-21.0	4.3	43.2	8.9
6.0	2.16	0.02	-17.0	2.0	36.7	4.3
7.0	2.31	0.02	-13.8	1.6	31.8	3.7

Table 2.1 details the results of the unfolding of CBD_{N2} in GuDCl. The most notable result was an increase in the unfolding midpoint of CBD_{N2} by 0.13 M per unit increase in pD. This indicates that, under denaturing conditions the stability of CBD_{N2} increases slightly with pD. However, the "m-value", a gauge of the protein surface area exposed in response to the presence of GuDCl (Creighton, 1992), decreases by 3.6 kJ/mol·[GuDCl] with each pD unit. Consistent with the common interpretation of m-values, this suggests that CBD_{N2} surface area exposure, upon unfolding, becomes less sensitive to the presence of GuDCl. This result is surprising, and a physical interpretation is hard to envision. However one might propose that if CBD_{N2} becomes less sensitive to the denaturing effects at higher pD, it is consistent that the unfolding midpoint would increase with pD.

The extrapolated values of $\Delta G_U(D_2O)$ indicated that the stability of CBD_{N2} actually decreases with increasing pD. The reverse trend of $\Delta G_U(D_2O)$ as compared to the midpoint value,

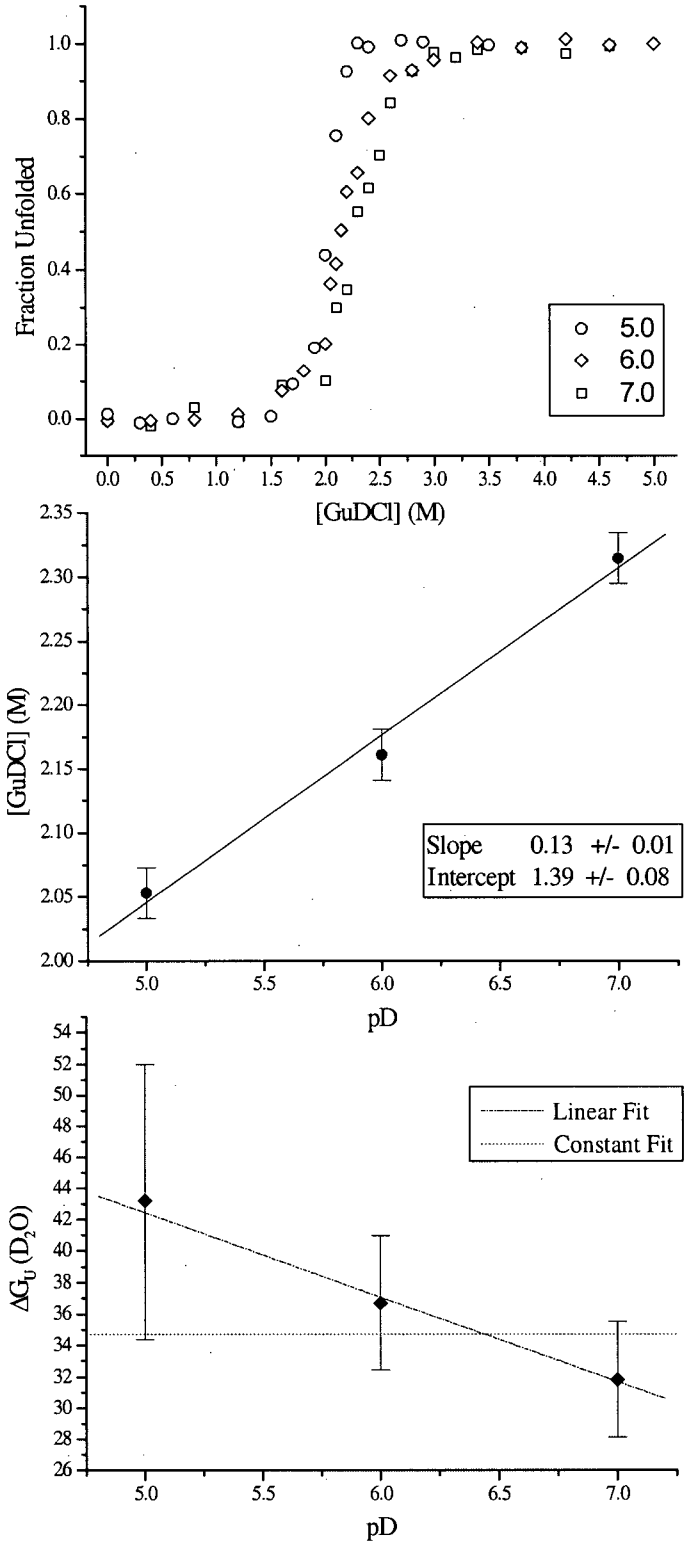


***Figure 2.3**: (a) The fraction unfolded of CBD_{N2} in the presence of GuDCl as a function of pH at 25 °C. (b) GuDCl unfolding midpoint as a function of pD. (c) $\Delta G_U(D_2O)$ as a function of pD.*

$[GuDCl]_{1/2}$, is due to the associated changes in m_U-values. However the significance of the trend is unclear based on the large errors associated with the extrapolation. Indeed the fitting error is so large that the $\Delta G_U(D_2O)$ could well be equal for the different pD values (Figure 2.3(c)).

2.2.3 Thermal Denaturation as a Function of pD.

Thermal denaturation experiments of CBD_{N2} were measured using CD spectroscopy at pD = 5.4, 6.0, 7.0, and 8.9. Below pD = 5.4, these experiments could not be performed as the protein aggregates. Table 2.2 shows the results of the nonlinear least squares fits to the temperature denaturation data. The range of 3.5 pD units did not suffice to show a change in melting temperature larger that 2 degrees.

Table 2.2: Thermodynamic Parameters from Thermal Denaturation of CBD_{N2} as a function of pD

pD	T_m (°C)	±	ΔH (kJ/mol)	±
5.4	67.4	0.1	600	29
6.0	67.0	0.1	480	22
7.0	68.0	0.5	350	35
8.9	65.8	0.2	590	58

The motivation for acquiring thermal denaturation data as a function of pD was to extract a value for the change in heat capacity during unfolding, ΔC_p, as required for extrapolation of ΔG_U to 25 °C (equation 2.16). Since the spread in T_m was so small and scattered, a linear fit of ΔH_m versus T_m was not viable for obtaining ΔC_p.

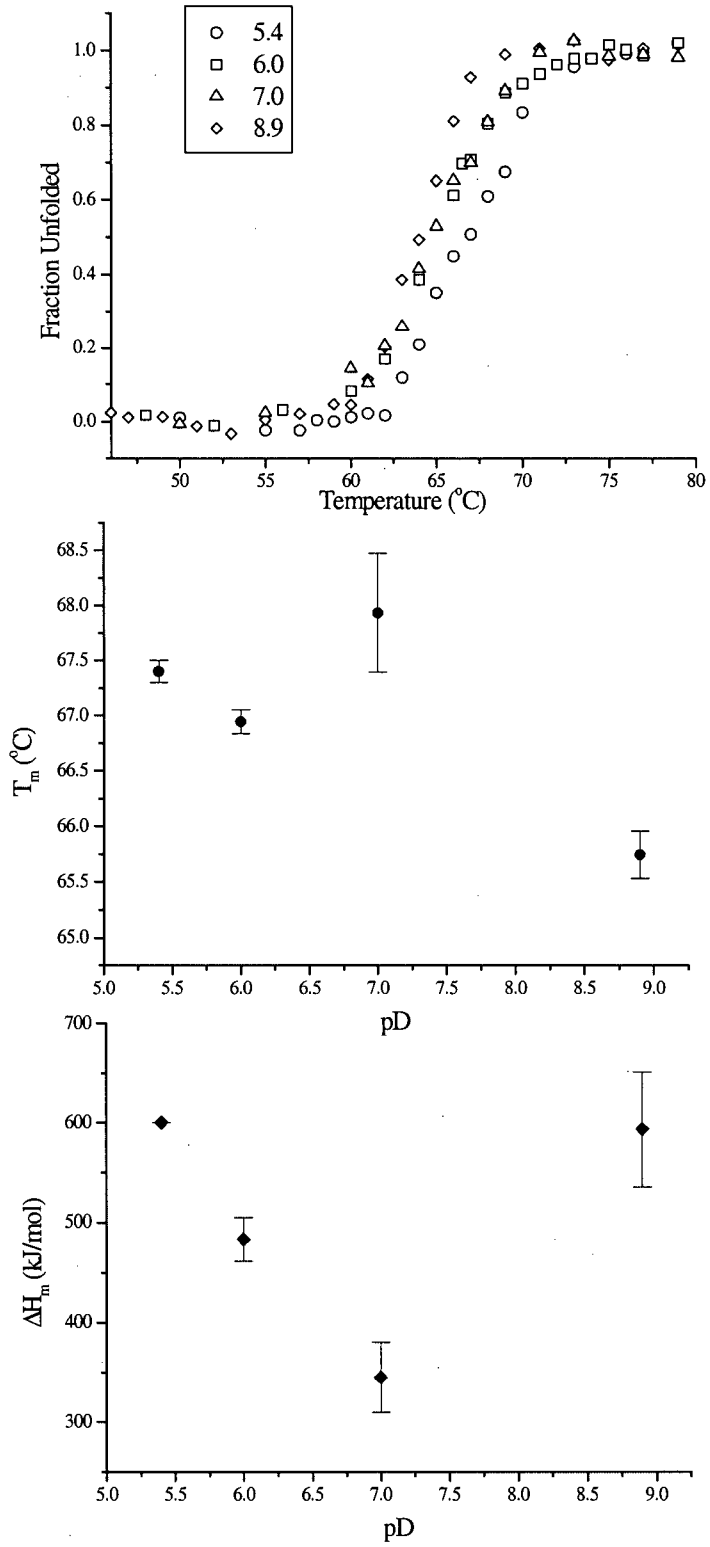


***Figure 2.4**: (a) The fraction unfolded during a temperature melt of CBD_{N2} as a function of pD. The dependence of (b) T_m and (c) ΔH_m on pD for CBD_{N2}.*

2.2.4 DSC of CBD_{N2} as a Function of pD

In collaboration with Dr. Louise Creagh the thermodynamics governing the unfolding of CBD_{N2} were measured as a function of pD using DSC. Thermograms were fit to a two-state model to extract ΔH_m and T_m. However, as also seen by the CD spectroscopic analysis, the observed unfolding midpoint temperatures varied by only 4 °C between pD 5.6 and pD 12.0. Again, since the spread in T_m was so small, there is little confidence in a linear regression in a plot of ΔH_m and T_m (Figure 2.6). Scatter in the data is attributed to ill-defined baselines.

Table 2.3: DSC Data for CBD_{N2} as a Function of pD

pD	T_m (°C)	±	ΔH_m (kJ/mol)	±
5.6	70.1	0.1	486	9
6	74.2	0.1	540	0
7	73.8	0.5	528	11
7.8	74.0	0.1	542	9
8.6	71.4	0.6	516	21
10	72.1	0.1	516	13
10.9	70.9	0.2	497	7
12	70.4	0.2	486	8
		ΔC_p	12.7	2.3

2.2.5 Thermal Denaturation as a Function of GuDCl to obtain ΔC_p

Given that obtaining ΔC_p from thermal denaturation profiles as a function of pD was met with little success, thermal denaturation experiments at a constant pD (5.4) with low, non-denaturing concentrations of GuDCl were performed in anticipation of a larger spread of T_m and ΔH values. To analyze this data it was necessary to assume that the ΔC_p was independent of [GuDCl].

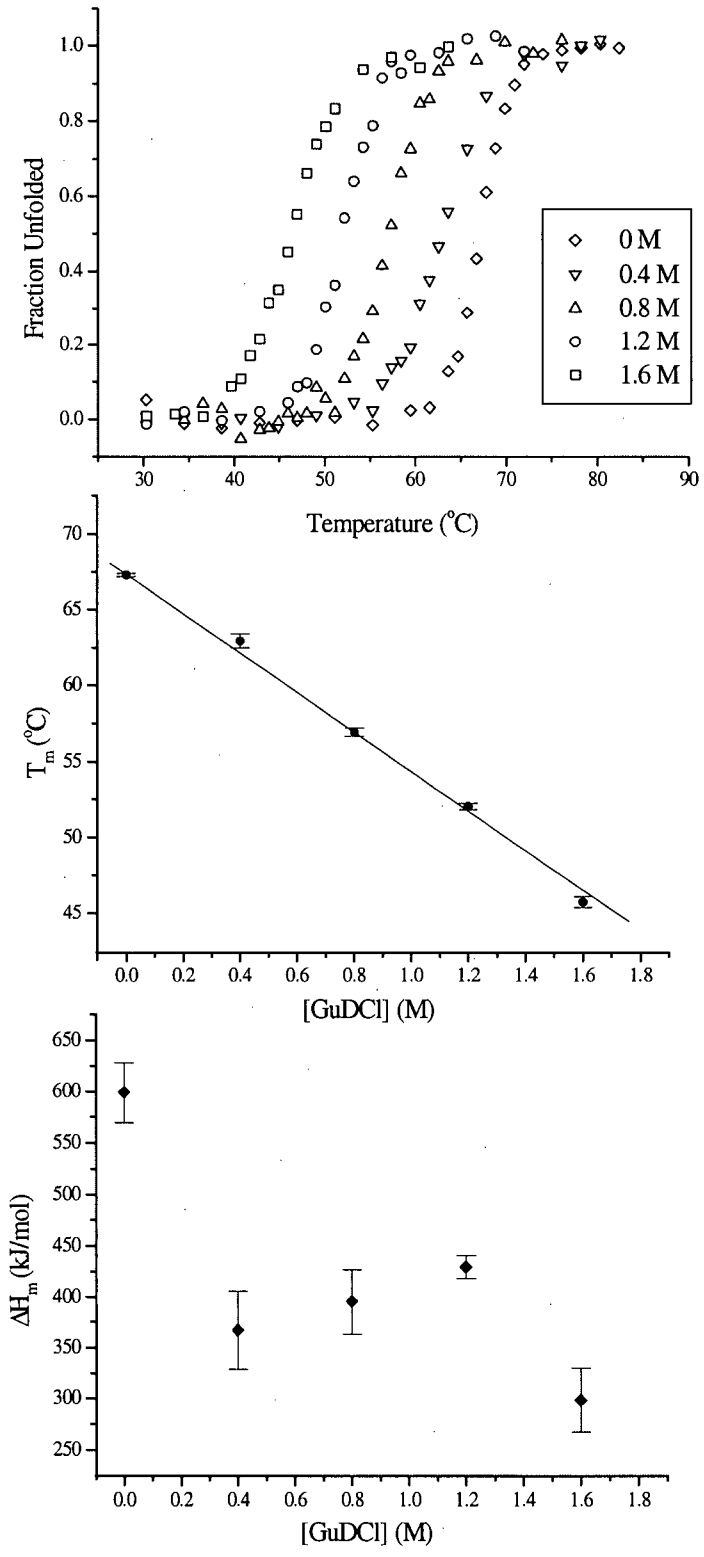


***Figure 2.5**: (a) The fraction unfolded during temperature denaturation of CBD_{N2} as a function of GuDCl at pD=5.4. The temperature unfolding midpoint (b) as a function of GuDCl at pD=5.4.*

Table 2.4 presents the results of the temperature denaturation experiments as a function of pD. As would be expected, the temperature unfolding midpoint, T_m, decreased by ~ 5 °C as the [GuDCl] increased by 0.4M. While the T_m values for this study were spread out and monotonically decreasing, the corresponding enthalpies did not have a corresponding decreasing trend, making linear regression analysis to obtain ΔC_p unproductive and statistically insignificant.

Table 2.4: Thermodynamic Parameters from Temperature Denaturation of CBD_{N2} as a function of [GuDCl] at pD=5.4

[GuDCl] (M)	T_m (°C)	±	ΔH_m (kJ/mol)	±
0	67.4	0.1	599	29
0.4	63.0	0.5	367	38
0.8	57.0	0.3	395	32
1.2	52.1	0.2	429	11
1.6	45.8	0.4	299	31

An alternative for obtaining ΔC_p was elicited from (2.11). Instead of obtaining the $\Delta G_U(T)$ using ΔC_p, the converse was tried. Prior to the temperature unfolding experiments as a function of denaturant, a simple GuDCl unfolding experiment was performed at pD=5.4. Thus $\Delta G([GuDCl])$ was known for each [GuDCl] used in the temperature study. By substituting ΔH_m and T_m, from Table 2.4 and $\Delta G([GuDCl])$ at each [GuDCl] into (2.11), 5 values for ΔC_p were obtained and averaged. This value for ΔC_p, as well as the DSC-ΔC_p and a rule-of-thumb determination are listed in Table 2.5.

2.2.6 Comparison of ΔC_p and ΔG_U Results

A major goal of the GuDCl and temperature denaturation experiments was to obtain a ΔG_U at 25°C and pD=6.0. The ΔC_p values determined from different methods were substituted, in conjunction with ΔH_m and T_m from the temperature melt at pD=6.0, into (2.11) to give ΔG_U detailed in table 2.5.

Table 2.5: Tabulation of ΔC_p and ΔG_U results for CBD_{N2}

	ΔC_p (kJ/mol· K)	±	$\Delta G_U(D_2O)$ (kJ/mol) at T=25°C and pD=6.0	±
Rule of Thumb	7.7 [a]	N/A	38.9 [b]	N/A
CD	8.5	2.0	36.7 [b]	6.1 [d]
DSC	12.7	1.4	27.6 [c]	5.1 [d]
CD (GuDCl Unfolding)	N/A		36.7	4.3

(a) As a rule of thumb one can estimate ΔC_p by multiplying the number of residues a protein has by 12 kcal/residue/mol/K.

(b) (2.11) was used to fit this data, and the values for ΔH_m and T_m were taken from the temperature denaturation values as a function of pD at pD=6.0

(c) Obtained using equation (2.11) and the values for T_m and ΔH_m obtained from the D.S.C. experiments performed in conjunction with Dr. Louise Creagh.

(d)

$$\delta\Delta G_U = \sqrt{[(\frac{d}{dT_m}\Delta G)\delta T_m)^2 + (\frac{d}{d\Delta H_m}\Delta G)\delta\Delta H_m)^2 + (\frac{d}{d\Delta C_p}\Delta G)\delta\Delta C_p)^2}$$

The "Rule of Thumb" and CD determination of ΔC_p are in better agreement than with the DSC results. While calorimetry (DSC) is generally thought to be the most cogent manner to obtain ΔC_p, the scatter within the DSC data and the extracted ΔC_p are both large. Figure 2.6 shows the slope of the line used to extract ΔC_p from the DSC data, along with another line representing the ΔC_p value obtained from CD experiments.

The agreement between ΔG_U obtained from GuDCl and temperature unfolding experiments using CD spectroscopy is surprising, given that the data used to obtain the respected

values of ΔG_U is completely independent. The ΔG_U (D_2O) value obtained from DSC experiments is extremely low.

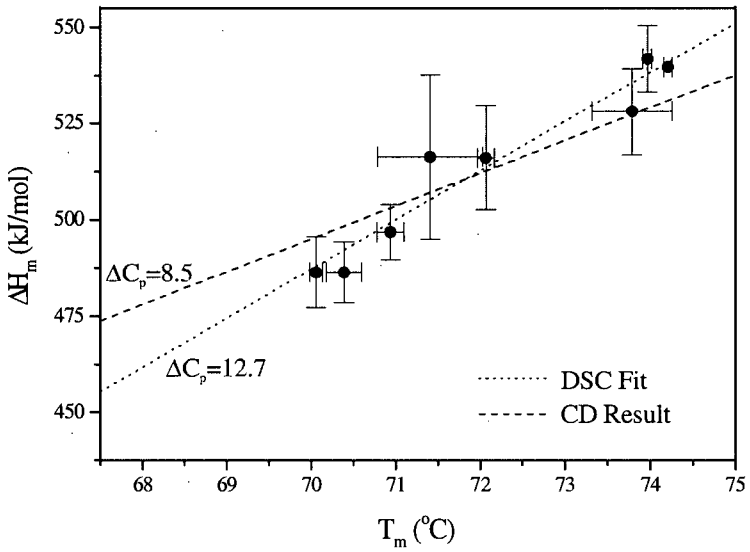


Figure 2.6: *Comparison of ΔC_p obtained from DSC and CD measurements*

2.2.7 Thermodynamic Conclusions

The intrinsic assumption in all the analysis was that CBD_{N2} unfolds in a 2-state transition. Given the general excellence of the fits of data to 2-state models, from both DSC and CD measurements, CBD_{N2} does not appear to have any detectable folding intermediate.

CBD_{N2} exhibits a surprising lack of pD dependent stability, evidenced by the narrow range of the temperature unfolding midpoint, T_m, in both DSC and CD unfolding experiments. This is likely a consequence of few titratable side-chains in the structure of CBD_{N2} (Brun, 2000). Given this robust characteristic, CBD_{N2} makes an excellent candidate for degradation procedures in the pulp and paper industry.

While CBD_{N2} displayed pD independence for temperature denaturation experiments, GuDCl denaturation experiments revealed a increase in the unfolding midpoint concentration by 0.13 M with every unit pD increase. These results suggest that the presence of GuDCl must perturb the protein in such a way as to introduce a pD sensitivity. If this pD sensitivity is truly a reflection of GuDCl sensitivity, then the extrapolated value of ΔG_U to zero denaturant ($\Delta G_U(D_2O)$) should be pD independent, so as to be consistent with the temperature denaturation experiments. Indeed, while there is a scatter in the ΔG_U (D_2O) from the GuDCl unfolding experiments, all the values are equal within their associated error.

2.3 Kinetic Results

2.3.0 Refolding of CBD_{N2} from GuDCl Denaturation

Refolding kinetics at pD=6.0 of CBD_{N2} into 0.11, 0.21, 0.58, and 0.74 M GuDCl from 5 M GuDCl were fit with the best residuals to a 3 independent exponential functions using Plotdata non-linear regression software.

$$y = A_1 e^{k_1 t} + A_2 e^{k_2 t} + A_3 e^{k_3 t} + y_\infty \tag{2.17}$$

Table 2.6: Refolding Rates for CBD_{N2} in Non-Denaturing Concentrations of GuDCl

[GuDCl] (M)	Refolding Rate (1/s)					
	Fast Phase	±	Medium Phase	±	Slow Phase	±
0.11	23.7	2.2	1.2	0.04	0.110	0.0019
0.21	5.4	0.5	0.48	0.018	0.032	0.0010
0.58	6.3	0.3	0.20	0.0058	0.0092	0.00035
0.74	7.9	0.4	0.14	0.0035	0.0063	0.00025

Despite the success of 2-state models in describing the thermal and chemical denaturation, three refolding phases were observed in the refolding kinetics of CBD_{N2}. The existence of multiple slower phases in the refolding kinetics of CBD_{N2} is likely a result of *cis / trans* isomerization of this 13 proline protein.

The native structure of CBD_{N2} places Pro 76 in the higher energy *cis* conformation, which likely relaxes to the lower energy *trans* conformation in the denatured protein. The slowest phase in the refolding kinetics is likely the required *trans* → *cis* conformation of Pro 76 to form native CBD_{N2}.

In general, not all Pro residues relax to the *trans* conformation in the denatured state. Instead an equilibrium will be reached with 6-38 % of the Xaa-Pro bonds in the *cis*

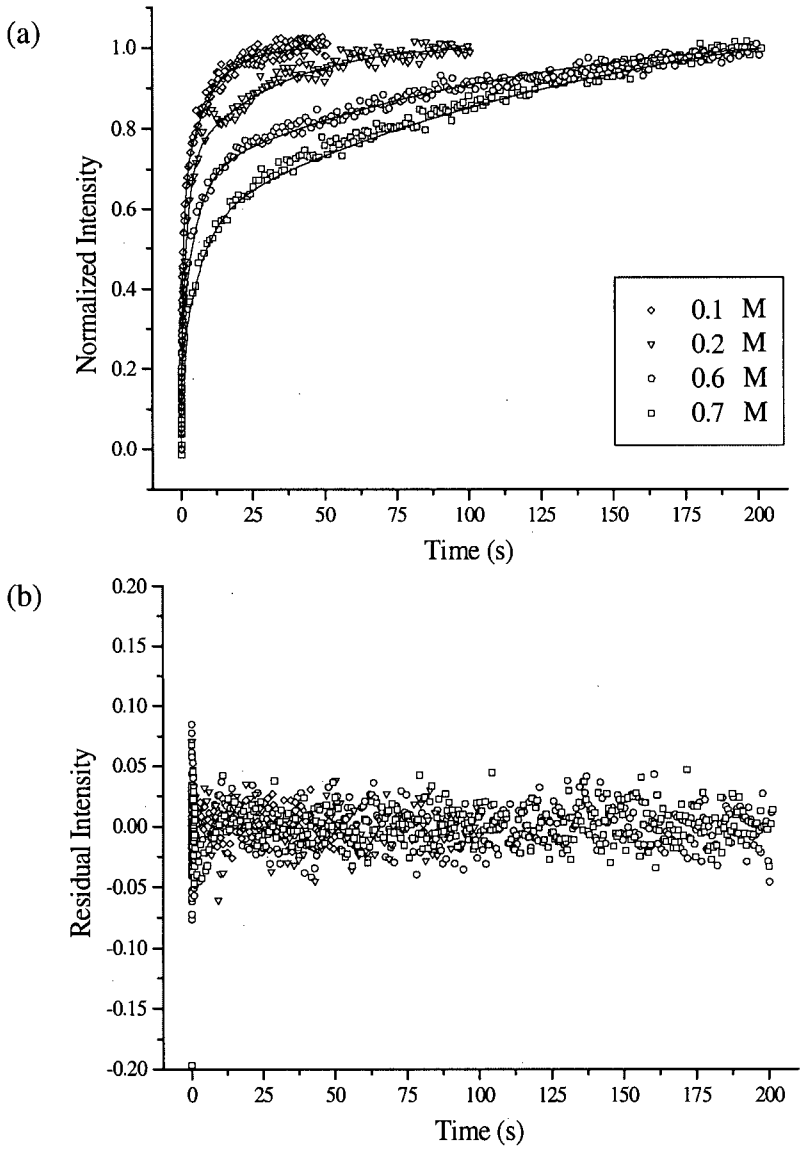


Figure 2.7: *Refolding of CBD_{N2} in non-denaturing concentrations of GuDCl. (a) Stopped-flow refolding traces and (b) residuals for the 3 exponential fit to this data.*

conformation, depending on the identity of the preceeding amino acid, Xaa (Huyghues-Despointes *et al*, 1999a). Xaa for Pro 76 is Thr 75, and has a comparitively low probability of maintaining the *cis* conformation in the denatured state (9%; Huyghues-Despointes *et al*, 1999a).

On the other hand, the Xaa residue of Pro120 is Phe119, which has a 23% probability of adopting the higher energy *cis* conformation in the unfolded state. Since refolding requires Pro120 to be in the *trans* conformation, it is possible that the medium refolding phase (Table 2.6) consists of the *cis*→*trans* conformation of Pro120 (Huyghues-Despointes *et al*, 1999a).

The salient feature of the proposed origin of the medium and slow refolding phases is that they are a consequence of allowing the prolines of CBD_{N2} to equilibrate in the denatured state. If indeed these latter phases in the refolding kinetics of CBD_{N2} are due to proline-isomerization, unfolding fluctuations in the native state likely refold according to the fast rate of Table 2.6, since by definition the protein is in equilibrium in the native conformation.

Chapter 3
NMR Experiments

3.0 Overview of the Native State HX Experiment

Amide hydrogen to deuterium exchange (HX) is measured quantitatively using 2D NMR HSQC spectroscopy. As a given ^{15}N-labelled amide hydrogen exchanges for a deuteron, it's signal intensity in a ^{1}H-^{15}N HSQC spectrum decreases. Fitting HSQC peak intensity versus time to the equation that describes a single exponential decay yields the first-order rate constant for exchange. The exchange rate of a given amide hydrogen in a folded protein is dependent upon the local or global fluctuations that lead to solvent exposure and/or disruption of hydrogen bonds. The presence of small concentrations of denaturant such as GuDCl destabilizes the protein enough to enhance amide exchange via fluctuations, but not enough to globally unfold the protein. According to models of this "native state" HX, the equilibrium constant responsible for exposing a given amide group to solvent for exchange can be extracted from the protection factor, a ratio of the observed rate constant for exchange in the protein relative to that of a reference unstructured peptide. A site-specific free-energy for the fluctuations leading to exchange can consequently be obtained by extrapolating to conditions of no denaturant.

3.1 Material and Methods

3.1.1 Experimental Conditions

The native state HX experiment requires that a protein remains globally folded through a GuDCl concentration range large enough for trends in the denaturant-dependent exchange to be elucidated. Thus one chooses solution conditions that maintain the protein in a stable folded state in as high of a [GuDCl] as possible. However, several other factors must also be considered when selecting solution conditions and NMR acquisition parameters. These include maintaining a soluble sample, with HSQC spectra that have clearly isolated peaks, and choosing a pH such that amide HX is in the EX2 limit and slow enough for reliable measurement of a first order decay. On the other hand, if the HX is too slow, then months pass before complete amide exchange has occurred, and perturbations to peak intensity including spectrometer instability, solvent evaporation, and protein precipitation decrease the accuracy of the exponential decay profiles. One solution for these latter problems is the acquisition of a reference 1D ^{1}H spectrum along with each HSQC spectrum. One assumes that a resolved upfield resonance in the 1D ^{1}H spectrum is exchange-insensitive and thus the area of the corresponding peak can be used as a normalizing factor for the associated HSQC spectrum.

In keeping with the goal of minimizing HX rates, the experimental temperature was chosen to be 25 °C rather than 35 °C (at which CBD_{N2} was previously characterized) (Brun *et al*, 2000). A water bath was maintained at 25 °C for sample storage between data acquisition. An even lower temperature was not chosen since this would require use of a cooling bath over a period of months and since NMR spectral quality quickly decreases due to slower tumbling of CBD_{N2} in solution.

The experimental pD of 6.0 was chosen based on several considerations. Since every unit increase in pD raises the first-order base catalyzed HX reaction 10-fold, which can lead to EX1 instead of EX2 exchange, the pD was chosen to be as low as possible. However temperature

denaturation of CBD_{N2} at pD=5.0 (pH=4.6) led to aggregation at temperatures higher than 40°C, preventing an independent measure of the stability of the protein. This likely reflects the minimal solubility of CBD_{N2} near its predicted isoelectric point (pI) of ~4.1. Since both GuDCl and temperature denaturation proved reversible at pD 6.0, this condition was selected as a compromise for the NMR HX experiments. Thermodynamic parameters of CBD_{N2} at this and other pD values are discussed in Chapter 2.

3.1.2 NMR Acquisition

All spectra were acquired on a Varian Unity 500 MHz NMR spectrometer equipped with a single axis gradient triple resonance probe. Experiments with $^{1}H^{N}$ detection utilized the enhanced sensitivity pulsed-field gradient approach (Kay *et al*, 1992; Muhandiram and Kay, 1994). ^{1}H was referenced with respect to an external DSS standard at 0.00 ppm. The ^{15}N referencing was to an external sample of 2.9 M $^{15}NH_4Cl$ in 1 M HCl at 24.93 ppm. The ^{1}H spectral width was 6500 Hz with 1024 complex points. The ^{15}N spectral width was 1600 Hz with 96 complex points. To observe the exchange of amides with limited protection, the first four HSQC spectra were acquired with 4 transients per free induction decay (FID) for an acquisition time of 15 minutes each. In order to improve signal-to-noise, all subsequent spectra comprised of 16 transients per FID, for a total experiment time of 1 hour. The peak intensities of the first four spectra were increased by a factor of 4 after processing.

1D ^{1}H spectra were acquired immediately after each HSQC spectrum so that the area under an exchange insensitive peak at –0.25 ppm could be used to normalize the amide intensity. 1024 complex points were acquired in 128 transients with a spectral width of 8000 Hz.

3.1.3 NMR 1D Data Processing and Analysis

^{1}H data was zero-filled to 2048 points prior to the Fourier transform. Only zero-order phase correction was performed, and a linear baseline subtracted from points in the region of a peak due to the 3 $H^{\delta 2}$ of Leu133 at – 0.23 ppm and the $H^{\beta 2}$ of Asn 47 at – 0.27 ppm. A Felix macro was used to determine the area of this peak by summing the intensity of each data point contributing to the peak. This area was used to normalize the intensity of the crosspeaks in the corresponding HSQC spectrum for spectrometer shift, protein degradation and solvent evaporation. Generally the correction adjusted the intensities of the HX time points by a factor between 0.96-1.2, although the range was extended as low as 0.5 and as high as 1.8. Appendix E presents the multiplication factor for all HSQC spectra.

3.1.4 NMR 2D Data Processing and Analysis

Approximately 30 HSQC spectra were taken for each of the 10 samples with concentrations of GuDCl ranging from 0 to 1.8 M at pD=6.0, and for the 5 samples at different pD values and [GuDCl]=1.6 M, to allow for the full time range of exchange. This amounted to an approximate total of 450 HSQC spectra. The HSQC data was processed using Felix95 software. The spectra were zero filled to 2048 points in the proton dimension and 1024 in the nitrogen dimension. Soft gaussian apodization (line broadening = -8 and Gaussian parameter = 0.1) in the ^{1}H dimension and sinebell squared apodization (window size = 96 points and phase shift = 80°) in the ^{15}N dimension were applied. Each spectrum was individually phase corrected in the ^{1}H dimension. 90 ° zero-order and 180° first-order phase corrections were performed in the ^{15}N dimension. A zeroth order polynomial for the ^{1}H dimension was also used to correct baselines in each spectrum.

3.1.5 Spectral Assignment

NMR spectral assignments were determined previously for CBD_{N2} at 35 °C, pH=6 (Brun, 2000). To extend these assignments to the conditions utilized for this study, HSQC spectra of CBD_{N2} were recorded as a function of temperature, GuDCl and pH.

i) Temperature Dependent Spectral Changes

HSQC spectra of CBD_{N2} were recorded at 35, 32, 30, 27, and 25 °C in order to obtain ^{1}H and ^{15}N chemical shift assignments for the protein at 25 °C. In general, some minor chemical shift variations are expected due to the temperature dependence of the D_2O lock. The average absolute value chemical shift in the ^{1}H dimension was 0.065 ± 0.02 ppm, and 0.094 ± 0.006 ppm in the ^{15}N dimension.

ii) GuHCl Dependent Spectral Changes

HSQC spectra of fully protonated CBD_{N2} were acquired with GuHCl at 0, 0.3, 0.6, 0.9, 1.2 and 1.5 M. CBD_{N2} is folded under these conditions. Amide residues exposed to the solvent tended to show the largest chemical shift changes. Interestingly, above average changes in chemical shifts were observed for the amides of many solvent exposed leucines, as well as the residues near or hydrogen bonded to these leucines. For example, the relatively large chemical shift change resulting from the GuHCl titration occurs for L102 in β-strand A4 (see figure 3.1a) and its hydrogen-bonding partner, M78, a member of β-strand A3. It is generally thought that GuHCl acts as a protein denaturant by reducing the energetically unfavorable exposure of hydrophobic residues to the polar solvent. Association of GuHCl with the amino acid leucine appears to be particularly favorable, consequently producing relatively large chemical shift changes to those groups near in space to solvent exposed leucines. The average absolute value chemical shift in the ^{1}H dimension was 0.080 ± 0.04 ppm, and 0.219 ± 0.019 ppm in the ^{15}N dimension.

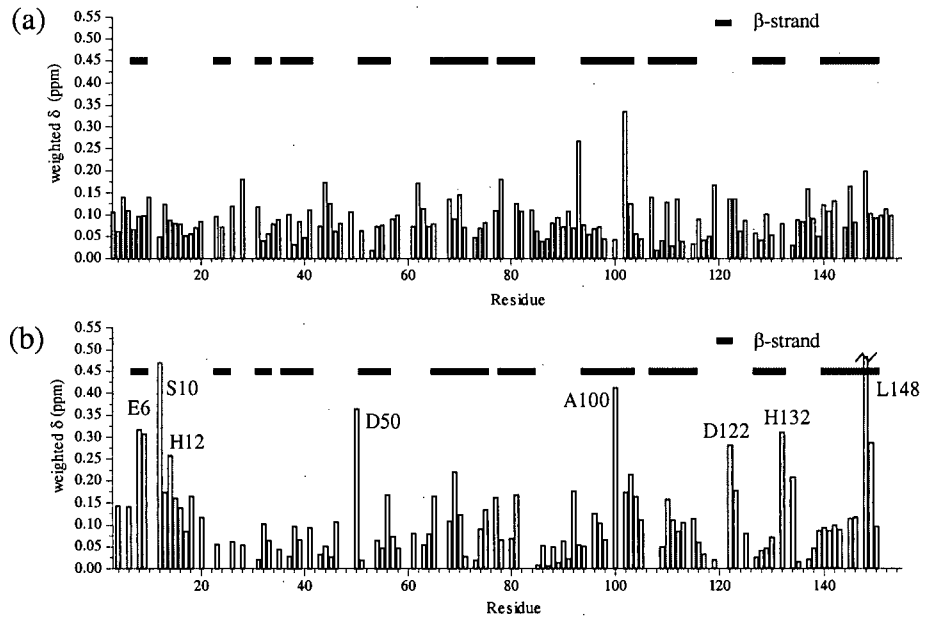


Figure 3.1: *Histogram of the weighted chemical shift difference* $\Delta\delta_w$[1] *between (a) 0-1.5M GuHCl and (b) pH 4-8 in the HSQC spectra of* CBD_{N2}. *Gaps are proline residues or residues with overlapping frequencies.*

[1] $\Delta\delta_w = \sqrt{(\Delta\delta_{^1H})^2 + (\frac{1}{6}\Delta\delta_{^{15}N})^2}$

iii) pH Titration

Thirteen HSQC spectra were acquired between pH 4 and 8, using HCl to initially reduce the pH from a value of 6, followed by the addition of NaOH to increase the pH. In general, amide groups of residues with or near solvent exposed titratible side chains (Asp, Glu, His) had chemical shifts showing the largest pH-dependence. The average absolute value chemical shift in the ^{1}H dimension was 0.091 ± 0.017 ppm, and 0.46 ± 0.06 ppm in the ^{15}N dimension (Fig. 3.1b).

3.1.6 Native State HX Protocol

Native state HX of CBD_{N2} was measured at pD=6.0 in the presence of 0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, and 1.8 M GuDCl. As well, exchange profiles of CBD_{N2} at 1.6 M GuDCl were measured at pD = 5.0, 5.5, 6.0, 6.5, and 7.0.

Generally 10 mg of ^{15}N labeled CBD_{N2} (Appendix B) in 500 µL of 50 mM potassium phosphate buffer (pH 6), 0.02 % sodium azide and 50 mM sodium chloride was lyophilized prior to being dissolved into and equivalent volume of D_2O and a specific GuDCl concentration.

GuDCl was obtained from GuHCl through 8 D_2O dilution cycles wherein 1 g of GuHCl was dissolved in 20 ml of D_2O followed by lyophilization. The difference between the index of refraction of a reference solution,[2] n_{ref}, and a solution containing GuDCl, n, was measured at room temperature on an Spectronic Instruments refractometer (model number 334610), and used to determine [GuDCl] within ~ 0.05M from (3.1) (Pace *et al*, 1992):

$$[\text{GuHCl}] = 57.147\,(n - n_{ref}) + 38.68\,(n - n_{ref})^2 - 91.6\,(n - n_{ref})^3 \qquad (3.1)$$

After resuspension of the lyophilized CBD_{N2}, the pD was quickly adjusted to 6.0 using DCl or NaOD and the sample briefly centrifuged to remove any undissolved protein.

---

[2] A reference sample was always taken prior to the addition of GuHCl in order to account for the extraneous refraction signal by H_2O, D_2O or buffer.

The solution was then transferred to an NMR tube and placed in the spectrometer with the probe temperature set to 25 °C. ^{1}H tuning and shimming was performed, followed by determination of the 90° pulse length for ^{1}H. The 90° pulse increased from 9 to 14 ms between [GuDCl]=0 and [GuDCl]=1.8 M. Generally the time between dilution into D_2O and the start of the first HSQC spectra was 20 minutes. Zero time for the exchange was set to be the point of protein rehydration in D_2O. The time point for exchange was taken to be the midpoint of the acquisition period of a given HSQC spectrum.

The final pD of all NMR samples, measured when exchange was complete or after 9 months, did not deviate by more than 0.1 unit from the initial value. The 90° pulse length was also re-measured for all NMR samples after exchange was complete. Deviations were restricted to 0.2 μs for all samples except for [GuDCl] = 0.4 M, where the deviation was 0.7 μs. The refractive index of HX samples, measured when exchange was complete, was slightly higher than at the time of sample preparation, due to the presence of buffer and protein. Accounting for these additional contributions, $[GuDCl]_{final} \sim [GuDCl]_{initial}$ as determined by refractometry.

3.1.7 Analysis of the HX Time Series

Peak intensities were measured using the titration module of Felix95. The intensity of a given peak was measured at the center of an encompassing box defined by a width of ~ 0.04 ppm in the ^{1}H dimension and ~ 0.03 ppm in the ^{15}N dimension. The shape of the peaks did not change or move significantly during the time course of the HX measurements. This scenario allows determination of the extent of exchange from the center intensity in the box that encompasses a peak. Decay rates obtained from fits to maximum intensities within a box were similar, within the error of the fit, to rates obtained from center intensities. The maximum determination however, had a consistently higher associated fitting error due to the effects of low signal-to-noise associated with amides nearly completely exchanged.

Before fitting was performed, the peak intensities of all amide resonances in a HSQC spectrum recorded at a given time point, were normalized to the most resolved upfield peak in the ^{1}H NMR spectrum of CBD_{N2}. The peak intensity of the first four time points, which corresponded to HSQC spectra taken with 4 instead of 16 transients, were also multiplied by a factor of 4. A single exponential was fit to the peak intensities of each resolvable amide groups in CBD_{N2}.

$$I(t) = I_o e^{-k_{ex}t} + I_\infty \tag{3.2}$$

The constant baseline value I_∞ represents the residual protons present in the sample, preventing complete NH → ND exchange. Nonlinear,least squared fits were performed in Origin. These fits were done individual residues at a time at each [GuDCl] to ensure quality of the data. Exchange rates with fitting errors much more than 10% of the fitted value were discarded.

3.2 Native State HX NMR Results

3.2.1 General Hydrogen Exchange Results

A total of 71 amide groups in CBD_{N2} had measurable hydrogen exchange at more than 2 different GuDCl concentrations. This is approximately half of the amide groups in the 153 residue protein, a typical fraction in native state hydrogen exchange experiments. A complete data set is precluded due to chemical shift overlap or exchange within the dead time of the experiment (~1.5 hours, since at least 5 time points are required for accurate fitting). Residues that exchanged faster than the dead time of the experiment were located in either the flexible loop regions, the N and C termini, or on the outer strands of the β-sheets with solvent-exposed amide groups. Exchange at [GuDCl] = 0 M occurred over 6 orders of magnitude with protection factors that ranged from ~10^2 to ~10^8.

The residues for which HX could be determined at 2 or more [GuDCl] are listed in Table 3.2. All elements of secondary structure had several probes, with the exception of β-strand B0. In the latter case, only 1 of the 3 amino acids that comprise this strand showed measurable protection. As [GuDCl] increased, individual amide HX rates likewise increased. This is illustrated in Figure 3.2 for Ala74, which is part of β-strand B3.

3.2.2 HX trends in β-strands

A common trend of fast exchange, or limited protection, exists among alternating residues in β strands at the edge of sheets A and B (Figure 3.3). Amide groups exposed to the solvent were not detectable after transfer to D_2O, while the intervening groups, intramolecularly H-bonded to the neighbor strand, exhibited significant protection from exchange.

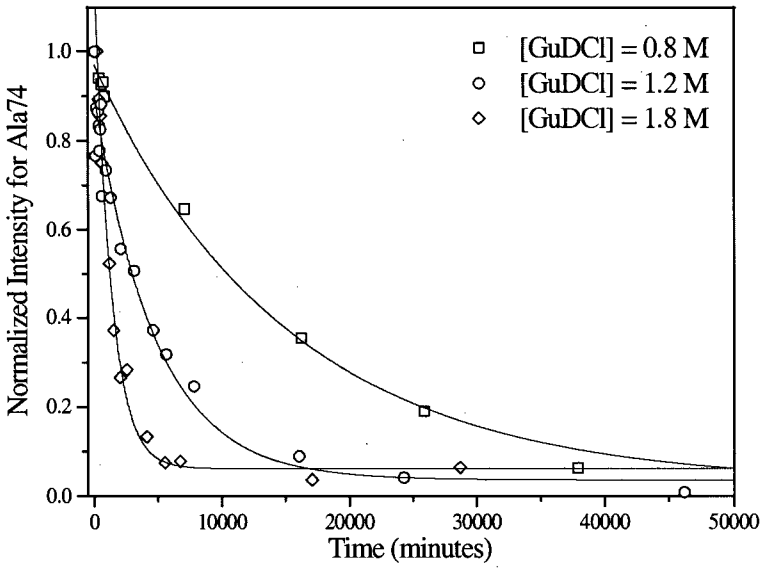


***Figure 3.2**: The exchange profile of Ala74 as a function of [GuDCl]*

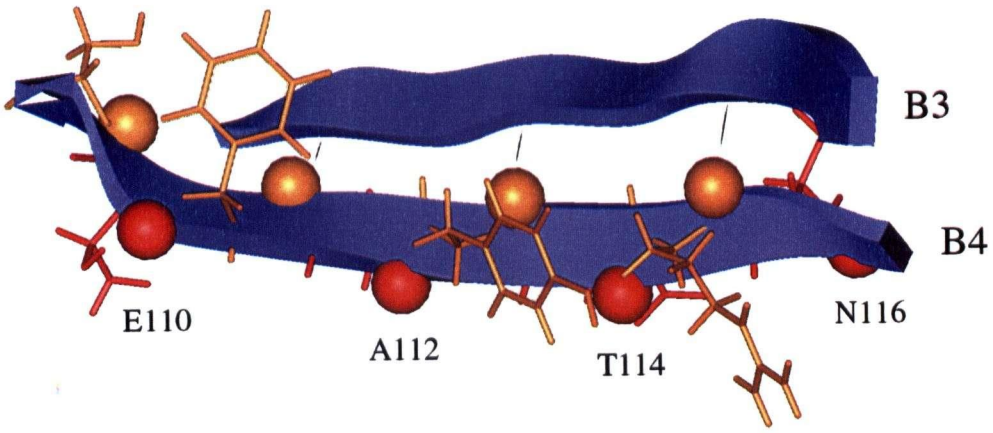


Figure 3.3 *β-strands B3 and B4. B4 is an edge strand of β-sheet B that displays measurable exchange for only alternating residues. β-strand B4 amide groups with measurable exchange are indicated by orange spheres, while the solvent exposed residues which exhibit fast exchange are indicated by red spheres. For reference, the orange side chains point in the direction of the hydrophobic core, while the red side chains point towards the protein exterior. This figure was created using the Biosym software, Insight 2000.*

This feature is a subtle indication of a several issues governing the analysis of HX in β-sheet proteins. Unlike α helices, where H-bonding is a sequence-local interaction between the i^{th} and i^{th}+4 residues, β sheets have non-local H-bonds. As well, backbone conformation in an α-helix is constrained such that most residues involved are H-bonded. Thus similar protection is expected for residues in the same α helix, since any structural fluctuations uniformly disrupt all the α-helical H-bonds. In a β-strand, the additional conformational flexibility does not force all amides to orient such that they form H-bonds. As well, only alternating residues form H-bonds with neighboring strands. Thus the fast exchange of a solvent exposed amide group does not necessarily indicate low stability in that β-strand. Indeed β-strand B4 shows an alternating pattern of fast exchange for solvent exposed residues, while interior facing residues are among the most protected in CBD_{N2}.

The previously reported secondary structure of CBD_{N2} was based on patterns of secondary chemical shifts, $^3J_{NH-H\alpha}$ scalar couplings, NOE-derived distance restraints, and hydrogen bonds identified through a qualitative HX analysis. The exchange patterns scrutinized in this study for the most part agree well with the previously reported secondary structure of CBD_{N2}. However, a few modifications to the secondary structure boundaries are suggested. For example, Tyr139 is among the most protected from HX, suggesting that it is the first residue in the β-strand B5, instead of the reported Glu140.

Several residues located at the boundary of β-strands exchanged in a GuDCl-insensitive manner, namely Leu9 from B0, Ser65 from B3 and Thr150 from B5. In contrast the interior residues of β–strands B3 and B5 contain some of the most protected residues in CBD_{N2} (Figure 3.8).

3.2.3 Highly Protected Residues

A residue was deemed a "slow exchanger" when there was no detectable change in protection level after 9 months in [GuDCl]=0.6 M. All slow exchangers were members of β-strands (including Tyr139 in B5). The sole exception is Gly85, a deeply buried residue in the A3-A4 loop.

Table 3.1: A comparison of sequence conservation in CBD_{N2}.

	Percent Conservation	
	4 most homologous	8 most homologous
All residues	52	33
Residues with HX information	56	36
Residues that are unexchanged after 9 months in [GuDCl]=1.6 M	80	55

Percent Conservation was calculated in the following manner: If the residue was 100%, 75%, or 50% conserved among the 4 (or 8) homologous sequences, it was assigned a conservation value of 100,75, or 50. Residues that were not conserved at all, or less that 50 % conserved were assigned a value of 0. Percent conservation was then calculated as the sum of the conservation value for a given set of residues divided the theoretical sum if all residues in the set were 100% conserved, multiplied by 100.

$$\%Conservation = \frac{Sum\_of\_conservation\_values}{Theoretical\_sum\_for\_100\%\_conservation} \cdot 100$$

In some proteins, residues that exhibit a high level of protection are also those that are highly conserved (Alexandrecu *et al*, 1999). This is also true for CBD_{N2}. 20 out of the 23 most protected residues in CBD_{N2} are conserved to some level among the 4 most homologous sequences (Figure 1.10), while 18 out of the 23 highly protected residues are conserved among the 8 most homologous sequences. Similar conservation was observed for the entire protein and among those residues that were probes in the native state HX experiment (Table 3.1). However significantly more percent conservation exists among those residues that are highly protected to exchange.

β-strands A2, B2, A3, B3, B4, A5, and B5, which contained these highly protected residues, were all members of the jellyroll structural component of CBD_{N2}. Boundary strands to the β sheet B0, A1, B1 and A4 had no residues that remained unexchanged after 9 months in 1.6 M GuDCl, suggesting that these residues exchange via local fluctuation unraveling at the N-terminus (for B0) and fraying of strands at the edges of the β-sheets (for A1, B1, and A4). B4 the analogous edge strand to A4 is a member of the most slowly exchanging strands by virtue of a single "slow exchanger", Tyr111, which forms a H-bond with Leu68 another "slow exchanger." The C-terminal strand, B5 is both the longest strand and among the most protected to exchange. Figure 3.4 shows the strands containing the most protected amides colored in blue and the remainder colored in orange.

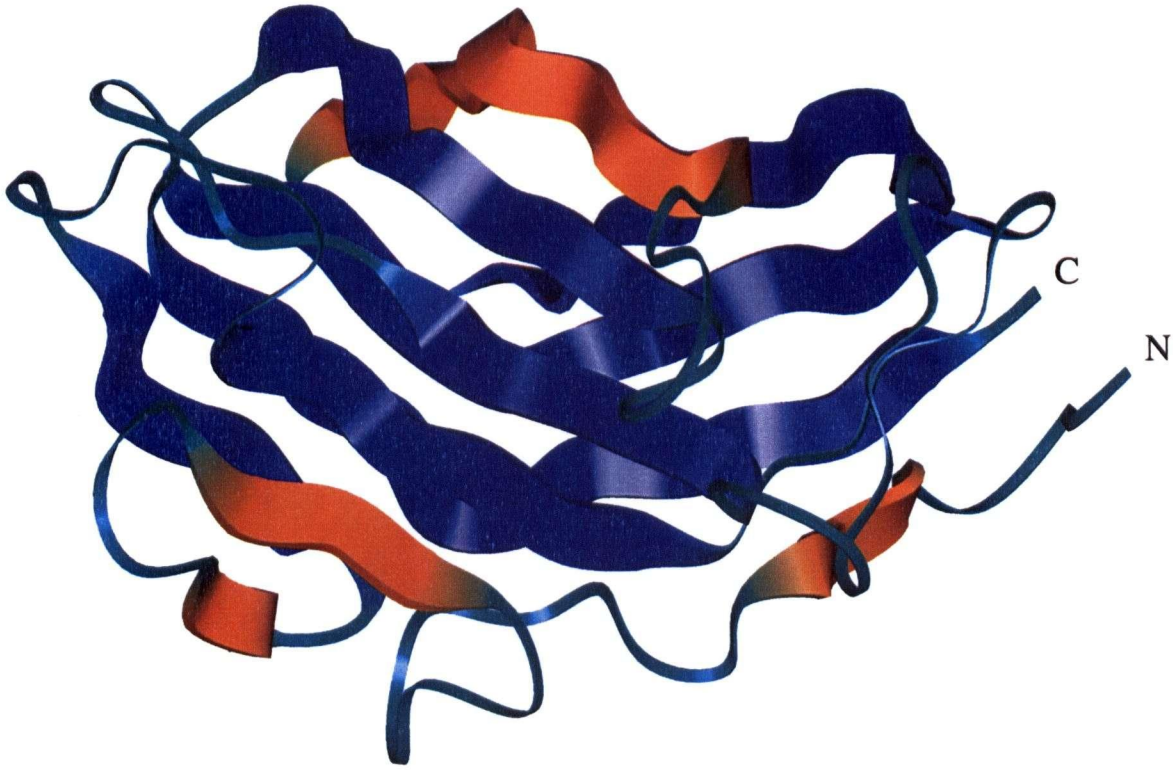


***Figure 3.4**: Top view of CBD_{N2}. β-strands containing residues that remain unexchanged after 9 months in 1.6 M GuDCl are colored blue, while less protected strands are colored orange. This figure was created using the Biosym software, Insight 2000.*

3.2.4 Determination of $\Delta G_{HX}(D_2O)$ and m_{HX}.

$$\Delta G_{HX} = -RT \ln(\frac{k_{ex}}{k_{\text{int}}}) \tag{3.3}$$

Free energy of HX, ΔG_{HX}, at a given [GuDCl], was determined according to (3.3), where the observed protein exchange rate k_{ex} was obtained for CBD_{N2} by NMR spectroscopy as described in section 3.1, and the intrinsic rates of a polypeptide chain, k_{int}, was calculated according to the method of Bai *et al* as discussed in detail in Appendix D (1993).

For almost all the residues measured, ΔG_{HX} had a simple linear dependence on [GuDCl], be it large or small, which could be fit to (3.4).

$$\Delta G_{HX} = \Delta G_{HX}(D_2O) + m_{HX}[GuDCl] \tag{3.4}$$

Figure 3.5(a) shows the exchange profiles of residues that have a mono-phasic dependence on [GuDCl]. Only 9 amide groups (Val31, Val58, Met78, Arg81, Phe119, His132, Glu140, Gln145, Leu148) in CBD_{N2} displayed bi-phasic exchange behavior: that is, exchange with little or no denaturant dependence at low [GuDCl], and yet large denaturant dependence at higher [GuDCl]. The transition occurred near 1.2 M GuDCl (Figure 3.5(b,c)). For those residues that displayed biphasic exchange the data were fit to a modified version of 3.4 (Chamberlin *et al*, 1996):

$$\Delta G_{HX} = -RT \ln(K_{subglobal} + K_{global}) \tag{3.5}$$

where the large denaturant dependent phase is described by:

$$K_{global} = \exp(\frac{\Delta G_{global}(D_2O) + m_{global}[GuDCl]}{-RT}) \tag{3.6}$$

and where $K_{subglobal}$ is described by:

$$K_{subglobal} = \exp(\frac{\Delta G_{sub}(D_2O) + m_{sub}[GuDCl]}{-RT}) \tag{3.7}$$

When there is no denaturant dependence in the subglobal phase, m_{sub} is set to zero in the fitting algorithm.

The entire set of fitted values for $\Delta G_{HX}(D_2O)$ and m_{HX} are listed in Table 3.2.

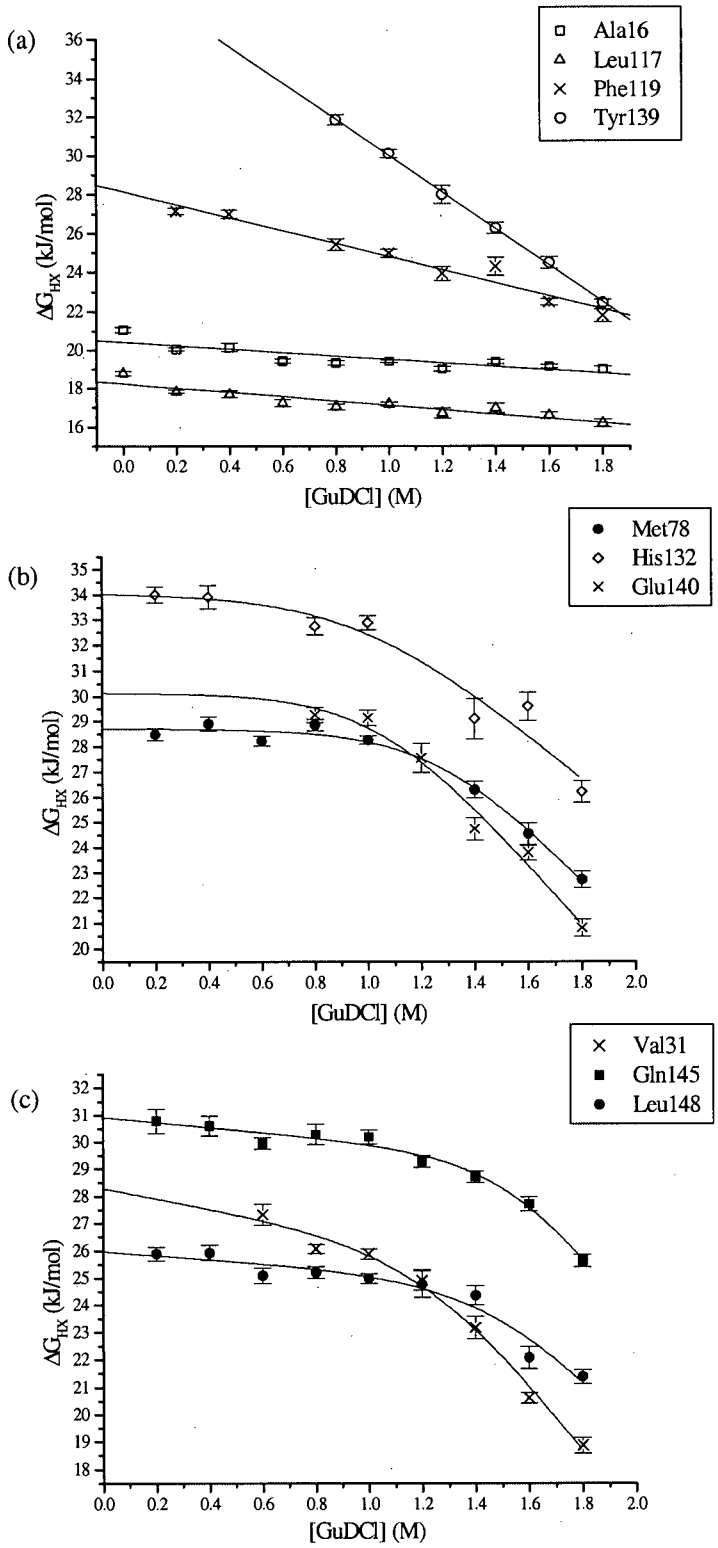


Figure 3.5: *The free energy of HX, ΔG_{HX}, as a function of [GuDCl] for CBD_{N2} residues that exchange (a) with a simple linear denaturant dependence, (b) with a denaturant independence at low GuDCl and a denaturant dependence at high GuDCl, and (c) with small denaturant dependence at low GuDCl and a larger denaturant dependence at high GuDCl.*

Table 3.2: Native State $\Delta G_{HX}(D_2O)$ and m values as fit to (3.4-3.7). Those values with no associated fitting error were obtained from a linear fit to two points.

AA	$\Delta G_{HX}(D_2O)$ (kJ/mol)	±	m (kJ/mol·M)	±	2°
L9	16.5	0.2	-1.3	0.3	B0
L10	21.8	0.3	-1.7	0.3	
F15	20.2	0.1	-1.5	0.1	
A16	20.3	0.2	-0.9	0.2	
S23	44.8	1.0	-11.1	0.8	A1
L24	18.2		-21.5		A1
S28	17.1		-0.7		
V31	42.5	6.8	-13.1	3.6	B1
	28.3	1.9	-1.8	3.9	
A33	19.3	0.5	-1.2	0.4	B1
M37	42.3	0.9	-9.5	0.6	B2
C38	45.1	0.3	-10.0	0.2	B2
V39	42.0	0.4	-10.9	0.3	B2
D40	42.3	0.2	-11.8	0.2	B2
G43	15.3	0.4	-1.2	0.3	
W49	14.8	0.3	-3.5	0.7	
A51	24.5	0.2	-2.8	0.1	A2
L53	40.0	0.8	-10.4	0.5	A2
V54	38.5	3.6	-10.7	2.3	A2
Y55	40.3	1.0	-10.4	0.8	A2
G57	28.0	0.2	-1.9	0.2	
V58	28.7	0.3	-3.3	0.3	
G61	24.7	0.3	-2.6	0.3	
E64	21.6	0.2	-2.0	0.2	
S65	15.6	0.9	-1.4	2.3	B3
L68	39.9	1.1	-11.3	0.7	B3
S69	42.6	0.8	-10.6	0.6	B3
F70	39.9	0.5	-8.9	0.4	B3
T71	40.4	0.4	-9.6	0.3	B3
S73	42.1	0.3	-9.4	0.2	B3
A74	36.3	0.5	-6.1	0.5	B3
T75	39.2	0.3	-8.7	0.2	B3
M78	43.9	2.1	-11.7	1.3	
	28.7	0.1	N/A		A3
R81	46.7	19	-12.5	10	A3
	27.2	0.4	-1.7	0.8	
V82	38.1	0.4	-9.5	0.3	A3
V84	36.3	0.9	-9.7	0.6	A3
G85	43.3	1.1	-10.9	0.8	
G87	24.4	0.6	-2.4	0.7	
G88	17.1	0.8	-0.4	0.7	
Y91	14.8	0.3	-2.8	0.4	
R92	21.9	0.2	-2.6	0.3	
A94	34.5	0.2	-5.8	0.2	A4
F95	35.0	0.3	-7.0	0.2	A4
Q97	20.4	0.3	-1.3	0.3	A4
A100	47.6	1.6	-12.3	1.2	A4
L102	22.2	0.2	-3.7	0.3	A4
T103	20.6	0.1	-3.1	0.2	A4
A107	37.6		-9.5		B4
R109	26.5	0.2	-2.2	0.2	B4
Y111	40.9	0.8	-11.1	0.6	B4
F113	40.0	0.8	-10.1	0.6	B4
S115	32.0	0.3	-3.1	0.3	B4
L117	18.3	0.2	-1.2	0.2	
F119	30.0	1.7	-4.4	1.0	
	28.7	1.7	N/A		
G127	32.8	1.6	-5.7	1.2	A5
Q128	42.8	1.3	-10.2	1.0	A5
V129	41.8	0.9	-11.9	0.7	A5
A130	42.8	0.4	-11.3	0.3	A5
H132	43.4	3.8	-9.2	2.4	A5
	34.0	0.7	N/A		
G134	21.6	0.1	-3.9	0.2	
K135	18.7	0.4	-2.9	0.4	
Y139	39.5	0.1	-9.4	0.1	
E140	43.1	3.2	-12.3	1.9	B5
	30.1	1.0	N/A		
F141	41.1	0.5	-10.8	0.4	B5
C142	46.9	0.7	-10.9	0.5	B5
S144	22.5	0.2	-0.4	0.2	B5
Q145	34.1	0.4	-4.0	0.3	B5
V146	33.7	0.4	-7.3	0.3	B5
L148	42.4	9.2	-11.4	5.1	B5
	26.0	0.5	-0. 7	1.0	
T149	41.0	0.6	-11.4	0.5	B5
T150	23.6	0.6	-1.8	0.5	B5

Figure 3.6 shows the distribution of $\Delta G_{HX}(D_2O)$ and m_{HX}-values for CBD_{N2}. There are two distinct exchange regimes: those amide groups that exchange via "local" fluctuations which are characterized by low m values (that is, small denaturant dependence) and a range of small $\Delta G_{HX}(D_2O)$, and those amide groups that exchange due to global unfolding. The latter exhibit large m_{HX} and $\Delta G_{HX}(D_2O)$ values which have a strong correlation. For global unfolding exchange, this correlation with $\Delta G_{HX}(D_2O)$ indicates that when a large change in solvent exposure (large m) occurs, there is a corresponding large free energy cost for the structural reorganization that allows exchange. Furthermore, the clustering among those residues that exchange globally in both sheets is indicative of a 2 state folding mechanism for CBD_{N2}.

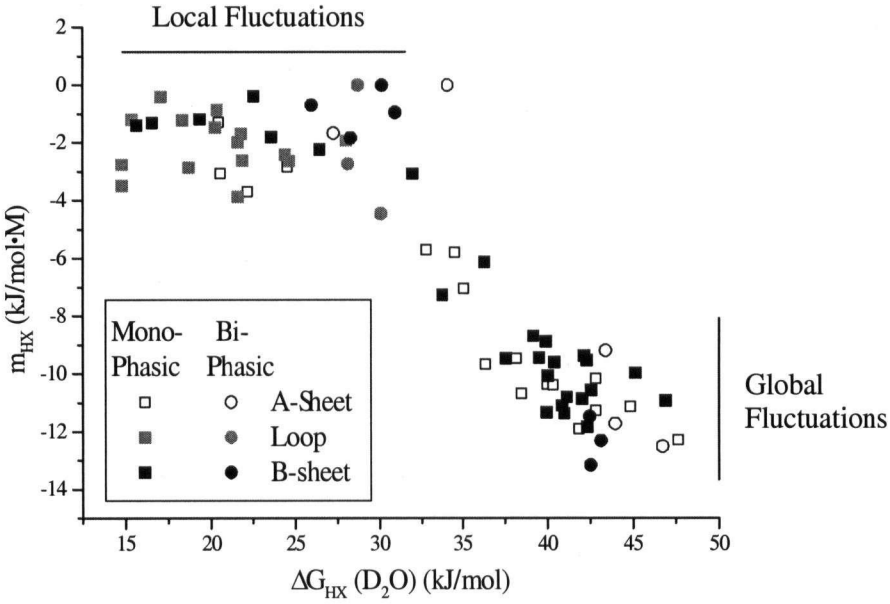


***Figure 3.6**: Distribution of $\Delta G_{HX}(D_2O)$ and m values for CBD_{N2}.*

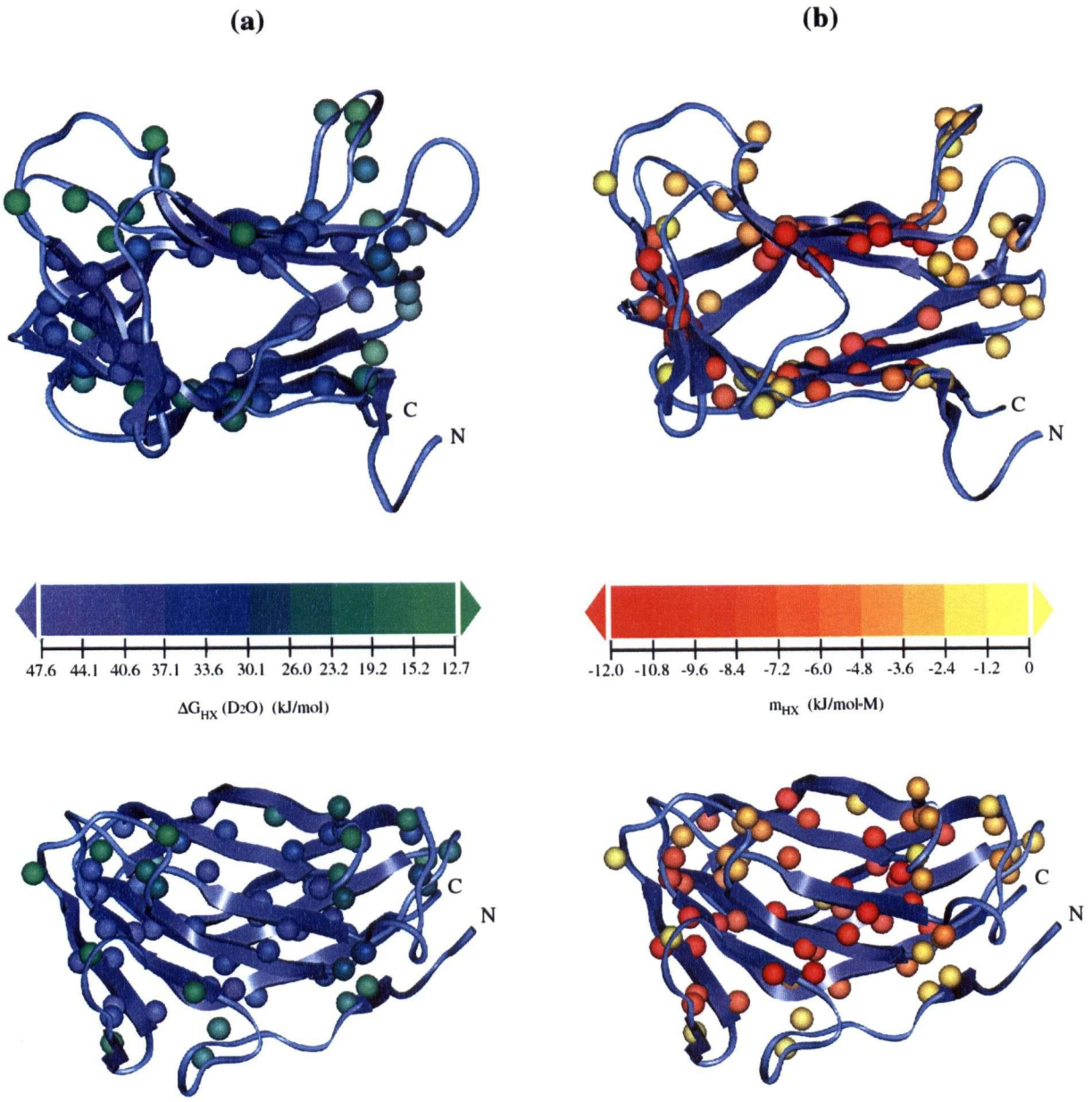


***Figure 3.7**: Side and top view of the (a) residue stability, $\Delta G_{HX}(D_2O)$, of CBD_{N2} to HX and (b) their relative m_{HX} values indicated by a color gradient. In the side view, the A-sheet is above the B-sheet. The A-sheet is in the foreground of the top view. This figure was created using Biosym – Insight 2000 software.*

3.2.5 Low Denaturant Dependent Exchange

Residues in CBD_{N2} displaying only low denaturant dependent exchange (small m_{HX} and small $\Delta G_{HX}(D_2O)$), even at the highest concentration of GuDCl, are listed in Table 3.3. Many of these residues have at least partially solvent-exposed side chains, and most are present in coil regions of CBD_{N2} (Figure 3.7 (a) green, (b) yellow) The average m_{HX} value for these residues is – 2.2± 0.2 kJ/mol·M and the average $\Delta G_{HX}(D_2O)$ is 20.2±0.7 kJ/mol.

Low denaturant dependent exchange is thought to take place during low-energy conformation fluctuations ("breathing"), which is common in native proteins. Dubbed "local fluctuation" exchange, it is traditionally expected to result from only small surface area reorganization of the protein. However, recent simulation work by Wooll *et al* suggest that an amide hydrogen which exchanges in a denaturant-independent manner does not necessarily do so as a sole result of small structural fluctuations but could exchange in this manner as a result of significant surface area exposure of the protein (~10%) (2000).

The amount of amino acid surface area exposed to the solvent was evaluated for those residues in CBD_{N2} that exchanged with a small denaturant dependence. Several of these residues were less solvent accessible than the average for highly protected residues. Furthermore, the percentage of solvent exposure was less than the percentage of amino acid surface area occupied by an amide hydrogen. Thus the exchange of Leu10, Phe15, Leu 102, and Gly134 likely requires significant structural fluctuations.

Table 3.3: Amide hydrogen and amino acid solvent exposure for residues that exchange in a denaturant independent way. The solvent exposure values listed here are averaged over all the CBD_{N2} NMR-based ensemble of structures. Highlighted in bold are those residues that exchange with low denaturant dependence, but which have less solvent accessible area exposed than a highly protected residue. Solvent exposure was measured using MOLMOL (Koradi, 2000).

AA	Accessible Surface Area (Å^2)	Total Surface Area(Å^2)	%Surface Area of ^{N}H relative to entire AA	% Solvent Exposure		2° structure
	AA			Residue	Particular AA average	
Leu9	76	303.7	7.5	24.9	10.8	B0
Leu10	**11**	**293.3**	**7.5**	**3.8**	**10.8**	
Phe15	**1**	**339.6**	**6.6**	**0.2**	**6.9**	
Ala16	26	**231.3**	9.6	11.2	16.5	
Leu24	34	295.6	7.5	11.6	10.8	A1
Ala33	56	229.8	9.6	24.3	16.5	B1
Ser65	39	240.4	9.3	16.1	17.9	-B3
Gly87	34	200.6	11.0	16.9	17.5	
Tyr91	154	354.2	6.3	43.6	18.6	
Arg92	170	**358.0**	6.5	47.6	29.3	
Gln97	54	306.4	7.2	17.5	17.4	A4
Leu102	**9**	**306.6**	7.5	**3.0**	**10.8**	**A4**
Leu117	74	304.8	**7.5**	24.2	10.8	
Asp122	115	266.0	8.2	43.1	36.9	
Gly134	**2**	**201.4**	**11.0**	**1.1**	**17.5**	
Lys135	67	336.5	6.5	20.0	20.0	
Ser144	18	240.1	9.3	7.4	17.9	B5
Thr150	21	264.9	8.5	7.7	18.0	-B5

Low denaturant dependence average	14.6
All CBD_{N2} AA average	19.0
Highly protected residues average	7.5

'-' indicates β-strand boundary

3.2.6 Large Denaturant Dependent Exchange.

Amide hydrogens that exchange with large denaturant-dependence are referred to as "global exchangers" since the fluctuations that lead to exchange require global unfolding of the protein, or $\Delta G_{HX}(D_2O) \sim \Delta G_U$.

A structural representation of CBD_{N2} (Figure 3.7) with the magnitude of $\Delta G_{HX}(D_2O)$ and m_{HX} displayed in a color gradient serves to illustrate that the most stable residues exist in the β-sheets. The average $\Delta G_{HX}(D_2O)$ for these residues is 41.9±0.4 kJ/mol while the average m_{HX} is –10.7±0.2 kJ/mol·M. While several trends exist within the global folding data, overall the HX behavior CBD_{N2} is consistent with a 2-state model in which global fluctuations involve the entire protein, rather than sub-domains, such as sheets A or B.

Figure 3.8 shows the values $\Delta G_{HX}(D_2O)$ and m mapped along the sequence of CBD_{N2}. In regions of β-strand secondary structure, $\Delta G_{HX}(D_2O)$ and m_{HX} have similar large values. This excludes the possibility that the individual β sheets have different stabilities (either within sheets A or B, or in comparison between sheets). This is not unexpected as it is unlikely that a jellyroll β-sandwich protein would have sub-domains with distinct stability given the non-local interactions that make up its secondary and tertiary structure. Solvent exposure of residues within a strand requires large-scale fluctuations involving the entire sheet. This likely requires disruption of the hydrophobic core of the protein and hence fluctuations in the opposing sheet as well.

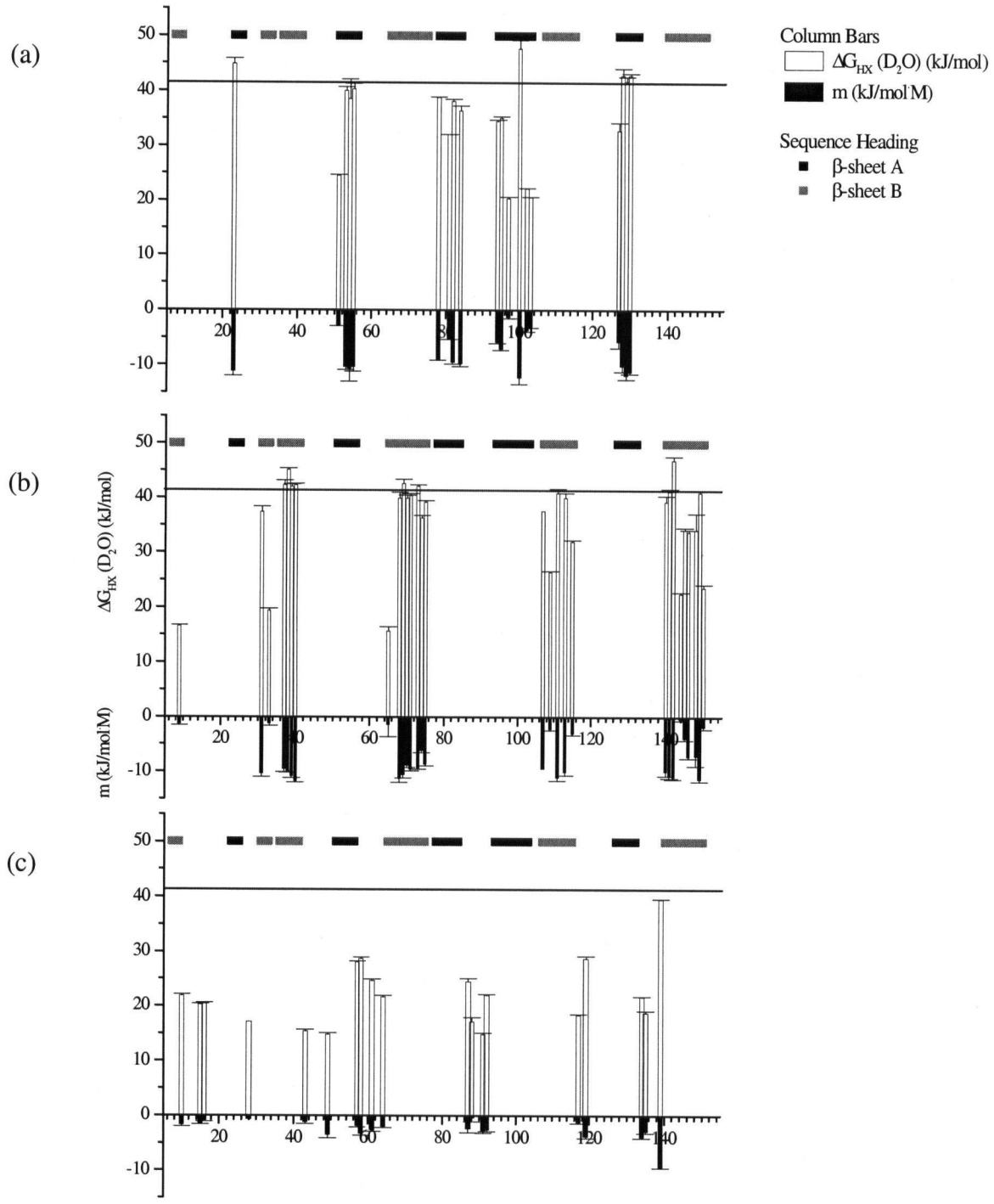


***Figure 3.8**: $\Delta G_{HX}(D_2O)$ and m for (a) β-Sheet A and (b) β-sheet B and (c) random coil regions mapped on the amino acid sequence.*

3.2.7 Comparison of $\Delta G_{HX}(D_2O)$ and ΔG_U

GuDCl and temperature unfolding experiments, as monitored by CD spectroscopy, gave a free energy of unfolding in D_2O for CBD_{N2} of 36.7±4.3 kJ/mol. Given the extrapolation involved, this value agrees well with the average ΔG_{HX} (D_2O) of 41.9±0.4 kJ/mol calculated from those residues which exchange globally. Upon closer inspection, there is a small overestimation of stability determined by HX methods. This originates because the structural fluctuations that give rise to HX occur on a time scale much faster than *cis-trans* proline isomerization. Thus in HX experiments, one measures the energy input required to not only to unfold the protein, but to unfold the protein to a higher energy *cis*-X-Pro denatured conformation. In contrast, during equilibrium unfolding experiments, denatured samples can undergo isomerization from a higher energy *cis*-X-Pro to a lower energy *trans*-X-Pro conformation. Thus in HX experiments of proteins with native *cis*-prolines, there is an overestimation of the energy of unfolding by the additional energy of isomerization. The energy adjustment is determined from the equilibrium constant (K_{PRO}) of isomerization of a given Xaa-Pro residue. K_{PRO} varies from 1.65 when Xaa is a Trp to 15.67 when Xaa is another Pro. CBD_{N2} has 13 prolines, of which one (Pro76) is in a *cis* conformation in the native state. The calculated adjustment is 9.4 kJ/mol, which results in a predicted unfolding ΔG_U ~ 46 kJ/mol in the absence of isomerization. Although this is slightly higher than the average value of 41.9 kJ/mol for HX global fluctuations, the correction is qualitatively consistent, and in agreement with a $\Delta G_{HX}(D_2O)$ of 47 kJ/mol for the most slowly exchanging residue, Cys142.

3.2.8 Stability in β-hairpins

β-hairpins are a structural motif consisting of two anti-parallel β-strands, typically joined by a loop of approximately 2-5 residues. The only motif among β sheets that are analogous to an α-helix, β-hairpins are secondary structure elements formed by sequence-local hydrogen bonds.

In CBD_{N2} there are 2 β-hairpins, while the rest of the β-sheets are comprised of non-sequence-local β-strand interaction. In both cases, the β-hairpins occur at the edge of the β-sheets of CBD_{N2}. The first β-hairpin is between B1 and B2 (Table 3.4), while the second is between A3 and A4 (Table 3.5).

Table 3.4: $\Delta G_{HX}(D_2O)$ and m_{HX} for the β-hairpin between B1 and B2.

AA	2° structure	$\Delta G_{HX}(D_2O)$ (kJ/mol)	±	m_{HX} (kJ/mol·M)	±
V31	B1	28.3	1.9	-1.9	2.9
		42.5	6.5	-13.1	3.6
A33	B1	19.3	0.5	-1.2	0.4
M37	B2	42.3	0.9	-9.5	0.6
C38	B2	45.1	0.3	-10.0	0.2
V39	B2	42.0	0.4	-10.9	0.3
D40	B2	42.3	0.2	-11.8	0.2

$\Delta G_{HX}(D_2O)$ of residues Met37, Val39, and Glu40 in B2 and the global unfolding phase $\Delta G_{HX}(D_2O)$ of Val31 in B1 seem to agree well, with and average $\Delta G_{HX}(D_2O)$ of 42.3 kJ/mol. Ala33 of B1 is 24% solvent exposed (higher than the CBD_{N2} average of 17%; Table 3.2), and so is not expected to show cooperative protection. Cys38 of B2, which forms a disulphide bond integral to the structure of CBD_{N2}, has a larger $\Delta G_{HX}(D_2O)$ than the average for this hairpin.

There is no consistent stability trend in the A3-A4 hairpin (Table 3.5). There is some agreement in the $\Delta G_{HX}(D_2O)$ and m_{HX} near the junction of the strands, but not before Val82 in A3 and not beyond F95 in A4.

Table 3.5: $\Delta G_{HX}(D_2O)$ and m for the β-hairpin between A3 and A4

AA	2° structure	$\Delta G_{HX}(D_2O)$ (kJ/mol)	±	m_{HX} (kJ/mol·M)	±
M78	A3	28.7	0.1	N/A	
		44.0	2.2	-11.7	1.3
N81	A3	27.2	0.4	-1.7	0.8
		46.6	19.1	-12.5	10.2
V82	A3	38.1	0.4	-9.5	0.3
V84	A3	36.3	0.9	-9. 7	0.6
A94	A4	34.5	0.2	-5.8	0.2
F95	A4	35.0	0.3	-7.0	0.2
Q97	A4	20.4	0.3	-1.3	0.3
A100	A4	47.6	1.6	-12.3	1.2
L102	A4	22.2	0.2	-3.7	0.3
T103	A4	20.6	0.1	-3.1	0.2

3.2.9 Hydrogen Bond Breakage

An amide group must be accessible to a catalyst for HX to occur. Since the encounter complex leading to exchange requires formation of a H-bond with a solvent, structural fluctuations that break internal H-bonds are required. Exchange profiles of residues that form cross-β-strand H-bonds often overlap, since structural fluctuations that break internal H-bonds are also those that promote exchange. Figure 3.9 highlights two such examples.

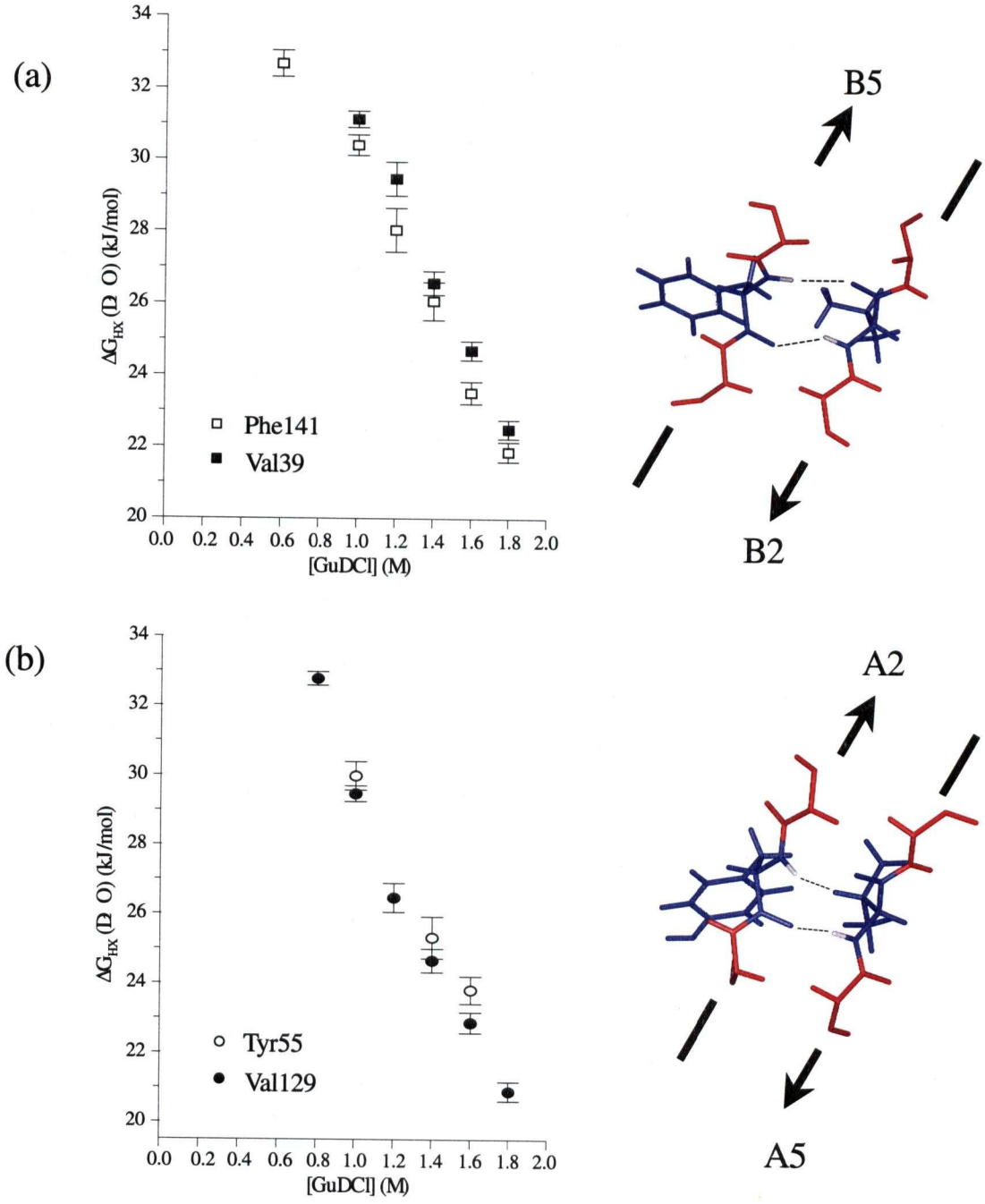


***Figure 3.9**: ΔG_{HX} profiles of residues that form H-bonds to each other.*

3.2.10 The Disulphide Bond

CBD_{N2} has a single disulphide bond between Cys38 and Cys142 that serves to cross-link the β-strands B2 and B5, both of which contain many residues exhibiting slow exchange via global fluctuations. Reduction of this disulphide bond leads to the unfolding of CBD_{N2} under otherwise benign conditions.

The amide groups of these two cysteine residues have exchange profiles that overlap (Figure 3.10). Their matching $\Delta G_{HX}(D_2O)$ and m_{HX} values are higher than any other residues in strand B2 and B5, and with the exception of Ala100, the rest of the protein. In addition, the stability of these residues is larger than the global stability of CBD_{N2}, measured by CD spectroscopy and corrected for proline isomerization. "Super" stability in the region of a disulphide bond has been reported previously, most recently by Laszlo et al (1999) in their HX study of human prion protein. This phenomenon is generally interpreted as an indicator of residual structure in the unfolded state. It is curious that the amide groups of these cysteines have the same stability pattern given that they are not involved in a mutual hydrogen-bond, but rather have amide groups that are oriented in opposite directions. Indeed, Cys38 forms a H-bond with Val31 of B1 while Cys142 forms a H-bond with Ser73 of B3; the profiles of Ser73 and Val31 do not overlap with the corresponding cysteines. Since the B2-B5 disulphide bond is close to the connection to the A2–A5 strands, any residual structure would likely contain hydrophobic contacts between the A and B sheet in this region.

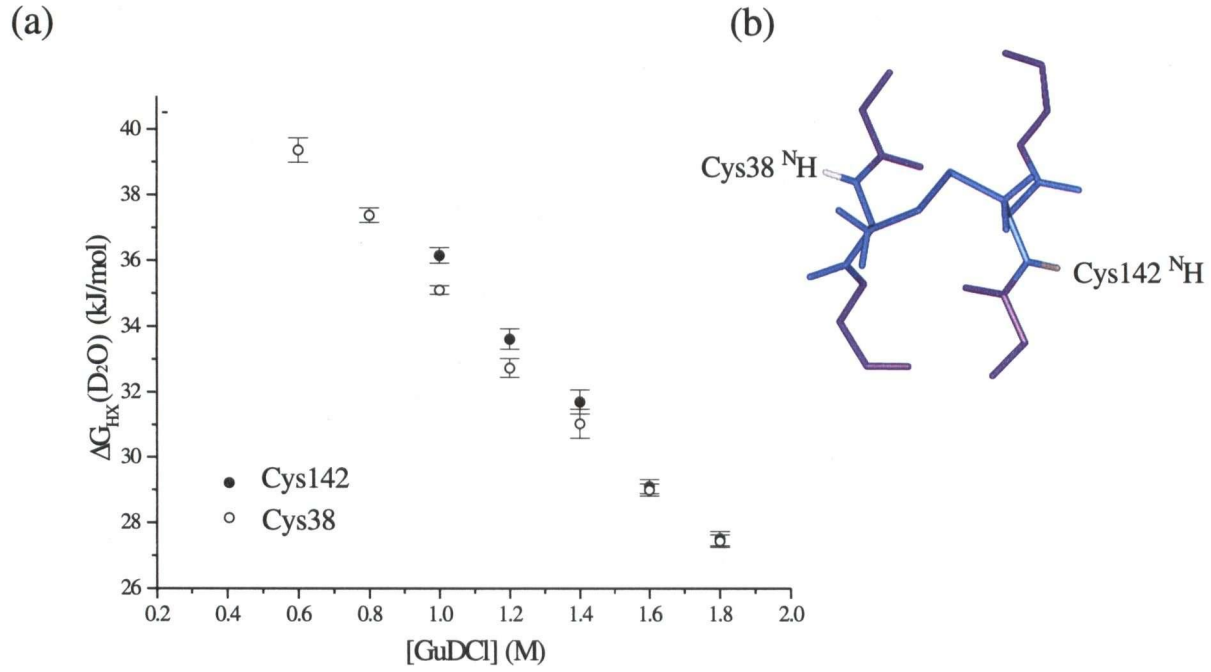


***Figure 3.10**: (a) Stability profile and (b) structure of Cys38 and Cys142, the residues comprising the cysteine bond in CBD_{N2}.*

3.2.11 Cellooligosaccharide Binding Cleft

Previous studies based on the measurement of chemical shift pertubation upon addition of cellooligosaccharide and intermolecular protein-sugar NOE interactions have identified the cleft across β-sheet A as the site of CBD_{N2} sugar binding (Figure 1.9) (Brun, 2000). Since structure and function are so intimately related in proteins, there was interest in exploring whether these residues displayed above average HX stability. Since backbone fluctuations that result in HX would also disrupt the orientation of side chains involved in sugar binding, do those amino acids integral to sugar binding display high protection? All residues except Trp49 and Tyr91 displayed protection of at least 32 kJ/mol, though usually closer to 40 kJ/mol. Both Trp49 and Tyr 91 are in loop regions flanking the binding cleft, and therefore would be unlikely to exhibit such a high degree of protection.

3.2.12 EX2 and pD Dependent HX

Extraction of residue specific stability from exchange profiles using equation 3.3 requires that exchange occur in the so-called EX2 regime. This necessitates that the rate of refolding (k_{cl}) be much greater that the intrinsic rate of exchange (k_{int}) such that $k_{ex}=K_{op}k_{int}$. Except under conditions of moderately high pH, an EX2 mechanism is generally expected. All experiments occurred at pH = 5.6 or pD = 6.0.

One method to establish that hydrogen exchange is occurring in the EX2 regime is to measure the overall amide exchange rate (k_{ex}) as a function of pD. Since the expression for k_{ex} in the EX2 regime has a linear dependence on k_{int}, this rate is expected to increase ~10 fold with every unit pH increment. In contrast, if exchange is occurring via the EX1 regime, then k_{ex} depends only on the rate of opening k_{op} and will display pH independence.

Since refolding rate is slowest at high [GuDCl], experiments as a function of pD (5.0, 5.5, 6.0, 6.5, and 7.0) were performed at 1.6 M GuDCl. If the rate of peptide exchange k_{int} was still

significantly smaller than the refolding rate at this GuDCl concentration, then EX2 exchange could be extrapolated to all smaller [GuDCl]. Surprisingly, instead of finding a linear increase in $\log(k_{ex})$ with pD (indicative of EX2) or a pD independence in the exchange rates (indicative of EX1), there was a decrease in k_{ex} for most residues in CBD_{N2} (Figure 3.11).

The only positive slope occurred for residues in loop regions, while all the residues involved in a β-strand displayed an unambiguous decrease in HX rate as a function of pD. It is striking that the slope of $\log(k_{ex})$ versus pD approaches –1 which is the inverse of the expected behavior (a slope of +1). Although surprising, this result can be explained by denaturation experiments (Chapter 2) that showed an increase of stability for CBD_{N2} in the presence of GuDCl as a function of pD. Therefore under these conditions, any increase in k_{int} due to an elevation in pD is offset by a larger increase in the stability of CBD_{N2}, culminating in a reduction in k_{ex}. Since GuDCl has such a perturbing effect, a similar experiment as a function of pD at 0 M GuDCl would ascertain EX2 exchange. Unfortunately, exchange occurs extremely slow (with time constants > 9 months for most residues) under non-denaturing conditions, making such an experiment extremely time consuming, and likely futile. Despite the unusual HX results, it is clear that EX1 is not the exchange regime. As well, stopped flow refolding experiments discussed in chapter 2 reveal refolding rates that are at least an order of magnitude larger than the internal peptide exchange rate (listed in Appendix D), the criteria for EX2 exchange.

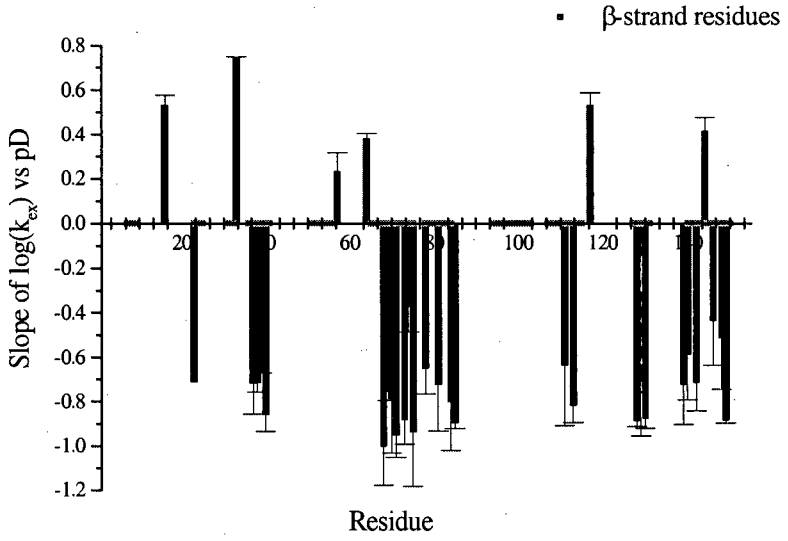


***Figure 3.11**: The linear fit of log(k_{ex}) vs pD at 25 °C mapped onto the CBD_{N2} sequence.*

3.3 Conclusions

Native state hydrogen exchange experiments in non-denaturing concentrations of GuDCl were performed on CBD_{N2}. Most residues exchanged with a mono-phasic dependence on denaturant, allowing simple extrapolation to $\Delta G_{HX}(D_2O)$. There appeared to be no difference in exchange patterns between sheet-A and sheet-B, suggesting a two-state folding mechanism, or a free-energy landscape that has no significant crevasses. The most slowly exchanging residues were confined to the β-strands contributing to the hydrophobic core of the protein.

Stabilities extracted from exchange rates showed linear dependence on [GuDCl]. The most protected residues had $\Delta G_{HX}(D_2O)$ that agreed with global free-energy of unfolding measured by CD denaturation experiments when accounting for the perturbation from proline isomerizations. The most protected residues had $\Delta G_{HX}(D_2O) \sim 42$ kJ/mol. Among these were the cysteines involved in the disulphide bond required for stabilizing the native state of CBD_{N2}. Residues that were hydrogen bonded together showed similar exchange profiles, highlighting that H-bond breaking fluctuations must occur before the encounter complex can be formed, and base catalysis can proceed. Some of the residues that exchanged with low denaturant dependence were significantly buried, and exchange likely occurred though fluctuations requiring significant structural reorganization. No structural domains of dissimilar stability were observed. These results support a two state folding model for CBD_{N2}.

Chapter 4
Concluding Remarks

The results from this work support a 2-state folding mechanism for CBD_{N2}. Global unfolding data described in detail in chapter 2, from both CD spectroscopy and DSC fit well to 2-state models. Proline isomerization, the probable basis for muliexponential refolding phases observed in stopped flow experiments, does not challenge the notion of 2-state folding. Instead it suggests the existence of small crevases in the conformational landscape, consistent with the "rough" landscape shown in Figure1.1.

It is, however, the native state HX experiments as a function of low concentrations of denaturant that provide the conclusive proof of two-state unfolding. Several proteins that exhibited 2-state unfolding as monitored by fluorescence and CD spectroscopy (Cytochrome C, Ribonuclease H) revealed several "partially unfolded forms" (PUFs) when probed by native state HX. Identification of these PUF's relied on relating similar extrapolated free-energies of unfolding ($\Delta G_{HX}(D_2O)$) to residues in similar structural elements. Often a PUF consisted of a single or group of α-helices or an entire β-sheet. At the beginning of this project, no all-β-sheet protein had revealed PUF's. It was unclear whether this was a consequence of some intrinsic characteristic of β-sheet proteins, or whether the proteins studied were too small or too unstable to provide enough probes for detection of PUF's.

CBD_{N2} is exempt from both these latter considerations since it is very stable, and has at least 2 HX probes, although generally 4-5, for each of the 11 β-strands that comprise this protein. Results from these native state HX experiments revealed predominantly mono-phasic exchange that occurred solely through global unfolding fluctuations or via small-scale low energy fluctuations. No difference in exchange patterns existed between sheet-A and sheet-B.

It has recently been suggested that PUF's exist regardless of whether they can be detected (Rumbley, 2001). The results of this thesis, and a survey of native state HX literature suggest an alternative view. Namely that PUF's will only be observed in regions of sequence-local secondary structure. That is, independent fluctuations of structural sub-domains occur only if the backbone is not constrained by non-local hydrogen bonds. α-helices, which contain sequence-local H-bonds, are often found to be PUFs.

Most β-sheet structures (with the exception of β-hairpins, and β-sheets composed of sequential β-hairpins) have non-local H-bonds. α-β-proteins exhibit PUFs which include entire β-sheets, or β-strands that are sequence-local to α-helices (Rnase H). In CBD_{N2}, the β-hairpin between B1 and B2 was the only structural element to display extrapolated stability within an extremely narrow range of values (although it was still in the global unfolding regime).

While this "sequence-local" hypothesis does not negate the significance of PUF's, it does suggest that all proteins do not necessarily fold in a series of sub-native structural domains. Other folding motivators, like hydrophobic collapse may initially play a more important role than the formation of secondary structure H-bonds.

Indeed, one might ask whether HX protection results from hydrophobic residue sequestering or maintenance of H-bonds?

Evidence in this thesis supports the significance of H-bonds in HX protection. Several CBD_{N2} residues involved in mutual H-bonds have completely overlapping HX profiles, with global $\Delta G_{HX}(D_2O)$. While the overlapping HX profiles suggest that H-bonds must be broken before HX can occur, it is the global $\Delta G_{HX}(D_2O)$ which indicates global unfolding fluctuations leading to exchange result from the H-bond breakage. If a single H-bond breakage was required for HX then measured rates would result in $\Delta G_{HX}(D_2O)$ values that were on the order of H-bond formation (~2-3kJ/mol). Often several H-bonds must be broken before a fluctuation can lead to

exchange. Is the measured ΔG_{HX} the sum of broken H-bonds, or are there further factors significant in preventing fluctuations leading to exchange?

An obvious additional contribution to HX protection could be the sequestering of hydrophobic residues, since fluctuations that expose the amide group to solvent catalysts, might also expose an energetically unfavorable hydrophobic side chain. In the case of a β-sandwich, residues along a strand have side chains whose orientation alternate between the hydrophobic interior and the solvent exposed exterior. If hydrophobic sequestering were indeed a significant factor in HX protection, one would expect to see an alternating pattern of stability along the sequence of a beta-strand. Fluctuations that disrupted only every second residue would require significant backbone strain, an improbable situation. Indeed Figure 3.8 shows no evidence of alternating patterns of stability along strands that form the β-sandwich, but also does not negate the importance of hydrophobic packing in stabilizing the tertiary contact between the 2 β-sheets of CBD_{N2}.

One possible way to establish the significance of hydrophobic interactions independent of backbone strain would be to do a native state HX experiment on a sequence-local β-sandwich - composed of 2 β-sheets that were both formed by consecutive β-hairpins, as opposed to CBD_{N2}, whose β-sheets are not sequence-local due to its jelly-roll topology. If hydrophobic interactions were not significant in HX protection, one expects the separate sheets would represent individual PUFs. If hydrophobic interactions were significant, both sheets would exhibit similar stability.

Native state HX data of stability mutants and stability homologues has yet to be potently exploited. Comparison of mesophilic and thermophilic homologues show a uniform increase in stability for all residues in the thermophilic protein. A point mutation that decreased the stability of CI2 showed effects only on the stability of residues that exchange through "global" fluctuations, while mutation that increased the stability of Rnase T1 resulted in an increase in

extrapolated $\Delta G_{HX}(D_2O)$ for all residues. With knowledge of residue specific stability of CBD_{N2}, it becomes simple to introduce expedient mutations.

One can envision mutations aimed to stabilize residues that exchange through local fluctuations, such that they exchange through global fluctuations. How would this affect the stability of other residues in the protein? How would this perturb the conformational landscape? Would this alter the folding pathways or kinetics of this protein? Would it be useful in industrial applications? Not all these questions can be answered by native state exchange experiments. However, a better understanding of how mutations affect the type of data obtained from native state HX could lead to better interpretations of this data in the context of a proteins conformational landscape. It is, after all, the conformational landscape that provides insight to protein folding pathways, and the topology of this landscape that will act as the reference for strategies aimed at the manipulation of these protein folding pathways.

Appendix A
Derivation of the general hydrogen exchange behavior of native proteins

Native state HX is often described by the following equation:

$$closed \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} open \xrightarrow{k_{int}} exchanged \qquad (A.1)$$

where k_{op} is the rate constant associated with the structural opening which can lead to exchange while k_{cl} is the refolding rate to the native state. k_{int} is the intrinsic amide exchange rate of an unstructured peptide.

The following chemical equations are obtained from the reaction described by (A.1).

$$\frac{d[closed]}{dt} = k_{cl}[open] - k_{op}[closed] \qquad (A.2)$$

$$\frac{d[open]}{dt} = k_{op}[closed] - (k_{cl} + k_{int})[open] \qquad (A.3)$$

$$\frac{d[exchanged]}{dt} = k_{int}[open] \qquad (A.4)$$

The method of eigenvalues can be used to solve a system of first order differential equations where there are n equations and n dependent variables. A matrix is set up where each row consists of one first order differential equation (ie (A.2), (A.3), or (A.4)), while the columns are filled by the coefficients of each variable. The eigenvalue, λ, is subtracted from the diagonal elements of this matrix. If the determinant of the matrix is set to zero, then n solutions of λ are obtained. In this case λ represents the overall exchange rate for a native protein.

$$\begin{array}{c} \\ \dfrac{d[closed]}{dt} \\ \dfrac{d[open]}{dt} \\ \dfrac{d[exchanged]}{dt} \end{array} \left| \begin{array}{ccc} [closed] & [open] & [exchanged] \\ -k_{op}-\lambda & k_{cl} & 0 \\ k_{op} & -(k_{cl}+k_{\text{int}})-\lambda & 0 \\ 0 & k_{\text{int}} & -\lambda \end{array} \right| = 0 \qquad \text{(A.5)}$$

The determinant of (A.5) is:

$$(k_{op}+\lambda)(k_{cl}+k_{\text{int}}+\lambda)(-\lambda)-(-\lambda)k_{op}k_{cl}=0 \qquad \text{(A.6)}$$

After some algebra, (A.6) reduces to:

$$-\lambda(\lambda^2-\lambda(k_{cl}+k_{\text{int}}+k_{op})+k_{op}k_{\text{int}})=0 \qquad \text{(A.7)}$$

The solutions for λ are:

$$\lambda=0,\lambda=\frac{(k_{cl}+k_{op}+k_{\text{int}})\pm\sqrt{(k_{cl}+k_{op}+k_{\text{int}})^2-4k_{\text{int}}k_{op}}}{2} \qquad \text{(A.8)}$$

The first solution is trivial and can be ignored. It is useful to simplify the second 2 equations of (A.8). The first step is to factor out $(k_{cl}+k_{op}+k_{\text{int}})^2$ from the root to obtain:

$$\lambda=\frac{(k_{cl}+k_{op}+k_{\text{int}})\pm(k_{cl}+k_{op}+k_{\text{int}})\sqrt{1-\dfrac{4k_{\text{int}}k_{op}}{(k_{cl}+k_{op}+k_{\text{int}})^2}}}{2} \qquad \text{(A.9)}$$

For a native protein, the closing rate is much faster than the opening rate, such that $k_{op} <<< k_{cl}$. This means that $\dfrac{4k_{\text{int}}k_{op}}{(k_{cl}+k_{op}+k_{\text{int}})^2}$ is much smaller than 1.

The binomial approximation states that $(1\pm\varepsilon)^k \sim 1\pm k\varepsilon$ if $1 >>> \varepsilon$. Thus the binomial approximation can be used to simplify the root in A.9.

$$\lambda = \frac{(k_{cl} + k_{op} + k_{\text{int}}) \pm (k_{cl} + k_{op} + k_{\text{int}})(1 - \frac{4k_{\text{int}}k_{op}}{2(k_{cl} + k_{op} + k_{\text{int}})^2})}{2}$$

$$\lambda = \frac{(k_{cl} + k_{op} + k_{\text{int}}) \pm ((k_{cl} + k_{op} + k_{\text{int}}) - \frac{2k_{\text{int}}k_{op}}{(k_{cl} + k_{op} + k_{\text{int}})})}{2} \quad \text{(A.10)}$$

A simple equation for HX in native proteins can therefore be obtained from the negative root of (A.10):

$$\lambda = \frac{k_{\text{int}}k_{op}}{(k_{cl} + k_{op} + k_{\text{int}})} \quad \text{(A.11)}$$

Appendix B
Protein Production and Purification

JM101 electro-competent cells were electroporated with a pTUG plasmid containing the gene encoding *Cellulomonas fimi* wild type CBD_{N2} gene, as well as the gene encoding for kanamycin resistance. Frozen stocks of electroporated cells were made from individual colonies on kanamycin agar plates and used within a 3 month period.

B.1 Unlabelled Protein Production

i) Expressing the CBDs in E. coli JM101

Overnight cultures with 50 μg/ml of kanamycin were directly inoculated with frozen stock cells and incubated at 30 °C.

Sterile TYP media (16 g of bacto-tryptone, 16 g yeast extract, 5 g NaCL, 2.5 g K_2HPO_4 (dibasic) per 1 litre of dH_2O) was inoculated with overnight cultures of the JM101 cells. After approximately 2.5 hours, when the media OD_{600} was greater than 0.5, protein expression was induced with the addition of 1 mM IPTG. The media was left in the shaker at 30 °C for 36 to 48 hours for maximal protein yield. No significant increase in protein yield resulted from the use of a fermentor as compared to simple shaker Feinback flasks.

ii) Protein Purification

CBD_{N2} is produced in JM101 cells as a secreted protein. The cells were centrifuged at 5000 rpm for 20 minutes and the supernatant retained. Attempts to extract protein in th periplasmic by osmotic shock treatment did not prove appreciably fruitful. 1M NaCl, 30 g of Avicel per 1litre, and 3-5 mg of EDTA was added to the supernatant and the slurry stirred at 4 °C for at least 4 hours for complete Avicel binding. The suspension was then vacuum filtered through glass fiber filter and allowed to air dry. 30 g of Avicel per 1litre was added to the flow-

through and stirred for at least 8 hours to allow binding of any remaining protein, followed by vacuum filtering.

The dried Avicel cakes were washed with with 500 mL 50 mM potassium phosphate buffer (pH 7), 1M NaCl. CBD_{N2} was then eluted with 500 mL of distilled H_2O (4° C). The conductivity of this solution was measured and diluted with distilled H_2O until the reading was less than 20 mS. This solution was filtered through 4 layers of glass fiber, and then through a 0.45 μm non-cellulose membrane to remove any residual Avicel. The supernatant was loaded on a Q-Sepharose column, equilibrated with 10 mM KPhos buffer at pH=6.0, and then eluted via a salt gradient using a 10 mM KPhos and 1 M NaCl buffer at pH=6.0. CBD_{N2} generally eluted from the column at about 55% NaCl. Purity was checked using SDS-PAGE and yields ranged between 60-120 mg/liter.

Protein concentration was measured from the absorption at 280 nm using an extinction coefficient of 20504 1/cm · M (Brun *et al*, 2000).

B.2 ^{15}N Labeled Protein Production

Labeled protein was produced in the same way as unlabeled with the following modifications. Overnight cultures in TYP media was centrifuged for 10 minutes at 6000 rpm and the supernatant disposed. Cells were re-suspended and then further inoculated in M9 minimal media (with 1g per litre of ^{15}N labeled Celltone). Typically the time taken to reach OD_{600}=0.5 was 3 hours.

M9 media recipe consists of 1 litre of dH_2O and 6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl and 1 g $^{15}[NH_4]Cl$. After autoclaving, the following aliquots of filter-sterilized solutions were added: 100μl 1M $CaCl_2$, 2 ml 1M $MgSO_4$, 1 ml 0.01M $FeCl_3$, 1 ml 1 mg/ml Thiamine, 20 ml 40% Glucose, and 1ml of 50 mg/ml kanamycin (Brun *et al*, 2000).

Appendix C
Derivation of ΔG_U as a function of ΔH_m, T_m, and ΔC_p

We want to derive (C.1) from the van't Hoff equation.

$$\Delta G(T) = \Delta H_m(1-\frac{T}{T_m}) - \Delta C_p(T_m - T - T\ln(\frac{T}{T_m})) \tag{C.1}$$

The van't Hoff equation is written as follows:

$$\frac{d}{d(\frac{1}{T})}(\ln K) = -\frac{\Delta H}{R} \tag{C.2}$$

where T is Temperature, K is the equilibrium constant for unfolding, and ΔH is the change in enthalpy during unfolding, and R is the gas constant.

We begin by rewriting $d(\frac{1}{T})$ as $-\frac{1}{T^2}dT$ and substituting this into (C.2) to obtain:

$$Rd(\ln K) = \frac{\Delta H}{T^2}dT \tag{C.3}$$

We want to integrate both sides from T to T_m, but must account for the fact that ΔH= ΔH(T).

Making use of the integration rule:

$$\int_a^b u dv = uv\Big|_a^b - \int_a^b v du$$

and setting $dv = \frac{dT}{T^2}$ and $u = \Delta H$ we can integrate (C.3) as follows:

$$R\int_T^{T_m} d(\ln K) = \frac{\Delta H}{T}\Big|_T^{T_m} - \int_T^{T_m}\frac{d\Delta H}{T} \tag{C.4}$$

By multiplying the 2nd term on the right of (C.4) by $\frac{dT}{dT}$ and by making use of the fact that $\Delta C_p = \frac{d\Delta H}{dT}$, (C.4) can be written as:

$$R\int_{T}^{T_m} d(\ln K) = \frac{\Delta H}{T}\Big|_{T}^{T_m} - \Delta C_p \int_{T}^{T_m} \frac{dT}{T} \tag{C.5}$$

Integrating (C.5) one obtains:

$$\frac{\Delta G(T_m)}{T_m} - \frac{\Delta G(T)}{T} = \frac{\Delta H_m}{T_m} - \frac{\Delta H}{T} - \Delta C_p \ln(\frac{T_m}{T}) \tag{C.6}$$

Making use of the fact that $\Delta G(T_m)=0$, and multiplying both sides of (C.6) by T:

$$-\Delta G(T) = \frac{T\Delta H_m}{T_m} - \Delta H - T\Delta C_p \ln(\frac{T_m}{T}) \tag{C.7}$$

Add and subtract ΔH_m from the right side of (C.7) to obtain:

$$-\Delta G(T) = \frac{T\Delta H_m}{T_m} - \Delta H - T\Delta C_p \ln(\frac{T_m}{T}) + \Delta H_m - \Delta H_m \tag{C.8}$$

Making use of the fact that $\Delta H_m = T_m \Delta C_p$ and $\Delta H = T\Delta C_p$ we can write (C.8) as:

$$-\Delta G(T) = \frac{T\Delta H_m}{T_m} - T\Delta C_p - T\Delta C_p \ln(\frac{T_m}{T}) + T_m \Delta C_p - \Delta H_m \tag{C.9}$$

Finally, grouping like terms together one obtains:

$$\Delta G(T) = \Delta H_m (1 - \frac{T}{T_m}) - \Delta C_p (T_m - T - T\ln(\frac{T}{T_m})) \tag{C.10}$$

Appendix D
Intrinsic peptide exchange rate for CBD_{N2}

Table D.1: Intrinsic H→D Amide Exchange Rates for CBD_{N2} at 25°C and pD=6.0.

AA	k_{int} (1/s)	AA	k_{int} (1/s)	AA	k_{int} (1/s)	AA	k_{int} (1/s)
S2	1.63E+01	D40	6.07E-02	V80	1.92E-02	P120	
L3	8.77E-02	L41	2.91E-02	R81	1.46E-01	P121	
D4	5.17E-02	P42		V82	5.54E-02	D122	4.82E-02
S5	2.59E-01	G43	1.79E-01	L83	3.19E-02	G123	2.06E-01
E6	1.03E-01	G44	4.60E-01	V84	2.06E-02	D124	1.24E-01
V7	2.36E-02	Q45	2.84E-01	G85	2.26E-01	A125	1.10E-01
E8	3.74E-02	G46	4.93E-01	E86	7.64E-02	P126	
L9	3.11E-02	N47	7.64E-01	G87	2.20E-01	G127	1.79E-01
L10	2.71E-02	P48		G88	4.60E-01	Q128	2.84E-01
P11		W49	3.74E-02	G89	4.60E-01	V129	5.29E-02
H12	6.07E-01	D50	6.50E-02	A90	2.47E-01	A130	1.21E-01
T13	9.62E-01	A51	1.10E-01	Y91	8.98E-02	F131	9.62E-02
S14	6.21E-01	G52	3.11E-01	R92	2.26E-01	H132	1.21E+00
F15	1.92E-01	L53	6.50E-02	T93	2.36E-01	L133	2.97E-01
A16	1.92E-01	V54	2.06E-02	A94	2.65E-01	G134	1.92E-01
E17	5.17E-02	Y55	6.50E-02	F95	9.62E-02	K135	2.26E-01
S18	2.77E-01	N56	5.80E-01	E96	5.93E-02	A136	2.20E-01
L19	8.77E-02	G57	6.50E-01	Q97	1.36E-01	G137	3.11E-01
G20	1.92E-01	V58	4.93E-02	G98	4.93E-01	A138	2.47E-01
P21		P59		S99	5.80E-01	Y139	8.98E-02
W22	3.74E-02	V60	1.92E-02	A100	3.34E-01	E140	5.80E-02
S23	3.04E-01	G61	2.26E-01	P101		F141	6.81E-02
L24	8.77E-02	E62	7.64E-02	L102	2.53E-02	C142	6.81E-01
Y25	5.54E-02	G63	2.20E-01	T103	8.77E-02	I143	8.98E-02
G26	3.49E-01	E64	7.64E-02	G104	4.93E-01	S144	2.31E-01
T27	2.10E-01	S65	2.77E-01	E105	7.64E-02	Q145	3.83E-01
S28	6.21E-01	Y66	1.79E-01	P106		V146	5.29E-02
E29	1.03E-01	V67	3.74E-02	A107	9.62E-02	S147	2.84E-01
P30		L68	3.19E-02	T108	1.42E-01	L148	8.77E-02
V31	1.92E-02	S69	2.42E-01	R109	3.19E-01	T149	8.77E-02
F32	6.97E-02	F70	1.92E-01	E110	8.57E-02	T150	2.26E-01
A33	1.92E-01	T71	1.63E-01	Y111	6.36E-02	S151	6.21E-01
D34	8.38E-02	A72	2.65E-01	A112	1.88E-01	A152	5.29E-03
G35	2.06E-01	S73	3.92E-01	F113	9.62E-02		
R36	2.97E-01	A74	3.34E-01	T114	1.63E-01		
M37	2.71E-01	T75	1.42E-01	S115	6.21E-01		
C38	7.64E-01	P76		N116	1.03E+00		
V39	9.62E-02	D77	3.34E-01	L117	9.19E-02		
		M78	1.08E-01	T118	8.77E-02		
		P79		F119	1.52E-01		

D.1 Code for the C-program that calculates intrinsic peptide exchange rates

```
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <math.h>

#define R 1.98709 /* kcal/molK */
#define arNT -1.32
#define brNT 1.62
#define alCT 0.96
#define blCT -1.8
#define T 298
#define Tref 278

/* Activation Energies are in kcal/mol */
#define EaA 14
#define EaB 17
#define EaW 19
#define kA 1.19
#define kB 9.90
#define kW -2.5
#define pD 6.0
#define pOD (15.65-6.0)

/* This program calculates the intrinsic exchange rate of a polypeptide
according to the Protein Structure Function and Genetics paper
by Bai et al vol 17 pp75-86 1993

It requires 2 input files: one with the aa seq in 3 letter code with modification to
the 3 letter code as described below called aaseq_letters.txt and a file with the
modification constants determine by Bai et al, both in the directory of compilation.
One must also specify an output file for the krc constants.

to compile: cc -lm krc.c -o krc.exe
then to run it just type krc.exe at the prompt

Some comments: It is important to properly define your chosen kA,kB,kW as well
as pOD (check your dissociation constant) and pD above.
The aa sequence file must be 3 letter coded and all cis PROLINES must have the
modified 3 letter code PRC instead of PRO. Trans proline maintain the 3 letter
PRO. Non Disulfide bonded Cysteines must have the 3 letter code CYN, whereas disulfide
Cysteines maintain the CYS 3 letter code.
All titratable side chains are assumed to be in charged form, and if the pH you
interested in is below 4.4 or above 6.5 then the bai_aa.txt must have the uncharged
data replaced, according to the Bai paper. NOte that calculating krc for proteins
with non bonded cysteines at a pH above 8.5 is impossible since data for charged
non-disulphide-cysteines is unavailable.
```

Each amino acid should be on a seperate line in the sequence file.
The aa sequence file must exist in the directory of the compiled program, and file name must be modified prior to compiling in the following code.

```
*/
main()
{
int flag=1;
float ar,al,bl,br,brnext,arnext,krc;
char AAL[4],AAR[4],BAI[4];
FILE *cfPtr,*cfPtrBAI,*cfPtrOUT;

/* AA SEQUENCE FILE NAME */
if ((cfPtr = fopen("aaseq_letters.txt","r"))==NULL)
        printf("Sequence file can't be opened, baby. \n");
else
        {
        if ((cfPtrBAI = fopen("bai_aa.txt","r"))==NULL)
                printf("File bai_aa.txt can't be opened, baby. \n");
        else
                {
                if ((cfPtrOUT = fopen("krc.txt","w"))==NULL)
                        printf("Output file can't be opened, baby. \n");
                else
                fputs(" AA BAI krc \n",cfPtrOUT);
                {
                do
                {
                rewind(cfPtrBAI);
                fscanf(cfPtr,"%s\n",AAL);
                if (feof(cfPtr))
                        flag=-1;

                switch (flag) {

                case 1:
                        ar=arNT;
                        br=brNT;
                        flag +=1;
                        break;

                case -1:
                        al=alCT;
                        /*printf("hi %f          %f\n",pow(10,5),(10^5));*/
                        bl=blCT;
                        krc= (pow(10,(kA+al+ar-pD))*(exp((-EaA/R)*((1/T)-(1/Tref)))) +
pow(10,(kB+bl+br-pOD))*(exp((-EaB/R)*((1/T)-(1/Tref)))) + pow(10,(kW+bl+br))*(exp((-
EaW/R)*((1/T)-(1/Tref)))));
```

```
                    krc=krc/60;
                    printf(" AA is %s and BAI is %s and krc is %f\n",AAL,BAI,krc);
                    fprintf(cfPtrOUT,"%s %s %f\n",AAR,AAL,krc);
                    break;
            default:

                    fscanf(cfPtrBAI,"%s%f%f%f%f\n",BAI,&al,&arnext,&bl,&brnext);

                    while (!feof(cfPtrBAI) && (strcmp(AAL,BAI) !=0))

      fscanf(cfPtrBAI,"%s%f%f%f%f\n",BAI,&al,&arnext,&bl,&brnext);

                    if (strcmp(AAL,BAI) == 0)
                    {
                        if ((strcmp(AAL,"PRO") == 0) || (strcmp(AAL,"PRC") == 0))
                            {
                            printf("AAL is %s BAI is %s\n",AAL,BAI);
                            fprintf(cfPtrOUT,"%s %s \n",AAR,AAL);
                            }
                        else
                            {
                            krc= (pow(10,(kA+al+ar-pD))*(exp((-EaA/R)*((1/T)-
(1/Tref)))) + pow(10,(kB+bl+br-pOD))*(exp((-EaB/R)*((1/T)-(1/Tref)))) +
pow(10,(kW+bl+br))*(exp((-EaW/R)*((1/T)-(1/Tref)))));
                            krc=krc/60;
                            printf(" AA is %s and BAI is %s and krc is
%f\n",AAL,BAI,krc);
                            fprintf(cfPtrOUT,"%s %s %f\n",AAR,AAL,krc);
                            }
                    } /*end of strcmp is */
                    else
                        printf("Couldn't find the AA %s\n",AAL);
                    ar=arnext;
                    br=brnext;
                    strcpy(AAR,AAL);
                    flag+=1;
                    break;

            } /*end of switch */
            } while (!feof(cfPtr)); /*end of while !feof cfPtr*/
            } /* end of file OUT else */
      } /* end of file cfPtrBAI selection loop ELSE */

      fclose(cfPtr);
      fclose(cfPtrBAI);

} /* END OF ELSE */
return 0;
} /*END OF MAIN*/
```

Appendix E
Normalization Factor for the HX Time Series from the 3 $H^{\delta 2}$ resonance of Leu133 and the $H^{\beta 2}$ resonance of Asn47.

[GuDCl] (M)	Time Series Data Point	Peak Area	Norm Factor
0.0	1-19	1.71E+06	1.00
	20	1.67E+06	0.98
	21	1.70E+06	0.99
	22	1.74E+06	1.02
	23	1.74E+06	1.02
	24	1.85E+06	1.08
	25	1.85E+06	1.08
	26	1.80E+06	1.05
	27	1.65E+06	0.96
	28	1.78E+06	1.04
	29	1.81E+06	1.06
	30	1.69E+06	0.99
	mean	1.75E+06	1.02
	stdev	6.83E+04	0.04
	sterr	1.97E+04	0.01
0.2	1-9	1.30E+06	1.00
	10-18	1.32E+06	1.02
	19	1.23E+06	0.95
	20	1.41E+06	1.08
	21	1.36E+06	1.05
	22	1.34E+06	1.03
	23	6.61E+05	0.51
	24	1.46E+06	1.12
	25	1.55E+06	1.19
	26	1.57E+06	1.21
	27	1.52E+06	1.17
	28	1.17E+06	0.90
	29	1.31E+06	1.01
	mean	1.32E+06	1.02
	stdev	2.33E+05	0.18
	sterr	6.46E+04	0.05
0.4	1-13	7.77E+05	1.00
	14-17	8.08E+05	1.04
	18	7.95E+05	1.02
	19	7.40E+05	0.95
	20	8.28E+05	1.07
	21	8.48E+05	1.09
	22	8.50E+05	1.09
	23	8.06E+05	1.04
	24	8.99E+05	1.16
	25	8.76E+05	1.13
	26	8.08E+05	1.04
	27	8.27E+05	1.06
	mean	8.22E+05	1.06
	stdev	4.32E+04	0.06
	sterr	1.25E+04	0.02
0.6	1-9	8.56E+05	1.00
	10-15	9.39E+05	1.10
	16-17	9.30E+05	1.09
	18	8.43E+05	0.98
	19	1.40E+06	1.64
	21	9.59E+05	1.12
	22	9.15E+05	1.07
	23	1.51E+06	1.76
	24	1.44E+06	1.68
	25	1.44E+06	1.68
	26	1.32E+06	1.54
	mean	1.14E+06	1.33
	stdev	2.75E+05	0.32
	sterr	8.28E+04	0.10
0.8	1-19	9.11E+05	1.00
	20	9.55E+05	1.05
	21	9.76E+05	1.07
	22	9.54E+05	1.05
	23	1.36E+06	1.49
	24	1.45E+06	1.59
	25	1.51E+06	1.66
	26	1.39E+06	1.53
	27	1.26E+06	1.38
	28	1.12E+06	1.23
	mean	1.19E+06	1.30
	stdev	2.32E+05	0.25
	sterr	7.33E+04	0.08

[GuDCl] (M)	Time Series Data Point	Peak Area	Norm Factor
1.0	1-17	7.63E+05	1.00
	18-20	8.05E+05	1.06
	21	8.27E+05	1.08
	22	8.41E+05	1.10
	23	8.66E+05	1.13
	24	8.06E+05	1.06
	25	8.72E+05	1.14
	26	8.65E+05	1.13
	27	8.61E+05	1.13
	28	1.37E+06	1.80
	29	1.43E+06	1.87
	30	1.01E+06	1.32
	31	1.31E+06	1.72
	32	1.27E+06	1.66
	mean	9.93E+05	1.30
	stdev	2.40E+05	0.31
	sterr	6.42E+04	0.08
1.2	1-15	3.73E+05	1.00
	16	4.89E+05	1.31
	17	4.89E+05	1.31
	18-19	4.99E+05	1.34
	20	5.17E+05	1.39
	21	5.50E+05	1.47
	22	5.52E+05	1.48
	23	5.64E+05	1.51
	24	5.60E+05	1.50
	25	5.04E+05	1.35
	mean	5.10E+05	1.37
	stdev	5.64E+04	0.15
	sterr	1.78E+04	0.05
1.4	1-15	6.29E+05	1.00
	16	7.90E+05	1.26
	17	7.59E+05	1.21
	18	7.55E+05	1.20
	mean	7.33E+05	1.17
	stdev	7.12E+04	0.11
	sterr	3.56E+04	0.06
1.6	1-6	9.37E+05	1.00
	7-16	9.73E+05	1.04
	17-21	9.59E+05	1.02
	22	1.04E+06	1.11
	23	1.02E+06	1.09
	24	1.01E+06	1.08
	26	1.04E+06	1.11
	27	1.02E+06	1.09
	28	1.01E+06	1.08
	29	1.04E+06	1.11
	30	1.01E+06	1.08
	mean	1.01E+06	1.07
	stdev	3.46E+04	0.04
	sterr	1.04E+04	0.01
1.8	1-17	7.97E+05	1.00
	18	7.85E+05	0.98
	21-22	8.02E+05	1.01
	23	6.74E+05	0.85
	mean	7.65E+05	0.96
	stdev	6.08E+04	0.08
	sterr	3.04E+04	0.04

BIBLIOGRAPHY

Alexandrescu, A. T., Jaravine, V. A., Dames, S. A. & Lamour, F. P. (1999). NMR hydrogen exchange of the OB-fold protein LysN as a function of denaturant: the most conserved elements of structure are the most stable to unfolding. *J Mol Biol* **289**(4), 1041-54.

Arrington, C. B. & Robertson, A. D. (1997). Microsecond protein folding kinetics from native-state hydrogen exchange. *Biochemistry* **36**(29), 8686-91.

Arrington, C. B. & Robertson, A. D. (2000). Microsecond to minute dynamics revealed by EX1-type hydrogen exchange at nearly every backbone hydrogen bond in a native protein. *J Mol Biol* **296**(5), 1307-17.

Arrington, C. B. & Robertson, A. D. (2000). Correlated motions in native proteins from MS analysis of NH exchange: evidence for a manifold of unfolding reactions in ovomucoid third domain. *J Mol Biol* **300**(1), 221-32.

Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**(1), 75-86.

Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1994). Protein stability parameters measured by hydrogen exchange. *Proteins* **20**(1), 4-14.

Bai, Y. & Englander, S. W. (1994). Hydrogen bond strength and beta-sheet propensities: the role of a side chain blocking effect. *Proteins* **18**(3), 262-6.

Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science* **269**(5221), 192-7.

Baker, D. (1998). Metastable states and folding free energy barriers [news]. *Nat Struct Biol* **5**(12), 1021-4.

Baker, D. (2000). A surprising simplicity to protein folding. *Nature* **405**(6782), 39-42.

Balbach, J., Forge, V., Lau, W. S., van Nuland, N. A., Brew, K. & Dobson, C. M. (1996). Protein folding monitored at individual residues during a two- dimensional NMR experiment. *Science* **274**(5290), 1161-3.

Bhuyan, A. K. & Udgaonkar, J. B. (1998). Two structural subdomains of barstar detected by rapid mixing NMR measurement of amide hydrogen exchange. *Proteins* **30**(3), 295-308.

Bork, P., Downing, A. K., Kieffer, B. & Campbell, A. D. (1996). Structure and distribution of modules in extracellular proteins. *Quaterly Reviews pf Biophysics* **29**(2), 119-167.

Branden, C. & Tooze, J. (1991). *Introduction to Protein Structure*. 1 edit, Garland Publishing Inc., New York.

Brun, E., Johnson, P. E., Creagh, A. L., Tomme, P., Webster, P., Haynes, C. A. & McIntosh, L.P. (2000). Structure and binding specificity of the second N-terminal cellulose- binding domain from Cellulomonas fimi endoglucanase C. *Biochemistry* **39**(10), 2445-58.

Chamberlain, A. K., Handel, T. M. & Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol* **3**(9), 782-7.

Chamberlain, A. K. & Marqusee, S. (1997). Touring the landscapes: partially folded proteins examined by hydrogen exchange. *Structure* **5**(7), 859-63.

Chamberlain, A. K. & Marqusee, S. (1998). Molten globule unfolding monitored by hydrogen exchange in urea. *Biochemistry* **37**(7), 1736-42.

Chamberlain, A. K., Fischer, K. F., Reardon, D., Handel, T. M. & Marqusee, A. S. (1999). Folding of an isolated ribonuclease H core fragment. *Protein Sci* **8**(11), 2251-7.

Chamberlain, A. K. & Marqusee, S. (2000). Comparison of equilibrium and kinetic approaches for determining protein folding mechanisms. *Adv Protein Chem* **53**, 283-328.

Chu, R. A., Takei, J., Barchi, J. J., Jr. & Bai, Y. (1999). Relationship between the native-state hydrogen exchange and the folding pathways of barnase. *Biochemistry* **38**(43), 14119-24.

Clarke, J. & Fersht, A. R. (1996). An evaluation of the use of hydrogen exchange at equilibrium to probe intermediates on the protein folding pathway. *Fold Des* **1**(4), 243-54.

Clarke, J., Itzhaki, L. S. & Fersht, A. R. (1997). Hydrogen exchange at equilibrium: a short cut for analysing protein-folding pathways? *Trends Biochem Sci* **22**(8), 284-7.

Clarke, J. & Itzhaki, L. S. (1998). Hydrogen exchange and protein folding. *Curr Opin Struct Biol* **8**(1), 112-8.

Clarke, J., Cota, E., Fowler, S. B. & Hamill, S. J. (1999). Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway. *Structure Fold Des* **7**(9), 1145-53.

Connelly, G. P., Bai, Y., Jeng, M. F. & Englander, S. W. (1993). Isotope effects in peptide group hydrogen exchange. *Proteins* **17**(1), 87-92.

Cota, E. & Clarke, J. (2000). Folding of beta-sandwich proteins: three-state transition of a fibronectin type III module. *Protein Sci* **9**(1), 112-20.

Coutinho, J. B., Gilkes, N. R., Warren, R. A., Kilburn, D. G. & Miller, R. C., Jr. (1992). The binding of Cellulomonas fimi endoglucanase C (CenC) to cellulose and Sephadex is mediated by the N-terminal repeats. *Mol Microbiol* **6**(9), 1243-52.

Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G. & Haynes, C. A. (1996). Binding of the cellulose-binding domain of exoglucanase Cex from Cellulomonas fimi to insoluble microcrystalline cellulose is entropically driven. *Proc Natl Acad Sci U S A* **93**(22), 12229-34.

Creamer, T. P. & Rose, G. D. (1994). Alpha-helix-forming propensities in peptides and proteins.

Proteins **19**(2), 85-97.

Creighton, T. E., Ed. (1992). Protein Folding. 1 edit. New York: W.H.Freeman and Company.

Creighton, T. E. (1993). *Proteins*. Second edit, W. H. Freeman and Company, New York.

Dabora, J. M. & Marqusee, S. (1994). Equilibrium unfolding of Escherichia coli ribonuclease H: characterization of a partially folded state. *Protein Sci* **3**(9), 1401-8.

Dabora, J. M., Pelton, J. G. & Marqusee, S. (1996). Structure of the acid state of Escherichia coli ribonuclease HI. *Biochemistry* **35**(37), 11951-8.

de Alba, E., Rico, M. & Jimenez, M. A. (1997). Cross-strand side-chain interactions versus turn conformation in beta- hairpins. *Protein Sci* **6**(12), 2548-60.

Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat Struct Biol* **4**(1), 10-9.

Englander, S. W. & Kallenbach, N. R. (1983). Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Quarterly Reviews of Biophysics* **16**, 521-655.

Englander, S. W. & Mayne, L. (1992). Protein folding studied using hydrogen-exchange labeling and two- dimensional NMR. *Annu Rev Biophys Biomol Struct* **21**, 243-65.

Englander, S. W., Sosnick, T. R., Englander, J. J. & Mayne, L. (1996). Mechanisms and uses of hydrogen exchange. *Curr Opin Struct Biol* **6**(1), 18-23.

Englander, S. W., Mayne, L., Bai, Y. & Sosnick, T. R. (1997). Hydrogen exchange: the modern legacy of Linderstrom-Lang. *Protein Sci* **6**(5), 1101-9.

Englander, S. W. (1998). Native-state HX. *Trends Biochem Sci* **23**(10), 378; discussion 379-81.

Fischer, K. F. & Marqusee, S. (2000). A rapid test for identification of autonomous folding units in proteins. *J Mol Biol* **302**(3), 701-12.

Fuentes, E. J. & Wand, A. J. (1998). Local stability and dynamics of apocytochrome b562 examined by the dependence of hydrogen exchange on hydrostatic pressure. *Biochemistry* **37**(28), 9877-83.

Fuentes, E. J. & Wand, A. J. (1998). Local dynamics and stability of apocytochrome b562 examined by hydrogen exchange. *Biochemistry* **37**(11), 3687-98.

Gerstein, M. & Levitt, M. (1997). A structural census of the current population of protein sequences. *Proc Natl Acad Sci U S A* **94**(22), 11911-6.

Ghaemmaghami, S., Fitzgerald, M. C. & Oas, T. G. (2000). A quantitative, high-throughput screen for protein stability. *Proc Natl Acad Sci U S A* **97**(15), 8296-301.

Gilkes, N. R., Jervis, E., Henrissat, B., Tekant, B., Miller, R. C., Jr., Warren, R. A. & Kilburn, D. G. (1992). The adsorption of a bacterial cellulase and its two isolated domains to crystalline cellulose. *J Biol Chem* **267**(10), 6743-9.

Goedken, E. R., Raschke, T. M. & Marqusee, S. (1997). Importance of the C-terminal helix to the stability and enzymatic activity of Escherichia coli ribonuclease H. *Biochemistry* **36**(23), 7256-63.

Grantcharova, V. P. & Baker, D. (1997). Folding dynamics of the src SH3 domain. *Biochemistry* **36**(50), 15685-92.

Hilser, V. J. & Freire, E. (1997). Predicting the equilibrium protein folding pathway: structure-based analysis of staphylococcal nuclease. *Proteins* **27**(2), 171-83.

Hollien, J. & Marqusee, S. (1999). Structural distribution of stability in a thermophilic enzyme. *Proc Natl Acad Sci U S A* **96**(24), 13674-8.

Hollien, J. & Marqusee, S. (1999). A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry* **38**(12), 3831-6.

Hosszu, L. L., Baxter, N. J., Jackson, G. S., Power, A., Clarke, A. R., Waltho, J. P., Craven, C. J. & Collinge, J. (1999). Structural mobility of the human prion protein probed by backbone hydrogen exchange. *Nat Struct Biol* **6**(8), 740-3.

Huyghues-Despointes, B. M., Scholtz, J. M. & Pace, C. N. (1999). Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat Struct Biol* **6**(10), 910-2.

Huyghues-Despointes, B. M., Langhorst, U., Steyaert, J., Pace, C. N. & Scholtz, J. M. (1999). Hydrogen-exchange stabilities of RNase T1 and variants with buried and solvent-exposed Ala --> Gly mutations in the helix. *Biochemistry* **38**(50), 16481-90.

Itzhaki, L. S., Neira, J. L. & Fersht, A. R. (1997). Hydrogen exchange in chymotrypsin inhibitor 2 probed by denaturants and temperature. *J Mol Biol* **270**(1), 89-98.

Jaravine, V. A., Rathgeb-Szabo, K. & Alexandrescu, A. T. (2000). Microscopic stability of cold shock protein A examined by NMR native state hydrogen exchange as a function of urea and trimethylamine N- oxide. *Protein Sci* **9**(2), 290-301.

Jerala, R. & Zerovnik, E. (1999). Accessing the global minimum conformation of stefin A dimer by annealing under partially denaturing conditions. *J Mol Biol* **291**(5), 1079-89.

Johnson, P. E., Creagh, A. L., Brun, E., Joe, K., Tomme, P., Haynes, C. A. & McIntosh, L. P. (1998). Calcium binding by the N-terminal cellulose-binding domain from Cellulomonas fimi beta-1,4-glucanase CenC. *Biochemistry* **37**(37), 12772-81.

Johnson, P. E., Brun, E., MacKenzie, L. F., Withers, S. G. & McIntosh, L. P. (1999). The cellulose-binding domains from Cellulomonas fimi beta-1, 4- glucanase CenC bind nitroxide spin-labeled cellooligosaccharides in multiple orientations. *J Mol Biol* **287**(3), 609-25.

Jonasson, P., Kjellsson, A., Sethson, I. & Jonsson, B. H. (1999). Denatured states of human carbonic anhydrase II: an NMR study of hydrogen/deuterium exchange at tryptophan-indole-H(N) sites. *FEBS Lett* **445**(2-3), 361-5.

Kay, L., Keifer, P. & Saarinen, T. (1992). *J Am Chem Soc* **114**, 10663-10665.

Keck, J. L. & Marqusee, S. (1996). The putative substrate recognition loop of Escherichia coli ribonuclease H is not essential for activity. *J Biol Chem* **271**(33), 19883-7.

Kern, G., Handel, T. & Marqusee, S. (1998). Characterization of a folding intermediate from HIV-1 ribonuclease H. *Protein Sci* **7**(10), 2164-74.

Koradi, R. (2000). MOLMOL 2.6 edit. Spectrospin, Faellanden.

Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* **55**(3), 379-400.

Li, R. & Woodward, C. (1999). The hydrogen exchange core and protein folding. *Protein Sci* **8**(8), 1571-90.

Liu, H., Farr-Jones, S., Ulyanov, N. B., Llinas, M., Marqusee, S., Groth, D., Cohen, F. E., Prusiner, S. B. & James, T. L. (1999). Solution structure of Syrian hamster prion protein rPrP(90-231). *Biochemistry* **38**(17), 5362-77.

Llinas, M. & Marqusee, S. (1998). Subdomain interactions as a determinant in the folding and stability of T4 lysozyme. *Protein Sci* **7**(1), 96-104.

Llinas, M., Gillespie, B., Dahlquist, F. W. & Marqusee, S. (1999). The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nat Struct Biol* **6**(11), 1072-1078.

Loftus, D., Gbenle, G. O., Kim, P. S. & Baldwin, R. L. (1986). Effects of denaturants on amide proton exchange rates: a test for structure in protein fragments and folding intermediates. *Biochemistry* **25**(6), 1428-36.

Marqusee, S. & Sauer, R. T. (1994). Contributions of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structure in lambda repressor. *Protein Sci* **3**(12), 2217-

25.

Mayo, S. L. & Baldwin, R. L. (1993). Guanidinium chloride induction of partial unfolding in amide proton exchange in RNase A [see comments]. *Science* **262**(5135), 873-6.

McCallister, E. L., Alm, E. & Baker, D. (2000). Critical role of beta-hairpin formation in protein G folding. *Nat Struct Biol* **7**(8), 669-73.

Merkel, J. S., Sturtevant, J. M. & Regan, L. (1999). Sidechain interactions in parallel beta sheets: the energetics of cross- strand pairings. *Structure Fold Des* **7**(11), 1333-43.

Minor, D. L. & Kim, P. S. (1994). Measurement of the beta-sheet-forming propensities of amino acids. *Nature* **367**(6464), 660-3.

Minor, D. L. & Kim, P. S. (1994). Context is a major determinant of beta-sheet propensity. *Nature* **371**(6494), 264-7.

Minor, D. L. & Kim, P. S. (1996). Context-dependent secondary structure formation of a designed protein sequence. *Nature* **380**(6576), 730-4.

Molday, R. S., Englander, S. W. & Kallen, R. G. (1972). Primary structure effects on peptide group hydrogen exchange. *Biochemistry* **11**(2), 150-8.

Moore, J. W. & Pearson, R. G. (1981). *Kinetics and Mechanism.* 3rd edit, John Wiley & Sons, Toronto.

Muhandiram, D. & Kay, L. (1994). *J Mag Res* **Series B 103**, 203-216.

Mullins, L. S., Pace, C. N. & Raushel, F. M. (1997). Conformational stability of ribonuclease T1 determined by hydrogen- deuterium exchange. *Protein Sci* **6**(7), 1387-95.

Myers, J. K. & Oas, T. G. (1999). Contribution of a buried hydrogen bond to lambda repressor folding kinetics. *Biochemistry* **38**(21), 6761-8.

Neira, J. L., Itzhaki, L. S., Otzen, D. E., Davis, B. & Fersht, A. R. (1997). Hydrogen exchange in chymotrypsin inhibitor 2 probed by mutagenesis. *J Mol Biol* **270**(1), 99-110.

Oberhauser, A. F., Marszalek, P. E., Carrion-Vazquez, M. & Fernandez, J. M. (1999). Single protein misfolding events captured by atomic force microscopy. *Nat Struct Biol* **6**(11), 1025-1028.

Pace, C. N., Shirley, B. A. & Thomson, J. A. (1992). Measuring the conformational stability of a protein. In *Protein Folding* (Creighton, T., ed.), pp. 309-329. W.H. Freeman, New York.

Parker, M. J., Dempsey, C. E., Lorch, M. & Clarke, A. R. (1997). Acquisition of native beta-strand topology during the rapid collapse phase of protein folding. *Biochemistry* **36**(43), 13396-405.

Parker, M. J. & Marqusee, S. (1999). The cooperativity of burst phase reactions explored. *J Mol Biol* **293**(5), 1195-210.

Parker, M. J. & Marqusee, S. (2000). A statistical appraisal of native state hydrogen exchange data: evidence for a burst phase continuum? *J Mol Biol* **300**(5), 1361-75.

Parker, M. J. & Marqusee, S. (2001). A Kinetic Folding Intermediate Probed by Native State Hydrogen Exchange. *J Mol Biol* **305**(3), 593-602.

Perrett, S., Clarke, J., Hounslow, A. M. & Fersht, A. R. (1995). Relationship between equilibrium amide proton exchange behavior and the folding pathway of barnase. *Biochemistry* **34**(29), 9288-98.

Plaxco, K. W., Spitzfaden, C., Campbell, I. D. & Dobson, C. M. (1997). A comparison of the folding kinetics and thermodynamics of two homologous fibronectin type III modules. *J Mol Biol* **270**(5), 763-70.

Qian, H. & Chan, S. I. (1999). Hydrogen exchange kinetics of proteins in denaturants: a generalized two-process model. *J Mol Biol* **286**(2), 607-16.

Ragona, L., Fogolari, F., Romagnoli, S., Zetta, L., Maubois, J. L. & Molinari, H. (1999). Unfolding and refolding of bovine beta-lactoglobulin monitored by hydrogen exchange measurements. *J Mol Biol* **293**(4), 953-69.

Ragona, L., Confalonieri, L., Zetta, L., De Kruif, K. G., Mammi, S., Peggion, E., Longhi, R. & Molinari, H. (1999). Equilibrium unfolding CD studies of bovine beta-lactoglobulin and its 14-52 fragment at acidic pH. *Biopolymers* **49**(6), 441-50.

Raschke, T. M. & Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions [published erratum appears in Nat Struct Biol 1997 Jun;4(6):505]. *Nat Struct Biol* **4**(4), 298-304.

Raschke, T. M. & Marqusee, S. (1998). Hydrogen exchange studies of protein structure. *Curr Opin Biotechnol* **9**(1), 80-6.

Raschke, T. M., Kho, J. & Marqusee, S. (1999). Confirmation of the hierarchical folding of RNase H: a protein engineering study [published erratum appears in Nat Struct Biol 2000 May;7(5):431]. *Nat Struct Biol* **6**(9), 825-31.

Richardson, J. S. (1977). beta-Sheet topology and the relatedness of proteins. *Nature* **268**(5620), 495-500.

Rumbley, J., Hoang, L., Mayne, L. & Englander, S. W. (2001). Inaugural Article: An amino acid code for protein folding. *Proc Natl Acad Sci U S A* **98**(1), 105-112.

Sadqi, M., Casares, S., Abril, M. A., Lopez-Mayorga, O., Conejero-Lara, F. & Freire, E. (1999). The native state conformational ensemble of the SH3 domain from alpha- spectrin. *Biochemistry* **38**(28), 8899-906.

Shtilerman, M., Lorimer, G. H. & Englander, S. W. (1999). Chaperonin function: folding by forced unfolding. *Science* **284**(5415), 822-5.

Skelton, N. J., Kordel, J., Akke, M. & Chazin, W. J. (1992). Nuclear magnetic resonance studies of the internal dynamics in Apo, (Cd2+)1 and (Ca2+)2 calbindin D9k. The rates of amide proton exchange with solvent. *J Mol Biol* **227**(4), 1100-17.

Spudich, G. & Marqusee, S. (2000). A change in the apparent m value reveals a populated intermediate under equilibrium conditions in escherichia coli ribonuclease HI [In Process Citation]. *Biochemistry* **39**(38), 11677-83.

Stryer, L. (1995). *Biochemistry*. 4rth edit, W.H. Freeman and Company, New York.

Taylor, J. R. (1982). *An Introduction to Error Analysis*. First edit, University Science Books, Mill Valley.

Tito, P., Nettleton, E. J. & Robinson, C. V. (2000). Dissecting the hydrogen exchange properties of insulin under amyloid fibril forming conditions: a site-specific investigation by mass spectrometry. *J Mol Biol* **303**(2), 267-78.

van Holde, K. E., Johnson, W. C. & Ho, P. S. (1998). *Physical Biochemistry*, Prentice Hall, Upper Saddle River.

Wildegger, G., Liemann, S. & Glockshuber, R. (1999). Extremely rapid folding of the C-terminal domain of the prion protein without kinetic intermediates. *Nat Struct Biol* **6**(6), 550-3.

Woodward, C. & Li, R. (1998). The slow-exchange core and protein folding. *Trends Biochem Sci* **23**(10), 379-81.

Wooll, J. O., Wrabl, J. O. & Hilser, V. J. (2000). Ensemble modulation as an origin of denaturant-independent hydrogen exchange in proteins. *J Mol Biol* **301**(2), 247-56.

Xu, Y., Mayne, L. & Englander, S. W. (1998). Evidence for an unfolding and refolding pathway in cytochrome c. *Nat Struct Biol* **5**(9), 774-8.

Yee, A., Szymczyna, B. & O'Neil, J. D. (1999). Backbone dynamics of detergent-solubilized alamethicin from amide hydrogen exchange measurements. *Biochemistry* **38**(20), 6489-98.

Yi, Q., Scalley, M. L., Simons, K. T., Gladwin, S. T. & Baker, D. (1997). Characterization of the free energy spectrum of peptostreptococcal protein L. *Fold Des* **2**(5), 271-80.