Changes in Biomolecular Conformation Seen by Small Angle X-ray Scattering

Sebastian Doniach

Departments of Applied Physics and Physics, Stanford University, Stanford California 94305

Received August 31, 2000

Contents

I. Introduction

The use of small-angle X-ray solution scattering (SAXS) as an analytical chemical tool for structural molecular biology is a mature field dating back to the work of Glatter and Kratky in the 1950s.[1] At that time, X-ray tube sources were used and it could take many hours or even days to collect a useful small-angle scattering profile of a protein in solution. The development of synchrotron radiation sources from multi-GeV electron and positron storage rings increased the available flux on the sample by factors of more than 10^5. Use of these sources coupled with electronic detectors made it possible to start collecting useful SAXS data on much more rapid time scales, which can now go down to tens of milliseconds and in special cases even faster than a millisecond.

Despite these enormous advances in data acquisition for SAXS, use of the technique in structural molecular biology has been relatively slow in developing. This may be attributed, in part, to the tremendous growth in the use of X-ray crystallography to determine biomolecular structures with Angstrom resolution and also the relative scarcity of synchrotron radiation centers with beamlines devoted to SAXS.

In this review we focus on three areas of recent progress in the application of SAXS to the study of changes of conformation of biomolecules.

Sebastian Doniach was born in Paris, France, in 1934 and educated in England where he received his B.A. degree in Mathematics from Cambridge University and his Ph.D. degree in Theoretical Physics from the University of Liverpool. After teaching in London University (at Queen Mary College and at Imperial College), he joined the Stanford University faculty in 1969. Doniach was Prinicpal Investigator of the proposal to utilize synchrontron radiation from the electron-positron storage ring SPEAR at the Stanford Linear Accelerator Center (SLAC). He was the founding director of what was subsequently named the Stanford Synchrontron Radiation Laboratory (SSRL), from 1973–78. He later served at the first chair of the SSRL division of the SLAC faculty after the two labs merged. Doniach has published extensively on theoretical physics of condensed matter and on experiemental and theoretical biophysics and biophysical chemistry.

10.1021/cr990071k CCC: $36.00
Published on Web 04/25/2001

These are as follows: (1) the use of SAXS to advance our understanding of the protein folding problem for small globular proteins using static X-ray measurements (we also mention recent work on RNA folding); (2) the use of SAXS to study the kinetics of protein folding through the use of time-resolved measurements (these techniques also help in observing the kinetics of protein misfolding leading to the formation of dimers and higher oligomers); and (3) recent advances in algorithmic methods for obtaining low-resolution three-dimensional density maps for protein complexes from the one-dimensional solution scattering data.

SAXS measurements provide global information on the size and shape of biomolecules in solution in contrast to other methods which generally measure local probes. This knowledge has proved to be very useful in characterizing global changes in size and shape of partially folded states of proteins and (very recently) of RNA.

One approach to understanding the nature of the free energy landscape of biomolecules is that of stabilizing partially folded intermediates states in thermodynamic equilibrium as a function of solvent conditions. For proteins, such conditions for destabilizing the native state include varying the pH, salt concentration, and/or denaturant concentration. In the case of RNA, the folded state of oligomers such as ribozymes may be destabilized by removal of Mg^{2+} counterions from the solution. For proteins, SAXS has proved to be a useful tool in helping to characterize the resulting molten globule states. At higher levels of denaturation and/or destabilization of protein native states by mutations, the ensemble of unfolded conformations may nevertheless contain considerable internal structure relative to the random coil state. Here again, SAXS provides structural information which is complementary to that obtained by other methods.

Studies of the kinetics of protein folding have been stimulated in recent years by the idea that transition states for folding of complex molecules such as a protein may comprise broad ensembles of states and lead to multiple parallel pathways for folding. Recently, experimental studies of folding using optical spectroscopy and hydrogen−deuterium exchange have been complemented by use of time-resolved SAXS. Time-resolved SAXS measurements provide information on global changes in size and shape of the protein as it goes through a number of conformational changes en route to the native state, in contrast to other methods which generally measure local probes. In addition, time-resolved measurements of forward scattering intensity for a protein undergoing folding or other conformational changes yield a direct way to measure changes in the hydration shell of the protein, thus yielding quantitative information on the changes in solvent accessible surface area of the protein as a function of degree of unfolding.

Protein refolding experiments often lead to self-association and formation of dimers and higher oligomers since, upon dilution of denaturant to trigger the refolding, exposed hydrophobic surfaces are no longer protected. Experiments where SAXS has proved to be useful in studying the resulting formation of dimers and higher oligomers will be briefly reviewed. There is a clear potential here for measurements which can help in understanding the processes of protein misfolding which lead to the formation of amyloid aggregates, of considerable importance in connection with a number of depository diseases including Alzheimer's, Parkinson's, and prion diseases.

SAXS has traditionally been used to provide geometrical characterization of protein complexes in solution, and a considerable amount of structural work is ongoing, which, however, we will not review here. Instead, we will focus on the substantial progress which has taken place over the past few years in developing algorithms for reconstruction of low-resolution three-dimensional density maps from the one-dimensional SAXS data. This capability, coupled with the use of improved synchrotron radiation X-ray sources and improved detection systems, makes SAXS a powerful tool for molecular structure determination at low resolution and for structural studies of conformational changes, which can be complementary to results obtained by electron microscopy.

II. What SAXS Measures for Macromolecules in Solution

The importance of SAXS for studying macromolecular structure is its ability to obtain global size and shape information for macromolecules in solution. This ability is particularly relevant to determining global information about partially folded or unfolded states of macromolecules, such as those that occur in protein folding, where X-ray crystallography cannot be applied and other techniques give local rather than global structural information. In this section we review some of the fundamentals of how structural information is obtained from the measured SAXS profile, with particular reference to the determination of changes in the structure of a protein which occur when it folds from a denatured or expanded state to its native conformation.

Protein folding has been studied by a wide variety of physical probes, including H−D exchange, UV circular dichroism, and various other optical probes. SAXS provides a complementary tool to investigate the degree of folding of proteins since it measures very directly the radius of gyration, R_g.

Measurement of SAXS for a protein in solution is, in principle, a very straightforward process. The set up for SAXS measurements at a synchrotron is illustrated in Figure 1. Unlike in crystallography, solution scattering is isotropic, so no particular sample handling is required. In practice, however, the need to avoid interparticle interference and radiation damage makes the technique subject to a variety of complications leading to poor signal/noise which are further discussed below. These complications lead to considerable sensitivity to beam drift due to temperature variations in the beam line optical elements which make background subtraction a particularly delicate issue.

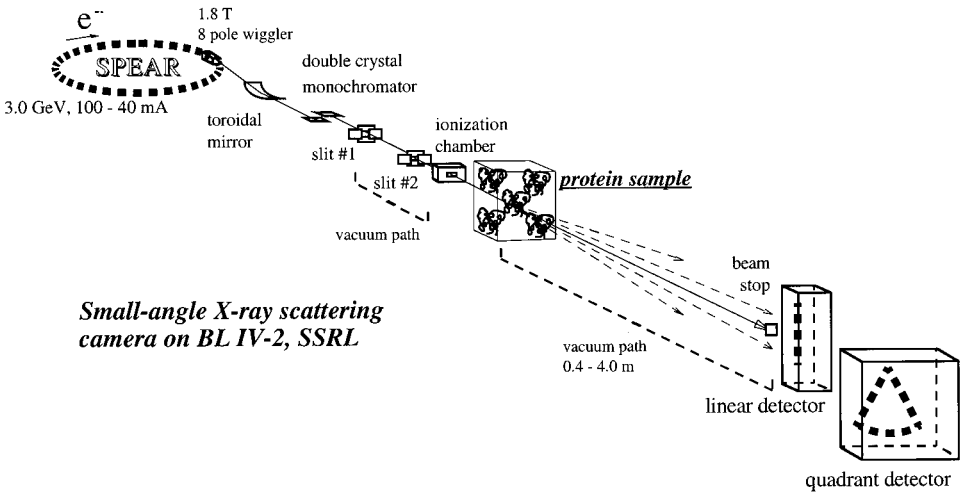

Figure 1. Schematic for beam line 4−2 at SSRL. Although the set up is simple in principle, parasitic scattering (slit scattering) and beam drift due to temperature variation in the beam line optical elements (monochromator, mirror) make background subtraction a delicate issue. (Figure courtesy of Hiro Tsuruta.)

The dependence of scattered intensity on the scattering wave vector, $s = (2/\lambda)\sin(\theta/2)$, where θ is the total scattering angle relative to the forward direction, for a molecule in solution, is given in the first approximation by the Debye formula

$$I(s) = \sum_i f_i^2 + \sum_{i<j} f_i f_j \frac{\sin(2\pi s r_{ij})}{2\pi s r_{ij}} \qquad (1)$$

where f_i is the atomic scattering factor for atom i. In practice f_i is well approximated for small angles of scattering by the electron number. $r_{ij} = \sqrt{|\vec{r}_i - \vec{r}_i|^2}$ is the distance between atoms i and j. This simplified formulation is particularly useful for rapid estimations of the SAXS profile of a molecule whose atomic coordinates are known.

The exact formulation of the scattering profile involves the contrast between the X-rays scattered from the molecule in the solution and the scattering from an equivalent volume of solvent displaced by the protein

$$I(\vec{q}) = \frac{1}{4\pi}\int d\Omega_{\vec{q}} \left| \int d^3 r e^{i\vec{q}\cdot\vec{r}} (\rho_{\text{protein}}(\vec{r}) - \rho_{\text{solvent}}(\vec{r})) \right|^2 \qquad (2)$$

where $2\pi s = |\vec{q}|$, $d\Omega_{\vec{q}}$ is the element of solid angle subtended by $\vec{q}$, and $\rho(\vec{r})$ is the electron density.

For proteins, the electron density is approximately 0.45 $e^-/Å^3$, while for water it is 0.33 $e^-/Å^3$. The square of the difference of these two numbers gives an estimate of the scattering strength for a protein in solution, which turns out to be on the order of 10% of what the scattering would be from a protein in vacuo. This is part of the reason that SAXS from biological molecules, particularly small proteins, has a relatively low counting rate. Furthermore, to obtain the SAXS profile for the monomeric species, interparticle interference effects need to be avoided, and this is achieved by going to relatively low concentrations of protein on the scale of 0.5 mM. This again reduces the counting rate obtainable for a given amount of radiation damage due to X-ray induced free radical formation. To compensate for this damage, the measurement of SAXS profiles for small proteins involves relatively large quantities of protein (on the scale of several milligrams) which are exposed in a flow cell. In the future it may be possible to limit the quantity of protein by measuring SAXS at low temperatures under vitreous ice conditions in order to minimize radiation damage as is done routinely in protein crystallography. However, this technique has not been developed to date for SAXS applications.

The other important consequence of eq 2 to be noted is that water molecules within 3 Å or so of the protein surface will generally be slightly more densely packed (on the order of 5−10%) than in the bulk solvent. Thus, the overall scattering profile for a protein in solution also takes account of the hydration shell. This point will be discussed further below. A very convenient piece of software, CRYSOL, due to D. I. Svergun and collaborators,[2] is available for evaluation of eq 2 given a pdb file, which includes estimates of the effects of the hydration shell.

In the small scattering angle limit, the scattering profile *I(s)* may be expanded to give the Guinier formula for the linear dependence of the log of I vs the square of the scattering vector, s^2

$$\ln(I(s)) \cong \ln(I(0)) - 4\pi^2 s^2 R_g^2/3 \qquad (3)$$

Thus, in the low-angle region the scattering profile $I(s)$ behaves like a Gaussian whose inverse width, $(2\pi s)^2 R_g^2/3$, is a model free measurement of the radius of gyration, R_g, where the only parameter is the X-ray wavelength.

It should be noted that under extreme denaturing conditions (i.e., in a "good" solvent from the viewpoint of polymer physics), the Guinier approximation starts to break down and the Debye approximation gives a better estimate of R_g

$$I(s) \cong I(0) 2(x - 1 + e^{-x})/x^2 \qquad (4)$$

where $x = (2\pi s)^2 R_g^2$ (see Calmettes et al.[3]).

An additional way of presenting SAXS data which has proved to be of semiquantitative usefulness for

partially folded or denatured states of proteins (or other macromolecules such as RNA) is a measure of compactness obtained from examination of the Kratky plot. Here the scattering profile intensity versus scattering wave vector is multiplied by the square of the scattering wave vector. As shown by Kratky, the scattering profile for a random coil state of a polymer falls off as $1/s$. This may be understood at a very basic level in terms of the scattering from narrow cylinders. Since the random coil is made up of quasirandomly oriented segments of the polypeptide chain whose length is on the order of the persistence length, roughly 8−10 Å, the interference pattern of the scattering from each segment cancels out due to their random orientation and overall scattering intensity falls as $1/s$. Hence, in the Kratky plot of $s^2I(s)$, the curve rises linearly at larger angles and so may be very clearly distinguished from that of a compact protein where, following the work of Porod, the clearly defined surface of the compact protein leads to a scattering profile which falls as $1/s^4$. The result is that native states or partially folded states with a relatively small amount of random coil disorder show a pronounced peak in the Kratky plot while fully denatured states of proteins, which are not constrained by disulfide bonds, show an upward rising feature in the Kratky plot (see Figure 2).

In addition to measuring R_g and giving indications of degree of compactness, the scattering profile may be expressed quite generally in terms of a pair distribution function *p(r)*

$$I(s) = \int_0^{r_{max}} dr p(r) \frac{\sin(2\pi sr)}{2\pi sr} \tag{5}$$

Within the Debye approximation, *p(r)* measures the distribution of pairwise distances within the molecule

$$p(r) = \sum_{i>j} f_i f_j \delta(r - |r_{ij}|) \tag{6}$$

so that r_{max} represents the longest interatomic distance in the molecule. More generally, *p(r)* is the density contrast pair correlation function of the molecule.

Derivation of *p(r)* from the measured intensity, *I(s)*, turns out to be a numerically ill-conditioned process. Because r_{max} is finite, it is formally possible to express *p(r)* as a discrete Fourier series. Since SAXS data falls off very fast with *s* (in the case of a compact molecular conformation), the number of terms which it is possible to define for such a Fourier series is, in principle, limited by Shannon's theorem. However. this reasoning leads to some ambiguity since expansion in a different set of basis functions than a Fourier series can lead to a *p(r)* function which differs quite a bit from that derived in terms of a finite Fourier sum. A very useful piece of software, GNOM due to Semenyuk and Svergun,[4] resolves this ambiguity by using a nonlinear least-squares approach. Here *p(r)* is represented as a set of bins, each of width d*r*, and the values of *p(r)* within each bin are treated as optimization variables. Since this can lead to more variables than can be justified by Shannon's theorem for scattering data measured of a give range, $s < s_{max}$,

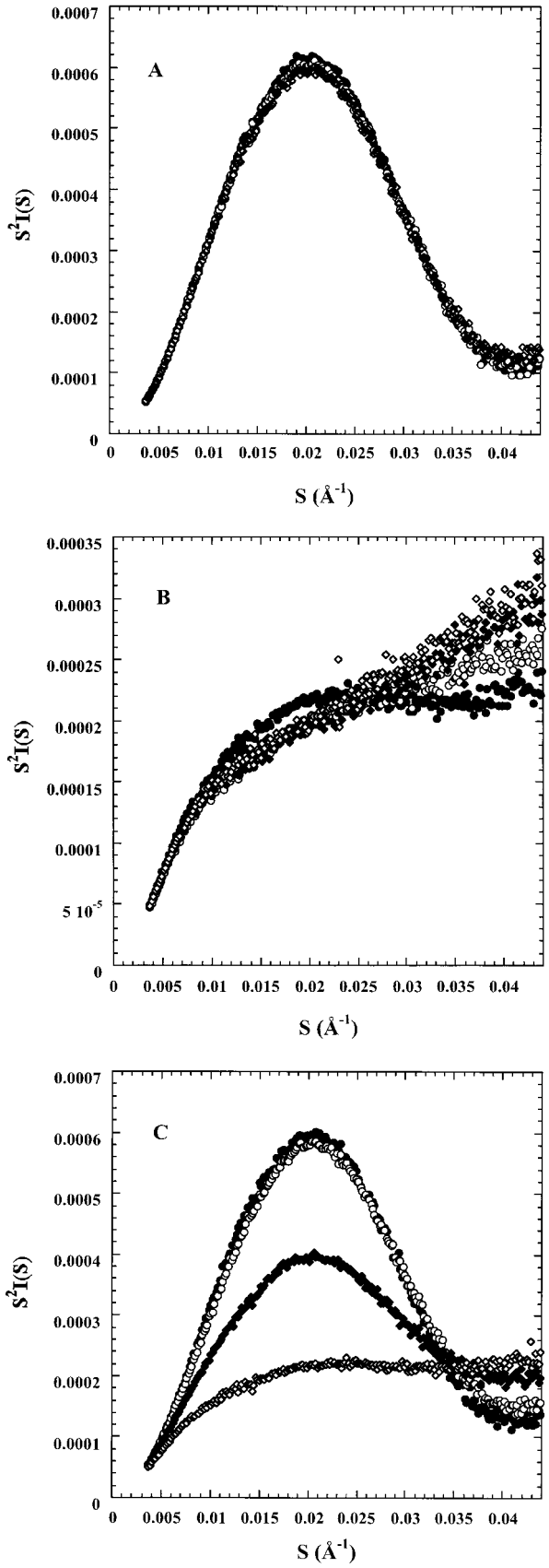

Figure 2. Kratky plots for cytochrome c^{27} showing cooperative unfolding. (A) Native conditions: 0.6 M (●), 1.2 M (○), and 1.8 M (◊) GdnHCl. (B) Denaturing conditions: 3.0 M (●), 3.6 M (○), 4.2 M (♦), and 4.8 M (◊) GdnHCl. (C) Cooperative unfolding region: 1.8 M (●), 2.2 M (○), 2.6 M (♦), and 3.0 M (◊) GdnHCl. (Reprinted with permission from ref 27. Copyright 1998 American Chemical Society.)

a regularization principle must be applied in which $p(r)$ is constrained to be a positive function and to have a gradient dp/dr, which is constrained to vary slowly. This approach to inverting an ill-conditioned integral transform[5] nevertheless contains some degree of arbitrariness in terms of the amount of constraint applied in the minimization and of the value of the parameter r_{max} used in defining the objective function. This degree of arbitrariness implies that the value of r_{max} is not very robust and may not be very well determined. Nevertheless, GNOM has proved to be a very useful tool for a semiquantitative analysis of SAXS data.

The fundamental question of how much structural information is encrypted in the SAXS data was discussed in terms of Shannon's theorem by Taupin and Luzzati.[6] They argued that for typical scattering profiles, about 15 basis functions could be combined linearly to represent a SAXS profile. More recently, advances in algorithmic decoding of SAXS data to give low-resolution three-dimensional density maps have called this linear limit into question. As discussed in section VIII below, such algorithms rely on nonlinear optimization and evidently extract quite a bit more structural information than is represented by the Shannon's theorem estimates based on linearity.

III. The Protein Folding Problem: Two-State Folding vs Partially Folded Intermediates in Thermodynamic Equilibrium

A. Overview

Because recent advances in applications of SAXS have been particularly relevant to improving our understanding of partially folded states of proteins which are central to the physical chemistry of protein folding, here we give a brief review of some of the basic issues in the protein folding problem before entering into details of the SAXS applications in this field.

The work of Anfinsen and others in the 1950s[7] on the reversible renaturation of denatured ribonuclease has led to a wealth of studies demonstrating that many, or perhaps most, small globular proteins may be reversibly denatured and renatured to the functional native state conformation. Despite the relatively small stabilization free energy on the scale of around 20 kcal/mol for a small globular protein of 100 residues or so, protein folding is highly cooperative. Therefore, it is natural to assume that the n residues in the primary amino acid sequence of a natural protein have been evolutionarily selected out of 20^n possible choices in sequence space so as to exhibit "foldability" as part of its functional attributes. Although protein folding in vivo is very often associated with the action of various chaperonins, in vitro experiments have generally demonstrated the relative thermodynamic stability of the natively folded state.

This idea of sequence selection for foldability has been born out by very simple theoretical models based on chains of beads placed on a lattice developed by Chan and Dill[8] and Shakhnovich and collaborators.[9] The gist of the theoretical results is that the selection of sequences which fold in such a way that the hydrophobic residues are in the interior of the protein and hydrophilic residues on the outside will, in general, lead to an "energy gap" separating the native state from the lowest free energy partially folded states. This may also be thought of in terms of the existence of a folding temperature which is higher than the temperature for being trapped into glassy intermediates, as developed by Wolynes and collaborators[10] and by Klimov and Thirumalai.[11]

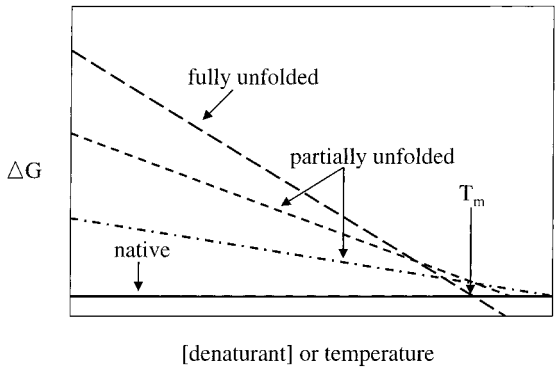

Figure 3. Schematic plot of ΔG for various thermodynamic states of a globular protein as a function of denaturant concentration or, equivalently, temperature.

One of the striking features of the thermodynamics of protein denaturation and renaturation is the fact that it may often be described with quite high accuracy by a two-state model in which the only important ensembles of protein conformation determining the thermodynamics of denaturation are an unfolded ensemble and the native ensemble. Fundamental understanding of this two-state nature for folding of small globular single-domain proteins comes from the use of NMR to measure hydrogen−deuterium exchange rates and thus determine protection factors for backbone amide protons for both the native state and a hierarchy of partially unfolded and fully unfolded states. The key observation, which comes from the work of Englander and his group,[12] is that the dependence of the relative free energy ΔG for partially unfolded states on denaturant concentration (or temperature for thermally induced denaturation) is generally highest for the fully denatured or random coil state. As shown schematically in Figure 3, the slope of the ΔG line for the quasirandom coil state (usually termed the m-value for chemical denaturation) is a measure of the entropy of the state which, by definition, is larger than that of the partially unfolded states. Hence, we can understand the two-state nature of denaturation in terms of sequence selection which guarantees that in the vicinity of the melting temperature the sequence of amino acids has gone through an evolutionary selection process which guarantees that ΔG for the partially folded states at T_m is finite, thus the so-called "energy gap" referred to above.

B. Partially Folded Intermediates

As emphasized above, it appears that protein sequences have been evolutionarily selected to ensure

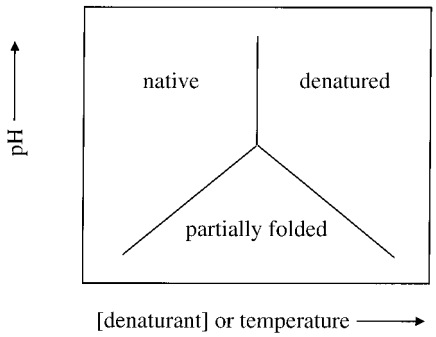

Figure 4. Schematic "phase diagram" for globular proteins as a function of solvent condition.

that only two thermodynamic states are important for protein folding under normal physiological conditions. However, by varying the chemical and/or physical condition away from the physiological, many studies have shown that partially folded states of proteins may be significantly populated. These have often been referred to as "molten globule" states of proteins.[13,14] Because structural details of such equilibrium states can be studied by a variety of physical techniques including SAXS, it is tempting to make the hypothesis that this structural information will be relevant to the kinetics of protein folding and that intermediates of the molten globule type may be populated (more or less briefly) along the folding "pathway". Thus, one of the most fundamental experimental goals in protein folding is that of comparing structural data for partially folded states which are stabilized in equilibrium, with similar data for on-pathway intermediates which are populated kinetically during the folding reaction.

Conditions which have been shown to destabilize the two-state nature of protein thermodynamics include lowering the pH, varying both pH and salt concentration,[15] varying pH and temperature, destabilizing the native state by mutations, in particular, deleting a few residues, or removal of the heme group in the case of apomyoglobin, and application of hydrostatic pressure. These destabilizing tendencies can be summarized in the form of a schematic "phase diagram" (see Figure 4). It is important to note that the protein is a small mesoscopic object, so that the usual concept of phase stability which applies in the thermodynamic limit of large systems is not strictly applicable to proteins. However, because of the cooperative nature of the folding leading to its two-state character, it does make sense to use approximate concepts from thermodynamics to classify the various observed phases of proteins in equilibrium.

An example of deviation from two-state folding under conditions of low pH where the Coulomb forces due to protonation of imidazole side chains of histidine residues will tend to stabilize a partially folded state is that of Barrick and Baldwin.[16] They studied the equilibrium states of apomyoglobin as a function of both pH and denaturant (urea) concentration using UV CD as a measure of helix content. For pH below about 5.5, they saw clear evidence of a three-state equilibrium in which native, unfolded, and partially folded states of apomyloglobin coexist. (See also a recent NMR study of residual structure in apomyoglobin at pH 4.[17])

IV. Use of SAXS to Study Partially Folded States of Proteins

A. Chemical Denaturation

As pointed out above, the thermodynamic state of a protein depends sensitively on environmental conditions including temperature, pressure, pH, salt concentration, and concentration of denaturants, such as urea, or guanidinium hydrochloride (Gdn-HCl). As these conditions are varied, two or more populations of protein conformations may coexist. Early studies of protein denaturation using SAXS include work by Damaschun et al. on apo-cytochrome *c*[18] and by Kataoka et al. on holo-cytochrome *c*.[19]

In the case of two-state folding, the SAXS profile will be a linear superposition of the native state profile, $I_{nat}(s)$ and of that of the unfolded state, $I_{und}(s)$.

$$I(s) = f_{nat}I_{nat}(s) + f_{unf}I_{unf}(s) \tag{7}$$

Hence, as the concentration of denaturant is varied in a titration series, the relative contributions of I_{nat} and I_{unf} to the scattering profile will vary.

In particular, we see that on expanding the equation for the case of a two-state system, in the Guinier region of small *s*, the square of the radii of gyration will vary as

$$R_g^2 = f_{nat}R_{g\text{—}nat}^2 + f_{unf}R_{g\text{—}unf}^2 \tag{8}$$

As the state of the protein is titrated by varying the denaturant concentration, the scattering profile is a linear superposition of a native fraction and a random coil fraction if the denaturation process is truly two-state. The two-state nature of such titration curves may be quantitatively analyzed in terms of a singular value decomposition (SVD) of a series of scattering profiles. If the scattering profile is a linear combination of two scattering curves each of which are independent of denaturant concentration, then by applying singular value decomposition to the series of scattering profiles measured in different denaturing concentrations, the occurrence of only two significant singular values will show that only two principle value components are present in the ensemble and that all higher order singular values only represent noise in the data. On the other hand, if there is a sizable population of partially folded intermediates between the native and the unfolded ensemble, then this will show up as a significant contribution to the third and possibly fourth singular values of the matrix of scattering profiles (see Appendix).

Application of this idea to the analysis of SAXS data was first tested by Chen et al.[20] in a study of hen egg white lysozyme (HEL) unfolding under acid conditions (pH 2.9). The authors studied the equilibrium state of lysozyme as a function of urea concentration both by SAXS and by far-UV CD at 222 nm, which gives a measure of the degree of α-helical character. It was found that the two curves differed

significantly, with the radius of gyration showing an increase at lower [urea] than did the far-UV CD. On performing an SVD analysis on the series of SAXS profiles at different [urea], clear evidence was seen for a third singular value, indicating the presence of three thermodynamically distinct states. By fitting to a three-state thermodynamic model, it was possible to track the appropriate linear combinations of the SVD orthogonal basis vectors (see Appendix) which corresponded to the SAXS profiles of the native, intermediate, and unfolded states, respectively. In this way it was possible to extract a SAXS profile for the intermediate alone. The R_g was found to be 19 Å, in contrast to that of the native R_g (15 Å) and of the denatured state (22 Å). (Note that owing to the four sulfur bridges in lysozyme, the "denatured" state is still quite compact relative to the ideal random coil state whose R_g would be expected to exceed 35 Å.)

Examination of the Kratky plot of the intermediate showed that while still showing some signs of random coil character in that the Kratky plot fell off slower than the native at larger angles, the intermediate is clearly quite compact. The measured R_g of the intermediate, coupled with the far-UV CD data is consistent with the "α-domain folded, β-domain unfolded" model proposed by Radford et al.[21] for the long-lived kinetic intermediate of HEL based on pulsed HD exchange measurements.

Thus, use of SVD combined with simple structural indicators such as the R_g and the Kratky plot can provide a unique tool with which to characterize the intermediate partially folded states of proteins.

B. Denaturation as a Function of Temperature and Pressure

In addition to the study of thermodynamically stable partially unfolded intermediates of globular proteins as a function of chemical denaturants (urea and guanidine hydrochloride) and lowered pH as discussed above, SAXS has also provided a tool to study the effects of other denaturing agents. Traditionally the thermodynamics of denaturation has been extensively studied using calorimetry. However, a major problem arises here: increasing temperature exposes hydrophobic regions of the protein, which then lead to aggregation in the absence of the stabilizing effects of chemical denaturants or of pH. SAXS has the advantage of not only measuring the size and shape of the protein being studied, but in the forward angle region can also give a measure of the pairwise probability distribution between monomers in a more concentrated solution as a result of interparticle scattering.

This has been worked out in the case of hen egg white lysozyme (HEL) by Arai and Hirai.[22] They show that as the pH is lowered, increased Coulomb repulsion between monomers of HEL leads to good temperature reversibility, even at 5% weight per volume, below pH 5. They also show that intramolecular structural changes corresponding to interdomain fluctuations start at temperatures quite a bit below the main unfolding transition as determined calorimetrically. Thus, even though the effects of these intramolecular changes are not directly observable in the differential scanning calorimetry thermograms, they can be monitored from the SAXS data, showing that the simple two-state model for denaturation is not applicable in thermally induced denaturation of HEL.

Hydrostatic pressure is also well-known as a denaturing environment. Work by Panick et al.,[23] using SAXS and FTIR techniques to look at pressure-induced unfolding of wild-type staphylococcal nuclease (SNASE), shows that pressures up to 3 kbar lead to an approximately 2-fold increase of the radius of gyration of the native protein from 17 to 34 Å. Furthermore, these changes could be observed on a time-dependent basis. Application of a pressure jump from 1 bar to 2.4 kbar at 20 °C led to an increase of R_g in a first-order manner over a time scale of approximately 30 min. The pair distribution function $p(r)$ under native conditions exhibits the maximum real space dimension, r_{max}, of 45 Å. Under pressure, this increases to about 70 Å at 2 kbar and about 112 Å at 3.25 kbar. Thus, at high pressure the $p(r)$ function is much more elongated (ellipsoidal) than in the normal state. However, the r_{max} value of 112 Å is smaller than would be expected for a fully extended random coil.

The time-resolved measurements show that although most of the expansion takes place on a faster time scale than the time resolution of the measurement (about 10 min per data point), continuing changes nevertheless occur at times longer than 30 min. Analysis of the FTIR spectral components reveals that the pressure-induced denaturation is accompanied by an increase in disordered and turn structures while the β-sheet and α-helix content drastically decreases. The pressure-induced denatured state above 3 kbar retains, nonetheless, some degree of β-like secondary structure, and the molecule cannot be described as a fully extended random coil. Thus, the pressure-induced state results in a profound disruption of the native structure and cannot directly correspond to a molten globule since there is very considerable destruction of the secondary structure in addition to that of tertiary packing.

It was observed that the pressure jump SAXS and FTIR experiments on SNASE show very similar relaxation times, indicating that no intermediate with significant secondary structure content was populated during the reaction. Thus, it would appear that formation of tertiary structure and secondary structure and collapse of the chain of SNASE in the folding process are all dependent upon the same rate-limiting step, which is a barrier crossing involving dehydration. This tends to contradict pictures of protein folding in which secondary structure formation precedes the formation of tertiary packing and in which equilibrium molten globules are hypothesized to resemble folding intermediates on the pathway to the native state.

C. Partially Folded Equilibrium Intermediates Induced by Helix-Stabilizing Agents

A class of interesting chemically induced changes in protein conformation are conversions from the

native fold to intermediates where the secondary structure character can be altered. These are induced by various alcohols, including trifluoroethanol (TFE) and methyl alcohol (MeOH). Efficiencies of destabilization of tertiary structures and stabilization of helical intermediates vary significantly depending on the protein and the kind of alcohol used. Two studies by Kamatari et al.[24] on methanol-induced denatured states of cytochrome *c* and by Hashino et al.[25] on trifluoroethanol- induced conformational transition of hen egg white lysozyme have used SAXS to probe the transforming effects of alcohols.

The effect of MeOH on cytochrome *c* studied by SAXS was found to be quite similar to that of the anion-induced molten globule state. Both states show a Kratky plot, indicating that the MeOH-induced state is a globular denatured state, which is quite compact (R_g of 17.7 Å in contrast with 32 Å for the urea denatured state). The CD spectrum showed that although specific tertiary interactions are absent in the intermediate state, they still retain secondary structure content similar to that of the native state. Increasing the alcohol concentration to 60% methanol led to a highly expanded but still strongly helical conformer without any tertiary structure, which shows Kratky plots quite similar to those observed for the acid-denatured state of cytochrome *c*.

The addition of TFE to lysozyme is observed to induce a cooperative transition to an intermediate state with an increased helical content and a radius of gyration which is 20% larger than that of the native while still retaining globularity. Singular value decomposition analysis of the Kratky plots of the SAXS profile indicates that two basis functions are sufficient to reproduce the observed data, so that the TFE induced transition can be described by a two-state model. Thus, the TFE-induced transition is more cooperative than the urea-induced denaturation discussed above,[20] in which a compact denatured intermediate was observed.

D. Nature of the Denatured State Ensemble

Recent studies have made it clear that the denatured states of proteins are not generally a simple random coil ensemble, as is often assumed. Instead, depending on denaturing conditions, significant correlations have been detected in denatured state ensembles. (For a review, see Shortle[26]).

As in the cases of the more compact partially folded intermediates discussed above, SAXS provides a powerful tool with which to characterize the more unfolded ensembles present in the so-called "denatured states" of globular proteins.

In one such study, Segel and collaborators[27] studied the unfolding of cytochrome *c* induced by guanidine hydrochloride. Above 6 M GdnCl, the Kratky plot showed a linear rise at large angles suggesting a random coil ensemble. However, the R_g of 32 Å does not correspond to what is estimated for a true random coil,[28] which is probably closer to 44 Å. The titration series was analyzed using SVD, and at least three singular value components were needed to represent the data. A fit to a three-state model led to partial

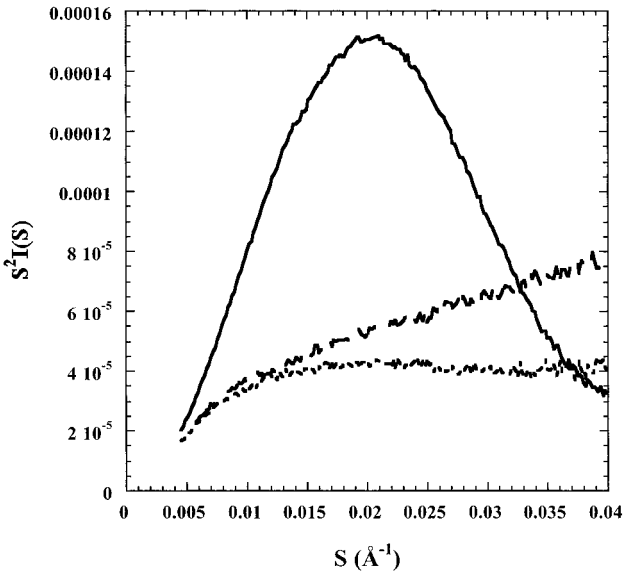

Figure 5. Scattering profiles for three thermodynamic states of cytochrome *c* extracted using SVD:[27] N (solid line), denatured (upper broken line), compact denatured (lower broken line). (Reprinted with permission from ref 27. Copyright 1998 American Chemical Society.)

occupancy of a second denatured ensemble at intermediate [GdnCl] of around 3 M. A Kratky plot of the scattering profile for this ensemble extracted by SVD showed a compact state with the plot falling at higher angles (Figure 5). Interestingly enough, however, R_g for the intermediate was found to be roughly the same as that for the 5 M denatured ensemble, indicating that even though the ensemble is compact by the Kratky plot, it still contains a very significantly extended fraction.

At a more quantitative level, the intramolecular $p(r)$ distribution measured by SAXS can, in principle, provide an experimental test of various models of the denatured ensemble obtained by NMR measurements. In two studies by Gillespie and Shortle[29,30] of a destabilized mutant of staphnuclease and by Zhang and Forman-Kay[31,32] of the drosophila kinase SH3 domain (drk SH3), measures of NOE distances and related indicators of local structure provide a basis for modeling an ensemble of relatively extended structures which still retain a fair amount of local correlations. For the destabilized staphylococcal mutant, Gillespie and Shortle found that the model ensemble showed striking similarity to the topology of the native state. For the drkSH3 domain on the other hand which undergoes a spontaneous conversion between the native state and a compact denatured ensemble, Zhang and Forman-Kay find that both non-native and native contacts occur in the exchange-unfolded state. It will be of interest to use the $p(r)$ determined by SAXS for these or other denatured states to test the model ensembles which come out of the NMR measurements.

Thus, there is increasing evidence that the nature of structural correlations in denatured ensembles is a complex issue depending on the protein in question, solvent condition, temperature, etc., and that seemingly denatured states of proteins may well contain significant residue−residue correlations which may

bias the folding process or possibly act as short-lived intermediates on a folding pathway. To help resolve these issues, SAXS can provide an important structural tool for studying intramolecular correlations in a manner complementary to the much more specific but local information obtainable from the NMR studies.

V. Kinetics of Protein Folding

A. Background

It's becoming increasingly clear from a theoretical point of view that the pathway for protein folding can be quite complex and may involve many parallel routes to the native state. Since folding of small proteins is very often two-state, there has generally been an assumption that there is a very significant free energy barrier to folding, defining a "transition state" ensemble through which the protein must pass in order to reach the native state. This follows the classical concept of nucleation of a first-order transition where a critical nucleus needs to be established in order for a raindrop to condense out of vapor.

Considerable progress has been made in studying the kinetics of folding as a function of single-point mutations in the amino acid sequence. Using a very simple Eyring-type formula, the two-state model suggests that the rate constant depends on the barrier height for forming the transition state. By studying the change of barrier height as a function of point mutations, Fersht and collaborators[33] were able to obtain estimates of which amino acids are involved in the transition state, hence, get some idea about the critical folding nucleus. It is becoming clear that there is quite a bit of heterogeneity—different proteins have different kinds of transition-state ensembles going from the rather broad ensemble of transition states for α-helix containing chymotrypsin inhibitor 2 (CI2)[34,35] to a fairly well-defined transition-state barrier for the β hairpin type src-SH3 domain protein.[36,37]

B. Time-Resolved Measurements of Protein Folding Using SAXS

The kinetics of how a protein evolves from a denatured state to its native state has been a subject of considerable study. Optical probes can measure at very short time scales from tens of nanoseconds in the case of temperature jump studies to seconds and more in the case of denaturant dilution experiments. Recently SAXS has been used as a means to determine changes in global size and shape in order to complement the local information resulting from the optical data.

Observation of protein folding using time-resolved SAXS has been difficult owing to the weak scattering signal and low X-ray intensity, even on synchrotron radiation beamlines. Work by Semisotnov et al.[38] on folding of proteins and model peptides was able to monitor the collapse of the molecule upon folding by observing changes in the integrated SAXS intensity but could not measure the change of R_g directly owing to insufficient X-ray flux.

Using X-rays from the wiggler beamline at SSRL[39] and with enhanced X-ray flux resulting from a use of a high-flux multilayer X-ray monochromator,[40] Eliezer et al.[41] were able to make a preliminary time-resolved measurement of the changes in R_g on folding of apomyoglobin allowing for a measurement of R_g within the first 100 ms of the initiation of folding. R_g for this collapsed state was found to be 23 Å, only 1 Å greater than the value of 22 Å for that of the refolded "molten globule" like state of the protein. By contrast, the unfolded state has an R_g of 34 Å.

In a more extended study, Chen et al.[42] were able to resolve changes in R_g on a time scale ranging from about 100 ms to seconds using stopped-flow mixing and signal averaging over many stopped-flow events. What was observed on folding lysozyme at pH 2.9 was that the folding process could be represented by a single rate constant on the time scale of 2 s^{-1}, although small changes were observed at a longer time scale consistent with proline isomerization taking place on a time scale of minutes.

C. Burst Phase Intermediates on the Folding Pathway Monitored by SAXS

One of the central issues in understanding the route to folding is that of initial hydrophobic collapse. Simple ideas from polymer physics suggest that when a polypeptide chain is taken from a good solvent (denaturing) to a poor solvent (renaturing), there will be an initial collapse of the chain to a relatively unstructured intermediate state that will then require considerable time searching in conformation space before nucleation of a transition state leading to folding. Evidence for such a hydrophobic collapse has been obtained from optical studies of tryptophan fluorescence changes during a refolding initiated by stopped-flow dilution of denaturant. Because of the dead time in stopped-flow mixers (on the order of a few milliseconds), what has characteristically been seen is a "burst phase" where the tryptophan fluorescence signal first observed after mixing has changed considerably from that of the denatured state. The final folding process may then occur a considerable time later, up to minutes in some cases.

SVD analysis of the series of time-resolved SAXS profiles for HEL obtained by Chen et al.[42] enabled the reconstruction of the SAXS profile for the intermediate formed at early time points immediately following the burst phase collapse. The resulting value for R_g of 19−20 Å and the form of the Kratky plot indicates a globular state which is relatively compact compared to the denatured state in 8 M urea with an R_g of 22 Å. This coupled with the considerable reduction of accessible surface area (ASA, see section V.G) shows that the compaction of HEL is about 30% complete immediately after the burst phase collapse. Nevertheless, the ASA is still 70% increased relative to the native, which is consistent with HD exchange data[21,43] indicating that the hydrophobic core residues are not completely shielded from solvent in the collapsed state and that the side chains are loosely packed with nonspecific tertiary interactions.

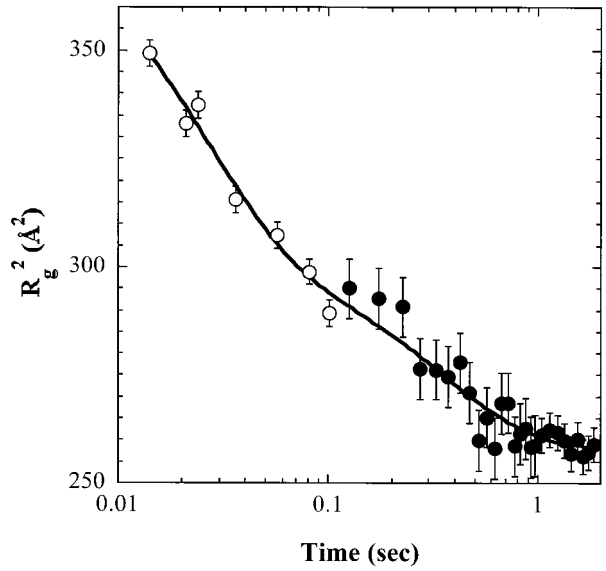

Figure 6. Time course of R_g for refolding of hen egg white lysozyme obtained by combining stopped flow data (●) with jet mixer data (○).[47] Final pH was 5.2.

D. SAXS Measurements of Folding at Millisecond Time Scales Using Continuous Flow Mixing

To get to time scales shorter than 100 ms, a continuous flow mixer was used in which the denatured protein solution and buffer were mixed under high pressure in a micrometer-sized volume, leading to turbulent mixing, and ejected as a jet of liquid from a nozzle.[44] Observation of SAXS a fixed distance away from the nozzle measures properties of the protein at a time delay relative to the time of mixing which is inversely proportional to the velocity of the jet. Thus, one can collect data at a fixed time interval from the mixing event with collection times of order minutes. This type of mixer has a long history.[45] Current designs can have a dead time on the order of 30 μs as observed in optical fluorescence studies.[46]

To have enough signal to observe protein folding with X-rays, the diameter of the final jet was increased to around 300 μm and a quartz capillary was used to avoid hydrodynamic instabilities in the jet which would otherwise lead to its break-up. In this setup the measurement dead time was around 10 ms owing to the decreased nozzle velocity of the emerging jet. To compare directly with earlier results of HD exchange by Radford et al.,[21] the experiment was carried out with a final pH of 5.2.

By combining the stopped flow data (Segel et al.[47]) with the jet mixer data, it was possible to measure R_g over a time scale ranging from 10 ms to 2 s (Figure 6). Evidence of parallel pathways for lysozyme folding was observed by fitting to a double exponential with rate constants of around 25 and 360 ms. When coupled with UV fluorescence measurements using an interrupted refolding technique (Kiefhaber [48]), a full SVD fit to the kinetics of the folding process could be made. (This could not be done from the SAXS data alone since it was necessary to know the native fraction in order to be able to separate the scattering profiles of the native from both that of the collapsed state and of the long-lived intermediate.) The fits to the data allowed a separation of the SAXS profile for the burst phase collapsed state, which was found to have an R_g of 19.6 Å, and for the long-lived intermediate, which was found to have an R_g of 17 Å.

The kinetics of lysozyme folding turns out to be strongly pH dependent with around 80% of the protein folding directly to the native at pH 2.9 while 20% goes through a long-lived intermediate. At pH 5.2, around 20% of protein folds directly to native while around 80% goes through the intermediate pathway.

Recently a novel continuous flow mixer was introduced in which the denatured protein and the diluting buffer are mixed by laminar flow in microchannels etched into a silicon wafer. Dilution of the denaturant occurs by diffusion rather than by turbulent mixing. In a preliminary experiment by Pollack et al.,[49] the folding of acid-denatured cytochrome *c* was observed by SAXS on a 500 μs time scale. It was concluded that a compact denatured state with an R_g of around 18 Å had formed by collapse from the denatured, random coil, state at times which were known from earlier optical studies to be on the order of 50 μs. Thus, although the burst phase was too fast to be observed directly, the existence of a collapsed intermediate could be seen, rather analogous to the lysozyme collapsed state discussed above.

E. Are Kinetic Intermediates on the Protein Folding Pathway Represented by Equilibrium Molten Globules?

Many workers who have studied equilibrium molten globule states have hypothesized that these represent local minima in the free energy surface which would be populated on the folding pathway.

This hypothesis is contradicted by the work on lysozyme. Segel et al.[47] found a radius of gyration of 19 Å for the collapsed kinetic burst phase intermediate relative to the 17 Å R_g for the long-lived intermediate. HD exchange work of Gladwyn and Evans[43] found that the burst phase intermediate retained very little α-helical structure. In contrast, the compact thermodynamic partially folded state observed by Chen et al.[42] in urea denaturation had an R_g of 19 Å yet still retained the α-helical fraction of the native as observed by far-UV CD. Thus, at least in this case, the free energy minimum represented by the compact urea-induced state does not appear to be the same as the long-lived kinetic intermediate observed on one of the two parallel pathways for lysozyme refolding.

F. Is a Burst Phase Collapsed State Present on the Folding Pathway of All Small Globular Proteins?

This commonly made assumption deriving from optical studies of folding of many small proteins is contradicted by the measurements on a number of two-state folding proteins which do not appear to show a burst phase.[50] A recent measurement by Plaxco et al.[51] shows that for a small protein L domain there is no detectable burst phase change of

radius of gyration (subject to measurement errors) relative to that of the guanidine denatured state, even though the folding is quite slow, on the order of 5 s. Thus, there is clear evidence in this case that no hydrophobic collapse occurred on removal of denaturant. This again suggests that the denatured state is by no means a random coil state and must contain significant internal correlations which prevent it from undergoing hydrophobic collapse under poor solvent conditions, possibly due to charged residues destabilizing the collapsed state.

G. Use of SAXS to Determine Changes in Hydration of Proteins along the Folding Pathway

In addition to a change of radius of gyration for lysozyme, Chen et al.[42] also observed a change in forward scattering intensity $I(0,t)$ as a function of time. Since the forward scattering depends on the contrast between the total electron number of the protein and that of the solvent it displaces (eq 2), the change in forward scattering amplitude signifies a change in total electron number since the protein is relatively rigid and should not alter its displaced volume by a significant amount. Thus, the 10% change observed by Chen et al. is likely to be the result of a change in the volume of the hydration shell of the protein on going from a denatured to a renatured state.

The importance of an average increase in density of the hydration layer relative to that in bulk water in determining SAXS intensity has been known for some time (Hubbard et al.[52]). As suggested by molecular dynamics studies of Gerstein, Tsai, and Levitt,[53] the surface layer of waters around a protein will have a density increase of 5−10%. Even though these waters are rapidly exchanging and hence may not all show up as crystallographically bound water, they will nevertheless alter the effective number of scattering electrons for the hydrated protein. Since the scattering depends on the contrast relative to bulk water, this relatively small change in water density can nevertheless lead to a relatively large change in scattering intensity since the latter goes as the square of the difference in electron number.

The protein hydration effect has been studied experimentally by Svergun et al.[54] by measuring both X-ray (SAXS) and neutron scattering (SANS) in H_2O and D_2O solutions for three different proteins. The key difference between SAXS and SANS arises from the fact that in H_2O proteins have a positive contrast (i.e., electron density) relative to bulk water, whereas the neutron scattering length of deuterium is much higher than that of hydrogen and the scattering length density of D_2O is higher than that of protein (negative contrast). Svergun et al. found that the SANS data are remarkably different from the SAXS data. This is particularly striking in the case of lysozyme, where the molecule apparently "shrinks" when going from H_2O to D_2O. Their results give an empirical measure of a first hydration shell which varies from protein to protein the case of the three proteins studied and whose density is on the order of 10% larger than that of the bulk solvent.

Thus, the effect of hydration on the forward scattering intensity, $I(0)$, gives us a measure of the accessible surface area (ASA) of the denatured protein relative to the ASAs for intermediate states. A simple estimate of the ASA for the denatured protein may be obtained by using the ASA for the central residue of a series of tripeptides. (However, this neglects possible reduction of ASA due to internal correlations in the denatured ensemble.) This was done for lysozyme and a result of 22 260 $Å^2$ was obtained, to be compared with 4484 $Å^2$ for the native state. This enables the determination of the ASA for the collapsed state first seen on renaturation in the experiments of Chen et al..[42] Apparently the ASA of the collapsed state, with a surface area of 7595 $Å^2$, is relatively close to that of the native compared to the ASA for the denatured protein.

H. What New Insights Has SAXS Provided for Understanding Protein Folding?

A principal goal of research in the protein folding problem is to seek out empirical rules, sometimes termed "the second genetic code", by which the primary polypeptide sequence dictates the three-dimensional fold of the native protein. As we have seen, evolutionary selection has narrowed enormously the possible space of primary sequences and "foldability" is an important driving force for this selection, in addition to selection for function in the biological system. The nature of this selection is a fundamental question.

Experimentally there have been two general approaches to investigating this question. One is to study the free energy landscape of the peptide chain by stabilizing partially folded intermediate states of proteins. These can vary all the way from compact molten globules which still retain a large fraction of the native structure to ensembles of largely unfolded conformations which characterize the various denatured states of proteins. The second general approach is to attempt to characterize the kinetics of the refolding process and, in particular, to obtain structural information on partially folded states which may be populated in the course of the refolding reaction.

Technological progress in recent years has enabled SAXS to provide a powerful tool which can be used in accomplishing each of these strategies. It is particularly useful in the protein folding problem because of its ability to give size and shape information for partially folded states. In this context, and taken as providing complementary information to the much more specific information provided by local probes, it can help delineate the free energy landscape of proteins and hence help elucidate which features are essential in determining the uniqueness and stability of the native conformation.

The main new information which SAXS has to offer for protein folding relative to that garnered with the traditional methods is that it gives a global measure of the size (R_g) and shape of the denatured and partially folded intermediates which either exist in thermodynamic equilibrium under appropriate sol-

vent conditions or are populated along the various folding pathways for a small globular protein in vitro.

As emphasized above, study of protein folding has traditionally been done by the use of local probes, as in the case of UV fluorescence from tryptophan subgroups, for example, or by the use of hydrogen−deuterium exchange coupled with NMR measurements. While NMR probes can give very important structural information in defining protection factors for the amide backbone and hence giving specific knowledge of which residues are buried in the course of the folding process, they give no indication of the global size and shape of the various folding intermediates. Thus, while HD exchange can localize backbone protection to certain residues, it can say very little about residues which remain exposed to solvent. As discussed above, NOEs and other NMR indicators of residue−residue correlations can help in constraining models of the ensemble of partially folded states. In this context, SAXS measurements of $p(r)$ can, in principle, provide tests for these structural models.

While SAXS is a relatively low-resolution measurement technique, it has been able to give precise answers to the questions of the degree of collapse of a protein which occurs during the folding process. Thus, using SAXS it has proved possible to measure directly changes in radius of gyration during the burst phase of folding of small proteins and to compare the size and shape of equilibrium molten globules with those of long-lived folding intermediates. An important example of this is the comparison of the initial collapsed state of HEL with that of the α-ordered, β-disordered long-lived intermediate. As emphasized above, the results obtained using SAXS tend to go against the hypothesis that molten globular equilibrium intermediates may also be populated as kinetic intermediates. This in turn calls into question the idea that a unique "free energy landscape" can serve to model both the thermodynamics and the kinetics of protein folding.

Interestingly enough, a lack of collapse for small proteins as seen by the lack of burst phase intermediates in optical measurements was also shown to be consistent with a lack of change of the radius of gyration for the small protein fragment protein-L on removing denaturant in a stopped-flow experiment. This highlights the issue of significant structural correlation in the denatured states of proteins. Here again, other techniques such as NMR and spin labeling have been used to show the existence of structural correlations in denatured states of proteins. SAXS measurements of denatured states of proteins have been able to distinguish unfolded random coil-like states from ensembles of partially folded states involving some degree of local structure, depending on denaturation conditions.

VI. RNA Folding Observed Using SAXS

SAXS has been used successfully over the years[55] to obtain low-resolution structural information on a variety of RNA oligonucleotides and RNA−protein complexes, among which the most notable are the various ribosomes.[56]

Interest in RNA folding has come to the fore in recent years following the discovery of ribozymes, i.e., RNA molecules that catalyze biochemical reactions.[57] Time-resolved oligonucleotide hybridization and hydroxyl-radical cleavage protection techniques that monitor local environments have been used to characterize the kinetic folding process of the *Tetrahymena* ribozyme.[58] The folding is found to be highly cooperative with Mg^{2+} concentration and also occurs in Ca^{2+} but not in Na^{+}, suggesting that it results from formation of a specific tertiary structure. Small-angle X-ray scattering (SAXS) can provide information complementary to these local techniques by providing data which allows, in a model-independent fashion, measurement of the overall size and shape of the RNA during folding.[72]

It has been proposed that the rate-limiting step for overall folding of the *Tetrahymena* ribozyme involves emergence from a kinetic trap, requiring the disruption of non-native or native structure. Conversely, it has also been suggested that a conformational search for long-range tertiary contacts limits overall folding. Time-resolved SAXS can help resolve these ambiguities since if the slow folding steps involve a search of conformational space by subdomains that have not yet formed stable tertiary structure, then a large fraction of the compaction would be expected to occur late in folding. Conversely, if a compact folding intermediate were formed quickly that must then rearrange or partially unfold in a slow step, much of the reduction in size would occur much earlier than completion of overall folding.

Russel et al.[59] recently reported an initial study in which time-resolved SAXS measurements were performed manually with an experimental dead time of about 1 min. Upon initiation of tertiary folding with the addition of Mg^{2+}, the R_g value was found to decrease from 74 to 51 Å with a minimum rate constant for the compaction of 3 min^{-1}, at least 20-fold faster than the overall rate constant for folding as determined from ribozyme functional activity.

This relatively rapid reduction in R_g relative to overall time for folding to the functional state shows that a compact intermediate or family of intermediates is formed early in folding, and the subsequent slow steps to form the native state result in only a small additional compaction. The intermediate is found to be nearly as compact as the final native form, so the slower folding steps are presumably rearrangements of this structure. These rearrangements may involve breaking and exchanging contacts from within the compact intermediate or transient opening of contacts that maintain the compactness of the intermediate, allowing less restricted motion of the core that can lead to a stable folded state. These results lead to a kinetic folding model in which an initial "electrostatic counterion mediated collapse" of the RNA is followed by slower rearrangements of elements which are initially mispositioned. Extending the SAXS measurements to shorter times will provide a critical evaluation of the rate at which the initial rapid electrostatic collapse starts the process of RNA folding.

VII. Protein–Protein Interactions, Formation of Dimers, and Higher Aggregates

SAXS is particularly sensitive to the formation of dimers and higher oligomers of biomolecules. Since the scattering intensity varies as the square of the molecular weight, the formation of a population of dimers from a monomer leads, in principle, to a factor of two increase in intensity (since there are one-half the number of dimers and each one scatters four times as strongly). Upon refolding proteins by dilution of denaturant, they are particularly susceptible to formation of dimers and higher oligomers since the hydrophobic interior of the protein is exposed in partially folded states before reaching the native state.

At micromolar concentrations where many optical studies of protein folding have been conducted, this is not a problem since the rate of dimer formation scales with the concentration. However, under millimolar concentrations used in SAXS experiments (and in many NMR experiments), formation of dimers and higher oligomers can be very significant and can make SAXS studies of the monomer folding process difficult or impossible, depending on the folding conditions.

The folding of cytochrome *c* is particularly susceptible to the formation of dimers and higher oligomers since in the native state the covalently bound heme group is strongly linked to a proximal methionine, residue number 80. As a result, dimers or higher oligomers can be formed by a domain swapping process in which the heme group of one molecule associates with the methionine of a second molecule. Such cross ligation processes can be to some extent inhibited by folding in the presence of imidazole, which preferentially binds to the heme, thus lowering the tendency for aggregation. However, a recent study by Segel et al.[60] showed that even in the presence of imidazole, transient dimers were formed in folding cytochrome *c* at millimolar concentrations which have a half-life that is strongly dependent on final denaturant and protein concentration. At low [GdnCl] of 0.6 M, dimers were observed to persist at times longer than 10 min, while for more elevated concentrations, shorter lifetimes were found of 14 s at a final [GdnCl] of 1.8 M and 0.6 s for 2.4 M before the protein finally reverted to the native state. In the last two cases, the final protein concentration was about 11 mg/mL. Similar results had been reported earlier for transient dimer formation in the folding of holomyoglobin by Eliezer et al.,[61] though for myoglobin, dimers were found to much less stable than for cytochrome *c*, presumably due to an absence of domain swapping.

Understanding protein misfolding which is of central importance in connection with amyloid disease and more generally in conformational changes associated with prion-like protein transformations[62] has the potential for elucidation by the use of SAXS. In preliminary studies, SAXS measurements on oligomerization of amyloid-producing proteins (Doniach and Fink, unpublished) information could be obtained on the kinetics of the oligomerization process and on the structure of the oligomers at low resolution. Thus, there is considerable potential for SAXS to help in understanding the early processes on the pathway to amyloid formation.

VIII. Reconstruction of Low-Resolution Three-Dimensional Electron Density Maps from One-Dimensional SAXS Data for Biomolecules

Despite the loss of angular information relative to protein crystallography for SAXS scattering profile, Svergun and Stuhrman[63] showed in 1991 that it is possible to obtain low-resolution three-dimensional density maps from one-dimensional SAXS data based on a spherical harmonic expansion. The essence of their method is based on an assumption that the electron density within a macromolecule such as a protein may be treated as a constant inside an envelope function defining the protein surface (in the case of DNA or RNA, such a density would differ from that of a protein). Then by expanding the function defining the molecular surface in spherical harmonics, Svergun and Stuhrman were able to show that a unique envelope function could be retrieved from the SAXS data using a nonlinear optimization algorithm.

The key to this discovery is that the method is nonlinear—for a linear expansion of the protein density in terms of a set of basis functions, there are an infinite number of solutions owing to the phase problem of X-ray crystallography. The remarkable thing about the work of Svergun and Stuhrman is that the nonlinear algorithm is capable of selecting a unique solution out of what is, in principle, an infinite number of feasible solutions. This result is essentially the same as that leading to the success of direct methods in protein crystallography in which the application of a physical constraint, namely, that the electron density is nonnegative, provides a filter which excludes solutions based on nonphysical phases. Software to execute the algorithm of Svergun and Stuhrman is now available.[64] More recently, Chacon et al.[65,66] were able to show that a three-dimensional density map could be obtained by a genetic algorithm in which a fixed number of point scatterers were rearranged in such a way as to simulate the correct scattering profile.

A new approach has been developed by Walther et al.[67] to allow for reconstruction of electron density maps for complex molecules in which the topology is not that of a sphere and where the restriction of a fixed number of point scatterers introduced by Chacon et al. is no longer necessary. In the algorithm of Walther et al., point scatterers are placed on a cubic lattice in a manner which conserves as well as possible the physical constraints of positivity and connectedness (implicit in Svergun and Stuhrman's assumption that the electron density interior to the protein is uniform). The essence of the algorithm of Walther et al. is to add nearest neighbors on the lattice to the ensemble of evolving point scatterers in such a way that the first choice for addition of neighbor is always a point scatterer with the least number of neighbors. In this way it is found that an initial ensemble is generated which very rapidly

leads to a radius of gyration comparable to that of the target molecule. Then as the algorithm proceeds it also enables the removal of point scatterers, as well as their addition, as a way to generate more complex topology than those derived from a sphere. It was found that by the use of this "give 'n take" algorithm that quite good density maps could be made for molecules as complex as the chaperonin protein assembly Gro-El, which is in the form of a barrel made up of two cylinders, each comprising seven identical proteins.

The resolution obtainable in such reconstruction methods is limited by the maximum scattering vector s_{max}, which is typically on the order of 0.04 Å^{-1}. In practice, it is found that lattice spacings on the order of 16 Å lead to a satisfactory reconstruction. In the application of this method, the spherical symmetry of the original data is broken by the algorithm and once broken allows for a low-resolution unique three-dimensional map. Each time the algorithm is run, as a result of the use of stochastic selection in choosing equivalent sites on the growing ensemble for where to add or subtract a point scatterer, each three-dimensional reconstruction differs in orientation relative to the underlying cubic lattice. By optimizing the overlap between succeeding reconstructions, a density map may be built up which is reasonably dense in point scatterers. In addition to Gro-El, this algorithm has been applied to a number of other protein samples, among which the DNA binding protein SAP of yeast (switch activated protein)[68] has revealed a highly elongated structure in which the DNA binding units form the head of a "broom" in which the handle is a coiled coil leucine zipper construction with, interestingly enough, an actin binding motif at the distal end relative to the DNA binding region. Svergun and colleagues applied their spherical harmonic-based algorithms to similar three-dimensional reconstructions of ribosomal proteins.[56,69] Software for execution of the algorithm of Walther et al. is freely available at http://cmpharm.ucsf.edu/~walther/saxs/.

It is interesting to compare the use of SAXS for three-dimensional density map measurements to these of electron microscopy. Although EM is capable of much higher than 15 Å resolution in the case of periodic samples, it often is limited in resolution by the need for computer averaging in the case of samples in which the target biomolecules are not arranged in a lattice. Here SAXS may have an advantage within the confines of relatively low resolution of being able to monitor changes in size and shape in real time and without the need for freezing and/or staining. In principle, by extending the angular range of the measurements, the resolution may be improved. However, since the scattering from compact molecules falls off as $1/s^4$, doubling the resolution would involve increasing the irradiation by a factor of 16. This would then lead to a corresponding increase in the amount of sample needed or possibly could be obtained by inhibiting radiation damage by cooling the sample into a vitreous ice state as is done in cryoelectron microscopy.

IX. Acknowledgment

The author expresses his indebtedness to and appreciation of support from Keith Hodgson over the years. He acknowledges the technical support of Hiro Tsuruta in building the SAXS capabilities at SSRL. The writing of this article was supported in part by the NIH through the Biotechnology Research Resource at SSRL.

X. Appendix: Singular Value Decomposition for SAXS Profiles

Given a series of M SAXS profiles $I_m(s_i)$ measured at N scattering vectors, s_i, one can use the singular value decomposition algorithm (SVD) to represent each profile as a different linear combination of a set of K orthogonal basis functions $u_m{}^i \equiv u_m(s_i)$. (Note that SVD is readily available either as C or FORTRAN code[70] or in numerical analysis packages such a MATLAB or MATHEMATICA. See Henry and Hofrichter[71] for a helpful introduction to the use of SVD for analysis of spectroscopic data.)

To do this, the series of scattering profiles are arranged to form a matrix, $A(s_i,m) = [I_1,I_2,...,I_M]$, in which each column of the matrix is one of the scattering profiles. SVD transforms A into a product

$$A = UWV^T \tag{A1}$$

in which W is a diagonal matrix whose elements, the "singular values" of A, are arranged in descending order. For SAXS data, it is found that the associated orthogonal basis vectors which are the columns, u_m, of the matrix U, very rapidly lose structure, so that after the third or fourth column, the profile $u_m(s_i)$ simply represents noise. The SAXS profiles are now given in terms of the SVD matrixes by $I(s,m) = \sum_k w_k v_k{}^m u_k(s)$, where the w_k are the singular values and $v_k{}^m$ are the elements of the coefficient matrix, V.

The power of the SVD analysis is that it becomes very evident how many basis functions are needed to represent the data by inspection of the drop off of the singular values and the form of their associated basis vectors. Thus, for example, if the data represents a superposition of SAXS profiles from just two thermodynamic or kinetic components, the third and subsequent basis profiles will readily be seen to just represent noise. Conversely, if more than two ensembles of conformations are involved, then it becomes apparent that use of only two basis functions is insufficient to represent the data. It should be noted that the amplitudes of the higher basis functions fall off very fast with each succeeding orthogonal basis vector, so that, in practice, SAXS data, which are quite noisy relative to optical spectroscopy data for instance, usually have a hard time going beyond three or at most four independent components of the signal.

If one knows the dependence of the SAXS profile on the external conditions such as denaturant concentration for a titration series or measurement time in the case of time-resolved SAXS, then this knowledge may be used to extract SAXS profiles for the

various individual ensembles of conformations contributing to the profile series.

Segel et al.[27] used this approach to extract profiles for three different thermodynamic states of cytochrome *c* (see Figure 2) from the GdnHCl titration series of SAXS profiles. The form of the SVD basis functions found for the titration data are shown in Figure 7. To determine the appropriate linear combinations of these basis functions which define the SAXS profiles for each of the three thermodynamic states populated in the titration series, a three-state thermodynamic model was used in which the fractional occupancy of each of the three states was expressed in terms of a denaturant-dependent free energy, relative to that of the native, for each of the two denatured ensembles in the form $\Delta G_\alpha = \Delta G_\alpha{}^0 - m_\alpha[GdnCl]$, where m_α is the "*m*-value" resulting from the association of the denaturant to the backbone residues of the protein which are exposed in the partially or fully unfolded states denoted by the suffix α.

The fractional occupancy for each of denatured states at denaturant concentration c_m may be represented by the fraction

$$f_\alpha(c_m) = \frac{K_\alpha}{1 + K_1 + K_2} \qquad \text{(A2)}$$

where

$$K_\alpha = \exp(-\Delta G_\alpha/RT) \qquad \text{(A3)}$$

Using eq A2, a nonlinear least-squares fitting procedure may be used to fit the data in terms of the thermodynamic parameters by expressing the scattering profiles of each of the thermodynamic components in terms the of the occupancies of the different states. The profiles of the individual components are expressed, in turn, in terms of the orthogonal basis functions coming from the SVD algorithm.

Thus

$$I(s,c_m) = \sum_\alpha f_\alpha(c_m) I_\alpha(s) \qquad \text{(A4)}$$

where, in turn, the $I_\alpha(s)$ are expressed as

$$I_\alpha(s) = \sum_k w_k b_k{}^\alpha u_k(s) \qquad \text{(A5)}$$

Here w_k are the elements of the matrix of singular values and $b_k{}^\alpha$ are a set of coefficients. These are then determined in the fit to the experimentally determined $v_k{}^m$ by minimizing

$$\chi^2 = \sum_k^K \sum_m^M \frac{w_k{}^2 (v_k{}^m - \sum_\alpha f_\alpha(c_m) b_k{}^\alpha)^2}{\sigma_k{}^2(m)} \qquad \text{(A6)}$$

with respect to the $b_k{}^\alpha$ and the ΔG_α, where *K* is the number of relevant SVD basis vectors, *M* is the number of profiles measured, and $\sigma_k{}^2(m)$ are the variances of the SVD coefficients. Since SVD is a linear operation, the $\sigma_k{}^2(m)$ are obtained as linear combinations of the variances for the SAXS data

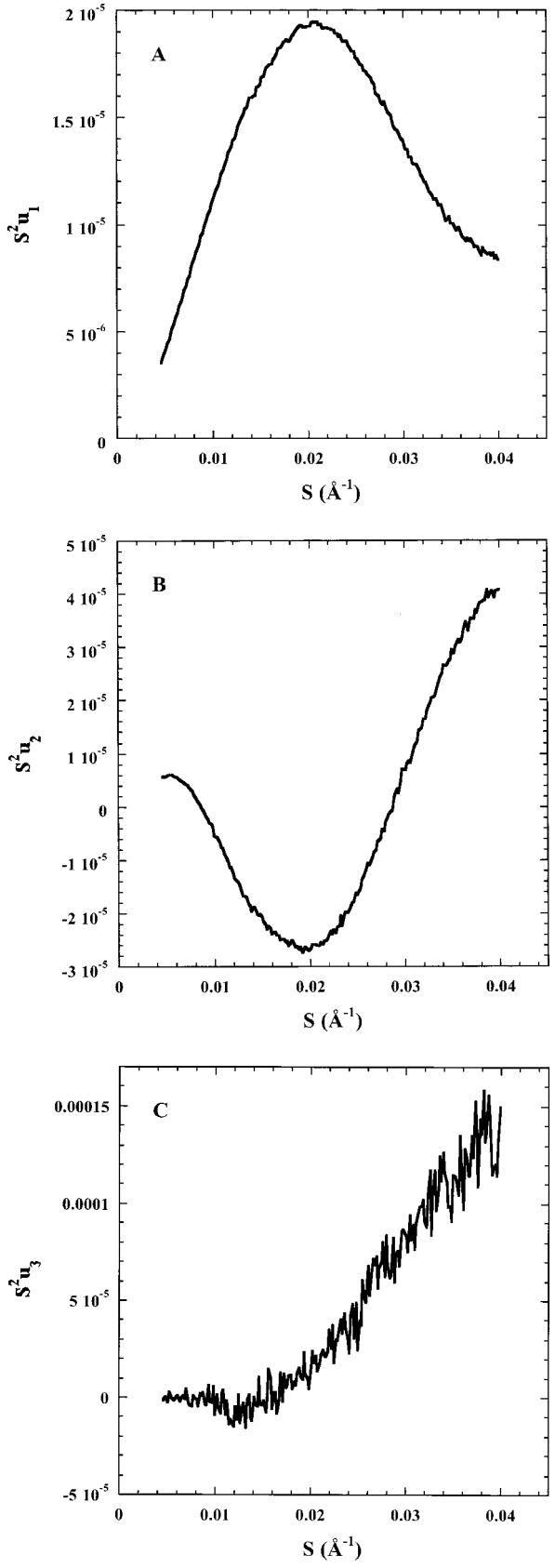

Figure 7. Kratky plots of singular value decomposition scattering basis functions for cytochrome *c* denturation series[27] (see Figure 5). It is important to realize that the scattering profiles of the various thermodynamic states are in each case different linear combinations of these orthogonalized basis functions: the basis functions are not, in themselves, scattering profiles. (Reprinted with permission from ref 27. Copyright 1998 American Chemical Society.)

points. From the fit to the thermodynamic parameters, Segel et al.[27] were able to extract the SAXS profiles for the individual thermodynamic components as given in Figure 5.

XI. References

(1) *Small-angle X-ray scattering*; Glatter, O., Kratky, O., Eds.; Academic: London, 1982.
(2) Svergun, D. I.; Barberato, C.; Koch, M. *J. Appl. Crystallogr.* **1995**, *28*, 768−773.
(3) Calmettes, P.; Durand, D.; Desmadril, M.; Receveur, V.; Smith, J. C. *Biophys. Chem.* **1994**, *53*, 105−114.
(4) Semenyuk, A.; Svergun, D. *J. Appl. Crystallogr.* **1991**, *24*, 537−540.
(5) Tikhonov, A. N.; Arsenin, V. Y. *Solutions of ill-posed problems*; Halsted Press ed.; Washington, Winston, NY, 1977.
(6) Taupin, D.; Luzzati, V. *J. Appl. Crystallogr.* **1982**, *15*, 289−300.
(7) Anfinsen, C. B. *Science* **1973**, *181*, 223−30.
(8) Chan, H. S.; Dill, K. A. *Proteins* **1998**, *30*, 2−33.
(9) Shakhnovich, E. I. *Curr. Opin. Struct. Biol.* **1997**, *7*, 29−40.
(10) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167−95.
(11) Klimov, D. K.; Thirumalai, D. *Proteins* **1996**, *26*, 411−41.
(12) Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W. *Science* **1995**, *269*, 192−7.
(13) Kuwajima, K. *Proteins* **1989**, *6*, 87−103.
(14) Ptitsyn, O. B.; Pain, R. H.; Semisotnov, G. V.; Zerovnik, E.; Razgulyaev, O. I. *FEBS Lett.* **1990**, *262*, 20−4.
(15) Uversky, V. N.; Karnoup, A. S.; Segel, D. J.; Seshadri, S.; Doniach, S.; Fink, A. L. *J. Mol. Biol.* **1998**, *278*, 879−94.
(16) Barrick, D.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 3790−6.
(17) Eliezer, D.; Chung, J.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2000**, *39*, 2894−901.
(18) Damaschun, G.; Damaschun, H.; Gast, K.; Gernat, C.; Zirwer, D. *Biochim. Biophys. Acta* **1991**, *1078*, 289−95.
(19) Kataoka, M.; Hagihara, Y.; Mihara, K.; Goto, Y. *J. Mol. Biol.* **1993**, *229*, 591−6.
(20) Chen, L.; Hodgson, K. O.; Doniach, S. *J. Mol. Biol.* **1996**, *261*, 658−71.
(21) Radford, S. E.; Dobson, C. M.; Evans, P. A. *Nature* **1992**, *358*, 302−7.
(22) Arai, S.; Hirai, M. *Biophys. J.* **1999**, *76*, 2192−7.
(23) Panick, G.; Malessa, R.; Winter, R.; Rapp, G.; Frye, K. J.; Royer, C. A. *J. Mol. Biol.* **1998**, *275*, 389−402.
(24) Kamatari, Y. O.; Konno, T.; Kataoka, M.; Akasaka, K. *J. Mol. Biol.* **1996**, *259*, 512−23.
(25) Hoshino, M.; Hagihara, Y.; Hamada, D.; Kataoka, M.; Goto, Y. *FEBS Lett.* **1997**, *416*, 72−6.
(26) Shortle, D. *FASEB J.* **1996**, *10*, 27−34.
(27) Segel, D. J.; Fink, A. L.; Hodgson, K. O.; Doniach, S. *Biochemistry* **1998**, *37*, 12443−51.
(28) Sosnick, T. R.; Trewhella, J. *Biochemistry* **1992**, *31*, 8329−35.
(29) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 158−69.
(30) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 170−84.
(31) Zhang, O.; Forman-Kay, J. D. *Biochemistry* **1995**, *34*, 6784−94.
(32) Zhang, O.; Forman-Kay, J. D. *Biochemistry* **1997**, *36*, 3959−70.
(33) Fersht, A. R. *Curr. Opin. Struct. Biol.* **1997**, *7*, 3−9.
(34) Jackson, S. E.; elMasry, N.; Fersht, A. R. *Biochemistry* **1993**, *32*, 11270−8.
(35) Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. *J. Mol. Biol.* **1995**, *254*, 260−88.
(36) Grantcharova, V. P.; Baker, D. *Biochemistry* **1997**, *36*, 15685−92.
(37) Grantcharova, V. P.; Riddle, D. S.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7084−9.
(38) Semisotnov, G. V.; Kihara, H.; Kotova, N. V.; Kimura, K.; Amemiya, Y.; Wakabayashi, K.; Serdyuk, I. N.; Timchenko, A. A.; Chiba, K.; Nikaido, K.; Ikura, T.; Kuwajima, K. *J. Mol. Biol.* **1996**, *262*, 559−74.
(39) Wakatsuki, S.; Hodgson, K. O.; Eliezer, D.; Rice, M.; Hubbard, S.; Gillis, N. H.; Doniach, S.; Spann, U. *Rev. Sci. Instrum.* **1992**, *63*, 1736−40.
(40) Tsuruta, H.; Brennan, S.; Rek, Z. U.; Irving, T. C.; Tompkins, W. H.; Hodgson, K. O. *J. Appl. Crystallogr.* **1998**, 672−82.
(41) Eliezer, D.; Jennings, P. A.; Wright, P. E.; Doniach, S.; Hodgson, K. O.; Tsuruta, H. *Science* **1995**, *270*, 487−8.
(42) Chen, L.; Wildegger, G.; Kiefhaber, T.; Hodgson, K. O.; Doniach, S. *J. Mol. Biol.* **1998**, *276*, 225−37.
(43) Gladwin, S. T.; Evans, P. A. *Folding Des.* **1996**, *1*, 407−17.
(44) Chan, C. K.; Hu, Y.; Takahashi, S.; Rousseau, D. L.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 1779−84.
(45) Regenfuss, P.; Clegg, R. M.; Fulwyler, M. J.; Barrantes, F. J.; Jovin, T. M. *Rev. Sci. Instrum.* **1985**, *56*, 283−90.
(46) Shastry, M. C.; Luck, S. D.; Roder, H. *Biophys. J.* **1998**, *74*, 2714−21.
(47) Segel, D. J.; Bachmann, A.; Hofrichter, J.; Hodgson, K. O.; Doniach, S.; Kiefhaber, T. *J. Mol. Biol.* **1999**, *288*, 489−99.
(48) Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9029−33.
(49) Pollack, L.; Tate, M. W.; Darnton, N. C.; Knight, J. B.; Gruner, S. M.; Eaton, W. A.; Austin, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10115−7.
(50) Schindler, T.; Herrler, M.; Marahiel, M. A.; Schmid, F. X. *Nat. Struct. Biol.* **1995**, *2*, 663−73.
(51) Plaxco, K. W.; Millett, I. S.; Segel, D. J.; Doniach, S.; Baker, D. *Nat. Struct. Biol.* **1999**, *6*, 554−6.
(52) Hubbard, S. R.; Hodgson, K. O.; Doniach, S. *J. Biol. Chem.* **1988**, *263*, 4151−8.
(53) Gerstein, M.; Tsai, J.; Levitt, M. *J. Mol. Biol.* **1995**, *249*, 955−66.
(54) Svergun, D. I.; Richard, S.; Koch, M. H.; Sayers, Z.; Kuprin, S.; Zaccai, G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2267−72.
(55) Leontis, N. B.; Moore, P. B. *Nucleic Acids Res.* **1984**, *12*, 2193−203.
(56) Fan, L. X.; Svergun, D. I.; Volkov, V. V.; Aksenov, V. L.; Algarov, C. C.; Selivanova, O. M.; Shcherbakova, I. V.; Koch, M. H.; Gilles, R.; Wiedenmann, A.; May, R.; Serdyuk, I. N. *J. Appl. Crystallogr.* **2000**, *33*, 515−518.
(57) Cech, T. R.; Bass, B. L. *Annu. Rev. Biochem.* **1986**, *25*, 599−629.
(58) Sclavi, B.; Sullivan, M.; Chance, M. R.; Brenowitz, M.; Woodson, S. A. *Science* **1998**, *279*, 1940−3.
(59) Russell, R.; Millett, I. S.; Doniach, S.; Herschlag, D. *Nat. Struct. Biol.* **2000**, *7*, 367−70.
(60) Segel, D. J.; Eliezer, D.; Uversky, V.; Fink, A. L.; Hodgson, K. O.; Doniach, S. *Biochemistry* **1999**, *38*, 15352−9.
(61) Eliezer, D.; Chiba, K.; Tsuruta, H.; Doniach, S.; Hodgson, K. O.; Kihara, H. *Biophys. J.* **1993**, *65*, 912−7.
(62) Fink, A. L. *Folding Des.* **1998**, *3*, R9−23.
(63) Svergun, D. I.; Stuhrmann, H. *Acta Crystallogr.* **1991**, *A47*, 736−744.
(64) Kozin, M.; Volkov, V.; Svergun, D. *J. Appl. Crystallogr.* **1997**, *30*, 811−815.
(65) Chacon, P.; Moran, F.; Diaz, J. F.; Pantos, E.; Andreu, J. M. *Biophys. J.* **1998**, *74*, 2760−75.
(66) Chacon, P.; Diaz, J. F.; Moran, F.; Andreu, J. M. *J. Mol. Biol.* **2000**, *299*, 1289−302.
(67) Walther, D.; Cohen, F. E.; Doniach, S. *J. Appl. Crystallogr.* **2000**, *33*, 350−363.
(68) Bada, M.; Walther, D.; Arcangioli, B.; Doniach, S.; Delarue, M. *J. Mol. Biol.* **2000**, *300*, 563−74.
(69) Svergun, D. *Biophys. J.* **1999**, *76*, 2879.
(70) Press: W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*, 2nd ed.; Cambridge University Press: Ithaca, NY, 1994.
(71) Henry ER.; Hofrichter, J. *Methods Enzymol.* **1992**, *210*, 129−192.
(72) Fang, X. W.; Littrell, K.; Yang, X.; Henderson, S. J.; Siefert, S.; Thiyagarajan, P.; Pan, T.; Sosnick, T. R. *Biochemistry* **2000**, *39*, 11107−13.

CR990071K