# STUDIES OF PROTEIN STRUCTURE IN SOLUTION AND PROTEIN FOLDING USING SYNCHROTRON SMALL-ANGLE X-RAY SCATTERING

Lingling Chen

Stanford Linear Accelerator Center
Stanford Synchrotron Radiation Laboratory
Stanford University, Stanford, California 94309
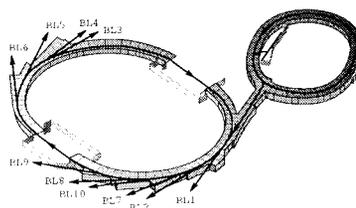
SLAC-Report-498
April 1996

*Ph.D. Thesis Stanford University, Stanford, CA 94309

STUDIES OF PROTEIN STRUCTURE IN SOLUTION AND PROTEIN FOLDING USING SYNCHROTRON SMALL-ANGLE X-RAY SCATTERING

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

By
Lingling Chen
April 1996

Abstract

Synchrotron small angle x-ray scattering (SAXS) has been applied to the structural study of several biological systems, including the nitrogenase complex, the heat shock cognate protein (hsc70), and lysozyme folding. The structural information revealed from the SAXS experiments is complementary to information obtained by other physical and biochemical methods, and adds to our knowledge and understanding of these systems.

Nitrogenase complex from *Azotobacter vinelandii*. The nitrogenase enzymatic system consists of two O_2-sensitive metalloproteins: the iron protein (Fe protein) and the molybdenum-iron protein (MoFe protein). The Fe protein transfers electrons from electron donors to the MoFe protein, and hydrolysis of ATP is required for the electron transfer. The MoFe protein contains a active site called FeMo cofactor where N_2 reduction occurs.

To investigate the effect of nucleotide (MgATP and MgADP) binding on the structure of the Fe protein, SAXS experiments have been carried out on the wild-type and a Nif⁻ mutant (UW91, which has an alanine157 mutation to serine) Fe proteins of nitrogenase from *Azotobacter vinelandii*. For the wild-type protein, the binding of MgATP induces a significant conformational change as evidenced by a decrease of about 2.0 Å in the radius of gyration (R_g). In contrast, the binding of MgADP to the wild type Fe protein does not detectably affect R_g. In the absence of nucleotides, the R_g for the UW91 mutant is indistinguishable from that of the wild type. However, unlike the wild type protein, the R_g of the UW91 Fe protein is not affected by the addition of MgATP. Because the UW91 protein has a normal [4Fe-4S] cluster and MgATP binding ability but both the electron transfer and MgATP hydrolysis are completely blocked, these x-ray scattering measurements suggest that a conformation different from that of the native state is required for the Fe protein to perform electron transfer to the MoFe protein. These results also support the hypothesis that residue alanine 157 is crucial for the iron protein to establish the electron-transfer-favored conformation induced by MgATP binding.

As a first step to study the interaction between the Fe protein and the MoFe protein during electron transfer, scattering data for the MoFe protein have also been taken, and the R_g is measured to be 39.8 ($\pm$ 0.4) Å. A FeMo cluster, a different form of

the FeMo cofactor, is mostly in homogeneous monomeric form in N-methylformamide solution, with a R_g of 3.6 ($\pm$ 0.4) Å. However, the FeMo cluster appears to highly aggregate in methyl-ethyl ketone solution.

The 70 kDa heat shock cognate protein hsc70. The bovine heat shock cognate protein (hsc70) is a constitutively expressed member of the molecular chaperones that are necessary for cell viability in all organisms. Solution x-ray scattering experiments have been carried out on recombinant bovine hsc70 (with 650 amino acid residues), a 60 kDa subfragment (residues 1-554) which has ATPase and peptide binding activities, a 44 kDa subfragment (residues 1-386) which has only ATPase activity, and a peptide binding fragment (residues 388-554). Modeling based on the steady-state values of R_g and P(r) functions show that the ATPase and peptide binding domains are oblate fragments, while hsc70 and the 60 kDa fragment are prolate and relatively elongated. R_g values of hsc70 and the 60 kDa fragments decrease significantly in the presence of MgATP relative to their values in the presence of MgADP (ΔR_g ~4-5 Å); in contrast, the R_g of the ATPase fragment remains essentially the same in the presence of either nucleotide. The kinetics of the change of R_g for hsc70 and the 60 kDa fragment under single-ATPase-cycle conditions show that the transition to the ATP-induced R_g occurs significantly more rapidly than ATP hydrolysis, while the reverse transition to the larger R_g value does not occur before product release. Altogether, the solution scattering data support a model in which a conformational change in hsc70 (presumably to the low peptide affinity state) is predicated on ATP binding, while the reverse transition is predicated on product release.

Equilibrium lysozyme unfolding. Equilibrium unfolding of hen egg lysozyme as a function of urea concentration at pH 2.9 and pH 2.0 has been studied by solution x-ray scattering. A difference in the cooperativity of the unfolding as monitored by the radius of gyration, R_g, is observed relative to that seen from the CD signal at 222 nm. This suggests the existence of a third scattering species, in addition to the native and the unfolded states. A singular value decomposition (SVD) analysis was applied to the scattering curves at different urea concentrations. This analysis shows clear evidence of a third basis component in the x-ray scattering curves, thus supporting the results of the R_g and CD measurements. The denaturant binding model was employed to estimate the thermodynamic parameters of denaturation for the intermediate and unfolded states. Use of these parameters to refine the SVD analysis allows us to reconstruct a scattering profile for the pure intermediate state. Simplified partially folded models based on the crystal

structure of hen lysozyme support a working model for the intermediate whose structure may be correlated with that of the kinetic intermediate found in the refolding pathway.

Kinetics of lysozyme refolding. The refolding process of lysozyme has been followed by time-resolved SAXS combined with the stopped-flow method. At the earliest measurable time (100 ms after initiation of refolding), R_g has decreased from 22 Å to ~ 19 Å. A fast process has been found to occur within the deadtime of the experiment (100 ms), which accounts for > 30% of the total change in R_g between the unfolded and the native states. The time constants (τ) for the slow process have been derived for different refolding conditions. The biphasic kinetics suggests that a folding intermediate forms within 100 ms in the pathway. The forward scattering intensity ($I(0)$) has been found to remain steady within experimental error throughout the refolding process, indicating that the change in R_g observed is due to the conformational change in the refolding process, rather than a change in aggregation state.

Acknowledgments

My graduate career at Stanford has been a tremendously rewarding experience. Many people have contributed in different ways to this work, and it is impossible to express my gratitude to everyone of them individually.

First, I would like to thank my research advisor, Keith Hodgson, for his generous support and encouragement throughout my research. Keith's patience and faith in my ability allowed me to develop independent thinking, and further my interest in biophysical field. I appreciate the opportunities he has provided to visit different laboratories including Photon Factory in Japan, and to present the projects that I worked on at conferences locally and internationally. I would like to acknowledge Seb Doniach for his insights which provide a step further in understanding experimental results. Seb's scientific enthusiasm and innovative ideas have been the inspirations to me during my research. I feel privileged to work under both Keith and Seb.

I would like to thank Britt Hedman and Pat Frank, two marvelous people in the Hodgson group. They are the resources that I can always rely on whenever I encountered difficulties, scientifically and personally. I am particularly grateful to Britt for proof reading this thesis, and for her perspectives and being a role model as a female scientist. I have benefited a great deal from Pat on a daily basis. He taught me step by step purification of nitrogenase, and is never tired of answering my constant questions. Pat also serves as my English tutor, commenting on my papers and part of this thesis, as well as teaching me American culture and history. I still owe him a jar of jelly worms from Xiamen (Fujian)!

Hiro Tsuruta and David Eliezer introduced me to small-angle x-ray scattering, and we had been working closely over the years. Hiro had always been the responsible personnel when I had emergency at the beamline. I would like to thank both of them for sharing with me the stressful and 'rewarding' experience with synchrotron experiments. It is a great pleasure to work with them. Dan Segel joined the group two years ago, and has made immediate contributions to SAXS. Dan is always happy to discuss any subjects and to offer helps. I would like to thank him for reading part of this thesis, for his American idioms, and most importantly for his friendship. I wish both him and Mike Bada, who just joined the group a few days ago, the best in their future experiments and successful graduate careers.

I would like to thank the rest of Hodgson members, past and present, Heather Bufford, Jason Chen, Jane DeWitt, Isaac Liu, Kristin Musgrave, Kent Nakagawa, Susan Shadle, Chrisie Stanfel, Grace Tan, Trevor Tyson, Erik Wasinger, Kendra Williams, Tami Westre, Holly Zhang, for creating a most harmonious and pleasant working environment. Being my neighbor, Chrisie has 'endured' the most of my complaints, and has always been willing to help me through those difficult moments. I have had a wonderful time working with members of Boxer group, sharing chemicals and instruments, and enjoying their great barbecue parties.

I am thankful to have the opportunity to work collaboratively with Burgess and McKay laboratories. Not only did they supply the high quality samples, but they also worked closely with me days and nights on the beamline, responsible for any sample emergency. Their pleasant cooperation has kept our projects going as smoothly as they could be. I would like to thank Dave McKay for his critical scientific attitude and his encouragement.

The SSRL staff have provided invaluable technical support. The x-ray group made extra efforts working on a stable and focused beam, and the SSRL machine shop did an excellent job on making sample cells and beamline components.

I would like to give special thanks to my previous advisors, Prof. Hui-Lin Wan, and Prof. Qi-Rui Tsai at Xiamen University, China. Prof. Wan guided me to the field of Physical Chemistry, and showed me the significance of doing research. I am indebted to his constant support, both psychological and academic, of my education at Xiamen University. I am grateful for their visits during my stay at Stanford.

A great deal of credits go to friends who have provided encouragement as well as entertainment all the way. I would like to thank my English tutor Jane Rebber who introduced me to American culture, and helped me to improve my English. Jian-Yi Lin and Leping Li are among the people who helped me to settle at Stanford. Ming-Qian Ma serves as my mentor in various aspects. Friends, far and close, Xiaoping Song, Jiahua Che, Kaiqin Lao, Huilin Zhou ... are truly dependable. A special thank goes to the Bromberg family, particularly our little friends, Leah, Luke, Nathaniel and Paulina for their hospitality. I will always cherish the friendships we share.

To my Mother, Xiuli Liao, and the memory of my Father, Bingnan Chen

Table of Contents

List of Tables

List of Figures

Chapter 5

List of Abbreviations

Å	Ångström
ADP	adenosine diphosphate
AMPPNP	5'-adenylyl-β, γ-imidodiphosphate
ANS	8-anilinonapthalene-1-sulfonate
ATP	adenosine triphosphate
CD	circular dichroism
Da	dalton
DOE	Department of Energy
GdmHCl	guanidinium hydrochloride
hsc70	70 kDa heat-shock cognate
I(0)	forward scattering intensity at zero scattering angle
MEK	methyl-ethyl ketone
MW	molecular weight
NIH	National Institutes of Health
NMF	N-methyl formamide
R_g	radius of gyration
SAXS	small-angle x-ray scattering
SSRL	Stanford Synchrotron Radiation Laboratory
SVD	singular value decomposition

Chapter 1

Introduction and Background

1.1. Scope and Organization of This Thesis

This dissertation focuses on the applications of synchrotron small-angle x-ray scattering (SAXS) to the study of conformation and structure of biological systems in solution. Chapter 1 provides a general introduction to the small-angle x-ray scattering technique, including a brief theoretical background, information content from SAXS, and experimental considerations. This is followed by four chapters describing detailed experiments, data analysis and results, focusing on the studies of three topics — the nitrogenase enzymatic system, the heat-shock cognate protein hsc70, and lysozyme folding. A brief introduction is provided in each of these chapters focusing on the aspects of the project to be studied. Chapter 2 presents the results of conformational studies of the Fe protein, the MoFe protein and a FeMo cluster of *Azotobactor vinelandii* nitrogenase. Chapter 3 describes the structural studies of a chaperone, hsc70, and its subfragments, concentrating on the development of low resolution models for hsc70 and the kinetics of conformational change in the ATPase cycle. Both Chapter 4 and Chapter 5 concerns the application of SAXS to studies of protein folding. Results of equilibrium unfolding of lysozyme are given in Chapter 4, and those of a time-resolved SAXS study of lysozyme refolding kinetics in Chapter 5.

1.2. Overview of X-ray Scattering Theory

When x-rays pass through matter, they interact with both electrons and nuclei. However, as we can see below, the scattering of x-rays by matter comes primarily from the interaction of x-rays with electrons. This is because the scattering length of a nucleus is less than 10^{-3} of that of an electron so that the nuclear scattering can be neglected.

Incident radiation can be represented by a plane monochromatic incident wave: $\bar{E}(\bar{r}) = E_0 \exp(2\pi i \bar{k}_0 \bullet \bar{r} - \omega t)$, where $\bar{E}$ is the electric field at position $\bar{r}$ with maximal amplitude E_0; $\bar{k}_0$ is the unit vector of the incident wave; and ω is the angular frequency. Suppose an incident x-ray beam ($\bar{k}_0$) impinges on a free electron positioned at the origin O, then the amplitude of the scattering wave ($\bar{k}$) is given by the following equation (Landau & Lifshitz, 1987):

$$E_s = -E_0 \frac{e^2}{mc^2} \frac{1}{r} \sin \psi \qquad [1]$$

where e and m are the charge and mass of the electron; c is the velocity of light; and ψ is the angle between the scattering beam and the direction of acceleration of the electron. Note that the phase of the scattering wave is shifted by 180° relatively to the incident beam.

For small-angle scattering, $\psi = 90^o$. If $b_x = e^2 m^{-1} c^{-2} \sin \psi = e^2 m^{-1} c^{-2} = r_0$, where $r_0 = 2.82 \times 10^{-13}$ cm — the Thomson electron radius, then b_x is called the x-ray scattering length of a free electron. The scattering cross section is defined as $i_x = b_x^2 = r_0^2 \sin^2 \psi = r_0^2$. For a single electron, $i_x = 7.96 \times 10^{-26}$ cm^2. From the above, it is thus clear that the scattering length is inversely proportional to the mass of the scatterer, and that the scattering amplitude by nuclei, which are more than 10^3 times heavier than electrons is negligible.

Now consider a free electron positioned at $\bar{r}$. The scattering wave becomes $E(\bar{S})e^{2\pi i \bar{S} \bullet \bar{r}}$, where $E(\bar{S})$ is the electric field of scattered radiation by a free electron at the origin with the amplitude E_s as defined in equation [1], $\bar{S} = \bar{k} - \bar{k}_0$, where $|\bar{S}| = 2 \sin \theta / \lambda$, and λ is the wavelength. The structure factor, $F(\bar{S})$, is defined as the ratio of the radiation scattered by any real sample to that scattered by a free electron at the origin. In the case of a free electron at $\bar{r}$, $F(\bar{S}) = e^{2\pi i \bar{S} \bullet \bar{r}}$. Because electrons are generally not localized, electron

density $\rho(\bar{r})$ in a volume element $d\bar{r}$ is used instead. Then the structure factor becomes $F(\bar{S}) = \rho(\bar{r})e^{2\pi i \bar{S} \cdot \bar{r}} d\bar{r}$.

For a sample (an atom or a molecule, for example) with a continuous electron density distribution, the structure factor is an integral over the entire sample:

$$F(\bar{S}) = \int \rho(\bar{r})e^{2\pi i \bar{S} \cdot \bar{r}} d\bar{r} \qquad [2a]$$

Note that usually the energy of the incident x-rays is much higher than the binding energy of the electrons so that electrons in the sample can be approximated as free electrons. The structure factor $F(\bar{S})$, and the electron density $\rho(\bar{r})$, are a Fourier pair.

$$\rho(\bar{r}) = (1/V) \int F(\bar{S})e^{-2\pi i \bar{S} \cdot \bar{r}} d\bar{S} \qquad [2b]$$

where V is the volume of the sample.

When the sample considered is an atom, the structure factor can be written as $F(\bar{S}) = f_a e^{2\pi i \bar{S} \cdot \bar{r}}$, where f_a is the atomic structure factor of atom a.

The above treatment is also valid for scattering of macromolecules in solution. However, the electron density, $\rho(\bar{r})$, is replaced by the excess electron density $\rho_{ex}(\bar{r})$, $\rho_{ex}(\bar{r}) = \rho(\bar{r})_{solute} - \rho(\bar{r})_{solvent} = \rho(\bar{r})_{solute} - \bar{\rho}_{solvent}$, where $\rho(\bar{r})_{solute}$ and $\rho(\bar{r})_{solvent}$ are the electron density of solute (macromolecule) and solvent, respectively. $\rho(\bar{r})_{solvent}$ can be approximated by the average value $\bar{\rho}_{solvent}$.

Experimentally, it is the scattering intensity $I(\bar{S})$ ($I(\bar{S}) = F(\bar{S})F^*(\bar{S})$) that is measured. The Fourier transform of $I(\bar{S})$ is called the Patterson function, $P(\bar{r})$. $P(\bar{r})$ is a self-convolution function: $P(\bar{r}) = F^{-1}(I(\bar{S})) = F^{-1}(F(\bar{S})F^*(\bar{S})) = \rho(\bar{r})*\rho(-\bar{r})$. Like the Fourier pair of $F(\bar{S})$ and $\rho(\bar{r})$, the scattering intensity and the Patterson function form another Fourier pair:

$$I(\bar{S}) = \int P(\bar{r})e^{2\pi i \bar{S} \cdot \bar{r}} d\bar{r} \qquad [3a]$$

$$P(\bar{r}) = (1/V) \int I(\bar{S})e^{-2\pi i \bar{S} \cdot \bar{r}} d\bar{S} \qquad [3b]$$

4

1.3. The Small-angle X-ray Scattering Technique

Small-angle x-ray scattering has become a standard experimental technique to study structural features of matter in the areas of condensed matter physics, polymer science, metallurgy and biophysics. The dimensions of the particles concerned range from 10 Å to 1000 Å. Since the x-ray wavelength (1-2 Å) is rather small compared to the size of the scatterers, significant scattering is only observed at small angular region. The term, small-angle x-ray scattering (SAXS), is thus derived. The typical angular range is about a few degrees in scattering angle.

1.3.1. Guinier Approximation

In solution, the sample is isotropic, *i.e.* each molecule can assume every possible orientation. The actual observed scattering intensity is the spatially averaged $I(\vec{S})$.

$$I(S) = \langle I(\vec{S}) \rangle = \int_\Omega I(\vec{S}) d\Omega = 4\pi \int_0^\infty P(r) \frac{\sin(2\pi Sr)}{2\pi Sr} dr \qquad [4]$$

where Ω is the solid angle, and $P(r) = r^2 \langle P(\vec{r}) \rangle$. $P(r)$ is called the distance distribution function.

The scattering intensity $I(S)$ can be further simplified, by substituting the Mclaurin series, $\sin(2\pi Sr)/2\pi Sr = 1 - (2\pi Sr)^2/6 + (2\pi Sr)^4/120 - \cdots$, into equation [4] and restricting to the first two terms for the small angle region where $2\pi SR_g < 1$, then:

$$I(S) = I(0)(1 - 4\pi^2 S^2 R_g^2/3) \qquad [5]$$

or

$$I(S) = I(0)\exp(-4\pi^2 S^2 R_g^2/3) \qquad [6a]$$

$$LnI(S) = LnI(0) - \tfrac{4}{3}\pi^2 R_g^2 S^2 \qquad [6b]$$

where

$$I(0) = 4\pi \int_0^\infty P(r)dr \qquad\qquad [7]$$

$$R_g^2 = \frac{1}{2} \int_0^\infty P(r)r^2 dr \bigg/ \int_0^\infty P(r)dr \qquad\qquad [8]$$

Equations [6a] - [6b] are different representations of the Guinier approximation (Guiner & Fournet, 1955), and are valid only in the small angle region, *i.e.* $2\pi SR_g < 1$. The forward scattering intensity $I(0)$ can be related to the total scattering length of the particle, and R_g is the radius of gyration of the particle (molecule) about its electron density center (Glatter & Kratky, 1982).

The Guinier approximation states that in the small angle region, the scattering intensity $I(S)$ can be approximated as a Gaussian distribution, and the width of the distribution is related to the radius of gyration of the particle. The scattering curve for a small particle with a small R_g is broad, while the curve for a large particle is narrow.

1.3.2. P(r) Function Calculation

The $P(r)$ function in solution scattering is similar to the Patterson function $P(\bar{r})$ in crystallography (see Section 1.3.1 for the definition of the $P(r)$ function). The $P(r)$ function describes the distribution of distances between the volume elements within the particle, and thus provides information regarding the geometry of the particle.

In principle, the $P(r)$ function can be calculated directly by Fourier transform of the scattering intensity $I(S)$ (refer to equation [3b]). However, in reality the scattering intensity $I(S)$ can not be taken at both the infinitely large angular region ($S \to \infty$) and the small angle region ($S \to 0$). This problem is called the termination effect. To reduce the termination effect in performing direct Fourier transform of $I(S)$, $I(S)$ is extrapolated to $S = 0$ by the Guinier approximation, and to the large angular range by Porod's law (Glatter & Kratky, 1982).

Another approach to calculate the $P(r)$ function is by the indirect Fourier transformation method (Glatter & Kratky, 1982). Briefly, the $P(r)$ function is approximated as a linear combination of a series of orthogonal functions $\varphi_\upsilon(r)$ in the range

6

of $0 \ll r \ll D$ with D being the maximal dimension of the particle: $P_A(r) = \sum_{\upsilon=1}^{N} c_\upsilon \varphi_\upsilon(r)$,

$P_A(r) \sim P(r)$. This function is subjected to a series of integral transformations $\hat{T}_1$ - $\hat{T}_4$, and the resultant function $\chi_\upsilon(S) = \hat{T}_4\hat{T}_3\hat{T}_2\hat{T}_1[P_A(r)]$ is compared with the experimental $I(S)$. The coefficients c_υ are determined by a weighted least squares approximation.

The indirect Fourier transform algorithm is implemented in the program GNOM (Semenyuk & Svergun, 1991). The P(r) functions in this dissertation are all calculated with the indirect Fourier transform method using GNOM.

1.3.3. Debye Formula

Instead of using a continuous electron density $\rho(\bar{r})$ to represent the particle, atoms in the particle can be treated as discrete scattering ensembles. The scattering length of the particle with N atoms can be written as: $F(\bar{S}) = \sum_{i=1}^{N} f_{a,i} e^{2\pi i \bar{S}\bar{r}}$, where $f_{a,i}$ is the atomic structure factor for atom i. The scattering intensity after averaging over all the orientations is:

$$I(S) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_{a,i} f_{a,j} \frac{\sin(2\pi S r_{ij})}{2\pi S r_{ij}} \qquad [9]$$

where r_{ij} is the distance between atoms i and j. Equation [9] is the Debye formula (Debye, 1915). Debye formula is very useful in the calculation of the scattering curve when the structure of the particle is known.

However, in solution scattering of biomolecules, solvent effect should be taken into account. To do this, the solvent inaccessible volume of the biomolecule needs to be evaluated. In the cube method (Fedorov & Denesyuk, 1978), the biomolecule is placed in a parallelepiped which is subdivided into small cubes with edges of 0.5-1.5 Å. The boundary of the biomolecule is determined by examining whether each cube belongs to the biomolecule or to the solvent (Hubbard, 1987). A more complicated method is to calculate the electron envelope of the macromolecule, and this algorithm is implemented in the program CRYSOL17 (Svergun, et al., 1995). Calculations of the scattering intensity in this dissertation are performed using the program CRYSOL17.

7

1.4. Information Content from Small-angle X-ray Scattering

Various structural information can be obtained from small-angle x-ray scattering data (Feigin & Svergun, 1987; Glatter & Kratky, 1982). Information including structural parameters derived in this dissertation is summarized in the following.

1.4.1. Size Information — Radius of Gyration

The radius of gyration (R_g) can be obtained directly from the scattering data by the Guinier approximation [6]. The maximal angular range S_{max} is constrained to the Guinier region, $2\pi SR_g < 1$. For globular-shaped protein molecules, it has been found that the linear region of $LnI(S)$ *vs.* S^2 in equation [6b] is extended outside the Guinier region, and a region of $2\pi SR_g < 1.3$ has been used to improve data fitting statistics (Olah, *et al.*, 1993). For elongated molecules, a deviation from the Guinier region will introduce systematic errors in R_g calculations. (See the heat-shock proteins in Chapter 3 for an example.) The minimal S is typically limited by the experimental setup. Parasitic scattering mainly from slits dominates the smallest resolvable angular range. A reliable S_{min} where parasitic scattering effects become negligible can be determined experimentally. S_{min} is also related to the distance between the sample and the detector (camera length); the longer the distance the smaller the S_{min} that can be achieved.

To minimize the interparticle interference (Glatter & Kratky, 1982), scattering curves are taken over several protein concentrations. R_g is then obtained by an extrapolation of apparent R_gs to zero protein concentration.

The radius of gyration can also be calculated from the distance distribution function ($P(r)$) according to equation [8]. The advantage of this approach is that the whole scattering curve including the Guinier region is used: the $P(r)$ function is obtained using $I(S)$ over the entire S range available. Since most interference, such as interparticle interference and contamination of large aggregates, distorts the scattering intensity predominantly at the small angle region within the Guinier region, by not restricting to the use of small angular region, this method should minimize the interference effect and give a better estimate of R_g (Glatter & Kratky, 1982).

If the structure of the molecule is known, R_g can be estimated from the atomic coordinates:

$$R_g^2 = \frac{\sum_i A_i[(x_i-x_c)^2+(y_i-y_c)^2+(z_i-z_c)^2]}{\sum_i A_i}$$ [10]

where A_i is the atomic number of the atom i, x_c, y_c, and z_c is the position of the electron center of the protein molecule.

1.4.2. Molecular Weight (MW)

Another direct structural parameter obtained from small angle x-ray scattering is the molecular weight (MW) of the particle. MW is directly related to the forward scattering intensity, $I(0)$ as: $I(0) \propto cM_r$, where c is the concentration in (g/cm^3). $I(0)$ is derived from the Guinier approximation of equation [6], or from the $P(r)$ function of equation [7]. To estimate the MW of the molecule, a sample with known MW is used as a standard for calibration:

$$\frac{(I(0)/c)_{c\to 0,\text{molecule}}}{(I(0)/c)_{c\to 0,\text{standard}}} = \frac{MW_{\text{molecule}}}{MW_{\text{standard}}}$$ [11]

where $(I(0)/c)_{c\to 0}$ is the concentration-normalized forward scattering intensity at infinitely dilute solution. In our experiments, lysozyme is frequently used as a standard sample.

1.4.3. Shape Information — P(r) Function and D_{max}

The distance distribution ($P(r)$) function can be calculated from the scattering curve using the direct or indirect Fourier transform methods. Since the $P(r)$ function represents the sample scattering in real space, the $P(r)$ function offers a more direct structural interpretation than the scattering intensity $I(S)$. Molecules with distinct structural features, such as dumbbell-shaped or highly elongated, show particular patterns in the $P(r)$ function (Glatter & Kratky, 1982).

A molecular parameter D_{max} can be obtained directly from the P(r) function. D_{max} is defined where P(r) first reaches zero (Glatter & Kratky, 1982). D_{max} describes the largest dimension of the molecule.

A further shape analysis involves an approximation of the molecule with simple geometrical models (Feigin & Svergun, 1987; Glatter & Kratky, 1982). A simple treatment is to model the whole molecule with a regular triaxial body, such as an ellipsoid, an elliptical cylinder or a prism. In Chapter 3 of this dissertation, rectangular prisms are used to represent the overall shapes of the ATPase and the peptide binding domains. A more complex approach is to model the molecule as a composite of regular triaxial bodies. This highly simplified modeling can, however, exclude some plausible models for the molecule (Wilbanks, et al., 1995).

1.4.4. Kratky Plot

The innermost part of the scattering curve, which contains information of the overall shape of the molecule, can be approximated as a Gaussian curve as in equation [6]. Due to the reciprocal relation, scattering at the large angular region is correlated to the structural features inside the molecule. Analysis of scattering intensity at the large angle region is complex, and depends on the structural detail of the molecule studied. However, some information can still be generalized.

It has been shown by Kratky (Glatter & Kratky, 1982) that for a chain-like molecule scattering intensity at the large angle is inversely proportional to S: $I \propto S^{-1}$. Thus, in the IS^2 vs S representation, or Kratky plot, a linear persistent increase is observed. On the other hand, for a globular-shaped molecule, scattering at the large angle follows Porod's law (Glatter & Kratky, 1982), that is $I \propto S^{-4}$, so that IS^2 decreases rapidly with S at large angle region. A well-defined peak in the Kratky plot is thus a characteristic of globular compact molecule. A coiled molecule and a globular molecule can accordingly be immediately distinguished in Kratky plots.

Kratky plots have been widely applied to the study of biological system, particularly in protein folding research (Kataoka, et al., 1993). Unfolded protein molecules can be treated as chain-like, while folded native molecules are compact and globular. Chapter 4 describes the application of Kratky plots in lysozyme folding studies.

1.5. General Experimental Aspects

All the x-ray scattering experiments described in this dissertation were performed on beamline 4-2 at the Stanford Synchrotron Radiation Laboratory (SSRL). Some of the experimental aspects are summarized in the following. Relevant information regarding small-angle x-ray scattering experiments on this beamline can be found in the references (Eliezer, 1994; Rice & Wakatsuki, 1991; Wakatsuki, *et al.*, 1992).

Fig. 1.1 shows a schematic drawing of the small-angle x-ray scattering experimental setup on beamline 4-2 at SSRL. A typical SAXS experiment setup includes: a (monochromatic) x-ray source, a collimation system, a sample cell, a scattering path and a detector.

Synchrotron radiation. All the scattering data in this dissertation were taken using a synchrotron x-ray source. A detailed description of the basic physics of synchrotron radiation can be found in the literature (Landau & Lifshitz, 1987; Winick & Doniach, 1982). Briefly, synchrotron radiation is the electromagnetic field radiated by relativistic accelerated charged particles. The charged particles are either electrons or positrons, and they are accelerated using a magnetic field. To recirculate the particles along a circular path, magnetic fields perpendicular to the horizontal orbital plane are used. Radiation emitted from such relativistic particles is highly collimated in the direction tangential to the circular orbital. This is the qualitative explanation for the extreme brightness of a synchrotron radiation source, compared with a conventional radiation generated in a x-ray tube. To further increase the brightness of radiation, wigglers and undulators are frequently used in synchrotron facilities.

Small divergence in the x-ray beam is the primary concern in small-angle x-ray scattering experiment. Although synchrotron radiation is highly collimated, the source is usually located far away (~ 26 m between the source and beamline 4-2, for example), consequently, a few milliradian emittant angle results in a few centimeter divergence at the experimental station. A mirror is thus used to focus the x-rays. Platinum (Pt) is commonly chosen as the coating material for mirrors because of its long lifetime. A monochromatic beam is another factor in SAXS experiment. Monochromatic x-rays are selected by monochromators. In our experiments, an energy of about 8980 eV is normally used,
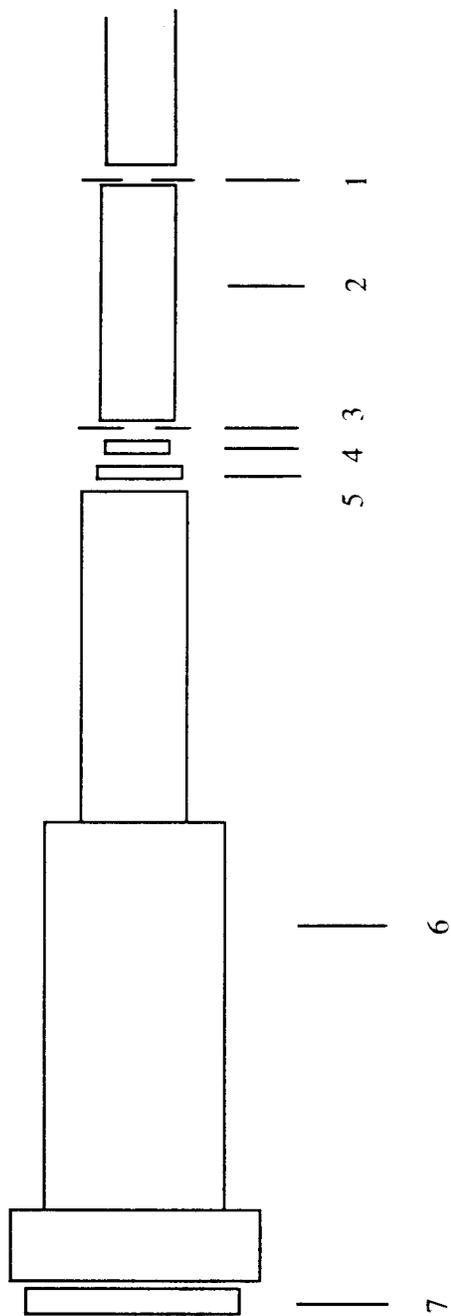
1 - defining slit

2 - collimation arm, ~ 1 m, under vaccum

3 - guard slit

4 - ionization chamber

5 - sample cell holder & sample cell

6 - scattering path, ~230 cm, under vaccum

7 - detector

Figure 1.1. Schematic drawing of the small-angle scattering experiment setup (sideview) at beamline 4-2 at SSRL.

obtained by using a double-crystal monochromator setup with Si [111] crystals. A pair of Mo/B$_4$C multilayers is used for the experiments with the broad-band pass beam.

Sample considerations. Solution sample has to be contained in a cell. The window material for the cell can be mica, quartz, or glass with 10 - 50 μm thickness. Two factors are considered in determining the optimum thickness of the sample. The scattering intensity increases linearly with the thickness (t), while the absorption of x-rays by the sample increases exponentially, $e^{-\mu t}$, with μ being the linear absorption coefficient of the sample. For an approximation, the maximum scattering intensity is achieved when the function $te^{-\mu t}$ is maximum, which occurs at $t = 1/\mu$. For dilute biological samples in water solution, absorption of the sample is mainly determined by the absorption of H$_2$O. For H$_2$O, $\mu = 1$ mm^{-1}, thus the optimum thickness of the sample is about 1mm.

The other consideration about biological samples concerns radiation damage. Biological samples may degrade when exposed to the highly intense synchrotron x-rays in ·the course of data collection. Changes in sample color or sample precipitation are the direct indications of radiation damage. Biological assay is normally carried out after the SAXS experiment if it is applicable to ensure that the sample is still biologically active. A further examination is to inspect the time course of scattering curves, particularly in the small angle region. There is no generalized method to prevent radiation damage to biological sample in a SAXS experiment, however, it has been found from our experience that addition of β-mercaptoethanol is effective in retarding protein sample deterioration in x-rays.

Detectors. Two kinds of detectors have been used in the experiments described in this dissertation. A linear position proportional counter is routinely used in our SAXS experiments. It consists of two electrodes housed in a pressurized cell. The anode is a 6 μm gold tungsten coated wire, and the cathode consists of an array of metal coated glass fiber attached to the delay lines. A high electric field (~1.7 keV) is applied between the anode and the cathode. A detecting gas, consisting of either Xe/CO$_2$ or Ar/CO$_2$, acts as a conversion medium for x-rays. The conversion efficiency of Xe/CO$_2$ is about 5 times that of Ar/CO$_2$, however, the cost of Xe/CO$_2$ is very high.

When a photon enters the detector chamber, it ionizes Xe or Ar in the conversion gas. Electrons thus generated are accelerated by the very high electric field to the anode wire. The energetic accelerated electrons in turn ionize other Xe or Ar atoms, and so on, triggering an avalanche and giving rise to a large amplification of the initial signal. When a

signal arrives at the anode, it results in a charge buildup on the cathode. The electric pulses are then picked up by the delay lines and propagate in both directions, and finally reach the ends of the delay lines at different times. The time difference between the delay lines is correlated to the position of the ionization event. The Xe^+ or Ar^+ ions are neutralized by the quenching gas CO_2. A quadrant detector, also used in some experiments, works in a similar fashion except that the detection area is larger and the delay signal detection is more sophisticated. A more detailed description of principles of both the linear and quadrant detectors can be found in Rice & Wakatsuki, 1991.

1.6. References

Debye, P. (1915). *Ann. Physik*, **28**, 809-823.

Eliezer, D. (1994) *Protein folding and protein metallocluster studies using synchrotron small angle x-ray scattering*. Ph.D. Thesis, Stanford University.

Fedorov, B. A., & Denesyuk, A. I. (1978). *FEBS Lett.*, **88**, 114-117.

Feigin, L. A., & Svergun, D. I. (1987). *Structure analysis by small-angle x-ray and neutron scattering*. New York: Plenum Press.

Glatter, O., & Kratky, O. (Ed.). (1982). *Small angle x-ray scattering*. New York: Academic press.

Guiner, A., & Fournet, G. (1955). *Small-angle scattering of x-rays* (C. B. Walker, Trans.). New York: John Wiley & Sons, Inc.

Hubbard, S. R. (1987) *Small-angle x-ray scattering studies of calcium-binding proteins in solution*. Ph.D. Thesis, Stanford Univiersity.

Kataoka, M., Hagihara, Y., Mihara, K., & Goto, Y. (1993). *J. Mol. Biol.*, **229**, 591-596.

Landau, L., & Lifshitz, E. M. (1987). In *The classical theory of fields*. Oxford: Pergamon.

Olah, G. A., Trakhanov, S., Trewhella, J., & Ouiocho, F. A. (1993). *J. Biol. Chem.*, **268**, 16241-16247.

Rice, M., & Wakatsuki, S. (1991). *A user's guide to the small-angle x-ray scattering/diffraction data acquisition system at SSRL*.

Semenyuk, A. V., & Svergun, D. I. (1991). *J. Appl. Crystallogr.*, **24**, 537-540.

Svergun, D. I., Barberato, C., & Koch, M. H. (1995). *J. Appl. Crystallogr*, submitted for publication.

Wakatsuki, S., Hodgson, K. O., Eliezer, D., Rice, M., Hubbard, S., Gillis, N., Doniach, S., & Spann, U. (1992). *Rev. Sci. Instrum.*, **63**, 1736-1740.

Wilbanks, S. M., Chen, L., Tsuruta, H., Hodgson, K. O., & McKay, D. B. (1995). *Biochemistry*, **34**, 12095-12106.

Winick, H., & Doniach, S. (Ed.). (1982). *Synchrotron radiation research.* New York: Plenum.

Chapter 2

Solution Small-angle X-Ray Scattering Study of the Components of the *Azotobacter vinelandii* Nitrogenase Enzyme System

2.1. Introduction and Background

Nitrogenase is the biological enzyme system that reduces atmospheric N_2 to ammonia, the main usable form of nitrogen for all organisms. The striking feature of this system is that the biological N_2 fixation occurs under normal atmospheric pressure and at normal temperature, whereas the equivalent industrial Haber-Bosch process requires high pressure (200 - 300 atm), elevated temperature (450 °C), and catalysts (Nielsen, 1968). The mechanism of nitrogenase, therefore, has been under extensive investigation since the mid 1960s. A wealth of information gathered from biochemistry, genetics, inorganic chemistry, and particularly structural biology, has expanded our understanding of many aspects of this biological N_2 fixation system.

One of the most intensively investigated nitrogenase systems is the molybdenum-containing nitrogenase from *Azotobacter vinelandii*, a free-living aerobic bacterium. The nitrogenase enzymatic system consists of two O_2-sensitive metalloproteins: the 60 kDa iron protein (Fe protein) and the 230 kDa molybdenum-iron protein (MoFe protein). Schematic representations of these two components are shown in Fig. 2.1. In the process of N_2-fixation, the Fe protein transfers one electron from an *in vivo* donor such as ferredoxin or flavodoxin to the MoFe protein. Then the MoFe protein utilizes electrons to reduce N_2 to ammonia (Burgess, 1983; Thorneley & Lowe, 1985). Hydrolysis of two MgATP molecules is required for one electron transfer from the Fe protein to the MoFe protein to occur. The overall stoichiometry of N_2 reduction by nitrogenase has been established as (Burgess, 1983):

$$N_2 + 8\,H^+ + 8\,e^- + 16\,MgATP \longrightarrow 2\,NH_3 + H_2 + 16\,MgADP + 16\,P_i$$
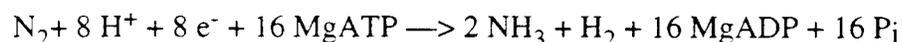
In addition to N_2, nitrogenase also supports reduction of C_2H_2 to C_2H_4, N_3^- to N_2 + NH_3, and RCN to RCH_3 + NH_3.

2.1.1. The Fe Protein

The Fe protein is a dimer of two identical subunits, which are covalently connected by a $[4Fe-4S]^{2+/1+}$ cluster at one end of the dimer. There are two nucleotide binding sites, one in each subunit (Burgess, 1993; Yates, 1991). Sequence comparisons (Robson, 1984), site-directed mutagenesis (Seefeldt, *et al.*, 1992) and x-ray crystallography (Georgiadis, *et al.*, 1992) are all consistent with a model whereby the terminal phosphate of

17

Mo nitrogenase

the Fe protein
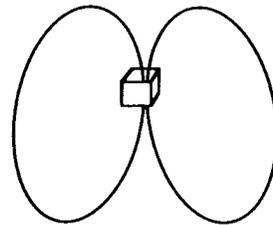
the MoFe protein

MW: 60 kDa

⊟ **4Fe:4S**
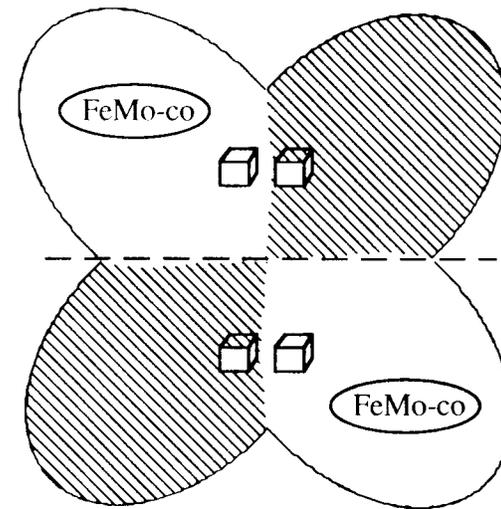
MW: 230 kDa

⊟ ⊟ **P-clusters**
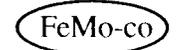
(FeMo-co) **1Mo:7Fe:9S:1homocitrate**

Figure 2.1. Schematic representations of the MoFe protein and the Fe protein of the Mo nitrogenase.

MgATP interacts with the "Walker's motif" sequence GXGXXGKS (Walker, *et al.*, 1982) which is located between residues 9-16 of the *Azotobacter vinelandii* Fe protein sequence. One of the most intriguing aspects of the Fe protein is that MgATP binding somehow influences the environment of the [4Fe-4S] cluster (Burgess, 1983; Burgess, 1993; Yates, 1991), even though the binding site for the terminal phosphate is located ~20 Å away (Georgiadis, *et al.*, 1992). Changes in the [4Fe-4S] cluster that occur upon MgATP binding include: 1) the reduction potential $E^{0'}$ is lowered by *ca.* 100 mV (Morgan, *et al.*, 1986; Zumft, *et al.*, 1974); 2) the S=1/2 EPR signal exhibited by $[4Fe-4S]^{1+}$ changes from a rhombic shape to a more axial shape (Orme-Johnson, *et al.*, 1972; Smith, *et al.*, 1973; Zumft, *et al.*, 1972); 3) the O_2 sensitivity increases; 4) large changes occur in the CD spectrum of the oxidized protein (Stephens, *et al.*, 1982); and 5) the cluster Fe atoms, which are inert to chelators in the absence of MgATP, become accessible to chelators in its presence (Walker & Mortenson, 1973; Walker & Mortenson, 1974). Thus, MgATP binding appears to induce a long-distance conformational change in the Fe protein. Mutagenesis data further suggest that this conformational change is a prerequisite for the next step in nitrogenase turnover, which is the binding of the Fe protein to the MoFe protein in such a way that electron transfer can occur (Gavini & Burgess, 1992; Seefeldt, *et al.*, 1992).

It is interesting that the effects of MgATP and MgADP are not identical. Binding of MgADP to the Fe protein causes changes in the CD spectrum and reduction potential $E^{0'}$ (Burgess, 1983; Morgan, *et al.*, 1986; Stephens, *et al.*, 1982; Zumft, *et al.*, 1974); but it is an inhibitor of the chelator reaction (Walker & Mortenson, 1973; Walker & Mortenson, 1974). Thus, even though the two nucleotides appear to bind to the same sites (Tso & Burris, 1973) they have different effects on protein structure. These changes in turn are important for function because MgATP is required for substrate reduction while MgADP inhibits substrate reduction.

A mutant variant of the *A. vinelandii* Fe protein, designated *A. vinelandii* UW91 (Shah, *et al.*, 1973), which has a serine substituted for an alanine at position 157 of the NifH sequence (Gavini & Burgess, 1992), has been purified and characterized. This mutant has a normal $[4Fe-4S]^{2+/1+}$ cluster, and normal MgATP binding ability but it is blocked for the MgATP-induced conformational change as measured by the chelation reaction described above (Gavini & Burgess, 1992). This altered Fe protein also cannot support substrate reduction (Gavini & Burgess, 1992).

2.1.2. The MoFe Protein

The MoFe protein is a heterotetramer $\alpha_2\beta_2$, with two types of metal clusters: two iron-molybdenum (FeMo) cofactors, and four [4Fe-4S] clusters (P-clusters) (Burgess, 1983). The FeMo cofactor is believed to be the substrate, N_2, binding and reducing site (Burgess, 1990). The P-cluster is proposed to bridge the electron transfer between the [4Fe-4S] cluster of the Fe protein and the FeMo cofactor of the MoFe protein (Holm, 1990). The MoFe protein has two Fe protein binding sites, and interacts only transiently with the Fe protein (Thorneley & Lowe, 1985). Although the Fe protein can bind MgATP by itself, the hydrolysis of ATP in the electron transfer from the Fe protein to the MoFe protein, only occurs within the Fe protein/MoFe protein complex. To ensure that electron transfer occurs in this binary protein complex, a correct orientation of the proteins to align the electron donor and acceptor sites must be maintained. So far there is no crystal structure available for this active complex. Several models, however, have been proposed based on cross-linking and mutagenesis studies (Thorneley, et al., 1993; Willing & Howard, 1990).

2.1.3. The FeMo Cofactor

The FeMo cofactor was first isolated in 1977 as a stable metallocluster from acid-denatured MoFe protein (Shah & Brill, 1977). It is extremely sensitive to oxygen; yet it is stable in N-methylformamide (NMF) at room temperature under anaerobic conditions. The isolated FeMo cofactor can be inserted back into cofactor-deficient MoFe protein, upon which the activity of the MoFe protein is restored. Since the cofactor is believed to be the active site for N_2 reduction, extensive research has been devoted to study the structure and function of this metal cluster, including great efforts in chemical modeling aspects. However, the isolated FeMo-cofactor has not yet been crystallized successfully. A major breakthrough came with the publication of the structure of cofactor inside the MoFe protein (Kim & Rees, 1992), shown in Fig. 2.2. It contains two subclusters: the $MoFe_3S_3$ and the Fe_4S_3 subclusters with a total of 1Mo:7Fe:9S, in addition to two protein ligands and a homocitrate. Structures of the cofactor in solution and inside the protein appear to be similar as determined by extended x-ray absorption fine structure (EXAFS), Mössbauer, and electron paramagnetic resonance (EPR) spectroscopies (Burgess, 1990; Conradson, et al., 1987; Hedman, et al., 1988; Rawlings, et al., 1978).
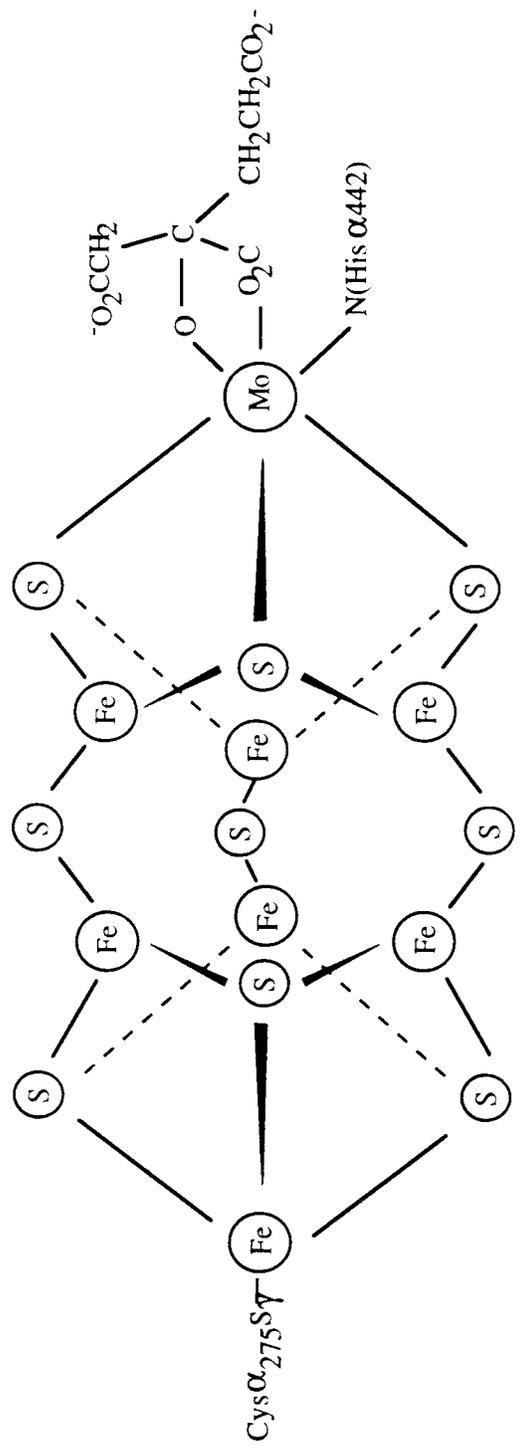
Figure 2.2. Current structural model for the FeMo cofactor suggested from the crystal structure of the MoFe protein (Kim & Rees, 1992).

21

2.1.4. The FeMo Cluster

A Mo-containing metal cluster (designated as FeMo cluster) has been isolated from the MoFe protein by a preparation method (Shah & Brill, 1981) different from that for the FeMo-cofactor. Instead of extraction into N-methylformamide (NMF) under alkaline condition (pH ~ 8.5), the extraction of the FeMo cluster is carried out at acidic pH (pH ~ 4) in methyl-ethyl ketone (MEK). This entity exhibits a very similar $S = 3/2$ EPR signal, a spectroscopic feature whose magnetic center is still not clear, to those of the isolated FeMo cofactor and the FeMo cofactor within the protein. The fact that the cluster inhibits the insertion of the cofactor back into cofactor-deficient MoFe protein indicates that the cluster binds to the same site as the cofactor (or nearby) in the MoFe protein (Burgess, 1990). However, they differ structurally. Circular dichroism indicates that the cluster does not contain homocitrate; both Fe analysis and EXAFS show that the Mo atom of the cluster has one less Fe atom in the first coordination shell and EXAFS in addition indicates a change in the long range order (Ma, *et al.*, 1994). Furthermore, this cluster can not activate the FeMo cofactor-deficient form of the MoFe protein.

It has been shown by small-angle x-ray scattering that FeMo cofactor in NMF solution is not homogenous (Eliezer, *et al.*, 1993), and this inhomogeniety may account for the failure in crystallizing the cofactor for structural determination. The cofactor exists in large aggregates (up to 25 Å in R_g) as well as in small oligomeric forms. The heterogeneity of the cofactor was suggested to be due to the presence of different ligation states of the endogenous homocitrate (Ma, *et al.*, 1994). Since homocitrate is not retained in the FeMo cluster, the comparison between the aggregation states of the cofactor and the cluster would provide immediate direct information on the effect of homocitrate on oligomerization. Efforts have been devoted to search for conditions to crystallize the cofactor outside the protein matrix for structural determination, and a homogenous, aggregate-free system is crucial for producing diffraction-quality crystals.

In this chapter, emphasis is placed on x-ray scattering studies of the Fe protein, which undergoes a conformational change triggered by MgATP. Some initial results of scattering studies of the MoFe protein as well as the FeMo cluster are also presented.

2.2. Experimental Procedures

2.2.1. The Fe Protein

The wild type and the *A. vinelandii* UW91 Fe proteins were purified as described elsewhere (Burgess, *et al.*, 1980; Gavini & Burgess, 1992). They were homogeneous by the criteria of one- and two-dimensional SDS-PAGE (sodium dodecylsulfate-polyacrylamide gel electrophoresis). The native protein had a specific activity of *ca.* 1,900 nmoles of H_2 evolved/min/mg protein and had no background activity. A complete characterization of the UW91 Fe protein, which is inactive, has been described (Gavini & Burgess, 1992).

All Fe protein samples were prepared anaerobically in a Vacuum Atmospheres glove box under dry nitrogen atmosphere. Stock Fe protein solutions (30 mg/mL in 0.05 M Tris-HCl, 0.5 M NaCl, 2 mM $Na_2S_2O_4$, pH 8.0) were first diluted to the appropriate concentrations with the same buffer. Where MgATP was needed, the solution was then made 2.5 mM in MgATP using a 25 mM stock solution. All samples were then adjusted to 5 mM in sodium dithionite by addition of the appropriate amount of a 300 mM pH-adjusted $Na_2S_2O_4$ stock solution. Both sample preparations and sample integrity tests (before/after SAXS experiments) were carried out by Dr. Narasaiah Gavini in Prof. Burgess' laboratory at UC-Irvine.

For the 1993 runs (see below), 5 μl of a 200 mM pH-adjusted $Na_2S_2O_4$ solution were added into each 495 μl protein sample solution just before the sample was loaded for scattering measurement. Based on the previous SAXS experience with the Fe protein, this was done to further ensure that the proteins were in a reducing environment. Identical samples, that were stored frozen at UC-Irvine for several weeks, were subsequently assayed for C_2H_2 reduction activity and were tested for the MgATP-induced conformation change by adding bathophenanthroline disulfonate (Ljones & Burris, 1978). All the wild-type samples showed C_2H_2 reduction activity within experimental error of the starting values. As previously reported (Gavini & Burgess, 1992), all native samples containing MgATP released their Fe atoms when the chelator was added, whereas the native samples without MgATP and all UW91 samples, did not.

23

Scattering experiments were performed on beamline 4-2 at Stanford Synchrotron Radiation Laboratory (SSRL) in May 1992, January and May 1993 using the SSRL small-angle x-ray scattering/diffraction system (Wakatsuki, *et al.*, 1992). Synchrotron radiation was provided by the storage ring SPEAR, which was operated at an electron energy of 3.0 GeV with a ring current of 40-90 mA. The synchrotron beam was monochromatized (to an energy of 8980 eV) by a double-crystal Si[111] monochromator. The typical beam size at the sample position was 1.0 mm (v) x 2.1 mm (h) for a flat-window cell in the 1993 runs, and 0.4 mm (v) x 3.0 mm (h) for capillary cells in the 1992 run. Scattering data were collected with a one-dimensional position-sensitive proportional counter (BioLogic, Grenoble, France) with a 200 mm long sensitive area. A short-path, open-air ionization chamber was placed immediately in front of the sample to record the incident x-ray intensity on the sample. The total flux on the sample was used to normalize scattering intensity against small fluctuations and intensity changes of the x-ray beam due to the decay of the electron current in SPEAR. The sample-to-detector distance was determined to be 264 ($\pm$ 1) cm, 265 ($\pm$ 1) cm, 230 ($\pm$ 1) cm for the May 1992, January and May 1993 runs, respectively.

Data were collected in a protein concentration series from 2 mg/ml to 12 mg/ml of five protein samples, wild-type without nucleotides, wild-type plus MgATP, wild-type plus MgADP, mutant UW91 without nucleotides and mutant UW91 plus MgATP. In the 1993 experiments, all data collection was performed using flat-window cells. These cells are made of Kel-F and have 25 μm thick, 5 mm diameter circular mica windows on each side, with a 1.2 mm light path. Two small holes from the top into the cell body allow for sample injection. The holes are anaerobically sealed by a threaded acrylic plug with two O-rings. The air tightness of the cells was satisfactory, as tested by the lack of change of the absorption spectra of a methyl viologen radical cation solution over a time longer than a typical SAXS measurement. During data collection, the cell was fit into a fixed cell holder with 6.3 μm thick polypropylene film on each side. Helium was continuously flowed through the holder to eliminate air scattering around the sample, to provide a micro-anaerobic environment around the sample cell, and to keep the window surface clean. The temperature was maintained at 20 °C by circulating an ethylene glycol solution through the sample holder. Three types of buffer solutions (without nucleotides, with 2.5 mM MgATP, and with 2.5 mM MgADP) were used as background solutions for the corresponding protein samples. Background scattering was recorded for every two protein samples. The exposure time for each buffer and protein solution was the same, 8 x 5 minutes.

For the 1992 run when capillaries were employed, the following procedure was used for every set of data collection. Background scattering from a 1.5 mm diameter x-ray glass capillary (Charles Supper Company) filled with the buffer solution was first recorded, and the buffer was subsequently removed. Then a *ca.* 80 µl aliquot of a protein solution was loaded, the capillary end was sealed with 5-minute epoxy glue (Devcon Corporation), and then data were collected. Every effort was taken to ensure that the capillary position was invariant with respect to the beam. The exposure time for each buffer and protein was 3 x 10 minutes. This capillary method was used initially until the reliable, anaerobic flat-window cells were developed and tested. Higher parasitic scattering from the round capillary wall and the resultant poor background subtraction make the capillary measurements less reproducible. For both the flat-window cell and the capillary experiments, all sample handling involving the proteins was carried out anaerobically in a Vacuum-Atmospheres dry N_2-filled glove box.

2.2.2. The MoFe Protein

The MoFe protein was obtained anaerobically during the purification of the wild type Fe protein (Burgess, *et al.*, 1980). It was purified to homogeneity as monitored by one- and two-dimensional SDS-PAGE. The specific activity of the MoFe protein was ~ 3000 nmol of H_2 evolved/min/mg protein. Samples for SAXS experiments were prepared by dilution from a stock MoFe protein solution (~ 40 mg/ml in 25 mM Tris-HCl, 250 mM, NaCl, 2 mM, $Na_2S_2O_4$, pH 7.4) with the same buffer to final concentrations of 2, 4, and 8 mg/ml, and were frozen in dry ice until the SAXS experiment. Representative samples were recovered anaerobically after data collection for activity assay, and their activities were found close to the initial value. All the sample preparations were performed by Dr. Narasaiah Gavini in Prof. Burgess' laboratory at UC - Irvine.

SAXS data were collected in May 93 on beamline 4-2 at SSRL. A flat-window sample cell was used for data collection. All sample manipulation procedures, as well as the beamline configurations, were similar to those of the Fe protein experiments unless otherwise stated. The sample-to-detector distance was 230 ($\pm$1) cm, which covered the angular range up to 0.04 Å^{-1} in S (S = 2 sinθ/λ). Background was recorded immediately before or after the sample measurement, with data collection time being 8 x 5 minutes.

2.2.3. The FeMo Cluster

The purification of the FeMo cluster was done by Dr. Li Ma in Prof. Burgess' laboratory at UC - Irvine according to reference (Ma, *et al.*, 1994). Briefly, inside the dry box, the MoFe protein (in 25 mM TrisHCl, 250 mM NaCl, 2 mM $Na_2S_2O_4$, pH 7.4) in a tube was placed underneath a layer of methyl ethyl ketone (MEK) which was presaturated with the same buffer. The tube, sealed with a serum stopper, was then removed from the dry box. Solution of 4 N HCl was added into the MEK layer, and the tube was subjected to centrifugation at 4 °C. The cluster which was extracted into the top MEK layer was removed inside the dry box. To transfer the cluster into the NMF solution, the MEK solution was diluted with NMF anaerobically, and MEK was distilled away under vacuum. The NMF solution contained 20 mM benzylthiol as the reducing agent, while the MEK solution contained saturated $Na_2S_2O_4$. The cluster samples were frozen at - 70 °C in a crimp-top v-shape vial with a silicon septum covered with several layers of Saran Wrap, and shipped in dry ice to Stanford University.

The stock cluster solution was concentrated when necessary and then diluted into different concentrations with the appropriate buffer, either MEK or NMF. In the first experiment of June 94, the cluster was prepared in both MEK and NMF. The stock MEK solution (containing ~ 1 mM FeMo cluster) was diluted to lower concentrations, making a total of three samples: ~ 1 mM, ~ 0.8 mM and ~ 0.6 mM. The stock NMF solution was first concentrated at 20-30 °C and 140 microns pressure, and then diluted to make three samples, 731 µM, 619 µM and 464 µM. In the June 1995 run, scattering data were collected only on the NMF solution because it was shown from the first run that the cluster predominantly formed large aggregates in the MEK solution. The concentration of the stock solution was 463 ($\pm$ 30) µM. A portion of this solution was retained as a scattering sample directly; the remainder was concentrated to dryness and redissolved in NMF to make a final concentration of 1.44 ($\pm$0.12) mM. Except for the MEK samples, the cluster concentrations were determined by iron analyses. All these sample manipulations were performed by Dr. Patrick Frank at Keith Hodgson's laboratory at Stanford University.

Since both MEK and NMF are organic solvents, they can dissolve the epoxy glue which adheres the mica windows to the Kel-F body of the flat-window cell. A special cell was therefore designed and constructed. The cell is built in the middle of a 40 x 40 x 20 mm nylon block. The cell is 5 mm in diameter and 1.0 mm thick with a total volume of ~

30 µl. Instead of gluing windows to the cell, circular precut 25 µm-thick 10 mm in diameter mica (Attwater & Sons Ltd., U.K.) sheets are attached to the cell surface via 0.1 mm-thick Teflon o-rings. The window is pressed firmly against the wall with a rubber gasket by a threaded nylon plug on each side. A hole of 6 mm in diameter is drilled through the plug in the middle, and coincides with the cell window. Like the flat-window cell, there are two tiny holes (1 mm in diameter) on the top for sample loading. The holes are sealed by a 2.25 mm-thick circular silicone septum and a threaded plug. The cell has a stem on the bottom of the block which fits into the regular cell holder.

Anaerobicity of the cell was tested by monitoring the time course of any change in the iron absorption spectrum of $(Et_4N)_2[Fe^{II}(S_2\text{-}o\text{-}xyl)_2]$ compound in N,N-dimethylformamide (DMF) solution. The spectrum was stable for at least 75 min, and appeared to change slightly over the rest of the measuring time, which was about 4 hours. An approximate 1-electron oxidation rate for the FeMo cofactor was derived to be about 96 nM/min by analogy with the Fe(II) model compound. Accordingly, even in the worst situation where data collection time was 5 hours and the sample was in the least concentration of 464 µM, only 6% of the cluster sample would be oxidized. The air-tightness is therefore satisfactory for our cluster scattering experiment. The cell design and anaerobicity test were done by Dr. Patrick Frank at Prof. Keith Hodgson's laboratory at Stanford University.

Since the cluster is such a small molecule, ~ 800 Da, much smaller than the usual small single domain proteins which are normally around 12 kDa in molecular weight, much longer data collection time is expected. In practice, particularly for the first sample, data were collected on one hour base. At the end of each hour, the sample was inspected by eye, and if there was no visible anomaly such as color change or precipitate, data collection was resumed. Then continuous data collections were set at the longest time available, 4 -5 hours. Background scattering was recorded either before or after the sample. Samples were recovered anaerobically for EPR inspection.

2.3. Data Analysis

Data were analyzed using the Guinier approach, as described in Chapter 1. Because the molecular weights of the scatterers vary from 220,000 Da to 60,000 Da to 800 Da, and consequently a great difference in their sizes or R_g is expected, the fitting range for

the Guinier approximation was chosen accordingly: 3.2×10^{-3} - 6.3×10^{-3} for the MoFe protein, 4.5×10^{-3} - 9.5×10^{-3} $\mathring{A}^{-1}$ for the Fe protein in S. For the FeMo cluster, S_{min} was selected where the parasitic scattering appears to be minimum, and S_{max} was chosen depending on the availability of data on higher angular range.

Radius of gyration can be calculated directly from the atomic coordinates:

$$R_g^2 = \frac{\sum_i A_i[(x_i - x_c)^2 + (y_i - y_c)^2 + (z_i - z_c)^2]}{\sum_i A_i} \qquad [1]$$

where A_i is the atomic number of atom i, x_c, y_c, and z_c are the coordinates of the electron density center of the protein molecule.

For protein molecules, scattering curves can also be calculated from atomic coordinates, including a 3 $\mathring{A}$ thick hydration layer of water around the protein molecule. A value of 0.03 $e^-/\mathring{A}^3$ is assumed to be the electron density contrast between the hydration layer and the bulk background. The calculations were performed using the program CRYSOL17 (Svergun, et al., 1995).

Although the detector response homogeneity was constantly checked during the experiment, in the data reduction, detector response correction was undertaken only on the most recent data, the FeMo cluster data. Before the 1994 run, a ^{55}Fe source was used to irradiate the detector, and as a result the response pattern was deformed somehow because of the point-like radiation source. Under these circumstances, the response pattern was taken only to ensure that no anomaly occurred to the detector particularly in the small angle regions during data collection. When an aberration was detected, the detector was cleaned until a reasonable response was obtained. In the recent runs, a more homogenous source of irradiation, fluorescence generated by direct x-ray beam hitting a copper plate, was utilized instead. The detector response homogeneity was reflected more precisely with this method, and the response pattern was used to correct any subtle inhomogeneity of the detector by a channel-by-channel division of the data by the detector response.

2.4. Results and Discussion

2.4.1. The Fe Protein

None of the samples measured showed a color change during/after the measurement which would have indicated sample oxidation/degradation. To further assure that no detectable radiation damage or aggregation occurred during the course of each experiment, an examination of normalized scattering curves from different cycles of data collection was carried out, particularly at the smallest angle region available. They overlapped within experimental error, suggesting that protein molecules remained in the same aggregation state throughout data collection process. A typical Guinier plot is shown in Fig. 2.3. The lack of upward curvature in the Guinier region indicates that no significant aggregation had taken place.

The apparent radii of gyration for each sample were derived from the Guinier approximation. The effect of concentration on radius of gyration was minimized as usual by the linear extrapolation of R_g *vs.* concentration (C) to zero protein concentration. The extrapolations are shown in Fig. 2.4.a-e. The nearly linear dependence of the apparent R_g on the protein concentration further demonstrates that the predominant aggregation state for each sample remained the same throughout the concentration range chosen. The forward intensity at zero scattering angle, I(0), which can be obtained from the extrapolation of Guinier plots to zero scattering angle, is roughly proportional to the square of the mass of the scatterer (Glatter, 1982). The ratio (I(0)/C) at zero concentration obtained for all five protein samples were similar. This result was expected because of the small mass of MgATP and MgADP compared to that of the protein molecule. Plots of (I(0)/C) *vs.* concentration for all samples were acceptably linear which is again consistent with no change of protein aggregation state within the concentration range tested. Again it is important to note that the lack of change in aggregation state described above was observed and reproduced over three independent experimental runs.

Independent tests of reproducibility. Despite the use of different sample cells, the results of the capillary method adopted in the May 1992 run were reproduced in the January 1993 data using the flat-window cell method. To further examine the reproducibility of data, a number of multiple measurements were carried out under the most favorable and reproducible experimental conditions (with the flat-window cells, and an additional amount
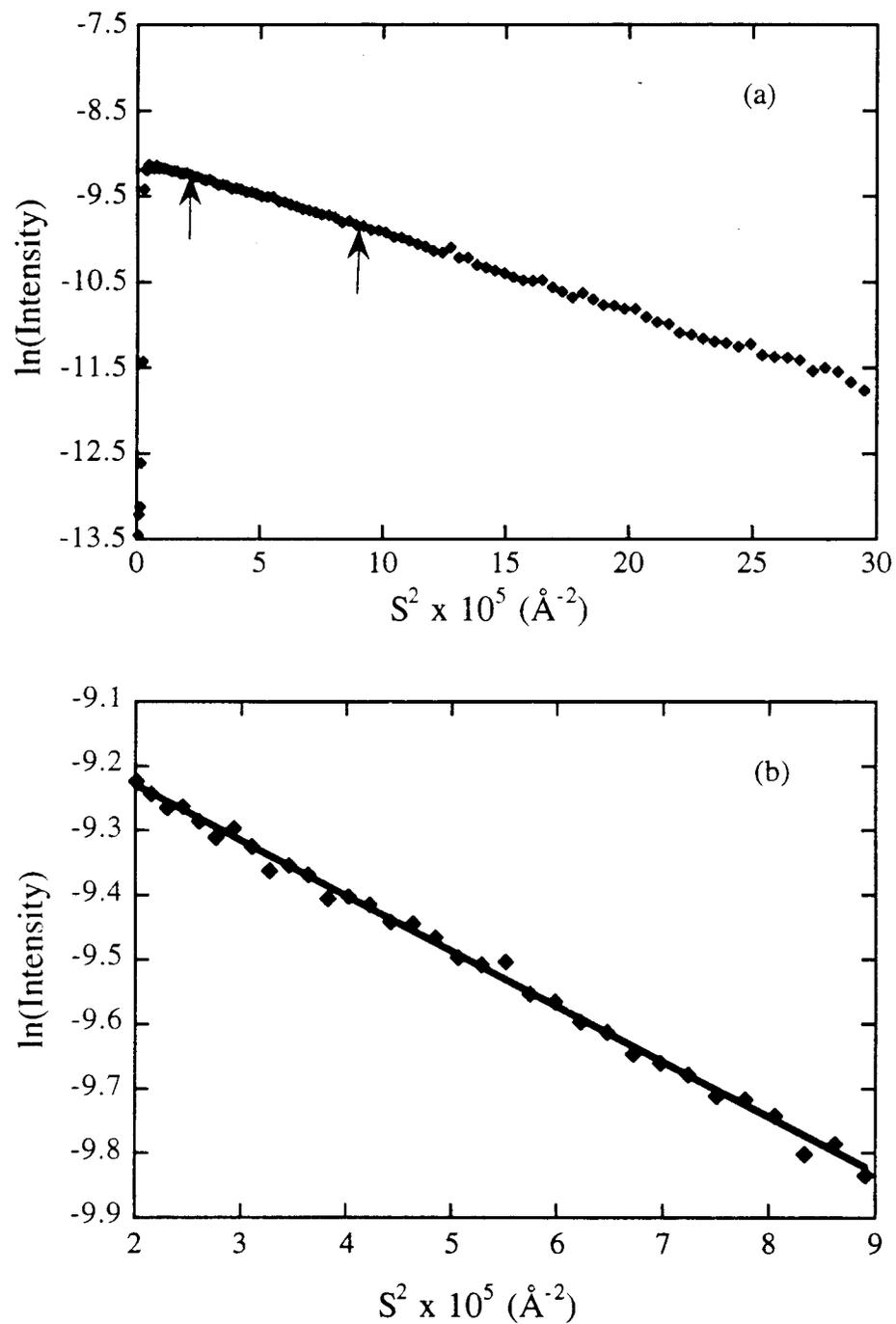
Figure 2.3. (a) A typical Guinier plot and region selected for Guinier fit (between arrows). (b) The expanded fitting region. Data were linearly fit by least squares methods. This data set is from the January 1993 run for wild-type Fe protein in the presence of MgATP at 12 mg/ml.
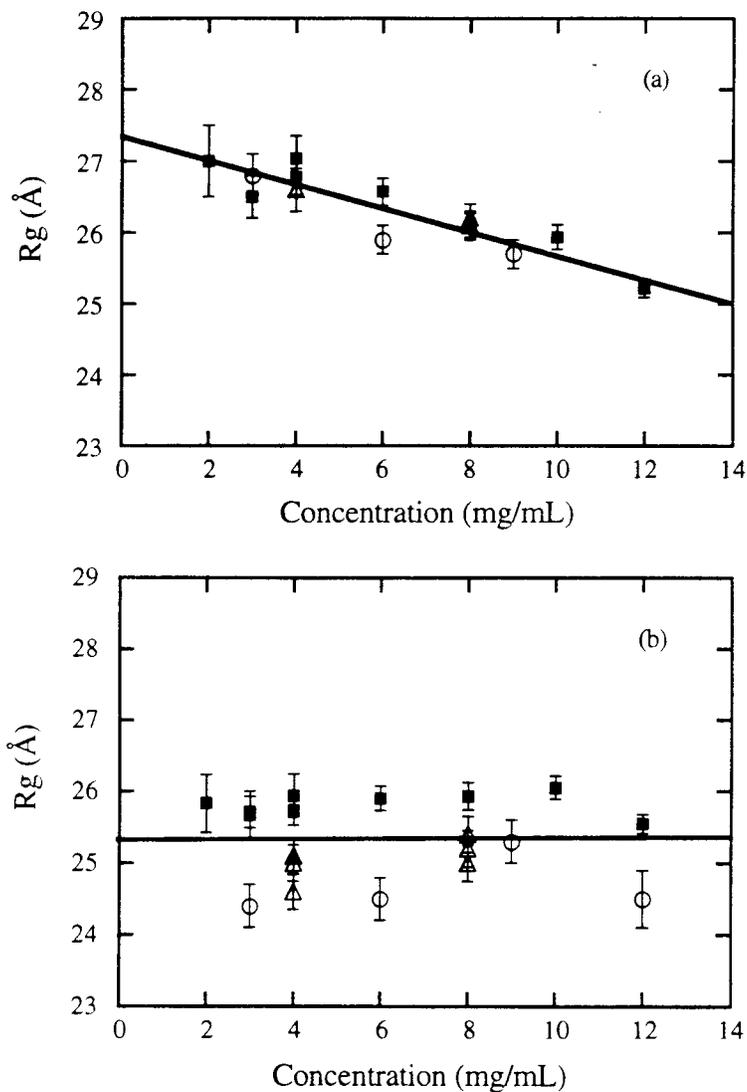
Figure 2.4. Radii of gyration as a function of (a) the wild-type Fe protein in the absence of nucleotides; (b) the wild-type Fe protein in the presence of MgATP; (c) the wild-type Fe protein in the presence of MgADP; (d) the mutant UW91 Fe protein in the absence of nucleotides; (e) the mutant UW91 Fe protein in the presence of MgATP. O, data of May 1992 run; ■, data of January 1993 run; Δ, data of May 1993. Least squares methods were employed in the extrapolation. The errors (shown as bars) are estimated from the propagation of all errors from counting statistics.
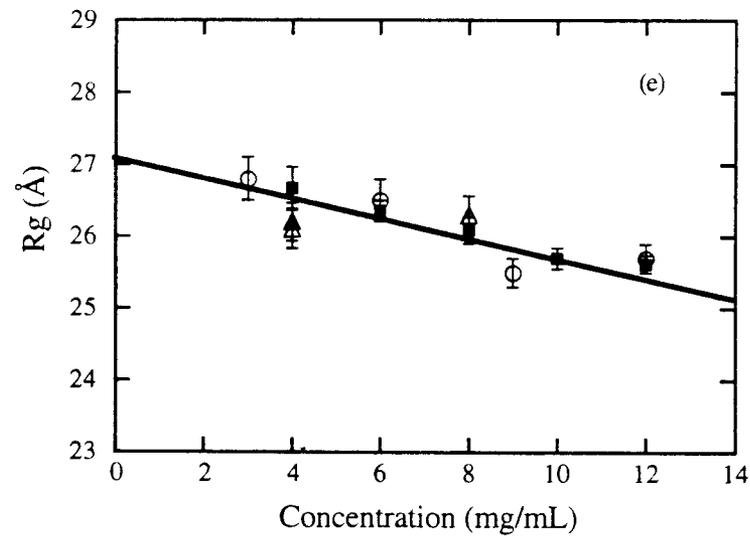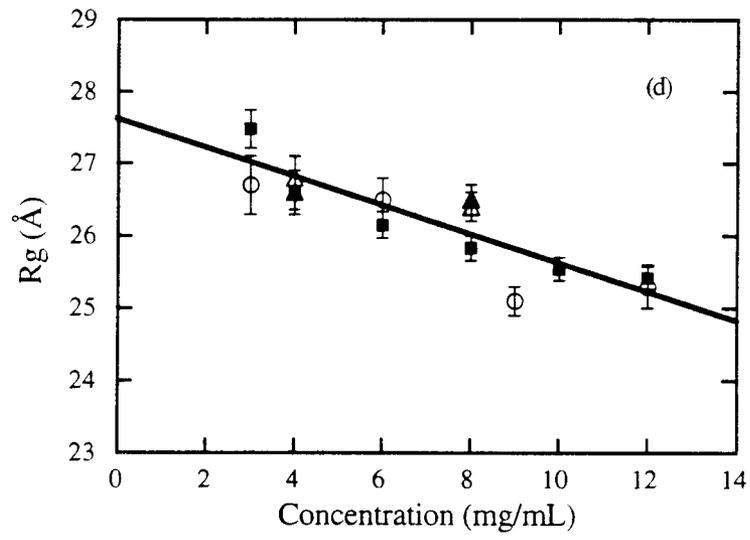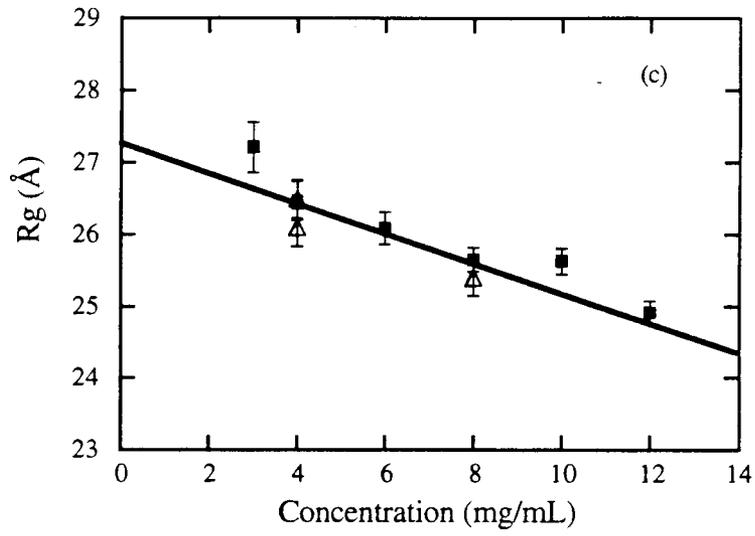
of $Na_2S_2O_4$ added just before the sample was exposed to x-rays). For example, for the wild-type without nucleotides sample, multiple measurements from two independent runs with flat-window cells gave R_g of 27.0 ± 0.3, 26.8 ± 0.3, 26.6 ± 0.3 Å for protein concentration at 4 mg/ml, and 26.1 ± 0.2, 26.1 ± 0.2, 26.2 ± 0.2 Å for 8 mg/ml. For the wild-type with MgATP sample, R_gs were measured to be 25.9 ± 0.3, 25.7 ± 0.2, 25.0 ± 0.3, 25.1 ± 0.2, 24.6 ± 0.3 Å for 4 mg/ml, and 25.9 ± 0.2, 25.4 ± 0.2, 25.2 ± 0.1, 25.0 ± 0.2 Å for 8 mg/ml. Comparable reproducibility was seen for the other three samples. All these data are shown in the Figures 2.4.a-e, from which it is very clear that the observed trends are reproducible and significant.

Changes in R_g observed upon addition of nucleotides. The final R_g values obtained from linear least-squares fits of the data points of three runs and subsequent extrapolations of R_g to zero protein concentrations are listed in Table 2.1. There are two major changes for the wild-type Fe protein associated with the binding of MgATP. First, the presence of MgATP reduces the radius of gyration of wild-type Fe protein by about 2.0 Å from 27.3 ± 0.2 Å to 25.2 ± 0.2 Å. The errors in Table 2.1 represent the statistical errors derived from the linear fit to the data. Although these errors are similar to those reported in similar studies in the literature (see McDonald, *et al.*, 1979; Pickover, *et al.*, 1979 for examples), our experience leads us to believe that these errors are underestimated due to systematic effects. A reasonable estimate for the maximum error is the biggest error in the individual measurements, in this case 0.5 Å for the 2 mg/ml samples. The true error probably lies between the values in Table 2.1 and this maximal value. However, even at the worst case of 0.5 Å error, the 2.0 Å difference is still outside the range of experimental error. In addition, a ~ 1 Å change in R_g resulting from protein conformational change is not uncommon in the literature, examples are SAXS studies of the substrate binding of hexokinase and phosphoglycerate kinase (McDonald, *et al.*, 1979; Pickover, *et al.*, 1979). It is therefore concluded that the 2.0 Å decrease in R_g detected by these measurements is experimentally significant and demonstrates a contraction of the wild-type Fe protein when MgATP is bound.

The second change concerns the dependence of the R_g on concentration. For four samples, wild-type without nucleotides, wild-type plus MgADP, UW91 without nucleotide and UW91 plus MgATP, there is an apparent decrease in R_g with increasing protein concentration. As shown in Fig. 2.4.b, the slope of R_g versus concentration for the wild-type protein in the presence of MgATP is significantly different from the other four samples. Despite run-to-run systematic errors, the different dependence of R_g on

Table 2.1. Radii of gyration for the wild type and mutant UW91 Fe protein in the absence and presence of nucleotides.

Fe protein	R_g in the absence of nucleotides (Å)	R_g in the presence of MgATP (Å)	R_g in the presence of MgADP (Å)
wild-type Fe protein	27.3 ± 0.2	25.2 ± 0.2	27.2 ± 0.2
mutant UW91	27.6 ± 0.3	27.1 ± 0.2	N.A.[a]

[a] N.A., not applicable.

These data are derived by the Guinier approximation followed by the linear extrapolation of R_g to zero protein concentration, as shown in Fig. 2.4. (a)-(e). The errors are derived from the least-squares linear extrapolation of R_g vs. concentration. (No data of mutant UW91 Fe protein in the presence of MgADP were collected because this mutant does not hydrolyze MgATP.) These results are derived from the combined data from all runs.

concentration for the wild-type Fe protein in the presence of MgATP was observed reproducibly in each independent run, indicating it is not an artifact. This slope will be discussed further later.

Table 2.1 shows that there is no detectable difference in the R_g observed between the wild-type and the UW91 mutant of Fe protein in the absence of added nucleotides. These data suggest that both proteins are initially in the same conformational state. There is also no detectable difference in the R_g observed when MgATP is added to the UW91 Fe protein, and when MgADP is added to the wild-type protein. These results confirm the conclusion drawn from chelation experiments that MgATP does not induce a detectable conformational change in the UW91 Fe protein (Gavini & Burgess, 1992), and that MgADP inhibits the chelation (Walker & Mortenson, 1973; Walker & Mortenson, 1974). All these data show that within the sensitivity of the SAXS experiment, the wild-type protein in the absence of nucleotide, the wild-type protein in the presence of MgADP, the UW91 protein in the absence of nucleotide and the UW91 protein in the presence of MgATP adopt a similar conformational state, which is different from that of wild-type protein in the presence of MgATP. More importantly, the lack of conformational change in the Fe protein mutant supports the hypothesis that a proper conformational state needs to be established for electron transfer to occur (Gavini & Burgess, 1992).

Possible nature of the MgATP-induced conformational change for the wild-type Fe protein. The SAXS measurements described above provide direct evidence that the wild-type Fe protein undergoes an overall contraction in the presence of MgATP. This conformational contraction is likely to be the source of the previously reported changes associated with the environment around [4Fe-4S] cluster following the binding of MgATP. These changes following the MgATP binding have been detected by EPR (Orme-Johnson, *et al.*, 1972; Smith, *et al.*, 1973; Zumft, *et al.*, 1972), electrochemistry studies (Morgan, *et al.*, 1986; Zumft, *et al.*, 1974), chelation studies (Walker & Mortenson, 1973; Walker & Mortenson, 1974) as well as oxygen sensitivity experiments.

Considering that the Fe protein consists of two monomers connected by a [4Fe-4S] cluster, and that the nucleotide binding sites are at the cleft between the two lobes of the protein molecule (Georgiadis, *et al.*, 1992), the present results could be accounted for by a model involving the hinged movement of the two subdomains of the Fe protein. A hinged motion of two domains of a protein molecule induced by binding substrate has been observed in hexokinase (McDonald, *et al.*, 1979) and phosphoglycerate kinase (Pickover,

et al., 1979), for example. In the case of hexokinase, the decrease of ~ 1 Å in R_g by SAXS is paralleled by the observations in a crystal structure study of a rotation of 12° of one lobe relative to the other and a closing of a cleft between two lobes with the movement of polypeptide backbone by 9 Å (McDonald, *et al.*, 1979).

For the Fe protein, the nucleotides are proposed to bind at one subunit with its phosphate groups and with the ribose and adenine extending across the subunit interface to interact with residues of the other subunit (Georgiadis, *et al.*, 1992). The binding of nucleotides might therefore result in narrowing the gap between the subunits. A decrease of 2.0 Å in R_g may originate from a several Å reduction in interdomain distance, and probably a rotation between domains. If this explanation is correct, however, the present results show that only MgATP, and not MgADP, can induce the closure of the gap. It may be that the additional contacts formed between protein domains and the γ-phosphate groups of ATP are required to bring the two subdomains closer to each other. Furthermore, this closure of domains induced by the binding of MgATP has to occur in such a way that the environment around the [4Fe-4S] cluster is altered to a state which is optimally configured to transfer electrons to the MoFe protein. The most straightforward hypothesis would be that the [4Fe-4S] cluster which is more buried in the protein matrix in the absence of MgATP becomes more exposed with the closure of the cleft, making it more accessible. Since it is proposed that the region of the Fe protein which includes the cluster is involved in forming the complex with the MoFe protein for electron transfer (Georgiadis, *et al.*, 1992), it is conceivable that the more exposed [4Fe-4S] cluster facilitates the effective transfer of electrons.

It is shown in Fig. 2.4.b that for the wild-type Fe protein in the presence of MgATP the radii of gyration are nearly independent of concentration. The effect of concentration-dependence of R_gs includes both interparticle interference and contamination of trace amounts of higher aggregation states (Hubbard, 1987). The nature of interparticle interference is such that it decreases the scattered intensity in the Guinier region (Glatter, 1982). On the other hand, because the forward intensity is roughly proportional to the square of molecular weight, the presence of higher aggregation states, *e.g.* dimers, will increase the intensity also in the Guinier region. Thus the lack of R_g concentration-dependence in the MgATP-bound wild-type protein could result from a cancellation between these two contributions. Since both the interparticle interference and dimerization originate from intermolecular forces, this cancellation may reflect a change in the intermolecular force of the MgATP-bound wild-type protein. The absence of this effect in

36

the MgATP-bound mutant UW91 protein further suggests that this change of intermolecular forces in the MgATP-bound wild-type Fe protein might be associated with the conformational change induced by MgATP. It is in any case quite clear that the concentration-dependence of solution aggregation of the Fe protein may be quite different for the MgATP-bound wild type, perhaps reflecting altered surface electrostatic properties.

Finally, we notice that at the highest scattering angle region available for these measurements ($S=0.025$ Å^{-1}), the scattering intensity from the wild-type Fe protein solution in the presence of MgATP is different from those of the other four systems. The higher angle region is sensitive to the internal structure of the protein molecule, *e.g.* the arrangement of secondary structure (Glatter, 1982). Therefore, it is worthwhile to extend our data collection to higher angle regions. This will enable more reliable use of the indirect transform method to obtain pair distribution (P(r)) functions, which have been found to be useful in studying subunit organization (Glatter, 1982). Since the crystal structure of the Fe protein has recently been solved (Georgiadis, *et al.*, 1992), it now becomes feasible to build a molecular model and alter it to investigate the effects of MgATP binding on the SAXS curves and P(r) functions. Such comparisons could provide molecular information on how MgATP triggers the Fe protein conformational change, which in turn is required for electron transfer.

2.4.2. The MoFe Protein

As an initial step to investigate the interaction between the MoFe protein and the Fe protein, scattering of the MoFe protein alone was measured. No sample color change was observed during or after data collection. In the Guinier plot, there is no obvious upswing at the small angle region, indicating that sample was largely aggregation free. Fig. 2.5. shows the concentration dependence of R_g for the MoFe protein. An extrapolated value of 39.8 ($\pm$0.4) Å is obtained, which is in good agreement with 39.2 Å calculated from the atomic coordinates including a 3 Å thick hydration layer, by CRYSOL17. A direct calculation from the coordinates by equation [1], however, gives a much lower value of 31.7 Å.
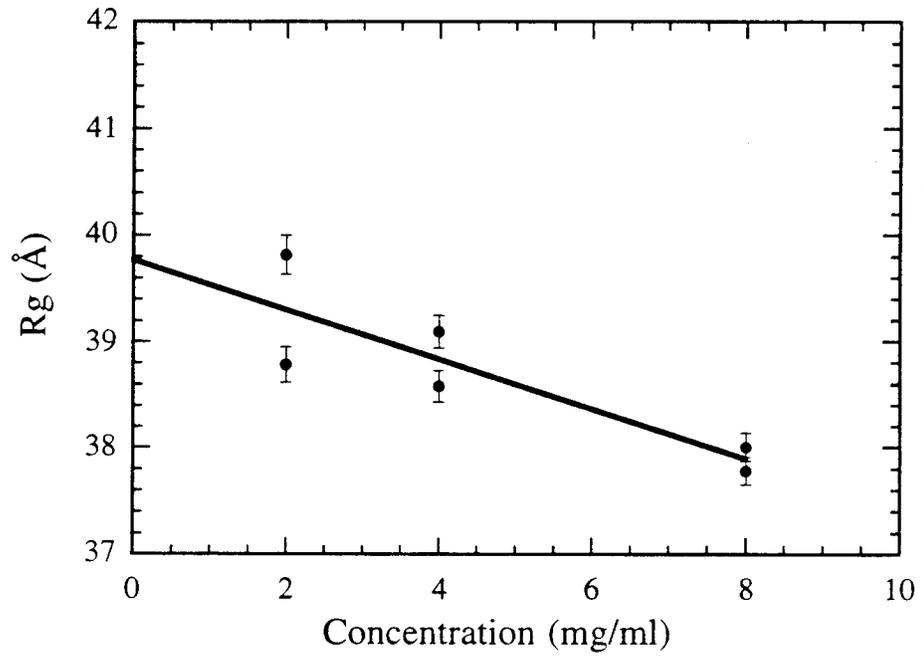
Figure 2.5. R_g as a function of the MoFe protein concentration and the extrapolation to zero protein concentration.

2.4.3. The FeMo Cluster

The color of the cluster solution varies from dark brown to light brown depending on the concentration. No sample color change was observed during/after data collection which normally lasted at least 2 hours.

In the first experiment of June 94, two solutions (in MEK and NMF) at a total of six concentrations were tested. For all three concentrations of the MEK solution, a significantly high counting rate was immediately observed when the sample was exposed to x-rays, indicating that the sample formed large aggregates in the solution. Scattering curves resemble those from highly aggregated poly-disperse systems. In Guinier plots, there are no linear regions available. Both the counting rate as well as the scattering curves did not show time dependence over the period of data collection (> 1 hour), suggesting that the aggregation was not due to radiation. Taken together, these observations indicate that the cluster form large nonspecific aggregates in MEK solution.

The typical counting rate in the first NMF experiment of June 94 was about 4,000 counts per second, for both cluster-containing samples and for background buffers. Interestingly, after normalizing for incident flux, the scattering curves showed a slight time dependence at the beginning of data collection: Scattering intensity tended to decrease, rather than increase which would be expected for the time- or radiation-dependent aggregation. About 60 minutes later, scattering curves became superimposable. This time dependence trend repeated in all three different concentrations, excluding the possibility of beamline or sample related artifacts. The cluster might still be cold when loaded into the SAXS cell since the sample had been frozen in dry ice. The slightly decrease in scattering intensity may be correlated with the sample warming-up process, which might be accelerated by the exposure to x-rays.

Because of the aggregation problem in the MEK solution, only NMF solutions were measured in the run of 1995. Limited by sample supply, only two concentrations were available. Data of the lower concentration (463 μM) appear to be problematic: the time course of scattering decrease was observed again for the first hour; a bubble introduced during loading inside the glove box was seen by x-rays during the data collection resulting in a background subtraction problem. Therefore, only data from the concentrated sample (1.44 mM) was analyzed. In this run, the detector was placed

asymmetrically with respect to the beam allowing data collection over a longer S range. The use of a longer fitting range for Guinier approximation would minimize the effect of any aggregates in solution on the fitting.

After background subtraction and detector response inhomogeneity correction, for some samples, the scattering intensity around the beamstop was still higher than in the other region. The Guinier plot also showed an incline in the smallest angular range for those samples. Therefore, trace amounts of aggregates may still have been present in the NMF solution. Alternatively, the high residual intensity may result from parasitic scattering. Except for this smallest range, data over a relatively large region was found to be reasonably linear (see Fig. 2.6.). In any case, data at the troublesome region could be excluded in the Guinier analysis, and appeared not to affect analysis results. Fitting results of both runs are summarized in Table 2.2.

The averaged R_g from all four measurements is 3.6 ($\pm$ 0.4) Å, the error being the standard deviation of the four measurements. R_g calculated from coordinates of the cofactor inside the MoFe protein is 2.93 Å. (The coordinates were taken from Brookhaven Protein Databank, entry PDB1MIN.ENT of the MoFe protein.) R_g calculated from coordinates is normally smaller than from the measurement done in solution because of solvation effects. For example, for the Fe protein and the MoFe protein, R_g calculated from atomic coordinates by equation [1] is 22.5 Å and 31.7 Å, whereas values of 27.2 ($\pm$0.3) Å and 39.8 ($\pm$0.4) Å respectively are obtained from SAXS. If a hydration layer of ~ 3 Å thick is considered, better agreements are obtained (26.5 Å and 39.2 Å for the Fe protein and the MoFe protein, respectively). A 3.6 Å R_g in solution is, therefore, rather reasonable for an estimate of 2.93 Å from coordinates without hydration layers, suggesting that the cluster is largely monomeric in NMF. Alternatively, a larger R_g may be correlated with the change in the Mo environment observed by EXAFS study (Ma, *et al.*, 1994). Compared with the FeMo cofactor in NMF solution or in the MoFe protein matrix, the Mo atom in the FeMo cluster has one less Fe atom and one less N/O atom in the first coordination shell, and does not show a contribution from a long distance of 3 iron atoms in the subcluster Fe_4S_3. Thus, the conformation of the cluster could be different from that of the cofactor. The implication of the loss of an iron atom in the perturbation of the bridge between the two subclusters and in the loss of the long range order, might suggest a partial opening of the structure in the cluster, which would result in a larger R_g for the cluster compared with the cofactor.
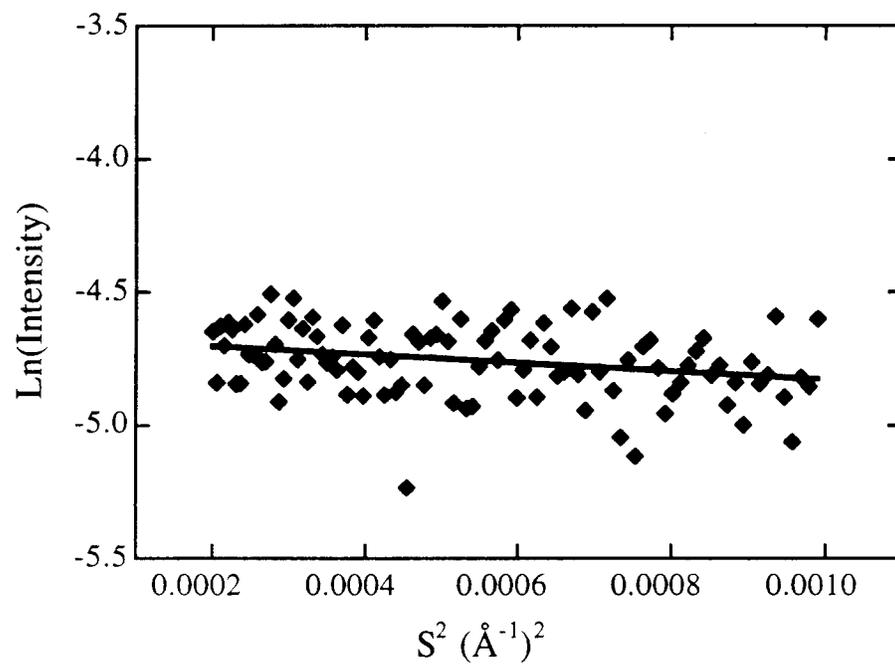
Figure 2.6. A representative Guinier plot of the FeMo cluster in NMF solution and the least-squares fit. The cluster concentration is 464 μM (June 94 data).

Table 2.2. Radii of gyration of the FeMo cluster calculated from Guinier approximation and the fitting range.

Run	Concentration	R_g (Å)	$S^2_{min} - S^2_{max}$
June of 94	464 μM	3.59 ± 0.59	0.0002 - 0.0010
	619 μM	4.16 ± 0.65	0.0002 - 0.0010
	731 μM	3.41 ± 1.64	0.0004 - 0.0012
June of 95	1.44 mM	3.39 ± 0.22	0.0003 - 0.0015

The correlation of the measured R_g for the cluster with the estimated R_g for the cofactor from the coordinates greatly contrasts with the much higher R_g (6.96 Å) obtained for the cofactor in the same solvent, NMF. A substantially higher intensity at the small angular region was observed in the cofactor samples, leading to a model of coexistence of small oligomers as well as large aggregates. As mentioned in section 2.1.4., compared with the cofactor, the FeMo cluster is obtained differently and exhibits some structural and functional differences. Results from SAXS experiments further demonstrate that the aggregation state in NMF solution also distinguishes the FeMo cluster from the FoMo cofactor. Therefore, it is suggested that some steps in the preparation, perhaps the loss of homocitrate, might be responsible for the change of the aggregation state.

The monodisperse monomeric state of the FeMo cluster also contrasts sharply with the cluster in MEK solution. In the latter, the cluster obviously formed large polydisperse aggregates such that there is no linear region in the Guinier plot. This is not unexpected because it is known that the cluster is more soluble in NMF than in MEK. The cluster can be concentrated up to 2-3 mM in NMF while the maximum concentration in MEK is ~ 1 mM. Considering its small size, a net charge of +2 renders the cluster a highly charged molecule. A much higher dielectric constant, 182.4 at 25 °C (Riddick & Bunger, 1970), makes NMF a better solvent for the cluster than MEK, which has a dielectric constant of 18.51 at 20 °C (Riddick & Bunger, 1970), which is even lower than H_2O, 80.10 at 20 °C (Weast, *et al.*, 1988). Obviously, NMF will separate the cluster molecules better, making a monodisperse solution, while the cluster tends to hold together in MEK, leading to polydispersity.

In summary, the FeMo cluster appears to be largely monomeric in NMF with a R_g of 3.6 ± 0.4 Å. On the contrary, the cluster aggregates significantly in the MEK solution.

2.5. Acknowledgments

2.6.   References

Burgess, B. K. (1983). *Advances in Nitrogen Fixation Research*. The Hague: Martinus-Nijhoff.

Burgess, B. K. (1990). *Chem. Rev.*, **90**, 1377-1406.

Burgess, B. K. (1993). *Molybdenum Enzyme, Cofactor and Models*. Washington D.C.: American Chemical Society.

Burgess, B. K., Jacobs, D. B., & Stiefel, E. I. (1980). *Biochim. Biophys. Acta.*, **614**, 196-209.

Conradson, S. D., Burgess, B. K., Newton, W. E., Mortenson, L. E., & Hodgson, K. O. (1987). *J. Am. Chem. Soc.*, **109**, 7507-7515.

Eliezer, D., Frank, P., Gillis, N., Newton, W. E., Doniach, S., & Hodgson, K. O. (1993). *J. Biol. Chem.*, **268**(28), 20953-20957.

Gavini, N., & Burgess, B. K. (1992). *J. Biol. Chem.*, **267**, 21179-21186.

Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J., & Rees, D. C. (1992). *Science*, **257**, 1653-1659.

Glatter, O. (1982). *Small-Angle X-ray Scattering*. New York: Academic Press.

Hedman, B., Frank, F., Gheller, S. F., Roe, A. L., Newton, W. E., & Hodgson, K. O. (1988). *J. Am. Chem. Soc.*, **110**, 3798-3805.

Holm, R. H. (1990). *Prog. Inorg. Chem.*, **38**, 1.

Hubbard, S. R. (1987) *Small-Angle X-ray Scattering Studies of Calcium-Binding Proteins in Solution*. Ph.D. Thesis, Stanford University.

Kim, J., & Rees, D. C. (1992). *Science*, **257**, 1677-1682.

Ljones, T., & Burris, R. H. (1978). *Biochemistry*, **17**, 1866-1872.

Ma, L., Gavini, N., Liu, H. I., Hedman, B., Hodgson, K. O., & Burgess, B. K. (1994). *J. Biol. Chem.*, **269**, 18007-18015.

McDonald, R. C., Steitz, T. A., & Engelman, D. M. (1979). *Biochemistry*, **18**, 338-342.

Morgan, T. V., Prince, R. C., & Mortenson, L. E. (1986). *FEBS Lett.*, **206**, 4-8.

Nielsen, A. (1968). *An Investigation of Promoted Iron Catalysts for the Synthesis of Ammonia*. Copenhagen: Jul. Giellerup Forlag.

Orme-Johnson, W. H., Hamilton, W. D., Jones, T. L., Tso, M.-Y. W., Burris, R. H., Shah, V. K., & Brill, W. J. (1972). *Proc. Natl. Acad. Sci. USA*, **69**, 3142-3145.

Pickover, A. C., McKay, D. B., Engelman, D. M., & Steitz, T. A. (1979). *J. Biol. Chem.*, **254**, 11323-11329.

Rawlings, J., Shah, V. K., Chisnell, J. R., Brill, W. J., Zimmerman, R., Munck, E., & Orme-Johnson, W. H. (1978). *J. Biol. Chem.*, **253**, 1001.

Riddick, J. A., & Bunger, W. B. (1970). *Techniques of Chemistry*. New York: Wiley-Interscience.

Robson, R. L. (1984). *FEBS Lett.*, **173**, 394-398.

Seefeldt, L. C., Morgan, T. V., Dean, D. R., & Mortenson, L. E. (1992). *J. Biol. Chem.*, **267**, 6680-6688.

Shah, V. K., & Brill, W. J. (1977). *Proc. Natl. Acad. Sci. USA*, **74**(8), 3249-3253.

Shah, V. K., & Brill, W. J. (1981). *Proc. Natl. Acad. Sci. USA.*, **78**, 3438-3440.

Shah, V. K., Davis, L. C., Gordon, J. K., Orme-Johnson, W. H., & Brill, W. J. (1973). *Biochem. Biophys. Acta*, **292**, 246-255.

Smith, B. E., Lowe, D. J., & Bray, R. C. (1973). *Biochem. J.*, **135**, 331-341.

Stephens, P. J., McKenna, C. E., McKenna, M., C., Nguyen, H. T., & Lowe, D. J. (1982). *Electron Transport and Oxygen Utilization*. New York: Elsevier Biomedical.

Svergun, D. I., Barberato, C., & Koch, M. H. J. (1995). *J. Appl. Crystallogr.*, submitted for publication.

Thorneley, R. N. F., Ashby, G. A., Fisher, K., & Lowe, D. J. (1993). Electron-transfer reactions associated with nitrogenase from *Klebsiella pneumoninae*. In Stiefel, E. I., Coucouvanis, D., & Newton, W. E. (Eds.), *Molybdenum Enzymes, Cofactors and Model Systems* (pp. 209-302). Washington, D. C.: American Chemical Society.

Thorneley, R. N. F., & Lowe, D. J. (1985). *Molybdenum Enzymes*. New York: Wiley.

Tso, M.-Y. W., & Burris, R. H. (1973). *Biochim. Biophys. Acta*, **309**, 263-270.

Wakatsuki, S., Hodgson, K. O., Eliezer, D., Rice, M., Hubbard, S., Gillis, N., Doniach, S., & Spann, U. (1992). *Rev. Sci. Instrum.*, **63**, 1736-1740.

Walker, G. A., & Mortenson, L. E. (1973). *Biochem. Biophys. Res. Commun.*, **53**, 904-909.

Walker, G. A., & Mortenson, L. E. (1974). *Biochemistry*, **13**, 2382-2388.

Walker, J. E., Saraste, M., Runswick, M. J., & Gay, N. J. (1982). *EMBO J.*, **8**, 945-951.

Weast, R. C., Astle, M. J., & Beyer, W. H. (Ed.). (1988). *Handbook of Chemistry and Physics* (69 ed.). Boca Raton, FL: CRC Press, Inc.

Willing, A., & Howard, J. B. (1990). *J. Biol. Chem.*, **265**, 6596-6599.

Yates, M. G. (1991). *Biological Nitrogen Fixation*. New York: Chapman and Hall.

Zumft, W. G., Cretney, W. C., Huang, T. C., Mortenson, L. E., & Palmer, G. (1972). *Biochem. Biophys. Res. Commun.*, **48**, 1525-1532.

Zumft, W. G., Mortenson, L. E., & Palmer, G. (1974). *Eur. J. Biochem.*, **46**, 525-535.

Chapter 3

Solution Small-angle X-ray Scattering Study of the Molecular Chaperone hsc70 and Its Subfragments

3.1. Introduction and Background

In response to unfavorable conditions such as heat shock, living organisms accelerate the rate of expression of some specific genes (heat shock genes); the products of these genes are referred to as heat shock proteins historically or stress proteins. Heat shock response was first observed in 1962 in *Drosophila hydei* where changes of puffing, or gene activity pattern, were induced by exposing salivary gland polytene chromosomes to evaluated temperature (Ritossa, 1962). However, it was the discovery of heat shock proteins in 1974 (Tissieres, *et al.*, 1974) that stimulated great interest from molecular biologists and launched the study of heat shock response. Since then knowledge in this field has been greatly accumulated and expanded, making the heat shock system one of the most stimulating and prosperous areas in molecular and cellular biology, and more recently in immunology and medicine.

Although heat shock proteins were initially recognized under heat shock condition, it was found later that organisms also enhance synthesis of these proteins in response to other general stress physiology, such as anoxia, chemical poisons and virus infection. It is further demonstrated that most of the heat shock proteins are also present under normal physiological conditions, but with less amount, and that they are crucial for cell viability (Morimoto, *et al.*, 1990). This subset of heat shock proteins, which are constuitively expressed, is referred to as heat shock cognates (hsc) to distinguish them from the stress induced heat shock proteins. The major heat shock proteins are classified into five subfamilies according to the approximate molecular weights (MW): 1) hsp100 with MW larger than 100 kDa; 2) hsp90 with MW of 83-90 kDa; 3) hsp70 with MW of 66-78 kDa; 4) hsp60 with MW of about 60 kDa; and 5) small heat shock proteins of 15-30 kDa. Strikingly, heat shock proteins are highly conserved within each class in evolution, for example, hsp70 family shares about 50% sequence conservation from *E. coli* to human.

Heat shock proteins were proposed to protect the cell against adverse environmental circumstances, and to facilitate the repair and recovery to normal cellular activities after the stress period (Nover, 1984). More specifically, since most inducers of stress response led to the accumulation of denatured proteins in the cell, heat shock proteins were implicated in functions dealing with this problem by preventing denatured proteins from misfolding or aggregation (Morimoto, *et al.*, 1990). Recent progress in this area reveals a better understanding of the biochemical properties of heat shock proteins

48

under both stress and normal conditions. They are involved in regulation of protein folding, protein synthesis, translocation of proteins across membranes, and assembly and disassembly of macromolecular complexes. All of these functions can be ultimately ascribed to the ability of heat shock proteins to recognize and interact with "unstructured" segments present in the other proteins. In particular, because of their roles in protein folding process, heat shock proteins are also called molecular chaperones (Morimoto, et al., 1994).

The 70 kDa heat shock proteins. Members of the 70 kDa heat shock protein family were identified during the first description of heat shock proteins in 1974 (Tissieres, et al., 1974). It was further found that hsp70 are the most abundant heat shock proteins in many organisms with most of them present in unstressed cells. The constutively-expressed members, or heat shock cognates, are one of the earliest studied and best characterized heat shock proteins. Bovine heat shock cognate protein hsc70 was initially isolated and characterized as a protein with an *in vitro* clathrin uncoating activity (Chappell, et al., 1986; Schlossman, et al., 1984). It has subsequently been implicated in transmembrane targeting of proteins (Chirico, et al., 1988; Deshaies, et al., 1988), and has been found associated with nascent polypeptides during translation (Beckman, et al., 1990; Langer, et al., 1992; Scherer, et al., 1990). In addition, *in vitro* studies have shown that the 70 kDa heat shock proteins bind both denatured proteins (but not their native counterparts) and some short (≥ 7 residues) synthetic peptides, and that these (poly)peptides are released in response to addition of MgATP (Blond-Elguindi, et al., 1993; Flynn, et al., 1989; Flynn, et al., 1991; Gragerov, et al., 1994).

The 70 kDa heat shock proteins have an ATPase activity, a peptide binding activity, and a coupling mechanism that regulates the peptide binding and release in a nucleotide-dependent manner (Palleros, et al., 1993). For bovine hsc70, the ATPase activity resides in an amino-terminal fragment of about 44 kDa with residues 1 to ~386 (Chappell, et al., 1987), and the peptide binding ability lies within an 18 kDa fragment immediately following the ATPase portion (Wang, et al., 1993). A construct consisting of the ATPase fragment and the peptide binding fragment, called the 60 kDa fragment, retains both the ATPase activity and the peptide binding ability (Chappell, et al., 1987). The function of the remainder ~ 10 kDa with residues 550-650 at the carboxyl terminal is not clear; however, this part of hsc70 has been implicated in interaction with other cellular components. Fig. 3.1 shows the functional arrangements within the primary

49

ATPase fragment
highly conserved in sequence

peptide binding fragment
relatively conserved in sequence

carboxyl terminal fragment
variable in sequence

N ——— C

amino acid sequence 1   ~ 385   ~ 550   ~ 650

the ATPase fragment (MW 44 kDa)   the peptide binding fragment (MW 18 kDa)

the 60 kDa fragment (MW 60 kDa)

the full-length hsc70 (MW 70 kDa)
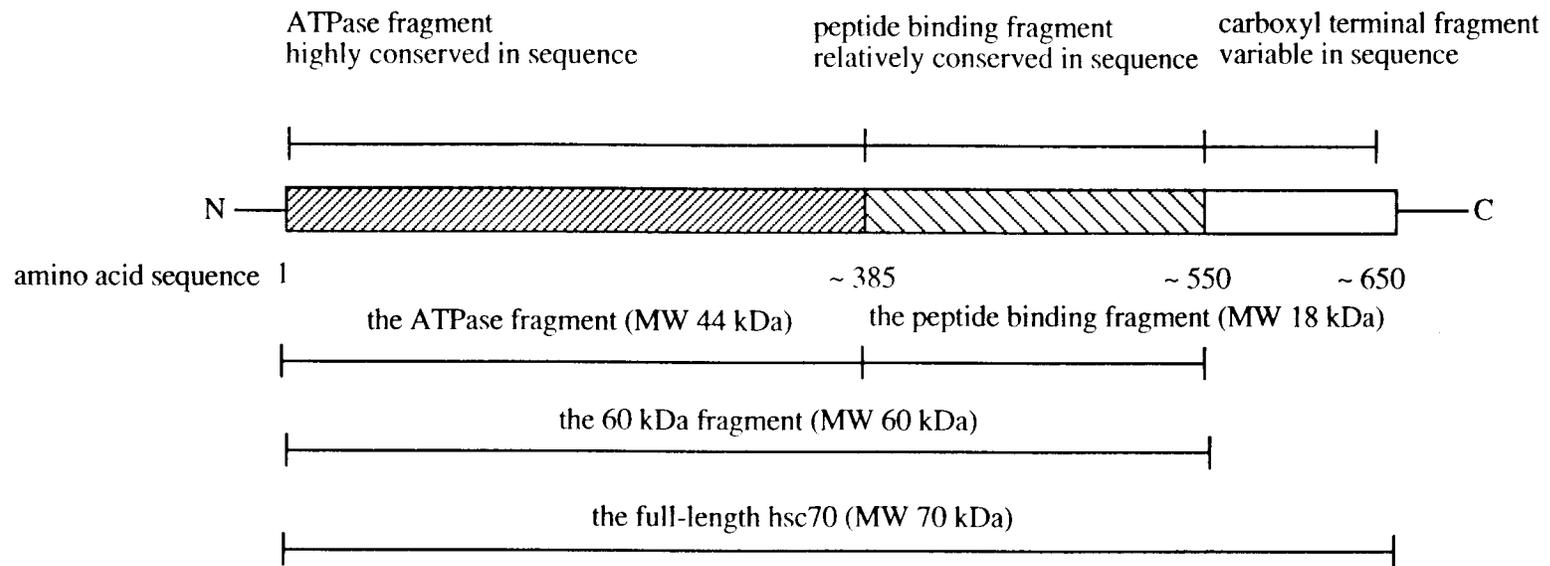
Figure 3.1. Schematic drawing of the functional organization of bovine hsc70 within the primary structure.

structure of bovine hsc70 (McKay, 1993). Interestingly, although hsp70 is highly conserved in amino acid sequence from bacteria to human, the conservation is not equally distributed over the sequence. As shown in Fig. 3.1, the most highly conserved region is in the ATPase fragment, followed by the less conserved peptide binding fragment. The carboxyl terminal segment is highly variable in sequence.

The turnover rate of ATP by hsp70 is relatively slow: 0.018 (mol ATP mol^{-1} hsc70 min^{-1}) for the bovine hsc70 (Ha & McKay, 1994), for example. The weak ATPase activity is enhanced severalfolds in the presence of stimulatory peptides. A maximal ATPase rate of ~ 1.0 (mol ATP mol^{-1} hsc70 min^{-1}) is reported for the bovine hsc70 in the presence of clathrin light chain peptides (Deluca-Flaherty, et al., 1990). It is interesting to note that this stimulated ATPase activity of the full-length hsc70 is similar to that of the ATPase domain alone, ~ 0.8 (mol ATP mol^{-1} ATPase fragment min^{-1}) under identical assay conditions (Deluca-Flaherty, et al., 1990). The C-terminal region including the peptide binding fragment appears to attenuate the ATPase activity of the full-length hsc70; binding of polypeptide substrates somehow restores the activity to a rate characteristic of the ATPase fragment of hsc70 alone. The stimulation of ATPase rate by peptides strongly indicates a close correlation between the ATPase and the peptide binding fragments of hsc70.

So far no crystal structures are available for the full-length hsp70 proteins. However, the ATPase domain of bovine hsc70 has been solved to 2.2 Å resolution, and the structure shows a striking similarity to the tertiary structure of actin (Flaherty, et al., 1990; Flaherty, et al., 1991). Determination of secondary structure of the peptide binding domain of rat hsc70 by NMR has also been reported (Morshauser, et al., 1995).

At least two distinct conformational states have been suggested in hsp70 proteins: one has relatively high affinity for peptides or denatured proteins, while the other has significantly reduced affinity. Proteolysis studies on DnaK, the E. coli representative of the hsp70 family, have shown different digestion patterns associated with the presence of MgATP and MgADP (Liberek, et al., 1991). Changes in intrinsic trytophan fluorescence spectra of DnaK have been observed between MgATP-bound form and MgADP-bound form (Palleros, et al., 1993). The MgADP-bound state of DnaK has a high affinity for denatured protein substrates; the addition of MgATP to the DnaK-MgADP-denatured protein complex leads to a release of the denatured protein substrate and shifts to the MgATP-bound conformation, suggesting a decrease in peptide binding affinity associated

51

with the MgATP-bound state (Palleros, *et al.*, 1993). It is further shown in the same study that the conformational transition triggered by MgATP requires the presence of K$^+$.

To develop a low resolution model for hsc70, to determine the magnitude of and the nucleotide requirements for the ATP-induced conformational change, and to attempt to correlate the conformational change with specific steps of the enzymatic ATPase cycle, we have carried out solution small-angle x-ray scattering studies of recombinant bovine hsc70 and its subfragments.

3.2. Experimental Procedures

3.2.1. Sample Preparations

Expression and purification of protein. Full-length recombinant bovine hsc70 (amino acid residues 1-650), which has ATPase and peptide binding activities, and the truncated fragment consisting of residues 1-386 ("ATPase domain") were expressed in *E. coli* as described elsewhere (Wilbanks, *et al.*, 1994). The plasmid for expression of the truncated fragment including residues 1-554 ("60 kDa fragment"), which also has ATPase and peptide binding activities, was subcloned from the pT7-7-derived plasmid used to express hsc70. This was achieved by replacing an EcoRI/SalI fragment including the codons for residues 540 to 650, termination signals and part of the polylinker with a synthetic oligonucleotide encoding residues 540-554 and a termination codon. Using the same procedure, the truncated fragment encoding residues 388-554 ("peptide binding domain") was subcloned from a pT7-7-derived plasmid that encodes the carboxyl terminal fragment of the protein, residues 388-650. Protein expression was induced in *E. coli* strain BL21(DE3) by addition of isopropylthio-β-D-galactoside (0.4 mM final concentration) to log phase cultures.

Purification schemes for hsc70 and the ATPase fragment were modifications of those published previously (O'Brien & McKay, 1993; Wilbanks, *et al.*, 1994). Cells were lysed by sonication, the lysate was cleared by centrifugation at ~16,000g and applied to a DE-52 anion exchange column (Whatman, Clifton, NJ) equilibrated in 25 mM Tris, 25 mM KCl, 0.1 mM PMSF, pH 7.0. hsc70 was eluted with 25 mM Tris, 150 mM KCl, 0.1 mM PMSF, pH 7.0; the ATPase fragment was eluted with a gradient of 25-200 mM KCl in the same buffer. Both proteins were then purified using ATP-agarose (Sigma, St.

Louis, MO) as published previously, except that hsc70 was applied to ATP-agarose by batch absorption with rocking for 16 hours at 4°C rather than by conventional loading of a pre-poured column. Chromatofocusing over a Mono-P column, followed by gel-filtration on a Superdex-75 column, was used as the final step of purification (both columns from Pharmacia, Uppsala, Sweden). Final buffer conditions, established on the gel filtration column, were 10 mM MOPS, 150 mM KCl, pH 7.0 for the ATPase fragment and 10 mM MOPS, 100 mM K(OAc), pH 6.5 for hsc70.

A variation of this protocol was used to purify the 60 kDa fragment. Since it was not retained on the DE-52 column, it was collected in the flowthrough, applied to a hydroxylapatite column (Bio-Rad, Richmond, CA) equilibrated in 20 mM potassium phosphate, 50 mM calcium phosphate, 0.1 mM PMSF, pH 7 and then eluted with a gradient of 20-200 mM potassium phosphate, also with calcium and PMSF at pH 7. Chromatofocusing and gel filtration were carried out as described above; the 60 kDa protein was stored in 10 mM MOPS, 150 mM KCl, pH 7.0.

Cells expressing the peptide binding domain were harvested and lysed in the same fashion, and the lysate was cleared by centrifugation. The supernatant fraction from that spin was brought to 40% saturation with $(NH_4)_2SO_4$, the resulting precipitate was removed by centrifugation, the supernatant fraction was brought to 60% saturation with $(NH_4)_2SO_4$ and the resulting precipitate again removed by centrifugation. The supernatant fraction, containing the majority of the peptide binding domain, was applied to phenyl-sepharose (Sigma, St. Louis, MO) equilibrated in 2 M $(NH_4)_2SO_4$, 20 mM bis-tris propane, pH 7, eluted with 0.6 M $(NH_4)_2SO_4$, 20 mM Bis-Tris propane, pH 7, and precipitated by addition of $(NH_4)_2SO_4$ to near saturation. Resuspended protein was filtered and chromatographed by gel-filtration on Superdex-75 in 10 mM MOPS, 100 mM K(OAc), pH 7.

Kinetic assays. Kinetic rate constants for ATP hydrolysis, product release, and steady-state ATPase turnover were determined for the 60 kDa fragment at 25 °C, in the presence of 10 mM MOPS, 4.5 mM $Mg(OAc)_2$, 150 mM KCl, pH 7, using methods described previously (Ha & McKay, 1994).

The rate constant for ATP hydrolysis was determined under single turnover conditions (1 or 2 µM 60 kDa fragment, 10 nM $\alpha-[^{32}P]$-ATP). Aliquots of the reaction mixture were withdrawn at specific time points, quenched with acid and neutralized, and

the α-^{32}P-labeled adenosine nucleotides were separated by thin layer chromatography. The relative amount of each radioactive nucleotide was determined with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA) system, and the fraction hydrolyzed was computed as the ratio of counts in the ADP spot to the sum of counts in both ADP and ATP spots (after correction for background). Data on the fraction of ATP hydrolyzed versus time were fit with a single exponential, yielding the hydrolysis rate constant.

The rates of release of products were determined using filter binding assays as described previously (Ha & McKay, 1994). To determine the rate of ADP release, protein was first incubated with stoichiometric amounts of α-$[^{32}P]$-ATP for sufficient time to allow complete hydrolysis (> 6 x hydrolysis half-life). Then, an excess of unlabeled ATP was added, and the amount of label released from the protein was followed as a function of time; concentrations after addition of unlabeled ATP were: 2 mM protein, 100 mM ATP. Aliquots were withdrawn from the reaction and filtered immediately through BA-85 nitrocellulose membranes (Schleicher & Schuell, Keene, New Hampshire); filters were then washed briefly with buffer, dried and counted by Cherenkov counting. A rate was inferred from a single exponential fit to the data. To determine the rate of release of P_i, reactions were initiated by mixing 10 nM of γ-$[^{32}P]$-ATP with 10 mM of protein, and the release of $^{32}P_i$ was monitored by filter binding. The rate constant for P_i release was extracted from the data using methods described previously which take into account the effect of ATP hydrolysis on the apparent release rate (Ha & McKay, 1994).

Steady state hydrolysis rates were determined in reactions containing 40 or 400 nM enzyme, 4 μM ^{32}P-α-ATP in the same buffer as for single turnover reactions. Separation and measurement of ATP and ADP were by TLC and PhosphorImager as described above. The hydrolysis rate was computed as the slope of a linear fit to data representing the fraction of ATP hydrolyzed versus time.

Preparation of protein samples for solution scattering. hsc70 and the peptide binding domain aggregate significantly when left in solution at high (~10 mg/ml) concentrations for an extended period of time (~days). To circumvent this problem, samples of these proteins used for scattering measurements were rendered aggregate-free by gel filtration within 24 hours of their use, and measurements were made at concentrations less than 4 mg/ml (for hsc70) or 8 mg/ml (peptide binding fragment). This precaution was not

54

necessary for the ATPase and 60 kDa fragments, since they had less propensity to aggregate. Protein concentrations were determined spectrophotometrically, using calculated molar extinction coefficients at λ=280 nm of 30,800; 23,700; 18,600 and 5,100 $(M^{-1}cm^{-1})$ for hsc70, the 60 kDa fragment, ATPase fragment, and peptide binding domain, respectively. The molecular weighs of these species were calculated to be 71.2; 61.0; 42.5 and 18.4 kDa, respectively. Samples were judged to be nucleotide-free based on an A_{280}/A_{260} ratio of >1.4. Where necessary, nucleotide was removed from protein by charcoal extraction as described previously (Ha & McKay, 1994).

Non-denaturing polyacrylamide gels were 7%, with 4% glycerol present (Kim, et al., 1992). Dynamic light scattering was measured on a Biotage dp801 Molecular Size Detector using samples filtered through 0.02 μm filters.

The above sample preparations and kinetic assays were performed by Dr. Sigurd Wilbanks at Prof. David McKay's laboratory at the Department of Structural Biology, Stanford Medical School.

Standard protein samples used in this study. Bovine serum albumin (from Sigma, St. Louis, MO), was dissolved in 50 mM Na-phosphate, 150 mM NaCl, pH 8.0 and further purified by gel filtration using Superose 6 (Pharmacia, Uppsala, Sweden). The solution was concentrated with Diaflo ultrafilter PM30 (Amicon, Danvers, MA) at 40 psi. The final protein concentration was determined to be 18.8 mg/ml using ε_{280}= 3.96 x 104 M^{-1} cm^{-1}.

Chicken egg-white lysozyme (Sigma, St. Louis, MO) was dialyzed against 40 mM NaOAc, 150 mM NaCl, pH 3.8, and the protein concentration was determined spectrophotometrically using ε_{280} = 2.64 $mg^{-1}cm^{-1}ml$.

3.2.2. X-ray Scattering Data Collection

X-ray solution scattering experiments were performed on beamline 4-2 of the Stanford Synchrotron Radiation Laboratory in March 1994, June 1994, and March 1995 using the resident SAXS camera (Wakatsuki, et al., 1992). The x-ray source (SPEAR) was operated at an energy of 3.0 GeV and a current of 60-100mA. The synchrotron beam was monochromatized (to an energy of 8980 eV) by a double crystal Si [111] monochromator. The beamsize at the sample position was 1.0 mm (v) x 2.1 mm (h). A

Pt-coated mirror was used to reject higher order harmonics in the March 1994 and June 1994 runs, and additionally as a focusing mirror in the March 1995 run. Incident beam intensity was monitored using an ionization chamber positioned immediately before the sample chamber. Both a linear detector (a one-dimensional position-sensitive proportional counter; manufactured by Bio-Logic, Grenoble, France) and a quadrant detector (developed at the EMBL, Grenoble) were used in separate experiments. The linear detector was filled with 80% xenon/20% carbon dioxide and was operated at 1.8 - 1.9 kilovolts during the experiments; the quadrant detector was filled with 70% argon/30% carbon dioxide and was operated at 3.4 kilovolts. Sample-to-detector distance was determined from the diffraction pattern of a cholesterol myristate powder sample, and was found to be 234 ± 1 cm for the linear detector and 238 ± 1 for the quadrant detector. The linear detector covered the scattering range $0.0002 < S < 0.003$ Å^{-1} ($S = 2\sin\theta/\lambda$, where 2θ = scattering angle and λ = x-ray wavelength) and the quadrant detector covered $0.0005 < S < 0.004$ Å^{-1}. Detector response was calibrated every six to twelve hours using a homogenous field of radiation, either from copper fluorescence (for the linear detector) or an [55]Fe source (for the quadrant detector).

Sample cells were Kel-F or Macor (Corning, Corning, NY) blocks with a 25-30 μl sample chamber, 1.2 mm path length and 10 μm thick mica windows. The temperature of the cell was controlled by circulating ethylene glycol from a temperature-controlled bath through the jacket of the cell holder.

During the March 1994 and June 1994 runs, data collection from each sample was limited to 10 minutes, divided into either 5 cycles of 2 minutes, or 10 cycles of 1 minute, since cumulative exposure times significantly greater than 10 minutes resulted in an upswing of the scattering curve at low angle, indicative of radiation-induced sample deterioration. During the March 1995 run, it was found that deterioration could be retarded by addition of 20 mM β-mercaptoethanol to the protein samples, allowing data collection time to be extended to 20 minutes. Background scattering was recorded either immediately before or immediately after each protein sample using the same sample cell filled with the corresponding buffer and nucleotide solution. Multiple samples were used under each set of conditions, and their results were averaged, in order to improve statistical precision in the final results. Representative samples that were recovered and assayed for ATPase activity after data collection were found to be fully enzymatically active.

56

3.3. Data Analysis

3.3.1. Data Reduction and Analysis

The SSRL software ANOM was used for data reduction (Rice & Wakatsuki, 1991). Each data cycle was normalized to the incident beam intensity and examined for obvious anomaly, such as failure of the system to record the data, or radiation-induced aggregation; anomalous sets of data amounted to fewer than 0.5% of all data sets and were discarded. Repetitive cycles of data collected on each sample were averaged and the measured background scattering profile was subtracted. A detector response correction was applied as a channel-by-channel division of the data by the detector response (integrated intensity for each channel) to a uniform field of radiation. For data collected with the linear detector, the two scattering curves from opposite sides of the direct beam were averaged.

Radii of gyration (R_g) were extracted from linear detector data with two different methods. First, a value of R_g was computed using the Guinier approximation (Glatter & Kratky, 1982),

$$LnI(S) = LnI(0) - \frac{4\pi^2 R_g^2}{3}S^2 \qquad [1]$$

where $I(S)$ is the scattered intensity at S, and $I(0)$ is the intensity at zero angle. This equation was fit to data over a range of $0.0032 < S < 0.0065$ Å^{-1} for samples of hsc70 or 60 kDa fragment, and over $0.0032 < S < 0.0095$ Å^{-1} for samples of ATPase fragment or peptide binding domain.

Second, the pair distribution function

$$P(r) = 8\pi r \int_0^\infty SI(S)\sin(2\pi Sr)dS \qquad [2]$$

was calculated using the indirect transformation method of Svergun, as implemented in the GNOM program package (Semenyuk & Svergun, 1991), with data over the angular range $0.0028 < S < 0.0280$ Å^{-1}. R_gs were also evaluated from the P(r) functions:

$$R_g^2 = \frac{\int_0^\infty P(r)r^2 dr}{\int_0^\infty P(r)dr} \qquad\qquad [3]$$

Data collected to higher angle on the quadrant detector were combined with data from equivalent samples collected on the linear detector by linearly scaling the two separate data sets to achieve the best fit in the range $0.008 < S < 0.012$ Å^{-1}. The P(r) function was then computed from the combined scattering curve with the GNOM program, using data over the angular range $0.0026 < S < 0.0280$ Å^{-1}. This function was used in the modeling studies.

In a monodisperse protein solution with a constant electron density contrast, the forward scattering intensity, I(0), is proportional to the square of the molecular weight (MW) of the protein molecule, $I(0) \propto C(MW)^2$, where C is protein concentration in molar concentration [M] (Glatter & Kratky, 1982). In a solution where a fraction (α) of protein molecules dimerize, the concentrations of the protein molecule in monomer (MW) and dimer (2MW) forms are $(1-\alpha)C$ and $(\alpha/2)C$ respectively. Hence $I(0) \propto \{(1-\alpha)C(MW)^2 + (\alpha/2)C(2MW)^2\}$ or $I(0) \propto (1+\alpha)MW^2$. For example, if 30% of protein monomers form dimers, a 30% increase in I(0) will be observed.

3.3.2. Computing Scattering Curves from Protein Coordinates

A scattering curve is computed from the atomic coordinates for the ATPase domain with the program CRYSOL (Svergun, *et al.*, 1995) since the structure of this domain has been solved. A 3 Å thick hydration layer, with a value of 0.03 e$^-$/Å^3 as the electron density contrast between this hydration layer and the bulk buffer, is included in the computing.

3.3.3. Computing Scattering Curves from Some Simple Triaxial Bodies

An initial and rudimentary approach to the shape analysis of the scatterer (protein molecule) is to model the molecule as some simple triaxial bodies. When the protein molecule consists of several domains or subunits, each domain can be modeled

individually, and the whole molecule is thus composed of several triaxial bodies in certain orientations. Since solution scattering is a low resolution technique, these rather crude models appear quite satisfactory to approximate the shapes of some protein molecules.

The scattering amplitude of some simple triaxial bodies, such as sphere, ellipsoid, cylinder and prism, can be computed analytically for any orientation (Glatter & Kratky, 1982). However, the spatial averaging, *i.e.*, the integration, has to be performed numerically. In the case of multibody modeling, the scattering amplitudes of individual domain are calculated analytically and added according to the particular orientation in the model. The scattering intensity is obtained by squaring the total amplitude and averaging over all orientations.

The following analysis is based on the model of prisms since rectangular prisms were used as rudimentary models for hsc70 and its subfragments in this study. For other triaxial bodies, the analysis is similar: only the scattering amplitude, F, needs to be replaced accordingly.

The scattering amplitude of a prism with edge lengths A, B, C is (Hiragi & Ihara, 1981):

$$F(A,B,C,S,\theta,\phi) = 8ABC\left[\frac{\sin(As_1\cos\phi)}{As_1\cos\phi}\right]\left[\frac{\sin(Bs_1\sin\phi)}{Bs_1\sin\phi}\right]\Psi_c \qquad [4]$$

where:

$s_1 = S\sin\theta$;

$s_2 = S\cos\theta$;

$\Psi_c = \sin(Cs_2)/(Cs_2)$;

θ and ϕ are the zenith and azimuth angles.

The scattering intensity of a prism is thus as follows:

$$I(S) = (V^2\bar{\rho}^2)\frac{\int_0^{2\pi}\int_0^{\pi}F(A,B,C,S,\theta,\phi)F^*(A,B,C,S,\theta,\phi)\sin\theta d\theta d\phi}{\int_0^{2\pi}\int_0^{\pi}\sin\theta d\theta d\phi}$$

$$= (V^2\bar{\rho}^2)\int_0^{2\pi}\int_0^{\pi}F^2(A,B,C,S,\theta,\phi)(\sin\theta/4\pi)d\theta d\phi \qquad [5]$$

$$= V^2\int_0^{2\pi}\int_0^{\pi}F^2(A,B,C,S,\theta,\phi)(\sin\theta/4\pi)d\theta d\phi$$

where V is the volume of the prism, $V = ABC$, and $\bar{\rho}$ is the average excess scattering density. For simplicity, $\bar{\rho}$ is taken as 1.

For the complex model of i prisms with scattering amplitude F_i's, the scattering intensity can be decomposed into two terms (Fujisawa, et al., 1990). In the first term, the calculation of the scattering intensity from each prism is performed independently in its local axes by equation [5], and the contribution from individual prism weighted by the excess scattering mass is added together. The second term concerns the interference between different prisms. The orientation of each prism can be viewed as a rotation of the local axes (θ_i, ϕ_i) with respect to the global axes (Θ, Φ) via three Euler angles (α_i, β_i, γ_i). Local axes are related to the global axes by:

$$\begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix} = T(\alpha_i,\beta_i,\gamma_i)\begin{pmatrix} \sin\Theta\cos\Phi \\ \sin\Theta\sin\Phi \\ \cos\Theta \end{pmatrix} \qquad [6]$$

where $T(\alpha_i, \beta_i, \gamma_i)$ is the orthogonal transformation matrix:

$$T(\alpha_i,\beta_i,\gamma_i) = \begin{pmatrix} \cos\alpha_i\cos\beta_i\cos\gamma_i & \sin\alpha_i\cos\beta_i\cos\gamma_i + \cos\alpha_i\sin\gamma_i & -\sin\beta_i + \cos\gamma_i \\ -\cos\alpha_i\cos\beta_i\sin\gamma_i & -\sin\alpha_i\cos\beta_i\sin\gamma_i + \cos\alpha_i\cos\gamma_i & \sin\beta_i\sin\gamma_i \\ \cos\alpha_i\sin\beta_i & \sin\alpha_i\sin\beta_i & \cos\beta_i \end{pmatrix}$$

$$[7]$$

The total scattering intensity from a complex model of i prisms can be written as:

60

$$I(S) = \sum_i (V_i^2 \bar{\rho}_i^2) \int_0^{2\pi} \int_0^{\pi} F_i^2(A,B,C,\theta,\varphi)(\sin\theta / 4\pi)d\theta d\varphi$$

$$+2\sum_i \sum_{j \neq i} (V_i\bar{\rho}_i)(V_j\bar{\rho}_j) \int_0^{2\pi} \int_0^{\pi} F_i(A,B,C,\theta,\varphi)F_j(A,B,C,\theta,\varphi)$$

$$\times \cos(2\pi S r_{ij} \sin\Theta \cos\Phi)(\sin\Theta / 4\pi)d\Theta d\Phi$$

[8]

where r_{ij} is the distance between prism i and j. Again, $\bar{\rho}_i$, $\bar{\rho}_j$ can be taken as 1 for simplicity.

In practice, the center of one of the prisms is taken as the origin of the global axes, and its local axes are lined up with the global axes.

The pair distribution (P(r)) function can be calculated analytically only for a few simple triaxial bodies (Glatter & Kratky, 1982). In this study, the P(r) functions were obtained from the computed scattering curves, either from equation [5] for the single body model or from equation [8] for the multi-body model, with indirect transformation method implemented in the GNOM program (Semenyuk & Svergun, 1991).

Radius of gyration can be derived from the computed scattering curves by the Guinier approximation, equation [1], or from the P(r) function, equation [3]. A direct calculation for a single body prism with edge lengths A, B, C is:

$$R_g^2 = \frac{A^2 + B^2 + C^2}{12}$$

[9]

R_g for the multi-body model is derived from parallel axis theorem:

$$R_{g,total}^2 = f_1 R_{g1}^2 + f_2 R_{g2}^2 + f_1 f_2 d^2$$

[10]

where f_1, f_2 are the mass fractions and R_{g1}, R_{g2} are the radii of gyration of prism 1 and prism 2, d is the distance between their centers of mass. Results from these three calculations should agree well.

3.4. Results

3.4.1. The Static Measurements

3.4.1.1. A Standard Globular Protein - BSA

Bovine serum albumin (BSA) is a frequently used globular protein with molecular weight (MW) of 66 kDa. BSA was chosen as a model protein initially for the purpose of testing the applicability of the quadrant detector in solution scattering, particularly the feasibility of merging data sets from the linear detector and the quadrant detector for computing the P(r) function. Because of the similarity in molecular weight, BSA is also a good reference in the size and shape analysis of the 60 kDa fragment and hsc70 proteins.

The scattering curves were recorded using both the linear and quadrant detectors. R_gs were calculated using the Guinier approximation, equation [1] applied to both linear and quadrant detector data. Fig. 3.2 shows the concentration dependence of R_g. The extrapolated value for R_g with the linear detector is 29.7 ($\pm$ 0.2) Å, which is in good agreement with 29.8 Å from literature (Anderegg, *et al.*, 1955).

The quadrant detector is known for improving the signal-to-noise ratio at high angular range which is necessary for computing the P(r) function. To test the reliability of the quadrant detector for solution scattering, as compared with the traditionally used linear detector, scattering data from BSA were also collected using a quadrant detector. Both the R_g at the individual protein concentration and the extrapolated value (30.0 $\pm$ 0.2 Å) using the quadrant detector are comparable to those using the linear detector (Fig. 3.2.a, b). Furthermore, the scattering curves with the quadrant detector superimpose satisfactorily with those using the linear detector (Fig. 3.3). Improved statistics with the quadrant detector is apparent from the figure in the wide angle region. However, in practice, the alignment of the quadrant detector to the direct beam is not trivial, and any slight imprecision in alignment may lead to distortions in the scattering profile predominantly occurring in the small angle region. Thus, by employing two detectors, it is reasonable to construct a complete model scattering curve where the low angle portion comes from data with the linear detector, and the wide angle part is taken from data using the quadrant detector. This combined scattering curve is then used to compute the P(r) function, equation [2].
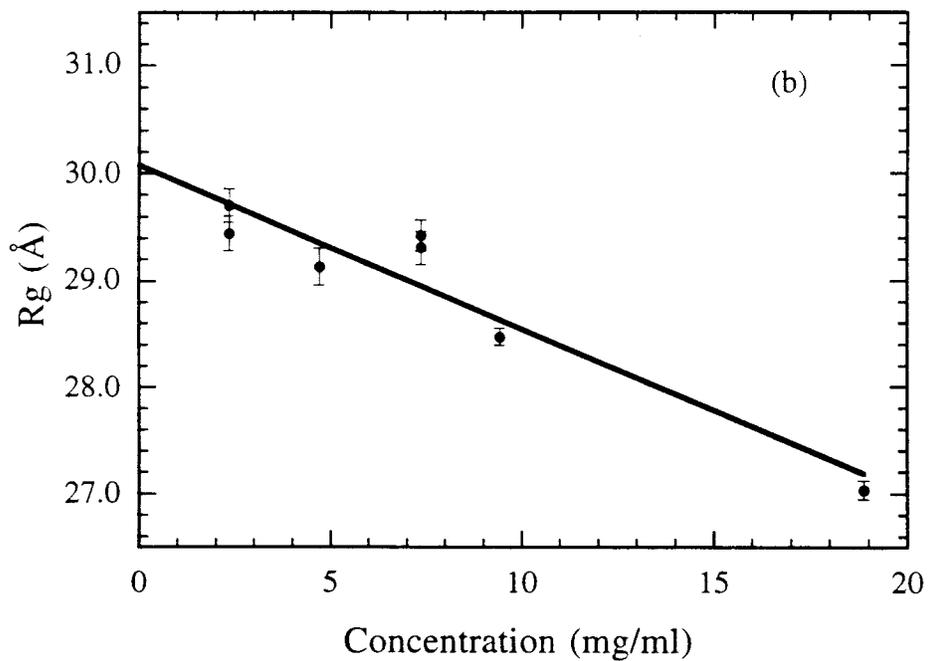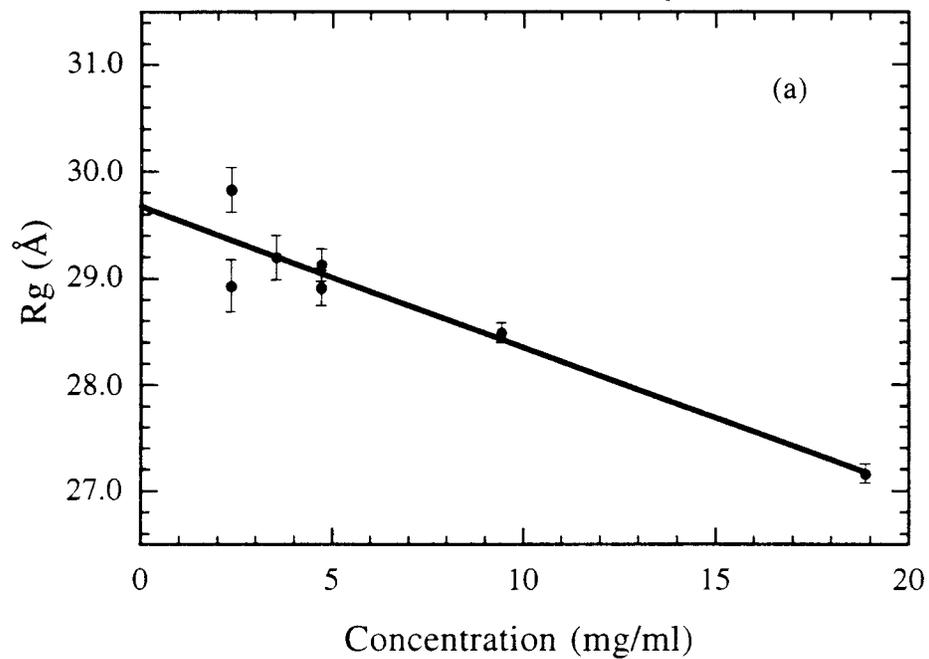
Figure 3.2. Concentration dependence of R_g and the extrapolation of R_g to zero protein concentration for BSA. a) with the linear detector; b) with the quadrant detector.
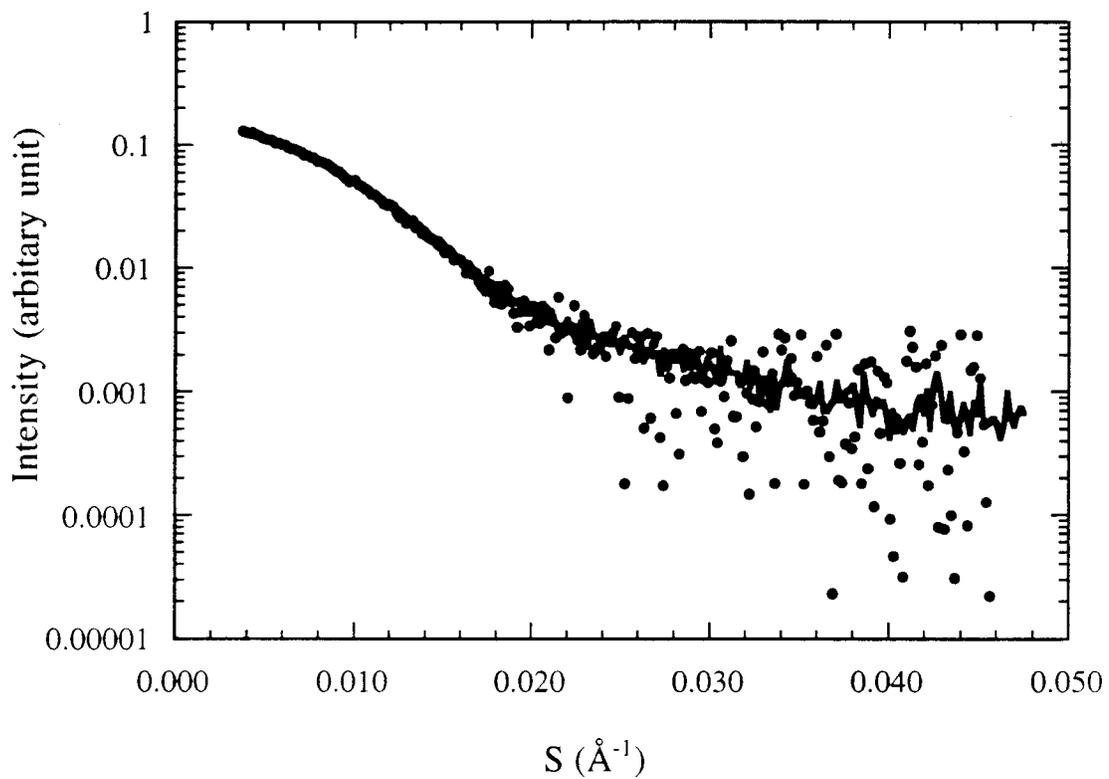
**Figure 3.3.** Representative scattering curves of BSA recorded with the linear detector (●) and the quadrant detector (—). The scattering curves are scaled to achieve the best fit in the range of $0.008 < S < 0.012$ Å^{-1}. The concentration for both measurements was 4.7 mg/ml.

A model scattering curve is produced by combining the linear detector data of 4.7 mg/ml sample with the quadrant detector data of 18.9 mg/ml sample. The merging range was around 0.017 Å^{-1} in S, beyond which the interparticle interference in the high concentration (18.9 mg/ml) sample was negligible, and could be neglected. The integrated scattering profile is shown in Fig. 3.4.a. Fig. 3.4.b shows the P(r) function calculated with the GNOM program (Semenyuk & Svergun, 1991) using this combined data set. R_g derived from the P(r) function [3] is 29.6 ($\pm$ 0.1) Å, in excellent agreement with the extrapolated R_g (29.7 $\pm$ 0.2 Å) from the Guinier approximation. The maximum length of the BSA molecule, D_{max}, is 95 ($\pm$ 5) Å, and the most populated distance, measured by the peak position, is 38 Å. The shape of the P(r) function is typical for a globular shape molecule.

3.4.1.2. The Guinier Analysis on hsc70 and Its Subfragements

Sample integrity. Fig. 3.5 shows the typical small angle x-ray scattering curves of the 60 kDa fragment. No significant nonspecific aggregation was observed for all the samples tested, as indicated by a lack of an upswing in the small angle region in the Guinier plot. The forwarded scattering intensity I(0), which is proportional to the protein molecular weight, was obtained using the Guinier approximation, equation [1]. None of the samples (hsc70, the 60 kDa fragment, the ATPase domain, the C-terminal domain, under various conditions tested) show significant concentration dependence of I(0), suggesting that protein molecules maintained the same aggregation state throughout the concentration range tested. The normalized I(0) values, I(0)/C, for the 60 kDa fragment are plotted in Fig. 3.6.

To evaluate the molecular weights of the samples, scattering data were also collected on lysozyme solutions during the measurements. Assuming the same partial specific volume for the hsc70 samples and lysozyme, molecular weights were estimated by comparing the ratios of I(0)/C of the hsc70 samples and that of lysozyme. The results are in reasonable agreement with calculations from amino acid residues, suggesting that the protein molecules were monomeric in solution. For example, MWs obtained from SAXS are ~72 kDa, ~54 kDa and ~ 40 kDa for hsc70, the 60 kDa fragment and the ATPase domain, while calculations from amino acid residues give 71.2 kDa, 61.0 kDa and 42.5 kDa respectively. Altogether, samples in our scattering studies were largely aggregation free, and in monomeric form.
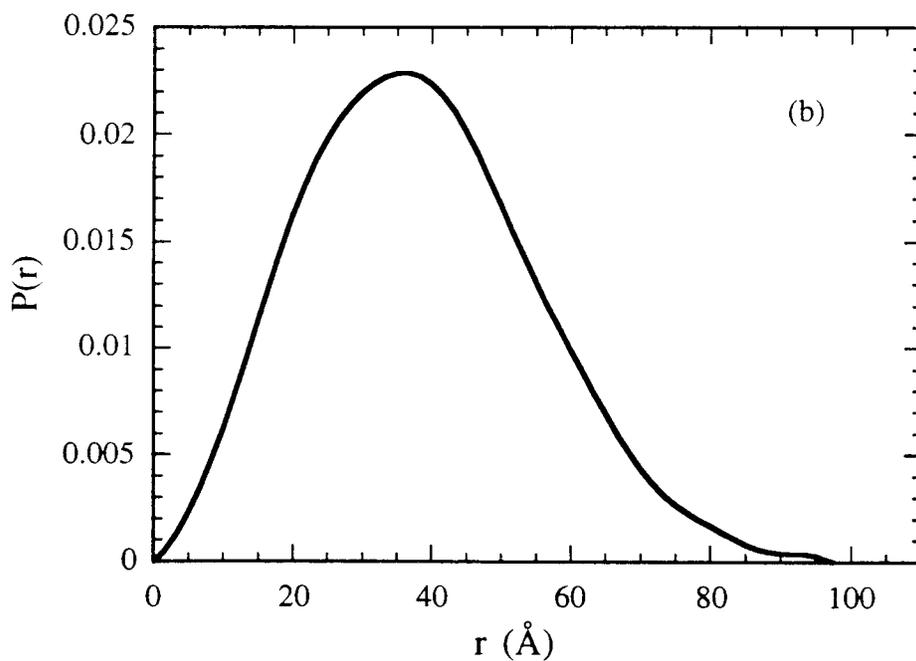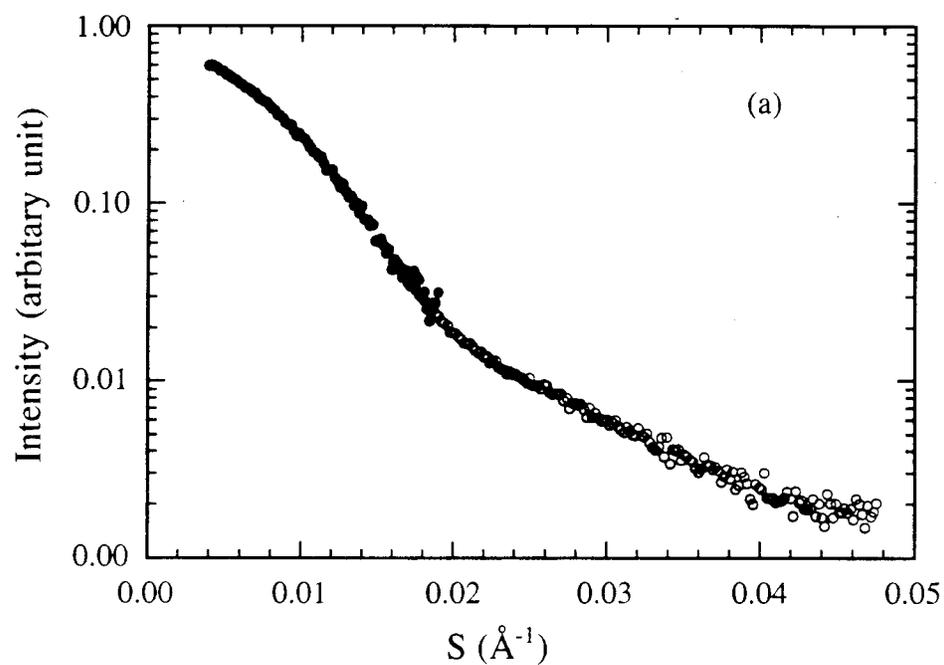
Figure 3.4. a) A combined scattering curve of BSA with the small angle range from the linear detector data (●), and the wide angle region from the quadrant detector (O). The linear detector data is from the 4.7 mg/ml data set, and the quadrant detector data is from the 18.8 mg/ml BSA data set. b) The P(r) function of this combined data.

66
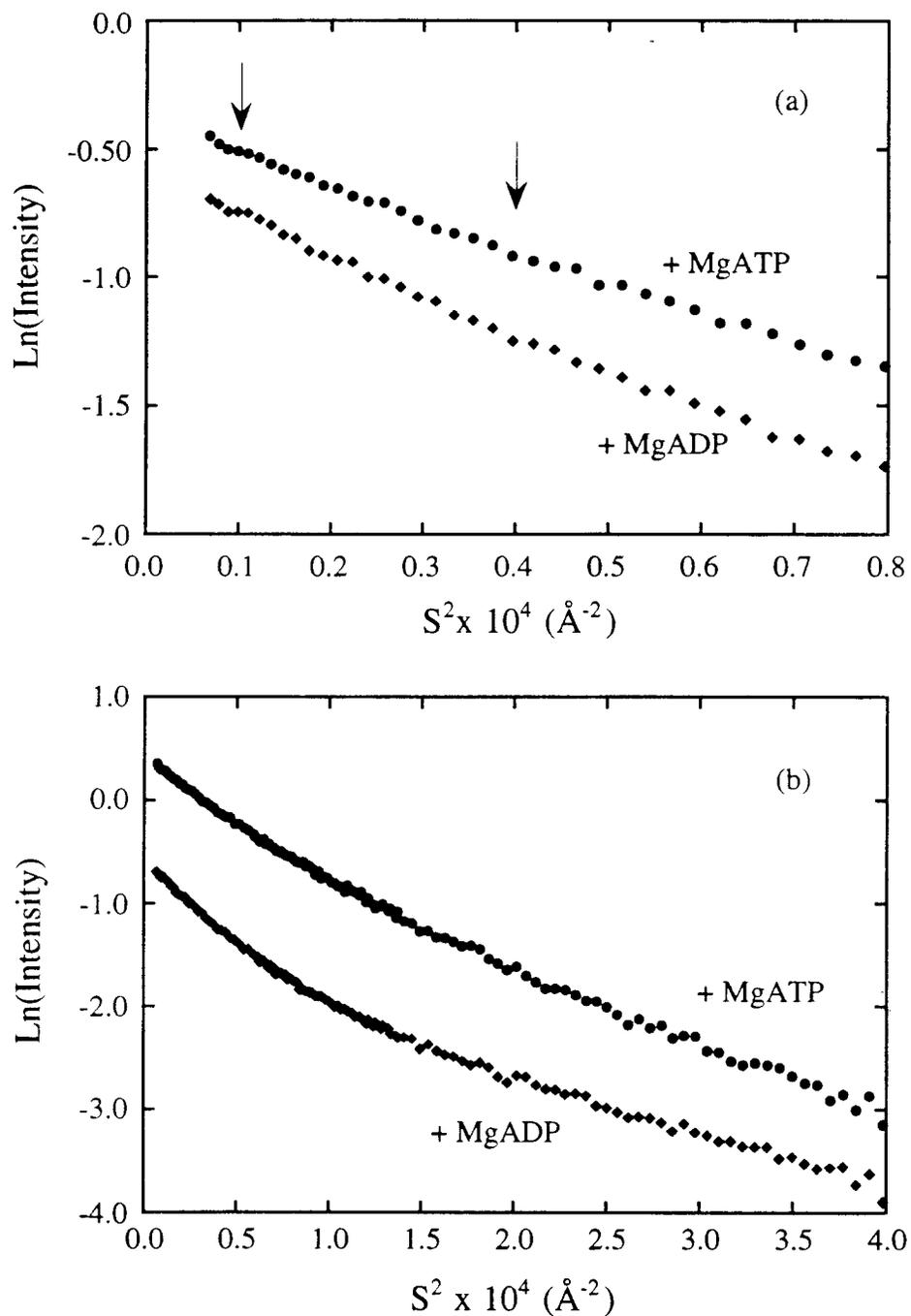
Figure 3.5. Representative scattering curves, in $\ln(I(S))$ versus S^2 for the 60 kDa fragment in the presence of MgATP (●) and MgADP (◆). (a) at low angle, and (b) an extended region. Arrows in panel a shows region used to compute R_g in the Guinier approximation. For clarity, the scattering curve collected with MgATP is shifted vertically from that collected with MgADP in both figures.
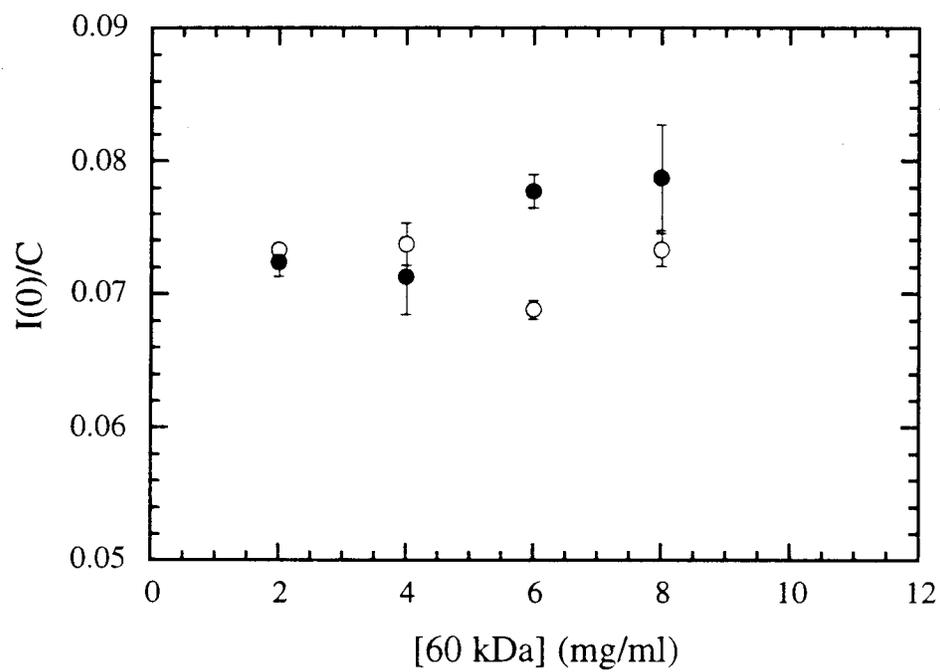
Figure 3.6. Forwarded scattering intensity normalized to protein concentration, I(0)/C, for the 60 kDa fragment in the presence of (●) MgATP and (O)MgADP.

Reproducibility of data. In order to assess the combined reproducibility of the sample preparation protocols and the experimental apparatus, complete datasets were taken on two separate runs for each of the protein constructs (Table 3.1). Pairs of R_g values determined on the same construct on different runs generally agreed within computed error (*i.e.* the standard deviation of the intercept value of a weighted linear fit to experimental values of R_g versus concentration); for example, the two values of R_g for the ATPase domain in the presence of MgADP differed by only 0.3 ± 0.3 Å. The greatest variation was observed with hsc70 in the presence of ADP, where differences between values from these two independent determinations of R_g were 2.6 ± 1.2 Å in the Guinier approximation, or 3.2 ± 0.7 Å by P(r) function analysis. This correlates with the fact that hsc70 shows the greatest propensity to aggregate in solution, and presumably reflects heterogeneity at the biochemical level, since variations of this magnitude did not occur with other samples. Because of this, an emphasis has been placed in the following discussion on using results from the constructs which are least variable in their behavior (ATPase and 60 kDa fragments) for quantitative modeling, and then extending these results to hsc70 as far as is supported by the data.

Steady-state radii of gyration. Fig. 3.7 shows the concentration dependence of R_g and the extrapolations to zero protein concentration. The extrapolation results, as well as results from P(r) function analysis (see Section 3.4.1.3), are summarized in Table 3.1.

Scattering data were collected on samples of hsc70 in the concentration range 1-4 mg/ml (14-57 μM) in the presence of 0.5 mM MgATP or MgADP. For MgATP, this concentration is well above the K_m of steady-state hydrolysis (~0.1 μM), and the nucleotide:protein ratio is such that less than 5% of the ATP will be hydrolyzed during the course of the measurements. For MgADP, this concentration is well above the K_d of the nucleotide (0.11 μM). In the presence of either MgATP or MgADP, the apparent R_g increases slightly with increasing protein concentration (Fig. 3.7.a), suggesting that a small amount of aggregation may be present at the higher protein concentrations.

The R_g measured in the presence of MgATP is substantially smaller than that measured in the presence of MgADP at all concentrations (Fig. 3.7.a and Table 3.1). Extrapolation of R_g to zero protein concentration gives values of ~36 Å in the presence of MgADP and ~33 Å in the presence of saturating amounts of MgATP, suggesting that MgATP induces a significant conformational change in hsc70, relative to its

69

Table 3.1. Steady-state structure parameters (Å) derived from solution scattering data for hsc70, the 60 kDa fragment, and the 44 kDa ATPase domain in the presence of MgADP and MgATP and the peptide-binding fragment.

protein	in the presence of 0.5 mM MgADP R_g Guinier	R_g P(r)	d_{max}	in the presence of 0.5 mM MgATP R_g Guinier	R_g P(r)	d_{max}	$R_g(ADP) - R_g(ATP)$ Guinier	P(r)
hsc70	34.9 ± 0.2^b	38.5 ± 0.5^b	130^d	32.8 ± 0.3^b	34.7 ± 0.8^b	120^d	2.1 ± 0.3^b	3.8 ± 0.9^b
	37.5 ± 1.0^c	41.7 ± 0.5^c		33.4 ± 0.2^c	34.5 ± 0.6^c		4.1 ± 1.0^c	7.2 ± 0.8^c
60 kDa	32.3 ± 0.1^b	34.0 ± 0.1^b	125^d	26.9 ± 0.3^b	28.7 ± 0.5^b	115^d	5.4 ± 1.0^b	5.3 ± 0.5^b
fragment	31.3 ± 0.5^c	34.5 ± 0.4^c		26.4 ± 0.9^c	29.6 ± 1.0^c		4.9 ± 1.0^c	4.9 ± 1.1^c
ATPase	22.8 ± 0.2^b	22.6 ± 0.2^b	70^d	22.6 ± 0.1^b	22.2 ± 0.2^b	70^d	0.2 ± 0.3^b	0.4 ± 0.2^b
domain	23.1 ± 0.2^e	22.9 ± 0.2^e						

	R_g Guinier	R_g P(r)	d_{max}
peptide-binding domain	18.3 ± 0.7^b	18.3 ± 1.0^b	$60 - 65^d$
	19.1 ± 0.3^e	19.8 ± 0.1^e	

[a] Values for Rg are extrapolations to zero protein concentration from values determined at several concentrations, either by Guinier analysis or from P(r) functions. [b] Data derived from March 1994 run. [c] Data derived from June 1994 run. [d] Derived from combined data of linear and quadrant detector, June 1994 run. [e] Data derived from March 1995 run.

Figure 3.7. The concentration dependence of steady-state radii of gyration computed by Guinier approximation for (a) Hsc70; (b) the 60 kDa fragment; (c) ATPase fragment; (d) the peptide binding domain. For a-c, ($\blacklozenge$), in the presence of MgADP; ($\bullet$), in the presence of MgATP; (O) in the presence of MgATP with Na$^+$ rather than K$^+$. Multiple measurements were taken at every concentration. Errors associated with individual measurement were derived from the least squares fitting of the data in the Guinier regions. The error bar shown at each concentration in the figures is chosen to be the larger value between the maximum of individual errors and the standard deviation from the mean value of different measurements. Solid and dashed lines represent the extrapolation to zero protein concentration in the presence of MgATP and MgADP, respectively.

(a)

(b)
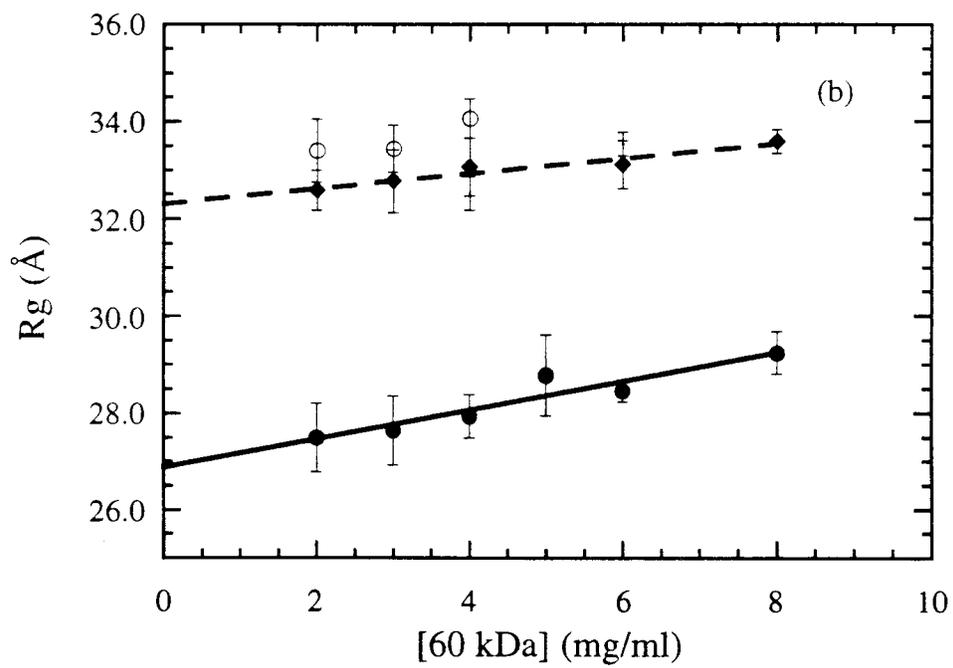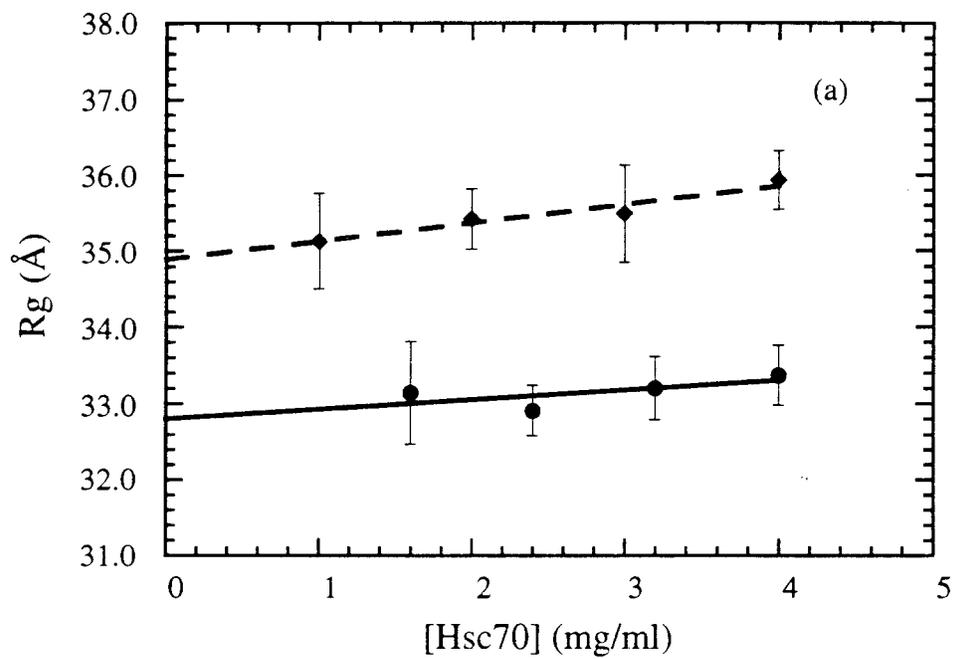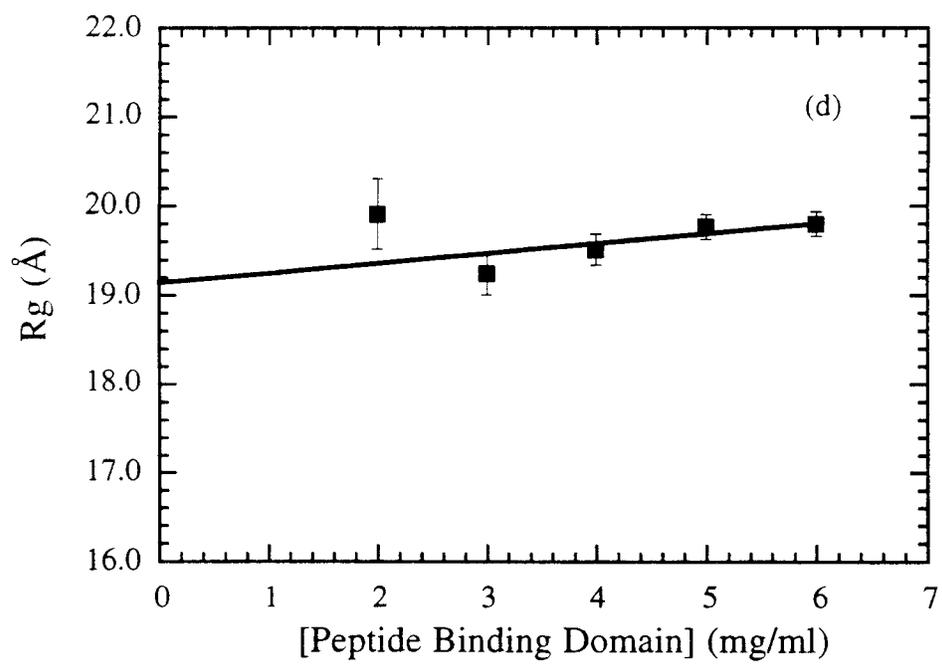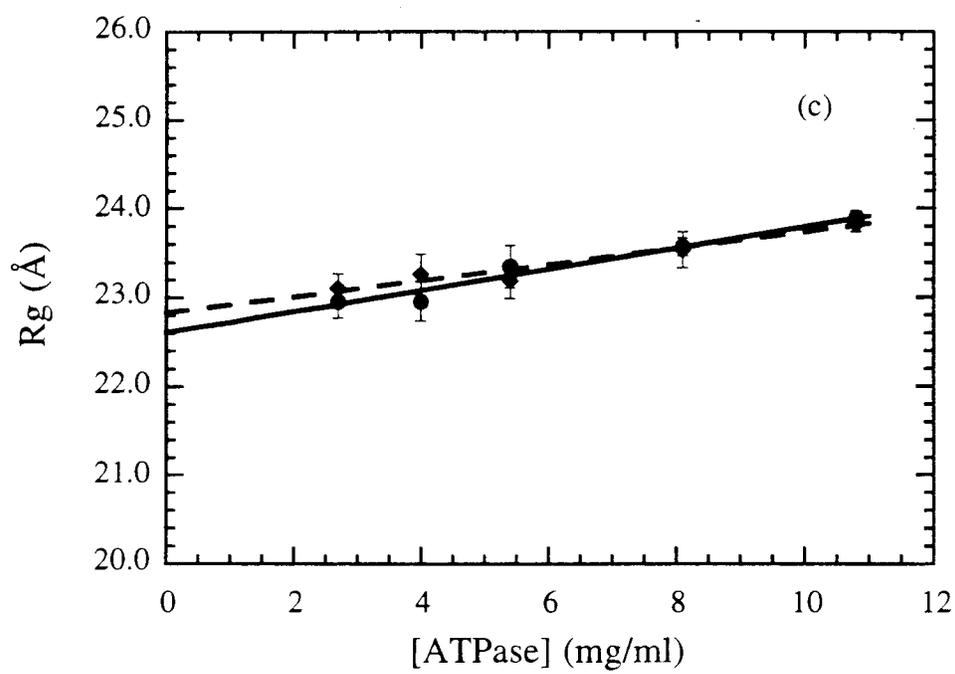
conformation in the presence of MgADP. The concentration-normalized values of $I(0)$ are the same in the presence of either nucleotide and are consistent with a molecular weight of ~ 70 kDa. Thus, under both conditions hsc70 is largely monomeric; the difference in R_g between that observed with MgATP and that observed with MgADP is not due to oligomerization.

Scattering curves were also recorded for hsc70 at 4 mg/ml in absence of nucleotide (Table 3.2). The computed R_g shows no statistically significant difference from that with MgADP present. Data were also collected in the presence of 1 mM MgAMPPNP, a slowly hydrolyzable analog of ATP. Again, the computed R_g is equal within experimental error to that observed with MgADP present. Only MgATP was able to induce the conformational change to the smaller R_g in hsc70.

To test whether the 60kDa fragment, which does not contain the ~ 100 amino acid residues at the carboxyl terminal of hsc70, undergoes a conformational change similar to that observed in hsc70, data were collected and analyzed in a similar manner from samples containing 1-8 mg/ml protein (Fig. 3.7.b). The MgATP-containing sample yields a substantially smaller R_g than the sample with MgADP (Table 3.1). Further, scattering curves were measured and R_gs were computed for several concentrations of 60 kDa fragment in the absence of nucleotide, in the presence of slowly-hydrolyzable ATP analogs AMPPNP and AMPPCP, and in the presence of MgATP but with Na^+ replacing K^+ in the sample buffer; representative values of R_g determined at 4.0 mg/ml protein concentration are summarized in Table 3.2. In all cases, the apparent R_g agreed within error to that observed with MgADP present. Thus, the transition to the smaller R_g requires MgATP (slowly hydrolyzable analogs can not substitute) and K^+ (Na^+ can not substitute). Since the 60 kDa fragment shows an ATP-induced conformational change similar to that of hsc70 and has less tendency to aggregate, many of the subsequent experiments were conducted on it in parallel with experiments on hsc70.

Both hsc70 and the 60 kDa fragment, particularly in the presence of MgADP and slowly hydrolyzable analogs and in the absence of nucleotides, do not appear to be globular in shape. Using 1.21 $Å^3$/dalton as partial specific volume, R_g of hsc70 and the 60 kDa fragment would be ~ 21 Å and ~ 20 Å if both were spherical. Even with a factor of 40% increase in volume to account for a hydration layer (Section 3.4.3.1) around the protein molecule in solution, R_g would only be increased to ~ 24 Å and ~ 22 Å for spherical hsc70 and the 60 kDa fragment. A more direct and sensible approach is to

Table 3.2. Representative R_g values (Å; computed in Guinier approximation) for hsc70, the 60 kDa fragment, and the 44 kDa ATPase fragment under different nucleotide conditions[a].

protein	nucleotides			no nucleotide	MgATP	
	MgADP	MgAMPPNP	MgAMPPCP		Na$^+$ buffer	K$^+$ buffer
hsc70	35.5 ± 0.6[b]	36.8 ± 0.3[b]	N.D.	37.4 ± 0.4[b]	N.D.	33.4 ± 0.4[b]
60 kDa fragment	33.1 ± 0.6[b]	32.5 ± 0.3[b]	32.7 ± 0.3[c]	32.6 ± 0.5[c]	34.1 ± 0.4[c]	27.9 ± 0.5[b]
ATPase domain	22.7 ± 0.2[b]	N.D.	N.D.		N.D.	22.5 ± 0.2[b]
	23.1 ± 0.1[d]			23.8 ± 0.1[d]		

[a] Values for hsc70 and the 60 kDa fragment were determined at a single protein concentration, usually 4 mg/ml, except for the 60 kDa fragment with MgAMPPNP and without nucleotides, where the concentration was 3 mg/ml. Values for R_g of the ATPase fragment are extrapolations to zero concentration. [b] Data derived from March 1994 run. [c] Data derived from June 1994 run. [d] Data derived from March 1995 run. Values not determined are labeled N.D.

compare R_g with a known globular-shaped protein molecule of a similar molecular weight. BSA has a MW of 66 kDa, between the 60 kDa fragment (MW ~60 kDa) and hsc70 (MW ~72 kDa). Compared with the R_g (29.6 ± 0.2 Å) of BSA (Section 3.4.1.1), the even smaller MW for the 60 kDa fragment appears to be rather elongated both in the presence of MgADP or slowly hydrolyzable analogs and in the absence of nucleotides, all having R_g of ~ 33 Å (Table 3.2). A similar conclusion is also reached for hsc70. More about shape analysis of the 60 kDa fragment and hsc70 will be discussed in Section 3.4.1.3.

To test whether the change in R_g, detected in both the 60 kDa fragment and hsc70, could be localized to the nucleotide binding domain of hsc70, data were collected and analyzed in the same way from samples containing 1-11 mg/ml ATPase domain (Fig. 3.7.c). Extrapolation of R_g to zero concentration yields values of 22.8 (± 0.2) Å in the presence of MgADP and 22.6 (± 0.2) Å in the presence of MgATP, revealing no detectable difference in conformation between the two states in solution. This is consistent with crystallographic results on wild-type and several mutant proteins which show the ATPase domain to have the same overall conformation in the presence of MgATP or MgADP (Flaherty, et al., 1995).

An increase of at least 0.7 ± 0.2 Å in R_g from the Guinier analysis was observed for the ATPase domain in the absence of nucleotides compared with those in the presence of MgADP (and MgATP) (Table 3.2). A somewhat large R_g in the absence of nucleotide suggests that a "hexokinase-like" opening of the nucleotide binding cleft may occur upon release of nucleotide (Bennett & Steitz, 1980; Fletterick, et al., 1975). An increment in the R_g of the 60 kDa fragment in absence of nucleotide was not observed; this is not unexpected, since (if the center-to-center distance of the ATPase and peptide binding domains remained essentially unchanged when nucleotide was released) only a 0.3 Å increase in R_g would be expected for this fragment for a conformational change that increased R_g of the ATPase fragment by 0.7 Å.

Scattering curves were also collected for the peptide binding domain (Fig. 3.7.d). Data from two different runs (June 1994 and March 1995) yielded R_g values ~19 Å.

### 3.4.1.3. The P(r) Function Analysis on hsc70 and Its Subfragments

Comparison with the Guinier analysis. A quadrant detector was employed in the data collection to improve statistics at wide angular range for P(r) function analysis. Radii of gyration were also evaluated from P(r) functions *via* equation [3], and are compared with those from the Guinier analysis in Table 3.1. For the smaller protein constructs (ATPase and peptide binding fragments), the maximum values of $2\pi SR_g$ used in the Guinier analysis ranged 1.1-1.3, and the R_g values computed (i) using the Guinier approximation and (ii) using the P(r) function agree within experimental error (Table 3.1). For the larger proteins (hsc70 and the 60 kDa fragment), R_gs computed from P(r) functions were systematically ~2 Å larger than those computed using the Guinier analysis, and maximum values of $2\pi SR_g$ used in the Guinier analysis ranged 1.1-1.5. For elongated molecules, such as the 60 kDa fragment and hsc70 in this study, the constraint on the range of S in the Guinier fitting ($2\pi SR_g \ll 1.0$) becomes more stringent . The disagreement between R_gs calculated by the two methods may be due to the use of a relatively generous Guinier region for the larger proteins, which was necessitated by the fact that restricting the analysis to smaller scattering angles reduced the number of data points included in the linear fit to fewer than 12 in some cases, at which point the errors in the detector response correction of individual channels significantly affected the calculated slope of the linear fit used for the Guinier analysis.

As discussed in Section 3.4.1.2, the largest discrepancy in data reproducibility was observed with hsc70 in the presence of MgADP, and the variation can be correlated with the fact that hsc70 shows greatest tendency to aggregate in solution (more about this will be discussed in Section 3.5.). Allowing for the variability of results on hsc70 in the presence of ADP, values of the change of R_g (ΔR_g) between different nucleotide (*e.g.* with MgATP versus with MgADP) conditions computed by the two different analyses (Guinier and P(r) analysis) agree within statistical error. Consequently, it appears that the ΔR_g values determined from the Guinier analyses of small angle data are as reliable as those computed by incorporating higher angle data into the P(r) function. In this study, Guinier values of R_g are used to parameterize the magnitudes and kinetics of nucleotide-induced conformational changes, as revealed by changes in R_g. The validity of using the P(r) function analysis of scattering data to higher angle has been established in the study of BSA (Section 3.4.1.1) where consistent R_gs derived from the Guinier approximation and from P(r) analysis are obtained. In addition, comparisons are made below (Section 3.4.3) between experimental P(r) functions for the ATPase fragment and those computed

from crystallographic coordinates; the agreement between the measured and computed distributions further supports the assertion that values of R_g derived from the P(r) function is accurate.

Calculation of P(r) functions. To acquire more information on the overall shape of hsc70 and its subfragments, we collected data from each construct at 4 mg/ml with both the linear detector (better resolution at low angle) and the quadrant detector (greater signal at higher angle) and used the combined scattering curves to calculate P(r) functions. (The validity of combining data from linear detector and quadrant detector for P(r) function calculation has been discussed in Section 3.4.1.1) In addition, P(r) functions were calculated from data collected solely on the linear detector, including data outside the Guinier region, for samples at concentrations other than 4mg/ml. P(r) functions computed using data from only the linear detector showed no significant differences, but associated with larger error bars, from those computed using combined data to higher angle from the linear and quadrant detectors.

The P(r) functions for the 60 kDa fragment and hsc70 (Fig. 3.8) share several features which are consistent with elongated proteins that undergo substantial conformational changes. First, the d_{max} values (120-130 Å for hsc70; 115-125 Å for the 60 kDa fragment) indicate that the molecules are elongated. As a comparison, BSA, a globular shape protein with similar MW, has a smaller d_{max}, ~95 Å. Second, the P(r) functions for both proteins in the presence of MgADP show distinct shoulders at 70-80 Å, suggestive of a bilobal shape. Again, the shape of the P(r) functions contrasts with a symmetric singular maximum P(r) function of a globular shape BSA (Fig. 3.4.b). Third, the P(r) functions for the proteins in the presence of MgATP reflect a substantial contraction relative to that for MgADP; the shoulders are much diminished and the d_{max} values are substantially smaller (Table 3.1).

The P(r) functions were also computed for the ATPase domain and the peptide-binding domain (Fig. 3.9), and the derived parameters are listed in Table 3.1. More about the P(r) functions of these two constructs will be discussed in the modeling section, Section 3.4.3.
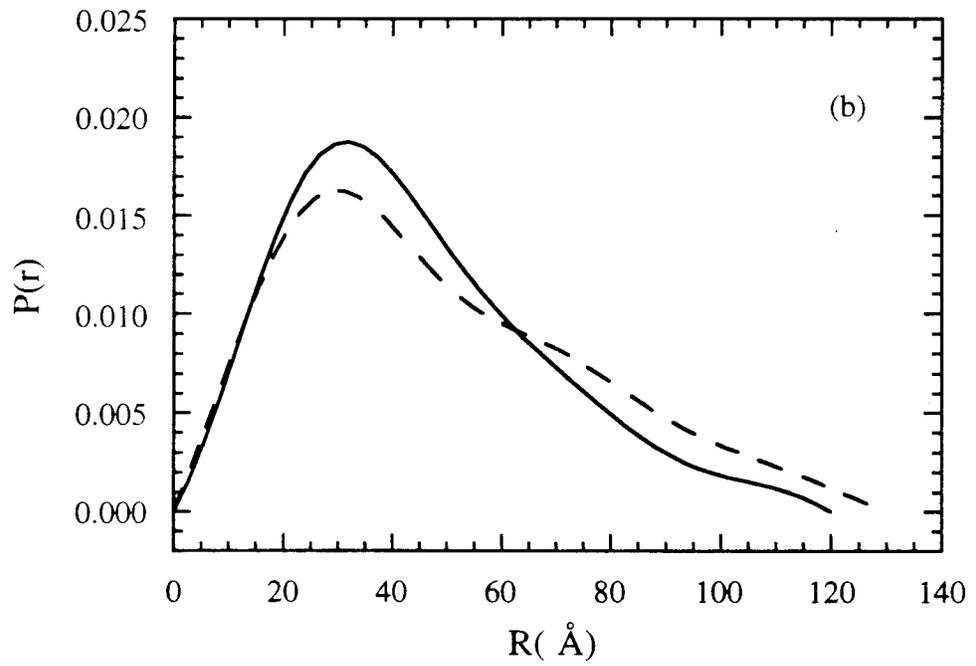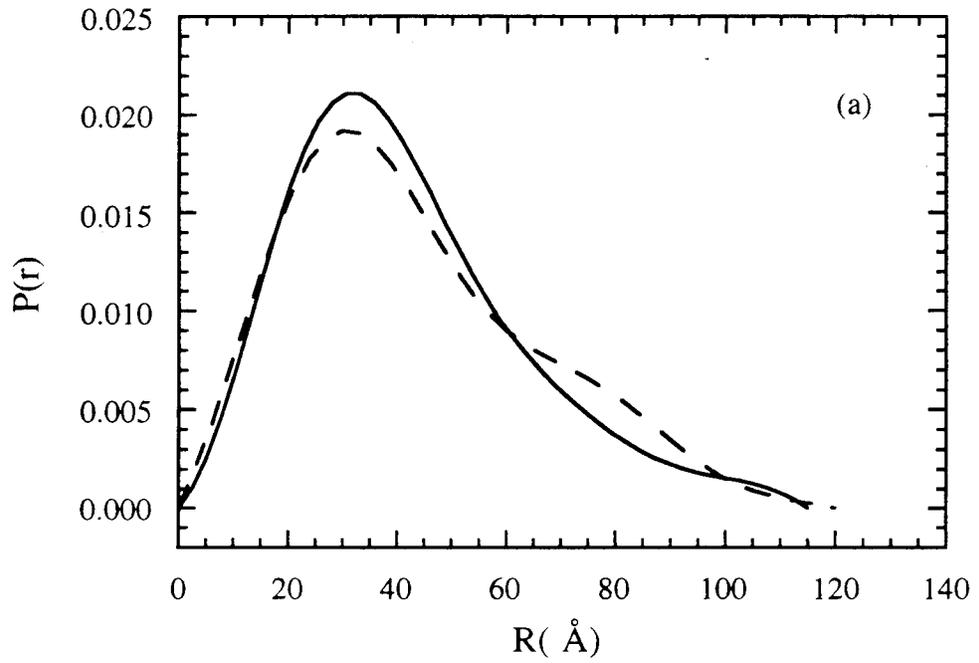
Figure 3.8. Distance distribution functions for (a) the 60kDa fragment; (b) the full-length hsc70, in the presence of MgATP (—), and in the presence of MgADP (- - -).
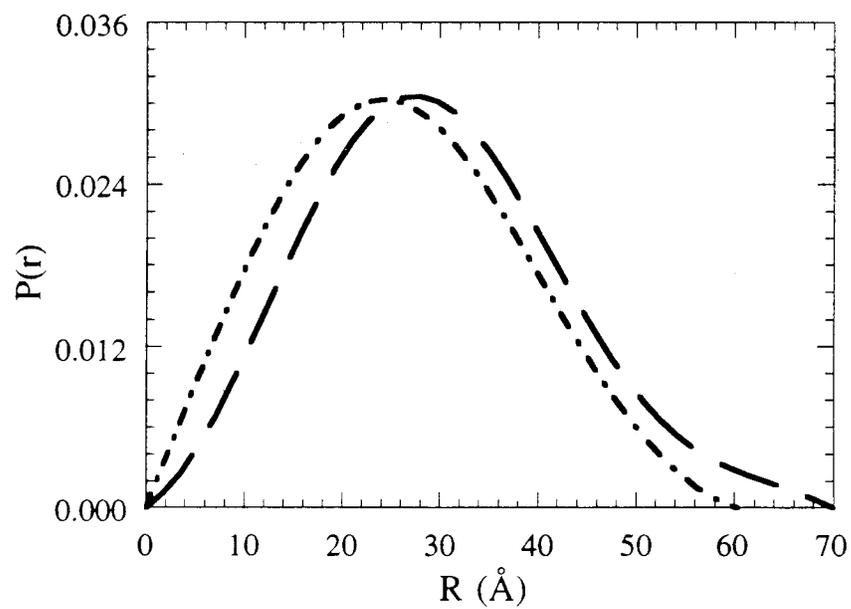
Figure 3.9. Distance distribution functions for the ATPase domain (———), and the peptide binding domain (— · —).

3.4.2. Kinetic Measurements

Kinetics of the ATPase cycle of the 60 kDa fragment. The rate of ATP hydrolysis was measured for the 60 kDa fragment under single turnover conditions, using experimental protocols and data analysis methods described previously (Ha & McKay, 1994). Fit of a single exponential to data for the fraction of ATP hydrolyzed versus time yielded a rate of 0.0017 ± 0.0001 sec^{-1} ($t_{1/2}$=7 min., Fig. 3.10.a). When protein is mixed with stoichiometric amounts of γ-[^{32}P]-ATP and the release of inorganic phosphate after binding and hydrolysis of the nucleotide was followed, it did not differ significantly from the rate of hydrolysis, demonstrating that phosphate release occurs rapidly after, and is significantly faster than, hydrolysis. The rate of dissociation of ADP from complexes of the 60 kDa fragment pre-incubated with α-^{32}P-ATP for a sufficiently long time to allow complete hydrolysis was measured to be > 0.008 sec^{-1}. The steady state ATPase rate was measured to be 0.0024 ± 0.0002 sec^{-1}. These data show that under the conditions of our experiments (T = 25 °C), ATP hydrolysis is the rate-limiting step of the ATPase cycle for the 60 kDa fragment. The impossibly low hydrolysis rate (in light of the steady state turnover rate) probably reflects experimental error and gives an estimate of the size of that error. The kinetic study was performed by Dr. Sigurd Wilbanks at Prof. McKay's laboratory of Department of Structural Biology, Stanford Medical School.

Kinetics of the change in R_g. To determine the time course of the change in R_g, data were collected for both hsc70 (March 1995) and the 60 kDa fragment (June 1994) after manual mixing of nucleotide-free protein with sub-stoichiometric amounts of MgATP (80-90%). Because radiation damage precluded data collection for longer than 10 minutes (for the 60 kDa fragment, June 1994) or 20 minutes (for hsc70, March 1995) from a single sample, longer time courses were assembled by transferring an aliquot from a master reaction to the sample cell every 12 minutes (the 60 kDa fragment) or 22 minutes (hsc70). The temperature of both cell and master reaction were maintained at 25°C. R_g values were computed for each one-minute or two-minute accumulation of data (Fig. 3.10). For both proteins, the minimum value of R_g had been reached at the earliest data collection time (~1.5 minutes). The minimum R_g is equal to the steady state value measured in the presence of excess MgATP. The observed R_g returned to the larger value, equal to (or larger than) that observed with excess MgADP in steady state measurements over the period of minutes.
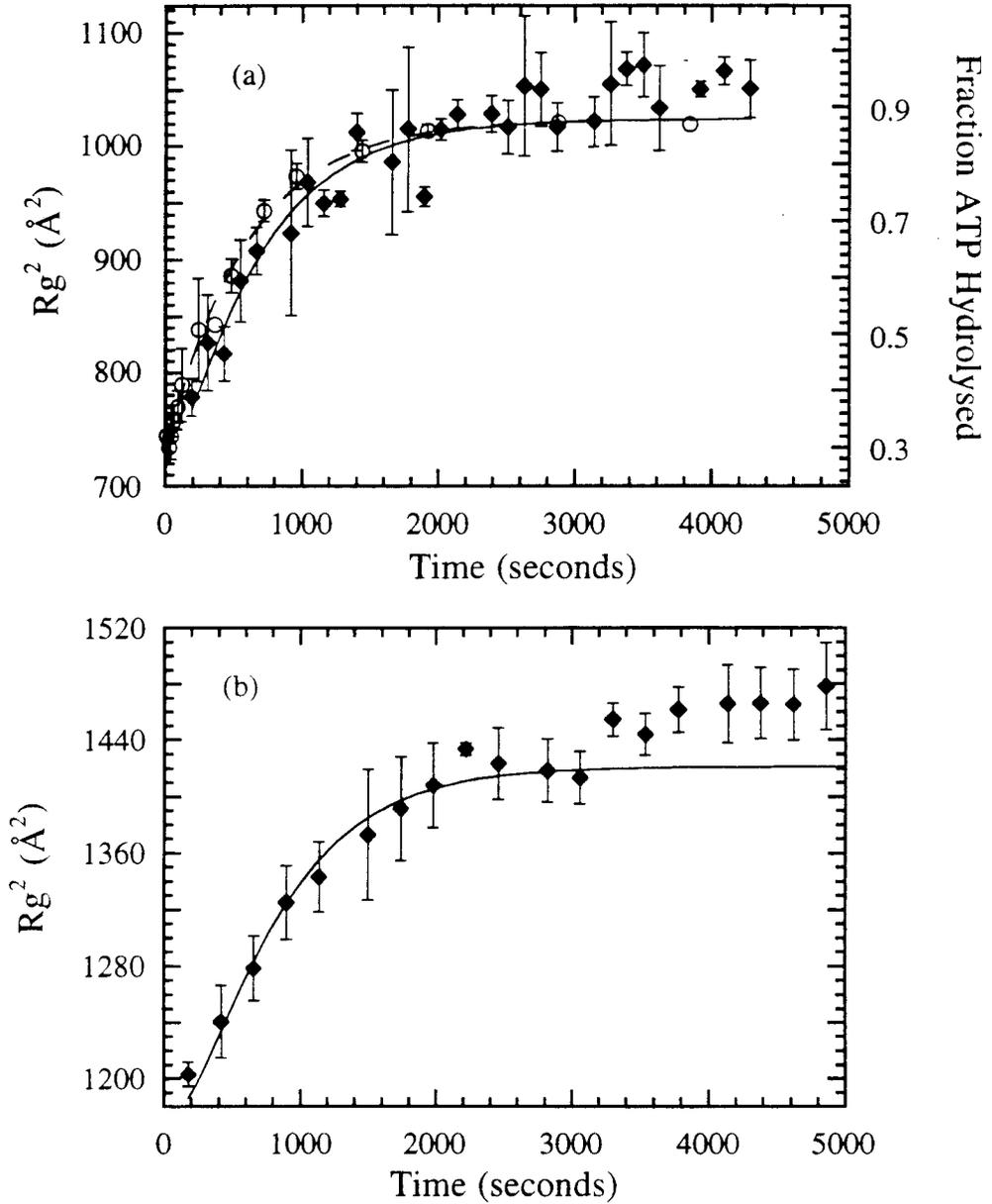
Figure 3.10. Time dependence of R_g^2 after mixing 80µM protein, 65-75µM ATP at 25 °C. R_g^2, computed in Guinier approximation, is shown as (◆); each time point represents the average of three separate experiments. (a) 60kDa fragment. Solid curve shows result of double exponential fit to the data. ATP hydrolysis under single-turnover conditions is shown in (O) with a single exponential fit (dashed curve). (b) Hsc70. Solid curve shows results of double exponential fit to the data, with initial and final values of the curve were constrained to be equal to steady state values of R_g in presence of ATP and ADP respectively, and with first rate constant equal to that of ATP hydrolysis.

The values of R_g^2 as a function of time were analyzed as a linear combination of the steady-state values of R_g^2 in the presence of MgADP and the presence of MgATP. For hsc70, estimating the rate of return to the larger R_g^2 by fitting a single exponential to the data gave a rate constant of 0.0012 ± 0.0001 sec^{-1} ($t_{1/2} = 9.6$ min.). Under the conditions of these experiments (T = 25 °C), the rate constant for ATP hydrolysis is 0.0030 sec^{-1} ($t_{1/2} = 3.8$ min.); the relaxation to the large R_g^2 is significantly slower than ATP hydrolysis. Therefore, the return of R_g^2 to the larger value was simulated as a two-step process, with the first step having the rate constant (k_{hyd}) of hydrolysis:

$$R_g^2(\text{apparent}) = R_g^2(\text{steady} - \text{stateADP}) - \Delta(R_g^2)\frac{k_{hyd}e^{-k_\beta t} - k_\beta e^{-k_{hyd}t}}{k_{hyd} - k_\beta}$$

where R_g^2 (steady state ADP) and ΔR_g^2 determined by steady state measurements under conditions identical (except for nucleotide concentration) to those of the kinetic experiment (Fig. 3.10). This yields a value 0.0019 ± 0.0003 ($t_{1/2} = 6.1$ min.) for the second rate constant, k_β.

Product release from hsc70 is sequential (Ha & McKay, 1994); P_i dissociates first with $k_{off}(P_i)$ of 0.0038 sec^{-1} ($t_{1/2} = 3.0$ min.); ADP then dissociates more rapidly with $k_{off}(ADP)$ of 0.029 sec^{-1} ($t_{1/2} = 0.4$ min.). The value of k_β determined here is significantly slower than the rates of product release. The possibility that this difference may be due to prep-to-prep variation in the protein cannot be ruled out; nor can it be ruled out that the difference is significant. Therefore, the return to the large R_g form of hsc70 after binding of MgATP may coincide with, or may be a relaxation process predicated on, product release; it does not coincide with, and is too slow to be correlated with, ATP hydrolysis.

This conclusion is corroborated by the kinetics of the change of R_g of the 60 kDa fragment. The return to a large R_g^2 could be fit as an exponential decay with rate constant of 0.0009 ± 0.0002 sec^{-1} ($t_{1/2}=13$ min.), to be compared with a rate of hydrolysis of 0.0017 sec^{-1} (Fig. 3.10.a). If a two-step process is used to model the data, using the ATP hydrolysis rate constant for the first step, then the second step has a rate constant 0.0070 ± 0.0020 sec^{-1} ($t_{1/2}= 1.6$ min.). For the 60 kDa fragment, release of both P_i and ADP are rapid compared to hydrolysis, so that it is not possible to discriminate between

hydrolysis and product release as concisely as with hsc70. However, it is clear that when a single ATPase cycle is initiated synchronously, the time required for the return to the larger R_g is not faster than the time required for hydrolysis and product release in the 60 kDa fragment.

At times longer than 50 minutes in the kinetic experiments, hsc70 reaches an R_g larger than that measured at steady state. It is not known whether this is due to aggregation of the sample, adoption of a third, even more extended conformation or some other phenomenon.

3.4.3. Modeling Based on P(r) Functions

3.4.3.1. Modeling the ATPase and the Peptide-binding Domains

P(r) functions were used to construct a rudimentary model for hsc70 and its subfragments. Initially, correspondence was established between the P(r) function derived from scattering data and P(r) functions computed from crystallographic coordinates of the ATPase fragment. The crystallographic model of the ATPase fragment included amino acid residues 4-381, one ADP molecule, one P_i, one Mg^{2+} ion, two K^+ ions, one Cl^- ion, and in some calculations, the 205 H_2O molecules that are within 3.0 Å of protein, nucleotide, or cofactor atoms. The value of R_g calculated from these coordinates, with or without the 205 water molecules, was 21.0 Å; the P(r) function computed from the coordinates showed significant disagreement with the function derived from experimental scattering data. However, addition of an appropriate layer of hydration (an approximately 3Å shell with an electron density contrast of 0.03 electrons/Å^3 greater than that of solvent) to the crystallographic model, using the method of Svergun as implemented in the CRYSOL program package (Svergun, et al., 1995), yielded good agreement between both R_g (22.5 Å) and the overall P(r) function (Fig. 3.11.a). We obtained a similar agreement between experimental and calculated P(r) functions only when a layer of hydration was added to the crystallographic structure of the molecule for a second test case, lysozyme, using data collected during these experiments (see Chapter 4).

A rectangular prism (or box) was chosen as a simple (described by three parameters) and convenient geometric shape for modeling the ATPase domain. The R_g of a box calculated with equation [9] was constrained to equal to the experimental value

84
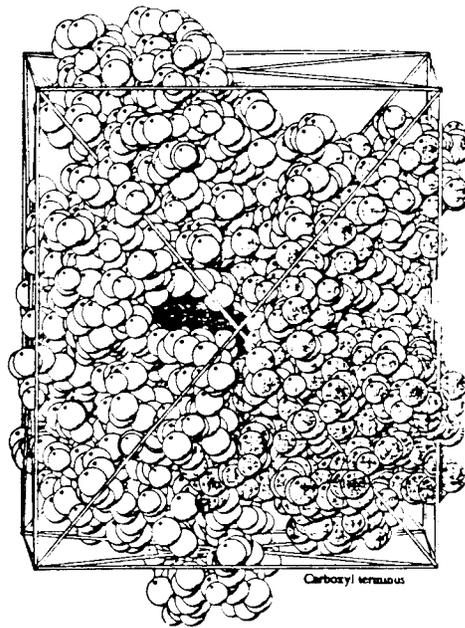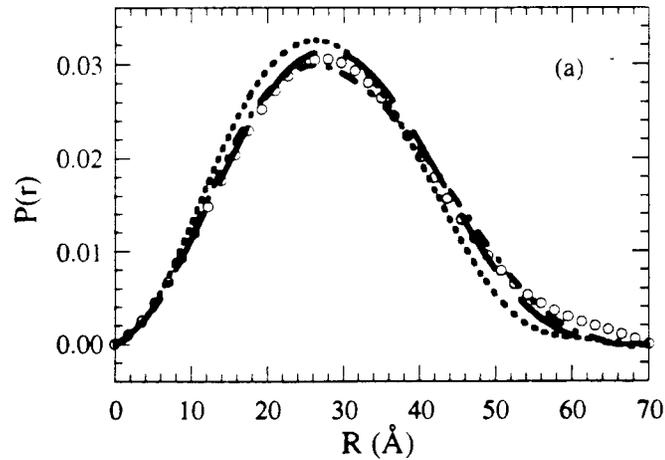
Figure 3.11. (a) Distance distribution functions for ATPase domain; (o), experimentally derived from 4mg/ml protein in the presence of MgADP; (---), computed from crystal coordinates; (— -), computed from the crystal coordinates with a hydration layer added; (— —), calculated from the box model. The experimentally derived P(r) functions in the presence of MgATP is practically superimposable with the experimental curve shown. (b) Superposition of the 56x47x26 Å box model of ATPase on a space filling model of the ATPase domain, drawn with Molscript (Kraulis, 1991). Residues 4-182 and 363-381 are shown in light gray, residues 183-362 in white and the bound reaction products in dark gray.

of R_g. The volume was set equal to the 1.4 x (protein molecular weight) x 1.21 $Å^3$/dalton (average protein specific volume). The use of a volume 40% larger than that computed from the molecular weight results from the increase in volume upon addition of the 3 Å hydration shell necessary to bring the R_g and P(r) functions calculated from "dry" atomic coordinates of the ATPase domain into agreement with SAXS data. Using these two constraints (R_g and volume), dimensions for a series of boxes were generated by allowing c to vary from 52 to 58 Å, varying the ratio of a:b from 1.0 to 2.5 in steps of 0.25. Scattering curves were generated, according to equation [5], for those rectangular prisms with the correct R_g and approximate d_{max} of the experimental value, and the P(r) functions were calculated with the GNOM program (Semenyuk & Svergun, 1991). The resulting P(r) functions were compared visually to the experimentally derived function in order to select an optimal set of dimensions. Typically, variations of ~10% in box dimensions, subject to the constraints described above, showed little apparent difference in agreement between calculated and experimental P(r) functions; larger variations showed clearly discernible differences.

The best dimensions found to model the ATPase domain in the presence of nucleotide were 56x47x26 $Å^3$. Superposition of such a box on the crystallographic structure encloses more than 90% of the protein atoms (Fig. 3.11.b). The P(r) function computed for these dimensions shows reasonable agreement with the experimental curve and with that computed from hydrated atomic coordinates (Fig. 3.11.a).

The same method was used to determine dimensions of a box to represent the peptide binding domain. The best values found for the peptide binding domain were 50x45x13 $Å^3$ (Fig. 3.12). These values support the qualitative judgment drawn from the shape of the P(r) function that the peptide binding domain is oblate rather than prolate.

3.4.3.2. Modeling the 60 kDa Fragment

Two different schemes were used to model the 60 kDa fragment. First, it was modeled as a monomeric protein in both the ATP-bound and ADP-bound forms by combining the ATPase and peptide binding domain models. This reduced the problem to determining (1) the separation between the two domains and (2) their relative orientations. The parallel axis theorem, equation [10], was used to derive an expression that determines the separation of the centers of mass of the ATPase and peptide binding domains, given the R_g values of both domains and the 60 kDa fragment. This analysis
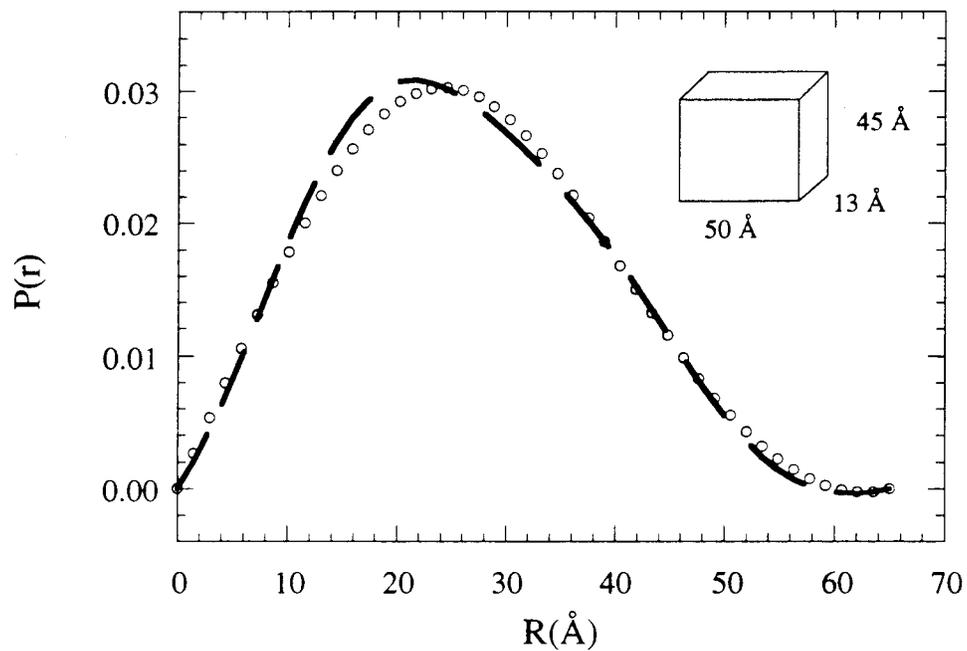
Figure 3.12. Distance distribution functions for the peptide binding domain; (o), experimentally derived from 4mg/ml protein; (— —), calculated from the 50x45x13 Å box model shown in insert.

yields a separation of 58 Å for the ADP-bound form and 44 Å for the ATP-bound form of the 60 kDa fragment.

To assess which relative orientations of the two domains were consistent with our data, a selection of orthogonal relative orientations was used with separations (d) of 42, 44, 46 Å and 56, 58, 60 Å to compute the scattering by equation [8], then calculate $P(r)$ functions. Contact or interpenetration between the two domains is not required in this analysis. Four orientations were examined, "T-shaped", "L-shaped", "face-to-face", and "elongated". Because of the unequivalent dimensions of the prisms, there were several domain placements for each orientation. Representative $P(r)$ function of each orientation is shown in Fig. 3.13. The $P(r)$ function for the ADP-bound form was particularly sensitive to alternative orientations of the two domains. In the case of the ATP-bound form, with a short separation of ~ 44 Å the two domains readily penetrated each other; the $P(r)$ functions appeared indistinguishable among different orientations if the two prisms overlapped. Thus, models were justified mainly based on the comparisons with the $P(r)$ function of the ADP-bound form, and were confirmed from the results of the ATP-bound state. Reasonable agreement was obtained with the "elongated" models where the short axes were parallel to each other and were perpendicular to the line through the centers of mass of the two domains (side-to-side, end-to-end or end-to-side models). In contrast, any models having the shortest axis of one domain parallel to the line through the centers of mass of the two domains ("T-shaped", "L-shaped" and face-to-face models) did not agree well. The ATP-induced conformational change was then modeled as a contraction between the domains.

In the second scheme, scattering data for the ADP-bound form of the 60 kDa fragment was modeled as resulting from aggregation of a molecule that retained the overall structure and dimensions of the ATP-bound form. Hypothetical scattering curves for a mixed population of monomers and dimers were computed, with the dimers formed by placing the monomers face-to-face, side-by-side, or end-to-end. It was found that when approximately 30% of the monomers formed end-to-end dimers, the $P(r)$ function calculated from the hypothetical curve also showed good agreement with the experimentally derived $P(r)$ function (Fig. 3.14). The R_g value calculated from the resultant model $P(r)$ function was 34.7 Å, in good agreement with the experimental value of 34.3 Å. However, the R_g value calculated from the Guinier analysis of the model scattering curves was significantly higher, 37.8 Å, reflecting a lack of agreement at low
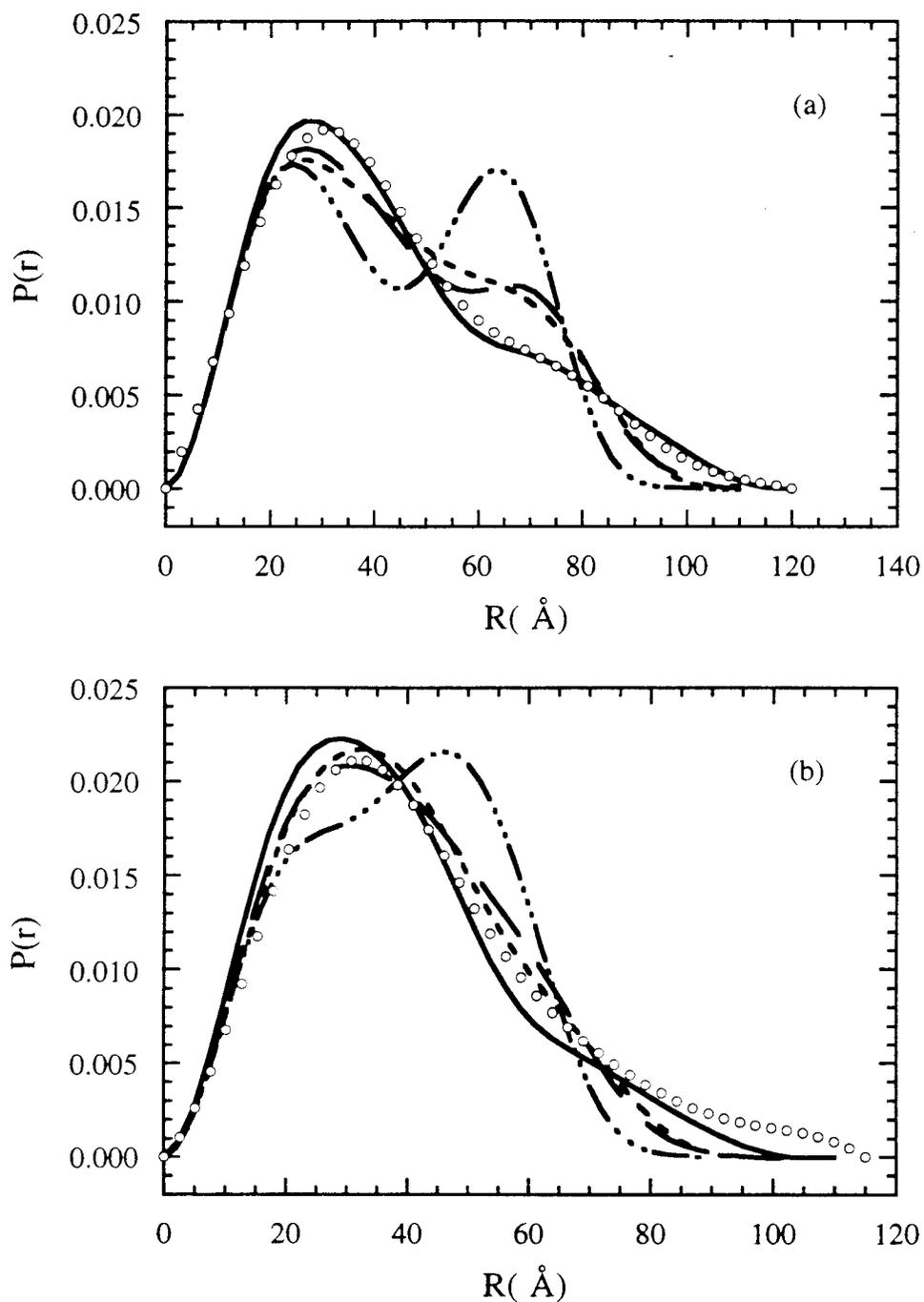
Figure 3.13. Distance distribution functions of the monomeric box models for the 60 kDa fragment (a) in the presence of MgADP with a separation of 58 Å; (b) in the presence of MgATP with a separation of 44 Å. (—), "elongated" model; (— —), "L-shaped" model; (- - -), "T-shaped" model; (— - - -), "paralelled" model. The experimentally derived P(r) functoins are shown in (O) for comparison.
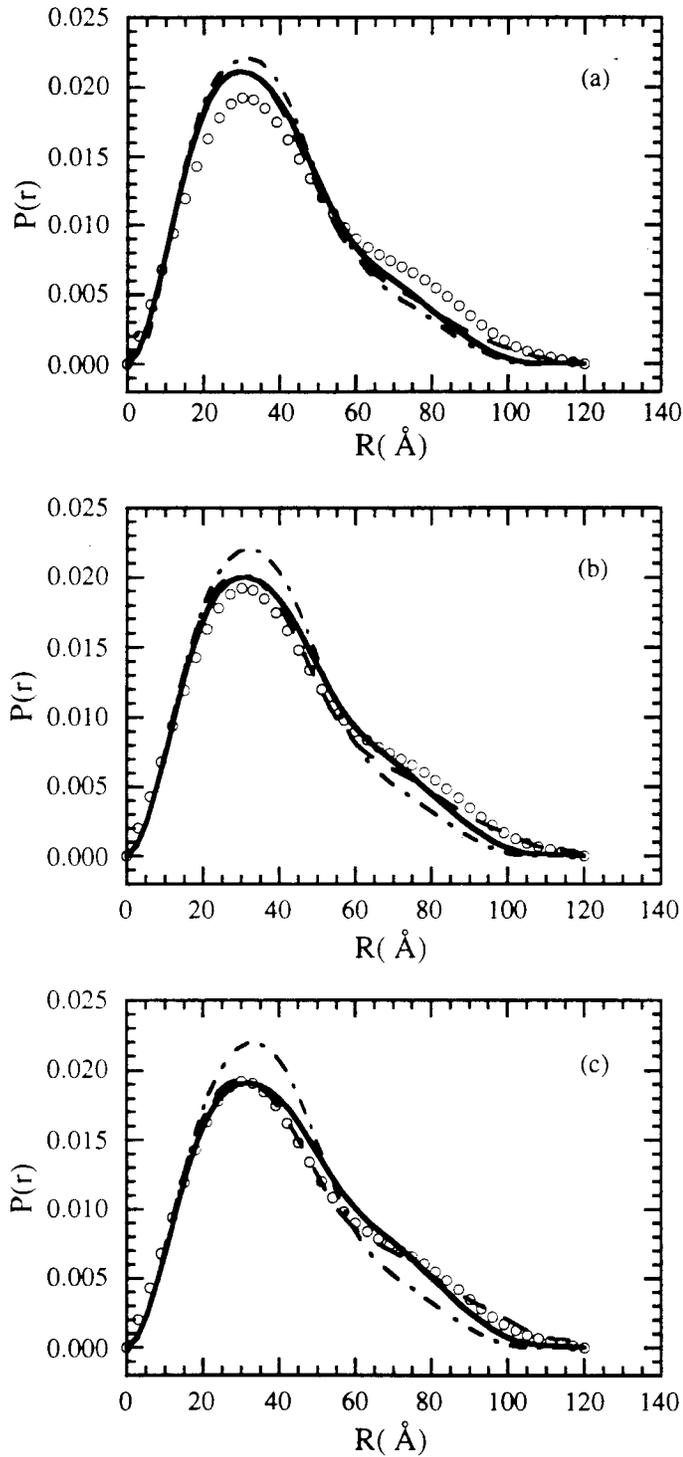
**Figure 3.14.** Distance distribution functions of the "partial dimer" box models for the 60 kDa fragment in the presence of MgADP with a separation of 58 Å. (a) 10% of the monomers form dimers; (b) 20% of the monomers form dimers; (c) 30% of the monomers form dimers. Dimers were formed —, end-to-end; – –, side-by-side; – - –, face-to-face. The experimentally derived P(r) functoins are shown in (O) for comparison.

angle between experimental scattering curves and those computed for a mixture of monomers and dimers.

In summary, agreement between experimental and model $P(r)$ functions for the ADP-bound form of the 60 kDa fragment can be achieved in two ways: either by inducing a large conformational change relative to the ATP-bound form and assuming that the solution is completely monomeric, or by introducing a significant fraction of dimerization in the solution. Arguments that may discriminate between these two alternatives are presented in the discussion.

3.4.3.3. Modeling the Full-length hsc70

To proceed the modeling to the full-length hsc70, the dimensions of the prism of the peptide binding domain were extended to account for the addition of the ~10 kD fragments. For convenience, only three situations were considered where an extension was made along one dimension in each case. Thus, prisms with 45x13x78 Å^3, or 45x22x50 Å^3, or 70x13x50 Å^3 were chosen to represent the carboxyl domain (residues 388-650) including the peptide binding and the ~10 kD fragments. As shown in Fig. 3.15, the best agreement on experimental and model $P(r)$ functions of both the ATP- and ADP-forms was with a box of 70x13x50 Å^3; this was achieved by increasing the intermediate dimension of the prism of the peptide binding domain by 25 Å.

3.5. Discussion

The results presented support the conclusions: (1) the ATPase and peptide binding fragments are relatively oblate; they are linked in a manner that forms a markedly elongated and prolate 60 kDa fragment and full length hsc70; (2) MgATP induces a substantial decrease in R_g in both hsc70 and the 60 kDa fragment, relative to their MgADP-bound forms; (3) the transition to the ATP-induced conformation occurs more rapidly than ATP hydrolysis, and presumably results from some step of ATP binding, while the rate of the reverse transition to the more elongated state correlates with product release.

Scattering curves measured under steady-state conditions for the 60 kDa fragment and its constituent domains provide the information necessary to construct "low
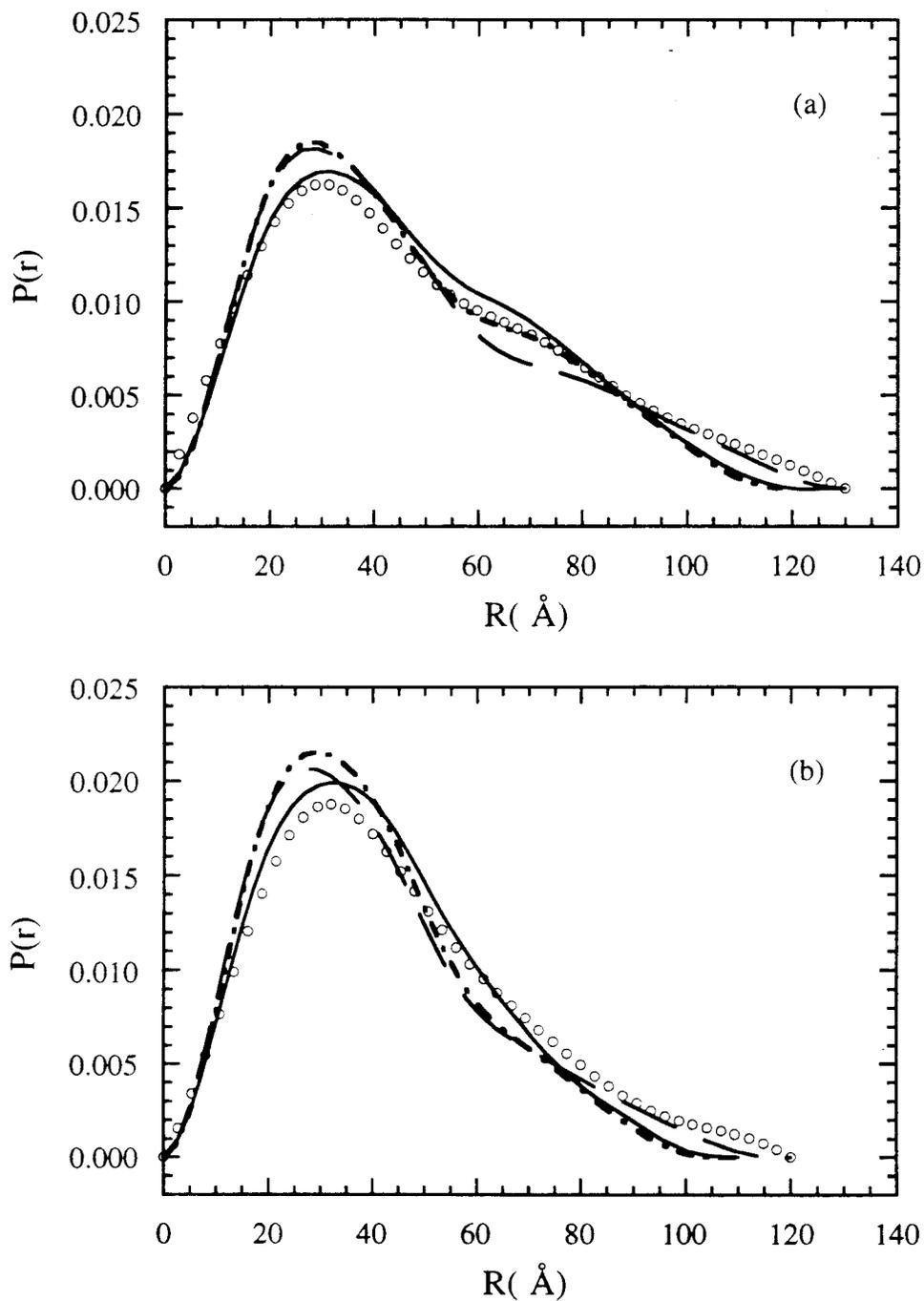
Figure 3.15. Distance distribution functions calculated for the box models for the full-length hsc70 in the presence of (a) MgADP; (b) MgATP. To make the box model for the carboxyl terminal fragment, the prism representing the peptide binding domain was enlarged along the direction of —, the longest dimension; — — the intermediate dimension; — -, the shortest dimension. The experimentally derived P(r) functions are shown in (O) for comparison.

resolution" models for these fragments. Reliability of the modeling methodology is evidenced by the agreement of the "box" model for the ATPase fragment with the overall dimensions of the crystallographic structure, and the agreement of both the box model and the crystallographic structure with the experimental scattering curves, as long as a ~3 Å layer of hydration is added to the crystallographic coordinates. These schematic models, taken in conjunction with the crystallographic structure of the ATPase fragment, show that a relatively flat peptide binding fragment probably interacts either end-to-end or side-to-side with the ATPase fragment in the ATP-induced form of the molecule. These results exclude models in which the peptide binding domain folds directly onto, and interacts extensively with, the largest faces of the ATPase domain.

Fig. 3.16 shows schemes of the two alternative hypotheses used to model the large increase in R_g that occurs going from the ATP-bound forms to the ADP-bound forms of the 60 kDa fragment of hsc70: (1) a large conformational change within monomeric molecules, versus (2) partial dimerization (*e.g.* ~30% of the protomers) of the molecules. It should be emphasized in this regard that the scattering curves do not display an upswing at low angle that would be indicative of nonspecific, heterogeneous aggregation (Fig. 3.5).

The partial dimerization model is appealing since hsc70 can be isolated in a dimeric form which can be dissociated with ATP (Freiden, *et al.*, 1992; Schlossman, *et al.*, 1984). However, there are several arguments against the dimerization model for the large ΔR_g we observe. First, partial dimerization would increase the average molecular weight for the ensemble of molecules, and this increase would be detectable as an increase in $I(0)/c$ (intensity at zero scattering angle over concentration). For example, if 30% of the molecules dimerize, an increase of ~30% in $I(0)/c$ is expected (see Section 3.3.1). However, although the sample-to-sample variation of the experimental values for this parameter is relatively large (~5-10 %) over extended periods of data collection, due to variations in beam intensity and detector response, we do not observe systematically larger values in the presence of ADP versus ATP on samples that are otherwise identical and on which measurements are made close in time (Fig. 3.6).

Secondly, if the larger R_g in the presence of ADP is the result of dimerization, this should be reflected in the dependence of R_g on protein concentration; R_g should increase significantly with increasing protein concentration and should extrapolate to approximately the same value at zero concentration as seen with ATP-bound protein.

ATP-Bound Form
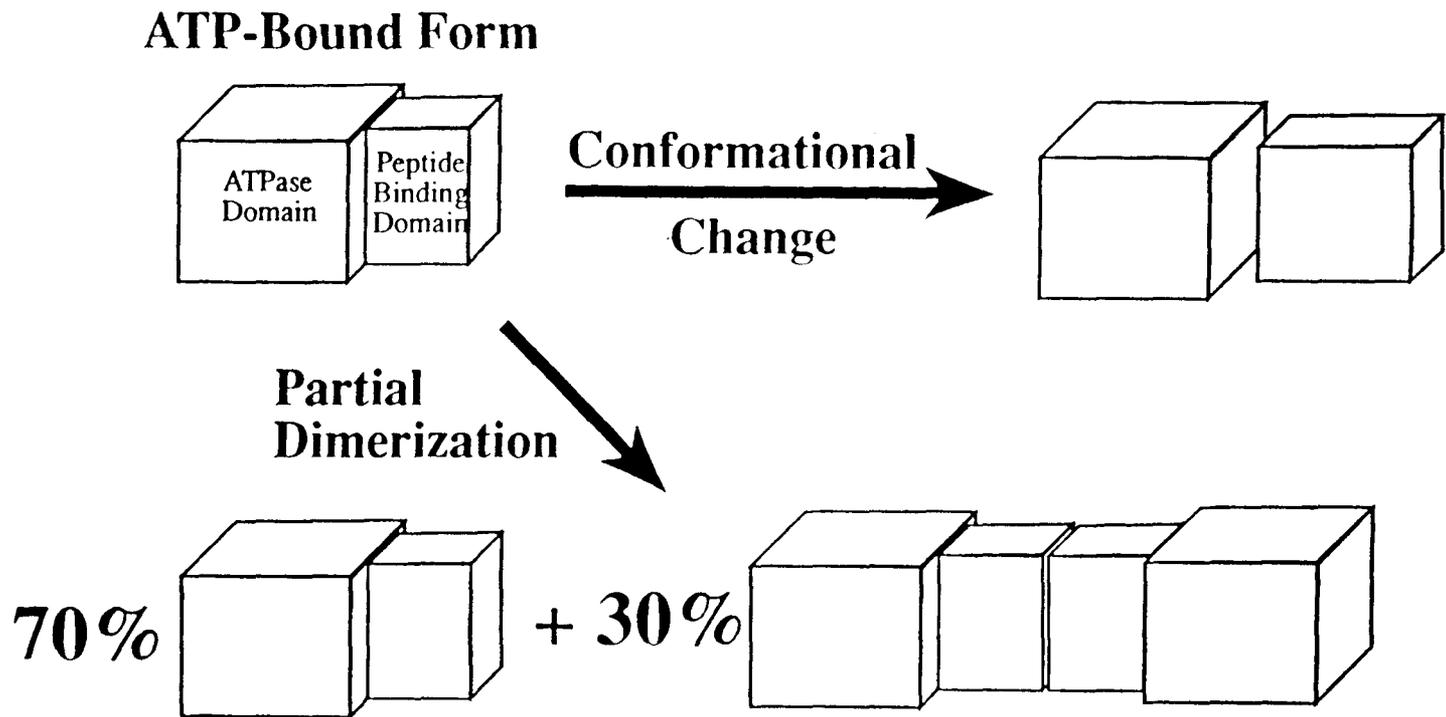
Figure 3. 16 Two alternative mechanisms for the MgATP-induced conformational change in the 60 kD fragment. In the monomeric model, the domain separation decreases from 58 Å to 44 Å; while in the "partial dimerization" model, 30% of the protomers form dimers.

However, the steady state R_g versus concentration has a similar, approximately level trend in the presence of both ADP and ATP for both hsc70 and the 60 kDa fragment (Fig. 3.7). Also, calculations show that the R_g computed in the Guinier approximation would be larger than that computed from the P(r) function; this is not the case (Table 3.1).

Thirdly, two different biochemical tests support our ability to isolate stable monomers of the 60 kDa fragment. Native gels of samples of the 60 kDa fragment in its ADP-bound form show only minor amounts (<1%) of dimer. hsc70 is more problematic; gels of hsc70 show significant aggregation after storage for extended periods of time. Dynamic light scattering from samples of monomeric hsc70 did not show polydispersity and showed a substantially smaller hydrodynamic radius compared to a sample purified as dimer.

Fourthly, when scattering curves were collected repetitively on samples of the 60 kDa fragment over many hours, the computed values of R_g remained stable. In contrast, experiments with hsc70 showed an increase in R_g by 20 hours, consistent with results of gel filtration showing that significant aggregation occurs with full-length protein after storage in solution for this amount of time.

Hence, although we can model the P(r) function of the ADP-bound 60 kDa fragment with a sample that is a mixture of monomers and dimers of the protein in its ATP-bound conformation, we do not see the effects we would expect from dimerization on values of I(0)/c, or on the concentration and time dependence of R_g. Nor do we see biochemical evidence for substantial aggregation of the 60 kDa fragment. It is notable in this regard that, taken by itself, the evidence on hsc70 is less compelling. While I(0)/c shows no dimerization of hsc70 in the presence of ADP, we see substantial variation in measured values of R_g as well as biochemical evidence for significant aggregation over time. We have tried to minimize the effect of this on our experiments by minimizing the amount of time between gel filtration of hsc70 samples and scattering measurements on them. The arguments discussed above, which rely more heavily on data from the 60 kDa fragment than on data from hsc70, favor a large, ATP-induced conformational change over partial dimerization as an explanation for the large ΔR_g that is observed in both proteins.

Regardless of which particular model ultimately proves to be correct, we discern two significantly different conformational states for both hsc70 and the 60 kDa fragment

by solution scattering. The state with the larger R_g is observed under several sets of conditions, including (i) absence of nucleotide, (ii) presence of saturating levels of MgADP, (iii) presence of slowly hydrolyzable ATP analogs, AMPPNP or AMPPCP, and (iv) with MgATP in the presence of 100 mM Na^+ rather than K^+ as a monovalent counter-ion. The other state, the ATP-induced state, is observed only in the presence of MgATP with K^+. This correlates with observations from other labs that MgATP induces a conformational change in hsp70 proteins, but slowly hydrolyzable ATP analogs fail to do so, and presence of K^+ is also required to achieve this conformational change (Banecki, *et al.*, 1992; Liberek, *et al.*, 1991; Palleros, *et al.*, 1992; Palleros, *et al.*, 1993).

Although these data demonstrate two different conformational states for hsc70 and the 60 kDa fragment, it should be noted that the values of steady-state R_g measured in the presence of saturating levels of MgATP represent a mixture of states. Under these conditions, all enzymatic intermediates of the ATPase cycle will be present, and the fraction of each intermediate in solution will be equal to the fraction of time that each molecule spends in that state during the enzymatic cycle. For the 60 kDa fragment, the kinetic constants for the ATPase cycle demonstrate that the majority of the molecules will have MgATP bound under the steady state conditions used in the scattering experiments. ATP binding is rapid ($\tau_{1/2}$ < 1 sec.), hydrolysis is the rate-limiting step of the cycle ($\tau_{1/2}$ = 400 sec.), and following hydrolysis, products are released rapidly ($\tau_{1/2}$ < 100 sec.), allowing ATP to re-bind. Thus greater than 80% of the 60 kDa fragment molecules will have ATP bound. By implication, since the structural transition to the ATP-induced conformation is apparently brought about by ATP binding, while the reverse transition is predicated on ATP hydrolysis or product release, greater than 80% of the molecules will be in the more compact, ATP-induced conformation in steady state experiments. This correlates with the observation that, when the transition is synchronized by mixing stoichiometric amounts of MgATP and 60 kDa protein, when essentially all of the protein should be in the ATP-induced conformation, the R_g at t = 0, (27.7 ± 0.5 Å, estimated from a double exponential fit to R_g versus time), is equal to that observed with MgATP in steady-state experiments (27.9 ± 0.5 Å).

For hsc70, the ATP hydrolysis and P_i release steps require approximately equal time ($\tau_{1/2}$ = 240 and 180 sec. respectively), while ADP release is relatively rapid ($\tau_{1/2}$ = 24 sec.). In the presence of excess MgATP, hsc70 will be an approximately 50:50 mixture of protein with ATP bound and protein with ADP + P_i bound. Depending on whether ATP hydrolysis, or alternatively, P_i or ADP release facilitates the transition from

the compact to the more elongated state, hsc70 would be either a ~50:50 mixture of protein in two different conformations, or alternatively, >90% in the ATP-induced form. The kinetics of the change of R_g after mixing stoichiometric amounts of hsc70 and MgATP show that the R_g observed at early times is equal within error to that observed with MgATP under steady state conditions, demonstrating that the transition to the elongated form is predicated on product release rather than ATP hydrolysis. The rate constant inferred from our data is consistent with this interpretation.

The conformational switch which induces the transitions between the two states presumably relies on the ability of nucleotide, which binds at the base of a cleft in the ATPase domain, to transmit a conformational shift to the periphery of this domain. However, no major change in R_g is observed between ATP-bound and ADP-bound 44 kDa ATPase fragment. This argues that any ATP-induced conformational change in this fragment is relatively subtle. An increase in R_g is observed in the absence of nucleotides, suggesting that the cleft of the ATPase fragment opens when nucleotide is released. However, this does not appear to correlate with the transition to or from the ATP-induced state.

When modeled as a large conformational change in a monomeric 60 kDa fragment, the change from the ATP-induced state to the more elongated ADP-bound state involves an increase of ~14 Å in the center-to-center distance between the two domains. This suggests that the ATPase and peptide binding domains disassociate and may have few if any noncovalent interactions in the more extended state. These conclusions can be extended to hsc70, whose behavior is similar to that of the 60 kDa fragment. The solution scattering data demonstrate that, at the level of substantial structural changes in the protein, the 60 kDa fragment mimics hsc70. A caveat to be emphasized at this point is that the carboxyl-terminal ~100 amino acid residues of hsc70 that are not present in the 60 kDa fragment may be involved in significant interactions to which the solution scattering experiments, being relatively coarse in resolution, are not sensitive.

The low peptide affinity state of DnaK characterized by Palleros *et al.* shares two significant features with the ATP induced state described here: (1) both are induced by ATP before hydrolysis occurs and (2) both appear to persist for a matter of minutes (Palleros, *et al.*, 1993). Although it is appealing and reasonable to suggest that the two different states observed in hsc70 and the 60 kDa fragment correspond to the high-

peptide-affinity and low-peptide-affinity states of the proteins, this has not been demonstrated directly.

A question remaining to be addressed is which chemical steps of the ATPase cycle transduce the conformational changes? The transition to the ATP-induced structure both in the 60 kDa fragment and in hsc70 occurs more rapidly than ATP hydrolysis and appears to be correlated with ATP binding. Crystallographic work on the wild-type ATPase fragment shows how both AMPPNP and ADP + P_i bind to this fragment. These structures, along with structures of several mutants with either ATP or ADP + P_i bound, show both a "pre-hydrolysis" state and the "post-hydrolysis" state for the nucleotide. There is no significant difference in the overall tertiary structure of the ATPase fragment between these two states. To the extent the crystallographic structure of the isolated ATPase fragment represents its structure in conjunction with the peptide binding domain in hsc70 and the 60 kDa fragment, this suggests that it must have an additional (that is, not yet crystallized) conformational state that is accessed by ATP but not by AMPPNP. It appears that AMPPNP is unable to stabilize the ATP-induced state in either crystal or solution studies. In the hydrolysis pathway proposed for ATP, the nucleotide first binds in the same manner as AMPPNP, and then rearranges to allow the phosphates to form a β,γ-bidentate complex with the Mg^{2+} ion in the active site. Hydrolysis then proceeds by an in-line attack mechanism. By inference, the postulated β, γ-bidentate complex of ATP is a candidate for a specific ATP conformation which provokes a structural shift which is then transduced from the nucleotide to the periphery of the ATPase fragment.

As for the step which precedes the converse structural transition, from the ATP-induced state to the more elongated state, our current data limit the possibilities to product release, but cannot discriminate between P_i or ADP release for this role; both are equally viable candidates for the biochemical step that triggers a conformational change. In the context of the structural similarity of the hsc70 ATPase fragment with actin, it is intriguing to see whether P_i release induces a subtle shift in the ATPase domain of hsc70 that destabilizes the interactions with the peptide binding domain, in the same manner that P_i release apparently induces a subtle structural change in actin that destabilizes the protein-protein interactions in the actin filament (Korn, *et al.*, 1987).

3.6. Acknowledgments

3.7. References

Anderegg, J. W., Beeman, W. W., Shulman, S., & Kaesberg, P. (1955). *J. Am. Chem. Soc.*, **77**, 2927-2937.

Banecki, B., Zylicz, M., Bertoli, E., & Tanfani, F. (1992). *J. Biol. Chem.*, **267**, 25051-25058.

Beckman, R. P., Mizzen, L. A., & Welch, W. J. (1990). *Science*, **248**, 850-854.

Bennett, W. S. J., & Steitz, T. A. (1980). *J. Mol. Biol.*, **140**, 211-230.

Blond-Elguindi, S., Cwirla, S. E., Dower, W. J., Lipshutz, R. J., Sprang, S. R., Sambrook, J. F., & Gething, M. J. (1993). *Cell*, **75**, 717-728.

Chappell, T. G., Konforti, B. B., Schmid, S. L., & Rothman, J. E. (1987). *J. Biol. Chem.*, **262**, 746-751.

Chappell, T. G., Welch, W. J., Schlossman, D. M., Palter, K. B., Schlesinger, M. J., & Rothman, J. E. (1986). *Cell*, **45**, 3-13.

Chirico, W. J., Waters, M. G., & Blobel, G. (1988). *Nature*, **332**, 805-810.

Deluca-Flaherty, C., McKay, D. B., Parham, P., & Hill, B. L. (1990). *Cell*, **62**, 875-887.

Deshaies, T. J., Koch, B. D., Werner-Washburne, M., Craig, E. A., & Schekman, R. (1988). *Nature*, **332**, 800-805.

Flaherty, K. M., DeLuca-Flaherty, C. M., & McKay, D. B. (1990). *Nature*, **346**, 623-628.

Flaherty, K. M., McKay, D. B., Kabsch, W., & Holmes, K. C. (1991). *Proc. Natl. Acad. Sci. USA*, **88**, 5041-5045.

Flaherty, K. M., Wilbanks, S. M., Deluca-Flanerty, C., & McKay, D. B. (1995). *J. Bio. Chem.*, **269**(17), 12899-12907.

Fletterick, R. J., Bates, D. J., & Steitz, T. A. (1975). *Proc. Natl. Acad. Sci. USA*, **72**, 38-42.

Flynn, G. C., Chappell, T. G., & Rothman, J. E. (1989). *Science*, **245**, 385-390.

Flynn, G. C., Pohl, J., Flocco, M. T., & Rothman, J. E. (1991). *Nature*, **353**, 726-730.

Freiden, P. J., Gaut, J. R., & Hendershot, L. M. (1992). *EMBO J.*, **11**, 63-70.

Fujisawa, T., Ueki, T., & Iida, S. (1990). *J. Biochem.*, **107**, 343-351.

Glatter, O., & Kratky, O. (Ed.). (1982). *Small Angle X-ray Scattering*. New York: Academic Press Inc.

Gragerov, A., Zeng, L., Zhao, X., Burkholder, W., & Gottesman, M. E. (1994). *J. Mol. Biol.*, **235**, 848-854.

Ha, J.-H., & McKay, D. B. (1994). *Biochemistry*, **33**, 14625-14635.

Hiragi, Y., & Ihara, S. (1981). *Acta Cryst.*, **A37**, 378-382.

Kim, D., Lee, Y. J., & Corry, P. M. (1992). *J. Cell Physiol.*, **153**, 353-361.

Korn, E. D., Carlier, M.-F., & Pantaloni, D. (1987). *Science*, **238**, 638-644.

Kraulis, P. (1991). *J. Appl. Cryst.*, **24**, 946-950.

Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M. K., & Hartl, F.-U. (1992). *Nature*, **356**, 683-689.

Liberek, K., Skowyra, D., Zylicz, M., Johnson, C., & Georgopoulos, C. (1991). *J. Biol. Chem.*, **266**, 14491-14496.

McKay, D. B. (1993). *Adv. Prot. Chem.*, **44**, 67-98.

Morimoto, R. I., Tissieres, A., & Georgopoulos, C. (Ed.). (1990). *Stress Proteins in Biology and Medicine*. New York: Cold Spring Harbor Laboratory Press.

Morimoto, R. I., Tissieres, A., & Georgopoulos, C. (Ed.). (1994). *The Biology of Heat Shock Proteins and Molecular Chaperones*. New York: Cold Spring Harbor Laboratory Press.

Morshauser, R. C., Wang, H., Flynn, G. C., & Zuiderweg, E. R. P. (1995). *Biochemistry*, **34**, 6261-6266.

Nover, L. (Ed.). (1984). *Heat Shock Response of Eukaryotic Cells*. New York: Springer-Verlay.

O'Brien, M. C., & McKay, D. B. (1993). *J. Biol. Chem.*, **268**, 24323-24329.

Palleros, D. R., Reid, K. L., McCarty, J. S., Walker, G. C., & Fink, A. L. (1992). *J. Biol. Chem.*, **267**, 5279-5285.

Palleros, D. R., Reid, K. L., Shi, L., Welch, W. J., & Fink, A. L. (1993). *Nature*, **365**, 664-666.

Rice, M., & Wakatsuki, S. (1991). *SSRL/SLAC user manual*.

Ritossa, F. (1962). *Experientia*, **18**, 571-573.

Scherer, P. E., Krieg, U. C., Hwang, S. T., Vestweber, D., & Schatz, G. (1990). *EMBO J.*, **9**, 4315-4322.

Schlossman, D. M., Schmid, S. L., Braell, W. A., & Rothman, J. E. (1984). *J. Cell Biol.*, **99**, 723-733.

Semenyuk, A. V., & Svergun, D. I. (1991). *J. Appl. Cryst.*, **24**, 537-540.

Svergun, D. I., Barberato, C., & Koch, M. H. (1995). *J. Appl. Crystallgr*, submitted for publication.

Tissieres, A., Mitchell, H. K., & Tracy, U. M. (1974). *J. Mol. Biol.*, **84**, 389-398.

Wakatsuki, S., Hodgson, K. O., Eliezer, D., Rice, M., Hubbard, S., Gillis, N., & Doniach, S. (1992). *Rev. Sci. Instrum.*, **63**, 1736-1740.

Wang, T., Chang, J., & Wang, C. (1993). *J. Biol. Chem.*, **268**(35), 26049-26051.

Wilbanks, S. M., DeLuca-Flaherty, C., & McKay, D. B. (1994). *J. Biol. Chem.*, **269**, 12893-12898.

Chapter 4

Equilibrium Studies of Lysozyme Unfolding by Solution X-Ray Scattering: Evidence for A Stable Folding Intermediate

4.1. Introduction and Background

The protein folding problem has been a major challenge in biophysical research. More specifically, the protein folding problem refers to the questions why and how a protein of given amino acid sequence folds into a specific three-dimensional native conformation. What determines which of the large number of possible conformations is adopted by a protein as its unique native conformation? How does the polypeptide chain search into the native conformation given the astronomical number of possible conformations that a given amino acid sequence composed of 20 amino acids can assume? What are the rules governing the relationship between amino acid sequence and three-dimensional structure? The ultimate goals in protein folding research are to predict three-dimensional structures from the primary sequences, to design proteins to achieve desired structures and therefore to perform specific functions.

The study of the protein folding problem began with Anfinsen's pioneering work in the late 1950's. In 1973, based on the observations that an unfolded polypeptide, ribonuclease A, could spontaneously refold to form a native protein with full biological activity *in vitro*, Anfinsen proposed that the necessary information required to define the three-dimensional structure of a protein is contained in the primary sequence of the polypeptide (Anfinsen, 1973). However, unlike the straightforward genetic DNA codes for protein synthesis, the folding codes for proteins appears formidably complex.

The theoretical and computational approaches to the folding problem can be roughly classified into two major categories. One is molecular dynamics simulation which is based on classsical statistics and quantum mechanics calculation (see Dill, *et al.*, 1995 for a recent review). Another approach is data-based predictive algorithms based on sequence homology in the protein data base (See Blundell & Johnson, 1993 for a recent review).

Experimentally, at present a more tractable problem is to study the folding mechanisms, *i.e.* the processes or pathways by which proteins adopt their native conformations. Much of the work in this field has been devoted to this kind of studies. Like any chemical reactions, understanding of a folding mechanism includes the elucidation of the pathway(s) between the native folded state and the unfolded state, the intermediates and the transition states along the folding pathway(s). Currently, there are two major generalized models for the folding mechanisms. The framework model suggests that an early intermediate with the framework of secondary structure is formed before a structure

with the native-like tertiary fold and stable secondary structure is formed (Ptisyn, 1973). The modular assembly model postulates that microdomains of a protein fold transiently and independently, and they diffuse, collide and coalesce into the stable native form (Karplus & Weaver, 1976; Kim & Baldwin, 1990).

Kinetics intermediates on the folding pathway have been identified mainly by kinetic CD, fluorescence and pulsed hydrogen labeling techniques (see Schmid, 1993 for a recent review). Early intermediates are formed within milliseconds after the initiation of refolding, and are normally beyond the time resolution of most techniques. Late intermediates, formed at a longer time scale, hundreds of milliseconds or even longer after refolding, reveal some structural information. However, structural characterizations of kinetic intermediates are intrinsically difficult due to their transient occurrence. An alternative is to investigate stable intermediates under different stages of equilibrium folding.

The first evidence for intermediate conformation between the rigid native structure and the random coil unfolded state was reported in 1967. It was found that temperature or acid denaturation of a protein transformed the native state to a state which could be further unfolded by stronger denaturants such as urea or guanidine hydrochloride (Aune, et al., 1967). The important demonstrations for the existence of intermediate folding species under equilibrium conditions came from equilibrium unfolding studies of α-lactalbumin where two distinct transitions were observed by far- and near-UV circular dichroism (Kuwajima, et al., 1976). The intermediate was found to have a native secondary structure but without a rigid tertiary structure.

The term "molten globule" was first proposed to describe some intermediate states at low pH (Ohgushi & Wada, 1983). The acid form is found to be nearly as compact as the native state with native-like secondary structure content, but without rigid tertiary structure. Since its emergence, the molten globule has attracted the most attention among various folding intermediate states. A range of physical techniques, particularly NMR and x-ray scattering, has been applied to characterize this type of intermediate, revealing different aspects of physical properties of this species. The current model for molten globule intermediates is that the molten globule state not only maintains some of the native-like secondary structure, but also preserves the native-like folding pattern; however, it lacks the tight packing of side chains and has looser structure than the native state (Ptisyn, 1993).

104

Study of equilibrium intermediates was initially motivated by the belief that these stable intermediates under different stages of equilibrium folding might mimic the transiently occurring kinetic counterparts in the folding pathway. Compact intermediates identified in the folding pathway are found to share similar structural features, *i.e.* substantial secondary contents and lack of rigid tertiary structure, to the equilibrium analog. Correlations between the kinetics and equilibrium intermediates have been observed in some of the most frequently studied proteins, such as α-lactalbumin, cytochrom c, hen lysozyme and interleukin-β. A kinetic intermediate in the apomyoglobin folding is found to have the same NH-protection pattern as an equilibrium molten globule (Jennings & Wright, 1993). It becomes established that the molten globule is the common kinetic intermediate on the folding pathway (Kim & Baldwin, 1990; Ptitsyn, 1995a).

In this chapter, the x-ray scattering technique has been applied to study urea-induced lysozyme unfolding under equilibrium conditions. A partially folded intermediate of hen lysozyme has been found which is induced by a combination of reduced pH and denaturant. The intermediate is relatively compact; however, there is a signature which suggests some random coil component.

Lysozyme. Hen egg-white lysozyme is one of the earliest characterized and most studied globular proteins. Its 129 amino-acid residues fold into two distinctive structural domains as shown in Fig. 4.1: the α domain consists of helices A-D and a 3^{10} helix, the β domain includes a long triple-stranded β sheet, a short double-stranded β sheet, a 3^{10} helix and a long loop (Blake, *et al.*, 1965; Radford, *et al.*, 1992). The structure of lysozyme is further stablized by the presence of four disulfide bonds: two (6-127, 30-115) within the α domain, one (76-94) between the α and β domains, and one (64-80) within the β domain. Two proline residues at positions 70 and 79 are *trans* in the native conformation. It is interesting to know that most issues on the folding (folding of α helices and β sheets, formation of disulfide bonds, and the effect of proline isomerization on folding) are present in such a small protein.

Kinetics and equilibrium aspects of lysozyme folding have been under extensive investigation (Dobson, *et al.*, 1994; Tanford, 1970). To decipher the folding pathway, it is important to characterize folding intermediates. For hen lysozyme, kinetic intermediates have been identified in the early stage (<100 ms) of the refolding pathway (Kato, *et al.*, 1981; Radford, *et al.*, 1992). The α domain of the intermediates appears to have some

Figure 4.1. Three dimensional structure of hen lysozyme (entry PDB6LYZ.ENT, Brookhaven Protein Data Bank). The four disulfide bridges are shown along with the corresponding Cysteine residues.

characteristics of a molten globule, *i.e.*, with an extensive secondary structure but a distorted, far from native-like tertiary structure, while the conformation of the β domain is still highly labile (Radford, *et al.*, 1992). Structural studies of these intermediates are difficult because of their transient occurrence.

An alternative approach is to study the partially folded states as evidence is growing that they may resemble kinetic intermediates (Dobson, 1994; Ptitsyn, 1995b). However, the equilibrium unfolding transition of lysozyme, mostly induced by thermal or GdmHCl denaturation, has been traditionally approximated by a two-state model due to the high unfolding cooperativity (Aune & Tanford, 1969). Stable partially folded intermediates were not visibly present under those circumstances. Recently, some stable intermediates have been reported on either lysozyme from other species, such as equine and human (Dael, *et al.*, 1993; Haezebrouck, *et al.*, 1995) as well as a homologous protein α-lactalbumin (Baum, *et al.*, 1989), or hen lysozyme using the mild denaturant trifluoroethanol (TFE) at pH 2.0 (Buck, *et al.*, 1993; Dael, *et al.*, 1993). Structures of these partially folded states, determined by NMR, far- and near-UV CD and fluorescence spectroscopies, show molten globule features. They have substantial α helical content and a distorted tertiary fold in the α domain and little structure in the β domain, and resemble that of the early kinetic intermediate in hen lysozyme. Additional evidence for intermediates is also documented in the literature: thermal unfolding curves of hen lysozyme are not identical, as observed by absorption at 292 nm and optical rotation at 546 nm at pH 2.2 (Wetlaufer, 1962), and by CD measurements at 222 nm and 270 nm at pH 1.0 (Haezebrouck, *et al.*, 1995). A population of ~25% of an intermediate was suggested from the latter study. An early NMR study found that in hen lysozyme, regions containing histidine and arginine residues unfolded at a lower urea concentration than portions with methionine at pH 2.8 in urea solution (Bradbury & King, 1969).

Compactness is one of the important properties pertinent to the structure of protein molecules in solution, and it also characterizes the degree of folding. Small angle x-ray scattering (SAXS) is one of the few techniques that provides a direct measurement of compactness since SAXS gives overall geometric information such as the size and shape of the molecule. Application of SAXS to the study of protein folding has been proven successful (Eliezer, *et al.*, 1993; Kataoka, *et al.*, 1995; Lattman, 1994), and information provided from such studies is complementary to that obtained by other techniques, such as CD, fluorescence and NMR, in the folding study.

4.2. Experimental Procedures

4.2.1. Sample Preparations

Hen egg-white lysozyme was purchased from the Sigma Chemical Company, St. Louis, MO. A lysozyme stock solution of ~ 70 g/l was dialyzed extensively at 4 ^{O}C against a large volume of buffer (100 mM Na-citrate, 100 mM NaCl at pH 2.9; 100 mM Na-phosphate, 100 mM NaCl at pH 2.0; or 100 mM Na-acetate, 100 mM NaCl at pH 3.8). The solution was concentrated back to ~ 70 g/l using an Amicon concentrator (Amicon, Danvers, MA) when necessary. The solution was ultra-centrifuged at 34,000 rpm in Beckman Ti50 rotor for at least 60 min at 10 °C to remove any protein precipitate and dust particles. The pH of a 10 M urea stock solution in the same buffer was adjusted to 2.9 or 2.0, and was filtered through a 0.2 μm filter. The titration samples of different urea concentrations were then prepared volumetrically from the stock solutions. All unfolded protein samples were allowed to equilibrate for at least five hours prior to both scattering and CD experiments. The final lysozyme concentrations, measured by UV absorption (ε_{280} = 2.64 mg/ml), were about 7 mg/ml for scattering experiments. All the samples containing urea were used within 24 hours. The lysozyme solutions were used within 48 hours. To prepare reduced unfolded lysozyme, the unfolded sample (in 8 M urea for more than 5 hours) was incubated in 80 mM dithiothrietol solution for up to 3 hours.

Heparin (from Calbiochem, San Diego, CA) was further purified by gel filtration using Sephacryl S-200. The buffer for heparin was 0.5 M NaCl, pH 2.5 under which conditions heparin is assumed to exist as a random coil by viscosity measurements (Liberti & Stivala, 1967).

Although GdmHCl is a stronger chemical denaturant for proteins than urea, it is not ideal for SAXS studies because of the relatively large absorption of x-rays by chlorine. As shown in Fig. 4.2, transmittance of x-rays decrease dramatically with increased GdmHCl concentration. The absorption of a 6 M GdmHCl solution is about 7.5 times higher than the same solution without GdmHCl. On the other hand, the presence of urea has little effect on absorption. Thus, urea is used instead as the denaturant in the equilibrium (this chapter) and kinetic (Chapter 5) studies.
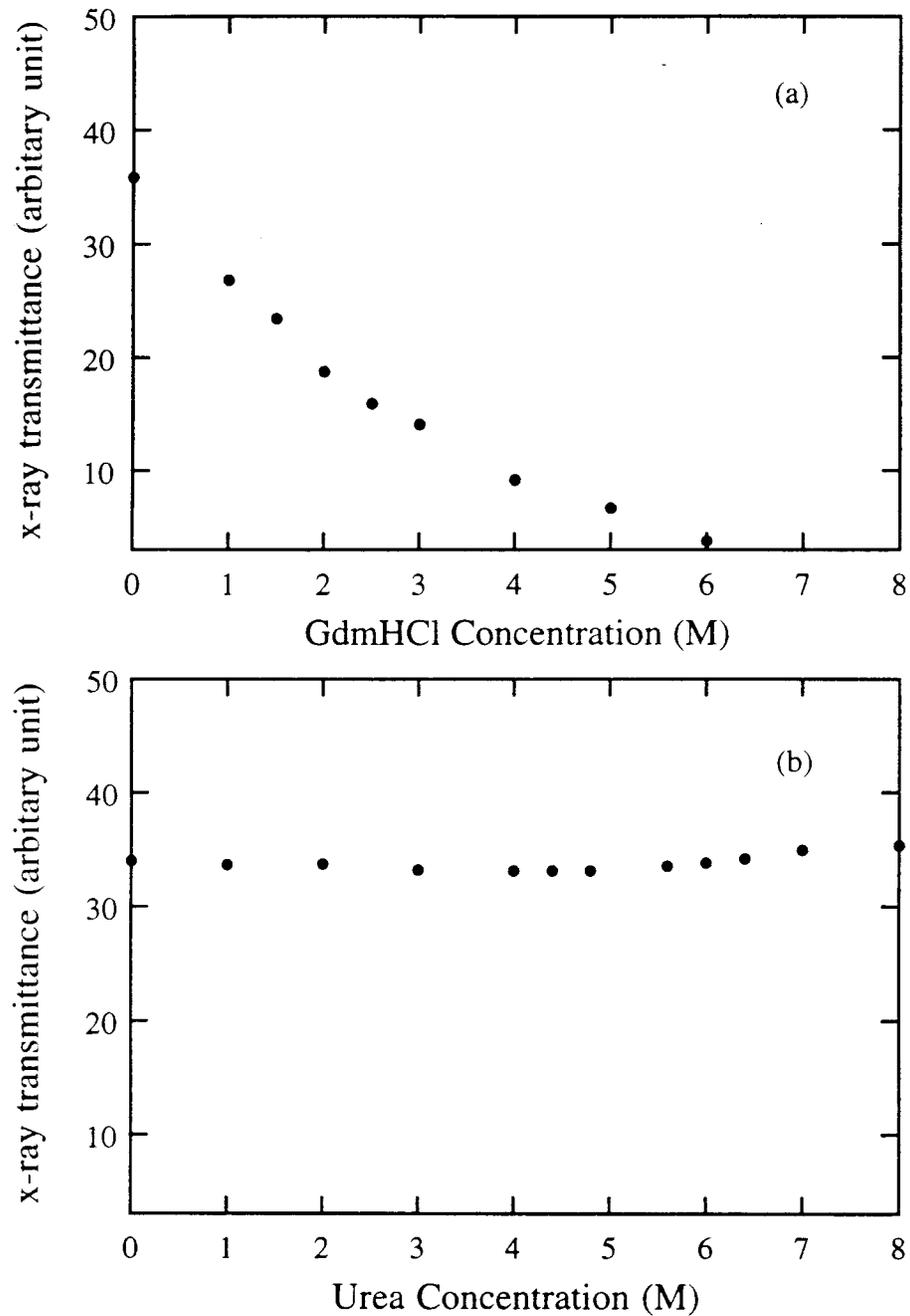
Figure 4.2. Absorption of x-rays by GdmHCl and urea solution. Urea is prefered for SAXS studies as it does not absorb much of the x-rays.

4.2.2. Data Collection

X-ray scattering data were collected at beamline 4-2 at Stanford Synchrotron Radiation Laboratory. The x-ray energy was selected by a pair of Si [111] monochromators at 8980 eV, or by a pair of Mo/B$_4$C multilayers (Stephenson, 1988). In the case of multilayers, aluminum foils were used to attenuate beam intensity when necessary. To monitor the fluctuations of the x-ray intensity, a short length ion chamber and a scintillation counter were placed before and after the sample cell, respectively. The linear position sensitive proportional counter for the scattering data was filled with either 70% Ar + 30% CO$_2$ or 80% Xe + 20% CO$_2$. When using the Ar mixture as detector gas and with monochromatic beam, typical counting rates were ~4,400 counts per second (cps) for buffer and 6,500 cps for the sample; while using Xe mixture as detector gas and with broad-band pass beam, counting rates increased to ~30,000 cps for buffer and ~50,000 cps for the sample. The camera length was calibrated to be 230 cm for the normal SAXS configuration and 150 cm for the short camera-length setup, using a cholesterol myristate sample.

At the early stage of these serial experiments, data were collected using static sample cells as described in Section 3.2.2. Depending on the intensity of the incident x-ray beam, the exposure time was limited to 8 or 10 minutes since a time dependence of the scattering profile at the small angle region was observed after ~12 minutes. For samples containing high concentration of denaturants, the difference in counting rate between the sample and the buffer decreased to less than 1,000 cps in some cases, hence a longer collection time was required to improve data statistics. A flow cell was then developed to extend data acquisition time to longer than 30 minutes by circulating protein solution continuously from a reservoir through a 1.3 mm scattering-path observation cell with 10 μm thick mica windows. The flow rate was adjusted in such a way that protein molecules would be exposed to the x-ray beam for less than one second during each circulation. The temperature of both the static and the flow cells was controlled by circulating ethylene glycol through the cell holder, and was set to be 20 °C for the cell, and also for the reservoir in the case of the flow cell.

CD data were measured at 222 nm at 20 °C with an Aviv 60DS spectropolarimeter using 1 mm-path cell. Protein concentration was determined to be ~ 40 μM (~ 0.6 mg/ml).

4.3. Data Analysis

4.3.1. Data Reduction and Analysis

The radius of gyration was calculated according to the Guinier approximation (Glatter & Kratky, 1981):

$$LnI(S) = LnI(0) - \frac{4\pi^2}{3} R_g^2 S^2 \qquad [1]$$

The fitting region is from 0.0045 to 0.01 $\AA^{-1}$ in S with $S = 2\sin\theta/\lambda$, λ being x-ray wavelength. I(0), the scattering intensity at zero angle, is proportional to the square of the excess electronic density between the protein molecule and the buffer background. As the urea concentration increases, the contrast in electron density becomes smaller, and I(0) decreases quadratically with urea concentration.

Distance distribution (P(r)) functions

$$P(r) = 8\pi r \int_0^\infty SI(S)\sin(2\pi Sr)dS \qquad [2]$$

were evaluated using the indirect transformation method implemented in the program GNOM (Semenyuk & Svergun, 1991). The P(r) function measures the density of intramolecular pair distance distributions within the scatterer, the protein molecule. The range used for the P(r) calculation was from 0.003 to 0.03 $\AA^{-1}$ in S. A scattering curve was computed from the coordinates of the crystal structure of lysozyme (entry PDB2LYZ.ENT in Brookhaven Protein Data Bank) using the program CRYSOL (Svergun, et al., 1995). The thickness of the hydration layer was set to one layer of H_2O molecules, and the excess electron density between this hydration layer and the buffer was taken to be 0.03 $e/\AA^3$.

4.3.2. Thermodynamic Analysis of the Denaturation Curve

The denaturation curve was analyzed with either a two-state or a three-state model. The following analysis was based on the three-state analysis, and could easily be applied to the two-state model. For an equilibrium ternary system consisting of native (N), intermediate (I) and unfolded (U) molecules, the measured amide CD signals (θ) and radius of gyration (R_g) were decomposed according to equations [3] and [4]:

$$\theta^k = f_N^k \theta_N + f_I^k \theta_I + f_U^k \theta_U \qquad [3]$$

$$R_{g,k}^2 = f_N^k R_{g,N}^2 + f_I^k R_{g,I}^2 + f_U^k R_{g,U}^2 \qquad [4]$$

where f_N^k, f_I^k and f_U^k are the fractions of the protein molecules in the native, intermediate and unfolded states respectively. At equilibrium, the fractions can be derived thermodynamically:

$$f_N^k = \frac{1}{1 + K_{N\to I}^k + K_{N\to U}^k} \qquad [5a]$$

$$f_I^k = \frac{K_{N\to I}^k}{1 + K_{N\to I}^k + K_{N\to U}^k} \qquad [5b]$$

$$f_U^k = \frac{K_{N\to U}^k}{1 + K_{N\to I}^k + K_{N\to U}^k} \qquad [5c]$$

where $K_{N\to I}^k$ and $K_{N\to U}^k$ are the equilibrium constants for native $\to$ intermediate and native $\to$ unfolded transitions at k M urea solution respectively; and $f_N^k + f_I^k + f_U^k = 1$. The equilibrium constants are related to the changes in Gibbs energy as:

$$\Delta G_I^k = -RT \ln K_{N\to I}^k \qquad [6a]$$

$$\Delta G_U^k = -RT \ln K_{N\to U}^k \qquad [6b]$$

4.3.3. Models of Denaturant Effects

$\Delta G_U^{H_2O}$, the free energy difference between the native and the unfolded state of a protein extroplated to zero denaturant concentration, is a good indication of the stability of a native protein. Estimates of $\Delta G_U^{H_2O}$ come from calorimetic measurements, hydrogen exchange studies, and mostly from analysis of the denaturation curve (Pace, 1986). To analyze the denaturation curve, three models are used to account for the effect of denaturants on protein stability.

Tanford's model. This model attributes the unfolding ability of denaturants to an increase in solubility of the constituent parts of proteins with increasing denaturant concentration (Tanford, 1970). Therefore, free energies of transfer between H_2O and denaturant (urea or GdmHCl), derived from solubility studies on amino acids and peptides, can be used quantitatively to describe unfolding of proteins by denaturants:

$$\Delta G_U = \Delta G_U^{H_2O} + \overline{\alpha} \sum_i n_i \delta g_{tr,i} \qquad [7]$$

where $\delta g_{tr,i}$ is the free energy for transfer of a group of type i from H_2O to denaturant, $\overline{\alpha}$ is the average fractional change in the degree of exposure for all groups when a protein unfolds, and n_i is the number of groups of type i in the protein. Values of both $\delta g_{tr,i}$ and n_i are available in the literature. $\overline{\alpha}$ is chosen so that the dependence of ΔG_U on denaturant concentration calcuated from equation [7] is equal to the experimentally derived value, the m value in equation [8] (see below).

Linear extrapolation model. This simple model (Greene & Pace, 1974) assumes that ΔG_U^k is linearly dependent on denaturant concentration, and that an estimate of $\Delta G_U^{H_2O}$ can be obtained by extrapolation of ΔG_U^k to zero denaturant concentration:

$$\Delta G_U^k = \Delta G_U^{H_2O} - m[\text{denaturant}] \qquad [8]$$

The assumption is based on the observation of the linear dependence of ΔG_U^k on denaturant concentration in the transition region. This method normally gives a lower value of $\Delta G_U^{H_2O}$ than the other two models described herein.

Denaturant binding model. This model is employed in our study to account for the effect of urea on protein stability. The model assumes that unfolding of a protein results from an increase in the number of denaturant binding sites on the unfolded (or intermediate) conformation compared to the native form (Aune & Tanford, 1969):

$$\Delta G_I^k = \Delta G_I^{H_2O} - \Delta n_I RT \ln(1 + ca) \qquad [9a]$$

$$\Delta G_U^k = \Delta G_U^{H_2O} - \Delta n_U RT \ln(1 + ca) \qquad [9b]$$

where $\Delta G_I^{H_2O}$ and $\Delta G_U^{H_2O}$ are the changes in free energy in the unfolding of the native to the intermediate and the native to the unfolded states respectively, extrapolated to zero denaturant concentration. Δn_I and Δn_U are the differences in the number of urea binding sites between the native and the intermediate states, and between the native and the unfolded states. R is the gas constant, and T is the temperature. c is the averaged .equilibrium binding constant at each site. a is the activity of urea which is related to the molar concentration, M, by the relation: $a = 0.9815(M) - 0.02978(M)^2 + 0.00308(M)^3$ (Pace, 1986).

Substitutions of the terms in equations [3] and [4] with corresponding terms from equations [5a-c], [6a-b] and [9a-b] give the following two equations:

$$\theta^k = \frac{(\theta_N + \theta_I e^{-\gamma} + \theta_U e^{-\rho})}{1 + e^{-\gamma} + e^{-\rho}} \qquad [10]$$

$$R_{g,k}^2 = \frac{(R_{g,N}^2 + R_{g,I}^2 e^{-\gamma} + R_{g,U}^2 e^{-\rho})}{1 + e^{-\gamma} + e^{-\rho}} \qquad [11]$$

where

$$\gamma = (\Delta G_I^{H_2O} - \Delta n_I RT \ln(1 + ca)) / RT$$

$$\rho = (\Delta G_U^{H_2O} - \Delta n_U RT \ln(1 + ca)) / RT$$

These two equations are used for simultaneous fit of the CD and $\dot{R}_g$ data using a nonlinear least squares method.

4.3.4. Singular Value Decomposition (SVD)

Singular value decomposition (SVD) (Golub & Reinsch, 1970; Press, *et al.*, 1992) has been successfully applied to the analysis of transient absorption spectra to resolve and identify kinetic intermediates involved in the process of photodissociation of carbon monoxide in hemoglobin and myoglobin (Hofrichter, *et al.*, 1983; Lambright, *et al.*, 1991). The method has also been utilized in x-ray scattering studies of structural transition of ATCase (Fetler, *et al.*, 1995). The significance of this technique is two-fold (Shrager & Hendler, 1982): 1) It does not require *a prior* knowledge of the models, *i.e.* the scattering curves of the components; 2) It is particularly powerful in decomposing complex spectra where components share overlapping features.

To determine the minimum number of basis scattering functions in the unfolding transition, the SVD method was applied to the analysis of the scattering curves, based on the supposition that the scattering curves are very likely to be of highly singular value. Unlike the R_g analysis in which only the scattering data from the small angle (Guinier) region is used, SVD is applied to the whole scattering space, including the Guinier region. The application of SVD to the analysis of titration data has been described (Fetler, *et al.*, 1995; Provencher & Blockner, 1983; Shrager & Hendler, 1982), thus only the salient features are summarized here.

A set of scattering profiles measured for a range of denaturant concentrations can be represented as an m x n matrix $A(S,k)$, with each column being the scattering profiles at k M urea concentration. According to the therom of SVD, any matrix may be written as the product of three matrices:

$$A(S,k) = U(S,k)SV^T$$

where $U(S, k)$ is an m x n matrix of orthogonal columns which form a complete set of basis scattering curves in terms of which the scattering spectrum at any urea concentration can be represented as a linear superposition. The column of $U(S, k)$ with the largest singular value is the best single-component representation of the $A(S, k)$ in the least square sense. The first two columns constitute the best two-component representation $A(S, k)$ and so on. S is an n x n diagonal matrix with non-negative elements (w) called the singular values, which describe the overall contribution of each basis scattering curve $U(S, k)$. V^T

115

is the transpose of an n x n orthonormal matrix V, whose columns represent the denaturant concentration depedence for each of the basic scattering curve. Each row of SV^T contains the set of urea-dependent linear superposition coefficients corresponding to each basis function.

The tremendous utility of SVD is realized when only a few components are found to be significant. In the present application, it is expected that only a few basic scattering functions will make significant contributions. The remains have singular values so close to zero that their contributions do not add to our structural knowledge, so they are neglected. The scattering curve at k M denaturant can then be approximated as a linear combination of the significant basis functions:

$$a_k(S) = \sum_{j=1}^{L} w_j V_{kj} U_j \qquad [12]$$

where L is the minimum number of components that are required to represent each of the scattering curves, terms with $j > L$ representing noise. $a_k(S)$, the column of matrix $A(S, k)$, represents the scattering intensity at k M urea, w_j is the singular value of the basis function U_j, and V_{kj} contains denaturant-dependent coefficients of the basis function U_j at k M urea.

It should be noted that the significant basis functions (or scattering curves) do not represent the scattering curves of molecular species. However, SVD does resolve the number of scattering distinct species (L) by determining the number of significant singular values contained in the data.

Background subtracted scattering data were first normalized to the same I(0) as with the native condition, in 0 M urea, to account for the decrease in electron density contrast due to the increase in urea concentration. Data were presented in $S^2 I$ *vs* S (Kratky form) and then smoothed with a three-point polynomial. The number of significant basis functions were selected based on the four criteria: significance of the singular value, the autocorrelation of the corresponding column U, the shape of the basis function, and the minimum fitting reduced-chi-square ($\overline{\chi}^2$).

116

$$\bar{\chi}^2 = \frac{1}{m \times (n-j)} \sum_k \sum_{i=1}^m \left(\frac{S_i^2 \times I_{exp}^k(S_i) - S_i^2 \times I_{cal,p}^k(S_i)}{S_i^2 \times \sigma_i} \right)^2.$$

$$= \frac{1}{m \times (n-j)} \sum_k \sum_{i=1}^m \left(\frac{I_{exp}^k(S_i) - I_{cal,p}^k(S_i)}{\sigma_i} \right)^2$$

[13]

where σ_i is the error associated with the counting statistics at S_i.

Since the scattering intensity is additive, the scattering intensity (I^k) of a solution at k M denaturant concentration with three species, native (N), intermediate (I) and unfolded (U) can be expressed as:

$$S^2 I^k = f_N^k S^2 I_N + f_I^k S^2 I_I + f_U^k S^2 I_U$$

[14]

where f_N^k, f_I^k and f_U^k are fractions of native (0 M urea), intermediate and unfolded (8M urea) species at k M urea solution respectively, and can be expressed as in equations [5a] to [5c]. According to SVD, the scattering curves can be approximated by the linear combination of the basis functions, so that equation [14] can be rewritten as:

$$S^2 I^k = f_N^k (b_1^N U_1 + b_2^N U_2 + b_3^N U_3) + f_I^k (b_1^I U_1 + b_2^I U_2 + b_3^I U_3)$$
$$+ f_U^k (b_1^U U_1 + b_2^U U_2 + b_3^U U_3)$$

[15]

where b_i^j (i = 1, 2, 3, and j denotes N, I, U for native, intermediate, and unfolded states, respectively) are the coefficients for U_1, U_2, and U_3 respectively. On the other hand, the SVD approximation of $S^2 I^k$ is:

$$S^2 I^k = b_1^k U_1 + b_2^k U_2 + b_3^k U_3$$

[16]

where b_1^k, b_2^k and b_3^k are the coefficients for U_1, U_2, and U_3 at k M urea concentration, derived from columns of SV. Since U_1, U_2, U_3 are linearly independent, the following equations are obtained by comparing equations [15] and [16]:

$$b_1^k = f_N^k b_1^N + f_I^k b_1^I + f_U^k b_1^U$$

[17a]

$$b_2^k = f_N^k b_2^N + f_I^k b_2^I + f_U^k b_2^U$$

[17b]

117

$$b_3^k = f_N^k b_3^N + f_I^k b_3^I + f_U^k b_3^U \qquad\qquad [17c]$$

Now, b_1^I, b_2^I and b_3^I, the coefficients of the intermediate, can be calculated by minimizing the following chi-square:

$$\chi^2 = \sum_k \sum_{i=1}^L \left(b_{i,obs}^k - b_{i,cal}^k \right)^2 \qquad\qquad [18]$$

The scattering curves of the intermediate can be constructed from the basis functions (U_1, U_2, U_3) and their corresponding coefficients (b_1^I, b_2^I and b_3^I).

4.4. Results

4.4.1. Sample Integrity

All data were normalized by the total flux of the incident x-ray beam on the cell, followed by background subtraction. To inspect radiation-induced sample aggregation during data collection, scattering curves from all the cycles (1 or 2 minutes per cycle) were overlaid and visually examined. A persistent growth in the scattering intensity was observed in the small angle range with the exposure time > 15 minutes with the monochromatic beam, or > 1 minute with the multilayer beam (~10 times more intense than the monochromatic beam). The native state was found to be more sensitive to the x-ray radiation than the denatured states. Therefore, for the monochromatic beam, data collection time was limited to 10 minutes; for the multilayer beam, in addition to a short data collection time, aluminum foils were used to attenuate the incident x-rays, and more recently, the solution was flowed continuously through the observation cell to minimize exposure of lysozyme to x-rays.

The overlapping scattering curves from different cycles were then averaged, and calibrated. None of the samples showed upward curvature at the smallest angular region available in the Guinier plot, indicating that the samples were largely homogenous and free of associative components (dimer or higher oligomers) in the solution. For the native state, measurements were carried out over a protein concentration series (from 3 mg/l to 44

mg/ml), and the concentration-normalized forward scattering intensity (I(0)/C) was found to decrease reasonably linearly with protein concentration. In our SAXS experiments, lysozyme is routinely used as a standard protein to estimate the molecular weight (M_r) of other proteins studied, such as heat shock proteins (Section 3.4.1.2), according to the following:

$$\frac{\left(I(0)/C\right)_{C\to 0,protein}}{\left(I(0)/C\right)_{C\to 0,lysozyme}} = \frac{MW_{protein}}{MW_{lysozyme}}$$

where $\left(I(0)/C\right)_{C\to 0}$ is the extrapolation of (I(0)/C) to infinitely dilute protein solution.

4.4.2. The Native State

The native state of lysozyme appears to be very stable over the pH range 1.2-11.3 (Imoto, *et al.*, 1972). However, lysozyme undergoes reversible self-association at pH > 4.5 with dimers being the predominant species (Bruzzesi, *et al.*, 1965). pH 3.8 of NaCl-NaOAc buffer has been commonly used to ensure a homogeneous monomeric lysozyme solution.

Fig. 4.3 shows the concentration dependence of R_g for lysozyme at pH 3.8. The extrapolated (to zero protein concentration) R_g (15.4 $\pm$ 0.3 Å) value is within the range of the published results (15 -16 Å) (Imoto, *et al.*, 1972). Although interparticle interference decreases scattering intensity at the small angle region, resulting in a reduced apparent R_g at solutions other than infinite dilution, the effect is not that significant at relatively dilute solutions: a decrease in R_g of ~ 3% at 6 mg/ml or ~4.5% at 8 mg/ml solutions. Because of this, in the equilibrium denaturation and kinetic refolding experiments, described in Chapter 5, measurements were carried out using lysozyme solutions of 6-8 mg/ml to improve data statistics.

The negligible interparticle interference effect at lysozyme concentrations lower than 8 mg/ml can be also seen from Fig. 4.4. At higher protein concentrations, such as 20 mg/ml or 44 mg/ml, a depressive trend is seen at the small angle region, due to the strong interparticle interference. The interference appears not to affect the scattering profile at the angular range $S > 0.015$ Å^{-1}. Hence, 0.015 Å^{-1} was used as the merging point where data
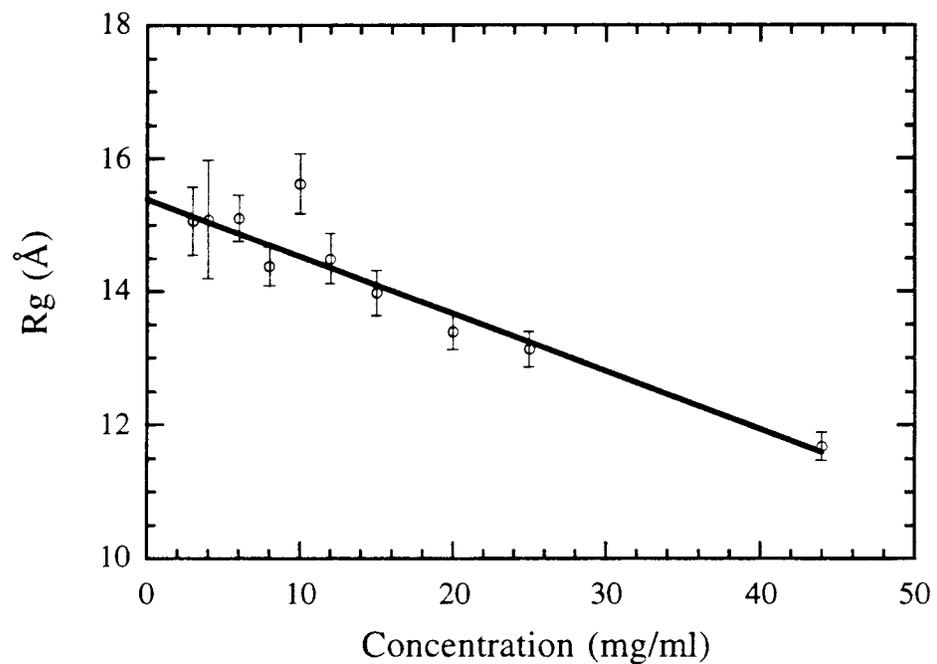
Figure 4.3. Concentration dependence of R_g of the native lysozyme at pH 3.8, and the extrapolation of R_g to zero protein concentration. Multiple measurements were carried out for each lysozyme concentration. The error bars represent the maximum value between the standard deviation of different measurements at the same lysozyme concentration and the fitting errors associated with each measurement derived from the Guinier fitting.
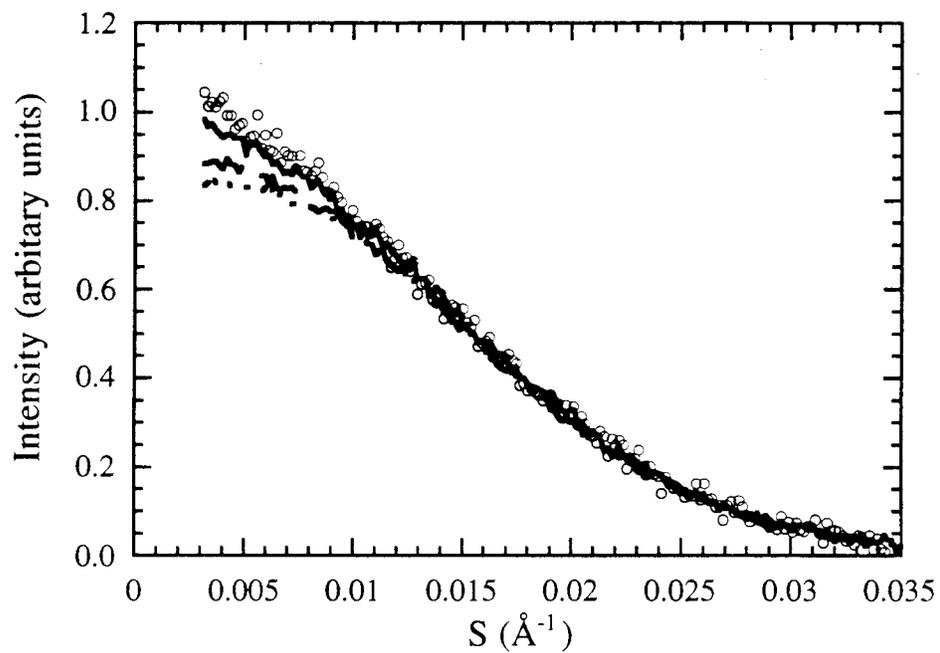
Figure 4.4. Concentration-normalized background-subtracted SAXS data for the native lysozyme at concentrations of 4 mg/ml (o), 8 mg/ml (—), 20 mg/ml (— —) and 44 mg/ml (- -). Interparticle interference effect is not significant below 8 mg/ml. All four scattering curves overlap at angular range $S > 0.015 \text{Å}^{-1}$, suggesting that interparticle effect can be neglected in the wide angle region.

of the low angular region from the dilute sample (4 mg/ml) overlapped with data of the wide angular region from concentrated solutions (20 mg/ml or 44 mg/ml), to calculate the P(r) function for the native state.

Fig. 4.5 shows the P(r) functions calculated from the x-ray solution scattering data and from crystal structure coordinates. The shape of the P(r) function computed from coordinates deviates substantially from that derived from the experiemental data: the P(r) function shifts to smaller distances with the peak positioned at ~17 Å, about 1 Å shorter than the experimental P(r) function. The maximum dimension of lysozyme, D_{max}, where P(r) reaches zero, in the crystal (42 $\pm$ 2 Å) is also smaller than that in solution (50 $\pm$ 2 Å). When a hydration layer of 3 Å was added around lysozyme, the calculated P(r) function became superimposable with the experimental P(r) function. A hydration layer was also found necessary for the ATPase domain of hsc70 (Section 3.4.3.1) to match the calculated P(r) function with that from x-ray solution scattering. The scattering curve thus calculated is in good agreement with the experimental data, as shown in the Kratky plots, Fig. 4.6. The well-defined peak in the Kratky plot indicates that the native lysozyme is a globular molecule in both solution and crystalline forms. Hence, by x-ray scattering technique, lysozyme seems to retain similar conformation in solution and in crystalline form.

Due to its unusual stability, unfolding of lysozyme by urea has to be conducted in reduced pH solutions. In our equilibrium denaturation and kinetic refolding studies, pH was reduced to pH 2.9 or pH 2.0, with Na-citrate ($NaC_6H_7O_6$) and Na-phosphate (NaH_2PO_4) replacing the NaOAc buffer, respectively. Interestingly, at similar lysozyme concentrations, R_g of lysozyme calculated from these three solutions showed some variations: 15.1 $\pm$ 0.4 Å for 6.0 mg/ml at pH 3.8, 15.4 $\pm$ 0.2 Å for 6.8 mg/ml at pH 2.9, and 14.2 $\pm$ 0.3 Å for 7.3 mg/ml at pH 2.0. Different counter ions (OAc^-, $H_2PO_4^-$, $C_6H_7O_6^-$) may be attributed to the observed variations in R_g; in acidic solution, lysozyme carries a net postive charge, for example, a net charge of +18 at pH 2.5 (Bruzzesi, *et al.*, 1965). Batch-to-batch difference of lysozyme sample can not be excluded. The overall scattering curves of these three solutions (Fig. 4.7) overlapped well, suggesting that the conformation of lysozyme in these three solutions was esentially the same, as judged by x-ray scattering.
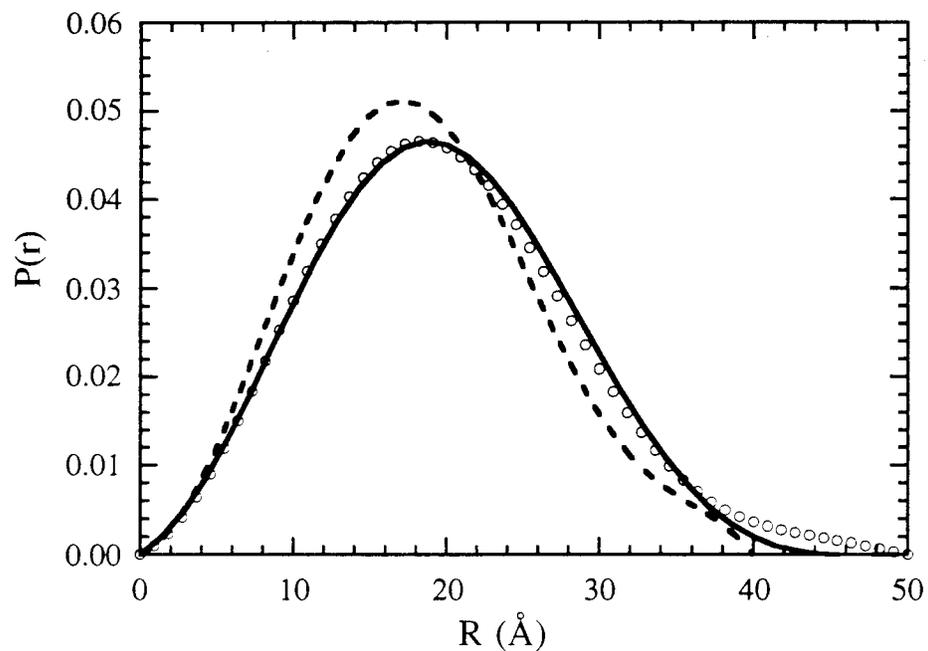
Figure 4.5. Comparisons between the experimental and the calculated P(r) functions of the native lysozyme. (o), derived from SAXS data; (- -), computed from crystal coordinates; (—), computed from crystal coordinates with a 3 Å hydration layer. To calculate the experimental P(r), data of the low angle region (S < 0.015 Å^{-1}) from a 4 mg/ml sample were combined with data of the wide angle region (S > 0.015 Å^{-1}) from a 44 mg/ml sample.
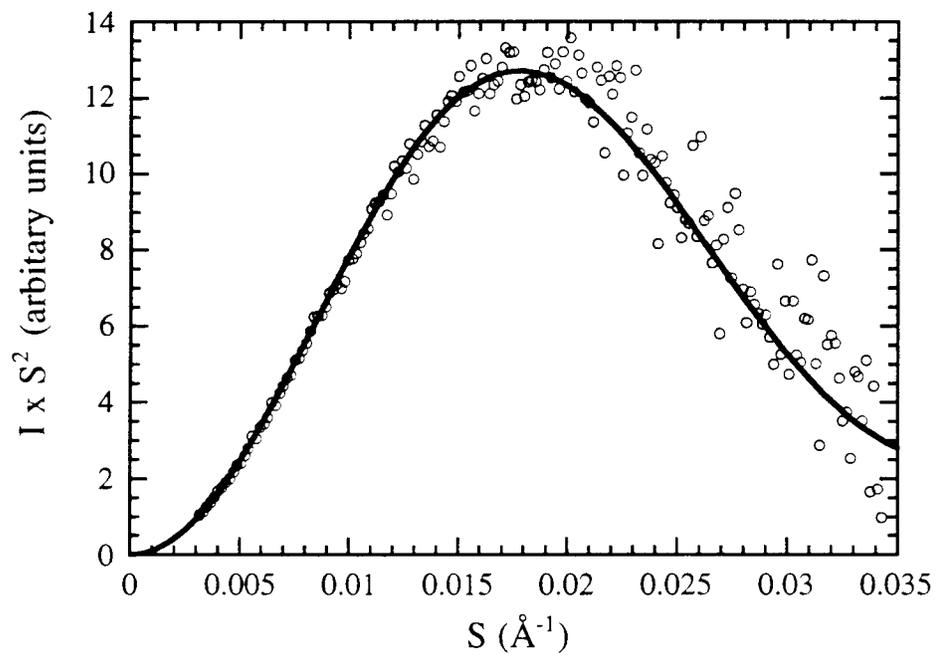
Figure 4.6. Kratky plots of the experimental data of native lysozyme (o), and calculated from crystal structure coordinates with a 3 Å hydration layer added (—). The data set shown is from the sample containing 4 mg/ml lysozyme.
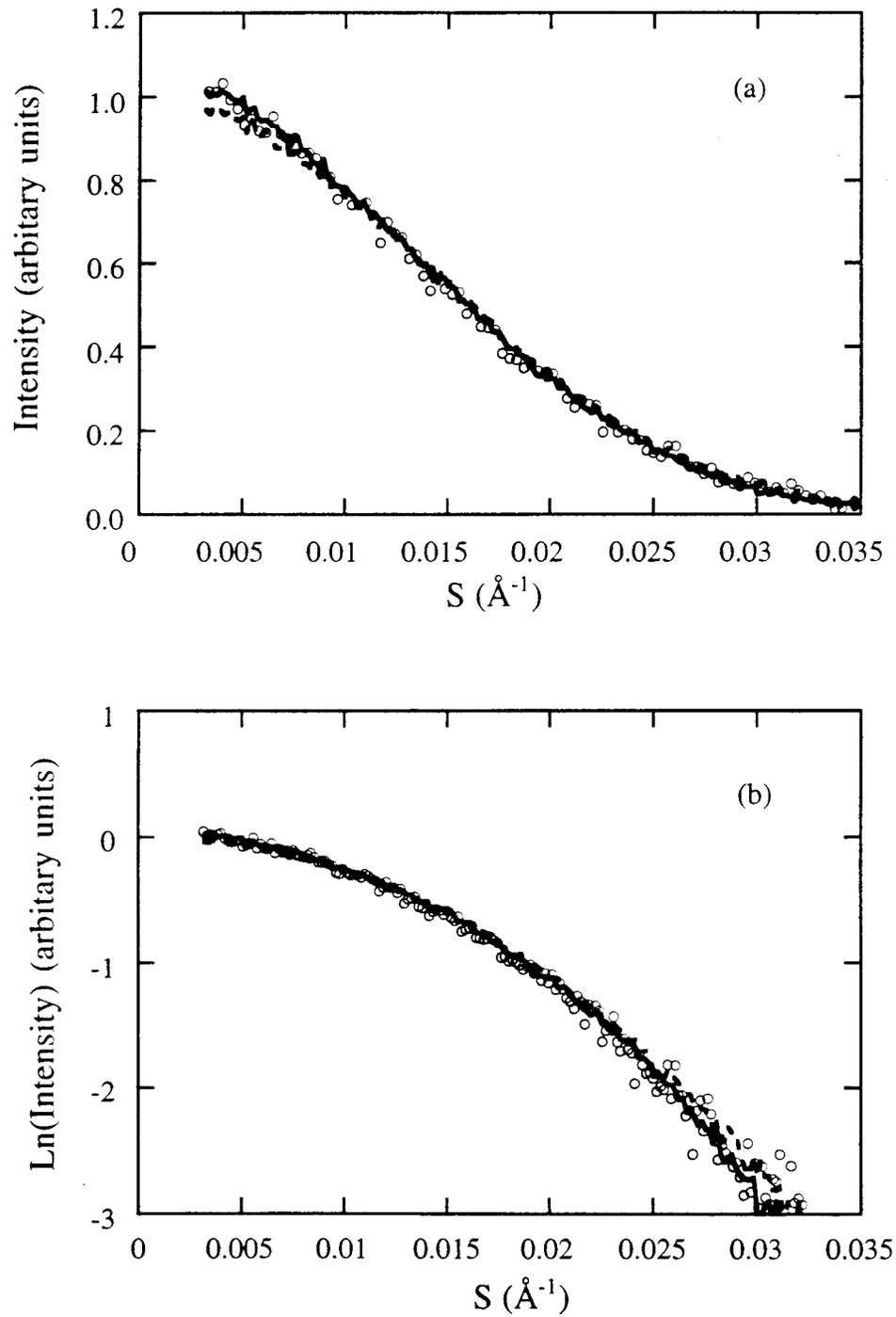
Figure 4.7. Comparisons of scattering curves of the native lysozyme at pH 3.8 (o), pH 2.9 (—) and pH 2.0 (– –). (a), in I(S) *vs.* S plot; (b) in LnI(S) *vs.* S plot. The differences between these three conditions are negligible.

4.4.3. The Unfolded State

The unfolded states of lysozyme were studied under several conditions: 8 M urea at pH 2.9, 8 M urea at pH 2.0, 8 M urea and 80 mM reductant, dithiothreitol (DTT) at pH 2.9, and 5 or 6 M GdmHCl at pH 2.9. Suprisingly, conformations of lysozyme in different nonreducing solutions appeared to be similar. For all three conditions (8 M urea pH 2.0, 8 M urea pH 2.0 and 6 M GdmHCl pH 2.9), R_g was found to be around 22 Å, and scattering profiles were superimposible (data not shown). An addition of reductant, DTT, decreased the compactness of unfolded lysozyme by an increase of ~ 2 Å in R_g resulting in a R_g of ~ 24 Å.

Heparin, a sulfated mucopolysaccharide, was used as a model for a random coil (Liberti & Stivala, 1967). Heparin is a particularly good model for comparison with lysozyme because it has a similar molecular weight (~13k) to that of lysozyme. Heparin has an R_g of 31.3 ± 0.2 Å by SAXS (Fig. 4.8), much larger than that of unfolded lysozyme, even in the dithiothreitol reduced form. The Kratky plot of heparin (Fig. 4.9.a) shows a plateau at higher angle region, which is a signature of a random coil. However, the Kratky plot of lysozyme in 8 M urea solution appears peaky, suggesting that lysozyme molecules in 8 M urea do not resemble random coils. This is plausible from a structural point of view. Lysozyme has four disulfide bonds distributed over the polypeptide chain (6-127, 30-115, 64-80, 76-94), and they remain intact under the conditions used in our present study, except for solution containing DTT. Structural constraints imposed by these four disulfide bonds restrict the freedom of polypeptide chain movement, resulting in a more compact ensemble of configuration than that expected for a random coil.

The distance distribution (P(r)) function for the unfolded state in 8 M urea without reductants is shown in Fig. 4.9.b. D_{max} is enlarged to ~ 70 Å, and the peak position is shifted to a larger distance, ~ 25 Å, upon unfolding in 8 M urea. As a comparison, D_{max} for heparin is ~ 120 Å (data not shown). The rather smaller D_{max} again suggests that lysozyme molecules in 8 M urea have some internal structure, and cannot be approximated as a random coil.

126
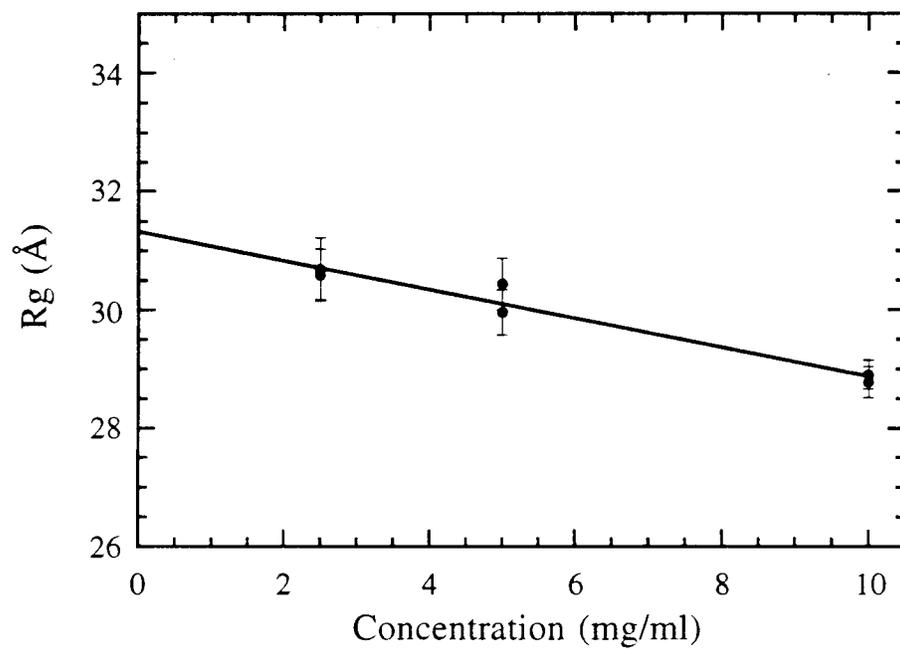
Figure 4.8. Concentration dependence of Rg of heparin at pH 2.5, and the extrapolation of R_g to zero heparin concentration.
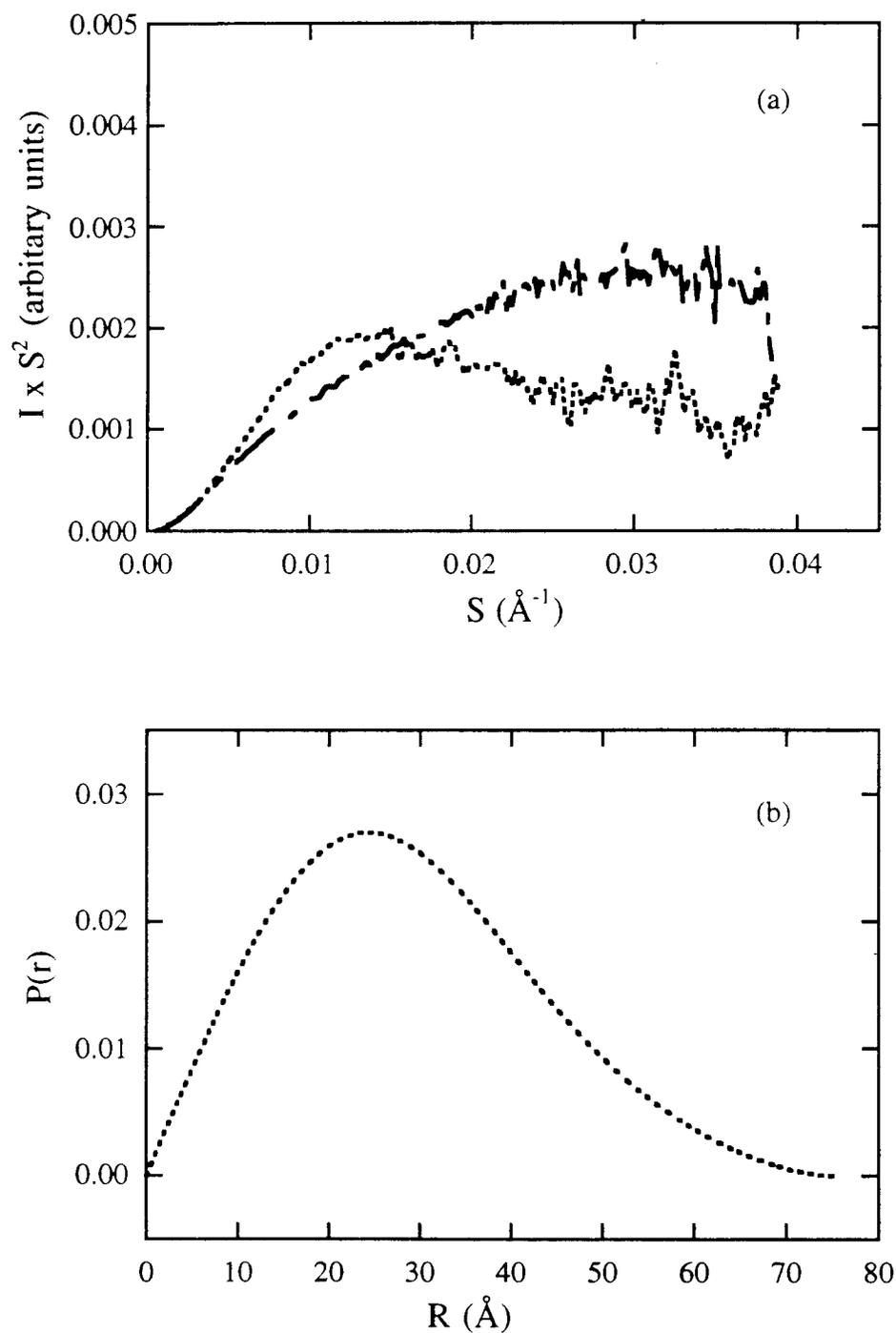
Figure 4.9. a) Kratky plots of a random coil model, heparin, (— – —) and lysozyme in 8 M urea solution pH 2.9 (– –). b) The P(r) function of lysozyme in 8 M urea pH 2.9. The P(r) function from the pH 2.0 sample overlaps with that of pH 2.9; data of the 6 M GdmHCl sample is too noisy to compute a reliable P(r) function.

4.4.4. Unfolding by Urea at pH 2.9

Like in the native state, a depressive trend at the small angle region was not seen in the Guinier plot, suggesting that the interparticle interference of lysozyme at ~7 mg/ml over the urea concentations tested was insignificant. The forward scattering intensity is proportional to the squares of both the electronic contrast between the protein molecule and the buffer background and the partial specific volume of the protein: $I(0) \propto (\rho_p - \rho_b)^2 \times V^2$. Since the difference in partial specific volume of lysozyme in 8 M urea (0.707 ml/g) versus in the native state (0.702 ml/g) is negligible (Prakash, *et al.*, 1981), I(0) is thus proportional to $(\rho_p - \rho_b)^2$. Fig. 4.10 shows the quadratic decrease in I(0) with the increase of urea concentration. After taking into account the urea effect and normalizing to protein concentration, I(0) values appear to be rather constant throughout the urea concentrations tested. The similar I(0) indicates that lysozyme remained unaggregated throughout the folded and unfolded transition.

Unfolding curves determined by amide CD and R_g. Fig. 4.11.a shows the change of apparent unfolded fraction (F_{app}) of lysozyme with urea concentration at pH 2.9 by CD and scattering measurements. The apparent fraction is defined as $F_{app} = (Y_{obs} - Y_N)/(Y_U - Y_N)$, where Y_{obs} is the experimental value of the physical parameter, $[\theta]_{222}$ or R_g, at certain urea concentration; Y_N and Y_U are values of the parameters in the native and unfolded form respectively. For pH 2.9, the transition in the CD is sharp, suggesting a highly cooperative unfolding process, while the transition in R_g is more gradual and thus less cooperative. It should be noted that the results of different unfolding behavior observed by amide CD and R_g have been reproduced in several independent experimental runs, ruling out the possibility of artifacts. In addition, the difference is clearly resolved at pH 2.0 (see Fig. 4.11.b). Due to limited amounts of beamtime, data at pH 2.0 were not as complete as those at pH 2.9, particularly in the post-transition region, resulting in less conclusive results. Therefore, the following analysis and discussion are focused on data at pH 2.9, and are supported by results at pH 2.0. Different unfolding behavior revealed by these two techniques are due to the fact that they probe different aspects of protein structure: amide CD at 222 nm mainly reflects the helical content, while R_g measures the size, and hence compactness of the molecule. The differences in the unfolding curves measured by different techniques suggest the presence of intermediate states. The observation that a structural transition measured by R_g occurs at lower urea concentration than that by CD, in conjunction with the assumption that α-helical

129
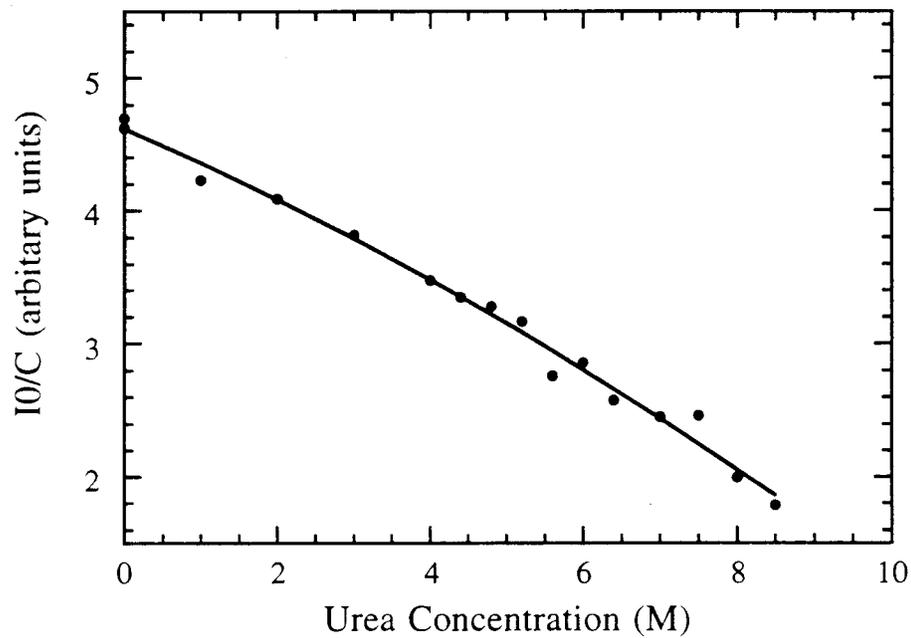
Figure 4.10. The concentration-normalized forward x-ray scattering intensity (I(0)/C) of lysozyme as a function of urea concentration at pH 2.9. The line is a second order polynomial fit to the data.
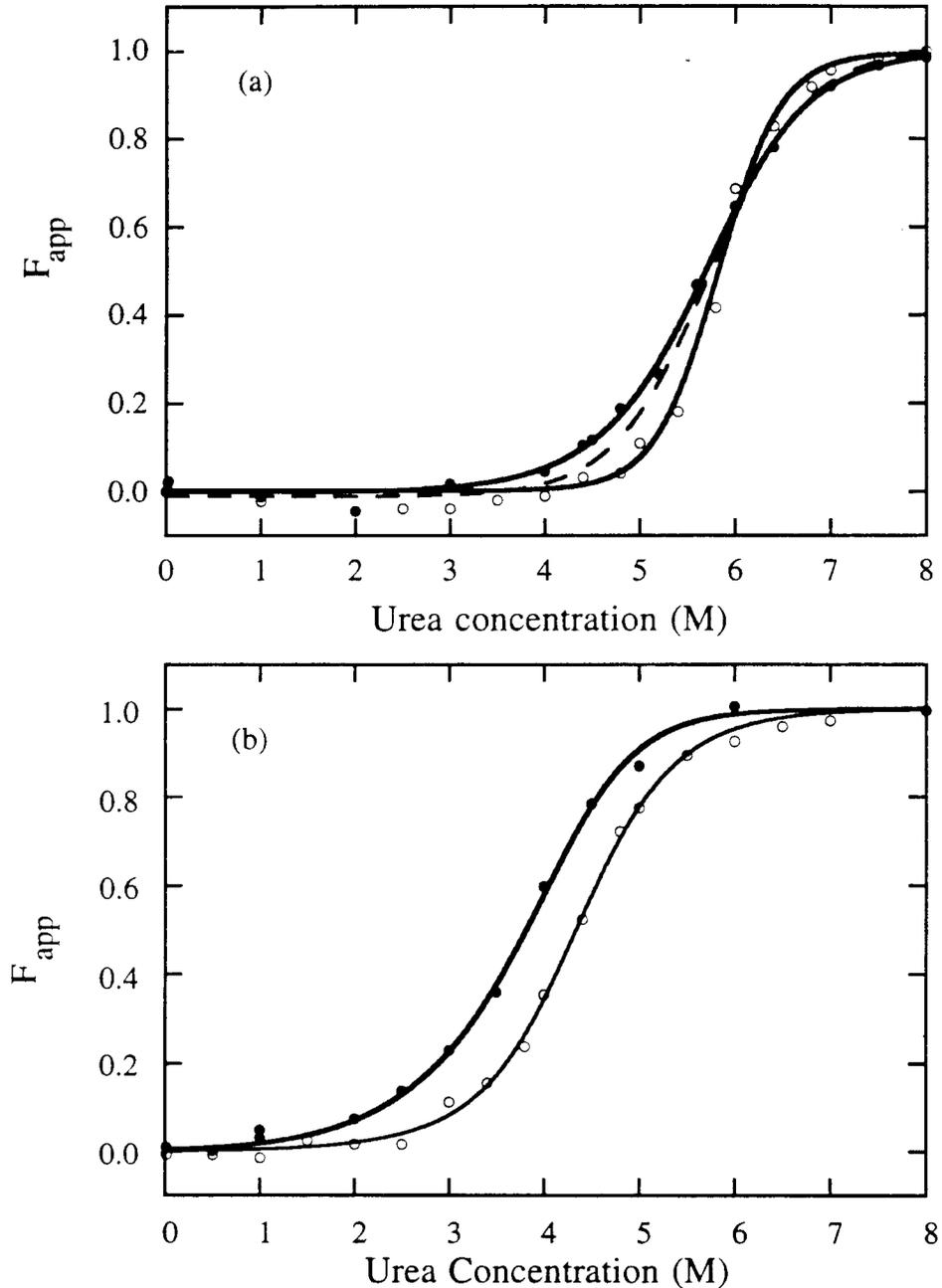
Figure 4.11. Changes of the apparent fraction, F_{app} (see text for definition), in the urea-induced lysozyme unfolding at 20 °C. (•), by R_g from x-ray scattering; (o), by amide CD θ_{222}. (a) pH 2.9. Solid lines are fitting results of scattering and CD data according to model 2 (see text); Dashed lines are the fits to model 1 (see text). The curves from the two sets of fits are almost identical and hence the two dashed lines overlay each other. (b) pH 2.0. Solid lines are preliminary fitting results according to model 2.

131

secondary structure persists in the intermediate state, leads to a working model that the intermediate has an extended R_g but retains native-like α helical content.

To estimate the thermodynamic parameters associated with unfolding process, a denaturant binding model (see Section 4.3.3) was used to account for the effect of urea on the equilibrium constants. A range of different urea binding constants (c) 0.1, 0.2, 0.4, 0.6, 0.8, 1.0 was tried, and it was found that the fractions of native, intermediate and unfolded states were not affected (within 1%) by the c value. The results presented here were from fits with c set to 0.1 (Pace, 1975). Initially, parameters for native, unfolding and intermediate states (R_g and θ), as well as thermodynamic parameters were not fixed during the fitting. After one round of nonlinear least squares fits, parameters (R_g and θ) of the native and unfolded states were fixed, the other parameters ($\Delta G_I^{H_2O}$, $\Delta G_U^{H_2O}$, Δn_I, Δn_U, R_g^I) were further refined. The refined parameters for the native and the unfolded states were found to be the same as values measured in 0 M and 8 M urea solutions, respectively. Data were simultaneously fit to two models: (1) both transitions in the CD and R_g curves taken as two-state; (2) the CD transition assumed to be two-state while the R_g transition as three-state. The fits are shown in Fig. 4.11.a, and the results are listed in Table 4.1. It is apparent from the figure that the best fit for both CD and R_g data is achieved with model (2). R_g for the intermediate state is found to be ~19.8 Å, as compared to 15.3 Å for the native and 21.9 Å for the unfolded.

Fractions of native, intermediate and unfolded states in different urea concentrations can be calculated by equations [5a-c]. Fig. 4.12 shows the fraction distributions based on R_g. A maximum for the intermediate distribution is found at 5.5-6 M urea concentration, where ~ 40% of the lysozyme molecules are partially unfolded. The unfolded species does not appear below ~ 5 M urea.

SVD analysis. To test the robustness of the SVD method in our studies, several circumstances have been inspected. Firstly, scattering data were also collected at pH 2.0 with the same camera length of $S_{max} = 0.04$ Å^{-1}, as well as at pH 2.9 but with an extended angular range of $S_{max} = 0.08$ Å^{-1}. Unfortunately, because of time constraints, experiments under these two situations were not performed in as many urea concentrations as in the case of pH 2.9 with $S_{max} = 0.04$ Å^{-1}. Nevertheless, three significant basis functions were always observed. Secondly, when a subset of scattering curves (six scattering curves out of a total of 13 curves) at some urea concentrations were selected and subjected to SVD analysis, a third basis function remained evident. Finally, SVD analysis was also carried

Table 4.1. Parameters for simultaneous fitting model 2 (2&3 state) and model 1 (2&2 state) to the CD (222 nm) and R_g unfolding curves in Fig. 4.11.a.

Model	$\Delta G_{N \to I}$ (kal/mol)	$\Delta n_{N \to I}$	$\Delta G_{N \to U}$ (kcal/mol)	$\Delta n_{N \to U}$	Native	Intermediate	Unfolded
					R_g (Å)		
model (2)	5.7 ± 0.3	23.6 ± 1.5	11.6 ± 0.5	48.8 ± 2.1	15.3^a	19.8 ± 0.2^a	21.9^a
					15.3 ± 0.2^b	N.A.	21.8 ± 0.3^b
					θ_{222} (mdeg cm^2 dmol^{-1} res^{-1})		
					-8991^a	N.A.	-2672^a
					-8979^b	N.A.	-2660^b
					R_g (Å)		
model (1)	N.A.	N.A.	8.0 ± 0.2	32.8 ± 0.8	15.5^a	N.A.	21.8^a
					15.3 ± 0.2^b	N.A.	21.8 ± 0.3^b
					θ_{222} (mdeg cm^2 dmol^{-1} res^{-1})		
					-9173^a	N.A.	-2538^a
					-8979^b		-2660^b

[a] used in the denaturation curve fitting.

[b] derived from x-ray scattering or CD measurements.
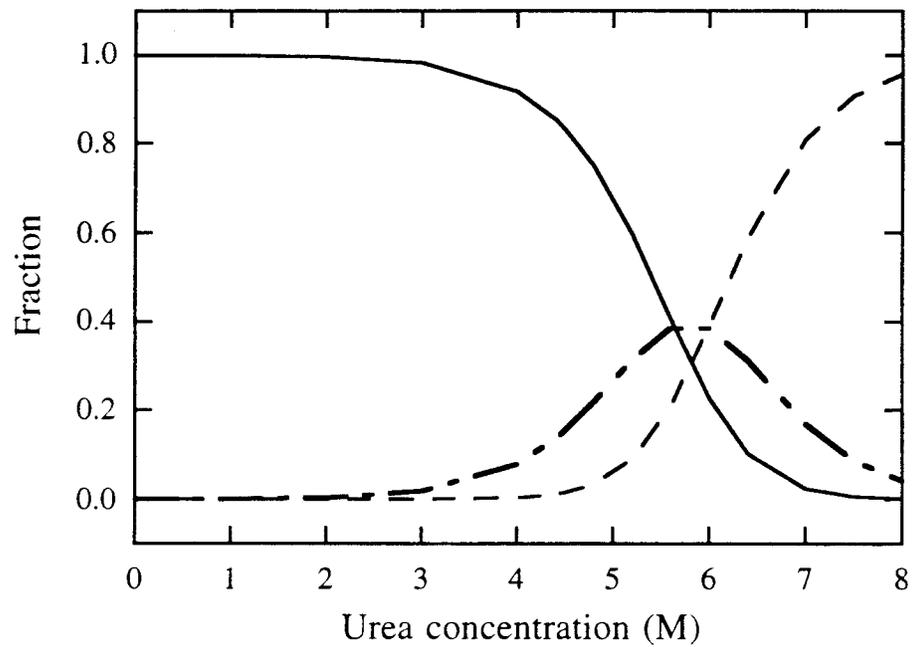
Figure 4.12. Population of native (—), intermediate (— - —) and unfolded states (– –) of lysozyme at pH 2.9 as a function of urea concentration as determined by R_g analysis.

out over raw data sets without smoothing and over data sets with 5-point smoothing. The results were the same as in the case of 3-point smoothing but with different noise levels.

The global reduced χ^2 values with L component approximation of Kratky data at pH 2.9 are 94, 16, and 1.14 for L=1, 2, 3, respectively. A third component is also found necessary to get reduced χ^2 to be around unity at pH 2.0. A typical three-component approximation and its residuals are shown in Fig. 4.13. The three basis functions weighted by their singular values are presented in Fig. 4.14. Although basis functions are not scattering patterns of specific molecules, some correlations are apparent. For example, the first basis function resembles characteristics of a globular protein, and can be related to scattering of the native state. The second and third basis functions are complex, yet they have certain features in the high angle region that are similar to scattering of random coils or the unfolded state. The observation of three significant scattering basis functions suggests a minimum of three unique scattering species; that is, there is an intermediate in addition to the native and unfolded forms.

To estimate b_1^I, b_2^I and b_3^I and construct the scattering curve for the isolated intermediate, three columns of V weighted by their singular values were simultaneously fit according to equation [17] with a set of thermodynamic parameters $\Delta G_I^{H_2O}$, $\Delta G_U^{H_2O}$, Δn_I and Δn_U. The best fits were achieved where $\Delta G_U^{H_2O}$ and Δn_U were in good agreement with those from denaturation curve analysis, although $\Delta G_I^{H_2O}$ and Δn_I were only half of the values derived from denaturation curve analysis. It should be noted that plots of b_i (i=1,2,3) versus urea concentrations (Fig. 4.15) are rather noisy, particularly the plot of b_3. On the other hand, both the R_g and CD denaturation curves (Fig. 4.11.a) are smooth, therefore parameters derived from them are considered to be accurate. Hence, data were fitted in such a way that the relevant thermodynamic parameters, $\Delta G_I^{H_2O}$, $\Delta G_U^{H_2O}$, Δn_I and Δn_U, were chosen to be close to those from the denaturation curve analysis. The fits for b_1 and b_2 were reasonable, whereas the result for b_3 was clearly less well defined. Again, the scattering data quality available at the present time may not be adequate for a reliable interpretation of the lysozyme unfolding mechanism taken on its own, but by considering it along with the analysis of the denaturation curves, the reliability of the fits is considerably increased.

A scattering curve for the intermediate was synthesized using the estimated b_1^I, b_2^I and b_3^I with equation [16]. The synthesized Kratky plot, shown in Fig. 4.16.a, appears to be a superposition of compact and random coil features (Doniach, *et al.*, 1995), and
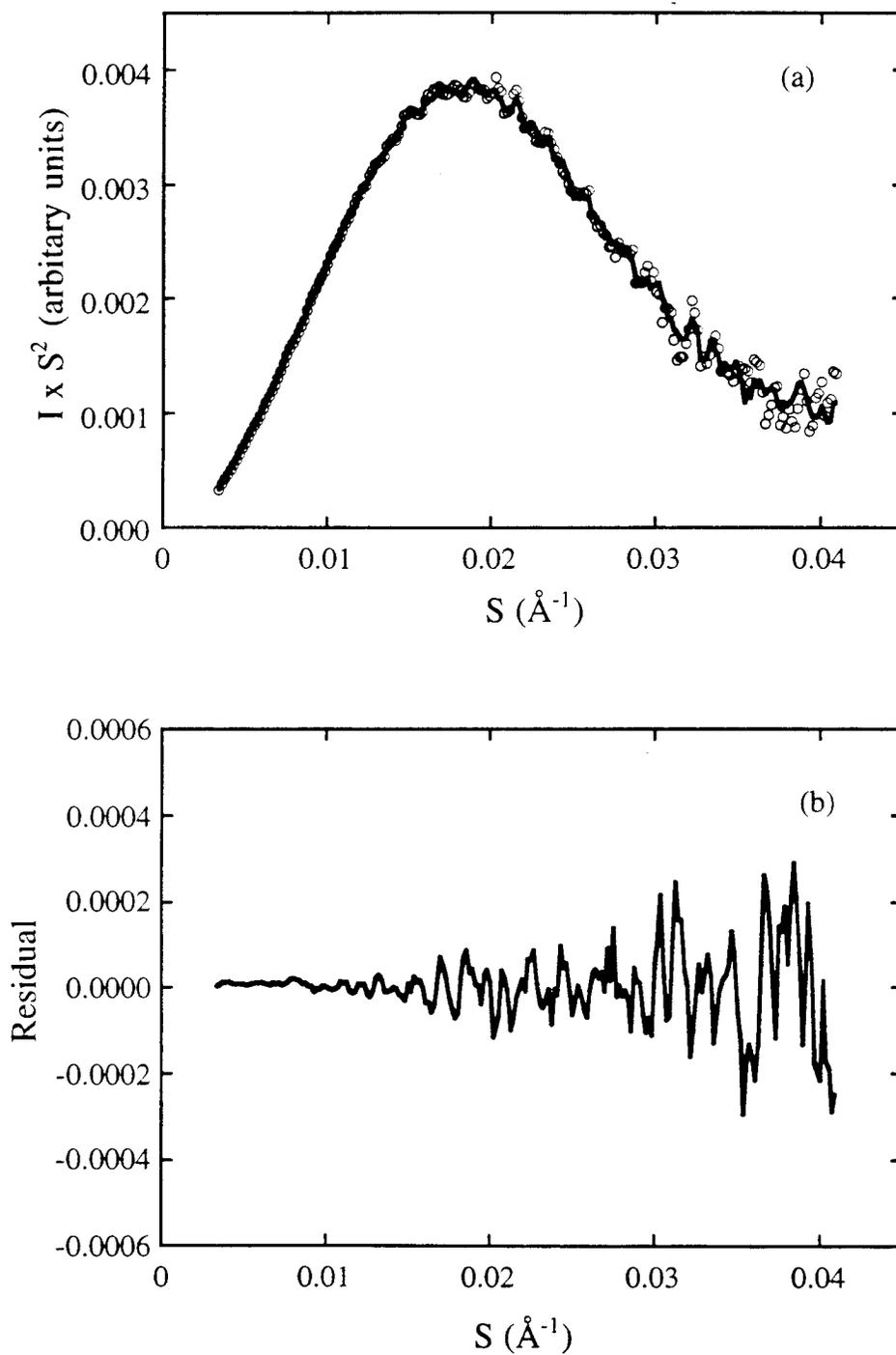
Figure 4.13. a) Representative Kratky plots of the data (o) and its three-component SVD approximation (—). The urea concentration is 4 M. b) The residuals, or differences between the experimental data and the calculation using three components.
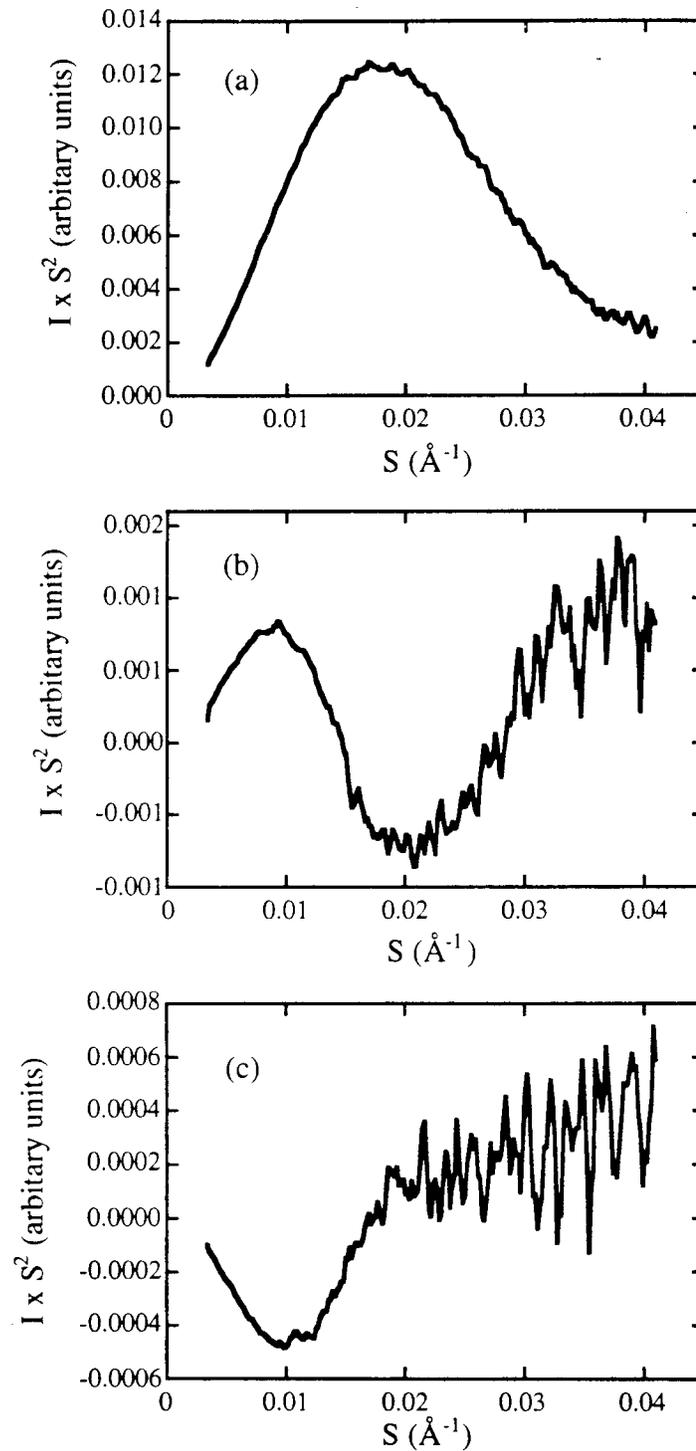
Figure 4.14. The orthogonal basis scattering functions (columns of matrix U) in Kratky form from SVD analysis. The curves are weighted by their singular values: a) the first function U1 by 0.118; b) the second function U2 by 0.004; c) the third function U3 by 0.002.
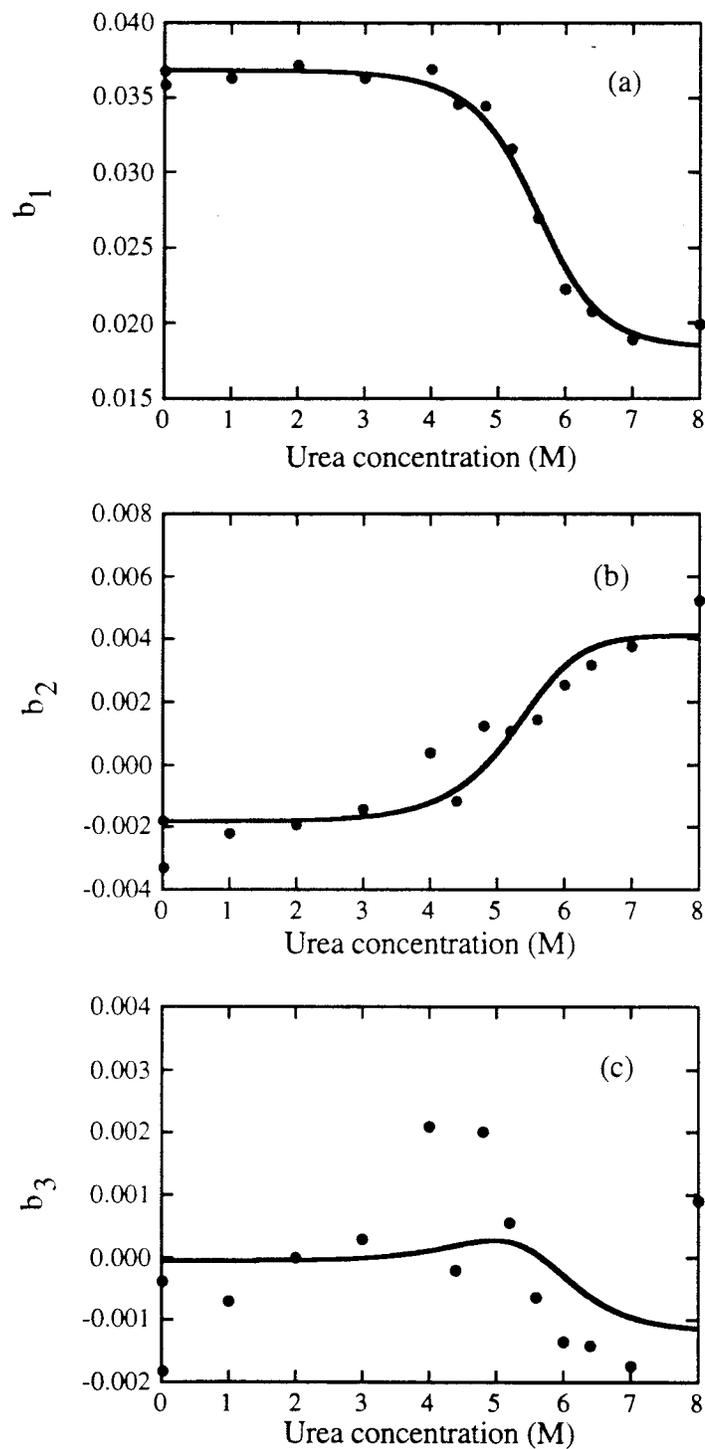
Figure 4.15. Singular value-weighted urea dependence of basis functions a) U1; b) U2; c) U3. The solid lines are results of simultaneous nonlinear least-squares fits with equation [17].
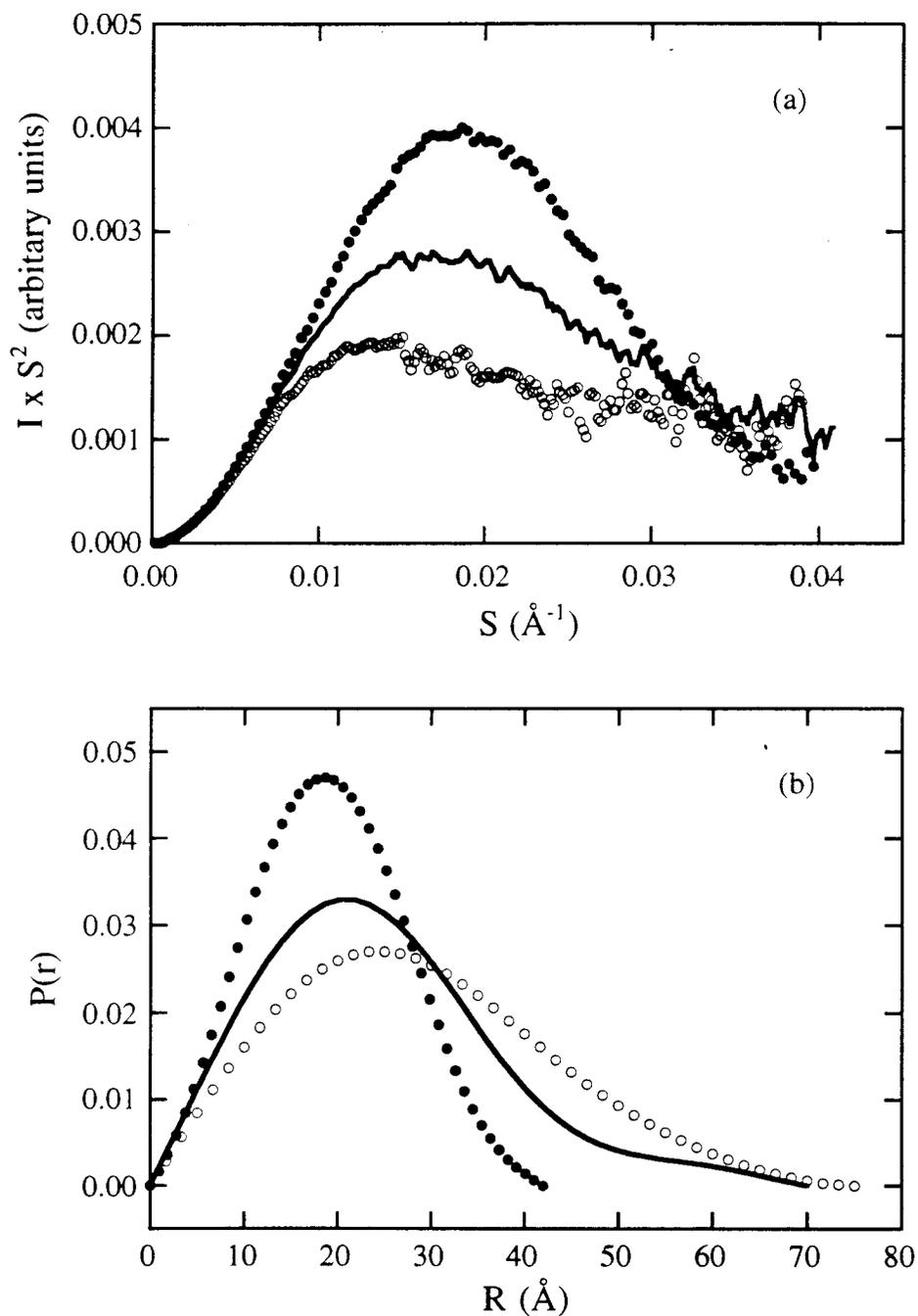
Figure 4.16. (a) The Kratky plot of a three-component SVD construction of the intermediate state (—). (b) The P(r) function for the synthesized intermediate state shown in (a). Data from both the native state (•), and the unfolded state in 8M urea pH 2.9 (o) are also included for comparison. Data shown in the Kratky plots were after with 3-point smoothing.

suggests that the globularity of the intermediate is closer to that of the native rather than to the unfolded state in 8 M urea. Fig. 4.16.b shows the P(r) function calculated from this constructed scattering curve. The peak position, ~20 Å, is close to that of the native, while the maximum distance of the intermediate, D_{max}, has extended to that of the unfolded state (~70 Å). The features in the P(r) function of the intermediate indicate again that the conformation of the intermediate is midway between those of the native and the unfolded states. The R_g for this intermediate is ~19 Å from the Guinier approximation, in good agreement with ~20 Å from the R_g denaturation curve analysis.

Structural models of the intermediate state. A kinetic intermediate has been identified in the refolding pathway of lysozyme which consists of a compact α domain showing H-D exchange protection factors of a molten globule character and a highly disordered β domain (Radford, *et al.*, 1992).

It is of interest to try and estimate whether our result of $R_g \sim 19\text{-}20$ Å is compatible with this structural model. To do this one can use the fact that the observed R_g may be decomposed into contributions from disjoint fragments of the protein. For simplicity each atom (not including hydrogens) in the protein is assumed to have the same scattering factor (this could easily be generalized to include more accurate estimates). One can write

$$R_g^2 = \frac{1}{n}\sum_i (\bar{r}_i - \bar{r}_{cm})^2$$

where n is the total number of atoms, with positions $\bar{r}_i$, and $\bar{r}_{cm}$ is the center of mass coordinate vector of the protein. Then for any disjoint subsets A and B, with n_a and n_b atoms

$$R_g^2 = \frac{1}{n}\{\sum_A (\bar{r}_i - \bar{r}_{cA})^2 + n_A(r_{cA} - r_{cm})^2 + \sum_B (\bar{r}_i - \bar{r}_{cB})^2 + n_B(r_{cB} - r_{cm})^2\}$$

$$= \frac{n_A}{n}\{R_{g,A}^2 + (r_{cA} - r_{cm})^2\} + \frac{n_B}{n}\{R_{g,B}^2 + (r_{cB} - r_{cm})^2\}$$

where $\bar{r}_{cA}$ is the center of mass of subset A and similarly for B. (This generalizes to an arbitrary number of disjoint subsets.) Using the atomic coordinates for native lysozyme these quantities have been evaluated for 3 subsets: $\alpha 1$ (residues 1-35), β (35-76) and $\alpha 2$ (76-129). The sets were divided at the α carbons. The resulting values are given in Table 4.2. In this approximation (for the scattering factors) the total R_g comes out to 13.9 Å,

somewhat smaller than 15.3 Å from solution scattering. A smaller computed R_g is expected because, in addition to the approximate nature of the calculation, the measured R_g from solution scattering includes the effects of the hydration layers formed around the lysozyme molecules in solution.

In order to estimate R_g for the intermediate partially folded state in the refolding pathway (Radford, *et al.*, 1992), one needs a model for an unfolded conformation of the β region. To do this realistically is a complex task which has not yet been done for a protein domain in contact with the rest of the protein in its native conformation. Therefore selected highly simplified models of the β domain were examined in order to get an estimate of the range of possible values of $R_{g,\beta}$ and $\left| \bar{r}_{c,\beta} - \bar{r}_{cm} \right|$ for the unfolded domain. To do this, the backbone was broken up into 10 segments of about 4 residues each. The idea here was that these fragments represent roughly the persistence length appropriate for disordering the polymer chain (Damaschun, *et al.*, 1991). The sum of the length of the fragments is found to be 104 Å. An extreme overestimate to the radius of gyration is obtained by taking an extended loop from residue $(C\alpha)_{35}$ to $(C\alpha)_{76}$. Taking this in the form of a circle oriented away from the protein center of mass it was found that $(R_{g,\beta}^{circle})^2 \cong 324$ Å^2 and $(\bar{r}_{c,\beta}^{circle} - \bar{r}_{cm})^2 \cong 533$ Å^2. This gives an upper boundary on R_g for the whole protein of $R_g^{total} < 20.4$ Å, which may correspond to 21-22 Å when considering the hydration factor. A second, somewhat more compact but still extended, model was taken to consist of a linear arrangement of four fragments (each 10.4 Å) oriented away from the center of mass then returning with an extra fold in the form of a zig-zag. This conformation gives $(R_{g,\beta}^{linear})^2 \cong 159$ Å and $(\bar{r}_{c,\beta}^{circle} - \bar{r}_{cm})^2 \cong 460$ Å. The resulting R_g estimate is R_g^{total} = 18.3 Å, which again may increase to 19-20 Å when a hydration layer is taken into account. Since even the second model is probably more extended than would be expected for a realistic random coil conformation of the β subset, it is seen that the observed value for R_g of the intermediate, ~ 19-20 Å, falls close to the upper end of the range of possible values for a model in which the $\alpha 1$, $\alpha 2$ subunits retain their native values. It is therefore concluded that the $\alpha 1$ and $\alpha 2$ components of the protein are likely to be somewhat disordered in the urea-induced intermediate state even though the CD evidence suggests that the α helical component of the intermediate is very close to that of the native. This may imply some disordering of the turns linking the α helical structures with some accompanying swelling of the $\alpha 1$, $\alpha 2$ components.

Table 4.2. Fragments of lysozyme, and their radii of gyration.

Subset	$\alpha 1$ (1-35)[a]	β (35-76)[a]	$\alpha 2$ (76-129)[a]
number of atoms	270	356	368
subset R_g (Å)[b]	9.8	10.4	13.2
$\left\| \vec{r}_{c,subset} - \vec{r}_{cm} \right\|^2$ (Å^2)[b]	69.0	109.0	19.1

[a] residue number.

[b] see text for discussion of how this value was determined.

4.5. Discussion

Our results from the x-ray solution scattering study provide evidence for a thermodynamically stable intermediate in the equilibrium unfolding of hen egg-white lysozyme by urea at pH 2.9. Three significant basis functions from the SVD analysis of the scattering curves (Kratky plots) suggest a minimum of three distinct scattering components in the urea-induced unfolding of hen lysozyme at pH 2.9. A third species is also implicated from the lack of coincidence of the denaturation curves measured by R_g and the CD signals at 222 nm; the transition in CD appears to be highly cooperative, whereas the cooperativity revealed by R_g tends to be less pronounced. The intermediate is assumed to have a similar CD signal at 222 nm to that of the native state. The radius of gyration of the intermediate is found to be around 19-20 Å, and a scattering curve for this intermediate is constructed.

Structures of native and unfolded states. The stability of hen egg-white lysozyme is remarkable. No significant conformational change has been observed over pH 1.2-11.3 in the absence of denaturant (Imoto, et al., 1972). The radius of gyration of lysozyme measured at pH 2.9, 15.3 Å, falls within the reported values of 14.3 to 16.0 Å (Imoto, et al., 1972). The P(r) function of lysozyme in pH 2.9 solution also agrees well with that computed from the crystal structure, so that within the resolution of the x-ray scattering data, the conformations of lysozyme in solution of pH 2.9 and in the crystalline form are essentially the same. Similar results are derived at pH 2.0. It is concluded therefore that lysozyme molecules exist in the native state under the conditions studied.

Thermodynamic standard functions of lysozyme over temperatures 0-100 $^{\circ}$C and pH 2-7 have been reported (Pfeil & Privalov, 1976), and the conformational stability of lysozyme ($\Delta G_U^{H_2O}$) can be calculated accordingly. At pH 3.0 and 20 $^{\circ}$C, $\Delta G_U^{H_2O}$ is ~ 12 kcal/mol by calorimetric measurements. Estimates of $\Delta G_U^{H_2O}$ from denaturation curve analysis, however, range from 6 kcal/mol to 14 kcal/mol in pH 3.0 to 6 at 25 $^{\circ}$C, depending on the method employed to extrapolate ΔG_U to zero denaturant concentration. Nevertheless, these values are reasonable for an estimate of 10-20 kcal/mol for small globular proteins. By using the denaturant binding model (which normally gives a higher estimate for $\Delta G_U^{H_2O}$ but which provides a better picture of the denaturant effect among other methods (Pace, 1986)), values of 11.6 ($\pm$ 0.5) kcal/mol and 8 ($\pm$ 0.2) mol/kcal are derived from 2&3-state analysis and 2&2-state analysis, respectively. A difference in

$\Delta G_U^{H_2O}$ is reported for the α subunit of trytophan synthase with ~ 12 kcal/mol from 3-state analysis and ~ 3.6 kcal/mol from 2-state analysis (Matthews & Crisanti, 1981). Although both analysis models yield a good estimate of $\Delta G_U^{H_2O}$ for lysozyme, SVD of the scattering curves supports the existence of a distinct scattering intermediate.

In the denaturant binding model, the protein unfolds as a result of an increase in urea binding sites in the unfolded state compared with the native state. In 8 M urea solution, a value of 77 mol urea per mol lysozyme is calculated by assuming two peptide bonds interacting with one urea molecule and one aromatic residue including histidine associating with one urea molecule (Prakash, et al., 1981), while 65 mol/mol is derived from partial specific volume measurements (Prakash, et al., 1981). Δn_U, the difference in the number of urea binding site between the unfolded state and the native state, is expected to be smaller because some of these binding sites are also accessible to urea in the native conformation. Both values of 49 mol/mol from the 2&3-state analysis and 33 mol/mol from the 2&2-state analysis appear to be plausible, but again the SVD analysis of the scattering data justifies the existence of an intermediate.

The rather small R_g, ~ 22 Å, of lysozyme in 8 M urea, when compared with 32 Å of a random coil model of similar molecular weight, heparin, suggests that the denatured lysozyme in 8 M urea pH 2.9 does not resemble a random coil. A slightly larger R_g, ~ 24 Å, is obtained for reduced unfolded lysozyme in 8 M urea with 80 mM DTT. The shape of the P(r) function of the denatured state as well as its markedly small maximum dimension, D_{max}, further support a departure from a random coil structure. Furthermore, the Kratky plot provides a powerful way to analyze the scattering data in distinguishing between coil and globular structures (Glatter & Kratky, 1981). Although the lack of a well-defined peak in the Kratky plot demonstrates a significant reduction in globularity in unfolded lysozyme, the absence of a plateau and the presence of a residual peak shape argue against a coil structure. The non-globularity of the disulfide-reduced unfolded lysozyme appears to increase further as judged from the Kratky plot (data not shown), but still deviates from that of a coil.

The degree of unfolding and the structure of the unfolded state are among the interesting issues in protein folding studies. Although far-UV CD and amide hydrogen exchange experiments suggest that lysozyme in 8 M urea pH 2.0 is largely unstructured (Buck, et al., 1994), evidence for nonspecific interactions and deviation from random coil is shown from ^{1}H NMR as well as photo-CIDNP spectra (Broadhurst, et al., 1991; Evans,

et al., 1991). Our scattering results clearly support a non-random coil structure in the denatured state of lysozyme in 8 M urea at pH 2.9. The residual structure appears to be only slightly affected by removal of disulfide bridges, arguing for existence of nonspecific interactions in addition to the geometric constraints imposed by covalent S-S bonds.

Evidence for an intermediate state. Cooperative unfolding does not necessary indicate a two-state transition with absence of equilibrium intermediates. The calculated three-state transition curves also appear to be quite cooperative, and there seems to be no clear boundary between two-state and three-state transitions (Lumry, *et al.*, 1966). A practical approach is to use different structural probes to follow the unfolding process under the same conditions, and the existence of stable intermediates may be inferred from the differences in the unfolding curves. In our study, a small yet visible difference in the unfolding curves of lysozyme in urea solution pH 2.9 is observed by CD 222 nm and R_g. This difference may reflect the occurrence of an intermediate as it usually does, or since the distinction is subtle, it might nevertheless be a result of poor accuracy and precision of the techniques as well as any variations associated with the sample (in our case this might result from the difference of lysozyme concentration ~ 0.6 mg/ml in CD versus ~ 7 mg/ml in scattering experiments). However, a strong argument for the existence of an intermediate comes from the SVD analysis of scattering data over the whole scattering angle region with the evidence of a third basis function. Our results stress the importance of applying various techniques to probe different aspects of protein structure.

An alternate to a three-state binding model for denaturation-induced unfolding has been put foward by Alonso and Dill (Alonso & Dill, 1991; Dill & Shortle, 1991). Their picture, which might be thought of as a "many-state" model is that there is a single free energy barrier for going from the native to an ensemble of disordered states which comprise structures varying from the compact conformation with a hydrophobic core to the fully unfolded conformation; however, the degree of unfolding of the disordered states is a continuous function of denaturant concentration. (See also the discussion in Doniach, *et al.*, 1995.) Although this provides a rather physically plausible scenario, the SVD analysis of our data shows that the signal/noise level attained in our measurements is probably not good enough to discriminate this continuous unfolding picture from a 3-state picture. In particular, Fig. 4.12 shows that at urea concentration between 6.5 M and 8 M the scattering profiles may be expressed as a linear superposition of the scattering profile from the partially folded intermediate and from the fully denatured state. If there is a gradual unfolding of the α domain over this concentration range, it would show up as additional

145

basis vectors in the SVD matrix. However, it was found that components of the SVD matrix with smaller singular values than the first three lie within the noise limits of the data, so they could not be discriminated for at the present signal/noise levels.

Structure of the intermediate. By assumption of a two-state approximation to CD measurements and the observation of unfolding at lower urea concentration detected by R_g than by CD, the intermediate has a similar CD signal at 222 nm to that of the native state. The compactness of this intermediate is characterized by 19-20 Å in R_g. Compared with the native (~ 15 Å) and the unfolded (~22 Å) states, the structure of the intermediate lies halfway between the native and the unfolded conformations, suggesting a great decrease in tertiary contacts in this intermediate. The constructed P(r) function for the intermediate shows that while the intermediate is swollen to such an extend that its maximum length is close to that in the unfolded state, the most populated intramolecular distances, measured by the peak position, remain similar to that in the native state. The intermediate must maintain some similar local structure present in the native conformation. A well-defined peak in the calculated Kratky plot indicates that the shape of the intermediate is still globular. Structural modeling suggests that the substantially enlarged R_g for the intermediate may not be solely attributed to a fully unfolded β domain; instead, it appears that the α domain (α1, α2 subsets) also expands to some extent. Taken together, our results suggest that the intermediate has native-like secondary structure for the α domain with a distorted compact fold, characteristic of a molten globule-like state, and an unfolded structure for the β domain.

The intermediary conformation of the partially folded state is also supported by the derived thermodynamic parameters: $\Delta G_I^{H_2O}$ is about half of $\Delta G_U^{H_2O}$. Partially folded states with intermediate stability are also found in urea-unfolding of trytophan synthase and β-lactamase (Matthews & Crisanti, 1981; Uversky & Ptitsyn, 1994). From the denaturant binding model, only half the urea binding sites in the unfolded state of lysozyme are available to urea in the transformation of N -> I; the other half becomes exposed in the unfolding of I -> U. Our results for lysozyme unfolding in urea solution pH 2.9 can be interpreted in the framework of the hierarchical unfolding model (Palleros, et al., 1993). In this model, the native conformation consists of structural units. The initial unfolding observed by R_g would be correlated with a breakdown of interunit interactions between the α domain and the β domain, and unfolding of the β domain. The further unfolding detected by the CD signal at 222 nm would be associated with the disruption of the α domain (including secondary interactions as well as the side-chain interactions within the

unit). The structure of the intermediate would be expected to maintain compactness (19-20 Å) to some extent since it retains some weak interactions besides containing substantial amount of secondary structure, particularly in the α domain.

The importance of our study is two-fold. First, although stable partially folded states for hen lysozyme are scarce, here we demonstrate that under appropriate conditions, intermediates can be accumulated at a level sufficient for characterization. Second, the partially compact, native-like secondary structure in the intermediate exhibits certain features of the early kinetic counterpart (Radford, *et al.*, 1992), and structural characterization of this intermediate may provide a model to help in elucidating the early and fast events in the protein folding process. Our recent time-resolved scattering studies show that the radius of gyration of the first resolved conformation (~ 100 ms after initiation of refolding in urea at pH 2.9) is ~19 Å, a value close to that of the intermediate (see Chapter 5).

4.6. Acknowledgments

4.7. References

Alonso, D. O. V., & Dill, K. A. (1991). *Biochemistry*, **30**, 5974-5985.

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

Aune, K. C., Salahuddin, A., Zarlenyo, M. H., & Tanford, C. (1967). *J. Biol. Chem.*, **242**, 4486-4489.

147

Aune, K. C., & Tanford, C. (1969). *Biochemistry*, **8**(11), 4586-4590.

Baum, J., Dobson, C. M., Evans, P. A., & Hanley, C. (1989). *Biochemistry*, **28**, 7-13.

Blake, C. C., Koening, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., & Sarma, V. R. (1965). *Nature*, **206**, 757-761.

Blundell, T. L., & Johnson, M. S. (1993). *Prot. Sci.*, **2**, 877-883.

Bradbury, J. H., & King, N. L. R. (1969). *Nature*, **223**, 1154-1156.

Broadhurst, R. W., Dobson, C. M., Hore, P. J., Radford, S. E., & Rees, M. L. (1991). *Biochemistry*, **30**, 405-412.

Bruzzesi, M. R., Chiancone, E., & Antonini, E. (1965). *Biochemistry*, **9**, 1796-1800.

Buck, M., Radford, S. E., & Dobson, C. M. (1993). *Biochemistry*, **32**, 669-678.

Buck, M., Radford, S. E., & Dobson, C. M. (1994). *J. Mol. Biol.*, **237**, 247-254.

Dael, H. V., Haezebrouck, P., Morozova, L., Arico-Muendel, C., & Dobson, C. M. (1993). *Biochemistry*, **32**, 11886-11894.

Damaschun, G., Damaschun, H., Gast, K., Gernat, C., & Eirwer, D. (1991). *Biochim. Biophys. Acta*, **1078**(2), 289-295.

Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D., & Chan, H. S. (1995). *Protein Science*, **4**, 561-602.

Dill, K. A., & Shortle, D. (1991). *Annu. Rev. Biochem.*, **60**, 795-825.

Dobson, C. M. (1994). *Curr. Biol.*, **4**(7), 636-640.

Dobson, C. M., Evans, P. A., & Radford, S. E. (1994). *Tibs*, **19**, 31-37.

Doniach, S., Bascle, J., Garel, T., & Orland, H. (1995). *J. Mol. Biol.*, **254**(5), 960-967.

Eliezer, D., Chiba, K., Tsuruta, H., Doniach, S., Hodgson, K. O., & Kihara, H. (1993). *Biophys. J.*, **65**, 912-917.

Evans, P. A., Topping, K. D., Woolfson, D. N., & Dobson, C. M. (1991). *Proteins: Struct. Funct. Genet.*, **9**, 248-266.

Fetler, L., Tauc, P., Herve, G., Moody, M. F., & Vachette, P. (1995). *J. Mol. Biol.*, **251**, 243-255.

Glatter, O., & Kratky, K. (Ed.). (1981). *Small-Angle X-Ray Scattering*. New York: Academic Press.

Golub, G. H., & Reinsch, C. (1970). *Numer. Math*, **14**, 403-420.

Greene, R. F., & Pace, C. N. (1974). *J. Biol. Chem.*, **249**, 5388-5393.

Haezebrouck, P., Joniau, M., Dael, H. V., Hooke, S. D., Woodruff, N. D., & Dobson, C. M. (1995). *J. Mol. Biol.*, **246**(3), 382-387.

Hofrichter, J., Sommer, J. H., Henry, E. R., & Eaton, W. A. (1983). *Proc. Natl. Acad. Sci. USA*, **80**, 2235-2239.

Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C., & Rupley, J. A. (1972). *Vertebrate Lysozyme*. In Boyer, P. D. (Eds.), *The Enzymes* (pp. 666-868). New York: Academic Press.

Jennings, P. A., & Wright, P. E. (1993). *Science*, **262**, 892-896.

Karplus, M., & Weaver, D. L. (1976). *Nature*, **260**, 404-406.

Kataoka, M., Nishii, I., Fujisawa, T., Ueki, T., Tokunaga, F., & Goto, Y. (1995). *J. Mol. Biol.*, **249**(1), 215-228.

Kato, S., Okamura, M., Shimamoto, N., & Utiyama, H. (1981). *Biochemistry*, **20**, 1080-1085.

Kim, P. S., & Baldwin, R. L. (1990). *Annu. Rev. Biochem.*, **59**, 631-660.

Kuwajima, K., Nitta, K., Yoneyama, M., & Sugai, S. (1976). *J. Mol. Biol.*, **106**, 359-373.

Lambright, D. G., Balasubramanian, S., & Boxer, S. G. (1991). *Chem. Phys.*, **158**, 249-260.

Lattman, E. E. (1994). *Curr. Opin. Struct. Biol.*, **4**, 87-92.

Liberti, P. A., & Stivala, S. S. (1967). *Arch. Biochem. Biophys.*, **119**, 510-518.

Lumry, R., Biltonen, R., & Brandts, J. F. (1966). *Biopolymers*, **4**, 917-944.

Matthews, C. R., & Crisanti, M. M. (1981). *Biochemistry*, **20**, 784-792.

Ohgushi, M., & Wada, A. (1983). *FEBS*, **164**, 21-24.

Pace, C. N. (1975). *Crit. Rev. Biochem.*, **3**, 1-43.

Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. In Colowick, S. P. & Kaplan, N. O. (Eds.), *Methods in Enzymology* (pp. 266-280). Orlando, Florida: Academic Press.

Palleros, D. R., Shi, L., Reid, K. L., & Fink, A. L. (1993). *Biochemistry*, **32**, 4314-4321.

Pfeil, W., & Privalov, P. L. (1976). *Biophys. Chem.*, **4**, 41-50.

Prakash, V., Loucheux, C., Scheufele, S., Gorbunoff, M. J., & Timasheff, S. N. (1981). *Arch. Biochem. Biophys.*, **210**(2), 455-464.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical Recipes in Fortran* (2nd ed.). Cambridge University Press.

Provencher, S. W., & Blockner, J. (1983). *J. Biochem. Biophys. Methods*, **7**, 331-334.

Ptisyn, O. B. (1973). *Dokl. Acad. Nauk SSSR*, **210**, 1213-1215.

Ptisyn, O. B. (1993). The molten globule state. In Creighton, T. E. (Eds.), *Protein folding* (pp. 243-300). New York: W. H. Freeman and Company.

149

Ptitsyn, O. B. (1995a). Molten Globule and Protein Folding. In Anfinsen, C. B., Edsall, J. T., Richards, F. M., & Eisenberg, D. S. (Eds.), *Advances in Protein Chemistry* (pp. 83-230). San Diego: Academic Press, Inc.

Ptitsyn, O. B. (1995b). *Curr. Opin. Struct. Biol.*, **5**, 74-78.

Radford, S. E., Dobson, C. M., & Evans, P. A. (1992). *Nature*, **358**, 302-307.

Schmid, F. X. (1993). Kinetics of unfolding and refolding of single-domain proteins. In Creighton, T. E. (Eds.), *Protein folding* (pp. 197-242). New York: W. H. Freeman and Company.

Semenyuk, A. V., & Svergun, D. I. (1991). *J. Appl. Crystallogr.*, **24**, 537-540.

Shrager, R. I., & Hendler, R. W. (1982). *Anal. Chem.*, **54**, 1147-1152.

Stephenson, G. B. (1988). *Nucl. Instrum. Meth. Phys. Res.*, **A266**(1-3), 447-451.

Svergun, D. I., Barberato, C., & Koch, M. H. (1995). *J. Appl. Crystallogr*, submitted for publication.

Tanford, C. (1970). Protein denaturation. Part C. Theoretical models of the mechanism of denaturation. In C. B. Anfinsen, J., Edsall, J. T., & Richards, F. M. (Eds.), *Adv. Prot. Chem.* (pp. 1-95). New York: Academic Press.

Uversky, V. N., & Ptitsyn, O. B. (1994). *Biochemistry*, **33**, 2782-2791.

Wetlaufer, D. B. (1962). Ultraviolet spectra of proteins and amino acids. In *Advances in protein chemistry* (pp. 303-390). New York: Academic Press, Inc.

Chapter 5

Time-resolved Small-angle X-ray Scattering Studies of Lysozyme Refolding

5.1. Introduction and Background

The complexity of kinetic study of protein folding arises from two major sources (Schmid, 1992). (1) The unfolded state of proteins is conformationally heterogeneous, which leads to multiple parallel refolding pathways. (2) Small globular proteins usually refold within seconds, and folding intermediates only form transiently in the refolding pathways.

To study the mechanism of folding reaction, two working models have been generalized. The framework model (Ptitsyn, 1973) suggests that a stable framework of secondary structure is formed before the native-like tertiary fold is formed. The modular assembly model (Karplus & Weaver, 1976; Kim & Baldwin, 1990) postulates that a protein folds by parts: the secondary and tertiary structures are formed at the same time within each folding units, or microdomains, transiently and independently, and these microdomains diffuse, collide and coalesce into the stable native conformation. Both models predict partially folded intermediates in the folding pathway.

Compact kinetic intermediates have been found in the refolding of a number of single-domain proteins. These include α-lactalbumin, ribonuclease A, apomyoglobin, cytochrome c, barnase, ubiquitin, and hen lysozyme. Kinetic intermediates comprises of "early" and "late" species. Early intermediates form rapidly in the millisecond time scale or faster, within deadtime of most techniques. Late intermediates accumulate prior to the rate-determining step of folding, and can be well populated to be detected. Much of structural information of folding intermediates comes from study of the late intermediates. However, it is very difficult to obtain detailed structure characterizations of these kinetic intermediates due to their transient nature.

In the study of equilibrium unfolding of bovine α-lactalbumin, a structural intermediate state was discovered (Kuwajima, et al., 1976). To describe the characteristics of this compact but with fluctuating intramolecular movement state, the term "molten globule" was proposed for this kind of folding intermediate state (Ohgushi & Wada, 1983). Later, it was found that an early transient intermediate in the α-lactalbumin refolding process had similar properties to this equilibrium molten globular state. Since then, evidence for identities between kinetic intermediates and equilibrium molten globule species has been found in several other proteins. A general role of molten globule forms involved on the folding pathway has thus been put forward (Kuwajima, 1989; Ptitsyn, 1995).

152

Structural characterizations of these stable equilibrium intermediates provide insights into the mechanism of protein folding.

Lysozyme folding. Lysozyme is among one of the well-studied proteins in folding studies. The kinetic aspect of the folding processes has been examined by a variety of physical techniques, including stopped-flow fluorescence, far- and near-UV CD and absorption spectroscopies, hydrogen exchange labeling, nuclear magnetic resonance spectroscopy (NMR) and electrospray mass spectrometry (ESMS) (Dobson, *et al.*, 1994). Information revealed from these complementary methods provides a rather detailed model for the folding pathway of lysozyme. Folding events along the pathway can be summarized to three stages. (1) At the earliest stage (t < 2 ms), the unfolded states have collapsed into states with substantial secondary structure. However, the structure formed is not persistent, and therefore these states are highly dynamic and fluctuating. This early step is not resolved by any methods employed so far, and is inferred from the stopped-flow far-UV CD and fluorescence experiments. (2) In the next stage with time constant $\tau < 100$ ms, the collapsed states are interconverting and secondary structure is stabilized. Hydrogen exchange experiments have resolved different kinetic behaviors for the α and β domains. A folding intermediate is implicated where the α domain has acquired persistent and native-like secondary structure, but the β domain is still largely unstable (Radford, *et al.*, 1992). The structure of the α domain at this stage resembles that of the molten globule state. (3) The native conformation is obtained at the last stage ($\tau \sim 350$ ms). Structure in the β domain is stabilized and α and β domains are organized into the final native state. Note that these refolding experiments were performed at pH 5.3.

In addition to these three stages, about 10% of the unfolded lysozyme molecules refolded in a slow process with time constant $\tau > 16$ s at pH 2.6 (Kato, *et al.*, 1982). The slow process is attributed to the interconversion of proline isomers in the unfolded states. A ratio of *trans* to *cis* isomers in the unfolded states is thus derived to be 9 : 1.

Recently, the role of intermediates in the folding pathway has been challenged. From interrupted refolding experiments, a fast folding pathway of lysozyme has been identified with time constant $\tau \sim 50$ ms (Baldwin, 1995; Kiefhaber, 1995). About 14% molecules refold in this process, while the majority of molecules, 86%, refold on a slower time scale that has been characterized as a three-stage process as described above. The two-state mechanism of the fast folding process suggests that the partially folded intermediates may be a result of kinetic traps in the folding pathway.

Information regarding the compactness of folding molecules is normally derived from the binding of hydrophobic fluorescent probes such as 8-anilinonapthalene-1-sulfonate (ANS), and intrinsic tryptophan fluorescence spectroscopy. However, these methods give only indirect information about protein compactness. Addition of ANS may complicate the already complex folding process; in fact, it has been found that ANS perturbs the refolding kinetics of lysozyme (Itzhaki, *et al.*, 1994). Exclusion of solvent around tryptophan residues may not reflect the overall compaction of the folding process.

Clearly, direct structural techniques are needed to probe the compactness of the folding molecule in the folding process. Small-angle scattering (SAXS) is one of these techniques. The radius of gyration (R_g) and distance distribution function ($P(r)$) derived from SAXS experiments, as described in detail in Chapter 1, give the measure of overall size and shape, thus compactness, of the protein molecule. Combined with the stopped-flow method, time-resolved SAXS can provide direct information on the kinetics of compactness in the folding process. The application of time-resolved SAXS technique to the study of lysozyme folding is described in this chapter.

5.2. Experimental Procedures

5.2.1. Sample Preparations

Hen egg-white lysozyme was purchased from the Sigma Chemical Company, St. Louis, MO. Lysozyme was directly dissolved in filtered 8 M urea solutions in 100 mM NaCl, 100 mM Na-citrate buffer for pH 2.9, or 100 mM NaCl, 100 mM Na-phosphate buffer for pH 2.0. The solutions were centrifuged at 3,000 rpm in a Beckman Ti 50 rotor for 60 minutes at 15 °C. Unfolding was allowed to equilibrate for at least 5 hours before the scattering experiments. Dilution buffers were 100 mM NaCl, 100 mM Na-citrate buffer at pH 2.9 for the final pH 2.9 condition, or 100 mM NaCl, 100 mM Na-phosphate buffer at pH 2.5 for the final pH 2.0 condition. Buffers were filtered through a 0.2 μm filter. The final lysozyme concentration was about 7 mg/ml. All the urea-containing samples were used within 24 hours.

5.2.2. Data Collection

Scattering data were collected at beamline 4-2 at Stanford Synchrotron Radiation Laboratory. The beamline configurations were basically the same as described in Section 4.2.2. A stopped-flow method was employed in the time-resolved scattering experiments. A stopped-flow rapid-mixer (Eliezer, 1994; Tsuruta, *et al.*, 1989) was used to replace the sample cell of the equilibrium study. Experiments were carried out over three independent runs, June 94, February 95, and June 95. Early data were collected using the rapid-mixer observation cell with 50 μm quartz windows, while the latest data were collected with the same cell but with 40 μm mica windows. The dilution ratios were 1:3.5 and 1:6 in volume. Refolding was initiated by rapidly diluting unfolded lysozyme in 8 M urea to a buffer of x M urea (x = 0, or 1 M depending on the mixing ratio), resulting in a strongly native condition (1 - 2 M urea). Stopped-flow data were collected in various time frames, ranging from 20 ms to 2 s, with a total time of 28 s as the longest period. A typical stopped-flow file contained data accumulated over 25 mixing repetitions. To improve the signal-to-noise ratio, hundreds of mixing events were measured for each condition. Control experiments were done by diluting sample solution (unfolded lysozyme in 8 M urea) with 8 M urea buffer without initiation of refolding, and data were collected in the same time-fashion as the corresponding stopped-flow experiments. The "initial" unfolded state was measured by diluting lysozyme in 8 M urea to 8 M urea buffer without invoking refolding, while the final folded state was measured by mixing lysozyme in 8 M urea with appropriate buffer and allowing folding to proceed for at lease 1 minute before data collection. Both the "initial" state and the final state were characterized immediately before or after stopped-flow experiments using the same stopped-flow cell for direct I(0) comparison. Temperature was set at 20 °C for most conditions used, however, experiments were also performed at 12 °C and 30 °C. The pH of the final condition varied within 0.1 unit.

5.2.3. Data Analysis

Data reduction. A typical stopped-flow experiment consisted of tens of data files totaling hundreds of mixing repetitions. Data of each stopped-flow file (25 mixing repetitions) were first normalized, background subtracted and calibrated. The calibrated data were carefully examined among files before they were combined together. It is not uncommon that data taken within the same time block (normally 6 hours) had to be grouped into two or even three sets. Time-dependence of parasitic scattering was apparent when

comparing the background data taken in the same block. For some unknown reason, a shift of scattering pattern by a few channels was also observed among the stopped-flow sample data, and between the stopped-flow and the static (background) data. For those time-dependent data sets, data were first divided into groups, and an appropriate background for each group was found to be either the background file taken immediately before or after the data group or a linear combination of these two background files. Data from different groups were not always combined. For data displaying a channel-shift, in some cases it was not possible to find the displacement, thus the stopped-flow sample data could not be combined or the sample data and the background data could not match. Data were not further analyzed in these cases. Fortunately, there were only a few files with this kind of anomaly.

Guinier analysis. Radii of gyration were derived from the Guinier approximation :

$$LnI(S) = LnI(0) - \frac{4\pi^2}{3} S^2 R_g^2 \qquad [1]$$

The fitting region, 0.0045 - 0.001 Å^{-1} in S, for all the analysis including the stopped-flow and the static data sets was chosen to be the same as that in the equilibrium unfolding studies described in Chapter 4. The forward scattering intensity, I(0), was also obtained from equation [1].

To estimate the correction factor for I(0) due to changes in solvent (urea) contrast, the electron density of lysozyme is approximated as follows. The partial specific volume, molecular weight and number of total electrons of lysozyme are 0.703 cm^3/g, 14,600, and 7,700, respectively. Then, the electron density of Lysozyme is:

$$[7,700/(14,600 \times 0.703)] \times 6.02 \times 10^{23}/10^{24} = 0.45 \quad (e^-/\text{Å}^3)$$

The concentration of H_2O is 55.55 M. Since the density of urea is 1.323 g/cm^3, the amount of H_2O displaced by urea in a n M urea solution is 45.35n g/l, or 2.52n mol/l. The number of electrons in one molecule of H_2O and urea is 10 and 32, respectively. The electron density of n M urea in H_2O solution is then:

$$[10 \times (55.5 - 2.52n) + 32n] \times 10^{-3} \times 6.02 \times 10^{23}/10^{24} = 0.334 + 0.00409n \quad (e/\text{Å}^3)$$

The increase in electron density due to the presence of buffer 100 mM NaCl, 100 mM Na-citrate is about 0.0095 e⁻/Å³. Therefore, the ratio of I(0) of lysozyme in 1 M urea (one of the final refolding conditions) to that in 8 M urea is $[(0.45-0.347)/(0.45-0.376)]^2 = 1.9$, so the I(0) in the refolding is about twice that of the control.

Kinetic analysis. The kinetics of changes in R_g as a function of time was fit into single exponential as formulated in equation [2]:

$$R_g^2(t) = Ae^{-kt} + R_g^2(eq) \qquad [2]$$

where $R_g^2(t)$ is the R_g^2 value measured at the refolding time t, $R_g^2(eq)$ is the equilibrium value of R_g^2 corresponding to the final native state, k is the rate constant (later referred to as time constant τ, $\tau=1/k$), and A is the amplitude. The refolding process could also be fit with two exponential terms. However, the quality of the current data did not allow a clear distinction between these two fitting procedures. In addition, the double-exponential approximation did not give consistent results for independent experiments performed under the same conditions. Therefore, the refolding process was treated as a one-step kinetic process and approximated with a single exponential.

5.3. Results

5.3.1. Sample Integrity

No evidence of nonspecific sample aggregation was observed for all the samples studied, including both the initial and final states and during the refolding kinetics, since if present it would show up in the small angle region in the Guinier plot. No clear depressive trend was displayed in the Guinier region, agreeing with the conclusion in Section 4.4.2 that interparticle interference can be ignored at this lysozyme concentration (7-8 mg/ml). For the static measurements on the initial and final states, data were collected up to 10 minutes and no time dependence in scattering pattern was observed. Since the longest data collection time in the stopped-flow experiments was 28 seconds, radiation-induced sample degradation for all samples can therefore be excluded.

5.3.2. The Initial and Final States

In all cases, R_g of lysozyme at the initial conditions (8 M urea at pH 2.9 or pH 2.0) from both the static and control experiments was found to be about 22 Å, indicative of an unfolded state of lysozyme at the beginning of the refolding experiments. The association state of lysozyme at the initial conditions can be inferred from a comparison of I(0) values of the initial and final states. For pH 2.9, a ratio of 1.83 is obtained between 1 M urea and 8 M urea samples, comparable to the predicted factor (1.9) due to a change in solvent contrast. The consistency in I(0) after correction for solvent contrast suggests that the unfolded lysozyme is largely monomeric in 8 M urea. Further supporting argument comes from the stopped-flow experiments and the control experiments. Fig. 5.1 shows the scattering curves of the last time-frame of the stopped-flow data and the corresponding control. Again, a factor of about 2 for I(0) is derived, suggesting the monomeric state of the unfolded lysozyme in the initial condition. I(0) for the initial and final states are comparable to those obtained for the control and the stopped-flow data, respectively, although values from the static data are systematically higher than those from the time-resolved data by 15%.

At the end of refolding, R_g of lysozyme (~ 15.3 Å) is in good agreement with that of the native state determined in Section 4.4.2. In addition, scattering curves of the last few time-frames superimpose with that of the final state, taken 1 minute after initiation of refolding. Therefore, within SAXS resolution, lysozyme has completely refolded to the folded state by the end of the refolding process followed by these time-resolved SAXS experiments.

5.3.3. The Refolding Kinetics

Several attempts were made to probe the early events in the refolding process by inducing the process at a lower temperature, or by refolding to a higher urea concentration. The time duration at the beginning of data collection were tried with 20 ms, 50 ms and 100 ms, and the relative signal-to-noise ratios for the 20 ms and 100 ms time-frame can be seen in Fig. 5.2. A deadtime of about 13 ms was estimated with the current SAXS setup; the deadtime includes both the inherent deadtime of the stopped-flow mixer and the time delay due to signal transfer among the computer, various electronic modules and the mixer. To obtain a reliable kinetic analysis on changes in R_g, the current smallest duration time-frame was found to be 100 ms with an accumulation of a few hundreds mixing repetitions. The
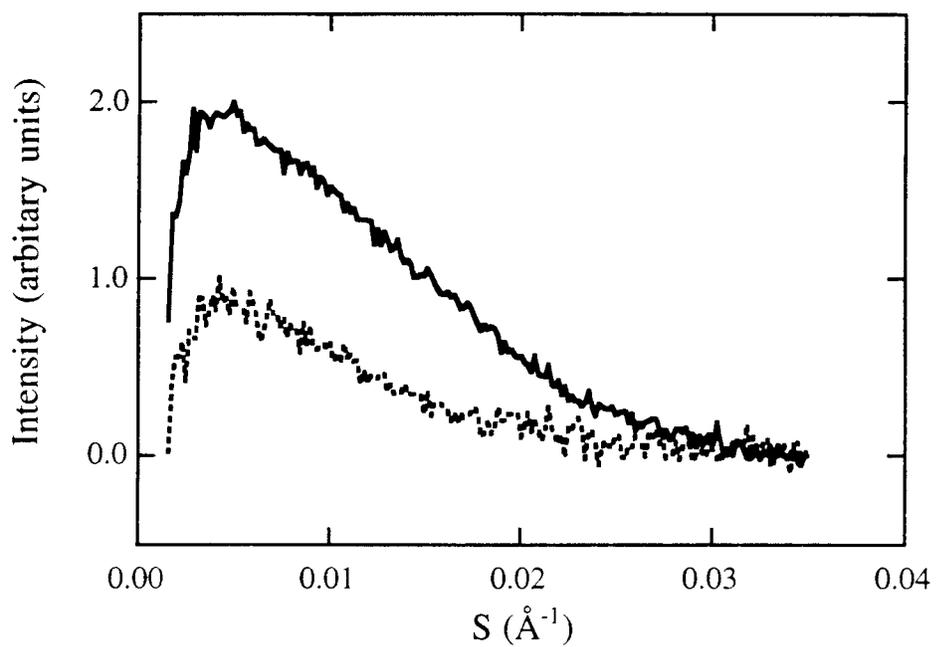
Figure 5.1. Scattering curves of the last time-frame of a stopped-flow (—) and the corresponding control experiments (- - -). The duration time is 2 s, and the number of mixing events is 700. The final refolding condition is 1.8 M urea. $I(0)$ of the unfolded state (in the control data) is about half of $I(0)$ of the refolded state (the final stage in the stopped-flow data).
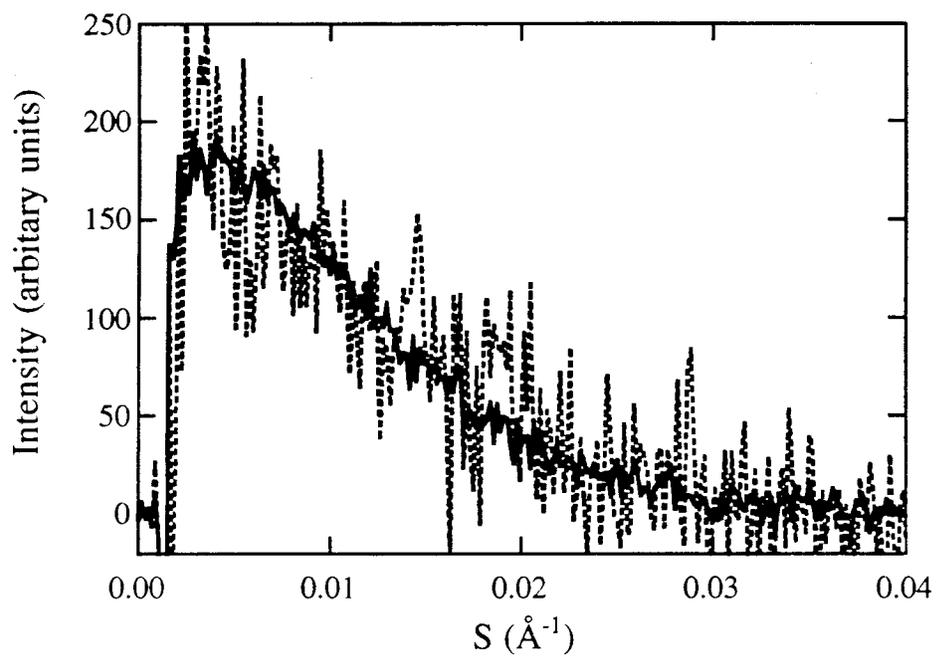
Figure 5.2. Comparisons of data statistics with a time-frame of 20 ms (- - -) and 100 ms (—). The accumulation is 305 mixing events. Data with 20 ms duration time is apparently too noisy for Guinier analysis.

use of more intense beam will certainly improve the statistics of 20 ms time frame and thus time resolution, as well as reduce the number of repetitions which becomes an important factor when experiments are constrained by sample supply .

Fig. 5.3 represents a typical time-course of R_g in the refolding with a final condition of 1.8 M urea at pH 2.9. At the earliest resolved time (~100 ms after initiation of refolding), R_g is found to be ~ 19 Å. A R_g of ~ 15 Å is reached about 5 s after the refolding, and is stabilized over the rest period of the collection time, suggesting that the major refolding kinetic events is complete within 5 seconds. Note that the initial state of lysozyme is verified to be in the unfolded state with a R_g of ~ 22 Å (see Section 5.3.2). Further, a R_g consistent with that of the unfolded state is obtained from the control data. The lack of time dependence of R_g in the control experiment (see Fig. 5.4) strongly suggests that the substantial reduction in R_g observed at the very first 100 ms of refolding is not due to a dilution from a relative concentrated lysozyme solution (~ 32 or ~ 56 mg/ml) to a final ~ 8 mg/ml condition. The time-course of R_g was fit with a single exponential according to equation [2], and is shown in Fig. 5.3. However, the single-phase kinetics does not account for the total amplitude; a loss of ~ 30 % of the total change in R_g occurs within the burst phase (< 100 ms). Therefore, this indicates that there is a faster phase occurring prior to the slower phase seen in time-resolved SAXS data. This result implies the existence of an early intermediate in the pathway.

The derived time constants from different experiments and under various refolding conditions are presented in Table 5.1. Refolding to a final condition of 1 M urea at pH 2.9 was repeated several times, and a time constant of ~ 1 s is consistently obtained from the data. As expected, refolding proceeds slower for transitions to 2 M than to 1 M (by a factor of 2). The refolding rate is further reduced again by about a factor of 2 with a decrease of 8 °C in temperature. The slowest process, ~ 9 s, is observed with transition to final urea concentration of 2 M at pH 2.0, a less favorable native condition just outside the transition region as seen in Fig. 4.11.b.

Fig. 5.5 shows the Kratky plots of the initial and the final time-frames. The presence of a peak in the Kratky plot indicates that at the early stage of refolding (~ 100 ms), lysozyme has collapsed into a globular structure. It is not surprising given that even in 8 M urea, lysozyme is not completely unfolded to a random coil structure as suggested by trace evidence of a peak in the Kratky plot (see Fig. 4.9.a). The well defined peak from the last time-frame is a characteristic of a globular structure.
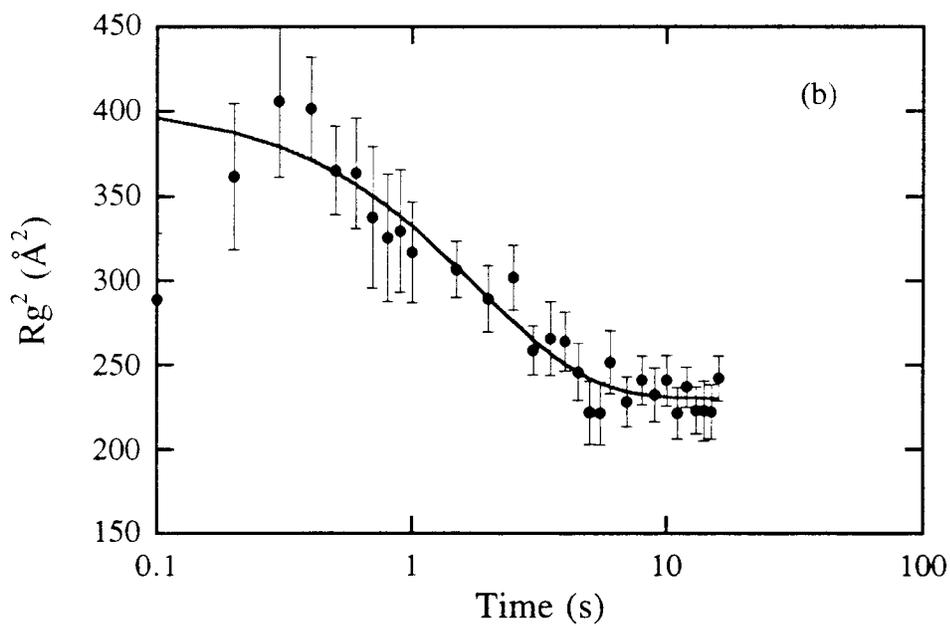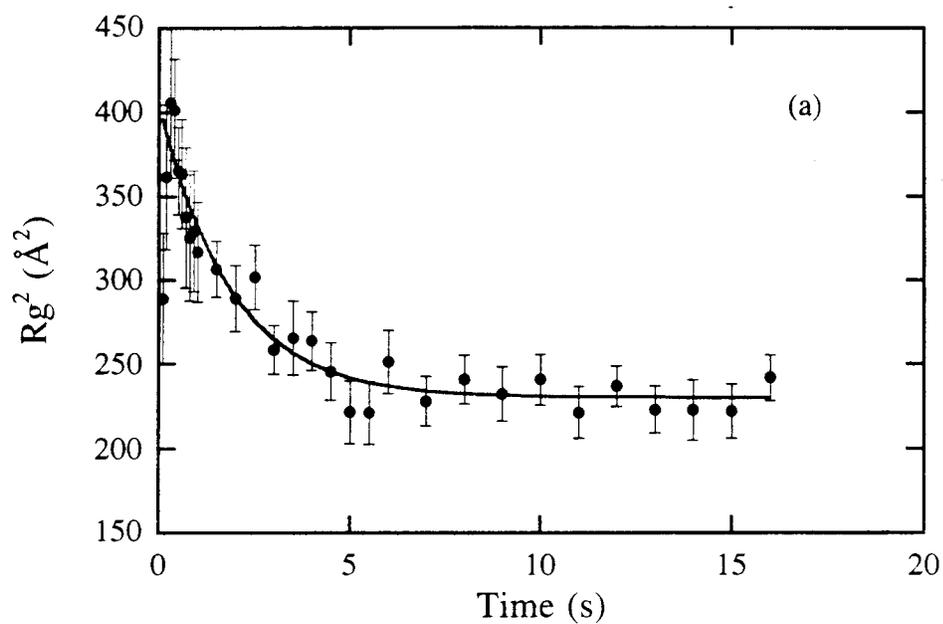
Figure 5.3. Time-course of Rg^2 in the refolding of lysozyme at pH 2.9. The transition is from 8 M urea to 1.8 M urea. (a) time is in linear scale; (b) time is in log scale for a better display of the early events.
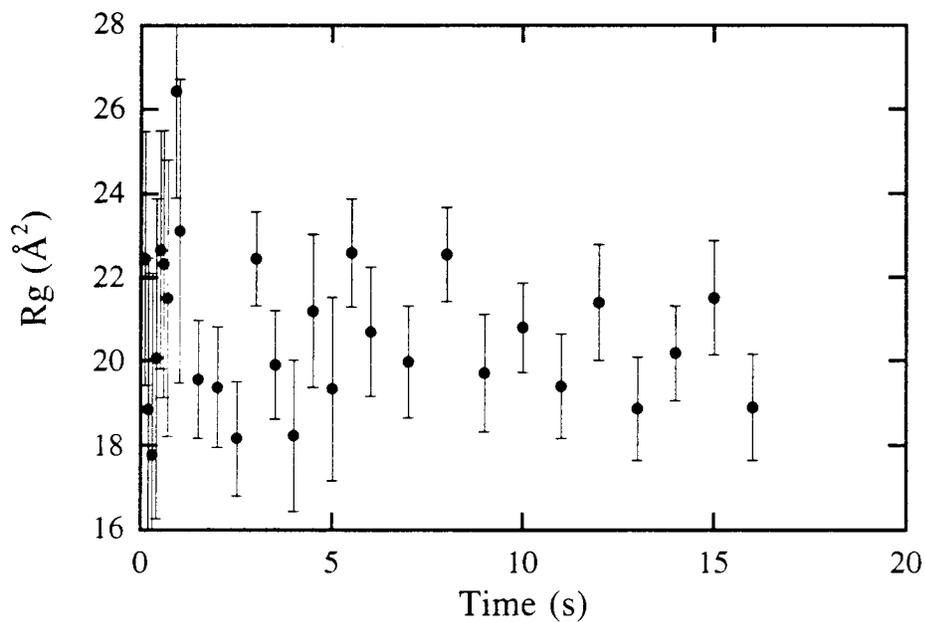
Figure 5.4. Time-course of R_g in a control experiment. The control experiment was done by diluting unfolded lysozyme in 8 M urea to a 8 M urea buffer without initiating refolding. R_g is steadily about 21 Å, comparable to that of the unfolded state from static measurements. The decrease in R_g during the refolding process is not observed in the control.

163

Table 5.1. Kinetic parameters derived for lysozyme refolding.

final refolding conditions	rate constant k (s^{-1})	time constant τ (s)	R_g (Å) for the final state	amplitude	numbers of repetitions
1 M urea, pH 2.9, 20 °C	0.97 ($\pm$ 0.1)	1.03 ($\pm$ 0.11)	15.6 ($\pm$ 0.1)	132.3 ($\pm$ 7.1)	265[b]
	1.38 ($\pm$ 0.2)	0.72 ($\pm$ 0.11)	15.6 ($\pm$ 0.1)	133.8 ($\pm$ 9.2)	400[b]
	0.91 ($\pm$ 0.15)	1.1 ($\pm$ 0.18)	15.0 ($\pm$ 0.1)	128.7 ($\pm$ 10.1)	150[b]
	1.0 ($\pm$ 0.03)	1.0 ($\pm$ 0.03)	15.5 ($\pm$ 0.1)	124.4 ($\pm$ 1.8)	225[b]
2 M urea, pH 2.9, 20 °C	0.54 ($\pm$ 0.05)	1.85 ($\pm$ 0.17)	15.2 ($\pm$ 0.1)	175.3 ($\pm$ 6.0)	725[a]
2 M urea, pH 2.9, 12 °C	0.25 ($\pm$ 0.1)	4.0 ($\pm$ 1.6)	15.6 ($\pm$ 0.4)	114 ($\pm$ 15.6)	405[a]
2 M urea, pH 2.0, 20 °C	0.11 ($\pm$ 0.02)	9.1($\pm$ 1.65))	14.7 ($\pm$ 0.3)	130.6 ($\pm$ 8.8)	705[a]

[a] data taken with monochromatic beam.
[b] data taken with multilayer beam.
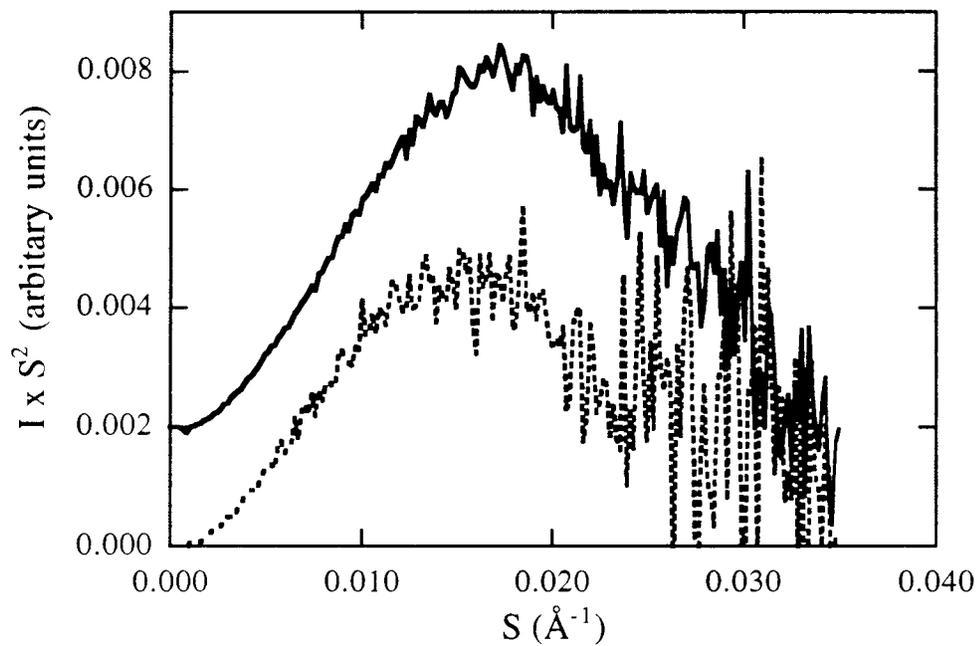
Figure 5.5. Kratky plots of the first 100 ms time-frame data (- - -) and the last 2 s time-frame data (—). The stopped-flow process was followed for 16 seconds. For clarity, Kratky plot of the 2 s duration data is shifted by 0.002 unit. The accumulation is 700 mixing events.

Immediately following the mixing, the background buffer remains the same throughout data collection. Consequently, I(0) should stay steady in the refolding process. I(0) values from the stopped-flow data are reasonably consistent with a variation of 10-15%, a greater deviation than normally observed in the static measurements. The mean I(0) is about 15% lower than I(0) of the final state in static measurement. (The mean I(0) of the control data is also 15% smaller than the corresponding I(0) of the initial state in the static measurement, see Section 5.3.2.) On the other hand, R_gs derived from time-resolved and static experiments agree well, and are reproduced between independent measurements. The reason for the inconsistency in I(0) is not clear so far. However, it might provide an estimate to the magnitude of deviation in I(0).

For some data sets, although noisy, a change in I(0) is observed for the first few time frames, including a step decrease of ~15%. Since the magnitude of the difference is comparable to that of the I(0) deviation, an interpretation of the change might not be meaningful. Despite a discrepancy in I(0) derived from the Guinier plot, scattering from the first and the last time-frames looks similar as seen in Fig. 5.5. The difference in I(0) derived from the Guinier plot might be attributed to the fitting error. Nor can it be excluded that the change is real. Better data quality and more measurements are necessary to justify a conclusive analysis. At the current stage, allowing a 15% deviation in I(0) derived from the Guinier plot, it appears that the association state of at least the majority of the lysozyme does not change throughout the refolding process, and is largely monomeric. The change in R_g can thus be correlated with conformational changes, rather than with a change in association state as observed in the case of myoglobin (Eliezer, *et al.*, 1993), of lysozyme in the refolding.

5.4. Discussion

Our time-resolved SAXS experiments provide direct measurements of compaction process in the folding pathway. At the earliest measurable time, 100 ms after initiation of refolding, R_g has decreased from ~ 22 Å of the unfolded state to ~ 19 Å. A single exponential fit to the data can not cover the total amplitude, suggesting that a faster separate process occurs within the deadtime (100 ms) of current experiment. Within experimental errors, I(0) remains consistent throughout the refolding, and is comparable to a monomeric

lysozyme. Therefore, the observed kinetics of changes in R_g corresponds to the time-course of conformational change of lysozyme in the refolding process.

For hen lysozyme refolding at 20 °C to a final concentration of 0.5 M GdmHCl at pH 5.2 (Itzhaki, *et al.*, 1994), two kinetic phases are observed in the binding of ANS with time constants of $\tau_1 = 23 \pm 5$ ms and $\tau_2 = 250 \pm 60$ ms, and in the intrinsic tryptophan fluorescence measurement with time constants of $\tau_1 = 25 \pm 3$ ms, $\tau_2 = 340 \pm 20$ ms. Our kinetics of changes in R_g is also biphasic. The burst phase, not resolved within our experimental deadtime of 100 ms, can be correlated with the fast process of ANS binding and tryptophan fluorescence, or the second stage of the folding event as described in Section 5.1. A substantial compaction occurs at this initial step with about 2-3 Å decrease in R_g. The second process ($\tau = 1$ s for the final condition of 1 M urea at pH 2.9) may also be correlated to the slow phase observed in ANS binding and intrinsic tryptophan fluorescence experiments based on the following reasons. (1) The refolding conditions used in our SAXS study are less favorable for folding. The ANS binding and intrinsic tryptophan fluorescence measurement were performed at a strongly native folding condition at pH 5.2, while SAXS experiments described here were carried out at pH 2.9 and pH 2.0. As expected, at a lower pH condition, the folding process will definitely slow down. A time constant of 1.3 s has been reported for lysozyme folding to a final concentration of 0.5 M GdmHCl at pH 2.6 at 25 °C (Kato, *et al.*, 1982). In our experiment, the refolding to a final concentration of 1.8 M urea at pH 2.0 is much slower, with $\tau = 9.1 \pm 1.6$ s (see Table 5.1). In addition, denaturants also affect folding. At pH 2.6 and 25 °C, refolding rate decreases with the increase of GdmHCl concentration, $\tau = 1.3, 4.3$ and 13.9 s for 0.5 M, 1.0 M and 1.5 M GdmHCl, respectively (Kato, *et al.*, 1982). The kinetics of changes in R_g is slower by a factor of 2 when refolding to a final concentration of 1.8 M urea than refolding to a final concentration of 1 M urea solution (see Table 5.1). (2) Although ANS binding and tryptophan fluorescence measurements suggest a time constant of ~ 300 ms for the last compaction process, hydrogen protection of some residues, mostly located in the β domain, occurs in a much slower phase ~400 - 800 ms (Radford, *et al.*, 1992). This suggests that the overall compactness in the folding process may not be complete at the time of the exclusion of ANS and the protection of tryptophan residues. A longer process observed by R_g is therefore not unexpected.

A folding intermediate was identified mainly by hydrogen exchange experiments in the second stage of refolding with $\tau < 100$ ms (see Section 5.1). The α domain of the folding intermediate has molten globule features, with persistent native-like secondary

structure and a loose tertiary fold, but the β domain is still largely unstructured. The biphasic kinetics observed by our time-resolved SAXS suggests that a folding intermediate has formed within 100 ms after initiation of refolding. The time scale of formation of this intermediate corresponds well with that of the intermediate reported previously, suggesting that these two species may be the same, *i.e.*, the intermediate found by SAXS has substantial native secondary structure in a distorted α domain, and an unstable β domain.

A stable intermediate has been resolved in the equilibrium unfolding study by SAXS (see Chapter 4) with R_g of 19-20 Å. The stable intermediate has a similar structure to that identified in the second stage of refolding. A rough estimate of R_g for the folding intermediate after the initial burst phase is about 19 ($\pm$ 1) Å (Table 5.1). The similar R_g values imply that the stable intermediate might resemble the folding intermediate in the refolding pathway.

5.5.   Acknowledgments

5.6.   References

Baldwin, R. L. (1995). *J. Bio. NMR*, **5**, 103-109.

Dobson, C. M., Evans, P. A., & Radford, S. E. (1994). *Tibs*, **19**, 31-37.

Eliezer, D. (1994) *Protein folding and protein metallocluster studies using synchrotron small angle x-ray scattering*. Ph.D. Thesis, Stanford University.

Eliezer, D., Chiba, K., Tsuruta, H., Doniach, S., Hodgson, K. O., & Kihara, H. (1993). *Biophysical J.*, **65**, 912-917.

Itzhaki, L. S., Evans, P. A., Dobson, C. M., & Radford, S. E. (1994). *Biochemistry*, **33**, 5212-5220.

Karplus, M., & Weaver, D. L. (1976). *Nature*, **260**, 404-406.

Kato, S., Shimamoto, N., & Utiyama, H. (1982). *Biochemistry*, **21**, 38-43.

Kiefhaber, T. (1995). *Proc. Natl. Acad. Sci. USA*, **92**, 9029-9033.

Kim, P. S., & Baldwin, R. L. (1990). *Annu. Rev. Biochem.*, **59**, 631-660.

Kuwajima, K. (1989). *Proteins*, **6**, 87-103.

Kuwajima, K., Nita, K., Yoneyama, M., & Sugai, S. (1976). *J. Mol. Biol.*, **106**, 359-373.

Ohgushi, M., & Wada, A. (1983). *FEBS*, **164**, 21-24.

Ptitsyn, O. B. (1973). *Dokl. Acad. Nauk SSSR*, **210**, 1213-1215.

Ptitsyn, O. B. (1995). Molten globule and protein folding. In Anfinsen, C. B., Edsall, J. T., Richards, F. M., & Eisenberg, D. S. (Eds.), *Advances in Protein Chemistry* (pp. 83-230). San Diego: Academic Press, Inc.

Radford, S. E., Dobson, C. M., & Evans, P. A. (1992). *Nature*, **358**, 302-307.

Schmid, F. X. (1992). Kinetics of unfolding and refolding of single-domain proteins. In Creighton, T. C. (Eds.), *Protein Folding* (pp. 197-242). New York: W. H. Freeman and Company.

Tsuruta, H., Nagamura, T., Kimura, K., Igarashi, Y., Kajita, A., Wang, Z.-X., Wakabayashi, K., Amemiya, Y., & Kihara, K. (1989). *Rev. Sci. Instrum.*, **60**, 2356-2358.