Stability and Dynamics Studies of Apo-azurin From *Pseudomonas aeruginosa*

Thesis by

Ji-Ye Luo

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Division of Chemistry and Chemical Engineering
California Institute of Technology
Pasadena, California

1997

(Submitted September 4, 1996)

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Steve Mayo for everything all these years.

I would also like to thank Dr. Sunney Chan, Dr. John Richards, Dr. Douglas Rees, and Dr. Eric Carreira who served on my committee.

Many thanks also go to current and past Mayo group members: Bassil Dahiyat, Huey-Jenn Chiang, Jane Takonuche, Alyce Su, Scott Ross, Elaine Marzloff, Chantal Morgan, Sandy Malakauskas, and Marie Ary; members of Dr. John Richards group: Claire Slutter, Rolf Langen, Cindy Kiaser, and Neil Farrow; and friends in the Department of Chemistry at Princeton University, particularly my advisor there, Dr. Steve Bernasek.

Last but not least, I would like to thank my family (my wife Xiaoyu and my son Sammy), my parents, and my sister and brother for all their support and encouragement for so many years.

ABSTRACT

The solution stability and dynamics of apo-azurin from *Pseudomonas aeruginosa* have been studied. First, the unfolding stability was measured for 20 apo-azurin mutants mutating at T124 in the context of a β-sheet host-guest environment. The unfolding free energy difference was taken as a measure of the relative β-sheet propensity for the 20 naturally occurring amino acids. These data, when combined with data from other experimental studies and compared to a statistical analysis of the protein structure data base, strongly support intrinsic secondary structure preference as the major determinant of β-sheet propensity. Particularly, residues Val, Thr, Ile, Phe, and Tyr as a group are intrinsically favored for β-sheet formation, and residues Pro, Gly, and Asp as a group are intrinsically disfavored. The relative contribution of the intrinsic propensity and local context to the β-sheet formation was further discussed with the results from mutating S34 to Val and Thr. The S34T mutant was more stable than the wild-type protein, but the S34V mutant was less stable. Secondly, partially perturbed states of different forms of azurin at low pH's were observed. Apo-azurin at pH 2.9 is a molten globule-like state, while holo-azurin at pH's as low as 2.6 is only partially perturbed. Third, backbone ^{1}H and ^{15}N chemical shift assignments and solution dynamics of apo-azurin were studied. The overall correlation time of apo-azurin at 30 °C was determined to be 5.8 ns, and order parameters were mostly at 0.8-0.95. Residues around the Cu(II) binding site exhibited low order parameters and significant ^{1}H chemical shift difference from those of holo-azurin. Last, the osmolyte stabilizing effect was studied on RNase A using Hydrogen/Deuterium exchange. Two types of H/D exchange behavior were observed. The H/D exchange rates of type I residues become slower in the presence of high concentrations of glycine, as expected from the global stability change; the H/D exchange rates of type II residues are not affected, however.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Chapter I

β-Sheet Propensities of the Naturally Occurring Amino Acids

SUMMARY

In this chapter, I describe my study of the β-sheet forming propensities of the 20 naturally occurring amino acids using apo-azurin as the host protein, a predominantly β-sheet protein with 128 amino acid residues. Saturation mutagenesis was carried out at the guest site (T124) in the context of a designed neutral local environment. The relative protein stabilities of all 20 mutant proteins were measured using circular dichroism. A relative stability scale for β-sheet formation was determined for the 20 naturally occurring amino acids based on the thermodynamics of thermal unfolding of the mutant proteins. It was assumed that this fully reflected the effect of the point mutation on β-sheet stability. The resulting propensities were then compared to previously published statistical and experimental data. The analysis revealed that there are amino acids which are always favored for β-sheet formation regardless of the context. These are the β-branched amino acids, Thr, Ile, and Val, and the aromatic amino acids, Tyr and Phe. Pro, Gly, and Asp have the lowest propensity for β-sheet formation. From these results, the 20 amino acids were divided into three groups: group I (Thr, Val, Ile, Tyr, and Phe), group III (Pro, Gly, and Asp), and group II (the other 12 amino acids). The ordering of relative stability among the groups was largely determined by context-independent intrinsic propensities. Within a group, however, the ordering was determined in large part by local context.

INTRODUCTION

Quantification of the thermodynamic scales for α-helix or β-strand formation is an important step in understanding the mechanisms involved in protein folding and in predicting the secondary structure of proteins (Fasman, 1989). Statistical surveys of proteins with known structure reveal a nonrandom distribution of occurrence of each amino acid in a particular secondary structure and imply a definite conformational preference (Chou and Fasman, 1974; Chou and Fasman, 1978; Fasman, 1989; Munoz and Serrano, 1994). It is now generally accepted that the 20 naturally occurring amino acids have different intrinsic propensities for α-helix formation. This conclusion is based on both statistical and experimental studies performed by several groups on different systems (Chakrabartty and Baldwin, 1993; Munoz and Serrano, 1994; Sueki *et al.*, 1984; Wojcik *et al.*, 1990). However, experimental (Kim and Berg, 1993; Minor Jr. and Kim, 1994a; Minor Jr. and Kim, 1994b; Otzen and Fersht, 1995; Smith *et al.*, 1994) and statistical (Chou and Fasman, 1974; Chou and Fasman, 1978; Fasman, 1989; Munoz and Serrano, 1994) studies on β-sheet forming propensities have not led to any consensus.

Chou and Fasman were the first to perform statistical studies on the occurrence frequencies of the 20 naturally occurring amino acids (Chou and Fasman, 1974). Using 19 known protein structures as the database, they found a nonrandom distribution of amino acids in different secondary structures. They updated their study later in 1978 using 29 known protein structures and again in 1989 using 64 known protein structures (Chou and Fasman, 1978; Fasman, 1989). As part of their work on *de novo* protein design, Sander and coworkers also studied the preference parameter for amino acids in various structural states (Sander *et al.*, 1992). They used 38 high resolution protein structures and provided their own definition of a β-sheet in terms of H-bonding. Recently, Munoz and Serrano studied the intrinsic secondary structure propensities of the amino acids using the statistical

$\psi - \phi$ matrices method (Munoz and Serrano, 1994). A much larger database of 279 known protein structures was used. Despite differences in database size and slight differences in the definition of a β-sheet, all these statistical studies revealed a nonrandom amino acid distribution with a basically similar propensity scale for β-sheet formation. More sophisticated statistical methods will probably be developed in the future as the protein structure database continues to expand.

Experimental studies have also been performed by several groups. Mutter and Altman used poly(ethylene glycol) as the host system to assess the β-sheet forming properties of Ala, Ile, and Leu (Mutter and Altman, 1985). Kemp later measured the propensity of Gly, Ala, Phe, and Val using a heteroaromatic template linked to short peptides (Kemp, 1990). These early experimental studies qualitatively established the partial ordering of β-sheet forming propensities for these amino acids: Gly < Ala < Leu < Val, Phe, Ile. An experimental attempt to quantify the β-sheet forming propensities for all 20 amino acids was first completed by Kim and Berg (Kim and Berg, 1993). They used a solvent-exposed position in an antiparallel β-strand of a zinc finger peptide as the guest site and assumed that the thermodynamics of metal binding fully reflected the peptide folding energy. The relative free energies of 20 amino acid mutants were determined by a competitive Co(II)-binding assay. A remarkably good relationship was observed with $P\beta$, the potential value calculated from the frequency of occurrence of the amino acids in β-sheet structures in the protein structure database. However, because their guest site was in an imperfect antiparallel β-sheet, their span of β-sheet propensities was small compared to the span obtained for α-helix propensities and for $P\beta$.

More recently, Kim's group and Regan's group used the IgG binding B1 domain of *streptococcal* protein G as the host protein and a solvent-exposed position (residue No. 53) as the guest site (Minor Jr. and Kim, 1994a; Smith *et al*, 1994). They chose a homogeneous sheet environment and minimized the neighboring side-chain's charge and

steric effects by engineering the neighboring environment to I6A, T44A, T51S, and T55S (Minor Jr. and Kim's study) or to I6A, T44A, T51, and T55 (Smith *et al.*'s study). The thermodynamic scales and β-sheet forming propensities obtained for the 20 amino acids in these two studies not only correlated well with each other, but also correlated well with statistical studies. In addition, the span of the β-sheet forming propensity scale was similar to that reported for the α-helix propensity scale and to the span obtained in statistical studies.

These studies provided a preliminary β-sheet forming propensity scale for the 20 naturally occurring amino acids. However, in a later study, Minor Jr. and Kim used a solvent-exposed position in the middle of an edge strand of a β-sheet as the guest site and found little correlation to their previous study in which the central strand was used. The results also didn't correlate well with the statistically determined propensities (Minor Jr. and Kim, 1994b). They suggested that context was a major determinant of β-sheet forming propensity. Most recently, Otzen and Fersht made mutations at six different positions in the β-sheet of chymotrypsin inhibitor 2 and measured the stability changes for Val, Ala, and Gly (Otzen and Fersht, 1995). They found that the absolute free energy changes varied greatly in different contexts, but that the ranking order in terms of stability remained the same (Val > Ala > Gly). In addition, when Val and Thr were interchanged, the wild type residue was always more stable, but by a different amount at different positions. This also led them to conclude that context is a major determinant of β-sheet stability.

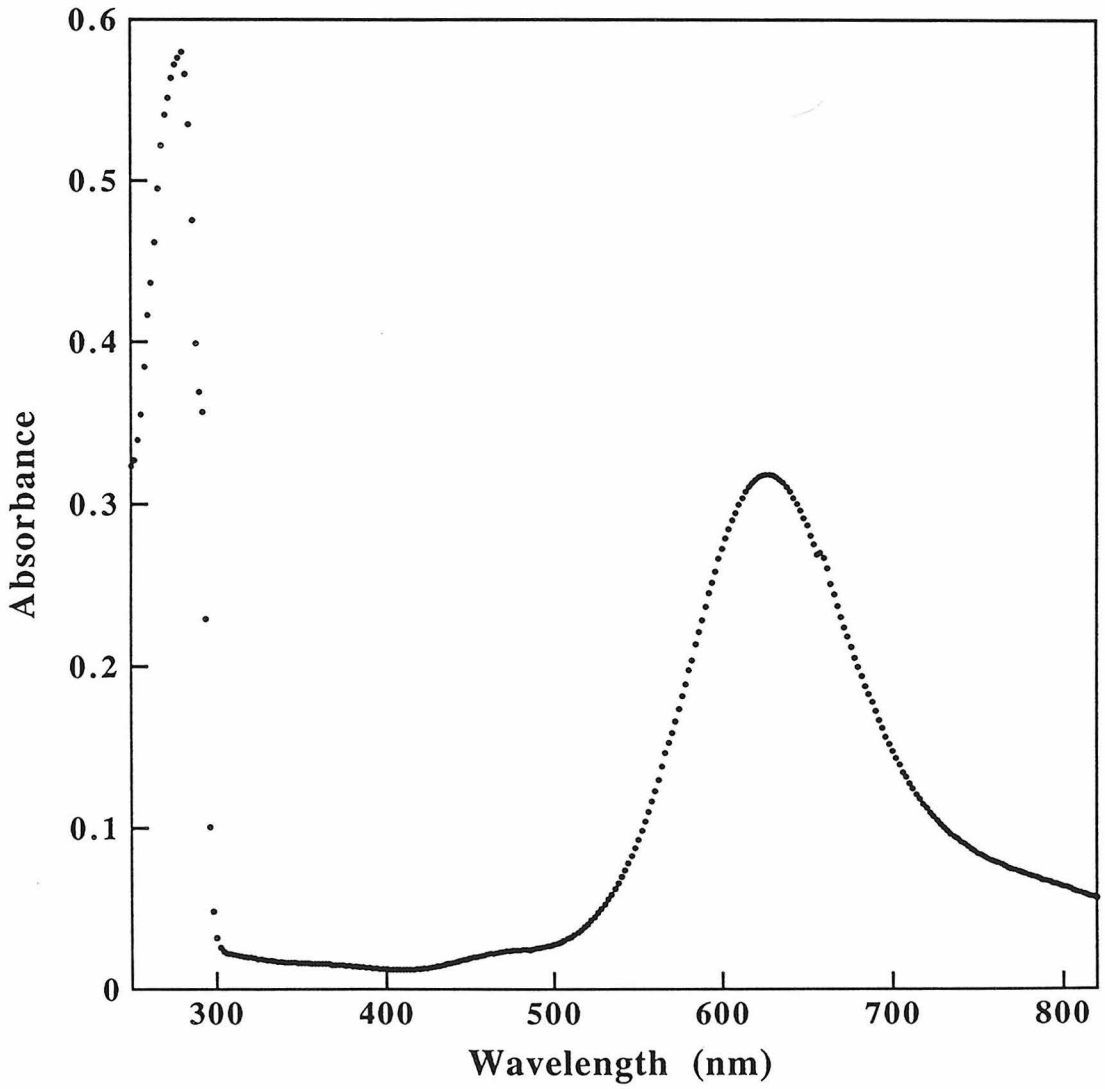
These statistical and experimental studies provide a contradictory view of the relative contributions of intrinsic propensity and local context to β-sheet formation. Additional experiments are needed in order to determine which of these factors is more important. The purpose of this study is to determine the importance of context and intrinsic propensity by obtaining additional experimental results using a totally different host-guest system. A thorough comparison of all the available data will then be presented.

MATERIALS AND METHODS

Protein Mutagenesis and Expression The *Pseudomonas aeruginosa* (*pa*) azurin gene was synthesized and cloned as described previously (Chang *et al.*, 1991). Site saturation mutagenesis was first tried with the Kunkel site-directed mutagenesis method using the available uracil containing single-strand DNA (Kunkel, 1985; Kunkel *et al.*, 1987). DNA oligos were synthesized with degenerate bases at residue 124. Of the 20 desired mutations, 14 were obtained with this method. The remaining mutations were achieved by the CloneTech Transformer Site-Directed Mutagenesis method (CloneTech Laboratories, 1993). The DNA was sequenced by the Perkin Elmer autosequencing protocol using -21M13 universal dye primer or by the traditional ^{35}S ATP method. The mutated genes were then cloned into pET9a for expression in the *E. coli* strain BL21. The cells were grown at 37 °C in LB with 50 μg/ml of kanamycin as the antibiotic. After reaching the logarithmic growth phase (OD_{600nm} ~1.0), the cells were induced with isopropyl-β-D-thiogalactopyranodide (IPTG) (0.5 mM) and allowed to grow for an additional 6–7 hours. $CuSO_4$ solution was then added to the protein solution after it was isolated by osmotic extrusion using 20% sucrose and 30 mM Tris·HCl at pH 8.0, as described previously (Chang *et al.*, 1991). Details of various steps in the mutagenesis experiments are described in the appendices for this chapter.

Isolated azurin was further purified with a home-packed Pharmacia SP Sepharose High Performance Ion Exchange column on a Pharmacia FPLC system. The purified holo-azurin had a UV-Vis absorption ratio A_{626nm}/A_{280nm} of 0.55-0.57 (Figure I.1).

Figure I.1. UV-Vis absorption spectrum of purified holo-azurin at room temperature.


0.6
0.5
0.4
0.3
0.2
0.1
0
Absorbance
300
400
500
600
700
800
Wavelength (nm)

Preparation of Apo-azurin It has been reported that the thermal unfolding of holo-azurin is irreversible (Engeseth and McMillin, 1986). I confirmed this observation (Figure I.2) and believe that the irreversibility is mainly caused by the presence of Cu(II) and the free thiol group of Cys-112 at elevated temperatures; the Cu(II) catalyzed radical reaction between Cu(II) and S^- of the Cys side-chain results in irreversible unfolding products (Luo *et al.*). I also found that the thermal unfolding of apo-azurin is reversible in the absence of Cu(II) under anaerobic conditions (Figure I.3).

A method has been reported for making apo-azurin from holo-azurin by using thio-urea and dialyzing it against reduced holo-azurin (Blaszak *et al.*, 1983). With this method, it takes more than 48 hours to completely eliminate Cu(II) in an inert atmosphere; it also results in a low recovery rate. I worked out an alternate fast method which is based on my observation that acid denatured azurin refolds reversibly. Azurin was first denatured with HCl to pH 1.3 to 1.6 in the presence of 2 times excess EDTA. The protein solution was allowed to stand at room temperature for 5 to 10 minutes. After holo-azurin's blue chromophore completely disappeared, the protein solution was renatured by quickly changing the pH to around 5 with NaOH. Apo-azurin was recovered by ultrafiltration with an amicon PK10 membrane. The complete disappearance of the Cu(II) absorption band for the prepared apo-azurin was checked by UV-Vis spectroscopy (Figure I.4). The yield was over 90%. The apo-azurin prepared by this method was reconstitutable with $CuSO_4$ to A_{280nm}/A_{626nm} of over 0.55. The homogeneity of both the apo-azurin and the reconstituted holo-protein were checked using a size-filtration column.

Figure I.2. CD spectra of holo-azurin at 25 °C before and after thermal unfolding.

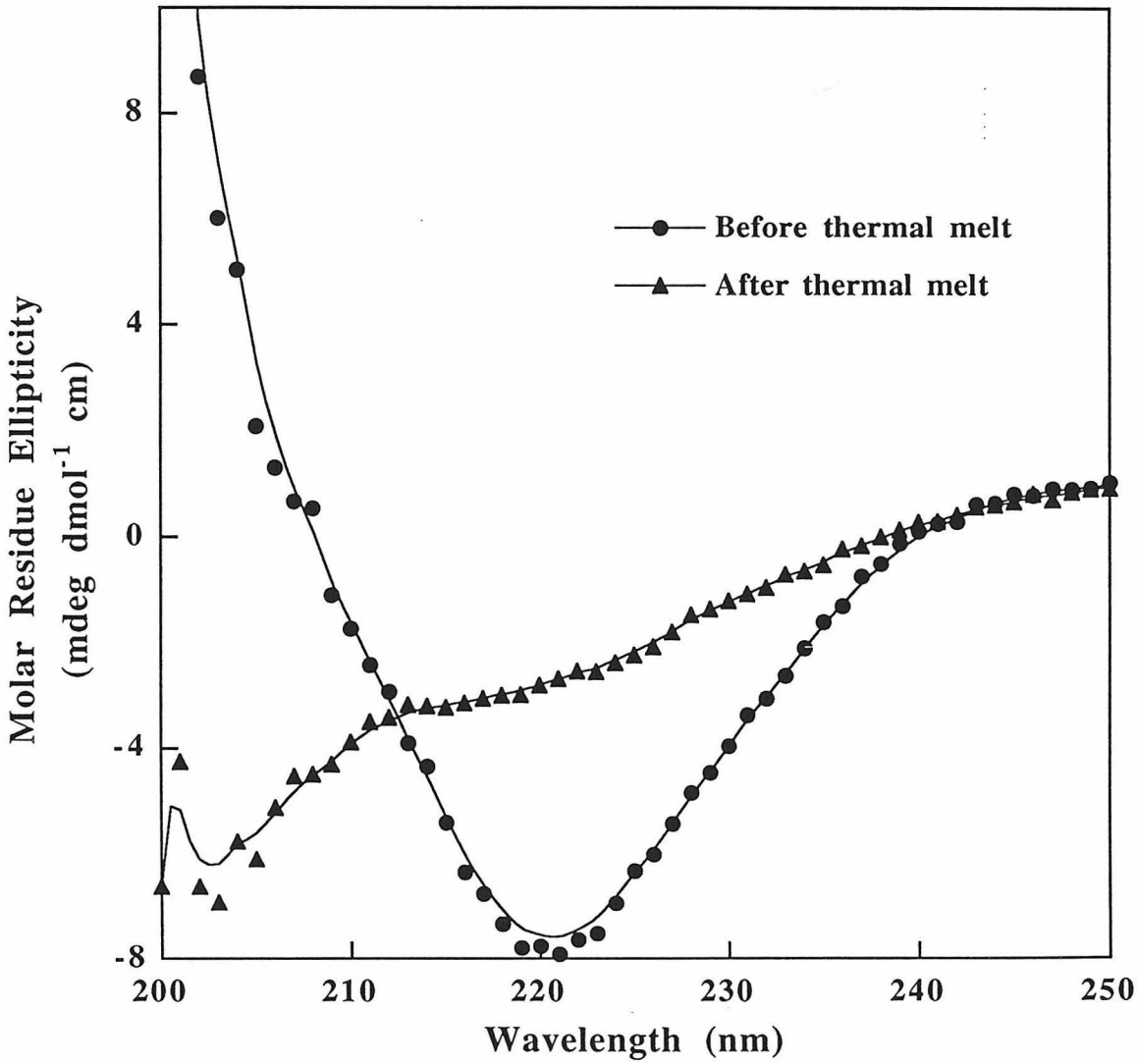

Molar Residue Ellipticity
(mdeg dmol-1 cm)
8
4
0
-4
-8
200
210
220
230
240
250
Wavelength (nm)
Before thermal melt
After thermal melt

Figure I.3. CD spectra of apo-azurin at 25 °C before and after thermal unfolding.

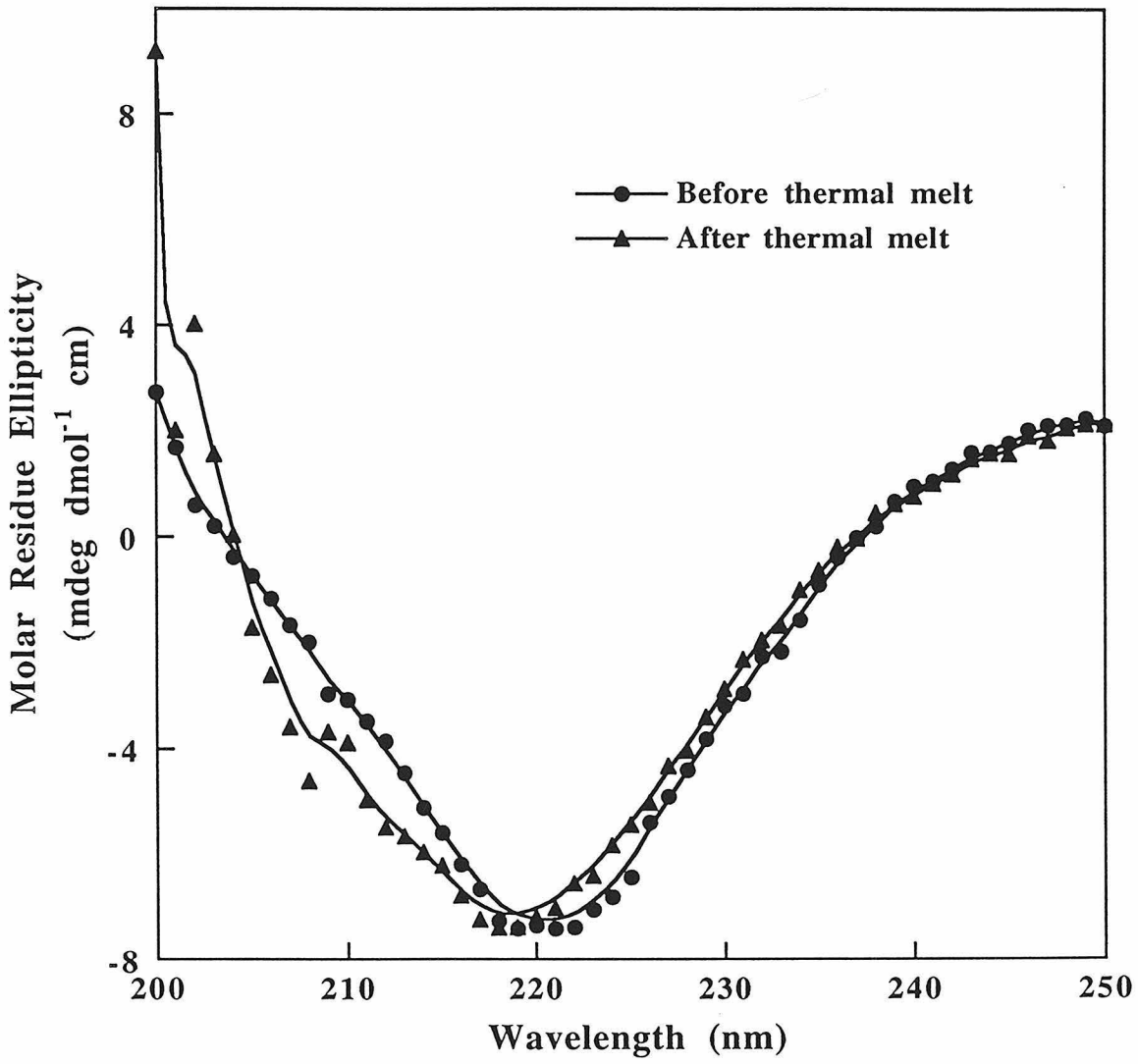
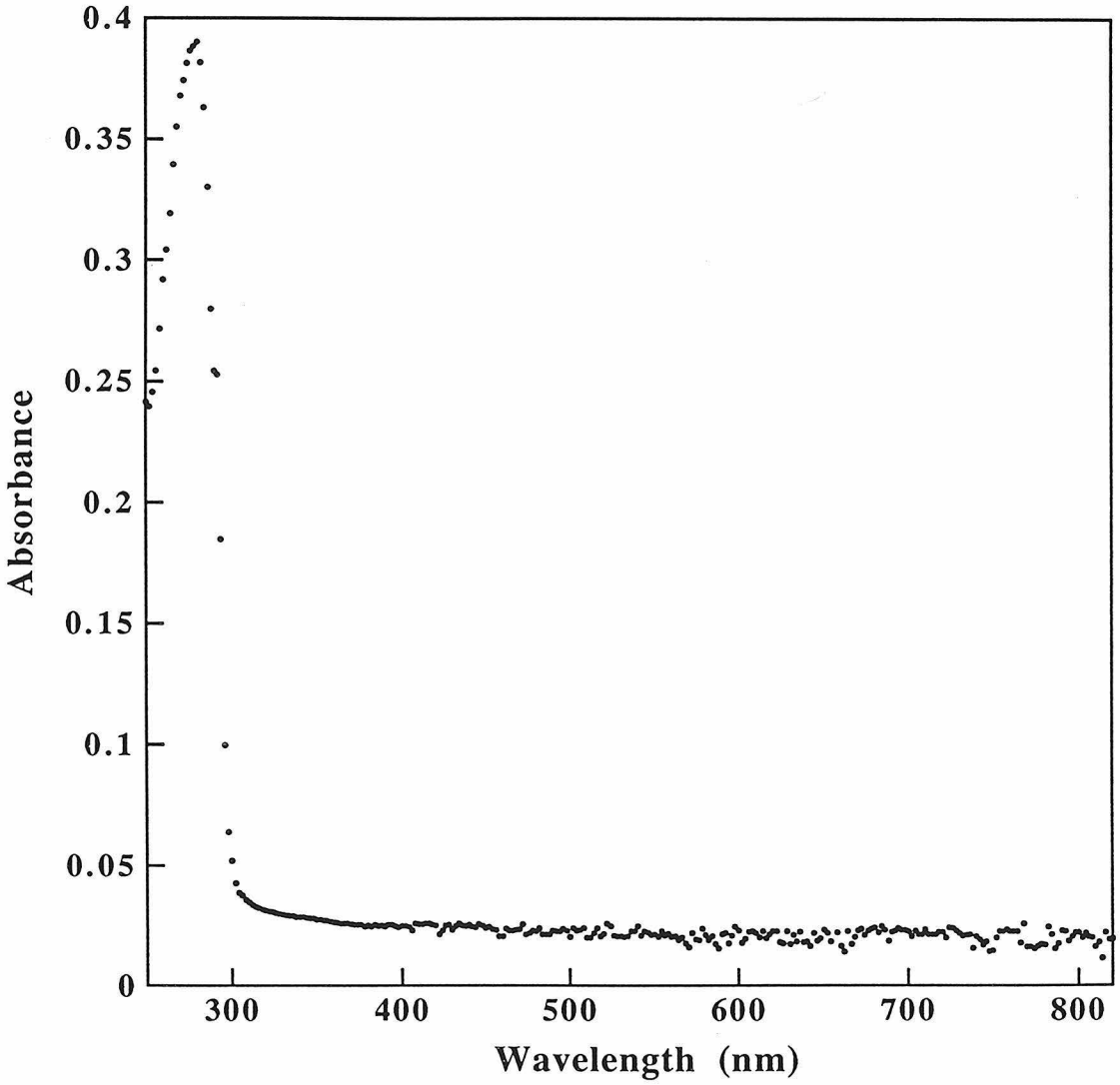

Before thermal melt
After thermal melt
Molar Residue Ellipticity
(mdeg dmol-1 cm)
8
4
0
-4
-8
200
210
220
230
240
250
Wavelength (nm)

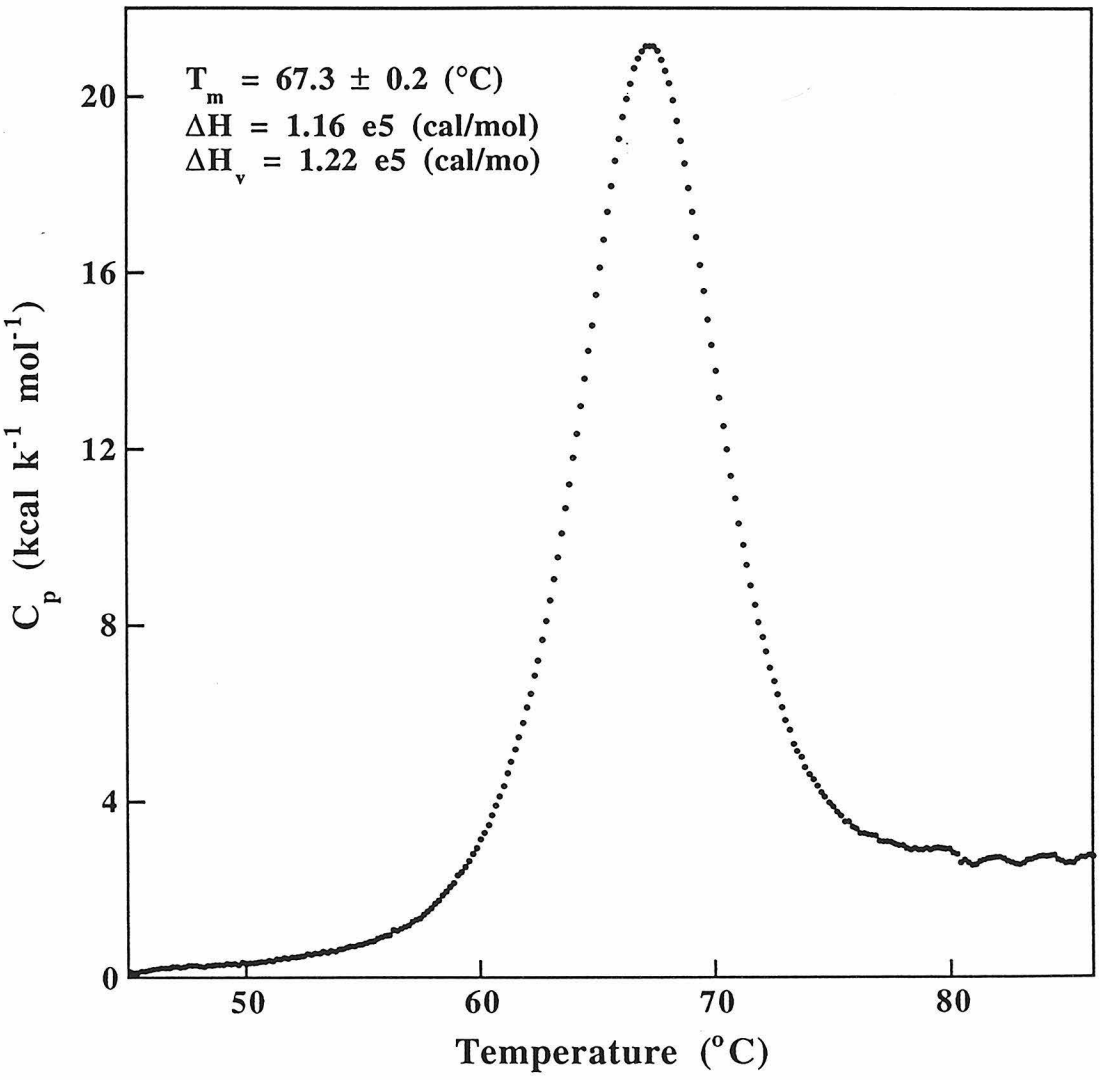
Figure I.4. UV-Vis absorption spectrum of purified apo-azurin at room temperature.


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Absorbance
300
400
500
600
700
800
Wavelength (nm)

Spectral Analysis Protein concentration and UV-Vis absorption characterization were measured on an HP8452A UV-Vis diode array spectrometer. Circular Dichroism (CD) spectra were measured on an AVIV 62DS spectrophotometer with 0.1 cm sample cell path length and a typical protein concentration of 35 μM. Thermal denaturation of proteins was monitored by recording the ellipticity of the proteins at 219 nm as a function of temperature on the AVIV spectrophotometer. A wavelength of 219 nm was chosen because it had the maximum ellipticity change in the protein unfolding reaction.

The thermal unfolding curves were analyzed according to a two-state transition model. The middle point transition temperature (T_m) was determined by taking the derivative of the thermal unfolding curve. The difference of unfolding free energy change, $\Delta\Delta G = \Delta G^* - \Delta G$, was calculated using Becktel and Schellman's approximation equation (Becktel and Schellman, 1987); for small $\Delta\Delta G$, $\Delta\Delta G = \Delta T \frac{\Delta H}{T_m}$. The ΔH value of wild type apo-azurin (119 kcal/mol), and the T_m of the Ala mutant (63.8 °C) were used as the reference. ΔH was measured by Differential Scanning Calorimetry (DSC) using a MicroCal MSC DSC calorimeter, and was analyzed with MSC Origin software. The measured calorimetric ΔH and the fitted van't Hoff ΔH were 116 kcal/mol and 122 kcal/mol respectively with a ratio of 0.95, suggesting a two-state unfolding reaction (Figure I.5). The average of these two ΔH's was used in the $\Delta\Delta G$ calculations.

Figure I.5. DSC measurement of the enthalpy change (ΔH) for the thermal unfolding of apo-azurin. The thermal unfolding scan was at a rate of 1 °C/min. from 20 °C to 90 °C. The T_m value was read from the top of the heat absorption peak with ± 2 °C accuracy. ΔH is the measured calorimetric enthalpy change, and ΔH_v is the fitted van't Hoff enthalpy change using the two-state unfolding reaction assumption. The ratio $\frac{\Delta H}{\Delta H_v}$ is 0.95, indicating a two-state unfolding reaction.


T_m = 67.3 ± 0.2 (°C)
ΔH = 1.16 e5 (cal/mol)
ΔH_v = 1.22 e5 (cal/mo)
C_p (kcal k^{-1} mol^{-1})
20
16
12
8
4
0
50
60
70
80
Temperature (°C)

RESULTS

Selection of the Guest Site and Host Environment *Pa* azurin was chosen as the host protein because it is a small β-sheet protein with 128 amino acids (Figure I.6) and I had previous experience with it. A guest site was selected following a strategy similar to that used by Minor Jr. and Kim and Smith *et al.* (Minor Jr. and Kim, 1994a; Smith *et al.*, 1994), i.e., the site had to be a solvent-exposed neutral residue, it had to be in the middle of a central β-strand, and it had to be surrounded by small neutral amino acid residues. Based on these criteria, T124 was selected. It is in the middle of the c-terminal strand, which itself is a central strand of a β-sheet. In addition, it is directly on a restriction enzyme (*Kpn I*) digestion site (see Appendix I-4), which facilitates protein mutagenesis. I also mutated Lys-122 to Ala to minimize side-chain steric and charge interactions. Figure I.7 shows a closer view of the guest site. The guest substitutes were achieved by site-directed mutagenesis. All 20 amino acids were substituted individually on the guest site in the host environment of A19, A122, M109, and T126 as described under Materials and Methods. The purified proteins all had A_{280nm}/A_{626nm} of 0.55-0.57, characteristic of purified wild type holo-azurin (van de Kamp *et al.*, 1990). The absorption spectra also had a shoulder peak at 292 nm. These absorption results indicate that the mutant proteins were pure and that the wild type structure was preserved.

Figure I.6. Molscript drawing of *pa* azurin crystal structure using the PDB file 4azu.pdb. The guest site (124) and the surrounding residues (A19, K122A, M109, and T126) are highlighted.

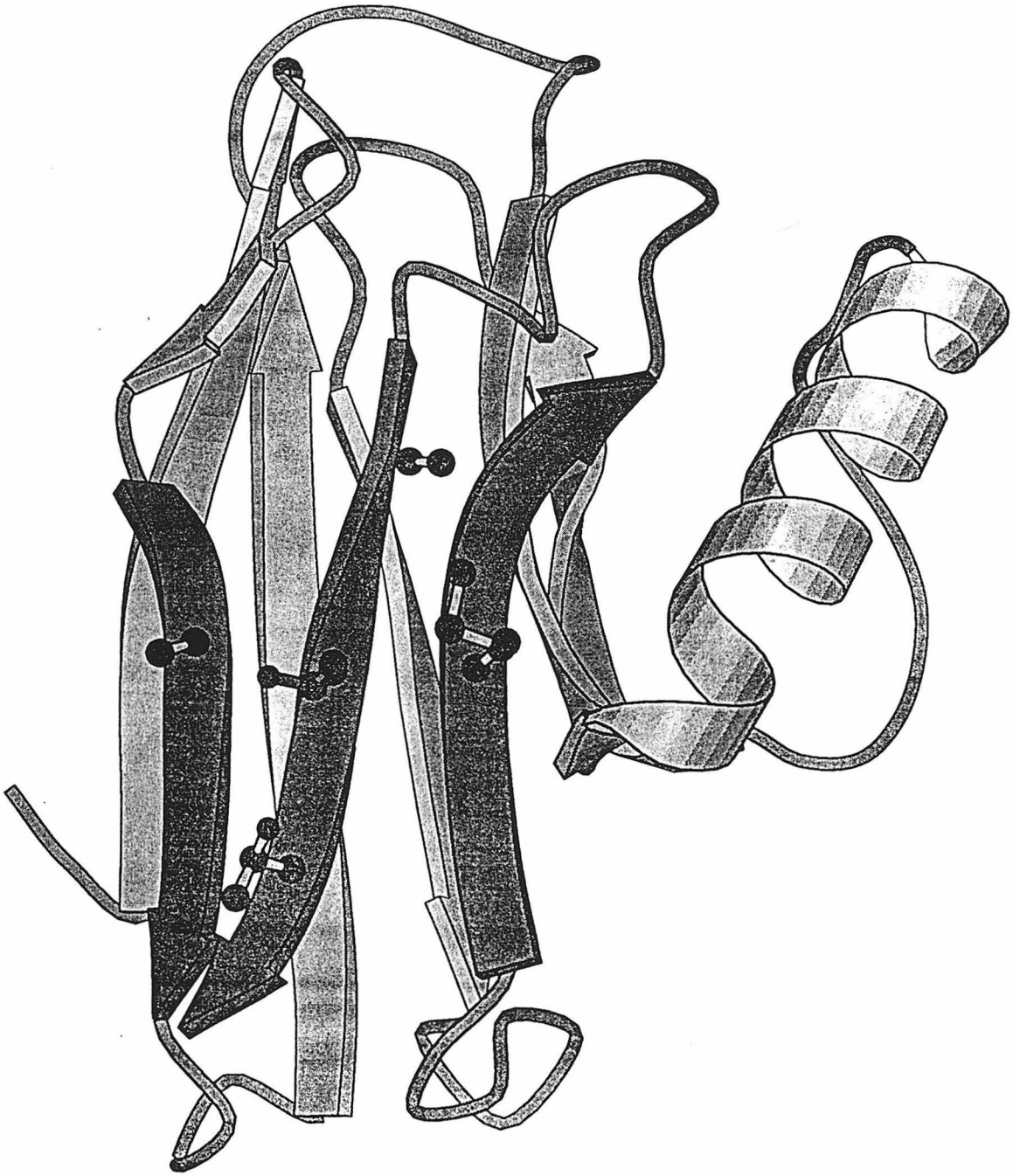
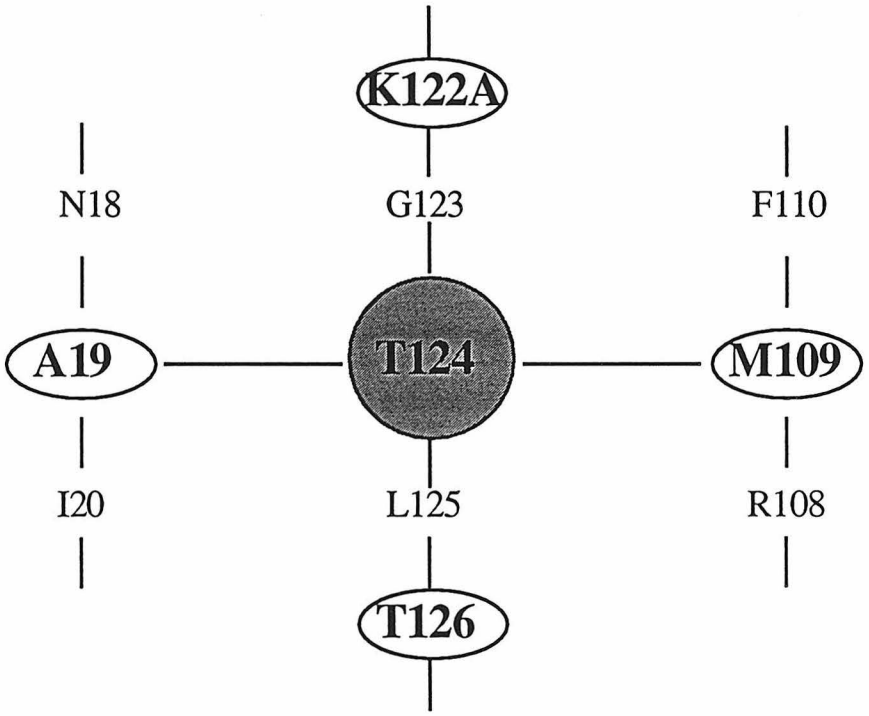


Figure I.7. A closer view of the guest site and surrounding residues.

A close view of the guest site



Stability Measurements Thermal unfolding stabilities were determined using CD. Typical unfolding curves are shown in Figure I.8. The relative difference of the unfolding free energy change, $\Delta\Delta G$, between a particular mutant and the reference (Ala mutant) was calculated using Becktel and Schellman's approximation (Becktel and Schellman, 1987). The measured T_m's and calculated $\Delta\Delta G$'s are listed in Table I.1. The validity of the approximation was checked using Minor Jr. and Kim's published data. As shown in Table I.2, the approximation is very good except when the ΔH of the perturbed protein is significantly lower than that of the wild type protein.

Table I.1 illustrates some notable points. First, there are definitely stability differences in the 20 apo-azurin mutants, with Thr being the most stable and Pro being the least stable. This is consistent with the other experimental results (see Table I.3). The wild type residue, Thr, is the most stable amino acid in my ranking. This may just be a coincidence, but it also suggests that for Thr neither the context change from wild type K122 to 122A in my host-guest system, nor the context change from my system to Minor Jr. and Kim's protein G and Kim and Berg's zinc-finger protein host-guest systems affect its stability ranking. Secondly, if Pro is excluded, the β-sheet forming propensity scale spans from Thr (1.38 kcal/mol) to Asp (−1.41 kcal/mol). This is similar to the range of values obtained by Minor Jr. and Kim and by Smith *et al.*, and is in line with the P_β spans obtained in statistical studies (see Table I.4).

Figure I.8. Typical thermal unfolding curves measured by Aviv 62A DS Circular Dichroism (CD) spectrometer at 219 nm.

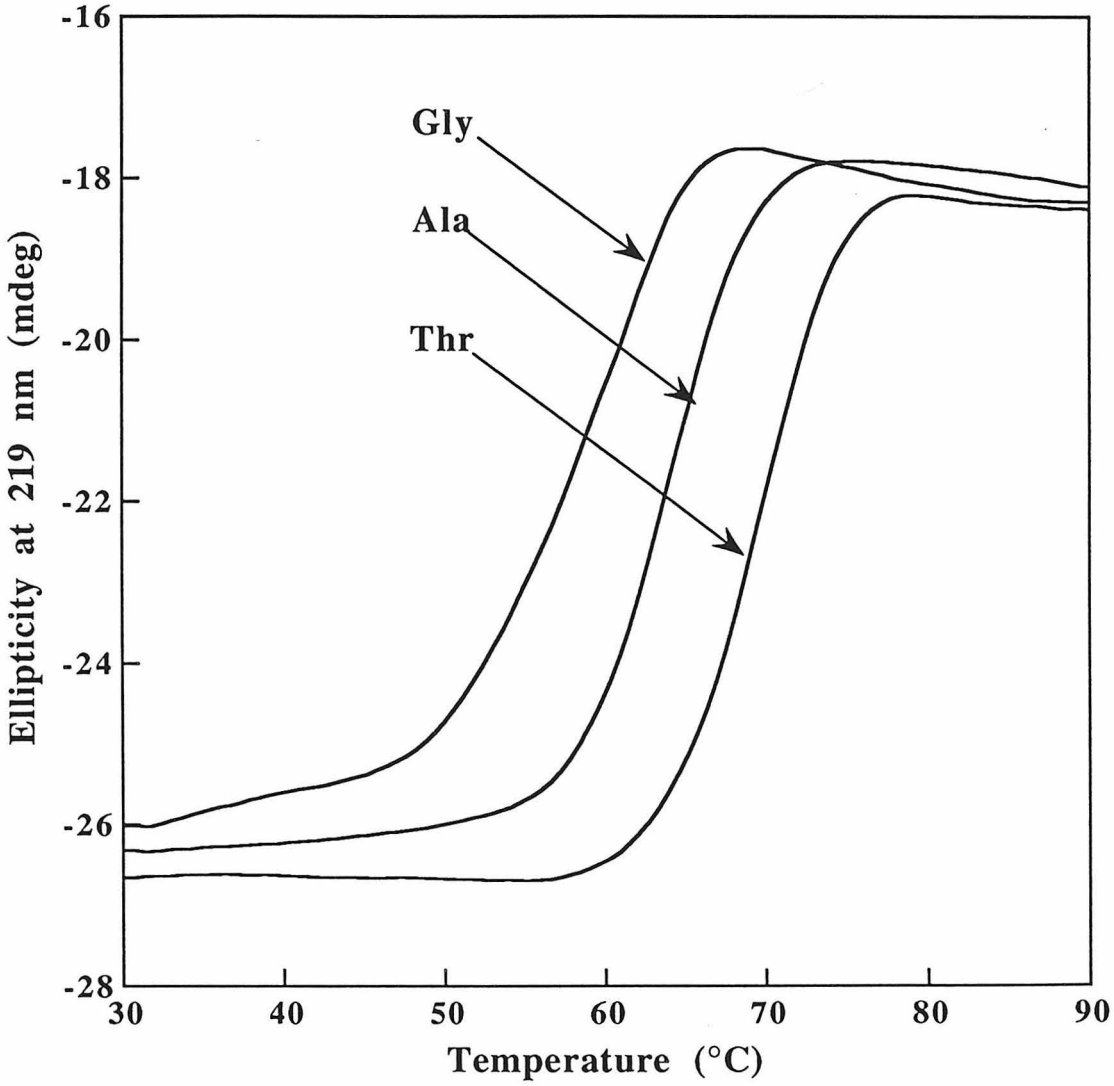

-16
-18
-20
-22
-24
-26
-28
Ellipticity at 219 nm (mdeg)
Gly
Ala
Thr
30
40
50
60
70
80
90
Temperature (°C)

Table I.1: Twenty Apo-azurin Mutants Listed in Descending Order of Thermal Stability

amino acid	T_m (°C) measured	$\Delta\Delta G$ (kcal/mol)
thr	67.7	1.38
val	66.5	0.95
tyr	66.2	0.85
ile	66.0	0.78
phe	65.9	0.74
cys	65.35	0.55
glu	65.3	0.53
gln	65.2	0.49
trp	65.0	0.42
lys	64.95	0.41
met	64.7	0.32
ser	64.65	0.30
asn	64.35	0.19
his	64.0	0.071
leu	63.9	0.035
arg	63.9	0.035
ala	63.8	0.00
gly	60.4	−1.20
asp	59.8	−1.41
pro	57.0	−2.40

Table I.2: Comparison of the Measured $\Delta\Delta G$ Values and the $\Delta\Delta G$ Values Calculated Using Becktel and Schellman's Approximation Method

amino acid	$\Delta\Delta G$ (kcal/mol) of Minor Jr. and Kim — measured*	$\Delta\Delta G$ (kcal/mol) of Minor Jr. and Kim — calculated**
thr	1.10	1.09
ile	1.00	1.02
tyr	0.96	0.96
phe	0.86	0.86
val	0.82	0.82
met	0.72	0.71
ser	0.70	0.70
trp	0.54	0.55
cys	0.52	0.53
leu	0.51	0.51
arg	0.45	0.46
lys	0.27	0.28
gln	0.23	0.23
glu	0.01	0.023
ala	0.0	0.00
his	−0.02	−0.057
asn	−0.08	−0.10
asp	−0.94	−1.00
gly	−1.20	−1.61
pro	< −4.0	< −4.0

* $\Delta\Delta G$ values reported by Minor Jr. and Kim (*Nature* **367**, 660–663, 1994).

** $\Delta\Delta G$ values calculated using Becktel and Schellman's approximation method and the average ΔH of 36 kcal/mol from ΔH values reported by Minor Jr. and Kim.

DISCUSSION

An analysis of the Ramanchandra plot indicates that there is more $\psi - \phi$ space permissible for β-sheet formation than for α-helix formation (Ramachandran *et al.*, 1963). This implies that there is more flexibility in conformation space for amino acids to adapt to a β-sheet than to an α-helix. We would therefore expect that the propensities to form β-sheets would be somewhat weaker than the propensities to form α-helices. In addition, since a β-sheet is formed by neighboring β-strands, there are likely to be more tertiary interactions involved in its formation than in that of an α-helix; that is, there should be more pair-wise side-chain interactions from neighboring strands or nearby residues. Thus, the local context could be a key determinant in the stability of β-sheet formation. This was discussed in part by Minor Jr. and Kim in a study where they used an exposed position on an edge strand of the β-sheet as the guest site (Minor Jr. and Kim, 1994b), and by Otzen and Fersht in their recent study (Otzen and Fersht, 1995). The results from Minor Jr. and Kim's study did not correlate well with their earlier work where the central strand was used as the guest site. They argued that unlike α-helix formation, β-sheet formation is determined in large part by local tertiary context, and not by intrinsic secondary structure preferences.

However, as was pointed out by Flory (Flory, 1969) and Chothia (Chothia, 1973), and more recently by Munoz and Serrano (Munoz and Serrano, 1994), the Ramanchandra plots show a non uniform distribution of dihedral angles of amino acids for β-sheets and support the notion that there are intrinsic differences for β-sheet formation. For example, the domain of angles permitted for Gly (for which side-chain R = H) is much greater than that for Ala. For residues such as Leu ($R = -CH_2CH(CH_3)_2$), Glu ($R = -CH_2CH_2CONH_2$), and Lys ($R = -CH_2CH_2CH_2CH_2NH_2$), which have longer side-chains but are of the form $-CH_2-R'$, the domain of angles differs little from those for

Ala (R = $-CH_3$). A side-chain which is branched at the β carbon diminishes the range of conformations available [e.g., Val (R = $-CH(CH_3)_2$)]. Such facts, pointed out by Ramachandran, Scheraga, and Flory more than 15 years ago, account for the intrinsic differences among different types of amino acids (Flory, 1969; Leach *et al.*, 1966; Ramakrishnan and Ramachandran, 1965).

Table I.3 lists the β-sheet propensities obtained in five experimental studies where different host-guest systems were used; Table I.4 shows the values obtained in different statistical studies. In order to evaluate the relative contributions of intrinsic property and local context, the different propensity data were plotted against each other (see Figures I.9 and I.10). Figure I.9 shows a good correlation between the propensities obtained by Minor Jr. and Kim (Minor Jr. and Kim, 1994a) and those obtained by Smith *et al.* (Smith *et al.*, 1994). This is not surprising since both of these studies used the same host protein and guest site and had a similar guest environment. I therefore took the average of these two propensity scales and used it in comparison with my data. The propensities from my study were plotted against these average propensities and the statistical propensities of Munoz and Serrano (see Figure I.10a and I.10b). Both plots show a good correlation, suggesting that intrinsic propensity is probably the dominant factor.

Table I.3: β-Sheet Propensities from Five Different Experimental Studies.

Luo *et al.*		Smith *et al.*		Minor and Kim (a)		Minor and Kim (b)		Kim and Berg	
amino acid	$\Delta\Delta G$*	amino acid	$\Delta\Delta G$*	amino acid	$\Delta\Delta G$*	amino acid	$\Delta\Delta G$*	amino acid	$\Delta\Delta G$*
thr	1.38	tyr	1.63	thr	1.10	thr	0.83	ile	0.21
val	0.95	thr	1.36	ile	1.00	ser	0.63	phe	0.20
tyr	0.85	ile	1.25	tyr	0.96	glu	0.31	val	0.16
ile	0.78	phe	1.08	phe	0.86	val	0.17	tyr	0.15
phe	0.74	trp	1.04	val	0.82	phe	0.16	thr	0.13
cys	0.55	val	0.94	met	0.72	tyr	0.11	trp	0.13
glu	0.53	met	0.90	ser	0.70	cys	0.08	leu	0.13
gln	0.49	ser	0.87	trp	0.54	ile	0.02	cys	0.12
trp	0.42	cys	0.78	cys	0.52	ala	0	met	0.11
lys	0.41	asn	0.52	leu	0.51	gln	0.04	his	0.11
met	0.32	leu	0.45	arg	0.45	his	−0.01	arg	0.09
ser	0.30	arg	0.40	lys	0.27	met	−0.02	lys	0.06
asn	0.19	gln	0.38	gln	0.23	asp	−0.10	asp	0.06
his	0.07	his	0.37	glu	0.01	trp	−0.17	glu	0.06
arg	0.04	lys	0.35	ala	0.00	leu	−0.24	gln	0.05
leu	0.04	glu	0.23	his	−0.02	asn	−0.24	ser	0.04
ala	0.00	ala	0.00	asn	−0.08	lys	−0.40	asn	0.03
gly	−1.20	asp	−0.85	asp	−0.94	arg	−0.43	ala	0.00
asp	−1.41	gly	−1.21	gly	−1.20	gly	−0.85	pro	−0.12
pro	−2.40	pro		pro		pro		gly	−0.35

*in kcal/mol

Table I.4: β-Sheet Propensities from Different Statistical Studies

Munoz & Serrano		Kabsch & Sander*		Chou & Fasman (1978)		Fasman (1989)	
val	−0.53	val	−0.53	val	0.87	val	0.85
ile	−0.48	ile	−0.50	ile	0.77	ile	0.78
tyr	−0.42	tyr	−0.42	tyr	0.64	thr	0.54
thr	−0.41	phe	−0.40	phe	0.55	tyr	0.52
cys	−0.41	cys	−0.35	trp	0.54	trp	0.45
phe	−0.39	trp	−0.33	leu	0.47	phe	0.44
trp	−0.31	thr	−0.30	cys	0.36	leu	0.38
his	−0.25	leu	−0.29	thr	0.36	cys	0.28
met	−0.25	met	−0.27	gln	0.27	met	0.22
leu	−0.21	his	−0.17	met	0.22	gln	0.21
arg	−0.19	arg	−0.10	arg	0.10	arg	0.15
ser	−0.19	ser	−0.093	asn	0.060	ser	0.15
lys	−0.14	lys	−0.054	his	0.040	gly	0.080
gln	−0.12	gln	−0.030	ala	0.00	his	0.040
asn	−0.063	ala	0.00	ser	−0.080	ala	0.00
glu	−0.016	glu	0.0050	gly	−0.080	lys	−0.060
ala	0.00	gly	0.024	lys	−0.090	asp	−0.13
asp	0.060	asn	0.12	pro	−0.28	asn	−0.13
gly	0.43	asp	0.19	asp	−0.29	pro	−0.17
pro	1.64	pro	0.33	glu	−0.46	glu	−0.28

*These are the statistical values by Munoz and Serrano (*Proteins: Structure, Function, and Genetics* **20**, 301-311, 1994) using Kabsch and Sander's β-sheet secondary structure definition.

Figure I.9. Correlation plot between the β-sheet propensities reported by Smith *et al.* (*Biochemistry* **33**, 5510-5517, 1994) and those obtained by Minor Jr. and Kim (*Nature* **367**, 660-663, 1994).

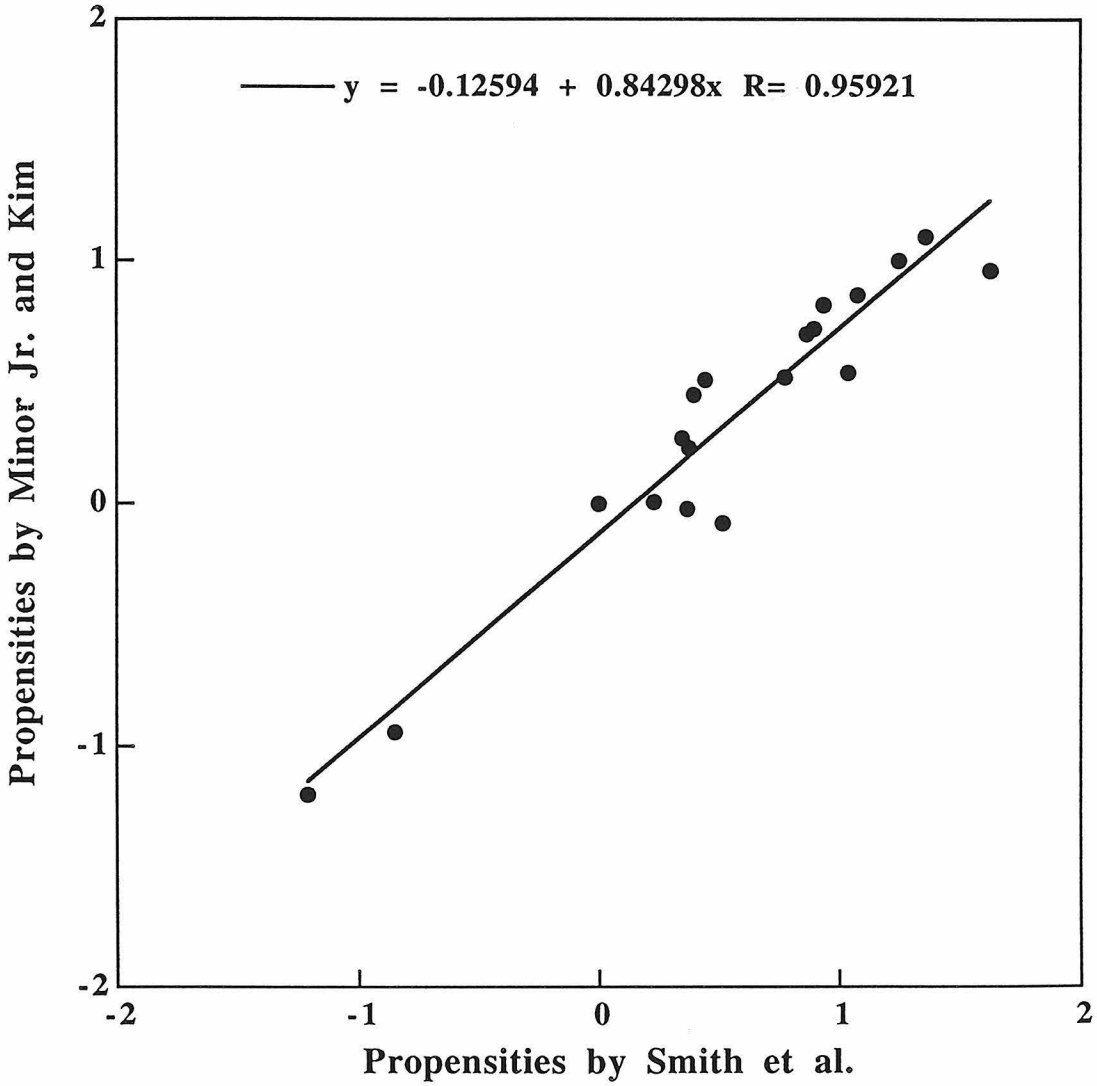
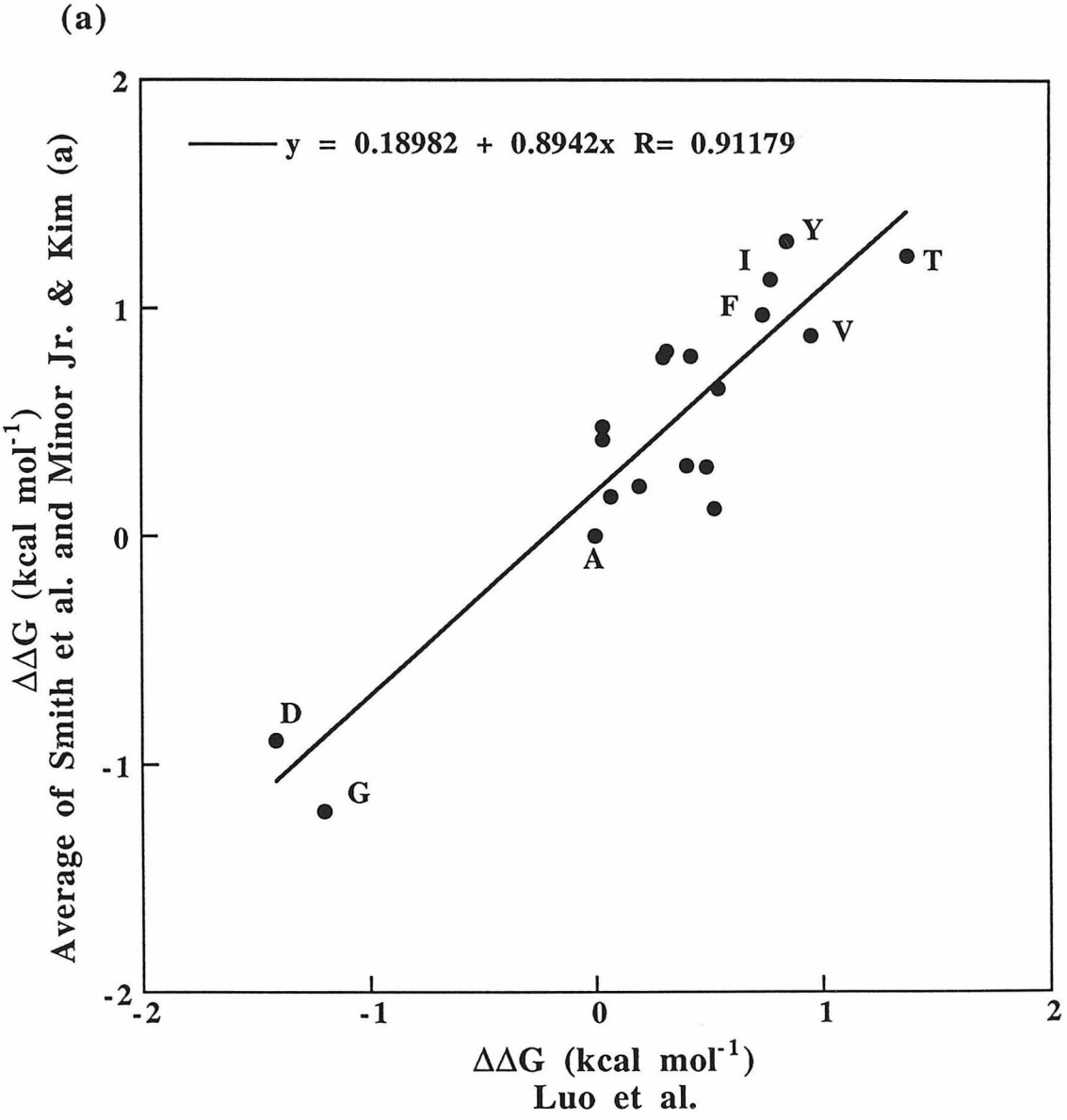

y = -0.12594 + 0.84298x R= 0.95921
Propensities by Minor Jr. and Kim
Propensities by Smith et al.
2
1
0
-1
-2
-2
-1
0
1
2

Figure I.10. (a) Correlation plot between the β-sheet propensities from my study and the average propensities shown in Figure I.9; (b) correlation plot between the β-sheet propensities from my study and those from the statistical study by Munoz and Serrano.


(a)
y = 0.18982 + 0.8942x R= 0.91179
ΔΔG (kcal mol-1)
Average of Smith et al. and Minor Jr. & Kim (a)
ΔΔG (kcal mol-1)
Luo et al.
2
1
0
-1
-2
-2
-1
0
1
2
Y
I
F
T
V
A
D
G

(b)

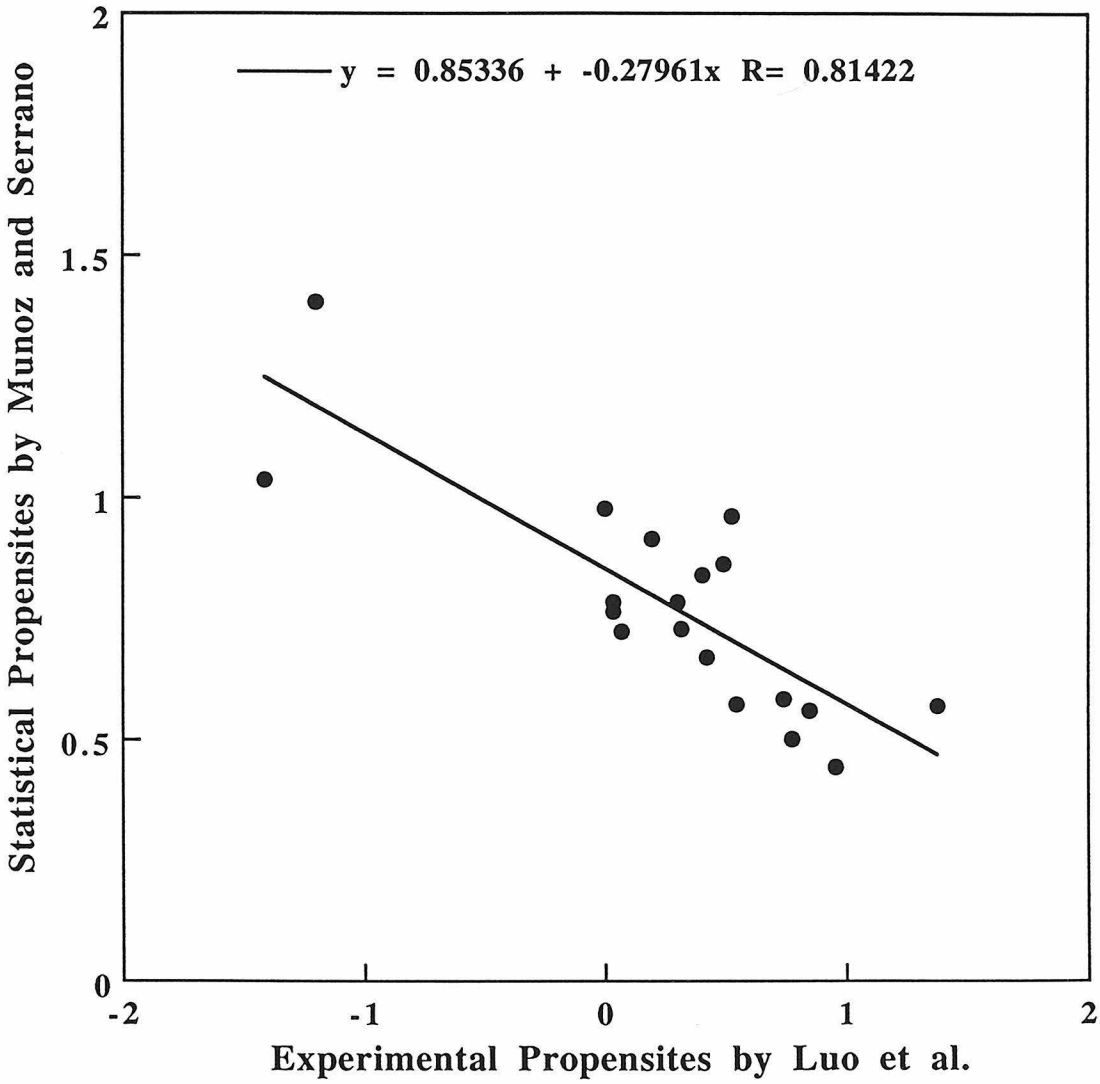

y = 0.85336 + -0.27961x R= 0.81422
Statistical Propensites by Munoz and Serrano
Experimental Propensites by Luo et al.
2
1.5
1
0.5
0
-2
-1
0
1
2

Table I.5: The Normalized Propensities of Table I.3.

Luo *et al.*		Smith *et al.*		Minor and Kim (a)		Minor and Kim (b)		Kim and Berg	
amino acid	$\Delta\Delta G^*$	amino acid	$\Delta\Delta G^*$	amino acid	$\Delta\Delta G^*$	amino acid	$\Delta\Delta G^*$	amino acid	$\Delta\Delta G^*$
thr	0.00	tyr	0.00	thr	0.00	thr	0.00	ile	0.00
val	0.15	thr	0.095	ile	0.043	ser	0.12	phe	0.018
tyr	0.19	ile	0.13	tyr	0.061	glu	0.31	val	0.054
ile	0.22	phe	0.19	phe	0.10	val	0.39	tyr	0.11
phe	0.23	trp	0.21	val	0.12	phe	0.40	thr	0.14
cys	0.30	val	0.24	met	0.17	tyr	0.43	trp	0.14
glu	0.30	met	0.26	ser	0.17	cys	0.45	leu	0.14
gln	0.32	ser	0.27	trp	0.24	ile	0.47	cys	0.16
trp	0.34	cys	0.30	cys	0.25	ala	0.48	met	0.18
lys	0.35	asn	0.39	leu	0.26	gln	0.49	his	0.18
met	0.38	leu	0.42	arg	0.28	his	0.50	arg	0.21
ser	0.39	arg	0.43	lys	0.36	met	0.51	lys	0.27
asn	0.42	gln	0.44	gln	0.38	asp	0.55	asp	0.27
his	0.47	his	0.44	glu	0.47	trp	0.60	glu	0.27
arg	0.48	lys	0.45	ala	0.48	leu	0.64	gln	0.29
leu	0.48	glu	0.49	his	0.49	asn	0.64	ser	0.30
ala	0.49	ala	0.57	asn	0.51	lys	0.73	asn	0.32
gly	0.92	asp	0.87	asp	0.89	arg	0.75	ala	0.38
asp	1.0	gly	1.0	gly	1.0	gly	1.0	pro	0.59
pro		pro		pro		pro		gly	1.0

*in kcal/mol

Table I.6: The Normalized β-Sheet Propensities of Table I.4.

Munoz & Serrano		Kabsch & Sander*		Chou & Fasman (1978)		Fasman (1989)	
val	0.00	val	0.00	val	0.00	val	0.00
ile	0.060	ile	0.051	ile	0.075	ile	0.062
tyr	0.12	tyr	0.17	tyr	0.17	thr	0.27
thr	0.13	phe	0.19	phe	0.24	tyr	0.29
cys	0.13	cys	0.26	trp	0.25	trp	0.35
phe	0.15	trp	0.29	leu	0.30	phe	0.36
trp	0.24	thr	0.33	cys	0.38	leu	0.42
his	0.29	leu	0.34	thr	0.38	cys	0.50
met	0.30	met	0.37	gln	0.45	met	0.56
leu	0.34	his	0.50	met	0.49	gln	0.57
arg	0.35	arg	0.60	arg	0.58	arg	0.62
ser	0.35	ser	0.61	asn	0.61	ser	0.62
lys	0.41	lys	0.67	his	0.62	gly	0.68
gln	0.44	gln	0.70	ala	0.65	his	0.72
asn	0.49	ala	0.74	ser	0.71	ala	0.75
glu	0.54	glu	0.75	gly	0.71	lys	0.81
ala	0.56	gly	0.78	lys	0.72	asp	0.87
asp	0.62	asn	0.90	pro	0.86	asn	0.87
gly	1.0	asp	1.0	asp	0.87	pro	0.90
pro		pro		glu	1.0	glu	1.0

Table I.3 and Table I.4 show that the absolute propensity values and the spans of the propensity scales vary in the different studies. In order to obtain fair averaging and a better comparison of plots, the propensity data from each of the studies were normalized (Munoz and Serrano, 1994). The $\Delta\Delta G$ for each amino acid was subtracted from the most stable $\Delta\Delta G$, then divided by the $\Delta\Delta G$ difference between the most stable $\Delta\Delta G$ and the least stable $\Delta\Delta G$ (excluding Pro). Thus, the data from each of the studies was normalized so that the propensity scale ranged from 0 (most stable) to 1 (least stable). The normalized data are listed in Table I.5 and Table I.6.

In Figure I.11, the average of the normalized experimental data is plotted against the normalized values of the most recent statistical data by Munoz and Serrano. A very good correlation is found. In fact, the correlation factor is better than that obtained when the experimental and statistical data from a single host-guest system are plotted against each other (Munoz and Serrano, 1994). This indicates the noise of the propensity data from different experimental studies gradually cancels each other as the number of experiments using different host-guest proteins increases. These results suggest that intrinsic propensity is the dominant factor in β-sheet formation.

Some additional features can also be seen from this plot. Gly and Asp are well separated from the rest of the amino acids, suggesting that these residues are disfavored for β-sheet formation. (Pro, also disfavored, is not shown on the plot.) When viewed from the vertical (experimental) axis, Val, Ile, Tyr, Thr, and Phe are clustered together and separated from the other residues, suggesting that these five residues are favored for β-sheet formation. However, when viewed from the horizontal (statistical) axis, Cys is also included, since it is at the same level as Tyr, Thr, and Phe. This difference in propensity for Cys can be attributed to differences between the statistical and experimental studies. In the experimental studies, only solvent-exposed free Cys's were examined, while in the statistical study, both disulfide bonded Cys's and free Cys's were included,

regardless of whether they were buried or exposed. Since disulfide bonded Cys's are usually favored for β-sheet formation in native proteins, the enhanced propensity for Cys that was observed in the statistical study is as expected. Finally, if the least favored amino acids (Pro, Asp, and Gly) and the most favored amino acids (Thr, Val, Ile, Tyr, and Phe) are excluded from the plot, the correlation between the data is poor. This suggests that for the rest of the amino acids, context is the major determinant in β-sheet formation.

All the statistical studies show Val as the most favored for β-sheet formation; Ile is second, and either Thr or Tyr is third (see Table I.4 or I.6 or Figure I.11). The experimental studies (see Table I.3 or I.5 or Figure I.11), on the other hand, show no such consensus; they agree only in their ranking of Thr as the most favored and Pro as the least favored. These differences between the statistical and experimental studies are probably due to the fact that all the experimental studies measured the surface-exposed amino acid residues while the statistical studies surveyed both the surface-exposed and the buried residues. The hydrophobic residues Val, Ile, and Tyr are favored over Thr for the buried positions in a β-sheet. However, the opposite is true for the surface-exposed positions. Considering this, it is not surprising that the hydrophobic residues Val and Ile have a greater statistical preference than the hydrophilic residue Thr, and that Thr exhibits a greater experimental preference than Val and Ile. Nevertheless, as shown in Figure I.11, both the statistical and the experimental studies are consistent in showing that Thr, Val, Ile, Tyr, and Phe are all favored for β-sheet formation, and that Pro, Gly, and Asp are all disfavored. Among the highly favored amino acids, Thr, Ile, and Val, are all β-branched, and Tyr and Phe have a phenyl group at the β-carbon. Chothia found that in Ramanchandra plots, the dihedral angles allowed for β-branched residues are limited to the ψ - ϕ angles which are ideal for forming parallel and antiparallel β-sheets (Chothia, 1973). This is probably the main reason why the β-branched amino acids are always favored for β-sheet formation regardless of context.

In light of these findings, I divided the 20 amino acids into three groups. The highly favored ones (Thr, Val, Ile, Tyr, and Phe) and the disfavored ones (Pro, Gly, and Asp) were categorized as group I and group III respectively. The rest of the amino acids were placed into group II. If the amino acids between the different groups are interchanged, the stability order of the resulting proteins should remain the same (group I > group II > group III). Experimental results from early studies by Mutter and Altman (Mutter and Altman, 1985) and by Kemp (Kemp, 1990), and the recent study by Otzen and Fersht (Otzen and Fersht, 1995) support this notion. On the other hand, if the amino acids belonging to the same group are interchanged, the relative stability should be context dependent. The wild type amino acid residue will be more stable than other residues. Otzen and Fersht showed that this is true when they interchanged Thr with Val at various positions in chymotrypsin inhibitor 2 (Otzen and Fersht, 1995). The notion that context is more important in determining stability within a group is also supported by the poor correlation for the same group residues from different studies.

The propensity results from Minor and Kim's recent study (Minor Jr. and Kim, 1994b) illustrate an exception to the general rules described above. Although all the amino acids of group I except Ile are on their favored amino acids list, Ser and Glu from group II are surprisingly among the top five favored amino acids. In addition, in their order list, the positively charged residues Lys and Arg are also shifted relatively to lower order, and negatively charged residues Asp and Glu are shifted upward in the order. I interpreted this as an example of a case where the context overwhelms the intrinsic propensity. In this case, the electrostatic potential surface of the molecule is significantly perturbed. As shown in Figure I.12, the electrostatic potential surface around the guest site in the wild type protein G is very negative. When the mutation replaces the negative wild type Glu-42 and Asp-46 to neutral Ala, the electrostatic potential surface is severely perturbed. In order to

compensate for the depletion of negative charges, negatively charged residues Glu and Asp are favored, and positively charged residues Lys and Arg are disfavored in the guest site.

After analyzing all the data, I conclude that both intrinsic properties and local context contribute to β-sheet formation. The β-branched amino acids, Thr, Ile, and Val, and the aromatic amino acids, Tyr and Phe, are intrinsically favored, while Pro, Gly, and Asp are intrinsically disfavored for β-sheet formation regardless of context. If the 20 naturally occurring amino acids are divided into three groups, intrinsic properties determine the relative stability ordering among these groups, but context determines the ordering within a group.

Figure I.11. Correlation plot of the β-sheet propensities: the average of the normalized experimental data is plotted against the normalized values of the most recent statistical data obtained by Munoz and Serrano. A single letter code is used to label the amino acids.

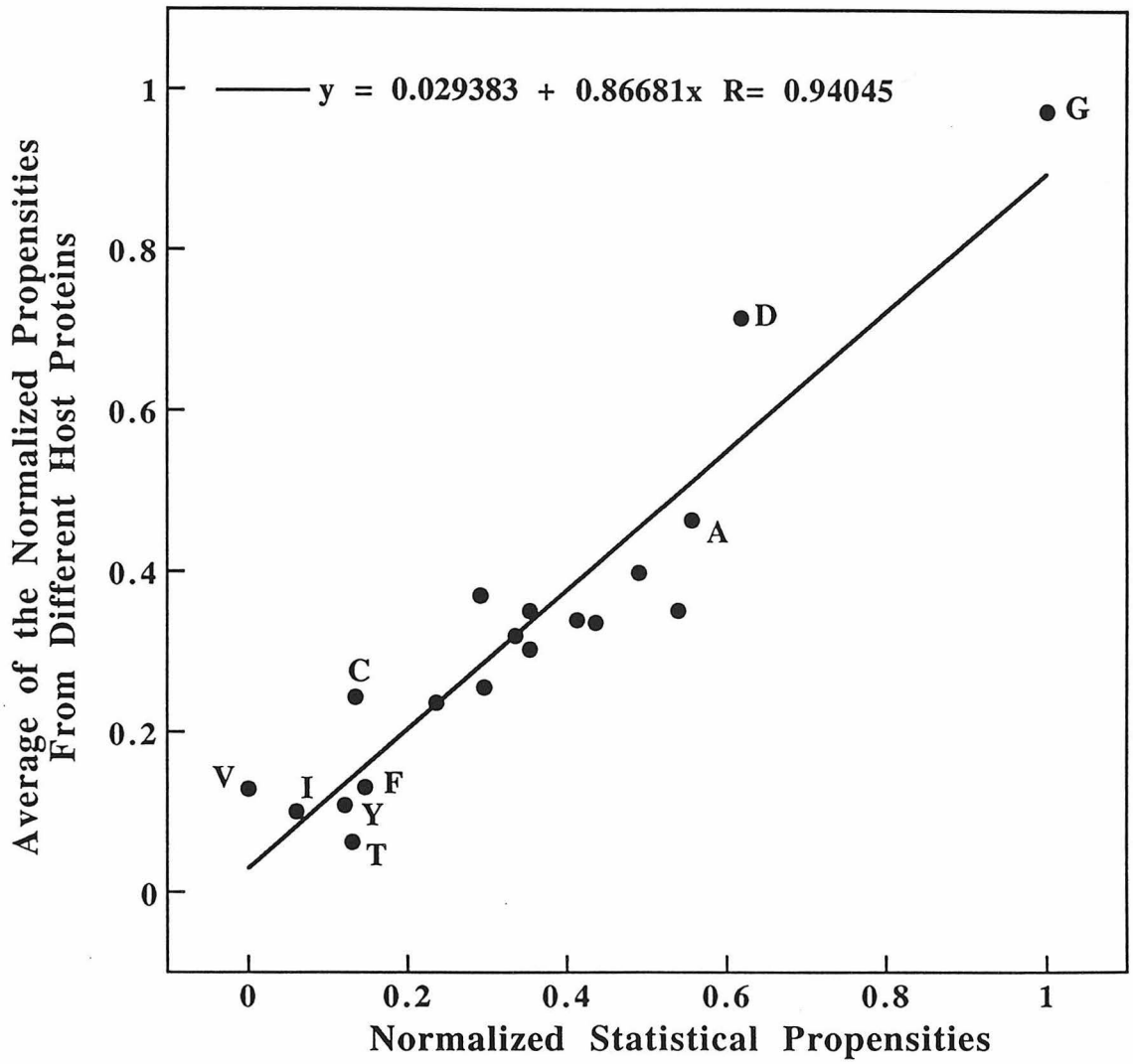
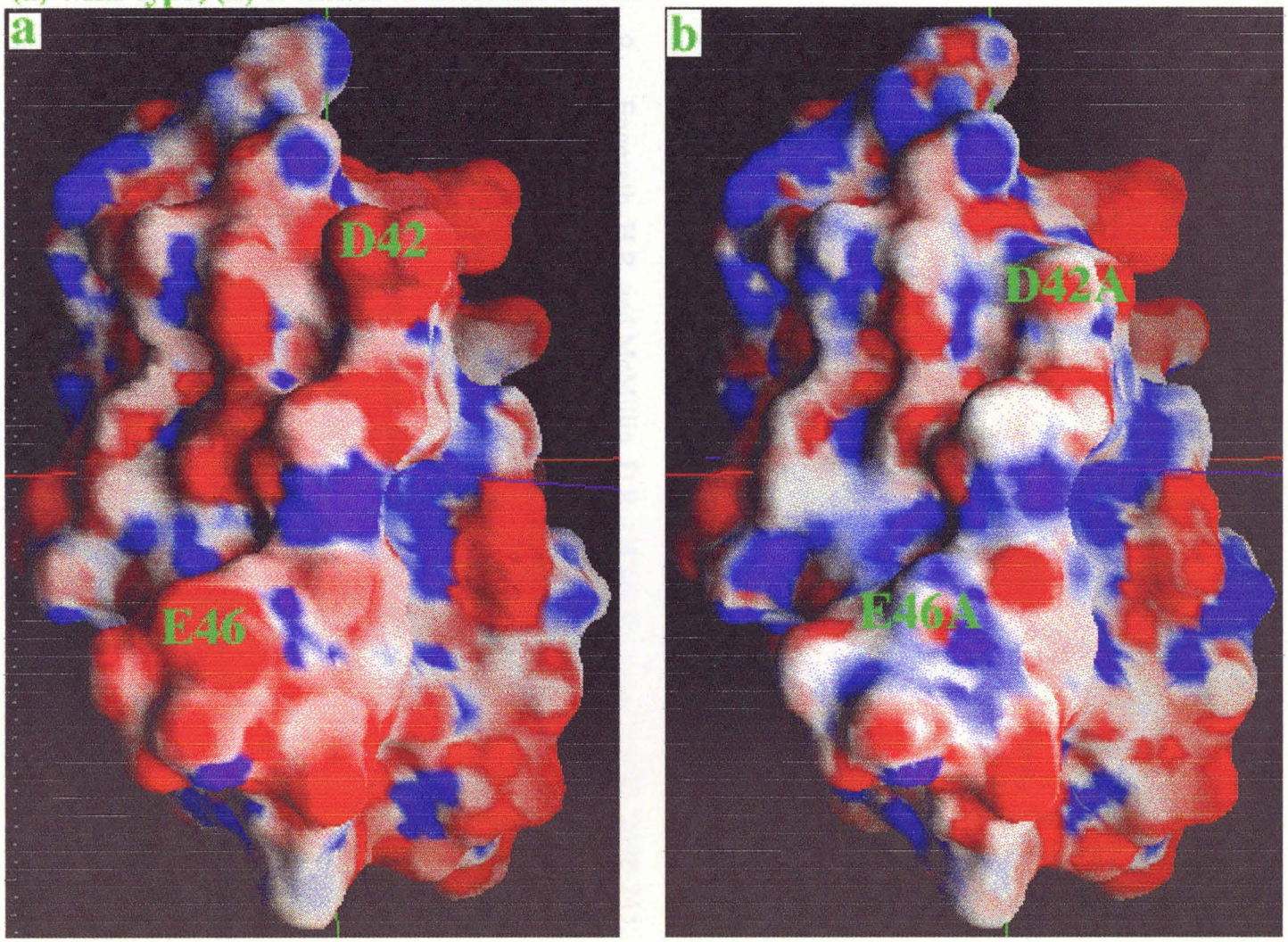

y = 0.029383 + 0.86681x R= 0.94045
Average of the Normalized Propensities From Different Host Proteins
Normalized Statistical Propensities
0
0.2
0.4
0.6
0.8
1
G
D
A
C
V
I
F
Y
T

Figure I.12. Color-coded electrostatic potential surface model for (a) the wild type IgG binding B1 domain of protein G and (b) its E42A/D46A double mutant. Red indicates negative potential and blue indicates positive potential. The electrostatic potential surface was calculated by GRASP (Biosym, Inc.) using DREIDING force field charges (Mayo *et al.* 1990).

Electrostatic potential surface of IgG binding domain of protein G.
(a) wild-type, (b) D42A/E46A double mutant



REFERENCES

1. Becktel, W.J. and Schellman, J.A. (1987) *Biopolymers* **26**, 1859-1877.

2. Blaszak, J.A., McMillin, D.R., Thorton, A.T., and Tennent, D.L. (1983) *J. Biol. Chem.* **258**, 9886-9892.

3. Chakrabartty, A., and Baldwin, R.L., in *Protein Folding: in vivo and in vitro*, J. L. Cleland Eds. (American Chemical Society, Washington, D.C, 1993) pp. 166-177.

4. Chang, T.K., Iverson, S.A., Rodrigues, C.G., Kiser, C.N., Lew, A.Y.C., Germanas, J.P., and Richards, J.H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1325-1329.

5. Chothia, C. (1973) *J. Mol. Biol.* **75**, 295-302.

6. Chou, P.Y. and Fasman, G.D. (1974) *Biochemistry* **13**, 211-222.

7. Chou, P.Y. and Fasman, G.D. (1978) *Adv. in Enzymol. Relat. Areas Mol. Biol.* **47**, 45-148.

8. CloneTech Laboratories, *Transformer Site-Directed Mutagenesis Kit (2nd version)* 1993.

9. Engeseth, H.R. and McMillin, D.R. (1986) *Biochemistry* **25**, 2448-2455.

10. Fasman, G.D., in *Prediction of Protein Structure and the Principles of Protein Conformation*, G. D. Fasman Eds. (Plenum, New York, 1989).

11. Flory, P.J. *Statistical Mechanics of Chain Molecules* (Oxford University Press, New York, 1969).

12. Kemp, D.S. (1990) *Trends Biotechnol.* **8**, 249-255.

13. Kim, C.A. and Berg, J.M. (1993) *Nature* **362**, 267-270.

14. Kunkel, T.A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 488-492.

15. Kunkel, T.A., Roberts, J.D., and Zakour, R.A. (1987) *Methods in Enzymol.* **154**, 367-382.

16. Leach, S.J., Nemethy, G., and Scheraga, H.A. (1966) *Biopolymers* **4**, 369.

17. Luo, J.-Y., Langen, R., and Mayo, S.L. *(unpublished results).*

18. Luo, J.-Y., Langen, R., Richards, J.H., and Mayo, S.L. *(to be published).*

19. Mayo, S.L., Olafson, B.D., and Goddard, W.A. (1990) *J. Phy. Chem.* **94**, 8897-8909.

20. Minor Jr., D.L. and Kim, P.S. (1994a) *Nature* **367**, 660-663.

21. Minor Jr., D.L. and Kim, P.S. (1994b) *Nature* **371**, 264-267.

22. Munoz, V. and Serrano, L. (1994) *Proteins: Structure, Function, and Genetics* **20**, 301-311.

23. Mutter, M. and Altman, K.-H. (1985) *Int. J. Pept.* **26**, 373-380.

24. Otzen, D.E. and Fersht, A.R. (1995) *Biochemistry* **34**, 5718-5724.

25. Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. (1963) *J. Mol. Biol.* **7**, 95-99.

26. Ramakrishnan, C. and Ramachandran, G.N. (1965) *Biophys. J.* **5**, 909.

27. Sander, C., Scharf, M., and Schneider, R., in *Protein Engineering: A Practical Approach*, Rees, A.R., Sternberg, M.J.E., Wetzel R. Eds. (IRL Press, Oxford University Press, 1992).

28. Smith, C.K., Withka, J.M., and Regan, L. (1994) *Biochemistry* **33**, 5510-5517.

29. Sueki, M., Lee, S., Powers, S.P., Denton, J.B., Konishi, Y., and Scheraga, H.A. (1984) *Macomolecules* **17**, 148-155.

30. Wojcik, J., Altmann, K.-H., and Scheraga, H.A. (1990) *Biopolymers* **30**, 121-134.

Chapter II

Mutating Ser-34 in Apo-azurin to Thr and Val

(Evaluating the Intrinsic Propensity Contribution and Local Context Contribution for β-sheet formation)

SUMMARY

In the last chapter, the 20 naturally occurring amino acids were divided into three groups in terms of their preferences for β-sheet formation. Group I residues, Thr, Val, Ile, Tyr, and Phe are intrinsically favored for β-sheet formation, group III residues are intrinsically disfavored, and group II residues rank in between. In this chapter, I present the results of further studies on this subject. I wanted to see if I could increase protein stability by taking wild-type non group I residues in the middle of a β-sheet and mutating them to the favored group I residues. After searching several common small and single domain proteins in the Protein Structure Data Bank, Ser-34 at the terminal end of a β-strand in azurin was chosen as the guest site for the mutagenesis study. Ser-34 was mutated to the favored group I residues, Thr and Val. The relative stabilities of the wild-type protein and the mutate proteins were then measured by circular dichroism. The thermal unfolding transition temperatures were 66.2 °C, 67.2 °C, and 63.5 °C for the wild-type apo-azurin, Thr-34 mutant, and Val-34 mutant respectively. The free energies of unfolding were also estimated. The Thr-34 mutant was 0.35 kcal/mol more stable than the wild-type protein, while the Val-34 mutant was 0.95 kcal/mol less stable. These results are discussed and the relative importance of local context vs. intrinsic propensity in β-sheet formation is evaluated.

INTRODUCTION

In Chapter I, I found that intrinsic secondary structure preference is the primary determinant for β-sheet propensity. The 20 naturally occurring amino acids were therefore divided into three groups. Group I residues (Thr, Val, Ile, Tyr, and Phe) are favored for β-sheet formation; group III residues (Pro, Asp, and Gly) are intrinsically disfavored; and group II residues rank in the middle of the other two groups. These results lead to a challenging question. Can we make a protein more stable than the wild-type by replacing a non group I wild-type residue in a β-sheet with a group I residue?

In order to answer this question and to further evaluate the contribution of local context to β-sheet formation, I performed the following study. First, I searched several common proteins for a non group I wild-type residue in the middle of a β-sheet. The following criteria had to be met: 1) the site had to be solvent exposed, 2) the residue had to be a small and neutral group II or group III residue, 3) there could be no side-chain H-bonding or salt bridge interactions to neighboring residues, and 4) the residue had to be in the middle of a β-sheet (i.e., in the middle of a central β-strand). Given these criteria, only Ala, Ser, and Gly were considered acceptable target residues for the search.

As the statistics suggest, there were not many group II residues in the middle of a β-sheet which matched the above requirements, and there were even fewer group III residues. Not surprisingly, not a single Gly or Ala was found which matched the criteria in the proteins searched (IgG binding domain of protein G, ubiquitin, staphylococcus nuclease, and azurin). Ser-34 in azurin was the only residue found which met most of the requirements. It is a neutral group II residue, solvent exposed, with no side-chain H-bonding or salt bridges to neighboring residues (see Figure II.1). The closest neighboring residues are Asn-32, Pro-36, Lys-92, and Gln-8. Although Ser-34 is at the terminal end of a β-strand, it does have i±1 backbone NH H-bonding to neighboring strands (Figure II.2).

Ser-34 in azurin was therefore chosen as the guest site and was mutated to the favored group I residues, Thr and Val.

Figure II.1. Molscript drawing of Ser-34 and its neighboring residues on the crystal structure of apo-azurin.

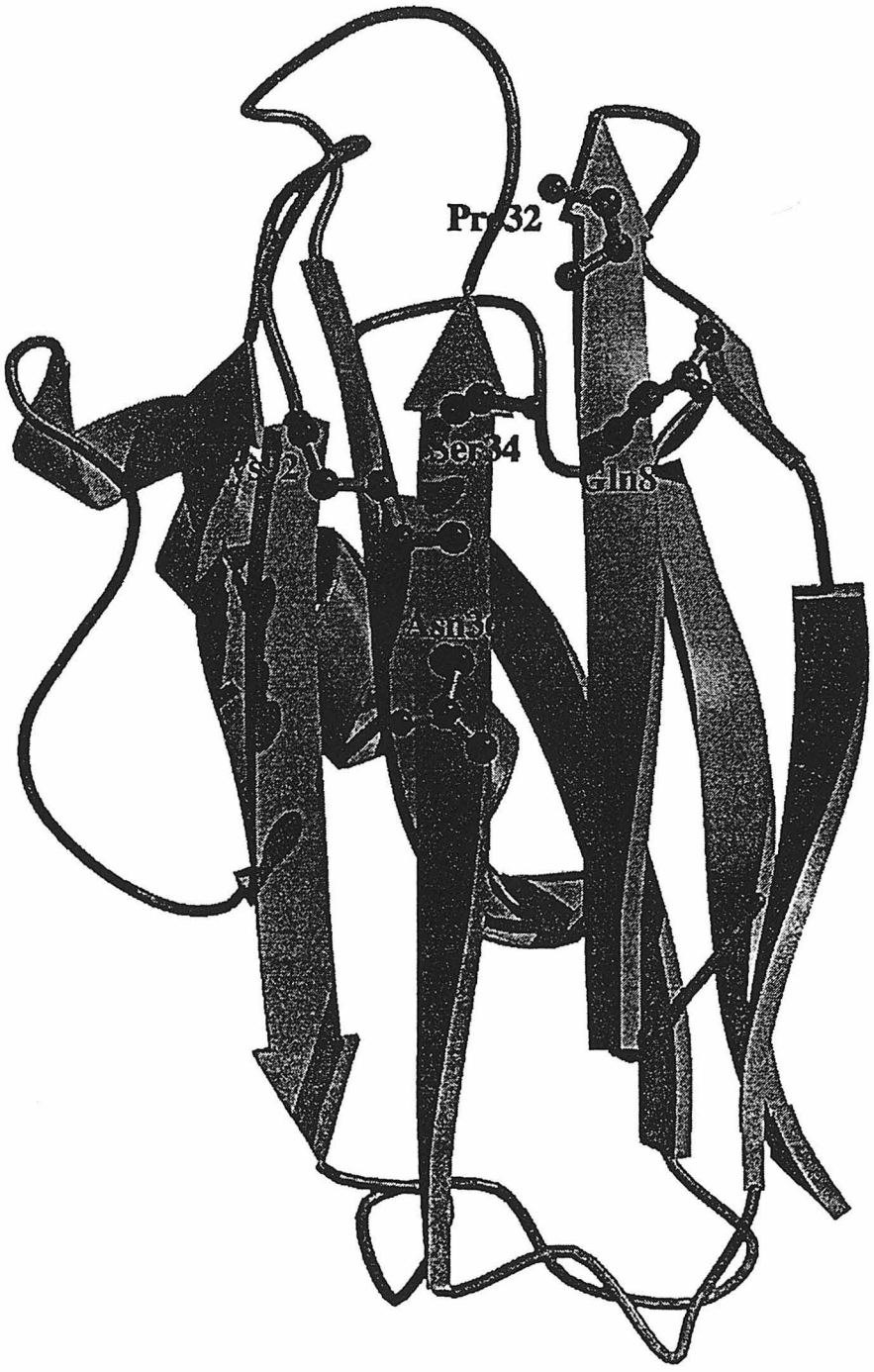
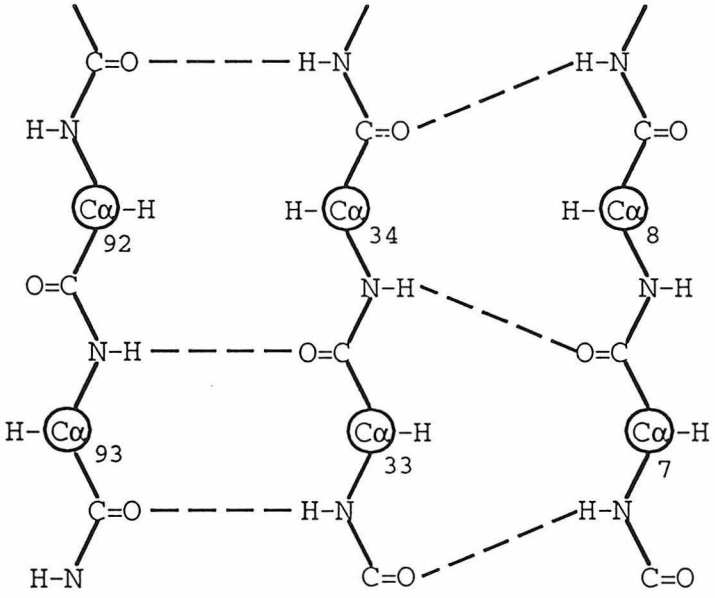

Pro32
Ser34
Gln8

Figure II.2. Close view of the neighboring backbone H-bonds (dash lines) around Ser-34 in azurin.


C=O
H-N
H-N
C=O
H-N
C=O
Cα-H
92
H-Cα
34
H-Cα
8
O=C
N-H
N-H
N-H
O=C
O=C
H-Cα
93
Cα-H
33
Cα-H
7
C=O
H-N
H-N
H-N
C=O
C=O

MATERIALS AND METHODS

The azurin gene cDNA (appendix I-4) shows two unique restriction digest sites close to Ser-34, *Bst EII* and *Hpa I* (see Figure II.3). Cassette mutagenesis method was therefore chosen to achieve the site-directed mutagenesis (McPherson, 1991).

```
31  32  33  34  35  36  37  38  39
Val Asn Leu Ser His Pro Gly Asn Leu
GTT AAC CTG TCT CAC CCA GGT AAC CTG
Hpa I                   BstE II
```

Figure II.3. The cDNA sequence nearby Ser34 (highlighted) and the two closest unique restriction sites.

ACT and GTC were used to code for Thr and Val respectively according to the codon usage in the previous gene construction. The four oligos shown below were ordered from Genosys Inc. for the mutagenesis experiments (see Figure II.4).

For Thr:

```
5' AAC CTG ACT CAC CCA G 3'          (34Tup)
3' TTG GAC TGA GTG GGT CCA TTG 5'    (34Tdown)
```

For Val:

```
5' AAC CTG GTC CAC CCA G 3'          (34Vup)
3' TTG GAC CAG GTG GGT CCA TTG 5'    (34Vdown)
```

Figure II.4. The four oligo's synthesized from Genosys Inc. (at 0.05 μmol scale). Labels at the end of each sequence are shown in parentheses.

The lyophilized oligos received from Genosys were dissolved in 200 μL of dH_2O and their concentrations were estimated using UV absorption at 260 nm. These oligo's were used directly without further purification in the subsequent reactions (5' phosphorylation, annealing, and ligation to the pET9a vector).

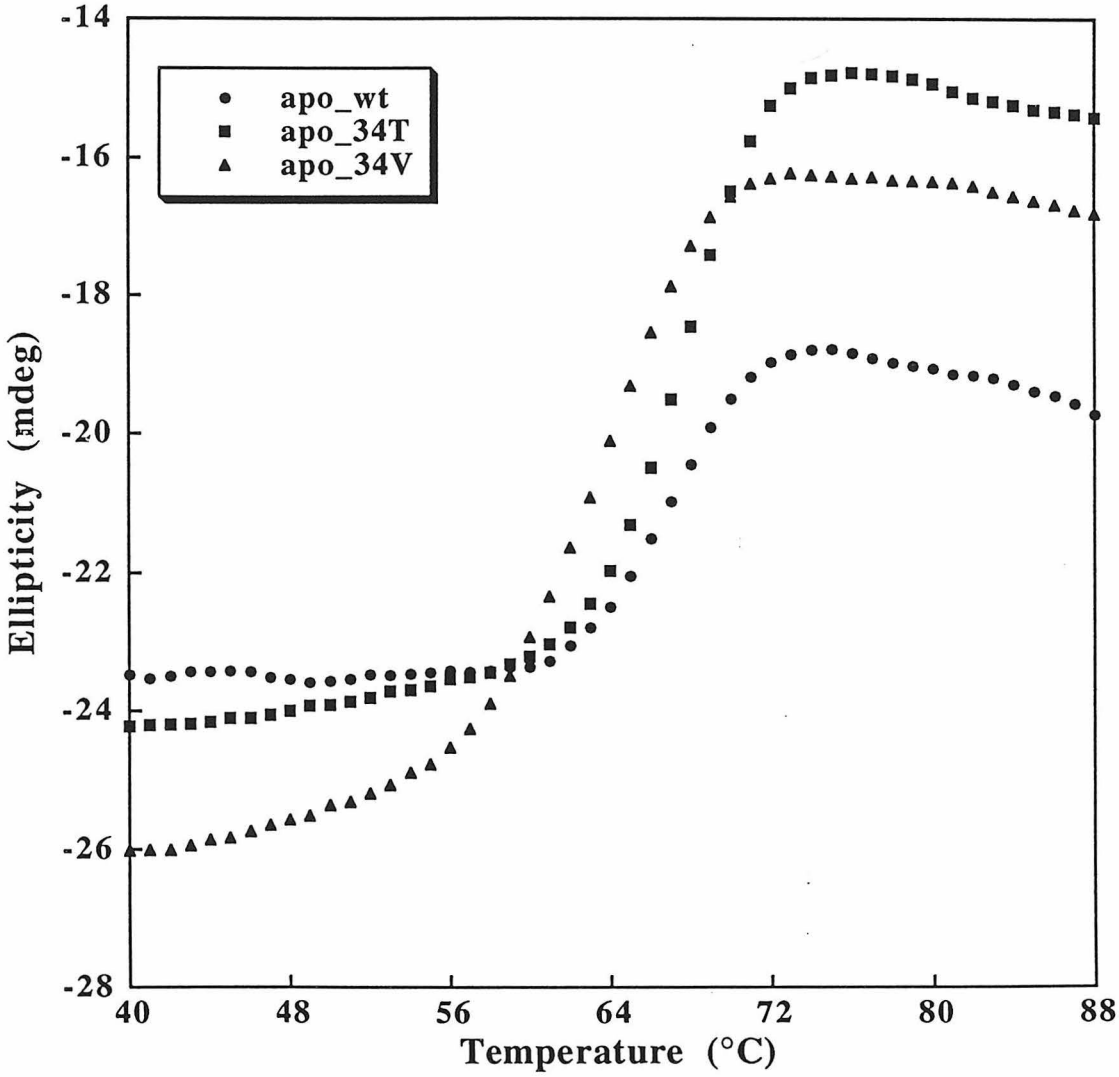
DNA reaction procedures were performed as described in the appendices (Appendices of Chapter I and II). The mutation sequences were identified by sequencing from both ends of the azurin gene. The positive mutant plasmids were transformed into the expression host cell BL21. Mutant proteins were overexpressed and purified using the same procedures as described in Chapter I.

RESULTS

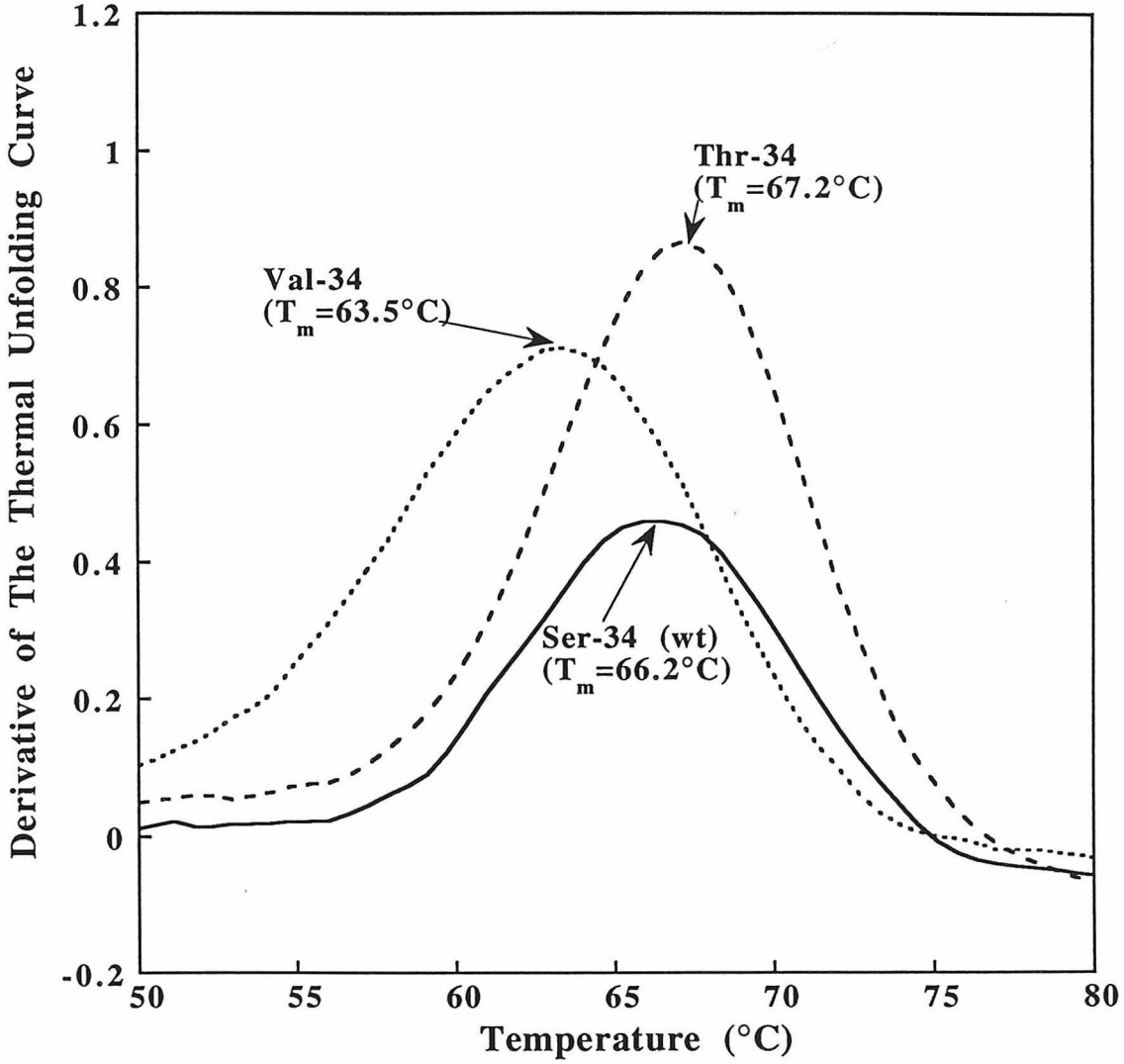
The thermal stability of holo-azurins and apo-azurins were measured by CD as described in Chapter I. In order to minimize the buffer difference between the wild-type and mutant proteins, the purified wild-type azurin from the previous study was run through the same purification scheme again as the newly prepared Thr-34 and Val-34 mutants. The same procedures were also used in preparing the mutant and the wild-type apo-azurins and in conducting the CD thermal unfolding measurements. The results are shown in Figure II.5. The T_m's were 67.2 °C and 63.5 °C for Thr-34 and Val-34 mutants respectively, compared to T_m of 66.2 °C for wild-type apo-azurin. In terms of free energy differences (Becktel and Schellman, 1987), Thr-34 was 0.35 kcal/mol more stable than the wild-type, and Val-34 was 0.95 kcal/mol less stable.

Figure II.5. (a) Thermal unfolding transitions of wild-type apo-azurin, Thr-34 mutant, and Val-34 mutant monitored by CD at 219 nm; (b) Smoothed derivative of the thermal unfolding curves in (a) for measuring the T_m.

(a)


apo_wt
apo_34T
apo_34V
-14
-16
-18
-20
-22
-24
-26
-28
Ellipticity (mdeg)
40
48
56
64
72
80
88
Temperature (°C)

(b)


Derivative of The Thermal Unfolding Curve
Temperature (°C)
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
50
55
60
65
70
75
80
Thr-34
(T_m=67.2°C)
Val-34
(T_m=63.5°C)
Ser-34 (wt)
(T_m=66.2°C)

DISCUSSION

This study shows that replacing Ser with Thr at position 34 makes apo-azurin more stable, but replacing it with Val does not. Initially, this result seems inconsistent with the conclusions of Chapter I, since both Thr and Val are intrinsically favored for β-sheet formation. If intrinsic propensity dominates β-sheet formation and enhancing β-sheet formation at position 34 also increases protein stability, then both Thr-34 and Val-34 mutants should be more stable than the wild-type. Thr-34 mutant was more stable, but Val-34 was not.

This may be explained, however, if we take the contribution of local context into account. As was pointed out in the introduction, position 34 is at the end of the β-strand (Figure II.1). The nearest neighbor residues are all very hydrophilic residues (Gln-8, Asn-32, Pro-36, and Lys-92). With such a hydrophilic local context, it is not surprising that a hydrophobic residue like Val is not favored at this position. In Val-34 mutant, the local context contribution favoring a hydrophilic residue overwhelms the intrinsic propensity contribution. On the other hand, Thr is hydrophilic, so that in the Thr-34 mutant, the local context contribution at position 34 is similar to that of the wild-type residue Ser. Therefore, in the Thr-34 mutant, the intrinsic propensity contribution still dominates, making Thr-34 more stable than the wild-type.

These results emphasize that, as pointed out in Chapter I, both intrinsic propensity and local context contribute to β-sheet formation. In most cases, where there are no significant local context preferences (like strong hydrophilic, hydrophobic, or electrostatic interactions), intrinsic propensity dominates the secondary structure formation of β-sheet. However, under extreme local context environments (like the host-guest site in this chapter and the host-guest site in Minor Jr. and Kim's second study), the local context contribution can overwhelm the intrinsic propensity contribution. Therefore, some caution should be

taken when evaluating the intrinsic propensity and local context contribution in β-sheet formation.

REFERENCES

1. Luo, J.Y., Langen, R., Richards, J.H., and Mayo, S.L. (to be published), or Chapter I and references therein.

2. Becktel, W.J. and Schellman, J.A. (1987) *Biopolymers* **26**, 1859-1877.

3. McPherson, M.J. (1991) "*Directed Mutagenesis, A Practical Approach*," IRL Press, Oxford.

4. Minor Jr., D.L. and Kim, P.S. (1994) *Nature* 367, 660-663.

Chapter III

Molten Globule State of *Pseudomonas aeruginosa* Apo-azurin

SUMMARY

I characterized a molten globule state of *Pseudomonas aeruginosa* apo-azurin, along with structurally perturbed states of holo-azurin at low pH. Studies using ANS binding and fluorescence measurement, circular dichroism, UV-Vis absorption, and ^{1}H-NMR spectroscopy indicate that apo-azurin is a molten globule at pH 2.9, while holo-azurins at pH's as low as 2.6 are only slightly perturbed. At pH 2.6, holo-azurins can bind the ANS, but retain native-like secondary and tertiary structures.

INTRODUCTION

It is widely believed that for single-domain proteins the folding reaction is usually a cooperative two-state reaction (Creighton, 1992). Fully cooperative two-state transitions provide no evidence on the mechanism of folding, which requires that the partially folded intermediates be identified and characterized. Although many kinetic folding studies report unstable transient intermediates in the folding pathway, their transient life-time prevents detailed studies on these kinetic intermediate states. The observation of an equilibrium intermediate state offers an alternate path for studying the folding reaction, although it has not yet been shown that an equilibrium intermediate state may be a kinetic intermediate state. Nevertheless, studying an equilibrium intermediate state always provides important and interesting insights into understanding how a particular secondary structure forms and how the protein folds.

An equilibrium intermediate state lacking native-like tertiary structure but having substantial secondary structure is usually called a molten globule state (Ptitsyn, 1992). Several native proteins have been shown to partially unfold to form equilibrium molten globules when exposed to specific environments (e.g., low pH or high anionic strength). The degree of partial unfolding of equilibrium molten globule states varies depending on the protein and the conditions. Highly unfolded to near native-like molten globule states have been found under different situations (Dolgikh *et al.*, 1985; Feng *et al.*, 1994; Redfield *et al.*, 1994; Ptitsyn, 1992), leading to the notion that proteins can exist in a wide range of partially denatured states (Redfield *et al.*, 1994).

It is generally believed that molten globule states are more commonly observed for α-helix proteins than for β-sheet proteins, since β-sheet formation requires the interaction of multiple neighboring strands. Most of the well-characterized molten globule states in the literature are for proteins with high percentages of α-helices in their structures. For

example, cytochrome *c* (Ohgushi and Wada, 1983), bovine carbonic anhydrase (Dolgikh *et al.*, 1984), α-lactalbumin (Dolgikh *et al.*, 1985), β-lactamase (Goto and Fink, 1989), interleukin-2 (Dryden and Weir, 1991), and apo-myoglobin (Barrick and Baldwin, 1993) are all α-helix dominated proteins. However, detailed secondary structure characterization reveals that the β-sheets present in these proteins are highly retained in their molten globule states. For example, Ptitsyn and coworkers found the α-helix and β-sheet contents to be about the same for the native and molten globule states of α-lactalbumin and bovine carbonic anhydrase B (Dolgikh *et al.*, 1984; Ptitisyn, 1992). Using infrared spectroscopy, they estimated that for α-lactalbumin the total content of α-helices differed from that of the native state by no more than 10%, and the β-sheet content differed by no more than 5% (Dolgikh *et al.*, 1985). These reports indicate that like α-helices, β-sheets can also retain their secondary structure in molten globule states without native-like rigid tertiary structure. This notion is further supported by recent reports of molten globule states of proteins with high β-sheet content, tumor-necrosis-factor (Hlodan and Pain, 1994) and spinach apo-plastocyanin (Tisi and Evans, 1993). Both of these proteins contain no α-helices.

Further studies on the equilibrium molten globule states formed by β-sheet dominated proteins would not only be helpful in understanding the mechanisms of folding for these proteins, but would also provide insight into β-sheet formation, which has not been as well studied and understood as α-helix formation. I report in this chapter my observation of a molten globule state of *Pseudomonas aeruginosa (pa)* apo-azurin, a β-sheet dominated protein, along with partially perturbed states of *pa* holo-azurins.

My interest in *pa* azurin was first motivated by the observation that although spinach apo-plastocyanin forms a molten globule state, the highly homologous pea apo-plastocyanin does not (Tisi and Evans, 1993). I wanted to find out whether the structurally homologous *pa* apo-azurin forms a molten globule state at low pH. All azurins and plastocyanins belong to the same protein family, type I blue copper proteins. Their

structures and biological functions are very similar. Their common structure can be described as an eight strand β-barrel. Azurin has an extra α-helix packed on one side of the β-barrel. The crystal structures of apo-azurin and apo-plastocyanin are also very similar to their holo counterparts, except for shifts in the copper ligands (Nar *et al.*, 1992b). I was also interested in *pa* azurin because an efficient expression system had been developed in *E. coli* (Chang *et al.*, 1991), and NMR assignments of reduced *pa* azurin were available (van de Kamp *et al.*, 1992).

MATERIALS AND METHODS

Holo-azurin and apo-azurin were prepared as described in Chapter I. The ammonium salt of ANS purchased from Sigma was re-crystallized before use in the ANS fluorescence measurements. All other materials were either reagent grade or ACS grade and were used directly.

CD spectra were recorded on a JASCO 600 or an AVIV 62A DS instrument at 22 °C unless otherwise specified. 1.0 mm and 10 mm pathlength cells were used for far-UV and near-UV CD measurements respectively. The protein concentrations for these measurements were 35 μM for far-UV and 135 μM for near UV. ANS fluorescence emission was measured at 480 nm with excitation at 400 nm using a SLM 8000C spectrofluorometer. Measurements were done at room temperature using a 1.0 cm pathlength cell. The protein concentration was 80 μM and the ANS concentration was 60 μM. UV-Vis absorption was measured on a Hewlett Packard 8452A Diode Array spectrophotometer. All protein concentrations were estimated by UV measurements at 280 nm with the extinction coefficient ε_{280} = 9,700 $M^{-1}cm^{-1}$ (Goldberg and Pecht, 1976). All pH measurements were made on an Accumet 10 with a glass electrode. Holo-azurin was reduced immediately before use (ascorbate was added at twice the protein concentration and it was kept at room temperature for about 30 minutes in a nitrogen environment).

NMR spectra were acquired on a Varian Unity plus 600 MHz instrument at 25 °C, and processed using Felix (Biosym, Inc.). The ^1H NMR spectra were acquired with 8 K complex points at 9640.9 Hz spectral width, zero-filled to 16 K complex points, and Fourier transformed without any apodization.

RESULTS

ANS Binding A standard technique for detecting the presence of partially unfolded proteins is a dye binding assay. Many dyes, such as 1-anilino-8-naphthalene sulfonic acid (ANS), are amphiphilic molecules — they can bind to hydrophobic-aqueous interfacial areas. These areas are usually not accessible for binding in native proteins in aqueous solution. However, when a native protein is perturbed (as in the molten globule state), some of its hydrophobic areas are exposed and dyes such as ANS can bind to it. This binding process causes a very significant change in the fluorescence emission spectroscopy of ANS. In aqueous solution, unbound ANS has a very low fluorescence intensity due to solvent quenching, but protein-bound ANS usually has significant fluorescence intensity due to shielding of the quenching pathways (Cantor and Schimmel, 1980). Thus, measurement of the fluorescence emission difference of ANS in the presence and absence of a protein has become a diagnostic test for the presence of perturbed protein states (Goto and Fink, 1989).

I employed this technique to determine the presence of partially perturbed states for azurin under different conditions. Three different forms of azurin were tested: oxidized holo-azurin, reduced holo-azurin, and apo-azurin. The results are shown in Figure III.1. There was very little change in the fluorescence intensity of ANS in the presence of the various proteins at pH 5.0. Similarly, little change in ANS fluorescence was observed in the presence of guanidine hydrochloride (GdnHCl) denatured protein. However, a significant increase in fluorescence intensity was observed at pH 2.6 for all forms of azurin and at pH 2.9 for apo-azurin as well, indicating the existence of protein-bound ANS and exposed hydrophobic areas in the proteins. ANS fluorescence intensity was enhanced three to six-fold for the oxidized holo-azurin, ten-fold for the reduced holo-azurin, and more than ten-fold for apo-azurin. The differences in the degree of enhancement may be

accounted for by two factors. First, the degree of partial perturbation may vary for the different protein forms; this could result in different degrees of ANS binding and shielding of the quenching pathways. Second, Cu(II) absorption may cause partial quenching of the ANS emission with oxidized holo-azurin, resulting in the relatively smaller enhancement seen for this form.

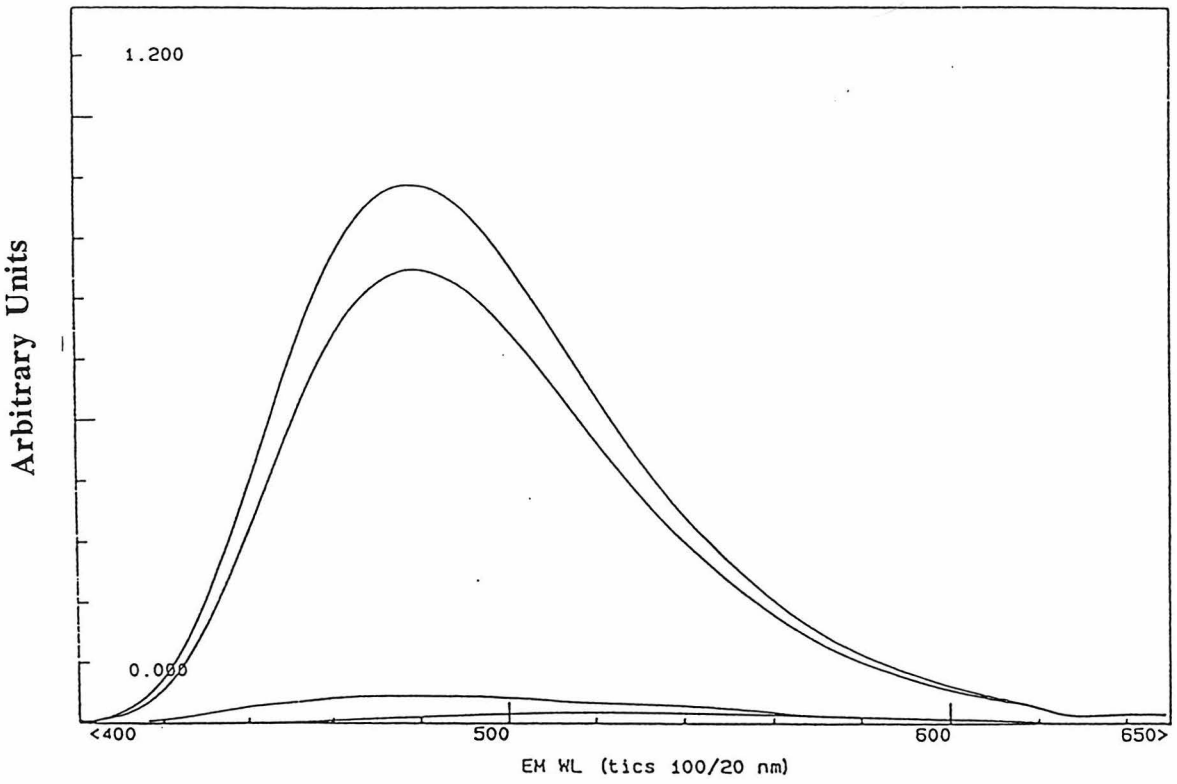
Circular Dichroism Measurements I also used circular dichroism (CD) to determine the extent of perturbation to the secondary structure and tertiary structure. Figures III.2 and III.4 show the CD measurement results. Figure III.2(a)(b)(c) shows the near-UV CD spectra of different forms of azurin at different pH's. Figure III.2a shows that there is a very substantial change for apo-azurin from pH 5.0 to pH 2.6. At pH 2.6, the near-UV signal is almost completely lost, like an unfolded state; while at pH 2.9, signals are still present, but significantly smaller than at pH 5.0. Figure III.2b and III.2c show a slight signal loss in the near-UV CD from pH 5.0 to pH 2.6 for both reduced and oxidized holo-azurin; there is more change for the reduced holo-azurin than for the oxidized form. The near-UV CD signal is due to absorption by aromatic chromophores in the vicinity of well-defined tertiary structures. The change in the near-UV CD signal therefore indicates a change in the tertiary structures around the aromatic chromophores. In *pa* azurin, there are a total of nine aromatic residues, one Trp (Trp-48), two Tyr's (Tyr-72 and Tyr-108), and six Phe's (Phe-15, 29, 97, 110, 111, and 114). Figure III.3 shows the distribution of these aromatic residues in azurin. The sharp positive peak around 292 nm is due to Trp-48, which is located in the center of the protein core. This peak decreases in intensity for apo-azurin from pH 5.0 to pH 2.9 and almost completely disappears at pH 2.6, but changes very little for both holo-azurin forms. This suggests that the packing environment of Trp is significantly perturbed for apo-azurin from pH 5.0 to pH 2.9 and is almost completely lost at pH 2.6, but is kept intact for both holo-azurins from pH 5.0 to

pH 2.6. The strong and broad negative peak from 274-280 nm is due to absorption by Phe and Tyr residues, which are distributed quite evenly around the protein. A change in the intensity of this peak indicates a change in the packing environment of Phe and Tyr residues or the whole protein. Apo-azurin shows a significant decrease in the intensity of this peak from pH 5.0 to 2.9, suggesting a significant loss of rigid packing for Phe and Tyr or for the whole protein. In contrast, small changes in this peak were observed for the holo-azurins, suggesting small perturbations in protein structure from pH 5.0 to 2.6. In summary, these results suggest that as the pH decreases different forms of azurin lose their well-defined tertiary structure to different extents. Apo-azurin shows a substantial loss of tertiary structure from pH 5.0 to pH 2.9, and almost a complete loss at pH 2.6, while the holo-counterparts are only slightly perturbed from pH 5.0 to pH 2.6.

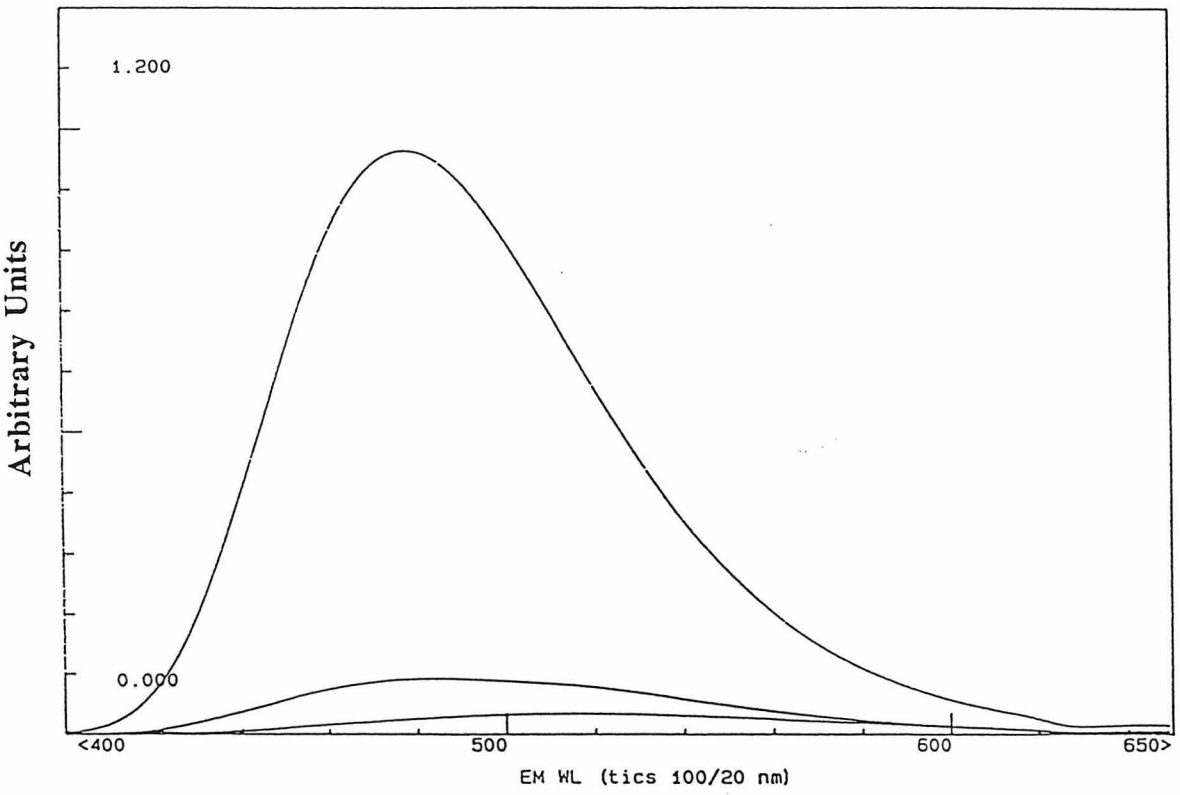
Relative to the near-UV CD signal described above, the far-UV CD signal results mainly from the secondary structures of a protein. Figure III.4 shows the far-UV CD measurements for different azurin forms at several pH's. There is little change for the holo-azurins as the pH changes from 5.0 to 2.6, indicating that the native secondary structures are retained at pH's as low as 2.6. In comparison, for apo-azurin, the far-UV CD signal is reduced at pH 2.9, and at pH 2.6, is about the same as the unfolded state. Considering both the far-UV and near-UV CD results, it is clear that apo-azurin is unfolded at pH 2.6. From pH 5.0 to 2.9, apo-azurin shows a signal loss in both the near-UV and far-UV CD spectra, but the well shaped CD signals at pH 2.9 indicates the presence of substantial secondary structures and partial tertiary structure.

Figure III.1. Fluorescence spectra of 60 μM ANS in the presence of 80 μM protein under different conditions. (a) from top to bottom: curves for apo-azurin at pH 2.9, pH 2.6, pH 5.0, and in the presence of 5.5 M GdnHCl; (b) from top to bottom: curves for reduced holo-azurin at pH 2.6, at pH 5.0, and in the presence of 5.5 M GdnHCl; (c) from top to bottom: curves for oxidized holo-azurin at pH 2.6, pH 5.0, in the presence of 5.5 M GdnHCl, and ANS in H_2O without protein. These spectra were measured with excitation at 400 nm and emission of 480 nm.

a



b



c

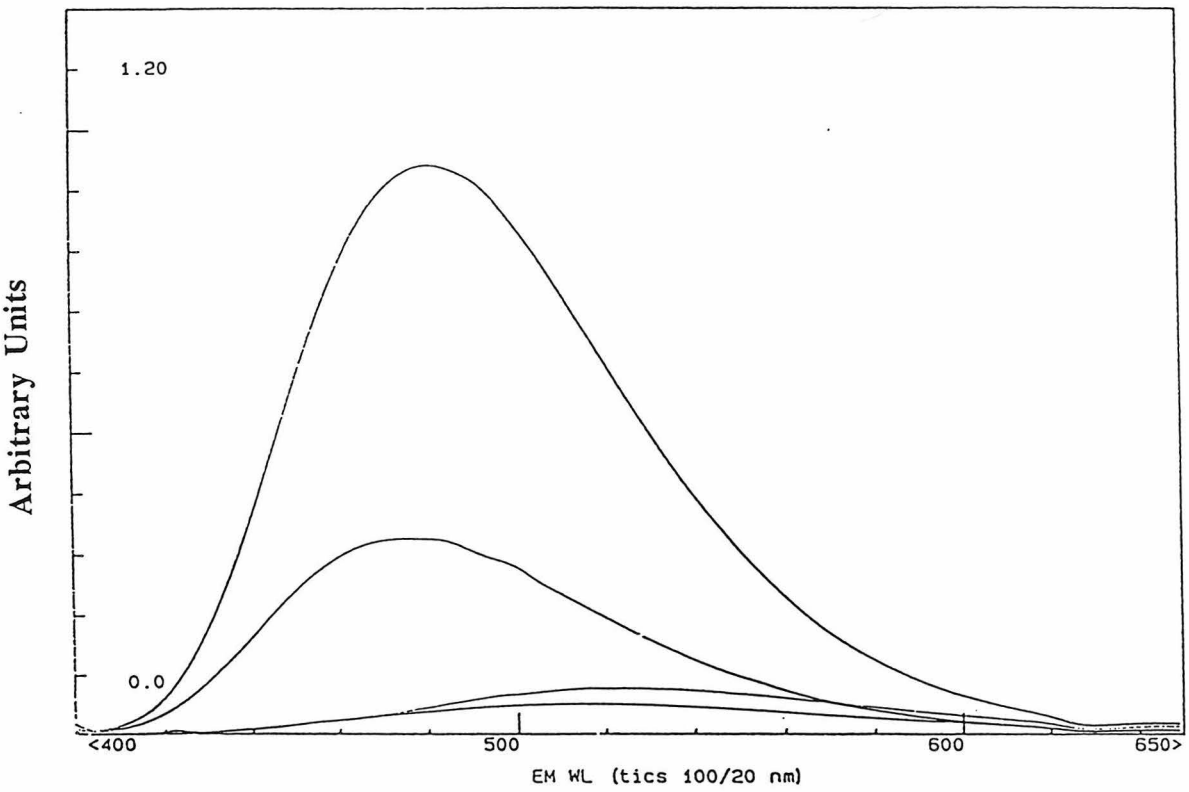
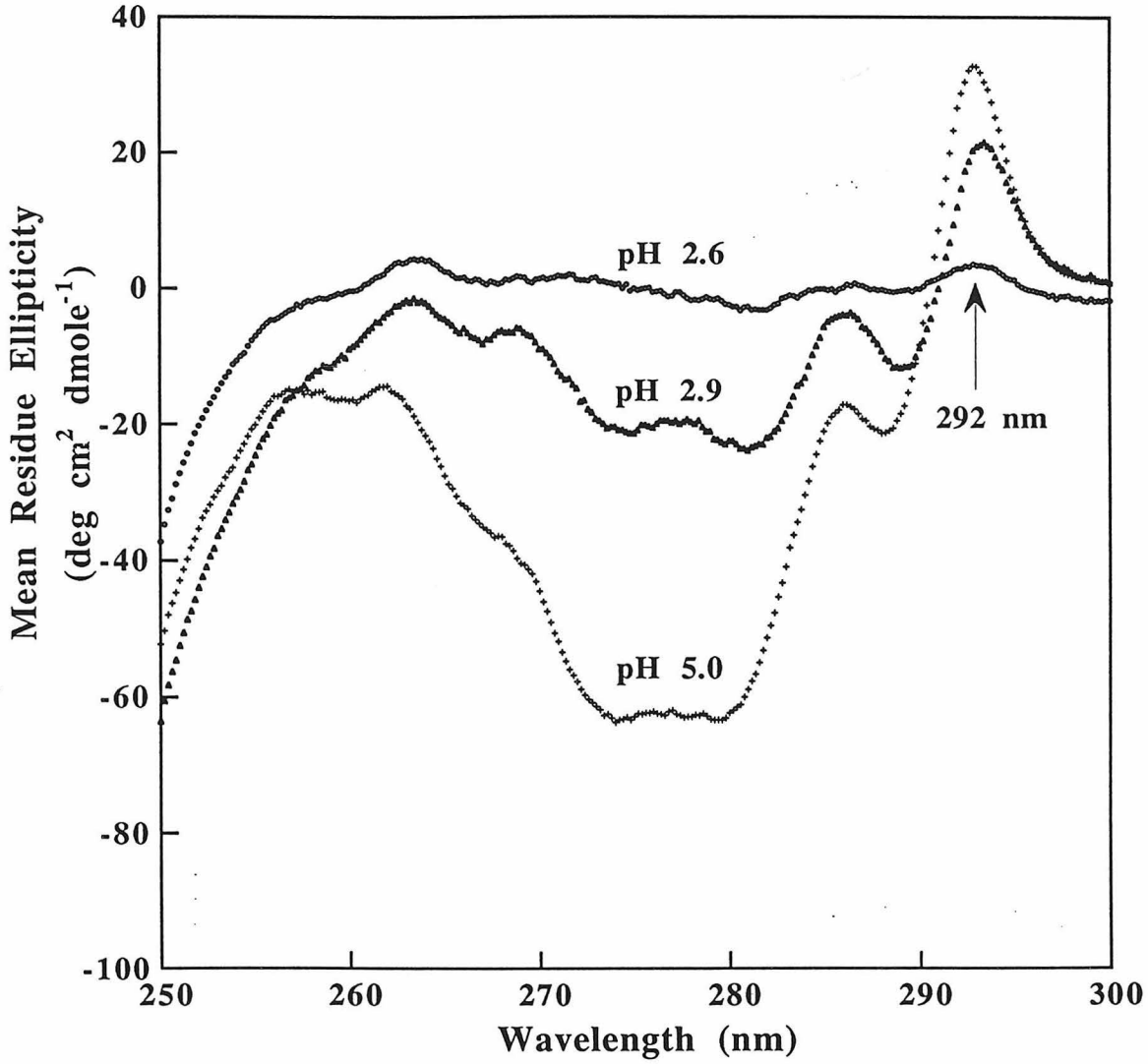
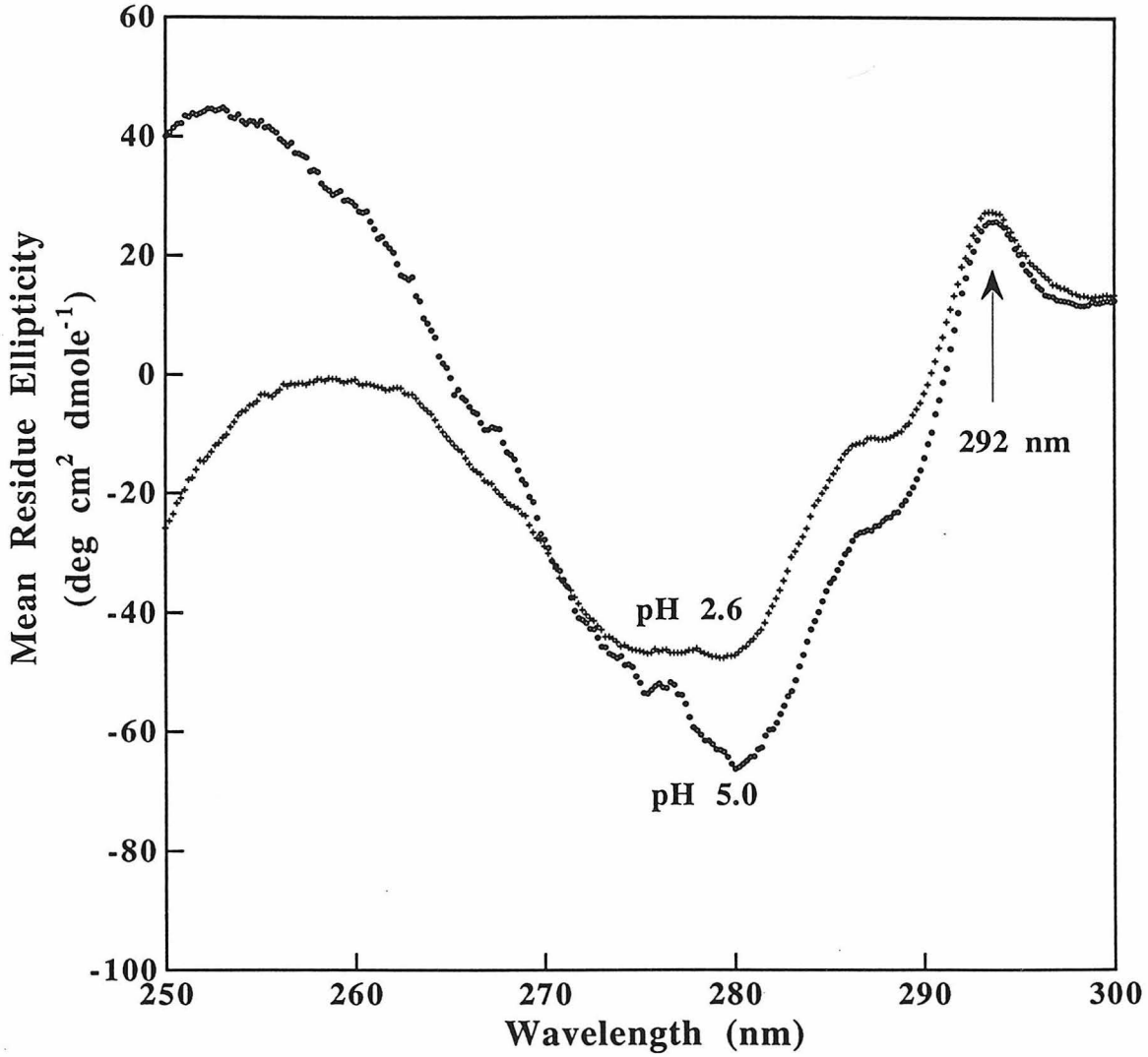


Figure III.2. Near-UV CD spectra of 145 μM protein, (a) apo-azurin at pH 5.0, pH 2.9, and pH 2.6; (b) reduced holo-azurin at pH 5.0 and pH 2.6; (c) oxidized holo-azurin at pH 5.0 and pH 2.6.

(a)


40
20
0
-20
-40
-60
-80
-100
250
260
270
280
290
300
Mean Residue Ellipticity
(deg cm^2 dmole^{-1})
Wavelength (nm)
pH 2.6
pH 2.9
pH 5.0
292 nm

(b)


60
40
20
0
-20
-40
-60
-80
-100
250
260
270
280
290
300
Mean Residue Ellipticity
(deg cm^2 dmole^{-1})
Wavelength (nm)
pH 2.6
pH 5.0
292 nm

(c)

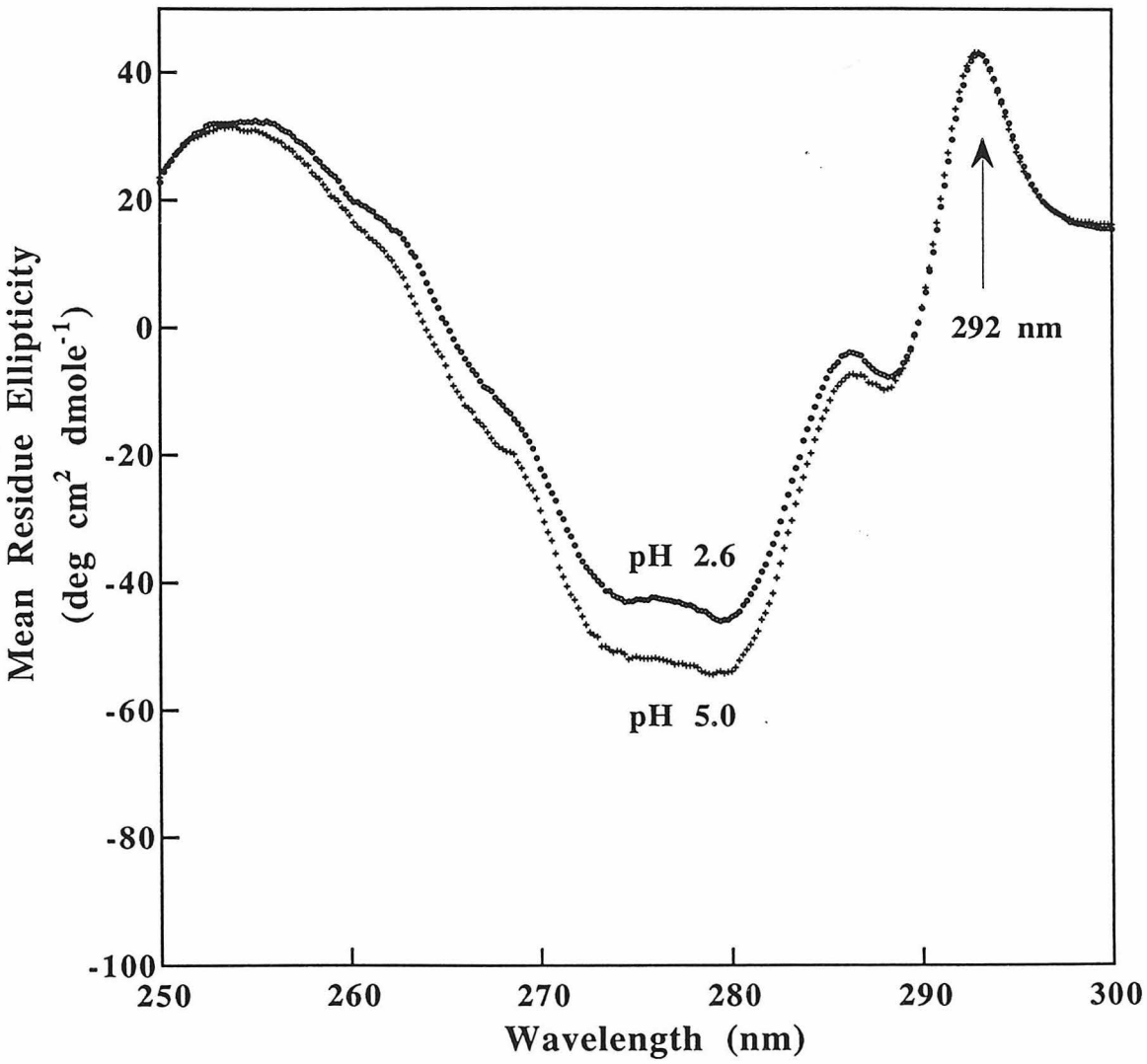

Mean Residue Ellipticity
(deg cm^2 dmole^{-1})
40
20
0
-20
-40
-60
-80
-100
250
260
270
280
290
300
Wavelength (nm)
292 nm
pH 2.6
pH 5.0

Figure III.3. Molscript drawing of holo-azurin crystal structure with side-chains of Trp, Phe, and Tyr residues highlighted.

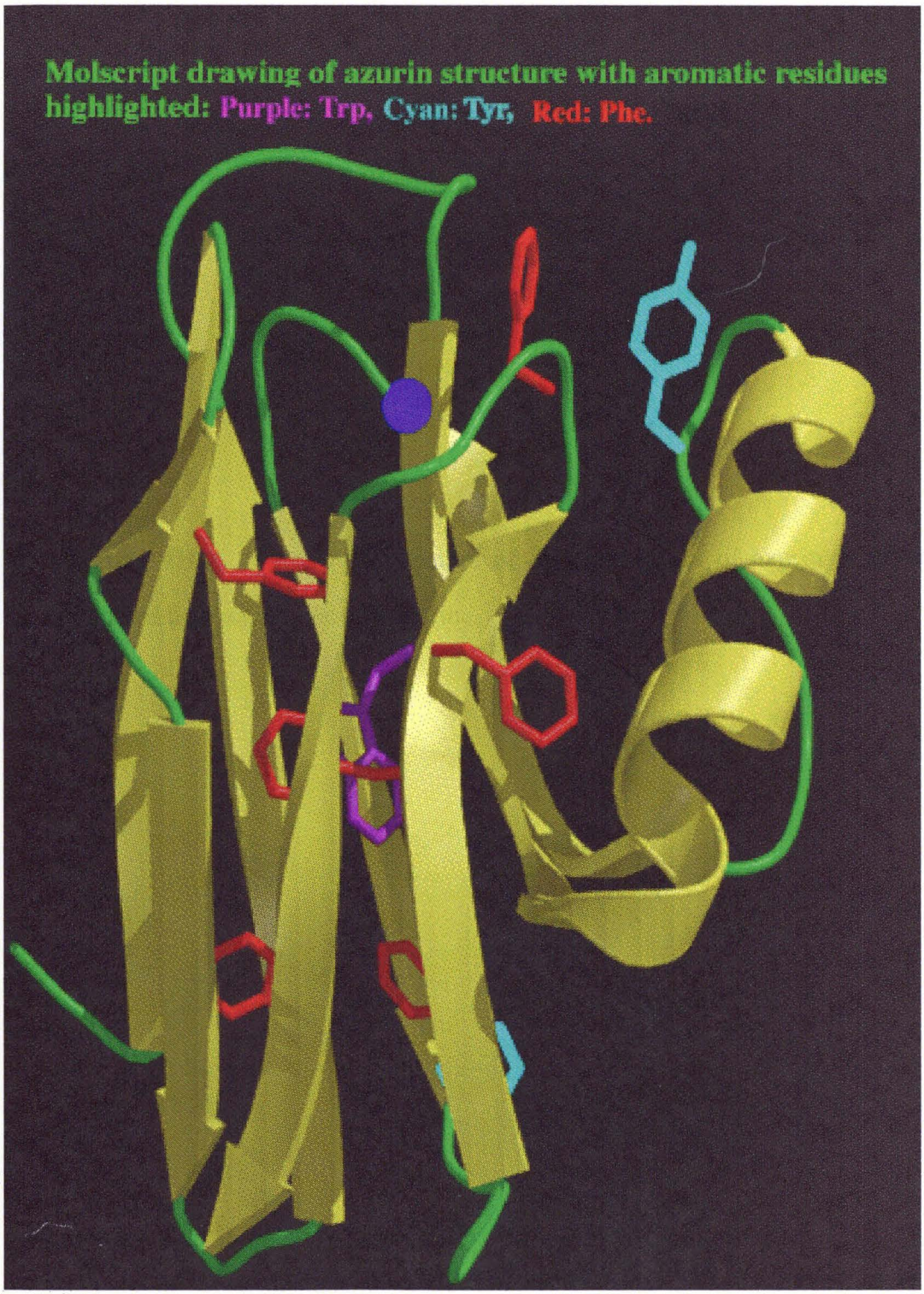
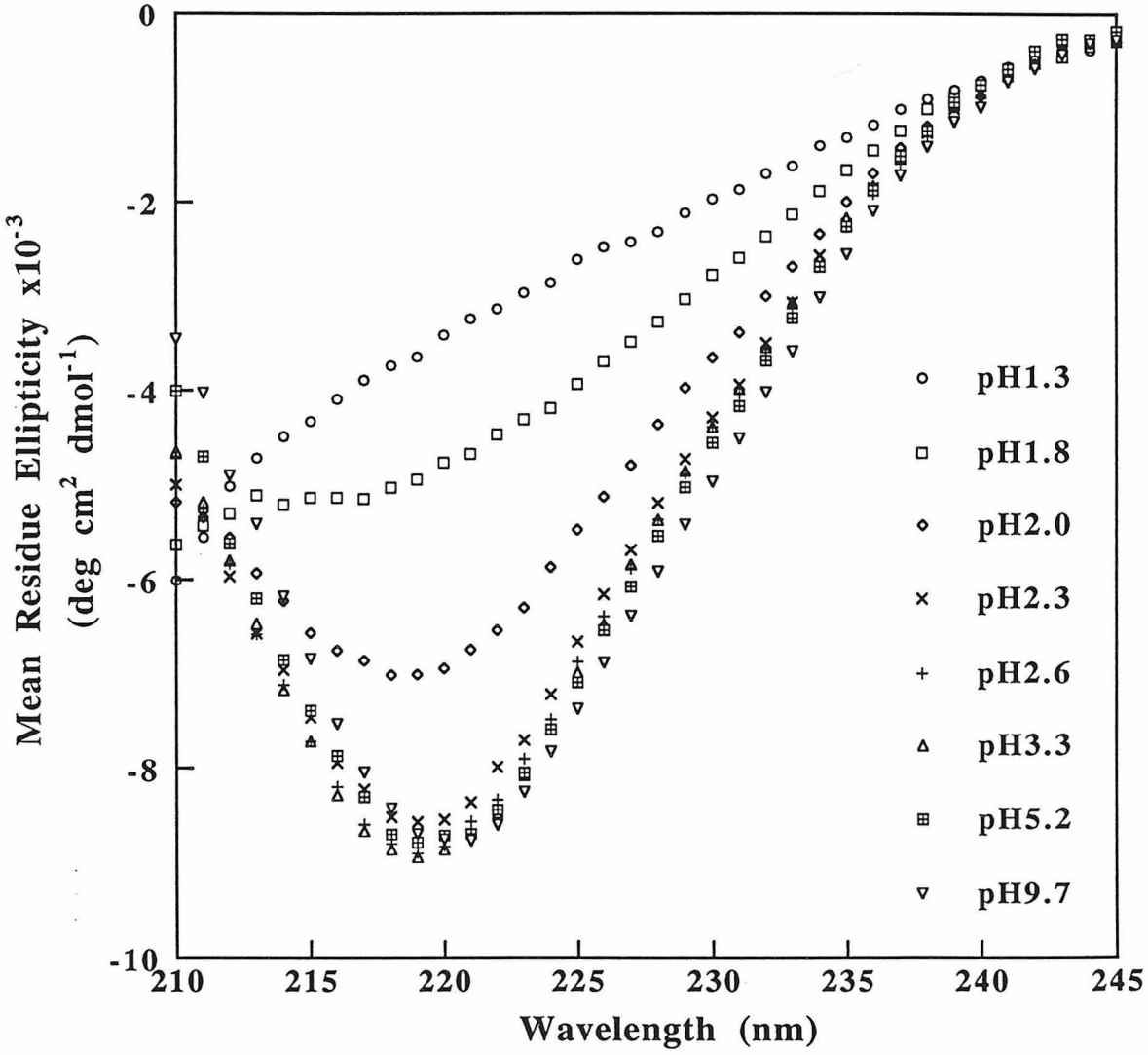
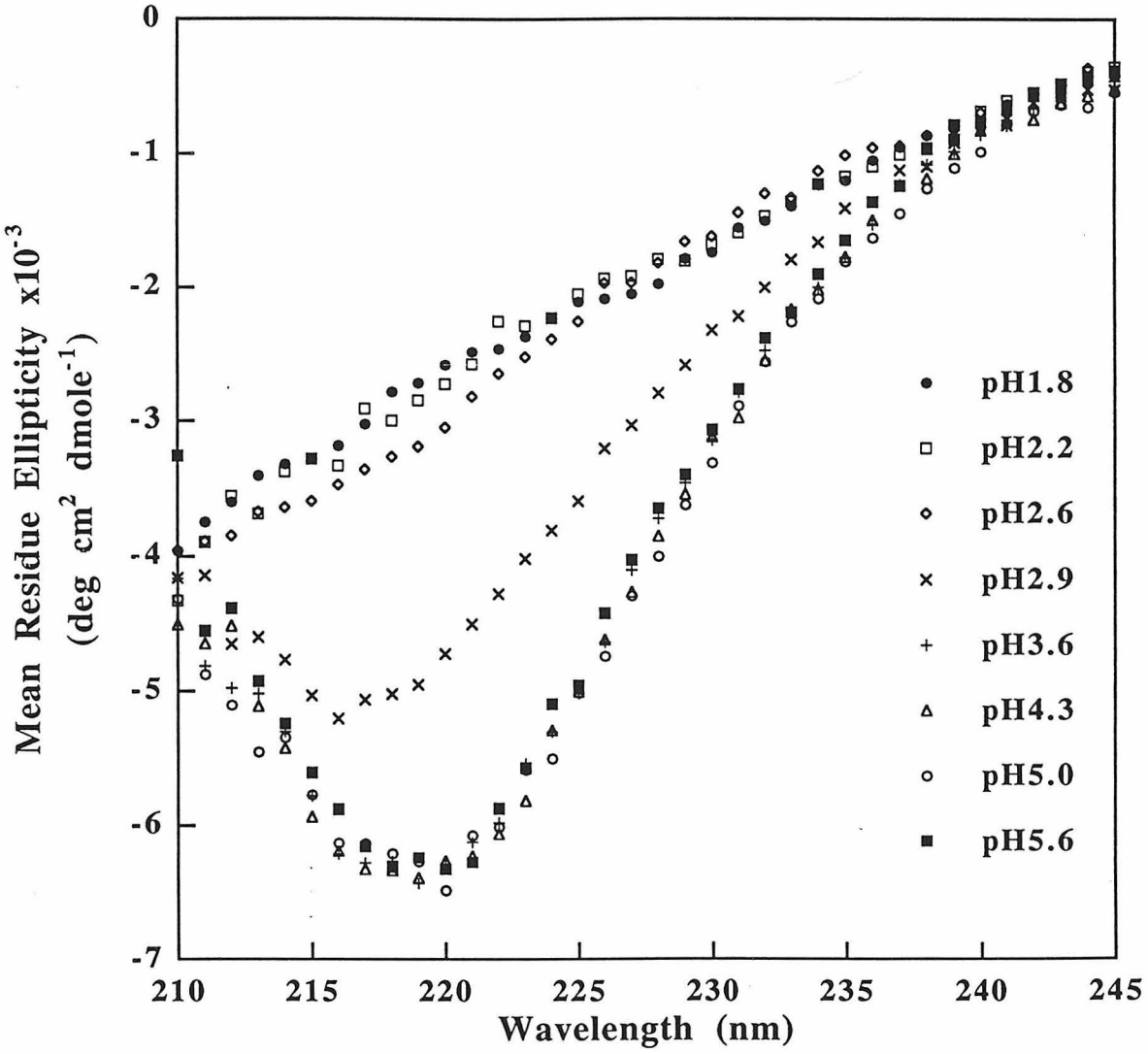

Molscript drawing of azurin structure with aromatic residues highlighted: Purple: Trp, Cyan: Tyr, Red: Phe.

Figure III.4. Far-UV CD spectra of different forms of azurin at several pH's, (a) oxidized holo-azurin; (b) apo-azurin.

(a)


Mean Residue Ellipticity x10-3
(deg cm2 dmol-1)
0
-2
-4
-6
-8
-10
210
215
220
225
230
235
240
245
Wavelength (nm)
pH1.3
pH1.8
pH2.0
pH2.3
pH2.6
pH3.3
pH5.2
pH9.7

(b)


Mean Residue Ellipticity x10^-3
(deg cm^2 dmole^-1)
0
-1
-2
-3
-4
-5
-6
-7
210
215
220
225
230
235
240
245
Wavelength (nm)
pH1.8
pH2.2
pH2.6
pH2.9
pH3.6
pH4.3
pH5.0
pH5.6

UV-Vis Absorption and ^1H NMR Measurements We observed that the protein solution of oxidized holo-azurin is still blue at pH 2.6, indicating the presence of the intact Cu center. Since the intense blue band absorption centered at 626 nm is due to the charge transfer of Cys-112 to Cu(II) in the unique ligand field in azurin (Solomon *et al.*, 1980), a change in this blue band absorption would indicate a change in the Cu binding environment. I used UV-Vis measurements to determine whether this blue band absorption changed from pH 5.0 to pH 2.6. Figure III.5 shows the results. Figures III.5a and III.5b show there is little change in the blue band absorption in terms of both intensity and wavelength from pH 5.0 to pH 2.6, suggesting little perturbation of the Cu(II) ligand field. The shoulder peak at 292 nm is believed to indicate the unique Trp-48 environment in *pa* azurin. The insets in Figure III.5a and III.5b show that this shoulder remains intact from pH 5.0 to pH 2.6 for holo-azurin, indicating that the packing environment in the vicinity of Trp-48 is barely changed from pH 5.0 to pH 2.6, as was suggested by the near-UV CD results. However, when the pH is reduced to 1.6, the acid denatured state, the absorption peak at 626 nm and the shoulder at 292 nm are both absent (Figure III.5c).

In contrast, UV-Vis measurements of apo-azurin produced different results (Figure III.6). The shoulder peak at 292 nm is present at pH 5.0 but diminishes at pH 2.9, suggesting the change or loss of the tertiary packing around Trp-48. These results are consistent with the near-UV CD results.

Nuclear Magnetic Resonance (NMR) spectroscopy was also used to qualitatively measure the side-chain flexibility of azurin under different conditions. Figure III.7 shows the ^{1}H-NMR spectra of reduced holo-azurin. The overall spectra for reduced holo-azurin at pH 5.0 and pH 2.6 are very similar, suggesting the native-like structure at pH 2.6. Figure III.8 shows the ^{1}H-NMR spectra of apo-azurin at different pH's. The spectra at pH 5.3 and pH 3.5 are very similar. At pH 2.9, however, the spectrum is much simpler than those

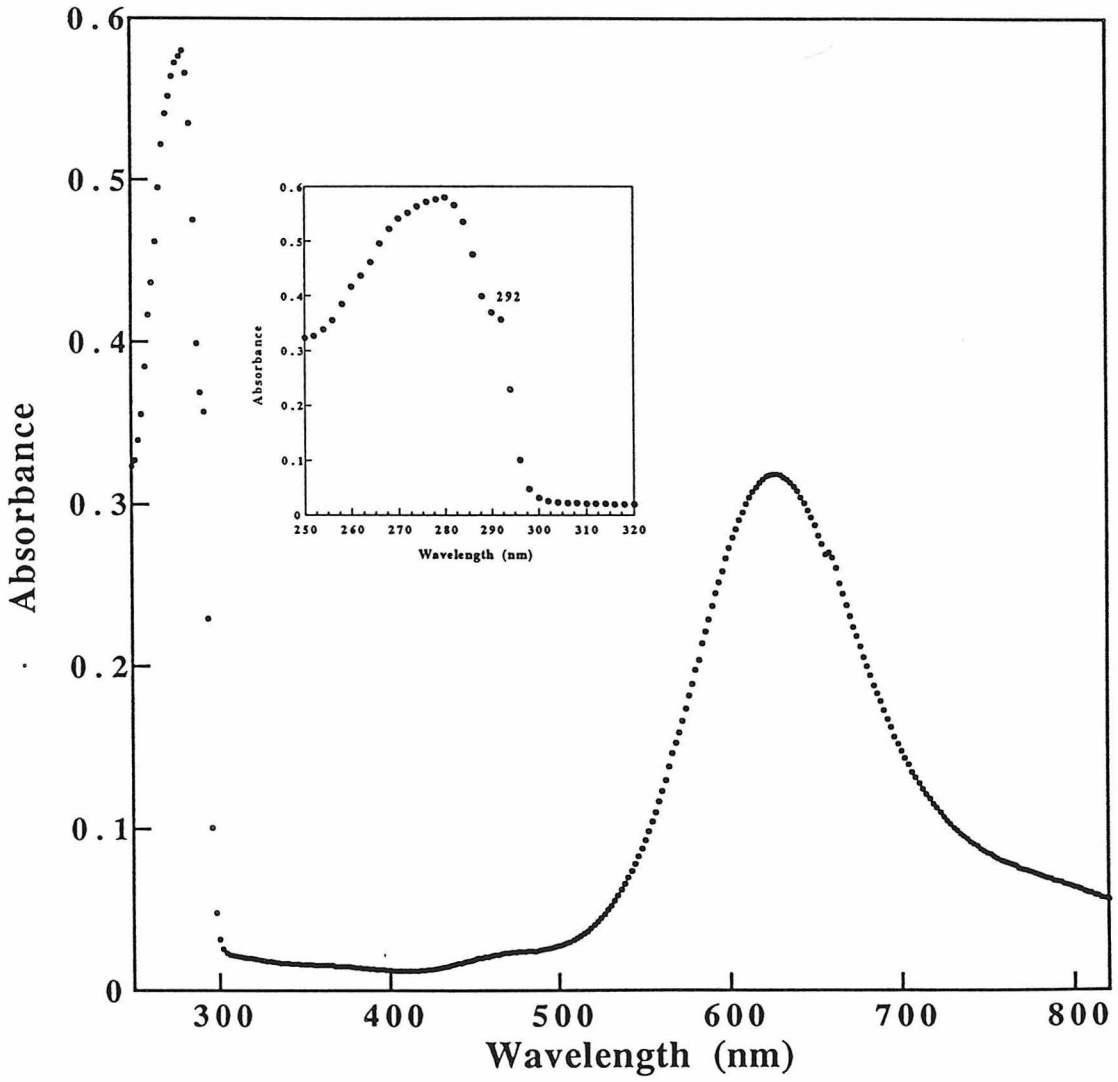
at pH 3.5 and pH 5.3, and is similar to the spectrum at pH 2.6, which resembles that of a totally unfolded state. This suggests that the rigid packing environment of the side-chains of apo-azurin is significantly perturbed at pH 2.9 and below (Ptitsyn, 1992).

Thermal Unfolding Finally, the relative stability of the different forms of azurin was determined using thermal unfolding measurements. Figure III.9 shows the results. Both oxidized and reduced holo-azurins have cooperative unfolding transitions at both pH 5.0 and pH 2.6. The T_m (mid-point unfolding transition temperature) values at pH 5.0 and pH 2.6 are 87 °C and 54 °C respectively for oxidized holo-azurin, and 84 °C and 48 °C respectively for reduced holo-azurin. In contrast, apo-azurin has a cooperative unfolding transition with a T_m of 68 °C at pH 5.0, but at pH 2.9 no longer shows a cooperative unfolding transition. This further supports the notion that apo-azurin is a molten globule state at pH 2.9.

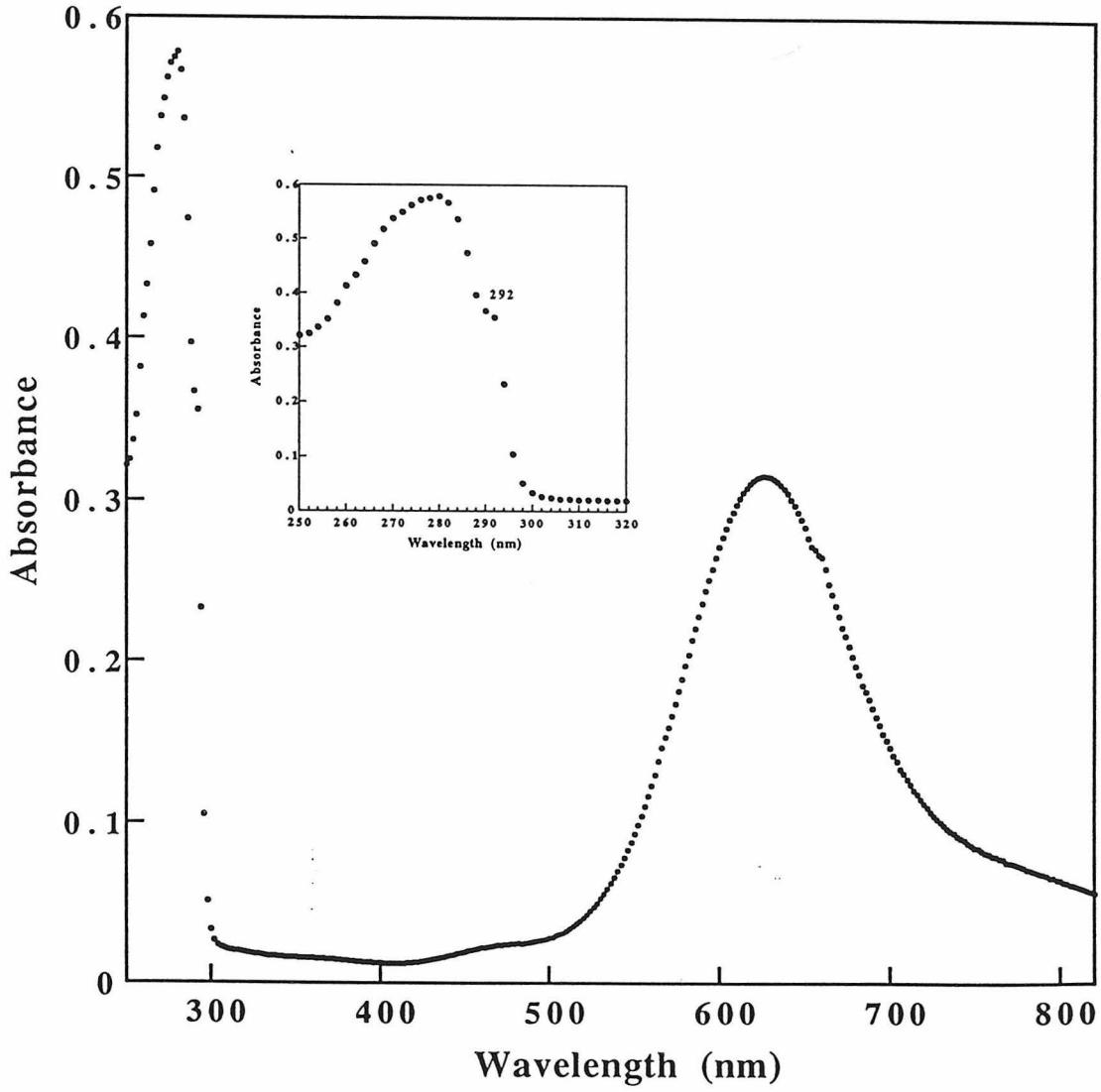
One extra note is that I only observed one thermal unfolding transition for apo-azurin at pH 5.5. This is in contrast to the two heat absorption peaks at 62 °C and 86 °C observed previously by Engeseth and McMillin in their calorimetry measurements (Engeseth and McMillin, 1986). I believe the presence of holo-azurin in their apo-azurin sample was the cause for the high temperature peak at 86 °C.

Figure III.5. UV-Vis absorption spectra of oxidized holo-azurin at different pH's. The insets show an expanded view of the spectra around 292 nm. (a) pH 5.0; (b) pH 2.6; (c) pH 1.6.

(a)


Absorbance
0.6
0.5
0.4
0.3
0.2
0.1
0
300
400
500
600
700
800
Wavelength (nm)
292
Absorbance
Wavelength (nm)

(b)


Absorbance
Wavelength (nm)
0.6
0.5
0.4
0.3
0.2
0.1
0
300
400
500
600
700
800
292
Absorbance
Wavelength (nm)
250
260
270
280
290
300
310
320

(c)

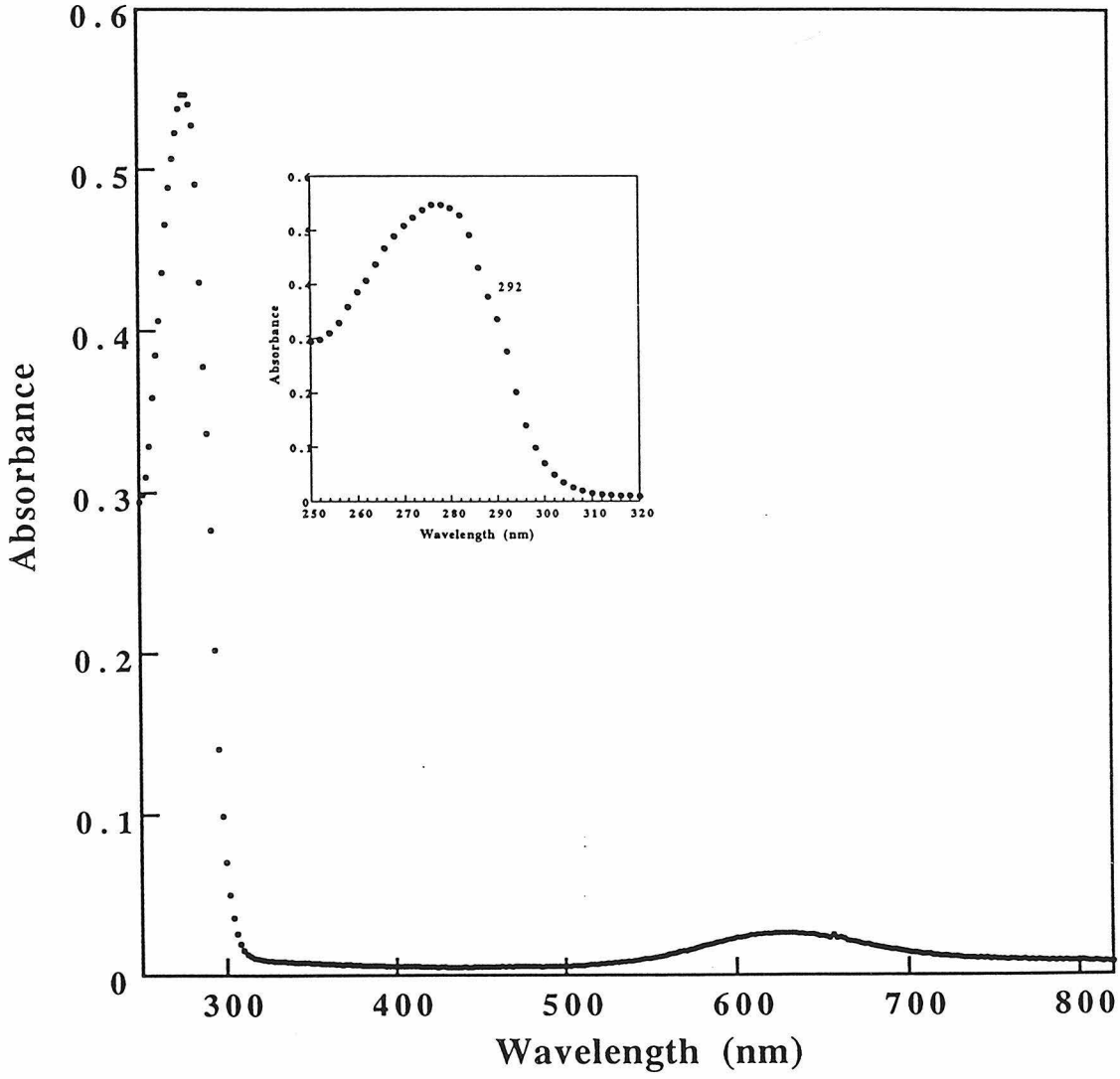
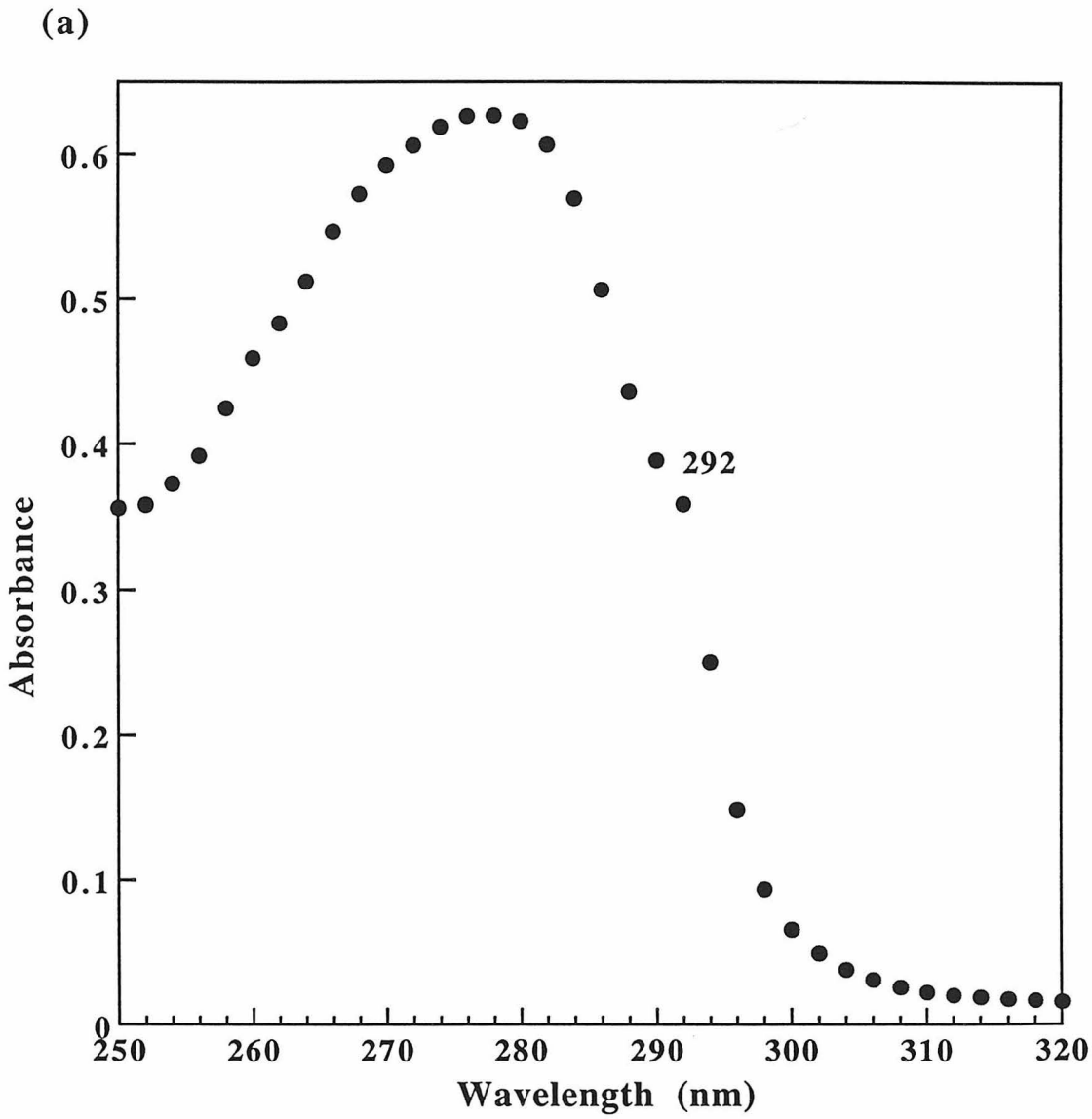

Absorbance
0.6
0.5
0.4
0.3
0.2
0.1
0
300
400
500
600
700
800
Wavelength (nm)
292
Absorbance
Wavelength (nm)

Figure III.6. UV-Vis absorption spectra around 292 nm of apo-azurin at (a) pH 5.0 and (b) pH 2.9.


(a)
Absorbance
0.6
0.5
0.4
0.3
0.2
0.1
0
292
250
260
270
280
290
300
310
320
Wavelength (nm)

(b)

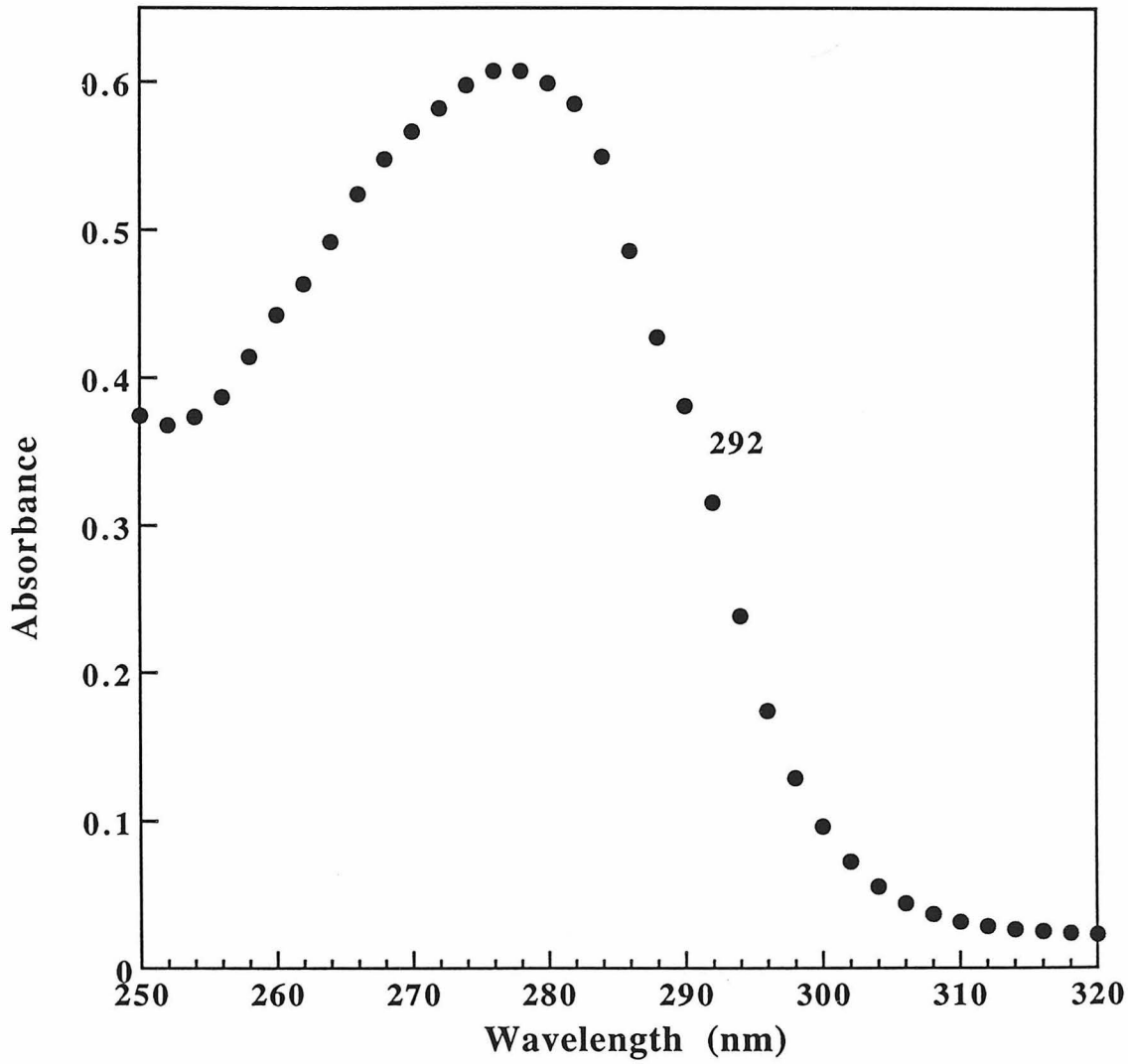

0.6
0.5
0.4
0.3
0.2
0.1
0
Absorbance
292
250
260
270
280
290
300
310
320
Wavelength (nm)

Figure III.7. ^{1}H-NMR spectra of the reduced holo-azurin at different pH's. From top to bottom: pH 5.0, pH 2.6, and pH 1.2 (totally unfolded).

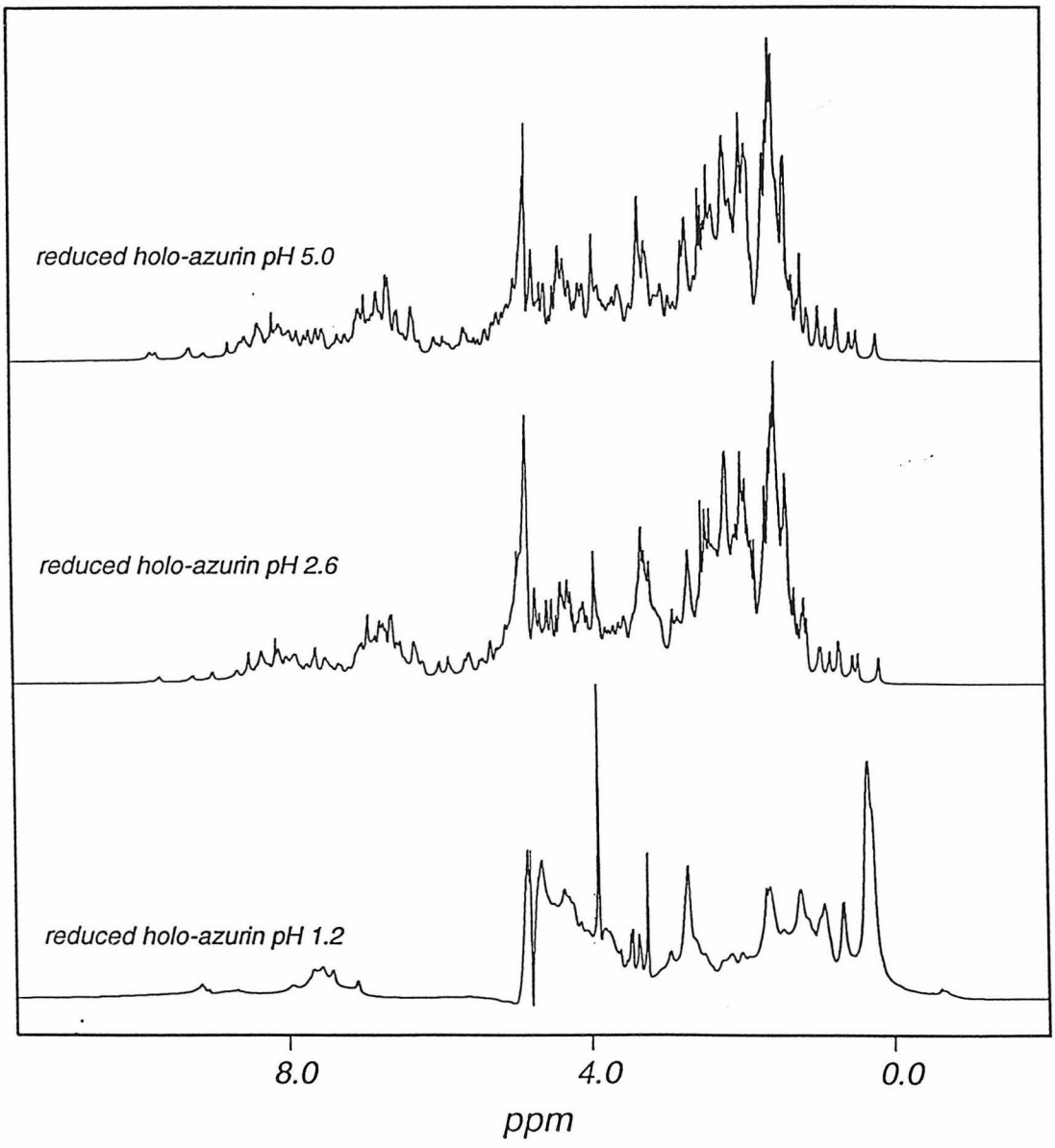
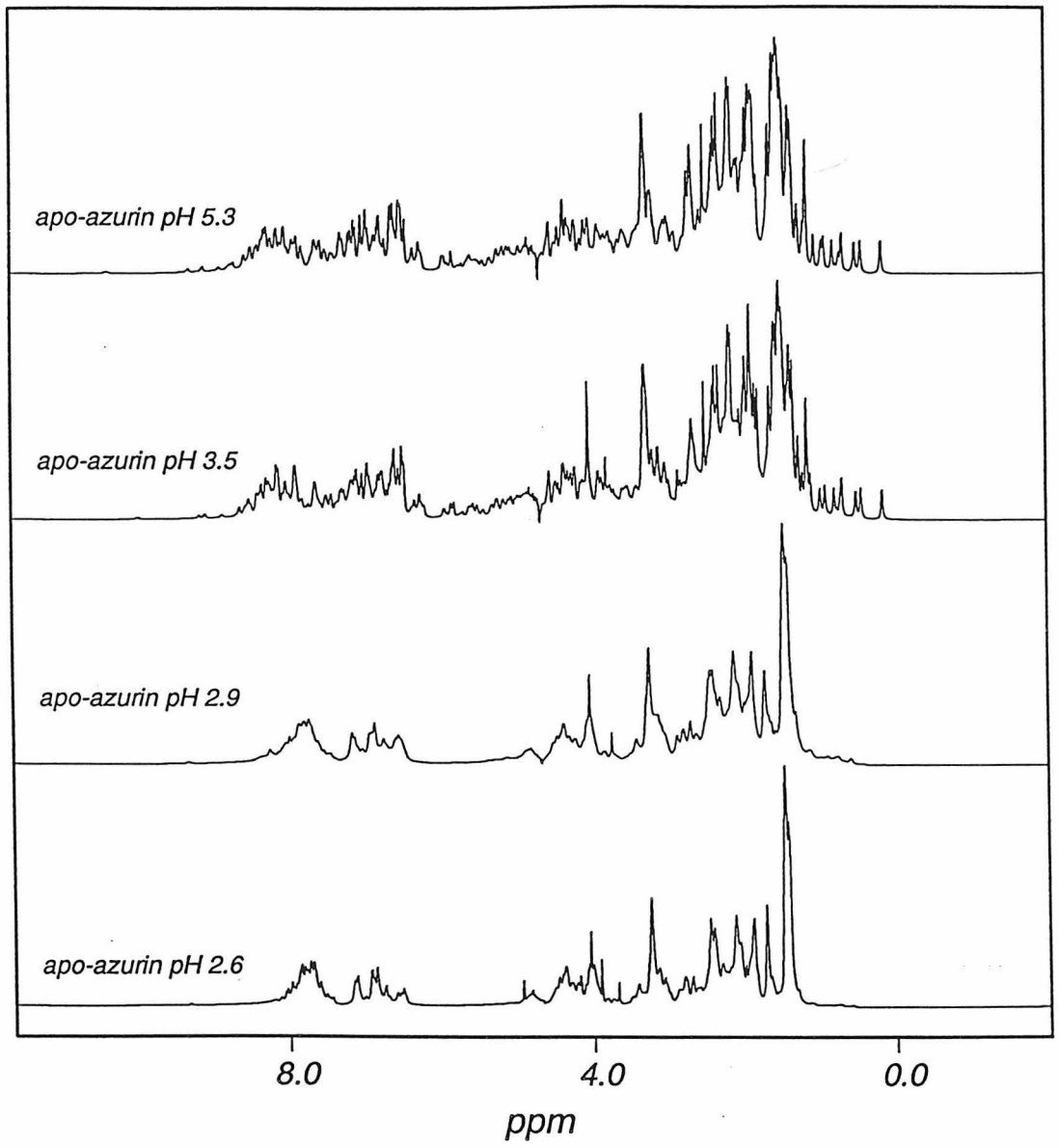

reduced holo-azurin pH 5.0
reduced holo-azurin pH 2.6
reduced holo-azurin pH 1.2
8.0
4.0
0.0
ppm

Figure III.8. (a) ^{1}H-NMR spectra of apo-azurin at different pH's, from top to bottom: pH 5.3, pH 3.5, pH 2.9, and pH 2.6. (b) The expanded amide region of spectra in (a) from chemical shift 5.5 ppm to 10 ppm.

(a)


apo-azurin pH 5.3
apo-azurin pH 3.5
apo-azurin pH 2.9
apo-azurin pH 2.6
8.0
4.0
0.0
ppm

(b)

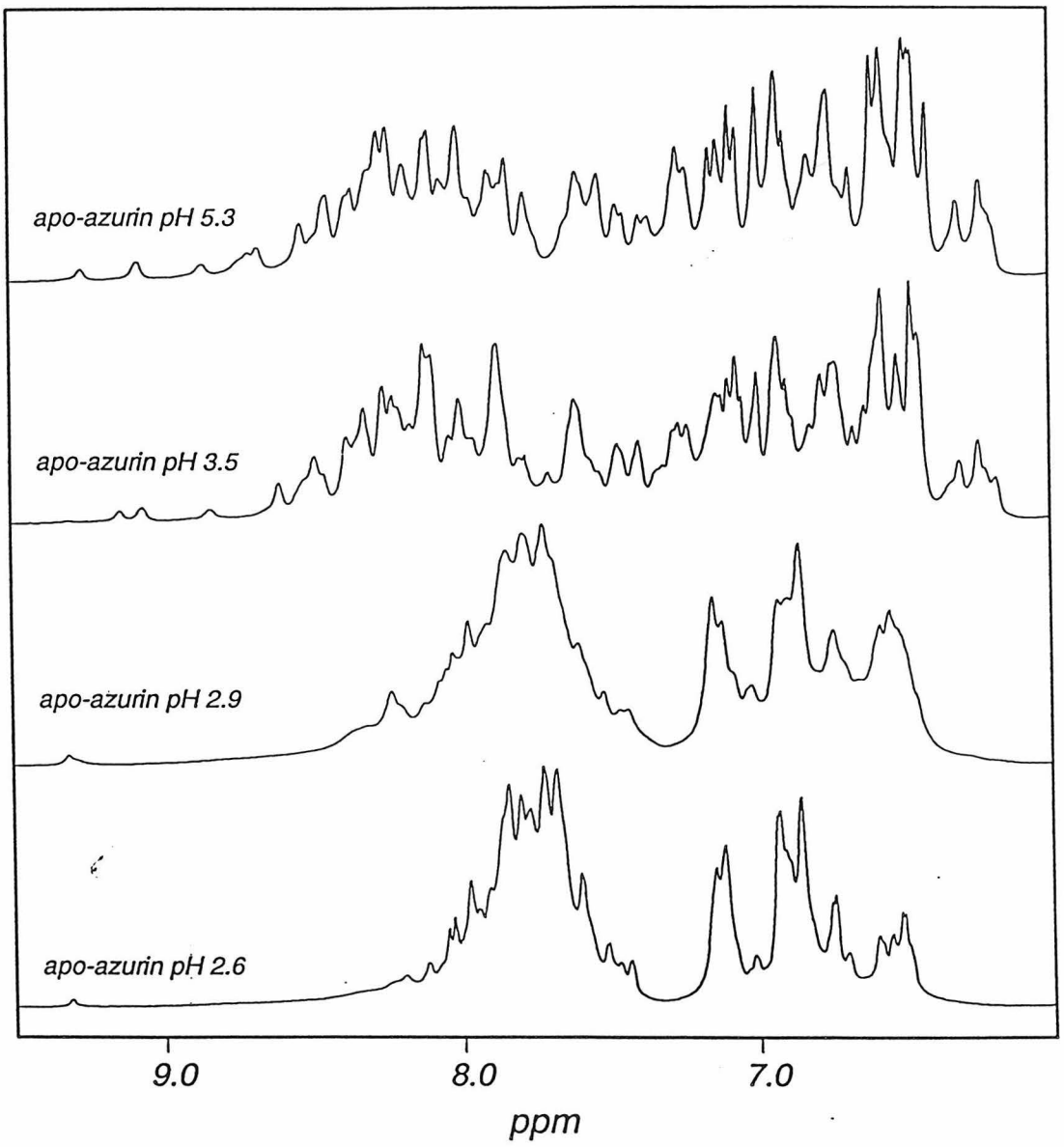
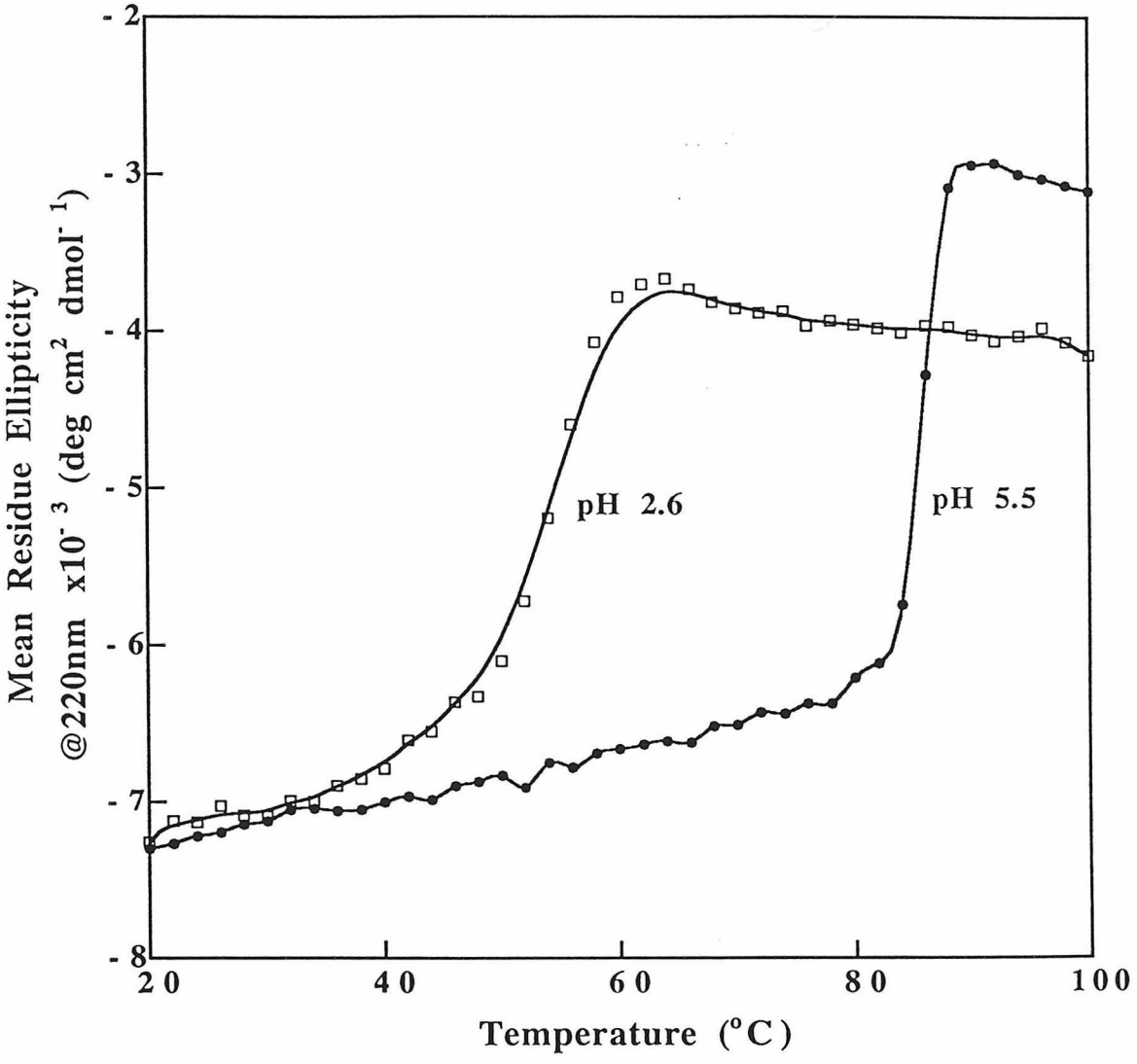
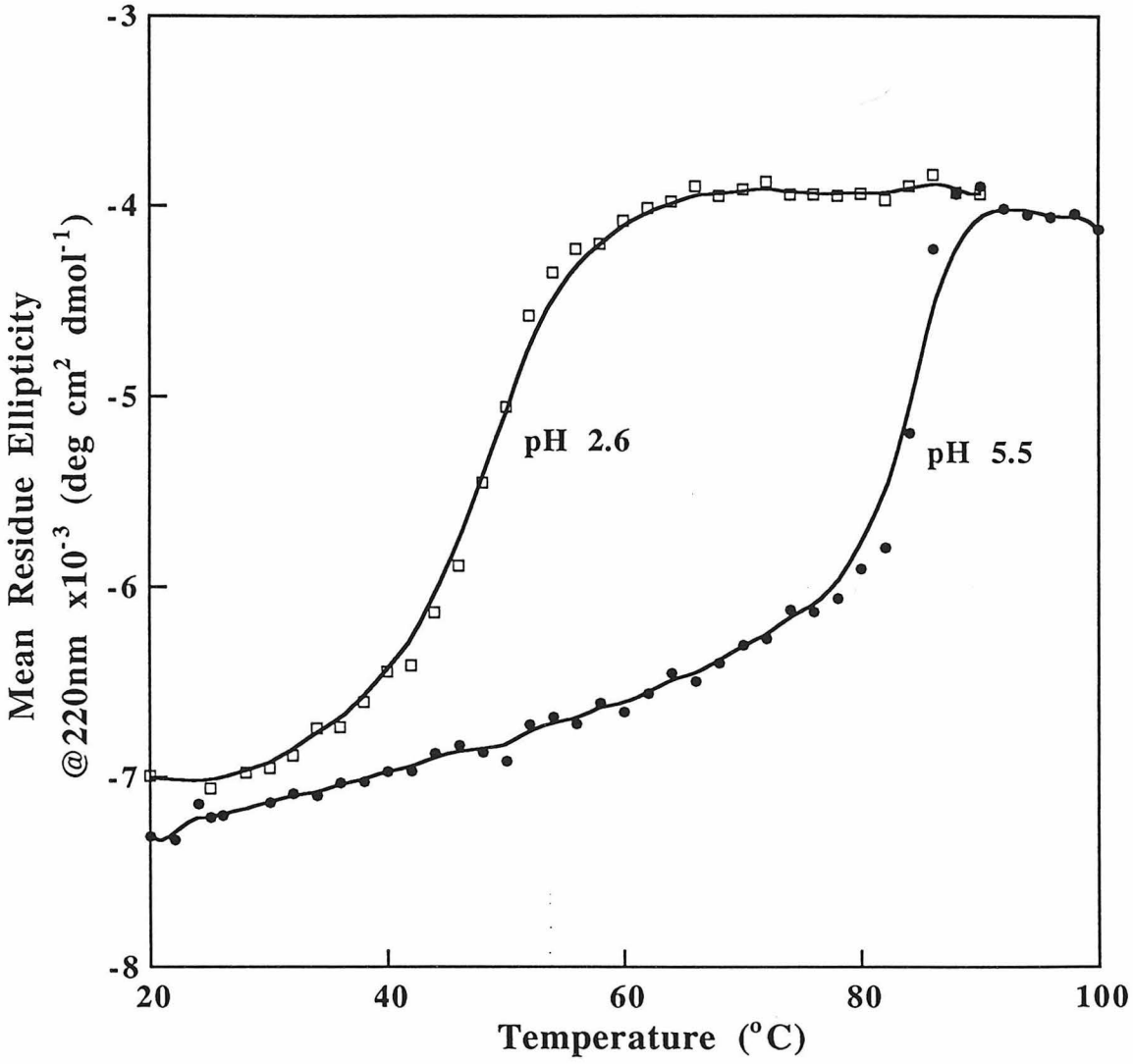

apo-azurin pH 5.3
apo-azurin pH 3.5
apo-azurin pH 2.9
apo-azurin pH 2.6
9.0
8.0
7.0
ppm

Figure III.9. Thermal unfolding curves measured by CD monitoring at 220 nm. (a) oxidized holo-azurin at pH 5.0 and pH 2.6; (b) reduced holo-azurin at pH 5.0 and pH 2.6; (c) apo-azurin at pH 5.0 and pH 2.9.

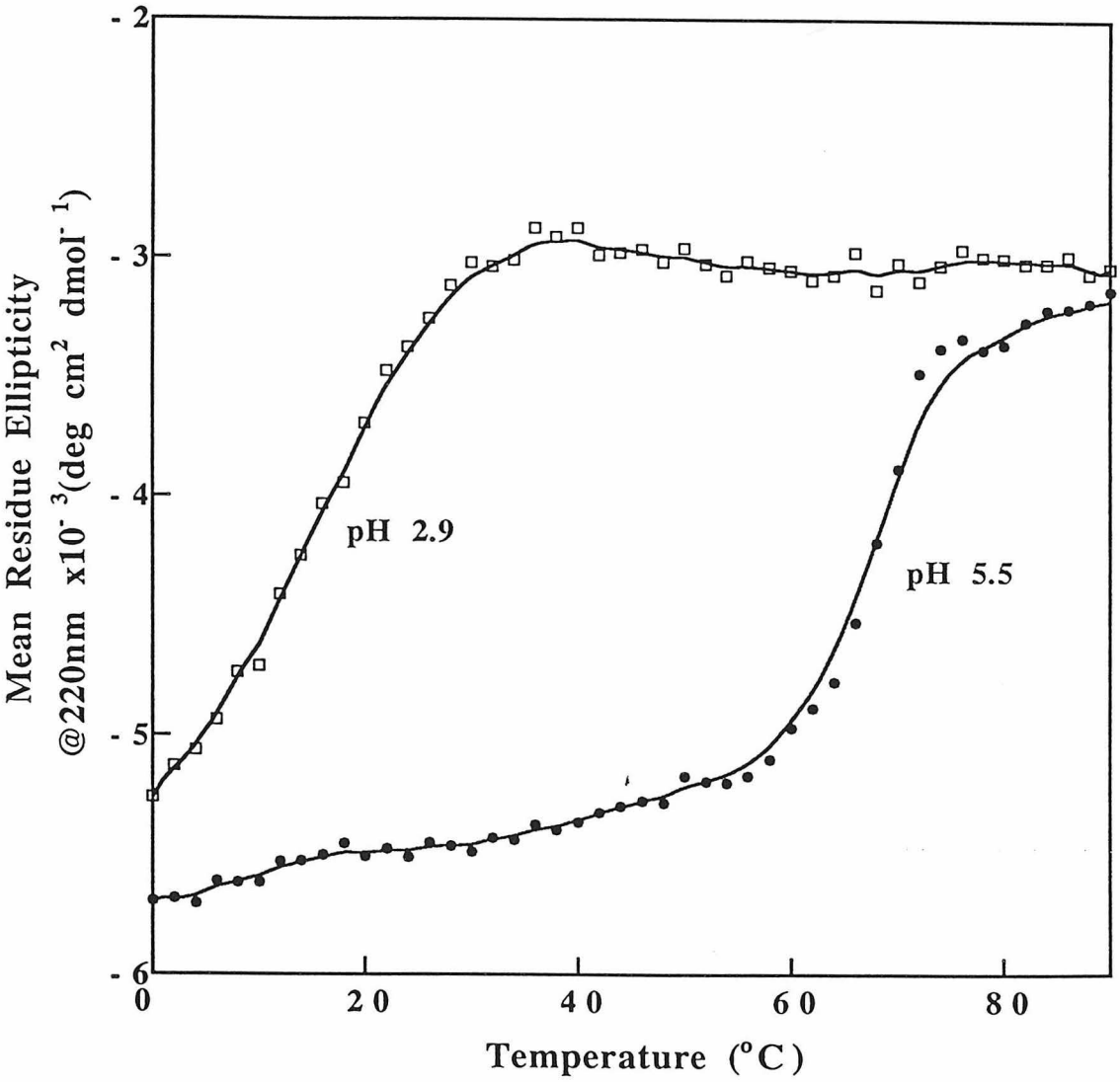
(a)


Mean Residue Ellipticity
@220nm x10⁻³ (deg cm² dmol⁻¹)
-2
-3
-4
-5
-6
-7
-8
20
40
60
80
100
Temperature (°C)
pH 2.6
pH 5.5

(b)


-3
-4
-5
-6
-7
-8
20
40
60
80
100
Mean Residue Ellipticity
@220nm x10-3 (deg cm2 dmol-1)
Temperature (°C)
pH 2.6
pH 5.5

(c)


- 2
- 3
- 4
- 5
- 6
0
20
40
60
80
Mean Residue Ellipticity
@220nm x10^{-3}(deg cm^2 dmol^{-1})
Temperature (°C)
pH 2.9
pH 5.5

DISCUSSIONS

Acid induced molten globule states have been reported for several proteins, and the role of this type of conformation in the mechanism and pathways of protein folding has been discussed (Ohgushi and Wada, 1983; Kuwajima, 1989; Ptitsyn, 1987; Goto and Fink, 1989). Although the mechanism of acid induced unfolding and the generality of the molten globule conformation under such conditions are not well understood, it is generally assumed that electrostatic repulsion causes the partial unfolding. We believe the molten globule conformation of apo-azurin at pH 2.9 also results from such electrostatic repulsion. Protonation of side chains results in a net positively charged protein whose electrostatic repulsion causes partial unpacking. According to the normal protonation pKa values (Stryer, 1988), very few side-chain protonation reactions occur at pH 3 or below, except for protonation of the terminal carboxyl group. This cannot be responsible for the partial unfolding transition observed for apo-azurin and holo-azurin, however, since there would be no difference between the two forms. The mechanism is more likely to involve protonation of the Cu(II) binding ligands. There are four side chains (Cys-112 S, His-46 and His-117 N, and Met-121 S) and one main chain carbonyl group of Gly-45 involved in Cu(II) binding. Although the solution structure of apo-azurin has not been determined, it is believed apo-azurin in solution is an equilibrium inter-conversion between two similar forms of apo-azurin determined in crystal structure (Nar *et al.*, 1992; Luo and Mayo). The unique Cu(II) ligand site buried inside azurin could cause the pKa's of both His-46 and His-117 protonation to be significantly below their normal value of 6.0. As was pointed out by Nar *et al.*, the pKa's for protonation of His-46 and His-117 could be less than 3 (Nar *et al.*, 1992). Therefore, apo-azurin's transition from the native-like state to the molten globule state at around pH 2.9 is most likely caused by the protonation of His-46 and His-117. Protonation of these two side-chains results in net positive charges close to

the Cu(II) binding site. The electrostatic repulsion between these positively charged side-chains leads to partial unfolding and formation of the molten globule conformation. For holo-azurin, however, due to the binding of Cu(II), an even lower pH is required for protonation of His-46 and His-117. Thus, these residues are not significantly protonated even at pH 2.6. Further lowering the pH (to 1.6) ultimately protonates them, and causes loss of Cu(II) binding, but also totally unfolds the protein.

Partially unfolded apo-azurin at pH 2.9 has all the characteristics of a molten globule: binding to ANS, substantial secondary structure but little tertiary structure, and no cooperative thermal unfolding transition. In comparison, although both the oxidized and reduced forms of holo-azurin can bind ANS at pH 2.6, indicating partial exposure of their hydrophobic surface, they essentially retain both their native-like secondary and tertiary structures, as revealed by various spectroscopic characterizations. Thus holo-azurins at pH 2.6 are not molten globules (Kim and Baldwin, 1990). Because of the Cu(II) binding of holo-azurin at pH 2.6, the partial solvent accessibility of hydrophobic regions is unlikely to be caused by the protonation of His-46 and His-117 as is the case for apo-azurin. Other mechanisms, such as the anion effects discussed by Goto *et al.*, may cause the ANS binding (Goto *et al.*, 1990). These include Debye-Huckel screening, the interaction with positive charges by ion-pair formation (or anion binding), and the effects of anions on the water structure. Apo-azurin can also bind ANS at pH 2.6, although CD measurements indicate it has neither tertiary or secondary structure. This suggests that the unfolding of apo-azurin at pH 2.6 may not be as complete as the unfolding of apo-azurin by 5.5 M GdnHCl.

The fact that the near UV CD signal of apo-azurin did not completely disappear at pH 2.9 suggests that some tertiary structure is still present. The extent to which the tertiary structure is retained and whether the observed molten globule conformation is also a kinetic intermediate state in the folding pathway at normal pH remain to be studied.

REFERENCES

1. Adman, E.T., (1991) *Advances in Protein Chemistry* **42**, 145-197.

2. Barrick, D. and Baldwin, R.L. (1993) *Protein Science* **2**, 869-876.

3. Cantor, C.R. and Schimmel, P.R. (1980) *Biophysical chemistry part* **II** New York: WH Freeman and Company, pp 433-465.

4. Chang, T.K., Iverson, S.A., Rodrigues, C.G., Kiser, C.N., Lew, A.Y.C., Germanas, J.P., and Richards, J.H. (1991) Proc. Natl. Acad. Sci. USA **88**, 1325-1329.

5. Dolgikh, D.A., Abaturov, L.V., Bolotina, I.A., Brazhnikov, E.V., Bushuev, V.N., Bychkova, V.E., Gilmanshin, R.I., Lebedev, Y.O., Semisotnov, G.V., Tiktopulo, E.I., and Ptitsyn, O.B. (1985) *Eur. J. Biophys* **13**, 109-121.

6. Dolgikh, D.A., Kolomiets, A.P., Bolotina, I.A., and Ptitsyn, O.B. (1984) *FEBS Letters* **165 (1)**, 88-92.

7. Dryden, D. and Weir, M.P. (1991) *Biochim E. Biophysica Acta* **1078**, 94-100.

8. Engeseth, H.R. and McMillin, D.R. (1986) *Biochemistry* **25**, 2448-2452.

9. Feng, Y., Sligar, S.G., and Wand, A.J. (1994) *Nature Structural Biology* **1**, 30-35.

10. Goldberg, M. and Pecht, I. (1976) *Biochemistry* **15**, 4197-4152.

11. Goto, Y. and Fink, A.L. (1989) *Biochemistry* **28**, 945-952.

12. Haynie, D.T. and Freire, E. (1993) *Proteins: Structure, Function, and Genetics* **16**, 115-140.

13. Hlodan, R. and Pain, R.H. (1994) *FEBS Letters* **343**, 256-260.

14. Kim, P.S. and Baldwin, R.L. (1990) *Ann. Rev. Biochem.* **59**, 631-660.

15. Luo, J.Y. and Mayo, S.L. (to be published) or Chapter IV of this thesis.

16. Nar, H., Messerschmidt, A., Huber, R., van de Kamp, M., and Canters, G.W. (1992) *J. Mol. Biol.* **221**, 765-772.

17. Nar, H., Messerschmidt, A., Huber, R., van de Kamp, M., and Canters, G.W. (1992) *FEBS Letters* **306**, 119-124.

18. Ohgushi, M. and Wada, A. (1983) *FEBS Letters* **164**, 21-24.

19. Ptitisyn, O.B. (1992) *Protein folding*. New York: WH Freeman and Company, Creighton TE, eds. pp 243-300.

20. Redfield, C., Smith, R.A.G., and Dobson, C.M. (1993) *Nature Structural Biology* **1**, 23-29.

21. Shepard, W.E.B., Anderson, B.F., Lewandowski, D.A., Norris, G.E., and Baker, E.N. (1990) *J. Amer. Chem. Soc.* **112**, 7817-7819.

22. Solomon, E.I., Hare, J.W., Dooley, D.M., Dawson, J.H., Stephens, P.J., and Gray, H.B. (1980) *J. Amer. Chem. Soc.* **102**, 168-178.

23. Tisi, L. and Evans, P.A. (1993) *Protein Engineering* **6(S)**, 24-24.

24. van de Kamp, M., Canters, G.W., Wijmenga, S.S., Lommen, A., Hilbers, C.W., Nar, H., Messerschmidt, A., and Huber, R. (1992) *Biochemistry* **31**, 10194-10207.

25. Weisman, J.S. and Kim, P.S. (1991) *Science* **253**, 1286-1393.

26. Weser, U. (1985) *Structure and Bonding* **61**, 145-160.

Chapter IV

^{1}H, ^{15}N Chemical Shift Assignments and Backbone Dynamics of *Pseudomonas aeruginosa* Apo-azurin by 2D NMR

SUMMARY

In the previous chapter, I found that apo-azurin [without Cu(II) binding] becomes less stable than the holo-azurin, and forms a molten globule conformation at acidic pH. In order to gain more insight into its solution properties, I decided to study apo-azurin using NMR technology. I started by isotope (^{15}N) labeling apo-azurin, and used 2D NMR to assign the ^{1}H and ^{15}N chemical shifts of the backbone. The relaxation properties of the backbone amide ^{15}N were then measured using inverse-detected 2D NMR. The relaxation data were analyzed with the model-free formalism of protein dynamics introduced by Lapari and Szabo, and the overall correlation time (τ_c) and the order parameter (S^2) for each residue were obtained. τ_c was determined to be 5.8 ns and S^2 mostly ranged from 0.8 to 9.5. Mapping of the dynamic order parameter (S^2) of each residue onto the apo-azurin backbone structure showed that S^2 were quite uniform in the well-defined β-sheet regions. Low order parameters were located in the turn or loop regions, and interestingly in the middle of α-helix as well. Low order parameters in the loop region near His-117 were also observed, consistent with the observation in apo-azurin's crystal structure.

INTRODUCTION

With the advance of modern Nuclear Magnetic Resonance (NMR) technology, NMR has become a common technique for studying the solution properties of proteins (Wuthrich, 1986; Clore and Gronenborn, 1993). Multi-dimensional NMR spectroscopies are now used to determine protein structure, to measure folding stability and kinetics, and to study protein dynamics. A common requirement for all NMR studies is the chemical shift assignment of the subject protein. The method most commonly used to obtain assignments is the sequential assignment strategy developed by Wuthrich (Wuthrich, 1986). In this chapter, I will describe my NMR study on apo-azurin using recently developed 2D NMR spectroscopies.

My study focused on the ^{1}H and ^{15}N chemical shift assignments and solution dynamics of apo-azurin backbone. The study was initiated for several reasons. First, the NMR chemical shift assignments of apo-azurin would be useful for future studies. Secondly, there is little residue coordinate difference in the x-ray crystal structure of apo-azurin in comparison to the holo counterpart (Nar *et al.*, 1992), but there is a significant stability difference in solution between apo-azurin and its holo counterpart. It would therefore be interesting to see the solution dynamics of apo-azurin at an atomic level. Third, the crystal structure of apo-azurin is very similar to that of holo-azurin. van de Kamp *et al.* has already done ^{1}H and ^{15}N chemical shift assignments of the backbone and side chains of reduced holo-azurin (van de Kamp *et al.*, 1992). We expect the overall NMR spectra of apo-azurin to be similar to that of reduced holo-azurin. These chemical shift assignments can therefore be used as a guide in assisting our apo-azurin assignments.

MATERIALS AND METHODS

^{15}N Labeling and Preparation of Apo-azurin ^{15}N labeling of azurin was achieved by expressing azurin in M9 minimal media using $^{15}NH_4Cl$ at concentration of 1g/L as the sole nitrogen source (appendix IV-1). The procedures for expression and isolation of ^{15}N labeled azurin were the same as those described in Chapter I for expression in regular medium. The expression level of azurin in minimal media is much lower than in the regular media. ^{15}N labeled azurin and apo-azurin were purified in the same way as the regular protein described in Chapter I. NMR samples were prepared from the purified and lyophilized apo-azurin powder stored at -20 °C.

NMR Data Collection NMR spectra were recorded on a Varian 600 MHz Unity Plus instrument. An inverse probe was used for both homonuclear ^1H-^1H spectra and heteronuclear ^1H-^{15}N spectra. Water suppression for 90% H_2O samples was primarily achieved by combining pulse-field gradient and continuous saturation irradiation at a low power level. Water-gated selective pulse was used in water-gated NOESY spectra. All 2D spectra were acquired at two different temperatures (30 °C and 40 °C) for final data processing and analysis. DQF-COSY spectra at lower temperatures (10 °C and 15 °C) were also acquired but were not used for final spectra analysis, because they were not as well resolved as the spectra at 30 °C and 40 °C. Apo-azurin samples aggregate gradually over time in the NMR tube at 30 °C or 40 °C. A fresh prepared sample usually can last about 48 hours in the probe at 30 °C and about 30 hours at 40 °C.

The ^1H chemical shifts in all spectra were referenced to HOD peak at the specific temperature according to the following formula:

$$\delta_{HOD} = 5.0252 - (0.01037 \times T) \qquad \text{where } T \text{ is temperature in } ^\circ C$$

^{15}N chemical shifts were referenced to pseudo zero chemical shift of ^{15}N using the Varian supplied "*Reference_N15*" methos in VNMR. (Note: in the rest of this chapter, all VNMR commands are printed in italics)

The acquisition parameters for various 2D NMR experiments are described below.

DQF-COSY. A pulse-field-gradient version of Multi-Quantum Filtered Correlation Spectroscopy (MQF-COSY) in phase sensitive mode was used for DQF-COSY unless otherwise specified. The parameter file was set up as "*gmqcosy*", then changed "seqfil" to "gmqcosyps". The typical parameters for the phase sensitive gradient MQF-COSY are listed below. (Note: The other 2D experiments I performed also use many of these parameters. The common parameters will not be listed again for these experiments.)

SAMPLE	OBSERVE	SPECTRUM	AUTOMATION
date =	tn = 'H1'	sw = 8100.4	alock = n
temp = 30	sfrq = 600.23	fb = 4400	in = n
solvent = D2O	tof = -1576.6	at = 0.126	spin = not used
PULSESEQUENCE	pw90 = 6.7	np = 2048	wshim = n
seqfil = 'gmqcosyps'	tpwr = 58	nt = 32	PROCESSING
d1 = 0.5	DECOUPLE	ct = 32	fn = 2048
satdly = 1.5	dn = 'H1'	ss = 32	sb = 0.060
satpwr = 0	dfrq = 600.230	gain = 36	sbs = not used
taud2 = 0	dof = -1576.6	2D SPECTRUM	gf = 0.040
tau1 = 0	dm = 'nnn'	sw1 = 8100.4	gfs = not used
GRADIENTS	dmm= 'c'	ni = 512	2D PROCESSING
gzlvl1 = 16000	dmf = 200	phase = arrayed(1,2)	fn1 = 2048
gt1 = 0.002460	dseq		sb1 = -0.010
grise = 0.000010	dres = 1.0		sbs1 = not used
gstab = 0	homo = 'y'	qlvl = 2	

At 40 °C, tof = dof = -1574.0. These numbers were determined through an array of numbers.

NOESY. For samples in 90% H_2O, both pulse-field-gradient NOESY (GTNNOESY) and water-gated NOESY (WATERGNOE) with mixing times of 100 ms and 150 ms were performed. WATERGNOE gave much better suppression of the HOD signal and the signal to noise ratio (S/N) than did GTNNOESY. Thus, only WATERGNOE spectra were used for NOESY spectra analysis. The WATERGNOE experiments were set up using the macro "*watergnoe*". Water suppression parameters and processing parameters for WATERGNOE are listed below.

WATER SUPPRESSION PARAMETERS

psel = 5210 #selective ^1H 90 degree time, determined from an array at power of pselvl

pselvl = -1 #power level for psel, (-2) or (-1)

phaseinc = -18 #0.5 degree phase shift for softpulse

gt1 = 0.0020 #gradient time, (0.002 sec)

glvl1 = 23500 #gradient level

gtm = 0.0020 #mix gradient time, (0.002 sec)

glvlm = -8000 #gtm level, (-8000)

PROCESSING PARAMETERS

sb = 0.06, sbs = -0.002, gf = 0.036

sb1= 'n', sbs1 = 'n', gf1 = 0.036

Some notes about WATERGNOE:

1. It is best to set up with waterg 1D and process with *wft2da('lfs')* (phase=1, 2). Set dm = 'nyny' and dm2 = 'nyny' to decouple ^{13}C and ^{15}N with labeled samples, and set dpwr, dmf and dpwr2, dmf2 accordingly.
2. When doing WATERGNOE experiments, the pulse sequence does not check the probe meltdown; one must doubly check appropriate parameters to make sure they are correct before execution.

TOCSY. TOCSY, TNTOCSY, and TOCSYSE were all tried for total correlation spectra. TOCSYSE spectra at 30 °C and 40 °C were used for spin correlation analysis, because TOCSYSE was found to be the best in terms of HOD suppression and S/N ratio (Griesinger *et al.*, 1988; Shaka *et al.*, 1988). Three different mixing times were used (35 ms, 50 ms, and 70 ms). The unique parameters for TOCSYSE are as follows:

---

pwspwr = 46

pws = 26.8 #90 degree pulse during mlev periods at power level of pwspwr, in the #range of 25 - 30 μs.

tnsat = 'yn' #use transmitter presaturation during delay d1.

satpwr = 0

sephase = arrayed (1,2)

Notes: for all Signal Enhanced (SE) experiments, the raw data (fid files) have to be converted before Fourier Transformation (FT). The procedure for conversion is: 1) use macro "*makesefid*" to convert a fid file consisting of two hyper complex datasets into a sensitivity enhanced fid file, 2) use "*wft2dse*" for FT. The macros and source code for the fid file conversion and FT are available from the user library in VNMR.

HSQC. GHSQCSE (Zhang *et al.*, 1994) was used for HSQC experiments because of its enhanced S/N ratio. GHSQCSE was set up by "*ghsqcse(15)*" macro. Channel I and channel III were used for ^{1}H and ^{15}N respectively. Some of the unique parameters for HSQCSE are listed below.

---

flippwr = 11 #power of flipback pulse

flippw = 2000 #length of flipback pulse

flipshap = 'gauss' #shapelib file to be used for flipback pulse

dfrq2 = '60.827' #the third channel

dn2 = 'N15'

dof2 = 960.0 (later 1600 was used)

dpwr2 = 40

pwxNlvl = 58 #power level for nitrogen hard pulses

pwxN15 = 43.0 #PW90 for nitrogen

dm2 = 'nnny'

dmf2 = 3846

BigT = 0.00150

BigT1 = 0.00050

d1 = 3.0

jxh = 91 #NH coupling constant

sw1 = 4600

GRADIENTS

gt0 = 0.0010, gzlvl10 = 8000, gt1 = 0.0005, gzlvl1 = 5000, gt2 = 0.0010,

gzlvl2 = 10000, gt3 = 0.0010, gzlvl3 = 30000, gt4 = 0.000105, gzlvl4 = 27605,

gt5 = 0.00050, gzlvl5 = 5000

GHSQC-T_1 and GHSQC-T_2. Pulse sequences for measuring ^{15}N T_1 and T_2 from Kay's group (Farrow *et al.*, 1994) were used. Channel I and channel III were used for ^{1}H and ^{15}N respectively as in the GHSQCSE experiments. Some of the unique parameters for ^{15}N-T_2 and ^{5}N-T_1 are listed below.

For GHSQC-T_2

pw_sl = 1920.0, tsatpwr = -16

tpwrsl = 8 #the power level for a 90 degree rectangle shaped hardpulse of 1920 μs

dpwr2 = 40, dhpwr2 = 58, pwn = 43, dmf2 = 3846, taua = 0.00225, taub = 0.00275, BigT1 = 0.00050, hscuba = 0.030

GRADIENTS

gt1 = 0.001, gzlvl1 = 5000, gt2 = 0.0005, gzlvl2 = 4000, gt3 = 0.001, gzlvl3 = 10000, gt4 = 0.0005, gzlvl4 = 8000, gt6 = 0.00125, gzlvl6 = 30000, gt7 = 0.0005, gzlvl7 = 4000, gt8 = 0.000125, gzlvl8 = 29000 (27800)

The T_2 relaxation time is approximately ncyc*15.8 ms.

For GHSQC-T_1

shss_pwr = 31 #power level for 180 degree SS-shaped pulse of 550 μsec, refer to SS-shaped pulse for setting up.

pw_shpss = 550 #180 degree SS-shaped pulse length

shape_ss = "Dual_ss_180_550" #the SS pulse shape file

The T_1 relaxation time is approximately ncyc*5.55 ms.

Data Processing NMR data were processed using VNMR from Varian, unless otherwise specified. Routinely, the data were zero-filled once in the t1 dimension, multiplied by a phase-shifted squared sine-bell apodization function to optimize the S/N ratio, Fourier transformed, and interactively phase corrected. If necessary, a low-frequency digital filter and linear backward or forward prediction were also applied to minimize the residual HOD signal and to increase the S/N ratio of the processed spectra. The zero or low frequency digital filtering is automatically included in the FT in VNMR once the proper parameters are set up. The parameters were set up by the macro *setss* (Appendix IV-2).

In the Pulse-Field-Gradient version of the DQF-COSY experiments, the raw data were also right-shifted and backward linearly predicted in order to compensate for the time delay in the last field-gradient pulse (between the detection pulse and the data acquisition). Because of this, in the experimental setup, the (gt1+grise) was set to an integral number of dwell time, i.e., (gt1+grise) = n* (1/sw), where "sw" is the spectral width. The integral number "n" is the number of points to be right-shifted and linearly back predicted. In VNMR, this is done by:

lsfid=-n #left shift negative "n" points.

setlpb(n) #linear predict backward "n+1" point.

{*setlpb(n)* is a macro composed of a chain of VNMR commands:

parlp lpalg='lpfft' lpfilt=32 lpnupts=128 strtlp=$1 strtext=$1+1 lpext=$1+1 proc='lp' lpopt='b'

(a similar macro *setlpf1(ni wanted, ni used)* for forward linear prediction was also set up for convenience,

parlp(1) lpalg1='lpfft' proc1='lp' lpopt1='f' lpfilt1=12 strtlp1=$2 lpnupts1=$2 strtext1=$2+1 lpext1=$1-$2 }

The processed data were exported to NMRCompass (MSI, Inc) for further data analysis. The export procedure involves three steps: 1) the processed NMR spectra are saved to the "phasefile" under VNMR/datdir directory using a train of commands in the format of a macro "*sv2disk*":

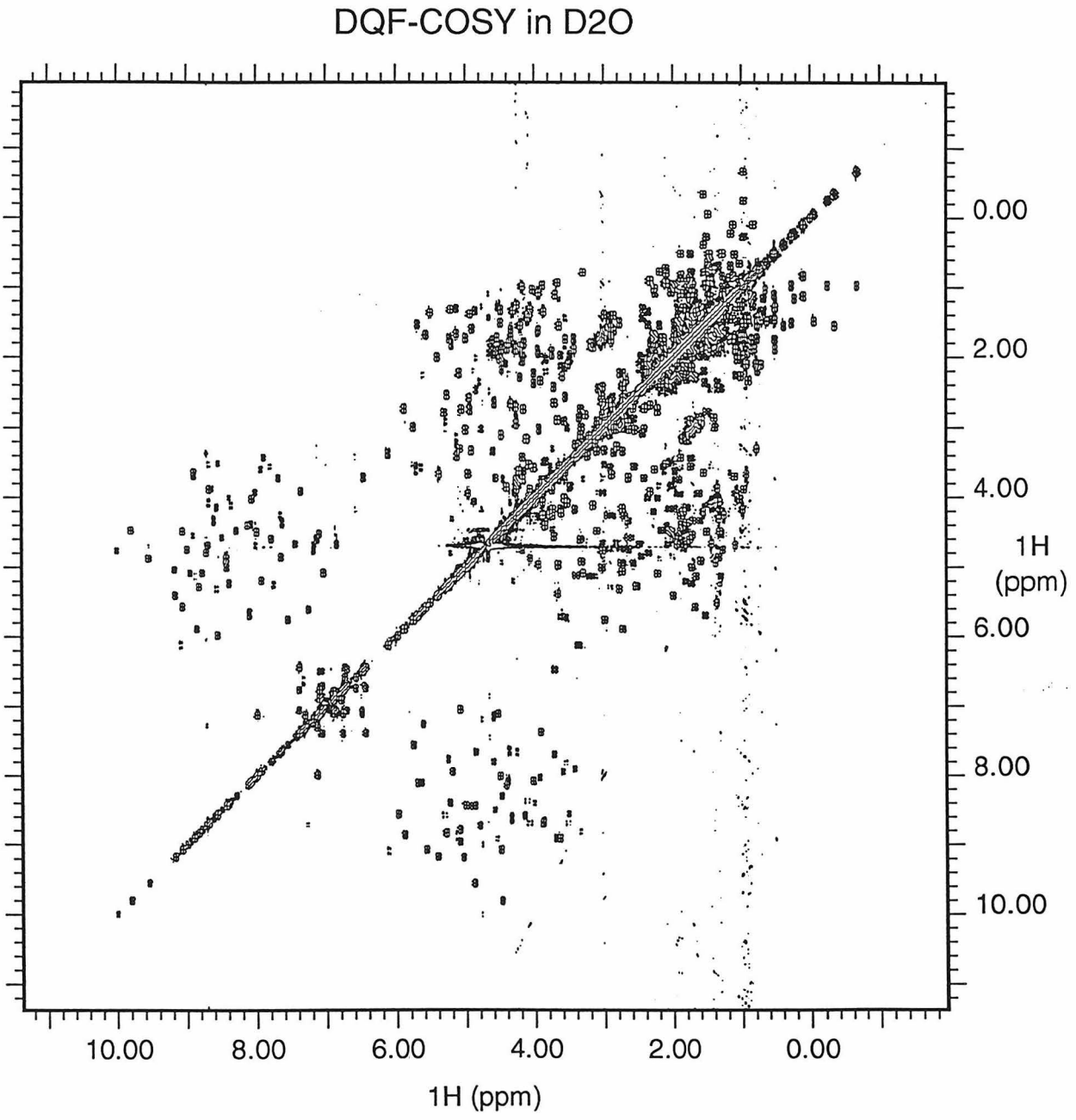
f trace='f1' dcon trace='f2' dcon flush

2) the associated parameters (spectra width and acquisition frequencies) are saved to a specified file using the command "*sv2d*" supplied by NMRCompass, 3) the "phasefile" and parameter file are then transferred to NMRCompass using the same name (the data file has the .dat extension).

RESULTS

Spin System Assignment Spin system assignment was started with the spectral analysis of 2D ^{1}H DQF-COSY and TOCSY spectra following the general procedure outlined by Wuthrich (Wuthrich, 1986). The spectral analysis was started from 2D ^{1}H DQF-COSY (Figure IV.1a and Figure IV.1b) and 50 ms mixing time TOCSY of fresh 99.9% D_2O samples at 30 °C. The fingerprint region spectra in D_2O are very much like those of the reduced holo counterpart (van de Kamp *et al.*, 1992). Some of the fingerprint region cross peaks were tentatively assigned by this peak pattern comparison. These peaks were then confirmed with spin system identification through the side-chains in both DQF-COSY and TOCSY spectra. These peaks (Q12, C26, T30, L33, S34, S51, S66, S89, M109, F110, C112, K122, T124) were labeled (see Figure IV.1b) and later used as the starting points for the sequential assignment analysis. Further spin system analysis was done on 2D ^{1}H DQF-COSY (Figure IV.2) and TOCSY (Figure IV.3) in 90% H_2O. A total of 122 NH-CαH cross peaks were found in the NH-CαH fingerprint region of DQF-COSY spectrum recorded at 40 °C in 90% H_2O (Figure IV.2). Because there are four prolines (no cross peaks) and 11 glycines (double cross peaks) among 128 residues, and the N-terminal NH is probably unobservable, the maximum number of cross peaks is 134. Following alignment of the fingerprint regions of DQF-COSY and TOCSY, 58 out of 122 were identified with their spin systems, 34 were ambiguous (including 11 Gly), and 30 were unidentifiable due to missing peaks or unresolved peaks.

Figure IV.1. DQF-COSY spectra of apo-azurin at 30 °C freshly prepared in 99.9% D_2O. (a) full spectrum; (b) the fingerprint region of (a).


DQF-COSY in D2O
0.00
2.00
4.00
1H
(ppm)
6.00
8.00
10.00
10.00
8.00
6.00
4.00
2.00
0.00
1H (ppm)

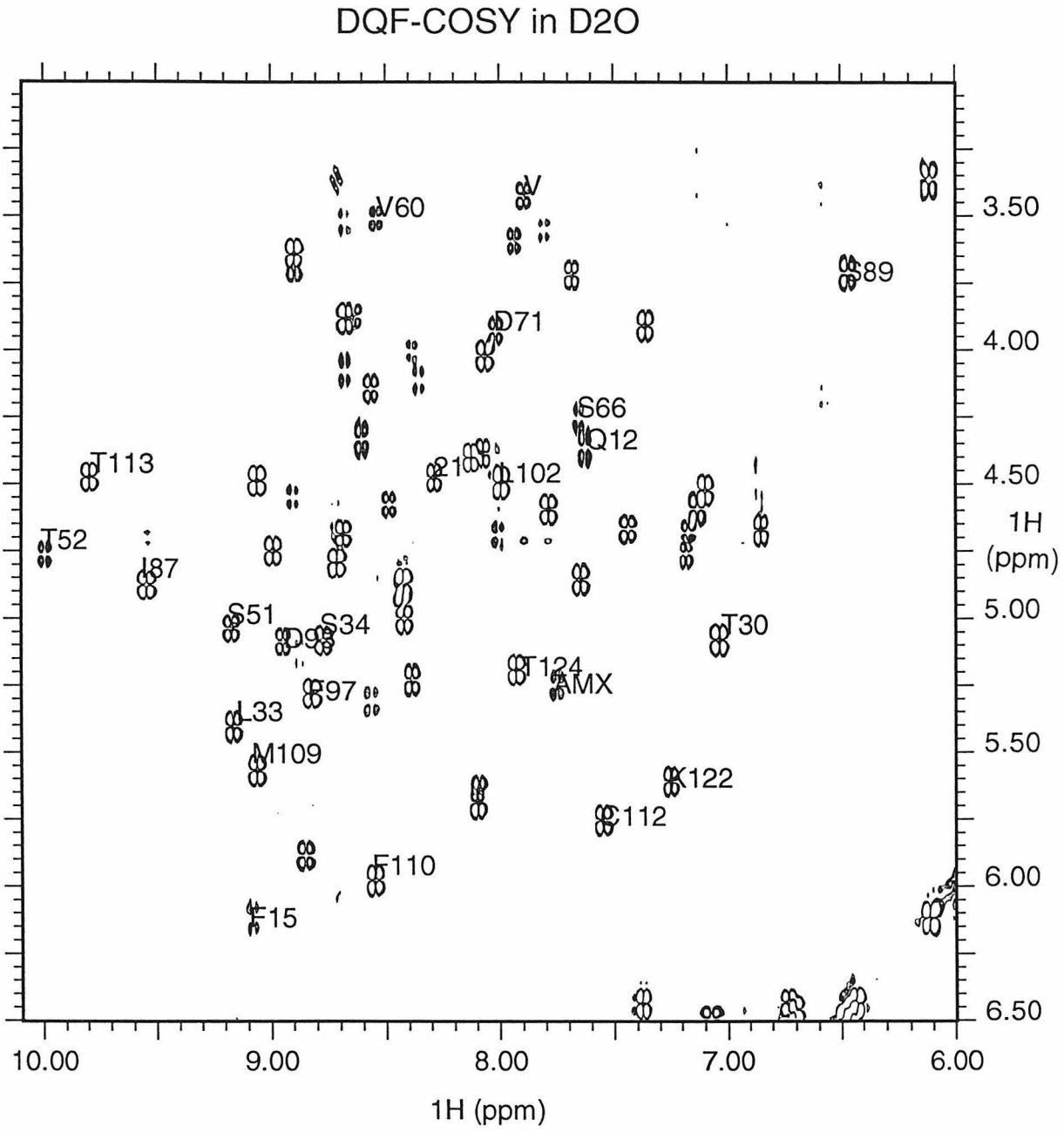

DQF-COSY in D2O
V60
D71
S66
Q12
T113
T102
T52
I87
S51
S34
T30
T124
AMX
I33
M109
K122
C112
F110
F15
3.50
4.00
4.50
5.00
5.50
6.00
6.50
1H (ppm)
10.00
9.00
8.00
7.00
6.00
1H (ppm)

Figure IV.2. Fingerprint region of DQF-COSY spectrum for apo-azurin in 90% H_2O recorded at 40 °C.

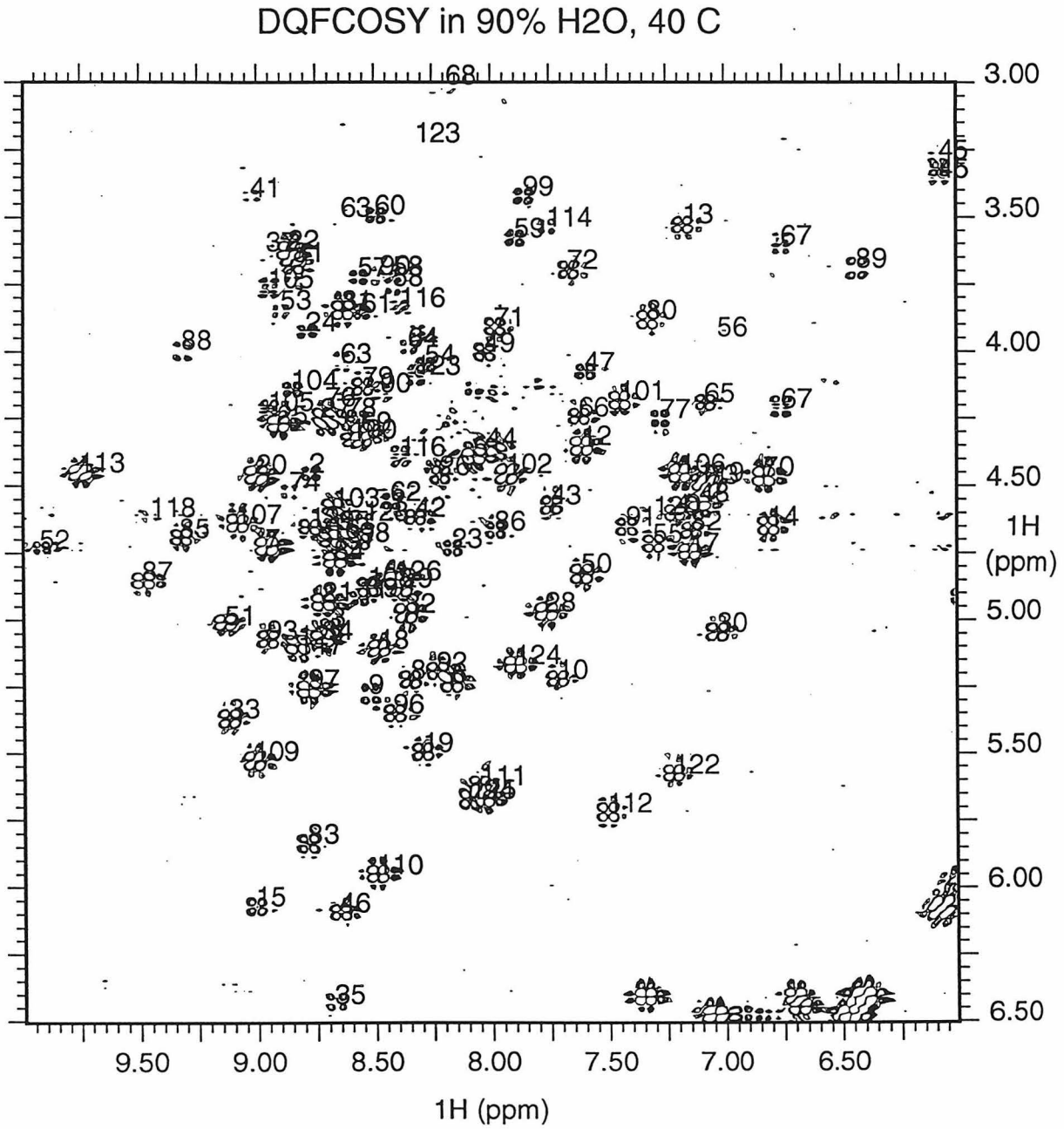
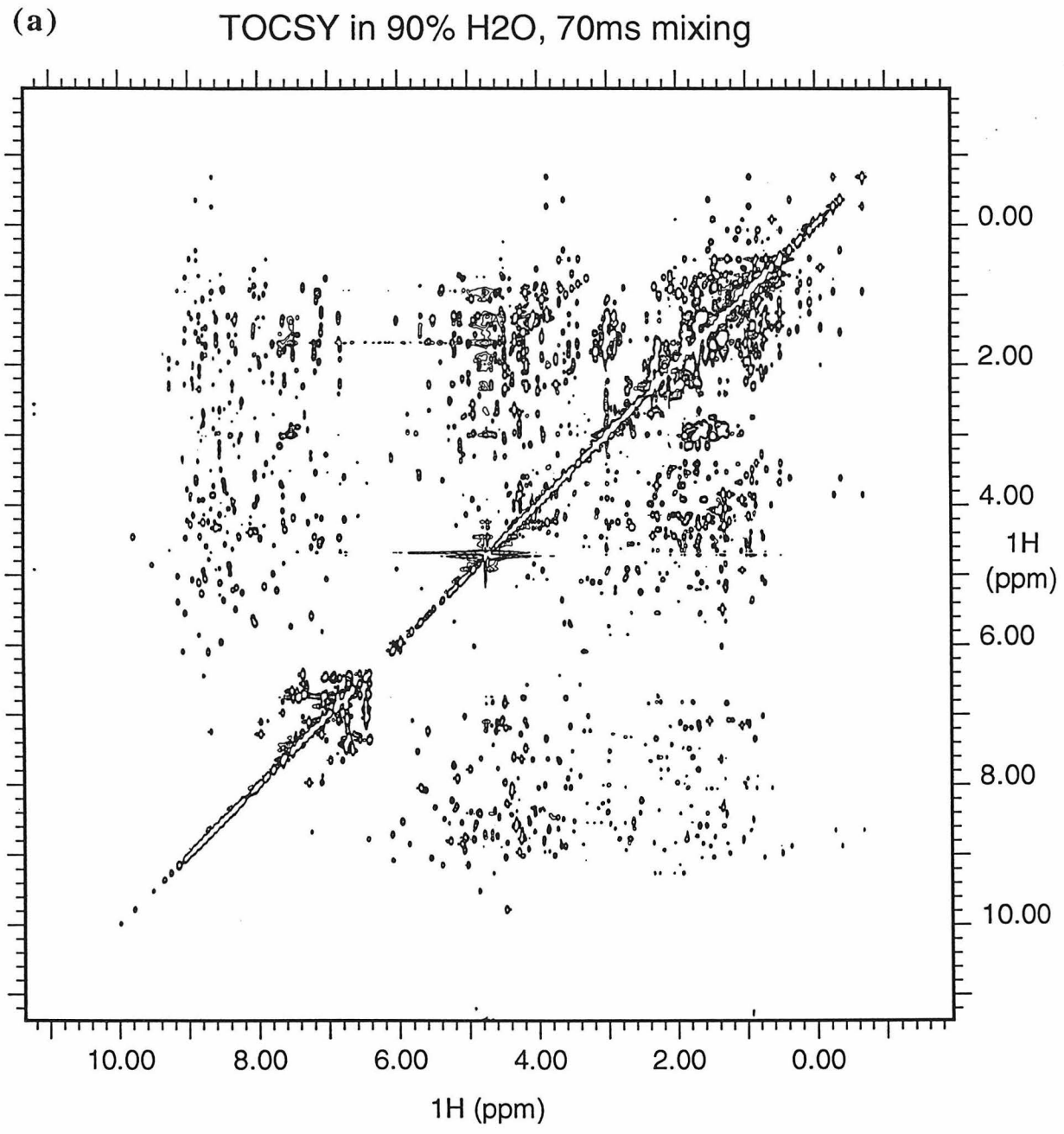
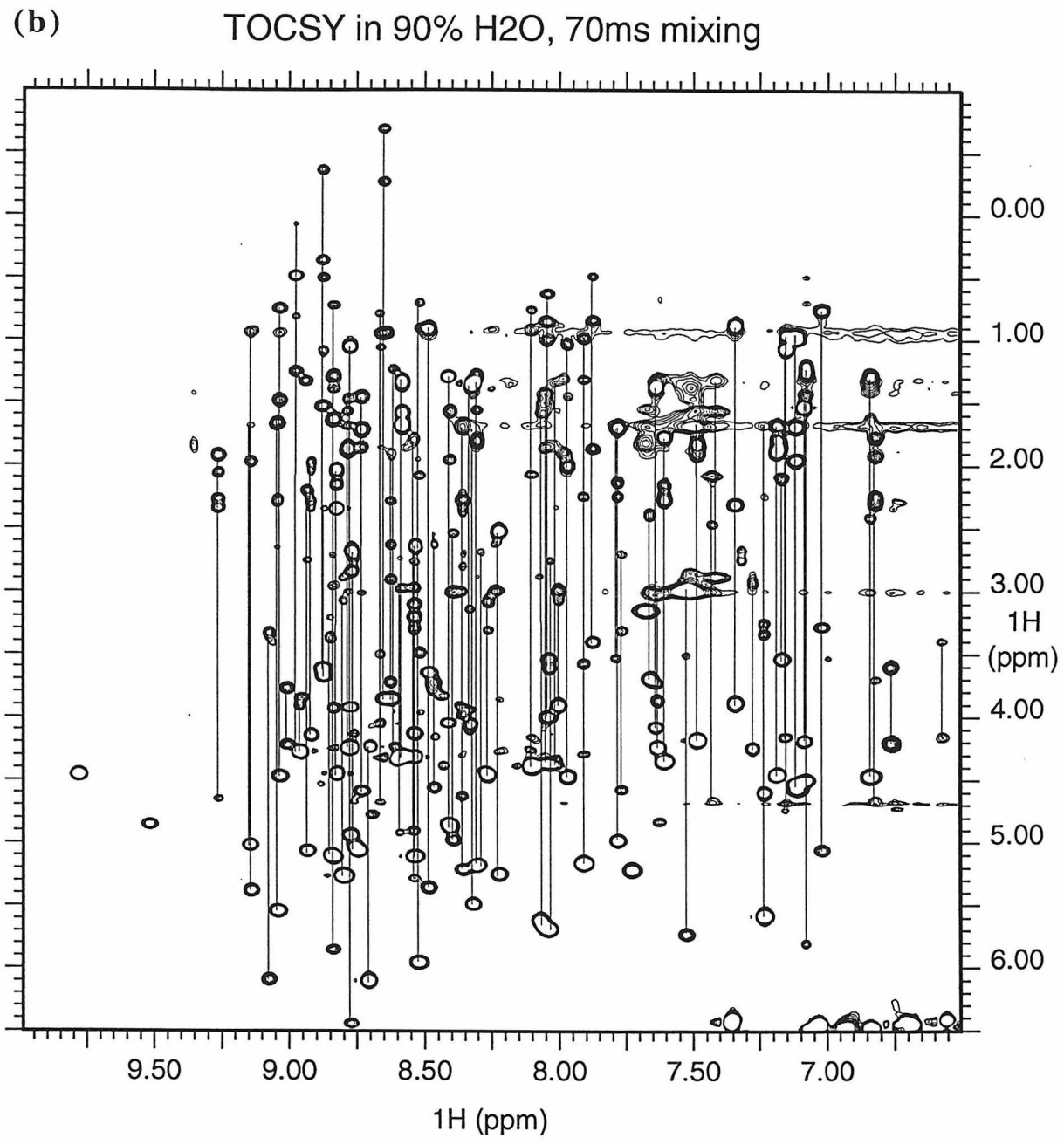

DQFCOSY in 90% H2O, 40 C
3.00
3.50
4.00
4.50
5.00
5.50
6.00
6.50
1H (ppm)
9.50
9.00
8.50
8.00
7.50
7.00
6.50
1H (ppm)

Figure IV.3. (a) TOCSY spectrum for apo-azurin in 90% H_2O recorded at 40 °C, (b) the fingerprint region of (a).


(a)
TOCSY in 90% H2O, 70ms mixing
0.00
2.00
4.00
1H
(ppm)
6.00
8.00
10.00
10.00
8.00
6.00
4.00
2.00
0.00
1H (ppm)


(b)
TOCSY in 90% H2O, 70ms mixing
0.00
1.00
2.00
3.00
1H
(ppm)
4.00
5.00
6.00
9.50
9.00
8.50
8.00
7.50
7.00
1H (ppm)

Sequential Assignment Sequential assignment was based on the fingerprint region spectra alignment of the WATERGNOE (Figure IV.4) and DQF-COSY (Figure IV.2) under the same conditions. Sequential connectivities were readily identified for residues 5-8, 17-23, 25-35, 47-50, 76-82, 92-99, 107-113, 122-127. They were much like those reported for reduced holo-azurin, corresponding to the eight β strands. The sequential connectivity information was feed back to the spin system assignments in the DQF-COSY spectrum to help confirm the assignments. Iterative sequential connectivity in the WATERGNOE spectrum and spin system assignments in DQF-COSY and TOCSY led to an almost complete assignment of all cross peaks in the DQF-COSY spectrum (all but five cross peaks). The assignments in the DQF-COSY spectrum were further confirmed by alignment with ^{1}H-^{15}N HSQC spectra (Figure IV.5).

Alignment between DQF-COSY and ^{1}H-^{15}N HSQC spectra was done as follows. Both spectra were displayed at the same spectra width on the ^{1}H axis. Most cross peaks in the DQF-COSY fingerprint region have one to one correspondence with the HSQC spectrum on ^{1}H axis. Because the ^{15}N chemical shifts are less dependent on the local environment, I expected the ^{15}N chemical shifts of apo-azurin to be similar to those of reduced holo-azurin. The ^{15}N chemical shift of each aligned peak was therefore compared to the ^{15}N chemical shift of holo-azurin; the peak was then assigned accordingly provided this was also consistent with the sequential connectivity assignment. All the peaks assigned in sequential assignment were thus confirmed; ambiguous peaks were also assigned using this alignment. The assignments were further confirmed for consistency in sequential connectivity searching in WATERGNOE spectra.

Higher dimensional NMR spectra (3D or 4D) are required for rigorous assignments of apo-azurin. However, due to the similarity between apo-azurin and holo-azurin, the iterative 2D spectral assignment presented here is adequate and valid for the backbone assignments. The results are presented in Table IV.1. When the ^{1}H chemical shift

assignments of apo-azurin are compared to those of reduced holo-azurin, the majority are within 0.1 ppm. ^{1}H chemical shifts of NH or C_{α}-H which differ by more than 0.1 ppm from those of the holo-azurin were found for the following residues: Asn-10 to Asn-16, Gly37 and Asn-38, Asn-42 to Leu-50, Asp-62 and Met-64, His-83 and Thr-84, Ile-87 to Lys-92, and Phe-111 to Met-121. In addition, I was not able to unambiguously identify the ^{1}H-^{15}N cross peaks of Gly-37, His-46, and Glu-91 in my ^{1}H-^{15}N HSQC spectra. The regions exhibiting backbone ^{1}H chemical shifts differences greater than 0.1 ppm compared to those of reduced holo-azurin are mapped on the apo-azurin structure (Figure IV.6). Most of these regions are located near the Cu binding ligands. This is discussed further in the Discussion section of this chapter.

Figure IV.4. Fingerprint region of WATERGNOE spectrum of apo-azurin in 90% H_2O recorded at 40 °C. Some of the sequential connectivities are shown.

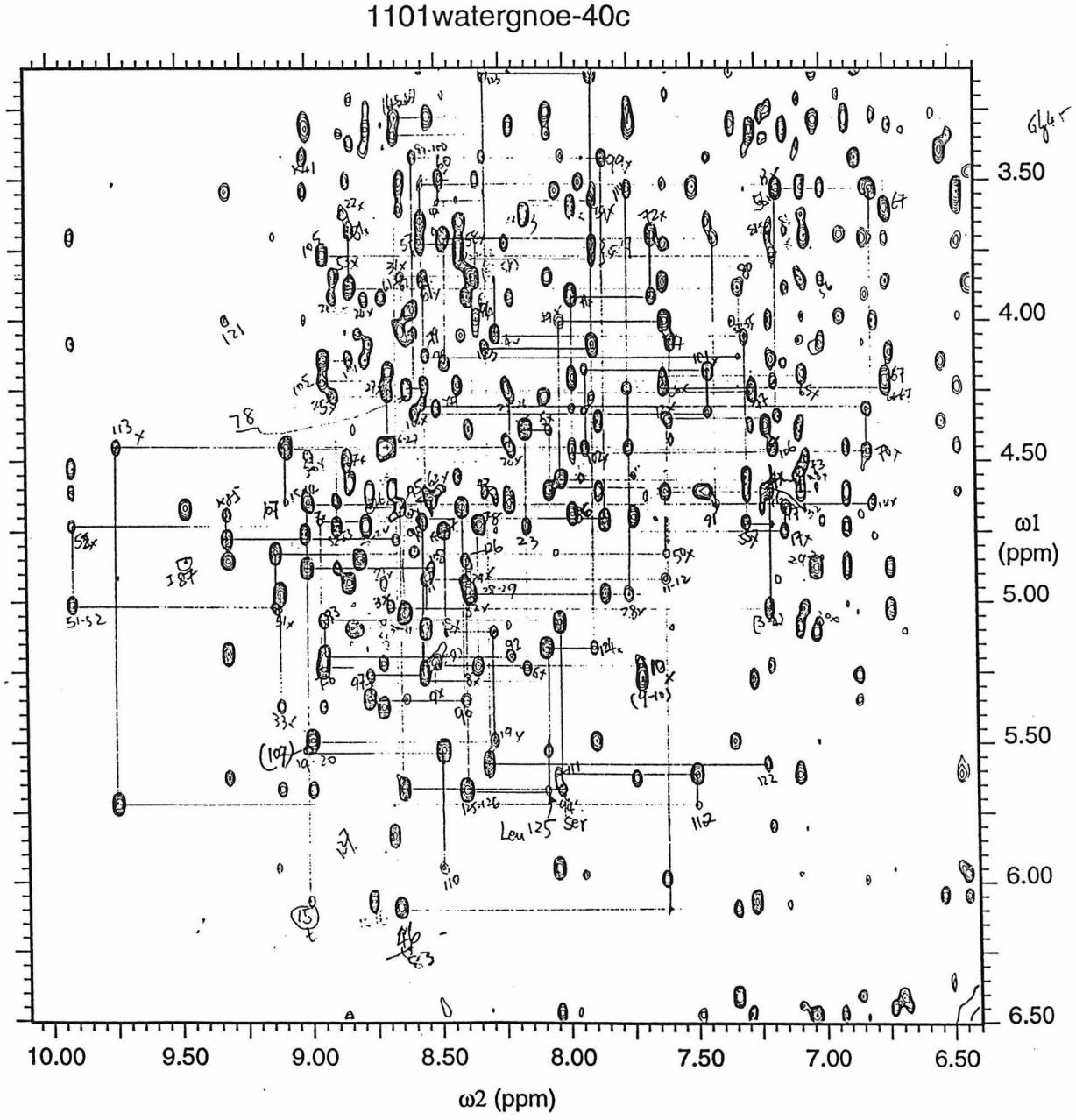

1101watergnoe-40c
10.00
9.50
9.00
8.50
8.00
7.50
7.00
6.50
ω2 (ppm)
3.50
4.00
4.50
5.00
5.50
6.00
6.50
ω1
(ppm)
Leu 125 Ser

Figure IV.5. ^{1}H-^{15}N HSQC spectrum of ^{15}N labeled apo-azurin recorded at 40 °C.

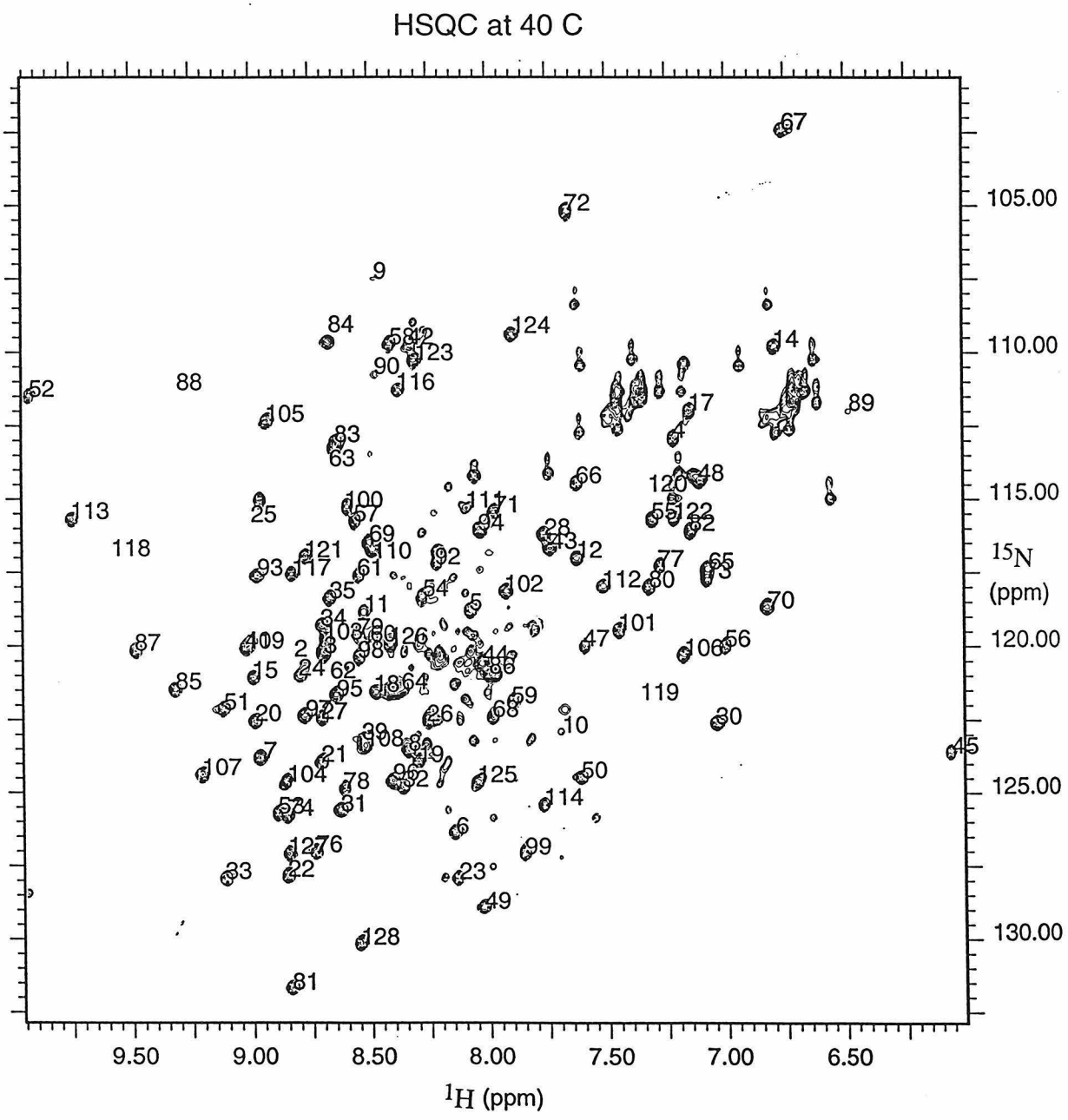

HSQC at 40 C
105.00
110.00
115.00
120.00
125.00
130.00
15N (ppm)
9.50
9.00
8.50
8.00
7.50
7.00
6.50
1H (ppm)

Table IV.1: ^{15}N and ^{1}H Chemical Shifts (ppm) for the Backbone of Apo-azurin from *P. aeruginosa* at pH 5.0, 40°C, in 2.5 mM Sodium Phosphate and 25 mM Sodium Chloride

residue	amide ^{15}N	amide NH ^{1}H	CαH ^{1}H
Ala-1			
Glu-2	120.7	8.78	4.45
Cys-3	120.2	8.70	5.02
Ser-4	112.9	7.20	4.60
Val-5	118.8	8.08	4.39
Asp-6	126.4	8.17	5.23
Ile-7	123.8	8.96	4.72
Gln-8	123.5	8.34	5.22
Gly-9	107.5	8.52	5.27
Asn-10	122.8	7.73	5.22
Asp-11	118.9	8.54	4.92
Gln-12	117.1	7.62	4.35
Met-13	114.1	7.17	3.53
Gln-14	109.8	6.81	4.65
Phe-15	121.1	9.01	6.09
Asn-16	117.4	8.64	4.67
Thr-17	112.0	7.15	4.75
Asn-18	121.6	8.48	5.11
Ala-19	124.0	8.32	5.50
Ile-20	122.6	9.05	4.47
Thr-21	124.0	8.78	4.93
Val-22	127.9	8.88	3.63
Asp-23	128.0	8.24	4.73
Lys-24	121.1	8.84	3.92
Ser-25	115.1	8.96	4.26
Cys-26	122.5	8.26	4.46
Lys-27	122.5	8.78	4.23
Gln-28	116.2	7.79	4.99
Phe-29	121.5	8.40	4.89

Table IV.1: (Continued)

residue	amide ^{15}N	amide NH ^{1}H	CαH ^{1}H
Thr-30	122.6	7.02	5.05
Val-31	125.6	8.65	3.85
Asn-32	124.9	8.41	4.99
Leu-33	127.9	9.14	5.38
Ser-34	119.3	8.76	5.05
His-35	118.5	8.78	6.45
Pro-36			
Gly-37		8.84	3.69
Asn-38	119.0	11.12	4.92
Leu-39	123.2	8.56	4.89
Pro-40			
Lys-41	120.1	9.02	3.42
Asn-42	109.7	8.33	4.61
Val-43	116.7	7.75	4.59
Met-44	120.6	8.03	3.99
Gly-45	123.6	6.08	3.27, 3.33
His-46		8.65	6.10
Asn-47	120.0	7.62	4.62
Trp-48	114.4	7.10	4.57
Val-49	128.9	8.03	4.00
Leu-50	124.5	7.62	4.83
Ser-51	122.2	9.15	5.01
Thr-52	111.5	9.93	4.73
Ala-53	125.7	8.90	3.85
Ala-54	118.4	8.29	4.05
Asp-55	115.7	7.31	4.71
Met-56	120.0	7.00	3.53
Gln-57	115.7	8.56	3.72
Gly-58	109.7	8.43	3.79
Val-59	121.9	7.90	3.58

Table IV.1: (Continued)

residue	amide ^{15}N	amide NH ^{1}H	CαH ^{1}H
Val-60	119.8	8.49	3.49
Thr-61	117.6	8.55	3.86
Asp-62	120.7	8.64	4.67
Gly-63	113.2	8.64	3.49, 4.04
Met-64	121.5	8.35	3.98
Ala-65	117.4	7.09	4.20
Ser-66	114.5	7.62	4.24
Gly-67	102.4	6.76	3.60, 4.21
Leu-68	122.4	7.98	2.76
Asp-69	116.5	8.56	4.30
Lys-70	118.7	6.83	4.46
Asp-71	115.5	7.99	3.91
Tyr-72	105.2	7.67	3.69
Leu-73	117.7	7.08	4.49
Lys-74	125.8	8.85	4.53
Pro-75			
Asp-76	127.0	8.69	4.19
Asp-77	117.3	7.28	4.25
Ser-78	124.8	8.62	4.25
Arg-79	119.6	8.54	4.13
Val-80	118.0	7.32	3.88
Ile-81	131.7	8.84	3.67
Ala-82	116.1	7.15	4.67
His-83	113.0	8.86	5.85
Thr-84	109.7	8.67	4.78
Lys-85	121.5	9.32	4.69
Leu-86	121.0	7.98	4.67
Ile-87	120.1	9.51	4.86
Gly-88	111.3	9.12	4.08, 3.70
Ser-89	112.0	6.43	3.68

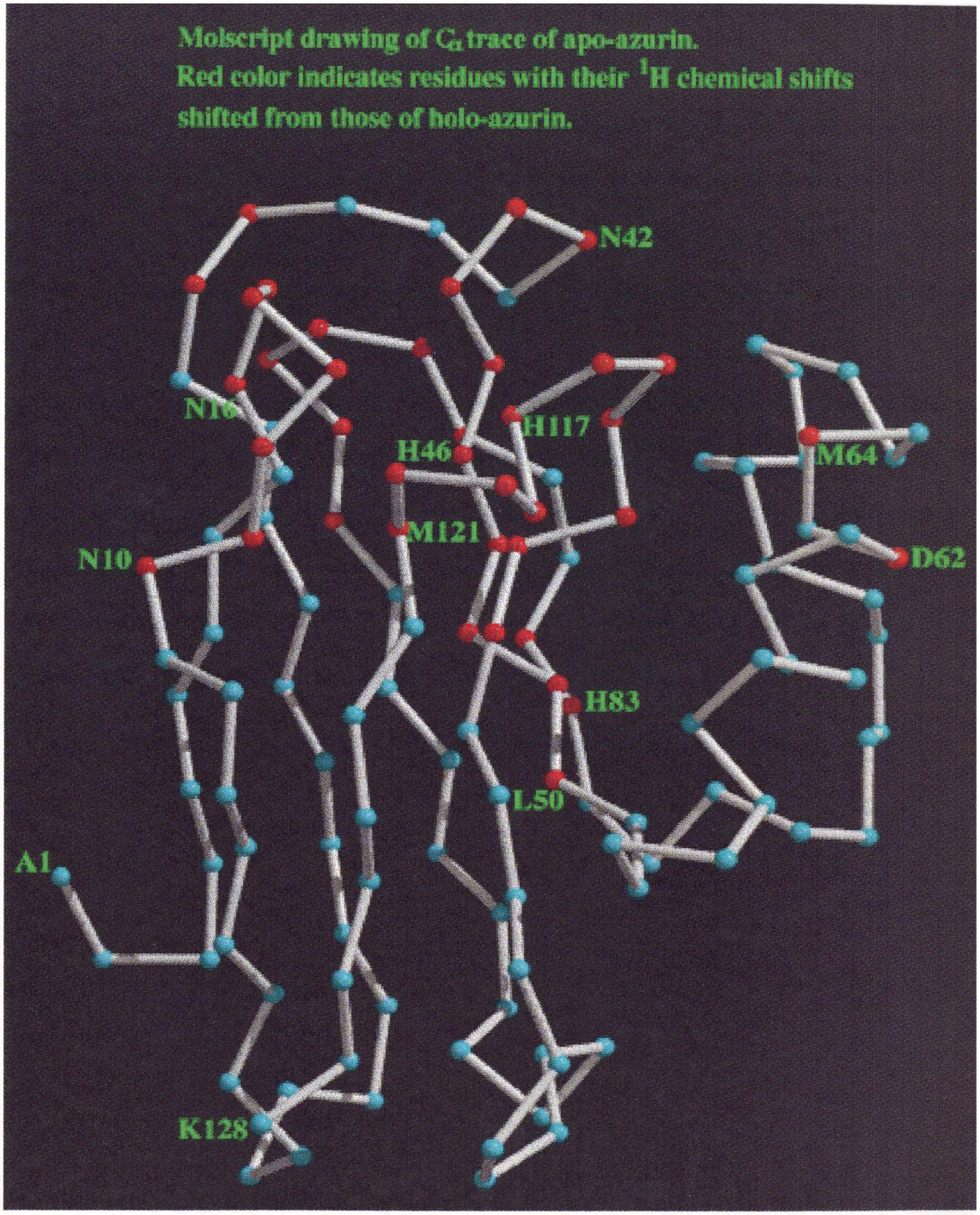
Table IV.1: (Continued)

residue	amide ^{15}N	amide NH ^{1}H	CαH ^{1}H
Gly-90	110.8	8.48	3.70, 4.15
Glu-91		7.42	4.65
Lys-92	117.2	8.22	5.19
Asp-93	117.7	8.96	5.07
Ser-94	116.1	8.02	5.67
Val-95	121.7	8.65	4.67
Thr-96	123.5	8.39	5.36
Phe-97	122.4	8.78	5.27
Asp-98	120.4	8.56	4.71
Val-99	127.1	7.86	3.42
Ser-100	115.3	8.59	4.33
Lys-101	119.5	7.46	4.17
Leu-102	118.2	7.93	4.45
Lys-103	119.8	8.68	4.58
Glu-104	119.8	8.86	4.15
Gly-105	112.3	8.94	3.76, 4.23
Glu-106	120.4	7.19	4.45
Gln-107	124.4	9.09	4.64
Tyr-108	123.4	8.53	4.87
Met-109	119.9	9.02	5.53
Phe-110	116.8	8.50	5.95
Phe-111	115.3	8.05	5.62
Cys-112	118.0	7.51	5.73
Thr-113	115.7	9.76	4.46
Phe-114	125.4	7.75	3.53
Pro-115			
Gly-116	111.3	8.38	3.83, 4.39
His-117	117.5	8.83	5.09
Ser-118	116.2	9.58	4.56
Ala-119	117.0	7.09	4.50

Table IV.1: (Continued)

residue	amide ^{15}N	amide NH ^{1}H	CαH ^{1}H
Leu-120	114.0	7.31	4.61
Met-121	116.9	8.78	4.66
Lys-122	115.6	7.23	5.58
Gly-123	110.3	8.324	3.12, 4.10
Thr-124	109.4	7.89	5.18
Leu-125	124.7	8.09	5.67
Thr-126	119.9	8.41	4.85
Leu-127	127.1	8.84	5.09
Lys-128	130.2	8.56	5.24

Figure IV.6. Molscript drawing of the C_α trace for apo-azurin structure; residues showing ^1H chemical shifts significantly shifted from those of holo-azurin are highlighted (red color).


Molscript drawing of C_α trace of apo-azurin.
Red color indicates residues with their ^1H chemical shifts
shifted from those of holo-azurin.
N42
N16
H117
H46
M64
M121
N10
D62
H83
L50
A1
K128

Backbone Dynamics In the model-free formalism of protein backbone dynamics introduced by Lipari and Szabo (Lipari and Szabo, 1982a; Lipari and Szabo, 1982b), the backbone dynamics of a protein can be described by three parameters: an overall correlation time τ_c, a generalized order parameter S^2, and an effective correlation time τ_e. The overall correlation time τ_c describes the overall motion, and both S and τ_e describe the rapid internal motion. The generalized order parameter S is a measure of the degree of spatial restriction of the internal motion, and the effective correlation time τ_e is a measure of the rate of the motion.

These parameters can be related to the experimentally measurable parameters: spin-lattice relaxation time T_1, spin-spin relaxation time T_2, and the NOE factor in NMR experiments (Abragam, 1961). When dipolar coupling to a directly bonded proton and chemical shift anisotropy are the dominant relaxation mechanisms (which is true under most circumstances), then T_1, T_2, and the NOE enhancement of an amide ^{15}N are determined by the spectral density function $J(\omega_i)$ as follows:

$$\frac{1}{T_1} = d^2\left[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)\right] + c^2 J(\omega_N) \qquad [1]$$

$$\frac{1}{T_2} = 0.5\,d^2\left[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)\right] + c^2\frac{\left[3J(\omega_N) + 4J(0)\right]}{6} \qquad [2]$$

$$\text{NOE} = 1 + T_1\left(\frac{\gamma_H}{\gamma_N}\right) d^2\left[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\right] \qquad [3]$$

where

$$d^2 = 0.1\,\gamma_H{}^2\,\gamma_N{}^2\left(\frac{h}{2\pi}\right)^2\left\langle r_{HN}{}^{-3}\right\rangle^2 \quad \text{and} \qquad [4]$$

$$c^2 = \left(\frac{2}{15}\right)\omega_N{}^2\left(\sigma_\parallel - \sigma_\perp\right)^2 \qquad [5]$$

$$\frac{h}{2\pi} = \frac{\text{Planck's constant}}{2\pi} = 1.054494\times10^{-27}\,\text{erg sec}$$

γ_H = gyromagnetic ratio of ^1H = $2.6753\times10^4 \frac{\text{rad}}{\text{sec } G}$

γ_N = gyromagnetic ratio of ^{15}N = $-2.71\times10^3 \frac{\text{rad}}{\text{sec } G}$

ω_H = nuclear ^1H Larmor frequency = $2\pi\left(600.13\times10^6\right)$ Hz

ω_H = nuclear ^{15}N Larmor frequency = $2\pi\left(60.83\times10^6\right)$ Hz

r_{HN} = NH bond length = 1.02 Å or 1.02×10^{-8} cm from neutron diffraction[1]

$\sigma_\parallel$ and $\sigma_\perp$ = parallel and perpendicular components of the axially symmetric ^{15}N chemical shift tensor

$\left(\sigma_\parallel - \sigma_\perp\right)$ = -160×10^{-6} [2]

$J(\omega_i)$ = spectral density function

Plugging in these values, we obtain the following:

$$c^2 = 0.498\times10^9\left(\frac{\text{rad}}{\text{sec}}\right)^2$$

$$d^2 = 0.519\times10^9\left(\frac{\text{rad}}{\text{sec}}\right)^2$$

In the isotropic motion model, assuming macromolecules undergoing isotropic tumbling and independent overall (τ_c) and internal (τ_e) motion, the spectral density function has the form

$$J(\omega_i) = \frac{S^2\tau_c}{\left(1+\omega_i^2\tau_c^2\right)} + \frac{\left(1-S^2\right)\tau}{\left(1+\omega_i^2\tau^2\right)} \qquad [6]$$

where $\tau = \frac{\tau_c\tau_e}{\left(\tau_c+\tau_e\right)}$

In the anisotropic model, the spectral density is given by (Lapari and Szabo, 1982)

$$J(\omega_i) = A\left\{\frac{S^2\tau_1}{\left(1+\omega_i^2\tau_1^2\right)} + \frac{\left(1-S^2\right)\tau_{1e}}{\left(1+\omega_i^2\tau_1^2\right)}\right\} +$$

---

[1]Clore *et al.*, 1990, and references therein
[2]Hiyama *et al.*, 1988

$$(1-A)\left\{\frac{S^2\tau_2}{\left(1+\omega_i{}^2\tau_2{}^2\right)} + \frac{\left(1-S^2\right)\tau_{2e}}{\left(1+\omega_i{}^2\tau_{2e}{}^2\right)}\right\} \qquad [7]$$

where $\frac{1}{\tau_{1e}} = \frac{1}{\tau_1} + \frac{1}{\tau_e}$ and $\frac{1}{\tau_{2e}} = \frac{1}{\tau_2} + \frac{1}{\tau_e}$

For a particular amide, the mixing parameter A represents the weights of the two modes of rotations represented by the two correlation times τ_1 and τ_2. When internal motions are much faster than overall motions, the above spectral density applies to both axially symmetric and general anisotropic motions (Lipari and Szabo, 1982ab; Tjandra *et al.*, 1995).

For most folded globular proteins, the isotropic motion approximation is valid. In addition, for most amino acid residues in a folded globular protein, it will also be true that $\tau_e << \tau_R$, i.e., the internal motion is much faster than the protein molecule's overall tumbling. The second term in equation [6] can thus be ignored to give the simplified equation [8].

$$J(\omega_i) = \frac{S^2\tau_R}{\left(1+\omega_i{}^2\tau_R{}^2\right)} \qquad [8]$$

Using the simplified spectral density function in equation [8], $\frac{T_1}{T_2}$ ratio can be written as equation [9], which is only dependent on τ_c and independent of both S and τ_e.

$$\frac{T_1}{T_2} = \frac{\left\{0.5d^2\begin{bmatrix}4J(0)+J(\omega_H-\omega_N)+3J(\omega_N)\\ +\,6J(\omega_H)+6J(\omega_H+\omega_N)\end{bmatrix} + c^2\left[3J(\omega_N)+4J(0)\right]\right\}/6}{d^2\left[J(\omega_H-\omega_N)+3J(\omega_N)+6J(\omega_H+\omega_N)\right] + c^2J(\omega_N)} \qquad [9]$$

The overall correlation time τ_c can therefore be determined from ^{15}N $\frac{T_1}{T_2}$ ratio according to equation [9].

With τ_c determined, which clearly must be the same for all residues, S^2 and τ_e can be determined for each residue by fitting the observed T_1 and T_2 data to equations [1] and [2] and simultaneously optimizing S^2 and τ_e. Since the accuracy of measuring NOE is much lower than that of T_1 and T_2, only T_1 and T_2 data were used here for analysis of the relaxation data. NOE data can be used to confirm the determined $S^2$and τ_e parameters by comparing the calculated NOE using equation [3] with the observed NOE.

In my study, a series of HSQC-T1 and HSQC-T2 were acquired with gradually increasing delay times for T_1 or T_2. In order to measure the relaxation time of each backbone amide ^{15}N, each resolved ^{15}N-^{1}H cross peak in the processed spectrum (usually the one with zero delay time) was first manually marked by the command "*mark*" in VNMR when displaying the spectrum in intensity mode.[3] Peak volumes on each spectrum were then calculated in VNMR using the macro "*getxpk_2*"(Appendix IV-3). The peak volumes were then exported along with the time series to Kaleidograph (Macintosh Software) and fitted with single exponential functions $I_{(t)} = I_{(0)} \exp\left(\frac{-t}{T}\right)$ to get T_1 or T_2 for each peak. Some typical data with the exponential fittings are shown in Figure IV.7. The measured T_1 and T_2 values along with the derived S^2 are listed in Table IV.2. The approximation method discussed by Habazettl and Wagner (Habazettl and Wagner, 1995) was used here in calculating τ_c and the order parameter (S^2) for each ^{15}N. The approximation is based on the fact that for macromolecules $\omega_0 \tau_c >> 1$, spectral density functions $J(\omega_H \pm \omega_N) << J(\omega_N)$ or $J(0)$. Using this approximation and plugging in the numbers above ($c^2 = 0.498\times10^9 \left(\frac{\text{rad}}{\text{sec}}\right)^2$, $d^2 = 0.519\times10^9 \left(\frac{\text{rad}}{\text{sec}}\right)^2$, and $\omega_N = 60.8\times10^6$ Hz), equation [9] can be further simplified to equation [10]

---

[3]The "*mark*" command used here and the peak intensity calculation only work when displaying the spectra in intensity mode; they do not work when displaying the spectra in contour mode.

$$\tau_c^2 = 5\left\{2.055\left(\frac{T_1}{T_2}\right) - 2.46\right\} \quad \text{where } \tau_c \text{ is in ns} \qquad [10]$$

τ_c for each residue can thus be simply calculated using equation [10]. The calculated τ_c ranged from 3.2 ns to 9.0 ns with an average of 5.8 ns. If the τ_c values which deviate more than 20% from the mean value are excluded (11 out of 99), the average τ_c was still 5.8 ns. This τ_c value along with the simplified spectral density function (equation [8]) was applied to equations [1] and [2] to calculate the order parameters (equations [11] and [12]).

$$S^2 = \frac{1}{T_1}\frac{\left(1+0.146\,\tau_c^2\right)}{\left(2.055\,\tau_c\right)} = 0.497\frac{1}{T_1} \qquad [11]$$

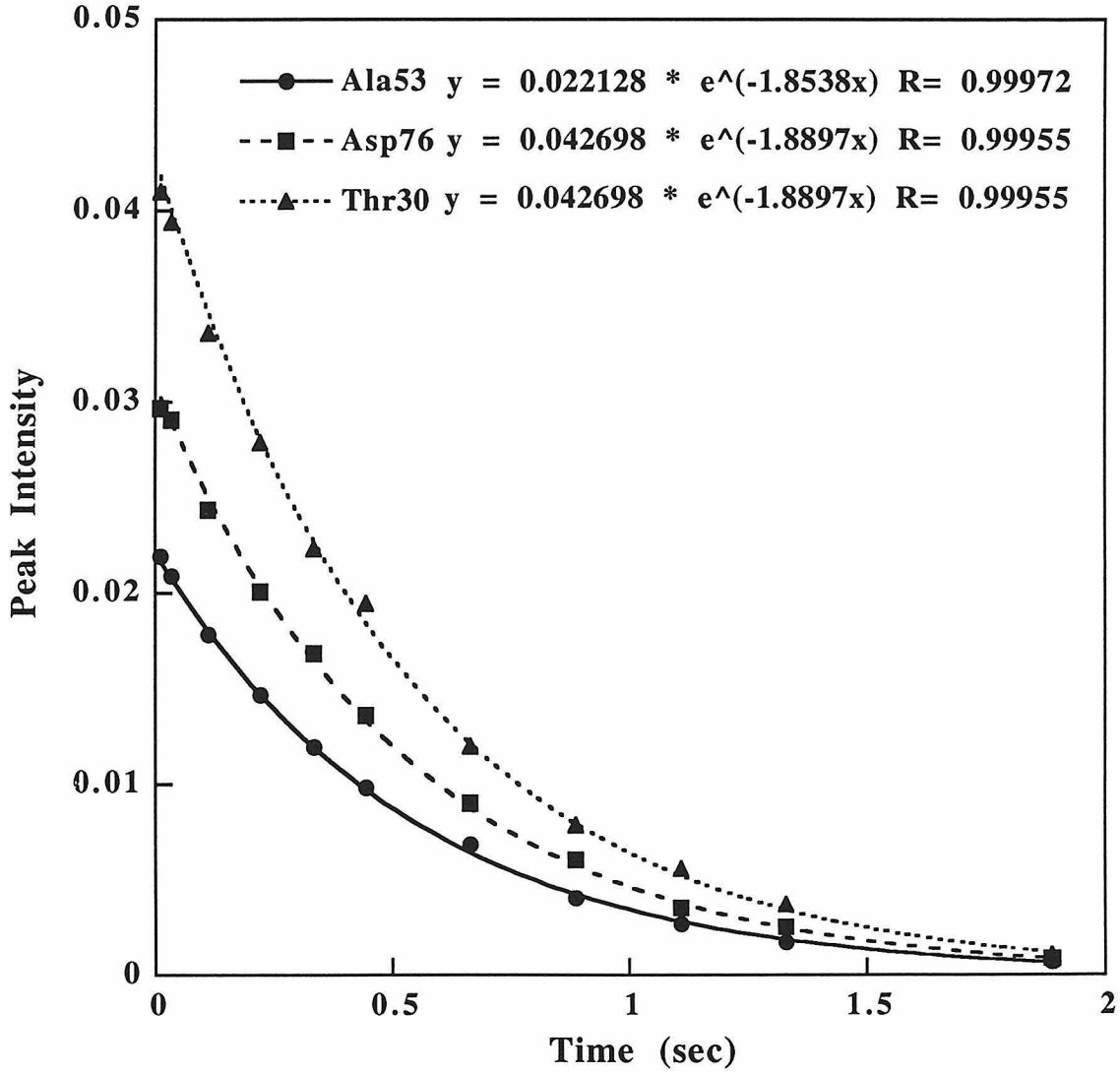
or

$$S^2 = \frac{1}{T_2}\frac{\left(1+0.146\,\tau_c^2\right)}{\left(2.46\,\tau_c + 0.2\,\tau_c^3\right)} = 0.1106\frac{1}{T_2} \qquad [12]$$

The order parameter (S^2) for each amide ^{15}N can thus be determined from either T_1 or T_2 data. The values reported in Table IV.2 are the average of the S^2 values calculated from T_1 and T_2 . For most residues, the S^2 values determined from T_1 or T_2 agree with each other fairly well. For a few residues, abnormally high S^2 values (over 1) result from either T_1 or T_2 data alone (probably due to experimental error). For these residues, only the acceptable S^2 values were used (without averaging). S^2 values are plotted against the residue numbers in Figure IV.8 to show the distribution of the low order parameter regions. The S^2 values are also mapped onto the backbone of the apo-azurin crystal structure (Figure IV.9) to show the location of the regions with low order parameters.

Figure IV.7. Typical plots for (a) T_1 and (b) T_2 measurements.

(a)


Ala53 y = 0.022128 * e^(-1.8538x) R= 0.99972
Asp76 y = 0.042698 * e^(-1.8897x) R= 0.99955
Thr30 y = 0.042698 * e^(-1.8897x) R= 0.99955
Peak Intensity
0.05
0.04
0.03
0.02
0.01
0
0
0.5
1
1.5
2
Time (sec)

(b)

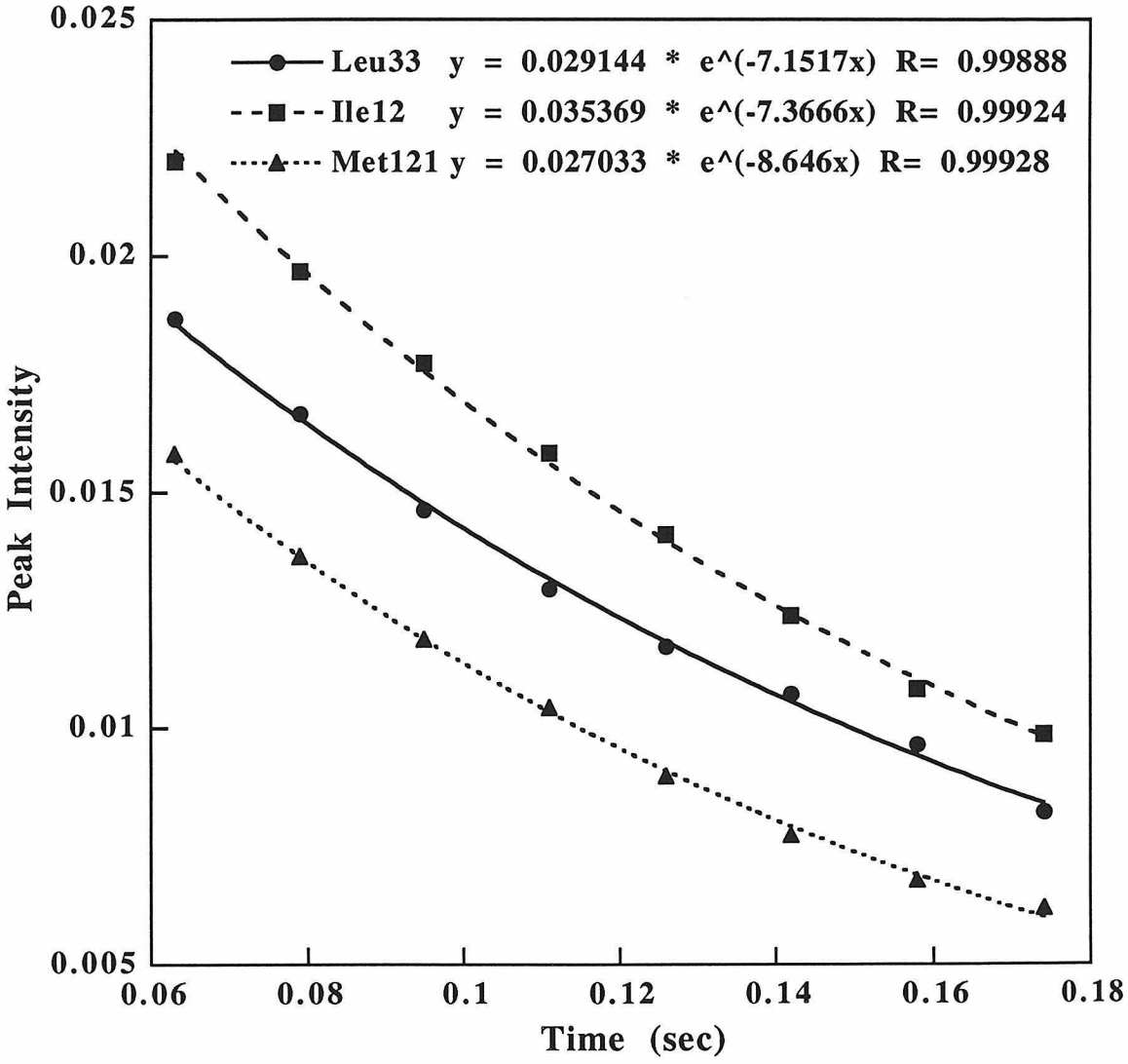

Leu33 y = 0.029144 * e^(-7.1517x) R= 0.99888
Ile12 y = 0.035369 * e^(-7.3666x) R= 0.99924
Met121 y = 0.027033 * e^(-8.646x) R= 0.99928
Peak Intensity
0.025
0.02
0.015
0.01
0.005
0.06
0.08
0.1
0.12
0.14
0.16
0.18
Time (sec)

Table IV.2: Relaxation Times T_1, T_2, and Order Parameters S^2 for Apo-azurin at 30 °C

residue number	T_1 (sec)	T_2 (sec)	S^2
3	0.617	0.153	0.76
4	0.538	0.0820	0.92
5	0.538	0.111	0.92
6	0.562	0.139	0.84
7	0.467	0.136	0.82
8	0.515	0.139	0.88
11	0.552		0.90
12	0.556	0.188	0.74
14	0.532	0.112	0.93
15	0.658	0.0883	0.76
16	0.503	0.0803	0.98
17	0.595	0.121	0.87
18	0.481	0.137	0.81
19	0.546		0.85
20	0.571	0.152	0.80
21	0.599	0.112	0.83
22	0.513	0.147	0.86
23	0.549	0.139	0.85
24	0.556	0.0604	0.89
25	0.476	0.120	0.92
26	0.459	0.122	0.91
27	0.485	0.136	0.82
28	0.513	0.142	0.87
30	0.526	0.0901	0.94
31	0.529	0.108	0.94
32	0.532	0.203	0.74
33	0.481	0.140	0.79
34	0.515	0.113	0.97
39	0.488	0.182	0.61
40	0.485	0.133	0.83

Table IV.2: (Continued)

residue number	T_1 (sec)	T_2 (sec)	S^2
42	0.503	0.127	0.87
43	0.538	0.0941	0.92
44	0.549	0.120	0.78
45	0.637	0.141	0.78
47	0.559	0.104	0.89
48	0.559	0.0976	0.89
49	0.556	0.106	0.89
50	0.503	0.117	0.94
51	0.483	0.126	0.88
52	0.513	0.165	0.82
53	0.541	0.136	0.87
54	0.552	0.104	0.90
55	0.552	0.124	0.90
56	0.543	0.103	0.91
57	0.535	0.148	0.84
58	0.508	0.134	0.90
59	0.532	0.111	0.93
60	0.543	0.111	0.91
61	0.826	0.115	0.78
63	0.465	0.135	0.82
64	0.546	0.249	0.68
66	0.556	0.108	0.89
67	0.595	0.177	0.73
68	0.568	0.0927	0.87
70	0.595	0.117	0.89
71	0.565	0.0898	0.88
72	0.571	0.110	0.87
73	0.549	0.100	0.90
74	0.495	0.138	0.80
76	0.529	0.140	0.87

Table IV.2: (Continued)

residue number	T_1 (sec)	T_2 (sec)	S^2
77	0.599	0.133	0.83
78	0.621	0.153	0.76
80	0.581	0.146	0.81
81	0.610	0.152	0.77
82	0.515	0.106	0.96
83	0.556	0.0947	0.89
84	0.510	0.143	0.87
85	0.541	0.103	0.92
86	0.538	0.129	0.89
87	0.602	0.0881	0.83
92	0.549	0.0910	0.90
93	0.637	0.130	0.82
94	0.565	0.105	0.88
95	0.552	0.117	0.90
96	0.562	0.118	0.91
97	0.532	0.159	0.82
99	0.485	0.158	0.70
100	0.526	0.118	0.94
101	0.474	0.137	0.81
102	0.581	0.108	0.86
103	0.559	0.142	0.83
104	0.581	0.119	0.89
105	0.562	0.111	0.88
106	0.585	0.0908	0.85
107	0.616	0.121	0.86
108	0.503	0.0864	0.98
109	0.521	0.102	0.95
111	0.515	0.128	0.91
112	0.592	0.0948	0.84
113	0.441	0.128	0.86

Table IV.2: (Continued)

residue number	T_1 (sec)	T_2 (sec)	S^2
114	0.562	0.163	0.78
116	0.585	0.177	0.74
117	0.495	0.150	0.74
121	0.654	0.0994	0.76
122	0.588	0.116	0.85
123	0.535	0.134	0.83
125	0.585	0.133	0.84
126	0.541	0.103	0.92
127	0.526	0.102	0.94
128	0.538	0.132	0.88

Figure IV.8. Plot of the order parameter (S^2) vs. the residue number.

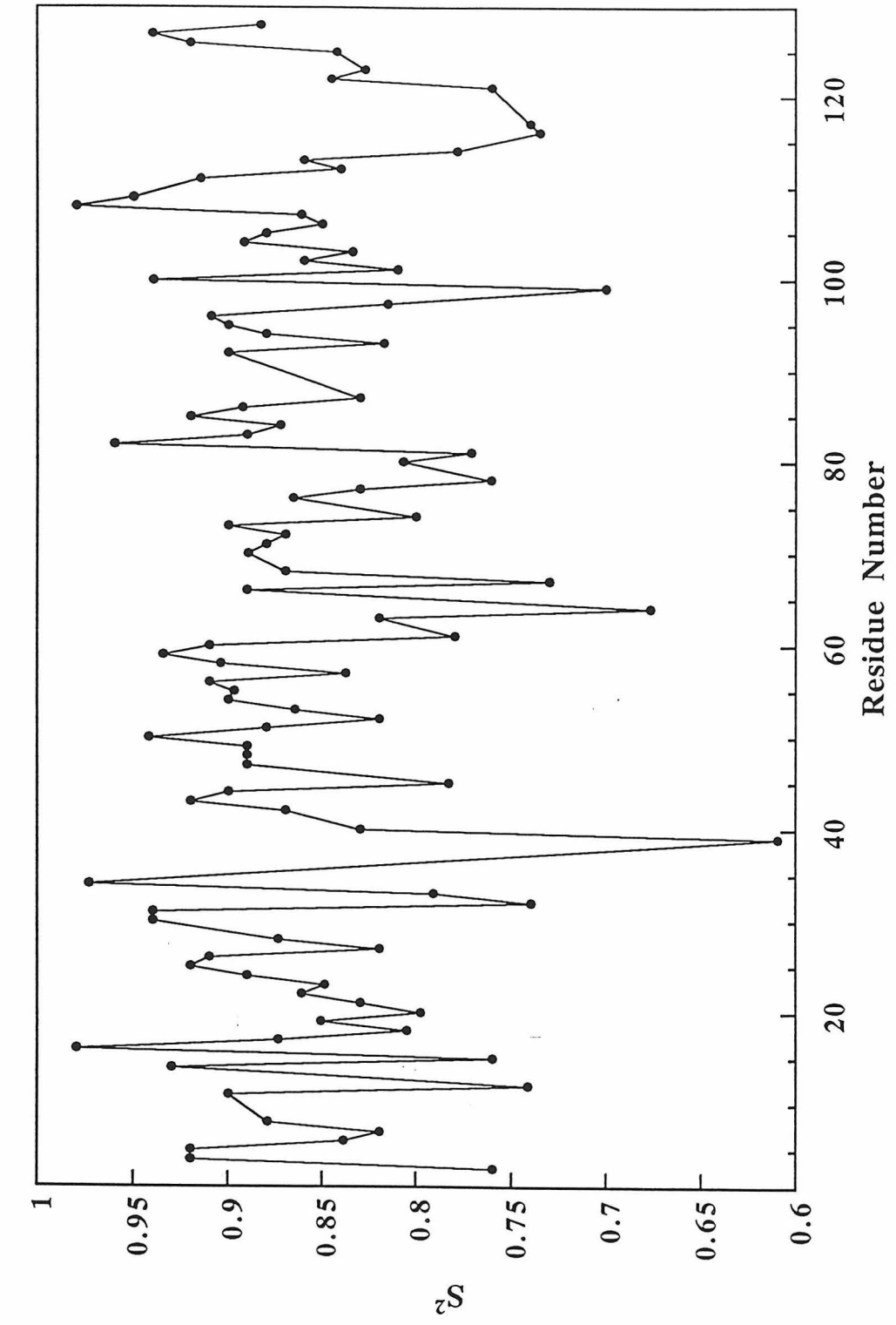
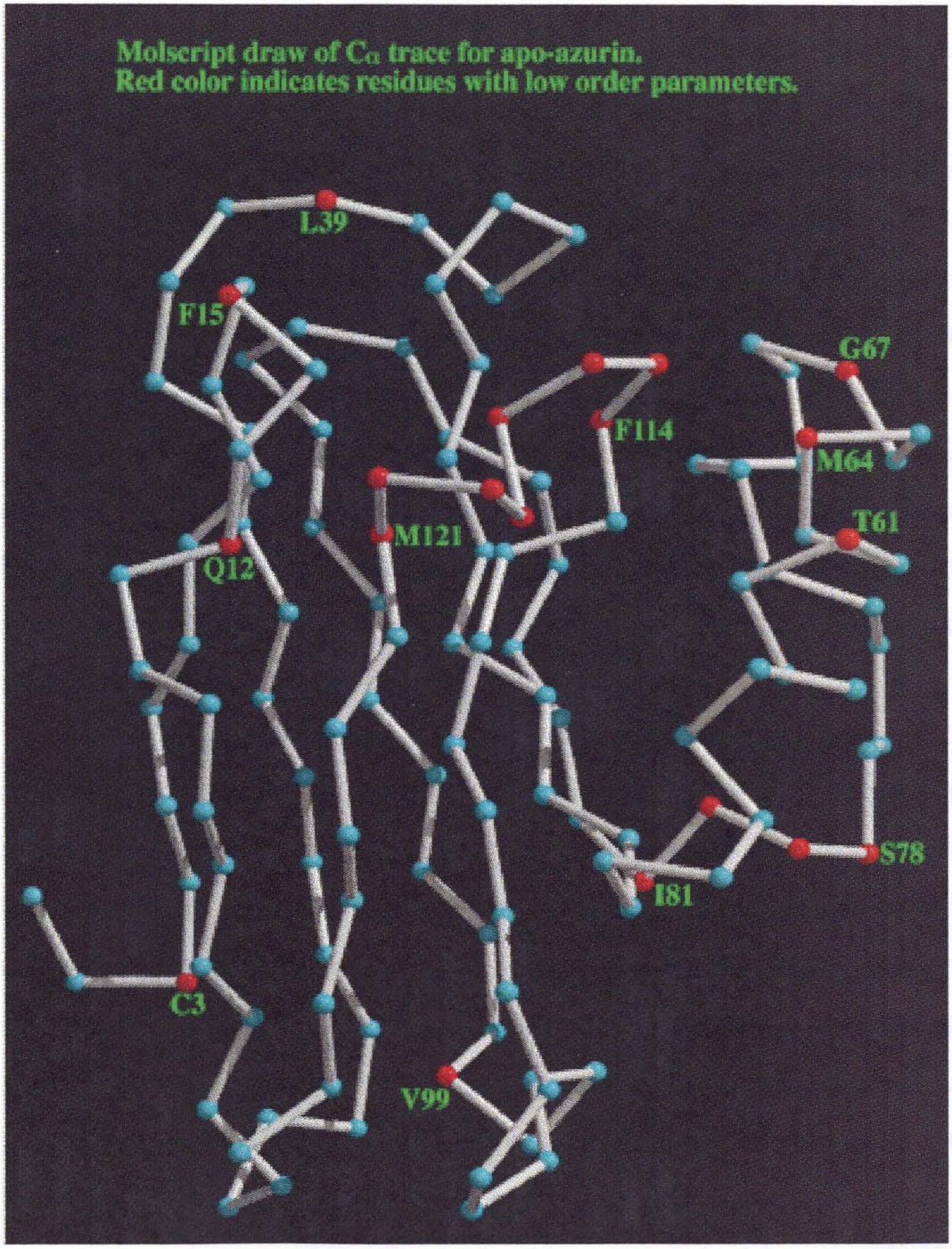

1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
S^2
20
40
60
80
100
120
Residue Number

Figure IV.9. Molscript drawing of the C_α trace for apo-azurin structure; residues showing low order parameters (below 0.8) are highlighted (red color).


Molscript draw of Cα trace for apo-azurin.
Red color indicates residues with low order parameters.
L39
F15
G67
F114
M64
M121
T61
Q12
S78
I81
C3
V99

DISCUSSIONS

Several interesting points are worth discussion here. (1) ^{1}H and ^{15}N chemical shifts of the apo-azurin backbone were assigned based on a series of 2D NMR spectra of apo-azurin where reported assignments for reduced holo-azurin were used as a reference. Differences in the backbone ^{1}H chemical shifts were relatively limited, as expected from the similarity of the crystal structures of apo- and holo-azurin. As shown in Figure IV.6, these differences were mainly limited to the regions near the Cu binding ligands. The chemical shift difference observed for residues Asn-10 to Asn-16 are believed to be due to the conformational change of the side chain of Met-13; this is caused by the 1.6 Å shift in nearby His-117 which can be seen in apo-azurin's crystal structure. Higher dimensional NMR spectra are needed to resolve the overlapping peaks in the aliphatic regions, to obtain more complete ^{1}H and ^{15}N chemical shift assignments, and to see the changes in the side chains.

(2) The relaxation properties of apo-azurin were measured. Since the overall correlation time τ_c is proportional to the molecular size, the 5.8 ns τ_c obtained for apo-azurin at 30 °C is in line with the τ_c values for similar size proteins reported in the literature. A τ_c of 3.3 ns was reported for the IgG binding domain of streptococcal protein G (56 mer) at 26 °C, 4.1-4.3 ns was found for Eglin C (69 mer), Calbindin D9K (75 mer), and Ubiquitin (76 mer), 6.5 ns for SH2-complex (a 12.3 kD protein with a 1.5 kD peptide) at 30 °C (Farrow *et al.*, 1994, and references therein), 8.13 ns for phosphotransfer domain of the Histidine Kinase CheA (134 mer) at 30 °C (Zhou *et al.*, 1995), and 8.3 ns for interleukin-1b (17.4 kD) (Clore *et al.*, 1990).

(3) The order parameters (S^2) for 99 residues were derived from the relaxation data analyzed with the Lapari and Szabo isotropic dynamic motion model. S^2 were typically in the range of 0.8 - 0.95, suggesting an overall restricted motion of backbone. The S^2

values are quite uniform in regions of well-defined β-sheet secondary structure. A few residues had S^2 values lower than 0.8. These were Cys-3, Gln-12 and Phe-15, Leu-39, Gly-67, Ser-78 to Ile-81, Val-99, and Phe-114 to Met-121, corresponding to either turn regions or the end of β-strand or α-helix secondary structures. Interestingly, Thr-61 and Met-64 in the middle of the α-helix also had low order parameters, suggesting a relatively flexible helix in apo-azurin. High mobility in the middle of α-helix has also been observed in Calmodulin by Barbato *et al.*(Barbato *et al.*, 1992).

(4) The loop containing three Cu binding ligands (Cys-112, His-117, and Met-121) exhibit low order parameters. This probably is because of the absence of Cu(II) center in apo-azurin. A flexible loop region nearby His-117 of apo-azurin in solution is consistent with the "swinging door" metal incorporation mechanism suggested by Nar *et al.* in their crystal structure study (Nar *et al.*, 1992). In the crystal structure of apo-azurin, two different molecular forms of apo-azurin are observed. One form (form 1) closely resembles the holo-azurin lacking copper. The other form (form 2) shows differences in the metal binding site region induced by the incorporation of a solvent molecule into the site. The positions of copper ligands His-46 and His-117 are shifted by 0.6 Å and 1.6 Å respectively. The large shift of His-117 is accompanied by shifts of the adjacent residues Phe-114, Pro-115, and Gly-116, as well as a conformational change of the side chain of the nearby residue Met-13. The 0.6 Å bend of His-46 side chain also rearranges its surrounding residues (Gln-8 to Asn-10). The rest of the azurin molecule is almost superimposable. In our NMR study, all the regions affected in the crystal structure exhibit low order parameters and significant ^{1}H chemical shift difference in solution as shown in Figure IV.6 and Figure IV.9. This suggests a fast equilibrium interconversion between the two forms observed in the crystal structure in solution. Although we were not able to directly measure the order parameter for Met-13, nearby residues Gln-12 and Phe-15 both

exhibited low order parameters, 0.74 and 0.76 respectively. Residue Gln-14 still had high order parameter 0.93 despite relative high mobility of nearby residues. This is probably because of the strong H-bonding formation between the amide proton of Gln-14 to carbonyl oxygen of Asn-10 in the turn.

REFERENCES

1. Abragam, A. *The Principles of Nuclear Magnetism* (Clarendon Press, Oxford, England, 1961).

2. Barbato, G., Ikura, M., Kay, L.E., Pastor, R.W., and Bax, A. (1992) *Biochemistry* **31**, 5269-5278.

3. Clore, G.M. and Gronenborn, A.M. (1993) *NMR of Proteins* (CRC Press, Boca Raton).

4. Clore, G.M., Driscoll, P.C., Wingfield, P.T., and Gronenborn, A.M. (1990) *Biochemistry* **29**, 7387-7401.

5. Farrow, N., Muhandiram, R., Singer, A., Pascal, S., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994) *Biochemistry* **33**, 5984-6003.

6. Griesinger, C., Otting, G., Wuthrich, K., and Ernst, R.R. (1988) *J. Am. Chem. Soc.* **110**, 7870-7872.

7. Habazettl, J. and Wagner, G. (1995) *J. Mag. Resonance, Series B* **109**, 100-104.

8. Hiyama, Y., Niu, C., Silverton, J.V., Bavaso, A., and Torchia, D.A. (1988) *J. Am. Chem. Soc.* **110**, 2378-2383.

9. Lipari, G. and Szabo, A. (1982a) *J. Am. Chem. Soc.* **104**, 4546-4559.

10. Lipari, G. and Szabo, A. (1982b) *J. Am. Chem. Soc.* **104**, 4559-4570.

11. Nar, H., Messerschmidt, A., Huber, R., van de Kamp, M., and Canters, G.W. (1992) *FEBS* **306**, 119-124.

12. Shaka, A.J., Lee, C.J., and Pines, A. (1988) *J. Magn. Reson.* **77**, 274.

13. Tjandra, N., Feller, S.E., Paster, R.W., and Bax, A. (1995) *J. Am. Chem. Soc.* **117**, 12562-12566.

14. van de Kamp, M., Canters, G.W., Wijmenga, S.S., Lommen, A., Hilbers, C.W., Nar, H., Messerschmidt, A., and Huber, R. (1992) *Biochemistry* **31**, 10194-10207.

15. Wuthrich, K. *NMR of Proteins and Nucleic Acids* (John Wiley and Sons, New York, 1986).

16. Zhang, O. *et al.* (1994) *J. Biomolecular NMR* **4**, 845-858.

17. Zhou, H., Lowry, D.F., Swanson, R.V., Simon, M.I., and Dahlquist, F.W. (1995) *Biochemistry* **34**, 13858-13870.

Chapter V

Structural Basis for Protein Stabilization by Osmolytes

SUMMARY

High concentrations of osmolytes has been proposed as a stress-adaptive mechanism for organisms exposed to high temperature, desiccation, or high levels of urea. I investigated this by studying the effect of glycine on the hydrogen/deuterium (H/D) exchange rates of individual protons in Ribonuclease A (RNase A) using 2D ^{1}H NMR. 37 amide protons in RNase A were measured in the presence of different concentrations of glycine. Two types of amide proton exchange behavior were observed. For type I amide protons, the H/D exchange rates become slower in the presence of high concentrations of glycine, as expected from global stability changes. The H/D exchange rates of type II amide protons, however, are not affected in the presence of 2.0 M glycine.

INTRODUCTION

It is well known that some naturally occurring organic solutes called osmolytes can significantly increase the stability of proteins without substantially altering their catalytic activity (Arakawa and Timasheff, 1985; Santoro *et al.*, 1992). It has been proposed that organisms and cellular systems use this as a mechanism to adapt to stresses such as high temperature, desiccation, and urea-concentrating environments (Yancey *et al.*, 1982; Timasheff, 1993). They can concentrate a particular osmolyte in their system without experiencing a deleterious effect in biological function. Studies from Timasheff's group revealed that the stabilization of the native structure of these proteins by the osmolytes is accompanied by the preferential hydration of the proteins (Arakawa and Timasheff, 1983). According to this mechanism, the unfavorable interactions of the osmolytes with the proteins (manifested by negative values of the preferential interaction parameter) is a reflection of an increase in the surface free energy of water induced by these additives, hence of the surface tension of water. This results in the exclusion of the additive from the water-macromolecule interface. By the Le Chatelier principle, the system should tend toward a state minimizing the additive effect, i.e., toward a minimization of the area of the water-protein interface. Globular proteins in the native state have a smaller surface area than in the denatured state. Hence, contact with these solvent systems should displace this native <=> denatured equilibrium to the left, resulting in the stabilization of the native structure.

Bolen and coworkers recently used Differential Scanning Calorimetery (DSC) to study the effects of glycine-based osmolytes in stabilizing Ribonuclease A (RNase A) and hen egg white lysozyme (Hew lysozyme) against thermal unfolding (Santoro *et al.*, 1992). They observed an increase of up to 22 °C in the T_m (the middle point unfolding transition temperature) of RNase A in the presence of high concentrations of sarcosine

(N-methyglycine), but the unfolding enthalpy changes were independent of T_m's. This suggested that the increase in RNase A stabilization with increasing osmolyte concentration is an entropy effect. More recently, Plaza del Pino and Sanchez-Ruiz studied the osmolyte effect on the heat capacity change for protein folding (Plaza de Pino and Sanchez-Ruiz, 1995). However, they found that the denaturation heat capacity increases with osmolyte concentration. This finding confirms the preferential hydration mechanism originally proposed by Timsheff and coworkers.

In this study, I used the hydrogen/deuterium (H/D) exchange technique to look at the effect of glycine on RNase A. My objective was to provide an assessment of the glycine stabilization in terms of a local effect by measuring H/D exchange rate changes for RNase A backbone amide protons. I hoped this would provide some insight into the relationship between local and global stabilizing effects.

I chose glycine as the model osmolyte and RNase as the model protein. Among all the glycine-based osmolytes, glycine exhibits the greatest effect at concentrations below 2.0 M. RNase A was chosen since it has been very well studied and has good thermal unfolding reversibility (Santoro *et al.*, 1992). In addition, Matthews and Leatherbarrow have recently reported that up to two molar concentration of glycine can be added to aqueous protein NMR samples without altering the NMR spectra and the folded three-dimensional structure (Matthews and Leatherbarrow, 1993).

MATERIALS AND METHODS

Bovine pancreatic RNase A was purchased from Sigma at the highest grade (type XII-A), and was further purified by a home-packed SP Sepharose ion exchange column (Pharmacia). RNase A thermal stability was measured on an Aviv 62DS CD spectrometer. The thermal unfolding of 35 μM RNase A at pH 5.0 was monitored at 220 nm and equilibrated for 3 minutes at every 2 °C step increment. NMR samples were prepared from desalted and lyophilized protein powder by dissolving in glycine D_2O solution (pD 5.0). 2D ^1H COSY spectra with spectral width of 9640 Hz and 1K complex points were acquired on a Varian Unity Plus 600 MHz NMR spectrometer at 34 °C. The data were then processed using Felix software (Biosym, Inc) on a Sun workstation. The macro used for automatic data processing and peak volume calculation in Felix is attached in Appendix V-I.

RESULTS

I first confirmed the stabilizing effect of glycine on RNase A. I measured the thermal unfolding of RNase A in the presence of various concentrations of glycine and found that the T_m of RNase A increased almost linearly with glycine concentration (Figure V.1). An increase in T_m of up to 11 °C was observed in the presence of 2.0 M glycine.

I also observed that the presence of 2.0 M glycine does not perturb the COSY spectrum of RNase A (Figure V.2). The H/D exchange rates of 37 amide protons of RNase A were then measured using 2D ^{1}H COSY at 34 °C under two different glycine concentrations, 0.1 M and 2.0 M. The time courses of the amide proton cross peak intensities were then fitted to time dependent single exponential functions. Figure V.3 shows the typical exchange rate fits. For residues like His12 and Gln60, the amide proton H/D exchange rates decreased in the presence of 2.0 M glycine; but for residues like Asp14 and Ser59, the exchange rates barely changed under the two different glycine concentrations. Table V.1 lists the ratios of the exchange rate for each amide proton under the two different glycine concentrations, and Figure V.4 shows the plot of the H/D exchange rates ratio vs. residue number.

Two types of backbone amide proton behavior were observed, as shown in Figures V.3 and V.4. Some amide protons (type I) exhibit a decrease in their H/D exchange rates in the presence of 2.0 M glycine, as one would expect from global stabilization; others (type II), however, show little change in their H/D exchange rates. Taking a 20% experimental variation into consideration, I placed residues that had H/D exchange rate ratios greater than 1.2 in the type I category; the rest were placed in type II. Figure V.4 shows that most residues belong to the type II category.

When these two types of amides were mapped to RNase A structure (Figure V.5), I found that most of the type I residues are located in α-helix segments with moderate local

stability (E9, R10, Q11, H12, M13, M29, M30, K31), and that the others are located in regions with low local stability (N44, F46, Q60, K61).

In order to see the global unfolding free energy change of RNase A in the presence of 2.0 M glycine, I also measured the GdnHCl unfolding of RNase A at 34 °C in the presence of 0.1 M glycine and 2.0 M glycine. Figure V.6 shows the results. The free energy of unfolding (ΔG) was calculated using equation [1] by assuming a two-state unfolding reaction:

$$\Delta G = -RT \ln K = -RT \ln\left[\frac{(y_f - y)}{(y - y_u)}\right] \qquad [1]$$

where K = equilibrium constant
y = observed ellipticity
y_f = pre - unfolding transition ellipticity
y_u = post - unfolding transition ellipticity

ΔG is known to be linear with GdnHCl concentration in the unfolding transition region,

$$\Delta G = \Delta G_{(0)} - m\,[\text{GdnHCl}] \qquad [2]$$

The unfolding free energy of RNase A in the absence of GdnHCl (ΔG(0)) can be thus derived from equation [2]. Figure V.7 shows such fittings. The unfolding free energies $\Delta G_{(0)}$ for RNase A at 34 °C in the presence of 0.1 M glycine and 2.0 M glycine are 6.82 kcal/mol and 8.21 kcal/mol respectively. This accounts for a global unfolding energy difference of 1.39 kcal/mol for RNase A in the presence of 2.0 M glycine and 0.1 M glycine. Given this, if a global unfolding event is involved in the H/D exchange, I would expect the high glycine concentration to cause all the amide proton exchange rates to decrease by a factor of 10 (Englander and Kallenbach, 1984). This is clearly not the case. The extent of stabilization of the amide proton H/D exchange varies in different parts of the protein. However, the maximum stabilization effect seen (on Met-13) is very close to the

extent of stabilization obtained globally. This suggests that different local "units" of a protein unfold independently during H/D exchange, and that such local unfolding events may or may not be coincident with the global unfolding event. This is consistent with the "local unfolding" model for H/D exchange (Englander and Kallenbach, 1984; Mayo and Baldwin, 1993).

From the GdnHCl denaturation measurement, it is also interesting to note that the *m* values in equation [2] in Figure V.7 are different. m = 2.40 kcal/mol in the presence of 2.0 M glycine and m = 3.07 kcal/mol in the presence of 0.1 M glycine. It is generally believed that the *m* value of a small protein is proportional to the hydrophobic surface area exposure in the protein unfolding. The smaller *m* value observed in the presence of 2.0 M glycine relative to that observed in the presence of 0.1 M glycine suggests a smaller hydrophobic surface area exposure during protein unfolding at the higher glycine concentration. This is also consistent with the preferential hydration model which states that increasing the osmolyte concentration will progressively make the unfolded state more compact (Arakawa and Timasheff, 1983). The intriguing question which still remains is how this kind of global effect can be reflected very differently at the local level, as observed in my study.

Figure V.1. Plot of the thermal unfolding temperature (T_m) of RNase A vs. glycine concentration. T_m's were determined by taking the derivative of the thermal unfolding curves monitored by circular dichroism.

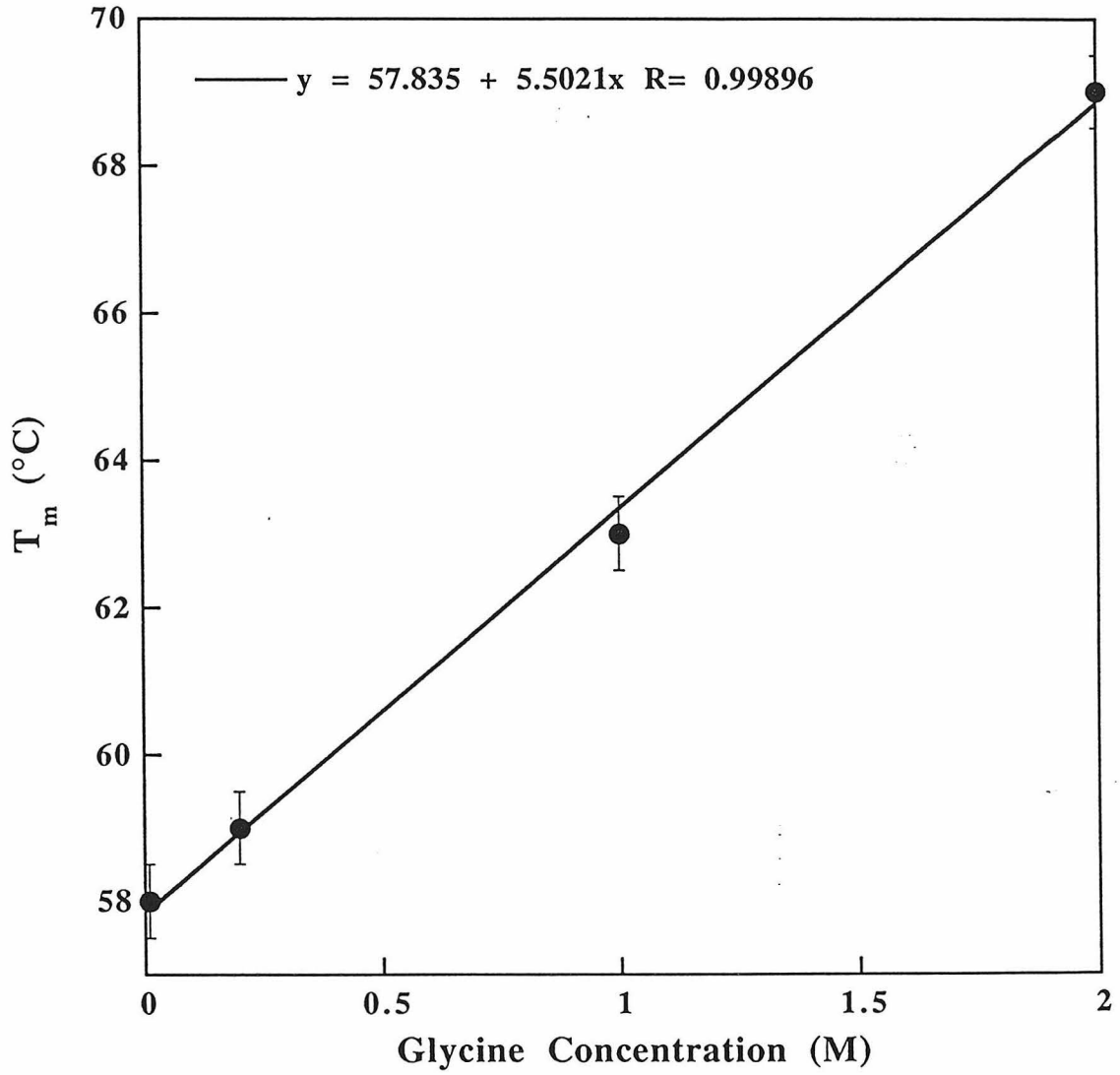
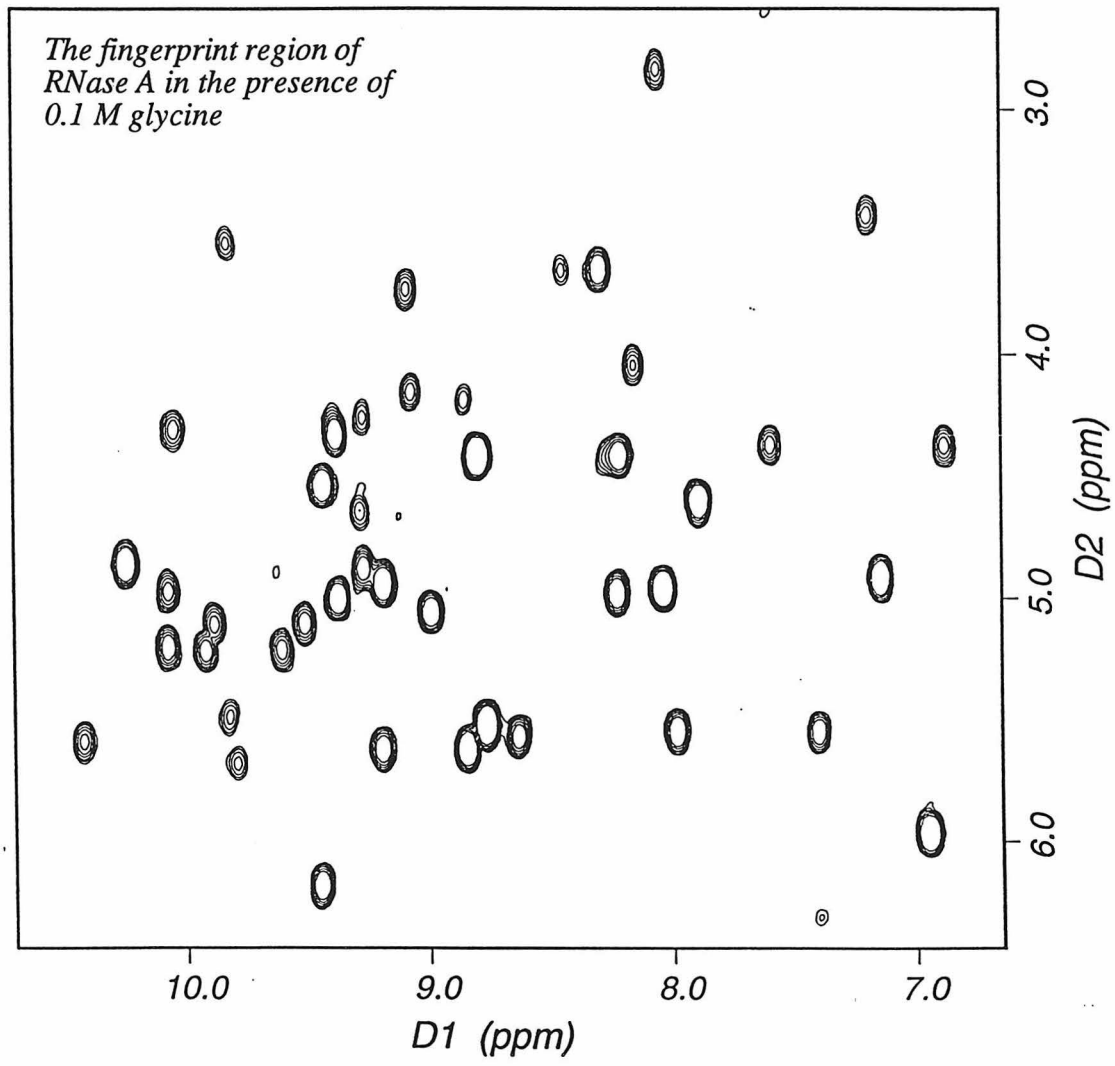

y = 57.835 + 5.5021x R= 0.99896
70
68
66
64
62
60
58
T_m (°C)
0
0.5
1
1.5
2
Glycine Concentration (M)

Figure V.2. Fingerprint region of the COSY spectra for RNase A (a) in the presence of 0.1 M glycine and (b) in the presence of 2.0 M glycine.

a


The fingerprint region of
RNase A in the presence of
0.1 M glycine
3.0
4.0
5.0
6.0
D2 (ppm)
10.0
9.0
8.0
7.0
D1 (ppm)

b

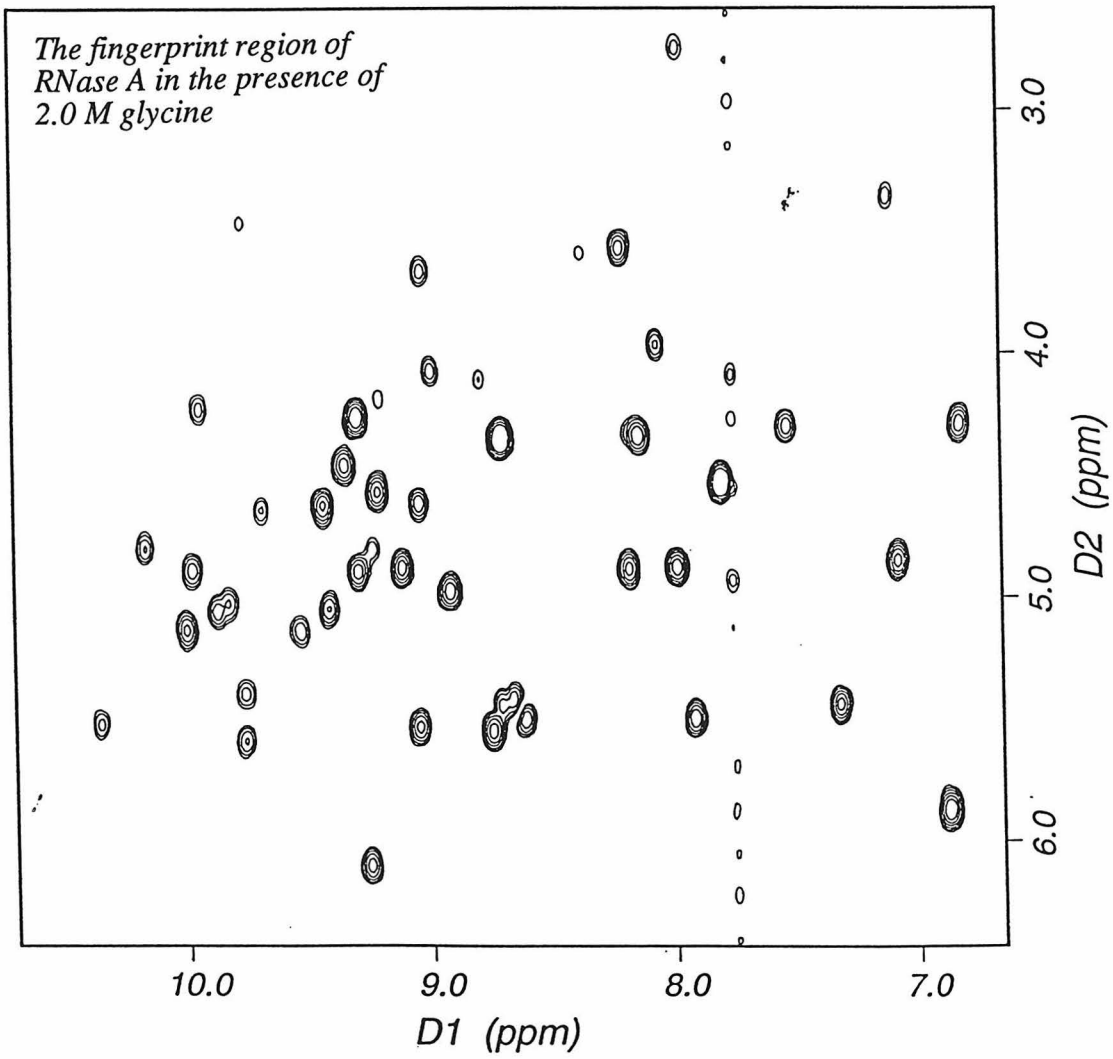
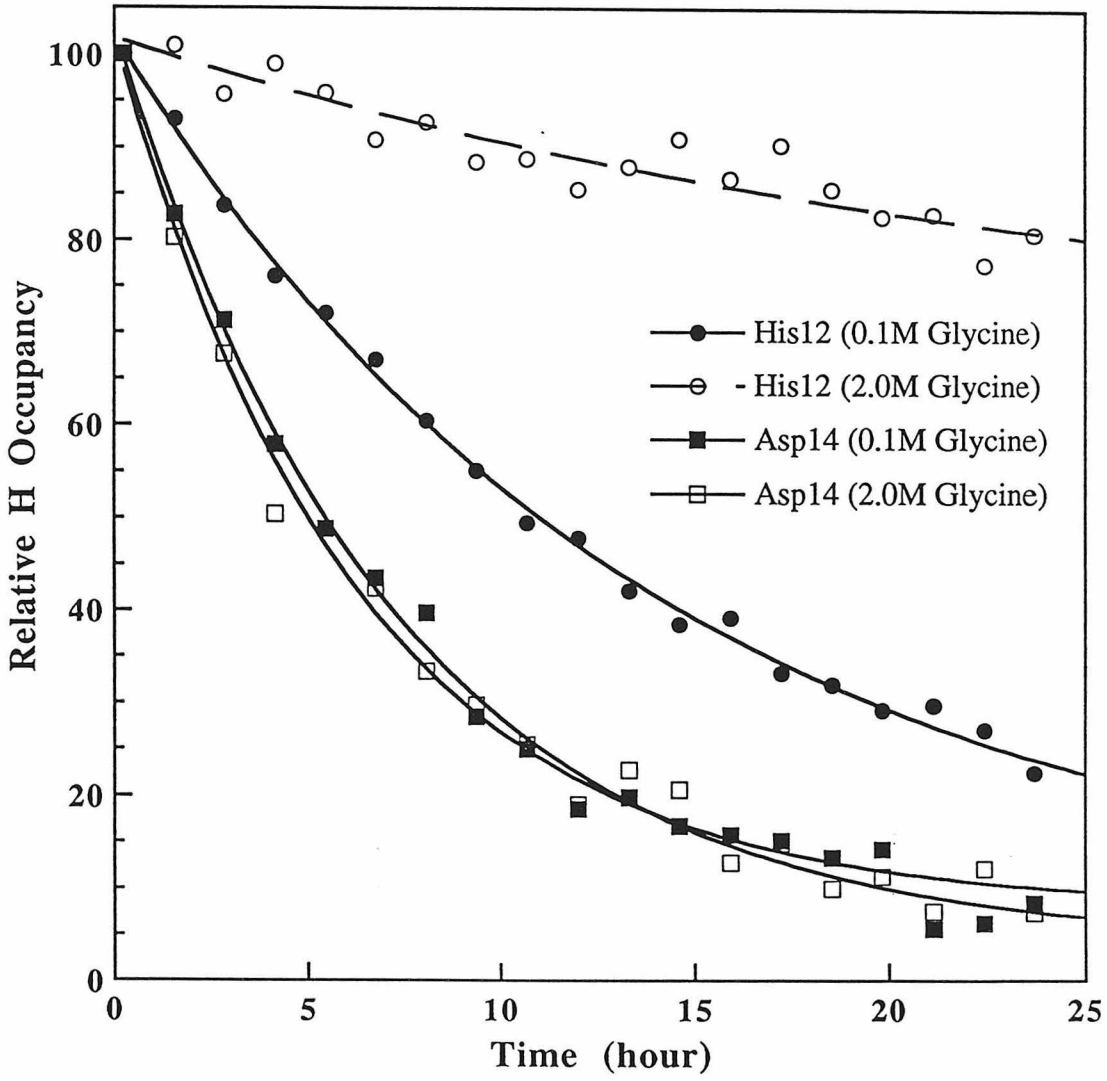

The fingerprint region of
RNase A in the presence of
2.0 M glycine
3.0
4.0
5.0
6.0
D2 (ppm)
10.0
9.0
8.0
7.0
D1 (ppm)

Figure V.3. Typical H/D exchange rate comparison in the presence of low (0.1 M) and high (2.0 M) glycine. (a) His-12 and Asp-14; (b) Ser-59 and Gln-60.

(a)


Relative H Occupancy
100
80
60
40
20
0
0
5
10
15
20
25
Time (hour)
His12 (0.1M Glycine)
His12 (2.0M Glycine)
Asp14 (0.1M Glycine)
Asp14 (2.0M Glycine)

(b)

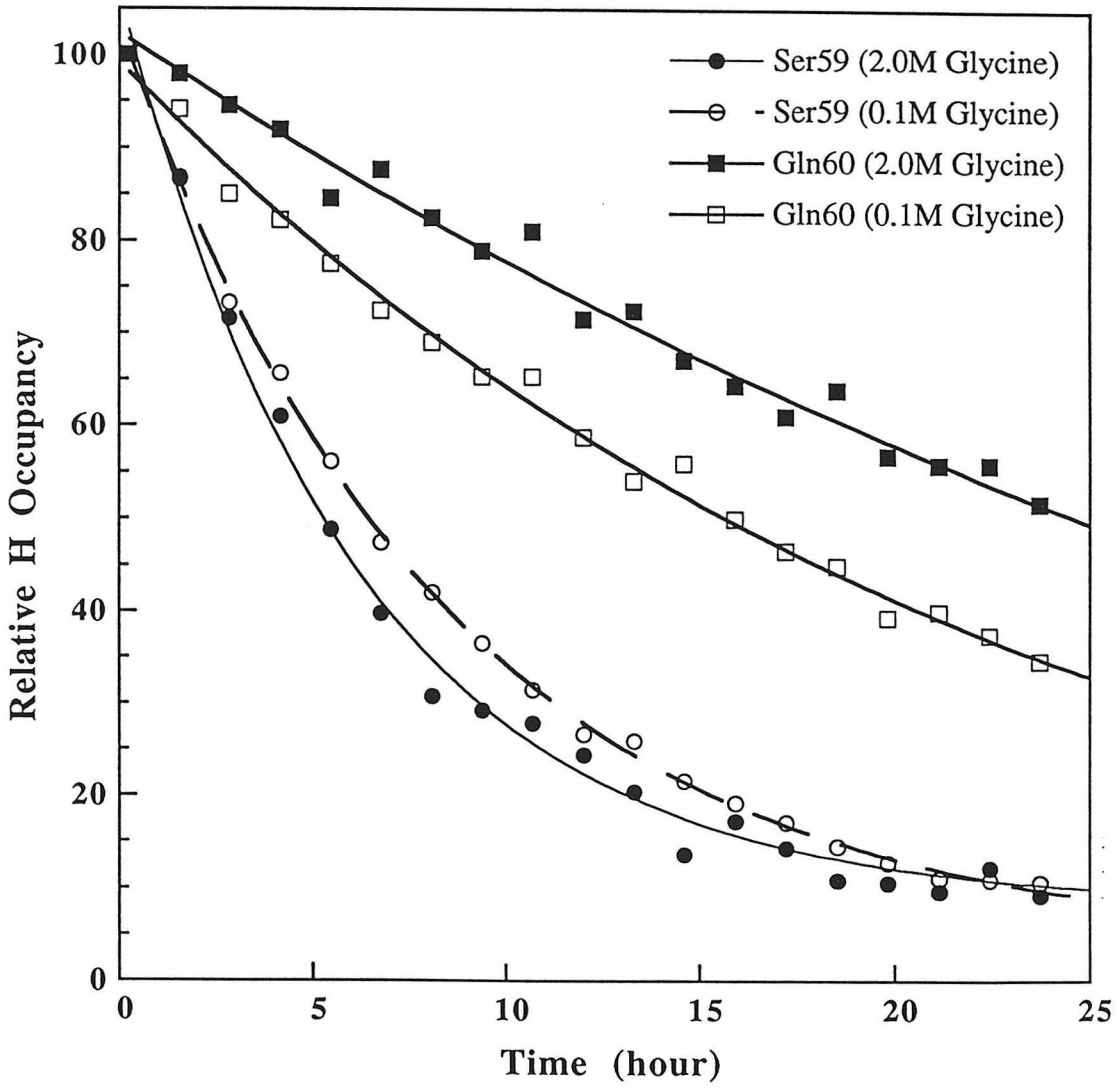

Ser59 (2.0M Glycine)
- Ser59 (0.1M Glycine)
Gln60 (2.0M Glycine)
Gln60 (0.1M Glycine)
Relative H Occupancy
100
80
60
40
20
0
0
5
10
15
20
25
Time (hour)

Table V.1: H/D Exchange Rates and Their Ratios in the Presence of 0.1 M and 2.0 M Glycine for Residues Measured in COSY Experiments.

residues	H/D exchange rate (hr^{-1}) in the presence of 0.1 M glycine	H/D exchange rate (hr^{-1}) in the presence of 2.0 M glycine	ratio of exchange rate (0.1 M vs. 2.0 M)
Glu-9	0.96	0.33	2.9
Arg-10	0.38	0.17	2.3
Gln-11	0.12	0.051	2.4
His-12	0.071	0.018	3.9
Met-13	0.020	0.0017	11
Asp-14	0.15	0.16	0.95
Met- 29	0.078	0.044	1.8
Met-30	0.58	0.23	2.5
Lys-31	0.29	0.15	1.9
Val-43	0.29	0.23	1.3
Asn-44	0.13	0.094	1.3
Phe-46	0.29	0.22	1.3
His-48	0.0072	0.0078	0.93
Glu-49	0.16	0.16	0.99
Val-54	0.0040	0.0046	0.87
Gln-55	0.0027	0.0027	1.0
Ala-56	0.029	0.040	0.73
Ser-59	0.12	0.15	0.78
Gln-60	0.046	0.027	1.7
Lys-61	0.11	0.090	1.2
Val-63	0.0023	0.0021	1.1
Thr-70	0.014	0.019	0.74
Cys-72	0.029	0.042	0.69
Gln-74	0.00053	0.00060	0.88
Met-79	0.0017	0.0015	1.2
Ile-81	0.0015	0.0018	0.83
Asp-83	0.17	0.23	0.74
Cys-84	0.0027	0.0029	0.93

Table V.1: (Continued)

residues	H/D exchange rate (hr^{-1}) in the presence of 0.1 M glycine	H/D exchange rate (hr^{-1}) in the presence of 2.0 M glycine	ratio of exchange rate (0.1 M vs. 2.0 M)
Tyr-97	0.26	0.31	0.84
Lys-98	0.0045	0.0064	0.71
Thr-100	0.040	0.036	1.1
Ala-102	0.017	0.020	0.85
Lys-104	0.0078	0.0033	2.4
Ile-107	0.012	0.0077	1.6
Val-108	0.0010	0.00095	1.1
Val-116	0.0022	0.0024	0.92
His-119	0.0080	0.0081	0.98

Figure V.4. Plot of the ratio of the H/D exchange rates in the presence of 0.1 M glycine and 2.0 M glycine vs. residue number. The minimum of the vertical axis is 1.2, so only ratios over 1.2 show up as the filled bars in the plot, indicating the slowdown of the H/D exchange rate for these residues.

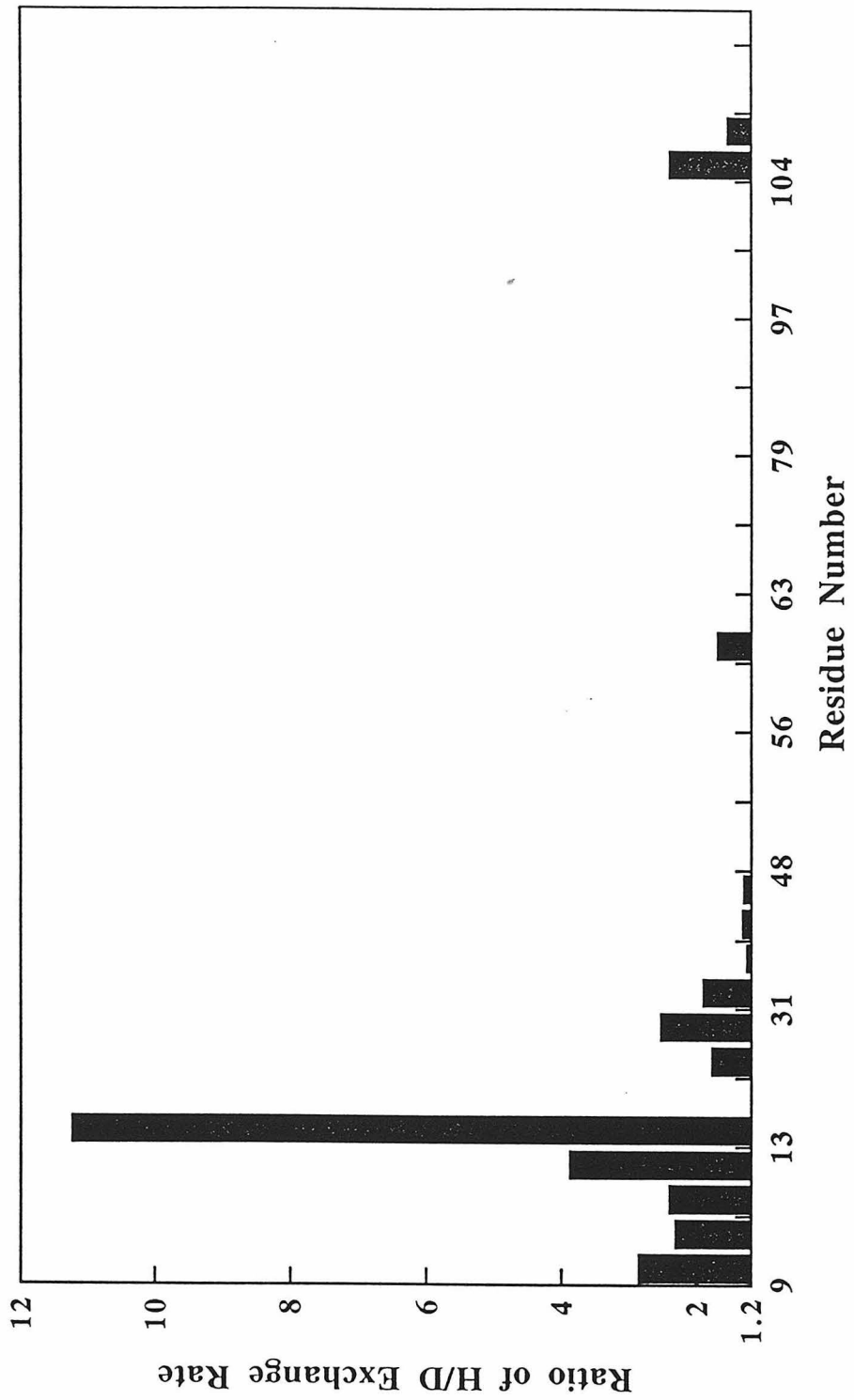

Ratio of H/D Exchange Rate
12
10
8
6
4
2
1.2
9
13
31
48
56
63
79
97
104
Residue Number

Figure V.5. Molscript drawing of RNase A structure. Residues showing a decrease in H/D exchange rates in the presence of 2.0 M glycine compared to 0.1 M glycine are highlighted.

Molscript drawing of RNase A crystal structure

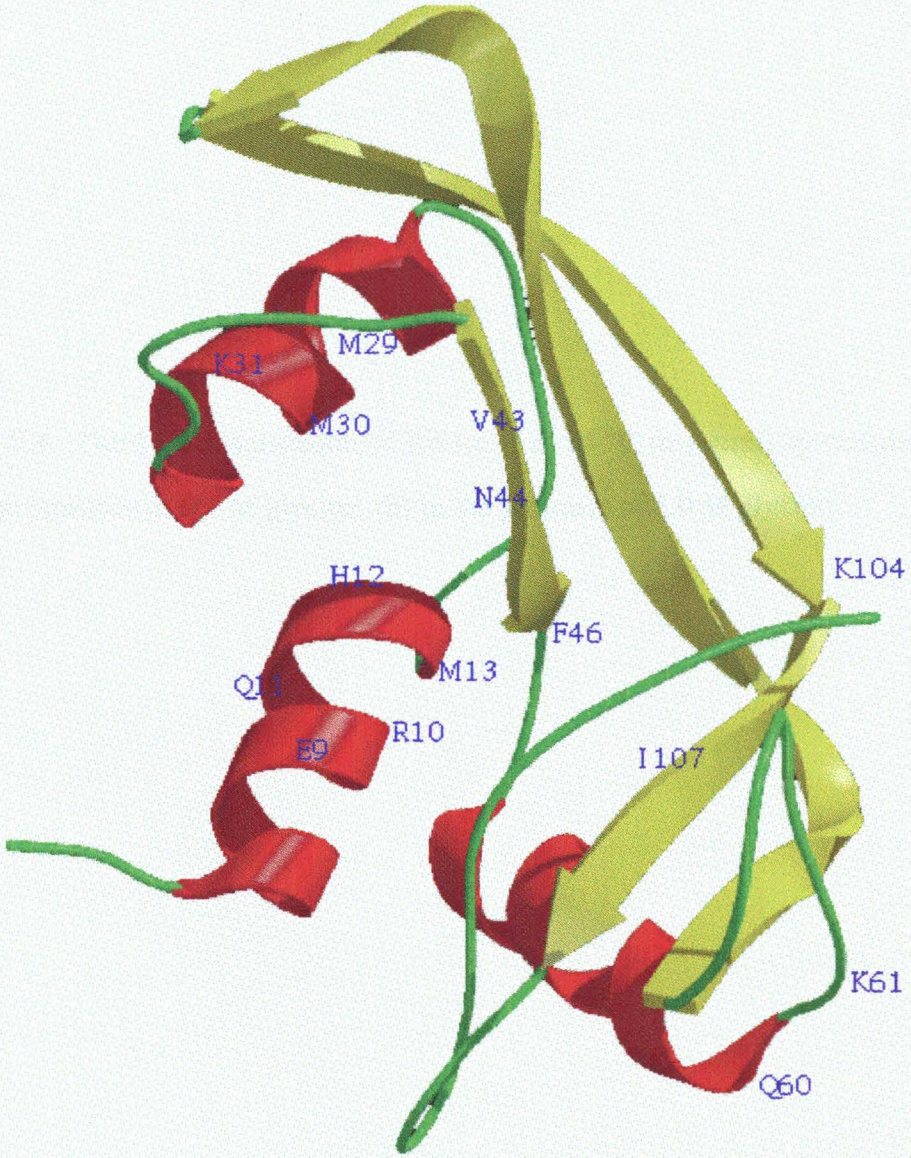
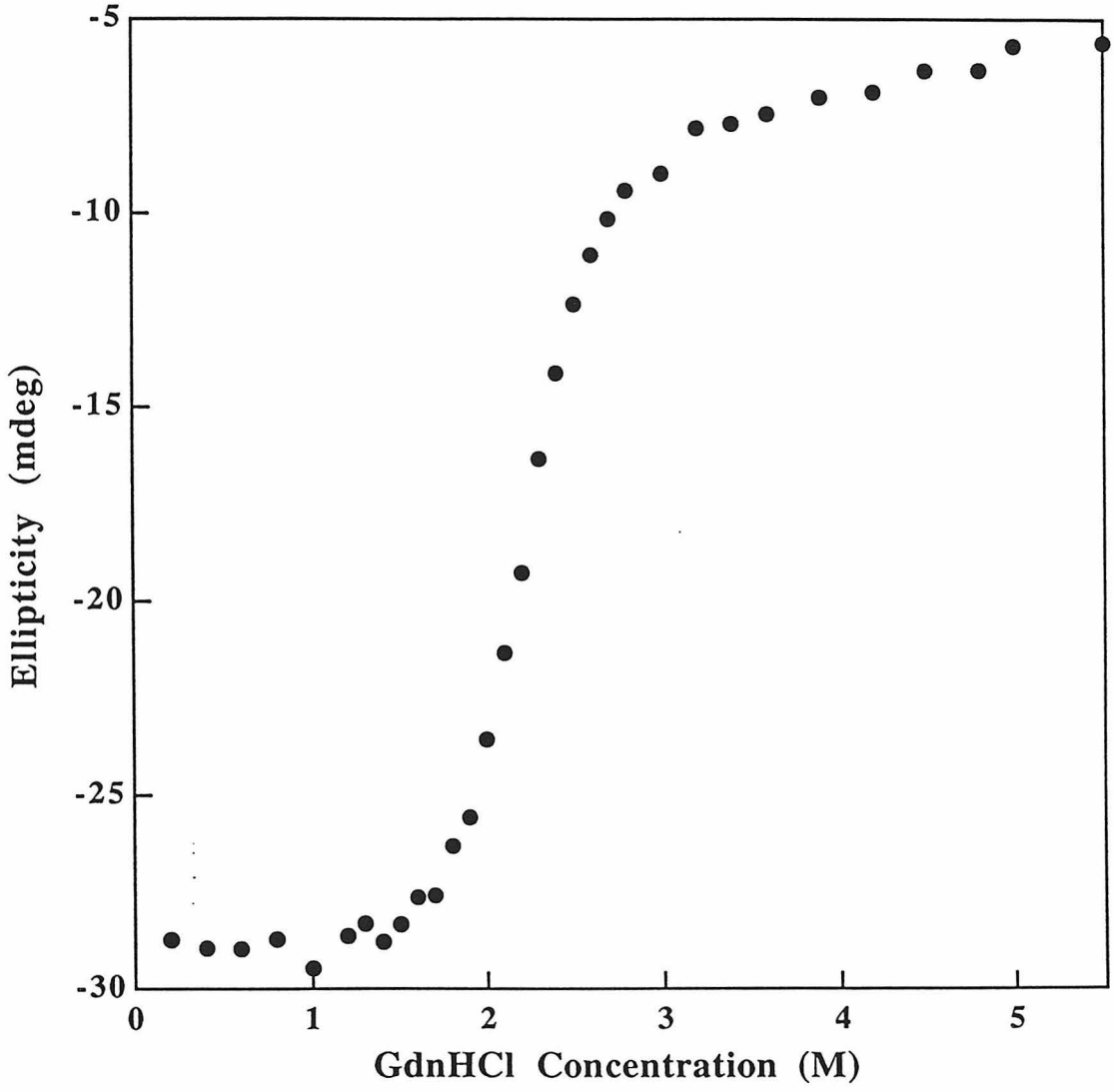


Figure V.6. GdnHCl denaturation of RNase A at 34 °C measured by circular dichroism, (a) in the presence of 0.1 M glycine; (b) in the presence of 2.0 M glycine.

(a)


-5
-10
-15
-20
-25
-30
Ellipticity (mdeg)
0
1
2
3
4
5
GdnHCl Concentration (M)

(b)

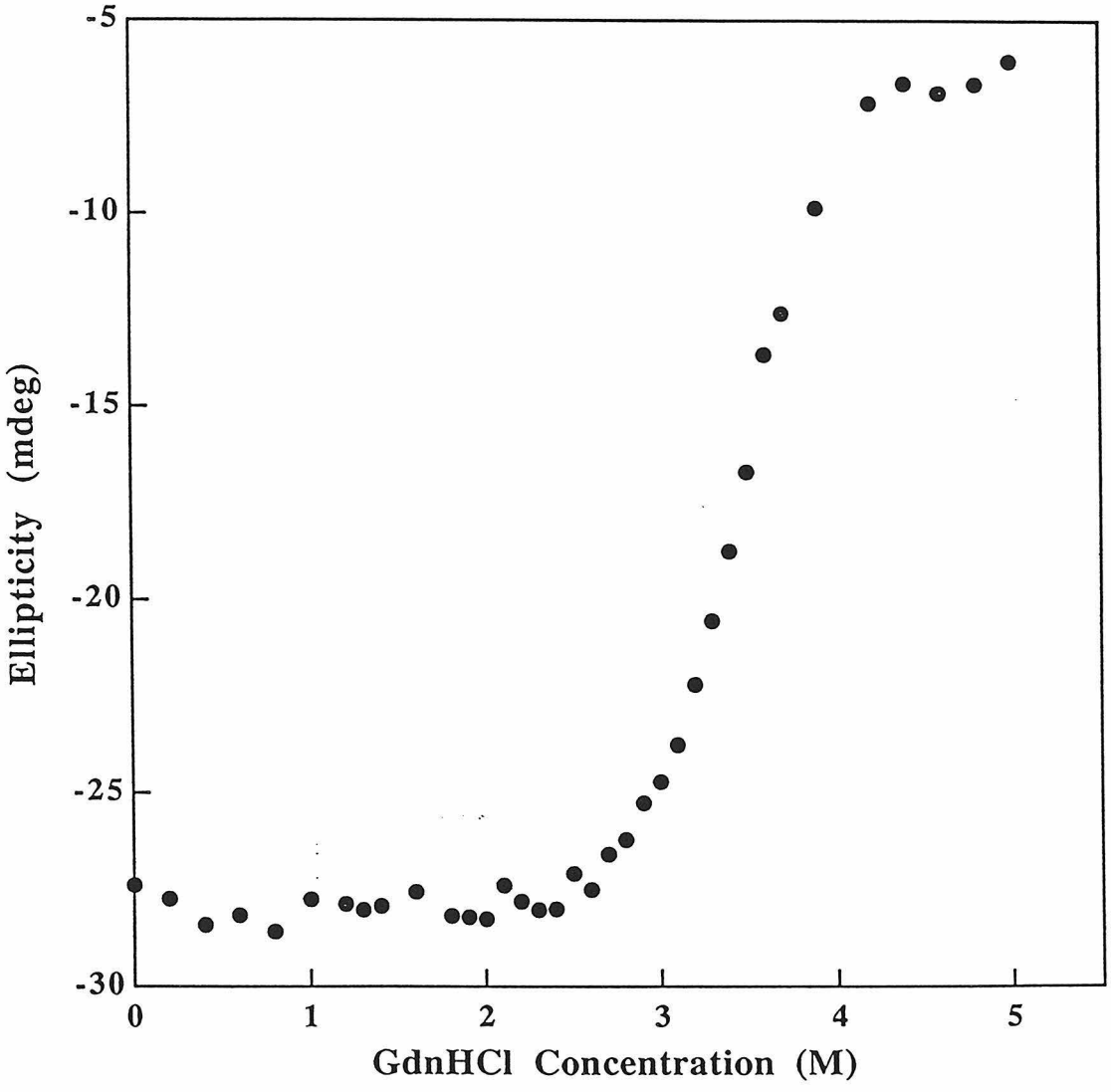
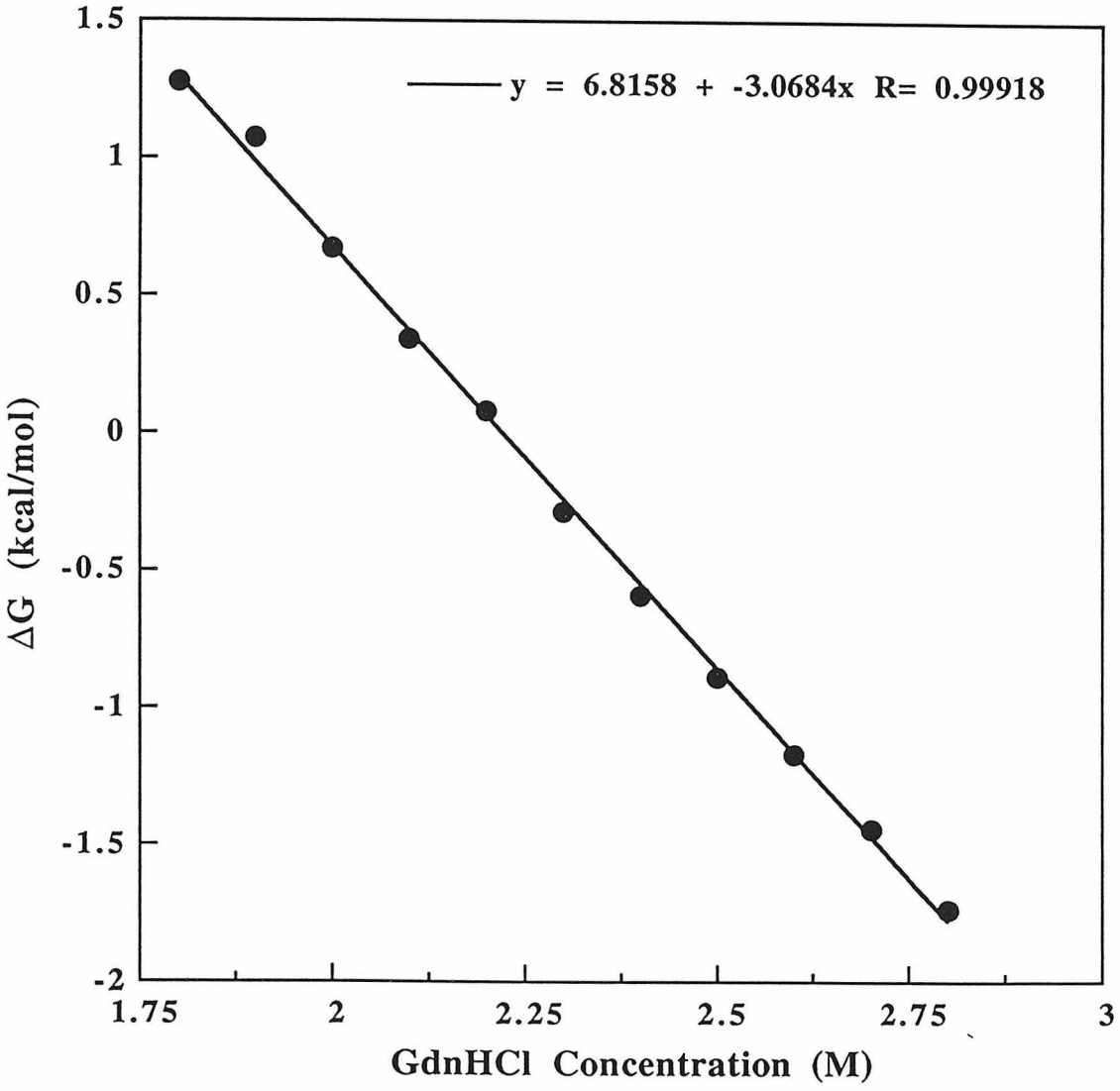
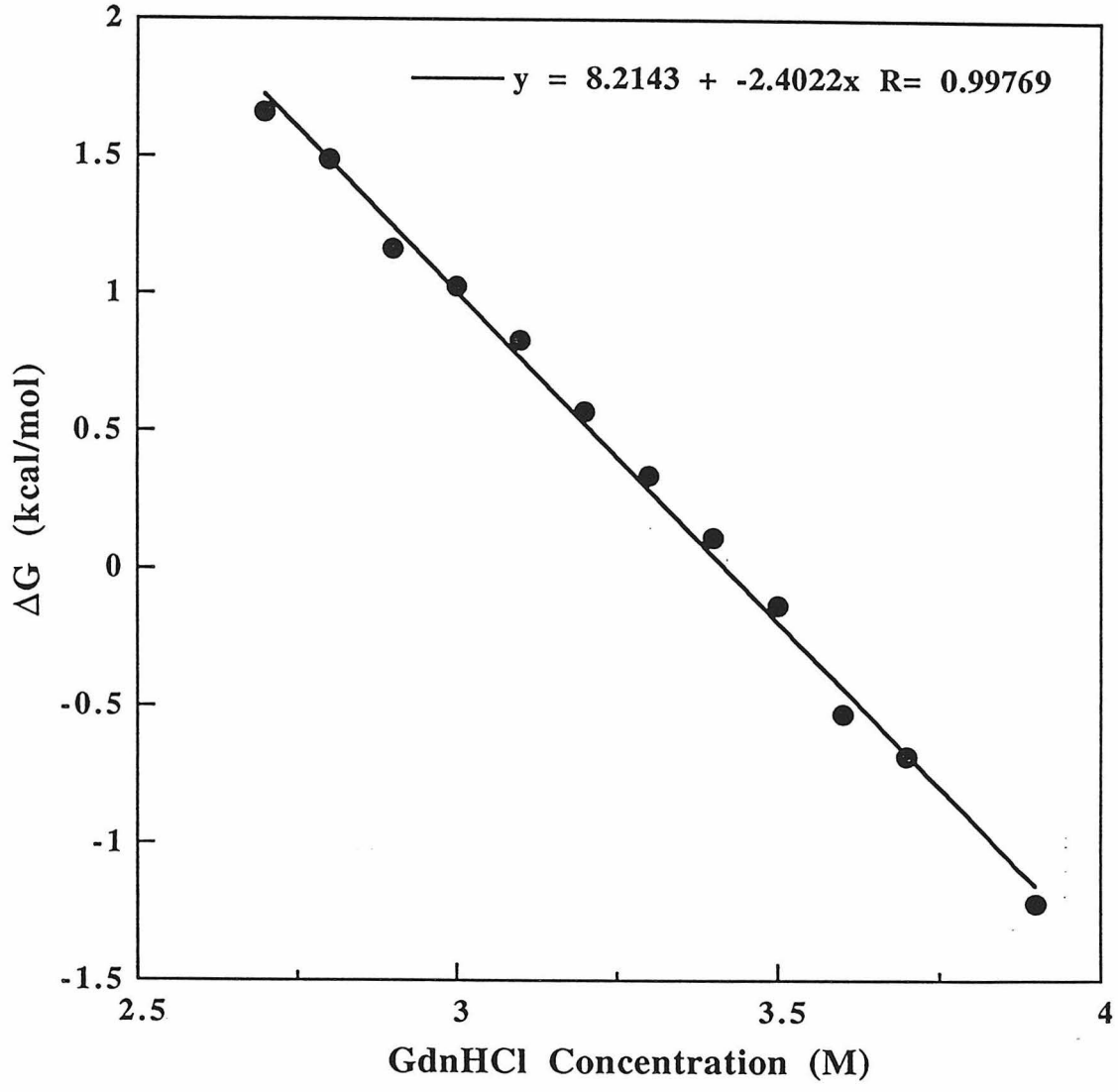

-5
-10
-15
-20
-25
-30
Ellipticity (mdeg)
0
1
2
3
4
5
GdnHCl Concentration (M)

Figure V.7. The unfolding free energy fits from the transition region of the GdnHCl denaturation curves in Figure V.6. (a) in the presence of 0.1 M glycine; (b) in the presence of 2.0 M glycine.

(a)


y = 6.8158 + -3.0684x R= 0.99918
1.5
1
0.5
0
-0.5
-1
-1.5
-2
ΔG (kcal/mol)
1.75
2
2.25
2.5
2.75
3
GdnHCl Concentration (M)

(b)


y = 8.2143 + -2.4022x R= 0.99769
2
1.5
1
0.5
0
-0.5
-1
-1.5
ΔG (kcal/mol)
2.5
3
3.5
4
GdnHCl Concentration (M)

REFERENCES

1. Arakawa, A. and Timasheff, S.N. (1983) *Arch. Biochem. Biophys.* **224,** 169-177.

2. Arakawa, A. and Timasheff, S.N. (1985) *Biophys. J.* **47**, 411-414.

3. Englander, S.W. and Kallenbach, N.R. (1984) *Q. Rev. Biophys.* **16**, 521-655.

4. Matthews, S.J. and Leatherbarrow, R.J. (1993) *J. Biomol. NMR* **3**, 597-600.

5. Mayo, S.L. and Baldwin, R.L. (1993) *Science* **262**, 873-876.

6. Pace, C.N. and Laurents, D.V. (1989) *Biochemistry* **28**, 2520-2525.

7. Plaza de Pino, I.M. and Sanchez-Ruiz, J.M. (1995) *Biochemistry.*

8. Santoro, M.M., Liu, Y., Khan, S.M.A., Hou, L.-X., and Bolen, D.W. (1992) *Biochemistry* **31**, 5278-5283.

9. Timasheff, S.N. (1993) *Annu. Rev. Biophys. Biomol. Struct.* **22**, 67-97.

10. Yancey, P.H., Clark, M.E., Hand, S.C., Bowlus, R.D., and Somero, G.N. (1982) *Science* **217**, 1214-1222.

APPENDICES

Appendices of Chapter I

Appendix I-1. Boiling miniprep protocol

The boiling miniprep method is used in all plasmids preparation in my experiments. The procedure is as follows:

1. Pick colony from LB/agar plate, grow overnight in ~3 ml LB/antibiotic media.
2. Spin down the cell in 2 ml eppendorf tube for 0.5 minute.
3. Suspend the cell pellet in 300 μl STET.
4. Add 20 μl (10mg/ml stock solution) lysozyme to the suspended cell, gentle mixing.
5. Wait at room temperature for 1 minute, then 2-3 minutes in boiling water.
6. Spin down the cell immediately for 5 minutes.
8. Pull out the cell debris (or transfer the solution to a clean tube)
9. Add 5-7 μl of RNase, 37 °C for over 15 minutes.
10. Add 300 μl of high salt (1:3 mixture of 8M NH_4Ac and isopropanol), mix by inversion to precipitate the DNA.
11. Centrifuge for 5-10 minutes.
12. Pour away the supernatant, wash the DNA with 70% ethanol.
13. Dry the DNA in air.
14. Dissolve the DNA with 50 μl of H_2O (if for next step sequencing, no EDTA here), and store at -20 °C for future use.
15. DNA is usually good enough for sequencing and concentration is about 0.2-1 μg/μl.

Appendix I-2. Protocol for preparation of the competent cell for electroporation

Two kinds of competent cells (DH10B and BL21) are used in the experiments described in Chapter I and Chapter II. They are prepared in the same way as the following:

---

1. A patch from old culture is grown overnight in 25 ml LB media overnight 37 °C.
2. Transfer 1 ml of overnight grown culture to 1L LB media, grow at 37 °C.
3. Collect the cell by centrifuging at 4-5k rpm for 10-15 min. when the OD_{600nm} of culture reaches >0.5.
4. Suspend the cell in ~100 ml of 10% glycerol (cold).
5. Centrifuge again for 10 minutes to collect the cell.
6. Wash the cell pellet again by suspending in ~100 ml of 10% glycerol.
7. Centrifuge at 5k rpm to get the cell pellet.
8. Resuspend the cell pellet in the leftover solution.
9. Transfer the cell in 180-200 μl aliquots into tubes pre-chilled in a dry-ice and ethanol bath.
10. Store at -70 °C for future use.

Appendix I-3. Clone the azurin gene with leader sequence into pET-9a

pET-9a is from Novagen. Azurin with leader sequence is cloned in pTZ-18u from previous work in Dr. John H. Richards' group.

---

1. Digest pTZ-18u-azurin with *Nde I* and *Bgl II* to get the gene coding for azurin and leader sequence.
2. Digest pET-9a with *Nde I* and *BamH I.*
3. Ligase the azurin gene to the above digested pET-9a vector.
4. The *Nde I* restriction site is restored, but the *BamH I* site on pET-9a is lost after ligation.

Appendix I-4. The cDNA sequence for *pa* azurin gene with the leader sequence

```
    Met Leu Arg Lys Leu Ala Ala Val Ser Leu Leu Ser Leu Leu
CAT ATG CTG CGT AAG CTG GCT GCA GTG TCT CTG CTG TCT CTG CTG
Nde I                  Pst I

                          1   2   3   4   5   6   7   8   9
Ser Ala Pro Leu Leu Ala Ala Glu Cys Ser Val Asp Ile Gln Gly
TCT GCT CCG CTG CTG GCT GCT GAA TGC TCC GTT GAT ATC CAG GGT
                           Bsm I          EcoR V

10  11  12  13  14  15  16  17  18  19  20  21  22  23  24
Asn Asp Gln Met Gln Phe Asn Thr Asn Ala Ile Thr Val Asp Lys
AAT GAT CAG ATG CAG TTC AAC ACC AAC GCC ATC ACC GTC GAC AAG
  Bcl I                                          Sal I

25  26  27  28  29  30  31  32  33  34  35  36  37  38  39
Ser Cys Lys Gln Phe Thr Val Asn Leu Ser His Pro GLy Asn Leu
AGC TGC AAG CAG TTC ACT GTT AAC CTG TCT CAC CCA GGT AAC CTG
                       Hpa I                  BstE II

40  41  42  43  44  45  46  47  48  49  50  51  52  53  54
Pro Lys Asn Val Met Gly His Asn Trp Val Leu Ser Thr Ala Ala
CCG AAG AAC GTT ATG GGT CAC AAC TGG GTT CTG TCC ACC GCG GCT
                                               Sac II

55  56  57  58  59  60  61  62  63  64  65  66  67  68  69
Asp Met Gln Gly Val Val Thr Asp Gly Met Ala Ser Gly Leu Asp
GAC ATG CAA GGC GTT GTC ACT GAC GGT ATG GCT AGC GGT CTG GAT
                                       Nhe I

70  71  72  73  74  75  76  77  78  79  80  81  82  83  84
Lys Asp Tyr Leu Lys Pro Asp Asp Ser Arg Val Ile Ala His Thr
AAA GAC TAC CTG AAG CCG GAT GAC TCT CGA GTT ATC GCC CAC ACC
                                   Xho I

85  86  87  88  89  90  91  92  93  94  95  96  97  98  99
Lys Leu Ile Gly Ser Gly Glu Lys Asp Ser Val Thr Phe Asp Val
AAG CTG ATC GGA TCC GGT GAA AAA GAC TCC GTT ACT TTC GAC GTT
            BamH I

100 101 102 103 104 105 106 107 108 109 110 111 112 113 114
Ser Lys Leu Lys Glu Gly Glu Gln Tyr Met Phe Phe Cys Thr Phe
TCC AAG CTT AAA GAA GGT GAA CAG TAC ATG TTC TTC TGC ACT TTC
    HinD III

115 116 117 118 119 120 121 122 123 124 125 126 127 128
Pro Gly His Ser Ala Leu Met Lys Gly Thr Leu Thr Leu Lys Trm
CCG GGT CAC TCC GCA CTG ATG AAA GGT ACC CTG ACT CTG AAA TAG
Xma I                          Kpn I

AGA TCT
Bgl II
```

Appendix I-5. Overexpression of azurin and azurin mutants

After identifying the correct sequence for the azurin in pET-9a vector by sequencing, the procedure for expression of azurin and mutants is as follows:

---

1. Transform the plasmid to competent cells (bacteria BL21) by electroporation.
2. Spread to LB/kanamycin plates after one hour incubation.
3. Pick colonies to LB/kanamycin media for growth.
4. IPTG induction after cell growth reaches the log phase (OD_{600nm}>0.6).
   (IPTG concentration: 200 mg/ml stock solution, 1 ml stock solution for 1L culture.)
5. Harvest cells after 5-6 hr. of IPTG induction by centrifuging at 5k rpm for 10 min.
6. Suspend cell pellets in 20% sucrose and shaking for 10-15 min.
7. Spin down the cells at 5k rpm for 10 min.
8. Suspend the cell pellets by shaking in ice-cold H_2O for >15 min.
9. Add a few drops or ~100 μl of 1M $CuSO_4$ stock solution (excess amount).
10. Collect supernatant by spin down the cell debris and centrifuge again (or filter) if necessary to get rid of cell debris left.
12. Desalt and concentrate the protein solution by Amicon 10K membrane.
13. Purification with a home pack SP Sepharose ion-exchange FPLC column with 20-25 mM NH_4Ac and 100-150 mM NH_4Ac pH 4.5 or equivalents as the loading and eluting buffer respectively.
14. Precaution: For purification of apo-azurin, there must be some EDTA (5-10 mM) in all buffers; otherwise, apo-azurin will bind trace metal ions presented in the process of purification. Rinsing the whole sample pass routes on FPLC with EDTA containing loading buffer is absolutely necessary before loading apo-azurin sample.

Appendix I-6. CloneTech Transformer Site-Directed Mutagenesis method

The CloneTech Transformer Site-Directed Mutagenesis is based on the method of Deng and Nickoloff (*Anal. Biochem.* 200, 81, 1992). The method works by simultaneously annealing two oligonucleotide primers to one strand of a denatured double-stranded plasmid. One primer introduces the desired mutation (referred to as the mutagenic primer). The second primer mutates a restriction site unique to the plasmid for the purpose of selection (referred to as the selection primer). In the case of the desired mutation on a unique restriction site, only one primer (the selection primer and mutagenic primer becomes the same one) is needed. The details can refer to the CloneTech's TransformerTM Site-Directed Mutagenesis Kit. I outline the procedures used in mutating the T124 of azurin as follows:

---

1. Denatured pET9a-azurin vector + oligo (for mutation)

   (pET9a-azurin vector is denatured either by Alkaline denaturation or boiling for 3 minutes.)

2. Annealing

10x annealing buffer	2.0 μl
plasmid DNA	2.0 μl
selection primer	2.0 μl
H_2O	14 μl (for total of 20 μl)

   incubate at 100 °C for 3 minutes,

   chill immediately in an ice-H_2O for 5 minutes.

3. Polymerization + ligation

   To the primer/plasmid reaction mixture, add:

10x synthesis buffer	3 μl
T_4 DNA polymerase	2-4 unit
T_4 DNA ligase	4 units
H_2O	5 μl (for total of 10 μl)

   mix well and centrifuge briefly,

   incubate at 37 °C,

   stop the reaction by heating at 70 °C for over 5 minutes,

   let the reaction mixture cool to room temperature.

4. Primary selection by restriction digestion with *Kpn I.*
5. Transform the digest product to DH10B.
6. Grow overnight for the first miniprep.
7. *Kpn I* digest of the plasmids from the first miniprep.
8. Transformation of the last *Kpn I* digest product to DH10B for selection.
9. Spread on LB/agar plates and overnight grow.
10. Pick colonies for second miniprep.
11. *Kpn I* digest to screen for positive mutations.
12. Pick positive mutations for sequencing.
13. Transformation of the right sequence to BL21 for next step protein prep.

Appendix I-7. Differential Scanning Calorimetry (DSC) experiment

DSC measurement of apo-azurin is done as follows:

---

1. Dissolve the lyophilized protein sample into buffer of 25 mM K_2HPO_4, 0.5 mM EDTA, pH5.5.
2. Dialyze the sample against the reference buffer (25 mM K_2HPO_4, 0.5 mM EDTA, pH5.5) overnight using 3K membrane.
3. Degas both the protein sample and reference buffer in a vacuum chamber for 5-10 minutes.
4. Estimate the protein sample by UV (using A_{280nm} and extinction coefficient ε_{280} = 9,700 $M^{-1}cm^{-1}$). The sample concentration in my measurement was 0.1375 mM.
5. Rinse the DSC cells by dH_2O first, then rinse the cells by reference buffer.
6. Fill the cells with buffer-buffer/sample, and apply a static N_2 pressure to the cell chamber.
7. Set water-bath temperature to 15 °C (5 degrees lower than the starting temperature of thermal scan).
8. Run a buffer-buffer first to get a good baseline from 20 °C - 90 °C at rate of 60 °C/hr. Repeat if necessary until a good baseline of buffer-buffer under the identical condition of running buffer-sample.
9. Data are analyzed by MicroCal Origin or KaleidoGraph (saved in ASCII format).

Appendices of Chapter II

Appendix II-1. 5' phosphorylation and annealing of oligos

8 λ 34Tup (or 34Vup) oligo

+ 8 λ 34Tdown (or 34Vdown) oligo

+ 3 λ (10x kinase buffer)

+ 3 λ (10x rATP)

+ 6 λ H2O (to make the total of 30 λ)

+ 2 λ of kinase

Mix well and sit at 37 °C for 1 hr.

Warm to 50-60 °C for a few minutes and cool down to room temperature to make sure of the correct annealing.

Store at -20 °C for future use.

Appendix II-2. Maxi plasmid preparation of pET9a-azurin

Because of the low copy number of plasmid with pET9a series, a maxi plasmid preparation of pET9a-azurin is done following the QIAGEN Midi/Maxi Plasmid/Cosmid Purification Protocol. The prepared DNA concentration is 1.0-2.0 μg/μl by running on 2% agarose gel with a known concentration of molecular weight marker.

---

1. Grow 500 ml of LB culture (of pET9a-azurin transformed DH10B bacteria).
2. Spin down the cells at 5K rpm.
3. Resuspend the bacterial pellet in 10 ml of Buffer P1 completely.

Ensure that RNase A has been added to Buffer P1.

4. Add 10 ml of Buffer P2, mix gently, and incubate at room temperature for 5 min.
5. Add 10 ml of chilled Buffer P3, mix immediately but gently, and incubate on ice for 15 or 20 min.
6. Centrifuge at over 20k x g for 30 min. at 4 °C. Collect supernatant promptly.
7. Centrifuge at over 20k x g for 15 min. at 4 °C. Collect supernatant promptly.
8. Equilibrate a QIAGEN-tip 500 by applying 10 ml Buffer QBT, and allow the column to empty by gravity flow.
9. Apply the supernatant from step 6 and 7 to the QIAGEN-tip and allow it to enter the resin by gravity flow.
10. Wash the QIAGEN-tip with 2 x 30 ml Buffer QC.
11. Elute DNA with 515 ml Buffer QF.
12. Precipitate DNA with 0.7 volumes of room-temperature isopropanol. Centrifuge immediately at over 15k x g for 30 min. at 4 °C, and carefully remove the supernatant.
13. Wash DNA with 5 ml of 70% ethanol, air-dry for over 5 min., and finally dissolve in a suitable volume (200λ of dH_2O + 150λ of TE) of buffer for store at -20 °C.

Appendix II-3. *Hpa I* and *Bst EII* digestion of pET9a-azurin plasmid

Because the optimal temperature for *Bst EII* is 60 °C and for *Hpa I* is 37 °C, the double restriction digests are done separately in two steps.

---

1. DNA digest with Hpa I and buffer A at 37 °C for 1 hr.;
2. spin off Hpa I and buffer A using amicon 100K tube and wash with dH_2O;
3. digest with Bst EII and buffer B at 60 °C for 1 hr.;
4. spin off Bst EII and buffer B using amicon 100K tube and wash with dH_2O;
5. repeat steps (1) to (4) to minimize the undigested DNA;
6. run the doubly digested DNA on 1% agarose gel with MW marker to check the molecular weight and estimate the DNA concentration;
7. store the doubly digested DNA at -20 °C for later use.

Appendices of Chapter IV

Appendix IV-1. M9 minimal medium recipe

For 1 L dH_2O, add,

200 mL	5x M9 salts
2 mL	1M $MgSO_4$
20 mL	20% glucose solution (or other appropriate carbon source)
0.1mL	1 M $CaCl_2$

5x M9 Salts is made by dissolving the following salts in 1 L dH_2O

64 g	$Na_2HPO_4 \cdot 7H_2O$
15 g	KH_2PO_4
2.5 g	NaCl
5.0 g	$^{15}NH_4Cl$

$MgSO_4$ and $CaCl_2$ solutions are prepared and sterilized by autoclaving separately. Glucose solution is sterilized by filtration through a 0.2 mm filter.

Appendix IV-2. VNMR macro "setss"

```
addpar('ss')
if $#=0 then ssfilter=100 ssorder='n' endif   /* for low frequency filter using the default 100 frequency width */
if $#=1 then ssfilter=$1 ssorder='n' endif    /*low frequency filtering using the entered ($1) frequency width */
if $#=2 then ssfilter=$1 ssorder=$2 endif     /*zero frequency filtering */
```

Appendix IV-3. VNMR macro "getxpk_t1/t2"

The following macro is modified from "getxpk" in the userlib in VNMR. The main modification is the data output format and the mixing time matching my experiments.

```
/*
"getxpk_t1/t2 -- get cross-peak volumes from a 2D heteronuclear data set for T1/T2
determination"
usage:
1) Join an experiment with a processed 2d data set;
2) prepare a peak list peaks.bin in the current experiment using the ll2d menu functions
from dconi;
3) cd to a directory in which you want to write all the volume integral information;
4) type getxpk('reference') to establish a file in this directory named xpk.ref which will be
a copy of the mark2d.out file and which will be used to establish the reference regions. The
macro checks if the file peaks.bin exists, and if it does, converts it into  an ASCII file
ll2d/mark2d.out.  This command will also create a reduced file containing only region
index and volume integral; this file will be named xpk.nnn, where nnn is the mixing time in
msecs.
5) Join other experiments or recall other 2D data sets into the same experiment, as desired,
and type getxpk_t1/t2; repeat step 6 for all mixing times; a series of xpk.nnn files will be
created.
94/03/14 - modified for VNMR 4.3, Adolf Gogoll, University of Uppsala, Sweden
*/
/*macro starts on next page*/
```

```
if $#>0 then     "verify existence of reference file"
        if $1='reference' then
                exists(curexp+'/ll2d/mark2d.out','file'):$e
                if $e then
                        rm(curexp+'/ll2d/mark2d.out')
                endif
                exists(curexp+'/ll2d/peaks.bin','file'):$e2
                if $e2 then
                        ll2d('writetext',curexp+'/ll2d/mark2d.out')
                else
                        write('error','No peak list found in the current experiment!')
                return(1)
                endif
                copy(curexp+'/ll2d/mark2d.out','xpk.ref')
                                        "copy to reference list in current working directory"
        endif
endif
write('line3','file  found')
exists('xpk.ref','file'):$y             "check current directory for reference pks"
if not $y then
  write('error','file xpk.ref does not exist in current directory')
  abort
endif
format(ncyc*mix,0,0):n1                         "generate file name using mixing time"
        "Here and following, mix=5.55ms for T1 and 15.8 ms for T2 are the cycle_time"
        "for HSQC-T1 and HSQC-T2 experiements respectively according to the pulse"
        "sequences"
$outfile='xpk.'+n1
$peaksout='peaks.'+n1
write('reset',$outfile)
write('file',$outfile,'Measured cross-peak volume integrals')
write('file',$outfile,'Mixing time = %6.1f msec\n',ncyc*mix)
write('file',$outfile,'Index\t\tVolume\t\t\theight')      "write both volume and height"
nrecords('xpk.ref'):$numpeaks           "$numpeaks is related to the number of peaks"
```

```
$numpeaks=($numpeaks-6)/5
mark('reset')
$i=1
lookup('file','xpk.ref','seek','0.00000000','skip')                    "position file"
repeat
lookup('read',4,'skip',7):$lof1,$hif1,$lof2,$hif2
   write('fileline',$peaksout,'peak label:\t')
   write('fileline',$peaksout,'%4d\n', $i)
   write('fileline',$peaksout,'peak positions:\t')
   write('fileline',$peaksout,'%3.4f\t%3.4f\t%3.4f\t%3.4f\n',$lof1,$hif1,$lof2,$hif2)
  cr1=$hif1*dfrq2 delta1=($hif1-$lof1)*dfrq2 cr=$hif2*sfrq delta=($hif2-$lof2)*sfrq
  mark(cr1,delta1,cr,delta):$height,$volume
   write('fileline',$peaksout, 'peak vol and height:')
   write('fileline',$peaksout,'\t%e\t%6.4f\n\n', $volume, $height)
   write('fileline',$outfile, '%4d',$i)
   write('fileline',$outfile,'\t\t%e\t\t%6.6f\n',$volume,$height)
                                                    "output both volume and height"
  $i=$i+1
until $i>($numpeaks)
cat($outfile)

/*macro end*/
```

Appendices of Chapter V

Appendix V-1. The Felix macro for automatic 2D COSY data processing and peak intensity calculation

Following is a Felix macro for automatic multiple data processing and peak volume integration for either Bruker or Varian MCOSY data files. It will process all the data files and pick the peaks specified by the pre-defined peak dba file, and save the volume in a specified slot in the fileinput.txt file which is saved in the current Felix directory. It is supposed that the data is acquired 2k total with 256 t1 increments, and processed as 1k in each dimension, but that can be changed if necessary. The macro syntax of Felix version 2.01 is followed.

```
c** auto.mac macro starts here
cl
cmx
get 'Which Machine:(bruker or varian)?' bv
get 'initial file number:(number)' file1
get 'end file number:(number)' file2
get 'dba file for peak pick:' dbaf
get 'contour level:(usu .02 is good)' contl
get 'Apodisation Fxn1:' fxn1
get 'Apodisation Fxn2:' fxn2
get 'peak halfwidth:' hfwd
dba file open &dbaf.dba                    ;open dba file:file.dba
define hafwid &hfwd
dba entity delete vol:volumes              ;delete previous volumes data in the dba
def count 0
for inp &file1 &file2                      ;the data files to be processed listed here
 eva count (&count+1)
 ty count=&count
```

```
wai 1                                   ;1 second delay to see the "count'
define input &inp
ty Building Matrix:&inp.mat...
bld &inp 2 1024 1024 0
mat &inp write
ty Doing T2 Transform...
def irow 1
for row 1 256
re &input
def datype 1
bc
&fxn1                                   ;ss 512 5 recommended
if &bv eqs bruker then                  ;determine which kind of fourier-ft
 bft                        ;this is the only difference between bruker and varian
else
 ft
eif
sep
def datype 0
def datsiz 1024
sto 0 &irow
eva irow (&irow+1)
def datsiz 2048
shl 1024
def datsiz 1024
sto 0 &irow
eva irow (&irow+1)
ty row=&row $
esc out
if &out ne 0 quit
next
ty Doing T1 Transform...
for col 1 1024
loa &col 0
```

```
 def datype 1
 &fxn2                                                    ;ss 256 10 recommended
 zf 1024
 ft
 ms
 rev
 sep
 def datype 0
 def datsiz 1024
 sto &col 0
 ty col=&col $
 esc out
 if &out ne 0 quit
 next
 cmx                                ;the transformation is finished here
 mat &inp.mat r                     ;open matrix for vol pick and measure at following
 lim 1 60 365                       ;to just show the fingerprint region in dimension 1
 lim 2 336 653                      ;to just show the fingerprint region in dimension 2
 define level &contl                      ;define level
 define nlevel 2                           ;number of level
 cp                                       ;contour plot
 drx xpk:peaks 2                          ;displays cross-peaks in the database
 if &count ge 7 and &count le 12 then
   eva count (&count-6)                   ;check if the vol slot# will exceed 6
 eif
 vol xpk:peaks vol:volumes &count 0.00   ;measure the volume
 dba entity write vol:volumes &dbaf_&inp.txt   ;generate ASCII file for output
 esc out                                   ;check for escape
 if &out ne 0 quit                         ;quit if escape
 ty &inp.mat is done, its volumes are saved in slot# &count !
 cmx
next
quit:
ty All Done....
```

```
dba file close &dbaf.dba                    ;close dba file
cmx
cl
ex &return
c** macro ends here
```