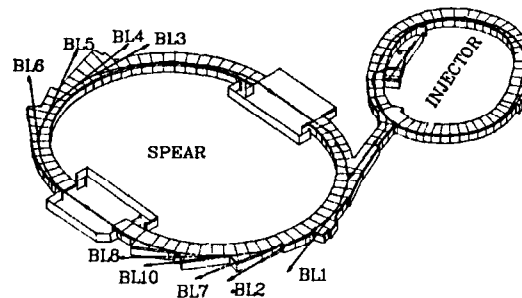


**PROTEIN FOLDING AND PROTEIN METALLOCLUSTER
STUDIES USING SYNCHROTRON SMALL ANGLER
X-RAY SCATTERING**

David Eliezer

SLAC-Report-448
June 1994



Prepared for the Department of Energy
under contract number DE-AC03-76SF00515
and the NIH, Biomedical Resource Technology Program,
Division of Research Resources

STANFORD LINEAR ACCELERATOR CENTER
STANFORD SYNCHROTRON RADIATION LABORATORY
Stanford University • Stanford, California

This document and the material and data contained therein, was developed under sponsorship of the United States Government. Neither the United States nor the Department of Energy, nor the Leland Stanford Junior University, nor their employees, nor their respective contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any liability or responsibility for accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use will not infringe privately-owned rights. Mention of any product, its manufacturer, or suppliers shall not, nor is it intended to, imply approval, disapproval, or fitness for any particular use. A royalty-free, nonexclusive right to use and disseminate same for any purpose whatsoever, is expressly reserved to the United States and the University.

SLAC-448
SLAC/SSRL-0085
UC-408
(SSRL-M)

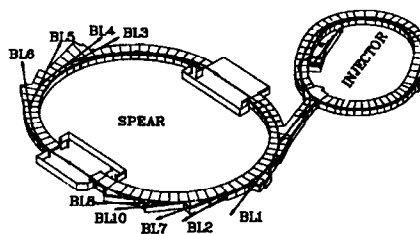
PROTEIN FOLDING AND PROTEIN METALLOCLUSTER STUDIES USING SYNCHROTRON SMALL ANGLE X-RAY SCATTERING

David Eliezer

*Stanford Linear Accelerator Center
Stanford Synchrotron Radiation Laboratory
Stanford University, Stanford, California 94309*

June 1994

Prepared for the Department of Energy under contract number DE-AC03-76SF00515 and
the NIH, Biomedical Resource Technology Program, Division of Research Resources



Printed in the United States of America. Available from the National Technical
Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield,
Virginia 22161

* Ph.D. thesis

**PROTEIN FOLDING AND PROTEIN METALLOCLUSTER STUDIES USING
SYNCHROTRON SMALL ANGLE X-RAY SCATTERING**

**A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY**

David Eliezer

June 1994

Abstract

Proteins, biological macromolecules composed of amino-acid building blocks, possess unique three dimensional shapes or conformations which are intimately related to their biological function. All of the information necessary to determine this conformation is stored in a protein's amino acid sequence. The problem of understanding the process by which nature maps protein amino-acid sequences to three-dimensional conformations is known as the protein folding problem, and is one of the central unsolved problems in biophysics today. The possible applications of a solution are broad, ranging from the elucidation of thousands of protein structures to the rational modification and design of protein-based drugs.

The scattering of X-rays by matter has long been useful as a tool for the characterization of physical properties of materials, including biological samples. The high photon flux available at synchrotron X-ray sources allows for the measurement of scattering cross-sections of dilute and/or disordered samples. Such measurements do not yield the detailed geometrical information available from crystalline samples, but do allow for lower resolution studies of dynamical processes not observable in the crystalline state. The main focus of the work described here has been the study of the protein folding process using time-resolved small-angle x-ray scattering measurements.

The original intention was to observe the decrease in overall size which must accompany the folding of a protein from an extended conformation to its compact native state. Although this process proved too fast for the current time-resolution of the technique, upper bounds were set on the probable compaction times of several small proteins. In addition, an interesting and unexpected process was detected, in which the folding protein passes through an intermediate state which shows a tendency to associate. This state is proposed to be a kinetic molten globule folding intermediate. The existence of such states has implications for the pathway followed by proteins as they fold.

Acknowledgments

A large number of people have made contributions, both direct and indirect, to this work, and it is not possible to thank them all here. I apologize to those who are not explicitly mentioned.

I would first like to thank my advisors, Seb Doniach and Keith Hodgson, for their patience and support during the often uncertain course of my research. I am grateful to Seb for his honesty and directness, which helped me to understand that character is as important in science as it is in other facets of life. I must also acknowledge Keith's unwavering support, both psychological and financial, of my numerous trips to Japan, without which I would not have been able to complete this thesis. Keith has always been encouraging, even where others may have balked in the face of initially scarce results.

In addition, I would like to thank Hiroshi Kihara, who has been an advisor away from home during my experiments at the Photon Factory. Kihara-san first suggested using myoglobin for the protein folding experiments. In addition to his invaluable contributions to my research, he was a gracious and generous host, and a good friend.

Next I would like to thank Britt Hedman and Patrick Frank. Britt is one of the most efficient and capable people I have encountered, and she has always been ready and able to find a 25th hour in any given day to help me with the emergency at hand. In addition to his crucial contributions to the FeMoco experiments, Pat has been a constant source of support and optimism during my stay at Stanford. He helped me from the brink of depression many times, and has always been ready with a kind word or an amusing anecdote.

I also owe a debt of gratitude to Yoshiyuki Amemiya. Amemiya-san acted as my host sponsor during my first visit to Japan, arranged for my first experiment at the Photon Factory, and introduced me to Kihara-san and to Hiro. Despite his amazingly busy schedule, he has been a constant help in all of our experiments at the PF. He is also an extremely nice individual, and is responsible for my best impressions of Japan, of Japanese culture, and of Japanese food.

One more Japanese friend is Hiro Tsuruta. Hiro has probably had the most direct impact on my work, being a close collaborator in all of the protein folding experiments, and the

expert on the stopped-flow rapid-mixers. He worked side-by-side with me for days at a time on beamlines at both SSRL and the PF, and shared in all of the 'pleasures' of synchrotron experiments. He is an extremely hard worker with an easy-going personality, which made working with him a pleasure.

I joined the Hodgson group in part because it was full of such nice people. Soichi, Trevor, Nanna, Mike, Jane, and Grace were all friends, with Soichi serving as mentor and tutor during my first few years. My remaining compatriots, Chrisie, Susan, Kent, Lingling, Tami, Isaac, Heather, Kendra, and Holly, have continued this tradition. I am glad to have known them, and wish them all speedy success. I would also like to thank Marybeth and Mike from the Bienenstock group for our shared adventures with the SAXS camera.

Other people who deserve mention are Michel Koch, who gave me some excellent advice about the facts of science, Pat Immormino, who helped me with innumerable emergency orders and packages, Debra Brodie, Lore Jung, and Marcia Keating, who helped me maintain contact with the physics department, Robert Mayer, who helped with many small and large tasks at SSRL, all of the other staff and scientists at SSRL who contributed something at some point, and Kent, Ross, Gideon, Chris, Tom, Max, and the rest of my entering class.

Many friends, near and far, including Lee-may, Denise, Alice, Julie, Eva, Brad, Greg, Eric, Alfred, Erik, Barbara, Anne, and Ariane, have helped me by listening to my complaints and listening to my complaints, and I want to thank them all.

Finally, I want to thank Eran and Ken and mom and dad for putting up with this graduate school business, and Debra for making the last three years the best three years.

Table of Contents

Abstract	iv
Acknowledgments	v
Table of Contents	vii
List of Tables	x
List of Figures	xi
Introduction	1
Small Angle Scattering as an Experimental Technique in Physics and Biophysics	1
Protein Folding	4
Chapter 1: The Small Angle X-ray Scattering Technique	7
General Overview	7
Relevant Theory	8
Experimental Details, General	12
The Synchrotron X-ray Source	12
Collimation	14
Solution Samples	15
Normalization Measurements	16
Detectors	18
Data Acquisition	20
Basic Data Analysis	21
Chapter 2: The Iron-Molybdenum Cofactor from <i>Azotobacter vinelandii</i> Nitrogenase	28
Motivation	28
Literature Review	29
Sample Preparation	31
FeMoco	31
Model Compounds	32
Sample Integrity	33

FeMoco	33
Model Compounds	34
Experimental Details, specific	34
Data Analysis and Results	39
Conclusions	43
Chapter 3: Protein Folding	56
Motivation	56
Literature Review	57
A: Time-Resolved Small Angle X-ray Scattering	60
Triggering by the Stopped-Flow Method	60
Stopped-Flow Rapid-Mixer Description	61
B: Ribonuclease A	65
Sample Preparation	65
Experimental Details, Specific	67
Analysis and Results	69
Conclusions	72
C: Myoglobin	86
Evidence for an Associative Folding Intermediate	86
Sample Preparation	86
Experimental Details, Specific	87
Analysis and Results	88
Discussion	92
Conclusions	94
Further Myoglobin Studies	96
Experimental Details, Specific	96
Analysis and Results	96
Conclusions	99
D: Cytochrome- <i>c</i>	124
Sample Preparation	124
Experimental Details, Specific	124
Analysis and Results	125
Discussion	128
Conclusions	131

Chapter 4: White Light SAXS	145
Motivation	145
Numerical Treatment	146
Experimental Details	148
Analysis and Results	149
Conclusions	151
Conclusion and Credits	160

List of Tables

Chapter 2:

Table 2.1: Observed Radii of Gyration for FeMoco and Models	48
Table 2.2: Average Observed and Calculated Radii of Gyration	49
Table 2.3: Observed and Calculated Mass Ratios	50

Chapter 3:

Table 3.C.1: Parameters from the Time-Course of the Forward Scattering from Myoglobin During Refolding	101
Table 3.C.2: Parameters from Temperature and Concentration Studies of the Refolding of Myoglobin	102
Table 3.D.1: Parameters from the Time-Course of the Forward Scattering from Cytochrome- <i>c</i> During Refolding	132

Chapter 4:

Table 4.1: White Light Radii of Gyration	154
--	-----

List of Figures

Chapter 1: SAXS

Figure 1.1: Diagram of a typical two-slit collimation system 27

Chapter 2: FeMoco

Figure 2.1: The structure of FeMoco 51

Figure 2.2: Scattering data from FeMoco and a model compound 52

Figure 2.3: FeMoco Guinier plots 53

Figure 2.4: Intensity contribution from homogeneous spheres 54

Figure 2.5: Maximum percentage contribution from homogeneous spheres 55

Chapter 3.A: Time-Resolved SAXS

Figure 3.A.1: Schematic of the stop-cock rapid-mixer 63

Figure 3.A.2: Circuit for remote control of the ruby-ball rapid-mixer 64

Chapter 3.B: Ribonuclease A

Figure 3.B.1: GuHCl-induced equilibrium unfolding 73

Figure 3.B.2: Extrapolation of the free energy of unfolding 74

Figure 3.B.3: SAXS data from February 1991 at 0, 4, and 8 M urea 75

Figure 3.B.4: SAXS data from November 1991 at 0, 4, and 8 M urea 78

Figure 3.B.5: R_g and I_0 from equilibrium unfolding studies 79

Figure 3.B.6: R_g from thermal denaturation at 0, 1, 2, and 4 M urea 80

Figure 3.B.7: R_g and I_0 during continuous thermal denaturation 82

Figure 3.B.8: R_g and I_0 as a function of refolding time 83

Figure 3.B.9: Comparison of the first and final data frames from refolding 85

Chapter 3.C: Myoglobin

Figure 3.C.1: CD data from equilibrium unfolding 103

Figure 3.C.2: R_g and I_0 from equilibrium unfolding studies 104

Figure 3.C.3: Representative time-courses of R_g during refolding 105

Figure 3.C.4: Time-courses of I_0 during refolding 106

Figure 3.C.5: First and final data frames from refolding	109
Figure 3.C.6: Time-courses of I_0 during control mixing experiments	110
Figure 3.C.7: Final condition after refolding and control experiments	111
Figure 3.C.8: Kratky plots from native and unfolded myoglobin	112
Figure 3.C.9: Kratky plots of initial and final data frames from refolding	113
Figure 3.C.10: Two possible folding pathways	114
Figure 3.C.11: Time-courses of R_g from temperature dependence studies	115
Figure 3.C.12: Time-courses of I_0 from January, 1993	116
Figure 3.C.13: Time-courses of I_0 from March, 1993	118
Figure 3.C.14: Apparent dissociation rate constant <i>versus</i> temperature	121
Figure 3.C.15: $-\ln K$ <i>versus</i> $1/T$	122
Figure 3.C.16: I_0 time-courses from concentration dependence studies	123
Chapter 3.D: Cytochrome-c	
Figure 3.D.1: R_g and I_0 from equilibrium unfolding	133
Figure 3.D.2: Data from unfolded protein at 5 different concentrations	134
Figure 3.D.3: R_g and I_0 time-courses from control mixing experiments	135
Figure 3.D.4: Representative R_g time-courses during refolding	136
Figure 3.D.5: I_0 time-courses during refolding	137
Figure 3.D.6: I_0 time-course during slow refolding process	140
Figure 3.D.7: Data from immediately- and long-after sample mixing	141
Chapter 4: White Light SAXS	
Figure 4.1: Attenuated white light energy distribution	155
Figure 4.2: Integrated monochromatic data <i>versus</i> scattering angle	156
Figure 4.3: Integrated monochromatic data <i>versus</i> $\langle K \rangle$	157
Figure 4.4: White light data from four proteins	158
Figure 4.5: Integrated monochromatic data from phosphorylase <i>b</i>	159

Introduction

The research presented in this dissertation is largely concerned with the application of small angle x-ray scattering (SAXS), a useful experimental technique in physics, material science, and biophysics, to problems in biophysics, most notably to the problem of protein folding.

Small Angle Scattering as an Experimental Technique in Physics and Biophysics:

Small angle scattering (both x-ray and neutron) is an experimental technique used in the study of a variety of problems in condensed matter physics and material science related mostly to properties of the statistical distributions (pair distribution functions, phase separations, density fluctuations, compressibilities, *etc.*) of particles in the solid or liquid states. In addition, small angle scattering has been found to be a useful tool for the study of physical properties (size, shape, and mass) of biological macromolecules in solution. As such, it is part of an array of techniques available in the rapidly expanding field of biophysics, which includes the application of physical techniques toward obtaining an understanding of the physical properties of biological systems.

Camera Development:

The fundamental requirement for any SAXS experiment is a SAXS camera, a system for irradiating a sample with a highly collimated beam of x-ray photons, and detecting the photons scattered at small angles to the incident beam efficiently. Although a biotechnology SAXS camera has existed at the Stanford Synchrotron Radiation Laboratory for many years, a large fraction of the effort involved in the research described herein was devoted to modifications and improvements of this experimental apparatus. A brief summary of these changes follows.

Because synchrotron experiments must be performed inside sealed hutches which are off-limit to experimenters during sample irradiation, much of the experimental apparatus needs to be remotely controllable. The development and maintenance of mechanical and electronic remote control systems for camera and sample alignment was thus an important part of these experiments.

The previous SAXS camera allowed for the study of solution samples only in quartz or glass capillaries. Capillaries are unsatisfactory as sample holders for a number of reasons. They are fragile, difficult to handle, difficult to remount accurately, difficult to align in the beam, and capable of reflecting the x-ray beam (thus endangering the detector and contaminating the data) from their upper or lower side walls. During the course of this work, solution sample cells employing flat quartz and later flat mica windows were incorporated into the SAXS camera. These sample holders do not have the problems listed above. For the experiments with the air-sensitive samples described in chapter two, however, capillaries had to be used because the flat window cells are not air-tight and are sensitive to the solvents used in these experiments. For these experiments, new capillary holders were developed in order to minimize the problems listed above.

Another important contribution to the current SAXS camera was the development of new and most importantly longer scattering paths. The original camera was limited to about an 80 cm sample-to-detector flight path. Three separate scattering path upgrades resulted in the current system, which allows anywhere from a 30 cm up to a 300 cm sample-to-detector evacuated flight path. In addition, a larger beamline hutch at SSRL was converted to a semi-permanent home for the camera in order to accommodate the now much longer apparatus. The new long flight path is crucial to the small angle resolution of the camera.

The previous SAXS camera used an old one-dimensional detector with a 10 cm by 1 cm active area, and without on-board preamplifiers. The shorter length of this detector limited the availability of data at higher scattering angles. In addition this detector was extremely difficult to calibrate, and was sensitive to noise because the signal preamplifiers had to be outside the experimental hutch, far from the detector. Two new detectors, a longer linear detector, and a 60° fan-shaped quadrant detector were added to the SAXS camera during this research. Both of these provide better data at higher scattering angles, and are less sensitive to noise because of on-board signal preamplifiers. The detectors and associated electronics were purchased separately, and a serious effort was required to test, adjust and calibrate the electronics in order to obtain optimal detector performance.

In addition to the above contributions, significant additions were made to the data acquisition and data analysis software packages which have seen ongoing development during the lifetime of the SSRL biotechnology SAXS camera.

Time-Resolved Technique:

The protein folding experiments discussed in chapter three rely on the stopped-flow time-resolved SAXS method, the capability for which was successfully incorporated into the SSRL SAXS camera. This process involved design modifications of and purchase of a stopped-flow rapid-mixing apparatus from a commercial manufacturer, design, assembly, and testing of the electronics required to operate the mixer in conjunction with sample exposure to x-rays, and the writing of the software necessary to drive the electronics, to collect time-resolved SAXS data, and to analyze the resulting data.

Broad-Bandpass Technique:

Chapter four discusses the development of another improvement to the SAXS technique. Many SAXS experiments, especially those involving weakly scattering samples such as dilute solutions of biological macromolecules, are limited by the available flux of x-ray photons. This research has opened a new avenue for mitigating this limitation, the use of very broad bandpass radiation from synchrotron sources. We have shown in a series of analytical and computational calculations that several of the basic SAXS parameters of interest, the radius of gyration and the forward scattered intensity, can be reliably determined from SAXS data collected using unmonochromatized 'white' light. This means that the photon flux for experiments interested in these parameters can be increased by two to three orders of magnitude. This represents a major advance for the study of samples which are not highly soluble (*i.e.* very dilute samples) and for time-resolved studies on a sub-millisecond time scale. This technique should have non-biological applications as well.

Anomalous Scattering Technique:

The anomalous scattering technique relies on the relatively large changes in atomic scattering lengths at or near resonance. Experiments performed at such energies can be used to isolate the scattering contributions of individual element components of heterogeneous systems. This technique has widespread applications in material science, and more limited applications in biophysics as well. Although none of the research discussed in this work employs this technique, the camera system used in these experiments was developed to be able to perform anomalous scattering experiments in preparation for an extension of the research described in chapter two on the

iron-molybdenum cofactor of the nitrogenase enzyme system. In particular, some effort was expended in developing hardware and writing and maintaining software to enable data collection at different energies around an absorption edge. This capability was effectively used in the dissertation research of two fellow physicists (Rice, 1994; Regan, 1994).

Protein Folding:

The protein folding problem is the problem of understanding the mapping between protein amino acid sequence and protein three-dimensional structure and understanding the details of the implementation of this mapping in nature. It is a problem which belongs to a number of disciplines, including mathematics, computer science, physics, chemistry, biology, and medicine. The problem was first posed by Anfinsen in 1973 when he showed that all of the information required to define a protein's three-dimensional structure was contained in the amino-acid sequence of that protein (Anfinsen, 1973). Since then, it has received ever increasing attention, but little progress has been made towards a solution. In the physics community, it has been considered as a variant of a random walk problem in statistical physics (Bascle, 1994; Warvari, 1972), as a constrained many body problem (Shakhnovich, 1991), as a disordered (glass-like) systems problem (Sasai, 1990), and even as a deterministic chaos problem (Böhm, 1991). Experimentally, however, the problem has been addressed almost exclusively by techniques (mostly spectroscopic) in physical chemistry and biochemistry. The research presented in this dissertation involves the application of an experimental technique from physics, small angle x-ray scattering, to the protein folding problem.

Recent progress in the field has indicated that one of the earliest events on the path of a protein molecule from an extended or unfolded conformation to a compact or folded conformation is the formation of secondary structure elements. In addition, it is obvious that somewhere along such a path, the protein molecule has to decrease in size, or undergo compaction. In order to appropriately model the folding process, it is necessary to know the temporal relation between the compaction event and the formation of secondary structure. Unfortunately, none of the methods employed in the literature to date are capable of measuring the size of a protein as it folds. Because small angle x-ray scattering is capable of measuring physical parameters such as the protein radius of gyration, which is a good indicator of size, it is a technique which is uniquely suited for the study of this problem at this time.

The results of the studies described in this thesis have added an important piece of information to the protein folding puzzle. It has long been suspected that a general kinetic intermediate on folding pathways is a state known as the molten globule state (Ptitsyn, 1987). This state is characterized in equilibrium as being compact and possessing some degree of native secondary structure, but missing all or most of the native tertiary contacts. Up until quite recently, however, there have been no conclusive observations of such a state during folding. Observations of early secondary structure formation abound, but no clear indications of compactness have accompanied these results. The myoglobin studies detailed in chapter three do provide a clear indication of compactness, and are one of the first pieces of direct evidence for a populated kinetic molten globule intermediate state on a protein folding pathway. Further evidence in support of this result has recently appeared in the literature (Jennings, 1993). Preliminary evidence is also presented for the existence of a kinetic molten globule state on the folding pathway of cytochrome-*c*.

One important outcome of this result is that the molten globule state now assumes the unique distinction of being a stable kinetic folding intermediate. This means that properties which belong to equilibrium molten globule states are likely to also belong to the equivalent kinetic molten globule folding intermediates. Because equilibrium molten globules are long-lived, they are far easier to study than kinetic intermediates, and provide the best opportunity yet to gather information about the location and type of structure formed early in the protein folding process.

Although none of the experiments described in this work succeeded in observing the compaction process of the proteins under study, they have established an experimental procedure by which it should be possible to observe this process. In the meanwhile, upper bounds have been set for the time-scale on which this process must occur under the conditions of the studies. From this upper-bound, it seems likely that at least some secondary structure formation occurs after protein compaction. Any correct theoretical models of the protein folding process will have to take these observations into account.

References:

Anfinsen, C.B. 1973. *Science*. 181:223-230.

Bascle, J., Doniach, S., and Orland, H. 1994. *J. Mol. Biol.* Manuscript under revision.

Böhm, G. 1991. *Chaos, Solitons, & Fractals*. 4:375-382.

Jennings, P.A. and Wright, P.E. 1993. *Science*. 262:892-896.

Ptitsyn, O.B. 1987. *J. Protein. Chem.* 6:273-293.

Regan, M. 1994. Ph.D. Dissertation. Department of Applied Physics. Stanford University.

Rice, M. 1994. Ph.D. Dissertation. Department of Electrical Engineering. Stanford University.

Sasai, M., and Wolynes, P.G. 1990. *Phys. Rev. Lett.* 65:2740-2743.

Shakhnovich, E., Farztdinov, G., Gutin, A.M., and Karplus, M. 1991. *Phys. Rev. Lett.* 67:1665-1668.

Warvari, H.E., and Scott, R.A. 1972. *J. Chem. Phys.* 57:1146-1153.

Chapter 1

The Small Angle X-ray Scattering Technique:

This chapter treats the reader to a review of the small angle scattering technique, including an overview of the literature, information content, and common applications of SAXS, a summary of the relevant theoretical treatment, a presentation of general experimental procedures, and a guide to basic SAXS data analysis. This chapter is written with the specific intent of making this thesis a useful and therefore somewhat self-contained text for future graduate students who may wish to continue the work started here. Very little if any of the material presented here is original, and an earnest attempt has been made to reference thoroughly. The content is not meant to be exhaustive, but rather to present the facets of the above subjects most relevant to the current work, and to provide the reader with citations to more in-depth treatments where necessary.

General Overview:

The documented use of SAXS begins in the 1930s, but the beginning of real interest in the field must be attributed to André Guinier and his derivation of Guinier's law (Guinier, 1939). The first textbook devoted to SAXS appeared in 1955 (Guinier, 1955) and in 1982 the classic text by Glatter and Kratky (Glatter, 1982) was published. In 1954, the SAXS technique was joined by SANS, small angle neutron scattering, with the publication of the Van Hove's neutron-scattering law (Van Hove, 1954). The difference of the interactions between photons and electrons and between neutrons and nuclei have made this technique a useful complement to SAXS. SANS has its own separate literature, which contains many interesting developments, but is not discussed here. In 1987 an excellent text devoted to small angle scattering by both x-rays and neutrons was published (Svergun, 1987).

The work in this dissertation has focussed on the derivation of two parameters from SAXS data, the radius of gyration of the scatterer and its relative molecular weight. The theoretical development of Guinier's law, which describes the extraction of these parameters from SAXS data is given in the following section. Additional information beyond these two parameters can be extracted from SAXS data in favorable circumstances. For homogeneous scatterers, this include the volume, mass, surface area,

and maximum dimension of the scatterer. Some information regarding the gross shape of the scatterer can also be obtained. In principle the characteristic function or distance distribution function is available as the Fourier transform of the measured intensity and contains all the available information. For polymer samples additional information such as the persistence length may be available. Unfortunately, all of these further parameters can only be determined from extremely high quality data. The studies herein focus on time-resolved measurements and on measurements from particularly weak scatterers, and data of sufficient quality to merit these more complex interpretations are not (yet) obtainable. The reader is referred to Russel, 1991 and Glatter, 1982 for further details on more sophisticated data analysis techniques.

Relevant Theory:

The interactions of photons with free electrons can be separated into coherent (elastic, Thompson, Rayleigh) scattering and incoherent (inelastic, Compton) scattering. When an electron is bound to an atom, there is also the possibility of Raman scattering, and photoabsorption (the absorption of a photon by core electrons which are then ejected from the atom). The hole left after photoabsorption is responsible for a host of subsequent effects, the most important of which in this context are fluorescence and the Auger effect. At energies between 5 and 20 keV, the photoabsorption cross-section is the largest of the three for most elements, and is roughly proportional to $CZ^4\lambda^3$ between absorption edges. However, the interaction which will be dealt with exclusively in this dissertation is elastic scattering.

The classical treatment of x-ray scattering begins with the scattering of an electromagnetic wave by a free electron (Thomson scattering). The assumption that the velocity of the electron is much less than the speed of light allows one to ignore the Biot-Savart force from the magnetic field, leaving only the oscillating electric field. The equation of motion is $m\mathbf{a} = e\mathbf{E}$, leading to a fluctuating dipole moment, $\mathbf{\ddot{d}} = e^2\mathbf{E}/m$ (from here on bolded characters represent vectors, and the same characters unbolded represent vector magnitudes). In the far field dipole approximation (Landau, 1987) the power emitted by an oscillating charge into a solid angle $d\Omega$ is $dI = (\mathbf{\ddot{d}} \times \mathbf{n})^2 d\Omega / 4\pi c^3$, which in this case equals $e^4(\mathbf{E} \times \mathbf{n})^2 d\Omega / 4\pi m^2 c^3$. Letting $\mathbf{E} = E_y \mathbf{e}_y + E_z \mathbf{e}_z$ and $\mathbf{n} = n_x \mathbf{e}_x + n_z \mathbf{e}_z$, the cross product term evaluates to $E_y^2 + E_z^2 \cos^2\theta$, where the incident wave propagates in the x direction and θ is the scattering angle between the directions of propagation of the scattered and incident waves. For random polarization of the incident wave, $E_y^2 = E_z^2 =$

$E^2/2$. The Poynting vector of the incoming wave is $I = cE^2/4\pi$, leading to a differential cross-section $d\sigma_e = dI/I = (e^2/mc^2)^2(1 + \cos^2\theta)d\Omega/2$. This is the Thomson equation. Integration over all angles yields $\sigma = 8\pi/3(e^2/mc^2)^2 = 8\pi r_e^2/3$, the Thomson cross-section, where r_e is the classical electron radius.

Because the electric field of the scattered wave is proportional to $(\ddot{\mathbf{d}} \times \mathbf{n}) \times \mathbf{n} = -\ddot{\mathbf{d}}$ (Landau, 1987) its phase is shifted by π from that of the incident wave. When a wave interacts with a system of many (non-interacting) electrons, the scattered wave from each electron is shifted by the same factor of π and hence this scattering process is coherent. If \mathbf{n}_0 is a unit vector in the direction of the incident wave, and \mathbf{n} is in the direction of an observer, the waves scattered from two different electrons separated by \mathbf{r} will travel paths of different length, that difference being equal to $(\mathbf{n} - \mathbf{n}_0) \cdot \mathbf{r}$. This leads to a phase difference of $\mathbf{K} \cdot \mathbf{r}$, where $\mathbf{K} = 2\pi(\mathbf{k} - \mathbf{k}_0) = 2\pi(\mathbf{n} - \mathbf{n}_0)/\lambda$ is the scattering (wave) vector, and the waves will interfere with each other. The differential scattering cross-section for such a system is $d\sigma_e |\sum e^{i\mathbf{K} \cdot \mathbf{r}}|^2$. In general, electrons are not localized, and this expression goes over to an integral over the electron density.

When the system of electrons is an atom, the electron density can be calculated, and the integral evaluated to give the atomic differential scattering cross-section. The above formula however is valid only when the incident wave energy is far from any of the binding energies of the electrons, and when second order scattering is neglected (the first Born approximation). When the incident wave energy approaches an electron binding energy, resonance scattering takes place, and the above treatment is not valid. Because the scattering cross-section is quite small however, it is almost always valid to neglect second order scattering. The results of detailed atomic cross-section calculations are tabulated in the literature (see for example McMaster, 1969). Frequently the atomic differential scattering cross-section is normalized by the single electron Thomson cross-section, giving the square of the atomic form factor. In the direction of the incident wave (forward scattering) the atomic form factor is just the number of electrons in the atom.

The above treatment is also valid when the system under consideration is a macromolecule in solution. In this case, however, it is convenient to write the electron density as $\rho = \rho_x + \rho_s$, where ρ_s is the electron density of the solvent and ρ_x is the 'excess' electron density of the molecule, defined by this equation. This is useful because ρ_s usually fluctuates on a length scale much smaller than the dimensions of the

macromolecule, and can therefore be treated as constant. In evaluating the double integral in the differential scattering cross-section, the term $\rho_{1x}\rho_{2x}$ is then written as $(\rho_{1x} + \rho_s)(\rho_{2x} + \rho_s)$. All of the cross terms except $\rho_{1x}\rho_{2x}$ contain a factor equaling the Fourier transform of ρ_s over the entire sample volume. This is effectively a delta function at zero-scattering angle, thus these terms do not contribute to scattering at detectable angles. This demonstrates that the addition or subtraction of a constant electron density does not change the scattering cross-section. Only the excess electron density of a macromolecule leads to (detectable) scattering.

We now have the macromolecular differential scattering cross-section

$$d\sigma \sim |\int \rho_x e^{i\mathbf{K}\cdot\mathbf{r}}|^2 = \int \rho(\mathbf{r}_1) e^{i\mathbf{K}\cdot\mathbf{r}_1} d\mathbf{r}_1 \int \rho(\mathbf{r}_2) e^{-i\mathbf{K}\cdot\mathbf{r}_2} d\mathbf{r}_2$$

Where from here on ρ refers to the (real) excess electron density. At this point, we must take into account that in solution samples, the various molecules are randomly oriented, and therefore the scattering cross-section must be spherically averaged over all possible orientations of \mathbf{r}_1 and \mathbf{r}_2 . This is equivalent to averaging over all possible orientations of \mathbf{K} . Only the factor $e^{\pm i\mathbf{K}\cdot\mathbf{r}}$ is effected, and becomes the Debye integral $4\pi \langle e^{\pm i\mathbf{K}\cdot\mathbf{r}} \rangle = \int e^{\pm i\mathbf{K}\cdot\mathbf{r}} d\Omega = \int e^{\pm iKr \cos\theta} \sin\theta d\theta d\phi = -2\pi(e^{-iKr} - e^{iKr})/iKr = 4\pi \sin Kr/Kr$, where $K = 4\pi \sin(\theta/2)/\lambda$ is the magnitude of the scattering vector \mathbf{K} , and θ is the scattering angle (K will be used in this manner throughout the remainder of this and the following chapters). We then have

$$d\sigma \sim \int \rho(\mathbf{r}_1) (\sin Kr_1/Kr_1) d^3r_1 \int \rho(\mathbf{r}_2) (\sin Kr_2/Kr_2) d^3r_2$$

We now confine our interest to small angles such that $Kr \ll 1$. In this case, we can expand $\sin(x)$ as $(x - x^3/6)$, and we get

$$d\sigma \sim \int \rho(\mathbf{r}_1) (1 - K^2 r_1^2/6) d^3r_1 \int \rho(\mathbf{r}_2) (1 - K^2 r_2^2/6) d^3r_2$$

letting $\int \rho(\mathbf{r}) d^3r = N_x$, the 'number' of excess electrons in the molecule, and defining $R_g^2 = \int r^2 \rho(\mathbf{r}) d^3r / \int \rho(\mathbf{r}) d^3r$ we get

$$d\sigma \sim N_x^2 (1 - K^2 R_g^2/6) (1 - K^2 R_g^2/6) = N_x^2 (1 - K^2 R_g^2/3) + \dots$$

This form is the first order expansion of an exponential, and we finally arrive at Guinier's

law (Guinier, 1939), which states that at small enough angles,

$$d\sigma = d\sigma_e N_x^2 e^{-K^2 R_g^2/3}$$

where R_g is the second moment of the excess electron density of the sample, known as the electronic radius of gyration. Guinier's law is quite general and applies to essentially any electron distribution, including compact proteins, extended polymers, and even density fluctuations in solids. Note that the third term in the above series expansion, geometrical considerations aside, is a factor of $5!/3! = 20$ smaller than the second term. Therefore, the Guinier approximation is typically valid at angles $KR \leq 1$. This has been confirmed empirically for globular scatterers. A good analytical treatment of the accuracy of Guinier's law as a function of KR is given by Svergun and Feigin (1987). It is also noteworthy that the forward scattered intensity, frequently (and henceforth) referred to as I_0 , is proportional to the square of the integral of the excess electron density. If the particle is reasonably homogeneous, this will be proportional to the square of the volume of the scattering molecule.

It is also possible to write the differential scattering cross-section as

$$d\sigma \sim \langle \int \int \rho(\mathbf{r})\rho(\mathbf{R}+\mathbf{r})d^3r e^{i\mathbf{K}\cdot\mathbf{R}}d^3R \rangle = 4\pi \int \gamma(R)(\sin KR/KR)R^2 dR$$

where $\mathbf{R} = \mathbf{r}_1 - \mathbf{r}_2$ and $\gamma(R) = \langle \int \rho(\mathbf{r})\rho(\mathbf{R}+\mathbf{r})d^3r \rangle$ is the characteristic function (or rotationally averaged autocorrelation or Patterson function) of the scatterer, and is related to R_g by

$$\begin{aligned} \int R^2 \gamma(R) d^3R &= \langle \int \int R^2 \rho(\mathbf{r})\rho(\mathbf{R}+\mathbf{r})d^3r d^3R \rangle \\ &= \langle \int \int (r_1^2 + r_2^2 - 2\mathbf{r}_1 \cdot \mathbf{r}_2) \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)d^3r_1 d^3r_2 \rangle \end{aligned}$$

(now because we work in center of mass coordinates $\int \mathbf{r}\rho(\mathbf{r})d^3r = 0$)

$$= 2R_g^2 \langle (\int \rho(\mathbf{r})d^3r)^2 \rangle = 2R_g^2 \int \gamma(\mathbf{r})d^3r$$

$$\text{so that } R_g^2 = \int R^2 \gamma(R) d^3R / 2 \int \gamma(R) d^3R$$

It is interesting to calculate some actual numbers for the scattering cross-section of a

typical protein in solution. In general, proteins are extremely compact molecules with a rather uniform electron density. For myoglobin for example we can use an electron density of about $0.45 \text{ e}^-/\text{\AA}^3$. For water, the electron density is about $0.33 \text{ e}^-/\text{\AA}^3$, and the excess electron density is about $0.12 \text{ e}^-/\text{\AA}^3$. The calculated specific volume (Watson, 1965) is $25,300 \text{ \AA}^3$, giving about 3000 excess electrons. In the forward direction, then, with $d\sigma_e = r_e^2 = 8 * 10^{-10} \text{ \AA}^2$ we get a differential cross-section of about 10^{-2} \AA^2 or 10^{-16} mm^2 . The number of myoglobin particles in a 1 mm^3 volume of 10 g/l protein is about $3.4 * 10^{14}$, giving a forward differential scattering cross-section from the entire sample of $3.4 * 10^{-2}$. As seen above this value falls off exponentially (with both angle and photon energy) at angles different from zero. For $d\theta = 1 \text{ mrad}$ this cross-section gives a scattering probability of about 10^{-5} . This is in reasonable agreement with the count rates observed at very low angles after including absorption effects and detector efficiency.

Experimental Details, General:

The Synchrotron X-ray Source:

In considering the experimental procedures used in this work, it is vital to keep in mind that all of the experiments discussed were performed using a synchrotron x-ray source. This section is devoted to characteristics of synchrotron sources which are of interest in SAXS experiments.

A detailed description of the basic physics of the production of synchrotron radiation by a charged particle undergoing circular motion can be found in Jackson, 1975 or Landau, 1987. Calculating the electric and magnetic fields from the Liénard-Wiechert potentials for such a point charge, one can show that at ultra-relativistic velocities, radiation is emitted which is confined almost entirely to the plane of the particle's motion, and to a direction tangent to the motion of the particle. In addition, the radiation is highly polarized in the plane of motion. The frequency spectrum of the emitted radiation consists of harmonics of the rotation frequency, ω_0 , but because the bulk of the emitted radiation is at frequencies much greater than the rotation frequency, the spectrum is well approximated by a continuous distribution. The spectrum exhibits a critical frequency, ω_c , above which the radiated intensity falls off exponentially with frequency. Below the critical frequency, the spectrum is quite flat, with a cube-root dependence on frequency. The critical frequency is typically given as $3\omega_0\gamma^3/2$. In practical units (keV) $h\omega_c$ can be

written as $2.22E^3/\rho$ or $0.0665BE^2$, where ρ is the radius of curvature in meters, B is the strength in kG of the magnetic field needed to cause the electron to circulate, and E is the total electron energy in GeV (Winick, 1982).

The situation in the actual electron storage rings (frequently termed synchrotrons after their older cousins) used to generate such radiation is quite a bit more complicated. In these machines a few to several hundred electron bunches, each consisting of around 10^{10} electrons, circulate in a roughly circular vacuum path (the ring) at a frequency of about 1 MHz providing a total current of a few hundred mA. The trajectory and chromaticity of the electrons is controlled by a series of magnets placed around the ring with various numbers of poles for bending and focussing the beam as necessary. The energy of the electrons is maintained by a synchronous microwave frequency electromagnetic field (typically at a few hundred MHz) generated in klystron resonant cavities coupled to the iris-loaded waveguide which contains the beam, *i.e.* the ring (Humphries, 1986).

The ring electrons do not move in a true circle, but rather in a series of straight line segments connected by corners (at bending magnet sites). Significant synchrotron radiation is produced only at these bending magnets, where the electrons are accelerated. In some of the straight sections, special insertion devices which cause local accelerations in the beam (wigglers and undulators) also produce radiation. At such points one or more evacuated metal pipes are used to provide a flight path for the radiation away from the ring housing and out to the experimental stations. These flight paths and associated optical elements are referred to as beamlines, and the final destinations of the radiation, large metal chambers in which experimental equipment can be set up, are called hutches. Access to the hutches is restricted by a safety interlock mechanism to insure that no people are inside when radiation is present. As a result, any experiments which require adjustments during irradiation must be remotely controllable.

Typical beamlines, including those used for SAXS applications, include two major optical components, mirrors and monochromators. Mirrors are used as focussing elements and monochromators are used to select a narrow distribution of photon energies from the broad synchrotron spectrum. Some monochromators are also used as focussing elements. X-ray mirrors rely on total external reflection at grazing incidence angles of the x-ray beam from a surface bent in such a way (elliptically or toroidally) as to provide focussing. They are typically constructed of a fused quartz substrate coated with thin layer of platinum or other element with the proper index of refraction (a lower index of refraction

allows for a larger critical angle of reflection, $\theta_c^2 \sim 2\delta$, where the index of refraction $n \sim 1 - \delta - i\beta$, see Agarwal 1979). Monochromators rely on Bragg's law of diffraction to select a narrow range of photon energies out of the continuous energy distribution emitted by the synchrotron. The actual bandpass of a monochromator is a function of the dynamical diffraction limit (the Darwin width, see Batterman, 1989) and the geometry of the entering beam (angular divergence, *etc.*, see Freund, 1993).

Collimation:

A SAXS camera requires a very tightly collimated x-ray beam in order to avoid contaminating the very weak scattered signal with unscattered photons. The collimation of the beam is the parameter which determines the smallest angles which can be resolved by the camera. The presence of any wings or tails on the spatial distribution of the main beam can greatly compromise the low angle limit of the camera. In principle, because of the high degree of initial collimation, and the presence of focussing elements, it should be possible to use a synchrotron x-ray beam without further collimation. In practice, however, x-ray mirrors do not provide a perfect focus, and beamline optical elements produce a significant amount of background scattering (from surface roughness, scattering, fluorescence, deformation due to heating, *etc.*). Therefore, it is crucial to provide a means for collimating the beam after it arrives in the experimental hutch.

The collimation system most frequently used in synchrotron SAXS cameras is the two slit system, illustrated in Figure 1.1. The main advantages of this system are that it is easily controlled remotely, is relatively easy to align, and is relatively inexpensive. The first set of slits define the size of the main beam. Direct beam photons striking each jaw of the first slit act as a new source of scattered photons. The second set of slits is positioned just outside the primary beam, so that they do not produce further primary beam scattering, and they define the largest angular divergence transmitted through the collimation system. For this system, the theoretical value of the smallest resolvable angle is determined by the size of the slit apertures and the distance between the two sets of slits. In the case of a highly divergent main beam it may not be realistic to position the second set of slits outside the beam. Instead, two slit planes are necessary to reduce the beam divergence to an acceptable value. Some direct beam photons will then strike the jaws of the second slits as well, scattering without further collimation. In this case it is necessary to use a third set of slits to limit the divergence of the photons scattered from the second set. The third set can be positioned just outside direct beam. In the case of a synchrotron source,

however, the beam is nearly perpendicular to the slit planes, and this approach is not generally required.

More sophisticated collimation systems exist which provide better collimation, but in general they are only suitable for cameras with an extremely stable beam. Typically the beam position in a synchrotron fluctuates with the ring parameters, which are often sensitive to outside influences such as insertion device settings, the time of day, or the season, and therefore these collimation systems are not well suited to the synchrotron environment. For a more thorough review of various collimation schemes in SAXS cameras, see Kratky, 1982.

The final component of any collimation system is the beam stop. This is typically a small flat or slightly concave piece of metal placed immediately in front of the detector in order to intercept the direct beam and prevent it from hitting the detector. The beam stop must be big enough to insure that none of the direct beam gets past (or through!) it. In addition, it should not be made out of a material which fluoresces at the intended x-ray energies.

Solution Samples:

Because the research in this dissertation centers around biological samples in solution, it is appropriate to discuss a few of the considerations involved with such samples. Typically, such samples are contained in small volume chambers with interfaces made of a thin x-ray transparent minimally scattering material. Some materials commonly used as window materials are quartz, mica, Kapton, and more recently sapphire. Thicknesses are typically between 10 and 50 microns. The absorption of such windows ranges between about 10 to about 50 % of the incident and scattered intensity. Because biological samples are typically suspended in water, the optimum thickness, t , (along the beam) of such samples is determined by the absorption properties of water. The total scattering cross-section is proportional to the amount of material in the beam and hence to t . Absorption, on the other hand, attenuates the incident and scattered radiation by a factor of $e^{-\mu t}$, where μ is the linear absorption coefficient of the sample. The maximum scattered intensity is obtained at the maximum of the function $te^{-\mu t}$, which occurs at $t = 1/\mu$. More formally, any thin section dx of the sample scatters $cIe^{-\mu x}dx$ photons, where I is the intensity incident on the sample, and c is a constant characteristic of the sample. These scattered photons are absorbed by the remaining sample, leading to another factor of $e^{-\mu(t-x)}$, giving a total scattered intensity of $cIe^{-\mu t}dx$. Integrating this quantity over the

sample thickness yields the above result. For water, μ is about 1 mm^{-1} at 8 keV, a typical energy in a SAXS experiment, and hence 1 mm is the widely accepted value for the optimum thickness of most dilute solutions of biological macromolecules.

For samples of optimal thickness, a fraction e^{-1} or about 0.37 of the incident beam intensity is absorbed by the sample volume. For a typical synchrotron flux of 10^{10} photons at 8 keV, $h\nu = 1.3 \times 10^{-15} \text{ J}$, and we have $4.9 \times 10^{-6} \text{ J/S}$ absorbed by the sample. The specific heat of water is $4.19 \text{ J/g}^\circ\text{K}$, so a 1 mm^3 sample would experience a temperature increase of about $1.17 \times 10^{-3} \text{ }^\circ\text{K/s}$. For irradiations longer than a few minutes, significant temperature changes can occur in samples, and it is therefore customary to use some means of temperature control in synchrotron SAXS experiments (see Stuhrmann, 1978, for a similar discussion).

Normalization Measurements:

Because the characteristics of synchrotron x-ray sources vary with time on a scale of picoseconds (due to the pulsed nature of the radiation from the circulating electron bunches) to seconds (due to small variations in the electron orbit) to hours (due to the gradual decay of the current in the storage ring), it is necessary to monitor the intensity of the collimated synchrotron x-ray beam during SAXS data collection in order to later normalize out any fluctuations. Typically, this is done by using an intensity monitor positioned just in front of the sample. The type of monitor most commonly used for this purpose is a type of x-ray detector known as an ionization chamber. An ionization chamber consists of an anode and a cathode plate a distance d apart, kept at a potential difference V and thereby creating a nearly uniform electric field V/d between them. When an x-ray photon passes through this space, which is filled with a gas, it may be (photo) absorbed, ionizing molecules of gas. Because of the electric field the anion and cation travel in opposite directions, creating a (small) ionization current, which can be measured at either the cathode or anode. Because the number of ionization events is proportional to the number of photons traversing the chamber, the measured current can be used to measure the photon flux entering and exiting the chamber. In order to convert the measured current into a photon flux, it is necessary to know the energy lost per ionization and the absorption coefficient of the gas in the chamber. The energy per ionization, E , is usually obtained from empirical tables (for example Curran, 1965), and the absorption coefficient, μ , is available from a number of sources ranging from purely ab-initio calculated values to purely empirical values. The number of photons absorbed

by the gas equals both $I_i(1 - e^{-\mu t})$, $I_f(e^{\mu t} - 1)$, and $I_m/(h\nu/E)$, where I_i is the photon flux incident on the chamber, I_f is the photon flux exiting the chamber, t is the length of the chamber, and I_m is the measured (electric) current. Using these relationships, it is possible to calculate the actual number of photons incident on the SAXS sample.

Typically, however, it is merely necessary to put the SAXS patterns on the same scale relative to other SAXS patterns. In addition, accurate measurement of the small currents generated in an ionization chamber (for example $e^{-\mu t}$ for 10 cm of N_2 at 8 keV is 0.92, meaning about $0.08 * 10^{10}$ photons/second are absorbed; the ionization energy is about 35 eV, meaning about 230 electrons per photon, or $1.8 * 10^{12}$ electrons/second are produced, giving a total current of 30 nA) is difficult. Therefore, it is common to simply measure a number proportional to the ionization chamber current during the course of the experiment, and normalize the data by this value. As long as the settings of the electronics are not changed, the proportionality constant remains the same. At synchrotrons, it has become conventional to convert the ionization chamber current to a voltage (usually using amplifiers produced by Keithley) and then convert this voltage to a (high) frequency signal. During the course of data measurement, the oscillating signal is used to increment a counter, effectively allowing for the integration of the ionization chamber signal over time.

A major disadvantage of ionization chambers is that the gas inside scatters a significant number of photons, increasing the experimental background. Although in the case of pre-sample chambers, much of this background is absorbed by the sample itself, it is still an undesirable feature. In the case of a very stable beam, the ionization chamber may be moved in front of the collimation arm. The collimated intensity will be proportional to the measured intensity. In the (normal) case of some beam instability, however, this method will not work, as the proportionality constant will change with beam position. The polypropylene film method described below is suitable for the measurement of the beam intensity in the evacuated path between the two collimation slits. Because the second set of slits is out of the main beam, a reasonably accurate measurement can be made. In this position, the thickness of the film can be increased somewhat to provide a higher count rate because the scattered photons will be collimated by the second slit.

In order to compare data from samples with different absorption coefficients, it is frequently desirable to normalize SAXS data not by the intensity, I , incident on the sample, but by $Ie^{-\mu t}$, the absorption corrected intensity. For this purpose, it is necessary

to measure the absorption corrected intensity (in general, samples are too complex to calculate μt). This is most commonly done using some form of intensity monitor to measure the intensity of the beam transmitted through the sample during the actual SAXS data collection. Although several methods have been employed, none are completely satisfactory. An ionization chamber placed after the sample usually causes an unacceptable amount of air scattering, and in addition, scattered photons from the sample may strike the anode and cathode plates directly, compromising the measurement accuracy. A fluorescent screen placed on the beam stop has been used, but it has been found that any extraneous structure in the beam stop causes a significant amount of scattering and significantly affects the quality of SAXS data at the smallest angles. A very thin film of polypropylene placed at a 45° angle to the beam in order to scatter photons into a scintillation counter has been found somewhat effective, but the count rates attained are quite low and the film must be placed far enough from the sample to limit the angular acceptance of the counter sufficiently to insure that only photons scattered by the film are detected. This is difficult due to the typically large size of the evacuated scatter path far from the sample. All of these options have been attempted in the course of this research, and all have been found wanting.

Detectors:

The number of commercially available position-sensitive x-ray detector technologies has grown quickly over the past few years. The most promising of these for SAXS measurements are gas filled chambers, imaging plates, and CCD (charge-coupled device) detectors. Gas filled chambers are by far the most common of these, and this work was performed almost entirely with such detectors. Therefore only they will be described here. For discussions of imaging plate and CCD detectors, see Ito, 1991 and Gruner, 1989.

As in an ionization chamber, gas filled position-sensitive proportional counter detectors rely on the ionization of gas atoms to produce an electronic signal. A noble gas is used to minimize the occurrence of non-photoabsorption interactions (low level molecular excitations). Unlike in an ionization chamber, however, the electric field in a proportional counter is set much higher, so that the initially produced electrons accelerate to the point where they are energetic enough to ionize other gas atoms, resulting in a multiplication or cascade effect. The initial number of ion pairs is determined, as above, by the ionization energy. The total number of electrons produced is determined by the amplification factor

of the detector (the number of electrons produced per primary ionization electron, typically around 10^4 to 10^5) which is a fixed function of the gas and the electric field. Because UV fluorescence from excited gas atoms can produce (far away) ionization events in the cathode or detector walls, a second polyatomic quench gas is typically included in the detector gas mixture in order to absorb the UV photons and also de-excite the noble gas atoms through collisions. Any electrons emitted by this (local) process are added to the current signal.

In order to create the high electric fields necessary in these detectors, very thin wires are used as anodes. Very near the wires, the electric field becomes very large. One-dimensional detectors use a single wire in the center of the detector. Electrons from a single photon event take the shortest path to reach the anode, where they are available as a current signal on the wire, with an amplitude that is proportional to the energy of the absorbed photon. In addition to the anode, current detectors employ a series of cathode strips perpendicular to the anode wire. These strips are connected to an inductance capacitance delay line (a coiled wire periodically connected to ground by capacitors), providing a capacitive coupling between the anode and cathode. The arrival of charge on the anode wire causes a buildup of charge on the cathode opposite the event location. This charge then flows in both directions along the delay line, and is detected at either end. The difference in the arrival times of the two signals at either end of the delay line allow the position of the event to be determined.

An important limitation on the dynamic range of these detectors is caused by the space charge effect. The time required for the positive ions generated during an avalanche to drift to the detector cathode is much longer than that taken by the electrons to reach the anode. If the local rate of events is high enough, positive charges will build up around the anode, decreasing the electric field seen by new electrons, and preventing further signals from reaching the anode. This is observed as a local saturation of the irradiated detector location. Because the drift time for the positive ions is on the order of microseconds, this limitation becomes important at **local** count rates greater than around 10^4 counts/second. In addition, at very high total count rates, the possibility of coincident events exists. Typical delay line lengths are less than $1 \mu\text{s}$, so this effect only becomes important at count rates of around 10^6 counts/second. Good discussions of position-sensitive proportional counters are given in Gursky, 1975 and Russel, 1991.

Data Acquisition:

At a synchrotron, the collection of actual SAXS data must be preceded by a calibration of the beam energy. This is typically performed by placing in the beam a filter made of an element with a known absorption edge in the vicinity of the target energy. Using the monochromator, the transmitted intensity of the beam through the filter is measured as a function of energy. The location of the absorption edge is then used to calibrate the measured spectrum and to position the monochromator at the desired energy.

During exposure of a SAXS sample to the x-ray beam, scattered photons propagate through the evacuated scattering path and reach the detector gas chamber, causing a signal at each end of the detector delay line. These signals are amplified, discriminated against noise, and combined to provide a location, or channel number for the event. A counter corresponding to that channel is incremented. At the end of the data collection, the resulting histogram is read by a computer and stored. A plot of the number of counts per channel *versus* channel number results in a SAXS pattern.

In general, there is a fixed time involved in processing each detector event, known as the detector dead-time, τ , during which no other events may be processed. The actual observed rate of events O , is related to the real rate of events, R , by $O = R(1-O\tau)$, so $\tau = 1/O - 1/R$. If $R = R_0 e^{-n\mu t}$, where n is a number of filters of thickness t and absorption coefficient μ , then for large n we have $\ln O = \ln R - n\mu t$. Extrapolation back to $n = 0$ yields a value for R , allowing for the calculation of τ . When experiments are performed at very high count rates, the detector dead time becomes important, and it is necessary to normalize the data by $1-O\tau$. During the exposure of the sample, the counter which measures the incident and/or transmitted beam intensities must also be gated on in order to provide the necessary flux normalization factor. The details of the electronics and software involved in the actual signal processing and collection vary with individual systems. A description of the basic details of the system used at SSRL during the course of these studies has been given elsewhere (Rice, 1991).

Typically, data collection for a sample is immediately preceded or followed by a background data collection. A background for solution samples consists of the exact sample holder used for the sample, filled with the exact solvent except for the presence of the actual sample itself. Because the background subtraction process (described below) is

crucial to the analysis of the data, it is imperative that the background data collection be performed with the utmost care. Poor background data can result from the movement of the sample holder or the beam between sample and background data collection and is one of the major sources of errors in SAXS measurements. It is also customary to subdivide the data collection time into fractional intervals. The data collected during each interval is referred to as a data cycle. The reason for collecting data in cycles is to look for any changes in the data as a function of sample exposure time. Such changes may be indicative of changes in the sample due to radiation damage or other factors (temperature changes, evaporation, window condensation, *etc.*).

Basic Data Analysis:

This section describes procedures common for the analysis of almost all SAXS data. The specific details of further data treatment will be presented in the experimental descriptions of the individual cases. The first step in the data analysis procedure is the normalization of the data. The data is normalized by the integrated transmitted beam intensity where available, or by the integrated incident beam intensity when absorption is not expected to play a significant role. The data should also be normalized for detector dead-time effects if high count rates were used. After normalization, it is necessary to compare the individual data cycles for each data set. If no changes are observed between the different cycles, they may be averaged together.

The next step is the background subtraction. In addition to scattering from the sample, SAXS data usually contains background scattering from the sample holder walls, from air, from vacuum windows, from fluctuations in the solvent density, from beamline components, and from the collimation system. All of this scattered intensity can be considered incoherent with respect to sample scattering, and is thus additive to the sample scattering. In order to extract the sample scattering alone, it is necessary to subtract the background. As mentioned above, the background scattering is measured under conditions identical to those of sample scattering, except for the absence of the scatterer. In the case of concentrated sample solutions, it is necessary to correct the background for the excess solvent present during the background measurement. This can be done by measuring the background without solvent I_{w0} and with solvent I_{ws} . The correct background is then $I_{w0} + f(I_{ws} - I_{w0})$ where f is the volume fraction of the solvent in the sample solution.

After background subtraction, it is customary to recalibrate the data, which at this stage is in the form of intensity *versus* channel number, into intensity *versus* K (the scattering vector magnitude). In order to do this, it is necessary to know θ and λ , the scattering angle and photon wavelength. λ is available from the initial energy calibration. To determine θ , one measures the distance between the detector and the sample, and the distance along the detector to the channel in question. The sample-to-detector distance can be measured physically, using a tape measure, or by measuring data from a sample with diffraction peaks at known values of K . Such a diffraction pattern can also be used to locate the center of the scattering patterns. The patterns themselves may not be entirely symmetric due to background scattering. The distance to any channel along the detector is calculated as the channel number times the distance per channel on the detector. The latter can be measured by placing a precision machined slotted mask in front of the detector during uniform irradiation.

The inherent statistical error in a SAXS measurement is $n^{1/2}$ where n is the number of counts measured at a particular detector channel. This is simply the error when measuring results of any Poisson process. The statistical error in a linear combination of two SAXS data sets will be greater than or equal to the error of the worse data set. Therefore background measurements must be made long enough to insure that the data quality is similar to that of data collected from the actual samples. Formally the difference of two Poisson distributions is not itself a Poisson distribution. For $I - kB$, the standard deviation is $(I + k^2B)^{1/2}$ where k is the ratio of the collection time for I to that for B (Wilson, 1992). The initial error estimate must be propagated through all further manipulations of the data. Unfortunately, it is only a measure of statistical error, and does not include any information about other random and systematic errors. The process of determining systematic errors is difficult and no foolproof approach exists. Frequently, the best estimates rely on the experience and intuition of the experienced experimenter.

The measured SAXS pattern is a function of the ideal scattering cross-section and several experimental variables. The most important of these are the finite spatial and spectral extent of the beam and the spatial extent of the detector. The beam-related effects can be accurately treated by considering the measured data as the convolution of the ideal data with the spatial and spectral distribution functions of the beam. Similarly the spatial extent of the detector results in a convolution of the ideal data with the detector aperture. In principle, it is necessary to deconvolve the measured data with these functions (which must be measured) in order to extract the actual scattering curve. In practice, these effects

are usually quite small, at most a few percent. Under the conditions used in the following experiments analysis of these effects showed them in most cases to be small compared with other experimental errors, and the data were not corrected. A good description of various data correction algorithms is given in Glatter, 1982.

When concentrated samples are used, the assumption that the scattering from individual solute particles is additive is no longer correct. Instead, interference effects are observed. These effects are most pronounced in the small angle region because of the large distances involved. Empirically, the most important effect is a decrease of the intensity scattered into the central Gaussian SAXS region, which results in shoulders or even peaks in the small angle region. A fully consistent theoretical treatment of interparticle interference effects has yet to be given (Svergun, 1987). Although the effects contain information about the interparticle distribution function and about interparticle forces, they lead to deviations from Guinier's law, and in general it is desirable to eliminate them. The common approach is to measure data at several different solute concentrations and extrapolate either the data or the extracted parameters to zero concentration. For typical protein samples, significant effects can be observed for concentrations as low as 5 mg/ml.

Although data analysis can proceed in many ways from here, the simple approach is to represent the normalized, background subtracted, calibrated data in a Guinier Plot of $\ln I$ versus K^2 . Because of the Gaussian form of the small angle data, the Guinier plot should exhibit a linear region at the smallest angles. The slope of this region should be just $-R_g^2/3$, and a simple linear regression can be used to extract R_g from the data. In addition, the extrapolation of the linear region to zero angle yields the forward scattered intensity, I_0 . This simple analysis technique is the basis for most of the results in this dissertation.

In this work, errors in the results of Guinier analyses are obtained from the unweighted linear least-squares fits to the Guinier plots. Weighted fits (propagating statistical errors) generally give the same result, but with smaller error estimates. Because other random and systematic experimental errors dominate in almost all cases, neither of these error estimates is particularly reliable for a single measurement. In most of the experiments in this work, many measurements are made to monitor the change in state of a sample as a function of small increments of some parameter (time, urea concentration, *etc.*). In such cases, nearby points are always expected to have similar values (all the processes are smooth) and the best indication of the error in the parameters is the magnitude of the

spread in the resulting plots. Consequently, data plots of this nature do not include error bars, unless the calculated errors were greater than those evident from the point spread. For values in the text and tables, the greater of the standard deviation of the mean of multiple measurements and the propagated statistical deviation of the average of these measurements is reported.

It is also important to remember that there is an inherent uncertainty involved with this method of analysis, because there is no way to determine a priori to what extent the data deviate from the Guinier approximation. On purely theoretical considerations, it is best to use the smallest angles possible. Experimental considerations, however, limit the smallest available angle, and it is frequently difficult to judge the quality of data gathered at the lowest angles due to errors introduced by parasitic scattering. In addition, it is necessary to have a reasonable number of data points in order to minimize the effect of noise on the fitting procedure. As a result, choosing the fitting range for analysis is somewhat of an art. Wherever fit results are presented in this work, the fitting range is also given. Invariably, using a lower angular range gives somewhat higher values for both R_g and I_0 , and vice-versa. The change in the values obtainable by going to slightly lower or slightly higher angle regions has been found in the worst cases to be on the order of 25% and 10% of the quoted value of R_g and I_0 , respectively. Because the purpose of most of the studies herein is to measure relative values (i.e ratios) of R_g and I_0 , this uncertainty is somewhat lessened by using the same fitting range for all of the data from any particular sample.

References:

- Agarwal, B.K. 1979. X-ray Spectroscopy, an Introduction. Springer-Verlag, Berlin. pp. 418.
- Batterman, B.W. 1989. in Applications of Synchrotron Radiation. Gordon and Breach, New York. pp. 535-577.
- Curran, S.C. and Wilson, H.W. 1965. in Alpha-, Beta-, and Gamma-Ray Spectroscopy (Siegbahn, K. ed.). North-Holland, Amsterdam. pp. 306.
- Freund, A. 1993. in Neutron and Synchrotron Radiation for Condensed Matter Studies. Springer-Verlag, Berlin. 1:79-93.
- Glatter, O. and Kratky, O. eds. 1982. Small Angle X-ray Scattering. Academic, New York. pp. 515.
- Gruner, S.M. 1989. Rev. Sci. Instrum. 60:1545-1551.
- Guinier, A. 1939. Ann. Phys. 12:161-237.
- Guinier, A. and Fournet, G. 1955. Small-Angle Scattering of X-rays. Wiley, New York. pp. 268.
- Gursky, H. and Schwartz, D. 1975. in X-ray Astronomy (Giacconi, R. and Gursky, H. eds.). Reidel, Boston. pp. 44-51.
- Humphries, S. 1986. in Principles of Charged Particle Acceleration. Wiley, New York. pp. 573.
- Ito, M. and Amemiya, Y. 1991. Nucl. Instrum. & Meth. A310:369-372.
- Jackson, J.D. 1975. Classical Electrodynamics. Wiley, New York. pp. 848.
- Kratky, O. 1982. in Small Angle X-ray Scattering (Glatter, O. and Kratky, O. eds.). Academic Press, London. pp. 53-83.

- Landau, L. and Lifshitz, E.M. 1987. *The Classical Theory of Fields*. Pergamon, Oxford. pp. 402.
- McMaster, W.H., Del Grande, N.K., Mallett, J.H., and Hubbell, J.H. *Compilation of X-ray Cross Sections*. U.S. Dept. of Commerce (NTIS), Springfield, VA. pp. 350.
- Rice, M. and Wakatsuki, S. 1991. *A User's Guide to the Small-Angle X-ray Scattering/Diffraction Data Acquisition System at SSRL*.
- Russel, T.P. 1991. in *Handbook on Synchrotron Radiation* (Brown, G.S. and Moncton, D.E. eds.). North-Holland, Amsterdam. 1:379-469.
- Stuhrmann, H.B. 1978. *Quarterly Reviews of Biophysics*. 11:71-98.
- Svergun, D.I. and Feigin, L.A. 1987. *Structure Analysis by Small-Angle X-ray and Neutron Scattering*. Plenum, New York. pp. 335.
- Van Hove, L. 1954. *Phys. Rev.* 95:249-262.
- Watson, H.C. 1965. in *Small-Angle X-ray Scattering* (Brumberger, H. ed.). Gordon and Breach, New York. pp. 267-275.
- Wilson, A.J.C. 1992. in *International Tables for Crystallography*. Kluwer, London. pp. 583-584.
- Winick, H. 1982. in *Synchrotron Radiation Research* (Winick, H. and Doniach, S. eds.). Plenum, New York. pp. 11-25.

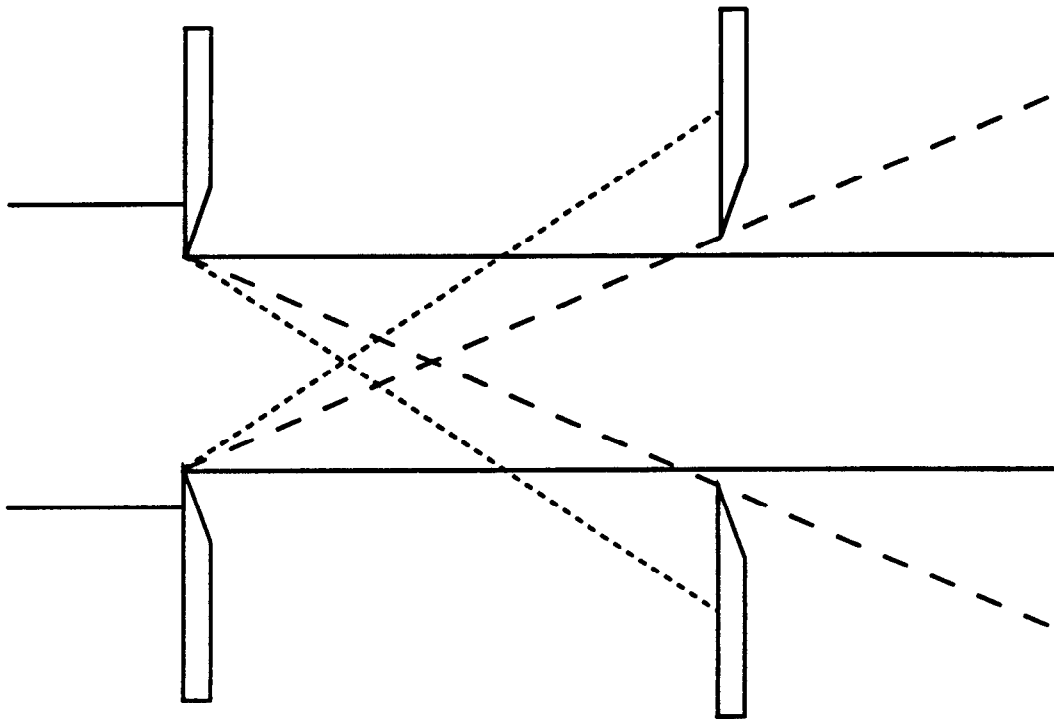


Figure 1.1: Diagram of a typical two-slit collimation system for a synchrotron SAXS camera. The smallest angle, measured from halfway between the slits, which is outside the area of parasitic scattering is $\alpha = \arctan(h/l)$ where $h-\Delta h$ is the first slit aperture size, h is the second slit aperture size (which must be slightly larger than the size of the direct beam at that point) and l is the distance between the two slits. For a sample just beyond the second slits and a sample to detector distance d , this corresponds to a scattering angle $\theta = \arctan(h/l + h/d)$. For $h = 1$ mm and $l = 1000$ mm (ignoring Δh), $\alpha = \arctan(.001) \approx 1$ mrad. For $d = l = 1000$ mm, $\theta \approx 2$ mrad. In reality, there is always some beam divergence, and Δh can't be completely ignored.

Chapter 2

The Iron-Molybdenum Cofactor from *Azotobacter vinelandii* Nitrogenase

This chapter describes SAXS studies of an important protein metallocluster, the iron-molybdenum cofactor from nitrogenase. This was a smaller side-project performed during the course of this research, which provided the author with a great deal of experience in SAXS data collection. The work is not directly related to the protein folding studies described later, but does illustrate a useful application of the SAXS technique to the study of a biologically important sample. The results are of significance to the study of nitrogen fixation.

Motivation:

The term "nitrogen fixation" refers to any reaction by which dinitrogen is converted to nitrogen covalently bound to any other atom. The existence of biological nitrogen fixation was first conclusively demonstrated by Hellriegel and Wilfarth in 1888 (Hellriegel, 1888). The nitrogenase enzyme complex, consisting of two proteins (Mortenson, 1965), is responsible for all biological reduction of dinitrogen. This process is the final step in the atmospheric nitrogen cycle, and is thus of critical importance to life on earth.

In addition to its importance as an enzyme on a vital biochemical pathway, nitrogenase also harbors a potential for industrial and agricultural application. The end product of the biological process is ammonia (Zelitch, 1951), an important industrial and agricultural product. Ammonia production worldwide is a multimillion dollar per year enterprise. Currently, commercial production of ammonia is accomplished through the Haber-Bosch process, first commercially implemented around 1913 (Jennings, 1989). This process proceeds by passing a 3:1 mixture of very pure H₂ and N₂ over an iron catalyst at about 450 °C and 200-300 atm (Nielsen, 1968; Jennings, 1989). The high temperatures and severe conditions involved in this process, as well as the scarcity of the necessary raw materials (typically methane from natural gas) make it energetically expensive. In contrast, nitrogenase reduces nitrogen *in-vivo* under conditions which are extremely mild by comparison (*e.g.* standard conditions, 25 °C and 1 atm), from abundant and essentially free raw materials (atmospheric N₂). Therefore, an understanding of the details of biological nitrogen fixation holds far more than a basic scientific attraction.

Literature Review:

The machinery for biological nitrogen fixation is genetically coded by a series of genes termed *nif* (for *nitrogen fixation*) genes. Genes *nif* H, and K and D, specify the main protein components; an α^2 iron protein (molecular weight 60,000 Dalton), known to function as an electron transfer protein, conveying the electrons necessary for the reduction process using the energy from the hydrolysis of ATP, and an $\alpha^2\beta^2$ molybdenum-iron protein (molecular weight 220,000 Dalton) which contains the site wherein the binding and reduction of dinitrogen occur.

In 1977 (Shah, 1977) a polynuclear metal cluster, termed the iron-molybdenum cofactor or FeMoco, was isolated from the iron-molybdenum protein of the nitrogenase complex. FeMoco was shown to be necessary for biological activity of the molybdenum-iron protein. MoFe protein is not able to reduce dinitrogen after the removal of the FeMoco cluster, but regains its activity after stoichiometric addition of FeMoco. Also upon the extraction of FeMoco from a poorly active molybdenum-iron protein mutant and its subsequent addition to normal but FeMoco deficient molybdenum-iron protein, it is found that the new reconstituted protein is poorly active, in the same manner as the original defective mutant (Hawkes, 1984). Also, single point mutants of the molybdenum-iron protein which alter the physical environment of FeMoco also alter substrate reduction properties (Scott, 1990). Finally, isolated FeMoco exhibits electrochemical behavior analogous to that observed within the native enzyme, including the three redox states, FeMoco(ox) (oxidized, S=0), FeMoco(s-r) (semi-reduced, S=3/2, EPR active), and FeMoco(red) (fully reduced, S=integer, substrate reducing state) separated by one electron each (Schultz, 1985 and 1988). These findings provided a firm basis for the hypothesis that the FeMoco site in the native MoFe protein is in fact the active site of the enzyme complex (Hawkes, 1984; Scott, 1990).

Despite extensive efforts to crystallize isolated FeMoco, to date no diffracting FeMoco crystals have ever been documented. As a result structural information about the nitrogenase active site was long confined to what could be learned from various forms of spectroscopy, most notably x-ray absorption spectroscopy (XAS) and electron paramagnetic resonance (EPR) spectroscopy.

In 1978, Hodgson and coworkers (Cramer, 1978) used the then new extended x-ray absorption fine structure (EXAFS) technique to demonstrate that the chemical

environment of the molybdenum in FeMoco was the same both in its protein bound form and in its isolated form. From further EXAFS data (Conradson, 1987), a first shell of three sulfurs at 2.37 Å, and a second shell of three irons at 2.68 Å were predicted around the lone molybdenum atom. A series of molybdenum-iron-sulfur and iron-sulfur model compounds were synthesized as possible models for FeMoco (Holm, 1986), but all of them were shown to possess different spectroscopic characteristics. Nevertheless, the crystal structures for these various compounds provided interesting insights into the chemistry of molybdenum-iron-sulfur and iron-sulfur compounds.

The biologically unique $S = 3/2$ EPR spectra of isolated FeMoco and FeMoco in its molybdenum-iron protein matrix are qualitatively similar, but the spectrum from the bound cofactor is much sharper than the isolated cofactor spectrum (Rawlings, 1978). Upon addition of benzenethiol to isolated FeMoco in NMF, the EPR signal is observed to sharpen and more closely approach the signal of bound FeMoco. The transition is complete upon the addition of one thiol per molybdenum, suggesting a single thiolate binding site, and the possibility of cysteine ligation when bound in the protein. The effect of cysteine-275 mutations on the molybdenum protein EPR spectra suggest it as a possible candidate (Scott, 1990). Such effects are also observed for histidine-195, and in addition, electron spin echo envelope modulation data show that a signal characteristic of nitrogen coordination is present for the molybdenum-iron protein (Thomann, 1987). Upon substitution of histidine-195, this characteristic signal disappears and enzyme activity is lost (Thomann, 1991). Evidence for nitrogen ligation is also provided by electron nuclear double resonance data (True, 1990).

Most recently, the structure of the molybdenum-iron protein from *Azotobacter vinelandii* was solved by Kim and Rees (Kim, 1992). The structure of FeMoco in its protein matrix was further clarified in following studies (Chan, 1993). The structure which emerged is illustrated in Figure 2.1 and consists of an iron-sulfur cubane and a molybdenum-iron-sulfur cubane, each missing a corner sulfur, linked together by three iron-sulfur-iron bridges. Cysteine-275 provides a ligand to the iron on the cubane corner opposite the molybdenum. Histidine 195 does not appear to be directly ligated to the cluster, although it is within a 5 Å distance.

Finally, it is worth noting that evidence exists for solution heterogeneity of FeMoco. Newton and coworkers (Newton, 1989) observed evidence from EPR and electrochemical studies that FeMoco(ox) appears to exist in three different forms, each corresponding to a

distinct FeMoco(s-r) form. The presence of the different forms is dependent on sample preparation and solvent conditions. Frank (1989) observed batch-dependent differences in the UV absorption spectrum of FeMoco, as well as history-dependent differences in the same sample (after various manipulations such as freezing, storage, and concentration).

Sample Preparation:

FeMoco:

Preparation of the molybdenum-iron protein and subsequent FeMoco isolation was carried out by Dr. William E. Newton as described by Burgess (1980) and Hedman (1988). A brief synopsis of the procedures follows.

Large scale cultures of *Azotobacter vinelandii* are grown in a large aerated fermentor, harvested using centrifugation and stored frozen at -80 °C. Crude extracts are prepared by washing with Tris-HCl to remove external contaminants followed by rupturing in a homogenizer.

All further steps are performed anaerobically using Schlenk techniques. All reagents used are rendered oxygen-free by standard Schlenk methods. *N*-methylformamide is purified by stirring overnight over sodium bicarbonate, followed by vacuum distillation from solid barium oxide, and is further degassed under vacuum and stored under a purified dinitrogen atmosphere.

The lysed cells are centrifuged to obtain the supernatant extract. Incubation with ribonuclease and deoxyribonuclease and brief heating of the extract followed by ultracentrifugation serve to remove contaminating proteins and nucleic acids from the supernatant. The molybdenum-iron protein fraction of the preparation is then separated from the iron protein fraction on a DEAE-cellulose column (monitored by absorbance at 405 nm), diluted, and repurified on a similar column. The peak fraction is concentrated by ultrafiltration. Dilution to the proper salt concentration is followed by crystallization. The crystals are washed and resuspended and the solution is frozen by allowing small drops to fall directly into liquid nitrogen.

FeMoco isolation proceeds with acid treatment of diluted protein solution, followed by neutralization and gentle centrifugation. The resulting pellet is washed in

N,N-dimethylformamide, vortexed, and centrifuged again. This resulting pellet is vortexed vigorously with *N*-methylformamide solution. Another centrifugation step results in a dark brown supernatant containing FeMoco. Concentration of the dilute solution is performed by high vacuum distillation using a large surface area trap while keeping the FeMoco solution in a room temperature water bath. The final FeMoco purification step, consisting of anaerobic gel filtration on a Sephadex G-25 column, followed again by concentration under high vacuum, was performed here at Stanford by Dr. Patrick Frank.

For the final SAXS sample preparation, all procedures involving FeMoco were performed in a Vacuum-Atmospheres glove box operating under a dinitrogen atmosphere of 1 part/million dioxygen or less. FeMoco(ox) samples were used as obtained from the final concentration step. Final sample concentrations ranged from 1 to 3 mM FeMoco (see below) and were verified directly by iron analyses in triplicate (except where noted) involving exposure of the analytical solution to air in the presence of excess *o*-phenanthroline and hydroxylamine·HCl, followed by absorbance measurements using $\epsilon(512) = 11,400 \text{ M}^{-1}\text{cm}^{-1}$ for the $\text{Fe}(\textit{o}\text{-phenanthroline})_3^{2+}$ complex in *N*-methylformamide. Concentrations were accurate to about 5%. FeMoco(s-r) samples were prepared using a 5 x stoichiometric excess of sodium dithionite, added as a 0.2 M solution in 1 mM NaOH. Benzenethiol, added as needed in 5 x stoichiometric excess, was used either as a 1% (97 mM) solution, or as a 250 mM solution, in *N*-methylformamide. The final samples were frozen under dry ice in crimp-top or screw-cap v-bottom vials under a silicon/Teflon septum, covered with several layers of Saran Wrap, and stored at -60 °C until use (several days later).

Model Compounds:

The details of the synthesis of the model compound studied are beyond the scope of this work. The inorganic complexes $((\text{C}_2\text{H}_5)_3\text{NCH}_2\text{C}_6\text{H}_5)_3[\text{Mo}_2\text{Fe}_6\text{S}_8(\text{SC}_2\text{H}_5)_9]$ and $((\text{CH}_3)_3\text{NCH}_2\text{C}_6\text{H}_5)_3[\text{Mo}_2\text{Fe}_7\text{S}_8(\text{SC}_2\text{H}_5)_{12}]$ were prepared by Dr. Patrick Frank using literature methods (Wolff, 1979 and 1980) and their identity and purity were established by elemental analyses and electronic spectra. The compound $((\text{C}_3\text{H}_7)_4\text{N})_6\text{Na}_4[\text{Fe}_{18}\text{S}_{30}]$ (see You, 1988 for details) was a kind gift of Professor Richard H. Holm of Harvard University, Department of Chemistry. Sample solutions of the molybdenum-containing dicubanes were prepared in *N,N*-dimethylformamide by dilution of a more concentrated stock solution. The sample solutions of the $[\text{Na}_2\text{Fe}_{18}\text{S}_{30}]^{8-}$ complex were prepared in

similar fashion in acetonitrile. Final sample concentrations ranged from 1 to 10 mM (see below). Concentrations were verified by measuring the absorbance of iron-complexed *o*-phenanthroline (using $\epsilon(514) = 11,600 \text{ M}^{-1}\text{cm}^{-1}$ in *N,N*-dimethylformamide, and $\epsilon(508) = 11,998 \text{ M}^{-1}\text{cm}^{-1}$ in acetonitrile) as described above. All model solutions were sealed in 1 ml crimp-top vials and stored frozen in dry ice prior to use.

Sample Integrity:

FeMoco:

Because FeMoco is extremely air-sensitive, it is very important to verify the integrity of FeMoco samples before and after any experimental procedures. In these studies, two procedures were used to evaluate the intactness of FeMoco samples before and after SAXS experiments. The $S = 3/2$ EPR signal is one signature of intact FeMoco, and EPR data collection was one procedure used to verify sample integrity. Assaying for biological activity is the most conclusive method for evaluating FeMoco samples, and this was the other procedure used with the SAXS samples. In brief, the activity assay involves adding increasing amounts of dilute FeMoco solution to bacterial cell extracts containing a FeMoco-deficient molybdenum-iron protein mutant, allowing a short period (typically 30 minutes) for the mutant to incorporate the added FeMoco, transferring the solution into an environment containing purified iron protein, an ATP source, dithionite, and C_2H_2 , (the ingredients necessary for enzyme activity) and measuring the rate of production of C_2H_4 .

All tests of FeMoco samples from these experiments were performed in the laboratory of Dr. William E. Newton. All the FeMoco preparations produced active FeMoco before the SAXS experiments. Immediately following SAXS experiments, the FeMoco containing capillaries (see below) were frozen in dry ice. The frozen samples were recovered several days later and prepared for EPR and activity analysis. The samples were stored anaerobically and frozen until the proper analysis could be performed. Unfortunately, both the EPR analysis and the activity verification procedures in Dr. Newton's laboratory experienced inconsistent results during the course of these experiments (both for our samples, and for samples from other experiments). In addition, the possibility of self-oxidation was not considered in many of the samples. Therefore, strict verification of sample integrity is not available for most of the FeMoco samples studied. However, several samples from early experiments, and all the samples from the last series of

experiments were verified to be active. Since the sample handling procedure was essentially the same for all samples, and was strictly adhered to, it is reasonable to assume that all of the FeMoco samples were intact. The good agreement of the SAXS data from different FeMoco samples during each experiment bolsters this assumption further.

Model Compounds:

Assessment of the integrity of model compound samples relied largely on the expertise of Dr. Patrick Frank with the preparation and handling of molybdenum-iron-sulfur model compounds. In all runs except the NSLS run (see below) the model compounds were visually determined to be intact. The color change noted below in the FeMoco samples of this run raised the possibility of sample warming during shipment. Therefore, in this case, sample integrity of the model compound stock solutions was examined both by means of UV/vis spectroscopy and by examining the availability, or unavailability, of cluster iron to capture by *o*-phenanthroline. The spectra of all three model compound stock solutions inspected after the NSLS run showed significant modifications from the expected spectra. All of the samples, however, were resistant to the action of *o*-phenanthroline, indicating that the structural integrity of the various clusters was maintained. Physical degradation of iron-sulfur clusters, and the molybdenum-containing homologues, in previous experience, has invariably produced iron readily available to uptake by *o*-phenanthroline. In the end, several samples which were run (see below) were rejected on the basis of the SAXS data itself, which in these cases were clearly indicative of highly aggregated samples. The R_g s obtained from the scattering profiles of all the remaining model compound samples were fully consistent with those calculated from the known crystal structures (see below).

Experimental Details, Specific:

SAXS data were collected on three different occasions (July 1991, August 1991, and April 1992) at SSRL on beamline 4-2 using the SSRL small angle scattering camera and also on one occasion (December, 1989) on beamline X14 at the National Synchrotron Light Source (NSLS).

In the December 1989 run at NSLS three FeMoco(ox) samples at concentrations of 3.05, 2.29, and 1.53 mM and two FeMoco(ox) + 2x benzenethiol samples at concentrations of 2.87 and 1.53 mM were examined. It should be noted that concentrations for this run

were not determined in triplicate. In triplicate concentration analysis of the FeMoco samples after the run showed somewhat higher concentrations (3.52, 2.84, and 1.92 mM, and 3.58 and 2.11 mM, respectively with standard deviations ranging from 2% to 15%). In addition, four Mo_2Fe_6 samples at concentrations of 6.4, 4.8, 3.2, and 1.6 mM, and four $\text{Fe}_{18}\text{S}_{30}$ samples at concentrations of 10.0, 5.0, 3.0, and 1.5 mM were run. An attempt was made to run some Mo_2Fe_7 samples as well, but they were run at the end of the experiment under a large time-pressure, and the data were found to be inconsistent.

Unfortunately, all the FeMoco samples in this run except one (3.05 mM FeMoco(ox)) were observed to be a light brown color upon arrival at the NSLS. At the time, this was thought to indicate that the samples had been exposed to oxygen, and were compromised. It is possible, however, that the samples had simply (and reversibly) self-oxidized, because FeMoco solutions are known to self-oxidize even when frozen (Schultz, 1985). This possibility is supported by the observation that a month later, when sodium dithionite was added to unused FeMoco samples left over from this run, they changed color from light brown back to black, the color of intact FeMoco solutions.

The SSRL small angle scattering camera was shipped in separate components to the NSLS and assembled on beamline X14. During the run the NSLS ring was filled to a current of about 200 mA and refilled at about 80 mA. The detector used was a Gabriel quadrant detector. The number of channels per mm was measured to be 4.64. Data were collected at a photon energy of 6912 eV. The sample to detector distance was 48.2 cm, as determined from a cholesteryl myristate powder diffraction pattern. The active area of the quadrant detector extends 15 cm from the detector center, and thus data were collected up to a K of 0.167 \AA^{-1} (where K is the magnitude of the scattering vector, as in Chapter 1). Photons scattered from the sample reached the detector through a helium-filled light-tight plexiglass drift-chamber. Incident beam intensity was measured using an ionization chamber, and transmitted beam intensity was measured using a phosphor-coated beam stop with a fiber optic cable to carry the emitted light to a photo-multiplier tube detector. Data were collected in one minute cycles with the number of cycles ranging from 15 to 60.

In all the runs 1 mm outer diameter quartz capillaries (Charles Supper Company, Natick MA) were used to contain the FeMoco and model compound samples. In this run, each capillary was first filled with about 20 μl of FeMoco solution in a nitrogen atmosphere glove box and sealed with fast drying epoxy. The capillary was then mounted in a

capillary holder and aligned in the beam. The position of the capillary in the sample holder was precisely marked using an indelible ink marker. After SAXS data collection, the capillary was removed and one end was broken off, using a bead of epoxy to keep cracks from propagating in the quartz. The remaining segment was filled with the appropriate buffer for a background measurement, the capillary was remounted in the holder, and background data were collected. All possible care was taken to remount the capillary in its initial position in order to insure a good background subtraction.

Measuring the sample data before the background data allowed for the pre-filling of subsequent capillaries, saving time during sample changes. However, the difficulty of breaking the capillaries open without cracking them considerably reduced or eliminated the time saved and led to the abandonment of this collection sequence in future experiments. In fact, data sets for the 6.4 mM Mo_2Fe_6 sample, and for the 10.0 mM and 1.5 mM $\text{Fe}_{18}\text{S}_{30}$ were of poor quality and did not scale properly with sample concentration, and were thus rejected as bad data. It is likely that the capillaries for these samples were either cracked, or were remounted in different positions for sample and background data collection.

Each FeMoco, model compound, and background data cycle was normalized by the integrated transmitted beam intensity measured during data collection. No time-dependent effects were observed, so cycles for the same sample were averaged together and the appropriate background data were subtracted. The resultant data were converted into K space.

In the July 1991 run, three FeMoco(ox) samples at concentrations of 1.93, 1.61, and 1.21 mM were run. In addition, three Mo_2Fe_7 samples were run, at concentrations of 6.17, 3.20, and 2.13 mM. In the August 1991 run, the same three FeMoco(ox) samples with 5x benzenethiol added, at final concentrations of 1.86, 1.55, and 1.15 mM, were run. In addition, one Mo_2Fe_7 sample at 6.04 mM and one Mo_2Fe_6 sample at 14.4 mM were run. However, the quality of the data from these models was low. R_g values are reported for these models below but they were not used in the final averages. All FeMoco samples in these two runs except the 1.15 mM sample in August remained black throughout handling and data collection. The remaining sample in August turned a light brown color, similar to that observed in the NSLS samples, prior to data collection. Again, (reversible) self-oxidation is one likely explanation, especially in view of the resultant data (see below).

These two runs used the SSRL small angle scattering camera on beamline 4-2. In the July and August runs SPEAR was typically filled to a current of about 45 (July) to 60 (August) mA and refilled at around 20 (July) to 30 (August) mA. A Gabriel linear detector, borrowed from Dr. Michel Koch of DESY, Hamburg, was used in both these runs. The active area of the detector was 20 cm by 1 cm. The number of channels per mm was measured to be 5.40 and the sample to detector distance (from cholesteryl myristate powder diffraction) was 128 cm in July and 127 cm in August. A photon energy of 7515 eV was used, but two data sets were also measured at 6900 eV, below the iron absorption edge, in order to quantitate the contribution of iron fluorescence to the data collected above the iron edge. The data collected at 6900 eV were indistinguishable from those collected at 7515 eV, demonstrating that iron fluorescence did not contribute significantly to the data. Data were measured out to K of 0.044 \AA^{-1} . For these two runs an adjustable length rectangular aluminum flight path was built in order to provide an evacuated drift-chamber for scattered photons. Unfortunately, either the welded seams or the o-ring grooves were defective, and only a very rough vacuum was achieved (about 1 Torr or 0.0013 atm). This was still a considerable improvement over the helium path used in the previous run however, since helium scattering is not negligible. The incident and transmitted beam intensities were measured by placing a thin film ($6.4 \mu\text{m}$) of polypropylene in the beam, at a 45° angle, before and after the sample. Scattering from each of the films was then measured in a direction normal to the beam using a scintillation counter photo-multiplier tube. The film before the sample was in air, and the film after the sample was in the evacuated drift chamber. Because of the much lower flux observed on beamline 4-2 during these two runs and the smaller active area of the linear detector (as compared to the quadrant detector), data were collected in one half hour or one hour cycles, with the number of cycles ranging from 3 to 12.

In these two runs, the capillaries used to contain the samples were epoxied into brass pins, allowing them to be mounted in a goniometer head. A specially machined brass plate interface, attached to one of the SAXS camera sample positioners, allowed the goniometer head, with the mounted capillary, to be removed, transported into the glove box for sample transfer, and remounted in precisely the same location. This time the capillary was first filled with solvent for the collection of background data. The solvent was then removed, and the capillary was filled with FeMoco or model compound solution (in the glove box) for sample data collection. This eliminated the need for breaking the capillaries, although it prevented the pre-filling of capillaries and thus increased the time necessary for sample changes.

Each FeMoco, model compound and background data cycle was again normalized by the integrated transmitted beam intensity, cycles from the same sample were averaged together, and the appropriate background data were subtracted. The resultant data were calibrated into K space.

In the April 1992 run, seven FeMoco samples were run: one FeMoco(ox) sample at 1.53 mM, two FeMoco(ox) + 5x benzenethiol samples at 1.48 and 1.30 mM, two FeMoco(s-r) samples at 1.51 and 1.30 mM, and two FeMoco(s-r) + 5x benzenethiol samples at 1.46 and 1.30 mM. In addition, two Mo₂Fe₇ samples were run at 5.40 and 1.50 mM. All FeMoco samples in this run remained a dark black color throughout handling and data collection.

The SSRL small angle scattering camera was again used at SSRL on beamline 4-2, with SPEAR filled to about 90 mA and refilled at about 40 mA. A new linear detector, designed by A. Gabriel and built by Biologic (Grenoble, France) was purchased for this run. The active area was again 20 cm by 1 cm, and the number of channels per mm was measured to be 3.54. A long sample to detector distance of 283 cm (as measured from cholesteryl myristate powder diffraction) and a photon energy of 8980 eV were used. Data were measured out to a K of 0.025 Å⁻¹. A new flight path made of commercially manufactured components was assembled for this run and allowed for evacuation down to a pressure of 0.01 Torr. The incident beam intensity was measured using a short ionization chamber, and no measurement of the transmitted beam intensity was made, in order to eliminate post-sample sources of scattering. Data were collected in 3 to 5 cycles of 30 minutes each.

In this run, the goniometer head capillary holder was replaced with yet another holder, which was mounted in place by sliding it into a slot machined in an aluminum piece. The capillary and holder could be removed as one unit, transferred to the glove box for a sample change, and returned to their original position. As usual, each data cycle was normalized, in this case by the integrated incident beam intensity, cycles from the same sample were averaged together, the appropriate background data were subtracted, and the resultant data were calibrated into K space.

Before proceeding to further analysis of the data, it was necessary to correct the NSLS data sets for the larger active area of the quadrant detector. It was originally believed that this could be done adequately by simply dividing the scattered intensity measured at any

detector channel by the length of the detector arc at that channel (this being simply the distance of the channel from the center of the detector multiplied by the angle spanned by the detector, 60°). Further experience with the detector, however, revealed that this procedure was not satisfactory, and large variations were observed from this predicted behavior, depending on the exact settings of the detector electronics. It was found that the most satisfactory way of correcting for the shape of the detector aperture was to collect a scattering pattern using a uniform radiation source (typically a point source placed sufficiently far from the detector) at the actual time (and hence with the same settings of the electronics) of data collection. Data can then be divided by the uniform irradiation pattern to correct for the detector shape.

Unfortunately, no such uniform irradiation pattern was collected during the NSLS run. Another method which can be used to correct for the detector aperture is to compare data collected for the same sample on both a linear and on the quadrant detector. The ratio of the linear detector data to the quadrant detector data should provide the pattern necessary for correcting other quadrant detector data sets. Unfortunately, the only data of high enough quality which was available for this procedure was the FeMoco data itself. Also, because of the greater angular range covered by the quadrant detector, it was necessary to splice together data from the first two SSRL runs and the last SSRL runs in order to obtain the necessary linear detector pattern. This spliced pattern was divided by a FeMoco data set from the NSLS run, and the resultant pattern was multiplied by the remaining NSLS data sets. Although the FeMoco data from the NSLS obviously could no longer be used, the model compound data obtained in this fashion were important in helping to validate the analysis procedure discussed below.

Data Analysis and Results:

As discussed in Chapter 1, solution SAXS data from monodisperse samples is well approximated by a Gaussian shape in the Guinier range, defined by $KR_g < 1$. The width of the Gaussian can be obtained from a straight line fit to a (Guinier) plot of $\log I$ versus K^2 (where I is the scattered intensity), and is inversely proportional to R_g . However, aggregation and, in the case where sample concentrations are high, inter-particle interference effects, can cause data to deviate from a Gaussian shape, and can skew estimates of R_g obtained from Guinier plots. These effects are most pronounced at very small angles, well within the Guinier range.

When the treated FeMoco and model compound SAXS data were represented in Guinier plots, an upturn of the data at the smallest angle regions (from the last SSRL run) was observed, giving evidence of some degree of aggregation in all of the samples, including the model compounds (Figure 2.2). Therefore, a fitting range was chosen so as to be within the Guinier region for an R_g value of around 5 Å (a value which was expected to be reasonable for most of the samples), but as much as possible outside the very small angle regions where the observed aggregation could introduce systematic errors. Only data from the NSLS run and the first two SSRL runs covered this range (0.13 to 0.20 Å⁻¹) of K space. Within this range, the data were observed to be satisfactorily linear, as required by Guinier's theory. Fits were performed using a linear least squares fitting procedure, and representative fits are shown in Figure 2.3. The fit results for each sample run are presented in Table 2.1. Averaged fit results are summarized in Table 2.2. Statistical errors in the data were obtained from the fitting procedure and propagated through the subsequent averaging steps. The resulting error estimates are included in the Tables. No significant dependence of the fit results on the different FeMoco and model compound concentrations used was observed, indicating that inter-particle interference was not significant. The observed aggregation is discussed further below.

The model inorganic clusters $[\text{Mo}_2\text{Fe}_6\text{S}_8(\text{SC}_2\text{H}_5)_9]^{3-}$ and $[\text{Mo}_2\text{Fe}_7\text{S}_8(\text{SC}_2\text{H}_5)_{12}]^{3-}$ were chosen for their similarity to FeMoco in composition, size range, and air sensitivity and because their crystal structures are known. They were in fact originally synthesized as possible models for the structure of FeMoco (Wolff, 1979 and 1980). Values of R_g for known structures can be obtained by direct calculation of the second moment of the electron density from the atomic coordinates. Such a calculation, however, does not take into account the contrast dependence of the scattering. In an attempt to do so, the number of electrons in each atom in the crystal structures of the molybdenum containing model compounds was decreased by 6 (the number of electrons in carbon, the main component of the background solvent). Because the atomic number density of the scattering molecules is probably higher than that of the solvent, and the solvent contains elements other than carbon, this is only an approximation. It was considered useful, however, because it effectively removed the exterior carbon atoms, which must be at very low contrast with the solvent, from the R_g calculation. The R_g values of the 6 Fe and 7 Fe compounds calculated in this fashion are 4.1 and 5.1 Å, respectively. The values of R_g obtained from the data are 4.22 and 4.53 Å, respectively, in reasonable agreement. The model compound $[\text{Na}_2\text{Fe}_{18}\text{S}_{30}]^{8-}$ was chosen for its different shape and size. The R_g value obtained by calculating moments of inertia from the molecular dimensions provided

in the publication of the structure (You, 1988) is 6.4 Å, and the measured experimental value is 6.78 Å, again in reasonable agreement. Table 2.2 lists the calculated and measured R_g s for all the model compounds. The good agreement between the experimental R_g values obtained for the model compounds and those calculated from the respective crystal structures is a good indication that the Guinier analysis used to extract the R_g s is valid, and that the effects of aggregation in these samples do not extend into the fitting range used. It seems likely that the aggregation detected in these samples consists of only trace amounts of large aggregates.

In all instances where FeMoco was studied in the presence of benzenethiol or sodium dithionite or both, the data and R_g values were almost identical to those from samples without any added thiol or dithionite. Therefore it was concluded that at the resolution of the experiments, the R_g of FeMoco was not dependent on the presence of either thiol or dithionite, and the various R_g values were averaged together to yield the value reported in Table 2.2. This final R_g value of 6.96 Å for FeMoco is surprisingly high, considering that its molecular weight, as calculated from its minimal composition, is significantly lower than that of any of the models involved, and that its overall shape in the protein (Kim, 1992) is similar to that of the dicubane model compounds. It is much higher than the value of 3.0 Å calculated from the coordinates taken from the molybdenum-iron protein crystal structure (Kim, 1992). In addition, the FeMoco SAXS data showed significantly stronger aggregation effects than the model compound data (Figure 2.2). These two observations lead to the hypothesis that unlike in the case of the model compounds, the aggregation effects observed in the FeMoco data are not caused merely by trace amounts of large aggregates. Instead it is believed that a distribution of smaller FeMoco oligomers is most likely responsible for the observed R_g and for the large deviation of the data from a Gaussian shape in the smaller angle regions. The following discussion supports this hypothesis.

The intensity scattered by various components in a solution increases as the square of the mass of the scatterer (ignoring the background solvent), but falls off much more sharply (as e^{-n^2}) with angle as the scatterer increases in size (as R_g goes to nR_g). Hence large aggregates in solution samples invariably produce more scattering in the very low angle regions than monomers do. At higher angles, however, the scatter is dominated by smaller components. It is therefore expected that the effect of the larger aggregates in our samples is negligible in the relatively higher angle region used for the Guinier analysis. This is verified by the model compound results, where in spite of the presence of trace

amounts of large aggregates, Guinier analysis yields R_g figures in good agreement with calculations. Qualitatively then, relatively small aggregates are necessary to explain the larger than expected R_g value measured for FeMoco.

In order to pursue this reasoning in a more quantitative manner, a model was applied using the analytical expression for the scattering from homogeneous spheres of specified R_g , first derived by Rayleigh (Rayleigh, 1911).

$$I(K) \sim [(\sin(2\pi KR) - 2\pi KR \cos(2\pi KR))/(2\pi KR)^3]^2$$

where R is the sphere radius. Using this formula, the theoretical scattering pattern for a sample consisting of spheres of any single size can be calculated. This pattern can be scaled so that it intersects the measured FeMoco SAXS data at one point, but does not exceed it at any point. It will then represent the maximum contribution that the presence, in the sample, of homogeneous spheres of that size could make to the measured data. Although the FeMoco aggregates are not likely to be exactly spherical or homogeneous, this model may still be useful for gauging the contribution of aggregates to the data.

Using successively larger spheres, it is straight-forward to estimate the maximum contribution that each particle size could make to the data (Figure 2.4). The contribution to the Guinier fitting region can then be calculated as the integrated intensity from the model normalized by the integrated intensity from the data (Figure 2.5). It is clear that according to this model the contribution from aggregates to the FeMoco data in the fitting region is limited to particles of less than about a $25 \text{ \AA} R_g$. In fact Figure 2.5 suggests that the major contribution to the data comes from particles of a 6 to about $12 \text{ \AA} R_g$. This bears out our expectation that the presence of smaller aggregates is necessary to explain the measured FeMoco R_g value.

Although more complex inverse transformation algorithms do exist for attempting to extract particle size distributions from SAXS data, they typically require shape information, which was not available, and more importantly very high quality data, which again is not available because of the low concentrations of and very weak scattering from the samples. Therefore, such methods were not viable in this study, and were not pursued.

Another parameter which can be derived from SAXS data is the forward scattered intensity (I_0). As noted in Chapter 1, I_0 is proportional to $c(\rho_s - \rho_b)^2 V^2$ (where c is the

concentration of the scattering molecule, V its volume, ρ_s its electron density, and ρ_b is the electron density of the background solvent). The ratio of I_0 for two monodisperse solutes of identical electron densities in identical background solvent equals the square of the ratio of their volumes, and hence masses. Thus measurements of I_0 can be used to determine the approximate mass of one solute relative to another similar solute. It is important to note, however, that I_0 is not independent of the aggregation state of the solute, because I_0 for an n -mer is proportional to $(nV)^2$ whereas I_0 for n monomers is proportional to nV^2 . Because of this fact, using I_0 to obtain a meaningful measure of the mass of solutes which aggregate is difficult.

Although the aggregation observed in the samples makes a precise analysis using I_0 impracticable, the Guinier fits do provide extrapolated values of I_0 . In the case of the model compounds, these values should be representative of the true monomeric I_0 because the data in the fitting range was judged to be reasonably free of contributions from aggregates. Unfortunately, I_0 values for the FeMoco samples represent only some weighted average of the I_0 s of the various oligomers contributing to the scattering in the fitting range. Nevertheless Table 2.3 lists the square root of the ratio of the concentration-normalized I_0 for each of the model compounds to that of FeMoco. This is a measure of the ratio of the apparent masses. Also listed are the value of this ratio that would be expected for monomeric samples. Contrary to the expected values, the apparent mass of the FeMoco samples is observed to be greater than that of the model compounds. This provides further evidence that the FeMoco samples are not monomeric. In fact the apparent mass of FeMoco is about 2.5 to 3 times greater than expected. This in turn supports the conclusion that a significant portion of the FeMoco aggregates are rather small. Note that the measured ratios of the masses of the various models to each other are in good agreement with the expected values. While this I_0 analysis should not be over-interpreted, it does support the conclusions drawn from the previous analysis.

Conclusions:

In summary, the model compound data can be well explained by assuming a monomeric sample with trace amounts of large aggregates. In contrast, the large measured value of R_g for FeMoco cannot be explained in the same way. A simple modeling analysis suggests that a significant presence of small FeMoco oligomers is necessary to explain the data. A unique distribution of particle sizes cannot be obtained from the data. However, it is concluded with some confidence that in *N*-methylformamide solution, FeMoco

largely exists in aggregated states of up to 25 Å in R_g . The average R_g value of the population is approximately 7 Å, a value significantly larger than what would be expected for a monomer. I_0 data are difficult to interpret for aggregate-containing samples, but do nevertheless support these conclusions.

Although the determination of R_g for monomeric FeMoco in solution did not prove feasible in these experiments, the results obtained are interesting for a number of reasons. The aggregation of FeMoco in solution may explain the failure of persistent widespread efforts to produce crystalline FeMoco samples suitable for crystallographic studies. It may also explain the noted slow electron exchange between FeMoco(ox) and FeMoco(s-r) in solution (Burgess, 1990), as well as the apparent existence in solution of multiple FeMoco species with different electrochemical properties (Schultz, 1985; Newton, 1989). The shielding of interior components of FeMoco aggregates against an electrode surface may explain the sluggish or incomplete electroactivity exhibited by FeMoco solutions at carbon electrodes (Schultz, 1985, 1988 and 1990). On the other hand these aggregates must be porous enough to admit small molecules such as benzenethiol, because the stoichiometry of thiol interaction with FeMoco under the above conditions is thought to be 1:1 (Burgess, 1980). It is also possible that the aggregate distribution of FeMoco in solution is not always the same. This may explain the batch- and history-related variation of FeMoco UV/visible spectra (Frank 1989) and the variation with pH of FeMoco EPR spectra (Burgess, 1990; Newton, 1989).

It should be possible in the future to use SAXS to screen FeMoco samples in order to search for conditions under which aggregation does not occur. In fact the studies of different oxidation and thiol-ligation states of FeMoco are the beginning of such a search. However the lack of any observed variation in the SAXS results for these samples suggests that previously observed effects on FeMoco(s-r) EPR spectra due to thiol ligation (Burgess, 1990) are not a result of some modification of aggregation behavior in solution.

There is a considerable interest in the solution chemistry of FeMoco as an inorganic heterometallic cluster. With the emergence of a FeMoco structure within the molybdenum-iron protein, for example, direct laboratory synthesis has become an immediate goal. The conditions necessary to induce catalytic activity in isolated FeMoco are also of great interest. Knowledge of the solution behavior of FeMoco is of vital importance in approaching such issues.

References:

- Burgess, B.K. 1990. *Chem. Rev.* 90:1377-1406.
- Burgess, B.K., Jacobs, D.B., and Stiefel, E.I. 1980. *Biochim. Biophys. Acta.* 614:196-209.
- Chan, M.K., Kim, J., and Rees, D.C. 1993. *Science.* 260:792-794.
- Conradson, S.D., Burgess, B.K., Newton, W.E., Mortenson, L.E., and Hodgson, K.O. 1987. *J. Am. Chem. Soc.* 109:7507-7515.
- Cramer, S.P., Gillum, W.O., Hodgson, K.O., Mortenson, L.E., Stiefel, E.I., Chisnell, J.R., Brill, W.J., and Shah, V.K. 1978. *J. Am. Chem. Soc.* 100:3814-3819.
- Frank, P., Gheller, S.F., Newton, W.E., and Hodgson, K.O. 1989. *Biochem. Biophys. Res. Commun.* 163:746-754.
- Hawkes, T.R., McLean, P.A., and Smith, B.E. 1984. *Biochem. J.* 217:317-321.
- Hedman, B., Frank, P., Gheller, S.F., Roe, A.L., Newton, W.E., and Hodgson, K.O. 1988. *J. Am. Chem. Soc.* 110:3798-3805.
- Hellriegel, H. and Wilfarth, H. 1888. *Beilageheft zu der Zeitschrift des Vereins für die Rübenzuckerindustrie des Deutschen Reiches.* 1-234.
- Holm, R.H. and Simhon, E.D. 1986 in *Molybdenum Enzymes* (Spiro, T.G. ed.) John Wiley & Sons, New York. pp. 1-87.
- Jennings, J.R., and Ward, S.A. 1989. in *Catalyst Handbook* (Twig, M.V. ed.). Wolfe, London. pp. 384-440.
- Kim, J., and Rees, D.C. 1992. *Science.* 257:1677-1682.
- Mortenson, L.E. 1965. in *Non-Heme Iron Proteins: Role in Energy Conversion* (San Pietro, A. ed.). Antioch Press, Yellow Springs, Ohio. pp. 243-259.

- Newton, W.E., Gheller, S.F., Feldman, B.J., Dunham, W.R., and Schultz, F.A. 1989. *J. Biol. Chem.* 264:1924-1927.
- Nielsen, A. 1968. *An Investigation of Promoted Iron Catalysts for the Synthesis of Ammonia.* Jul. Gjellerup Førlag, Copenhagen. pp. 165
- Rawlings, J., Shah, V.K., Chisnell, J.R., Brill, W.J., Zimmerman, R., Münck, E., and Orme-Johnson, W.H. 1978. *J. Biol. Chem.* 253:1001-1004.
- Rayleigh, L. 1911. *Proc. R. Soc. Lond. A*-84:25-46.
- Schultz, F.A., Gheller, S.F., Burgess, B.K., Lough, S., and Newton, W.E. 1985. *J. Am. Chem. Soc.* 107:5364-5368.
- Schultz, F.A., Gheller, S.F., and Newton, W.E. 1988. *Biochem. Biophys. Res. Commun.* 152:629-635.
- Schultz, F.A., Feldman, B.J., Gheller, S.F., and Newton, W.E. 1990. *Inorg. Chim. Acta.* 170:115-122
- Scott, D.J., May, H.D., Newton, W.E., Brigle, K.E., and Dean, D.R. 1990. *Nature.* 343:188-190.
- Shah, V.K. and Brill, W.J. 1977. *Proc. Natl. Acad. Sci. U.S.A.* 74:3249-3253.
- Thomann, H., Morgan, T.V., Jin, H., Burgmayer, S.J.N., Bare, R.E., and Stiefel, E.I. 1987. *J. Am. Chem. Soc.* 109:7913-7914.
- Thomann, H., Bernardo, M., Newton, W.E., and Dean, D.R. 1991. *Proc. Natl. Acad. Sci. USA.* 88:6620-6623.
- True, A.E., McLean, P., Nelson, M.J., Orme-Johnson, W.H., and Hoffman, B.M. 1990. *J. Am. Chem. Soc.* 112:651-657.
- Wolff, T.E., Berg, J.M., Hodgson, K.O., Frankel, R.B., and Holm, R.H. 1979. *J. Am. Chem. Soc.* 101:4140-4150.

Wolff, T.E., Power, P.P., Frankel, R.B., and Holm, R.H. 1980. *J. Am. Chem. Soc.* 102:4694-4703.

You, J.-F., Snyder, B.S., and Holm, R.H. 1988. *J. Am. Chem. Soc.* 110:6589-6591.

Zelitch, I., Rosenblum, E.D., Burris, R.H., and Wilson, P.W. 1951. *J. Biol. Chem.* 191:295-298.

Table 2.1:
Observed Radii of Gyration for FeMoco and Models.

Sample	Concentration (mM)	Observed R_g (Å)	$\sigma(R_g)$ (Å)
FeMoco	1.93	6.51	0.18
	1.61	7.83	0.14
	1.21	6.65	0.22
	1.86	6.12	0.29
	1.55	7.97	0.36
	1.15	6.68	0.66
$[\text{Mo}_2\text{Fe}_7\text{S}_8(\text{SC}_2\text{H}_5)_{12}]^{3-}$	6.17	4.45	0.26
	6.04	4.14	0.24
	3.20	4.72	0.23
	2.13	4.41	0.28
$[\text{Mo}_2\text{Fe}_6\text{S}_8(\text{SC}_2\text{H}_5)_9]^{3-}$	14.4	3.99	0.17
	4.80	4.31	0.41
	3.20	4.04	0.43
	1.60	4.31	0.48
$[\text{Na}_2\text{Fe}_{18}\text{S}_{30}]^{8-}$	5.00	6.79	0.25
	3.00	6.76	0.25

Table 2.2:
Average Observed and Calculated Radii of Gyration.

Sample	Observed $\langle R_g \rangle$ (Å)	$\sigma(\langle R_g \rangle)$ (Å)	Calculated R_g (Å)
FeMoco	6.96 ^a	0.31	3.0 ^d
$[\text{Mo}_2\text{Fe}_7\text{S}_8(\text{SC}_2\text{H}_5)_{12}]^{3-}$	4.53 ^b	0.15	5.1 ^e
$[\text{Mo}_2\text{Fe}_6\text{S}_8(\text{SC}_2\text{H}_5)_9]^{3-}$	4.22 ^b	0.25	4.1 ^e
$[\text{Na}_2\text{Fe}_{18}\text{S}_{30}]^{8-}$	6.78 ^c	0.18	6.4 ^f

^a An average of six samples.

^b An average of three samples.

^c An average of two samples.

^d Calculated using approximate atomic coordinates for FeMoco inside the molybdenum-iron protein obtained from the protein crystal structure determination (Kim, 1992).

^e Calculated using crystal structure atomic coordinates (Wolff, 1979 and 1980).

^f Calculated using approximate dimensions obtained from the crystal structure determination (You, 1988).

**Table 2.3:
Observed and Calculated Mass Ratios.**

Sample	M/M(FeMoco) observed	M/M(FeMoco) expected
FeMoco	$1.00 \pm .05^a$	1.00^d
$[\text{Mo}_2\text{Fe}_7\text{S}_8(\text{SC}_2\text{H}_5)_{12}]^{3-}$	$0.70 \pm .07^b$	1.80^e
$[\text{Mo}_2\text{Fe}_6\text{S}_8(\text{SC}_2\text{H}_5)_9]^{3-}$	$0.54 \pm .04^b$	1.58^e
$[\text{Na}_2\text{Fe}_{18}\text{S}_{30}]^{8-}$	$1.06 \pm .06^c$	2.96^e

^a An average of six samples.

^b An average of three samples.

^c An average of two samples.

^d Calculated from the molecular formula for FeMoco in (Kim, 1992) using molybdenum, iron and sulfur atoms.

^e Calculated from the molecular formula using molybdenum, iron, sodium, and sulfur atoms.

The observed ratio is just the square root of the ratio of the concentration-normalized forward scattering (I_0) of each of the models to that of FeMoco. The errors represent the spread in the measurements. The expected ratio is the ratio of the molecular weight of each of the models to that of FeMoco. Only atoms heavier than oxygen are included, in order to approximate the effect of the background solvent (which consists largely of carbon, nitrogen, and oxygen).

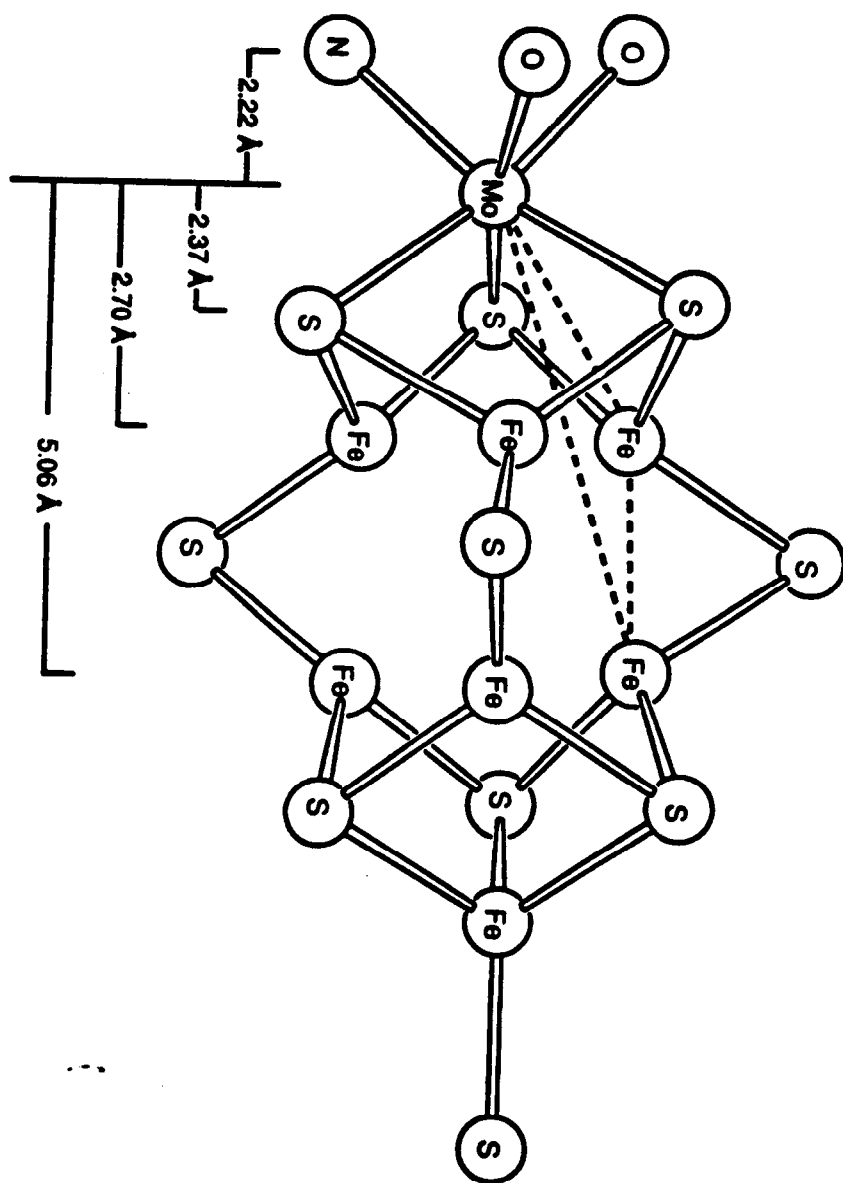


Figure 2.1: The structure of FeMoco (from Kim, 1992, Chan, 1993, and a private communication from D.C. Rees)

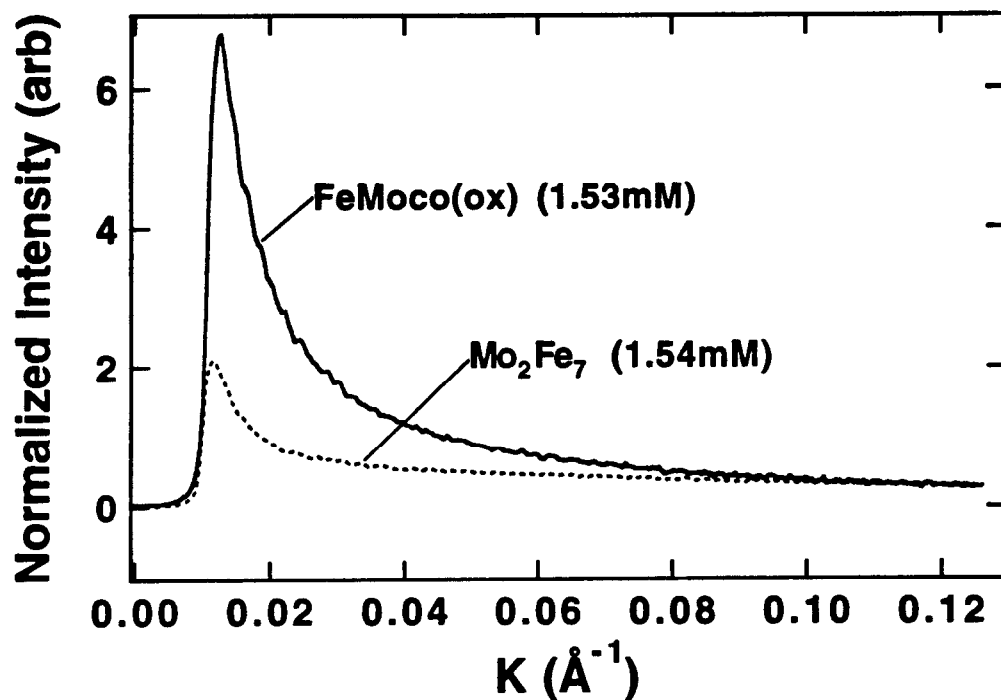


Figure 2.2: Scattering data from equal concentrations of FeMoco and the $[\text{Mo}_2\text{Fe}_7\text{S}_8(\text{SC}_2\text{H}_5)_{12}]^{3-}$ model compound at angles smaller than those used for the Guinier fits. The sharp rise of the scattering at small angles indicates the presence of trace amounts of aggregates. The presence of more and/or larger aggregates in the FeMoco samples is readily apparent from the much stronger scattering at these low angles.

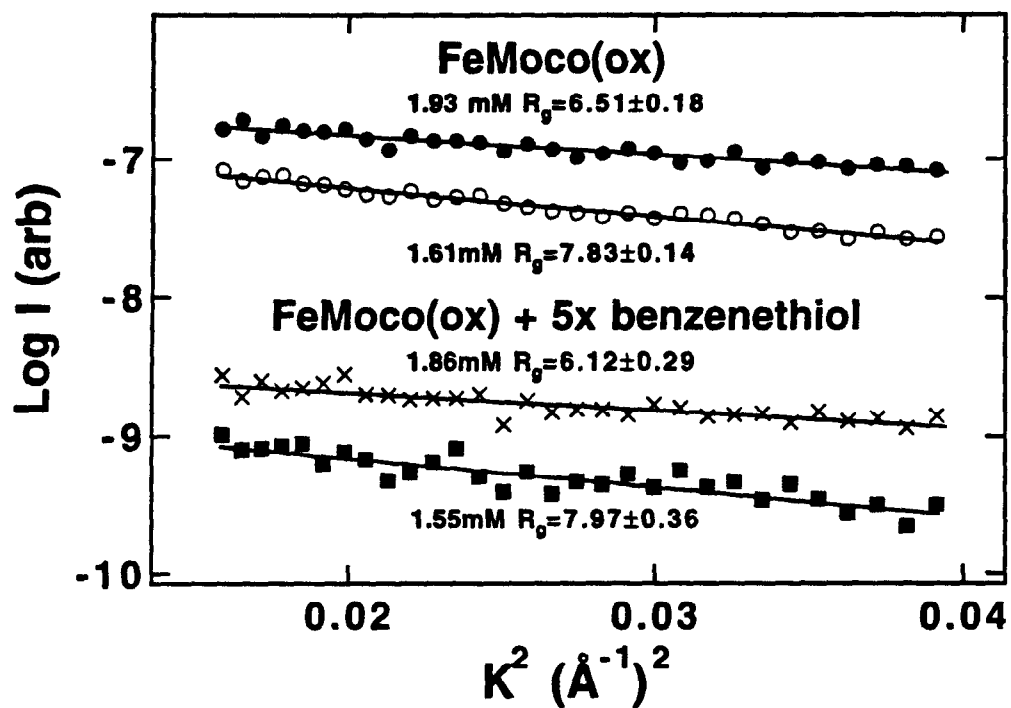


Figure 2.3: Normalized background subtracted SAXS data from four different FeMoco samples represented in Guinier plots ($\log I$ versus K^2) along with the fits from which R_g values were determined. The plots are shifted arbitrarily along the ordinate for clarity of presentation.

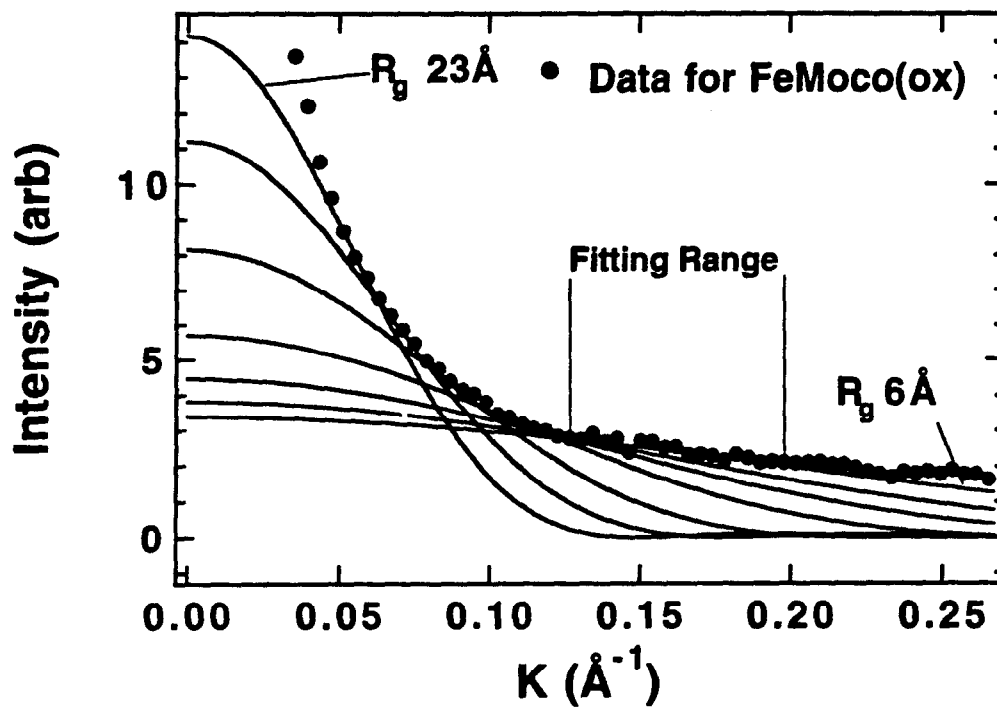


Figure 2.4: Maximum possible contributions from homogeneous spheres of varying R_g to the measured intensity. The scattering calculated for each sphere size may not exceed the measured scattering at any point.

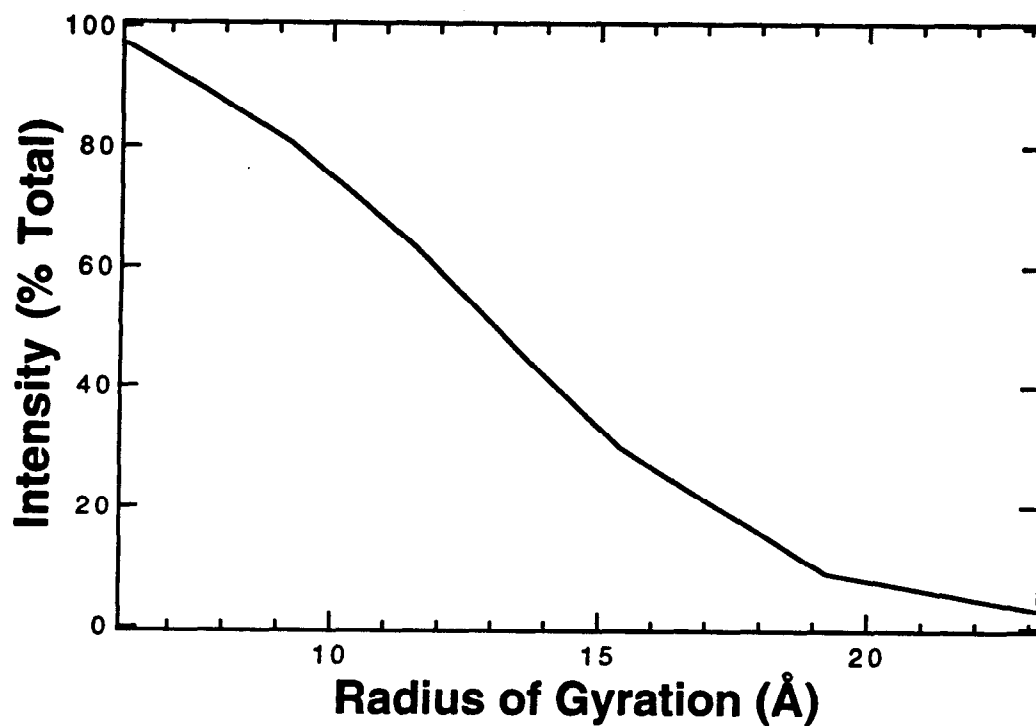


Figure 2.5: Maximum percentage contribution from homogeneous spheres of varying R_g to the FeMoco(ox) scattering data in the range used for the Guinier fits. Calculated as the integral, over the fitting range, of the maximum possible sphere model intensity divided by the integral, over the same range, of the measured intensity.

Chapter 3

Protein Folding

Motivation:

The protein folding problem is one of the most important problems in biophysics today. The problem has been defined in two ways:

1. To implement a homomorphic mapping from protein amino acid sequences to (correct) protein three dimensional structures.
2. To understand in detail nature's implementation of this mapping.

The second formulation, which may or may not subsume the first, is the one addressed in this research.

Because proteins are involved in almost every facet of biological function, and because over the past several decades it has become clear that the physical form or structure of proteins is intimately related to their function, the determination of protein structures has become the focus of a large research effort. Currently, atomic resolution protein structures can be determined only by crystallographic or in favorable cases nuclear magnetic resonance techniques. Both of these avenues require a tremendous expenditure of time and effort per structure determination, and indeed the total number of structures thus determined to date is only about 500 (Chan, 1993). In contrast, protein sequencing (the determination of a protein's amino acid sequence) is a relatively simple process, and tens of thousands of protein sequences are known (Chan, 1993). Therefore, the successful implementation of a mapping from sequence to structure will immediately make available the structures of thousands of proteins which might otherwise not have been known for years or decades.

In addition, because function is tied to structure, a solution to the protein folding problem promises to revolutionize the fields of protein engineering and rational drug design. Once it is possible to predict the effects of various changes in sequence on the resultant structure, it should begin to be possible to increase (or decrease) enzyme efficiencies, to alter substrate specificities, to alter allosteric mechanisms, to design antigen specific

antibodies, and to perform a host of other precise modifications with a reasonably confident prediction of the effects. In some ways, a solution to the protein folding problem offers a pathway to surgery on the molecular level.

Literature Review:

Although observations of 'denaturation', or large changes in the properties of proteins upon changes in temperature or environment, were made early in the century, and were attributed to the likely expansion or contraction of polymer chains, the 'official' origin of the protein folding problem can be traced back to the work of Anfinsen, who in 1973 formulated the hypothesis, based on experiments with ribonuclease A, that all of the information necessary to determine a protein's three-dimensional structure is contained within the protein's amino acid sequence. In these experiments it was observed that ribonuclease A, which contains 4 disulfide bonds, when transferred from an unfolding environment to native conditions, in the absence of any reducing agents, regains enzymatic activity. When transferred from an unfolding environment to native conditions in the presence of a reducing agent, enzymatic activity is again regained upon subsequent removal of the reducing agent. If, however, the reducing agent is removed before transfer into native conditions, only a small fraction of enzymatic activity is recovered. Full enzymatic activity can be regained even in this case if trace amounts of reductant are again added to the sample. The interpretation of these results was that unless incorrect disulfide bonds are allowed to form, ribonuclease A will assume a single, 'correct' fold or topology. In addition, even if the wrong disulfides are allowed to form, any perturbation of these wrong bonds will cause the protein to interconvert to the 'correct' fold, which must therefore be thermodynamically stable.

The question of whether or not a protein's native conformation represents a global or local energy minimum has yet to be definitively answered. Several examples have been discovered in which proteins fold to a form which is not the most stable form, and require 'help' to overcome some kinetic barrier to the final native form (Baker, 1992, and Mottonen, 1992). Such examples illustrate the possibility that the native state itself may be separated by a kinetic barrier from some other state of greater thermodynamic stability. Most researchers, however, believe that the native state energy is in fact the global minimum for a protein.

The (Gibbs) free energy of unfolding (the energy of the unfolded state minus that of the folded state) of many proteins has been measured by various methods, and is typically on the order of only ten kcal/mole (Karplus, 1992). This energy is composed of the entropy gain associated with the greater number of states available to the disordered unfolded state, the entropy loss due to the ordering of solvent molecules around exposed protein side-chains, the enthalpy of breaking protein-protein and forming protein-solvent Van der Waals contacts, the enthalpy of breaking protein-protein and forming protein-solvent hydrogen bonds, and the enthalpy due to electrostatic interactions (see Privalov, 1992 for a review). The sum of these relatively large and poorly understood effects results in the marginal stability of the native state of the protein under physiological conditions. Detailed modeling of the energetics of proteins is therefore easily susceptible to small errors.

Several attempts to model the folding process at varying degrees of detail exist in the literature. Extensive statistical mechanical modelling by Karplus and Shakhnovich and others has led to various postulates regarding the nature of the potential energy surface of protein molecules. In addition, atomic or near atomic level simulations, both exhaustive (using short peptides) and statistical (using lattice models and/or Monte Carlo methods) have been implemented. Although interesting results have been obtained, to date, no one has successfully produced the native fold of a protein starting from an unfolded initial configuration. An exact simulation is very likely an intractable computational problem and may never be solvable on human time scales. However, an exact simulation is probably not necessary, as the detailed atomic motions of the folding process are of little interest. What is necessary is to determine what level of detail is sufficient to account for the uniqueness of the final folded state of proteins. (see Karplus, 1992 for a review of these topics).

The protein folding process is highly cooperative, and data from equilibrium studies where a protein is studied under gradually stronger unfolding conditions typically exhibit a 'two-state transition' in which the property under observation changes rapidly (over a small change in conditions) from that of the initial (folded) state to that of the final (unfolded) state. Indeed, for many purposes the protein folding process can be treated as a first order phase transition. Under most circumstances, partially folded states intermediate between the folded and unfolded state are not populated at equilibrium.

Over the past decade, however, mounting evidence has indicated that the folding process occurs through a series of kinetically accumulated intermediate states (see for example Kuwajima, 1988, Udgaonkar, 1988, and Roder 1988). This evidence has largely been in the form of time-resolved studies in which some property of the protein was observed to change in two or more well separated phases with distinct rate constants during folding. Any information which can be gained about such intermediate states will obviously be of great help in determining the sequence of events on the folding pathway, and thus the characterization of folding intermediates is currently at the center of many experimental efforts. Because such intermediates are very short lived, however, the task is a difficult one.

In parallel with evidence for kinetic folding intermediate states, an exception to the two-state equilibrium behavior has emerged. It has been found that under the proper conditions, many proteins do populate an equilibrium state which is different in many ways from both the native and unfolded state. The general characteristics of such states seem to be a degree of compactness much greater than that of the unfolded state, and only slightly less than that of the native state, the presence of secondary structure, again in excess of that present in the unfolded form yet less than that of the native form, and the almost complete absence of tertiary interactions. Although these properties hold to different extents in the different cases studied, these states have been lumped together under the appellation of 'molten globule state' (see Ptitsyn, 1992 for a recent review).

The emergence of the molten globule as a third stable equilibrium folding state was accompanied by experiments which supported a three-state model (as opposed to the previous two-state model) of equilibrium unfolding. The question was raised as to whether or not this third equilibrium state was at all relevant to kinetic folding pathways. Some postulate that it is merely a curiosity, typically observed under conditions far from those *in-vivo*, and therefore an irrelevant artifact. Others argue that the molten globule must also be a kinetic folding intermediate, and that as such it should be studied in detail, since it provides the only opportunity to characterize such an intermediate over long time scales (Ptitsyn, 1987 and 1990).

In this chapter experiments are described which attempt to determine the size (through the radius of gyration) as a function of folding time for three small single-domain globular proteins in order to search for, and characterize the size of, any distinguishable kinetic folding intermediates. Small single-domain globular proteins can be expected to exhibit

the simplest folding behavior, and are thus frequently the subjects of folding studies. The proteins studied here are bovine pancreatic ribonuclease A, horse skeletal muscle myoglobin, and horse heart cytochrome-*c*. Ribonuclease A is a digestive enzyme belonging to the class of nucleases. Its function is to catalyze the cleavage of bonds between sugar groups in RNA. Myoglobin is an oxygen binding heme protein which supplies oxygen to muscle tissue. Cytochrome-*c* is an electron carrier which transports the electrons necessary for the final reduction of O₂ during oxidative phosphorylation in cell mitochondria.

3.A: Time-Resolved Small Angle X-ray Scattering:

Although useful experiments investigating the folding of proteins can be performed under equilibrium conditions, the most interesting experiments are those which observe the dynamic folding process. This involves the collection of time-resolved data, and in order to collect such data, modifications were needed to the basic SAXS data acquisition method described in the introduction. This section describes the modifications which were used for experiments discussed in the remainder of this chapter.

Triggering by the Stopped-Flow Method:

In order to collect time-resolved data (using any probe) from a process, it is necessary to synchronize the collection of the data with some point in the process. Typically, the start of the process is chosen as the point of synchronization. As long as the time between the initiation of data collection and the start of the process is known, however, any point can be chosen. In the case of protein folding, several methods have been employed by investigators to initiate or trigger the folding of protein samples. These all involve rapidly changing the protein environment from one in which the unfolded state is energetically favored to one in which the folded state is favored. The change can be in any of the thermodynamic parameters of system, including the temperature, pressure, and chemical potential.

The stopped-flow method, used in these studies, relies on the rapid mixing of fixed volumes of solutions to cause a rapid change in the chemical environment (and the associated chemical potential) seen by the unfolded protein molecules. The observation of events at the early stages of folding by this method is inherently limited by the minimum time required for the complete and homogeneous mixing of the sample

solutions, known as the mixing dead-time.

In a closely related technique, the continuous-flow method, the solutions are mixed together continuously and transported from the mixing location at a constant velocity, so that observation of the mixed solution at any fixed point in space corresponds to a fixed point in the time elapsed since mixing. This technique, however, typically requires even larger amounts of sample than the stopped-flow technique, and was not employed here.

Stopped-Flow Rapid-Mixer Description:

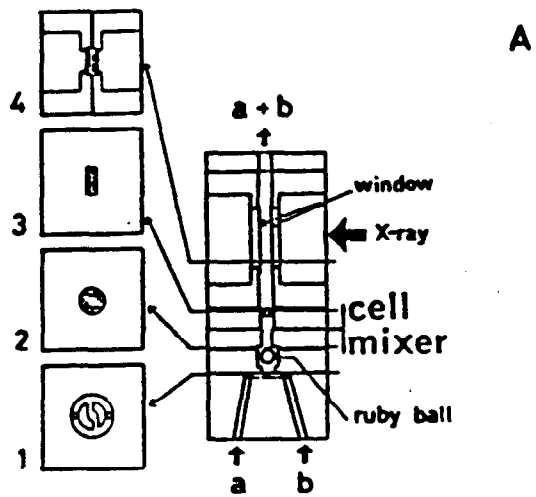
The stopped-flow rapid-mixers used in these experiments were built by T. Nagamura and Unisoku Corporation (Osaka, Japan). A schematic of the stop-cock mixer is presented in Figure 3.A.1 (from Tsuruta, 1990, with permission) and a detailed description of the apparatus has been published (Tsuruta, 1989). A rotating ceramic cock is used to alternately connect two syringe chambers to two sample reservoir chambers or to a two stage mixing chamber and observation cell. Mixing ratios of approximately 1:1, 1:3.5, and 1:6 in volume are available by using syringes of different diameters. Each syringe body is made of glass, with a rubber O-ring at the end of each plunger to provide a seal. The syringe plungers are retracted and pushed using pressurized gas controlled by a pneumatic valve. The mixing chamber consists of a two-jet stage followed by a turbulent stir ball-mixing stage. The observation cell is machined from ceramic, and the 3.5 mm wide by 2.5 mm high by 1 mm thick volume is sealed on either side by an epoxied 50- μ m thick quartz-window. All mixer components are mounted in an aluminum block which is thermostated using a Peltier device, in turn cooled by a circulating bath. The stop-cock mixer was used in all time-resolved experiments at the Photon Factory.

A similar mixer, the ruby-ball mixer, was used for time-resolved experiments at SSRL. The main difference between this mixer and the above description is that the stop cock is replaced by a set of one-way valve chambers, which allow in-flow only from the mixer reservoirs, and allow out-flow only to the mixing chamber. This one-way valve action is implemented by a small ruby ball trapped in a small chamber with two openings. One of the openings is constructed so that the ruby ball seals it when pressed against it by an incipient flow. Flow through the other opening is unrestricted.

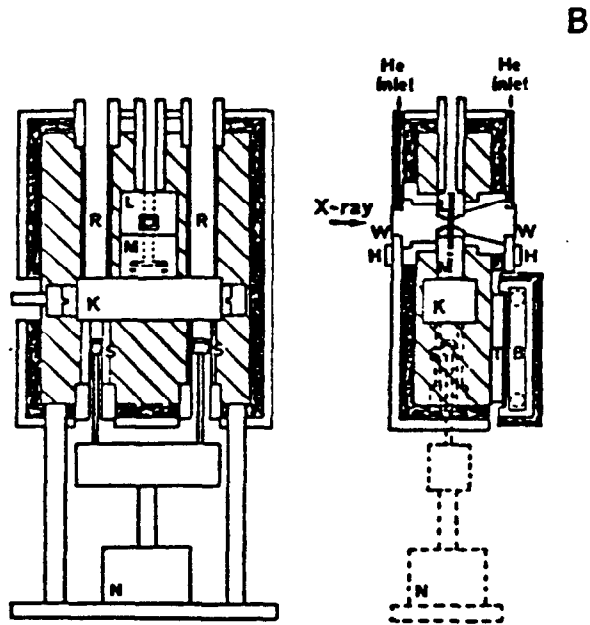
In the course of this research, it was necessary to integrate the ruby-ball mixer into the SSRL SAXS camera. This involved interfacing the control unit supplied by the

manufacturer with the SSRL camera control hardware and software. The necessary software modifications were easy due to the general nature of the existing data acquisition system. It was simply necessary to send the appropriate CAMAC commands to start each data collection cycle by giving a push command to the mixer, to increment a histogramming memory frame pointer for each data frame, to read out the integrated beam intensity after each data frame, and to read the final data at the end of the cycle and issue a pull command to the mixer to reload the syringes for the next cycle. One complication, however, was the need to account for the CPU processing time of the computer used to run the data acquisition software. This was accomplished by using an oscilloscope to monitor the triggering signals and insure that they occurred at the correct frequency.

In order to allow the CAMAC electronic modules to communicate with the controller supplied with the mixer, it was necessary to build a simple signal processing circuit. The circuit uses a monostable multivibrator IC chip (a 'one-shot') to convert a short (32 ns) NIM pulse generated by a software driven CAMAC RTC-014 clock module (by Ortec) into a longer (30 ms) pulse used to drive an electromechanical relay switch which generates the button push (*i.e.* contact) needed to make the mixer pull or push the syringes. Although a more elegant system can easily be (and should be) designed, a basic understanding of this circuit is necessary in order to operate the current system, and therefore a schematic is included in Figure 3.A.2. The circuit was designed with the help of and subsequently constructed by Alan Swithenbank of SSRL.



A



B

Figure 3.A.1: Schematic of the stop-cock mixer used in the time-resolved protein folding experiments. The turning stop-cock (K) alternately connects the syringe chambers (S) to the sample reservoirs (R) or to the mixing cell, which consists of a two-jet stage (1) followed by a turbulent stir stage (2) and finally the 50 μm quartz-window observation chamber. The unit is cooled by a Peltier device (T) in turn cooled by a circulating bath flow (B). Figure taken (with permission) from Tsuruta, 1990.

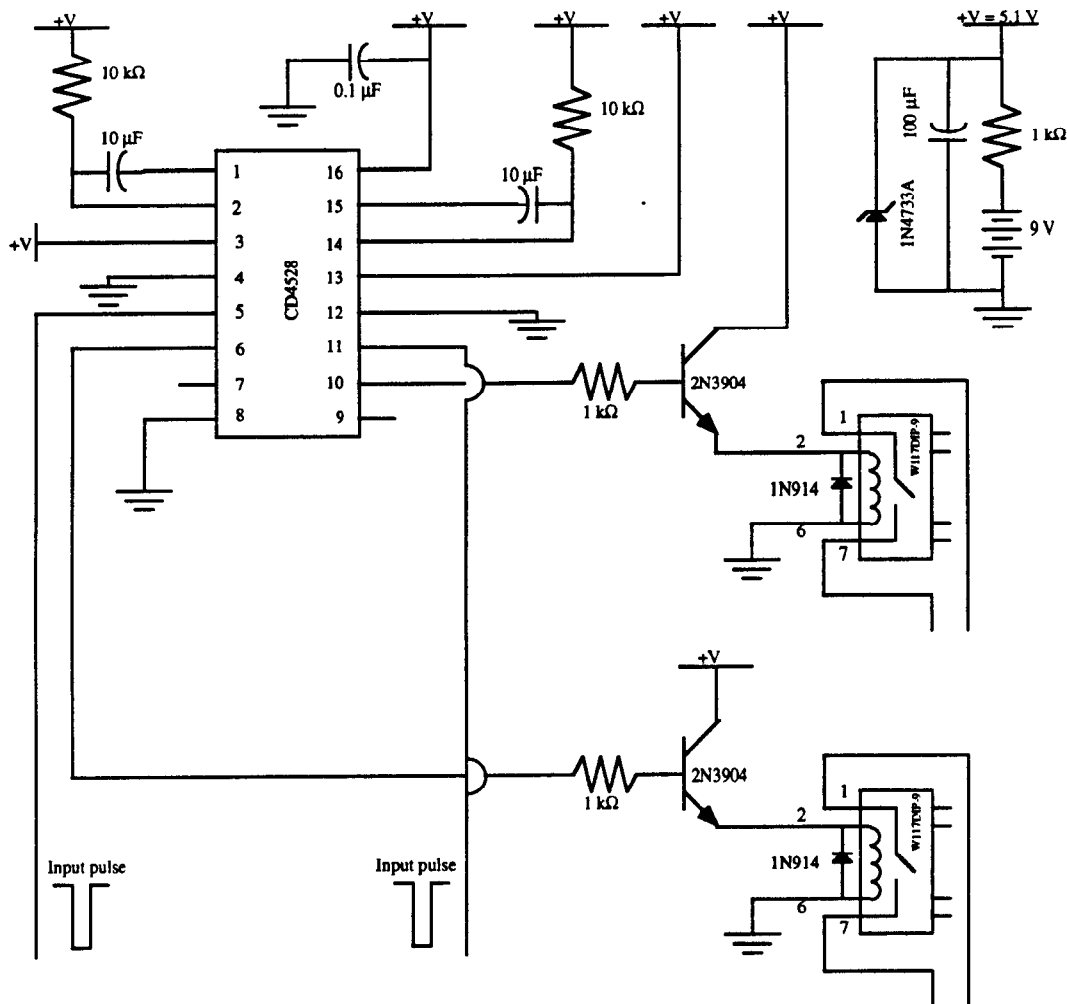


Figure 3.A.2: Circuit for converting 32 ns NIM pulses to long (30 ms) pulses for triggering relays to 'push' the rapid mixer controller buttons. The circuit is run off a 9 V battery regulated to 5.1 V with a 1N4733A Zener diode. The falling edge signal is converted by a monostable multivibrator (National Semiconductor CD4528C 'one-shot') to a ~30 ms high level pulse which triggers an electromechanical relay (Magnecraft W117DIP-9) to make the required contact. The length of the output pulse is controlled by the RC components connected to the 4528 and is specified as $0.2R(\text{k}\Omega)C(\text{pF})\ln(V_{\text{dd}}-V_{\text{ss}})$. The output pulse is used to drive a 2N3904 transistor in order to insure adequate power to the relay. The 1N914 diodes across the inner relay pins is recommended by the manufacturer to prevent voltage transients from causing multiple trigger events. Two 100 and 0.1 μF capacitors across the regulated voltage are used to eliminate transients (the circuit currently resides on an unshielded breadboard in the synchrotron hutch).

3.B: Ribonuclease A:

Sample Preparation:

Purified lyophilized bovine pancreatic ribonuclease A (Type XII-A) was purchased from Sigma Chemical Corporation (P.O. Box 14508, St. Louis, MO 63178). The protein was further purified chromatographically (by Dr. Virginia Robbins using the method of Garel, 1976) to remove residual contaminants, and was stored lyophilized at -5 °C. All samples were prepared gravimetrically, and because of the large number of samples and the time-pressure of synchrotron experiments, protein concentrations were only occasionally checked, using an extinction coefficient of $9800 \text{ M}^{-1}\text{cm}^{-1}$ at 278 nm. Protein concentrations are estimated to be accurate to within 5% of the values reported. Ribonuclease A is extremely soluble, but whenever possible SAXS samples were centrifuged in a desktop microfuge at 15000 rpm for 20 minutes or longer to remove any aggregates. No significant precipitate was ever observed.

Experiments on ribonuclease A were performed at conditions corresponding to earlier studies of ribonuclease A refolding using pulse labelling NMR techniques (Udgaonkar, 1988 and 1990) and equilibrium fluorescence measurements (Pace, 1990). Initially, conditions were chosen to correspond to the pulse conditions of the NMR experiments, specifically, 10 °C, 26 mM sodium formate, 0.1 M glycine, 0.4 M sodium sulfate, pH 9.0. In order to determine the unfolding transition midpoint and the protein stability, protein samples were prepared for UV absorption measurements at 0.04 g/l ribonuclease A and GuHCl (guanidine-hydrochloride or guanidinium, a protein denaturing agent) concentrations ranging from 0 to 8 M in 0.25 M increments. After several trials where large pH fluctuations were observed, it was found that under the above conditions the pH was highly temperature dependant. Therefore, all solutions had to be titrated to pH 9 at 10 °C, rather than at room temperature. The final solution samples were allowed to equilibrate overnight. Sample absorbance at 287 nm was used to probe the equilibrium unfolding transition. A Cary spectrophotometer modified to include a circulating bath temperature regulator was used to obtain the measurements. Samples remained in the instrument until the absorbance was observed to stabilize prior to recording a final value. Unfortunately, results remained time-dependant for long times (minutes). The resultant data is plotted in Figure 3.B.1. The transition midpoint was around 4M GuHCl, but the data were quite noisy. To determine a value for ΔG , the free energy of unfolding, the equilibrium constant of unfolding, $K = f_d/f_n$, was determined at each GuHCl

concentration as $(A - A_n)/(A_d - A)$, where A , the measured absorption, equals $f_n A_n + f_d A_d$, f_n and f_d are the native and denatured protein fractions and A_n and A_d are the absorption of the native and denatured state, extrapolated from the data outside the transition region. Linear extrapolation (Figure 3.B.2) of $\Delta G = -RT \ln K$ from the transition region to 0 M denaturant (Pace, 1986) gave a value of 7.14 ± 0.24 kcal/mole, leading to a value of $3.06 \cdot 10^{-6}$ for K .

Because of the temperature dependence of the glycine buffer, sample preparation and handling was judged to be too difficult and no SAXS data were collected at the pulse conditions. Instead, all further experiments were performed in 50 mM formate buffer at pH 4. This corresponds roughly to the refolding conditions used by Udgaonkar (1990). In the initial time-resolved experiments GuHCl was used as the denaturing agent. Solutions were prepared volumetrically using enzyme grade GuHCl. The large hygroscopicity of GuHCl and the lack of a refractometer for measuring the index of refraction of the solutions in order to determine the absolute GuHCl concentration introduced an uncertainty into this value. The error, however, was judged to be sufficiently small to not compromise the results.

In these initial SAXS experiments (see below), a very low signal to noise ratio was observed due to the relatively large absorption of the heavier chlorine atoms in the concentrated GuHCl solutions. As a result, samples for later experiments were prepared in urea (also a denaturing agent), which is composed of lighter atoms. Unfortunately, one complication of working with urea is its slow decomposition into cyanate and ammonium ions in solution (Pace, 1986). Cyanate can bind to the amino groups of proteins. In order to prevent possible modification of the protein structure and/or behavior as a result of such binding, it is necessary to work only with fresh (less than 24 hours old) urea solutions. All urea containing samples below were used within 24 hours of preparation. The midpoint of the urea unfolding transition of ribonuclease A under the chosen conditions, at 25 °C, is known from UV fluorescence measurements to be at 4 M urea (Pace, 1990).

Ribonuclease cleaves RNA by severing the phosphodiester bond between the 5' oxygen of one pyrimidine sugar and the 3' phosphate of the following sugar. The free 3' phosphate group forms a 2'-3' cyclic phosphate intermediate, which is then hydrolyzed back to a 3' only group. Isolated cytidine 2'-3' cyclic monophosphate (2'-3' CMP) is an analog for a ribonuclease A substrate intermediate, and is hydrolyzed by the enzyme. As

a result, enzyme activity, and therefore sample integrity, can be evaluated by measuring the rate of 2'-3' CMP hydrolysis by small added volumes of ribonuclease A samples (Crook, 1960). The UV absorption difference at 292 nm of $570 \text{ M}^{-1} \text{ cm}^{-1}$ between the 2'-3' and 3' forms provides a convenient monitor. Although absolute activity can be determined using the enzyme and substrate concentrations and the empirically determined reaction rate, it is frequently sufficient to compare the reaction rate observed for the sample in question to that of a sample of known integrity under identical conditions. This procedure was attempted for irradiated samples, and is described below.

Experimental Details, Specific:

Initial experiments were performed in July of 1990 at the Photon Factory in Japan. Static data were collected for 0.67, 1.3, and 6 mM ribonuclease A in 0.44 M, 0.9 M, and 4 M GuHCl. Collection times were 300 or 600 seconds for sample and background. The samples were contained in a flat 50- μm thick quartz-window cell. The sample-to-detector distance was 249 cm and the photon energy was 8250 eV. Time-resolved data were collected using the stop-cock rapid-mixer with the same sample-to-detector distance and photon energy. Conditions were 1:3.5 mixing of 6 mM protein in 4 M GuHCl with 0 M GuHCl buffer at 10 °C. Data were accumulated using various frame length structures, including 20 repetitions of (25 x 50 ms frames, 25 x 100 ms frames, 20 x 200 ms frames, 20 x 1 s frames, and 4 x 10 s frames), 10 repetitions of (10 x 10 ms, 10 x 20 ms, 10 x 50 ms, and 10 x 100 ms) and 1 repetition of 40 x 10 s.

In February of 1991 at the Photon Factory static data were collected from three concentrations of ribonuclease A, 26, 13, and 6.5 g/l, in 0 through 4 M GuHCl (one molar increments and one point at 2.5 M) and in 0 through 8 M urea (one molar increments plus 3.5 and 4.5 M) at 25 °C. Samples were contained in 20- μm thick quartz-window cells. A sample-to-detector distance of 111 cm and a photon energy of 8250 eV were used. Data were collected in single 600 (900) second cycles for each sample (background). Time-resolved data were collected with the stop-cock mixer at the same energy and sample-to-detector distance for 1:3.5 mixing of 55 g/l protein in 6 M urea with 0 M urea buffer at 4 °C and of 55 g/l protein in 5 M urea, 20% ethylene glycol with 0 M urea, 20% ethylene glycol buffer at -5 °C. The presence of ethylene glycol was necessary in order to prevent the freezing of the protein solution. At 4 °C, 31 cycles of 25 x 20 ms, 20 x 50 ms, 20 x 100 ms, 20 x 500 ms, and 9 x 1 s were collected, and at -5 °C, 23 cycles of the same time-structure were collected.

Following these experiments, an attempt was made to evaluate the potential for radiation damage to the samples by measuring the enzymatic activity of exposed samples compared to that of unexposed samples. Unfortunately, the attempt was inconclusive. Of the samples examined more than half showed an activity level within a few percent of that of unexposed samples, but several samples showed a decrease in activity of more than 10%, and a few samples even showed an increase in activity of a few percent! Obviously, the accuracy of these measurements is in question. The probable reasons for these problems are that the large dilutions are required from the concentrated SAXS samples (tens of mgs/ml) to the very dilute activity samples ($\mu\text{gs/ml}$) were too difficult to perform accurately under the time pressure and with the available equipment in Japan at that time. In addition, the UV absorption activity measurements are highly temperature sensitive, and the stability of the available spectrophotometer temperature-control system was questionable. Because most samples showed good activity, and because ribonuclease A is known to be extremely resilient, it was thought that radiation damage would not be an issue in these studies. However, subsequent analysis of the data raised the possibility of radiation induced disulfide bond reduction (see below).

The next experiments were performed at the Photon Factory in November of 1991. Static data only were collected for 20 g/l and 4 g/l ribonuclease A in 0 through 8 M urea (half molar increments except 0.5 and 7.5 M) at 25 °C. A sample-to-detector distance of 107 cm was used for the 20 g/l data and of 229 cm for the 4 g/l data. The photon energy was 8250 eV. Single cycles of 1500 seconds were used for samples and backgrounds at 20 g/l and single cycles of 2700 seconds were used for samples and backgrounds at 4 g/l. Samples were contained in 20- μm thick quartz-window sample cells.

In 1992, two sets of experiments (April and August) were performed at SSRL in order to study the thermal denaturation of ribonuclease A. 5 g/l protein samples were studied in 0, 1, 2, and 4 M urea at temperatures from 10 to 75 °C. The samples were contained in 10- μm thick quartz-window cells which were heated using a circulating water bath. The temperature of the sample holder corresponding to the bath temperature was calibrated prior to the experiment. The sample-to-detector distance was 284 cm in April and 263 cm in August. The photon energy was 8980 eV. Data were collected in 3 to 6 five or ten minute cycles for each temperature, as well as in 25 to 30 one-minute cycles for experiments in which the temperature was allowed to change continuously.

In July of 1992 a final time-resolved experiment was performed at the Photon Factory in which 50 g/l ribonuclease A in 7 M urea were mixed using the stop-cock mixer at a 1:6 ratio with 0 M urea buffer at 7.5 °C. Data were collected for 15 cycles of 94 x 20 ms. The sample-to-detector distance was 233 cm, and the photon energy was 8250 eV.

Analysis and Results:

All of the data collected were normalized by the integrated incident beam intensity, background subtracted, and calibrated. In the case of the concentration series of February 1991, the background subtraction was first performed for the highest protein concentration data. The subtraction for the lower protein concentrations was then performed by minimizing the (least squares) difference between the resultant data and the higher concentration data at angles past those at which interparticle interference effects were present (see Hubbard, 1987). Plots of the data from three protein concentrations at 0, 4, and 8 M urea from February 1991 are shown in Figure 3.B.3. Interparticle interference effects are clearly visible as a decrease in the low angle intensity for the higher protein concentrations. After the background subtraction, the February data were extrapolated to zero protein concentration in order to eliminate these effects. It was found, however, that results from the extrapolated data only differed by a few percent from those of the lowest concentration data. Therefore only two concentrations were collected in November, and no extrapolation was used for these data. Instead, after background subtraction, the 20 g/l and 4 g/l data were spliced at the point where interparticle interference effects were no longer evident. Data from November 1991 are shown in Figure 3.B.4.

The results of a Guinier analysis of the two equilibrium urea unfolding experiments are presented in Figure 3.B.5 as plots of R_g and I_0 versus urea concentration. The fitting range used was $K = 0.028$ to $K = 0.049 \text{ \AA}^{-1}$ for all except the 0 M urea data, for which an upper value of $K = 0.056 \text{ \AA}^{-1}$ was used (K is again the magnitude of the scattering vector, as in Chapter 1). This was the lowest angular range which gave reasonably consistent results. Despite the noise in the data, it is immediately apparent from both experiments that the sharp unfolding transition which was expected around the 4 M urea midpoint is absent. Instead, a gradual transition is observed which appears to be almost complete by 4 M urea. The I_0 data show a possible increase between 0 and 1 M urea, but this is more likely due to experimental error. These results are very puzzling, and no conclusive explanation has been found, although several possible ones are discussed below. A

two-state folding transition is clearly observed under these conditions when optical probes (UV absorbance, fluorescence, and circular dichroism) are used.

Thermal denaturation data were only collected at one concentration (5 g/l), and the results of a standard Guinier analysis (again using a fitting range of $K = 0.28$ to 0.049 or 0.056 \AA^{-1}) are presented in Figures 3.B.6 and 3.B.7. A clear two-state transition is evident, as expected, in the 0 M urea data. No significant change in I_0 accompanies the transition. A similar, though less obvious, transition is apparent in the 2 M urea data. The 1 M data does not show a transition, probably because a high enough temperature was not reached. A transition is also absent in the 4 M data implying that the protein is already unfolded. Note, that the R_g values at 25 °C at 0, 1, 2, and 4 M urea are in good agreement with the results from the 1991 equilibrium urea unfolding data. Thus, the pre-transition increase in R_g during urea unfolding at 25 °C has been reproduced in three independent experiments.

It is interesting, to compare the thermal denaturation data with similar data collected around the same time by Sosnick (1992) at Los Alamos using an x-ray generator source. These results indicate that in the absence of denaturant, under non-reducing conditions, the R_g for thermally denatured ribonuclease A is around 20 Å. This is far less than the value of around 28 Å observed in the 1992 SSRL data. Under reducing conditions, however, Sosnick observed a denatured R_g value of 28 Å. This suggests the possibility that the disulfide bonds of the samples studied at SSRL and the Photon Factory were actually reduced by the high x-ray flux. The reduction of disulfide bonds by synchrotron radiation has been documented, especially in the presence of formate ions (Koch, 1991 and Yokoya, 1991). Because the flux produced by an x-ray generator is several orders of magnitude lower than that produced by a synchrotron, this effect might have been absent in the Los Alamos data. This hypothesis is further supported by the fact that the thermally induced transitions observed at SSRL were not reversible upon recooling of the sample. In the Los Alamos data, the thermal transition under non-reducing conditions was reversible, but the transition under reducing conditions was irreversible.

The possibility that ribonuclease A disulfides can be reduced by a high x-ray flux suggests that this effect may be responsible for the unexpected behavior of R_g in the urea induced unfolding transition. Although the reduced protein remains compact in the absence of chemical denaturants, it seems quite plausible that it is less stable in this condition, and is thus susceptible to unfolding by lower concentrations of denaturant. The

observed urea unfolding data would be consistent with a broad transition centered around a 1 to 2 M urea concentration.

Several other possibilities can be suggested for the lack of a clear transition in the urea unfolding R_g data. First of all, it is possible that this data is simply a result of the failure of the two-state model. Although optical and other probes show a two-state transition, they are only sensitive to local environments in the protein, and it is possible that an overall expansion of the protein can take place without disturbing the local environment. In this scenario, the protein may start to swell long before the unfolding transition midpoint.

R_g values derived from SAXS data of polydisperse systems will be heavily biased in favor of the R_g value of the largest scatterers in the sample. Therefore, it is possible that the small amounts of unfolded protein which are to be expected as the equilibrium between the folded and unfolded state begins to shift bias the data sufficiently to wash out the sharp two-state transition.

Denaturants like GuHCl and urea bind to proteins in solutions. Studies exist which show a significant increase in the binding of denaturants to protein during unfolding, and such binding begins long before the unfolding transition point is reached. In fact, published curves of such binding bear a considerable resemblance to the R_g data (Lee, 1974). If denaturant molecules are locally denser when bound to the protein than when free in the solvent, they will add to the excess electron density of the protein, and will thus increase its R_g . The binding of urea to ribonuclease A may be responsible for the pre-transition increase observed in R_g . However, such an effect should also be apparent in I_0 , and even if the change observed in I_0 at 1 M urea is not due to experimental error, it is much smaller than and not well correlated with the change in R_g .

A common problem with many studies of proteins is the tendency of proteins to aggregate in solution. In addition, this tendency is known to be enhanced by exposure to intense radiation. Although ribonuclease A is particularly well-behaved, some evidence is available in the literature for ribonuclease A dimerization. Protein aggregation, radiation induced or not, is clearly a possible cause of the pre-transition increase in R_g . Because aggregation can effect R_g more strongly than I_0 , it is possible that the increase in I_0 is in fact an indication of sample aggregation. Dimerization of 10% of the sample would lead to a 10% increase in I_0 , but could lead to a larger increase in R_g .

The time-resolved data collected in 1990, 1991, and 1992 were all subjected to standard Guinier analysis as above. Unfortunately, the statistics accumulated were not sufficient to obtain very meaningful R_g values, as can be seen from Figure 3.B.8a. Plots of I_0 were of sufficient quality to indicate that no change in I_0 occurred during the observation times (Figure 3.B.8b). By overlaying data from the initial and final frames, it was apparent that the large difference observed between the (contrast corrected) static data from folded and unfolded states was absent (Figure 3.B.9). It was concluded that compaction of the protein must have occurred within the time-resolution of the experiment, implying an upper bound of 20 ms for the compaction of ribonuclease A under the various conditions used.

Conclusions:

Although these experiments did not succeed in observing the compaction process of ribonuclease A, a valuable piece of information has been added to the picture of how this protein folds. It has been established that the compaction process must occur within the first 20 ms of refolding of this protein under conditions similar to those in which secondary structure formation occurs into the 100 ms time scale. It is therefore apparent that at least some secondary structure formation occurs concurrently with and/or after the compaction of the protein. The fact that no observable compaction event was noted, despite the reputed existence of three different unfolded species of ribonuclease A (different proline isomers, Udgaonkar, 1990) which fold at very different rates (a few ms, hundreds of ms, and tens of s), would seem to imply that even the slowest folding of the unfolded species compacts very quickly.

The equilibrium experiments also failed to establish SAXS as a technique for monitoring equilibrium unfolding transitions. One possible explanation for this failure would be the reduction of the disulfide bonds in the protein by the high synchrotron x-ray flux. This explanation could be tested by measuring the urea unfolding transition of ribonuclease A under reducing conditions using another probe and comparing these results with the SAXS results. Another possible explanation is a partial aggregation of the protein at low urea concentrations. Better SAXS data is needed to determine the behavior of I_0 more precisely in order to test the validity of this explanation.

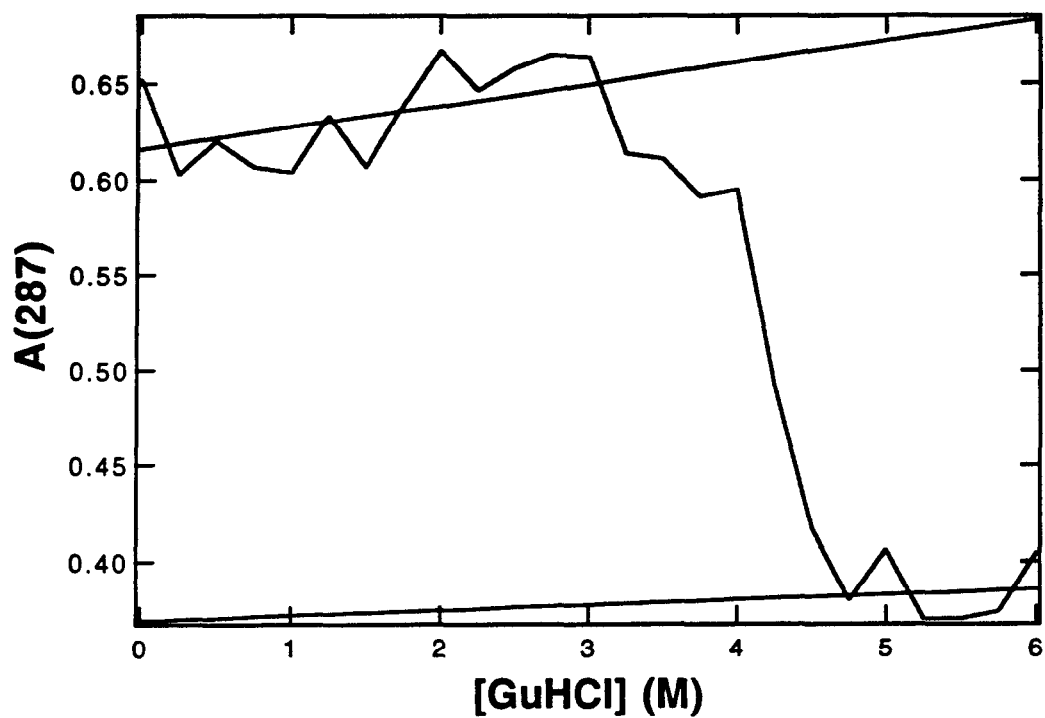


Figure 3.B.1: GuHCl-induced equilibrium unfolding of 0.04 g/l ribonuclease A in 26 mM formate, 0.1 M glycine, 0.4 M sulfate, pH 9, monitored by UV absorbance at 287 nm. The lines are linear regression fits to the pre- and post-transition baselines.

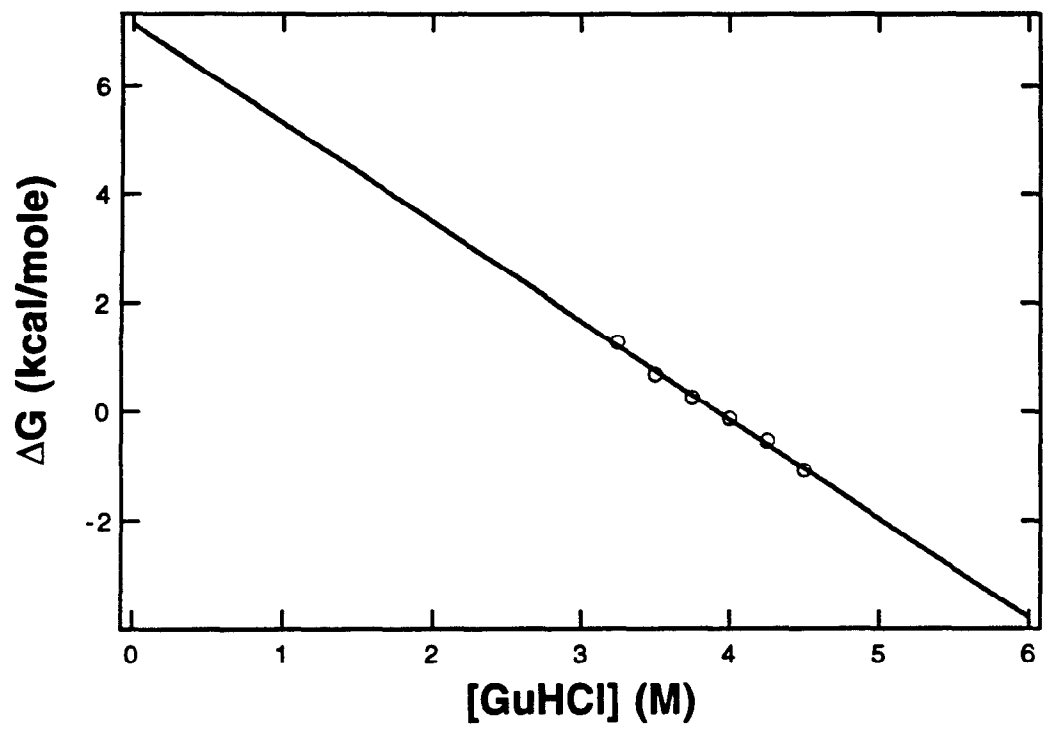


Figure 3.B.2: Extrapolation of the free energy of unfolding from the transition region to 0 M GuHCl. At each data point the equilibrium constant K was calculated using values from a polynomial fit to the data in Figure 3.B.1 in the transition range.

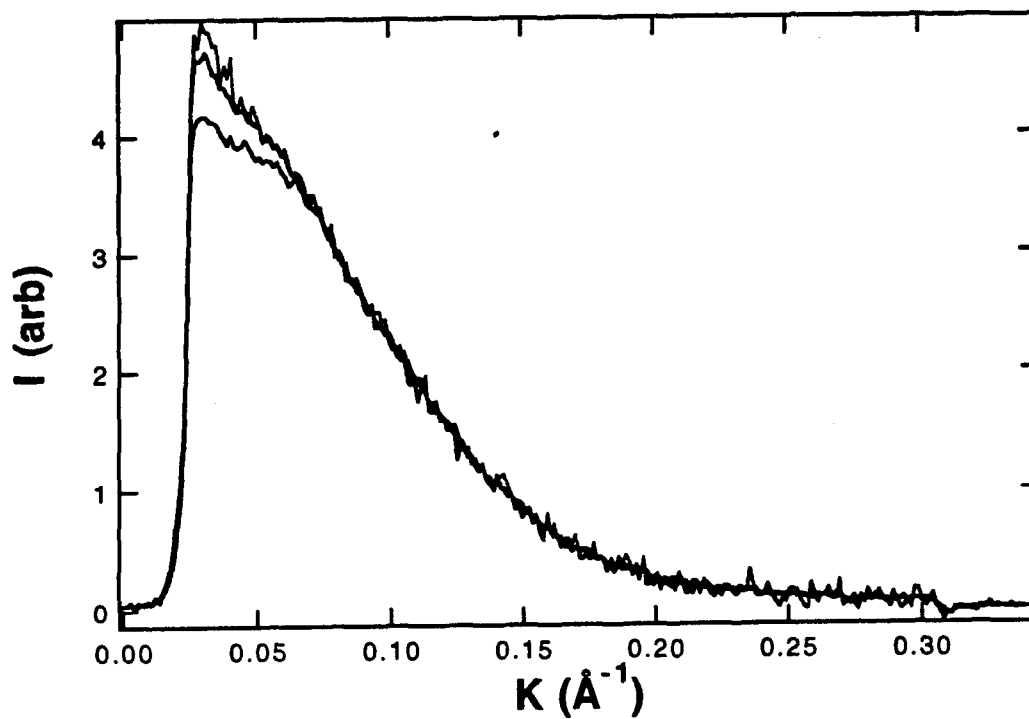


Figure 3.B.3a: Normalized background-subtracted calibrated SAXS data from February 1991 for 26, 13, and 6.5 g/l ribonuclease A in 50 mM formate, 0 M urea, pH 4.

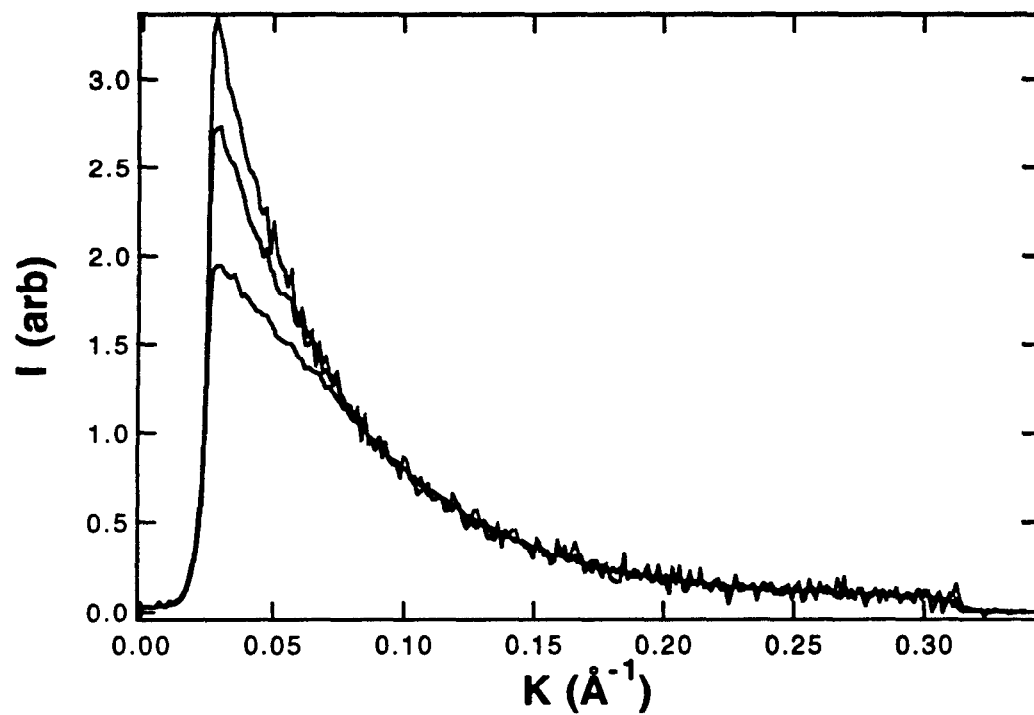


Figure 3.B.3b: Normalized background-subtracted calibrated SAXS data from February 1991 for 26, 13, and 6.5 g/l ribonuclease A in 50 mM formate, 4 M urea, pH 4.

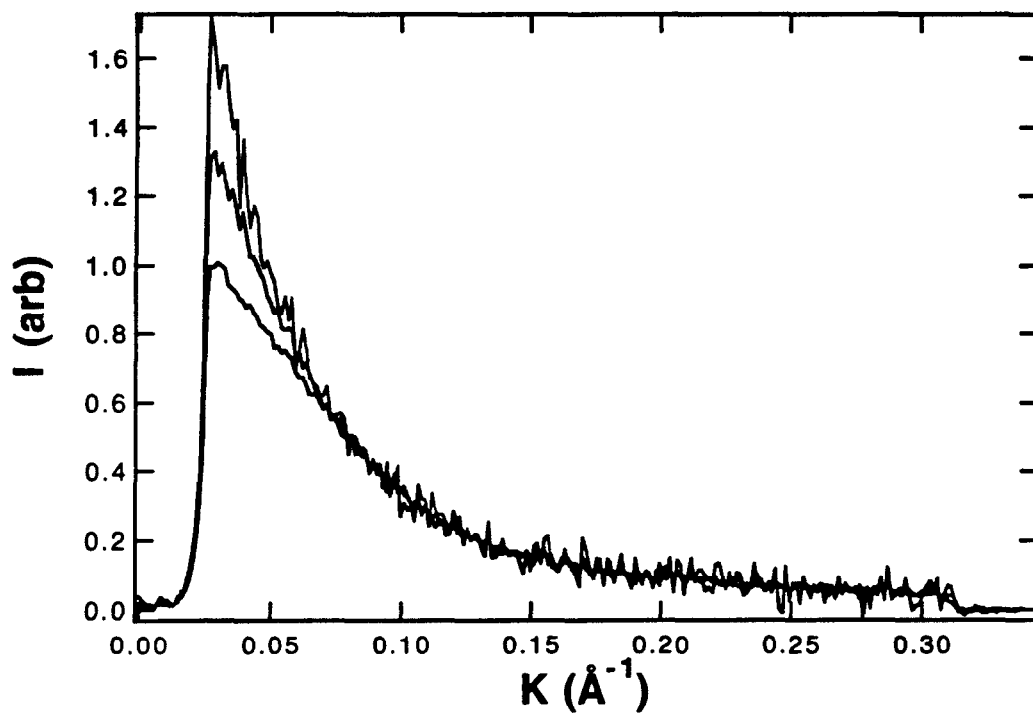


Figure 3.B.3c: Normalized background-subtracted calibrated SAXS data from February 1991 for 26, 13, and 6.5 g/l ribonuclease A in 50 mM formate, 8 M urea, pH 4.

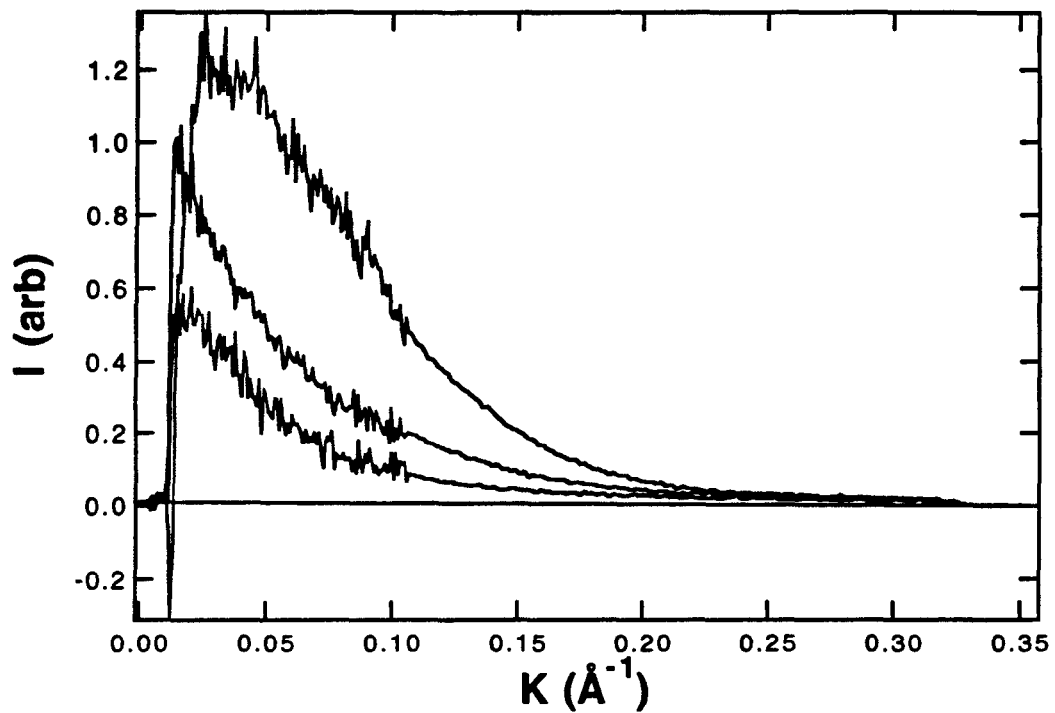


Figure 3.B.4: Normalized background-subtracted calibrated SAXS data from November 1991 spliced from 20 g/l (higher angles) and 4 g/l (lower angles) ribonuclease A in 50 mM formate, pH 4.0, and 0 (upper), 4 (middle), and 8 (lower) M urea.

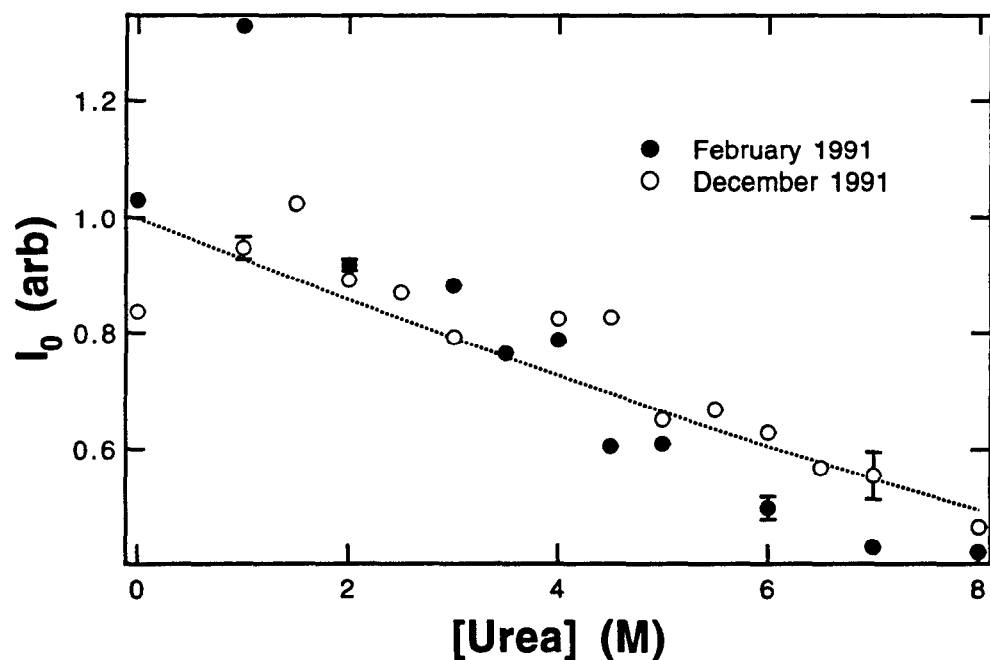
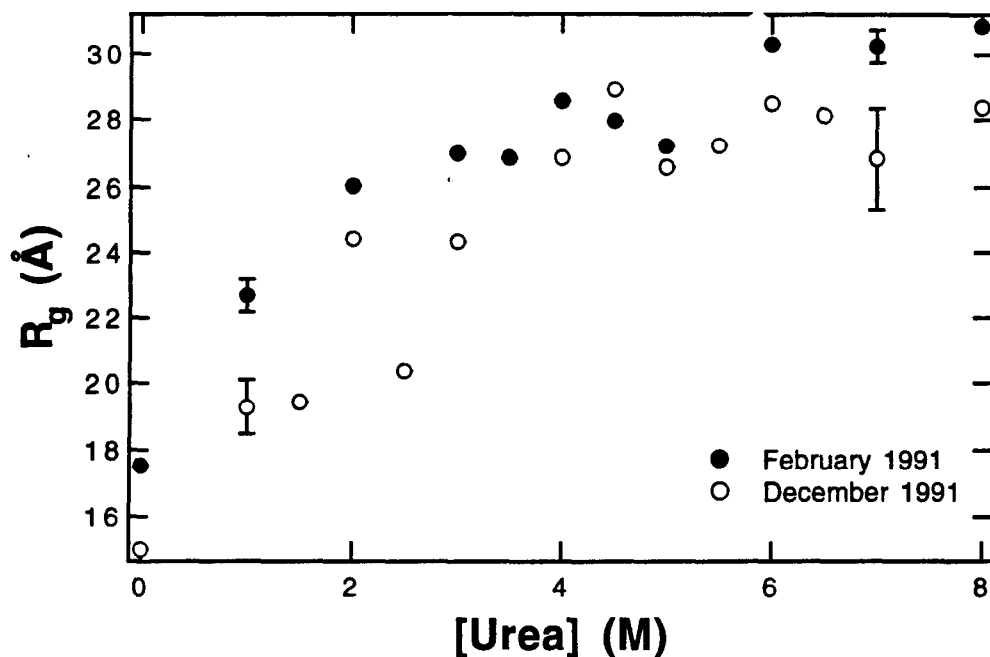


Figure 3.B.5: R_g and I_0 values from the February and November 1991 ribonuclease A data. An sharp unfolding transition is expected at 4 M urea. Instead R_g displays a very broad transition between 0 and 4 M urea. The dotted line in the I_0 plot represents the expected decrease in I_0 due to the changing solvent electron density, assuming a (constant) $0.45 \text{ e}^-/\text{\AA}^3$ for the protein, $0.333 \text{ e}^-/\text{\AA}^3$ for water, and $0.00432 \text{ e}^-/\text{\AA}^3/\text{M}$ for urea (see myoglobin section). The data is in rough agreement except between 0 and 1 M urea where I_0 shows a possible increase. If this increase is real, it may indicate solvent binding or sample aggregation.

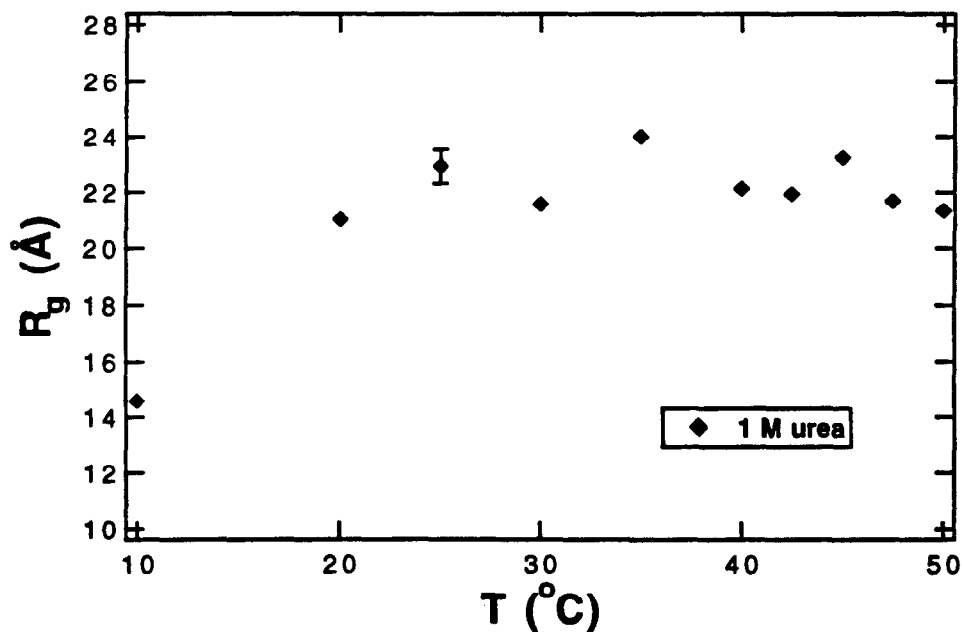
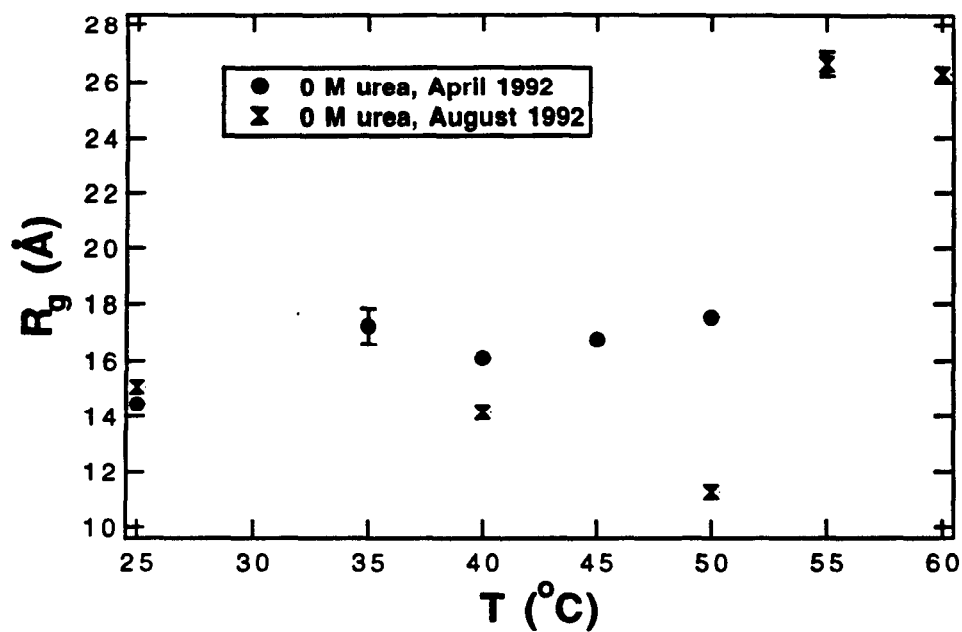


Figure 3.B.6a,b: R_g values measured during the thermal denaturation of 5 g/l ribonuclease A in 50 mM formate, pH 4.0, and a) 0 M urea, b) 1 M urea. A transition is clearly evident around 50 to 55 °C in the 0 M urea data. No transition is observed in the 1 M urea data, indicating it probably occurs above 50 °C. Note the room temperature value of R_g in 1 M urea agrees with that in Figure 3.B.5.

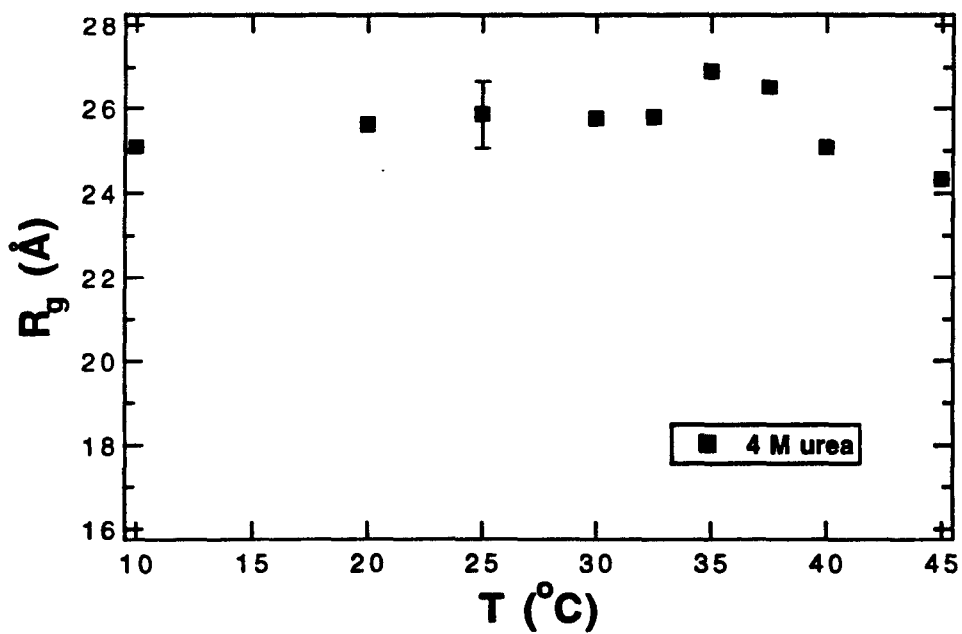
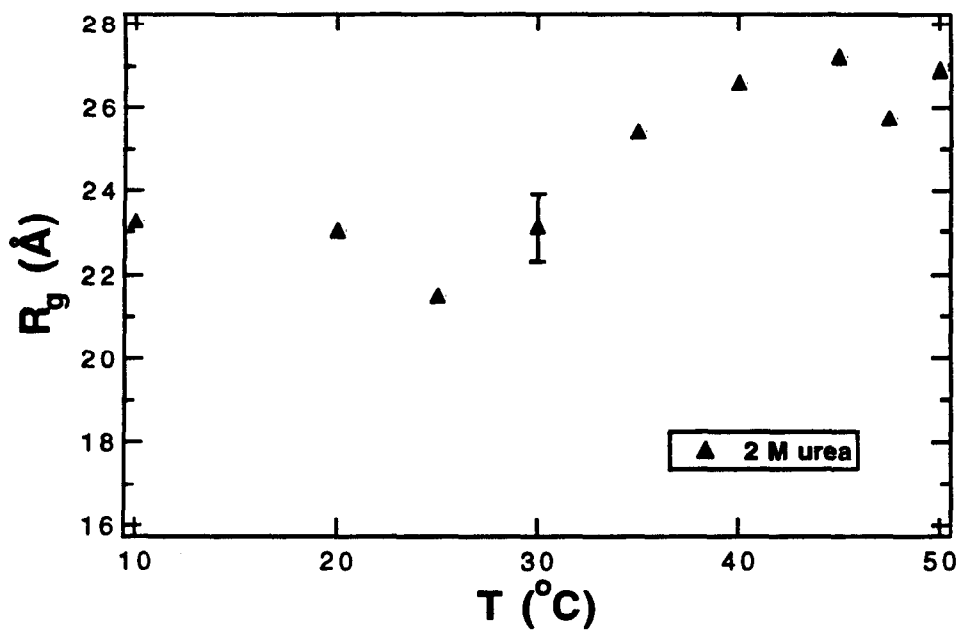


Figure 3.B.6c,d: R_g values measured during the thermal denaturation of 5 g/l ribonuclease A in 50 mM formate, pH 4.0, and c) 2 M urea, d) 4 M urea. The 2 M urea data show a transition around 35 °C, whereas the 4 M urea data show no transition, indicating the protein is probably already unfolded. Again the room temperature values of R_g agree with those in Figure 3.B.5:

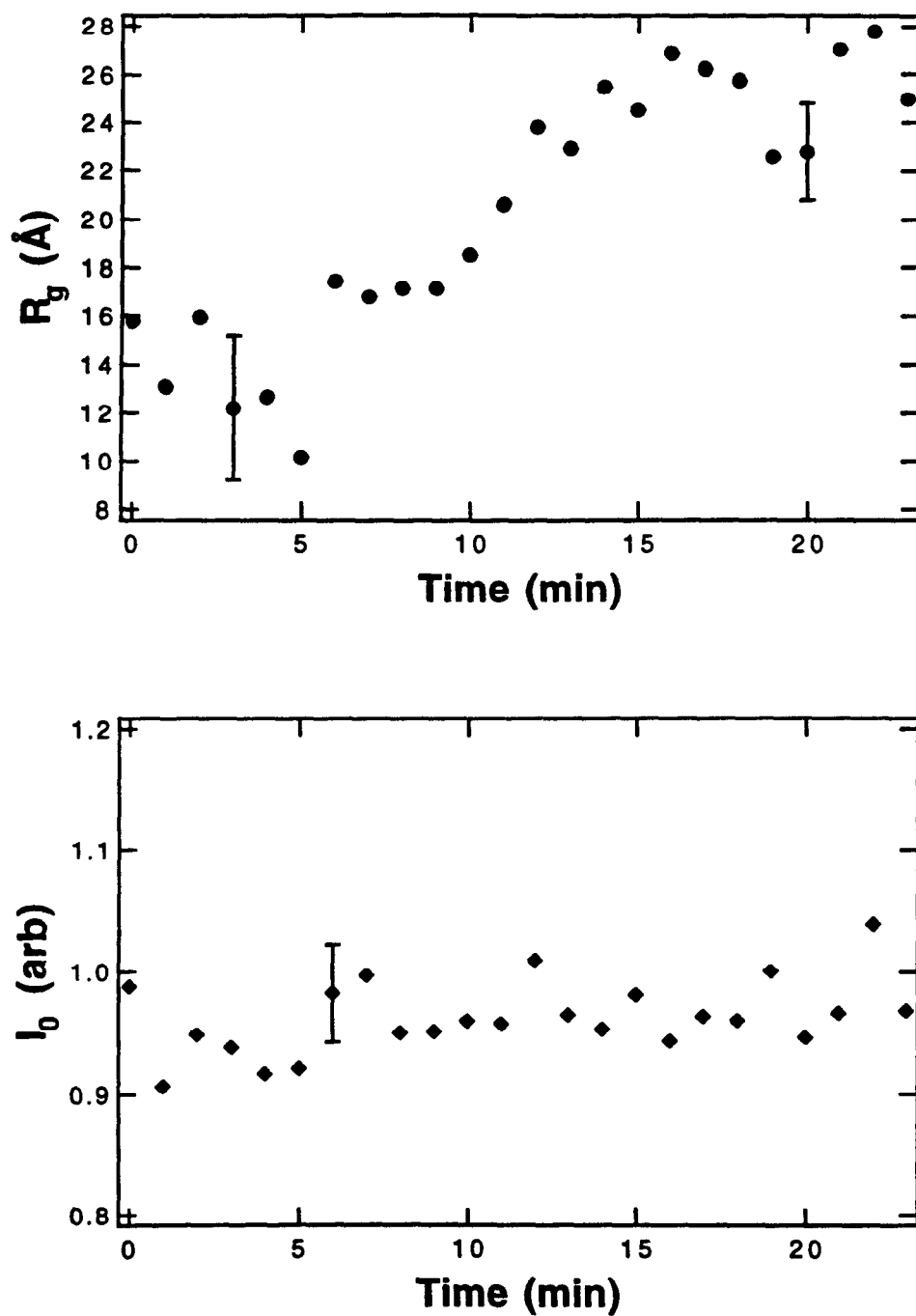


Figure 3.B.7: R_g and I_0 during the thermal denaturation of 5 g/l ribonuclease A in 50 mM formate, pH 4.0 (0 M urea). The sample was equilibrated to 25 °C at the beginning of the experiment. Using a circulating bath, the sample was then heated to 75 °C. During heating, SAXS data was collected in one-minute cycles. The data points are an average of two consecutive experiments. A transition is observed in R_g at about 10 minutes. No significant change is observed in I_0 .

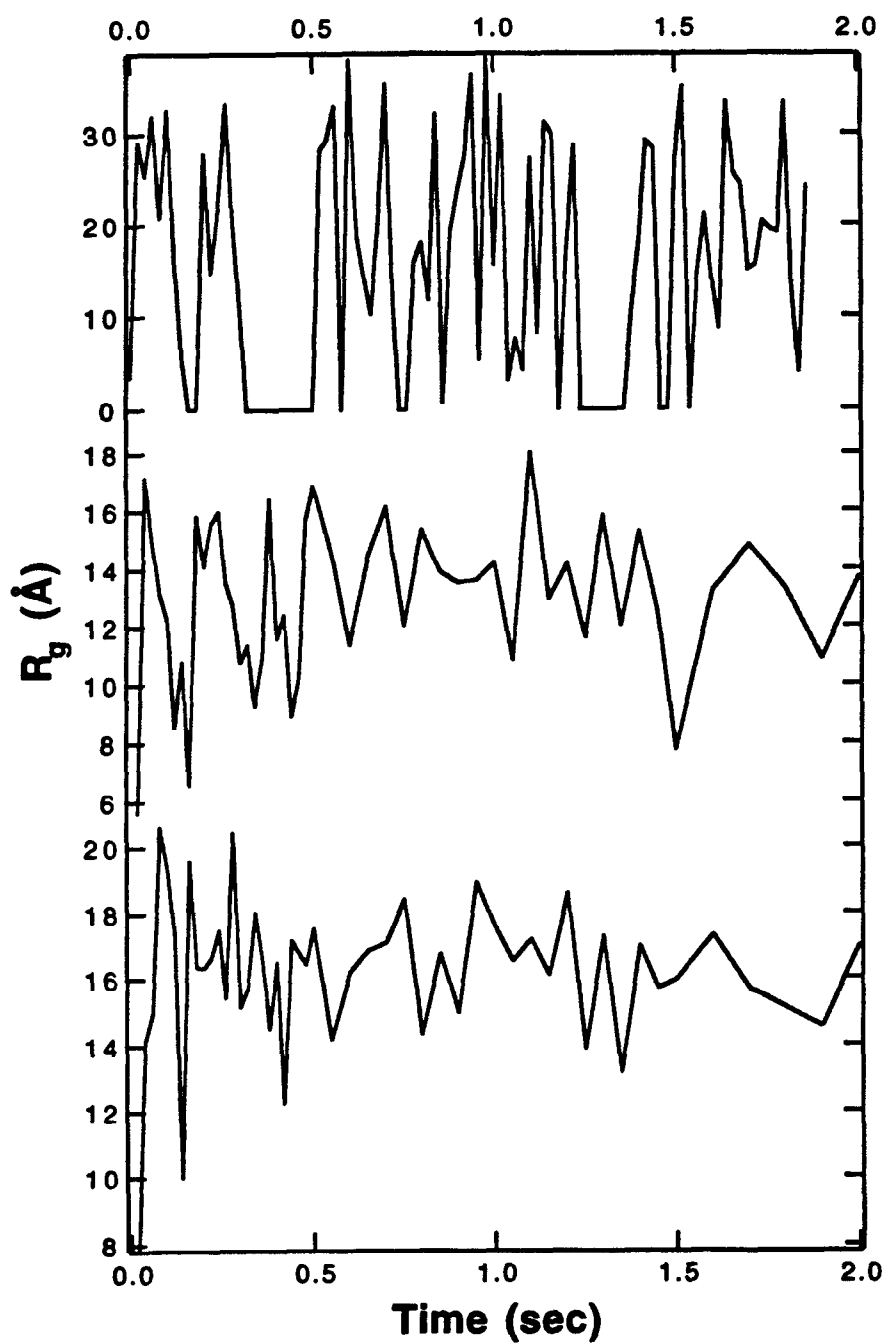


Figure 3.B.8a: R_g as a function of the refolding time of ribonuclease A in 50 mM formate, pH 4.0 when diluted from 7 to 1 M urea at 7.5 °C (upper), from 6 to 1.3 M urea at 4 °C (middle) and from 5 to 1.1 M urea at -5 °C (in the presence of 20 % ethylene glycol). No significant trend is observed.

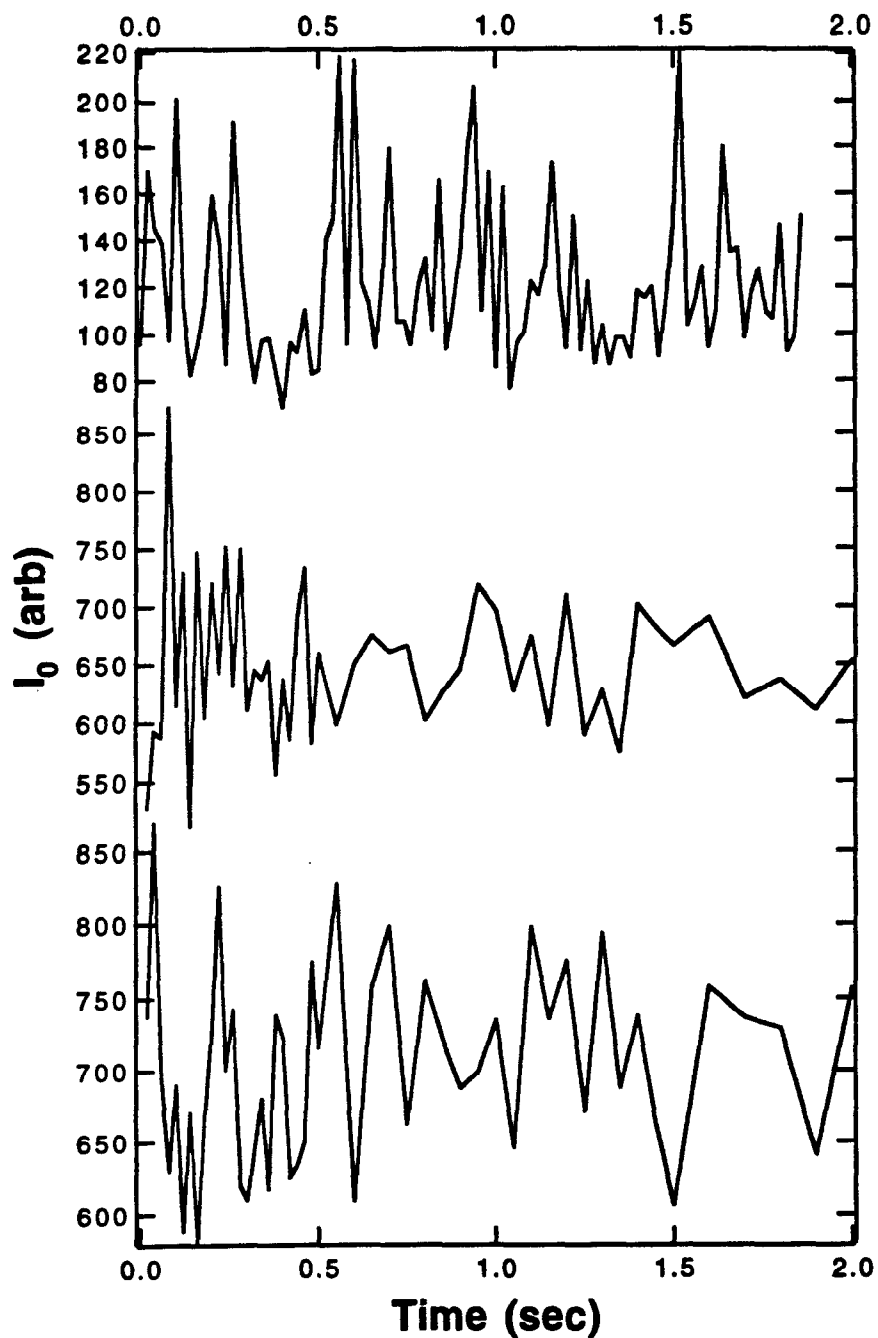


Figure 3.B.8b: I_0 as a function of the refolding time of ribonuclease A in 50 mM formate, pH 4.0 when diluted from 7 to 1 M urea at 7.5 °C (upper), from 6 to 1.3 M urea at 4 °C (middle) and from 5 to 1.1 M urea at -5 °C (in the presence of 20 % ethylene glycol). No significant trend is observed.

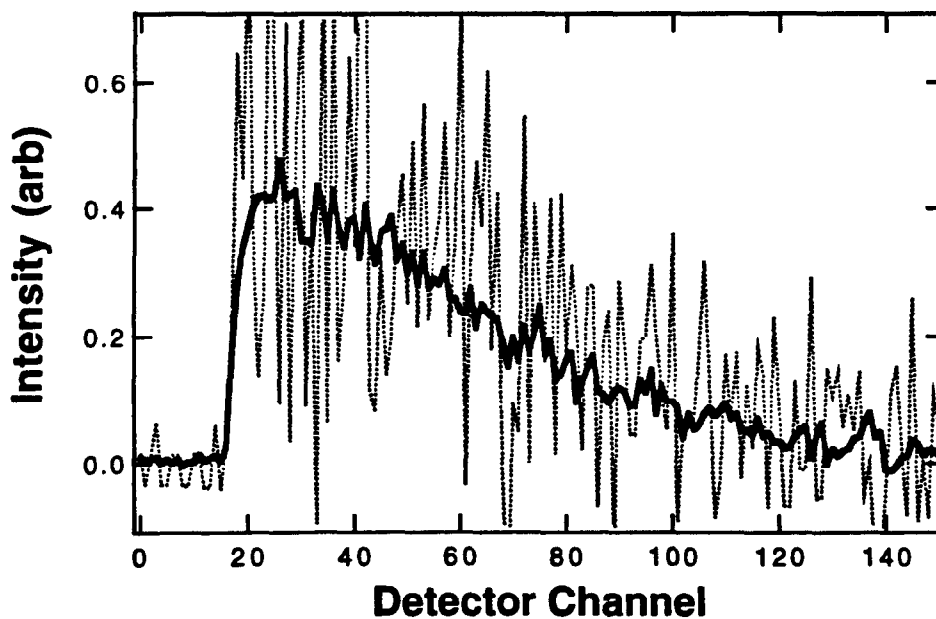
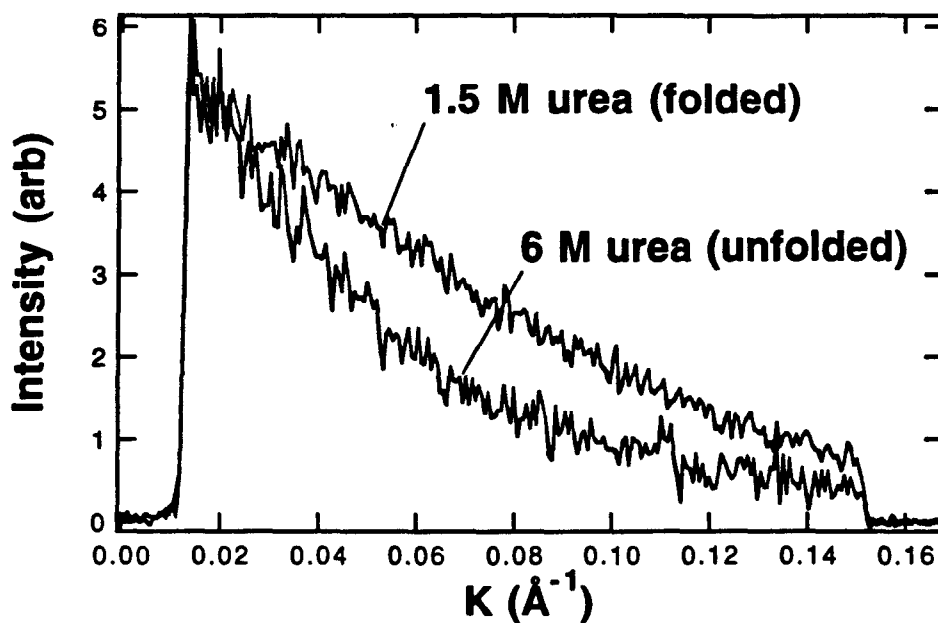


Figure 3.B.9: A comparison of the difference between the SAXS patterns of unfolded and folded ribonuclease A and the difference between the first 20 ms data frame and one of the last 1 s data frames from the refolding of ribonuclease A at 4 °C, showing that the protein is already folded by the time the first data frame is collected. Similar results were obtained under the other experimental conditions.

3.C: Myoglobin:

Evidence for an Associative Folding Intermediate:

Sample Preparation:

Horse skeletal muscle myoglobin was purchased from the Sigma Chemical Company, St. Louis, MO (crystallized and lyophilized, salt-free, 95-100% pure) and stored frozen until use. The purchased protein was not purified further.

The conditions used for the study of myoglobin were developed in collaboration with K. Chiba of the Tokyo Institute of Technology. Initially, the chosen conditions were pH 9.1, 0.1 M NaCl, 50 mM borate, 10 x stoichiometric excess CN^- (added as NaCN), and 5 M GuHCl (the denaturing agent). The presence of cyanide throughout the studies was necessary in order to prevent the potential aggregation of free or exposed heme groups. The cyanide tightly binds the heme iron on either side of the heme, preventing iron-iron interactions (Shack, 1947). As discussed in the ribonuclease A studies, however, high concentrations of GuHCl are non-ideal for SAXS studies because of the relatively large absorption cross-section of chlorine. Therefore, the denaturing agent was switched to urea. As in the case of ribonuclease A, all urea-containing solutions were used within 24 hours of preparation in order to avoid urea decomposition. In addition, the borate buffering agent was switched to bis-tris-propane, a dual pK_a (6.8 and 9.0) biological buffer, which was found to be more soluble than borate.

Protein samples were prepared gravimetrically, without further verification of the absolute protein concentration. The estimated accuracy is about 5%. This level of uncertainty in the protein concentration does not impact the reported results (see below). All protein samples prepared were subjected to centrifugation in order to remove any undissolved or aggregated protein. It was found that the occasional presence of ethylene glycol (see below) resulted in more precipitate formation during centrifugation. In the absence of ethylene glycol, little or no precipitate was typically observed upon centrifugation. When possible, the samples were spun at 15000 rpm for 30 minutes or longer in a temperature controlled microfuge. Occasionally, only an older centrifuge was available, in which samples could only be spun at up to 4000 rpm. All unfolded protein samples were allowed to equilibrate for at least one hour prior to data collection.

In order to determine the urea concentration of the equilibrium unfolding transition midpoint under these conditions, circular dichroism measurements were made at 222 nm using 25 μ M myoglobin in urea concentrations between 0 and 8 M. The resulting equilibrium unfolding curve is presented in Figure 3.C.1. It should be mentioned that this curve is not in agreement with CD data collected under similar conditions in Japan by Chiba. Although the discrepancy has not yet been explained, the data collected at Stanford were found to be reproducible in a repetition of the experiment. On the other hand, the data collected in Japan are in much better agreement with the SAXS equilibrium unfolding studies (presented below). One possible cause of these observations may be a dependence of protein stability on protein concentration.

Experimental Details, Specific:

In February of 1991, initial data were collected at the Photon Factory using borate buffer and GuHCl. As discussed above, these conditions were found to be non-ideal. The first indications of a signal were observed, however, and provided encouragement for further experiments.

In November of 1991, the first experiment using urea was run at the Photon Factory. 68 g/l myoglobin in 8 M urea, 10% ethylene glycol, 0.1 M NaCl, 40 mM NaCN, 50 mM bis-tris-propane, pH 9.1 were mixed with equivalent 0 M urea buffer using the stop-cock mixer at a 1:3.5 ratio at 8 °C. The ethylene glycol was used in anticipation of the need to go to sub-zero temperatures. Except for the amount of ethylene glycol, all the following experiments were performed using the same buffer conditions, which will not hereafter be explicitly stated. Data were collected in 10 cycles of 20 x 20 ms, 20 x 50 ms, 20 x 200 ms, 20 x 500 ms, and 14 x 2 s. A total of 100 cycles were collected. The photon energy was 8250 eV and the sample-to-detector distance was 230 cm.

In April of 1992, the ruby-ball stopped-flow rapid-mixer was commissioned at SSRL. Because of occasional urea precipitation at 8 °C in the previous experiment, the ethylene glycol content was raised to 40%, as it was thought this would improve the solubility of urea. This turned out to be an erroneous assumption. 68 g/l myoglobin in 7 M urea were mixed at a 1:6 ratio with 0 M urea at 8 °C. Data were collected in cycles of 10 x 200 ms, 10 x 400 ms, and 10 x 800 ms. 65 cycles total were collected at an energy of 8980 eV. In May of 1992, further data were collected, under identical conditions. A total of 270 cycles of 10 x 200 ms, 10 x 400 ms, and 10 x 800 ms were collected at 8980 eV and a

sample-to-detector distance of 267 cm.

In July of 1992, at the Photon Factory, static data were collected from 5 g/l myoglobin in 0 through 8 M urea in 1 M increments under the same conditions, including 40% ethylene glycol. Samples were contained in a 20- μm thick quartz-window cell. The sample-to-detector distance was 231 cm and the photon energy was 8250 eV. Time-resolved data were also collected using the stop-cock mixer. At 1:6 mixing, 60 g/l myoglobin in 7 M urea both with and without 40% ethylene glycol were mixed with 0, 1, and 2 M urea buffer at 15 and 8 °C. The sample-to-detector distance was 233 cm and the photon energy was 8250 eV. Data were collected in 5 cycles of 94 x 50 ms, 100 ms, or 200 ms at a time for a total of 50 to 75 cycles. 50 cycles of 50 x 20 ms and 44 x 50 ms were also collected using an initial concentration of 8 M urea in order to check for any difference in the results due to a higher initial urea concentration. No significant difference was noted (see below).

Analysis and Results:

All data collected were normalized, background subtracted, and calibrated. Interparticle interference effects were ignored throughout this analysis, since protein concentrations were always below 10 g/l, and because the focus was not to obtain absolute values of R_g , but rather to observe changes in R_g , which should be less influenced by such effects. The R_g results of a Guinier analysis of the July equilibrium unfolding data (fits from $K = 0.028$ to $K = 0.062$) are presented in Figure 3.C.2a, along with results from further equilibrium data collected in March, 1993 in the absence of ethylene glycol (see below). The fitting range was chosen as the smallest range which gave consistent results for all the data (time-resolved and static). The consequences of this choice are discussed further below. As can be seen, the two data sets are in good agreement, although the noise level in general is high. Because the earlier data were collected in the presence of 40% ethylene glycol, the agreement of the two data sets indicates that ethylene glycol does not significantly alter the stability of the protein. Despite the disagreement with the CD data, the myoglobin equilibrium SAXS data, unlike those in the case of ribonuclease A, do display a reasonably abrupt unfolding transition around 6 M urea, and there is no evidence of aggregation at intermediate urea concentrations.

As mentioned in Chapter 1, I_0 is proportional to the square of the excess electron density of the scattering protein over the solvent electron density, or $[(\rho_p - \rho_s)V]^2$, where ρ_p is

the protein electron density, ρ_s is the solvent electron density, and V is the (specific) volume of the protein in the solvent. In order to use the equilibrium unfolding data to determine if any solvent binding effects are observed in the case of the unfolded protein, it is first necessary to correct for the effects of the different contrast levels between the protein and the solvent in solutions of different urea concentrations. To do this we use the above equation, where to a first order approximation we hold the protein volume and the protein electron density fixed throughout the unfolding transition. While this may seem unreasonable at first, in fact measurements of total volume changes of protein solutions upon unfolding of the protein fraction reveal that the change in the specific volume of the protein is small (Privalov, 1988), meaning that the solvent meshes effectively with the exposed surface of the unfolded protein. It appears plausible that the correct picture to keep in mind is one of an unfolding chain of fixed diameter and length (and therefore volume) and constant internal electron density. With these constraints, it becomes simple to make numerical predictions for the decrease of I_0 with increasing urea concentration.

The concentration of liquid water is about 55.5 M. The amount of water displaced by urea in n M urea solution is approximately 44.7n g/l or about 2.48n moles/l (CRC, 1987). The number of electrons in a molecule of water or urea is 10 and 32 respectively. The electron density of n M urea is then about $(10(55.5-2.48n) + 32n) * N_A \text{ e}^-/\text{liter} = 0.6(0.555 + 0.0072n) \text{ e}^-/\text{\AA}^3 = 0.333 + 0.00432n \text{ e}^-/\text{\AA}^3$. The effect of the buffer components must also be accounted for. The electron density increase due to NaCl, NaCN, and bis-tris-propane is about $.046 \text{ e}^-/\text{\AA}^3$. The electron density of myoglobin has been given in the literature as $0.45 \text{ e}^-/\text{\AA}^3$ (Stuhrmann, 1980). Therefore the ratio of I_0 in 8 M urea to that of I_0 in 0 M urea is about $[(0.45-0.411)/(0.45-0.379)]^2 = 0.30$, and we should expect roughly a 70% decrease in I_0 .

In Figure 3.C.2b I_0 is plotted as a function of urea concentration. The better data set (collected without ethylene glycol) has been normalized by the value at 0 M urea of a quadratic fit to the data, and the remaining data were visually scaled to the former. In addition, the behavior of I_0 predicted by the above treatment is shown. The good agreement between the expected and observed decrease in I_0 upon urea denaturation indicates that once solvent contrast effects are accounted for, no change or only a small positive change should be observed in I_0 upon refolding. This conclusion is further supported by the results of the thermal denaturation of ribonuclease A, where no significant change in I_0 is observed as a result of unfolding in a constant contrast environment.

In analyzing the time-resolved data, it was found that as in the case of ribonuclease A the accumulated statistics were not sufficient to produce reliable plots of R_g as a function of time. Two representative plots are shown in Figure 3.C.3 to illustrate the high noise level. In contrast, plots of I_0 versus time were much cleaner. This is probably because the slope of a Guinier plot is more sensitive to noise than the intercept. The time-courses of I_0 from all conditions without ethylene glycol are shown in Figure 3.C.4 (because of the apparent tendency of concentrated unfolded myoglobin to aggregate in the presence of ethylene glycol, data from ethylene glycol containing samples were not treated further). As can be seen in these plots, unlike in the case of ribonuclease A, a definite signal was observed. This change is also clearly visible in a plot of the first time frame versus the last time frame of data collection, as demonstrated for one case in Figure 3.C.5.

In contrast, the results from control experiments in which unfolded protein was diluted without initiating refolding do not show any obvious time-dependent signal. The best control data were collected after the above experiments (see below), and are presented in Figure 3.C.6 as plots of I_0 versus time. These controls rule out the possibility that the large signal in the folding experiments is caused by the dissociation, upon dilution, of protein aggregates *already present* at the high protein concentrations (up to 70 g/l) of the initial solutions. The dissociation of such aggregates should have been observed even upon dilution without refolding. It is also possible to rule out the presence of initial aggregates which dissociate only upon dilution *and* refolding. Examination of the (constant) value of I_0 in the controls reveals that it is around one third that of the final I_0 value in the refolding experiments. The equilibrium I_0 data have established about a 70% decrease in I_0 upon unfolding due to solvent contrast effects. Thus after correction for solvent contrast, I_0 in the controls equals I_0 of the final state of the refolding experiments, proving that the initial state has the same molecular weight as the final state, and is therefore not dimerized.

Unfortunately, the accumulated statistics in the control data were not sufficient to determine the time-course of R_g . Static data collected after the control mixing, however, verify that the protein was indeed unfolded, with an R_g of around 30 Å. When these data are compared with static data collected at the end of refolding experiments (Figure 3.C.7), it can again be seen that I_0 of the initial condition is about the same as I_0 of the final condition, after a factor of 2 correction for the change in solvent contrast.

In order to estimate the R_g values during the initial time frames of the refolding experiments, the five first frames of each time-resolved data set were averaged and fit. The resulting values were again averaged (they were reasonably similar) to give a value of 32 ± 3 Å. This value was verified in later experiments (see below) and is consistent with that of either a small oligomer or an unfolded state, but not with that of the native state (18 ± 1 Å). The R_g from the final time-frames (20 ± 2 Å) is in reasonable agreement with the native state value, although the existence of a second smaller and slower kinetic phase is difficult to rule out due to the noise level. The observed change in R_g is consistent with both a monomer-refolding process and a dimer- or multimer-dissociation process.

It has been shown that Kratky plots (in which K^2I is plotted *versus* K) can be used to distinguish between coil-like and globular scatterers (Kataoka, 1993, and Ramakrishnan, 1991). Kratky plots of static data from unfolded myoglobin have a shape characteristic of coil-like scatterers, and different from the shape of Kratky plots of data from native myoglobin or other globular scatterers (Figure 3.C.8). In order to obtain information about the compactness of myoglobin during the course of folding, Kratky plots of initial and final data frames were examined. Some of these plots are shown in Figure 3.C.9. In all cases, initial time-frame Kratky plots do not show the coil-like characteristic shape of unfolded myoglobin. Instead, they show a shape characteristic of a globular scatterer. This directly implies that immediately after the mixing of the solutions, the protein is no longer an unfolded monomer. This result means that the observed change in R_g can not be caused by a monomer-refolding process. This leads to the necessary conclusion that the signal is generated by a dimer- or multimer-dissociation process.

This conclusion is further supported by the observation of a large decrease in I_0 concomitant with the decrease in R_g . After the initial mixing dead-time (of 10 ms), the solvent electron density around the protein no longer changes, and if the sample were monomeric at all times, no changes in I_0 should be observed, as pointed out above in the discussion of the equilibrium data. Any changes in I_0 must be a result of a change in the total (excess) electron density of the scatterer, and hence of its mass. Because the observed change in I_0 is rather large, it seems most likely that it is caused by a dissociation (since I_0 decreases) of protein oligomers.

It should be pointed out that there is a possible alternative explanation for the Kratky plot and I_0 data. It is possible that the high initial value of I_0 is due to a large amount of urea remaining bound to the unfolded protein throughout the mixing process. The urea density

'inside' the protein would then be significantly higher than the 'outside' post-mixing urea density, and could contribute much more to I_0 and R_g under these conditions than it would in the equilibrium studies, where the 'outside' urea density could not be far different from that of the bound urea. As the protein folds, the urea would be displaced leading to a loss of both mass and size. In addition, the urea could 'fill in' the spaces in the unfolded protein, making it appear globular rather than coil-like. Although difficult to exclude experimentally, this scenario seems implausible because it is unlikely that the urea would remain bound to the protein for so long (exchange rates for solvent molecules bound to proteins are quite high) and because the observed signals seem far too slow to correspond to the compaction of the protein.

Because no protein aggregates or multimers are observed in either the initial or final conditions of the experiments, the observed multimer dissociation implies that said multimers must form during the course of the experiment (apparently within the mixing dead-time). Neither the initial (unfolded) nor the final (folded) state associate, and therefore the presence of the multimers implies **the existence of a transient association-prone intermediate on the refolding pathway of myoglobin**. This is the major result of these studies.

Discussion:

A transition from oligomer to monomer does not involve a change in electron density, and hence the ratio of I_0 for an oligomer of n monomers to I_0 for n individual monomers is $(nV)^2/nV^2$ which is just n . Our results show that, within the rather large error margin, I_0 is observed to decrease by a factor close to 2 (1.87 ± 0.08 on average, see Table 3.C.1). This leads to the further hypothesis that the signal observed in I_0 is engendered by the dissociation of the entire intermediate population from dimers to monomers. If this were the case, the concentration of dimer in the sample at any time, normalized by the initial protein concentration, would be

$$[D](t)/[M](0) = (I_0(t)/I_0(\infty) - 1)/2$$

where $[D]$ represents the dimer concentration and $[M](0)$ the initial (before mixing) monomer concentration. Additional support for this hypothesis is provided by the fact that the time-course of I_0 for all samples, although noisy, is well fit by a single exponential, as would be proper for a first order dissociation process. The measured rate

constant is the apparent dissociation rate constant, k_{app} . Despite the aesthetic appeal of the complete dimerization hypothesis, it should be emphasized that the quality of the data is not sufficient to exclude the possibility of smaller amounts of trimers and higher order oligomers, and a second or higher order dissociation process.

Many possible folding pathways may be consistent with the above result, but only two possibilities, which seem most probable, are discussed here. In the first pathway the protein undergoes a conformational change while in the dimerized state, and then proceeds to dissociate and continue refolding. This pathway illustrated schematically in Figure 3.C.10a. U and N represent the unfolded and native state respectively, I represents the intermediate whose existence we have inferred, and I' represents I after a further conformational change.

Referring to the figure, we make the assumption that $k_i \gg k_a, k_n$, *i.e.* that the initial concentration of I equals the initial protein concentration. Because dimerization is observed to be rapid and complete, we also assume that $k_a \gg k_n$. If the above assumptions hold, then either k' or k_d are rate determining in this pathway, and if they are well separated, k_{app} , the rate constant measured for the decay of I_0 , will likely equal the lesser of the two. If $k_d \gg k'$ then k_{app} represents the rate of conversion from I to I'. If $k' \gg k_d$ then k_{app} represents the rate of dissociation of the I' dimer.

A second possible pathway is presented in Figure 3.C.10b. Here no conformational change takes place in the dimerized state. We make the same assumptions about k_i and k_a as above. Then if $k_d \gg k_n$, I and its dimer are in rapid equilibrium and the transition from I to N is the rate determining step for dimer dissociation.

If instead k_n is of the order of or greater than k_d then both k_n and k_d will contribute to the reaction rate, and k_{app} should be proportional to their product.

Unfortunately, further information is required to discriminate between the above pathways and interpretations. Note that other possible intermediates on the above pathways are implicitly included in U and N, because our probe is not able to distinguish among them. They are either too short-lived or do not significantly differ in size or shape from U or N.

Under the above conditions the dimerizing species is not long-lived enough to measure its association equilibrium constant. It is possible, however, that this intermediate is the same one detected by Phillips in a thermal denaturation study of myoglobin (Phillips, 1988), and that it may be possible to study its properties under those conditions, where a much longer lifetime was observed. It is also possible that this intermediate is similar to an equilibrium molten globule intermediate (see below) which is known to exist for apomyoglobin, and is known to be prone to associate (Hughson, 1990). In the meanwhile, the dimerization of the protein appears to be completed within the dead-time of the rapid-mixer, which is 10 ms. Using this observation one can set a lower limit of about $0.1 \text{ ms}^{-1}\text{mM}^{-1}$ or $10^5 \text{ s}^{-1}\text{M}^{-1}$ for the association rate constant, k_a . Assuming that the measured apparent rate constant k_{app} is of the order of the dissociation rate constant, the equilibrium association constant K equals k_a/k_{app} and is in the vicinity of 10^5 M^{-1} under our various conditions. These values of k_a and K are similar in magnitude to values reported in the literature for similar processes in carbonic anhydrase B, bovine growth hormone, and lactate dehydrogenase (Cleland, 1992, Brems, 1988, and Kiefhaber, 1991). Although this similarity is encouraging, it should be emphasized that the numeric K value obtained is only a rough estimate.

Conclusions:

These results add to a growing body of evidence that transient association may be a commonly occurring process on protein folding pathways (Cleland, 1992, Brems, 1988, and Mitraki, 1989). Most notably, bovine carbonic anhydrase B was recently shown, using high-performance liquid chromatography, to undergo a dimerization process during refolding (Cleland, 1992).

It is proposed that the dimerizing intermediate is globular and that much of the native myoglobin helical structure has formed. It would then fit the description of a molten globule state. Molten globule states are compact globular intermediates with secondary structure and fluctuating tertiary structure, and are suspected to be a common motif in protein folding pathways (Ptitsyn, 1987). Globularity is supported by the Kratky plot analysis, and the fast formation of secondary structure has been observed in CD studies by K. Chiba (private communication). This is also consistent with results recently reported for a kinetic apomyoglobin folding intermediate (Jennings, 1993). It can be speculated that in this intermediate the helices have not yet arranged themselves correctly. Because the helices are highly amphipathic, this could leave the extremely hydrophobic interior

regions of the helices exposed, and inter-protein helix interactions could lead to association and an accompanying stabilization of the intermediate state. A reorganization of the helices might be responsible for the ensuing dissociation.

Further Myoglobin Studies:

This section describes preliminary results from attempts to learn more about the thermodynamic properties of the myoglobin dimer, and to explore the ability of the stopped-flow time-resolved SAXS technique to resolve changes in the parameters observed under different conditions. Further experiments were performed to study the temperature dependence of the myoglobin dissociation process, and an attempt was also made to look for any concentration dependence of the apparent dissociation rate constant. Unfortunately, only a narrow range of concentrations could be studied because of the decrease in signal strength with decreasing concentration (the initial experiments were performed near the limit of solubility of the protein).

Experimental Details, Specific:

In January of 1993 at SSRL, the ruby-ball mixer was used to measure time-resolved data from the 1:6 mixing of 64 g/l myoglobin in 8 M urea with 1 M urea buffer at 20, 16, 12, and 8 °C, from 1:6 mixing of 70, 64, 58, and 52 g/l myoglobin in 8 M urea with 1 M urea at 10 °C, and from 1:6 mixing of 64 g/l myoglobin in 8 M urea with 8 M urea at 20 °C. Data were collected in cycles of 10 x 50, 75, or 100 ms, 10 x 150, 225, or 300 ms, and 10 x 300, 450, or 600 ms. Between 250 and 500 cycles were collected for folding conditions, and 175 cycles were collected for non-folding (control) conditions. The sample-to-detector distance was 254 cm and the photon energy was 8980 eV.

In March of 1993, at the Photon Factory, static data were recorded from 10 g/l myoglobin in 0, 2, and 4 through 8 M (in half molar increments) urea, using a 25- μ m thick mica-window cell. The sample-to-detector distance was 251 cm and the photon energy 8250 eV. Time-resolved data were collected from 70 g/l myoglobin in 8 M urea mixed at a 1:6 ratio with 1 M urea at 20, 16, 12, 10, and 8 °C. A control was also collected for mixing with 8 M urea at 12 °C. Data were collected in 60 to 300 cycles of 94 x 50, 75, or 100 ms for folding conditions, and in 10 cycles of 94 x 75 ms for non-folding (control) conditions.

Analysis and Results:

All of the data were normalized, background subtracted, and calibrated. R_g and I_0 were determined in the same fashion and with the same fitting range as for the previous

myoglobin data, and the same decrease as before was observed in both parameters. An attempt was also made to examine the effect of the chosen fitting range. In the case of most of the data, a significant increase in both R_g and I_0 was observed in the unfolded or dimerized state upon use of lower angle fitting ranges. Some increase was also occasionally observed in R_g and I_0 from the native state. This is surprising, because the Guinier range of the smaller native state should extend further than that of either an unfolded or dimerized state. In addition, the magnitude of the observed increase varied considerably between data sets, indicating that the data in that range were not reliable. As mentioned in Chapter 1, this is probably a result of parasitic scattering at the lowest angles observed. The fitting range which was chosen yields results which are in reasonable agreement between data sets, indicating the data is relatively free of parasitic scattering effects. An evaluation of the lower angle fits to the 'best' data sets indicates that a maximum increase of about 20% may be expected in R_g and I_0 of the unfolded or associated protein. No such increase is expected for the native protein. This does not change any of the conclusions presented here. In fact, a reasonable increase of the initial-to-final I_0 ratio in the time-resolved measurements would strengthen the complete dimerization model.

Because of the larger number of accumulations in these experiments, particularly for the SSRL data, the R_g time-courses were somewhat cleaner, as can be seen from the example shown in Figure 3.C.11a. These verify that the initial R_g value is around 30 Å and the final value is around 20 Å. The R_g time-courses from the Photon Factory data were noisier but the best of these (Figure 3.C.11b) is also in reasonable agreement with the quoted initial and final R_g values. The time-courses of R_g follow those in the I_0 data, but because I_0 is a linear function of the dimer and monomer concentrations, and R_g is not, the kinetic analysis was again performed using the I_0 data. The time-courses of I_0 , along with fits and fit results, are presented for all experiments in Figure 3.C.12 (January data), Figure 3.C.13 (March data), and Table 3.C.2. The data were again well fit by a single exponential. The rate constants obtained from the two different experiments are in reasonably good agreement.

There remain sizeable (up to 48% in one case) deviations from a factor of 2 in the initial-to-final ratio of I_0 . The average of the ratios from the all of the temperature and concentration dependence measurements is 1.78 ± 0.04 , significantly lower than the anticipated value of 2. It is possible that the 20% shortfall arises from the choice of the fitting range, as discussed above. Alternately, it may be that the difference is real. In this

case, the complete dimerization-dissociation hypothesis would be wrong, and should probably be replaced either by a partial dimerization model, or by a model in which complete dimerization is followed by bi-phasic dissociation, with a slower small amplitude (about 10%) second phase. A slow second phase could be responsible for the slightly high final value of R_g , as mentioned previously. A bi-phasic folding model is illustrated in the case of cytochrome-*c* in the next chapter. Higher quality data at lower scattering angles will be required to discriminate between the various possibilities. It is important to keep in mind, however, that the argument demonstrating the existence of an associative kinetic (molten globule) intermediate remains valid under any of these scenarios.

It is interesting to note that in several of the data sets, the first data frame shows a much lower I_0 and R_g value. Although one should be very wary of overinterpreting data from one (short) data frame, this may be an observation of the intermediate prior to (or during) dimerization. If verified, this would provide concrete evidence that the dimers are in fact transient (not present in the initial sample). It also indicates that the association process may not be completely lost in the mixing dead-time. An attempt to collect better data at these very fast times (perhaps using the method outlined in the next chapter) may provide interesting results.

It is possible to fit the temperature dependence of observed rate constant to an Arrhenius equation, $k = Ae^{-E/RT}$ (Wayne, 1969). The results are plotted in Figure 3.C.14. The value of the activation energy E is 25 kcal/mole. It is not clear how to interpret this value. It may represent the energy barrier associated with the I to I' transition suggested in the first pathway model (see above). Perhaps this activation energy is increased by the dimerizing interactions. If we assume that in the dimer I is stabilized relative to I' by the dissociation energy ΔG , then the activation energy for conversion of monomeric I to I' could be $E - \Delta G$, which (using ΔG from below) is around 18 kcal/mole. Although it may be a coincidence, it is worth noting that this energy is close to that predicted for the energy barrier between *cis* and *trans* proline isomers. Myoglobin has four proline side-chains.

In spite of the possible observation of the association process, there still does not appear to be any reasonable means of extracting an association rate constant from these data, and the approximation of $10^5 \text{ s}^{-1}\text{M}^{-1}$ used in the preceding section is used here as well, without any estimate of its temperature dependence. The free energies corresponding to

the equilibrium dissociation constant calculated at the different temperatures under these assumptions and approximations are tabulated in Table 3.C.2. It can be seen that they are comparable to the free energies of unfolding of typical proteins (around 10 kcal/mole) and this may explain the long lifetime of the dimer.

Finally, it is possible, using the van't Hoff equation, $\Delta H = -Rd(\ln K)/d(1/T)$, to determine approximately the enthalpy of dissociation from the temperature dependence of the equilibrium constant. $-R\ln K$ is plotted *versus* temperature in Figure 3.C.15. Rather than attempting a numerical differentiation with so few data points, the data were reasonably fit with a straight line. The slope, 36 ± 3 kcal/mole, is a measure of the (positive) enthalpy of dissociation. The entropy of dissociation can also be evaluated, and comes out to 0.1 ± 0.01 kcal/mole/°K, showing that the entropy increases upon dissociation. Note that as long as the association constant k_a is assumed to be temperature independent, the above values are independent of the actual value used for k_a , since $d(\ln K) = d(\ln k_d - \ln k_a) = d(\ln k_d)$.

The I_0 time-courses from the concentration study are shown in Figure 3.C.16, and the fit results are in Table 3.C.2. No significant concentration dependence of either the initial-to-final I_0 ratio or of the apparent rate constant is observed. In order to observe concentration dependent effects due to the second order association process it is probably necessary to change the concentration more drastically, which is unfortunately not possible under the current experimental limitations.

Conclusions:

The further studies described do not greatly extend or alter the results of the initial myoglobin studies. They do establish that the observed initial-to-final I_0 ratio is in fact less than 2, but this does not necessarily rule out the complete dimerization model, since the true ratio could be somewhat higher. The other explanations offered should also be considered, and in particular the possibility of a slow final folding phase should be investigated, especially in light of the cytochrome-*c* results presented below. The well-resolved and reproducible rate constants measured in the temperature dependence experiments illustrate the capabilities of the time-resolved SAXS technique. Unfortunately, the interpretation of this temperature dependence is quite limited in scope at this stage. It does suggest the possible involvement of a proline isomerization event in the folding process. The limitation imposed on the technique by the narrow range of

(very high) usable sample concentrations is evident in the concentration study. This limitation could be eased by a more intense source. One method of achieving this with current machines is presented in the next chapter.

In none of the myoglobin data is there any observation of the compaction of the protein. The earliest observed times vary from condition to condition, but it seems safe to set an upper bound of around 50 ms for the compaction time. Unlike in the case of ribonuclease A, it is not the lack of any difference between the initial and final data frames that leads to this conclusion. Rather, it is the fact that Kratky plots of even the earliest data frames show that the protein is already globular, not coil-like. It seems reasonable to expect that the compaction processes of most small globular proteins occur on similar time scales, and it is therefore not too surprising that the observation of compaction was not possible in the case of myoglobin either. The observation under similar conditions by circular dichroism of secondary structure formation on the 100 to 200 ms time scale illustrates that in myoglobin, as in ribonuclease A, a significant amount of secondary structure formation occurs after the compaction event.

**Table 3.C.1:
Parameters from the Time-Course of the Forward
Scattering from Myoglobin During Refolding**

T (°C)	Final [Urea] (M)	$k_{app}(S^{-1})$	$I_0(t_i)/I_0(t_f)$
15±0.1	1.0	2.19±.43	1.86±0.24
15±0.1	1.9	1.27±.18	1.84±0.18
15±0.1	2.7	0.31±.04	1.63±0.12
15±0.1	1.1	2.43±.47	2.08±0.23
8±0.1	1.0	0.73±.11	1.93±0.14

For all samples the myoglobin concentration was 60 g/l before mixing and 0.86 g/l after mixing. The initial urea concentration was 7 M in all samples except for the fourth, where it was 8 M. The apparent dissociation rate constant, k_{app} , and the ratio of the forward scattering at the initial and final time-frames, $I_0(t_i)$ and $I_0(t_f)$, were determined from single exponential fits to the data. Fits and errors were calculated using the program Igor (WaveMetrics, Lake Oswego, OR). The ratio $I_0(t_i)/I_0(t_f)$ should be 2 if the protein is completely dimerized within the mixing dead-time.

**Table 3.C.2:
Parameters from Temperature and Concentration
Studies of the Refolding of Myoglobin**

T (°C)	initial [mb] (g/l)	$k_{app}(S^{-1})$	ΔG (kcal/mole)	$I_0(t_i)/I_0(t_f)$
20±0.1	64	2.43±.14	6.19	1.81±0.07
16±0.1	64	1.17±.09	6.52	1.74±0.06
12±0.1	64	.580±.054	6.83	1.70±0.07
8±0.1	64	.147±.039	7.50	1.81±0.35
20±0.1	70	1.90±.21	6.33	1.52±0.11
16±0.1	70	1.20±.10	6.51	1.71±0.07
12±0.1	70	.400±.039	7.04	1.92±0.08
10±0.1	70	.260±.032	7.23	1.93±0.12
8±0.1	70	.151±.009	7.48	2.02±0.07
10±0.1	70	.225±.021	7.31	1.84±0.10
10±0.1	64	.307±.023	7.14	1.79±0.06
10±0.1	58	.301±.029	7.15	1.67±0.08
10±0.1	52	.392±.035	7.00	1.73±0.07

For all samples the initial and final urea concentrations were 8 M and 2 M, respectively. The SSRL temperature study data are presented first, followed by the Photon Factory temperature study data, followed by the SSRL concentration study data. The apparent dissociation rate constant, k_{app} , and the ratio of the forward scattering at the initial and final time-frames, $I_0(t_i)$ and $I_0(t_f)$, were determined as in Table 3.C.1. ΔG was determined as $-RT\ln K$, where K , the dissociation equilibrium constant, was estimated as $k_{app} \times 10^{-5}$.

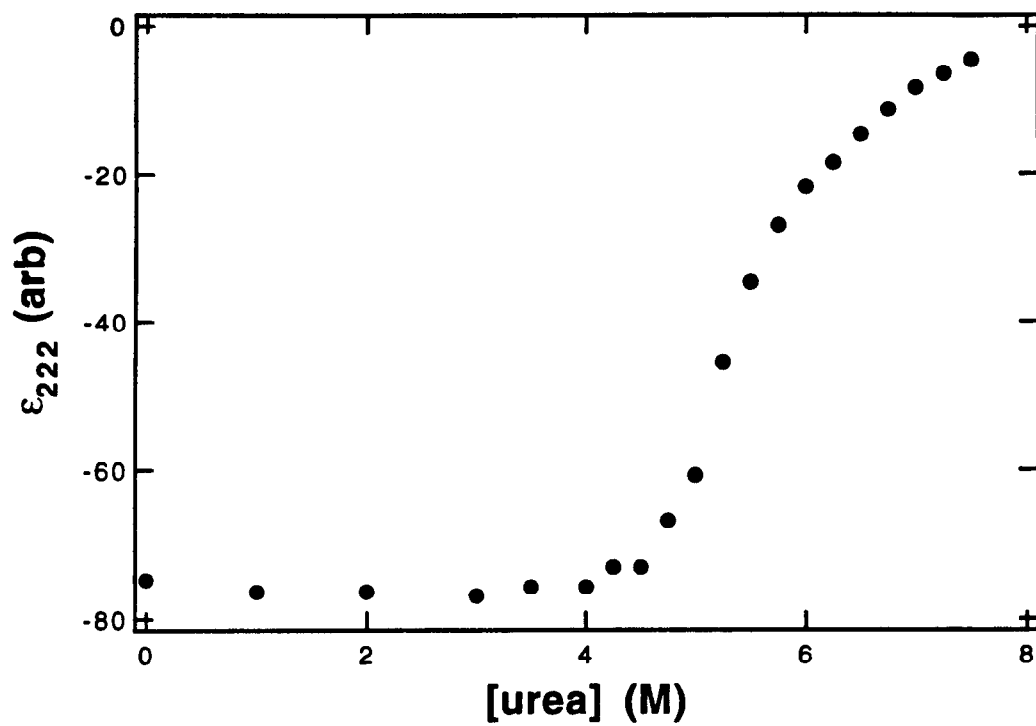


Figure 3.C.1: Circular dichroism from the equilibrium unfolding of myoglobin at 222 nm. Sample conditions were 25 μM myoglobin in 0.1 M NaCl, 40 mM NaCN, 50 mM bis-tris-propane, pH 9.1.

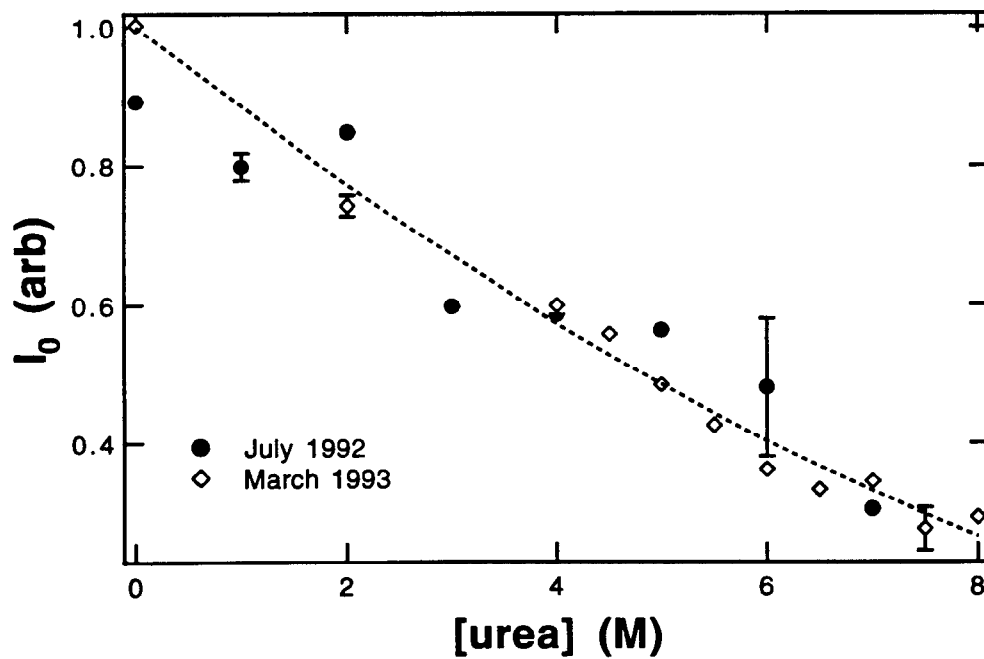
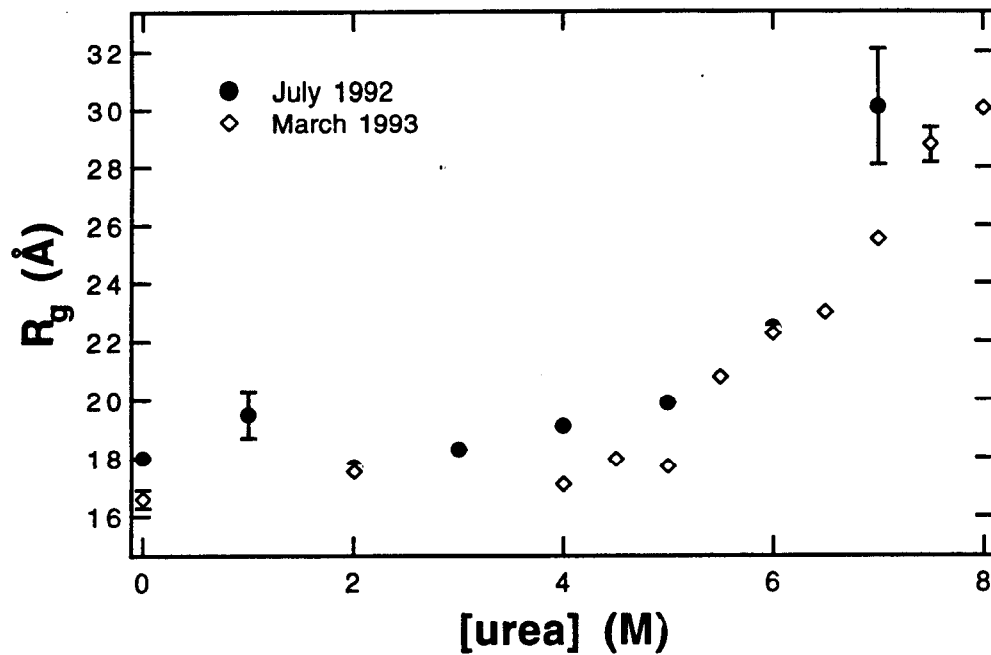


Figure 3.C.2: SAXS data from the equilibrium unfolding of myoglobin. a) R_g as a function of urea concentration shows a transition around 6 M urea, about 1 M urea higher than the transition observed in the CD data. b) I_0 shows a decrease with increasing urea concentration, caused entirely or mostly by the increasing solvent electron density. The dashed line is the predicted behavior.

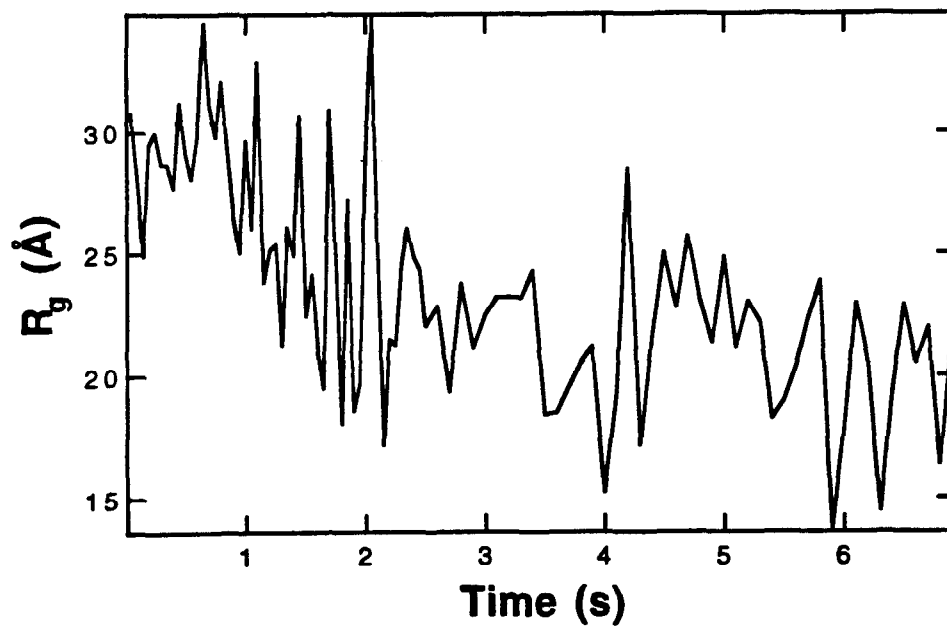
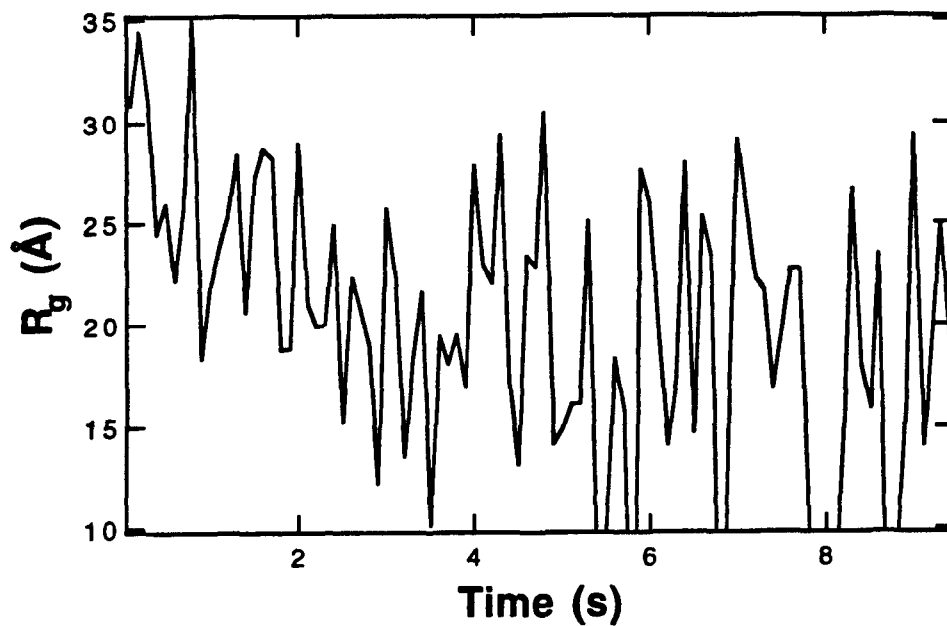


Figure 3.C.3: Time-course of R_g during the refolding of myoglobin. Although a change is evident, the noise is too high to make any further interpretation. The initial conditions are 60 g/l myoglobin in 7 M urea, 0.1 M NaCl, 40 mM NaCN, 50 mM bis-tris-propane, pH 9.1. The final urea concentrations are 1.9 M (upper) and 1 M (lower). The temperature was 15 °C (upper) and 8 °C (lower).

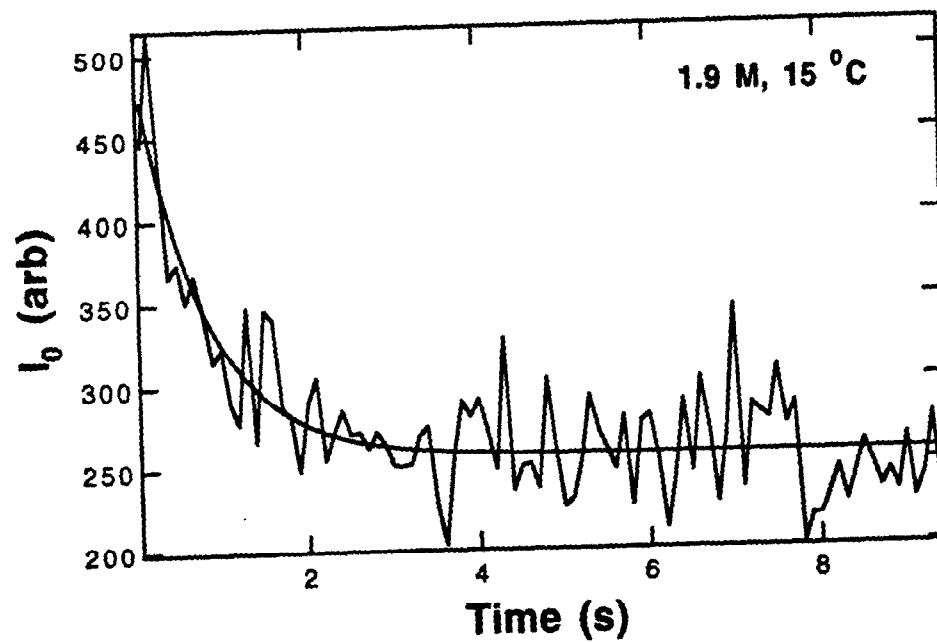
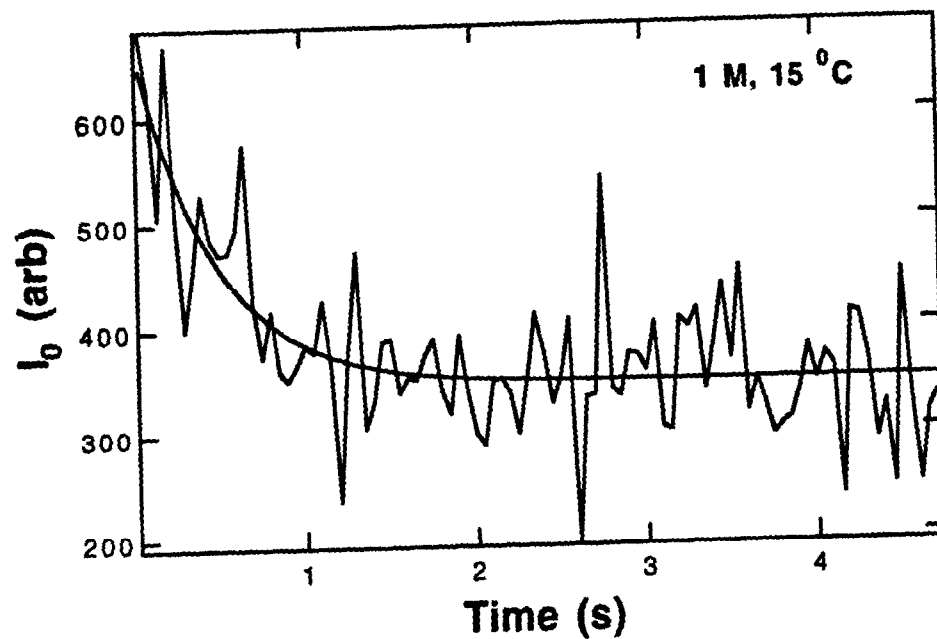
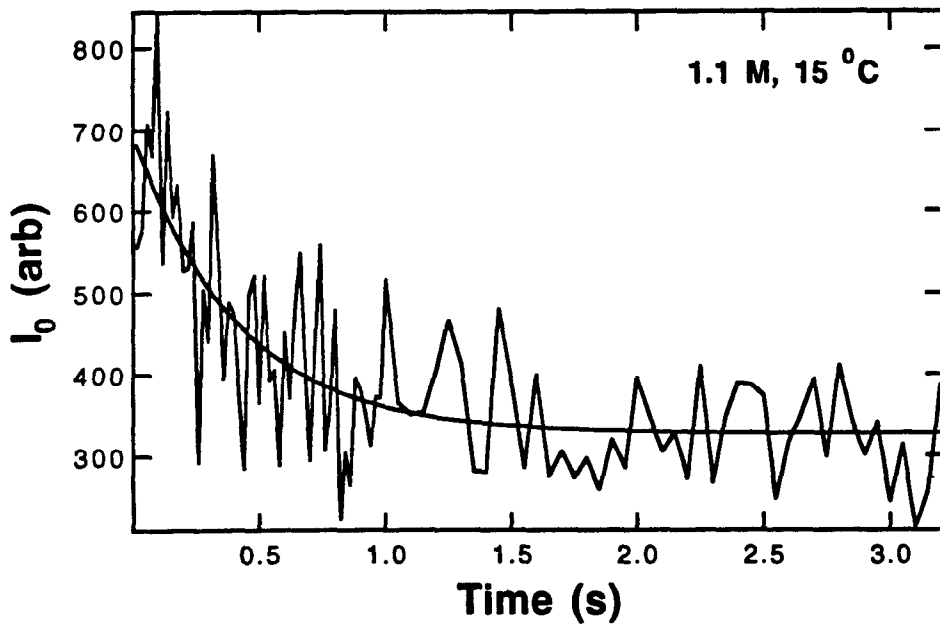
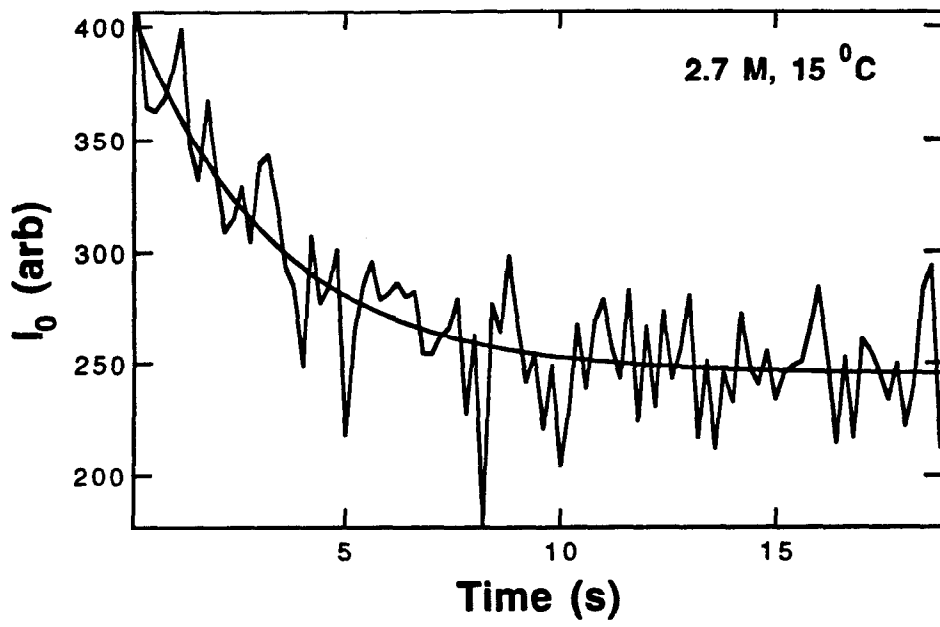
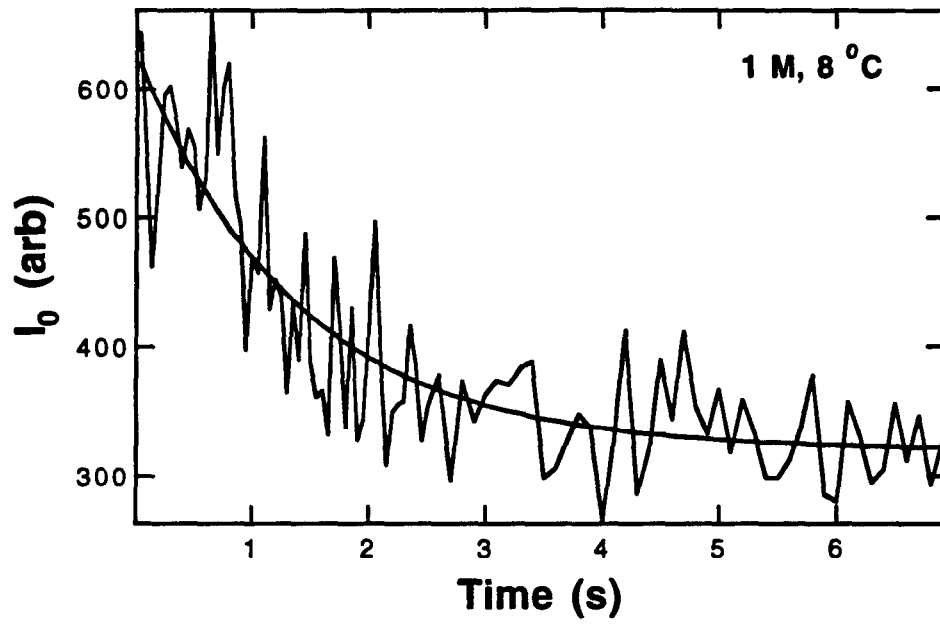


Figure 3.C.4: Time-courses of I_0 during the refolding of myoglobin, with single exponential fits. A clear change is visible, and the magnitude of the change is nearly a factor of 2 in all cases. The buffer conditions are as in Figure 3.C.3. The initial urea concentration is 7 M in all but the fourth plot, for which it is 8 M. The final urea concentrations are, in order of presentation, 1, 1.9, 2.7, 1.1, and 1 M. The temperature, in the same order is 15, 15, 15, 15, and 8 °C.





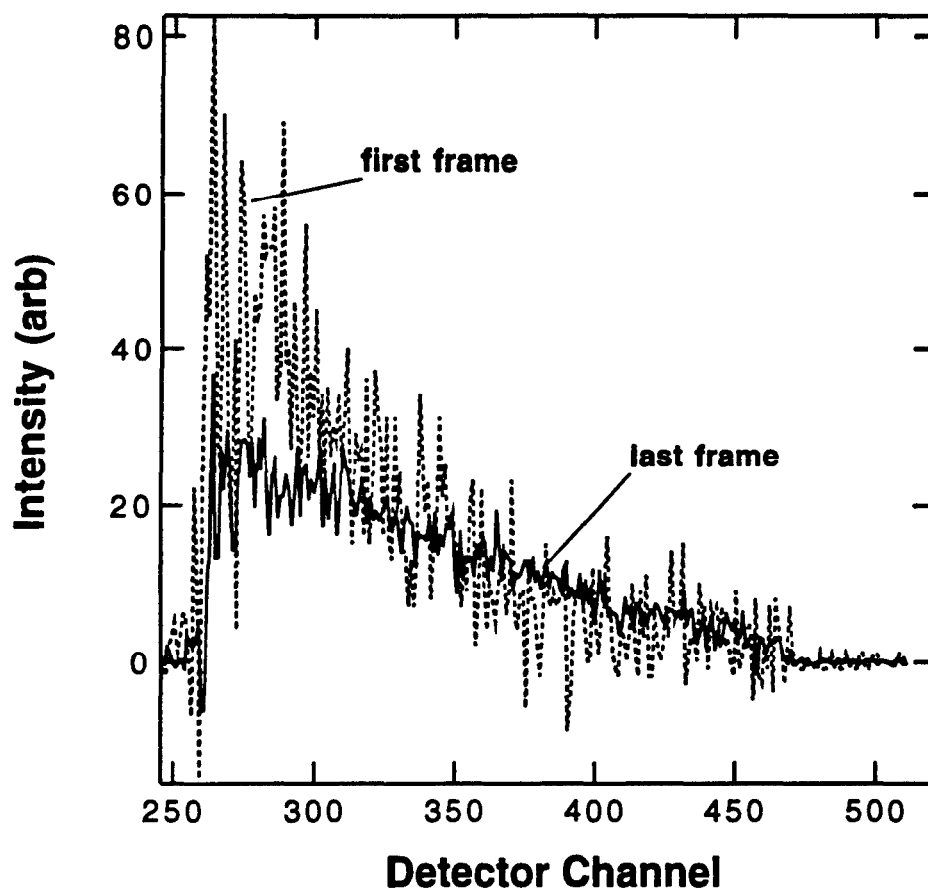


Figure 3.C.5: The first and last time frame from the 1.9 M final urea concentration refolding experiment. The change in I_0 is clearly visible.

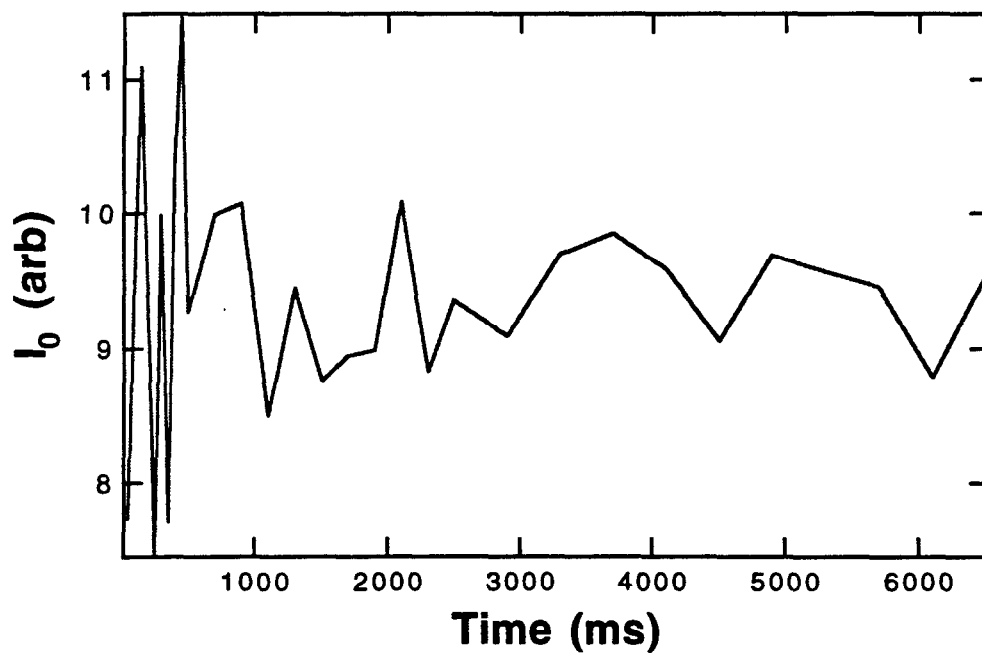
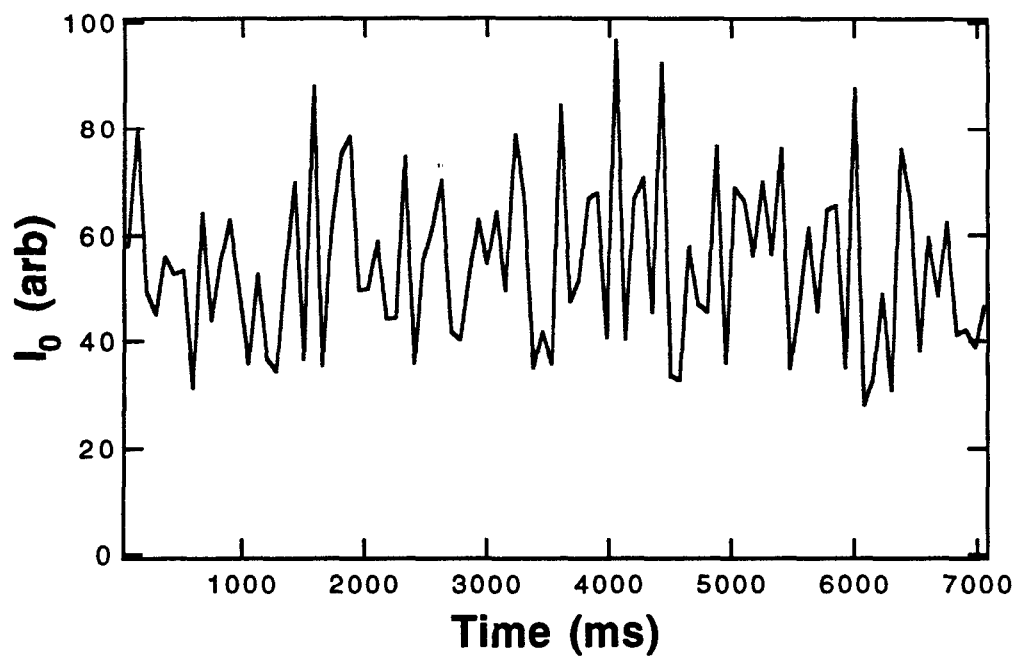


Figure 3.C.6: Time-courses of I_0 during control experiments in which myoglobin in 8 M urea was mixed with 8 M urea, so that no refolding took place. The type of signal evident in the refolding experiments is clearly absent in these data, despite the high noise level. The final values of I_0 from corresponding folding experiments (see Figures 3.C.12 and 3.C.13) are around 160 for the upper plot, and around 25 for the lower plot, showing that I_0 for the initial condition is about the same as that for the final condition, after correction for solvent contrast.

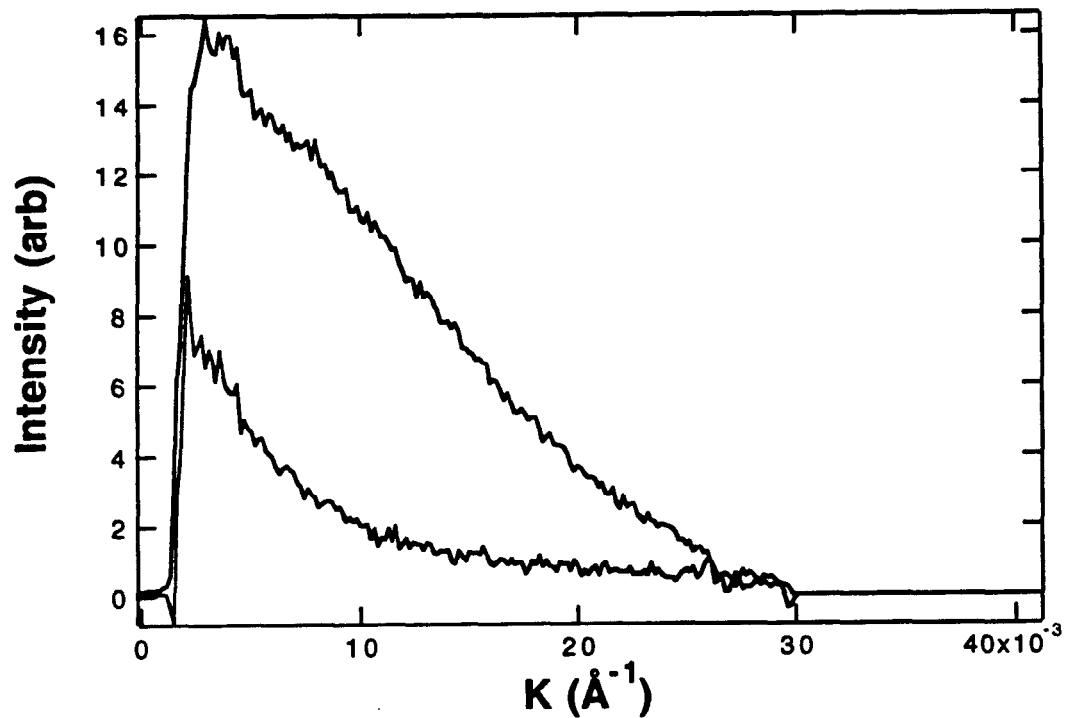


Figure 3.C.7: A comparison of the scattering pattern collected after completion of a refolding experiment (upper plot) with that collected after completion of a control experiment (lower plot). I_0 unfolded is just around one half of I_0 refolded, as predicted from the solvent contrast calculation. This provides crucial evidence that at the beginning of the refolding experiment, the molecular weight of the scatterer is the same as it is at the end, meaning that the protein is not associated to begin with, and the association observed during refolding is transient.

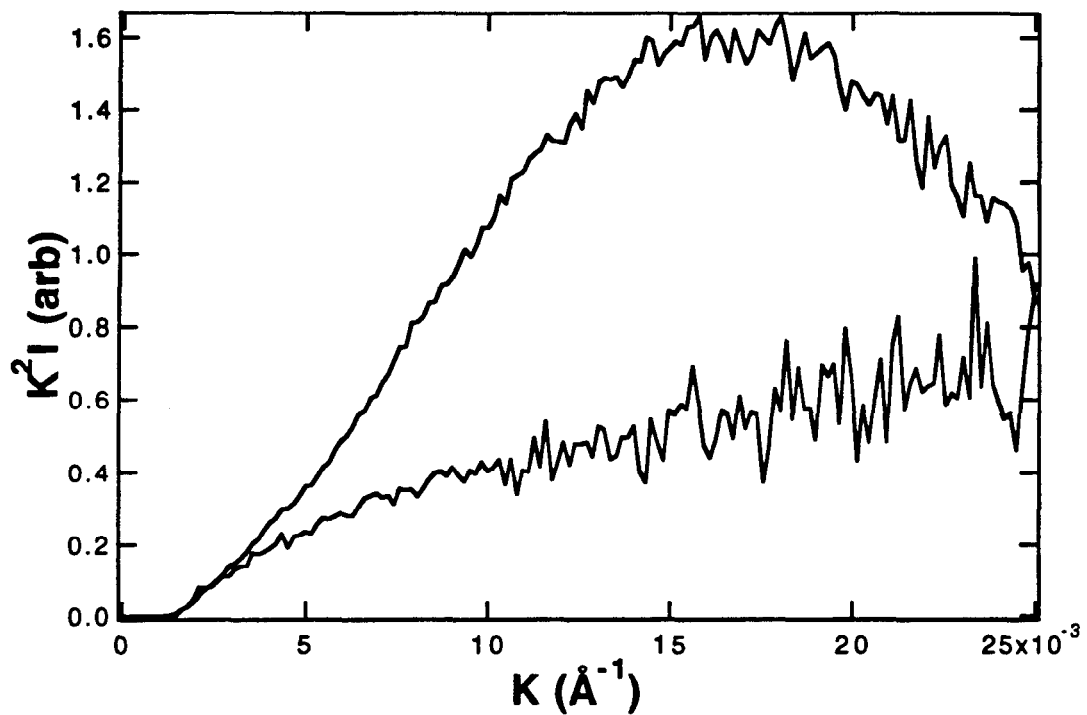


Figure 3.C.8: Kratky plots from urea-unfolded and native myoglobin. The data are the same as those in Figure 3.C.7, but plotted as $K^2 I$ versus K . The broad peak in the native data indicates that the scattering molecule is compact. The persistent linear rise of the unfolded data is characteristic of a random coil molecule.

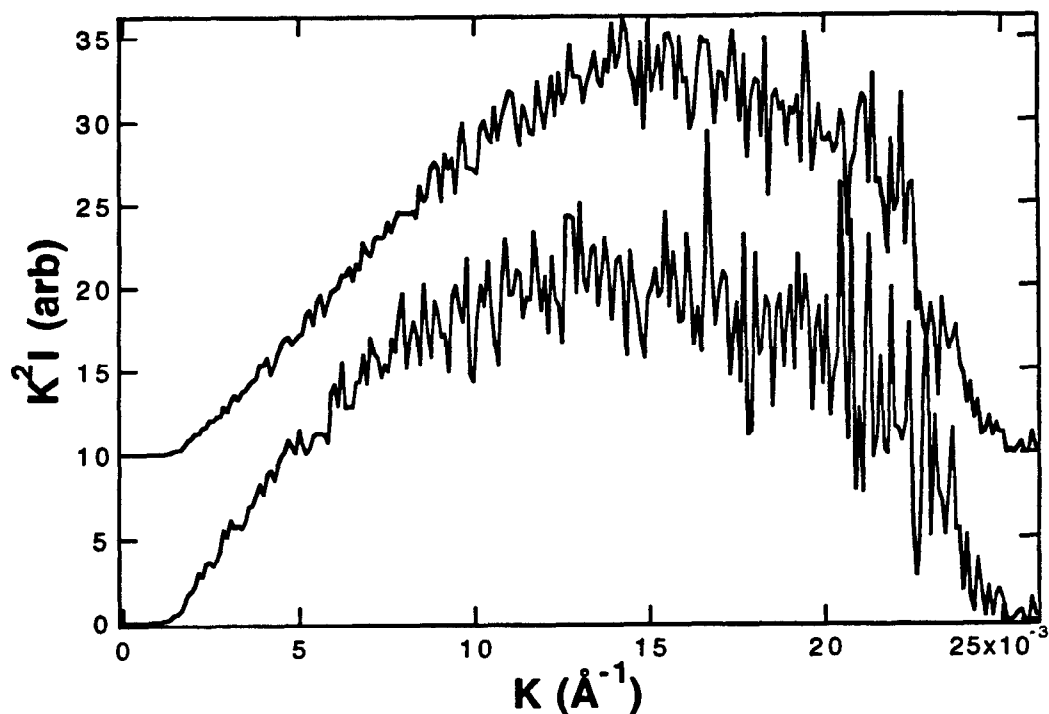


Figure 3.C.9: Kratky plots averaged over the first 0.8 seconds (lower) and the last 4.4 seconds (upper) after mixing. The plot for the last 4.4 seconds has been shifted arbitrarily along the ordinate for clarity of presentation. The sample conditions were 7.0 M initial and 1.0 M final urea concentration, 8.6 g/l final protein concentration and 8 °C. Averaging was necessary in order to obtain reasonable statistics, but the averaged plots overlaid quite well onto the unaveraged plots of the first time-frame and last time-frame, respectively. In addition, later data confirm this result even without frame averaging. The time-course of I_0 is shown in Figure 3.C.4. The total exposure time was 6.9 seconds, and the determined reaction rate constant was 0.73 s^{-1} (half-life 0.95 seconds). The lack of the characteristic persistent linear rise in the initial time-frame plot indicates that the protein loses its random coil character within the mixing dead-time. The same result was observed under all other conditions.

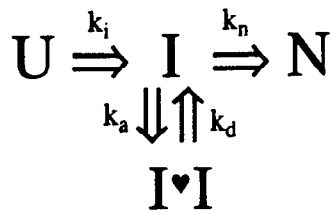
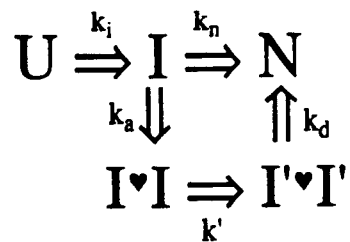


Figure 3.C.10: Two possible folding pathways consistent with the proposed transient dimerization of myoglobin during refolding. **a)** The unfolded state U folds quickly to the (molten globule) intermediate I which quickly dimerizes. A conformational change to I' causes a dissociation of the dimer, followed by the completion of folding. **b)** Here the final folding step only takes place from monomeric I, and is in competition with the dimerization process.

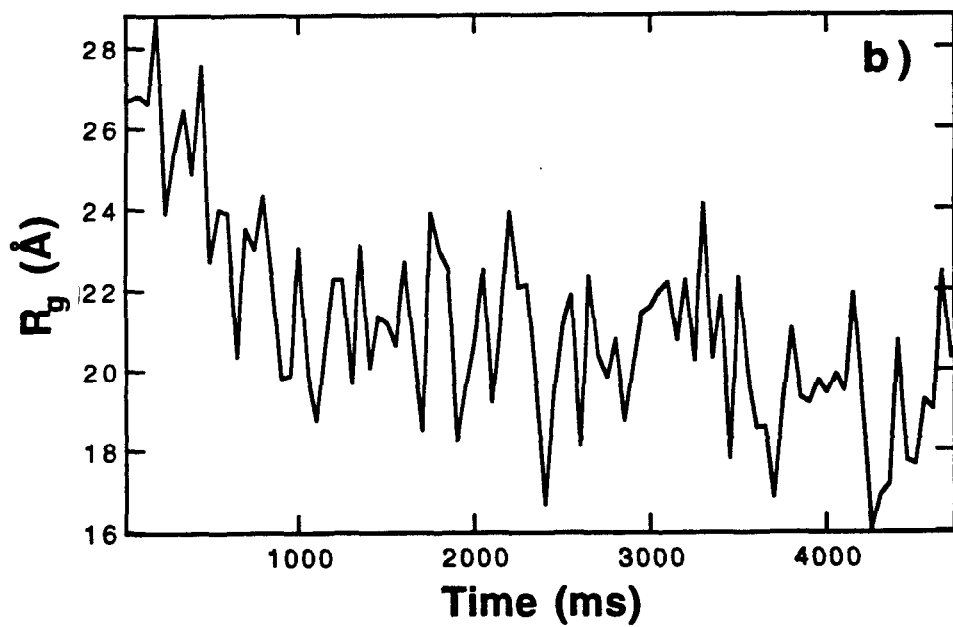
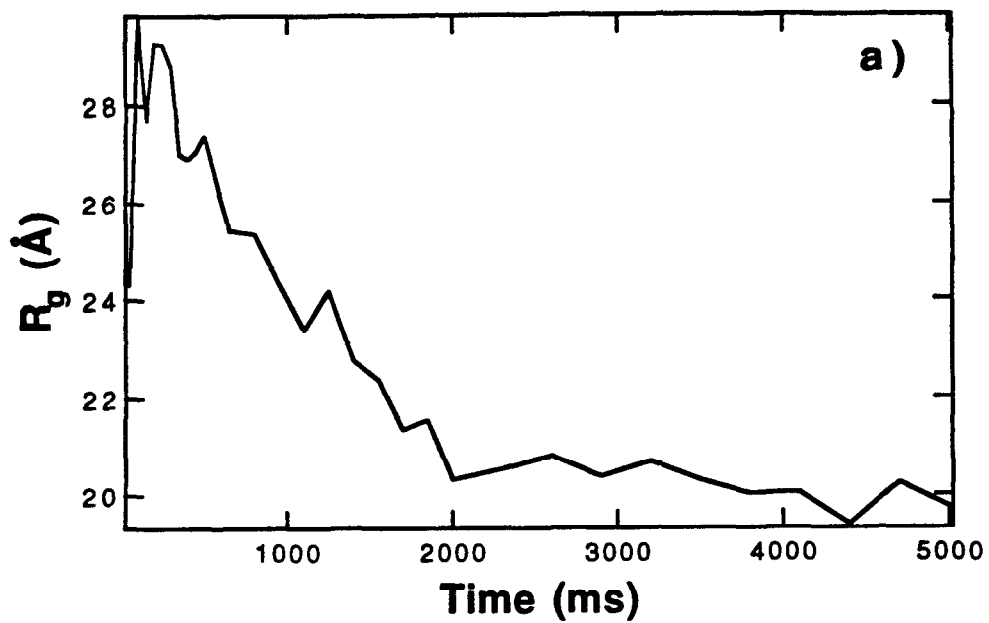


Figure 3.C.11: Time-courses of R_g during myoglobin refolding from 8 M to 2 M urea at a) 16 °C (January data) and b) 20 °C (March data).

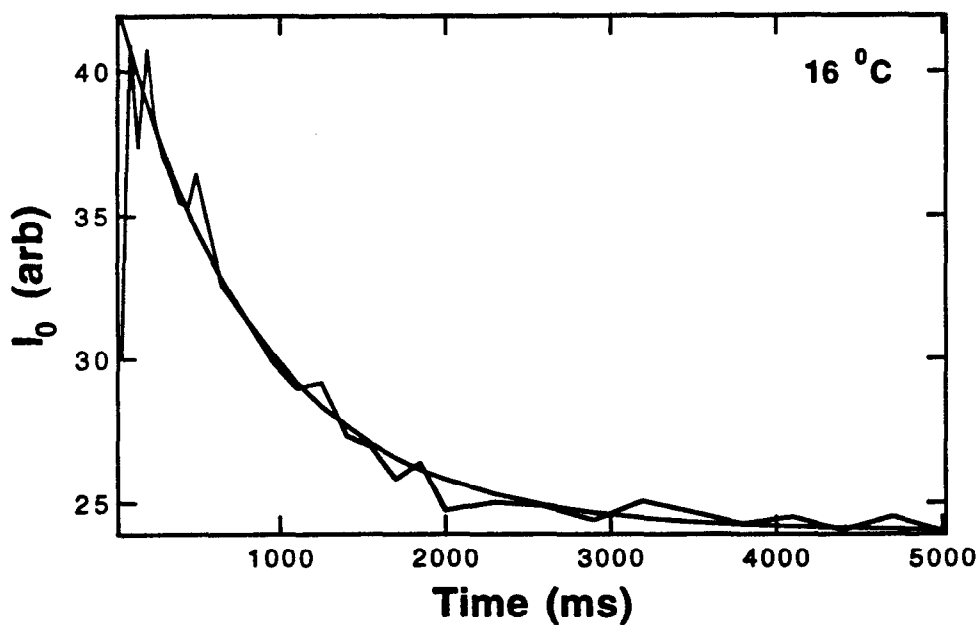
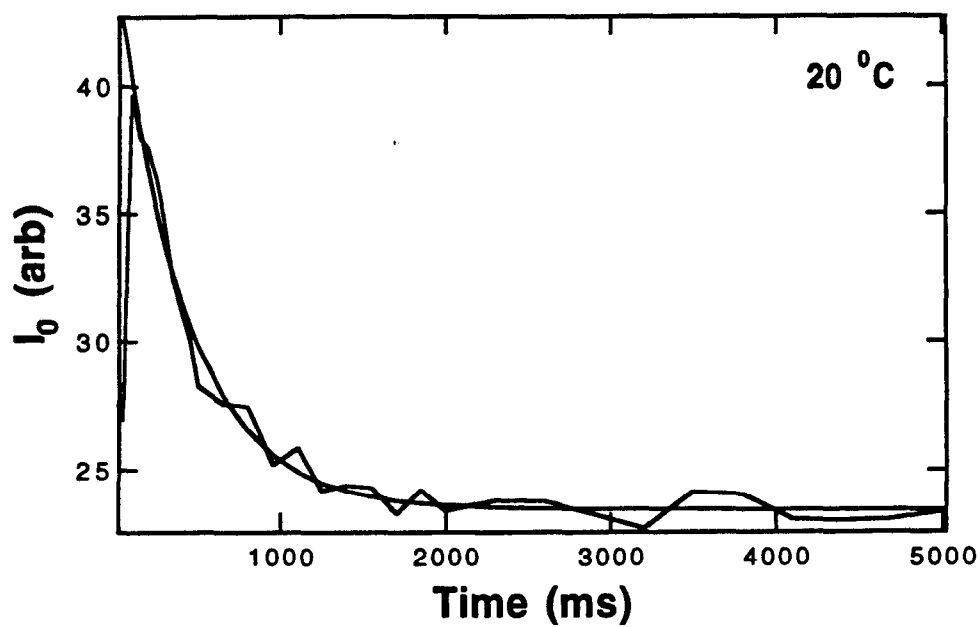
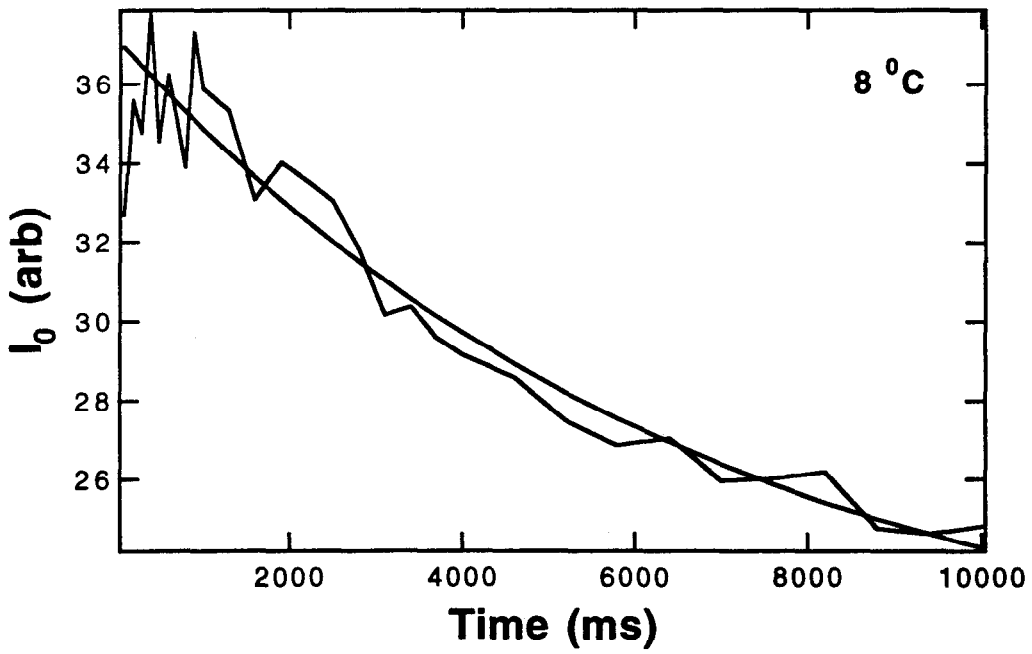
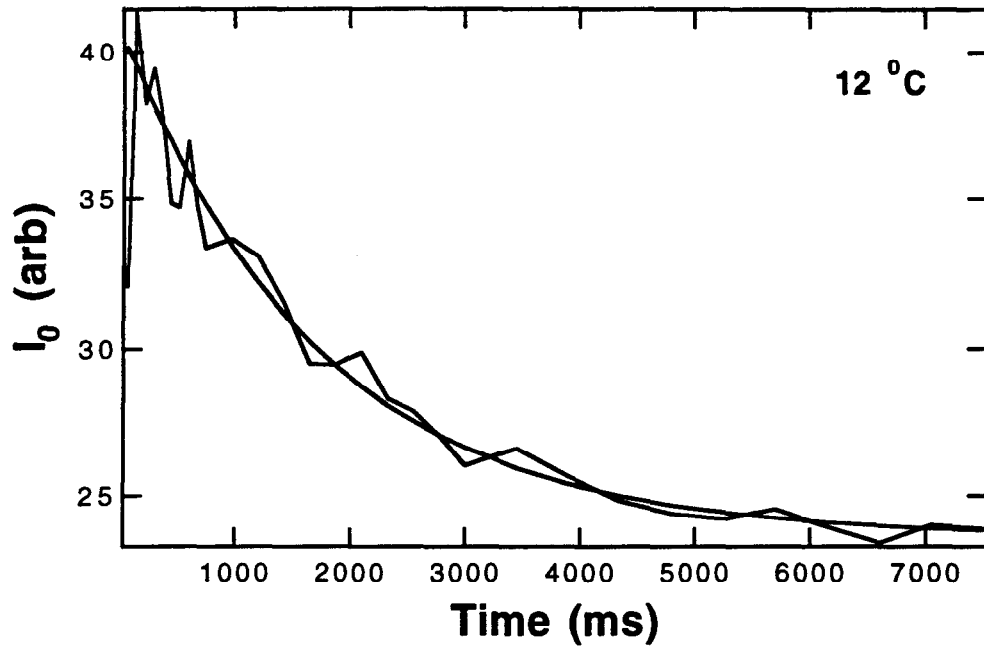


Figure 3.C.12: Time-courses and fits of I_0 during myoglobin refolding from 8 M to 2 M urea from January data. The temperatures in order of appearance are 20, 16, 12, and 8 °C. The fit results are given in Table 3.C.2.



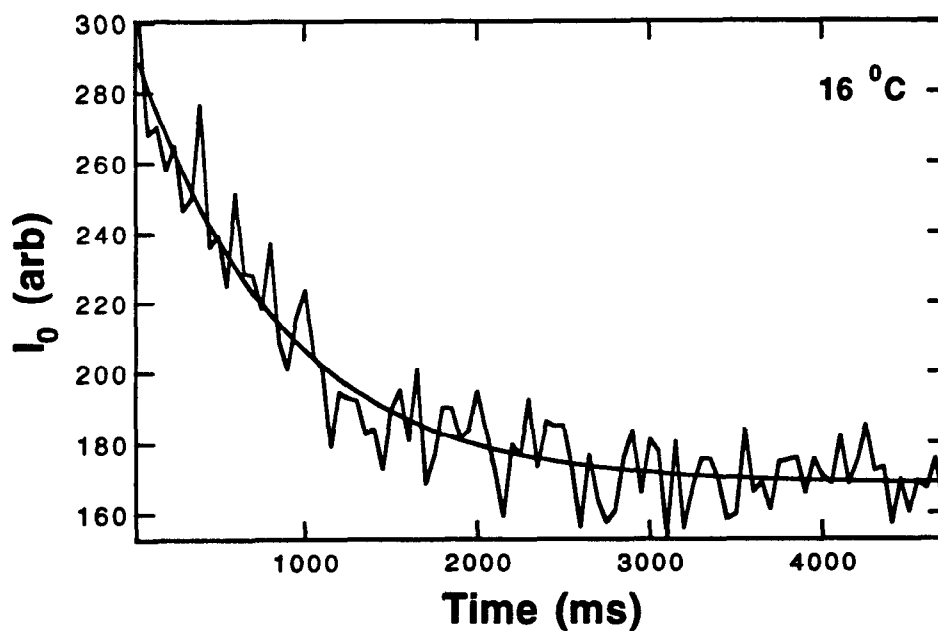
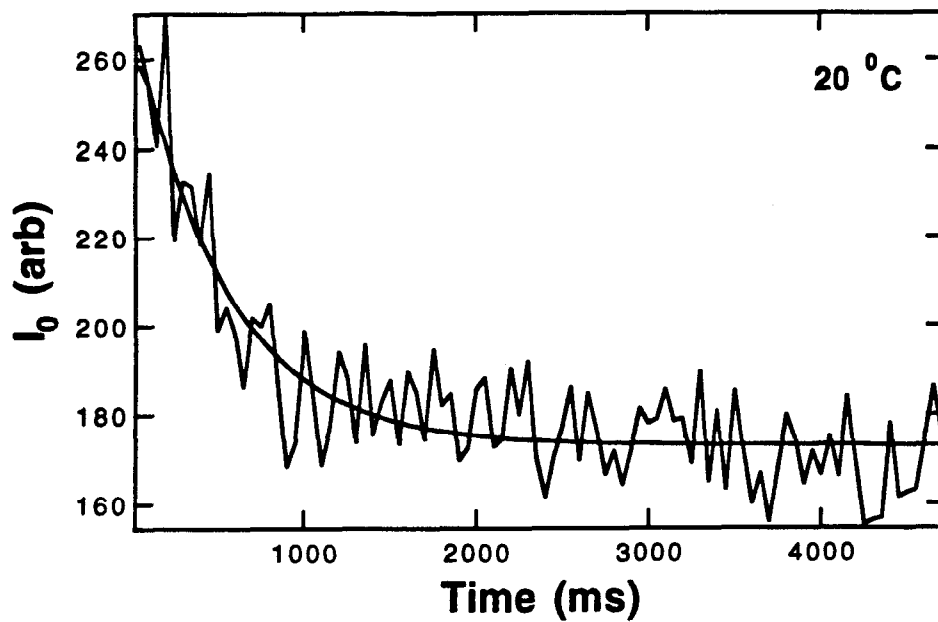
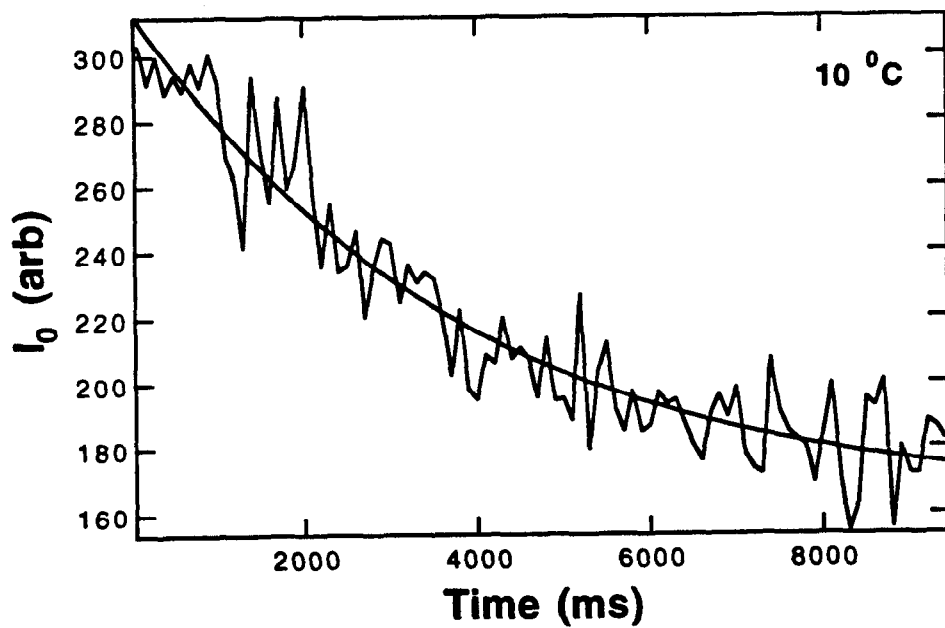
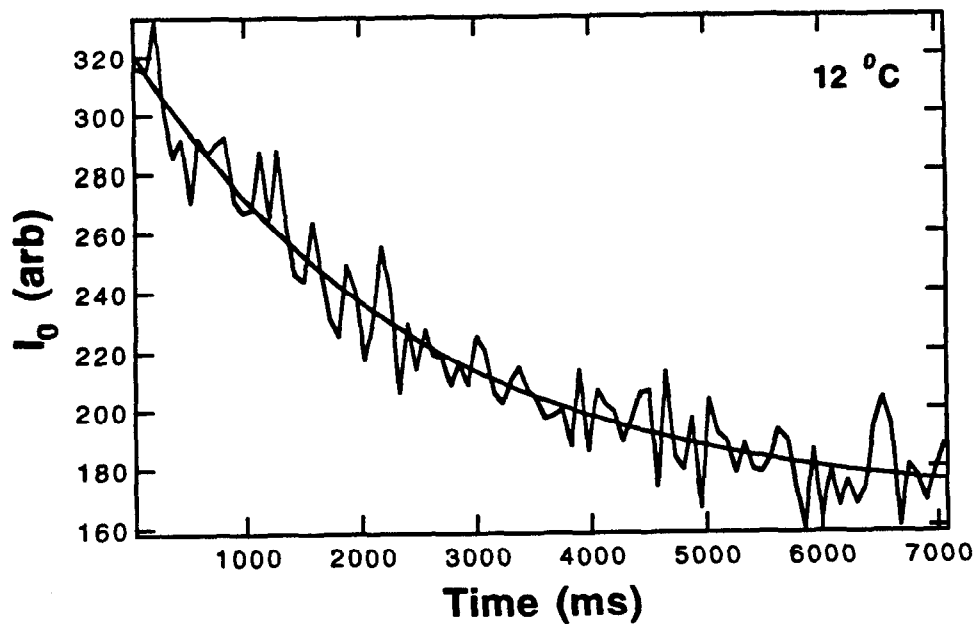
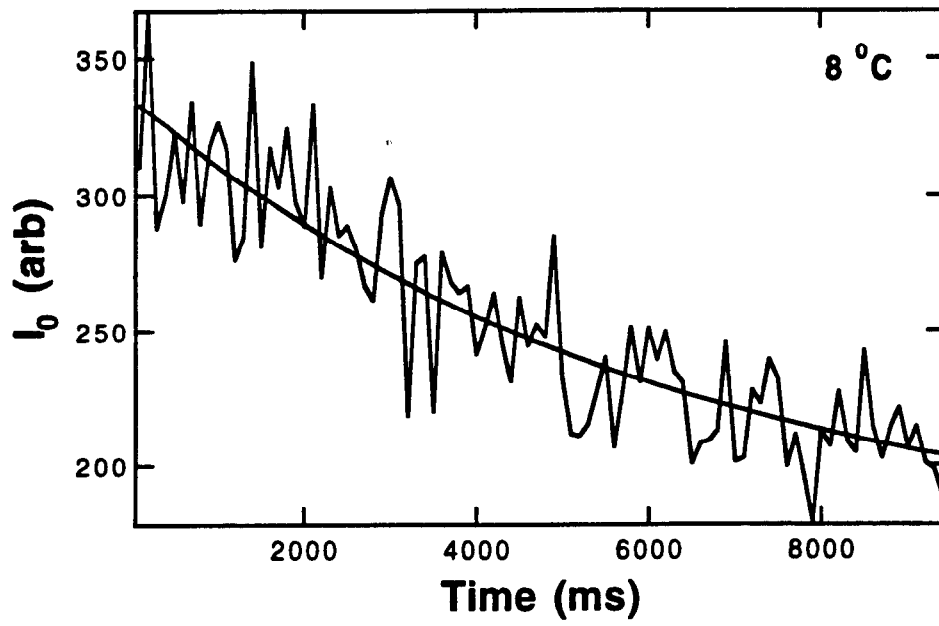


Figure 3.C.13: Time-courses and fits of I_0 during myoglobin refolding from 8 M to 2 M urea from March data. The temperatures in order of appearance are 20, 16, 12, 10, and 8 °C. These fit results also are given in Table 3.C.2.





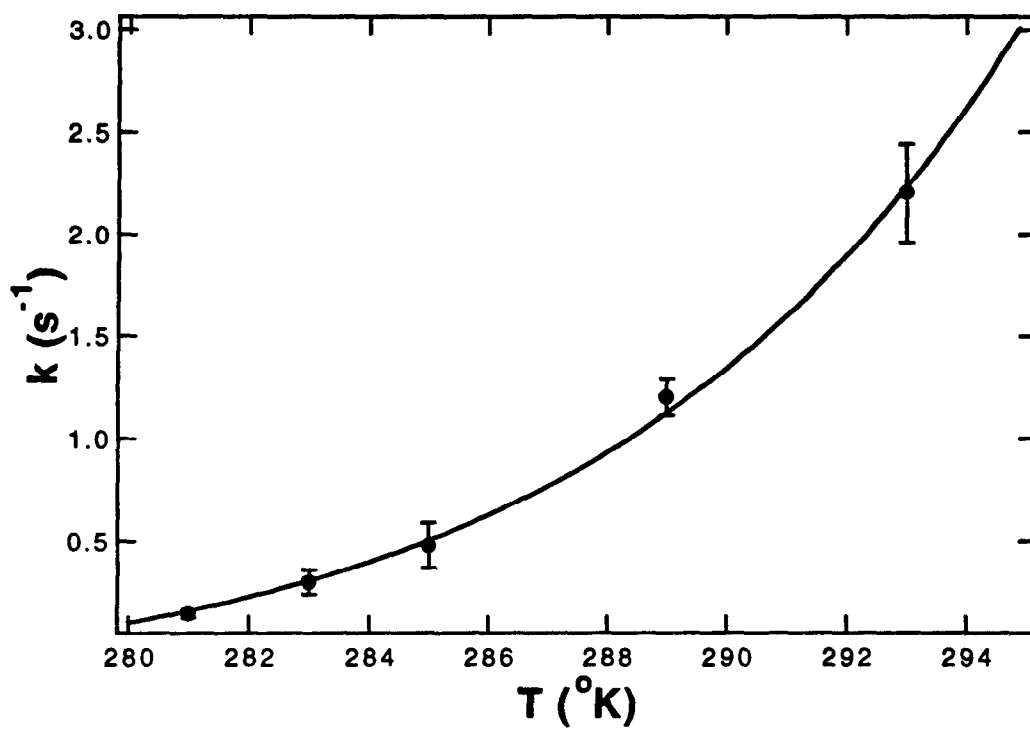


Figure 3.C.14: The apparent dissociation rate constant plotted as a function of temperature, along with a fit to $Ze^{-E/RT}$. The determined value of E is 25 kcal/mole.

test

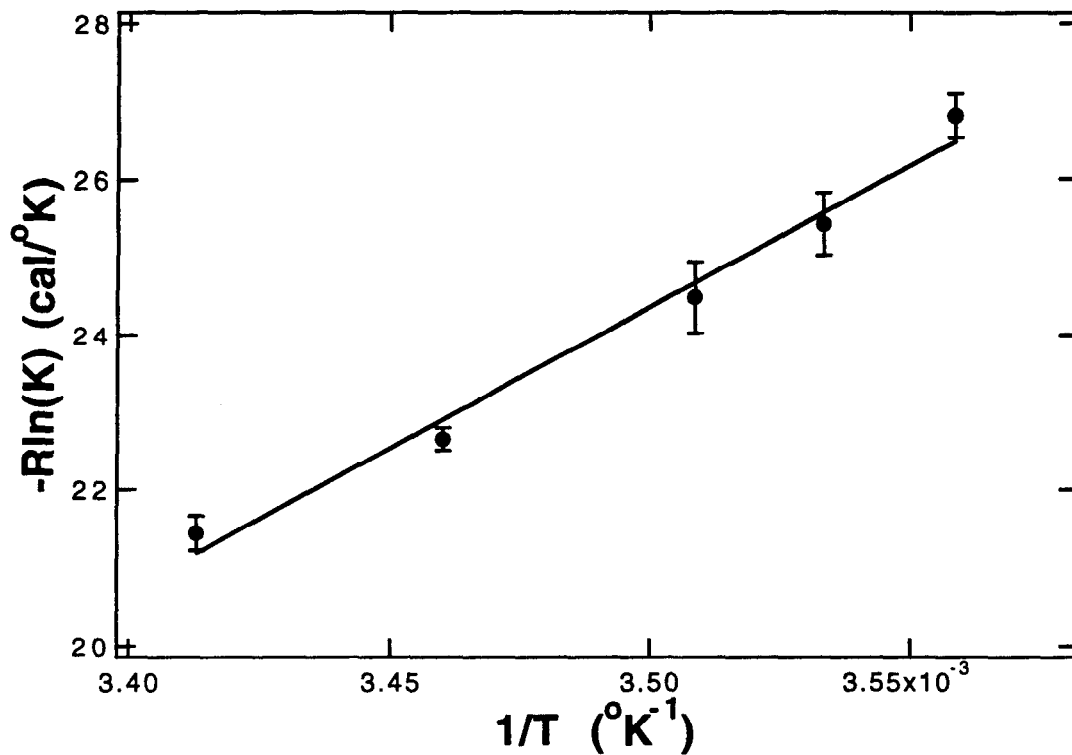


Figure 3.C.15: $-R\ln K$ versus $1/T$. K is the estimated dissociation equilibrium constant. The slope of the linear fit gives an estimate of 36 kcal/mole for the enthalpy of dissociation.

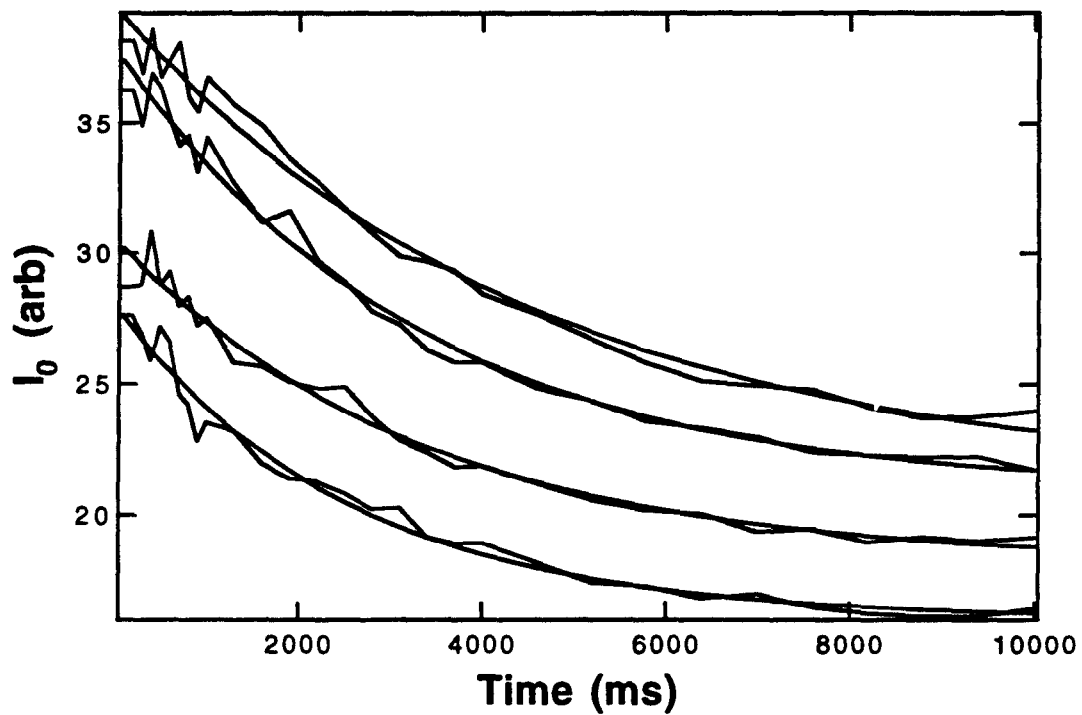


Figure 3.C.16: Time-courses and fits of I_0 during myoglobin refolding from 8 M to 2 M urea at four different initial protein concentrations (70, 64, 58, and 52 g/l in order from the upper to the lower curve).

3.D: Cytochrome-*c*:

Sample Preparation:

Horse heart cytochrome-*c* was purchased from Sigma (minimum 95% pure, prepared without the use of tricarboxylic acid) and used without further purification. Samples were prepared gravimetrically, and occasional verification of the protein concentration (using an extinction coefficient of $106000 \text{ M}^{-1}\text{cm}^{-1}$ at 408 nm) showed that the reported concentrations are again accurate to about 5%. All protein solutions were centrifuged at around 15,000 rpm for 30 minutes or longer prior to use in order to remove any aggregates. Unfolded protein solutions were allowed to equilibrate for a minimum of 1 hour prior to use. The experimental conditions were chosen to be as similar as possible to those used by Roder (1988) and Elöve (1992), namely 100 mM acetate at pH 6.3. Unfortunately, these authors used GuHCl as the denaturing agent in their studies, and as mentioned several times already, this is a poor choice for SAXS experiments. As usual, urea was substituted for GuHCl, and all solutions were used within 24 hours of preparation to prevent effects from the decomposition of urea. Because near pH 7, the midpoint of the equilibrium unfolding transition of cytochrome-*c* has been reported to be around 7.5 M urea (Tsong, 1976) an initial urea concentration of 9 M was chosen, to insure that the protein was indeed unfolded in the initial state. A rather high final concentration of 3 M urea was chosen in an attempt to slow down the folding process. The data below verify the protein to be in the native state at this urea concentration. No independent measurement of equilibrium unfolding under these conditions was made.

Experimental Details, Specific:

In January of 1993, initial data were collected at SSRL. Using the ruby-ball mixer, data were collected from the 1:6 mixing of 70 g/l cytochrome-*c* in 9 M urea, 100 mM acetate, pH 6.3 with 2 M urea equivalent buffer at 12 °C. The sample-to-detector distance was 254 cm and the photon energy was 8980 eV. Data were collected in cycles of 10 x 500 ms, 10 x 1000 ms, and 10 x 2000 ms, and 210 cycles total were collected. 10 cycles of 10 x 100 ms, 10 x 500 ms, and 10 x 2000 ms were collected from mixing with 9 M urea for a control.

In June of 1993, further data were collected at SSRL. 76 g/l cytochrome-*c* in 9 M urea were mixed at a 1:6 ratio with 2 M urea using the ruby-ball mixer at 20, 18, 16, and 14 °C.

The sample-to-detector distance was 225 cm and the photon energy was 8980 eV. Data were collected in cycles of 10 x 100 ms and 20 x 500 ms. Between 150 and 350 cycles total were collected at each temperature. 30 cycles of 10 x 100, 10 x 500, and 10 x 2000 ms were collected from mixing with 9 M urea for a control. Because data from the final (post-mixing) condition was not consistent with the protein being in the native state, a longer collection of 70 one-minute cycles was also undertaken to check for slower processes.

In December of 1993, a final data set was collected at the Photon Factory. 45 g/l cytochrome-*c* in 9 M urea were mixed using the stop-cock mixer at a 1:3.5 ratio with 1.3 M urea at 20 °C. The lower initial protein concentration was used to address the possible aggregation effects observed in the earlier data (and discussed below). The sample-to-detector distance was 212 cm and the photon energy was 8250 eV. Data were collected in cycles of 50 x 50 and 44 x 100 ms. 330 folding cycles and 20 control cycles total were collected. In order to compare I_0 of the final state to that of the native state, both mixer syringes were also loaded with directly prepared 10 g/l protein in 3 M urea and the scattering of the mixed sample was measured. This was necessary because it is nearly impossible to correctly compare I_0 values measured in different sample cells.

In addition, static data were collected from 10 g/l protein in 0, 3, 6, 7, 8, and 9 M urea. Data were also collected from 64 g/l protein in 9 M urea diluted (manually) to 9 g/l in 3 M urea both immediately before and 11 hours before measurement, and from some mixed sample from the folding measurements which was stored for 24 hours after mixing (this is the only sample which was not used less than 24 hours after initial preparation). Finally, data were collected from 5, 10, 20, 40, and 60 g/l protein prepared directly in 8.6 M urea in order to check for concentration dependent aggregation in the initial condition. All static data were collected using a 25- μ m thick mica-window cell. Between 3 and 6 cycles of 300 seconds were collected for each condition. In no case were any significant time-dependent effects observed on this (300 second) time scale.

Analysis and Results:

The results of a Guinier analysis of the equilibrium unfolding data (fitting range $K = 0.028$ to 0.062 \AA^{-1}) are presented in Figure 3.D.1. This range was chosen partly because it was used in the myoglobin studies, but also because again it was the lowest range in which results were reasonably reproducible between experiments, indicating that parasitic

scattering effects were no longer present. It is clear that the results from the equilibrium data are consistent with an unfolding transition in the neighborhood of 6 M urea. The dependence of I_0 on urea concentration is reasonably fit by the previous model using $0.43 \text{ e}^-/\text{\AA}^3$ as the electron density of cytochrome-c. Also plotted in the figure are the two data points obtained from the dilution of 64 g/l unfolded protein to 9 g/l and 3 M urea both immediately prior to and 11 hours prior to the measurement. It appears that the sample which has been allowed to fold for 11 hours is in the same state as the directly prepared 3 M urea sample, but the newly mixed sample is not! It has both a higher R_g and a higher I_0 value than the directly prepared sample. This implies the existence of a longer (than 1200 seconds, the time for static data collection) time scale folding process. It also explains the observation in January of 1993 that the final state after the refolding experiments was not the native state. This was initially suspected to be a result of the presence of irreversible aggregation in the samples.

The static data from the initial condition of the refolding experiments at five different protein concentrations, normalized by the respective concentration, are shown in Figure 3.D.2. The effects of interparticle interference are clear at the higher concentrations. Apart from these effects, however, the curves overlay quite nicely. Therefore, there is no evidence for any aggregation at any of the concentrations, implying that at the initial condition of the mixing experiments the protein is monomeric.

This important point is verified by the comparison of the data from the directly prepared 3 M urea sample with the static data collected after the end of the control mixing experiments. The I_0 ratio of the native state to that of the initial state (the protein remains in the initial state during the control mixing experiments) is 1.95 ± 0.1 , in excellent agreement with the value of 1.96 ± 0.1 observed in the equilibrium data due to the different solvent contrast. Thus, no aggregation is present in the initial sample after dilution. The control experiment from December (the other controls were too noisy to be useful) also shows no obvious change in the behavior of I_0 or R_g as function of time (Figure 3.D.3) indicating that no aggregation is present in the undiluted initial sample either. However, because the signal in these refolding experiments is much smaller than that observed for myoglobin, this control is too noisy to be conclusive, and further control mixing measurements should be performed.

The refolding data were analyzed in the same fashion as the static data, using the same fitting range. R_g time-courses were noisy, and not in very good agreement between the

various conditions. The initial value varied from 25 to 30 Å, and the final value was around 20 Å, significantly higher than that of the native state (13 ± 1 Å). This is consistent with the observation above that after mixing, the protein does not immediately return to the native state. Representative plots are presented in Figure 3.D.4. Plots of I_0 time-courses are shown in Figure 3.D.5, along with single exponential fits, for all 6 data collections. The fit results are summarized in Table 3.D.1. A definite decrease is observed in the magnitude of I_0 , but unlike in the case of myoglobin, this decrease is only a factor of 1.27 ± 0.04 , nowhere near 2. In the figure containing the Photon Factory data, however, the I_0 value of the directly prepared native protein (in 3 M urea) is also plotted. It is evident that at the end of the time-resolved measurement, the protein has not achieved its native value of I_0 . Once again, it appears as though a slower process is present. Moreover, comparison of I_0 from initial data frames with I_0 for the directly prepared native protein reveals that their ratio is 1.86 ± 0.06 , reasonably close to 2 after all. These combined observations lead to the hypothesis that cytochrome-*c*, like myoglobin, undergoes a complete dimerization event during refolding, but that unlike myoglobin, the subsequent dissociation of these dimers takes place on (at least) two well-separated time scales.

Verification of this hypothesis is available from the I_0 data of the 70 one-minute cycles after mixing (Figure 3.D.6). Although very noisy, these data show a definite decrease in I_0 over the course of the measurement. Unfortunately, the quality of the data is insufficient to uniquely determine an exponential. On the assumption that only two processes are involved, and that therefore the total decrease in the slower phase must be a factor of $2/1.3 = 1.54$, the baseline of the fit to the data was fixed appropriately. The time constant determined under these conditions is 0.007 ± 0.001 minutes, corresponding to a half-life of 100 minutes. The resultant fit is presented in the figure.

Further verification is provided by the static data measured from the final state of the refolding experiments. These data always give a lower value of I_0 than the final value from the kinetics data. For the three experiments, the ratio of I_0 from 20 minutes of final data to I_0 of the final kinetic data frames was 0.93, 0.90, and 0.84. If we take the average value, 0.89 ± 0.03 , to represent the decay of I_0 in 10 minutes after mixing, it is in reasonable agreement with the value predicted using the 0.007 minute time constant (which is 0.93). Again this presupposes only two processes. It is entirely possible that the 10% decrease is a separate intermediate rate process, and/or that the slow process consists of more than one phase.

Finally, there is the mixed sample which was collected after the stopped-flow experiments in December. Data measured from this sample on the next day are compared with data collected immediately after mixing, and with data from directly prepared 10 g/l protein in 3 M urea in Figure 3.D.7. It is quite clear that during the intervening time, the mixed sample has completed the transition to the native state.

Discussion:

The most natural interpretation of the above results is similar to that in the case of myoglobin. In particular, it follows from the same reasoning that transient association also occurs on the refolding pathway of cytochrome-*c*, and that therefore this protein too folds through one, or several, associative folding intermediates. Once again circular dichroism studies by other workers support the presence of early secondary structure in this intermediate, and a Kratky analysis (not shown) confirms the compact nature of the intermediate. The conclusion must be that this kinetic intermediate is a molten globule.

Similarly, the fact that the total observed decrease in I_0 is close to a value of 2 indicates that the protein probably associates into dimers. In this case, however, the return of I_0 to its native value follows a bi-phasic behavior, with a fast phase accounting for 45% of the total change and a much slower phase for the remaining 55% (assuming complete dimerization). The only cause of changes in I_0 in these studies has been shown to be the dissociation of protein oligomers. Because a dimer can not dissociate partially, accepting the dimerization hypothesis forces the conclusion that not all the dimers are equal. Some dissociate quickly, and others much more slowly. Because the value of I_0 normalized to that of the final (monomeric) sample is $(1-f) + 4f/2 = 1+f$, where f is the dimer fraction, the fraction of fast- and slow-dissociating dimer populations must equal the fractional I_0 change observed in the corresponding phases, namely 45% and 55%.

At first this seems like an unlikely supposition, as there appears to be nothing in the experiments to produce such heterogeneity. In fact, however, the unfolded state of cytochrome-*c* has been observed to be heterogeneous by many other investigators, including Tsong (1976), Ridge (1981), Brems (1983) and most recently Roder (1988) and Elöve (1992) using many probes, including UV absorbance, tryptophan fluorescence, circular dichroism, and NMR pulse labelling. Many of these authors have suggested that the heterogeneity is caused by a population of slowly isomerizing *cis* conformation proline residues. Cytochrome-*c* has four prolines, all in the *trans* conformation in the

native state. The probability of finding an unconstrained proline in the *cis* conformation has been estimated to be between 0.1 and 0.3 (for other amino acids, this probability is around 0.001). In at least one study of cytochrome-*c* refolding (Ridge, 1981) bi-phasic kinetics were observed, and proline isomerization was strongly implicated as the cause of the slower kinetic phase by 'double-jump' experiments in which the amplitude of this phase was found to increase with the time the sample was allowed to remain unfolded. All prolines in the freshly unfolded protein must be in the correct conformation, but as time passes, some fraction will isomerize to the incorrect conformation. This fraction will reach a constant (equilibrium) value after enough time, with one hour typically being sufficient. All samples in the current study were allowed to unfold for over one hour. Unfortunately, no double-jump experiments have been attempted yet.

As discussed by Levitt (1981) the fractional population of the fastest folding state, with four prolines in the correct (in this case *trans* conformation) is about $(.8)^4 = 0.41$. It has been suggested that one of the prolines in cytochrome-*c*, however, does not effect the folding process (Brems, 1983). In this case, the fast folding fraction would be $(.8)^3 = 0.51$. Both of these values are reasonably close to the observed value of the fastest phase in the SAXS study. It is important to note, however, that these numbers are quite sensitive to the value assumed for the probability of the *trans* configuration. Nevertheless, it appears possible that the observed fast phase is produced by those unfolded molecules in the most favorable conformation, and the remaining slow phase by all the rest of the molecules. No data is available at times sufficiently large to determine if the slow phase is indeed a single phase, or a combination of several phases.

Because the second phase observed is much much slower than any observed in other studies, and because proline isomerization typically occurs on a seconds or minutes (not hours) time scale, it must be the case that interactions within the dimer act to stabilize the proline isomers (although it should be pointed out that the rather high final urea concentration is also likely to significantly slow folding in general). In return, the incorrect proline isomers must stabilize the dimer in order to cause the bi-phasic behavior. Such stabilization is possible to imagine if the proline in question is involved directly at a specific dimerization site. As with the case of myoglobin, it could be that the dissociation of the dimers is caused by some conformational change. It may be that this change can not proceed before the isomerization step.

It is worth noting that the 45% amplitude of the fastest phase is equal to that observed for the fastest phase in the CD studies of Elöve (1992) under the same conditions. It is difficult to say whether this is significant. It is possible that the amplitude of the initial CD phase represents the complete CD signal of a single fraction of the protein population. This is not in agreement, however, with pulse labeling NMR studies of Roder (1988), again under the same conditions, which show only 20% of the protein folding on the same (fastest) time scale. It seems likely that the CD signal is a linear combination of a large fraction or all of the total CD signal of the fastest folding population and a smaller fraction of the total CD signal of slower folding populations, in which case the agreement of the fast phase amplitudes observed by CD and SAXS is coincidental.

Two possibilities are suggested here to explain the difference between the 45% fast-folding population observed in the SAXS study and the 20% fast-folding population observed in the NMR study under nearly identical conditions. First, the NMR study used GuHCl as the denaturing agent. It may be possible that the difference in denaturant leads to different populations of fast- and slow-folding species. This possibility could be tested by repeating the NMR experiments using urea, or the SAXS experiments using GuHCl, but both of these are difficult propositions, because the NMR experiments require long times, and the SAXS experiments require high contrast.

The NMR studies also suggested the presence of two distinct (intermediate and truly slow) slow folding species, with populations of around 40% each. The suggestion of complete dimerization in the SAXS study assumes about a 15% error in the 1.86 ± 0.06 value of the initial-to-final I_0 ratio. As discussed in the myoglobin studies, there is a valid reason to believe that this value is somewhat lower than the true value, because of the effect of the fitting range. Unfortunately, only one measurement is available. Suppose, however, that this value is approximately correct. This would mean that only 86% of the protein was dimerized, the other 14% having folded too fast to allow for dimerization. The observed factor of 1.3 decrease of the fast phase would then correspond to 43% of the total protein, and the remaining slow phase would constitute the final 43%. This scenario would be quite consistent with the NMR pulse labeling data. More and better SAXS data would be needed to test this possibility.

Proline isomerization, of course, is not the only possible explanation for the observed bi-phasic dimer dissociation. One obvious alternative is that the nature of the dimerizing interaction might be different in the two populations. For example, it may be possible

that two specific dimerization sites exist. In this case, there may be no correlation at all between the phase amplitudes observed using SAXS and those observed using other techniques.

Conclusions:

Although once again the initial collapse of the protein in question apparently occurs too quickly to be observed on the available time scale (in this case at 50 ms), this study does provide compelling evidence for the existence of a kinetic molten-globule folding intermediate in yet another protein, horse heart cytochrome-*c*. The potential for transient association on refolding pathways is again illustrated, and dimerization in particular is implicated. In this case, however, dissociation proceeds in two distinct phases, the amplitude of one of which corresponds well to a refolding phase observed using circular dichroism. Agreement with NMR pulse labeling studies is reasonable if the complete dimerization model is relaxed to allow for about a 20% protein fraction which folds too fast to dimerize. The distinct phases are hypothesized to result from a heterogeneous population of unfolded states differing in the isomerization state of one or more proline residues. If this is the case, the *cis* conformation of the involved prolines must somehow be responsible for the tremendous extension of the lifetime of the longer-lived dimer fraction.

The three studies presented in this chapter have illustrated the utility of the stopped-flow time-resolved SAXS technique in the study of the protein folding problem. The demonstrated existence of kinetic molten-globule folding intermediates is a small but important step in the progress towards a fuller understanding of the protein folding process. The upper bounds set on the compaction processes of the three proteins examined may also prove useful in discriminating between various theoretical or computational folding models. It is likely that further application of this method to the study of other proteins will result in further interesting insights into the protein folding problem.

**Table 3.D.1:
Parameters from the Time-Course of the Forward
Scattering from Cytochrome-c During Refolding**

T (°C)	init [cyt-c] (g/l)	k_{app} (S ⁻¹)	$I_0(t_i)/I_0(t_f)$
20±0.1	76	0.62±0.16	1.26±0.14
18±0.1	76	0.38±0.06	1.28±0.06
16±0.1	76	0.44±0.08.	1.27±0.07
14±0.1	76	0.31±0.06	1.25±0.07
12±0.1	45	0.26±0.02	1.30±0.07
20±0.1	70	0.92±0.14	1.23±0.10

For all samples the initial and final urea concentrations were 9 M and 3 M, respectively. The apparent rate constant, k_{app} , and the ratio of the forward scattering at the initial and final time-frames, $I_0(t_i)$ and $I_0(t_f)$, were determined from single exponential fits to the data. Fits and errors were calculated using the program Igor (WaveMetrics, Lake Oswego, OR).

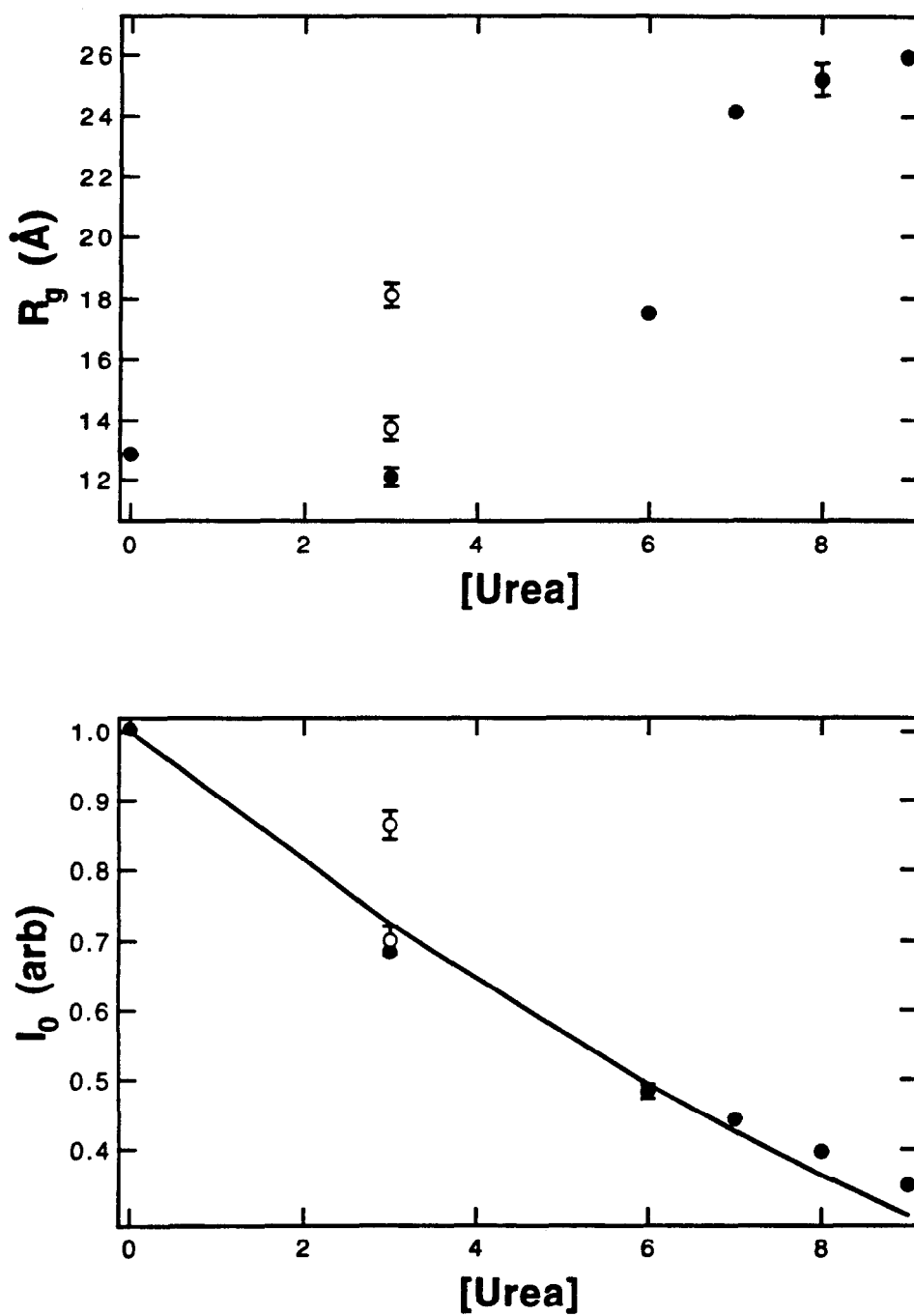


Figure 3.D.1: R_g and I_0 from the equilibrium unfolding of cytochrome-c. The solid circles are from samples directly prepared at 10 g/l. The open circles are from samples prepared as 64 g/l protein in 9 M urea and diluted to 9 g/l in 3 M urea either 11 hours (lower) or immediately (upper) before measurement.

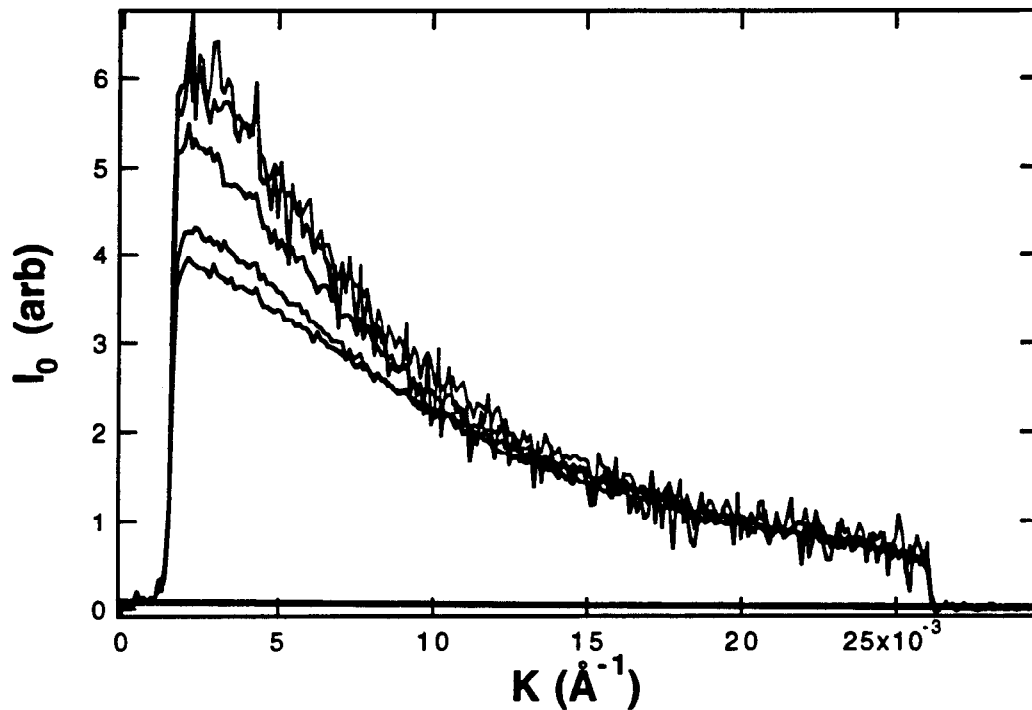


Figure 3.D.2: Concentration-normalized background-subtracted calibrated SAXS data from 5, 10, 20, 40, and 60 g/l cytochrome-*c* in 8.6 M urea (in order from the upper to the lower curve). Interparticle interference is responsible for the decrease in intensity at low angles with increasing concentration. No indications of aggregation are visible.

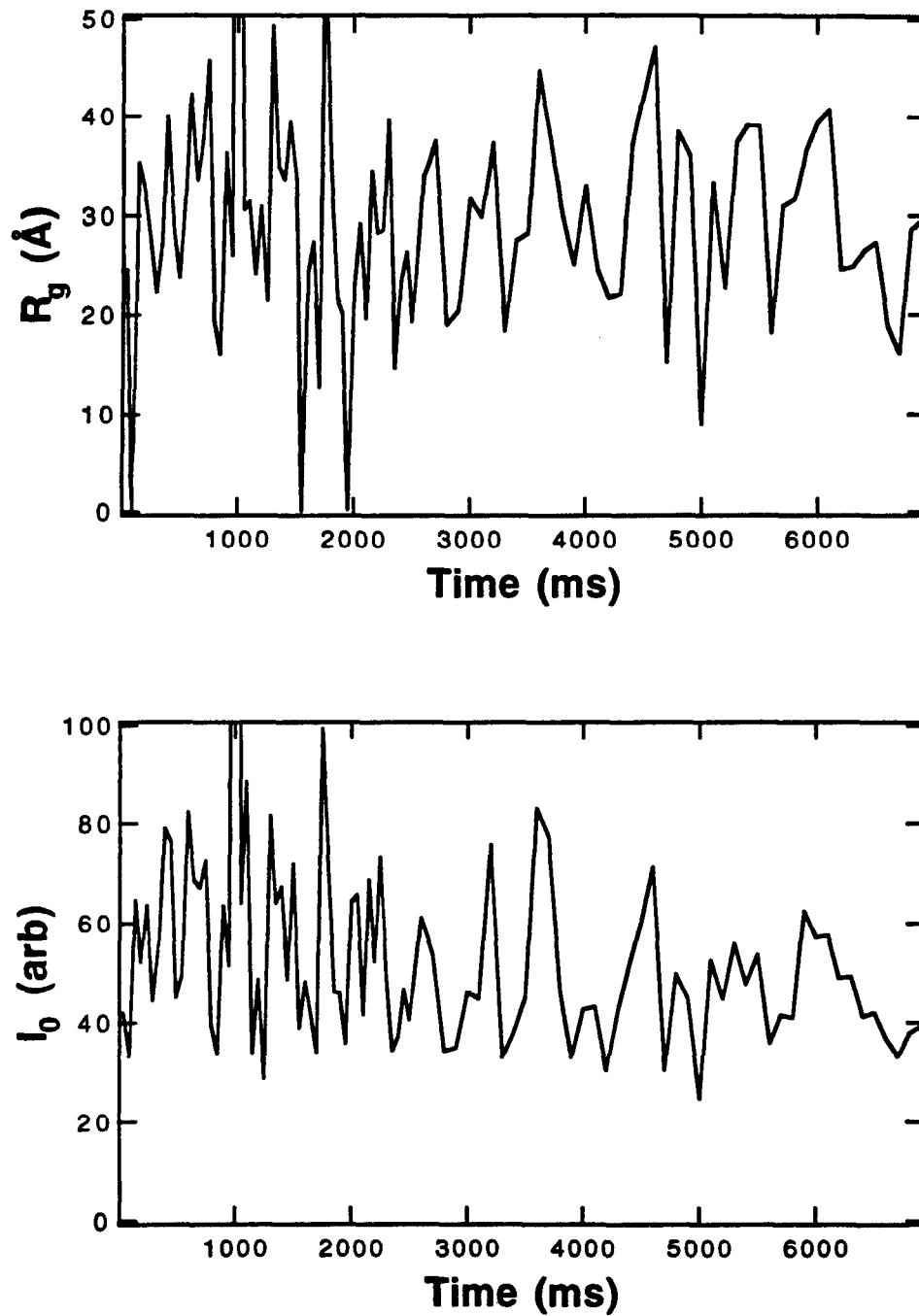


Figure 3.D.3: R_g and I_0 as a function of time after mixing in a control experiment in which cytochrome-*c* unfolded in 9 M urea was diluted without changing the urea concentration. No obvious time-dependent process is observed.

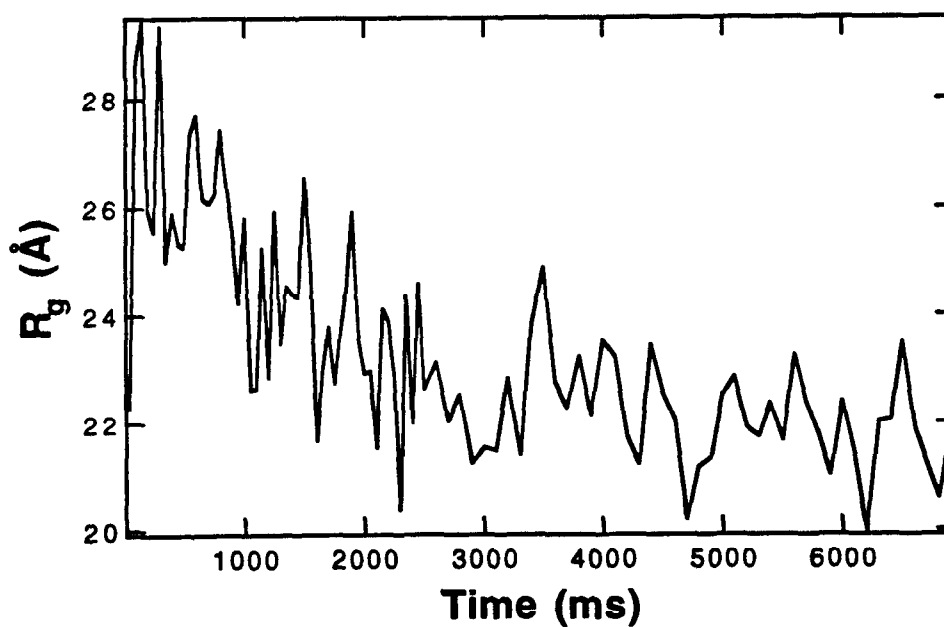
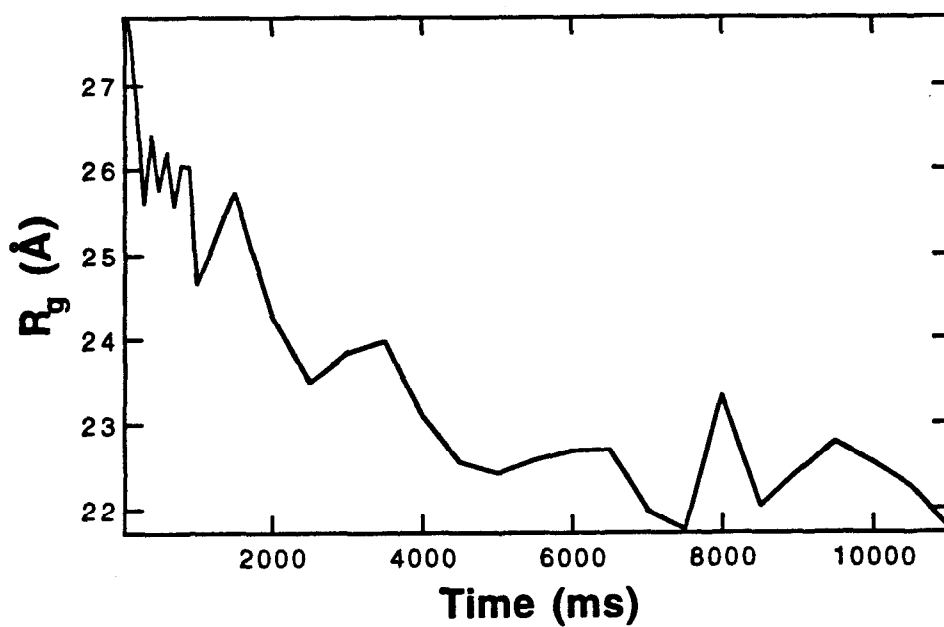


Figure 3.D.4: Representative plots of R_g during the refolding of cytochrome-*c* from 9 M to 3 M urea at 14 °C (upper) and at 20 °C (lower).

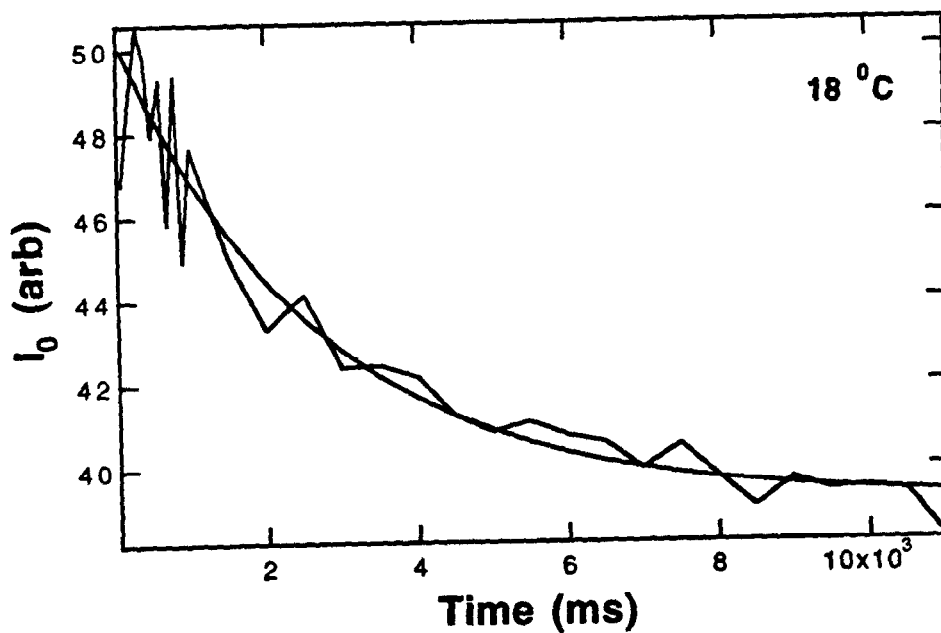
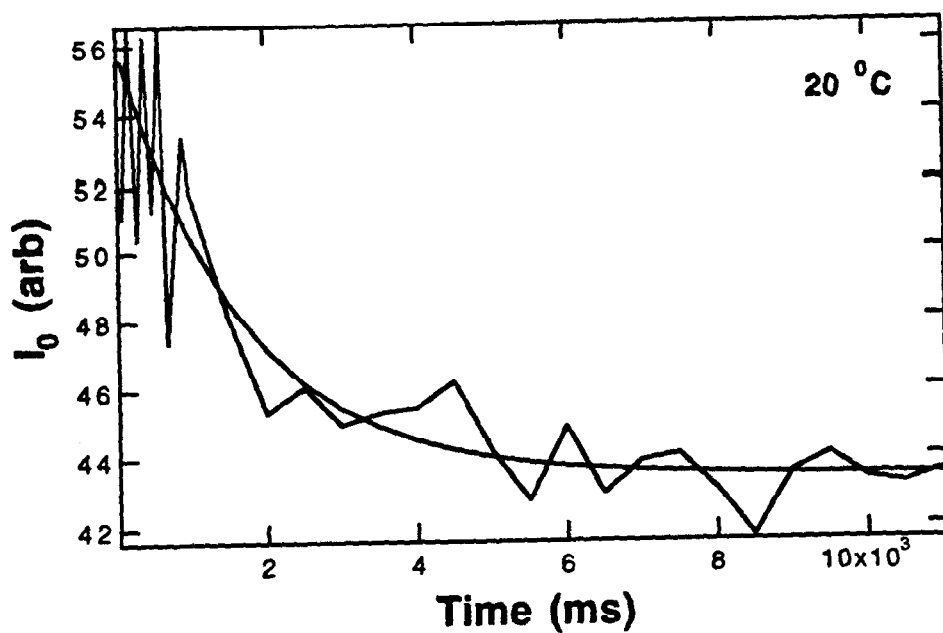
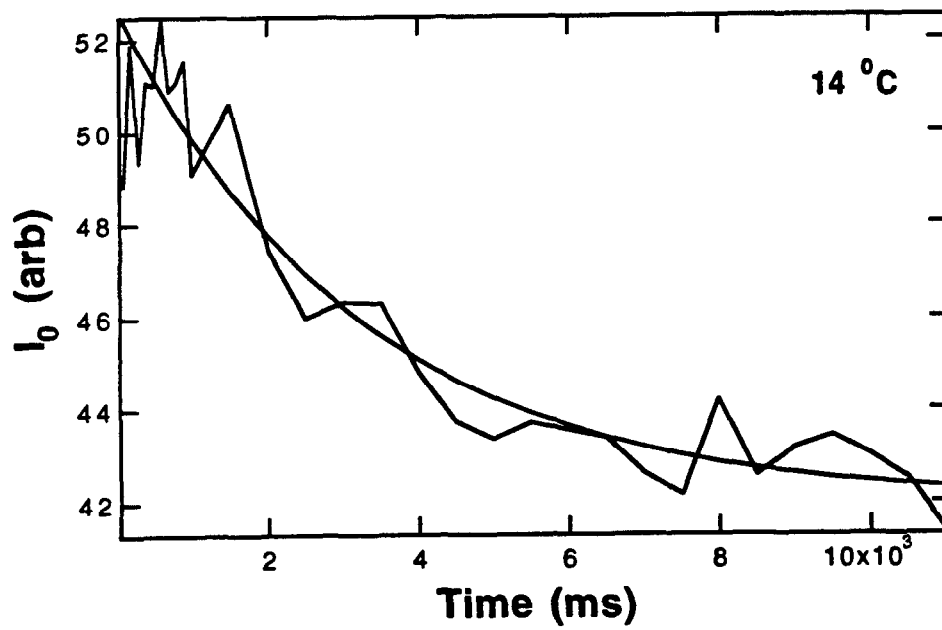
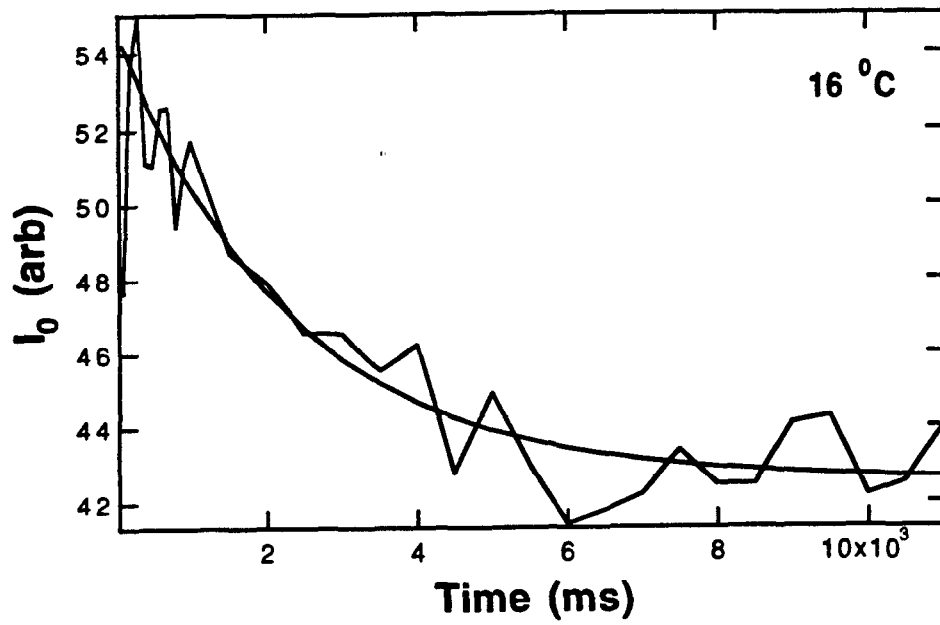
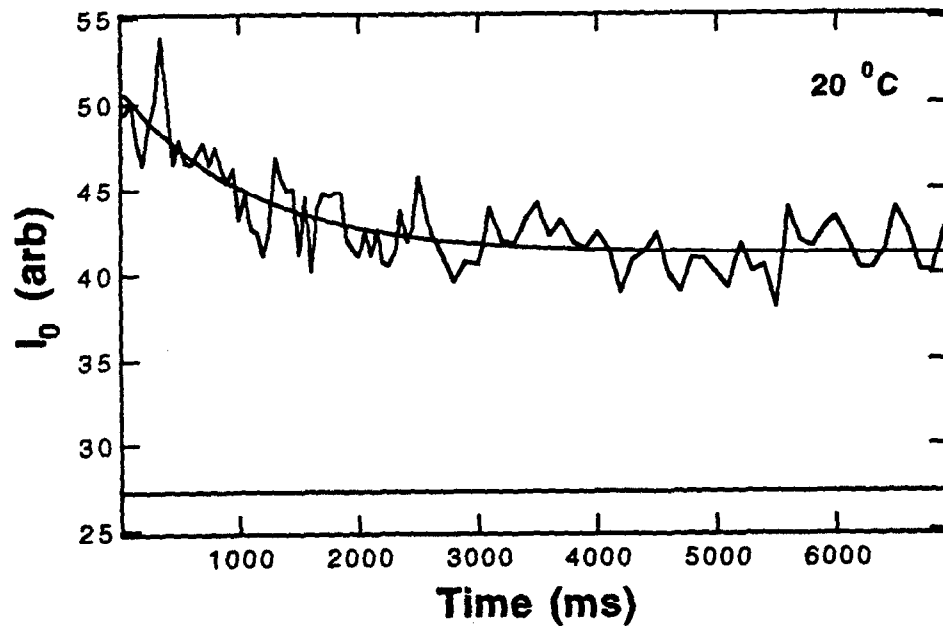
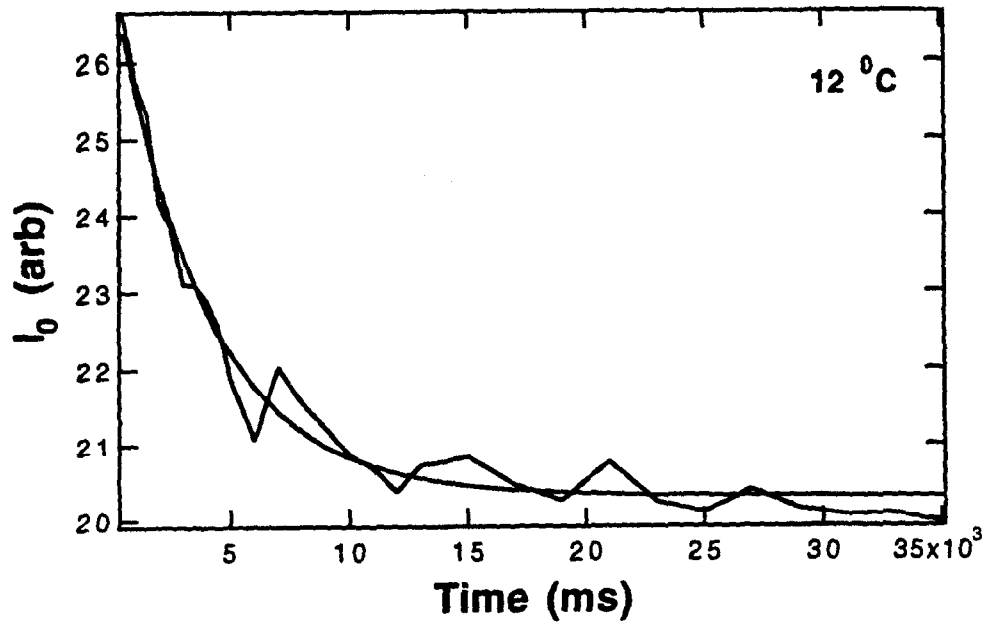


Figure 3.D.5: I_0 (with single exponential fits) during the refolding of cytochrome-*c* from 9 M to 3 M urea at (in order of appearance) 20, 18, 16, 14, 12, and again 20 °C. The final plot also shows the baseline value of I_0 from fully native protein prepared directly in 3 M urea. Fit results are given in Table 3.D.1.





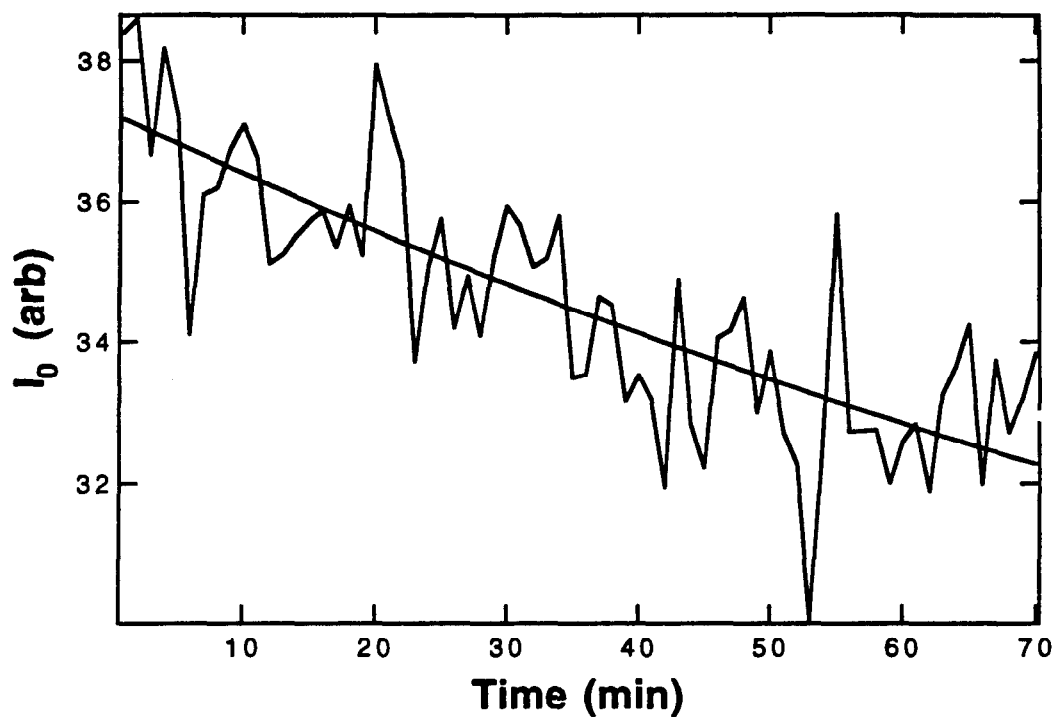


Figure 3.D.6: I_0 during 70 one-minute data cycles collected from the refolding of cytochrome-c from 9 M to 3 M urea at 20 °C. The displayed fit is a single exponential obtained by fixing the baseline value at $1/1.54 = 0.65$ times the initial value of 0.37.

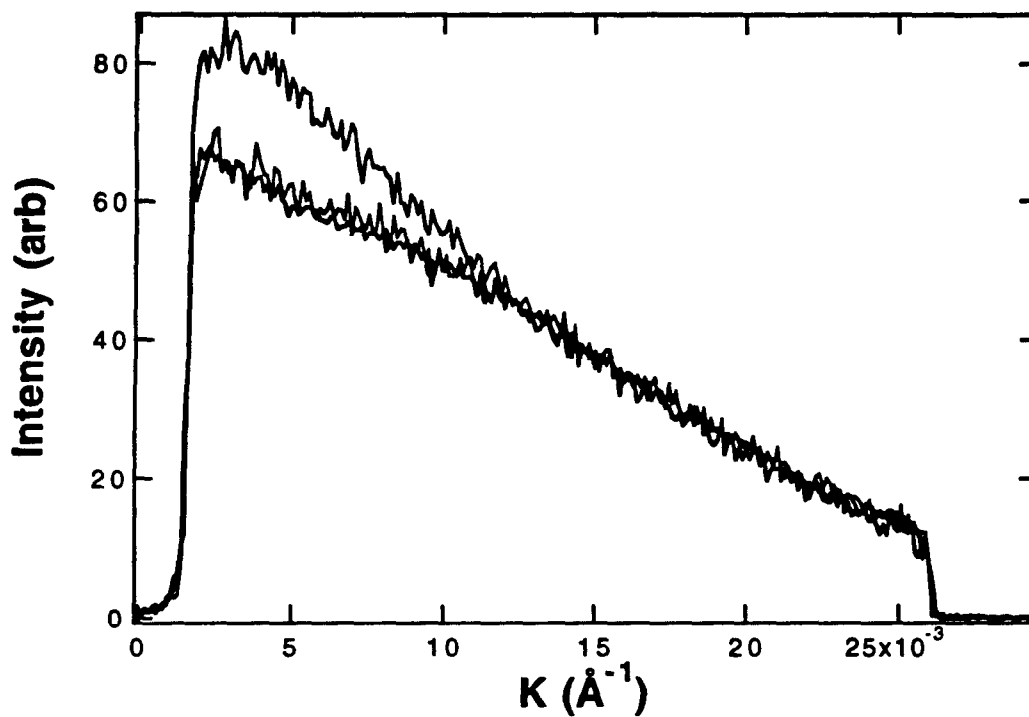


Figure 3.D.7: Normalized background-subtracted calibrated SAXS data collected immediately after the completion of a refolding experiment (upper), from sample mixed in refolding experiments and stored for 24 hours, and from protein prepared directly in 3 M urea. It clear that the mixed sample completes the transition to the native state during the storage period.

References:

- Anfinsen, C.B. 1973. *Science*. 181:223-230.
- Baker, D., Sohl, J.L., and Agard, D.A. 1992. *Nature*. 356:263-265.
- Brems, D.N. and Stellwagen, E. 1983. *J. Bio. Chem.* 258:3655-3660.
- Brems, D.N., Plaisted, S.M., Havel, H.A., and Tomich, C.-S.C. 1988. *Proc. Natl. Acad. Sci. USA*. 85:3367-3371.
- Chan, H.S. and Dill, K.A. 1993. *Physics Today*. 46:24-32.
- Cleland, J.L., and Wang, D.I.C. 1992. *Biotechnol. Prog.* 8:97-103.
- Crook, E.M., Mathias, A.P., and Rabin, B.R. 1960. *Biochem. J.* 74:234-238.
- CRC Handbook of Chemistry and Physics. 1987. CRC Press, Boca Raton, Florida. D-266.
- Elöve, G.A., Chaffotte, A.F., Roder, H., and Goldberg, M.E. 1992. *Biochemistry*. 31:6876-6883.
- Garel, J.-R. 1976. *Eur. J. Biochem.* 70:179-189.
- Hubbard, S.R. 1987. Ph.D. Thesis. Department of Applied Physics. Stanford University.
- Hughson, F.M., Wright, P.E., and Baldwin, R.L. 1990. *Science*. 249:1544-1548.
- Jennings, P.A. and Wright, P.E. 1993. *Science*. 262:892-896.
- Karplus, M. and Shakhnovich, E. 1992. in *Protein Folding* (Creighton, T.E. ed.). Freeman, New York. 127-195.
- Kataoka, M., Hagihara, Y., Mihara, K., and Goto, Y. 1993. *J. Mol. Biol.* 229:591-596.

- Kiefhaber, T., Rudolph, R., Kohler, H.-H., and Buchner, J. 1991. *Bio-Technology*. 9:825-829.
- Koch, C.J. and Raleigh, J.A. 1991. *Arch. Biochem. Biophys.* 287:75-84.
- Kuwajima, K., Sakuraoka, A., Fueki, S., Yoneyama, M., and Sugai, S. 1988. *Biochemistry*. 27:7419-7428.
- Lee, J.C. and Timasheff, S.N. 1974. *Biochemistry*. 13:257-265.
- Levitt, M. 1981. *J. Mol. Biol.* 145:251-263.
- Mitraki, A. and King, J. 1989. *Bio-Technology*. 7:690-697.
- Mottonen, J., Strand, A., Symersky, J., Sweet, R.M., Danley, D.E., Geoghegan, K.F., Gerard, R.D., and Goldsmith, E.J. 1992. *Nature*. 355:270-273.
- Müller, K. 1982. in *Small Angle X-ray Scattering* (Glatter, O. and Kratky, O. eds.). Academic Press, London. 215-236.
- Pace, C.N. 1986. *Methods Enzymol.* 131:266-280.
- Pace, C.N., Laurents, D.V., and Thomson, J.A. 1990. *Biochemistry*. 29:2564-2572.
- Phillips, J.C., LeGrand, A.D., and Lehnert, W.F. 1988. *Biophys. J.* 53:461-464.
- Privalov, P.L., Bendsko, P., Pfeil, W., and Tiktopulo, E.I. 1988. *Biophys. Chem.* 29:301-307.
- Privalov, P.L. 1992. in *Protein Folding* (Creighton, T.E. ed.). Freeman, New York. 83-126.
- Ptitsyn, O.B. 1987. *J. Prot. Chem.* 6:273-293.
- Ptitsyn, O.B., Pain, R.H., Semisotnov, G.V., Zerovnik, E., and Razgulyaev, O.I. 1990. *FEBS Lett.* 262:20-24.

- Ptitsyn, O.B. 1992. in *Protein Folding* (Creighton, T.E. ed.). Freeman, New York. 243-300.
- Ramakrishnan, V., Patthy, L., and Mangel, W.F. 1991. *Biochemistry*. 30:3963-3969.
- Ridge, J.A., Baldwin, R.L., and Labhardt, A.M. 1981. *Biochemistry*. 20:1622-1630.
- Roder, H., Elöve, G.A., and Englander, S.W. 1988. *Nature*. 335:700-704.
- Shack, J., and Clark, W.M. 1947. *J. Biol. Chem.* 171:143-187.
- Sosnick, T.R. and Trehwella, J. 1992. *Biochemistry*. 31:8329-8335.
- Stuhrmann, H.B. 1980. in *Synchrotron Radiation Research* (Doniach, S. and Winick, H. eds.). Plenum Press, New York. 513-531.
- Tsong, T.Y. 1976. *Biochemistry*. 15:5467-5473.
- Tsuruta, H., Nagamura, T., Kimura, K., Igarashi, Y., Kajita, A., Wang, Z.-X., Wakabayashi, K., Amemiya, Y., and Kihara, K. 1989. *Rev. Sci. Instrum.* 60:2356-2358.
- Tsuruta, H. 1990. Ph.D. Thesis. Department of Materials Science. Hiroshima University.
- Udgaonkar, J.B. and Baldwin, R.L. 1988. *Nature*. 335:694-699.
- Udgaonkar, J.B. and Baldwin, R.L. 1990. *Proc. Natl. Acad. Sci. USA*. 87:8197-8201.
- Wayne, R.P. 1969. in *Chemical Kinetics* (Bamford, C.H. and Tipper, C.F.H. eds.). Elsevier, New York. 189-301.
- Yokoya, A., Kobayashi, K., Usami, N. and Ishizaka, S. 1991. *J. Radiat. Res.* 32:215-223.

Chapter 4

White Light SAXS

This chapter describes preliminary results from an attempt to evaluate a method for greatly increasing the photon flux available for synchrotron SAXS experiments. The experimental results are encouraging, although the overall quality of the data collected are quite poor due to the various difficulties encountered. Further studies will be required before an adequate assessment of the technique can be made.

Motivation:

The weakness of the scattering interaction between x-rays and biological molecules in (monomeric) solutions, discussed in Chapter 1, is the most serious obstacle in the application of small angle x-ray scattering to the study of biological samples. The most straightforward way to overcome this obstacle is to increase the available flux of x-rays. This is the reason that synchrotrons are required for many SAXS experiments. The flux from a monochromatic synchrotron beam, however, is still not sufficient for some experiments. Described here is a method for increasing the flux available for some synchrotron SAXS experiments by a factor of up to a thousand. In particular this method may make two classes of previously impossible experiments feasible, namely time-resolved experiments requiring the measurements of R_g and I_0 on a sub-millisecond time-scale, and static SAXS experiments attempting to measure R_g (and I_0) for samples which are only available at sub-millimolar concentrations.

This method involves using a very broad energy bandpass, or white light, for the proposed SAXS experiments. Conventional synchrotron SAXS experiments typically use an energy bandpass of 10 to 100 eV, depending on the monochromatization system. In the white light SAXS technique, a band pass of 1000 to 10000 eV or more may be used, increasing the available flux by up to three orders of magnitude. The only requirements for the implementation of this method are a synchrotron beamline capable of providing white (unmonochromatized) radiation, and a slightly modified synchrotron SAXS camera. The idea of using white light in SAXS experiments was originally suggested by Stuhrmann (1978), but it appears that no treatment or attempted implementation has been reported prior to the work described here.

Numerical Treatment:

The validity of the white light technique rests on the fact that small angle scattering data collected using very broad bandpass radiation still obey Guinier's law (*i.e.* are still Gaussian) when the angles involved are 'sufficiently small'. The dependence of measured intensity on the scattering angle θ in any SAXS experiment using a distribution of photon energies $P(\omega)$ is just the integral over that distribution of the monochromatic intensity.

$$I(\theta) = \int I(\omega, \theta) P(\omega) d\omega$$

At sufficiently small angles, as discussed in Chapter 1, the classic result by Guinier tells us that the monochromatic intensity will take the form of a Gaussian, $e^{-K^2 R_g^2/3}$. It is easy to show that to first order, substituting Guinier's result into the above integral yields the following expression, which is also a Gaussian, independently of the exact nature of the photon energy distribution.

$$I(\theta) = e^{-R_g^2 \int 16\pi^2 \sin^2(\theta/2) P(\omega) d\omega / 3\lambda^2}$$

By using a generalized definition of the monochromatic scattering vector magnitude, $K = 4\pi \sin(\theta/2)/\lambda$, we can rewrite this expression as

$$I(\langle K \rangle) = e^{-R_g^2 \langle K \rangle^2 / 3} \quad \text{where } \langle K(\theta) \rangle^2 = 16\pi^2 \sin^2(\theta/2) \int P(\omega) d\omega / \lambda^2$$

We see that the integrated intensity still obeys Guinier's law if we choose the right definition for $\langle K \rangle$. It is crucial to note that the 'correct' value of $\langle K \rangle$ depends only on the energy distribution of the incoming photons, and is thus the same for all samples studied during a particular experiment. This means that $\langle K \rangle$ can be determined by calibration of data from a well characterized sample to monochromatic data. Of course, $\langle K \rangle$ could also be determined by direct measurement of the photon energy distribution, but this tends to be much harder.

In order to evaluate the usefulness of the above observation it is necessary to know what angles are 'sufficiently small'. Stuhrmann (1978) gives the deviation of R_g measured using a Gaussian energy distribution, using the central value for K , and using the standard Guinier region of $K R_g < 1$, as $R^2 = [1 - (\Delta\lambda/\lambda)^2/2] R_{\text{exp}}^2$ (R is the true value and R_{exp} is

the measured value), showing that even for large widths the error is only a few percent. Doniach (1991) has also shown that for a flat energy distribution and Gaussian shaped or spherical scatterers, the deviations of very broad bandpass data from a Gaussian in the Guinier region of corresponding monochromatic data lead to only a few percent errors in R_g . Such analytical treatments are useful for establishing the validity of the white light approach.

As a complement to the above results a simple program was written to integrate arbitrary SAXS patterns over arbitrary energy distributions. The energy distribution and SAXS data are input at discrete values of E and K. The SAXS data are decalibrated from K to θ at each of the discrete energy values. Each pattern is then interpolated, using a spline fit, onto to the θ values of the lowest energy. Integration at each θ value which is available at all energies is performed using a simple extended trapezoidal rule algorithm, and the integrated data are calibrated using $\langle K \rangle$ as calculated from the energy distribution. The splining and integration algorithms were used as given in the book Numerical Recipes (Press, 1992).

Two different photon energy distributions were chosen to perform model integrations, a flat energy distribution, and an energy distribution which was calculated using a spectrum measured on an actual beamline (4-2) at SSRL, attenuated by the calculated absorbance of a 250 μm aluminum filter and the calculated sample absorbance. The attenuation factors were calculated using the Absorption program, written by Sean Brennan of SSRL, and based on cross-section calculations by Cromer, Liberman, and McMasters. Lower and upper limits of 8 and 12 keV were used to reflect the absorption and mirror cutoffs on beamline 4-2. The calculated flux gain for the two distributions (compared to the normal bandpass of 10 eV) is a factor of 400 for the flat distribution, and 15 for the attenuated distribution. The flat energy distribution is actually a good approximation for the direct synchrotron radiation beam, before any optical elements. The attenuated energy distribution is a model for the actual energy distribution expected in a real experiment, with the addition of an absorber to provide a low-energy cutoff and to allow for longer (and thus more easily controlled) sample exposure times, without worrying about excessive heating of and ensuing damage to the samples. A plot of the second energy distribution is presented in Figure 4.1. The calculated values of $\langle E \rangle$ for the flat and attenuated distributions are 10106 and 9873 eV, respectively.

The results of the numerical integration of an actual (measured) monochromatic SAXS pattern, collected for ribonuclease A, over the above energy distributions, using a 10 eV step-size, are presented in Figures 4.2 and 4.3. Figure 4.2 compares the monochromatic scattering data with the predicted broad bandpass data as a function of detection angle. Note that the experimental noise evident in the monochromatic data is smoothed out in the integration process. Figure 4.3 compares the data sets after calibration into reciprocal space, where K was used for the monochromatic data, and $\langle K \rangle$ was used for the very broad bandpass data. This figure illustrates how small the differences between the monochromatic and white light data sets are at these angles. The R_g values obtained from fits to these data are 15.48 for the monochromatic data, 14.24 for the flat distribution white light data, and 14.87 for the attenuated distribution white light data, using a K or $\langle K \rangle$ range of 0.038 to 0.063 \AA^{-1} . Part of the discrepancy between these values can probably be attributed to the 0.4 \AA standard deviation in the R_g value from the (noisy) monochromatic data. This procedure has also been carried out for several other proteins, with similar results.

The above analytical and numerical treatments make it clear that angles which are routinely accessed in monochromatic SAXS experiments are in fact already 'sufficiently small' to allow for the extraction of reasonably accurate R_g values from white light SAXS data. Measurements of I_0 ratios should clearly also be unaffected by the use of white light. Therefore, this technique holds much promise for providing a higher flux for time-resolved and other SAXS experiments.

Experimental Details:

In order to further test the validity of this approach, a series of preliminary measurements were performed at SSRL in March of 1992 on beamline 10-2, which is capable of providing white light. The beamline mirror was used to provide a high energy cutoff of around 22 keV and no filters were used. The power of the full white beam was quite high, and to prevent extreme heating of the samples, the beam had to be apertured using slits placed before the focussing mirror. A standard two-slit collimation arm was used, but because of the high flux and higher energy photons, a large amount of parasitic scattering was observed in the data. Future experiments should use a minimum of three collimating slits.

Concern over possible damage from the high flux to the conventional gas-filled proportional counter detector normally used in SSRL SAXS experiments led to the use of imaging plates (essentially sheets of plastic film) for data collection. Unfortunately, the available system (manufactured by Fuji Film Corporation) has never before been used for SAXS measurements, and some uncertainty remains regarding its adequacy. Several different imaging plates were employed to save time, and the reproducibility of data between different plates has not been tested. In addition, plates had to be removed and replaced between sample exposures. A carefully constructed (by Paul Phizackerley of SSRL) cartridge was used to insure, as much as possible, an identical placement and replacement of each plate, but this was still a less satisfactory arrangement than a single fixed detector.

Samples were contained in 20- μ m thick quartz-window cells. A series of proteins were studied in order to provide a test of the energy calibration procedure. Data were collected from ribonuclease A in 50 mM formate buffer, pH 4.0, myoglobin and hemoglobin in 50 mM phosphate buffer, pH 6.0, and phosphorylase *b* in 50 mM pipes, 5 mM mercaptoethanol, pH 6.8. All proteins were used as purchased from the Sigma Chemical Company without further purification. Samples were prepared at concentrations of 5 and 1 g/l. Exposure times ranged from 0.1 to 1.5 seconds, and were generated using an electronically controlled lead shutter borrowed from Paul Phizackerley of SSRL. Such short times were used in order to minimize the potential for sample damage from the intense beam. Even so, small amounts of visible aggregation were observed on the sample cell windows after a few of the longer exposures. This observation makes it likely that all of the samples experienced at least some degree of damage, and probably some accompanying aggregation.

Analysis and Results:

Data were read from the imaging plates using a scanner manufactured by Fuji Film Corporation, and stored as two-dimensional matrices of intensities. Normalization measurements could not be performed as usual because no ionization chamber was used in the experiments, in an attempt to minimize the air-path of the beam. Instead, the synchrotron ring current was recorded during each exposure, and its value multiplied by the exposure time was used to normalize the data. After normalization, background data were subtracted from sample data. This subtraction relied on the mechanical alignment of the imaging plates, as the two data matrices were simply subtracted at each position.

After background subtraction, the center of the scattered intensity was determined as the point around which the data were most symmetric, and the data were circularly averaged around this center.

The normalized subtracted averaged data from all samples showed a large amount of residual scattering at low angles. The source of this scattering could be sample aggregation, beam fluctuations between sample and background data collection, or inadequate background subtraction of the strong parasitic scattering. All three of these factors probably played some role. A second indication of the poor quality of the data was the failure of data collected from the same sample at different (and even the same) concentrations to overlap when normalized by the sample concentration. The data collected at the lower sample concentrations (1 g/l) showed significantly poorer background subtractions and was discarded from further analysis. The scattered intensity also did not correlate well with the mass of the samples, and the shapes of the scattering curves varied considerably.

During the circular averaging process, an evaluation was made of the lowest angle at which data along different directions overlapped. A plot of the data beyond this point for all of the more concentrated samples is presented in Figure 4.4. Despite all of the above comments the higher molecular weight scatterers are readily identifiable by their higher narrower shape. With this encouraging observation in mind, the search for a proper value of $\langle E \rangle$ was begun. Because the Guinier region of smaller proteins extends to higher angles, and because more previous data is available for the two smaller proteins studied (ribonuclease A and myoglobin) these were used to calibrate the data. A value of 16 keV for $\langle E \rangle$ was found to produce reasonable R_g s for both ribonuclease A (14 Å) and myoglobin (17 and 19 Å) from a standard Guinier analysis using a $\langle K \rangle$ fitting range of 0.049 to 0.077 Å⁻¹. R_g values using the same fitting range on monochromatic data from these proteins are 15 and 17 Å. The R_g values obtained from the white light hemoglobin and phosphorylase *b* data were 22 and 25 Å, and 27 Å, respectively. These values are much lower than the true R_g values, but this is expected because the angular fitting range is too high for a reliable Guinier analysis of either of these larger proteins.

Monochromatic SAXS data from phosphorylase *b* give an R_g of 45 Å using a K range of 0.020 to 0.034 Å⁻¹. When the monochromatic phosphorylase *b* data is analyzed using the above $\langle K \rangle$ range, the apparent R_g value is 34 Å, closer to the value obtained using the same range on the white light data. In Figure 4.5, the monochromatic data from

phosphorylase *b* are compared with the same data integrated over a flat spectrum from 8 to 23 keV and calibrated with $\langle E \rangle = 16093$ eV. Because of the large energy range used, and the larger size of the protein, the calculated white-light pattern differs significantly from the monochromatic pattern. Fitting the calculated white light data using the lower angle fitting range (actually, a slightly higher range, 0.025 to 0.034, must be used, because the calculated white-light data start at the same value of θ , and hence at a larger value of $\langle K \rangle$) gives an R_g of 44 Å in good agreement with 45 Å. Amazingly, fitting the calculated white light data with the $\langle K \rangle$ range used in fitting the experimentally determined white light data gives an R_g of 28 Å, quite close to the value of 27 Å obtained from the experimental data. Thus the measured data appear to be in agreement with the numerical calculation. When the same integration and calculation is performed for ribonuclease A (using $\langle K \rangle = 0.065$ to 0.077), the predicted R_g value is 14 Å, still in (very) good agreement with the measured 14 Å.

Conclusions:

The numerical results presented and the referenced analytical calculations clearly show that white light SAXS measurements can in principle yield reasonably accurate R_g values. The poor quality of the data collected in the initial studies casts some doubt on any conclusions which may be drawn from the analysis. Nevertheless, one can not overlook the fact that the R_g values measured for three proteins of different sizes are in quite reasonable agreement with the predicted values, once the appropriate factors are taken into account. Considering the unfavorable conditions under which the data were collected, these results are most encouraging, and support the validity of the simple numerical and analytical models. Considerable optimism seems warranted for future experiments performed under more optimal conditions.

One serious problem which still needs to be addressed is the prevention of and/or evaluation of sample damage. This is a long-standing issue relating to all synchrotron experiments (monochromatic as well as white light), but the delicate samples and extremely high flux make it especially important in white light protein studies. The most reasonable way to combat the effect of the increased flux in these experiments is to reduce the total exposure time. Although sample damage is known to be a function of both integrated and instantaneous flux, this should still help considerably. Time-resolved studies of sub-millisecond processes inherently incorporate this approach, and are therefore particularly suitable for the white light method. In contrast, studies of very

dilute samples require longer exposure times, and would therefore require more sophisticated methods for controlling sample damage.

References:

Doniach, A. 1991. A Note on Laue Small Angle X-ray Scattering (LSAXS). Private Communication.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. 1992. Numerical Recipes. Cambridge, New York. pp. 963.

Stuhrmann, H.B. 1978. Quarterly Reviews of Biophysics. 11:71-98.

Table 4.1:
White Light Radii of Gyration.

Sample	Exposure (s)	R_g (Å)	σ
myoglobin	0.1	16.98	0.32
myoglobin	0.2	18.97	0.22
hemoglobin	0.5	25.23	0.32
hemoglobin	0.1	22.13	0.25
phosphorylase <i>b</i>	0.2	26.68	0.16
ribonuclease A	0.5	14.22	0.09

Results from Guinier fits to the normalized, background-subtracted averaged white light data over $\langle K \rangle = 0.049$ to 0.077 \AA^{-1} . All samples were prepared at 5 g/l. All background exposures were recorded for 1.5 seconds.

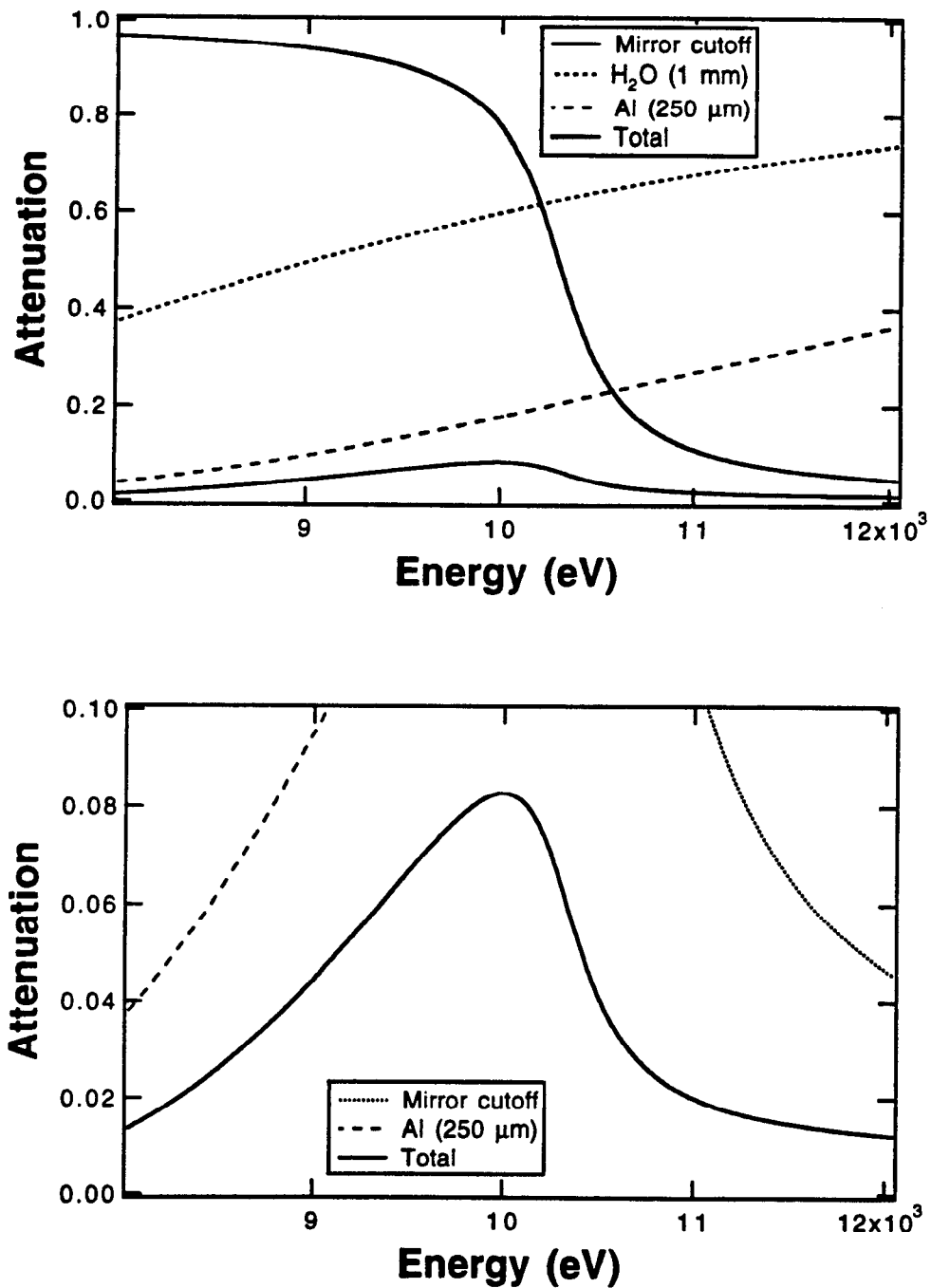


Figure 4.1: The attenuated photon energy distribution calculated using a measured spectrum from SSRL beamline 4-2, the calculated absorption of a 250 μ m aluminum filter, and the calculated sample absorption (approximated as 1 mm of H₂O). The lower panel is simply an enlargement of the upper panel.

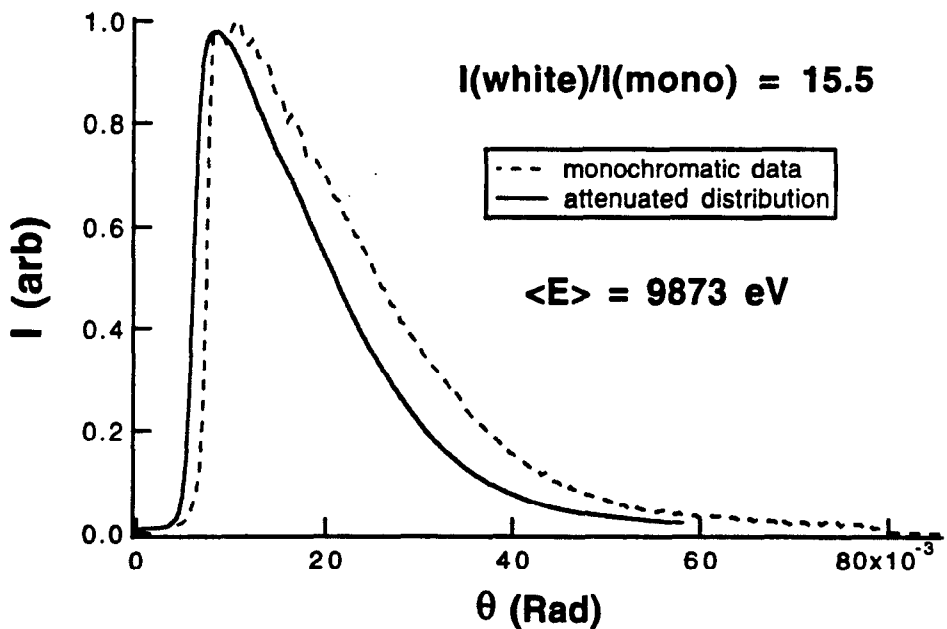
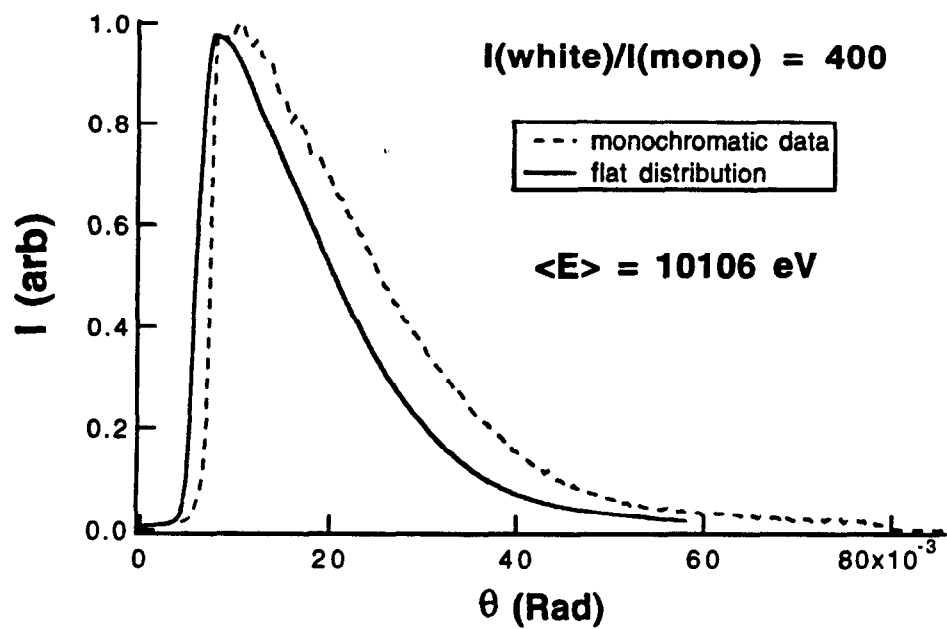


Figure 4.2: Monochromatic and calculated white light data as a function of scattering angle for ribonuclease A. The calculation procedure is described in the text. Both a flat and an attenuated energy distribution were used.

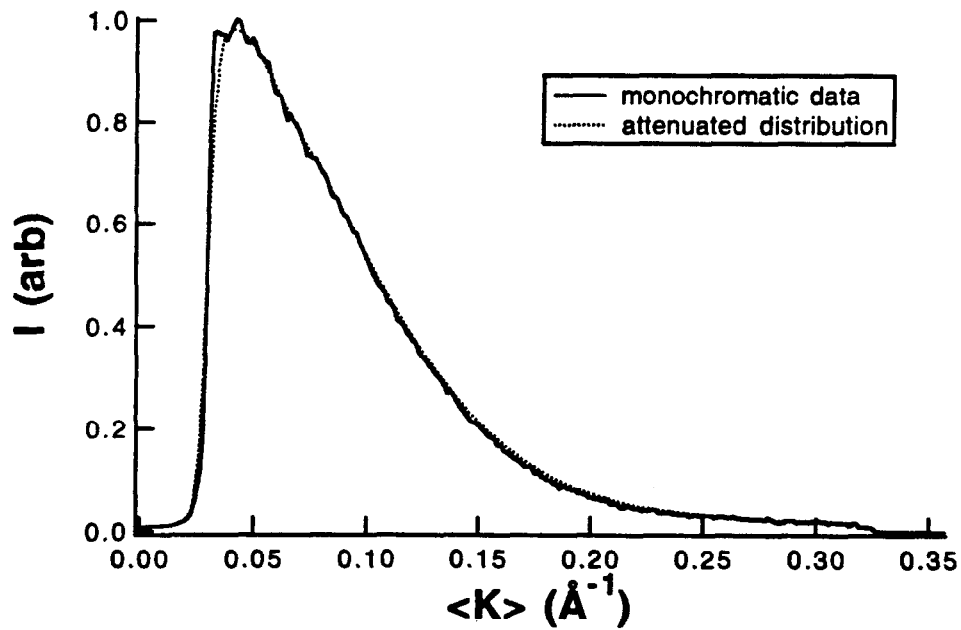
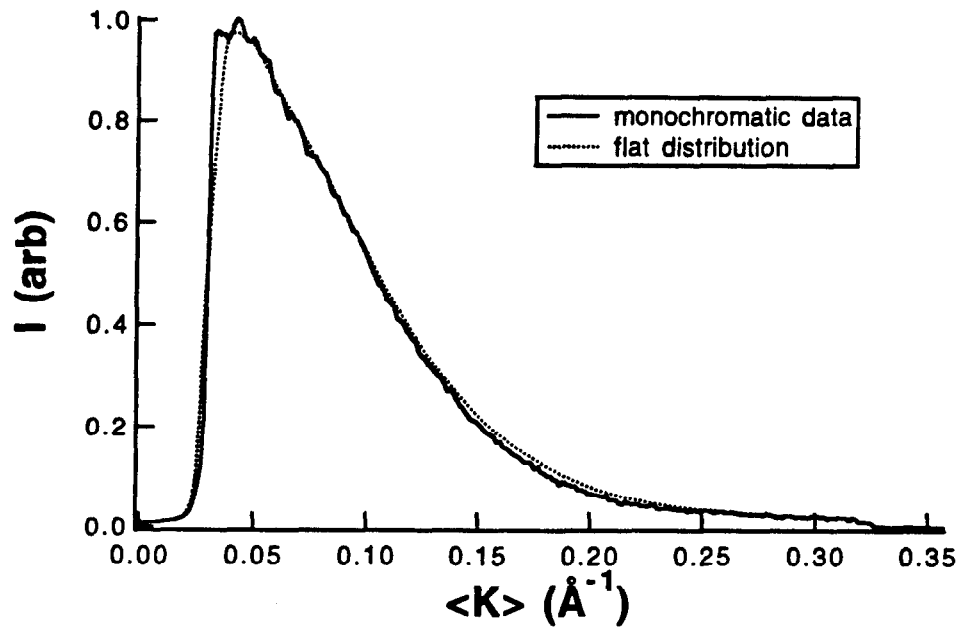


Figure 4.3: Monochromatic and calculated white light data as a function of $\langle K \rangle$ for ribonuclease A. These are the same data as in Figure 4.2.

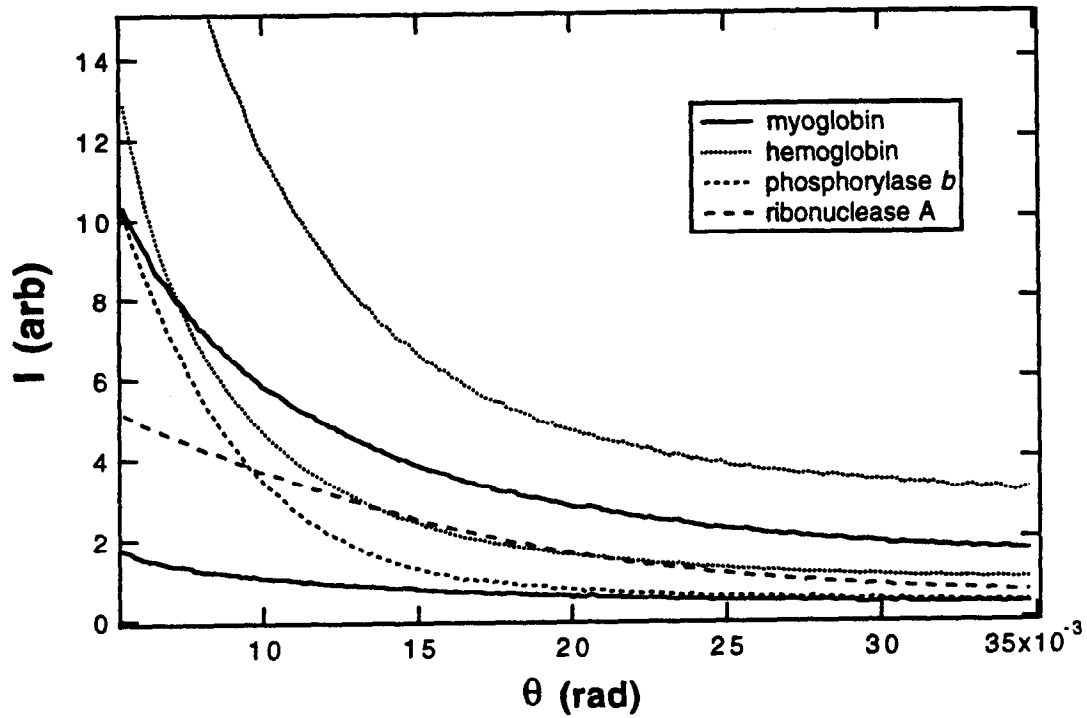


Figure 4.4: White light SAXS data measured on SSRL beamline 10-2 for four proteins. The lowest angle at which the data were judged to be of a reasonable quality was around 5 milliradians. The results of a Guinier analysis of these data are given in Table 4.1.

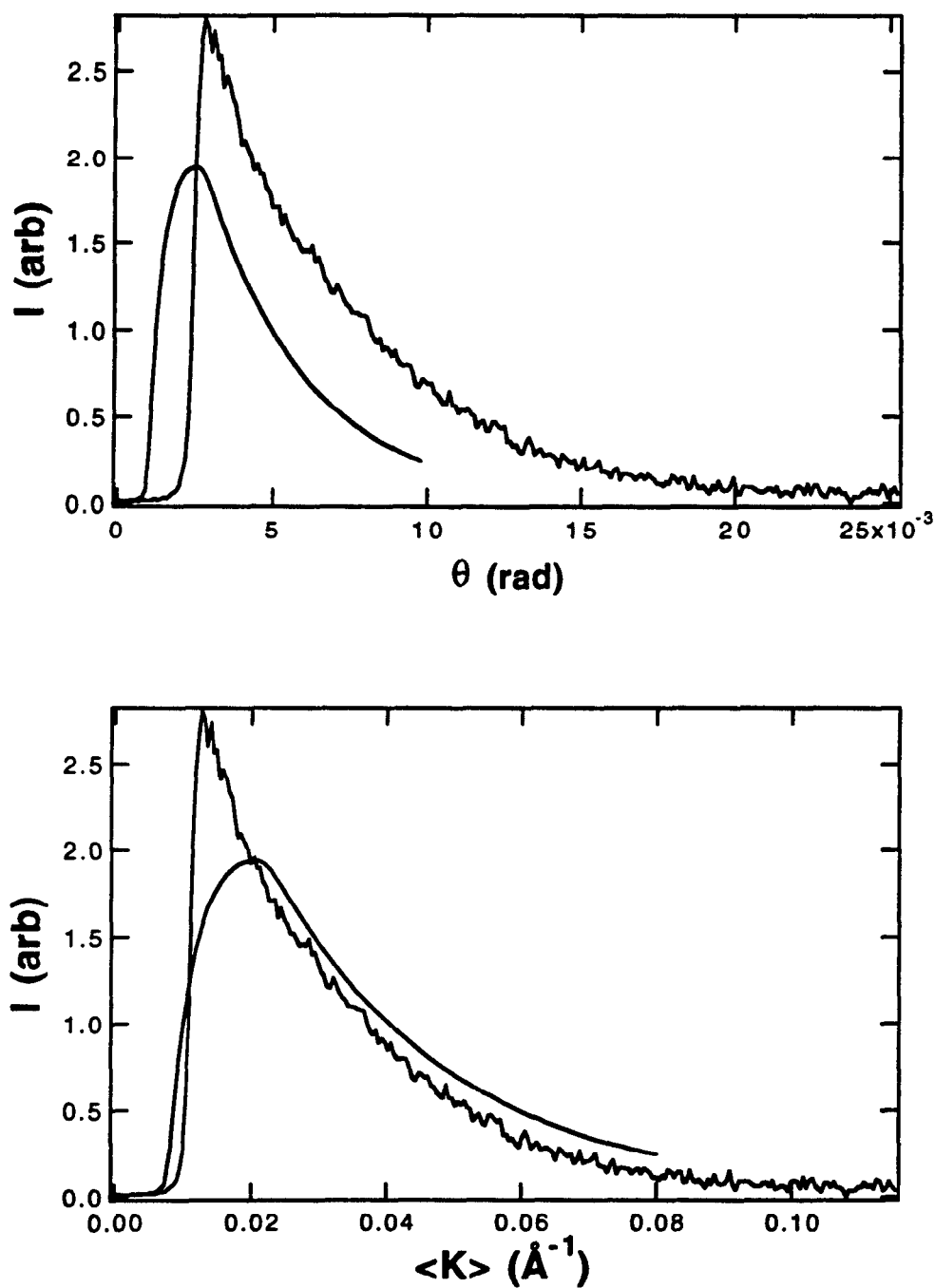


Figure 4.5: Monochromatic and calculated white light data for phosphorylase *b*. The calculated data were obtained using a flat photon energy distribution from 8 to 23 keV to model the energy distribution on SSRL beamline 10-2. The white light data were calibrated from θ (upper plot) to $\langle K \rangle$ (lower plot) using $\langle E \rangle = 16093$ eV.

Conclusion and Credits

The accurate measurement of small-angle x-ray scattering is a difficult experimental challenge, with a relatively low ratio of information gained to effort expended. In general either crystallographic or a variety of spectroscopic techniques can yield far more detailed information. Nevertheless, there are certain situations for which SAXS, and particularly synchrotron SAXS, is uniquely suitable. This dissertation illustrates the utility of the technique in two such situations, the study of the size of dilute non-crystallizable materials in solution, and the study of dynamic processes in solution. SAXS is also uniquely sensitive to aggregation or association phenomena, and the studies herein have illustrated the ubiquity of such phenomena in solution samples, and the importance of the ability to detect them.

The FeMoco studies described in chapter 2 were technically challenging for three primary reasons: the extreme sensitivity of the sample itself to the preparation and handling process, the lack of flat window cells capable of containing the sample in an oxygen free environment, and the weakness of the scattering, resulting from the small size and limited solubility of the sample. In spite of these difficulties, the results revealing the aggregation of FeMoco in solution will be valuable in the interpretation of past and future data on the behavior of this important protein metallocluster. In addition, the experiments may serve as a protocol for the testing of FeMoco and other solution samples for the presence of aggregates in order to determine conditions suitable for crystallization. The studies have also illustrated the feasibility of using SAXS to obtain information about even very small samples.

The time-resolved protein folding studies are the main focus of and the greatest success of my graduate work. Despite the failure to observe the actual collapse of any of the proteins studied, an interesting and unexpected process, transient dimerization, was discovered during the refolding of both myoglobin and cytochrome-*c*. This discovery implies the presence of kinetic molten-globule folding intermediates in the refolding pathways of these two proteins. This result has long been suspected for proteins in general, but supporting evidence has been difficult to obtain. In addition, the observed bi-phasic nature of the dissociation kinetics of the cytochrome-*c* folding intermediate verifies previous reports of heterogeneity in the unfolded state of this protein. Because of the specific sensitivity of the SAXS forward scattering data to one parameter (molecular weight) SAXS measurements of the fractional populations of the different unfolded states

may prove less ambiguous than those determined by spectroscopic probes, which are sensitive to a number of different effects. The success of these pioneering studies illustrates the great potential of the time-resolved SAXS technique for the study of the protein folding problem.

The white light SAXS studies, although quite preliminary, provide a tantalizing glimpse of what may be accomplished using this technique. All of the data presented in that section were collected in under 10 seconds total of actual sample exposure. Although experiments in the immediate future should concentrate on using a lower flux to produce higher quality data and verify the validity of the method, it seems reasonable to hope that later experiments can be performed with a thousand-fold increase in time-resolution, probing processes on the tens of microseconds time scale. Studies of the protein folding problem on this time-scale should provide very interesting information indeed, perhaps including at last the much sought after measurements of the time scale of the protein compaction process. Unfortunately, there remains a serious obstacle to such studies beyond verification of the validity of the white light approach. In order to study sub-millisecond processes it is necessary to have a sub-millisecond trigger. Rapid-mixing seems to be limited by mixing dead-times of about 1 ms. Laser or capacitive-discharge induced temperature jumps may be feasible, but may cause unacceptable levels of sample damage. Pressure jump is an attractively noninvasive trigger, but may not be compatible with the thin-window requirement of SAXS sample containers.

In relation to the FeMoco studies, credit should be given to William E. Newton for the isolation of the FeMoco samples, to Patrick Frank for the preparation and handling of the samples, to Nanna Gillis for originally starting the project and for the collection of the NSLS data, and to Hirotsugu Tsuruta for extensive help with upgrading the SSRL biotechnology SAXS camera.

Credit for contributions to the protein folding work should be given to Robert L. Baldwin for originally suggesting the approach and for assistance with and advice on the ribonuclease A studies, to Virginia Robbins for the purification of ribonuclease A, to Hiroshi Kihara for suggesting and assisting in the myoglobin studies and for providing the stop-cock rapid-mixer, to Kaori Chiba for providing the conditions initially used (but later modified) in the myoglobin experiments, to Yoshiyuki Amemiya for help with the SAXS instrumentation at the Photon Factory, to Gennady V. Semisotnov for useful discussions regarding the interpretation of the data, to Hirotsugu Tsuruta for help in using both of the

stopped-flow rapid-mixers and for help with the data collection, and to Lingling Chen for the development of the mica-window sample cells.

Credit should be given to Hirotsugu Tsuruta for help with the experimental setup, sample preparation, and data collection, and to Lingling Chen for help with sample preparation in the white light experiment.