Spring 6-14-2013

Structural Studies of Proteins in Solution: Cytochrome-C and Mitochondrial RHO

Anthony Banks
DePaul University, anthonybanks85@gmail.com

Recommended Citation

Banks, Anthony, "Structural Studies of Proteins in Solution: Cytochrome-C and Mitochondrial RHO" (2013). *College of Science and Health Theses and Dissertations*. 47.
https://via.library.depaul.edu/csh_etd/47

STRUCTURAL STUDIES OF PROTEINS IN SOLUTION: CYTOCHROME-C AND MITOCHONDRIAL RHO

A Thesis
Presented in
Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE

March, 2013

BY
Anthony Banks

PHYSICS DEPARTMENT
College of Liberal Arts and Sciences
DePaul University
Chicago, Illinois

TABLE OF CONTENTS

TABLE OF CONTENTS – *Continued*

LIST OF FIGURES

LIST OF FIGURES – *Continued*

LIST OF FIGURES – *Continued*

LIST OF FIGURES – *Continued*

LIST OF FIGURES – *Continued*

LIST OF FIGURES – *Continued*

LIST OF FIGURES – *Continued*

LIST OF FIGURES – *Continued*

ABSTRACT

Small Angle X-Ray Scattering (SAXS) is a rapidly developing technique for determining the structures of proteins in solution. This thesis presents new solution structures of two proteins: cold denatured equine cytochrome-c, and mitochondrial rho (miro). Cytochrome-c is a model system for protein folding and stability studies; this work demonstrates the interplay of thermodynamic forces in determining the global shape of unfolded proteins. Whereas denatured cytochrome-c is not able to form crystals in principle, miro is a newly purified protein of which no crystal structure is available in any conditions. We have used our SAXS data to verify conjectures for miro's structure based solely on protein sequence analysis.

Acknowledgements

For Jo, Gerard, Katie, and Quinn Banks

Agus go raibh míle maith agaibh speisialta do na daoine seo

Dr. Landahl
Dr. Sarma
Dr. Pando
Dr. Rice
Christopher Asta
Matthew Andorf
Julian Klosowiak
Trevor GrandPre
The Rice Lab
The BioCAT-18 Staff
Every other member of the Landahl Lab

CHAPTER 1

Introduction

This experiment concerns itself with two proteins. The first of these is equine cytochrome-c, a well-documented and widely studied protein. The second protein is mitochondrial-rho (miro), a protein involved in the movement of mitochondria in human cells. Experimentation on the miro protein is in collaboration with the Rice Lab at Northwestern University. While both proteins are studied via small-angle x-ray scattering (SAXS), the motivations for such studies are separate. Building from previous work (Elmer, 2010), our purpose for using cytochrome-c is to gain further understanding of the process of protein folding. The purpose for using miro is to learn more about its size, shape, and structure. Eventually, we also hope to be able to study protein conformational change in miro using similar techniques for studying protein folding in cytochrome-c. SAXS measurements are performed at Argonne National Laboratory, where one dimensional scattering patterns are obtained. Data analysis involves reconstructing from these 1D patterns three dimensional structures of the proteins.

1.1 Protein Folding

Proteins are a vital component of any organism. Inside the cell, DNA indirectly encodes amino acid sequences, from which ribosomes construct proteins (Nelson, 2005). How these amino acid sequences fold into stable, functional components of an organism is the core of the protein folding problem. In this thesis, we observe protein unfolding as a result of several parameter variations, and examine the macromolecular structure of the protein in its stable, folded state, and its unfolded state.

Examination of this data requires comparison to a theoretical model that accounts for the parameters varied.

Folding and unfolding of proteins in the cell occurs spontaneously and begins during construction by the ribosomes. This is a result of the environment in which the proteins are contained and the electrostatic forces of the amino acids themselves. The amino acids reach a final state; should this state be a stable, functional conformation, the protein is said to be in the native state. If the requirements of stability and functionality are not met, the protein is said to be in the denatured state. In this experiment, proteins are studied in terms of their distribution between these two states, i.e. a greater proportion of proteins in the unfolded state indicates denaturation. Measurement of the average size of a protein distribution allows determination of the proportion of folded to unfolded protein. Protein folding is accomplished here using several experimental parameters.

1.2 Prior Studies

Previous studies of protein folding and structure have been carried out using a variety of techniques. These studies may fall into categories of experimental and computational.

Experimental techniques involving nuclear magnetic resonance (NMR) have been used to study the structure of folded and unfolded proteins. This technique can also be used to study different sections of the same protein, thereby providing information on secondary structure. As a drawback, protein folding occurs over a much smaller timescale than that of a NMR measurement process. Fluorescence studies or more specifically Forster resonance energy transfer has also been used to characterize the local structure of proteins. This form of fluorescence exposure testing provides fluorescence absorption information for different areas of a protein. Another established technique for protein folding study is circular dichroism (CD).

Alpha helices and beta sheets show different absorption patterns when exposed to circularly polarized UV light, thus providing secondary structural information. CD studies may also give information on global structure, of significance since protein folding depends on both local and long range bonding.

Computational research in protein folding has attempted to begin with an amino acid sequence and explain how it eventually results in its conformation of lowest free energy. It has been theorized (Dobson, 1998) that a sequence of 100 amino acids may have up to 10^{18} possible final conformations. Modern computer programs are capable of calculating this final state of lowest free energy within two days, but proteins find this state in less than a second on average. Through studies of protein fragments, it has been shown that there are certain amino acid sequences that invariably result in three dimensional structures. Results of such studies have been used with the Protein Data Bank (PDB) to give faster conformational calculations (Pain, 2000). One way by which the number of calculations is reduced is the assumption that the amino acid sequence undergoes hydrophobic collapse upon leaving the ribosome. Calculation time may be reduced further by noting that specific amino acid sequences are associated with several secondary structures, and by considering certain tertiary structures for given sequences. The study of protein folding may provide answers as to how a protein finds its lowest free energy state in such a short time, and may also help in validating algorithms as described above and increasing their efficiencies.

1.3 Equine Cytochrome-C

Cytochrome-c is a protein found in numerous organisms and plays an important role in the transportation of electrons in mitochondria. Equine cytochrome-c is chosen for study as its size is well known and its genomic sequence has previously been documented (Bilsel, 2006). Further, the folding of cytochrome-c has been previously studied using a variety of techniques, and it is also inexpensive. Full details of this protein may be found in the Protein Data Bank (ID: 1HRC). The amino acid sequence for equine cytochrome-c is shown in figure 1.1 (Margoliash, 1961), and the corresponding structure in figure 1.2, determined by crystallography. In this experiment, SAXS measurements are performed on cytochrome-c over a range of temperatures, such that the protein may be observed in its native and denatured states, and information regarding size and shape are obtained. The use of SAXS as an experimental technique is necessary since denatured proteins cannot be crystallized.

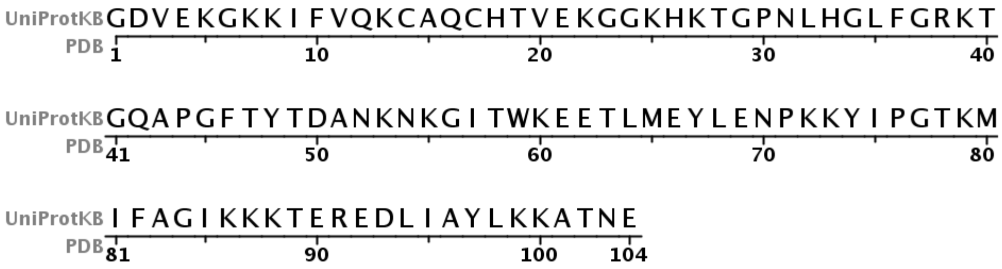


Figure 1.1: Amino acid sequence for equine cytochrome-c (PDB ID: 1HRC).

1.4 Mitochondrial-Rho

Mitochondria take up sources of energy (glucose, fats) in the cell and convert them to usable energy, or adenosine triphosphate (ATP). ATP is used to carry out almost all cell operations. In order to deliver ATP to areas of the cell in most need of it, mitochondria are transported throughout the cell by molecular motors called kinesin. Kinesin transports the mitochondria throughout the cell by moving along microtubules, which are hollow tubes made from long chains of tubulin protein. This is an important function in the cell, since an inability to properly distribute ATP can lead to dysfunctional cells. Kinesin connects to mitochondria through the mitochondrial rho (miro) protein. The structure of miro is a hypothesis based on

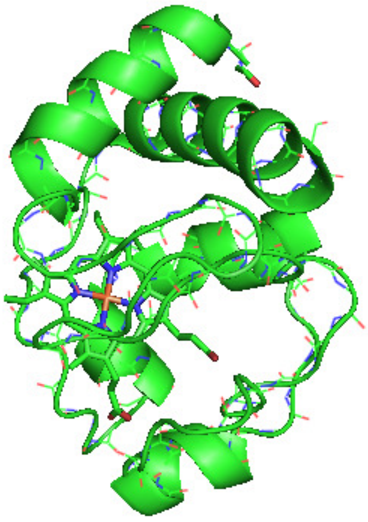


Figure 1.2: Equine cytochrome-c crystal structure. Taken from the RCSB Protein Data Bank (ID: 1HRC) and viewed using the application PyMOL. For the purpose of scale, this structure has a maximum diameter of ∼ 34Å.

amino acid sequence only, and may be described as follows. The mitochondrion connects to miro which is then connected to kinesin via an adapter protein, milton (Glater, 2006). The miro protein consists of a GTPase domain, followed by two ef hands, followed by a second GTPase domain. The GTPase domain is so called as it belongs to a family of enzymes that can bind guanosine triphosphate (GTP). Ef hands are calcium binding proteins. It is suggested that miro is involved in the regulation of the kinesin motor by its ability to bind GTP and Ca^{2+}, and is thereby integral in regulating mitochondrial movement in the cell. Low levels of Ca^{2+} are consistent with freely moving mitochondria, whereas high levels of Ca^{2+} cause the transport of mitochondria to cease. An outline of the miro protein and this process can be seen in figure 1.3.

In this experiment two variations of the miro protein, provided by the Rice Lab of

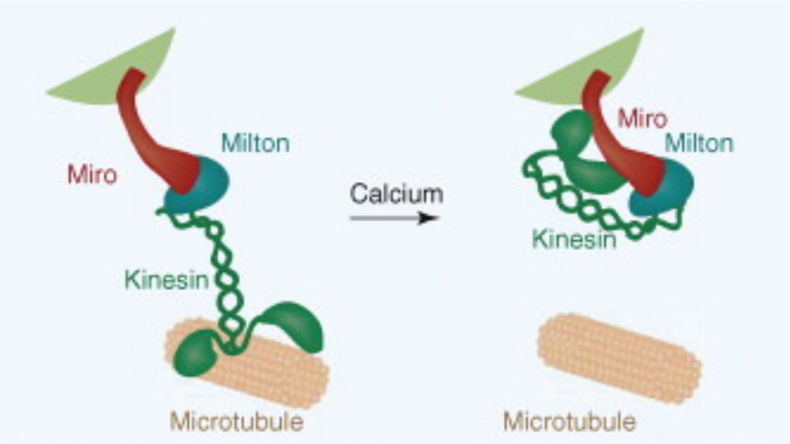


Figure 1.3: On the left, miro in a low-calcium environment allowing ample transport of mitochondria. On the right, a calcium-rich environment signifying a halt in mitochondrial transportation (MacAskill, 2009)

Northwestern University, are examined. One of these is the protein as described above, with 2 GTPase domains connected by two ef hands. The other is the same protein but without a second GTPase domain. The former we refer to as miro-L (as it is longer), and the latter (and shorter of the two) we refer to as miro-S. The Rice Lab is currently processing crystallography results for the miro-S protein and it is hoped these and our own SAXS experiments on miro-S will compliment one another and provide further information and insight into the structure of these proteins.

1.5 Small Angle X-Ray Scattering

A technique first developed in the 1940s, small angle x-ray scattering (SAXS) can be used to determine the size and shape of small objects (less that 2 microns across). SAXS measures electron density, and since the electron density of a protein should decrease as it unfolds, it follows that SAXS may be used to measure the size and shape of proteins that change shape, and thus determine in which state, native or denatured, the proteins reside. The use of SAXS for determining complex structures such as proteins, however, has required three more recent developments: bright 3rd generation synchrotron sources such as Argonne's Advanced Photon Source, new era x-ray detectors and high-speed computers, and new computational algorithms.

1.6 Data Analysis Process

A significant portion of this thesis is devoted to the process of analyzing the data. SAXS measurements are reduced to one dimensional scattering patterns in inverse space, where Guinier analysis is carried out. By performing a special, undersampled, indirect Fourier transform, this data is then analyzed in real space using the pair-distance distribution function, a measure of the amount of pairs of electrons separated by a given distance. The next step in the analysis is obtaining a low-resolution three dimensional reconstruction of the protein, by using a simulated annealing algorithm on a dummy atom model. The quality of the reconstruction can then be checked by carrying out a Fourier transform back to inverse space; should this unique solution match the SAXS measurements, the reconstruction may be deemed successful. In this sense the data analysis follows a pattern: SAXS data analysis in inverse space → analysis in real space → 3D reconstruction → reconstruction analysis in inverse space. This cycle of analysis allows a measure of consistency in results through the use of several analytical methods.

CHAPTER 2

Cytochrome-C Protein Stability Theory

2.1 Enthalpy, Entropy, and the Gibbs Free Energy

In this chapter we outline a theory to predict the average size of a distribution of proteins, based on the population levels of the protein in two states, the native (folded) and denatured (unfolded). This theory was developed by Dill (Dill, 2009) and applied to cytochrome-c by Elmer (Elmer, 2010). There are three components to this theory. Firstly, denaturation of the protein may be achieved by varying the thermodynamic conditions of our protein solution. Changing the temperature can induce structural change, which indicates a change in state. The Gibbs and Helmholtz free energies of the system may be used to describe such a state change. In order to do this we must begin by considering the entropy and enthalpy of the system.

Enthalpy is defined as the heat transfer of a closed system under constant pressure, and may be expressed as

$$\Delta H = C_P \Delta T \tag{2.1}$$

where C_P is the specific heat at constant pressure, defined as

$$C_P = \left(\frac{\partial H}{\partial T}\right)_P = \left(\frac{\partial U + p\partial V}{\partial T}\right)_P \tag{2.2}$$

and U is the internal energy of the system, V is the volume, and H is the enthalpy. With $\Delta T = T - T_0$, we have a relationship between enthalpy and temperature,

$$H(T) = C_P(T - T_0) + H(T_0) \tag{2.3}$$

The temperature at which enthalpy is zero, and has no effect on the system's ability to change, we denote T_h. For his ideal thermal protein model for cytochrome-c, Robertson (Robertson, 1997) measured this temperature T_h as $373K$. For $T > T_h$, positive enthalpy resists change in the system, and for $T < T_h$, negative enthalpy makes system change more likely.

Entropy provides a description of the order of a system on the microscopic level, i.e. as a system becomes more disordered, entropy increases. By holding constant the number density, N and volume, V we may define entropy in terms of temperature and internal energy, thus also providing a definition for temperature,

$$\Delta S = \frac{\Delta U}{T} \tag{2.4}$$

Any variable prefixed by a Δ indicates a transition at constant temperature from an unfolded to folded state. By incorporating the definition of specific heat at constant volume, we integrate equation 2.4 to obtain

$$S(T) = S(T_0) + C_P ln\Big(\frac{T}{T_0}\Big) \tag{2.5}$$

The temperature at which entropy is zero, and has dependance only the system's chemical structure, we denote T_S. For cytochrome-c, T_S is $385K$ (Robertson, 1997). As is the case for enthalpy, for $T < T_S$ the system resists change, and for $T > T_S$ positive entropy encourages system change, i.e. favors increasing disorganization.

In a closed system, the Gibbs free energy, G, is the amount of energy available for use without a change in system pressure. Enthalpy and entropy may be related through the Gibbs equation (Dill, 2003),

$$G = H - TS \tag{2.6}$$

Although the Gibbs free energy is defined for a given temperature, the enthalpy and entropy components have their own dependencies on temperature. Thus the Gibbs free energy has an indirect temperature dependence,

$$(\Delta G)_T = \Delta H(T) - T\Delta S(T) \tag{2.7}$$

Substituting in equations 2.3 and 2.5 to equation 2.7 gives,

$$(\Delta G)_T = \Delta C_P(T - T_h) - T\Delta C_P ln\Big(\frac{T}{T_0}\Big) \tag{2.8}$$

where an arbitrary initial temperature T_0 has been chosen to simplify the expression. Folded protein has a smaller specific heat than unfolded protein; therefore ΔC_P has a negative value. A graphical representation of equation 2.8 is given in figure 2.1.

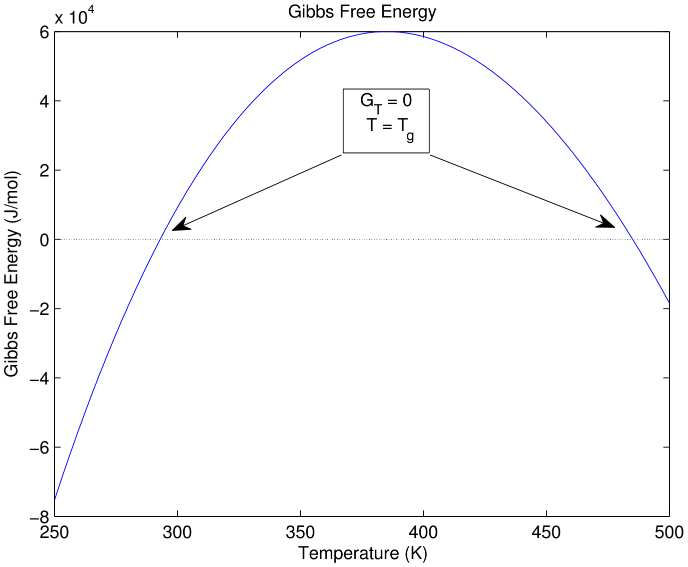


Figure 2.1: Gibbs free energy in Joules per mole as a function of temperature in Kelvin. This graph has been extended down to $250K$ to match experimental conditions, and extended up to $500K$ to show the location of the higher T_g.

Over the range of temperatures from 250K to 500K the change in Gibbs free energy is zero at two points, denoted T_g, at which it has no effect on the likelihood of system change. The system is most stable at postitive values for the Gibbs free energy, specifically at the apex of the $(\Delta G)_T$ curve. At this temperature we would expect the protein to be in its native state. Deviation from this temperature toward less positive $(\Delta G)_T$ values indicates a propensity toward instability and protein folding. Further deviations to regions of negative $(\Delta G)_T$ values indicates instability and an inclination to protein unfolding. It follows that we should observe both cold and warm unfolding of the protein. Previously, we chose an arbitrary value for T_0 to simplify our expression for $(\Delta G)_T$. As we approach $(\Delta G)_T = 0$ we must take into account T_g in the terms for enthalpy and entropy. Replacing T_0 with T_g in equations 2.3 and 2.5 and substituting into equation 2.7 we get

$$\Delta H(T_g) = T_g \Delta S(T_g) \tag{2.9}$$

We can now rewrite the changes in enthalpy and entropy as

$$\Delta H(T) = \Delta H(T_g) - \Delta C_P(T_g - T) \tag{2.10}$$

and

$$\Delta S(T) = \frac{\Delta H(T_g)}{T_g} - \Delta C_P ln\Big(\frac{T_g}{T}\Big) \tag{2.11}$$

Graphing these expressions (see figure 2.2), we can infer that it is likely to find the protein in its native state at temperatures within the envelope defined by the intersections of the enthalpy and temperature-scaled entropy curves. Outside this envelope, i.e. at temperatures at which the Gibbs free energy is negative, we would expect the protein to be undergoing phase change to its unfolded state.

We may relate the quantities T_h, T_S, and T_g by incorporating the limits of the native and denatured states of the protein. When enthalpy and entropy are zero, equations 2.10 and 2.11 give expressions for T_h and T_S respectively. The expression for T_S is

$$T_S = \frac{T_g}{\exp(\frac{\Delta H(T_g)}{T_g \Delta C_P})} \tag{2.12}$$

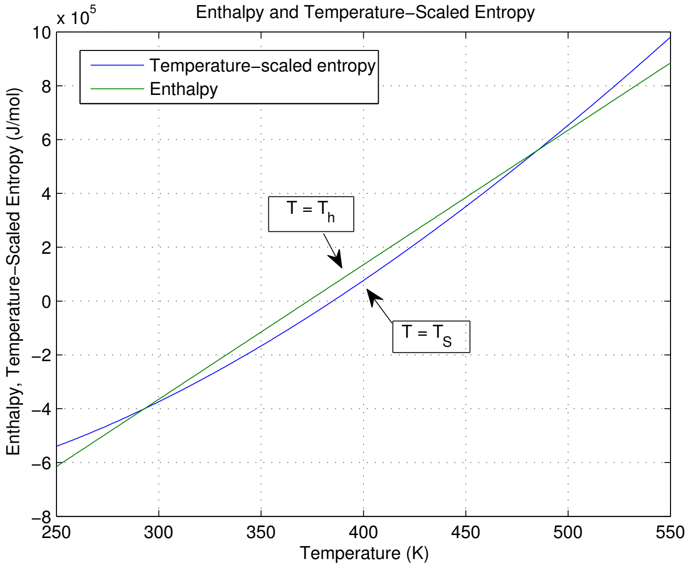


Figure 2.2: Temperature-scaled entropy and enthalpy in Joules per mole, graphed against temperature in Kelvin. T_h is $373K$ and T_S is $385K$.

The denominator above may be eliminated by assuming $\Delta H << T_g \Delta C_P$ and expanding the exponent (Robertson, 1997), giving

$$T_S \cong \frac{T_g}{1 + \frac{\Delta H(T_g)}{T_g \Delta C_P}} \tag{2.13}$$

To simplify, we substitute equation 2.10 (with enthalpy equal to zero) into the denominator, to get

$$T_S \cong \frac{T_g^2}{2T_g - T_h} \tag{2.14}$$

We have now defined a relationship between the temperatures at which enthalpy, entropy, and the Gibbs free energy vanish. Graphically, we may observe these variables by combining figures 2.1 and 2.2.

Figure 2.3 demonstrates the relationship of enthalpy, entropy and the Gibbs free energy to temperature. Recall that the system is most stable for positive Gibbs free energy, with peak stability occurring at the apex of the curve. For negative

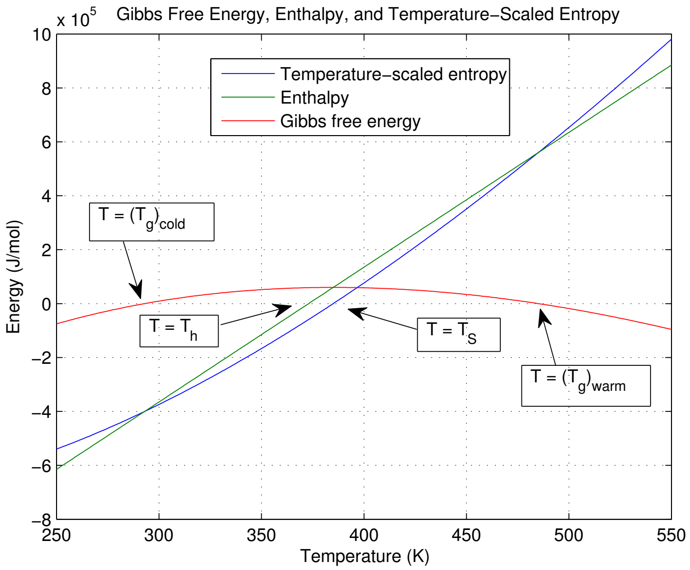


Figure 2.3: Gibbs free energy and its components (enthalpy and temperature-scaled entropy) in Joules per mole, graphed against temperature in Kelvin.

Gibbs free energy the system is unstable and we would expect a protein at these temperatures to be unfolded. For enthalpy, the system favors change for $T < T_h$, i.e. we would expect to find the protein in its folded state at such temperatures, whereas for positive enthalpy, system change is resisted and we would expect to find the protein in its unfolded state. The opposite is true for entropy, with respect to temperature. Figure 2.2 portrays these properties quite clearly. Temperatures at $T < (T_g)_{\text{cold}}$, i.e. $T < 299K$, and $T > (T_g)_{\text{warm}}$, i.e. $T > 475K$, correspond to temperatures lying outside the envelope defined by the enthalpy and entropy curves. Outside this range of temperatures we would expect the system to tend toward instability, i.e. we would expect the protein to be unfolded.

2.2 Neutral Folding Free Energy

For a closed system at constant volume, the Helmholtz free energy, F, is the amount of energy available for use. The change in Helmholtz free energy will be used to characterize the energy required to change a protein from its denatured to native state, i.e.

$$\Delta F_{\text{fold}} = F_{\text{native}} - F_{\text{denatured}} \tag{2.15}$$

Henceforth, this change in Helmholtz free energy, ΔF, will be referred to as the folding free energy. The Gibbs and Helmholtz free energies are closely related; respectively, they describe isobaric and isometric systems, the relationship given by, (Dill, 2003)

$$F = G + PV \tag{2.16}$$

In this experiment, the product of the changes in pressure and volume is $\Delta P \Delta V \cong 1 Jmol^{-1}$. However, based on the accepted value for the specific heat of cytochrome-c ($5.3 kJK^{-1}mol^{-1}$) (Robertson, 1997), the energy required for a temperature change of $1K$ is $5300 Jmol^{-1}$. Since this is far greater than the product of the changes in pressure and volume, the folding free energy may be approximated as the Gibbs free energy, allowing us to express the folding free energy in familiar terms, i.e.

$$\Delta F = \Delta H - T \Delta S \tag{2.17}$$

for a static temperature measurement. It should be noted that this is the neutral folding free energy of our system, i.e. since our system is macromolecular in nature, there is a considerable charge distribution we must account for. This contribution will be addressed later. This macromolecular structure also has effects on enthalpy and entropy which are now considered.

Our expression for enthalpy is subject to change from the effects of molecular packing and denaturant concentration. From the ideal thermal protein model (Dill, 2009) our expression for enthalpy becomes

$$\Delta H(T) = \Delta H + g(c)N = \Delta C_P(T - T_0) + (g_0 + m_1c)N \tag{2.18}$$

Here, the term g_0 represents an average packing energy dependent on the denaturant concentration, m_1c is a product of concentration c and a denaturant-specific folding parameter m_1, and N is the number of amino acids in the protein. Note that equation 2.18 contains conflicting terms with regard to total enthalpy. Since g_0N is always negative it decreases total enthalpy, thereby encouraging folding of the protein. The term m_1c is always positive however and therefore acts to increase enthalpy. Equation 2.18 demonstrates the balance between the concentration and packing energy; they will cancel each other out at T_h. Should the change in enthalpy at T_h not be exactly zero, it may be be incorporated into the change in entropy.

Our expression for entropy, equation 2.5, must now also be modified to account or structural interactions. The total change in entropy becomes (Dill, 2009)

$$\Delta S = -kNln(z) + \Delta C_P ln\Big(\frac{T}{T_S}\Big) \tag{2.19}$$

where we have again chosen an arbitrary T_0 in equation 2.5 such that our expression is simplified. In equation 2.19 z is the number of rotational isomers possible around the carbon backbone per amino acid of the protein, and k is Boltzmann's constant.

Having defined new terms for the enthalpy and entropy we may now form a full expression for the folding free energy. By substituting equations 2.18 and 2.19 into equation 2.17 we obtain the desired expression for neutral folding free energy:

$$\Delta F(T)_{\text{neutral}} = gN + TkNln(z) + \Delta C_P(T - T_h) - T\Delta C_P ln\Big(\frac{T}{T_S}\Big) \tag{2.20}$$

Graphing equation 2.20 we find that the neutral folding free energy causes unfolding of the protein, albeit only at extreme temperatures; $\sim 290K$ for cold denturation, $\sim 850K$ for warm denaturation.

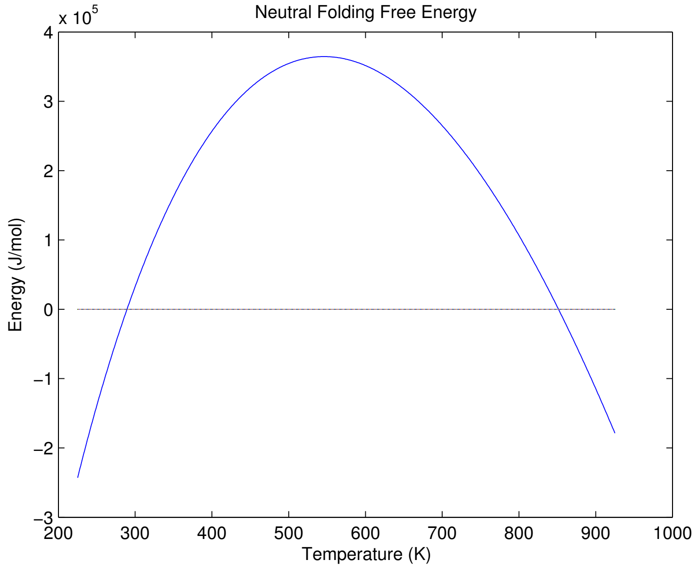


Figure 2.4: Neutral folding free energy in Joules per mole for cytochrome-c with a $2M$ guanidine HCl denaturant, as a function of temperature in Kelvin. In equation 2.20, $\Delta C_P = 5kJmol^{-1}K^{-1}$, $N = 104$, $z = 7.54$, and $g = -4.81kJmol^{-1}$

As previously mentioned, equation 2.20 does not account for the charge distribution present in our system. The folding free energy may now be adapted to include this contribution.

2.3 Electrostatic Folding Free Energy

The macromolecular nature of our protein results in charge interactions between the protein and solution, and also affects the charge distribution throughout the solution. The charge interactions contribute to the free energy that causes protein folding, and along with the solution's charge distribution, makes up the electrostatic component of the folding free energy, $\Delta F_{\text{electrostatic}}$. The total folding free energy now becomes

$$\Delta F_{\text{total}} + \Delta F_{\text{electrostatic}} \tag{2.21}$$

In order to successfully add these components, they must contribute on the same scale. The distance at which the charge interactions are comparable in scale to the contributions from thermal energy is the Bjurrem length, l_b. This is given by (Dill, 2009)

$$l_b = \frac{4\pi\epsilon_0 e^2}{\epsilon kT} \tag{2.22}$$

$\Delta F_{\text{electrostatic}}$ contains contributions from pH level and salt concentration, c_s. Salt concentration is incorporated with the Bjurrem length through the linearized Poisson-Boltzmann constant, κ,

$$\kappa^2 = 2c_s l_b \tag{2.23}$$

The contribution from pH level is given by the net charge resulting from charge contributions of basic and acidic amino acids. These contributions are given by

$$Q_b = \sum_i^b \frac{10^{pk_i \,-\, pH}}{1 + 10^{pk_i \,-\, pH}} \tag{2.24}$$

and

$$Q_a = \sum_i^a \frac{10^{pH \,-\, pk_i}}{1 + 10^{pH \,-\, pk_i}} \tag{2.25}$$

where b and a are the amounts of basic and acidic amino acids, respectively, and pk_i are the proton dissociation constants for the amino acids. Specific pk_i can be found in the appendix of Elmer (Elmer, 2010). Net charge is then

$$Q = Q_b - Q_a \tag{2.26}$$

This treatment is necessary for both native and denatured states, giving a net charge contribution of Q_n and Q_d, respectively. Following a Born and Debye-Huckel treatment of electrostatics (Dill, 2009, and Elmer, 2010), the electrostatic energy is given by

$$\Delta F_{\text{electrostatic}} = kT\left(\frac{Q_n^2 l_b}{2R_n(1+\kappa R_n)} - \frac{Q_d^2 l_b}{2R_d(1+\kappa R_d)}\right) \tag{2.27}$$

where R_n and R_d represent the radius of gyration for the native state and denatured state. The neutral and electrostatic components of the folding free energy may now be combined to obtain a full expression. From equations 2.21, 2.20, and 2.27 we obtain

$$\begin{aligned}\Delta F_{\text{total}} &= gN + TkNln(z) + \Delta C_P(T - T_h) - T\Delta C_P ln\left(\frac{T}{T_S}\right) \\ &+ kT\left(\frac{Q_n^2 l_b}{2R_n(1+\kappa R_n)} - \frac{Q_d^2 l_b}{2R_d(1+\kappa R_d)}\right)\end{aligned} \tag{2.28}$$

We now have a complete model for the folding free energy of our protein which can be used to model populations in the two states and predict the radius of gyration of the protein at given conditions.

CHAPTER 3

Small Angle X-Ray Scattering Theory

We outline here the theory of small angle x-ray scattering. This theory begins by calculating the scattering from a single electron and extending this to a continuous charge distribution by considering the phase difference between numerous scattered waves. The scattering intensity is obtained via Fourier transform, and the Guinier approximation, i.e. the dependence of intensity on the radius of a spherical particle at low angles, is shown. Following this, further theory relating to three-dimensional reconstruction from one-dimensional scattering patterns via indirect Fourier transform, simulated annealing, and the Debye formula is presented.

3.1 Scattering Angle

We first evaluate the case of a single x-ray photon scattering from a single electron. Consider the photon propagating along the x-axis toward an electron at the origin, O. The photon, denoted by the unit vector $\frac{\vec{S_i}}{\lambda}$ is deviated from its original path by angle 2θ; the scattered photon is denoted by the unit vector $\frac{\vec{S_o}}{\lambda}$. Note the vectors are normalized by the x-ray wavelength. The initial photon and scattered photon are related through the vector

$$\vec{Q} = \frac{\vec{S_i} - \vec{S_o}}{\lambda} \tag{3.1}$$

This relationship may be seen in figure 3.1. By considering the midpoint of $\vec{Q}$, we may express its magnitude as

$$Q = \frac{2\sin\theta}{\lambda} \tag{3.2}$$

The angular deviation in the intensity of the incident photon is described by the vector $\vec{Q}$. We now consider multiple scattering centers and scattered intensities from which interference arises.

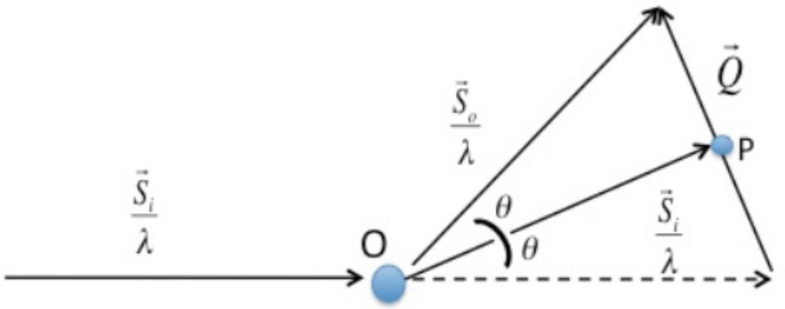


Figure 3.1: X-ray scattering for a single electron (Elmer, 2010).

3.2 Phase Shift and Interference

Let us consider the case of a two electron system in which the electrons are separated by a distance x. Supposing two identical, parallel x-rays are incident on each electron, the resulting scattered x-rays should also be identical. Whether these x-rays are in phase depends on the distance, x, between them (see figure 3.2). This figure describes how the phase shift may be defined;

$$\varphi = |\vec{PB}| - |\vec{AO}| = \vec{-x} \cdot -\vec{S}_o + \vec{-x} \cdot \vec{S}_o = -2\pi\vec{x} \cdot \vec{Q} \tag{3.3}$$

Euler angles are used to express the amplitude of a scattered x-ray in terms of the phase shift,

$$A = A_0 e^{i\varphi} \tag{3.4}$$

This expression may then be extended to the case of multiple x-rays incident on an electron density, $\rho(\vec{x})$, causing multiple interference events, j,

$$A = A_0 \sum_j \rho(\vec{x}) e^{i\varphi_j} \tag{3.5}$$

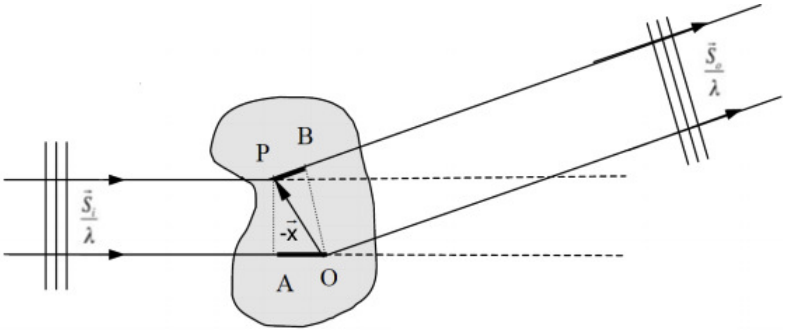


Figure 3.2: X-ray scattering for a pair of electrons. Directional inversion has been used on the vector $\vec{x}$ so that it and the vector $\vec{Q}$ point in the same general direction.

The assumption that multiple interference patterns add holds true for small-angle scattering, where observations are made in the far field. Considering a continuous, homogenous distribution of scattering centers, the amplitude becomes

$$A = A_0 \int \rho(\vec{x}) e^{-2\pi i \vec{x} \cdot \vec{Q}} dv \tag{3.6}$$

3.3 Scattering Intensity

Intensity is defined as the modulus of the amplitude squared, i.e.

$$I = |A|^2 \tag{3.7}$$

In this experiment our protein is suspended in solution, and so we can express the density, $\rho(\vec{x})$ as $\rho - \rho_0$, where ρ_0 is the density of the medium. Intensity now becomes

$$I = \left| (\rho - \rho_0) \int e^{-2\pi i \vec{x} \cdot \vec{Q}} dv \right|^2 \tag{3.8}$$

We now make the substitution

$$\Sigma(\vec{Q}) = \int e^{-2\pi i \vec{x} \cdot \vec{Q}} dv \tag{3.9}$$

which we refer to as the form function. Note that equation 3.9 closely resembles a Fourier transform. The intensity may be expressed as

$$I(\vec{Q}) = (\rho - \rho_0)^2 |\Sigma(\vec{Q})|^2 \tag{3.10}$$

As an example of such a scattering relationship we consider the specific case of a radially symmetric sphere. To find the form function $\Sigma(\vec{Q})$ we must integrate

$$\Sigma(\vec{Q}) = \int \sigma(\vec{x}) e^{-2\pi i \vec{x} \cdot \vec{Q}} dv \tag{3.11}$$

where $\sigma(\vec{x})$ is the form factor for the sphere, i.e. for $|\vec{x}| \leq a$, $\sigma(\vec{x}) = 1$, and for $|\vec{x}| > a$, $\sigma(\vec{x}) = 0$, where a is the radius of the sphere. Equation 3.11 evaluates to, (Guinier 1994)

$$\Sigma(\vec{Q}) = \frac{4}{3}\pi a^3 \left[3\frac{\sin(2\pi a Q) - 2\pi a Q \cos(2\pi a Q)}{(2\pi a Q)^3} \right] \tag{3.12}$$

From equation 3.10 the scattering intensity is now given by

$$I(\vec{Q}) = (\rho - \rho_0)^2 V^2 \left[3\frac{\sin(2\pi a Q) - 2\pi a Q \cos(2\pi a Q)}{(2\pi a Q)^3} \right]^2 \tag{3.13}$$

where V is the volume of the sphere. Figure 3.3 shows a plot of equation 3.13. A further way to view this relationship is through a log-scaled plot, which can be seen in in figure 3.4. Since the analysis of experimental data will involve log-scaled plots, it is prudent to include an example of the relationship here.

We now have an expression for scattering intensity on a global scale. Since in this experiment we studied two different proteins, which may exhibit varying form under denaturation, it is necessary to extend this expression to objects of any shape. To begin, we note that for very small angles, the vector $\vec{Q}$ is perpendicular to the direction of the incident photon beam, $\vec{S_i}$. Let us denote this direction of $\vec{Q}$ as $\vec{D}$. We now suppose $\vec{x}$ to be any vector in our object space. If we take the projection of $\vec{x}$ onto $\vec{D}$ to be x_D then the dot product $\vec{Q} \cdot \vec{x} = Q x_D$. We also define a term $\sigma(x_D)$ as a cross-section of our object at a distance x_D from the origin; $\sigma(x_D)$ lies on a plane normal to both $\vec{S_i}$ and $\vec{D}$, with the origin chosen as the object's center of mass. Equation 3.9 may now be written as

$$\Sigma(\vec{Q}) = \int e^{-2\pi i Q x_D} \sigma(x_D) dx_D \tag{3.14}$$

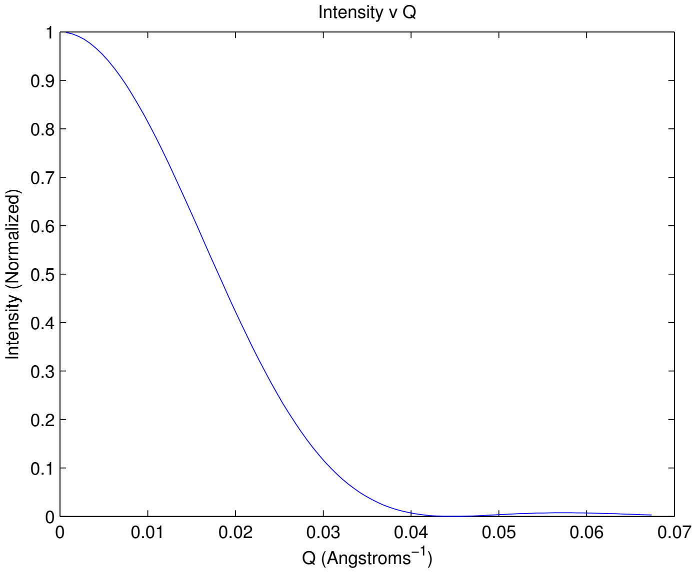


Figure 3.3: Intensity diffracted from a homogeneous sphere of radius $a = 16$Å, density 4 times that of the surrounding solution, with an x-ray wavelength of $\lambda = 10$Å.

From our definition of $\sigma(x_D)$, the object volume is given by

$$\int \sigma(x_D) dx_D = V \tag{3.15}$$

and from our choice of origin we also note that

$$\int x_D \sigma(x_D) dx_D = 0 \tag{3.16}$$

We now expand the exponential in equation 3.14 to the Q^2 order,

$$\Sigma(\vec{Q}) = \int \sigma(x_D) dx_D + 2\pi i Q \int x_d \sigma(x_D) dx_D - 2\pi^2 Q^2 \int x_D^2 \sigma(x_D) dx_D \tag{3.17}$$

From equations 3.15 and 3.16, equation 3.17 becomes

$$\Sigma(\vec{Q}) = V - 2\pi^2 Q^2 V \int \frac{1}{V} x_D^2 \rho(x_D) dx_D \tag{3.18}$$

We denote the remaining integral in equation 3.18 by R_D^2, where R_D is the average inertial distance along $\vec{D}$ (Guinier 1994). Using this substitution and approximating

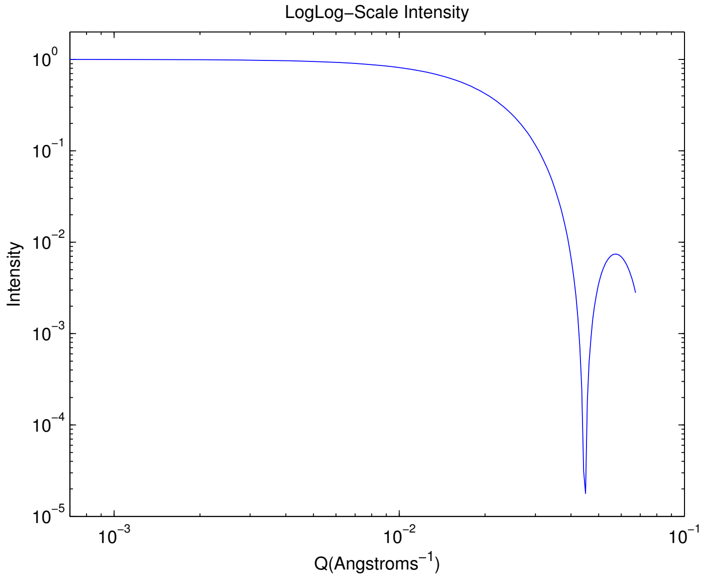


Figure 3.4: Log-scaled intensity diffracted from a homogeneous sphere of radius $a = 16$Å, density 4 times that of the surrounding solution, with an x-ray wavelength of $\lambda = 10$Å.

back to exponential form, 3.18 may be expressed as

$$\Sigma(\vec{Q}) = Ve^{-2\pi^2 Q^2 R_D^2} \tag{3.19}$$

We may now substitute equation 3.19 into equation 3.10 to obtain a corresponding scattering intensity equation

$$I(\vec{Q}) = n^2 e^{-4\pi^2 Q^2 R_D^2} \tag{3.20}$$

where $n = (\rho - \rho_0)V^2$. This expression is now expanded to three dimensions by considering two other axes, $\vec{U}$ and $\vec{V}$, both orthogonal to $\vec{D}$. In such a system, a point is now described by the coordinates x_D, x_U, x_V, at a distance of $r^2 = x_D^2 + x_U^2 + x_V^2$ from the origin. From equation 3.18 we can express this as the sum of the squares of the average inertial distances along each axes, i.e by defining a value R_g as

$$R_g^2 = R_D^2 + R_U^2 + R_V^2 = \frac{\int r^2 dv}{V} \tag{3.21}$$

This value R_g is the radius of gyration of the object. Furthermore, it is noted that under rotation about the origin, R_g^2 is unchanged, implying equality between R_D^2, R_U^2, and R_V^2. Thus, the radius of gyration is given by $R_g^2 = 3R_D^2$. From equation 3.20 we obtain a final expression for scattering intensity,

$$I(\vec{Q}) = n^2 e^{\frac{4\pi^2 Q^2 R_g^2}{3}} \tag{3.22}$$

We now have an expression for scattering intensity that applies to an object of any shape, and from which we can extract valuable information. Note that up to this point we have assumed our object, or particle, to be homogeneous. While this is often the case with the proteins under consideration, it is necessary to extend our treatment to the case of heterogenous particles. Supposing these particles are made up of atoms of atomic number f_k at distances r_k from the origin, and are oriented at random, the average scattering power per particle may be given by the Debye formula,

$$I(Q) = \sum\sum f_k f_j \frac{\sin(2\pi Q r_{kj})}{2\pi Q r_{kj}} \tag{3.23}$$

We may now perform a power series expansion in Q,

$$I(Q) = \sum\sum f_k f_j - \sum\sum f_k f_j \frac{4\pi^2 Q^2 r_{kj}^2}{6} + ... \tag{3.24}$$

The first term is n^2, the total number of electrons in the particle. To evaluate the second term we expand r_{kj}^2,

$$r_{jk}^2 = (\vec{r}_j - \vec{r}_k)^2 = r_k^2 + r_j^2 - 2\vec{r}_j \cdot \vec{r}_k \tag{3.25}$$

and

$$\sum\sum f_k f_j r_{kj}^2 = \sum\sum f_k f_j r_k^2 + \sum\sum f_k f_j r_j^2 - 2\sum f_k \vec{r}_k \cdot \sum f_j \vec{r}_j \tag{3.26}$$

From the first two terms of equation 3.26,

$$\sum f_j \sum f_k r_k^2 = n \sum f_k r_k^2 \tag{3.27}$$

and furthermore, if we choose the origin such that $\sum f_k \vec{r}_k = 0$, then the third term of equation 3.26 is zero. For this to remain consistent with our expression for intensity in equation 3.22, we see that the radius of gyration is given by

$$R_g^2 = \frac{\sum f_k r_k^2}{\sum f_k} \tag{3.28}$$

In the case of homogeneous particles, this more general form of the radius of gyration reverts to the form given previously.

3.4 Population Modeling an Protein Size Distribution

Before describing the analysis of the intensity patterns outlined above, we return momentarily to the case of our ideal thermal model for cytochrome-c, as described in Chapter 2. Although there are known intermediate states of equine cytochrome-c, the two state model outlined here is sufficient for describing denaturation of the protein. This model makes the assumption that at any condition, the total protein population is comprised of native and denatured protein, i.e.

$$p_n + p_d = 1 \tag{3.29}$$

where the populations of native (p_n) and denatured (p_d) protein change based on temperature and denaturant concentration. As a sum to unity, equation 3.29 may be expressed using Boltzmann factors

$$\frac{e^{-x}}{1+e^{-x}} + \frac{1}{1+e^{-x}} = 1 \tag{3.30}$$

By using the Boltzmann factors, we express the variable in terms of its probability of being in one state or the other. The dependence of temperature and concentration now come from a substitution for the variable x. The use of Boltzmann factors is appropriate for our thermodynamic system, and so in place of x we use the folding free energy (equation 2.28) divided by energy per mole, i.e. $x = \frac{\Delta F(T)}{RT}$, so that our populations becomes

$$p_n(T) = \frac{e^{\frac{-\Delta F(T)}{RT}}}{1 + e^{\frac{-\Delta F(T)}{RT}}} \tag{3.31}$$

and

$$p_d(T) = \frac{1}{1 + e^{\frac{-\Delta F(T)}{RT}}} \tag{3.32}$$

We have now obtained an expression for the population levels of our protein in the folded and unfolded states. We now incorporate the expression for radius of gyration from equation 3.28. In the case of our protein distribution consisting of two states which sum to unity, equation 3.28 becomes

$$R_g^2(T) = R_n^2 p_n(T) + R_d^2 p_d(T) \tag{3.33}$$

Figure 3.5 shows plotted radii of gyration for a range of temperatures and denaturant concentrations.

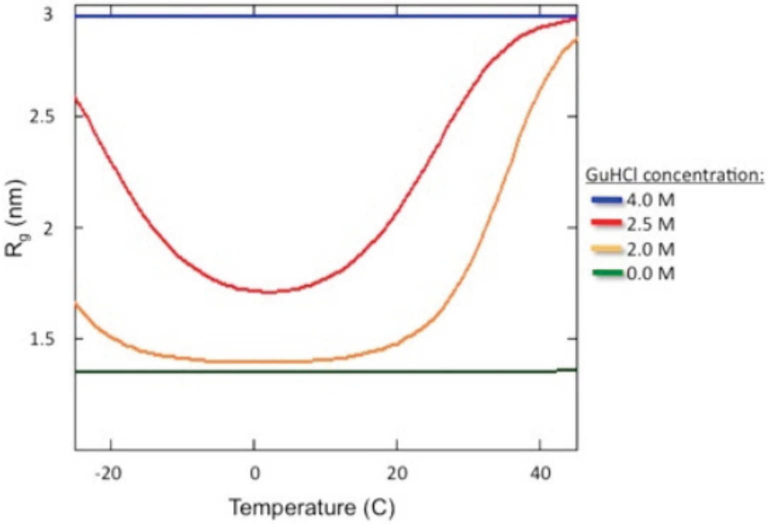


Figure 3.5: Theoretical calculation of radius of gyration (nm) versus temperature (^oC) for cytochrome-c in guanidine HCl solutions of concentration $0M$, $2M$, $2.5M$ and $4M$ (Elmer 2010).

.

3.5 Guinier Analysis

Following the description by Guinier (Guinier, 1994) we can analyze further the expression for scattering intensity. Since in this experiment we are interested in determining protein size and the change thereof under different conditions, we use our data with equation 3.22 to extract a value for the radius of gyration of our protein. Taking the natural log of both sides of equation 3.22 gives us

$$\ln(I) = \ln(n^2) - \frac{4}{3}\pi^2 Q^2 R_g^2 \tag{3.34}$$

From equation 3.34 we can see that by plotting $\ln(I)$ against Q^2, we should obtain a linear relationship, the slope of which provides a value for R_g, i.e.

$$R_g = \frac{3m}{4\pi^2} \tag{3.35}$$

where m is the slope. As an example of this, we use the intensity curve from figure 3.3. Since this curve represents a sphere of radius $a = 16$Å, we may use equation 3.21 to obtain its radius of gyration. By expressing the volume element as $dv = 4\pi r^2 dr$, equation 3.21 evaluates to

$$R_g = \sqrt{\frac{3}{5}}a \tag{3.36}$$

Thus from a plot of equation 3.34 we should obtain a R_g close to 12.39Å. A plot of $\ln(I)$ against Q^2 can be seen in figure 3.6. Such a plot is referred to as a Guinier plot.

Since the larger scattering angles will determine the size of the object, the region of the Guinier Plot we are interested in is that close to the y-axis, i.e. the low-Q region. In this region the curve exhibits a linear relationship up to $Q^2 \approx 0.5\text{x}10^{-3}\text{Å}^{-2}$. By performing a best fit line on this region of the curve, specifically up to $Q^2 = 0.4\text{x}10^{-3}\text{Å}^{-2}$, the radius of the gyration was found to be $R_g = 12.74$Å. This is reasonably close to 12.39Åand may be improved by limiting the extent of the linear fit to lower values of Q^2.

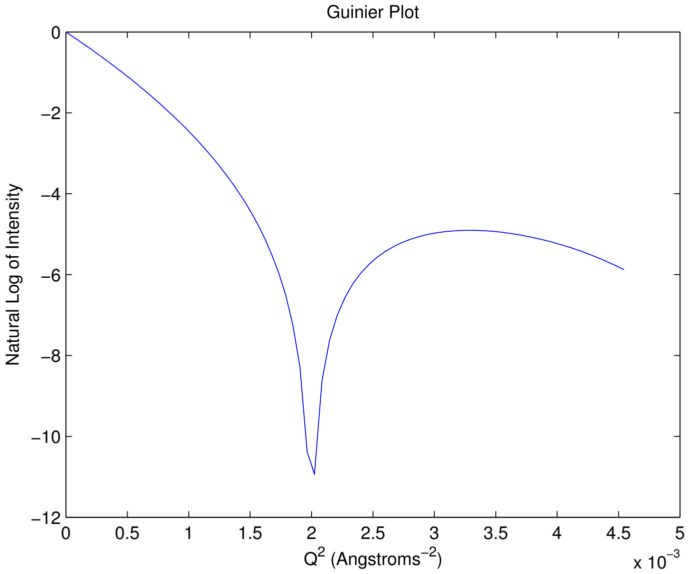


Figure 3.6: Guinier plot for the intensity diffracted from a homogeneous sphere of radius $a = 16$Å, density 4 times that of the surrounding solution, with an x-ray wavelength of $\lambda = 10$Å

.

We may also determine object shape from what is referred to as a Kratky plot. This is a plot of $Q^2 I$ against Q and provides insight into the degree of compactness of an object. Since the shapes of the proteins studied in this experiment are somewhat known, the Krakty plot serves as a useful tool for checking the quality of data throughout. A sample of various Kratky plot behaviour can be seen in figure 3.7.

Kratky plots that exhibit a clear peak and then decay are indicative of compact objects. Decreasing rate of decay following an initial peak indicates a decrease in compactness of the object. In figure 3.7 random walk (worm-like chain) behaviour is shown between q_1 and q_2. In the context of this experiment, one might expect a combination of the rod and worm-like chain curves from Kratky analysis of an unfolded protein, as opposed to the green curve which is more characteristic of a

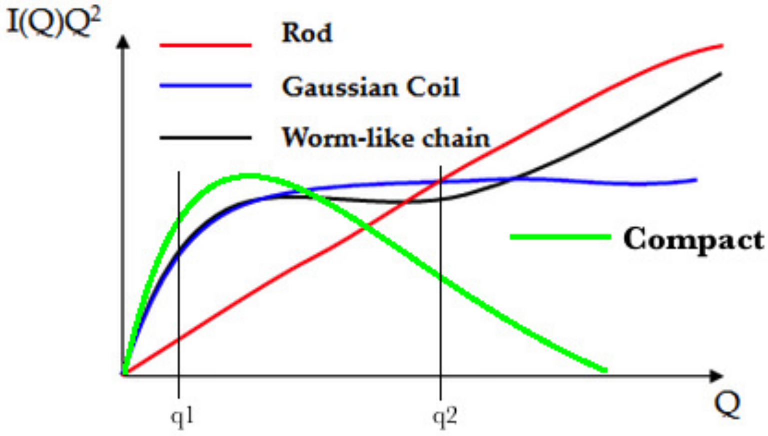


Figure 3.7: Kratky Plot with Q^2I on the y-axis and Q on the x-axis. This figure shows four curves. The green curve is typical of a compact object, characterized by the clear peak and decay. A quicker rate of decay indicates a more compact object. Thin rod-like behavior is characterized by steady increase and worm-like chain behaviour is characterized by a plateau prior to $q2$. Courtesy of $http://www.softmatterresources.com/small-angle-neutron-scatter/small-angle-scattering/kratky-plot.html$

.

compact, folded protein. As for the thin rod-like behaviour at high-Q, this for example may be indicative of protein structure over distances less than the persistence length, i.e. very low-degree of compactness. The persistence length may be thought of as the distance a polymer must travel in a straight line before turning, thereby quantifying the flexibility of the polymer. Persistence length can be extracted from the Kratky plot by performing a linear fit to the high-Q region (Brulet, 1996). The high-Q region ($Qp \geqslant 4; p =$ persistence length) behaves as:

$$I(Q) = \frac{\pi}{QL} + \frac{2}{3Q^2pL} \tag{3.37}$$

where L is contour length (i.e. the maximum length of the polymer chain). Phrasing equation 3.37 in terms of the Kratky plot gives

$$Q^2I(Q) = \frac{Q\pi}{L} + \frac{2}{3pL} \tag{3.38}$$

thereby providing the desired linear relationship. By dividing the intercept by the slope, one may obtain a value for the inverse persistence length, $Q_p(\sim \frac{2}{3\pi p})$. The Kratky plot corresponding to the scattering intensity in figure 3.3 is shown in figure 3.8. This is what we would expect from a perfect sphere.

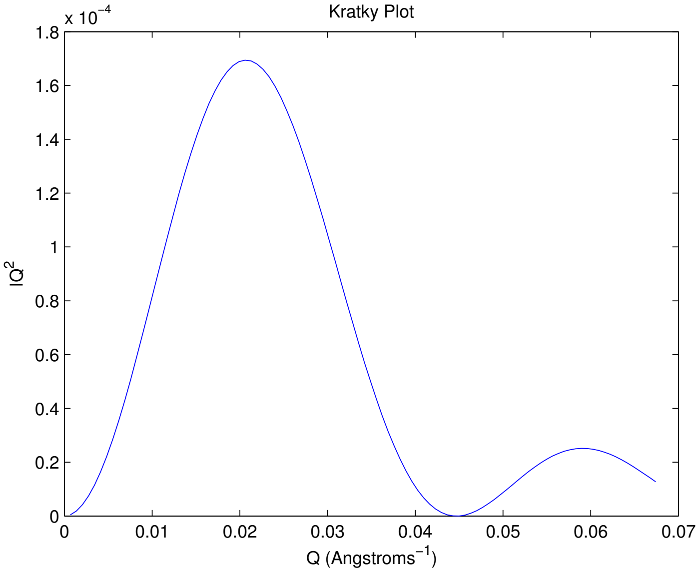


Figure 3.8: Kratky plot for the intensity diffracted from a homogeneous sphere of radius $a = 16\text{Å}$, density 4 times that of the surrounding solution, with an x-ray wavelength of $\lambda = 10\text{Å}$.

.

Up until now, we have made interpretations and extracted information by working in inverse space, i.e. by observing intensity as a function of $\vec{Q}$. In order to obtain a reconstruction from our scattering data, it is necessary to transfer back to real space. The first step in this process is obtaining the pair-distance distribution function.

3.6 The Pair-Distance Distribution Function

The pair-distance distribution function, which we denote $p(r)$, may be generally defined as the distribution of scatterers (electrons) a distance r apart. More specifically, if $\gamma_0(r)$ is the probability of finding a pair of electrons a distance r apart, then the number of pairs a distance r apart is proportional to $r^2V\gamma_0(r)$ (Vachette, 2010). From this we may define the pair-distance distribution function as

$$p(r) = \rho^2\gamma_0(r)Vr^2 \tag{3.39}$$

One can histogram all pairs of scatterers for each separation, r, to obtain the $p(r)$ function. In terms of scattering intensity $I(Q)$, $p(r)$ may be given in the general form of a Fourier transform (Vachette, 2010),

$$p(r) = \frac{r^2}{2\pi^2}\int_0^\infty Q^2 I(Q)\frac{sin(Qr)}{Qr}dQ \tag{3.40}$$

Solving $p(r)$ in this manner is impractical however, since we only know $I(Q)$ for a small range of Q; truncated data means greater effect from experimental errors. The solution to this is the indirect Fourier transform (Glatter, 1997),

$$I(Q) = 4\pi\int_0^{R_{\max}} p(r)\frac{sin(Qr)}{Qr}dr \tag{3.41}$$

where $R_{\max}$ is the maximum diameter of the object under investigation, and must be estimated a priori. A sample of some $p(r)$ functions are shown in figure 3.9.

In this experiment the program GNOM (Svergun, 1992) was used to calculate $p(r)$ functions from $I(Q)$ datasets. GNOM uses the regularization technique to find a stable solution while minimizing the effect from experimental errors. GNOM also chooses an optimal sampling rate at which to solve the Fourier transform, so as to avoid under/oversampling. The pair-distance distribution function can also be used as an alternative to calculate the radius of gyration of the object under investigation. As a general definition, the radius of gyration is given by the root mean square of the distance between pairs of scatterers (Vachette, 2010)

$$R_g^2 = \frac{\int_0^{R_{\max}} r^2 p(r)dr}{2\int_0^{R_{\max}} p(r)dr} \tag{3.42}$$

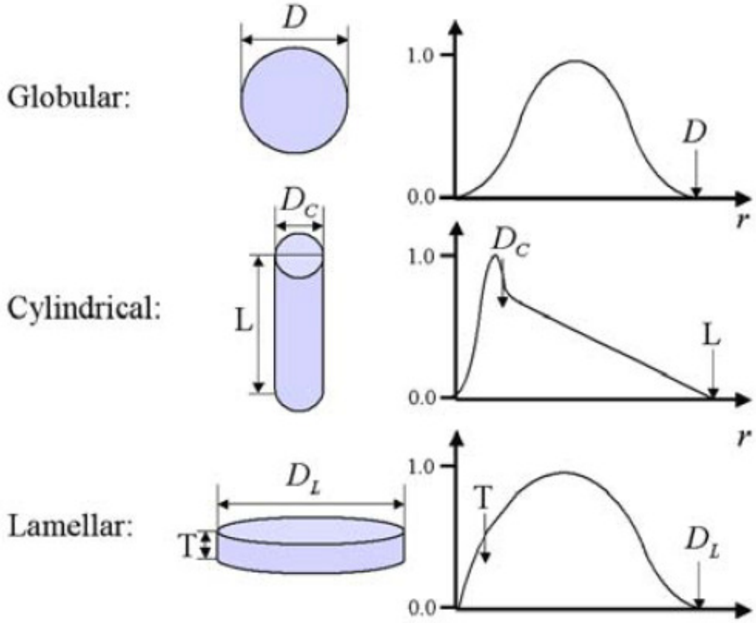


Figure 3.9: An example of pair-distance distribution functions for several shapes. The maximum particle diameter, $R_{\max}$ is represented by D, L, and D_L for the Globular, Cylindrical, and Lamellar examples respectively (Schnablegger, 2011).

.

Obtaining the radius of gyration using the $p(r)$ function is preferable to performing Guinier analysis, as it uses the full set of data, i.e. the entire Q-range, in its calculation.

At this point in the theory we have sufficient information to begin a three-dimensional reconstruction.

3.7 Reconstruction via Simulated Annealing

The program DAMMIN (Svergun, 1999) is used to obtain a reconstruction of the protein shape in this experiment. DAMMIN is specifically designed to develop low-resolution shapes of biological macromolecules in solution from SAXS measurements. The program initiates a user defined search volume (radius $R_V > \frac{1}{2}R_{\max}$), filled with N densely packed spheres (dummy atoms) of radius r_0 ($r_0 << R_V$), with fixed spatial positions. This model can now be described by a vector X with N components. Since this model is designed for proteins in solution, it is multiphase; each dummy atom is assigned an index X_j indicating to which phase it belongs ($X_j = 0$ for solution, $X_j = k$ for protein, where $0 < k < K$). The scattered intensity from such a model is given by (Svergun, 1999)

$$I(Q) = 2\pi^2 \sum_{l=0}^{\inf} \sum_{m=-1}^{l} \left[\sum_{k=1}^{K} [\Delta\rho_k A_{lm}^{(k)}(Q)]^2 + 2 \sum_{n>k} \Delta\rho_k A_{lm}^{(k)}(Q) \Delta\rho_n [A_{lm}^{(n)}(Q)]^* \right] \quad (3.43)$$

where $\Delta\rho_k$ is the contrast of a dummy atom of the $k_t h$ phase, and $A_{lm}^{(k)}(Q)$ are the representations through spherical harmonics of the scattering amplitude $A_k(Q)$, with $A_{lm}^{(k)}(Q)$ given by

$$A_{lm}^{(k)}(Q) = i^l \sqrt{\frac{2}{\pi}} f(Q) \sum_{j=1}^{N_k} j_l(Qr_j) Y_l^* m(\omega_j) \quad (3.44)$$

where r_j and ω_j are the polar coordinates of the dummy atoms, j_l is the spherical Bessel function and $f(s)$ is the scattering from a single atom. For a given dummy atom model, equations 3.43 and 3.44 give the corresponding scattering intensity curve. After this initial step, the problem becomes finding a vector X so as to minimize a target function $f(X)$ which describes the difference between scattering calculated from the dummy atom model and that of experimental data. This difference is given by

$$\chi^2 = \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{N(i)} \left[\left(I_{exp}^{(i)} \left(Q(_j) - \frac{I^{(i)}(Q_j)}{\sigma(Q_j)} \right) \right) \right]^2 \quad (3.45)$$

so that the target function is given by

$$f(X) = \chi^2 + \alpha P(X) \quad (3.46)$$

where $P(X)$ is a looseness penalty (user-defined or automatic) and $\alpha(> 0)$ is the weight of this penalty. This target function is minimized using simulated annealing.

Simulated annealing is a global optimization method, the classic example of which is the travelling salesman problem. If a salesman must visit a number of cities in a region, and the cost of travel is directly related to the distance between the cities, a simulated annealing algorithm can be used to compute the order in which cities should be visited such that cost is minimized. A description of this method, which encompasses the context of this experiment, can be found in Kirkpatrick et al. (Kirkpatrick, 1983). The simulated annealing algorithm mimics the process of atomic displacement through change in temperature of a metal. For DAMMIN, it is as follows:

1: Initially DAMMIN begins with a random configuration X_0 at a high temperature T_0, so that $T_0 = f(X_0)$.

2: Next, a new configuration $f(X')$ is obtained by randomly changing the phase of a random dummy atom, and the difference, $\Delta = f(X') - f(X)$ is calculated.

3: If $\Delta < 0$, the new configuration $f(X')$ is accepted. If $\Delta > 0$, $f(X')$ is accepted based on the probability $e^{\frac{-\Delta}{T}}$. Should the probability dictate that $f(X')$ be unacceptable, step 2 is repeated until $f(X')$ is accepted.

4: Temperature, T is now held constant for either $10N$ successful reconfigurations, or $100N$ reconfigurations. The system is then cooled to $T' = 0.9T$. This process is repeated until $f(X)$ is minimized.

Note that due to the random and probabilistic nature of this algorithm, the eventual solution produced is not unique. In practice, it is prudent to run the process several times to ensure consistency of the solution.

Furthermore, since the simulated annealing algorithm does not give a unique solution, it must be checked against the original scattering intensity.

3.8 Reconstructed Scattering Intensity

There are two options available for comparing the experimental scattering intensity to the scattering intensity from our DAMMIN reconstruction. The first option, conveniently, is to study one of the DAMMIN output files, which contains 4 columns of data: values of Q, experimental scattering and the error therein, and simulated scattering from the DAMMIN reconstruction. The simulated scattering is calculated from equations 3.43 and 3.44 for the case when $f(X)$ is minimized. The second option is to make use of the Debye formula for spherical scatterers (Debye, 1915). We use this formula through the online program FoXS (Fast X-ray Scattering), a method for computing theoretical scattering profiles from protein structures. FoXS is available freely from the University of California San Francisco (Schneidman-Duhovny, 2010). This uses the Debye formula (essentially equation 3.41) to calculate a scattering pattern from the measurable p(r) function of the DAMMIN reconstruction. In carrying out this step the theory has now come full circle. The process of obtaining a 3D reconstruction from 1D scattering data is referred to as the inverse problem. The reverse of this, calculating scattering from the reconstructed object, we refer to as the forward problem. Note that the forward problem has a unique solution, whereas the inverse is an optimization technique, based on initial user estimates.

A full walkthrough of the analysis procedure, particularly that of the GNOM-DAMMIN-FoXS steps can be found in the Data Analysis chapter, where a control data set of equine cytochrome-c is used to validate the choice of method of analysis.

CHAPTER 4

Experimental Setup

This chapter will describe the physical setup used to perform SAXS experiments on our proteins. An overview of the x-ray generation will be presented, followed by a description of the protein sample preparations. Following this, the loading of protein samples and absorption of scattered x-rays will be discussed.

4.1 X-Ray Source and Apparatus

The x-ray radiation used in this experiment is generated at the Advanced Photon Source (APS) at Argonne National Laboratory (ANL). An overview of the APS synchrotron system can be seen in figure 4.1.

Electron production takes place in the linear accelerator (LINAC) seen inside the ring of figure 4.1. A cathode heated to $\sim 1100^oC$ emits electrons which are accelerated through alternating electric fields to $450MeV$, i.e. travelling at $> 99.999\%$ the speed of light. From here, the electrons are injected into the booster/injector synchrotron, where in a ring of electromagnets they are accelerated to $7GeV$, or a velocity $> 99.999999\%$ the speed of light. The electrons are then injected to the storage ring where they are focused into a narrow beam by an electromagnetic field. The storage ring contains 40 sectors, 35 at which a variety of scientific experiments are carried out. At each sector, a system of alternating magnets (undulator), oscillates the electron beam to produce radiation which is Doppler shifted and collimated to obtain the desirable photon energy.

This experiment took place at BioCAT 18-ID-D, the Biophysics Collaborative Access Team sector. Sector 18 utilizes x-ray radiation over a range of energies, from

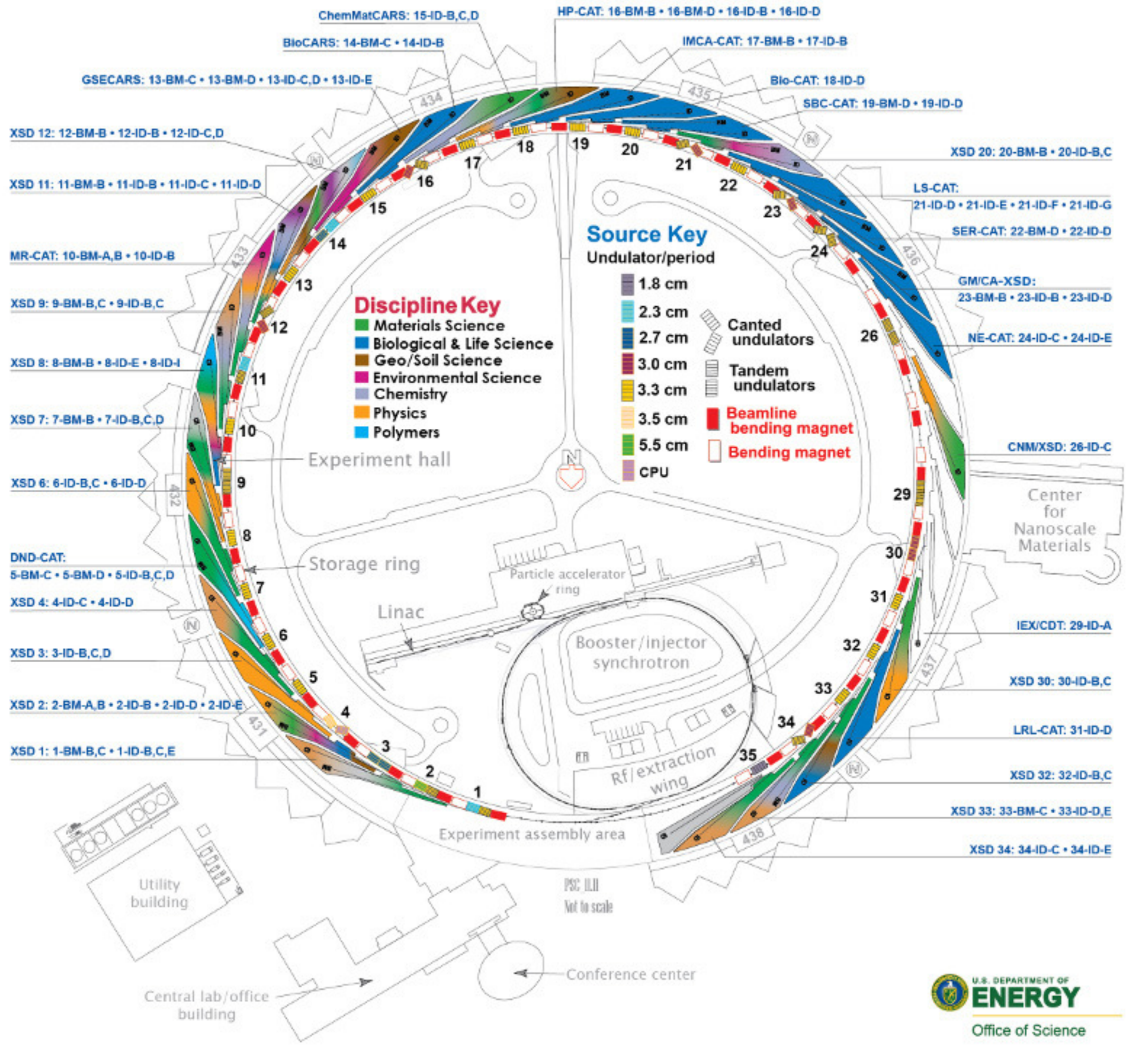
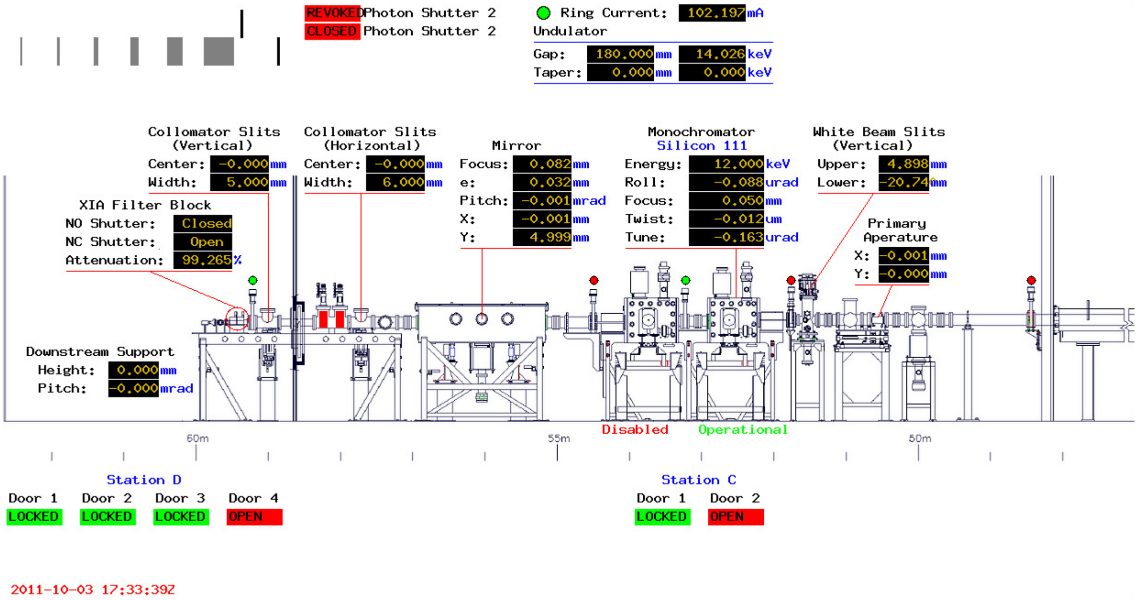


Figure 4.1: Overview of the Advanced Photon Source at Argonne National Laboratory. Courtesy of the U.S. Department of Energy, Office of Science.

.

$3.5keV$ to $35keV$. In this experiment, the x-ray energy was in the range of $10keV$. Experiments at Sector 18 take place in a hutch, an overview of which can be seen in figure 4.2.

The x-ray beam propagates in vacuum through two silicon monochromators so that


REVOKED Photon Shutter 2
CLOSED Photon Shutter 2

Ring Current: 102.197 mA
Undulator
Gap: 180.000 mm 14.026 keV
Taper: 0.000 mm 0.000 keV

Collomator Slits (Vertical)
Center: -0.000 mm
Width: 5.000 mm

Collomator Slits (Horizontal)
Center: -0.000 mm
Width: 6.000 mm

Mirror
Focus: 0.082 mm
e: 0.032 mm
Pitch: -0.001 mrad
X: -0.001 mm
Y: 4.999 mm

Monochromator
Silicon 111
Energy: 12.000 keV
Roll: -0.088 urad
Focus: 0.050 mm
Twist: -0.012 um
Tune: -0.163 urad

White Beam Slits (Vertical)
Upper: 4.898 mm
Lower: -20.74 mm

XIA Filter Block
NO Shutter: Closed
NC Shutter: Open
Attenuation: 99.265 %

Primary Aperature
X: -0.001 mm
Y: -0.000 mm

Downstream Support
Height: 0.000 mm
Pitch: -0.000 mrad

Station D
Door 1 LOCKED
Door 2 LOCKED
Door 3 LOCKED
Door 4 OPEN

Station C
Door 1 LOCKED
Door 2 OPEN

2011-10-03 17:33:39Z

Figure 4.2: Overview of the BioCAT 18-ID-D experimental hutch as of October 3rd, 2011. Courtesy of $http://www.bio.aps.anl.gov/igor/bag.html$.

.

it is almost monoenergetic. From here the beam is focused using a toroidal mirror to an area no less than $\sim 60\mu m^2$. Collimator and guard slits can be used to further focus the beam prior to interaction with the experimental sample. Samples are held in $150\mu m$ diameter capillary held in a brass mount. Attached to this sampling mount are a suction pump and refrigerating/heating bath. The suction pump is used to move the sample up and down within the capillary so that no one part of the sample is overexposed to the x-ray beam. This mechanism is controlled by a Microlab 600 pump and computer. A Thermo Neslab RTE-740 Digital Plus Refrigerated and Heating bath is used to control the sample temperature. This bath is filled with antifreeze and distilled water to prevent freezing and calcium build up. Also attached to the brass sampling mount is a temperature probe; data collection corresponds to the temperature reported by this probe. In order to maintain sample temperature during data collection the brass sampling mount and hose are insulated using styrofoam casing (see figure 4.4). At temperatures below freezing, helium is pumped into the brass mount so that ice crystals do not form in the air. The x-ray beam leaves the vacuum chamber before interacting with the sample and re-enters another vacuum chamber of length 1 meter following scattering. Scattered x-rays are detected by a MAR 165 CCD detector. The detector pixel size is $80 \mu m^2$ which make up 2048 x 2048 pixel area of diameter $1.62m$. The sample-to-detector distance was $1.539m$. The focused x-ray beam has a size (prior to use of slits) of $150\mu m$ x $50\mu m$ (horizontal x vertical), and a flux of 2 x $10^{13} photons/sec$ at $10keV$. The x-ray energy used in this experiment was $12keV$. An outline of this experimental setup can be seen in figure 4.3.

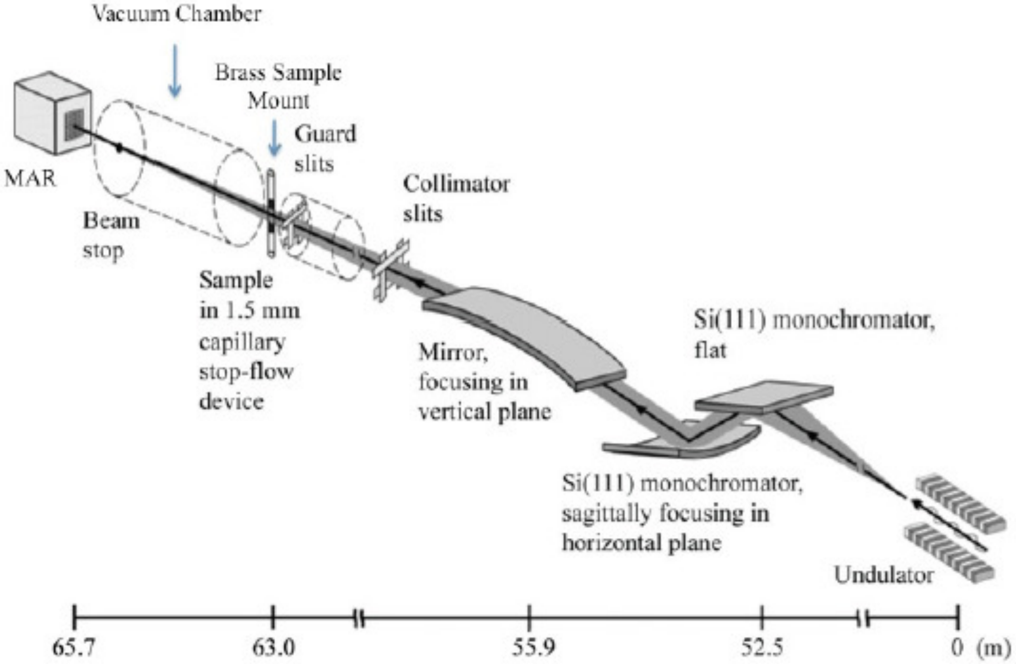


Figure 4.3: The experimental setup for x-ray focusing and data collection (Fischetti, 2004).

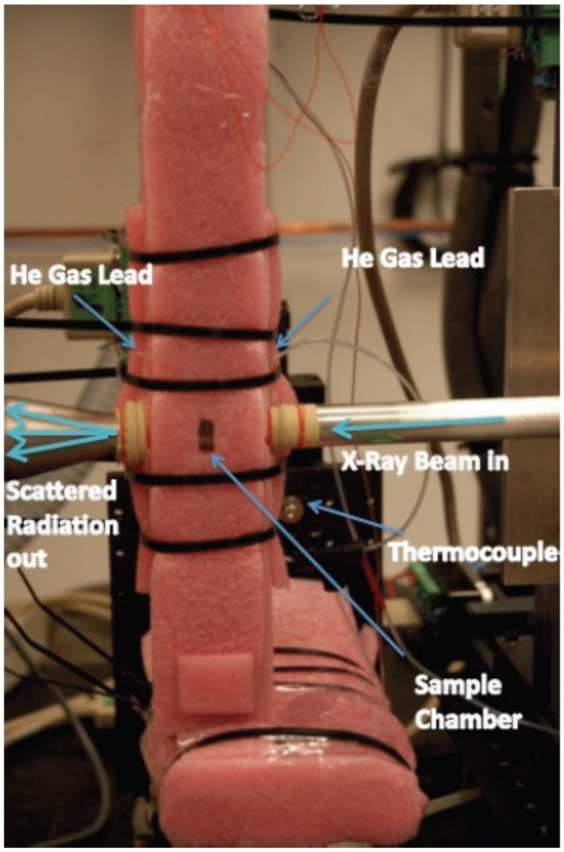


Figure 4.4: The insulated brass sampling mount.

4.2 Sample Preparations

Cytochrome-c is prepared in solution that allows examination of the protein over the chosen temperature range, and that reflects the naturally occurring environment of the protein. The solution consists of a number of components in order to achieve the goals above. Firstly, it contains a mixture of water and ethylene glycol (MEG) so that the solution will not freeze. Also, ethylene glycol does not affect the scattering of x-rays, and does not react with the protein. 100ml of this mixture is prepared, 55ml of which is H_20, 45ml of which is MEG. Next, a buffer is needed to control the pH level of the solution. Consistent pH level is necessary so that any changes in the protein are a result of temperature only. The buffer consists of $0.5844g$ of NaCl at $0.15M$ concentration and $10ml$ of Tris buffer at $0.1M$ concentration. These amounts result in a pH level of ~ 3.5 so that radius of gyration measurements can be compared to literature values (Akiyama, 2002, Pollack, 1999). This mixture of H_20, MEG, NaCl, and Tris acts as the backbone to the eventual solution which contains the protein and denaturant. In preparing the denaturant solution, $31.5g$ of crystalline guanidine HCl (GuHCl) is added to a $50ml$ conical, which is then filled to the meniscus with the H_20/MEG/NaCl/Tris backbone solution. This results in a bulk $6.6M$ guanidine HCl solution which is used throughout the experiment to denature the protein. The protein solution is made in a similar manner. For performing SAXS measurements, protein concentration should be at least 2 milligrams of protein per milliliter of solution. With this in mind a solution of $4mg/ml$ is ideal for taking data. To achieve such concentrations, a bulk solution of $40mg/ml$ is prepared which is then diluted to $4mg/ml$ for experimentation. $50mg$ of cytochrome-c, in cyrstalline form, is added to a $2ml$ conical which is then filled to the meniscus with the buffered H_20/MEG/NaCl/Tris backbone, thereby providing ample protein solution for experimentation. A more detailed treatment of the sample preparation process has been carried out by by Elmer (Elmer, 2010). Bulk preparation of the protein and denaturant solutions ensures a coherent data collection process. With separate protein and denaturant solutions, SAXS measurements may be performed at various levels of denaturant concentration. In performing such measurements, both background (with denaturant) and sample (with protein and denaturant) data collections are carried out. For this reason, a different background solution is required depending on the level of denaturant concentration. In this experiment, three levels of denaturant are of interest. These are the $0M$, $2.5M$, and $4M$ cases. The $0M$ and $4M$ solutions allow us to observe the protein in its native and denatured states respectively, while the $2.5M$ solution allows us to observe change in protein conformation by varying the temperature of the solution. For each data run, be it

$4mg/ml$ Protein in Dentaurant			
	Volume of Solution		
Solution	$0M$ GuHCl	$2.5M$ GuHCl	$4M$ GuHCl
$40mg/ml$ Cytochrome-C	$10\mu l$	$10\mu l$	$10\mu l$
$6.6M$ GuHCl	$0\mu l$	$37.5\mu l$	$60\mu l$
H_20/MEG Buffered Mix	$90\mu l$	$52.5\mu l$	$30\mu l$

Table 4.1: Volumes of solutions for cytochrome-c measurements at various denaturant concentrations.

Background Denaturant Solution			
	Volume of Solution		
Solution	$0M$ GuHCl	$2.5M$ GuHCl	$4M$ GuHCl
$6.6M$ GuHCl	$0\mu l$	$37.5\mu l$	$60\mu l$
H_20/MEG Buffered Mix	$100\mu l$	$62.5\mu l$	$40\mu l$

Table 4.2: Volumes of solutions for background measurements at various denaturant concentrations.

background or sample, $100\mu l$ of solution are required. The protein and background recipes for each level of denaturant are shown in tables 4.1 and 4.2 respectively.

The buffer solution for performing SAXS measurements on miro is more complicated. While a concise version is outlined here, a full description of the recipe is given by Northwestern University's Rice Lab (manuscript in preparation). It consists of $25mM$ HEPES buffer, $0.5mM$ of the reducing agent TCEP, 5% sucrose, $100mM$ of potassium chloride (KCl), $1mM$ of the chelating agent EGTA, and $2\mu M$ of guanosine triphosphate (GTP). This acts as the background solution to which miro is added during experimentation.

4.3 Sample Loading and Cleaning

In each experimental run, data collection is first performed on the background solution. Prior to this however, the sample mount is cleaned using a five-step process.

For this process, five 50*ml* conicals are prepared; three of H_20/MEG mix, one of isopropyl alcohol, and one of a 20% bleach solution. Each 50*ml* conical is loaded into the sample mount and using the MicroLab 600 pump system, 3*ml* of each solution are pumped through a hose attached to the capillary in the following order: H_20/MEG - isopropyl alcohol - H_20/MEG - 20% bleach - H_20/MEG. Following the final H_20/MEG cleaning and removal of its conical, the pump is allowed to continue until the system is evacuated. Once this cleaning process is finished, the background solution is loaded, by inserting the 2*ml* conical of 100μl solution into the brass sampling mount. The capillary hose is placed in the conical such that it reaches the bottom and the sample is pumped into the capillary by selecting the GL-Protein option on the MicroLab 600 interface. Before beginning the pumping sequence that takes place throughout each data collection, the MAR 165 detector is triggered by exposing the sample to the beam. This cleaning process is performed after each collection of protein data. It is not necessary to clean the system after each background run since the background data will be subtracted from the protein data during the data analysis.

4.4 Data Acquisition

After cleaning and sample loading the hutch is evacuated according to safety procedure, the lead doors are closed, and the x-ray shutter is opened. On the user computers outside the hutch, a filename is chosen corresponding to the protein conditions. The filename should reflect the level of denaturant and temperature conditions. For each set of temperature/denaturant conditions, 15 background and 15 sample scattering images are recorded (20 and 20 for miro). For background data collection, the filename is numbered 1. Successive filenames are updated by the software to reflect each image acquisition. Similarly for protein sample data collection, the file is numbered 16 (21 for miro). This identifies background images as numbered 1 through 15 (20 for miro) and protein sample images as numbered 16 through 30 (21 through 40 for miro). Two computers are used in data collection. One to acquire detector images and another to process and save the data. After choosing the filename, a baseline signal is acquired using the processing computer. This computer is now ready for data collection. The imaging computer is now used to trigger the detector. At this point, the externally controlled pumping process should also begin. Data can now be acquired. The GL-Protein setting in the MicroLab 600 pump is designed such that the sample is pumped continuously throughout the data collection process, and deposited back into the conical upon completion. Once the collection run is finished the x-ray shutter is closed and the sample can be removed, with cleaning carried out if necessary. A quick preliminary analysis of the data is carried out on-site to ensure quality of data collected. Should the data be problematic (e.g. presence of air bubbles, ice deposits) another run may be performed. Otherwise, the next desired conditions are achieved, and collection continues.

The process outlined above is repeated until all desired variations in protein conditions are met and data acquired is of acceptable quality. Following this the data is ready to be analyzed.

CHAPTER 5

Data Analysis

5.1 Introduction

The following chapter describes the analysis of data acquired at Sector 18 of Argonne National Laboratory's Advanced Photon Source between December 16th and December 21st of 2009. This procedure could also be adopted to other samples of protein in solution. There are several steps in the analysis procedure. First, the raw data is reduced to obtain a one dimensional scattering curve. Guinier analysis is then performed on the dataset. These first two steps were performed on-site at the APS and were carried out again at a later date. The third step is using an indirect Fourier transform to obtain the pair-distance distribution function for the dataset. Following this, a simulated annealing procedure is performed on the data to obtain a three dimensional reconstruction of the protein. The previous two steps are carried out using the ATSAS software package, from the EMBL Hamburg Biological Small Angle Scattering group. Once a 3D structure has been obtained, its corresponding scattering curve may be calculated and compared to the original scattering input data using the FoXS server, courtesy of the University of California San Francisco. This analysis procedure can be seen in the flowchart of figure 5.1. As a final step in the analysis, we view our 3D structure in the application PyMol, where it may be compared to the known crystallographic structure for the protein, if applicable. In describing the analysis procedure, a sample dataset for cytochrome-c in the native state will be used throughout. Since this protein has a known, well-studied structure, it will act as our control data.

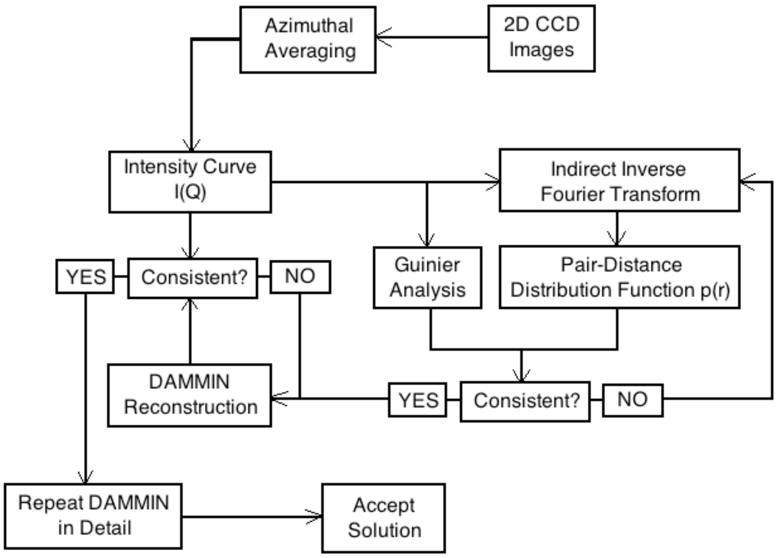


Figure 5.1: Flowchart of the data analysis process. Consistency between the Guinier and $p(r)$ analyses refers to production of a well-behaved $p(r)$ curve which has an R_g as close as achievable to that of Guinier analysis. Consistency in the DAMMIN reconstruction refers to multiple, identical DAMMIN runs, each producing a similar $I(Q)$ curve which is as close as achievable to that of the original data.

5.2 Data Reduction

For each data acquisition, 15 background images and 15 images with the protein included are recorded by the detector. Each image is the result of a 1 second exposure to the full intensity beam. The images are of the form shown in figure 5.2.

This data is analyzed in the program IgorPro using a Macro created by Liang Guo, at Sector 18 of the APS. Seen in the lower half of figure 5.2 is the beam stop, which is in place to protect the detector from damage by unscattered x-rays. The 2D radially symmetric intensity images are azimuthally averaged at each radial distance, to give an intensity at each radial distance. The beam stop in figure 5.2 is automatically subtracted from the data at this point. Plotting this reduced data using IgorPro

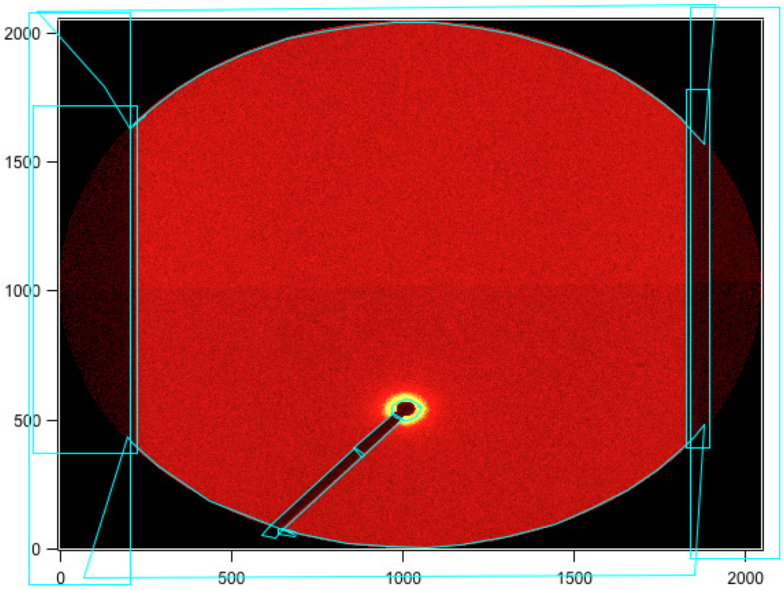


Figure 5.2: Raw data as acquired by the MAR 165 CCD detector.

gives figure 5.3.

The upper curve in figure 5.3 is comprised of the 15 images taken with the protein present in the solution. The lower curve consists of the 15 images without the protein, i.e. the background data. In some cases, the curve will contain outliers to the data. The primary cause of outliers in the data are major inhomogeneities in the sample, such as air bubbles. Outliers may also be caused by dust and in the case of temperatures at or below freezing, ice. These may be removed in IgorPro prior to proceeding. In the case of the cytochrome-c data in figure 5.3 this step is not necessary. An example of scattering data in which outliers should be removed is shown in figure 5.4.

One can see in figure 5.4 that the individual curves do not form a cohesive bunch as in the case of figure 5.3. In this situation the data is trimmed by removing individual curves. Once the outliers have been removed, the background and protein data must be averaged, with the average background data then subtracted from the

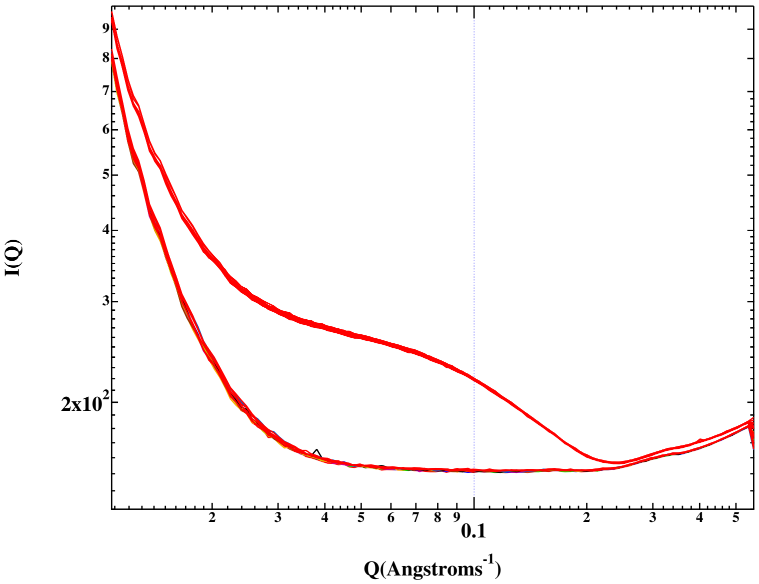


Figure 5.3: Log-log plot of raw reduced scattering data for cytochrome-c in a $0M$ guanidine HCl solution at 0^oC. Error bars due to x-ray counting statistics have been omitted.

average protein data. This background subtracted curve is the final result of this step of the data reduction process. Figure 5.5 shows a background subtracted curve corresponding to figure 5.3.

5.3 Guinier Analysis

The scattering pattern in figure 5.5 shows similar characteristics to the pattern of a homogeneous sphere, as seen in figure 3.4. Before performing Guinier analysis, protein shape may be investigated by utilizing a Kratky plot. In the case of our sample cytochrome-c, figure 5.6 gives further information on the protein shape.

Figure 5.6 clearly shows the globular nature of the protein in this solution. Since a Kratky plot is very sensitive to the behaviour of polymer chains, it will be performed for each dataset to ensure consistency and check for deviations from expected behaviour. In this way, the Kratky plots are also useful in deciding if a dataset is

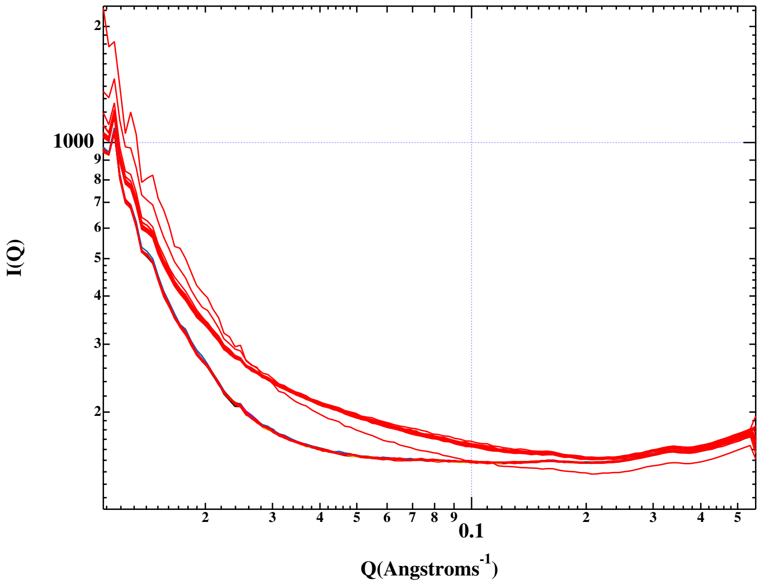


Figure 5.4: Log-log plot of raw reduced scattering data for cytochrome-c in a $2.5M$ guanidine HCl solution at 50^oC. Error bars due to x-ray counting statistics have been omitted.

considered bad or faulty. The size of the protein may be calculated by obtaining its radius of gyration from a Guinier plot, as outlined in Chapter 3. For our sample cytochrome-c, the Guinier plot takes the form of figure 5.7.

As described in chapter 3, the initial portion of this plot is of interest when obtaining the radius of gyration, R_g. An IgorPro macro is used to perform a line of best fit and calculate a value for R_g. For our sample cytochrome-c the radius of gyration is $R_g = 13.47 \pm 0.01$Å, which agrees with literature value (Hsu, 2007). This radius of gyration is recorded and is used to ensure consistency during later steps of the analysis.

5.4 The Pair-Distance Distribution Function

Evaluation of the pair-distance distribution function, $p(r)$ is a necessary step in obtaining a 3D reconstruction from our 1D dataset. However, the result of this

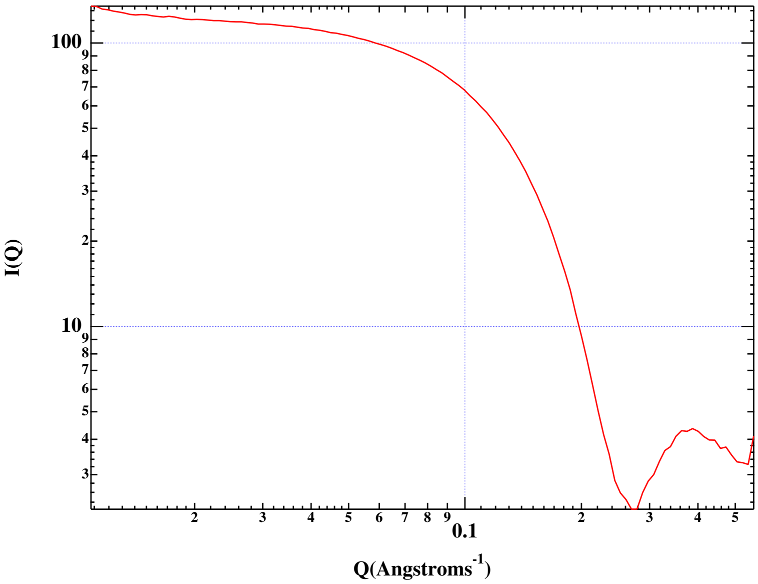


Figure 5.5: Log-log plot of averaged background subtracted data for cytochrome-c in a $0M$ guanidine HCl solution at 0^oC. Error bars due to x-ray counting statistics and standard deviation from averaging have been omitted.

step is valuable in and of itself. Prior to this section, analysis has been carried out in IgorPro macros available from Sector 18 of the APS. To acquire the pair-distance distribution function, we utilize the ATSAS software package from the EMBL Hamburg Biological Small Angle Scattering group. Specifically, we use the program GNOM. GNOM reads in one dimensional scattering curves and evaluates the pair-distance distribution function by carrying out an indirect Fourier transform. Several input parameters are required of the user by GNOM. After choosing the input and assigning an output file name, the user has the choice to omit a number of initial and final data points. Based on the quality of the scattering curve, these parameters may or may not have to be adjusted from the default (zero). The most significant parameter the user must define is the $R_{\max}$ input, i.e. the maximum particle diameter. This will effect the radius of gyration which GNOM calculates from the $p(r)$ function. In the case of our sample dataset, we saw previously that

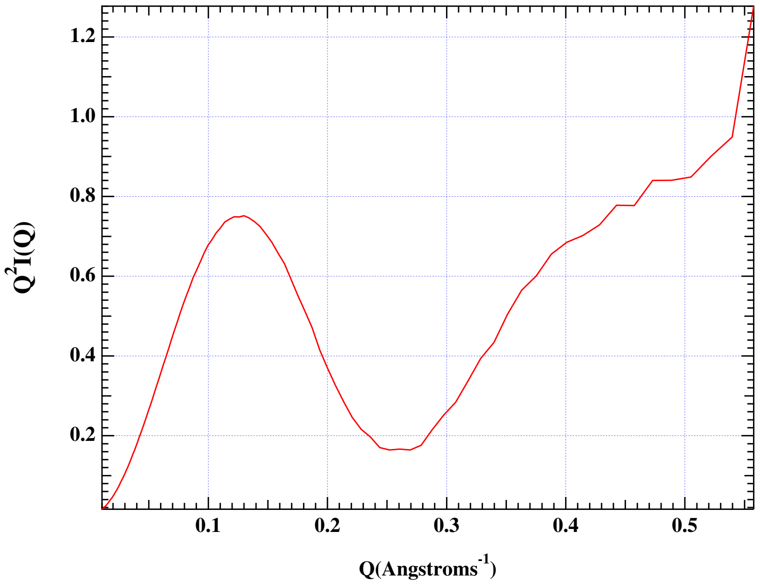


Figure 5.6: Kratky plot of background subtracted data for cytochrome-c in a $0M$ guanidine HCl solution at 0^oC. This plot indicates a compact structure.

the protein has a somewhat spherical shape. In order to decide a value for $R_{\max}$ we calculate the radius from our value of R_g obtained during Guinier analysis, using the formula (see chapter 3)

$$R = \sqrt{\frac{5}{3}} R_g \tag{5.1}$$

and doubling the result. All other other parameters may be left as their default values. For a detailed description of the GNOM analysis, and a sample of its output, one should see the Appendix. For a detailed description of the GNOM program in general, the manual is available on the EMBL Hamburg website. Should the input parameters be satisfactory, GNOM outputs a $p(r)$ curve for the dataset. For our sample cytochrome-c, this is shown in figure 5.8

The $p(r)$ curve in figure 5.8 has a radius of gyration of $R_g = 13.00 \pm 0.02$Å, which is also consistent with literature values (Hsu, 2007). This value is calculated using equation 3.42. It is immediately noticeable that this value is significantly different

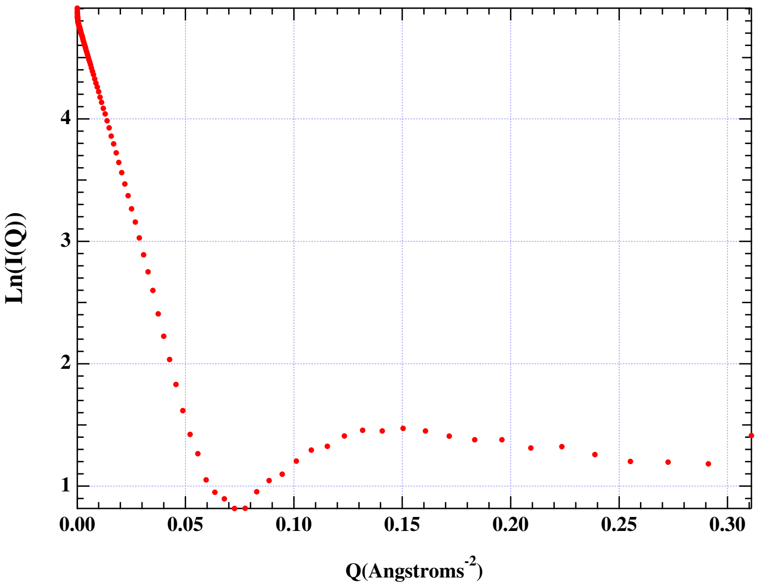


Figure 5.7: Guinier plot for cytochrome-c in a $0M$ guanidine HCl solution at 0^oC. Error bars from counting statistics have been omitted. The slope of the low-Q region is used to calculate the radius of gyration.

than that obtained via Guinier analysis ($R_g = 13.47 \pm 0.01$Å). The reason for this discrepancy likely lies in the fact that Guinier analysis is carried out at low-Q values only, i.e. it is a small-angle approximation, reliable only below $R_gQ_{max} \approx 1.3$. The calculation of radius of gyration from the $p(r)$ curve uses the entire Q-region. Given that for SAXS measurements greater error occurs at high-Q, it is unsurprising that the two values differ. While both values fall within accepted literature range, the $p(r)$-calculated value is likely more accurate since it uses the entire $I(Q)$ curve. The output $p(r)$ curve should be smooth, continuous, decaying to zero, with low error, and no oscillations. If the output curve does not display these characteristics, or the resulting radius of gyration disagrees drastically with the Guinier analysis, the GNOM procedure should be repeated using a different input for $R_{\max}$ until a satisfactory result is obtained.

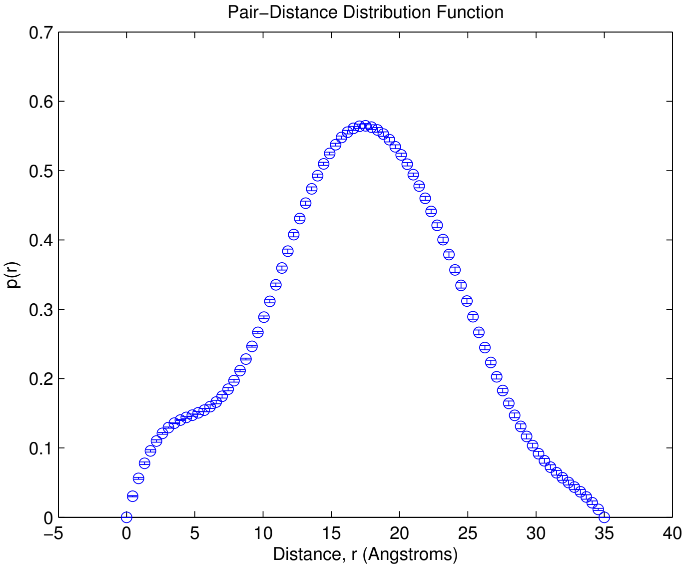


Figure 5.8: Pair-distance distribution function plot for cytochrome-c in a $0M$ guanidine HCl solution at 0^oC. Error bars are contained within the circular data point markers.

5.5 Three Dimensional Reconstruction

Reconstruction of the protein shape is carried out in the program DAMMIN (Svergun, 1999), a part of the ATSAS software package. DAMMIN uses a simulated annealing algorithm to determine protein shape from a dummy atom model. The input data for DAMMIN is the output from GNOM, i.e. the $p(r)$ data and associated parameters. As with GNOM, several input criteria must be defined by the user. The most important of these is choosing an initial shape (i.e. dummy atom model) for the protein. Based on the analysis of our sample cytochrome-c dataset, it is sensible in this case to choose a spherical starting model. The maximum diameter of this sphere is dictated by the parameters of the input $p(r)$ data (another reason why choosing an appropriate $R_{\max}$ input is important in GNOM). In the case of our sample dataset, all other parameters may remain as their default values. For greater detail regarding this procedure and for a sample output, please see the Appendix.

For greater detail on the DAMMIN program itself, one should see the DAMMIN manual online at the EMBL Hamburg website. DAMMIN begins its reconstruction from the user defined dummy atom model. A scattering intensity curve is calculated for this model and compared to the actual scattering data as described in Chapter 3. The simulated annealing algorithm relocates dummy atoms within the model and modifies the temperature, all while making further comparisons to the experimental data. This process is repeated by DAMMIN until the discrepancy between the dummy atom model's scattering curve and the original data is minimized. The output of DAMMIN is a PDB file which contains coordinates for the locations of the dummy atoms in the final reconstructed model, along with the associated parameters for the model. Since this solution is not unique, i.e. there may be more than one reconstruction which gives a scattering curve in agreement with the input data, the DAMMIN process is repeated several times. A snapshot of the DAMMIN graphics window when it has finished its reconstruction can be found in figure 5.9.

Since the DAMMIN algorithm has converged on a solution in this case, the red line in figure 5.9 corresponding to the scattering curve of the final dummy atom model is obscuring visibility of the line of best fit of the input scattering data. The output PDB file gives a radius of gyration for this solution as $R_g = 13.44\text{Å}$, in very close agreement with that of the GNOM output, and in good agreement with the Guinier Analysis. Using the application PyMOL, the PDB file may be viewed in its 3D form (see figure 5.10).

5.6 Comparison of Scattering Intensities

In this section we make use of the FoXS (Fast X-Ray Scattering) server, a freely available online application from the University of California San Francisco. The FoXS server allows the user to select a PDB file from the RCSB Protein Data Bank

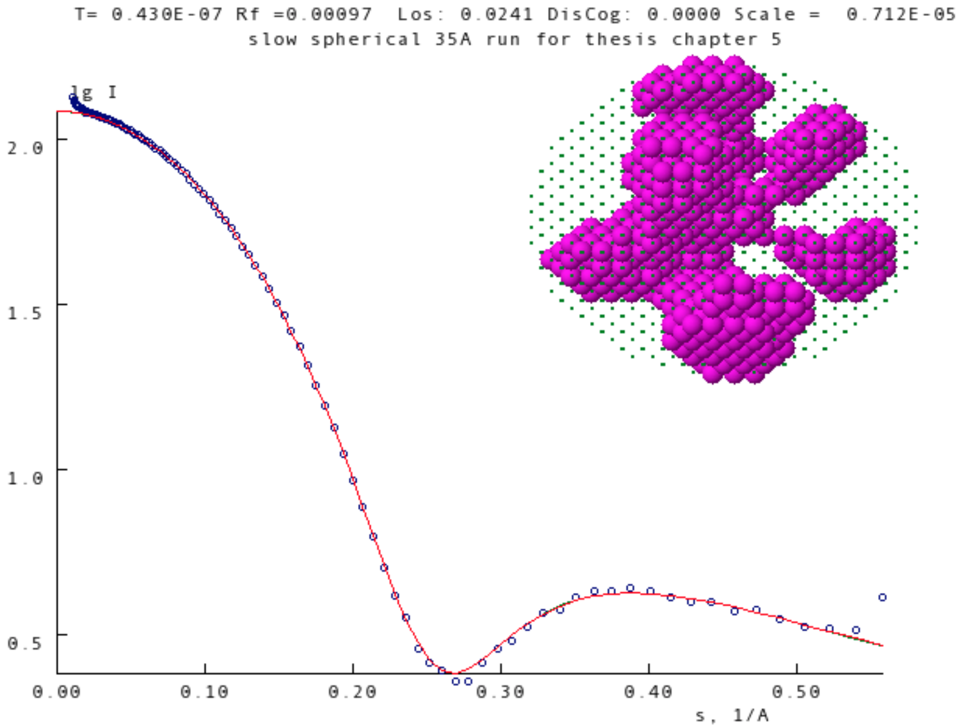


Figure 5.9: The graphics window displayed by DAMMIN during each step of the simulated annealing process. This image is of the final reconstruction, showing the final dummy atom model with its corresponding scattering curve (red line), and the input scattering data (blue) with best fit line (green).

or upload a PDB file of their own, and compare the corresponding scattering intensity to a SAXS experimental data file, also uploaded by the user. The FoXS server uses the Debye formula for spherical scatterers in computing a scattering profile from the input PDB file. The server outputs a text file for this scattering pattern and also plots this pattern with the uploaded SAXS data, providing a chi-squared value to quantify the level of agreement between the two. In the case of our sample cytochrome-c, both the results from x-ray crystallography and our DAMMIN reconstruction are compared to the original SAXS data. The x-ray crystallography data are found on the Protein Data Bank website under the identifier 1HRC (Bushnell, 1990). The output text files from the FoXS server are loaded into MATLAB where

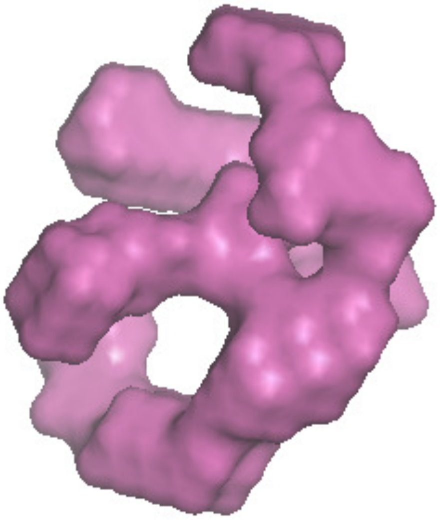


Figure 5.10: Three dimensional reconstruction of cytochrome-c in a $0M$ GuHCl solution at 4^oC by the program DAMMIN. For scale, this object has a maximum diameter of 34.69Å.

scattering curves are constructed.

Figure 5.11 shows good agreement between the three sets of data, with some discrepancy occurring at high and low ranges of Q. While the original SAXS data appears to differ at very low Q, it is the agreement between our reconstruction and crystallography results that is of more interest in this range. At high Q, we may expect some disagreement since the resolution of our SAXS measurements is not as fine as that of crystallography. The area of high Q may be observed in greater detail by comparing the Kratky plots corresponding to each of the three datasets. This may be seen in figure 5.12.

In quantifying the level of agreement between each dataset, FoXS reports a chi-

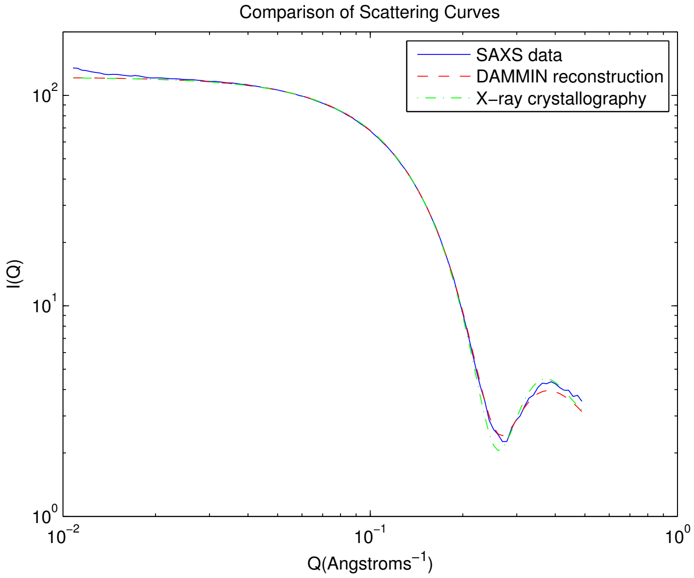


Figure 5.11: Comparison of intensity scattering curves for SAXS measurements, DAMMIN reconstruction, and x-ray crystallography (PDB ID: 1HRC) for native cytochrome-c.

squared value of 1.37 for the SAXS data and DAMMIN reconstruction, and a chi-squared of 1.5 for the SAXS data and crystallography results. A value of 1.5 or better is considered in literature to be a sign of sufficient agreement with crystal structure. Investigation of figure 5.12 provides evidence as to where the disagreement may occur. The difference in data points at higher values of Q, particularly for $Q > 0.2\text{Å}^{-1}$ is a likely cause of a high chi-squared value. As previously mentioned, the nature of the SAXS experiment is not favorable to determining great detail, and as such it is more important that our results be in good agreement at low to mid-Q values. What we may take from this section of the analysis is that the overall procedure is a reliable method for determining protein shape and size. To further bolster this conclusion we may use PyMOL to compare our data to accepted structures.

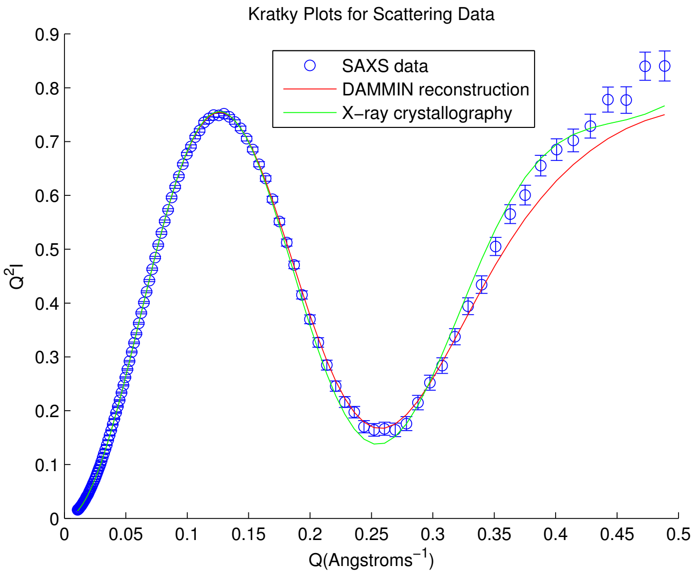


Figure 5.12: Comparison of Kratky plots for SAXS measurements (with error bars), DAMMIN reconstruction, and x-ray crystallography (1HRC) for native cytochrome-c.

5.7 Comparison of Three Dimensional Structure

While there are many ways to view a reconstruction in PyMOL, it is most useful in this case to compare the reconstruction to the known crystallographic structure, 1HRC, used in the previous section. Figures 5.13 and 5.14 show the x-ray crystallography results for equine cytochrome-c, and those results “docked” with our reconstruction of the protein shape, respectively.

As can be seen in figure 5.14 our reconstructed shape for cytochrome-c in the native state is consistent with the accepted structure based on x-ray crystallography. We do not expect the structure of 1HRC to fit exactly with our native-state ctytochrome-c reconstruction, since we are comparing a crystallized protein with one in solution. Even though thermodynamic conditions are such that the protein in solution is as close as possible to its native state, the details are not expected to be an exact match.

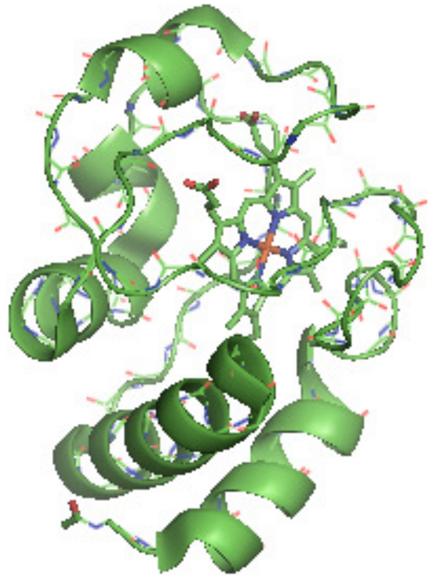


Figure 5.13: 3D structure of equine cytochrome-c, a result of x-ray diffraction on the crystallized protein. Available from the RCSB Protein Data Bank (ID: 1HRC). For scale, the above object has a maximum diameter of $\sim 34\text{Å}$.

What is expected is that the two structures have a similar size (both in maximum diameter and radius of gyration), and produce a similar scattering intensity. While not a quantifiable method of measuring agreement, the PyMOL comparison ensures that our structure agrees with accepted results. The above procedure has now been shown to be a viable method for determining a 3D structure of our protein based on 1D scattering intensity measurements. This method has been carried out over numerous sets of data for cytochrome-c, with results of similar quality. These results may be found in the Appendix. We may now use this method to analyze the warm

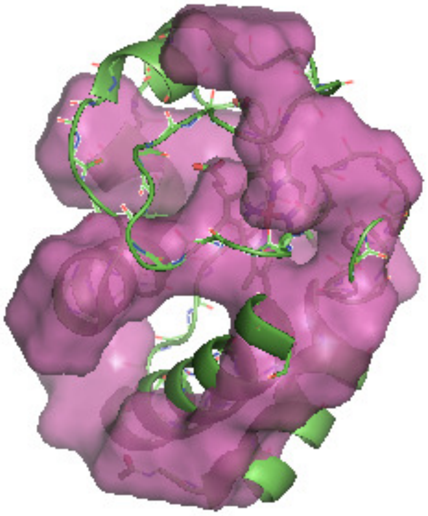


Figure 5.14: Structure from figure 5.14 docked with our reconstructed cytochrome-c model. (5.10)

and cold denatured states of cytochrome-c and the miro-S and miro-L proteins. Results for these measurements are outlined in succeeding chapters.

CHAPTER 6

Cytochrome-C Results

In this chapter we review the results for cytochrome-c, at denaturant concentrations of $0M$, $2.5M$, and $4M$ over a temperature range of -24^oC to 50^oC. The $0M$ and $4M$ measurements, representing the native and denatured states respectively, each give consistent values for the radius of gyration over the the temperature range. The $2.5M$ measurements allow observation of cold and warm denaturation over the given temperature range. Expectations for values of radius of gyration can be seen in figure 3.5. The radius of gyration is determined in two ways. Firstly via Guinier analysis, and secondly via analysis of the pair-distance distribution function. Consistency in these values indicates a successful Fourier transform into real space. Following this, reconstructions of the proteins are presented. For the $0M$ and $4M$ cases, one reconstruction of each is presented, as their consistency makes multiple presentations redundant. A reconstruction will be presented at each temperature for the $2.5M$ concentration. Reconstructions are deemed successful on two criteria. Firstly, their radius of gyration should be consistent with that of the Guinier and $p(r)$ analysis. Secondly, the program FoXS is used to calculate the scattering pattern for each reconstruction and ensure these patterns are consistent with the SAXS measurements at the corresponding conditions.

6.1 Inverse Space Guinier Analysis

We begin this chapter by presenting results for Guinier analysis in inverse space. These results are obtained from SAXS intensity measurements. Scattering intensity for all three denaturant concentrations of cytochrome-c can be seen in figures 6.1, 6.2, and 6.3.

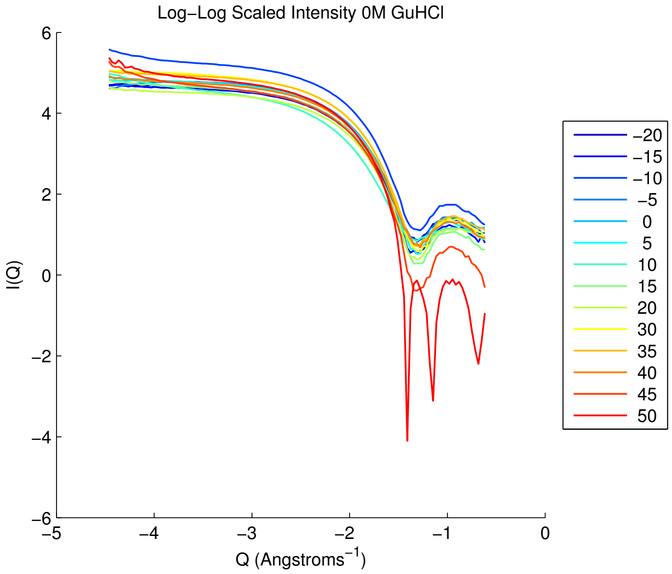


Figure 6.1: Log-scaled scattering intensity curves for cytochrome-c at $0M$ denaturant concentration over a temperature range of -20^oC to 50^oC.

.

As expected, the scattering curves for different temperatures have similar shapes for the $0M$ and $4M$ cases. In both cases, the curves that differ most significantly from those at other temperatures are the extreme high and low temperature curves. Furthermore, for each of the three concentrations, a greater difference between scattering curves is seen in the high-Q range. These discrepancies at high-Q do not affect the results of Guinier analysis, since such analysis is confined to the low-Q region, i.e. in the range of $1.0 < R_gQ_{max} < 1.3$. Results of pair-distance distribution function analysis, however, incorporate the entire Q-range and as such a difference from Guinier analysis results is expected. While the intensity curves are useful when compared to each other, corresponding Kratky plots are far more useful for investigating the shape of the protein, and for checking whether the data is of adequate quality or not. For instance, since we know that cytochrome-c in the native state

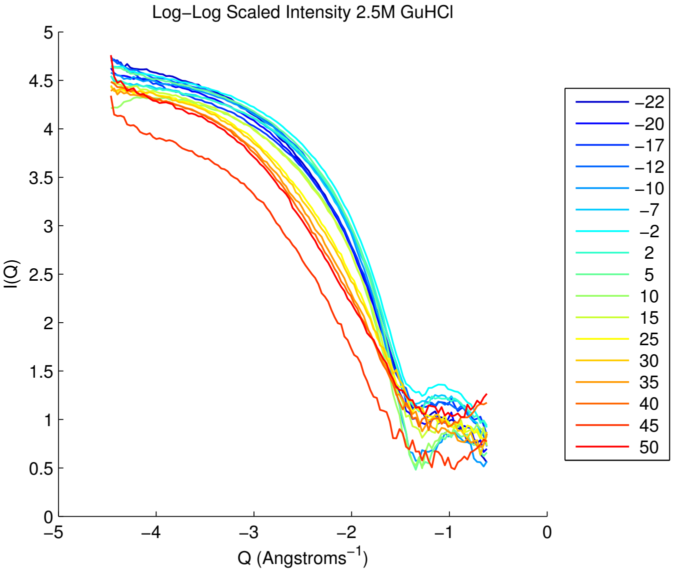


Figure 6.2: Log-scaled scattering intensity curves for cytochrome-c at $2.5M$ denaturant concentration over a temperature range of -22^oC to 50^oC.

.

is globular in nature, Kratky plots for the $0M$ data should exhibit with increasing Q an initial peak followed by a decay and subsequent increase, indicating the overall shape, random walk behaviour, and thin rod-like behaviour (characterized by the persistence length). In the $4M$ case however, the Kratky plot should not exhibit such a pronounced initial peak, since the protein is denatured and less globular in shape, i.e. unfolded. Kratky plots for all temperatures at each concentration can be seen in figures 6.4, 6.5, and 6.6.

The shapes of the Kratky plots are consistent with our expectations. For the $0M$ data, an initial peak is clearly present across all temperatures, indicative of the protein in its native, folded state. The large differences between curves of various temperature in figure 6.4 are expected due to the high sensitivity of the Kratky plot at high-Q values. This graph confirms the presence of cytochrome-c in the native state. In figure 6.6 the opposite situation is apparent for the $4M$ data. Kratky plots

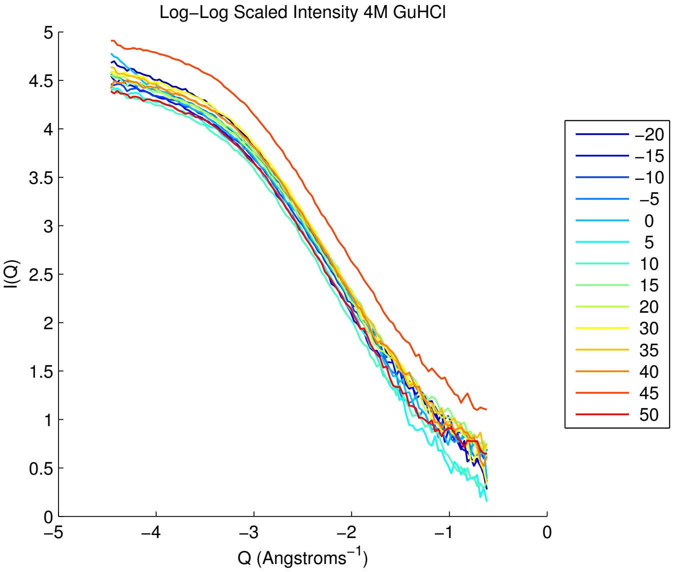


Figure 6.3: Log-scaled scattering intensity curves for cytochrome-c at $4M$ denaturant concentration over a temperature range of -20^oC to 50^oC.

.

for all temperatures show no initial peak, instead increasing gradually with Q, and exhibiting random walk characteristics followed by thin rod-like behavior at high-Q. Again, this is as expected, with persistence length characterizing the Kratky plot at high-Q (shorter real-space length scale). With the exception of the 45^oC data, the difference at high-Q between plots of various temperatures is relatively small, indicating a consistency in structure across all temperatures.

The case of most interest however, is the $2.5M$ data, seen in figure 6.5. From this figure we can infer that at different temperatures, the protein has significantly different shape. At high temperature, for example, there is clear lack of evidence for a globular structure, indicating the protein in its denatured state. The temperatures at which the most globular behavior is exhibited are those in the lower-middle section of our range. Specifically, protein in the -10^oC to 5^oC range shows the most pronounced initial peaks. Structures with the least globular shape appear

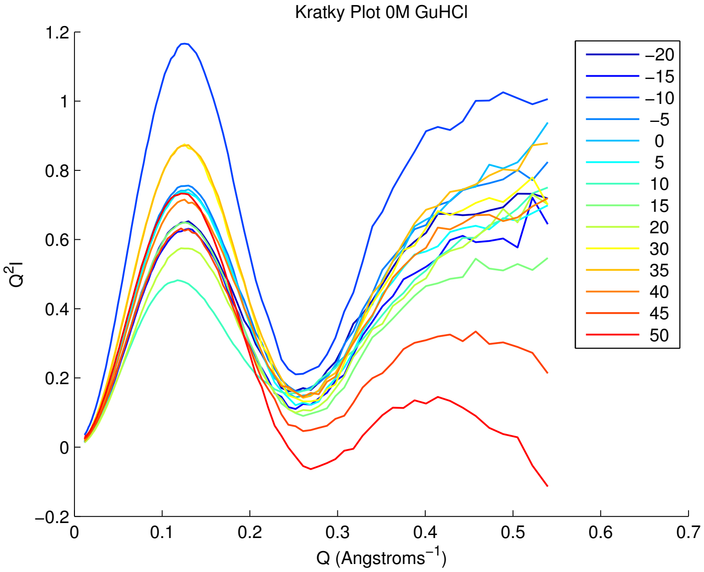


Figure 6.4: Kratky plots of $Q^2I(Q)$ against Q for cytochrome-c at $0M$ denaturant concentration over a temperature range of -20^oC to 50^oC.

.

at $T > 25^oC$, evidenced by a lack of initial peak. Between these two structural extremes, several plots (-20^oC to -12^oC and 10^oC to 15^oC) are present. These plots indicate a globular nature, albeit smaller than that of the low-middle temperature range.

Investigating the Kratky plots in this qualitative manner provides one with a good expectation for the results of subsequent Guinier analysis. Guinier analysis of the data takes place at low-Q. Figures 6.4 and 6.6 show little variation in shape at low-Q, particularly when compared to figure 6.5. Thus, we would expect little change in the size of our protein distribution at the $0M$ and $4M$ concentrations. For the $2.5M$ concentration, the variation at low-Q indicates a change in protein size across the temperature range. The values for the radius of gyration from Guinier analysis at each concentration can be seen in figure 6.7.

Figure 6.7 contains values resulting from repeated analysis by Elmer (Elmer, 2010)

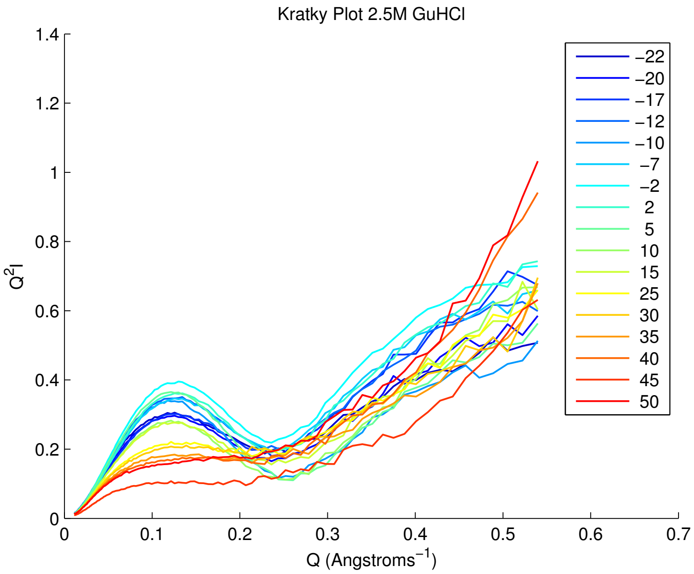


Figure 6.5: Kratky plots of $Q^2I(Q)$ against Q for cytochrome-c at $2.5M$ denaturant concentration over a temperature range of -22^oC to 50^oC.

.

and the author. These results are consistent with expectations stemming from observations of the Kratky plots, with the low-middle temperature range of the $2.5M$ data having smaller radii of gyration than the data from lower and higher temperature. Expected consistency is also evident for the $4M$ and, in particular, the $0M$ data. Even with the significant error associated with several of the $4M$ values, these results are consistent with literature values (Pollack, 1999, Hsu, 2007). Errors on data points of figure 6.7 arise from performing a low-Q linear fit to the Guinier plots of the data in IgorPro, as described in Chapter 5. With the Guinier analysis results now complete, results of the data in real space using the pair-distance distribution function are now be presented.

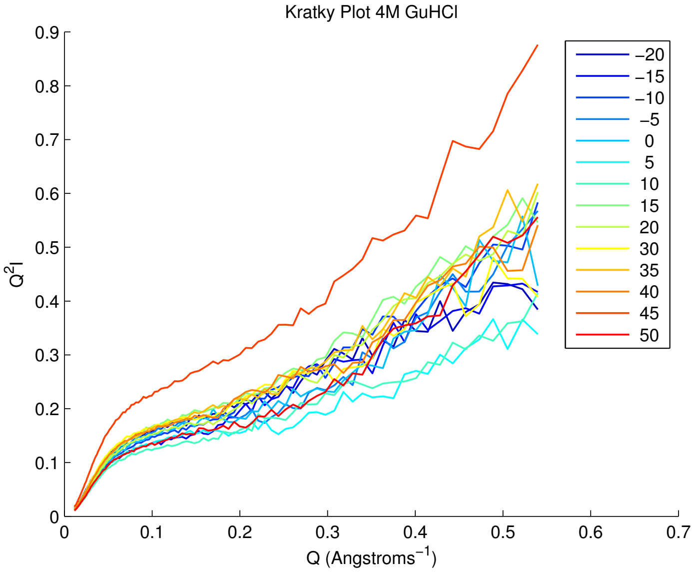


Figure 6.6: Kratky plots of $Q^2I(Q)$ against Q for cytochrome-c at $4M$ denaturant concentration over a temperature range of -20^oC to 50^oC.

.

6.2 Real Space $p(r)$ Analysis

Following the Fourier transform procedure described in Chapter 3, a pair-distance distribution function is produced for each dataset, as described in the example in Chapter 5. Given that the $p(r)$ function is obtained from the scattering intensity $I(Q)$, we should see a consistency in the radii of gyration obtained from the $p(r)$ curves and Guinier analysis. Furthermore, calculation of the radius of gyration from the $p(r)$ curve uses the entire Q-range, and thus one would expect the result to be more reliable than that of Guinier analysis, as previously mentioned. For the $2.5M$ case, $p(r)$ curves can be seen in figure 6.8.

The $p(r)$ curves in figure 6.8 are those that have met the criteria for an acceptable pair-distance distribution function as outlined in Chapter 5, i.e. no oscillations, smooth decay to zero, low error. Error bars, on the order of 10^{-3}, have been omitted

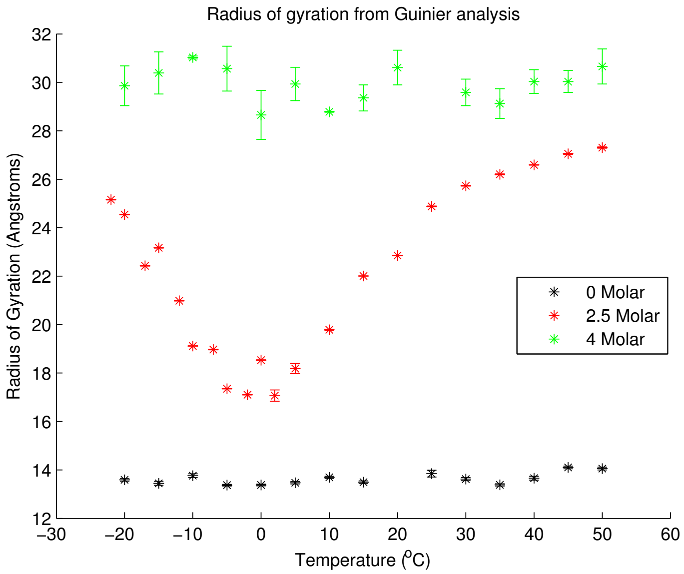


Figure 6.7: Radius of Gyration from Guinier Analysis for Cytochrome-C at $0M$, $2.5M$, and $4M$ denaturant concentrations over a temperature range of -22^oC to 50^oC

.

from figure 6.8 for the sake of presentation. We can expect a folded protein to have a greater peak on its $p(r)$ curve since a greater number of pairs of particles will be separated by a shorter distance, r, apart. With this in mind, the shape of the $p(r)$ curves is consistent with results thus far. As with the Kratky plots in figure 6.5, protein samples in the low-middle temperature range of figure 6.8 correspond to proteins that are more globular in structure, while samples at the highest temperatures have lower peaks on their $p(r)$ curves, indicating the same protein is now less globular, i.e. unfolded. Missing from figure 6.8 are the protein samples at -7^oC and 50^oC. Using the GNOM application on the 7^oC sample produced a $p(r)$ curve that never exceeded $7x10^{-3}$ on its y-axis and thus this data was considered erroneous. The protein sample at 50^oC presented a different problem in the GNOM process. In this case, the indirect Fourier transform procedure could not adequately sample the scattering pattern without producing oscillations in the

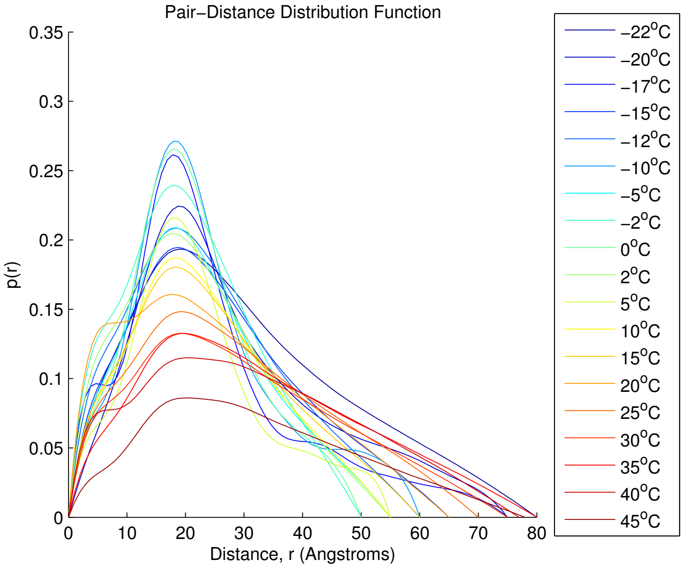


Figure 6.8: Pair-distance distribution curves for cytochrome-c in a $2.5M$ denaturant solution over a temperature range of -22^oC to 45^oC.

.

$p(r)$ curve. This is likely an example of oversampling, which occurs when the original dataset is being sampled at a frequency higher than that of the Nyquist frequency. This example of oversampling can be seen in figure 6.9. Despite numerous attempts to modify the sampling rate in the indirect Fourier transform using GNOM, the $2.5M$ 50^oC data consistently resulted in $p(r)$ curves with high amounts of oscillation, i.e. an ideal sampling frequency could not be obtained. Data such as this was not considered for further analysis.

As described in Chapter 3, the $p(r)$ function can be used to calculate a radius of gyration for the protein by integrating over the entire r-range (see equation 3.42). Applying equation 3.42 to the entire set of pair-distance distribution functions, i.e. for each molarity and at each temperature, gives the results seen in figure 6.10.

There are several notable aspects to the values in figure 6.10. Although the -7^oC and 50^oC data from the $2.5M$ dataset did not produce an acceptable $p(r)$ curve,

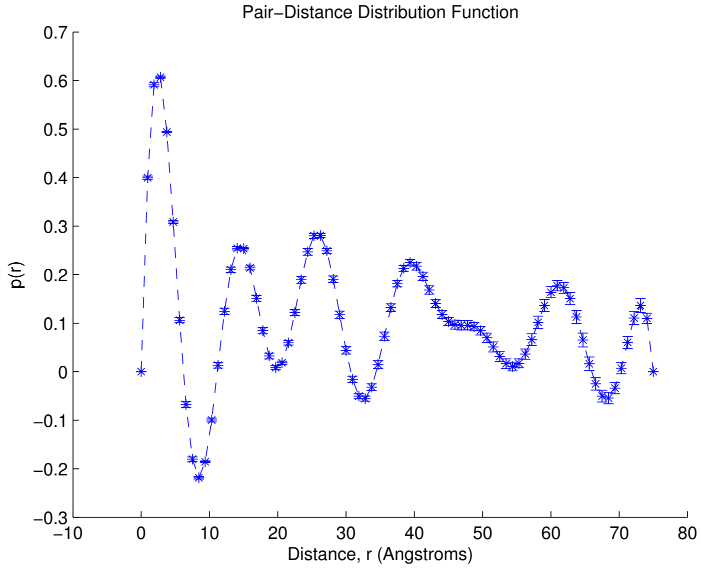


Figure 6.9: Pair-distance distribution curve for cytochrome-c in a $2.5M$ denaturant solution at a temperature of 50^oC. This is an example of oversampling in the indirect Fourier transform.

.

GNOM still gives values for the radius of gyration in these cases, and these values are consistent with those of Guinier analysis. Also notable from figure 6.10 is the lower error, particularly at for $4M$ calculations, when compared to radii of gyration calculated from Guinier analysis. A comparison of values from Guinier and $p(r)$ analysis may be seen in figure 6.11. Errors have been omitted from figure 6.11 for the sake of presentation. A comparison of errors from both methods of calculation can be seen in figure 6.12, each presented as a percentage of the respective calculated radius of gyration.

As can be seen from figure 6.11, results from Guinier and $p(r)$ analysis are quite consistent with each other at $0M$ denaturant concentration. At the $2.5M$ concentration, the $p(r)$ analysis gives a lower value for radius of gyration for all temperatures except the -7^oC data. It should be noted that this was a dataset which previ-

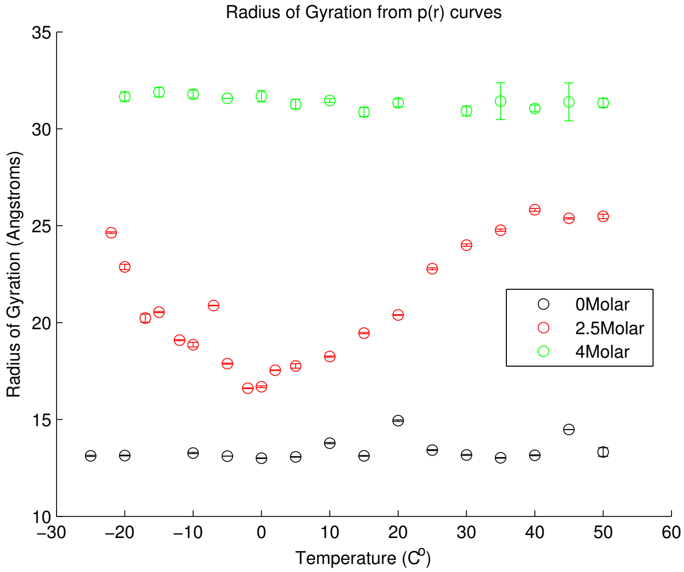


Figure 6.10: Radius of Gyration from $p(r)$ analysis for cytochrome-c at $0M$, $2.5M$, and $4M$ denaturant concentrations over a temperature range of -22^oC to 50^oC.

.

ously gave an unacceptable $p(r)$ curve. Interestingly, the converse is true for the $4M$ concentration, where the $p(r)$ analysis gives consistently higher values. At this concentration, the results from $p(r)$ analysis are more consistent with results from wide angle x-ray scattering (WAXS) experiments (Pollack, 1999), than with previous results from SAXS experiments (Hsu, 2007). Figure 6.12 also reveals some interesting insights. For the majority of radius of gyration calculations, the percent error is low ($< 1\%$). Highest errors occur most notably for the $4M$ Guinier results, while the $p(r)$ results show significantly less error at this concentration. Only three calculations from the $p(r)$ analysis show a percent error greater than 1%; 50^oC at $0M$, and 35^oC and 40^oC at $4M$. Error is also low across the majority of results from the $2.5M$ data from both methods of calculation.

With acceptable results obtained for the radius of gyration of cytochrome-c at various denaturant conditions, reconstructions based on this data are now presented.

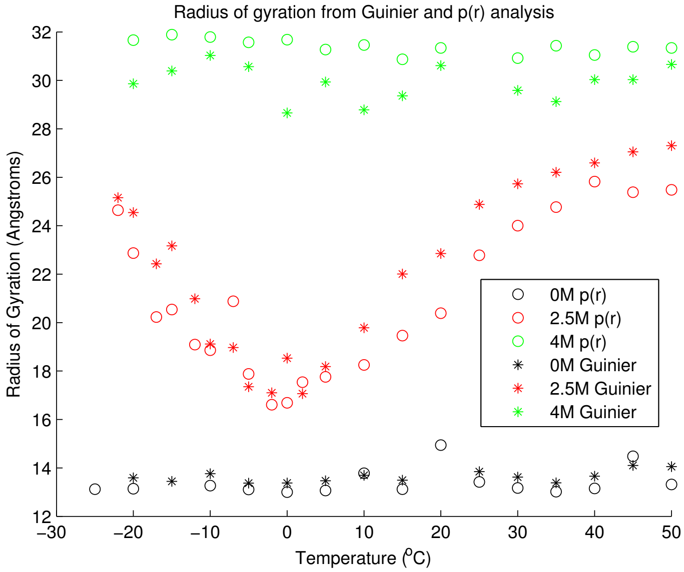


Figure 6.11: Radius of Gyration from Guinier and $p(r)$ analysis for Cytochrome-C at $0M$, $2.5M$, and $4M$ denaturant concentrations over a temperature range of -22^oC to 50^oC. Error bars ommited for clarity (see previous plots).

.

6.3 Protein Reconstruction

All reconstructions presented in this section are produced according to the example given in Chapter 5. The bulk of the reconstructions are from the $2.5M$ dataset, since at this denaturant concentration change in the protein's conformation is observed. Reconstructions of this dataset will include a reconstruction from each of the $0M$ and $4M$ datasets, thereby providing a full spectrum of cytochrome-c as it changes from native to denatured state.

DAMMIN produces a value for the radius of gyration from its reconstruction by calculating a unique scattering pattern, finding the $p(r)$ function from equation 3.40, and calculating R_g via equation 3.42. Results of this calculation for the $2.5M$ dataset corresponding to the images in figures 6.13 to 6.18 is shown in figure 6.19 along with results for the radius of gyration from Guinier and $p(r)$ analysis.

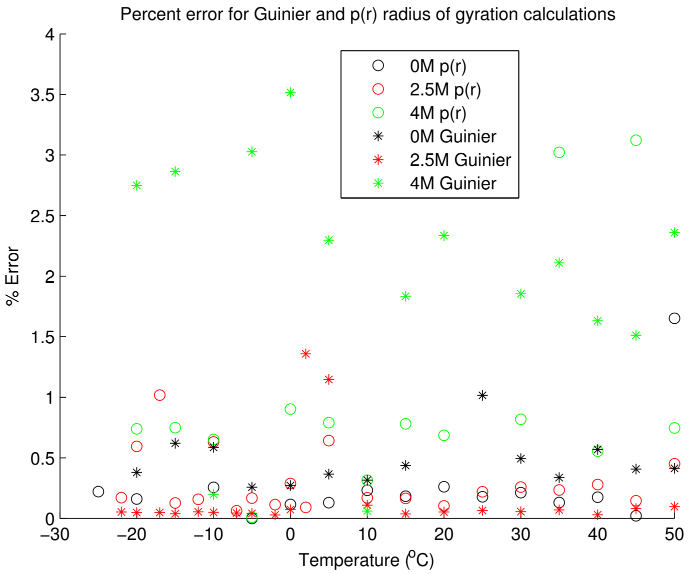


Figure 6.12: Percent errors from radius of gyration calculation from Guinier and $p(r)$ analysis for cytochrome-c at $0M$, $2.5M$, and $4M$ denaturant concentrations over a temperature range of -22^oC to 50^oC.

.

Figures 6.13 to 6.18 provide visual evidence of the thermodynamically and chemically induced folding/unfolding of equine cytochrome-c. For the $2.5M$ reconstructions, the proteins are clearly more elongated and spread out at the extreme high and low temperatures. The notable outlier is again the -7^oC trial, which did not produce an acceptable $p(r)$ curve previously. The reconstruction of this data (figure 6.15, protein 3a) does not fit visually with surrounding reconstructions. As with the calculation of R_g from $p(r)$ analysis of this dataset, the R_g reported by DAMMIN nevertheless meets expectations, as seen in figure 6.19. In figure 6.19 we see that the R_g values from DAMMIN are reasonably similar to those resulting from $p(r)$ analysis, but differ slightly with results from Guinier analysis.

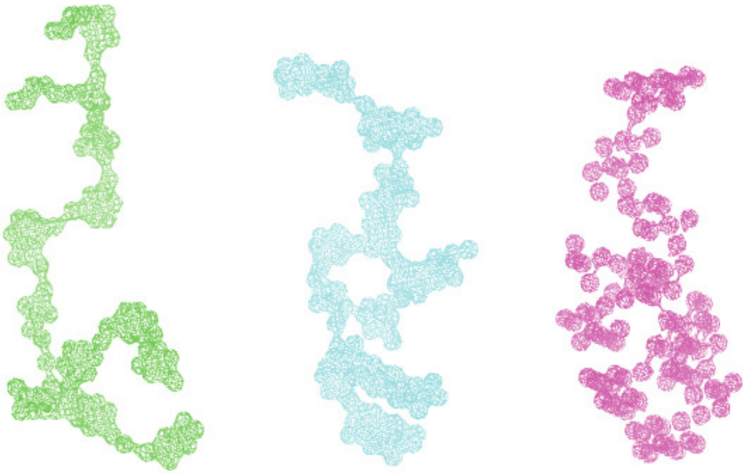


Figure 6.13: From left to right: Cytochrome-c in 1a: $4M$ denaturant concentration at 0^oC, 1b: $2.5M$ denaturant concentration at -22^oC, 1c: $2.5M$ denaturant concentration at -20^oC. For scale, the maximum diameter of protein 1a is 85.94Å.

.

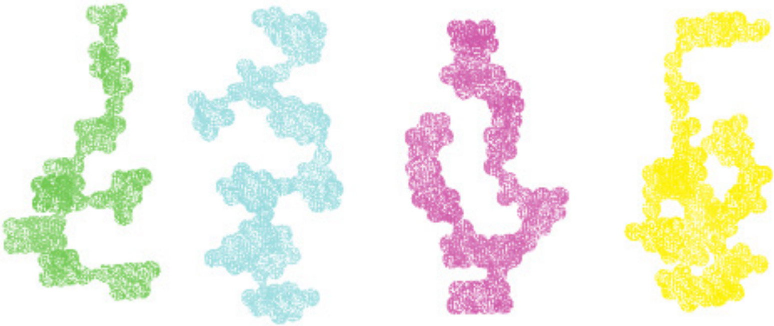


Figure 6.14: From left to right: Cytochrome-c in $2.5M$ denaturant concentration at 2a: -17^oC, 2b: -15^oC, 2c: -12^oC, 2d: -10^oC. For scale, the maximum diameter of protein 2a is 56.53Å.

.

6.4 Reconstruction Quality

Figure 6.19 provides us with some evidence of successful reconstructions based on the similarities from separately calculated values for the radii of gyration. This alone is not sufficient to say that these reconstructions are correct, or useful for

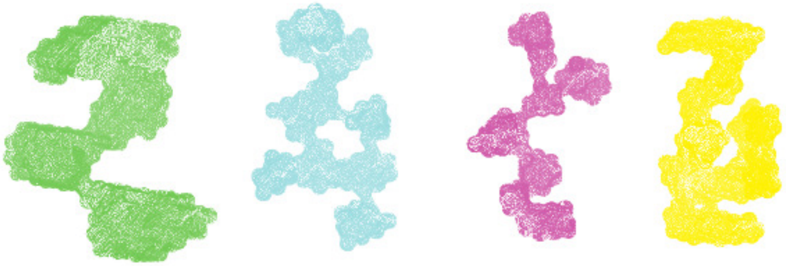


Figure 6.15: From left to right: Cytochrome-c in $2.5M$ denaturant concentration at 3a: -7^oC, 3b: -5^oC, 3c: -2^oC, 3d: 0^oC. For scale, the maximum diameter of protein 3a is 54.55Å.

.

Figure 6.16: From left to right: Cytochrome-c in 4a: $0M$ denaturant concentration at 0^oC, 4b: $2.5M$ denaturant concentration at 2^oC, 4c: $2.5M$ denaturant concentration at 5^oC, 4d: $2.5M$ denaturant concentration at 10^oC. For scale, the maximum diameter of protein 4a is 34.69Å.

.

that matter, since two different scattering curves may produce a similar R_g. Since the scattering pattern from a reconstruction is unique, it is then compared to the experimental scattering pattern at the corresponding conditions. Agreement of the scattering patterns indicates an acceptable reconstruction. Using the process outlined in Chapter 5, the program FoXS is used to calculate the scattering pattern from each reconstruction and compare it to its experimental partner. A chi-function

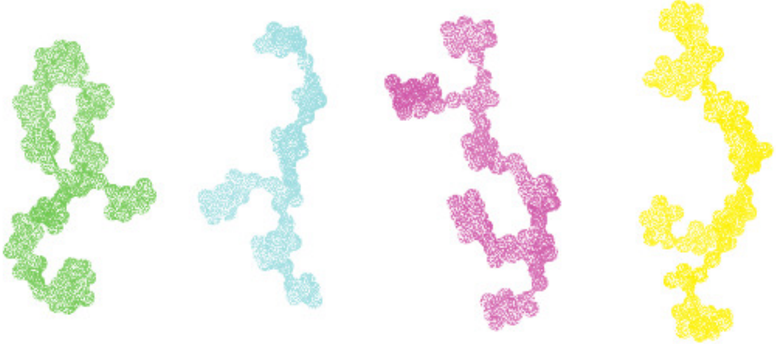


Figure 6.17: From left to right: Cytochrome-c in 2.5M denaturant concentration at 5a: 15^oC, 5b: 20^oC, 5c: 25^oC, 5d: 30^oC. For scale, the maximum diameter of protein 5a is 58.70Å.

.

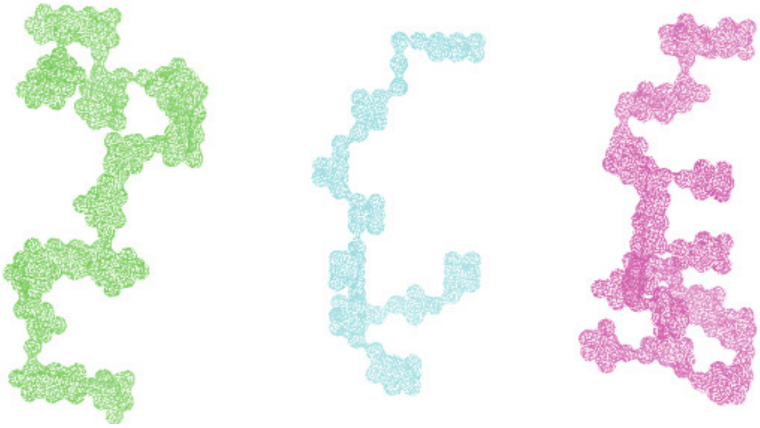


Figure 6.18: From left to right: Cytochrome-c in 2.5M denaturant concentration at 6a: 35^oC, 6b: 40^oC, 6c: 45^oC. For scale, the maximum diameter of protein 6a is 81.38Å.

.

is used by FoXS to quantify agreement between the two curves, by the formula

$$\chi^2 = \sum_{k}^{Q} \left(\frac{I_{exp}(Q_k) - cI_{DAM}(Q_k)}{\sigma_{exp}(Q_k)} \right)^2 \tag{6.1}$$

where I_{exp} are the experimental data points, I_{DAM} are the corresponding points on

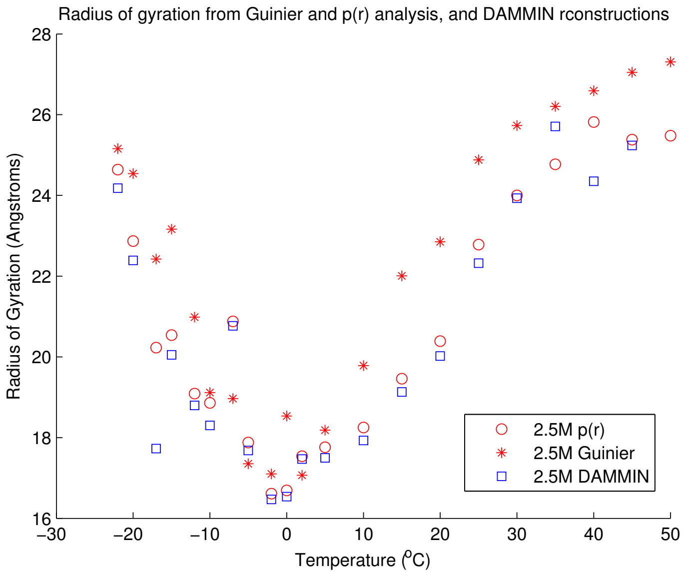


Figure 6.19: Radius of gyration from Guinier, $p(r)$, and DAMMIN analysis for cytochrome-c in a $2.5M$ denaturant concentration over a temperature range of -22^oC to 50^oC.

.

the DAMMIN reconstructed scattering curve, c is a scaling parameter, and σ_{exp} are the experimental errors.

In surveying figures 6.20 to 6.25, the notable outlier is again the $2.5M$ -7^oC data. The disparity between the reconstructed scattering intensity and the experimental data for this temperature is obvious in figure 6.22, subplot 3a, and is made quantifiably clear by its disagreement of $\chi = 21.5$. All other DAMMIN-reconstructed intensities show varying levels of agreement with their experimental counterparts, ranging from $\chi = 1.02$ to $\chi = 5.31$. It is evident from the figures that most disagreement arises from either the high-Q or low-Q regions, or both. Disagreement at low-Q is reflected in figure 6.19 as the difference between R_g calculated from Guinier analysis and from DAMMIN reconstructions, since Guinier analysis uses only this region to determine R_g. Disagreement at high-Q is to be somewhat expected since

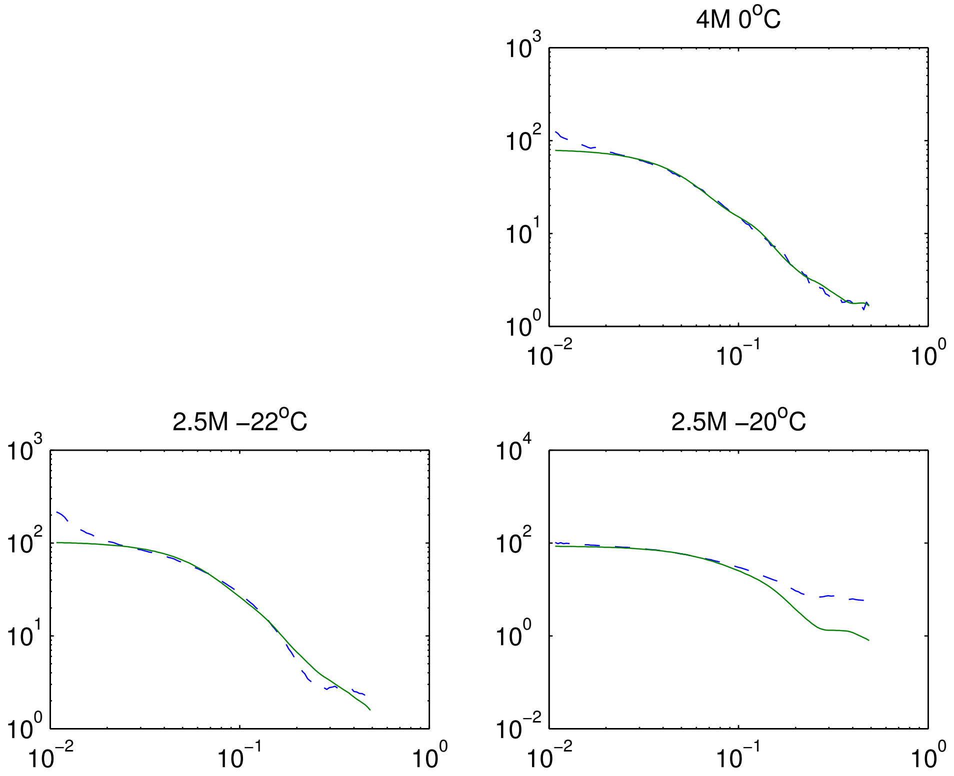


Figure 6.20: Scattering curves from SAXS experimentation and DAMMIN reconstructions for cytochrome-c in 1a: $4M$ denaturant concentration at 0^oC, 1b: $2.5M$ denaturant concentration at -22^oC, 1c: $2.5M$ denaturant concentration at -20^oC. $\chi = 1.24$, 5.31, 1.62 respectively.

.

resolution at such small angles can be difficult to achieve using the SAXS method.

6.5 Cold and Warm Denaturation

Observed in the previous section were both cold and warm denatured states of cytochrome-c. At both extremes, i.e. -22^oC and 45^oC, the proteins are the same size; we know this not only from the maximum diameter of each, but also from their similar radii of gyration (24.64Åand 25.48Årespectively, from $p(r)$ analysis).

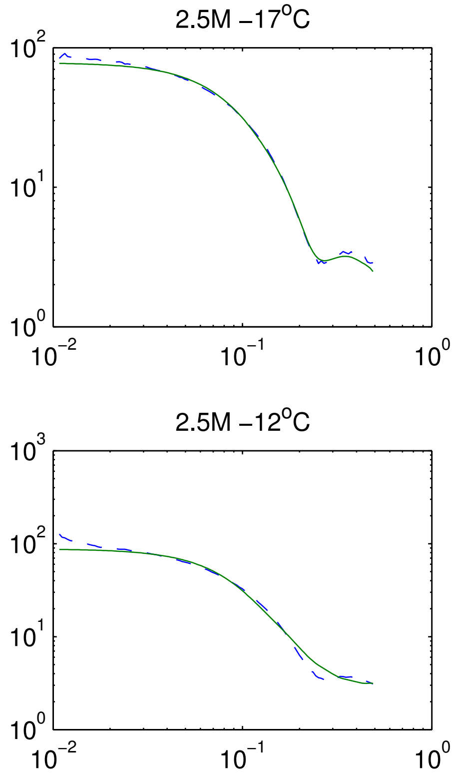


Figure 6.21: Scattering curves from SAXS experimentation and DAMMIN reconstructions for cytochrome-c in a $2.5M$ denaturant concentration at 2a: $-17^{o}C$, 2b: $-15^{o}C$, 2c: $-12^{o}C$, 2d: $-10^{o}C$. $\chi = 1.63$, 3.89, 3.67, 1.56 respectively.

.

However, it is important to note that the driving force behind denaturation in each case is not the same. As discussed in Chapter 2, at low temperature enthalpy dominates the process of denaturation, whereas at high temperature the same is true of entropy. Since entropy is the driving force at warm temperature, one would expect more possible configurations of the warm denatured protein. In the case enthalpy-driven cold denaturation there are less possible configurations the protein may take. Furthermore, the hydrophobic effect is not present in the case of cold denaturation and thus the protein backbone is prone to stiffening, thereby imposing a lower limit

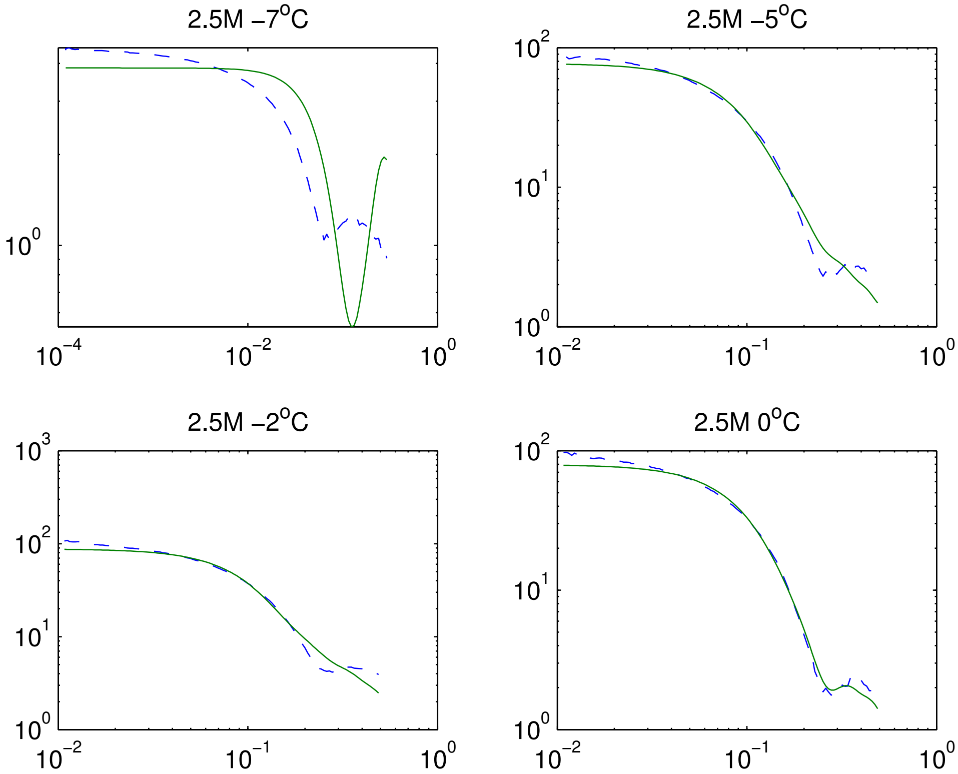


Figure 6.22: Scattering curves from SAXS experimentation and DAMMIN reconstructions for cytochrome-c in a $2.5M$ denaturant concentration at 3a: -7^oC, 3b: -5^oC, 3c: -2^oC, 3d: 0^oC. $\chi = 21.5, 3.93, 5.15, 2.72$ respectively.

.

on the total possible configurations the protein may assume in this case. This difference in driving force can be seen in figure 6.26 from the two state protein stability theory outlined in Chapter 2. This is a plot of the various components of the total neutral folding free energy (equation 2.20). One can see the decrease in enthalpy and increase in entropy (solvation entropy) as temperature increases. The amount of possible configurations the protein may assume in each denatured state is also reflected by the persistence length. The persistence length may be thought of as the minimum distance the protein chain must travel in a straight line due to limitations

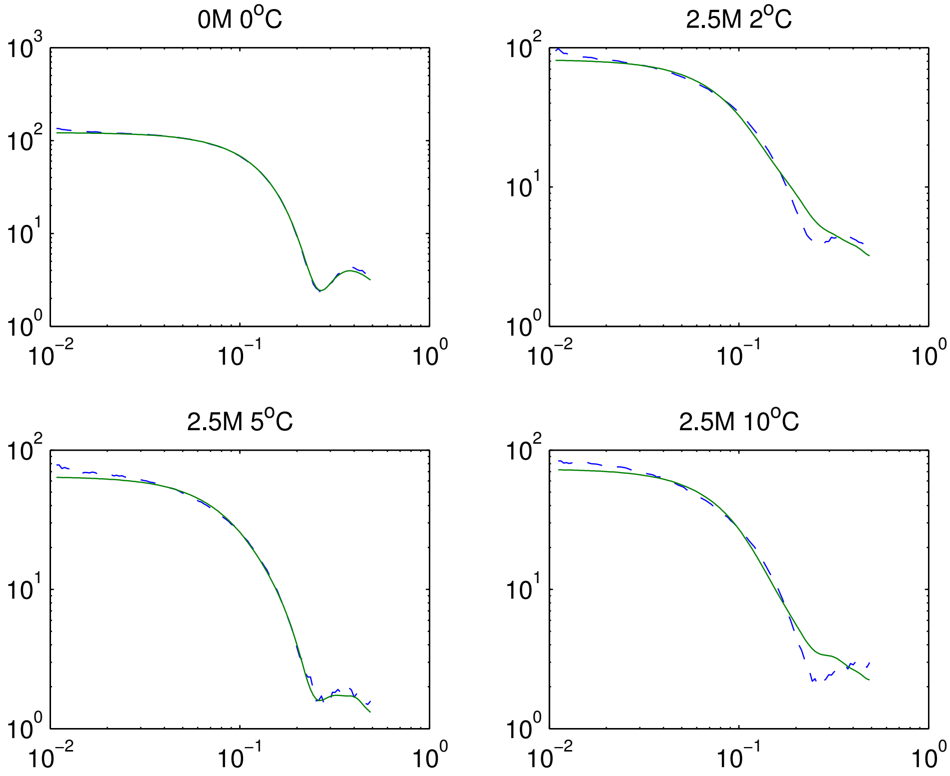


Figure 6.23: Scattering curves from SAXS experimentation and DAMMIN reconstructions for cytochrome-c in 4a: $0M$ denaturant concentration at 0^oC, 4b: $2.5M$ denaturant concentration at 2^oC, 4c: $2.5M$ denaturant concentration at 5^oC, 4d: $2.5M$ denaturant concentration at 10^oC. $\chi = 1.37$, 4.96, 1.49, 3.27 respectively.

.

on its flexibility. One would expect that if more configurations of the protein are possible in the case of entropy-driven warm denaturation, then the protein should also exhibit a shorter persistence length to allow for such configurations. These expectations are also reflected in the two-state stability theory and can be seen as the "chain entropy" component of the neutral folding free energies in figure 6.26, i.e. with more chain entropy free energy, more configurations are possible. As mentioned in Chapter 3, the high-Q region of the Kratky plot is characterized by the

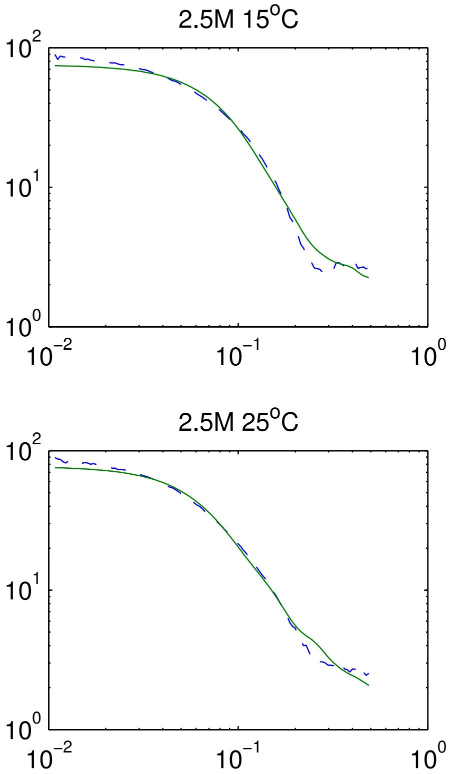


Figure 6.24: Scattering curves from SAXS experimentation and DAMMIN reconstructions for cytochrome-c in a $2.5M$ denaturant concentration at 5a: $15^{o}C$, 5b: $20^{o}C$, 5c: $25^{o}C$, 5d: $30^{o}C$. $\chi = 3.42$, 5.19, 2.17, 1.41 respectively.

.

persistence length of the protein. Using Porod's law, the persistence length may be obatined by performing a linear fit to this high-Q region (see figure 6.27). Results for the persistence length of our protein in $2.5M$ solution at each temperature can be seen in figure 6.28. In this figure, the persistence length is represented as persistence Q($\mathrm{\AA}^{-1}$); its tendency to increase with increasing temperature reflects a decrease in persistence length with increasing temperature. Lastly, this effect may be seen visually by comparing side by side the DAMMIN reconstructions of both cold and warm denatured ctyochrome-c (figure 6.29). Here one can see the longer

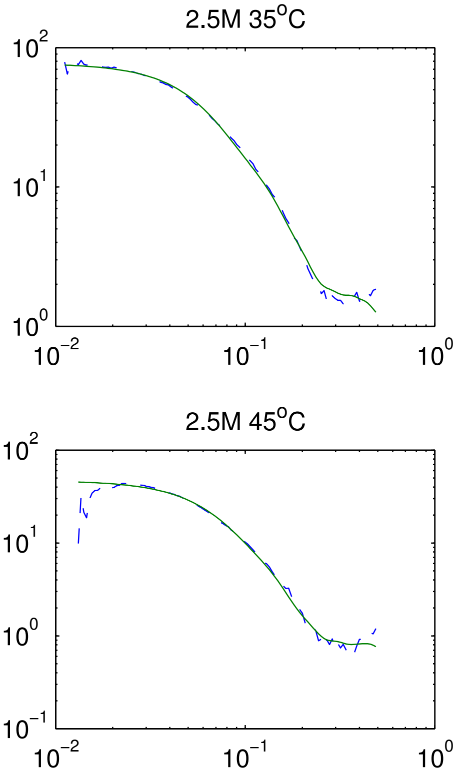


Figure 6.25: Scattering curves from SAXS experimentation and DAMMIN reconstructions for cytochrome-c in a $2.5M$ denaturant concentration at 6a: 35^oC, 6b: 40^oC, 6c: 45^oC. $\chi = 1.44$, 2.89, 1.6 respectively.

.

persistence length in the cold denatured protein on the left hand side. As a crude form of measurement, one can count the amount of straight line sections and find that there are a lesser amount in the cold denatured protein. One can also use in PyMOL the measurement, wizard which allows measurement of distances between two atomic locations, to verify this.

$$\Delta F_{neutral}(T) = N\left[g + m_1 c + RT\ln(z) + \Delta C_p\left(T - T_h\right) - T\Delta C_p \ln\left(\frac{T}{T_s}\right)\right]$$

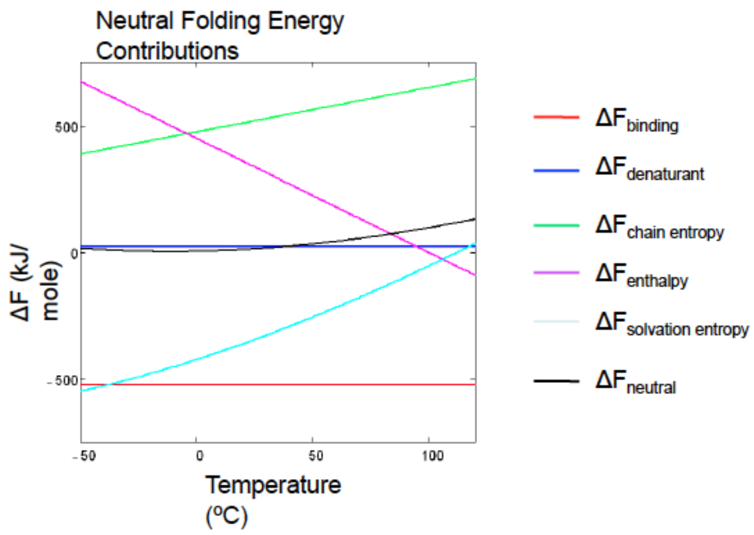


Figure 6.26: Individual components of the total neutral folding free energy. From the two state protein stability theory described in Chapter 2 (Landahl et al. 2013).

.

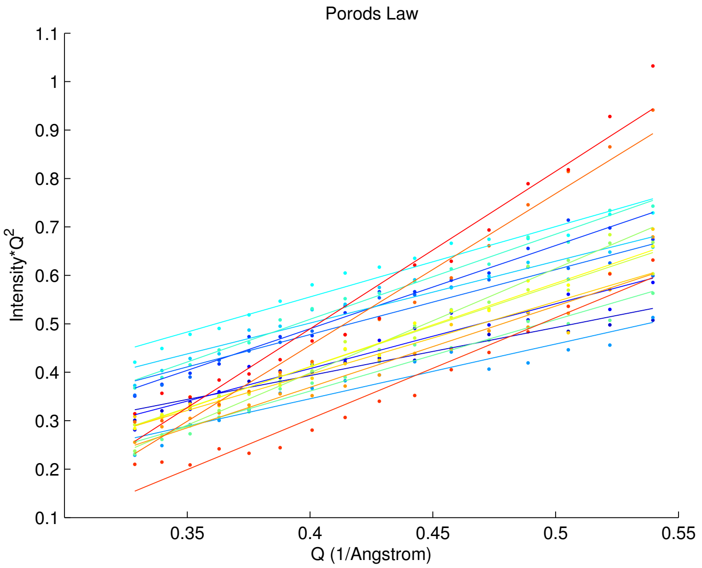


Figure 6.27: Kratky plot demonstrating Porod's law for cytochrome-c in a $2.5M$ GuHCl solution over a temperature range of -22^oC (dark blue) to 45^oC (dark red).

.

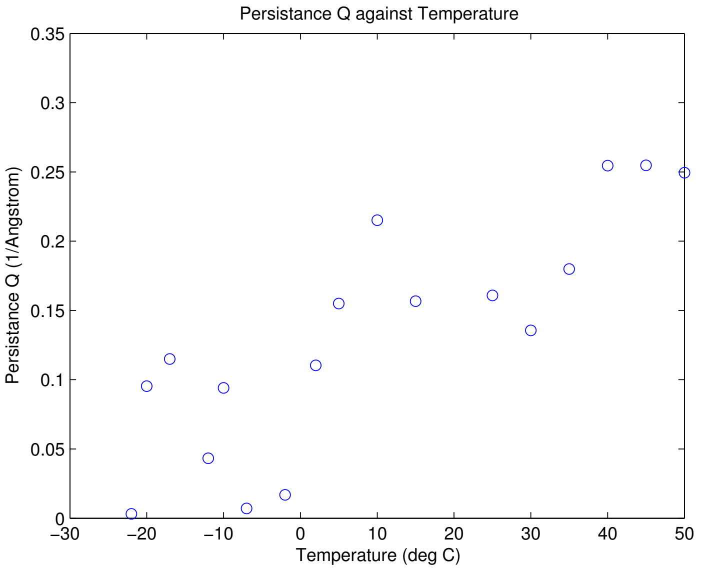


Figure 6.28: Persistence Q(Å^{-1}) for cytochrome-c in a $2.5M$ GuHCl solution at various temperatures (Landahl et al. 2013)

.

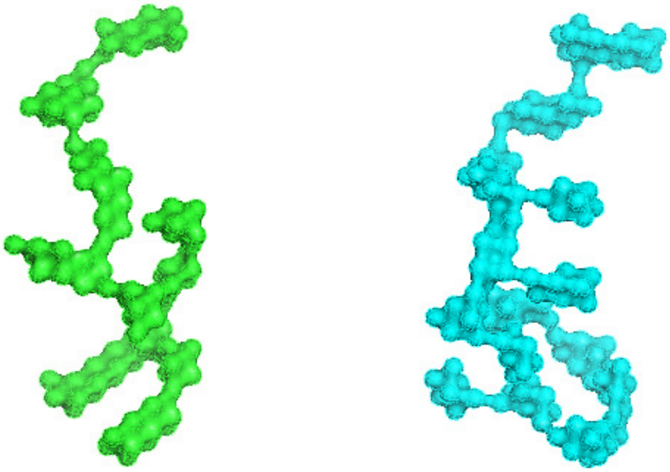


Figure 6.29: Left hand side: Cold denatured cytochrome-c at -22^oC. Right hand side: Warm denatured cytochrome-c at 45^oC. Both in a $2.5M$ GuHCl solution.

.

CHAPTER 7

Mitochondrial Rho Results

In this chapter we present the results for SAXS experiments on the mitochondrial rho (miro) protein. This chapter is divided into two parts, for the short (miro-S) and long (miro-L) versions of miro. Analysis is carried out on the raw data in the same manner as that of cytochrome-c, with some minor exceptions. Hence, results are presented in a similar manner to Chapter 6. Miro was studied under a variety of chemical conditions, and since this is the first experiment of its kind to be performed on this protein, results were not obtained as easily as with cytochrome-c. For both miro-S and miro-L we present here a single dataset and reconstruction. These two were part of the select few datasets that resulted in satisfactory pair-distance distribution functions and DAMMIN reconstructions. It is known from amino acid sequence analysis that miro consists of two GTPase domains and an ef hand motif. Unlike cytochrome-c, we do not have a crystal structure with which to compare our experimental reconstructions. In order to check that our reconstructions are correct we construct a somewhat crude estimation of miro using two different proteins for which we have accepted crystal structures. In place of the GTPase domain(s) of our miro, we use human rac1, a GTP binding molecule (PDB ID: 1MH1). Taking the place of the ef hand in our miro is one half of the crystal structure of human centrin 2 (HsCen2), an ef hand calcium binding protein (PDB ID: 2OBH). This protein consists of 2 ef hand domains which are easily separated using PyMOL. Apart from using 1MH1 and 2OBH to check the quality of our reconstructions, we also compare the amino acid sequence of these proteins to the amino acid seqeunce for miro-S and miro-L, provided by the Rice Lab. By doing this we can ensure that any structural inconsistencies evident in our reconstructions are reflected in the differences between the amino acid sequences. Before presenting results for miro-S and miro-L we first show the structure of 1MH1, 2OBH, and the combinations of both which will be used for comaparison to our results. Also presented prior to experimental results are the amino acid sequence comparisons.

Shown in figure 7.1 and figure 7.2 are the crystal structures of 1MH1 (Hirshberg, 1997) and 2OBH (Charbonnier, 2007) respectively.

Figure 7.1: Crystal structure from x-ray diffraction for the human rac1 protein (PDB ID: 1MH1). For the purpose of scale, this structure has a maximum diameter of ~ 55Å.

.

In figure 7.2, one can see two of the same structure, each defined by a long helix with several small helices at each end. These are the two ef hand domains. By removing one of these domains in PyMOL and importing the other into our 1MH1 figure we make a mock construction of miro-S, seen in figure 7.3. In order to construct a mock

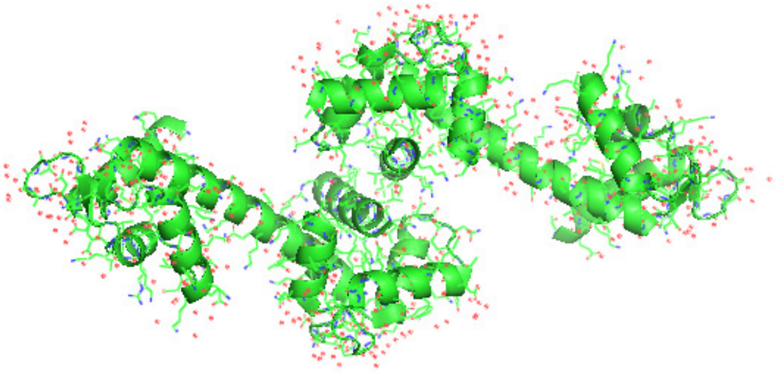


Figure 7.2: Crystal structure from x-ray diffraction for the human centrin 2 protein (PDB ID: 2OBH). For the purpose of scale, this protein has a maximum diameter of ~ 115Å.

.

miro-L we make a copy of 1MH1 and import it into our mock miro-S structure. This can be seen in figure 7.4.

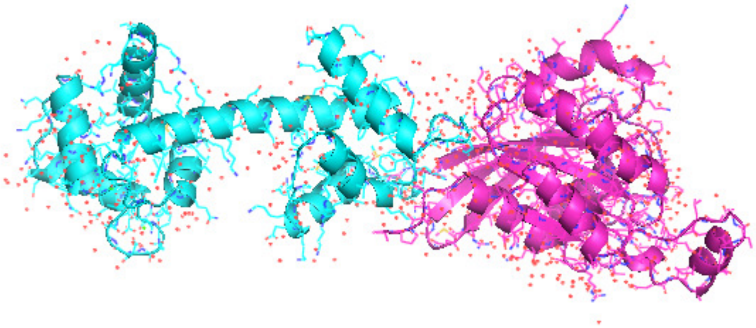


Figure 7.3: Mock miro-S structure built from crystal structures of the 1MH1 protein and an ef hand domain from the 2OBH protein.

.

When reconstructions for miro-S and miro-L are obtained they are "docked", i.e. imported and overlayed with the crystal structures in figures 7.3 and 7.4 respectively. In building these mock structures it should be noted that the alignment of 1MH1

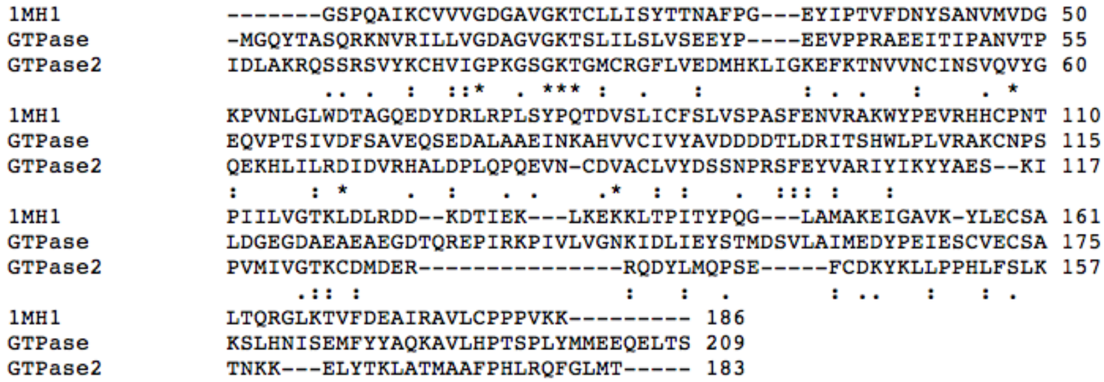
Figure 7.4: Mock miro-L structure built from crystal structures of two 1MH1 proteins and an ef hand domain from the 2OBH protein.

.

with respect to the ef hand domain of 2OBH is arbitrary, since in solution the ef hand and GTPase domain(s) of miro are not joined together in a rigid manner. As such, the 1MH1 structure(s) may be rotated and realigned in PyMOL when docking the miro reconstructions.

To compare the amino acid sequence of miro with those of 1MH1 and 2OBH, the sequences are separated. In figure 7.5, the amino acid sequence for 1MH1 is aligned with the amino acid sequences for both GTPase domains in miro-L. In figure 7.6 the amino acid sequence for one ef hand domain of 2OBH is aligned with the amino acid sequence of the miro ef hand domain. These sequences and alignments are courtesy of the Rice Lab at Northwestern University.

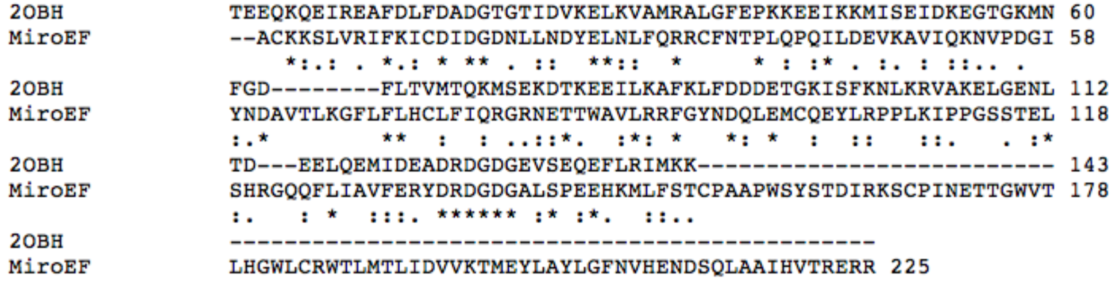
Figures 7.5 and 7.6 list the amino acid sequence for the relevant structures with a running total count of amino acids at the end of each line. While not of concern in this experiment it should be noted that the dots and asterisks included in these figures are representative of various levels of agreement between the compared amino acid sequences. The most notable aspect of these figures is the variation in sequence length between 1MH1 and GTPase/GTPase2 and particularly between 2OBH and MiroEF. The difference between the sequence length of 2OBH and the miro ef hand



```
1MH1            -------GSPQAIKCVVVGDGAVGKTCLLISYTTNAFPG---EYIPTVFDNYSANVMVDG 50
GTPase          -MGQYTASQRKNVRILLVGDAGVGKTSLILSLVSEEYP----EEVPPRAEEITIPANVTP 55
GTPase2         IDLAKRQSSRSVYKCHVIGPKGSGKTGMCRGFLVEDMHKLIGKEFKTNVVNCINSVQVYG 60
                       .. .  :  ::*  . *** :  .   :       : . .   :    . *
1MH1            KPVNLGLWDTAGQEDYDRLRPLSYPQTDVSLICFSLVSPASFENVRAKWYPEVRHHCPNT 110
GTPase          EQVPTSIVDFSAVEQSEDALAAEINKAHVVCIVYAVDDDDTLDRITSHWLPLVRAKCNPS 115
GTPase2         QEKHLILRDIDVRHALDPLQPQEVN-CDVACLVYDSSNPRSFEYVARIYIKYYAES--KI 117
                :     : *    .  :   . .    .*  : :   .  ::: :   :
1MH1            PIILVGTKLDLRDD--KDTIEK---LKEKKLTPITYPQG---LAMAKEIGAVK-YLECSA 161
GTPase          LDGEGDAEAEAEGDTQREPIRKPIVLVGNKIDLIEYSTMDSVLAIMEDYPEIESCVECSA 175
GTPase2         PVMIVGTKCDMDER---------------RQDYLMQPSE-----FCDKYKLLPPHLFSLK 157
                     .:: :                   :   :  .       : ..   :   : .
1MH1            LTQRGLKTVFDEAIRAVLCPPPVKK--------- 186
GTPase          KSLHNISEMFYYAQKAVLHPTSPLYMMEEQELTS 209
GTPase2         TNKK---ELYTKLATMAAFPHLRQFGLMT----- 183
```

Figure 7.5: Amino acid sequence alignment for 1MH1 and both GTPase domains of miro-L.

.



```
2OBH            TEEQKQEIREAFDLFDADGTGTIDVKELKVAMRALGFEPKKEEIKKMISEIDKEGTGKMN 60
MiroEF          --ACKKSLVRIFKICDIDGDNLLNDYELNLFQRRCFNTPLQPQILDEVKAVIQKNVPDGI 58
                    *:.: . *.: * ** . ::  **::  *     * : :* . :. : ::.. .
2OBH            FGD--------FLTVMTQKMSEKDTKEEILKAFKLFDDDETGKISFKNLKRVAKELGENL 112
MiroEF          YNDAVTLKGFLFLHCLFIQRGRNETTWAVLRRFGYNDQLEMCQEYLRPPLKIPPGSSTEL 118
                :.*        **  :  : ..::*.  :*: *   *: *  :  ::   ::.   . :*
2OBH            TD---EELQEMIDEADRDGDGEVSEQEFLRIMKK-------------------------- 143
MiroEF          SHRGQQFLIAVFERYDRDGDGALSPEEHKMLFSTCPAAPWSYSTDIRKSCPINETTGWVT 178
                :.   : *  :::. ****** :* :*.  ::..
2OBH            -----------------------------------------------
MiroEF          LHGWLCRWTLMTLIDVVKTMEYLAYLGFNVHENDSQLAAIHVTRERR 225
```

Figure 7.6: Amino acid sequence alignment for 2OBH and the ef hand domain of miro.

.

domain represents an 82 amino acid linker about which nothing is currently known. Thus when we obtain reconstructions for miro-S and miro-L, we expect to see excess volume in the ef hand region when compared to our mock protein models.

7.1 Miro-S

The miro-S presented here is in a solution as described in Chapter 3, but with an additional $3mg/ml$ Ca^{2+}, since this protein was more stable under these conditions.

The protein sample is at a temperature of 4^oC. The scattering pattern for this protein can be seen in figure 7.7.

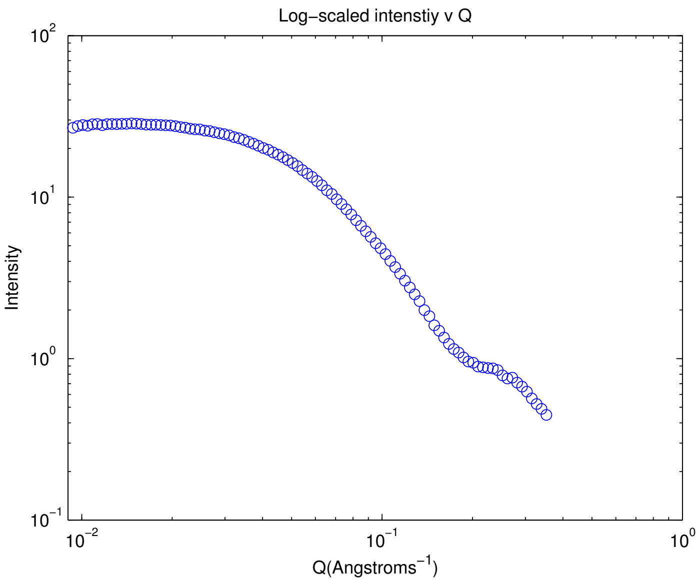


Figure 7.7: Scattering intensity curve from SAXS for miro-S in a $3mg/ml$ Ca^{2+} solution at 4^oC.

The flatness in the low-Q region of the log-log scaled intensity plot of figure 7.7 indicates some globular structure to the protein, while its steady decrease going toward high-Q is an indicator of elongated behavior for scattering at smaller angles. The size of this protein is estimated, as before, by calculating its radius of gyration from a Guinier plot of $ln(I)$ against Q^2 (see figure 7.8).

Taking the slope of the low-Q region (i.e. $Q_{max}R_g < 1.3$) gives a radius of gyration for this protein sample of 27.55±0.18Å. As in Chapter 5, we investigate the structure of the protein using a Kratky plot of Q^2I against Q (see figure 7.9).

The Kratky plot of figure 7.9 is consistent with our observations of the Guinier plot in figure 7.8, in that the initial peak followed by an increase at high-Q indicates

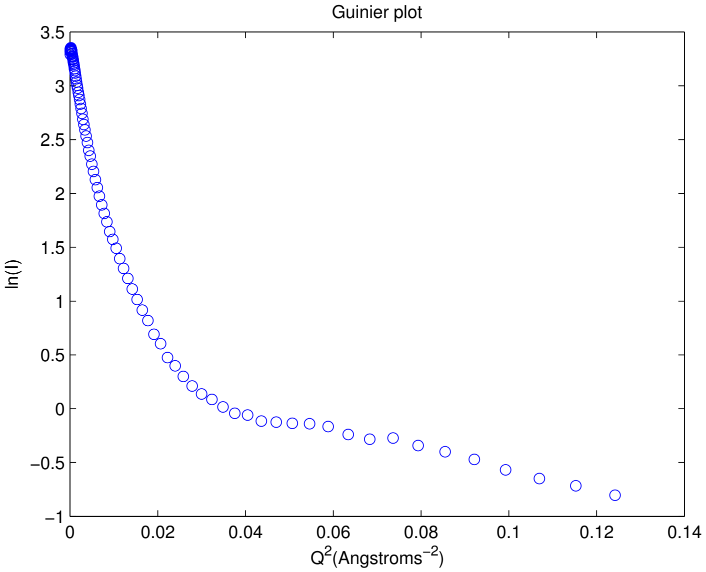


Figure 7.8: Guinier plot for miro-S in a $3mg/ml$ Ca^{2+} solution at 4^oC. $R_g = 27.55 \pm 0.18$Å.

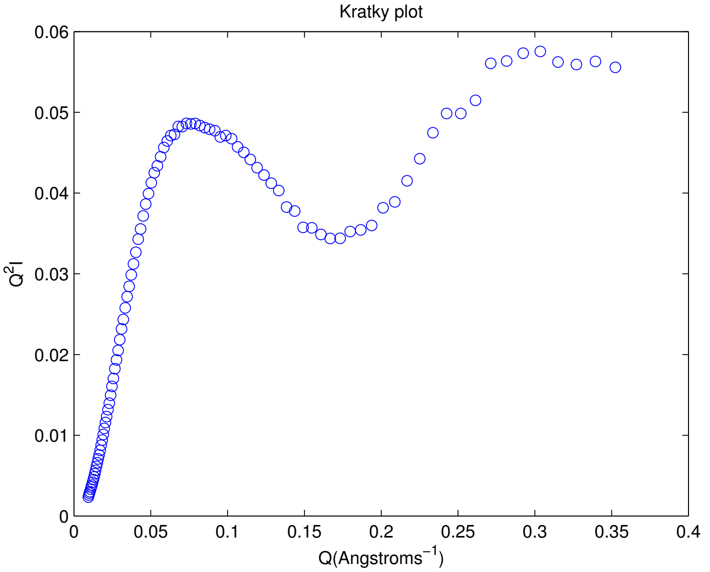


Figure 7.9: Kratky plot for miro-S in a $3mg/ml$ Ca^{2+} solution at 4^oC.

a globular overall structure with thin rod-like behavior on the smaller scale. We now evaluate for this miro-S sample the pair-distance distribution function ($p(r)$ curve) using the program GNOM. Initially, a satisfactory $p(r)$ curve could only be obtained using a maximum particle diameter of 85Å, giving an $R_g = 27.95 \pm 0.08$Å. This curve smoothly decayed to zero and had low error; however, based on our knowledge of the ef hand and GTPase domain makeup of miro-S, we expected a larger maximum diameter to be necessary for carrying out a reconstruction. Performing a reconstruction based on this $p(r)$ function revealed the problem. The DAMMIN result did not resemble a structure that contains this GTPase domain, i.e. there was no evidence of a globular structure. Instead, our reconstruction resembled something more akin to the denatured cytochrome-c reconstructions seen in Chapter 6. Numerous attempts were made at reconstructing the protein by varying parameters in DAMMIN. Eventually, we realized that the high-Q region of our data was having a dominant effect on our attempts to reconstruct. This may have been due to contaminants in the protein, a feature which would be evident at high-Q. To fix this problem, we returned to the Krakty plot of figure 7.9 and eliminated the high-Q data. Specifically, of the 98 data points in the Kratky plot, the final 18 were omitted. The Kratky plot is an appropriate point in our analysis to do this since it is very sensitive at high-Q. Since we are primarily interested in obtaining an estimate of the overall shape of the miro-S protein, this sacrifice of data was deemed a necessary step. The essential effect of this data omission is that the resulting reconstruction is now of lower resolution. In choosing an appropriate cut-off point in the Kratky plot, it was decided that the minimum of the post-peak region was suitable, since the initial peak is what describes the object shape. In short, greater detail was sacrificed for more accurate estimation of overall shape. In figure 7.10, the low-resolution Kratky plot is shown.

Using only the first 80 of the 98 points from our $I(Q)$ data, a satisfactory $p(r)$ curve was obtained, which consisted of low error, no oscillations, and slow decay to zero.

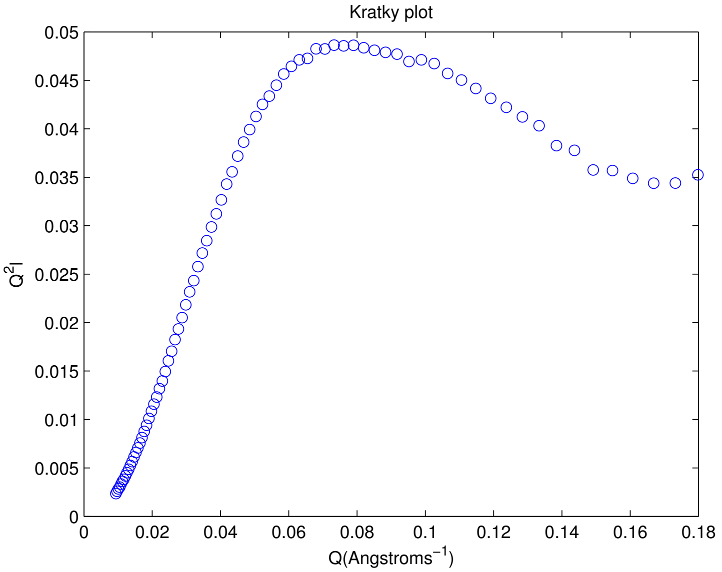


Figure 7.10: Low-resolution (using 80 of 98 data points) Kratky plot for miro-S in a $3mg/ml$ Ca^{2+} solution at 4^oC.

This was the largest number of data points that still allowed production of satisfactory curve. Evaluating the $p(r)$ function on this low-resolution data also allowed us to push the maximum diameter for our reconstruction to a more appropriate 100Å. The $p(r)$ curve for this data can be seen in figure 7.11 and has a corresponding radius of gyration of $R_g = 28.67 \pm 0.12$Å.

With a satisfactory $p(r)$ function now evaluated, a low-resolution reconstruction can now be obtained. This DAMMIN reconstruction can be seen in figure 7.12 and has a radius of gyration of 28.45Å.

In an effort to check the quality of this reconstruction, two tasks are now carried out. Firstly, the reconstruction is docked in PyMOL with the mock miro-S structure in figure 7.3. The reconstruction is expected to have a larger volume in the ef hand region than the mock structure, due to the presence of an excess 82 amino acid linker in the ef hand region of miro. This docked reconstruction, a combination of

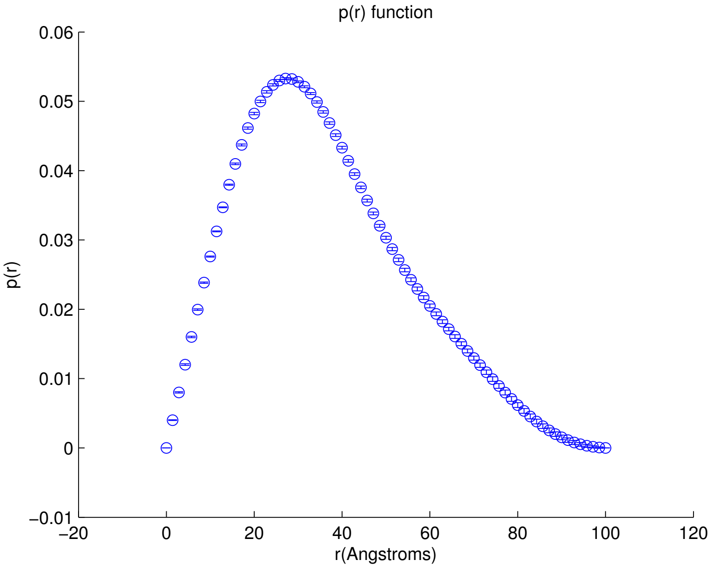


Figure 7.11: Low-resolution pair-distance distribution function for miro-S in a $3mg/ml$ Ca^{2+} solution at 4^oC. $R_g = 28.67 \pm 0.12$Å.

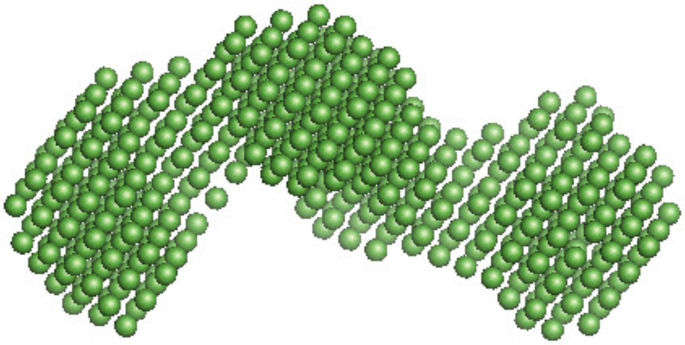


Figure 7.12: Low-resolution DAMMIN reconstruction of miro-S in a $3mg/ml$ Ca^{2+} solution at 4^oC. Radius of gyration $R_g = 28.45$Å. For the purpose of scale this reconstruction has a maximum diameter of 93.53Å.

figures 7.12 and 7.3, can be seen in figure 7.13.

Evident from figure 7.13 is the expected excess volume of the DAMMIN reconstruction in the ef hand region. Also noticeable is the slightly larger size of 1MH1

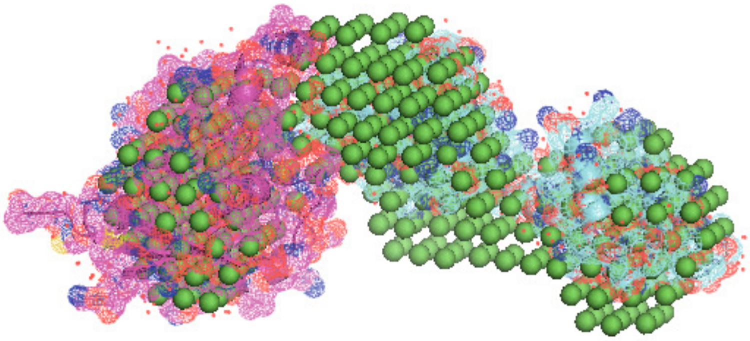


Figure 7.13: Low-resolution DAMMIN reconstruction of miro-S in a $3mg/ml$ Ca^{2+} solution at 4^oC, docked with the mock miro-S structure consisting of 1MH1 and an ef hand domain from 2OBH.

compared to the volume of this region in the DAMMIN reconstruction. This volume can be better estimated by performing an even lower resolution reconstruction; while a reconstruction using 60 of 98 data points produces a GTPase region volume large enough to fully encompass the 1MH1, significant detail and volume is lost outside of this region, and the reconstruction does not resemble at all the mock structure. Regarding figure 7.13, an estimate was made of the total excess volume of the DAMMIN structure. Using the online application VADAR (Willard, 2003), volumes were calculated for 1MH1 ($V_{1MH1} = 23,571\text{Å}^3$) and one half of 2OBH ($V_{efhands} = 21,513\text{Å}^3$). The volume of our miro-S SAXS reconstruction is $V_{miro-S} = 59,488\text{Å}^3$. The excess volume of our SAXS structure compared to the mock structure is then $V_{esaxs} = 14,403\text{Å}^3$. Knowing the volume and chain length of one half of 2OBH allows calculation of the average volume per amino acid in this structure,

$$V_{aa} = \frac{V_{efhands}}{143} = 150.4\mathring{A}^3 \tag{7.1}$$

From figure 7.6 we can then estimate the total excess volume of our SAXS structure based on the extra 82 amino acids evident from sequence analysis. This excess volume is $V_{eaa} = 82\text{x}150.4 = 12,336\text{Å}^3$. This is slightly less than V_{esaxs}, with the difference being $2,067\text{Å}^3$ or, based on V_{aa}, about 14 amino acids. Since nothing is known about the 82 amino acid linker in miro, it may be the case that these amino acids have an average volume greater than V_{aa}. Indeed, if the excess volume is involved in Ca^{2+}-binding, the amino acids here may be larger, since 4 of the 5 and 7 of the 10 strongest Ca^{2+}-binding amino acids have a volume greater than V_{aa} (Ho, 2007). Most importantly however, this excess volume is certainly enough to hold an additional 82 amino acids.

Next, the quality of this reconstruction is checked using the online application FoXS, by comparing the unique scattering pattern from the reconstruction to the scattering pattern obtained via SAXS. The FoXS scattering comparison can be seen in figure 7.14.

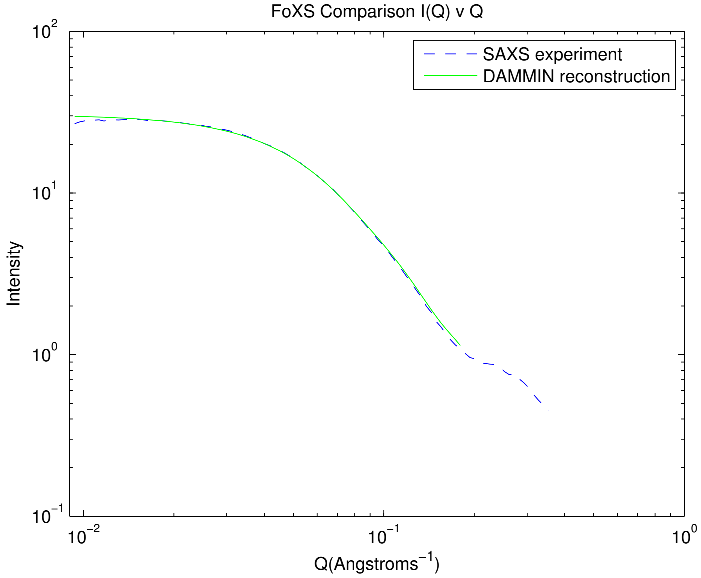


Figure 7.14: FoXS comparison of log-scaled scattering curves from low-resolution DAMMIN reconstruction and SAXS experiment for miro-S in a $3mg/ml$ Ca^{2+} solution at 4^oC. $\chi = 0.95$.

The FoXS comparison in figure 7.14 reports an excellent agreement between the scattering pattern from the low-resolution DAMMIN reconstruction and the experimental data from SAXS. This, along with visual inspection of figure 7.13, is an indicator of an acceptable reconstruction.

7.2 Miro-L

Similarly to miro-S, miro-L is examined in a solution as described in Chapter 4, but with an additional $30mg/ml$ of Ca^{2+}. The scattering pattern from SAXS experimentation on this protein sample is shown in figure 7.15.

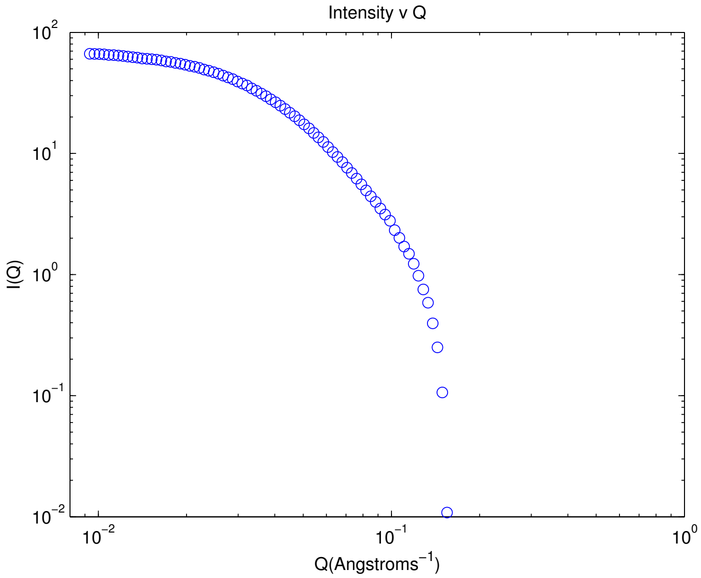


Figure 7.15: Scattering intensity curve from SAXS for miro-L in a $30mg/ml$ Ca^{2+} solution at $4^{o}C$.

The size of this protein is estimated as usual by calculating the slope of the low-Q region of a Guinier plot (see figure 7.16).

The data in figure 7.16 gives a radius of gyration for this protein of $R_g = 44.3 \pm 0.3$Å.

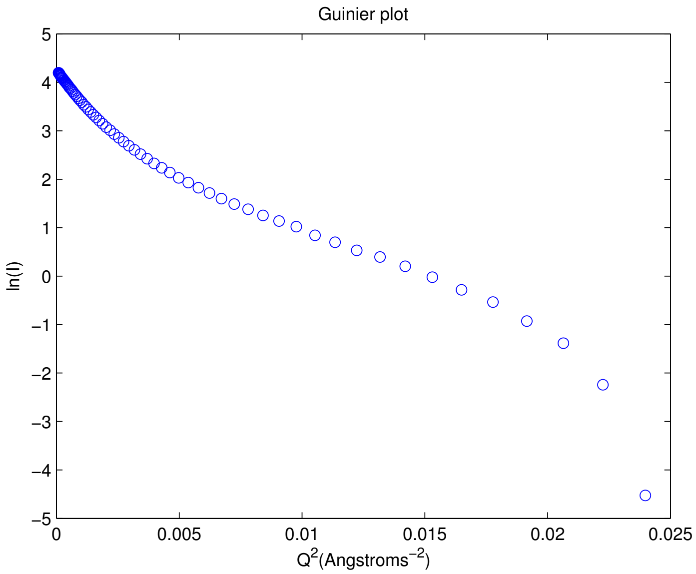


Figure 7.16: Guinier plot for miro-L in a $30mg/ml$ Ca^{2+} solution at 4^oC. $R_g = 44.3 \pm 0.3$Å.

The overall structure of this miro-L sample is again investigated using a Kratky plot, seen in figure 7.17.

When comparing the Kratky plots for miro-L and miro-S, the most obvious difference is the lack of high-Q data for miro-L. This serves as evidence that the miro-S sample may have been contaminated, although several trials of miro-L also had Kratky plots dominated by high-Q data. Another possible reason for the lack of high-Q data is that due to the large size of the miro-L protein, our technique is simply not capable of resolving at such a small length scale. Again, our primary goal in this experiment is an estimation of overall shape, and so a lack of detail at high-Q is not a major concern. More significantly, a clear and clean peak is seen on the Kratky plot, indicating a large, globular structure. Since the high-Q data is not dominating in this case, it is possible to perform a reconstruction using the full dataset, and not have to reduce the resolution any further. The size of this protein is further clarified by evaluating as usual its $p(r)$ function. This can be seen in figure 7.18.

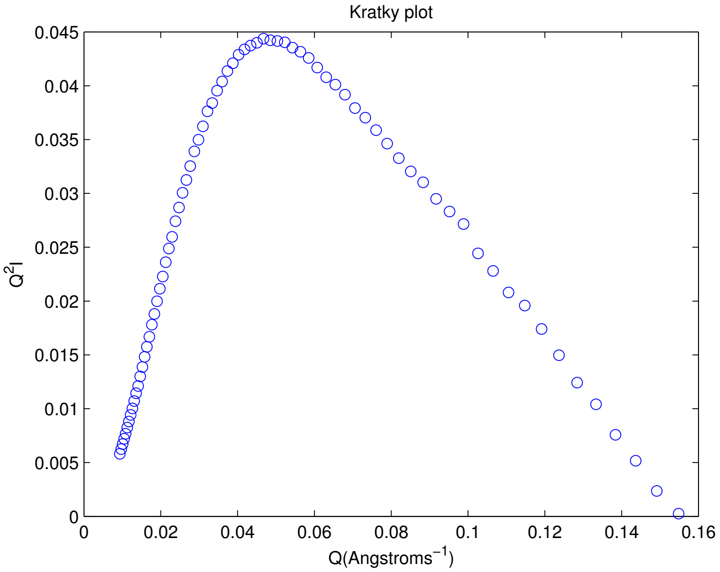


Figure 7.17: Kratky plot for miro-L in a $30mg/ml$ Ca^{2+} solution at 4^oC.

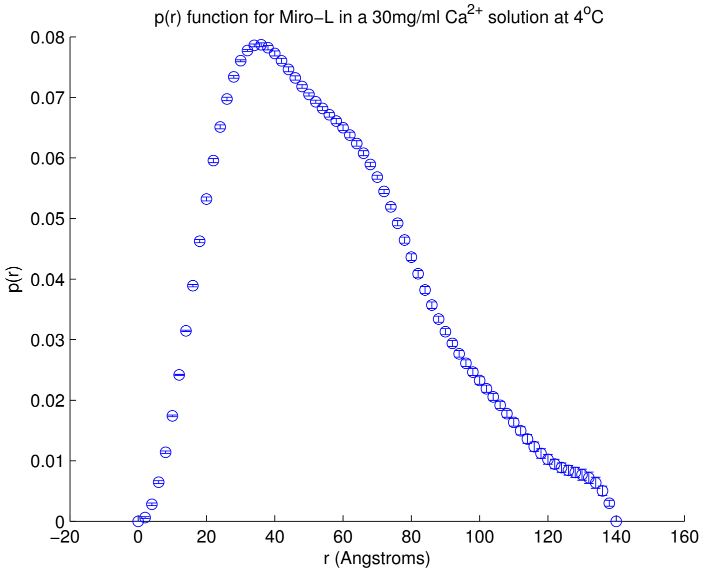


Figure 7.18: Pair-distance distribution function for Miro-L in a $30mg/ml$ Ca^{2+} solution at 4^oC. $R_g = 44.42 \pm 0.18$Å.

The low error and lack of oscillations in figure 7.18 are indicators of a successful $p(r)$ function evaluation. The radius of gyration calculated from figure 7.18, $R_g = 44.42 \pm 0.18$Åis also a close match to that obtained via Guinier analysis. This $p(r)$ curve was one of the few successful attempts carried out on the miro-L data. In most cases, the $p(r)$ curves oscillated wildly and had significant error. This was likely due to the greater size of miro-L, e.g. a maximum diameter of 140Åin figure 7.18, which caused sampling problems with the indirect Fourier transform used by GNOM. The DAMMIN reconstruction for this miro-L data can be seen in figure 7.19. This reconstruction reports a radius of gyration of $R_g = 44.19$Åand has a maximum diameter of 140Å. In order to check the accuracy of this reconstructed shape we dock it in PyMOL with the structure seen in figure 7.4. As in the case of miro-S, we expect the reconstruction from SAXS to cover a larger volume than that of our mock structure, due to the missing 82 amino acid linker in the ef hand domain. The docked structure can be seen in figure 7.20.

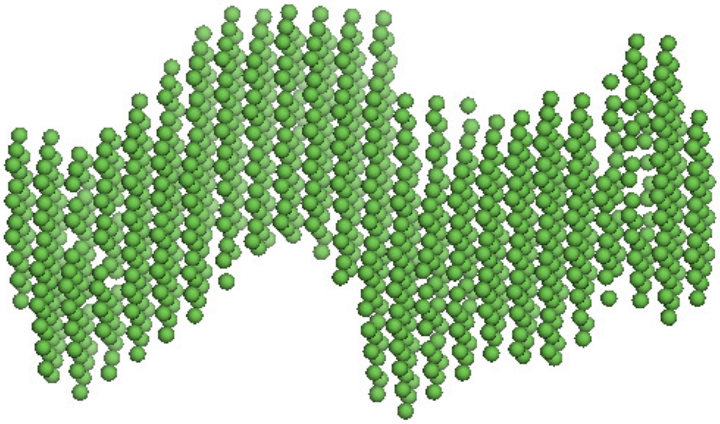


Figure 7.19: DAMMIN reconstruction of miro-L in a $30mg/ml$ Ca^{2+} solution at 4^oC. Radius of gyration $R_g = 44.19$Å. For the purpose of scale this reconstruction has a maximum diameter of 140Å.

It is once again worth noting that the orientation of the GTPase domains with

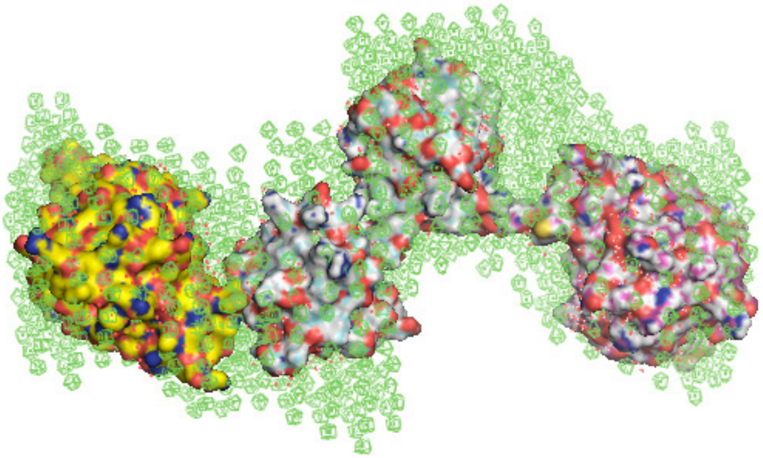


Figure 7.20: DAMMIN reconstruction of miro-L in a 30mg/ml Ca^{2+} solution at 4^oC, docked with the mock miro-L structure consisting of two 1MH1 structures and an ef hand from 2OBH.

respect to the ef hand region of miro-L is not specific. Therefore when docking the 2OBH and 1MH1 structures with our DAMMIN reconstruction, the 1MH1 and 2OBH are rotated and shifted in such a way that they best fit the DAMMIN reconstruction. Also, figure 7.20 confirms our expectations of excess volume of the DAMMIN reconstruction when compared to the mock structure. Attempts to quantify the excess volume in miro-L were not successful. The volume of the structure in figure 7.19 is reported by DAMMIN as $267,000Å^3$. This is significantly greater than expected and corresponds to an excess volume of $198,345Å^3$ when compared to the mock miro-L structure. This excess volume alone corresponds to the volume of almost 3 mock miro-L structures. It does not take much investigation of figure 7.20 to see that the SAXS reconstruction does not cover almost 4 times the volume of the mock structure. In this situation, DAMMIN has clearly overestimated the volume of the reconstruction.

The quality of this reconstruction is now checked by carrying out the FoXS comparison. In figure 7.21 the comparison between the experimental scattering intensity and the scattering pattern from the DAMMIN reconstruction is shown.

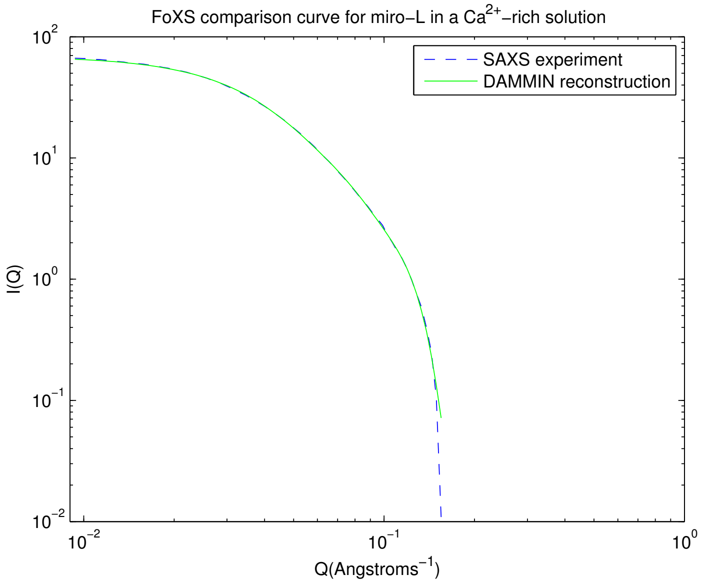


Figure 7.21: FoXS comparison of log-scaled scattering curves from DAMMIN reconstruction and SAXS experiment for miro-L in a $30mg/ml$ Ca^{2+} solution at $4^{o}C$. $\chi = 0.95$.

The FoXS comparison in figure 7.21 reports a remarkably similar agreement to that seen in the comparison for miro-S, with a fit of $\chi = 0.95$. This level of agreement combined with the visual match offered by figure 7.21 is indicative of a successful reconstruction of miro-L. As mentioned earlier, several other attempts were made at reconstructing other datasets of miro-L but in many cases, problems were encountered with the indirect Fourier transform used by GNOM when calculating the $p(r)$ function. In other cases the data was dominated by high-Q data, most evident in the Kratky plots, possibly due to contamination of the protein sample. In the case of both miro-S and miro-L, the best reconstructions were obtained from data that was taken earlier in the acquisition timeline, with the quality of data generally declining as more and more data was taken.

7.3 Calcium Ion Levels in the Miro Solution

As mentioned in Chapter 1, Ca^{2+} is suspected to play a key role in the movement of mitochondria throughout the cell (Rice, 2006). Transport of mitochondria is expected to cease in the presence of high Ca^{2+} levels. One of the possibilities in this situation is that miro undergoes a conformational change in the presence of high Ca^{2+} concentration. In this experiment, miro-S was studied in both Ca^{2+}-rich environments (as seen in figure 7.12) and in Ca^{2+}-free environments. The results of the Ca^{2+}-free SAXS studies can be seen in figures 7.22, 7.23, 7.24, 7.25, 7.26, 7.27, 7.28, and 7.29. The Ca^{2+}-rich results are also included in these figures for the sake of comparison.

In figure 7.22, the absence of Ca^{2+} results in a lower scattering intensity, but suggests a similar overall shape. The Guinier plots for both miro-S samples are very similar at low-Q, with the Ca^{2+}-free sample resulting in a radius of gyration calculation of $R_g = 27.01 \pm 0.28$Å. This is similar to the Ca^{2+}-rich sample in figure 7.8, suggesting a comparable size. Examining the Kratky plots for both samples (figure 7.24) suggests a slightly less globular structure for miro-S in the Ca^{2+}-free solution. The $p(r)$ curves for both samples, in figure 7.25, also show results one would expect.

The Ca^{2+}-rich sample has a greater scattering intensity and consequently has a greater peak on the $p(r)$ curve. Despite the much higher peak of this sample, the shapes of the $p(r)$ curves suggests the overall conformation of both structures is not dissimilar. Furthermore, the radius of gyration calculated from the Ca^{2+}-free $p(r)$ curve is $R_g = 27.24 \pm 0.3$Å, again comparable to the calculation from the Ca^{2+}-rich solution from figure 7.11. As with the Ca^{2+}-rich sample, the possible contamination of the Ca^{2+}-free protein solution caused the initial attempt at a DAMMIN reconstruction to be dominated by high-Q data. For this reason, the Ca^{2+}-free $p(r)$ curve is a low-resolution dataset, obtained by using only the first 70 of 98 data points available from the initial $I(Q)$ data. This is a significant amount

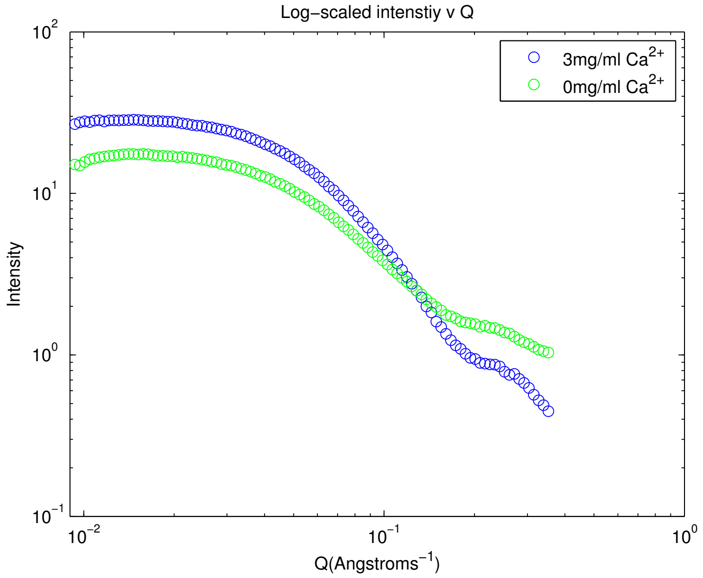


Figure 7.22: Log-scaled scattering intensity curves for miro-S in both $3mg/ml$ and $0mg/ml$ Ca^{2+} solutions at $4^{o}C$.

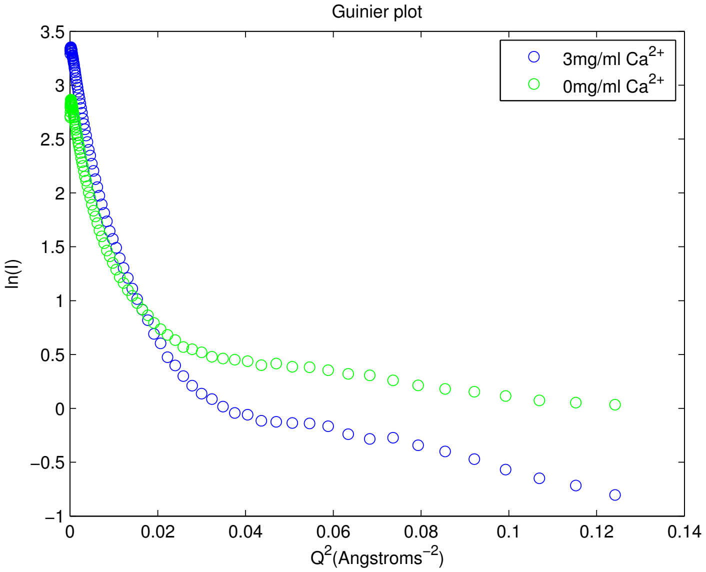


Figure 7.23: Guinier plots for miro-S in both $3mg/ml$ and $0mg/ml$ Ca^{2+} solutions at $4^{o}C$.

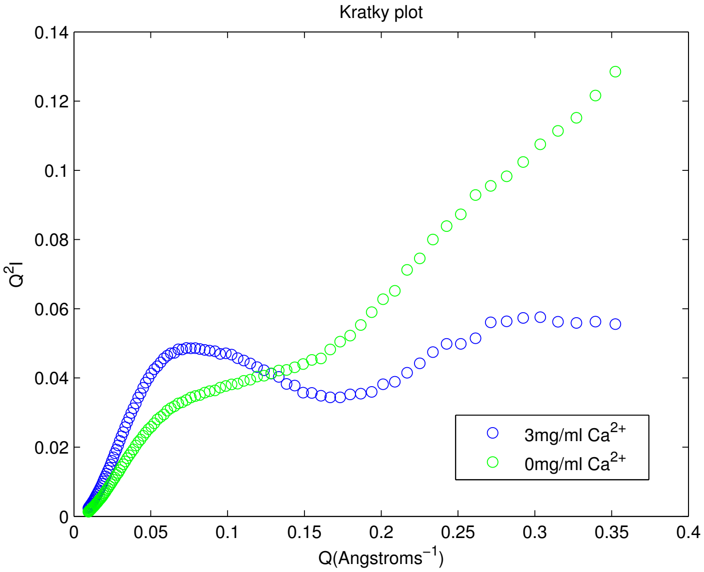


Figure 7.24: Kratky plots for miro-S in both $3mg/ml$ and $0mg/ml$ Ca^{2+} solutions at 4^oC.

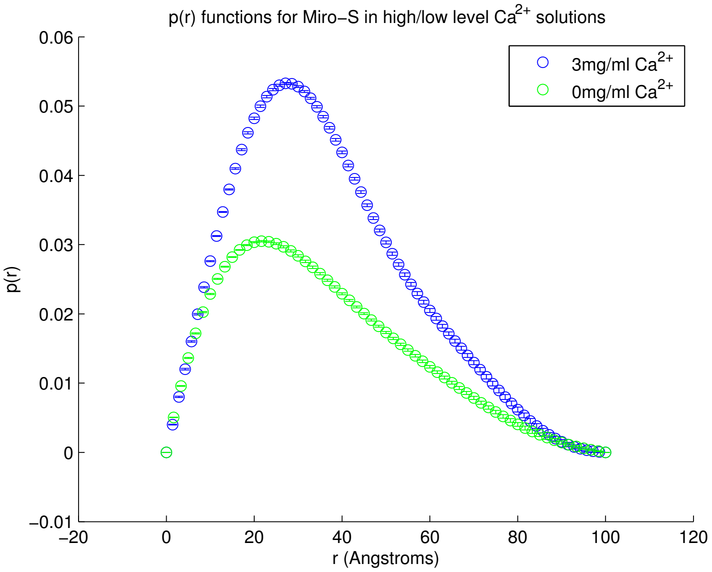


Figure 7.25: Pair-distance distribution functions for miro-S in both $3mg/ml$ and $0mg/ml$ Ca^{2+} solutions at 4^oC.

of data to omit, but it is necessary to obtain an estimation of the overall shape of the protein. The reconstruction of this protein and its comparison with the mock miro-S structure can be seen in figures 7.26 and 7.27.

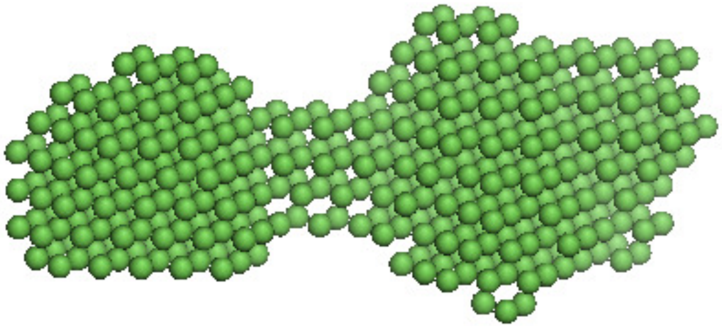


Figure 7.26: Low-resolution (70 of 98 data points from Kratky plot) DAMMIN reconstruction for miro-S in a $0mg/ml$ Ca^{2+} solution at 4^oC. For the purpose of scale, this structure has a maximum radius of 93.22Å.

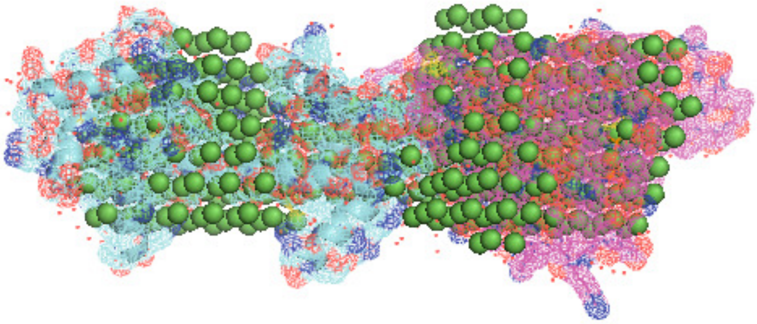


Figure 7.27: Low-resolution DAMMIN reconstruction for miro-S in a $0mg/ml$ Ca^{2+} solution at 4^oC docked with the 1MH1 and 2OBH structures.

It is quite significant that we do not see any major structural difference on the macromolecular level in miro-S when it is in a Ca^{2+}-rich environment. Although at

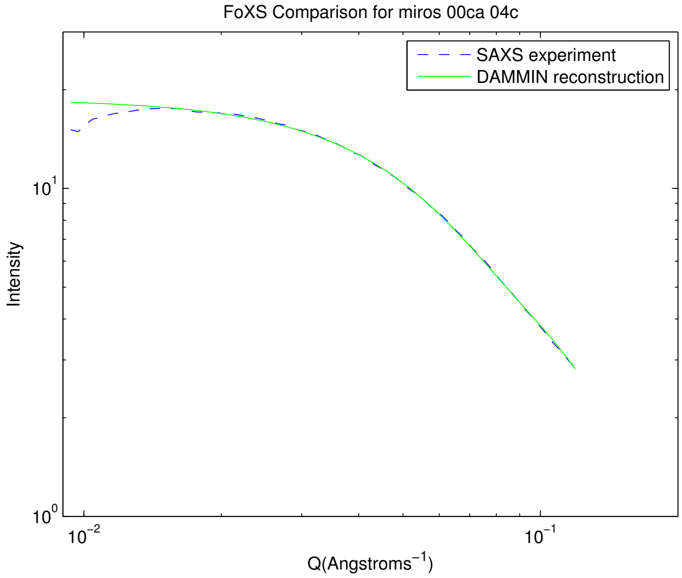


Figure 7.28: Low-resolution FoXS comparison of log-scaled scattering curves from DAMMIN reconstruction and SAXS experiment for miro-S in a $0mg/ml$ Ca^{2+} solution at 4^oC. $\chi = 0.86$

first glance the reconstructions from both Ca^{2+}-free and Ca^{2+}-rich solutions appear markedly different, one must recall that in solution, the GTPase domain acts as though hinged to the ef hand motif. In this sense, the orientation of the GTPase domain relative to the ef hand region is not specific. This is evidenced by the similar maximum diameter of both structures, 93.53Åand 93.22Åfor the Ca^{2+}-rich and Ca^{2+}-free samples, respectively. The two reconstructions can be seen docked together in figure 7.29. Although producing similar structures, we are inclined to favor results associated with the Ca^{2+}-rich data, since less data points were sacrificed in obtaining a reconstruction.

The same anlaysis is now presented for the case of miro-L in a Ca^{2+}-free solution. In figure 7.30 one can see a comparison of scattering intensities for miro-L in both Ca^{2+}-rich and Ca^{2+}-free solutions. This result differs from the results for miro-S, in that both samples of miro-L have similar intensity at low-Q. Similarly to the case

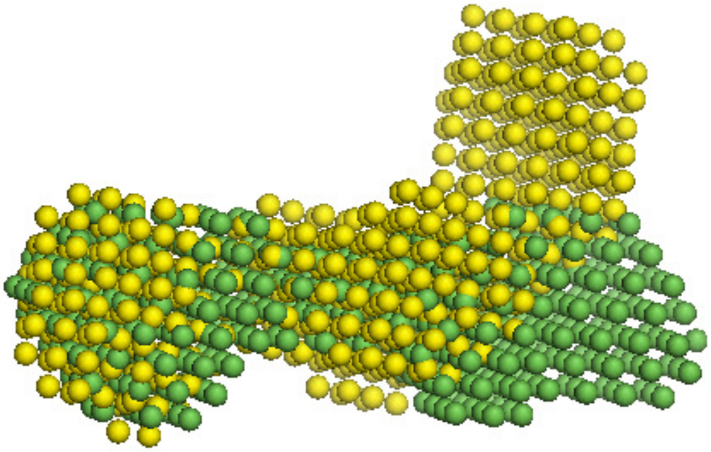


Figure 7.29: Low-resolution DAMMIN reconstructions for Miro-S in solutions with (yellow) and without (green) $3mg/ml$ Ca^{2+}.

of miro-S however, we do see greater intensity at high-Q for the Ca^{2+}-free dataset. Once again this is an indicator of similar size and shape, with details differing on a shorter (high-Q) scale.

Figure 7.31 also indicates similar size of both miro-L structures, with the low-Q regions of the Guinier plots exhibiting similar behavior. Guinier analysis of the Ca^{2+}-free data gives a radius of gyration $R_g = 42.35 \pm 0.28$Å. This is slightly less than the 44.3 ± 0.3Åobtained for the Ca^{2+}-rich solution of miro-L, indicating a slight difference in size between the two protein samples.

The Kratky plots in figure 7.32 indicate proteins of a similar shape, with the Ca^{2+}-free protein offset due the greater intensity detected at high-Q.

Examining both $p(r)$ curves in figure 7.33 further reinforces the claim of similar shape. The radius of gyration calculated from the $p(r)$ curve of the Ca^{2+}-free protein is 42.47 ± 0.12Å, again lower than its Ca^{2+}-rich counterpart (44.42 ± 0.18Å). Despite this difference, it is consistent with results of Guinier analysis for both protein

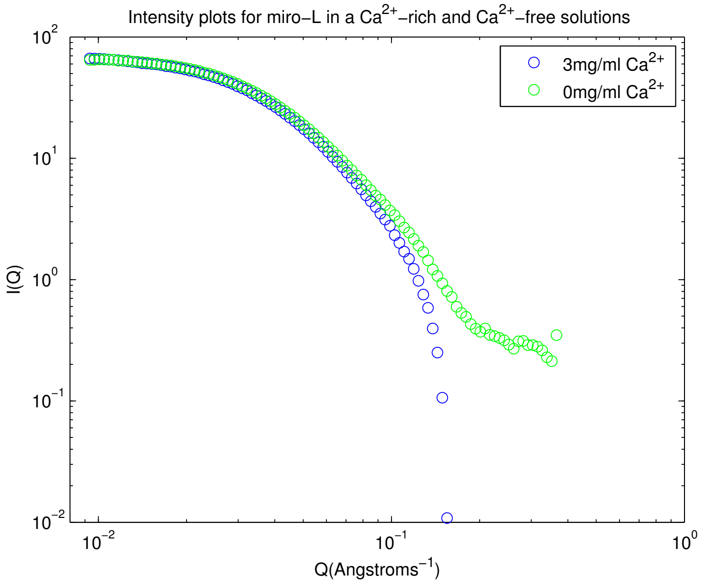


Figure 7.30: Log-scaled scattering intensity curves for miro-L in both $3mg/ml$ and $0mg/ml$ Ca^{2+} solutions at 4^oC.

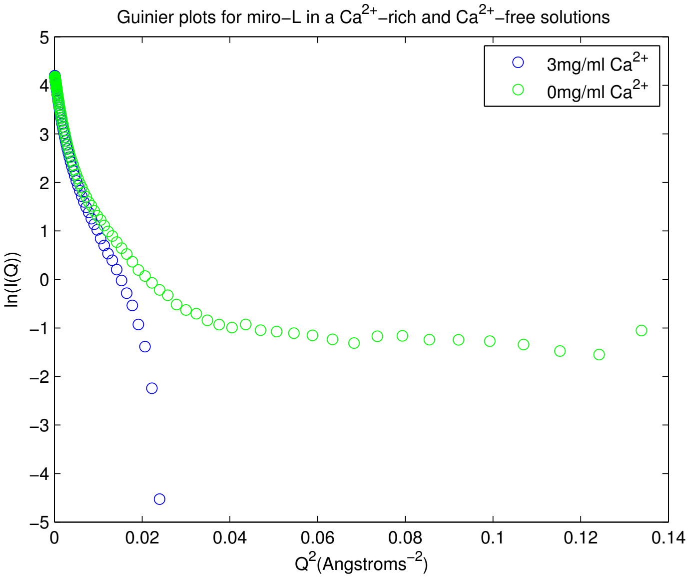


Figure 7.31: Guinier plots for miro- in both $3mg/ml$ and $0mg/ml$ Ca^{2+} solutions at 4^oC.

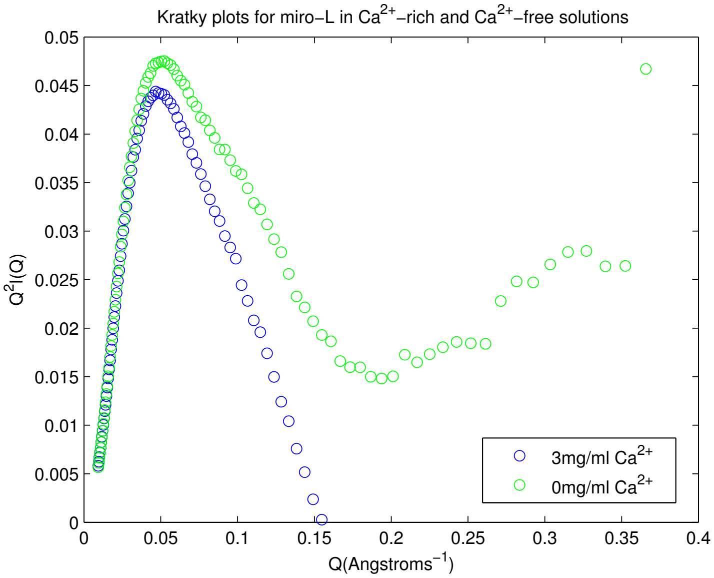


Figure 7.32: Kratky plots for miro-L in both $3mg/ml$ and $0mg/ml$ Ca^{2+} solutions at 4^oC.

samples. The higher $p(r)$-calculated R_g for the Ca^{2+}-rich sample may be due to the choice of a greater maximum diameter (R_{max}) for that sample. One will recall, R_{max} is chosen such that the $p(r)$ curve contains the desired characteristics outlined in Chapter 5, and thus it is not always possible to choose the same R_{max} for two samples of the same protein.

The DAMMIN reconstruction for our Ca^{2+}-free sample of miro-L can be seen in figure 7.34. The radius of gyration reported by DAMMIN for this object is identical to that caluclated from its $p(r)$ curve, $R_g = 42.47$.

Furthermore, its structure, while seemingly dissimilar on a high-Q scale, retains the overall shape observed in the reconstruction of our Ca^{2+}-rich sample. This variation at high-Q reflects the differences observed in the Guinier and Kratky plots of both miro-L samples. The reconstruction from the Ca^{2+}-free sample can be seen docked with our mock miro-L object in figure 7.35.

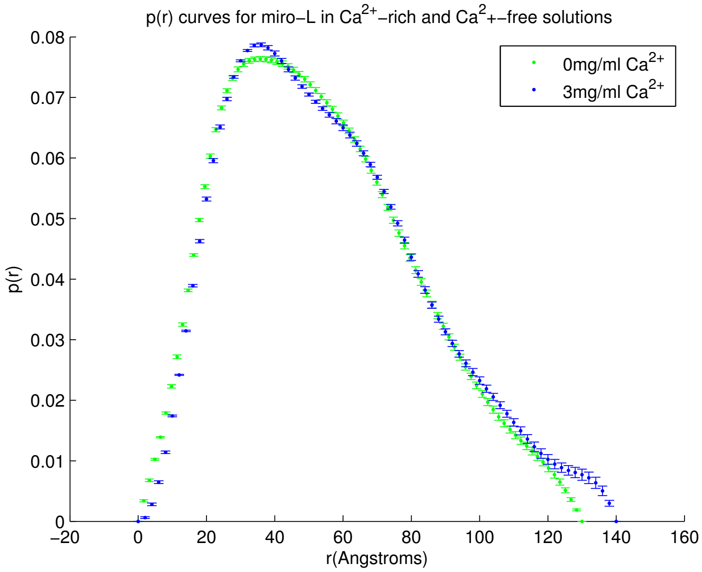


Figure 7.33: Pair-distance distribution functions for miro-L in both $3mg/ml$ and $0mg/ml$ Ca^{2+} solutions at 4^oC.

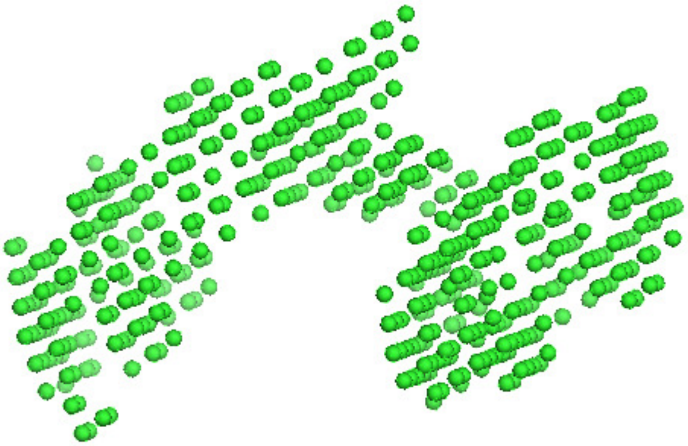


Figure 7.34: DAMMIN reconstruction for miro-L in a $0mg/ml$ Ca^{2+} solution at 4^oC. For the purposes of scale, this structure has a maximum radius of 93.22Å.

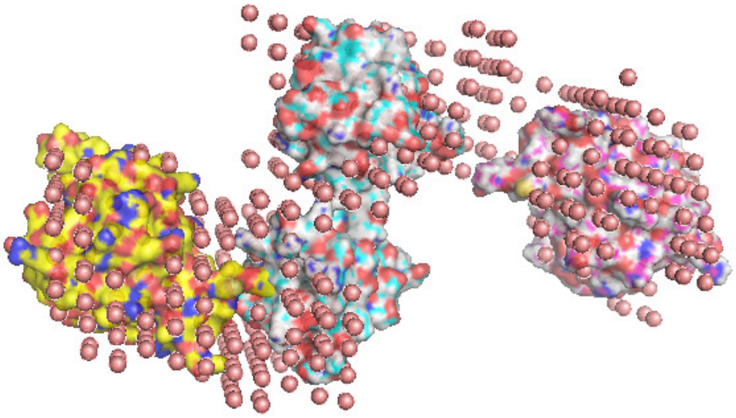


Figure 7.35: DAMMIN reconstruction for miro-L in a $0mg/ml$ Ca^{2+} solution at 4^oC docked with the 1MH1 and 2OBH structures.

One can again see that the reconstruction fits the expected form of miro-L, with excess volume appearing in the ef hand region of the SAXS data. The similarity in overall shape when compared to the Ca^{2+}-rich sample can be seen in figure 7.37.

As with the case of miro-S, the amount of structural change on the macromolecular level is not significant. Rather, the addition of Ca^{2+} to our miro protein solution appears, if anything, to affect the protein on the high-Q scale, indicating small scale change within the protein. Change such as this is not easily detectable by SAXS since its observation requires a higher resolution.

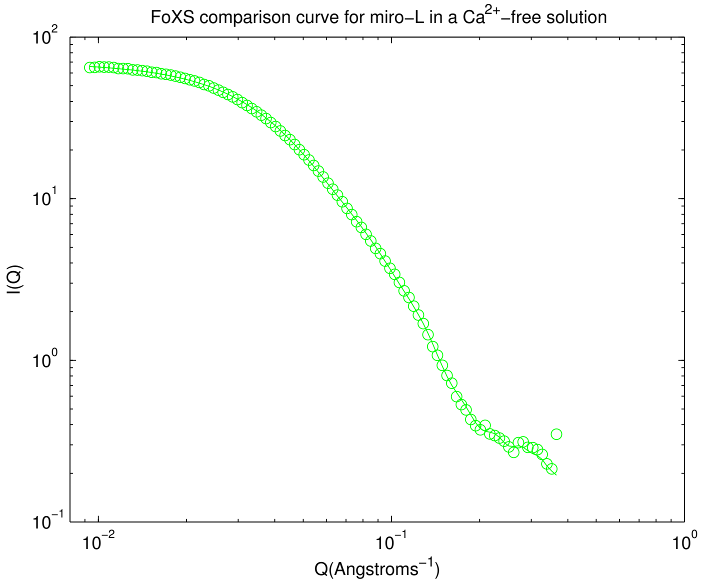


Figure 7.36: FoXS comparison of log-scaled scattering curves from DAMMIN reconstruction and SAXS experiment for miro-L in a $0mg/ml$ Ca^{2+} solution at 4^oC. $\chi = 0.94$

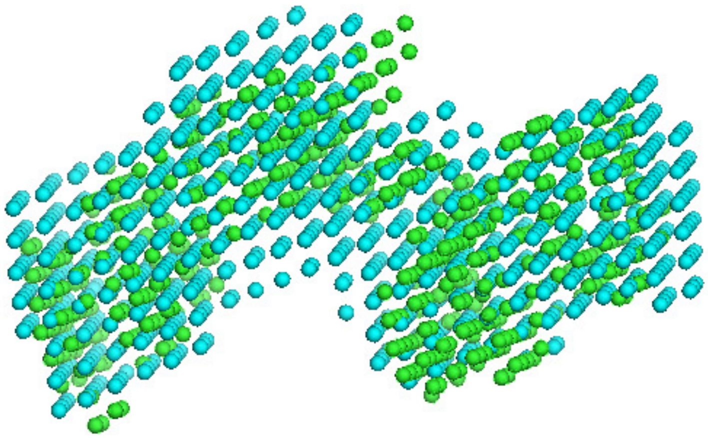


Figure 7.37: DAMMIN reconstructions for miro-L in both $0mg/ml$ Ca^{2+} (green spheres) and $3mg/ml$ Ca^{2+} (blue spheres) solutions at 4^oC.

CHAPTER 8

Conclusion

8.1 Cytochrome-c

Cytochrome-c is an ideal protein for demonstrating the roles of various thermodynamic forces in protein denaturation. In this experiment we have obtained for the first time direct visualization of cold denatured equine cytochrome-c through the use of SAXS. In doing so we have confirmed the adequacy of the two state protein stability theory (Dill, 2009) extended to the case of cytochrome-c (Elmer, 2010) by verifying data collected by Elmer in 2009. In verifying this data we have shown the indirect Fourier transform and simulated annealing process of obtaining a 3D reconstruction to be appropriate in the case of cytochrome-c. Reconstruction of SAXS data for the protein in its native state shows a global shape consistent with that of x-ray crystallography studies. Reconstructing the denatured protein, both at warm and cold temperatures, sometimes proved problematic due to issues in carrying out the indirect Fourier transform on the scattering data. The roles of various thermodynamic forces are evident when comparing the cases of warm and cold denaturation. Both denatured proteins were shown to have the same size, but the forces driving denaturation in each case we suggest to be different. In the case of warm denaturation the thermodynamic process is entropy driven. Theory suggests the protein has more possible configurations in this case, and experimental evidence in the form of shorter persistence length is consistent with this expectation. The converse is expected in the case of cold denaturation. Here, the thermodynamic process is enthalpy driven and less possible configurations are suggested. The longer persistence length calculated in this case suggests the same. For this reason, warm and cold denaturation of cytochrome-c, despite being described by the same total free folding energy, result in different unfolded structures of similar size.

Future work on this topic includes time-resolved studies of the protein folding process. Since conformational changes occur over nanoseconds or microseconds, it is desirable to capture images of the protein at a corresponding frequency, so that the unfolding process may be observed directly. In this method, a high power laser is used to cause a temperature jump in the protein solution, thereby inducing the folding/unfolding process. Work such as this, if successful, may also aid in verifying computational models of the protein folding process. As a more direct result of this project, implications may be important in future protein folding experiments where different thermodynamic conditions are altered to induce protein unfolding.

8.2 Mitochondrial-Rho

This experiment has also provided for the first time an observation of the global structure of the miro protein. Prior to these SAXS studies of miro, suggestions of the protein's structure came from analysis of the amino acid sequence. The results of this experiment have verified inferences based on such analysis, specifically with regard to an unknown 82 amino acid chain within the ef hand region of the protein. Sequence analysis suggested two GTPase domains connected by an ef hand motif, a structure verified by SAXS analysis and protein reconstruction. The size of the miro protein (particularly miro-L) unfortunately caused problems when implementing the indirect Fourier transform during the analysis process. It is also possible that contamination of the protein sample led to corrupt data which could not be reconstructed. SAXS studies of the miro protein also showed no change in its global structure when additional calcium ions were added to the protein solution. Since miro is a calcium binding protein it has been suggested that Ca^{2+} may be involved in stopping mitochondrial movement within the cell by altering miro's structure.

This experiment has shown that should this be true, the change to miro's structure is likely on the local scale, since the addition of Ca^{2+} to both miro-S and miro-L did not show any significant change to the global structure.

Future work is already underway in further determining the structure of miro. The Rice Lab at Northwestern University is currently in the process of developing a crystallized miro-S sample and further SAXS experiments on the protein are also upcoming. Experiments in the immediate future will be focused on using SAXS results to complement results from x-ray crystallography on miro-S.

APPENDIX A - MATLAB CODES

GRAPHS FOR CHAPTER 2:

```
% Thesis chapter 2 graphs. Anthony Banks, 8/29/12

clear all
close all

% Parameters
Cp = 5000;          % Specific Heat (J/mol*K)
T = 225:1:925;      % Temp (K)
Th = 373;           % Temp for H=0 (K)
Ts = 385;           % Temp for S=0 (K)
% y = -0.8e6:1e4:8e6;
% ys = -4e3:10:2e3;
% yg = -2.5e5:1e3:1e5;
g0 = -1200*4.184;   % Avg Packing Energy (J/mol)
m1 = 25*4.184;      % Concentration-Energy Factor (J/mol)
N = 104;            % Number of Amino Acids
z= 7.54;            % Number of Rotational Isomers
c = 2;              % Denaturant Concentration (molar)
R = 8.314;          % Universal Gas Constant (J/mol*K)
kb = 1.38e-23;      % Boltzmann Constant (J/K)

% Functions
HT = Cp.*(T-Th);    % Enthalpy (assuming H(Th) is zero)
ST = Cp.*log(T./Ts); % Entropy (assuming S(Ts) is zero)
G = HT - T.*ST;     % Gibbs free energy, (J/mol)
g = g0 + m1*c;      % Packing Energy (cal/mol)

% Neutral Folding Free Energy
F = g*N + T.*R*N*log(z) + Cp.*(T-Th) - T.*Cp.*log(T./Ts);

% Electrostatics
pH = 0.0214.*T + 9.89;              % pH level of solution
eps = 7e-4.*(T.^2) - 0.79.*T + 251; % Dielectric Constant of Water
lb = 1.39e-4./(eps.*R.*T);          % Bjerrum Length
kp = sqrt(2*c.*lb);                 % Poisson-Boltzmann Constant
pKb = [11,9];
pKa = [2.4,2.9,6.4];            % Protein Dissociation Constants

% Basic and Acidic Charge Contributions
```

```
for i = length(pKb)
    for j = length(T)
        Qb(i,j) = (log10(pKb(i)-pH(j)))/(1+log10(pKb(i)-pH(j)));
    end
end
for i = length(pKa)
    for j = length(T)
        Qa(i,j) = (log10(pH(j)-pKa(i)))/(1+log10(pH(j)-pKa(i)));
    end
end

Qn = sum(Qb) - sum(Qa); % Net Charge

% Gibbs Free Energy Curve
figure (1)
plot(T,G,T,0)
title('Gibbs Free Energy')
xlabel('Temperature (K)')
ylabel('Gibbs Free Energy (J/mol)')

zero = 0*T;
% Enthalpy and Temperature-Scaled Entropy
figure (2)
plot(T,T.*ST,T,HT)
title('Enthalpy and Temperature-Scaled Entropy')
xlabel('Temperature (K)')
ylabel('Enthalpy, Temperature-Scaled Entropy (J/mol)')

% Plotting the above graphs together
figure (3)
plot(T,T.*ST,T,HT,T,G)
title('Gibbs Free Energy, Enthalpy, and Temperature-Scaled Entropy')
xlabel('Temperature (K)')
ylabel('Energy (J/mol)')

% Neutral Folding Free Energy Curve
figure (4)
plot(T,F,T,0)
title('Neutral Folding Free Energy')
xlabel('Temperature (K)')
ylabel('Energy (J/mol)')
```

GRAPHS FOR CHAPTER 3:

```
close all
clear all

%% Small Angle X-Ray Scattering

%  Plotting the intensity of scattered x-rays from a homogeneous sphere
%  Anthony Banks, 2/7/12. Revised 10/30/12

%% Fixed Parameters

p = 4;                  % density of sphere (kg/m^3)
p0 = 1;                 % density of water (kg/m^3)
% Rg = 13;                % radius of gyration of sphere (A)
a = 16;                 % radius of sphere
Rg = sqrt(3/5)*a;       % radius of gyration of sphere (A)
L = 10;                 % x-ray wavelength (A)
maxth = 0.006*180/pi;   % max angle

%% Optional Inputs

%a = input('sphere radius = ');          % radius
%maxth = input('max angle = ');          % max angle
%L = input('x-ray wavelength = ');       % wavelength

%% Variables

j = maxth/100;                          % step size
th = 0:j:maxth;                         % angle (rad)
Q = 2.*sin(th)./L;                      % distance from center
f = 2*pi*a.*Q;                          % variable such that PHI = PHI(f)
PHI = 3.*((sin(f)-f.*cos(f))./(f.^3));  % variable such that I = I(PHI)

I = (((p-p0)*((4/3)*pi*(a^3)))^2).*((PHI).^2);  % intensity
I = I./(((p-p0)*((4/3)*pi*(a^3)))^2);           % normalized intensity

rmax = 200;
r = rmax/length(Q):rmax/length(Q):rmax;
int = (Q.^2).*I.*sin(Q.*r)./(Q.*r);
pr = ((r(2:length(r)).^2)./(2*pi^2)).*trapz(Q(2:length(Q)),int(2:length(int)));

%% Graphs
```

```
% Scattering
figure (1)
plot(Q,I)
title 'Intensity v Q'
xlabel 'Q (1/A)'
ylabel 'Intensity (Normalized)'

Q2 = Q.^2;
gQ2 = Q2(2:(0.3*(length(th)-1)));
lnI = log(I);
glnI = lnI(2:(0.3*(length(th)-1)));

% Gunier
figure (2)
plot(Q2,lnI)
title 'Guinier Plot'
xlabel 'Q^2 (1/A^2)'
ylabel 'Natural Log of Intensity'

% Guinier Region
figure (3)
plot(gQ2,glnI)
title 'Low-Q Region of Guinier Plot'
xlabel 'Q^2 (1/A^2)'
ylabel 'Natural Log of Intensity'

p2 = polyfit(gQ2,glnI,1);
Rg2 = sqrt((-p2(1)*3)/(4*(pi^2)));      % Rg from Guinier Plot

% Kratky
figure (4)
plot(Q,I.*(Q.^2))
title 'Kratky Plot'
xlabel 'Q (1/A)'
ylabel 'IQ^2'

% LogScale
figure (5)
loglog(Q,I)
title 'LogLog-Scale Intensity'
ylabel 'Intensity'
xlabel 'Q'
```

```
CHAPTER 5 GRAPHS:

clear all
close all

% Program to take input data and generate p(r) curve for analysis section
% of thesis. Data is of folded Cytochrome-C. Anthony Banks 9/13/12

% Load Data
rawdata = '/Users/anthonybanks/Desktop/Research/Thesis/thch5prdata.txt';
fid = fopen(rawdata);
dat = textscan(fid, '%f %f %f');
r = dat{1};
p = dat{2};
err = dat{3}.*ones(size(r));

% Plot p(r) Curve
figure (1)
errorbar(r,p,err,err,'o')
title('Pair-Distance Distribution Function')
xlabel('Distance, r (Angstroms)')
ylabel('p(r)')

Rg = sqrt(trapz(r,p.*(r.^2))/(2*trapz(r,p)));

% Program to take input scattering data from SAXS measurements, x-ray
% crystallography, and reconstructed protein and compare for analysis
% section of thesis. Data is for folded Cytochrome-C

% Load XRD data and SAXS data
rawdata2 = '/Users/anthonybanks/Desktop/Research/Thesis/cytcXRDwithSAXSdata.txt';
fid2 = fopen(rawdata2);
dat2 = textscan(fid2, '%f %f %f');
q2 = dat2{1};
expI2 = dat2{2};
xrdI2 = dat2{3};
q2b = q2;
% (1:length(q2)-1);
expI2b = expI2;
% (1:length(expI2)-1);
```

```
% Load DAMMIN reconstrcuted data and SAXS data
rawdata3 = '/Users/anthonybanks/Desktop/Research/Thesis/cytcDAMwithSAXSdata.txt';
fid3 = fopen(rawdata3);
dat3 = textscan(fid3, '%f %f %f');
q3 = dat3{1};
expI3 = dat3{2};
damI3 = dat3{3};

% Plot scattering curves for the three datasets
figure (2)
loglog(q2,expI2,'b',q2,damI3,'r--',q2,xrdI2,'g-.')
title('Comparison of Scattering Curves')
ylabel('Scattering Intensity, I(Q)')
xlabel('Radial Distance, Q(1/A)')

rawdata4 = '/Users/anthonybanks/Desktop/Research/Thesis/M00C2.dat';
fid4 = fopen(rawdata4);
dat4 = textscan(fid4, '%f %f %f');
q4 = dat4{1};
I4 = dat4{2};
q4 = q4(1:length(q4)-4);
I4 = I4(1:length(I4)-4);
errxp = dat4{3};
errxp = errxp(1:length(errxp)-4).*(q4.^2);

% Kratky plot for three datasets
figure (3)
hold on
errorbar(q4,(q4.^2).*I4,errxp,errxp,'o')
plot(q2,(q2.^2).*damI3,'-r',q2,(q2.^2).*xrdI2,'-g')
hold off
title('Kratky Plots for Scattering Data')
xlabel('Radial Distance, Q(1/A)')
ylabel('I*Q^2')


CHAPTER 6 GRAPHS:

% SAXS_data_Rg_lb.m
% Import, process, and plot reduced SAXS and calculate Radius of Gyration, Rg
% By Eric Landahl, 11/9/10 revised to make better plots 11/18/10
% Modified to import all data at 0, 2, 2.5, and 4 M 1/10/11
% Modified to calculate P(r) and persistance length lb 6/7/11
```

```
% Porod's Law calculation of lb commented out 6/7/11

%% Import Data

clear all;

fconc = '25';  % concentration for filename

if fconc == '0'
    numtemps = 14; % for 4.0 M % number of temperature runs
    ffront='/Nanosecond T-Jump Small Angle X-Ray Scattering/Text/Data/';
    ftemp1 = -[20 15 10 5]; % for 2.5 M
    ftemp2 = [0 5]; % for 2.5 M
    ftemp3 = [10 15 20 30 35 40 45 50]; % for 2.5 and 2.0 M
    fopt = 'M';  % optional filename adddition for ftemp3
    foptT = ''; % optional filename addition
end

if fconc == '2'
    numtemps = 15; % for 2.0 M
    ffront='/Nanosecond T-Jump Small Angle X-Ray Scattering/Text/Data/M';
    ftemp1 = -[24 20 15 10 5]; % for 2.0 M
    ftemp2 = [0 5]; % for 2.0 M
    ftemp3 = [10 15 25 30 35 40 45 50]; % for 2.5 and 2.0 M
    fopt = 'M'; % optional filename adddition for ftemp3
    foptT = 'T'; % optional filename addition
end

if fconc == '25' % for 2.5 M
    numtemps = 17; % for 2.5 M % number of temperature runs
        ffront='/Nanosecond T-Jump Small Angle X-Ray Scattering/Text/Data/M';
    ftemp1 = -[22 20 17 12 10 7 2]; % for 2.5 M
    ftemp2 = [2 5]; % for 2.5 M
    ftemp3 = [10 15 25 30 35 40 45 50]; % for 2.5 and 2.0 M
    fopt = '';  % optional filename adddition for ftemp3
    foptT = 'T'; % optional filename addition
end

if fconc == '4'
    numtemps = 14; % for 4.0 M % number of temperature runs
    ffront='/Nanosecond T-Jump Small Angle X-Ray Scattering/Text/Data/';
    ftemp1 = -[20 15 10 5]; % for 2.5 M
    ftemp2 = [0 5]; % for 2.5 M
```

```
    ftemp3 = [10 15 20 30 35 40 45 50]; % for 2.5 and 2.0 M
    fopt = 'M';  % optional filename adddition for ftemp3
    foptT = ''; % optional filename addition
end

fend='C1.txt';

x_0 = 0;  % first data point (normally zero)
x_f = 118; % last data point (normally 100)

for t = 1:length(ftemp1);
    fname = [ffront 'M' fconc foptT '_' num2str(-ftemp1(t)) fend];
    fid = fopen(fname, 'r');
    jnk = textscan(fid, '%f %f %f',1,'headerlines',2);
    data = textscan(fid, '%f %f %f');
    x(:,t)=data{1};
%     error(:,t)=data{3};
    q = data{2};
end

for t = length(ftemp1)+1:length(ftemp1)+length(ftemp2);
    fname = [ffront 'M' fconc foptT num2str(ftemp2(t-length(ftemp1))) fend];
    fid = fopen(fname, 'r');
    jnk = textscan(fid, '%f %f %f',1,'headerlines',2);
    data = textscan(fid, '%f %f %f');
    x(:,t)=data{1};
%     error(:,t)=data{3};
%    q = data{2};
end

for t = length(ftemp1)+length(ftemp2)+1:numtemps;
    fname = [ffront fopt fconc foptT num2str(ftemp3(t-(length(ftemp1)+length(ftemp2)))) fend];
    fid = fopen(fname, 'r');
    jnk = textscan(fid, '%f %f %f',1,'headerlines',2);
    data = textscan(fid, '%f %f %f');
    x(:,t)=data{1};
%     error(:,t)=data{3};
end

ftemp = [ftemp1 ftemp2 ftemp3];

%% Remove highest q points (data is useless due to noise)
x_old = x;
```

```
q_old = q;
clear x;
clear q;
for i = 1:(x_f-1)
    x(i,:) = x_old(i,:);
    q(i) = q_old(i);
end
q = q';

%% Normalize
 %for j = 1:numtemps
 %     x(:,j)=x(:,j)/max(x(:,j));
 %end

%% Kratky Plots

figure(1);clf;hold on;
colr = colormap(jet(numtemps));
for t=1:numtemps
    plot(q,(q.^2).*x(:,t),'Color',colr(t,:),'LineWidth',1)
end
title('Kratky Plot 2.5M GuHCl')
xlabel('Q (1/Angstrom)')
ylabel('Intensity * Q^2')
legend(num2str(ftemp'))
hold off

%% Intensity Plots
figure(31)
hold on
for t=1:numtemps
plot(log(q),log(x(:,t)),'Color',colr(t,:),'LineWidth',1)
end
title('Log-Log Scaled Intensity 2.5M GuHCl')
xlabel('Q (1/Angstrom)')
ylabel('Intensity')
legend(num2str(ftemp'))
hold off

%% Calculate Rg
minpt = 15; % lowest q index for Gunier plot nom. 15
maxpt = 60; % highest q index for Gunier plot nom. 60
xg = x(minpt:maxpt,:);
```

```
qg = q(minpt:maxpt);
figure(2);clf;hold on;
for t = 1:numtemps
    plot(qg.^2,log(xg(:,t)),'.','Color',colr(t,:))
    fit = polyfit(qg.^2,log(xg(:,t)),1);
    m(t)=fit(1); % Slope (related to radius of gyration)
    b(t)=fit(2); % Intercept (related to particle volume)
end
Rg = sqrt(-3*m)/10; % Radius of gyration in nm
title('Gunier Plot 2.5M GuHCl')
xlabel('Q^2 (1/Angstrom^2)')
ylabel('log(Intensity)')
legend(num2str(ftemp'))
for t = 1:numtemps
   plot(qg.^2,m(t)*qg.^2+b(t),'Color',colr(t,:))
end
hold off
figure(3);clf;hold on;
plot(ftemp,Rg,'o')
xlabel('Temperature (deg C)')
ylabel('Rg (nm)')
title('Protein Size vs. Temperature at 2.5M GuHCl')

%% Plot Kratky with similar Rg

figure(4);clf;hold on;
colr = colormap(jet(numtemps));
t1=4;
plot(q,(q.^2).*x(:,t1),'Color',colr(t1,:),'LineWidth',2)
t2=11;
plot(q,(q.^2).*x(:,t2),'Color',colr(t2,:),'LineWidth',2)
title([num2str(Rg(t1)) 'A at ' num2str(ftemp(t1)) 'C and ' num2str(Rg(t2)) 'A at ' num2str(ftemp
(t2)) 'C'])
xlabel('Q (1/Angstrom)')
ylabel('Intensity * Q^2')
%legend(num2str(ftemp'))
hold off

figure(5);clf;hold on;
colr = colormap(jet(numtemps));
t1=3;
plot(q,(q.^2).*x(:,t1),'Color',colr(t1,:),'LineWidth',2)
t2=12;
```

```
plot(q,(q.^2).*x(:,t2),'Color',colr(t2,:),'LineWidth',2)
title([num2str(Rg(t1)) 'A at ' num2str(ftemp(t1)) 'C and ' num2str(Rg(t2)) 'A at ' num2str(ftemp
(t2)) 'C'])
xlabel('Q (1/Angstrom)')
ylabel('Intensity * Q^2')
%legend(num2str(ftemp'))
hold off

figure(6);clf;hold on;
colr = colormap(jet(numtemps));
t1=6;
plot(q,(q.^2).*x(:,t1),'Color',colr(t1,:),'LineWidth',2)
t2=10;
plot(q,(q.^2).*x(:,t2),'Color',colr(t2,:),'LineWidth',2)
title([num2str(Rg(t1)) 'A at ' num2str(ftemp(t1)) 'C and ' num2str(Rg(t2)) 'A at ' num2str(ftemp
(t2)) 'C'])
xlabel('Q (1/Angstrom)')
ylabel('Intensity * Q^2')
%legend(num2str(ftemp'))
hold off

%% Porod's Law
minpt = 102; % lowest q index for Porod's Law
maxpt = 117; % highest q index for Porod's Law
xp = x(minpt:maxpt,:);
qp = q(minpt:maxpt);
figure(7);clf;hold on;
for t = 1:numtemps
    plot(qp,xp(:,t).*qp.^2,'.','Color',colr(t,:))
    fit = polyfit(qp,xp(:,t).*qp.^2,1);
    a(t)=fit(2); % offset
    b(t)=fit(1); % 1/q
 %   c(t)=fit(1); % 1/q^2
end
title('Porods Law')
xlabel('Q (1/Angstrom)')
ylabel('Intensity*Q^2')
%legend(num2str(ftemp'))
for t = 1:numtemps
   plot(qp,a(t)+b(t).*qp ,'Color',colr(t,:))
end
% hold off
% figure(8);clf;hold on;
```

```
% for t = 1:numtemps
%     plot(ftemp(t),a(t),'o','MarkerEdgeColor',colr(t,:))
% end
% ylabel('Dimensionality')
% xlabel('Temperature (deg C)')
% title('Porod Analysis at 2.5M GuHCl')
persist=-a./b; % persistance q
figure(8);clf;
t = 1:numtemps;
plot(ftemp(t),persist(t),'o')
xlabel('Temperature (deg C)')
ylabel('Persistance Q (1/Angstrom)')

% %% Distance Distribution function P(r)
% figure(9);clf;
% for t = 1:numtemps
%     rpts=100;
%     dmax=70;
%     for rnum = 1:rpts
%         r = dmax*rnum/rpts; % Was 50
%         ptemp=(r/(2*pi^2))*(q(1)*x(1,t).*sin(q(1)*r))*q(1);
%         ptemp=0;
%         for qnum = 15:80 % Was 40:100
%             dq = q(qnum)-q(qnum-1);
%             ptemp = ptemp+(r/(2*pi^2))*(q(qnum)*x(qnum,t).*sin(q(qnum)*r))*dq; % radial dist
 func.
%         end
%         p(rnum,t)=ptemp;
%         rr(rnum)=r;
%     end
% end
% figure(9);hold on;
% plot(rr,p)
% ylim([0 max(max(p))])
% xlabel('r (Angstrom)');
% ylabel('p(r)');
% title('Distance Distribution Function');
% for t = 1:numtemps
%     rgtemp1=0;
%     rgtemp2=0;
%     for rnum = 1:rpts
%         if p(rnum,t) > 0  % only positive values of p(r) contribute to Rg
%             rgtemp1=rgtemp1+p(rnum,t)*rr(rnum)^2;
```

```
%             rgtemp2=rgtemp2+2*p(rnum,t);
%         end
%     end
%     rg(t)=sqrt(rgtemp1/rgtemp2);
% end
% figure(10);clf;hold on;
% plot(ftemp,rg,'o')
% xlabel('Temperature (deg C)')
% ylabel('Rg (Angstrom)')
% %% Calculate lb using curve fit to p(r)
% figure(12);clf(12); % Clear figure showing fits
% for t = 1:numtemps
%     [po(t) lb(t) p_fit r_fit]=WLC_Fit(p(:,t),rr,rg(t));
% end
% figure(11);clf;hold on;
% plot(ftemp,lb,'o')
% xlabel('Temperature (deg C)')
% ylabel('Persistance Length (Angstrom)')
%

clear all
close all

% Program to read in radii of gyration from cytochrome-C p(r) curves and
% plot for varying molarity and temperature. Also contains Rg values from
% IgorPro and from MATLAB analysis. Anthony Banks, 12/10/12

%% Import p(r) Data
rawdata = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/GNOMRgvals.txt';
fid = fopen(rawdata);
dat = textscan(fid, '%f %f %f');
t = dat{1};
prrg = dat{2};
err = dat{3};

%% Sort p(r) data
prt0 = t(1:15);
prt2 = t(16:28);
prt25 = t(29:49);
prt4 = t(50:63);

prrg0 = prrg(1:15);
prrg2 = prrg(16:28);
```

```
prrg25 = prrg(29:49);
prrg4 = prrg(50:63);

err0 = err(1:15);
err2 = err(16:28);
err25 = err(29:49);
err4 = err(50:63);

prpct0 = 100*err0./prrg0;
prpct25 = 100*err25./prrg25;
prpct4 = 100*err4./prrg4;

%% Results from MATLAB analysis (SAXS_data_Rg_lb.m)

Rg0 = [1.37506077630418,1.35709391066313,1.45614379848387,1.29461257698835,1.3662574
9672085,1.36423217170344,1.69609348568154,1.38957499914510,1.35533910803646,1.
43538811383953,1.35477087688931,1.39652672045133,1.56870709952835,1.55905997
170966];
t0 = [-20,-15,-10,-5,0,5,10,15,20,30,35,40,45,50];
Rg25 = [2.12816569327235,2.05818532930598,2.01523819142115,1.94402827023630,1.9117183
7852353,1.86170973371064,1.84122299550482,1.81203048758621,1.87578767178838,1.
97012340621021,2.04802378239003,2.26906271329393,2.31415044090789,2.36767554
631922,2.44174320564040,2.48782097794133,2.47419020505677];
t25 = [-22,-20,-17,-12,-10,-7,-2,2,5,10,15,25,30,35,40,45,50];
Rg4 = [2.65050639049563,2.64546749338987,2.60272884428079,2.57759493866481,2.5933746
1306441,2.58502639490038,2.62689185940457,2.51837829693407,2.57376658665074,2.
57947808578814,2.58022687473143,2.56147756603217,2.57992933669984,2.60640941
440917];
t4 = [-20,-15,-10,-5,0,5,10,15,20,30,35,40,45,50];

%% Results from IgorPro analysis

iprg0 = [13.5890000000000,13.4480000000000,13.7630000000000,13.3700000000000,13.377000
0000000,13.4720000000000,13.6930000000000,13.4920000000000,13.8480000000000,1
3.6190000000000,13.3830000000000,13.6570000000000,14.1020000000000,14.0560000
000000];
ipt0 = [-20,-15,-10,-5,0,5,10,15,25,30,35,40,45,50];
iperr0 = [0.0514790000000000,0.0833520000000000,0.0805660000000000,0.0343460000000000,
0.0361910000000000,0.0492840000000000,0.0432250000000000,0.0587980000000000,
0.140690000000000,0.0672320000000000,0.0449690000000000,0.0777510000000000,0
.0574710000000000,0.0580460000000000];
iprg25 = [25.1580000000000,24.5420000000000,22.4240000000000,23.1660000000000,20.984000
0000000,19.1170000000000,18.9680000000000,17.3510000000000,17.1010000000000,1
```

```
8.5340000000000,17.0680000000000,18.1840000000000,19.7840000000000,22.0070000
000000,22.8500000000000,24.8800000000000,25.7320000000000,26.2040000000000,26
.5920000000000,27.0480000000000,27.3070000000000];
ipt25 = [-22,-20,-17,-15,-12,-10,-7,-5,-2,0,2,5,10,15,20,25,30,35,40,45,50];
iperr25 = [0.0134050000000000,0.0117110000000000,0.0106620000000000,0.0090211000000000
0,0.0112600000000000,0.00946470000000000,0.00851170000000000,0.0070144000000
0000,0.00498300000000000,0.0134020000000000,0.231970000000000,0.208460000000
000,0.0215020000000000,0.00806270000000000,0.0120430000000000,0.016069000000
0000,0.0141530000000000,0.0181810000000000,0.00786960000000000,0.02198400000
00000,0.0265180000000000];
iprg4 = [29.8620000000000,30.3900000000000,31.0320000000000,30.5660000000000,28.657000
0000000,29.9350000000000,28.7850000000000,29.3610000000000,30.6110000000000,2
9.5870000000000,29.1230000000000,30.0330000000000,30.0330000000000,30.6570000
000000];
ipt4 = [-20,-15,-10,-5,0,5,10,15,20,30,35,40,45,50];
iperr4 = [0.820940000000000,0.870399900000000,0.0613330000000000,0.925420000000000,1.00
750000000000,0.687680000000000,0.0174390000000000,0.538790000000000,0.714590
000000000,0.548631000000000,0.614500000000000,0.489920000000000,0.4541800000
00000,0.723060000000000];

ippct0 = 100*iperr0./iprg0;
ippct25 = 100*iperr25./iprg25;
ippct4 = 100*iperr4./iprg4;

%% Rgs from DAMMIN Reconstructions

tD = [-22,-20,-17,-15,-12,-10,-7,-5,-2,0,2,5,10,15,20,25,30,35,40,45];
rgD = [24.18,22.39,17.73,20.05,18.8,18.3,20.77,17.68,16.47,16.54,17.47,17.5,17.93,19.13,20.02,22
.32,23.93,25.71,24.35,25.24];

%% Plot Data
figure (1)
title 'Radius of Gyration from p(r) curves'
xlabel 'Temperature (C^o)'
ylabel 'Radius of Gyration (Angstroms)'
hold on
errorbar(prt0,prrg0,err0,'ko')
% errorbar(prt2,prrg2,err2,'bo')
errorbar(prt25,prrg25,err25,'ro')
errorbar(prt4,prrg4,err4,'go')
hold off

figure (2)
```

```
hold on
plot(t0,10.*Rg0,'bo')
plot(t25,10.*Rg25,'ko')
plot(t4,10.*Rg4,'ro')
title 'Radius of gyration from MATLAB analysis'
xlabel 'Temperature (^oC)'
ylabel 'Radius of Gyration (Angstroms)'
hold off

figure (3)
hold on
errorbar(ipt0,iprg0,iperr0,'k*')
errorbar(ipt25,iprg25,iperr25,'r*')
errorbar(ipt4,iprg4,iperr4,'g*')
title 'Radius of gyration from Guinier analysis'
xlabel 'Temperature (^oC)'
ylabel 'Radius of Gyration (Angstroms)'
hold off

figure (4)
hold on
plot(prt0,prrg0,'ko')
% plot(prt2,prrg2,err2,'bo')
plot(prt25,prrg25,'ro')
plot(prt4,prrg4,'go')
plot(ipt0,iprg0,'k*')
plot(ipt25,iprg25,'r*')
plot(ipt4,iprg4,'g*')
title 'Radius of gyration from Guinier and p(r) analysis'
xlabel 'Temperature (^oC)'
ylabel 'Radius of Gyration (Angstroms)'
hold off

figure (5)
hold on
plot(prt0,prpct0,'ko')
plot(prt25,prpct25,'ro')
plot(prt4,prpct4,'go')
plot(ipt0,ippct0,'k*')
plot(ipt25,ippct25,'r*')
plot(ipt4,ippct4,'g*')
title 'Percent error for Guinier and p(r) radius of gyration calculations'
xlabel 'Temperature (^oC)'
```

```
ylabel '% Error'
hold off

figure (6)
hold on
% plot(prt0,prrg0,'ko')
% plot(prt2,prrg2,err2,'bo')
plot(prt25,prrg25,'ro')
% plot(prt4,prrg4,'go')
% plot(ipt0,iprg0,'k*')
plot(ipt25,iprg25,'r*')
% plot(ipt4,iprg4,'g*')
plot(tD,rgD,'bsq')
title 'Radius of gyration from Guinier and p(r) analysis, and DAMMIN rconstructions'
xlabel 'Temperature (^oC)'
ylabel 'Radius of Gyration (Angstroms)'
hold off

clear all
close all

% Program to take in scattering profiles from FoXS and plot DAMMIN I(q)
% with experimental I(q). Anthony Banks, 01/03/13

%% Get Data

rawdata_22 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
MM25T_22C1/foxs_22.txt';
fid_22 = fopen(rawdata_22);
dat_22 = textscan(fid_22, '%f %f %f','headerlines',1);
Q_22 = dat_22{1};                       % Scattering vector
Iexp_22 = dat_22{2};                    % Scattering from theoretical sphere
Isim_22 = dat_22{3};                    % FoXS-calculated scattering from DAMMIN

rawdata_20 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
M25T_20C1/foxs_20.txt';
fid_20 = fopen(rawdata_20);
dat_20 = textscan(fid_20, '%f %f %f','headerlines',1);
Q_20 = dat_20{1};                       % Scattering vector
Iexp_20 = dat_20{2};                    % Scattering from theoretical sphere
Isim_20 = dat_20{3};                    % FoXS-calculated scattering from DAMMIN

rawdata_17 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
```

```
MM25T_17C1/foxs_17.txt';
fid_17 = fopen(rawdata_17);
dat_17 = textscan(fid_17, '%f %f %f','headerlines',1);
Q_17 = dat_17{1};                    % Scattering vector
Iexp_17 = dat_17{2};                 % Scattering from theoretical sphere
Isim_17 = dat_17{3};                 % FoXS-calculated scattering from DAMMIN

rawdata_15 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
M25T_15C1/foxs_15.txt';
fid_15 = fopen(rawdata_15);
dat_15 = textscan(fid_15, '%f %f %f','headerlines',1);
Q_15 = dat_15{1};                    % Scattering vector
Iexp_15 = dat_15{2};                 % Scattering from theoretical sphere
Isim_15 = dat_15{3};                 % FoXS-calculated scattering from DAMMIN

rawdata_12 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
MM25T_12C1/foxs_12.txt';
fid_12 = fopen(rawdata_12);
dat_12 = textscan(fid_12, '%f %f %f','headerlines',1);
Q_12 = dat_12{1};                    % Scattering vector
Iexp_12 = dat_12{2};                 % Scattering from theoretical sphere
Isim_12 = dat_12{3};                 % FoXS-calculated scattering from DAMMIN

rawdata_10 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
MM25T_10C1/foxs_10.txt';
fid_10 = fopen(rawdata_10);
dat_10 = textscan(fid_10, '%f %f %f','headerlines',1);
Q_10 = dat_10{1};                    % Scattering vector
Iexp_10 = dat_10{2};                 % Scattering from theoretical sphere
Isim_10 = dat_10{3};                 % FoXS-calculated scattering from DAMMIN

rawdata_7 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
MM25T_7C1/foxs_7.txt';
fid_7 = fopen(rawdata_7);
dat_7 = textscan(fid_7, '%f %f %f','headerlines',1);
Q_7 = dat_7{1};                    % Scattering vector
Iexp_7 = dat_7{2};                 % Scattering from theoretical sphere
Isim_7 = dat_7{3};                 % FoXS-calculated scattering from DAMMIN

rawdata_5 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
M25T_5C1/foxs_5.txt';
fid_5 = fopen(rawdata_5);
dat_5 = textscan(fid_5, '%f %f %f','headerlines',1);
```

```
Q_5 = dat_5{1};                      % Scattering vector
Iexp_5 = dat_5{2};                   % Scattering from theoretical sphere
Isim_5 = dat_5{3};                   % FoXS-calculated scattering from DAMMIN

rawdata_2 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
MM25T_2C1/foxs_2.txt';
fid_2 = fopen(rawdata_2);
dat_2 = textscan(fid_2, '%f %f %f','headerlines',1);
Q_2 = dat_2{1};                      % Scattering vector
Iexp_2 = dat_2{2};                   % Scattering from theoretical sphere
Isim_2 = dat_2{3};                   % FoXS-calculated scattering from DAMMIN

rawdata0 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
MM25T0C1/foxs0.txt';
fid0 = fopen(rawdata0);
dat0 = textscan(fid0, '%f %f %f','headerlines',1);
Q0 = dat0{1};                      % Scattering vector
Iexp0 = dat0{2};                   % Scattering from theoretical sphere
Isim0 = dat0{3};                   % FoXS-calculated scattering from DAMMIN

rawdata2 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
MM25T2C1/foxs2.txt';
fid2 = fopen(rawdata2);
dat2 = textscan(fid2, '%f %f %f','headerlines',1);
Q2 = dat2{1};                      % Scattering vector
Iexp2 = dat2{2};                   % Scattering from theoretical sphere
Isim2 = dat2{3};                   % FoXS-calculated scattering from DAMMIN

rawdata5 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
M25T5C1/foxs5.txt';
fid5 = fopen(rawdata5);
dat5 = textscan(fid5, '%f %f %f','headerlines',1);
Q5 = dat5{1};                      % Scattering vector
Iexp5 = dat5{2};                   % Scattering from theoretical sphere
Isim5 = dat5{3};                   % FoXS-calculated scattering from DAMMIN

rawdata10 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
M25T10C1/foxs10.txt';
fid10 = fopen(rawdata10);
dat10 = textscan(fid10, '%f %f %f','headerlines',1);
Q10 = dat10{1};                      % Scattering vector
Iexp10 = dat10{2};                   % Scattering from theoretical sphere
Isim10 = dat10{3};                   % FoXS-calculated scattering from DAMMIN
```

```
rawdata15 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
M25T15C1/foxs15.txt';
fid15 = fopen(rawdata15);
dat15 = textscan(fid15, '%f %f %f','headerlines',1);
Q15 = dat15{1};                       % Scattering vector
Iexp15 = dat15{2};                    % Scattering from theoretical sphere
Isim15 = dat15{3};                    % FoXS-calculated scattering from DAMMIN

rawdata20 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
M25T20C1/foxs20.txt';
fid20 = fopen(rawdata20);
dat20 = textscan(fid20, '%f %f %f','headerlines',1);
Q20 = dat20{1};                       % Scattering vector
Iexp20 = dat20{2};                    % Scattering from theoretical sphere
Isim20 = dat20{3};                    % FoXS-calculated scattering from DAMMIN

rawdata25 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
M25T25C1/foxs25.txt';
fid25 = fopen(rawdata25);
dat25 = textscan(fid25, '%f %f %f','headerlines',1);
Q25 = dat25{1};                       % Scattering vector
Iexp25 = dat25{2};                    % Scattering from theoretical sphere
Isim25 = dat25{3};                    % FoXS-calculated scattering from DAMMIN

rawdata30 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
M25T30C1/foxs30.txt';
fid30 = fopen(rawdata30);
dat30 = textscan(fid30, '%f %f %f','headerlines',1);
Q30 = dat30{1};                       % Scattering vector
Iexp30 = dat30{2};                    % Scattering from theoretical sphere
Isim30 = dat30{3};                    % FoXS-calculated scattering from DAMMIN

rawdata35 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
M25T35C1/foxs35.txt';
fid35 = fopen(rawdata35);
dat35 = textscan(fid35, '%f %f %f','headerlines',1);
Q35 = dat35{1};                       % Scattering vector
Iexp35 = dat35{2};                    % Scattering from theoretical sphere
Isim35 = dat35{3};                    % FoXS-calculated scattering from DAMMIN

rawdata40 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
M25T40C1/foxs40.txt';
```

```
fid40 = fopen(rawdata40);
dat40 = textscan(fid40, '%f %f %f','headerlines',1);
Q40 = dat40{1};                      % Scattering vector
Iexp40 = dat40{2};                   % Scattering from theoretical sphere
Isim40 = dat40{3};                   % FoXS-calculated scattering from DAMMIN

rawdata45 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/25M BSUB DATA/
M25T45C1/foxs45.txt';
fid45 = fopen(rawdata45);
dat45 = textscan(fid45, '%f %f %f','headerlines',1);
Q45 = dat45{1};                      % Scattering vector
Iexp45 = dat45{2};                   % Scattering from theoretical sphere
Isim45 = dat45{3};                   % FoXS-calculated scattering from DAMMIN

rawdata0M0 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/0M BSUB DATA/
M0OC2 Test Runs/foxs0M0.txt';
fid0M0 = fopen(rawdata0M0);
dat0M0 = textscan(fid0M0, '%f %f %f','headerlines',1);
Q0M0 = dat0M0{1};                      % Scattering vector
Iexp0M0 = dat0M0{2};                   % Scattering from theoretical sphere
Isim0M0 = dat0M0{3};                   % FoXS-calculated scattering from DAMMIN

rawdata4M0 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/CYTC/4M BSUB DATA/
M4OC1 Test Runs/foxs4M0.txt';
fid4M0 = fopen(rawdata4M0);
dat4M0 = textscan(fid4M0, '%f %f %f','headerlines',1);
Q4M0 = dat4M0{1};                      % Scattering vector
Iexp4M0 = dat4M0{2};                   % Scattering from theoretical sphere
Isim4M0 = dat4M0{3};                   % FoXS-calculated scattering from DAMMIN

chiD = ((Isim_20 - Iexp_20).^2)./Iexp_20;
chiD = sum(chiD);

%% Plots

colr = colormap(jet(22));

figure (1)
subplot(2,2,2),loglog(Q4M0,Iexp4M0,'--',Q4M0,Isim4M0,'-','color',colr(1,:))
subplot(2,2,3),loglog(Q_22,Iexp_22,'--',Q_22,Isim_22,'-','color',colr(2,:))
subplot(2,2,1),loglog(Q_20,Iexp20,'--',Q_20,Isim_20,'-','color',colr(3,:))

figure (2)
```

```
subplot(2,2,1),loglog(Q_17,Iexp_17,'--',Q_17,Isim_17,'-','color',colr(4,:))
subplot(2,2,2),loglog(Q_15,Iexp_15,'--',Q_15,Isim_15,'-','color',colr(5,:))
subplot(2,2,3),loglog(Q_12,Iexp_12,'--',Q_12,Isim_12,'-','color',colr(6,:))
subplot(2,2,4),loglog(Q_10,Iexp_10,'--',Q_10,Isim_10,'-','color',colr(7,:))

figure (3)
subplot(2,2,1),loglog(Q_7,Iexp_7,'--',Q_7,Isim_7,'-','color',colr(8,:))
subplot(2,2,2),loglog(Q_5,Iexp_5,'--',Q_5,Isim_5,'-','color',colr(9,:))
subplot(2,2,3),loglog(Q_2,Iexp_2,'--',Q_2,Isim_2,'-','color',colr(10,:))
subplot(2,2,4),loglog(Q0,Iexp0,'--',Q0,Isim0,'-','color',colr(11,:))

figure (4)
subplot(2,2,1),loglog(Q0M0,Iexp0M0,'--',Q0M0,Isim0M0,'-','color',colr(12,:))
subplot(2,2,2),loglog(Q2,Iexp2,'--',Q2,Isim2,'-','color',colr(13,:))
subplot(2,2,3),loglog(Q5,Iexp5,'--',Q5,Isim5,'-','color',colr(14,:))
subplot(2,2,4),loglog(Q10,Iexp10,'--',Q10,Isim10,'-','color',colr(15,:))

figure (5)
subplot(2,2,1),loglog(Q15,Iexp15,'--',Q15,Isim15,'-','color',colr(16,:))
subplot(2,2,2),loglog(Q20,Iexp20,'--',Q20,Isim20,'-','color',colr(17,:))
subplot(2,2,3),loglog(Q25,Iexp25,'--',Q25,Isim25,'-','color',colr(18,:))
subplot(2,2,4),loglog(Q30,Iexp30,'--',Q30,Isim30,'-','color',colr(19,:))

figure (6)
subplot(2,2,1),loglog(Q35,Iexp35,'--',Q35,Isim35,'-','color',colr(20,:))
subplot(2,2,2),loglog(Q40,Iexp40,'--',Q40,Isim40,'-','color',colr(21,:))
subplot(2,2,3),loglog(Q45,Iexp45,'--',Q45,Isim45,'-','color',colr(22,:))
```

CHAPTER 7 GRAPHS

```
%Program for analyzing miros data from March 13th 2012 experiment at the
%APS. I(q) data comes from 2D intensity images that have been reduced to I(q)
%curves and then filtered for quality. p(r) data comes from GNOM output files
%Anthony Banks 8/18/2012

clear all
close all

%% Load I(q) data

run1 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_S/miros_03ca_04c_1_1/
miros_03ca_04c_1_1.dat';
```

```
fid1 = fopen(run1, 'r');
jnk1 = textscan(fid1, '%f %f %f',1,'headerlines',1);
dat1 = textscan(fid1, '%f %f %f');

run2 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_S/miros_00ca_04c_3_1/
miros_00ca_04c_3_1.dat';
fid2 = fopen(run2, 'r');
jnk2 = textscan(fid2, '%f %f %f',1,'headerlines',1);
dat2 = textscan(fid2, '%f %f %f');

%% Data separation

Qrun1 = dat1{1};
Irun1 = dat1{2};
erun1 = dat1{3};

Qrun2 = dat2{1};
Irun2 = dat2{2};
erun2 = dat2{3};

%% Load p(r) data

load /Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_S/
miros_03ca_04c_1_1/80_98LR100A1pr.txt
load /Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_S/
miros_00ca_04c_3_1/70_98LR100Apr.txt

p1 = X80_98LR100A1pr(:,2);
p2 = X70_98LR100Apr(:,2);

r1 = X80_98LR100A1pr(:,1);
r2 = X70_98LR100Apr(:,1);

errp1 = X80_98LR100A1pr(:,3);
errp2 = X70_98LR100Apr(:,3);

%% Load FoXS data

foxs1 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_S/miros_03ca_04c_1_1/
foxs80_98LR100A1.txt';
fid1 = fopen(foxs1, 'r');
jnk1 = textscan(fid1, '%f %f %f',1,'headerlines',1);
foxsdat1 = textscan(fid1, '%f %f %f');
```

```
foxq1 = foxsdat1{1};
saxsI1 = foxsdat1{2};
damI1 = foxsdat1{3};

foxs2 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_S/miros_00ca_04c_3_1/
foxs70_98LR100A2.txt';
fid2 = fopen(foxs2, 'r');
jnk2 = textscan(fid2, '%f %f %f',1,'headerlines',1);
foxsdat2 = textscan(fid2, '%f %f %f');
foxq2 = foxsdat2{1};
saxsI2 = foxsdat2{2};
damI2 = foxsdat2{3};

%% Plot data

%loglog plots
figure(1)
loglog(Qrun1,Irun1,'bo',Qrun2,Irun2,'go')
title ('Intensity plots for Miro-S in high/low level Ca^2^+ solutions')
xlabel ('Q(cm^-^1)')
ylabel ('Intensity')

%regular I(q) plots
figure(2)
plot(Qrun1,Irun1,'b',Qrun2,Irun2,'g')
title ('intensity plots')
xlabel ('q(1/A^o)')
ylabel ('intensity')

%guinier plots
figure(3)
plot(Qrun1.^2,log(Irun1),'bo',Qrun2.^2,log(Irun2),'go')
title('Guinier plots for MiroS in high/low level Ca^2^+ solutions')
xlabel('Q^2(cm^-^2)')
ylabel('ln(I)')

%kratky plots
figure (7)
plot(Qrun1,(Qrun1.^2).*Irun1,'bo',Qrun2,(Qrun2.^2).*Irun2,'go')
%plot(Qrun1(1:80),(Qrun1(1:80).^2).*Irun1(1:80),'bo')
title('Kratky plots for Miro-S in high/low level Ca^2^+ solutions')
xlabel('Q(cm^-^1)')
ylabel('Q^2I')
```

```
%p(r) curves
figure(4)
hold on
errorbar(r1,p1,errp1,'bo')
errorbar(r2,p2,errp2,'go')
title ('p(r) functions for Miro-S in high/low level Ca^2^+ solutions')
xlabel ('r')
ylabel ('p')
hold off

%FoXS plot for 03ca Miros
figure(5)
loglog(foxq1,saxsI1,'b--',foxq1(1:80),damI1(1:80),'b-')
title('FoXS Comparison for miros 03ca 04c')
xlabel('Q')
ylabel('I')

%Foxs plot for 00ca Miros
figure(6)
loglog(foxq2,saxsI2,'b--',foxq2,damI2,'b-')
title('FoXS Comparison for miros 00ca 04c')
xlabel('Q(cm^-^1)')
ylabel('Intensity')


%Program for analyzing mirol data from March 13th 2012 experiment at the
%APS. I(q) data comes from 2D intensity images that have been reduced to I(q)
%curves and then filtered for quality. p(r) data comes from GNOM output files
%Anthony Banks 8/18/2012

clear all
close all

%% Load I(q) data

run3_1 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
mirol_run3_00ca1mg_1/mirol_run3_00ca1mg_1.dat';
fid31 = fopen(run3_1, 'r');
jnk31 = textscan(fid31, '%f %f %f',1,'headerlines',1);
dat31 = textscan(fid31, '%f %f %f');
%
% run3_2 = '/Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
```

```
mirol_run3_00ca1mg_2/mirol_run3_00ca1mg_2.dat’;
% fid32 = fopen(run3_2, ’r’);
% jnk32 = textscan(fid32, ’%f %f %f’,1,’headerlines’,1);
% dat32 = textscan(fid32, ’%f %f %f’);

run4_1 = ’/Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
mirol_run4_30ca1mg_1/mirol_run4_30ca1mg_1.dat’;
fid41 = fopen(run4_1, ’r’);
jnk41 = textscan(fid41, ’%f %f %f’,1,’headerlines’,1);
dat41 = textscan(fid41, ’%f %f %f’);

% run4_2 = ’/Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
mirol_run4_30ca1mg_2/mirol_run4_30ca1mg_2.dat’;
% fid42 = fopen(run4_2, ’r’);
% jnk42 = textscan(fid42, ’%f %f %f’,1,’headerlines’,1);
% dat42 = textscan(fid42, ’%f %f %f’);
%
% run13_1 = ’/Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
mirol_run13_00ca05mgalf_1/mirol_run13_00ca05mgalf_1.dat’;
% fid131 = fopen(run13_1, ’r’);
% jnk131 = textscan(fid131, ’%f %f %f’,1,’headerlines’,1);
% dat131 = textscan(fid131, ’%f %f %f’);
%
% run23_1 = ’/Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
mirol_run23_0ca1mg_1/mirol_run23_0ca1mg_1.dat’;
% fid231 = fopen(run23_1, ’r’);
% jnk231 = textscan(fid231, ’%f %f %f’,1,’headerlines’,1);
% dat231 = textscan(fid231, ’%f %f %f’);
%
% run24_1 = ’/Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
mirol_run24_300ca1mg_1/mirol_run24_300ca1mg_1.dat’;
% fid241 = fopen(run24_1, ’r’);
% jnk241 = textscan(fid241, ’%f %f %f’,1,’headerlines’,1);
% dat241 = textscan(fid241, ’%f %f %f’);
%
%% Data separation

Qrun31 = dat31{1};
Irun31 = dat31{2};
erun31 = dat31{3};
%
% Qrun32 = dat32{1};
% Irun32 = dat32{2};
```

```
% erun32 = dat32{3};
%
Qrun41 = dat41{1};
Irun41 = dat41{2};
erun41 = dat41{3};
%
% Qrun42 = dat42{1};
% Irun42 = dat42{2};
% erun42 = dat42{3};
%
% Qrun131 = dat131{1};
% Irun131 = dat131{2};
% erun131 = dat131{3};
%
% Qrun231 = dat231{1};
% Irun231 = dat231{2};
% erun231 = dat231{3};
%
% Qrun241 = dat241{1};
% Irun241 = dat241{2};
% erun241 = dat241{3};

%% Load p(r) data

load /Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/mirol_run3_00ca1mg_1/
mirol_run3_1p(r).txt
% load /Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
mirol_run3_00ca1mg_2/mirol_run3_2p(r).txt
load /Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/mirol_run4_30ca1mg_1/
mirol_run4_1p(r).txt
% load /Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
mirol_run4_30ca1mg_2/mirol_run4_2p(r).txt
% load /Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
mirol_run13_00ca05mgalf_1/mirol_run13_1p(r).txt
% load /Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
mirol_run23_0ca1mg_1/mirol_run23_1p(r).txt
% load /Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
mirol_run24_300ca1mg_1/mirol_run24_1p(r).txt
%
p31 = mirol_run3_1p_r_(:,2);
% p32 = mirol_run3_2p_r_(:,2);
p41 = mirol_run4_1p_r_(:,2);
% p42 = mirol_run4_2p_r_(:,2);
```

```
% p131 = mirol_run13_1p_r_(:,2);
% p231 = mirol_run23_1p_r_(:,2);
% p241 = mirol_run24_1p_r_(:,2);
%
r31 = mirol_run3_1p_r_(:,1);
% r32 = mirol_run3_2p_r_(:,1);
r41 = mirol_run4_1p_r_(:,1);
% r42 = mirol_run4_2p_r_(:,1);
% r131 = mirol_run13_1p_r_(:,1);
% r231 = mirol_run23_1p_r_(:,1);
% r241 = mirol_run24_1p_r_(:,1);

err31 = mirol_run3_1p_r_(:,3);
err41 = mirol_run4_1p_r_(:,3);

%% FoXS data
run4_1foxs = '/Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
mirol_run4_30ca1mg_1/mirol30cafoxs.txt';
fid41foxs = fopen(run4_1foxs, 'r');
jnk41foxs = textscan(fid41foxs, '%f %f %f',1,'headerlines',1);
dat41foxs = textscan(fid41foxs, '%f %f %f');

q41foxs = dat41foxs{1};
I41foxsexp = dat41foxs{2};
I41foxsdam = dat41foxs{3};

run3_1foxs = '/Users/anthonybanks/Desktop/Research/ANALYSIS/MIRO_L/
mirol_run3_00ca1mg_1/mirol0cafoxs.txt';
fid31foxs = fopen(run3_1foxs, 'r');
jnk31foxs = textscan(fid31foxs, '%f %f %f',1,'headerlines',1);
dat31foxs = textscan(fid31foxs, '%f %f %f');

q31foxs = dat31foxs{1};
I31foxsexp = dat31foxs{2};
I31foxsdam = dat31foxs{3};

%% Plot data

%loglog plots
figure(1)
loglog(Qrun41,Irun41,'bo',Qrun31,Irun31,'go')
title ('Intensity plots for miro-L in a Ca^2^+-rich and Ca^2^+-free solutions')
xlabel ('Q(Angstroms^-^1)')
```

```
ylabel ('I(Q)')

% %regular I(q) plots
% figure(2)
% title ('intensity plots')
% xlabel ('q(1/A^o)')
% ylabel ('intensity')
% plot(Qrun31,Irun31,'b',Qrun32,Irun32,'g',Qrun41,Irun41,'r',Qrun42,Irun42,'y',Qrun131,Irun131,'c',
Qrun231,Irun231,'m',Qrun241,Irun241,'k')

% gunier plots
figure(3)
plot(Qrun41.^2,log(Irun41),'bo',Qrun31.^2,log(Irun31),'go')
title ('Guinier plots for miro-L in a Ca^2^+-rich and Ca^2^+-free solutions')
xlabel ('Q^2(Angstroms^-^2)')
ylabel ('ln(I(Q))')

% kratky plot
figure (5)
plot(Qrun41, Irun41.*(Qrun41.^2),'bo',Qrun31, Irun31.*(Qrun31.^2),'go')
title ('Kratky plots for miro-L in Ca^2^+-rich and Ca^2^+-free solutions')
xlabel ('Q(Angstroms^-^1)')
ylabel ('Q^2I(Q)')

%p(r) curves
figure(4)
title ('p(r) curves for miro-L in Ca^2^+-rich and Ca^2+-free solutions')
xlabel ('r(Angstroms)')
ylabel ('p(r)')
hold on
errorbar(r31,p31,err31,'g.')
% plot(r32,p32,'g')
errorbar(r41,p41,err41,'b.')
% plot(r42,p42,'y')
% plot(r131,p131,'c')
% plot(r231,p231,'m')
% plot(r241,p241,'k')
hold off

figure (6)
loglog(q41foxs,I41foxsexp,'bo',q41foxs,I41foxsdam,'b')
title('FoXS comparison curve for miro-L in a Ca^2^+-rich solution')
xlabel('Q(Angstroms^-^1)')
```

```
ylabel('I(Q)')

figure (7)
loglog(q31foxs,I31foxsexp,'go',q31foxs,I31foxsdam,'g')
title('FoXS comparison curve for miro-L in a Ca^2^+-free solution')
xlabel('Q(Angstroms^-^1)')
ylabel('I(Q)')

%% Optional individual plots

% loglog(Qrun31,Irun31,'b')
% loglog(Qrun32,Irun32,'r')
% loglog(Qrun41,Irun41,'g')
% loglog(Qrun42,Irun42,'y')
% loglog(Qrun131,Irun131,'c')
% plot(Qrun231,Irun231,'m')
% plot(Qrun241,Irun241,'k')
```

APPENDIX B: DATA COLLECTION/ANALYSIS

IGORPRO ANALYSIS

This will serve as a step-by-step process for taking the raw data (2D TIF images) and accompanying LOG files, and processing them to obtain background-subtracted scattering intensity curves (I v Q). NOT described here is how to create a new experiment file, since this depends on the detector and its distance from the sample, along with other parameters.

STEP 1: REDUCE THE DATA

-In IgorPro click on ‘‘BioCAT’’ in the toolbar.
-Now click ‘‘SAXS Data Reduction’’
-This will present a window describing the parameters. These should be fixed from the start and should remain unchanged for the experiment, with the exception of ‘‘Number of data runs to read’’; the number (X) here should correspond to the number of snapshots taken by the camera during each data collection run. Enter this and click ‘‘Continue’’. More parameters are presented that should again be fixed values. Just be sure that the ‘‘Q vector scale desired’’ is set to ‘‘log scale’’. Click ‘‘Continue’’.
-Choose the log file corresponding to the data you wish to reduce.
-Now choose the TIF image corresponding to the first snapshot.
-IgorPro will reduce it and the next (X-1) files.

STEP 2: PLOT THE REDUCED DATA

-Click ‘‘Plot Fits’’ on the toolbar.
-Now go to ‘‘LogLog Plots’’ -> ‘‘LogLogPlot’’
-Enter the number X in ‘‘Number of data sets to plot’’
-In the ‘‘Select the Q data’’ drop-down menu choose the file corresponding to TIF image you chose previously. It will have the same name as the TIF file but with a ‘q’ in front.
-Make sure to choose ‘‘yes’’ for ‘‘Log x-axis?’’ and ‘‘Log y-axis?’’
-Choose ‘‘Lines’’ for ‘‘Lines or Markers?’’
-Choose ‘‘Yes’’ for ‘‘Even Decades?’’
-Choose ‘‘Inverse Centimeters’’ for ‘‘Intensity Units’’
-Choose ‘‘Yes’’ for ‘‘Sequence of plots with same base’’

STEP 3: SUBTRACT THE BACKGROUND

-If the data collection has been successful there will be a clear distinction in the I(Q) curves between the background and the data. Double-click on the graph legend and delete it, its not necessary here. Then maximize the graph.
-Press CTRL-A to autoscale the graph
-Outliers can be removed by right-clicking on the lines and clicking ‘‘Remove’’
-You may also zoom in by clicking-and-dragging a selection box over an area, clicking in this area and clicking ‘‘expand’’.

-Before subtracting the background, both it and the data must be averaged. Click-and-drag a box to select the background curves, then hit CTRL-P.
-You will be prompted: ‘‘Weighting average by uncertainty?’’ (Yes), and ‘‘Name of averaged data file:’’ (The name will be automatic, but add a ‘b’ at the end for background).
-Repeat this averaging process for the data curves, but add a ‘d’ to the ‘‘Name of averaged data file’’
-To subtract the background hit CTRL-Q
-Again you will be prompted to choose the background and data files. Choose from the drop down menus the files you named in the previous step.
-The scaling factor should be 1; Click ‘‘Continue’’.
-Now go back to ‘‘Plot Fits’’ -> ‘‘LogLog Plots’’ -> ‘‘LogLogPlot’’
-This time the ‘‘Number of data sets to plot’’ is 1. Other inputs should be the same as in step 2.
-Choose the appropriate Q data from the drop down menu (it will have the same name as in previous steps but with ‘‘bsub’’ at the end)
-This will plot the background subtracted I(q) curve

STEP 4: SAVE THE DATA
-Below the graph, in a toolbar are a circle and square, next to which are the letters A and B respectively.
-Click-and-drag the circle and place it at the beginning of the curve.
-Click-and-drag the square and place it at the end of the curve.
-In the main toolbar, do ‘‘Argonne SAXS’’ -> ‘‘Save data from plot to file’’
-Choose ‘‘Yes’’ for ‘‘Change output file name?’’
-Enter an appropriate file name (should indicate protein conditions) in the text box below (the more abbreviated the better) and
click ‘‘Continue’’.
-Save the file to a location of your choice.

Note: It is useful to catalogue this graph for future reference. To do so, copy and paste the graph to a word document.

STEP 5: GUINIER ANALYSIS
-Do: ‘‘Plot fits’’ -> ‘‘Special plots’’ -> ‘‘Make Gunier plot’’
-For ‘‘Enter the data you wish to plot’’ choose the ‘bsub’ file you plotted in step 3
-For ‘‘Lines?’’ choose ‘‘Markers’’
-For ‘‘Type of Guinier Plot’’ choose ‘‘Standard Compact Object’’
-For ‘‘How Many Waves?’’ choose 1
-For ‘‘Intensity Units’’ choose ‘‘Inverse Centimeters’’
-Click Continue to get your Guinier plot
-Zoom in on straight-line portion of the beginning of the Guinier plot
-The idea here is to isolate the linear relationship at the start of the curve; do this by placing the circle and square at the beginning and end of this straight-line portion.
-Click ‘‘Plot fits’’ -> ‘‘Special plots’’ -> ‘‘Perform Guinier fit’’

-The coefficients of the chosen line will be displayed. The thing to note here is that the Qmax*Rg value should be between 1 and 1.3
-Record the Rg and again save a copy of the graph for future reference.

STEP 6: CREATE A FOLDER
Each individual .dat file resulting from step 4 should be placed in its own folder PRIOR to the GNOM-DAMMIN analysis. This isnt essential, but its a good thing to do to stop files from cluttering up later on

THE GNOM PROCEDURE

-Select data file: Input chosen .dat file from designated folder
-Name output file: Identical to input file but with .out extension instead of .dat. Also indicate the Rmax chosen (see below). e.g. if Rmax = 64A, call output file filename_64A.out
-Rmax for evaluating p(r): Maximum diameter of particle
-Take Rg value from IgorPro analysis (slope of Gunier plot)
-Get corresponding radius by R = sqrt(5/3)Rg for sphere, or Rg = 1/5(a^2 + b^2 + c^2), for ellipsoid, where a,b,c are the semi-axes lengths.
-Rmax = 2R for sphere, or 2a for ellipsoid (assuming a is longest axis)
-GNOM output, i.e. p(r) curve should be smooth, decaying to zero, with low error and no oscillations.

If p(r) curve is bad, rerun GNOM procedure and choose different Rmax (perhaps less).
Do this by answering ‘‘Next Data Set?’’ as ‘‘yes’’ and renaming output file according to new Rmax. Repeat this procedure until the output p(r) curve meets the desired characteristics, the resulting Rg is close to that of the IgorPro analysis, and the error on this Rg is low.

NOTE: There may be datasets that incur Fourier transform sampling problems when processed through GNOM. These are usually characterized by a significant dip near the beginning of the p(r) curve, and large oscillations at higher Q. In such cases, reducing the Rmax may result in a better curve, but this Rmax typically ends up being smaller than the actual particle diameter. As one might imagine, this problem is typical of larger particles. This can also be overcome by significantly reducing the ‘‘number of points in real space’’ input, but when this is done it usually results in a loss of detail and significant error. Alternatively, one can omit high Q data points (also a user input in GNOM), and a low -resolution reconstruction can be carried out instead.

-Once the p(r) curve has been produced, record the Rg and error from the .out file
-For every other input, hit enter.

THE DAMMIN PROCEDURE

-Mode: Choose ‘‘fast’’ mode
-Log file name: Name the file such that it matches the .out filename, but also indicate the shape and mode (e.g. for a fast spherical reconstruction, I usually add FS in front of the Rmax value), and use a .log extension. Note: There’s a limit on how many characters the log filename can have, you might have to get creative with abbreviations.
-Enter project description: This should include all details of the dataset, i.e. protein_concentration_temperature_shape_Rmax_mode.
-For every other input, hit enter.
-The simulated annealing procedure will take a half hour to an hour in fast mode. Once it is finished, there will be multiple files in the folder with the GNOM .out file. The one of most interest is that with the -1.pdb extension. Open this with a text application, such as a Notepad, or TextEdit. Before the columns of data there are some parameters. One of these will be radius of gyration. Record this value. It should be close to the Rg from Guinier analysis and the p(r) curve.

THE FOXS PROCEDURE

-Go online to the FOXS server, at http://modbase.compbio.ucsf.edu/foxs/index.html
-Here, upload the -1.pdb and original .dat file (the input for GNOM) using the ‘‘choose file’’ option. The ‘‘Experimental profile’’ is where the .dat file should be uploaded.
-Upload both and click ‘‘Submit form’’.
-This will show 2 graphs. On the left, the original experimental data and on the right, the data with the scattering profile the .pdb gives (this is a unique solution).
-A good match and a low chi-squared is an indicator that the Fourier transform and simulated annealing procedure has found a good solution based on the experimental data.
-Click on ‘‘Experimental profile fit file’’. This links to an html of the data, which should be saved to a .dat file in the same folder as the other files. Identify it as the FOXS data file.
-If the graphs dont match and there is a large chi-squared, the DAMMIN portion of this analysis should be repeated, until a satisfying fit is reached.
-Once FOXS gives a good match, you can rerun DAMMIN in slow mode (takes 6-8 hours usually)

THE PYMOL PROCEDURE

-Open the .pdb file in PYMOL.
-There are a number of tasks one can carry out in PYMOL, but one of the easiest is distance measurement. Click on the ‘‘Wizard’’ tab, then ‘‘Measurement’’. Now click on two atoms to get the distance between them. At the very least this is useful for putting a scale on the image.
-To the right of the image is a column with the filename. To the right of this one can click

the letters A,S,H,L,C. Clicking on ‘S’ allows one to change the display. ‘‘Mesh’’ or ‘‘Surface’’ are good for studying overall structure, and ‘‘Spheres’’ is also good.

-Moving structures in PYMOL: You can double-click on a structure to bring up a list of options, one of which is ‘‘drag object coordinates’’. Click this. Now, hold shift and click-and-hold the left mouse button; moving the mouse now rotates the structure. You can also hold shift and click-and-hold the right mouse key; moving the mouse up and down moves the structure along the axis that goes into the monitor. Also, by holding ALT and the left mouse button and moving the mouse you can move the structure with respect to the background. This is all useful for docking reconstructions with crystal structures.

APPENDIX C: GNOM OUTPUT

```
          ####    G N O M   ---   Version 4.5a revised 09/02/02     ####

                                         02-Oct-2012   11:15:16
          ===    Run No   1   ===
 Run title:  Q_data I_data Err_data


   *******    Input file(s) : M00C2.dat
          Condition P(rmin) = 0 is used.
          Condition P(rmax) = 0 is used.

         Highest ALPHA is found to be   0.5349E+02

  The measure of inconsistency AN1 equals to    0.2315E+00
     Alpha    Discrp  Oscill  Stabil  Sysdev  Positv  Valcen    Total
  0.5349E+05 73.4101  1.0624  0.6736  0.0336  1.0000  0.8922  0.39845
  0.1070E+05 37.1732  1.0285  0.3404  0.0336  1.0000  0.9044  0.40204
  0.2140E+04 15.7305  1.0117  0.1479  0.0336  1.0000  0.9215  0.46143
  0.4279E+03  6.5894  1.0667  0.0676  0.0336  1.0000  0.9394  0.59723
  0.8559E+02  2.1754  1.1592  0.0318  0.0336  1.0000  0.9541  0.64718
  0.1712E+02  0.8809  1.2037  0.0111  0.3025  1.0000  0.9600  0.71455
  0.3423E+01  0.6722  1.2267  0.0144  0.3361  1.0000  0.9629  0.73385
  0.6847E+00  0.5972  1.2644  0.0107  0.4034  1.0000  0.9648  0.71943
  0.1369E+00  0.5779  1.3096  0.0203  0.3697  1.0000  0.9642  0.70037
  0.2739E-01  0.5656  1.4537  0.0297  0.4034  1.0000  0.9603  0.64599
  0.5478E-02  0.5582  1.8135  0.0795  0.4706  0.9999  0.9551  0.45513
  0.1096E-02  0.5455  3.6986  0.2728  0.4706  0.9952  0.9314  0.22992
  0.2191E-03  0.5257  7.0315  0.3025  0.5042  0.9653  0.8467  0.17627
  0.4382E-04  0.5181  8.3991  0.1068  0.5042  0.9316  0.6339  0.23120
  0.8764E-05  0.5170  9.6876  0.2755  0.4706  0.9119  0.6015  0.10735
  0.1753E-05  0.5161 12.6464  0.5124  0.4370  0.8538  0.6088  0.07616
  0.3506E-06  0.5153 15.2193  0.5629  0.4370  0.7959  0.3904  0.06166
  0.7011E-07  0.5143 14.7138  1.0763  0.4706  0.7550  0.3577  0.05810
  0.1402E-07  0.5107 12.5461  0.8587  0.4706  0.7146  0.3544  0.05626
  0.2805E-08  0.5035 12.7197  0.5003  0.5042  0.7026  0.3545  0.05444
  0.5609E-09  0.4972 15.6705  0.6759  0.4874  0.6823  0.3648  0.05283
  0.1122E-09  0.4827 16.5323  0.9699  0.5378  0.7330  0.3515  0.00000

         ***  Golden section search to maximize estimate ***
```

```
   Alpha    Discrp  Oscill  Stabil  Sysdev  Positv  Valcen   Total
0.6331E+01  0.7234  1.2170  0.0122  0.3361  1.0000  0.9619  0.73677
0.9257E+01  0.7657  1.2120  0.0109  0.3025  1.0000  0.9612  0.73474
0.5006E+01  0.7020  1.2204  0.0132  0.3361  1.0000  0.9623  0.73627
0.7319E+01  0.7381  1.2151  0.0117  0.3697  1.0000  0.9616  0.73650
0.5788E+01  0.7149  1.2182  0.0126  0.3361  1.0000  0.9620  0.73669
0.6692E+01  0.7288  1.2162  0.0120  0.3361  1.0000  0.9618  0.73673
0.6117E+01  0.7201  1.2175  0.0124  0.3361  1.0000  0.9619  0.73676
0.6466E+01  0.7254  1.2167  0.0122  0.3361  1.0000  0.9618  0.73676
0.6248E+01  0.7221  1.2172  0.0123  0.3361  1.0000  0.9619  0.73677
0.6382E+01  0.7241  1.2169  0.0122  0.3361  1.0000  0.9618  0.73677
0.6299E+01  0.7229  1.2171  0.0123  0.3361  1.0000  0.9619  0.73677
0.6280E+01  0.7226  1.2171  0.0123  0.3361  1.0000  0.9619  0.73677
0.6311E+01  0.7231  1.2170  0.0123  0.3361  1.0000  0.9619  0.73677
0.6292E+01  0.7228  1.2171  0.0123  0.3361  1.0000  0.9619  0.73677
0.6299E+01  0.7229  1.2171  0.0123  0.3361  1.0000  0.9619  0.73677



           ####           Final results           ####

Parameter    DISCRP    OSCILL    STABIL    SYSDEV    POSITV    VALCEN
Weight        1.000     3.000     3.000     3.000     1.000     1.000
Sigma         0.300     0.600     0.120     0.120     0.120     0.120
Ideal         0.700     1.100     0.000     1.000     1.000     0.950
Current       0.723     1.217     0.012     0.336     1.000     0.962
               - - - - - - - - - - - - - - - - - - - - - - - - -
Estimate      0.994     0.963     0.990     0.000     1.000     0.990

 Angular   range    :     from    0.0108   to    0.5577
 Real space range   :     from      0.00   to     35.00

 Highest ALPHA (theor) :   0.535E+05                JOB = 0
 Current ALPHA         :   0.630E+01   Rg :  0.130E+02   I(0) :   0.122E+03

          Total  estimate : 0.737  which is  A REASONABLE  solution

     S          J EXP       ERROR       J REG       I REG

 0.0000E+00                                      0.1216E+03
 0.3853E-03                                      0.1216E+03
 0.7707E-03                                      0.1216E+03
 0.1156E-02                                      0.1216E+03
```

```
0.1541E-02                                      0.1216E+03
0.1927E-02                                      0.1216E+03
0.2312E-02                                      0.1216E+03
0.2697E-02                                      0.1216E+03
0.3083E-02                                      0.1216E+03
0.3468E-02                                      0.1215E+03
0.3853E-02                                      0.1215E+03
0.4239E-02                                      0.1215E+03
0.4624E-02                                      0.1215E+03
0.5010E-02                                      0.1215E+03
0.5395E-02                                      0.1214E+03
0.5780E-02                                      0.1214E+03
0.6166E-02                                      0.1214E+03
0.6551E-02                                      0.1213E+03
0.6936E-02                                      0.1213E+03
0.7322E-02                                      0.1213E+03
0.7707E-02                                      0.1212E+03
0.8092E-02                                      0.1212E+03
0.8478E-02                                      0.1211E+03
0.8863E-02                                      0.1211E+03
0.9248E-02                                      0.1210E+03
0.9634E-02                                      0.1210E+03
0.1002E-01                                      0.1209E+03
0.1040E-01                                      0.1209E+03
0.1079E-01  0.1348E+03  0.8943E+01  0.1208E+03  0.1208E+03
0.1115E-01  0.1345E+03  0.7727E+01  0.1208E+03  0.1208E+03
0.1153E-01  0.1316E+03  0.5832E+01  0.1207E+03  0.1207E+03
0.1192E-01  0.1310E+03  0.4670E+01  0.1207E+03  0.1207E+03
0.1232E-01  0.1294E+03  0.4076E+01  0.1206E+03  0.1206E+03
0.1274E-01  0.1286E+03  0.3531E+01  0.1205E+03  0.1205E+03
0.1317E-01  0.1276E+03  0.3460E+01  0.1204E+03  0.1204E+03
0.1361E-01  0.1261E+03  0.3146E+01  0.1204E+03  0.1204E+03
0.1407E-01  0.1255E+03  0.2816E+01  0.1203E+03  0.1203E+03
0.1454E-01  0.1259E+03  0.2857E+01  0.1202E+03  0.1202E+03
0.1504E-01  0.1257E+03  0.2467E+01  0.1201E+03  0.1201E+03
0.1554E-01  0.1247E+03  0.2353E+01  0.1200E+03  0.1200E+03
0.1607E-01  0.1240E+03  0.2304E+01  0.1199E+03  0.1199E+03
0.1661E-01  0.1235E+03  0.1994E+01  0.1197E+03  0.1197E+03
0.1717E-01  0.1241E+03  0.1998E+01  0.1196E+03  0.1196E+03
0.1775E-01  0.1235E+03  0.1850E+01  0.1195E+03  0.1195E+03
0.1835E-01  0.1222E+03  0.1727E+01  0.1193E+03  0.1193E+03
0.1897E-01  0.1213E+03  0.1629E+01  0.1192E+03  0.1192E+03
0.1961E-01  0.1209E+03  0.1579E+01  0.1190E+03  0.1190E+03
```

```
0.2027E-01  0.1210E+03  0.1498E+01  0.1188E+03  0.1188E+03
0.2095E-01  0.1208E+03  0.1367E+01  0.1186E+03  0.1186E+03
0.2166E-01  0.1204E+03  0.1319E+01  0.1184E+03  0.1184E+03
0.2239E-01  0.1198E+03  0.1235E+01  0.1182E+03  0.1182E+03
0.2314E-01  0.1199E+03  0.1220E+01  0.1180E+03  0.1180E+03
0.2392E-01  0.1194E+03  0.1168E+01  0.1178E+03  0.1178E+03
0.2473E-01  0.1189E+03  0.1083E+01  0.1175E+03  0.1175E+03
0.2557E-01  0.1188E+03  0.1043E+01  0.1172E+03  0.1172E+03
0.2643E-01  0.1187E+03  0.1018E+01  0.1169E+03  0.1169E+03
0.2732E-01  0.1180E+03  0.9488E+00  0.1166E+03  0.1166E+03
0.2824E-01  0.1175E+03  0.9523E+00  0.1163E+03  0.1163E+03
0.2919E-01  0.1166E+03  0.8364E+00  0.1159E+03  0.1159E+03
0.3018E-01  0.1165E+03  0.8716E+00  0.1155E+03  0.1155E+03
0.3120E-01  0.1164E+03  0.8159E+00  0.1151E+03  0.1151E+03
0.3225E-01  0.1159E+03  0.7697E+00  0.1147E+03  0.1147E+03
0.3334E-01  0.1152E+03  0.7624E+00  0.1142E+03  0.1142E+03
0.3446E-01  0.1147E+03  0.7157E+00  0.1137E+03  0.1137E+03
0.3563E-01  0.1144E+03  0.7090E+00  0.1132E+03  0.1132E+03
0.3683E-01  0.1137E+03  0.6818E+00  0.1126E+03  0.1126E+03
0.3807E-01  0.1128E+03  0.6426E+00  0.1120E+03  0.1120E+03
0.3935E-01  0.1125E+03  0.6179E+00  0.1114E+03  0.1114E+03
0.4068E-01  0.1115E+03  0.6145E+00  0.1107E+03  0.1107E+03
0.4205E-01  0.1109E+03  0.5978E+00  0.1100E+03  0.1100E+03
0.4347E-01  0.1098E+03  0.5892E+00  0.1093E+03  0.1093E+03
0.4494E-01  0.1087E+03  0.5663E+00  0.1085E+03  0.1085E+03
0.4645E-01  0.1082E+03  0.5596E+00  0.1076E+03  0.1076E+03
0.4802E-01  0.1072E+03  0.5467E+00  0.1067E+03  0.1067E+03
0.4964E-01  0.1062E+03  0.5324E+00  0.1057E+03  0.1057E+03
0.5132E-01  0.1050E+03  0.5147E+00  0.1047E+03  0.1047E+03
0.5305E-01  0.1038E+03  0.5012E+00  0.1036E+03  0.1036E+03
0.5484E-01  0.1028E+03  0.4967E+00  0.1025E+03  0.1025E+03
0.5669E-01  0.1015E+03  0.4810E+00  0.1013E+03  0.1013E+03
0.5860E-01  0.9988E+02  0.4650E+00  0.1000E+03  0.1000E+03
0.6058E-01  0.9873E+02  0.4530E+00  0.9867E+02  0.9867E+02
0.6262E-01  0.9733E+02  0.4399E+00  0.9725E+02  0.9725E+02
0.6473E-01  0.9570E+02  0.4263E+00  0.9575E+02  0.9575E+02
0.6692E-01  0.9399E+02  0.4184E+00  0.9417E+02  0.9417E+02
0.6917E-01  0.9235E+02  0.4010E+00  0.9250E+02  0.9250E+02
0.7151E-01  0.9055E+02  0.3908E+00  0.9075E+02  0.9075E+02
0.7392E-01  0.8869E+02  0.3807E+00  0.8892E+02  0.8892E+02
0.7641E-01  0.8694E+02  0.3685E+00  0.8699E+02  0.8699E+02
0.7899E-01  0.8496E+02  0.3596E+00  0.8498E+02  0.8498E+02
0.8165E-01  0.8279E+02  0.3466E+00  0.8287E+02  0.8287E+02
```

```
0.8441E-01  0.8044E+02  0.3347E+00  0.8066E+02  0.8066E+02
0.8726E-01  0.7830E+02  0.3289E+00  0.7837E+02  0.7837E+02
0.9020E-01  0.7567E+02  0.3171E+00  0.7598E+02  0.7598E+02
0.9324E-01  0.7313E+02  0.3077E+00  0.7350E+02  0.7350E+02
0.9639E-01  0.7080E+02  0.3001E+00  0.7093E+02  0.7093E+02
0.9964E-01  0.6818E+02  0.2887E+00  0.6827E+02  0.6827E+02
0.1030E+00  0.6514E+02  0.2821E+00  0.6552E+02  0.6552E+02
0.1065E+00  0.6249E+02  0.2733E+00  0.6270E+02  0.6270E+02
0.1101E+00  0.5949E+02  0.2656E+00  0.5980E+02  0.5980E+02
0.1138E+00  0.5683E+02  0.2567E+00  0.5684E+02  0.5684E+02
0.1176E+00  0.5371E+02  0.2506E+00  0.5382E+02  0.5382E+02
0.1216E+00  0.5069E+02  0.2410E+00  0.5074E+02  0.5074E+02
0.1257E+00  0.4740E+02  0.2341E+00  0.4764E+02  0.4764E+02
0.1299E+00  0.4453E+02  0.2266E+00  0.4450E+02  0.4450E+02
0.1343E+00  0.4136E+02  0.2205E+00  0.4136E+02  0.4136E+02
0.1388E+00  0.3821E+02  0.2127E+00  0.3822E+02  0.3822E+02
0.1435E+00  0.3518E+02  0.2070E+00  0.3510E+02  0.3510E+02
0.1483E+00  0.3205E+02  0.1994E+00  0.3202E+02  0.3202E+02
0.1534E+00  0.2914E+02  0.1942E+00  0.2900E+02  0.2900E+02
0.1585E+00  0.2618E+02  0.1879E+00  0.2606E+02  0.2606E+02
0.1639E+00  0.2351E+02  0.1843E+00  0.2322E+02  0.2322E+02
0.1694E+00  0.2066E+02  0.1918E+00  0.2050E+02  0.2050E+02
0.1751E+00  0.1799E+02  0.1951E+00  0.1792E+02  0.1792E+02
0.1810E+00  0.1564E+02  0.1982E+00  0.1549E+02  0.1549E+02
0.1871E+00  0.1345E+02  0.1999E+00  0.1325E+02  0.1325E+02
0.1934E+00  0.1110E+02  0.2011E+00  0.1120E+02  0.1120E+02
0.1999E+00  0.9251E+01  0.2030E+00  0.9354E+01  0.9354E+01
0.2067E+00  0.7651E+01  0.2024E+00  0.7729E+01  0.7729E+01
0.2136E+00  0.6240E+01  0.1993E+00  0.6328E+01  0.6328E+01
0.2208E+00  0.5038E+01  0.1973E+00  0.5154E+01  0.5154E+01
0.2283E+00  0.4149E+01  0.1945E+00  0.4204E+01  0.4204E+01
0.2360E+00  0.3543E+01  0.1925E+00  0.3473E+01  0.3473E+01
0.2439E+00  0.2862E+01  0.1907E+00  0.2946E+01  0.2946E+01
0.2521E+00  0.2586E+01  0.1854E+00  0.2608E+01  0.2608E+01
0.2606E+00  0.2450E+01  0.1789E+00  0.2437E+01  0.2437E+01
0.2694E+00  0.2266E+01  0.1728E+00  0.2408E+01  0.2408E+01
0.2785E+00  0.2270E+01  0.1676E+00  0.2493E+01  0.2493E+01
0.2879E+00  0.2595E+01  0.1622E+00  0.2664E+01  0.2664E+01
0.2976E+00  0.2847E+01  0.1576E+00  0.2890E+01  0.2890E+01
0.3076E+00  0.3000E+01  0.1529E+00  0.3145E+01  0.3145E+01
0.3180E+00  0.3337E+01  0.1487E+00  0.3404E+01  0.3404E+01
0.3287E+00  0.3649E+01  0.1449E+00  0.3645E+01  0.3645E+01
0.3397E+00  0.3762E+01  0.1408E+00  0.3852E+01  0.3852E+01
```

```
0.3511E+00  0.4096E+01  0.1367E+00  0.4016E+01  0.4016E+01
0.3630E+00  0.4290E+01  0.1334E+00  0.4131E+01  0.4131E+01
0.3752E+00  0.4266E+01  0.1301E+00  0.4196E+01  0.4196E+01
0.3878E+00  0.4358E+01  0.1272E+00  0.4215E+01  0.4215E+01
0.4008E+00  0.4264E+01  0.1249E+00  0.4193E+01  0.4193E+01
0.4143E+00  0.4090E+01  0.1227E+00  0.4136E+01  0.4136E+01
0.4282E+00  0.3974E+01  0.1207E+00  0.4052E+01  0.4052E+01
0.4426E+00  0.3971E+01  0.1189E+00  0.3946E+01  0.3946E+01
0.4575E+00  0.3713E+01  0.1178E+00  0.3822E+01  0.3822E+01
0.4729E+00  0.3757E+01  0.1170E+00  0.3683E+01  0.3683E+01
0.4887E+00  0.3518E+01  0.1172E+00  0.3534E+01  0.3534E+01
0.5051E+00  0.3327E+01  0.1186E+00  0.3378E+01  0.3378E+01
0.5221E+00  0.3306E+01  0.1268E+00  0.3219E+01  0.3219E+01
0.5396E+00  0.3260E+01  0.1687E+00  0.3061E+01  0.3061E+01
0.5577E+00  0.4107E+01  0.1526E+02  0.2908E+01  0.2908E+01

         Distance distribution  function of particle



     R          P(R)      ERROR

0.0000E+00  0.0000E+00  0.0000E+00
0.4375E+00  0.3027E-01  0.8502E-03
0.8750E+00  0.5620E-01  0.1415E-02
0.1312E+01  0.7794E-01  0.1711E-02
0.1750E+01  0.9575E-01  0.1763E-02
0.2188E+01  0.1100E+00  0.1608E-02
0.2625E+01  0.1210E+00  0.1292E-02
0.3062E+01  0.1294E+00  0.8838E-03
0.3500E+01  0.1357E+00  0.5496E-03
0.3938E+01  0.1403E+00  0.6841E-03
0.4375E+01  0.1440E+00  0.1146E-02
0.4812E+01  0.1473E+00  0.1635E-02
0.5250E+01  0.1506E+00  0.2060E-02
0.5688E+01  0.1546E+00  0.2378E-02
0.6125E+01  0.1597E+00  0.2567E-02
0.6562E+01  0.1662E+00  0.2614E-02
0.7000E+01  0.1745E+00  0.2520E-02
0.7438E+01  0.1847E+00  0.2298E-02
0.7875E+01  0.1970E+00  0.1975E-02
0.8312E+01  0.2115E+00  0.1602E-02
0.8750E+01  0.2281E+00  0.1278E-02
0.9188E+01  0.2466E+00  0.1168E-02
```

```
0.9625E+01  0.2668E+00  0.1368E-02
0.1006E+02  0.2886E+00  0.1762E-02
0.1050E+02  0.3115E+00  0.2200E-02
0.1094E+02  0.3353E+00  0.2598E-02
0.1138E+02  0.3595E+00  0.2910E-02
0.1181E+02  0.3838E+00  0.3108E-02
0.1225E+02  0.4077E+00  0.3181E-02
0.1269E+02  0.4309E+00  0.3129E-02
0.1312E+02  0.4530E+00  0.2963E-02
0.1356E+02  0.4737E+00  0.2712E-02
0.1400E+02  0.4927E+00  0.2422E-02
0.1444E+02  0.5098E+00  0.2160E-02
0.1488E+02  0.5248E+00  0.2012E-02
0.1531E+02  0.5375E+00  0.2042E-02
0.1575E+02  0.5478E+00  0.2244E-02
0.1619E+02  0.5556E+00  0.2544E-02
0.1662E+02  0.5610E+00  0.2861E-02
0.1706E+02  0.5640E+00  0.3136E-02
0.1750E+02  0.5646E+00  0.3327E-02
0.1794E+02  0.5628E+00  0.3413E-02
0.1838E+02  0.5588E+00  0.3387E-02
0.1881E+02  0.5527E+00  0.3257E-02
0.1925E+02  0.5446E+00  0.3045E-02
0.1969E+02  0.5345E+00  0.2792E-02
0.2012E+02  0.5227E+00  0.2555E-02
0.2056E+02  0.5092E+00  0.2409E-02
0.2100E+02  0.4941E+00  0.2413E-02
0.2144E+02  0.4777E+00  0.2577E-02
0.2188E+02  0.4600E+00  0.2854E-02
0.2231E+02  0.4411E+00  0.3174E-02
0.2275E+02  0.4212E+00  0.3477E-02
0.2319E+02  0.4004E+00  0.3719E-02
0.2362E+02  0.3790E+00  0.3874E-02
0.2406E+02  0.3569E+00  0.3930E-02
0.2450E+02  0.3345E+00  0.3890E-02
0.2494E+02  0.3119E+00  0.3770E-02
0.2538E+02  0.2892E+00  0.3598E-02
0.2581E+02  0.2668E+00  0.3412E-02
0.2625E+02  0.2448E+00  0.3255E-02
0.2669E+02  0.2233E+00  0.3165E-02
0.2712E+02  0.2026E+00  0.3157E-02
0.2756E+02  0.1829E+00  0.3221E-02
0.2800E+02  0.1643E+00  0.3325E-02
```

```
0.2844E+02  0.1470E+00  0.3429E-02
0.2888E+02  0.1311E+00  0.3500E-02
0.2931E+02  0.1165E+00  0.3517E-02
0.2975E+02  0.1034E+00  0.3481E-02
0.3019E+02  0.9177E-01  0.3410E-02
0.3062E+02  0.8144E-01  0.3338E-02
0.3106E+02  0.7231E-01  0.3306E-02
0.3150E+02  0.6420E-01  0.3341E-02
0.3194E+02  0.5688E-01  0.3440E-02
0.3238E+02  0.5007E-01  0.3560E-02
0.3281E+02  0.4345E-01  0.3626E-02
0.3325E+02  0.3665E-01  0.3549E-02
0.3369E+02  0.2931E-01  0.3237E-02
0.3412E+02  0.2102E-01  0.2598E-02
0.3456E+02  0.1138E-01  0.1545E-02
0.3500E+02  0.0000E+00  0.0000E+00
        Reciprocal space: Rg =   13.00     , I(0) =   0.1216E+03
   Real space: Rg =   13.00 +- 0.015  I(0) =   0.1216E+03 +-  0.1356E+00
```

APPENDIX D: DAMMIN OUTPUT

```
 Project description:   CytC M00C2 fast sphere recon with Rmax 35A
 j:  81 T: 0.254E-06 Suc:     4 Eva:  5912047 CPU:  0.844E+03 SqF: 0.0171
  Rf: 0.00228 Los:0.0477 Dis:0.0000 Per:  0.5604 Sca: 0.752E-05
 Computation mode ...................................... : Fast
 Number of particle atoms .............................. : 323
 Number of graphs ...................................... : 1
 Anisometry penalty .................................... : 0.0
 Point symmetry group .................................. : P1
 DAM packing radius .................................... : 1.300
 Average volume per atom ............................... : 12.44
 Total excluded DAM volume ............................. : 4017.
 Particle center:   1.2276   0.5514  -0.4326
 Radius of gyration .................................... : 12.98
 Maximum diameter ...................................... : 34.69
 DAM shape anisometry .................................. : 0.2282
 DAM non-prolateness ................................... : 0.0
 DAM non-oblateness .................................... : 1.424e-2
ATOM      1  CA   ASP    1        0.000   0.000 -12.869  1.00 20.00 0 2 201
ATOM      2  CA   ASP    1        0.000   2.600 -12.869  1.00 20.00 0 2 201
ATOM      3  CA   ASP    1        0.000   7.800 -12.869  1.00 20.00 0 2 201
ATOM      4  CA   ASP    1        0.000  10.400 -12.869  1.00 20.00 0 2 201
ATOM      5  CA   ASP    1        2.600  -2.600 -12.869  1.00 20.00 0 2 201
ATOM      6  CA   ASP    1        2.600   0.000 -12.869  1.00 20.00 0 2 201
ATOM      7  CA   ASP    1        2.600   2.600 -12.869  1.00 20.00 0 2 201
ATOM      8  CA   ASP    1        5.200  -5.200 -12.869  1.00 20.00 0 2 201
ATOM      9  CA   ASP    1        5.200  -2.600 -12.869  1.00 20.00 0 2 201
ATOM     10  CA   ASP    2        5.200   0.000 -12.869  1.00 20.00 0 2 201
ATOM     11  CA   ASP    2       -1.300  -1.300 -11.031  1.00 20.00 0 2 201
ATOM     12  CA   ASP    2       -1.300   1.300 -11.031  1.00 20.00 0 2 201
ATOM     13  CA   ASP    2       -1.300   3.900 -11.031  1.00 20.00 0 2 201
ATOM     14  CA   ASP    2       -1.300   6.500 -11.031  1.00 20.00 0 2 201
ATOM     15  CA   ASP    2       -1.300   9.100 -11.031  1.00 20.00 0 2 201
ATOM     16  CA   ASP    2       -1.300  11.700 -11.031  1.00 20.00 0 2 201
ATOM     17  CA   ASP    2        1.300  -3.900 -11.031  1.00 20.00 0 2 201
ATOM     18  CA   ASP    2        1.300  -1.300 -11.031  1.00 20.00 1 2 201
ATOM     19  CA   ASP    2        1.300   1.300 -11.031  1.00 20.00 1 2 201
ATOM     20  CA   ASP    3        1.300   3.900 -11.031  1.00 20.00 0 2 201
ATOM     21  CA   ASP    3        1.300   6.500 -11.031  1.00 20.00 0 2 201
ATOM     22  CA   ASP    3        1.300   9.100 -11.031  1.00 20.00 0 2 201
ATOM     23  CA   ASP    3        1.300  11.700 -11.031  1.00 20.00 0 2 201
```

```
ATOM     24  CA  ASP    3        3.900  -6.500 -11.031  1.00 20.00 0 2 201
ATOM     25  CA  ASP    3        3.900  -3.900 -11.031  1.00 20.00 0 2 201
ATOM     26  CA  ASP    3        3.900  -1.300 -11.031  1.00 20.00 1 2 201
ATOM     27  CA  ASP    3        3.900   1.300 -11.031  1.00 20.00 0 2 201
ATOM     28  CA  ASP    3        3.900   3.900 -11.031  1.00 20.00 0 2 201
ATOM     29  CA  ASP    3        3.900   9.100 -11.031  1.00 20.00 0 2 201
ATOM     30  CA  ASP    4        3.900  11.700 -11.031  1.00 20.00 0 2 201
ATOM     31  CA  ASP    4        6.500  -6.500 -11.031  1.00 20.00 0 2 201
ATOM     32  CA  ASP    4        6.500  -3.900 -11.031  1.00 20.00 0 2 201
ATOM     33  CA  ASP    4        6.500  -1.300 -11.031  1.00 20.00 0 2 201
ATOM     34  CA  ASP    4        6.500   1.300 -11.031  1.00 20.00 0 2 201
ATOM     35  CA  ASP    4       -2.600 -13.000  -9.192  1.00 20.00 0 2 201
ATOM     36  CA  ASP    4       -2.600  10.400  -9.192  1.00 20.00 0 2 201
ATOM     37  CA  ASP    4       -2.600  13.000  -9.192  1.00 20.00 0 2 201
ATOM     38  CA  ASP    4        0.000  -2.600  -9.192  1.00 20.00 0 2 201
ATOM     39  CA  ASP    4        0.000   0.000  -9.192  1.00 20.00 0 2 201
ATOM     40  CA  ASP    5        0.000   2.600  -9.192  1.00 20.00 0 2 201
ATOM     41  CA  ASP    5        0.000   7.800  -9.192  1.00 20.00 0 2 201
ATOM     42  CA  ASP    5        0.000  10.400  -9.192  1.00 20.00 1 2 201
ATOM     43  CA  ASP    5        0.000  13.000  -9.192  1.00 20.00 1 2 201
ATOM     44  CA  ASP    5        2.600  -5.200  -9.192  1.00 20.00 0 2 201
ATOM     45  CA  ASP    5        2.600  -2.600  -9.192  1.00 20.00 0 2 201
ATOM     46  CA  ASP    5        2.600   0.000  -9.192  1.00 20.00 0 2 201
ATOM     47  CA  ASP    5        2.600   2.600  -9.192  1.00 20.00 0 2 201
ATOM     48  CA  ASP    5        2.600   7.800  -9.192  1.00 20.00 0 2 201
ATOM     49  CA  ASP    5        2.600  10.400  -9.192  1.00 20.00 1 2 201
ATOM     50  CA  ASP    6        2.600  13.000  -9.192  1.00 20.00 0 2 201
ATOM     51  CA  ASP    6        5.200 -10.400  -9.192  1.00 20.00 0 2 201
ATOM     52  CA  ASP    6        5.200  -7.800  -9.192  1.00 20.00 0 2 201
ATOM     53  CA  ASP    6        5.200  -5.200  -9.192  1.00 20.00 0 2 201
ATOM     54  CA  ASP    6        5.200  -2.600  -9.192  1.00 20.00 0 2 201
ATOM     55  CA  ASP    6        5.200   0.000  -9.192  1.00 20.00 0 2 201
ATOM     56  CA  ASP    6        5.200   2.600  -9.192  1.00 20.00 0 2 201
ATOM     57  CA  ASP    6        5.200   7.800  -9.192  1.00 20.00 0 2 201
ATOM     58  CA  ASP    6        5.200  10.400  -9.192  1.00 20.00 0 2 201
ATOM     59  CA  ASP    6       -3.900 -14.300  -7.354  1.00 20.00 0 2 201
ATOM     60  CA  ASP    7       -3.900 -11.700  -7.354  1.00 20.00 0 2 201
ATOM     61  CA  ASP    7       -3.900  11.700  -7.354  1.00 20.00 0 2 201
ATOM     62  CA  ASP    7       -3.900  14.300  -7.354  1.00 20.00 0 2 201
ATOM     63  CA  ASP    7       -1.300 -14.300  -7.354  1.00 20.00 1 2 201
ATOM     64  CA  ASP    7       -1.300 -11.700  -7.354  1.00 20.00 0 2 201
ATOM     65  CA  ASP    7       -1.300   9.100  -7.354  1.00 20.00 0 2 201
ATOM     66  CA  ASP    7       -1.300  11.700  -7.354  1.00 20.00 1 2 201
```

```
ATOM     67  CA  ASP     7        -1.300  14.300  -7.354  1.00 20.00 0 2 201
ATOM     68  CA  ASP     7         1.300 -14.300  -7.354  1.00 20.00 1 2 201
ATOM     69  CA  ASP     7         1.300 -11.700  -7.354  1.00 20.00 0 2 201
ATOM     70  CA  ASP     8         1.300  -3.900  -7.354  1.00 20.00 0 2 201
ATOM     71  CA  ASP     8         1.300  -1.300  -7.354  1.00 20.00 0 2 201
ATOM     72  CA  ASP     8         1.300   1.300  -7.354  1.00 20.00 0 2 201
ATOM     73  CA  ASP     8         1.300   9.100  -7.354  1.00 20.00 0 2 201
ATOM     74  CA  ASP     8         1.300  11.700  -7.354  1.00 20.00 1 2 201
ATOM     75  CA  ASP     8         1.300  14.300  -7.354  1.00 20.00 0 2 201
ATOM     76  CA  ASP     8         3.900 -14.300  -7.354  1.00 20.00 0 2 201
ATOM     77  CA  ASP     8         3.900 -11.700  -7.354  1.00 20.00 0 2 201
ATOM     78  CA  ASP     8         3.900  -9.100  -7.354  1.00 20.00 0 2 201
ATOM     79  CA  ASP     8         3.900  -3.900  -7.354  1.00 20.00 0 2 201
ATOM     80  CA  ASP     9         3.900  -1.300  -7.354  1.00 20.00 0 2 201
ATOM     81  CA  ASP     9         3.900   1.300  -7.354  1.00 20.00 0 2 201
ATOM     82  CA  ASP     9         3.900   9.100  -7.354  1.00 20.00 0 2 201
ATOM     83  CA  ASP     9         3.900  11.700  -7.354  1.00 20.00 0 2 201
ATOM     84  CA  ASP     9         6.500 -14.300  -7.354  1.00 20.00 0 2 201
ATOM     85  CA  ASP     9         6.500 -11.700  -7.354  1.00 20.00 0 2 201
ATOM     86  CA  ASP     9         6.500  -9.100  -7.354  1.00 20.00 0 2 201
ATOM     87  CA  ASP     9         6.500   6.500  -7.354  1.00 20.00 0 2 201
ATOM     88  CA  ASP     9         6.500   9.100  -7.354  1.00 20.00 0 2 201
ATOM     89  CA  ASP     9        -2.600 -15.600  -5.515  1.00 20.00 0 2 201
ATOM     90  CA  ASP    10        -2.600 -13.000  -5.515  1.00 20.00 0 2 201
ATOM     91  CA  ASP    10        -2.600  10.400  -5.515  1.00 20.00 0 2 201
ATOM     92  CA  ASP    10        -2.600  13.000  -5.515  1.00 20.00 0 2 201
ATOM     93  CA  ASP    10        -2.600  15.600  -5.515  1.00 20.00 0 2 201
ATOM     94  CA  ASP    10         0.000 -15.600  -5.515  1.00 20.00 1 2 201
ATOM     95  CA  ASP    10         0.000 -13.000  -5.515  1.00 20.00 0 2 201
ATOM     96  CA  ASP    10         0.000  -2.600  -5.515  1.00 20.00 0 2 201
ATOM     97  CA  ASP    10         0.000  10.400  -5.515  1.00 20.00 0 2 201
ATOM     98  CA  ASP    10         0.000  13.000  -5.515  1.00 20.00 0 2 201
ATOM     99  CA  ASP    10         0.000  15.600  -5.515  1.00 20.00 0 2 201
ATOM    100  CA  ASP    11         2.600 -15.600  -5.515  1.00 20.00 1 2 201
ATOM    101  CA  ASP    11         2.600 -13.000  -5.515  1.00 20.00 0 2 201
ATOM    102  CA  ASP    11         2.600  -2.600  -5.515  1.00 20.00 0 2 201
ATOM    103  CA  ASP    11         2.600  10.400  -5.515  1.00 20.00 0 2 201
ATOM    104  CA  ASP    11         2.600  13.000  -5.515  1.00 20.00 0 2 201
ATOM    105  CA  ASP    11         5.200 -15.600  -5.515  1.00 20.00 0 2 201
ATOM    106  CA  ASP    11         5.200 -13.000  -5.515  1.00 20.00 0 2 201
ATOM    107  CA  ASP    11         5.200 -10.400  -5.515  1.00 20.00 0 2 201
ATOM    108  CA  ASP    11         7.800   5.200  -5.515  1.00 20.00 0 2 201
ATOM    109  CA  ASP    11         7.800   7.800  -5.515  1.00 20.00 0 2 201
```

```
ATOM    110  CA  ASP   12      -11.700   1.300  -3.677  1.00 20.00 0 2 201
ATOM    111  CA  ASP   12      -11.700   3.900  -3.677  1.00 20.00 0 2 201
ATOM    112  CA  ASP   12      -11.700   6.500  -3.677  1.00 20.00 0 2 201
ATOM    113  CA  ASP   12       -9.100   1.300  -3.677  1.00 20.00 0 2 201
ATOM    114  CA  ASP   12       -9.100   3.900  -3.677  1.00 20.00 0 2 201
ATOM    115  CA  ASP   12       -9.100   6.500  -3.677  1.00 20.00 0 2 201
ATOM    116  CA  ASP   12       -3.900 -14.300  -3.677  1.00 20.00 0 2 201
ATOM    117  CA  ASP   12       -1.300 -16.900  -3.677  1.00 20.00 1 2 201
ATOM    118  CA  ASP   12       -1.300 -14.300  -3.677  1.00 20.00 0 2 201
ATOM    119  CA  ASP   12       -1.300  -1.300  -3.677  1.00 20.00 0 2 201
ATOM    120  CA  ASP   13       -1.300  14.300  -3.677  1.00 20.00 0 2 201
ATOM    121  CA  ASP   13       -1.300  16.900  -3.677  1.00 20.00 1 2 201
ATOM    122  CA  ASP   13        1.300 -16.900  -3.677  1.00 20.00 1 2 201
ATOM    123  CA  ASP   13        1.300 -14.300  -3.677  1.00 20.00 0 2 201
ATOM    124  CA  ASP   13        1.300  -3.900  -3.677  1.00 20.00 0 2 201
ATOM    125  CA  ASP   13        1.300  -1.300  -3.677  1.00 20.00 0 2 201
ATOM    126  CA  ASP   13        1.300  11.700  -3.677  1.00 20.00 0 2 201
ATOM    127  CA  ASP   13        1.300  14.300  -3.677  1.00 20.00 0 2 201
ATOM    128  CA  ASP   13        1.300  16.900  -3.677  1.00 20.00 0 2 201
ATOM    129  CA  ASP   13        3.900 -14.300  -3.677  1.00 20.00 0 2 201
ATOM    130  CA  ASP   14        6.500   3.900  -3.677  1.00 20.00 0 2 201
ATOM    131  CA  ASP   14        6.500   6.500  -3.677  1.00 20.00 0 2 201
ATOM    132  CA  ASP   14        9.100   3.900  -3.677  1.00 20.00 0 2 201
ATOM    133  CA  ASP   14        9.100   6.500  -3.677  1.00 20.00 0 2 201
ATOM    134  CA  ASP   14       11.700   3.900  -3.677  1.00 20.00 0 2 201
ATOM    135  CA  ASP   14      -10.400   0.000  -1.838  1.00 20.00 0 2 201
ATOM    136  CA  ASP   14      -10.400   2.600  -1.838  1.00 20.00 0 2 201
ATOM    137  CA  ASP   14      -10.400   5.200  -1.838  1.00 20.00 0 2 201
ATOM    138  CA  ASP   14      -10.400   7.800  -1.838  1.00 20.00 0 2 201
ATOM    139  CA  ASP   14       -7.800   0.000  -1.838  1.00 20.00 0 2 201
ATOM    140  CA  ASP   15       -7.800   2.600  -1.838  1.00 20.00 0 2 201
ATOM    141  CA  ASP   15       -7.800   5.200  -1.838  1.00 20.00 0 2 201
ATOM    142  CA  ASP   15       -7.800   7.800  -1.838  1.00 20.00 0 2 201
ATOM    143  CA  ASP   15       -5.200  -2.600  -1.838  1.00 20.00 0 2 201
ATOM    144  CA  ASP   15       -5.200   0.000  -1.838  1.00 20.00 0 2 201
ATOM    145  CA  ASP   15       -5.200   2.600  -1.838  1.00 20.00 0 2 201
ATOM    146  CA  ASP   15       -2.600 -15.600  -1.838  1.00 20.00 0 2 201
ATOM    147  CA  ASP   15       -2.600 -13.000  -1.838  1.00 20.00 0 2 201
ATOM    148  CA  ASP   15       -2.600  -2.600  -1.838  1.00 20.00 0 2 201
ATOM    149  CA  ASP   15       -2.600   0.000  -1.838  1.00 20.00 0 2 201
ATOM    150  CA  ASP   16       -2.600  15.600  -1.838  1.00 20.00 0 2 201
ATOM    151  CA  ASP   16        0.000 -15.600  -1.838  1.00 20.00 0 2 201
ATOM    152  CA  ASP   16        0.000 -13.000  -1.838  1.00 20.00 0 2 201
```

```
ATOM    153  CA  ASP   16        0.000  -5.200  -1.838  1.00 20.00 0 2 201
ATOM    154  CA  ASP   16        0.000  -2.600  -1.838  1.00 20.00 0 2 201
ATOM    155  CA  ASP   16        0.000   0.000  -1.838  1.00 20.00 0 2 201
ATOM    156  CA  ASP   16        0.000  13.000  -1.838  1.00 20.00 0 2 201
ATOM    157  CA  ASP   16        0.000  15.600  -1.838  1.00 20.00 0 2 201
ATOM    158  CA  ASP   16        7.800   5.200  -1.838  1.00 20.00 0 2 201
ATOM    159  CA  ASP   16       10.400   2.600  -1.838  1.00 20.00 0 2 201
ATOM    160  CA  ASP   17       10.400   5.200  -1.838  1.00 20.00 0 2 201
ATOM    161  CA  ASP   17       13.000   0.000  -1.838  1.00 20.00 0 2 201
ATOM    162  CA  ASP   17       13.000   2.600  -1.838  1.00 20.00 0 2 201
ATOM    163  CA  ASP   17       13.000   5.200  -1.838  1.00 20.00 0 2 201
ATOM    164  CA  ASP   17       15.600   0.000  -1.838  1.00 20.00 0 2 201
ATOM    165  CA  ASP   17      -11.700   1.300   0.000  1.00 20.00 0 2 201
ATOM    166  CA  ASP   17      -11.700   3.900   0.000  1.00 20.00 0 2 201
ATOM    167  CA  ASP   17      -11.700   6.500   0.000  1.00 20.00 0 2 201
ATOM    168  CA  ASP   17       -9.100   1.300   0.000  1.00 20.00 0 2 201
ATOM    169  CA  ASP   17       -9.100   3.900   0.000  1.00 20.00 1 2 201
ATOM    170  CA  ASP   18       -9.100   6.500   0.000  1.00 20.00 0 2 201
ATOM    171  CA  ASP   18       -6.500  -1.300   0.000  1.00 20.00 0 2 201
ATOM    172  CA  ASP   18       -6.500   1.300   0.000  1.00 20.00 0 2 201
ATOM    173  CA  ASP   18       -6.500   3.900   0.000  1.00 20.00 0 2 201
ATOM    174  CA  ASP   18       -6.500   6.500   0.000  1.00 20.00 0 2 201
ATOM    175  CA  ASP   18       -3.900  -1.300   0.000  1.00 20.00 0 2 201
ATOM    176  CA  ASP   18       -3.900   1.300   0.000  1.00 20.00 0 2 201
ATOM    177  CA  ASP   18       -3.900  16.900   0.000  1.00 20.00 0 2 201
ATOM    178  CA  ASP   18       -1.300 -16.900   0.000  1.00 20.00 0 2 201
ATOM    179  CA  ASP   18       -1.300 -14.300   0.000  1.00 20.00 0 2 201
ATOM    180  CA  ASP   19       -1.300  -3.900   0.000  1.00 20.00 0 2 201
ATOM    181  CA  ASP   19       -1.300  14.300   0.000  1.00 20.00 0 2 201
ATOM    182  CA  ASP   19       -1.300  16.900   0.000  1.00 20.00 0 2 201
ATOM    183  CA  ASP   19        1.300 -14.300   0.000  1.00 20.00 0 2 201
ATOM    184  CA  ASP   19        1.300  -3.900   0.000  1.00 20.00 0 2 201
ATOM    185  CA  ASP   19        1.300  -1.300   0.000  1.00 20.00 0 2 201
ATOM    186  CA  ASP   19        1.300  14.300   0.000  1.00 20.00 0 2 201
ATOM    187  CA  ASP   19        1.300  16.900   0.000  1.00 20.00 0 2 201
ATOM    188  CA  ASP   19        3.900  -3.900   0.000  1.00 20.00 0 2 201
ATOM    189  CA  ASP   19        9.100   3.900   0.000  1.00 20.00 0 2 201
ATOM    190  CA  ASP   20        9.100   6.500   0.000  1.00 20.00 0 2 201
ATOM    191  CA  ASP   20       11.700  -1.300   0.000  1.00 20.00 0 2 201
ATOM    192  CA  ASP   20       11.700   1.300   0.000  1.00 20.00 0 2 201
ATOM    193  CA  ASP   20       11.700   3.900   0.000  1.00 20.00 1 2 201
ATOM    194  CA  ASP   20       11.700   6.500   0.000  1.00 20.00 0 2 201
ATOM    195  CA  ASP   20       14.300  -1.300   0.000  1.00 20.00 0 2 201
```

```
ATOM    196  CA  ASP   20      14.300   1.300   0.000  1.00 20.00 0 2 201
ATOM    197  CA  ASP   20      14.300   3.900   0.000  1.00 20.00 0 2 201
ATOM    198  CA  ASP   20     -10.400   2.600   1.838  1.00 20.00 0 2 201
ATOM    199  CA  ASP   20     -10.400   5.200   1.838  1.00 20.00 0 2 201
ATOM    200  CA  ASP   21      -7.800   0.000   1.838  1.00 20.00 0 2 201
ATOM    201  CA  ASP   21      -7.800   2.600   1.838  1.00 20.00 0 2 201
ATOM    202  CA  ASP   21      -7.800   5.200   1.838  1.00 20.00 0 2 201
ATOM    203  CA  ASP   21      -2.600 -15.600   1.838  1.00 20.00 0 2 201
ATOM    204  CA  ASP   21      -2.600  15.600   1.838  1.00 20.00 0 2 201
ATOM    205  CA  ASP   21       0.000 -15.600   1.838  1.00 20.00 0 2 201
ATOM    206  CA  ASP   21       0.000  -5.200   1.838  1.00 20.00 0 2 201
ATOM    207  CA  ASP   21       0.000  -2.600   1.838  1.00 20.00 0 2 201
ATOM    208  CA  ASP   21       0.000   0.000   1.838  1.00 20.00 0 2 201
ATOM    209  CA  ASP   21       0.000  13.000   1.838  1.00 20.00 0 2 201
ATOM    210  CA  ASP   22       0.000  15.600   1.838  1.00 20.00 0 2 201
ATOM    211  CA  ASP   22       2.600  -5.200   1.838  1.00 20.00 0 2 201
ATOM    212  CA  ASP   22       2.600  -2.600   1.838  1.00 20.00 0 2 201
ATOM    213  CA  ASP   22      10.400   2.600   1.838  1.00 20.00 0 2 201
ATOM    214  CA  ASP   22      10.400   5.200   1.838  1.00 20.00 0 2 201
ATOM    215  CA  ASP   22      13.000   0.000   1.838  1.00 20.00 0 2 201
ATOM    216  CA  ASP   22      13.000   2.600   1.838  1.00 20.00 0 2 201
ATOM    217  CA  ASP   22      13.000   5.200   1.838  1.00 20.00 0 2 201
ATOM    218  CA  ASP   22      15.600   0.000   1.838  1.00 20.00 0 2 201
ATOM    219  CA  ASP   22      -1.300 -16.900   3.677  1.00 20.00 0 2 201
ATOM    220  CA  ASP   23      -1.300 -14.300   3.677  1.00 20.00 0 2 201
ATOM    221  CA  ASP   23      -1.300  11.700   3.677  1.00 20.00 0 2 201
ATOM    222  CA  ASP   23      -1.300  14.300   3.677  1.00 20.00 1 2 201
ATOM    223  CA  ASP   23      -1.300  16.900   3.677  1.00 20.00 1 2 201
ATOM    224  CA  ASP   23       1.300 -14.300   3.677  1.00 20.00 0 2 201
ATOM    225  CA  ASP   23       1.300  -3.900   3.677  1.00 20.00 0 2 201
ATOM    226  CA  ASP   23       1.300  11.700   3.677  1.00 20.00 0 2 201
ATOM    227  CA  ASP   23       1.300  14.300   3.677  1.00 20.00 0 2 201
ATOM    228  CA  ASP   23       1.300  16.900   3.677  1.00 20.00 0 2 201
ATOM    229  CA  ASP   23      11.700   3.900   3.677  1.00 20.00 0 2 201
ATOM    230  CA  ASP   24      14.300   1.300   3.677  1.00 20.00 0 2 201
ATOM    231  CA  ASP   24      14.300   3.900   3.677  1.00 20.00 0 2 201
ATOM    232  CA  ASP   24      -2.600 -15.600   5.515  1.00 20.00 0 2 201
ATOM    233  CA  ASP   24      -2.600 -13.000   5.515  1.00 20.00 0 2 201
ATOM    234  CA  ASP   24      -2.600  10.400   5.515  1.00 20.00 0 2 201
ATOM    235  CA  ASP   24      -2.600  13.000   5.515  1.00 20.00 0 2 201
ATOM    236  CA  ASP   24      -2.600  15.600   5.515  1.00 20.00 0 2 201
ATOM    237  CA  ASP   24       0.000 -15.600   5.515  1.00 20.00 0 2 201
ATOM    238  CA  ASP   24       0.000 -13.000   5.515  1.00 20.00 0 2 201
```

```
ATOM    239  CA  ASP    24       0.000  10.400   5.515  1.00 20.00 0 2 201
ATOM    240  CA  ASP    25       0.000  13.000   5.515  1.00 20.00 1 2 201
ATOM    241  CA  ASP    25       0.000  15.600   5.515  1.00 20.00 1 2 201
ATOM    242  CA  ASP    25       2.600 -15.600   5.515  1.00 20.00 0 2 201
ATOM    243  CA  ASP    25       2.600 -13.000   5.515  1.00 20.00 0 2 201
ATOM    244  CA  ASP    25       2.600  13.000   5.515  1.00 20.00 0 2 201
ATOM    245  CA  ASP    25       2.600  15.600   5.515  1.00 20.00 0 2 201
ATOM    246  CA  ASP    25      10.400   2.600   5.515  1.00 20.00 0 2 201
ATOM    247  CA  ASP    25      -3.900 -14.300   7.354  1.00 20.00 0 2 201
ATOM    248  CA  ASP    25      -3.900 -11.700   7.354  1.00 20.00 0 2 201
ATOM    249  CA  ASP    25      -1.300 -14.300   7.354  1.00 20.00 0 2 201
ATOM    250  CA  ASP    26      -1.300 -11.700   7.354  1.00 20.00 0 2 201
ATOM    251  CA  ASP    26      -1.300  11.700   7.354  1.00 20.00 0 2 201
ATOM    252  CA  ASP    26      -1.300  14.300   7.354  1.00 20.00 0 2 201
ATOM    253  CA  ASP    26       1.300 -14.300   7.354  1.00 20.00 1 2 201
ATOM    254  CA  ASP    26       1.300 -11.700   7.354  1.00 20.00 0 2 201
ATOM    255  CA  ASP    26       1.300  11.700   7.354  1.00 20.00 0 2 201
ATOM    256  CA  ASP    26       1.300  14.300   7.354  1.00 20.00 0 2 201
ATOM    257  CA  ASP    26       3.900 -14.300   7.354  1.00 20.00 1 2 201
ATOM    258  CA  ASP    26       3.900 -11.700   7.354  1.00 20.00 0 2 201
ATOM    259  CA  ASP    26       6.500  -1.300   7.354  1.00 20.00 0 2 201
ATOM    260  CA  ASP    27       6.500   1.300   7.354  1.00 20.00 0 2 201
ATOM    261  CA  ASP    27       9.100  -1.300   7.354  1.00 20.00 0 2 201
ATOM    262  CA  ASP    27       9.100   1.300   7.354  1.00 20.00 0 2 201
ATOM    263  CA  ASP    27      -5.200   0.000   9.192  1.00 20.00 0 2 201
ATOM    264  CA  ASP    27      -2.600 -13.000   9.192  1.00 20.00 0 2 201
ATOM    265  CA  ASP    27      -2.600 -10.400   9.192  1.00 20.00 0 2 201
ATOM    266  CA  ASP    27      -2.600   0.000   9.192  1.00 20.00 0 2 201
ATOM    267  CA  ASP    27      -2.600   2.600   9.192  1.00 20.00 0 2 201
ATOM    268  CA  ASP    27       0.000 -13.000   9.192  1.00 20.00 0 2 201
ATOM    269  CA  ASP    27       0.000 -10.400   9.192  1.00 20.00 0 2 201
ATOM    270  CA  ASP    28       0.000   0.000   9.192  1.00 20.00 0 2 201
ATOM    271  CA  ASP    28       0.000   2.600   9.192  1.00 20.00 0 2 201
ATOM    272  CA  ASP    28       2.600 -13.000   9.192  1.00 20.00 0 2 201
ATOM    273  CA  ASP    28       2.600 -10.400   9.192  1.00 20.00 0 2 201
ATOM    274  CA  ASP    28       2.600   0.000   9.192  1.00 20.00 0 2 201
ATOM    275  CA  ASP    28       2.600   2.600   9.192  1.00 20.00 0 2 201
ATOM    276  CA  ASP    28       5.200 -13.000   9.192  1.00 20.00 0 2 201
ATOM    277  CA  ASP    28       5.200 -10.400   9.192  1.00 20.00 0 2 201
ATOM    278  CA  ASP    28       5.200   0.000   9.192  1.00 20.00 0 2 201
ATOM    279  CA  ASP    28       5.200   2.600   9.192  1.00 20.00 0 2 201
ATOM    280  CA  ASP    29       7.800  -2.600   9.192  1.00 20.00 0 2 201
ATOM    281  CA  ASP    29       7.800   0.000   9.192  1.00 20.00 0 2 201
```

```
ATOM    282  CA  ASP   29       7.800   2.600   9.192  1.00 20.00 0 2 201
ATOM    283  CA  ASP   29      10.400   0.000   9.192  1.00 20.00 0 2 201
ATOM    284  CA  ASP   29      10.400   2.600   9.192  1.00 20.00 0 2 201
ATOM    285  CA  ASP   29      -6.500  -1.300  11.031  1.00 20.00 0 2 201
ATOM    286  CA  ASP   29      -6.500   1.300  11.031  1.00 20.00 0 2 201
ATOM    287  CA  ASP   29      -3.900  -1.300  11.031  1.00 20.00 0 2 201
ATOM    288  CA  ASP   29      -3.900   1.300  11.031  1.00 20.00 1 2 201
ATOM    289  CA  ASP   29      -3.900   3.900  11.031  1.00 20.00 0 2 201
ATOM    290  CA  ASP   30      -1.300 -11.700  11.031  1.00 20.00 0 2 201
ATOM    291  CA  ASP   30      -1.300  -1.300  11.031  1.00 20.00 0 2 201
ATOM    292  CA  ASP   30      -1.300   1.300  11.031  1.00 20.00 0 2 201
ATOM    293  CA  ASP   30      -1.300   3.900  11.031  1.00 20.00 0 2 201
ATOM    294  CA  ASP   30       1.300 -11.700  11.031  1.00 20.00 0 2 201
ATOM    295  CA  ASP   30       1.300  -1.300  11.031  1.00 20.00 0 2 201
ATOM    296  CA  ASP   30       1.300   1.300  11.031  1.00 20.00 0 2 201
ATOM    297  CA  ASP   30       1.300   3.900  11.031  1.00 20.00 0 2 201
ATOM    298  CA  ASP   30       3.900 -11.700  11.031  1.00 20.00 0 2 201
ATOM    299  CA  ASP   30       3.900  -1.300  11.031  1.00 20.00 0 2 201
ATOM    300  CA  ASP   31       3.900   1.300  11.031  1.00 20.00 0 2 201
ATOM    301  CA  ASP   31       6.500  -1.300  11.031  1.00 20.00 0 2 201
ATOM    302  CA  ASP   31       6.500   1.300  11.031  1.00 20.00 1 2 201
ATOM    303  CA  ASP   31       9.100  -1.300  11.031  1.00 20.00 0 2 201
ATOM    304  CA  ASP   31       9.100   1.300  11.031  1.00 20.00 0 2 201
ATOM    305  CA  ASP   31      11.700  -1.300  11.031  1.00 20.00 0 2 201
ATOM    306  CA  ASP   31      11.700   1.300  11.031  1.00 20.00 0 2 201
ATOM    307  CA  ASP   31      -5.200  -2.600  12.869  1.00 20.00 0 2 201
ATOM    308  CA  ASP   31      -5.200   0.000  12.869  1.00 20.00 0 2 201
ATOM    309  CA  ASP   31      -5.200   2.600  12.869  1.00 20.00 0 2 201
ATOM    310  CA  ASP   32      -2.600  -2.600  12.869  1.00 20.00 0 2 201
ATOM    311  CA  ASP   32      -2.600   0.000  12.869  1.00 20.00 1 2 201
ATOM    312  CA  ASP   32      -2.600   2.600  12.869  1.00 20.00 0 2 201
ATOM    313  CA  ASP   32       0.000   0.000  12.869  1.00 20.00 0 2 201
ATOM    314  CA  ASP   32       0.000   2.600  12.869  1.00 20.00 0 2 201
ATOM    315  CA  ASP   32       2.600   0.000  12.869  1.00 20.00 0 2 201
ATOM    316  CA  ASP   32       2.600   2.600  12.869  1.00 20.00 0 2 201
ATOM    317  CA  ASP   32       5.200   0.000  12.869  1.00 20.00 0 2 201
ATOM    318  CA  ASP   32       7.800   0.000  12.869  1.00 20.00 0 2 201
ATOM    319  CA  ASP   32      10.400   0.000  12.869  1.00 20.00 0 2 201
ATOM    320  CA  ASP   33      -3.900  -1.300  14.708  1.00 20.00 0 2 201
ATOM    321  CA  ASP   33      -3.900   1.300  14.708  1.00 20.00 0 2 201
ATOM    322  CA  ASP   33      -1.300  -1.300  14.708  1.00 20.00 0 2 201
ATOM    323  CA  ASP   33      -1.300   1.300  14.708  1.00 20.00 0 2 201
```

APPENDIX E: COMPREHENSIVE RESULTS

A note on filename nomenclature:

For cytochrome the filenames are as follows:

MolarityTemperatureRun#

An underscore before the temperature indicates a negative temperature

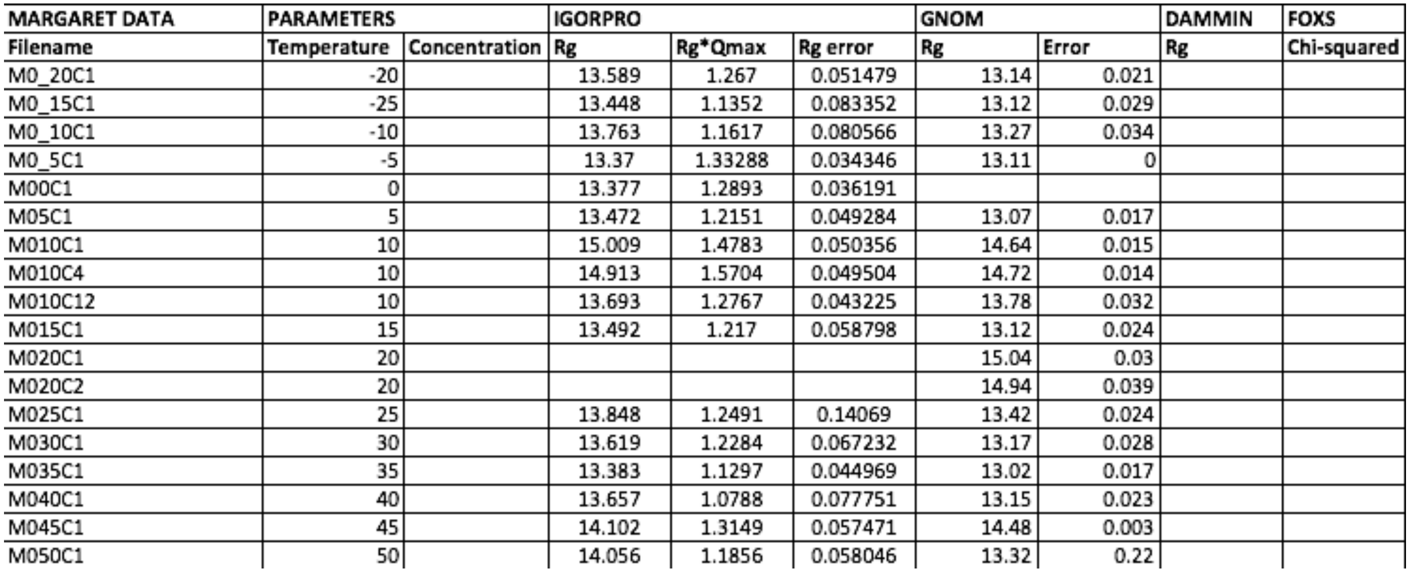
e.g. M0_20C1 means 0 Molar GuHCl solution at -20 degrees celcius, first run

For miro the filenames are as follows:

Protein_Calcium ion concentration_Temperature_Run#_Analysis attempt

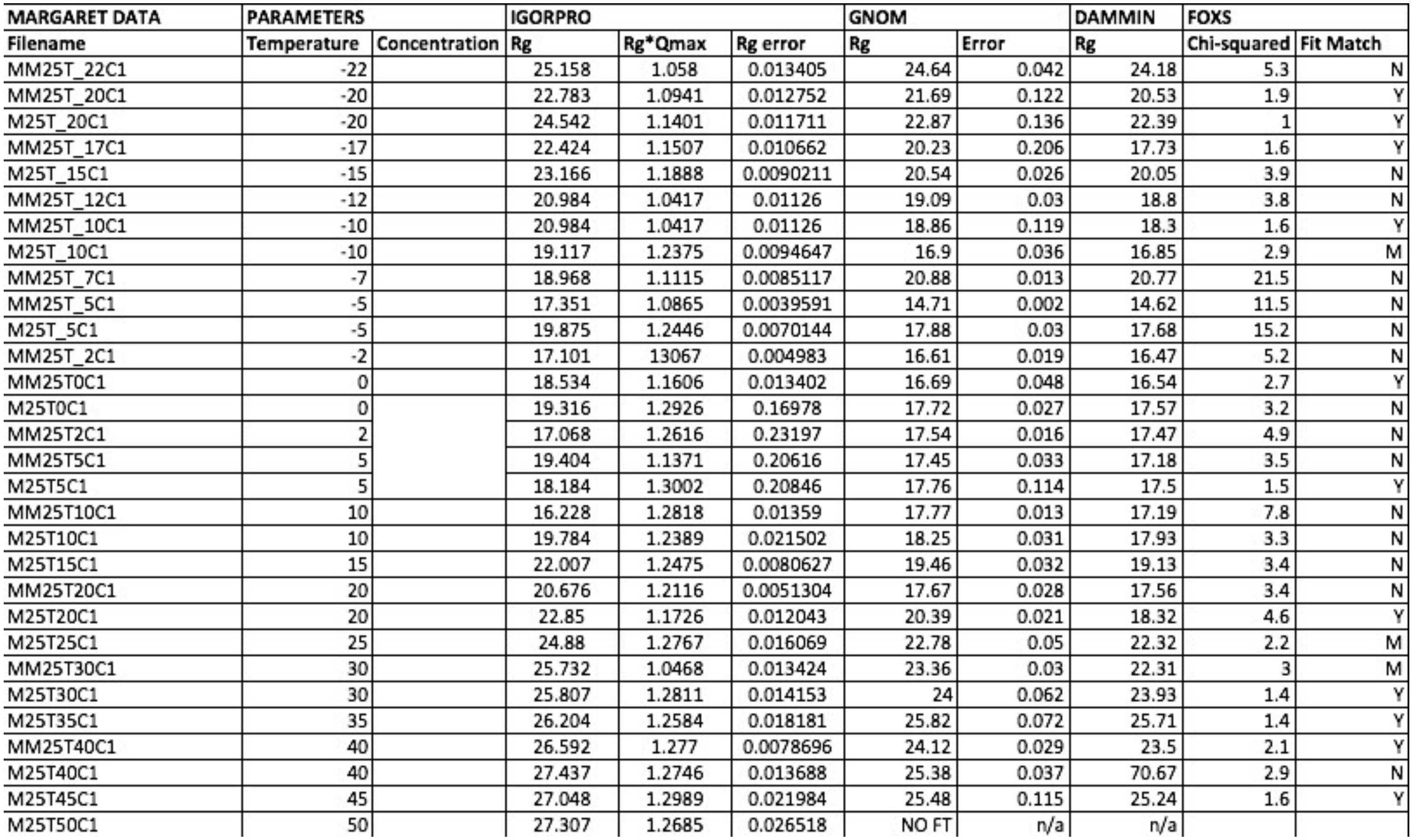
If no temperature is present, the temperature is 4 degrees celcius.

e.g. miros_00ca_04c_3_1 means miro-S in a 0mg/ml calcium ion solution, 3rd data collection, 1st analysis attempt



MARGARET DATA	PARAMETERS		IGORPRO			GNOM		DAMMIN	FOXS
Filename	Temperature	Concentration	Rg	Rg*Qmax	Rg error	Rg	Error	Rg	Chi-squared
M0_20C1	-20		13.589	1.267	0.051479	13.14	0.021		
M0_15C1	-25		13.448	1.1352	0.083352	13.12	0.029		
M0_10C1	-10		13.763	1.1617	0.080566	13.27	0.034		
M0_5C1	-5		13.37	1.33288	0.034346	13.11	0		
M00C1	0		13.377	1.2893	0.036191				
M05C1	5		13.472	1.2151	0.049284	13.07	0.017		
M010C1	10		15.009	1.4783	0.050356	14.64	0.015		
M010C4	10		14.913	1.5704	0.049504	14.72	0.014		
M010C12	10		13.693	1.2767	0.043225	13.78	0.032		
M015C1	15		13.492	1.217	0.058798	13.12	0.024		
M020C1	20					15.04	0.03		
M020C2	20					14.94	0.039		
M025C1	25		13.848	1.2491	0.14069	13.42	0.024		
M030C1	30		13.619	1.2284	0.067232	13.17	0.028		
M035C1	35		13.383	1.1297	0.044969	13.02	0.017		
M040C1	40		13.657	1.0788	0.077751	13.15	0.023		
M045C1	45		14.102	1.3149	0.057471	14.48	0.003		
M050C1	50		14.056	1.1856	0.058046	13.32	0.22		

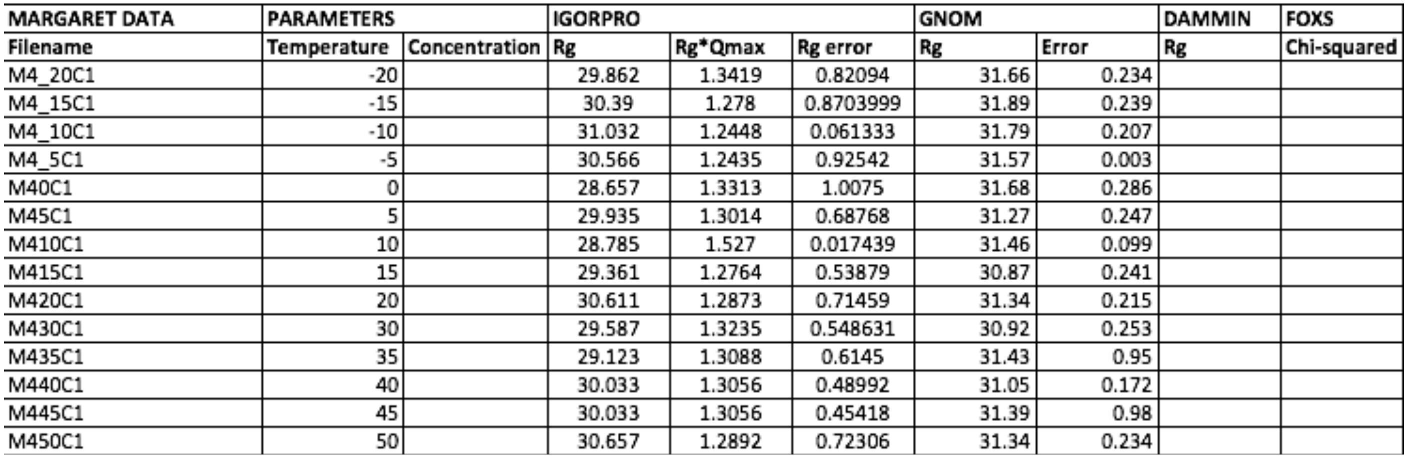
Figure 1: Results for cytochrome-c in a 0M GuHCl solution



MARGARET DATA	PARAMETERS		IGORPRO			GNOM		DAMMIN	FOXS	
Filename	Temperature	Concentration	Rg	Rg*Qmax	Rg error	Rg	Error	Rg	Chi-squared	Fit Match
MM25T_22C1	-22		25.158	1.058	0.013405	24.64	0.042	24.18	5.3	N
MM25T_20C1	-20		22.783	1.0941	0.012752	21.69	0.122	20.53	1.9	Y
M25T_20C1	-20		24.542	1.1401	0.011711	22.87	0.136	22.39	1	Y
MM25T_17C1	-17		22.424	1.1507	0.010662	20.23	0.206	17.73	1.6	Y
M25T_15C1	-15		23.166	1.1888	0.0090211	20.54	0.026	20.05	3.9	N
MM25T_12C1	-12		20.984	1.0417	0.01126	19.09	0.03	18.8	3.8	N
MM25T_10C1	-10		20.984	1.0417	0.01126	18.86	0.119	18.3	1.6	Y
M25T_10C1	-10		19.117	1.2375	0.0094647	16.9	0.036	16.85	2.9	M
MM25T_7C1	-7		18.968	1.1115	0.0085117	20.88	0.013	20.77	21.5	N
MM25T_5C1	-5		17.351	1.0865	0.0039591	14.71	0.002	14.62	11.5	N
M25T_5C1	-5		19.875	1.2446	0.0070144	17.88	0.03	17.68	15.2	N
MM25T_2C1	-2		17.101	13067	0.004983	16.61	0.019	16.47	5.2	N
MM25T0C1	0		18.534	1.1606	0.013402	16.69	0.048	16.54	2.7	Y
M25T0C1	0		19.316	1.2926	0.16978	17.72	0.027	17.57	3.2	N
MM25T2C1	2		17.068	1.2616	0.23197	17.54	0.016	17.47	4.9	N
MM25T5C1	5		19.404	1.1371	0.20616	17.45	0.033	17.18	3.5	N
M25T5C1	5		18.184	1.3002	0.20846	17.76	0.114	17.5	1.5	Y
MM25T10C1	10		16.228	1.2818	0.01359	17.77	0.013	17.19	7.8	N
M25T10C1	10		19.784	1.2389	0.021502	18.25	0.031	17.93	3.3	N
M25T15C1	15		22.007	1.2475	0.0080627	19.46	0.032	19.13	3.4	N
MM25T20C1	20		20.676	1.2116	0.0051304	17.67	0.028	17.56	3.4	N
M25T20C1	20		22.85	1.1726	0.012043	20.39	0.021	18.32	4.6	Y
M25T25C1	25		24.88	1.2767	0.016069	22.78	0.05	22.32	2.2	M
MM25T30C1	30		25.732	1.0468	0.013424	23.36	0.03	22.31	3	M
M25T30C1	30		25.807	1.2811	0.014153	24	0.062	23.93	1.4	Y
M25T35C1	35		26.204	1.2584	0.018181	25.82	0.072	25.71	1.4	Y
MM25T40C1	40		26.592	1.277	0.0078696	24.12	0.029	23.5	2.1	Y
M25T40C1	40		27.437	1.2746	0.013688	25.38	0.037	70.67	2.9	N
M25T45C1	45		27.048	1.2989	0.021984	25.48	0.115	25.24	1.6	Y
M25T50C1	50		27.307	1.2685	0.026518	NO FT	n/a	n/a		

Figure 2: Results for cytochrome-c in a 2.5M GuHCl solution

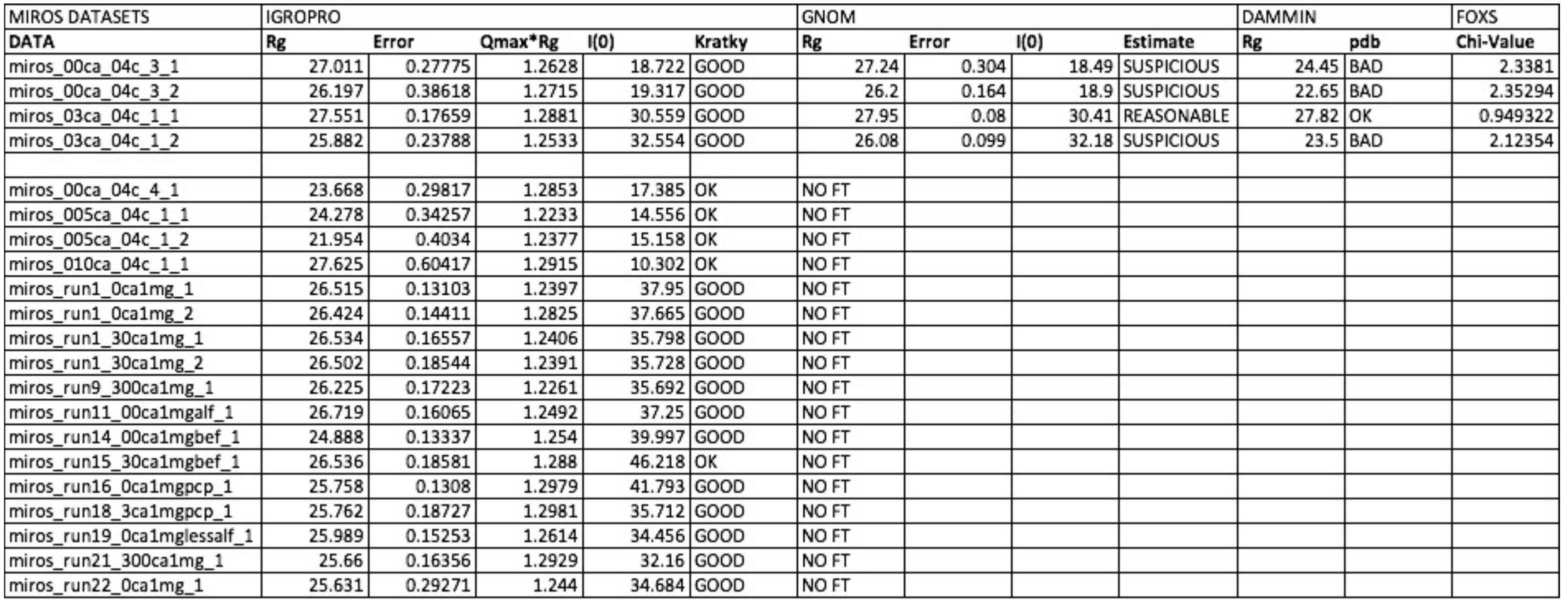
.



MARGARET DATA	PARAMETERS		IGORPRO			GNOM		DAMMIN	FOXS
Filename	Temperature	Concentration	Rg	Rg*Qmax	Rg error	Rg	Error	Rg	Chi-squared
M4_20C1	-20		29.862	1.3419	0.82094	31.66	0.234		
M4_15C1	-15		30.39	1.278	0.8703999	31.89	0.239		
M4_10C1	-10		31.032	1.2448	0.061333	31.79	0.207		
M4_5C1	-5		30.566	1.2435	0.92542	31.57	0.003		
M40C1	0		28.657	1.3313	1.0075	31.68	0.286		
M45C1	5		29.935	1.3014	0.68768	31.27	0.247		
M410C1	10		28.785	1.527	0.017439	31.46	0.099		
M415C1	15		29.361	1.2764	0.53879	30.87	0.241		
M420C1	20		30.611	1.2873	0.71459	31.34	0.215		
M430C1	30		29.587	1.3235	0.548631	30.92	0.253		
M435C1	35		29.123	1.3088	0.6145	31.43	0.95		
M440C1	40		30.033	1.3056	0.48992	31.05	0.172		
M445C1	45		30.033	1.3056	0.45418	31.39	0.98		
M450C1	50		30.657	1.2892	0.72306	31.34	0.234		

Figure 3: Results for cytochrome-c in a 4M GuHCl solution

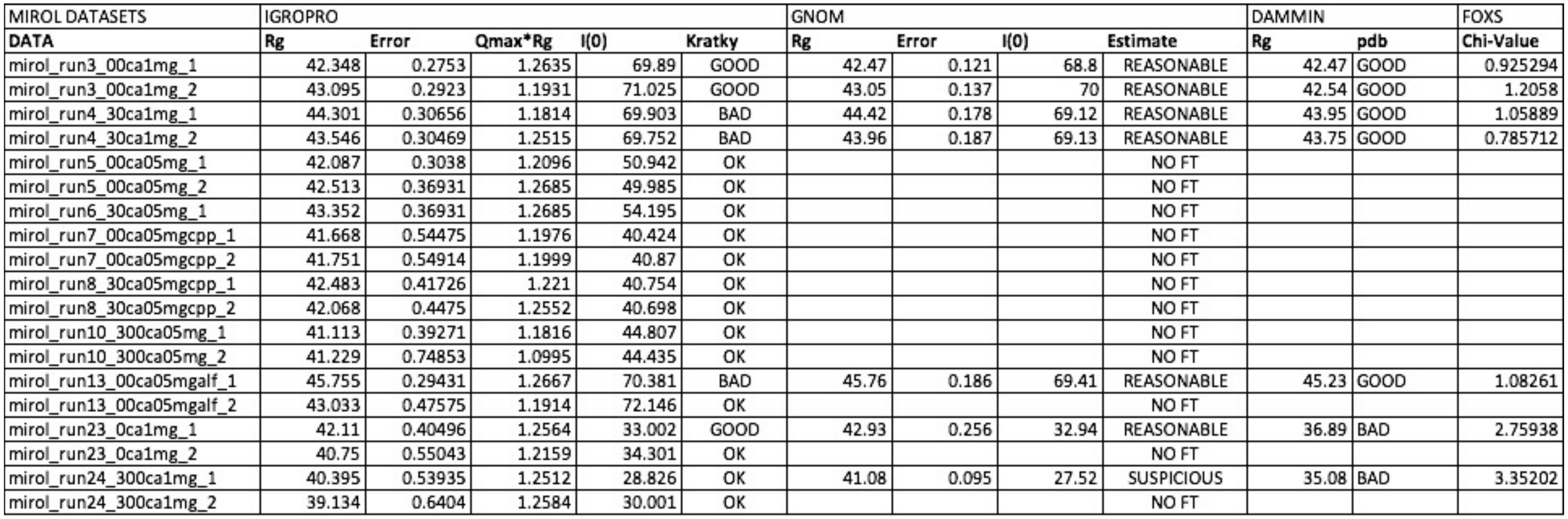
.



MIROS DATASETS	IGROPRO					GNOM				DAMMIN		FOXS
DATA	**Rg**	**Error**	**Qmax*Rg**	**I(0)**	**Kratky**	**Rg**	**Error**	**I(0)**	**Estimate**	**Rg**	**pdb**	**Chi-Value**
miros_00ca_04c_3_1	27.011	0.27775	1.2628	18.722	GOOD	27.24	0.304	18.49	SUSPICIOUS	24.45	BAD	2.3381
miros_00ca_04c_3_2	26.197	0.38618	1.2715	19.317	GOOD	26.2	0.164	18.9	SUSPICIOUS	22.65	BAD	2.35294
miros_03ca_04c_1_1	27.551	0.17659	1.2881	30.559	GOOD	27.95	0.08	30.41	REASONABLE	27.82	OK	0.949322
miros_03ca_04c_1_2	25.882	0.23788	1.2533	32.554	GOOD	26.08	0.099	32.18	SUSPICIOUS	23.5	BAD	2.12354
miros_00ca_04c_4_1	23.668	0.29817	1.2853	17.385	OK	NO FT						
miros_005ca_04c_1_1	24.278	0.34257	1.2233	14.556	OK	NO FT						
miros_005ca_04c_1_2	21.954	0.4034	1.2377	15.158	OK	NO FT						
miros_010ca_04c_1_1	27.625	0.60417	1.2915	10.302	OK	NO FT						
miros_run1_0ca1mg_1	26.515	0.13103	1.2397	37.95	GOOD	NO FT						
miros_run1_0ca1mg_2	26.424	0.14411	1.2825	37.665	GOOD	NO FT						
miros_run1_30ca1mg_1	26.534	0.16557	1.2406	35.798	GOOD	NO FT						
miros_run1_30ca1mg_2	26.502	0.18544	1.2391	35.728	GOOD	NO FT						
miros_run9_300ca1mg_1	26.225	0.17223	1.2261	35.692	GOOD	NO FT						
miros_run11_00ca1mgalf_1	26.719	0.16065	1.2492	37.25	GOOD	NO FT						
miros_run14_00ca1mgbef_1	24.888	0.13337	1.254	39.997	GOOD	NO FT						
miros_run15_30ca1mgbef_1	26.536	0.18581	1.288	46.218	OK	NO FT						
miros_run16_0ca1mgpcp_1	25.758	0.1308	1.2979	41.793	GOOD	NO FT						
miros_run18_3ca1mgpcp_1	25.762	0.18727	1.2981	35.712	GOOD	NO FT						
miros_run19_0ca1mglessalf_1	25.989	0.15253	1.2614	34.456	GOOD	NO FT						
miros_run21_300ca1mg_1	25.66	0.16356	1.2929	32.16	GOOD	NO FT						
miros_run22_0ca1mg_1	25.631	0.29271	1.244	34.684	GOOD	NO FT						

Figure 4: Results for miro-S

.



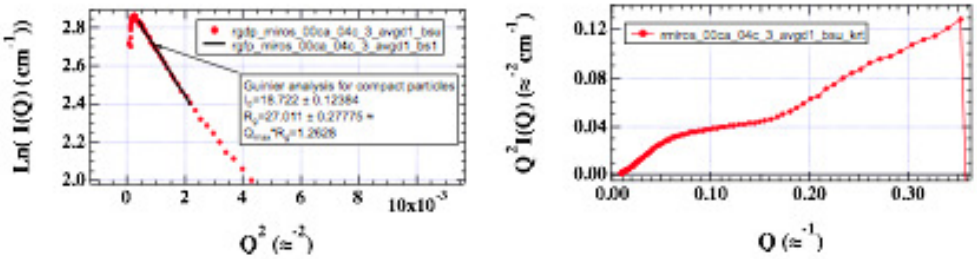
MIROL DATASETS	IGROPRO					GNOM				DAMMIN		FOXS
DATA	**Rg**	**Error**	**Qmax*Rg**	**I(0)**	**Kratky**	**Rg**	**Error**	**I(0)**	**Estimate**	**Rg**	**pdb**	**Chi-Value**
mirol_run3_00ca1mg_1	42.348	0.2753	1.2635	69.89	GOOD	42.47	0.121	68.8	REASONABLE	42.47	GOOD	0.925294
mirol_run3_00ca1mg_2	43.095	0.2923	1.1931	71.025	GOOD	43.05	0.137	70	REASONABLE	42.54	GOOD	1.2058
mirol_run4_30ca1mg_1	44.301	0.30656	1.1814	69.903	BAD	44.42	0.178	69.12	REASONABLE	43.95	GOOD	1.05889
mirol_run4_30ca1mg_2	43.546	0.30469	1.2515	69.752	BAD	43.96	0.187	69.13	REASONABLE	43.75	GOOD	0.785712
mirol_run5_00ca05mg_1	42.087	0.3038	1.2096	50.942	OK				NO FT			
mirol_run5_00ca05mg_2	42.513	0.36931	1.2685	49.985	OK				NO FT			
mirol_run6_30ca05mg_1	43.352	0.36931	1.2685	54.195	OK				NO FT			
mirol_run7_00ca05mgcpp_1	41.668	0.54475	1.1976	40.424	OK				NO FT			
mirol_run7_00ca05mgcpp_2	41.751	0.54914	1.1999	40.87	OK				NO FT			
mirol_run8_30ca05mgcpp_1	42.483	0.41726	1.221	40.754	OK				NO FT			
mirol_run8_30ca05mgcpp_2	42.068	0.4475	1.2552	40.698	OK				NO FT			
mirol_run10_300ca05mg_1	41.113	0.39271	1.1816	44.807	OK				NO FT			
mirol_run10_300ca05mg_2	41.229	0.74853	1.0995	44.435	OK				NO FT			
mirol_run13_00ca05mgalf_1	45.755	0.29431	1.2667	70.381	BAD	45.76	0.186	69.41	REASONABLE	45.23	GOOD	1.08261
mirol_run13_00ca05mgalf_2	43.033	0.47575	1.1914	72.146	OK				NO FT			
mirol_run23_0ca1mg_1	42.11	0.40496	1.2564	33.002	GOOD	42.93	0.256	32.94	REASONABLE	36.89	BAD	2.75938
mirol_run23_0ca1mg_2	40.75	0.55043	1.2159	34.301	OK				NO FT			
mirol_run24_300ca1mg_1	40.395	0.53935	1.2512	28.826	OK	41.08	0.095	27.52	SUSPICIOUS	35.08	BAD	3.35202
mirol_run24_300ca1mg_2	39.134	0.6404	1.2584	30.001	OK				NO FT			

Figure 5: Results for miro-L

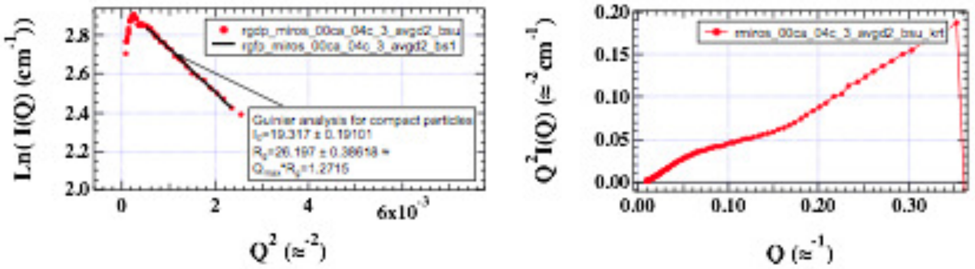
.

MIROS GUNIER AND KRATKY PLOTS

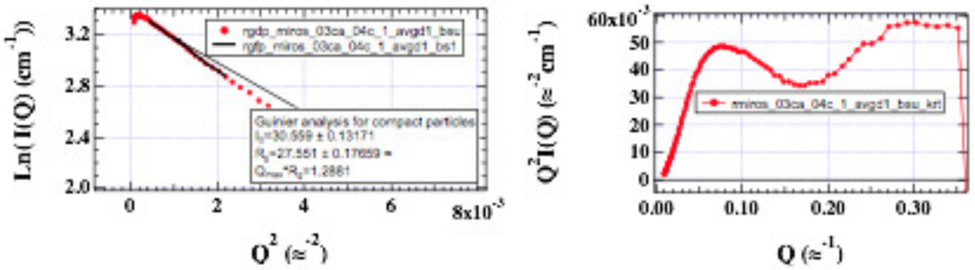
miros_00ca_04c_3_1



miros_00ca_04c_3_2



miros_03ca_04c_1_1



miros_03ca_04c_1_2

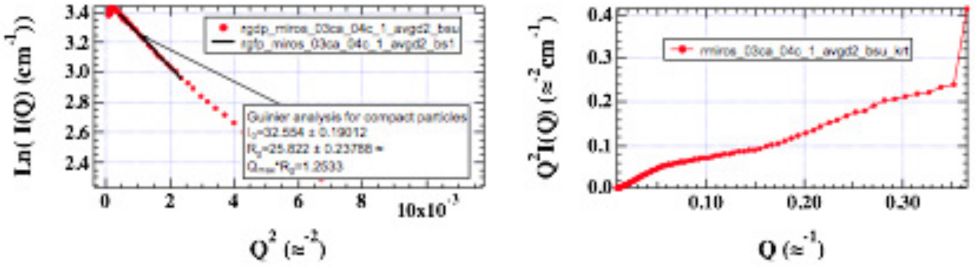


Figure 6: Guinier and Kratky plots for miro-S

.

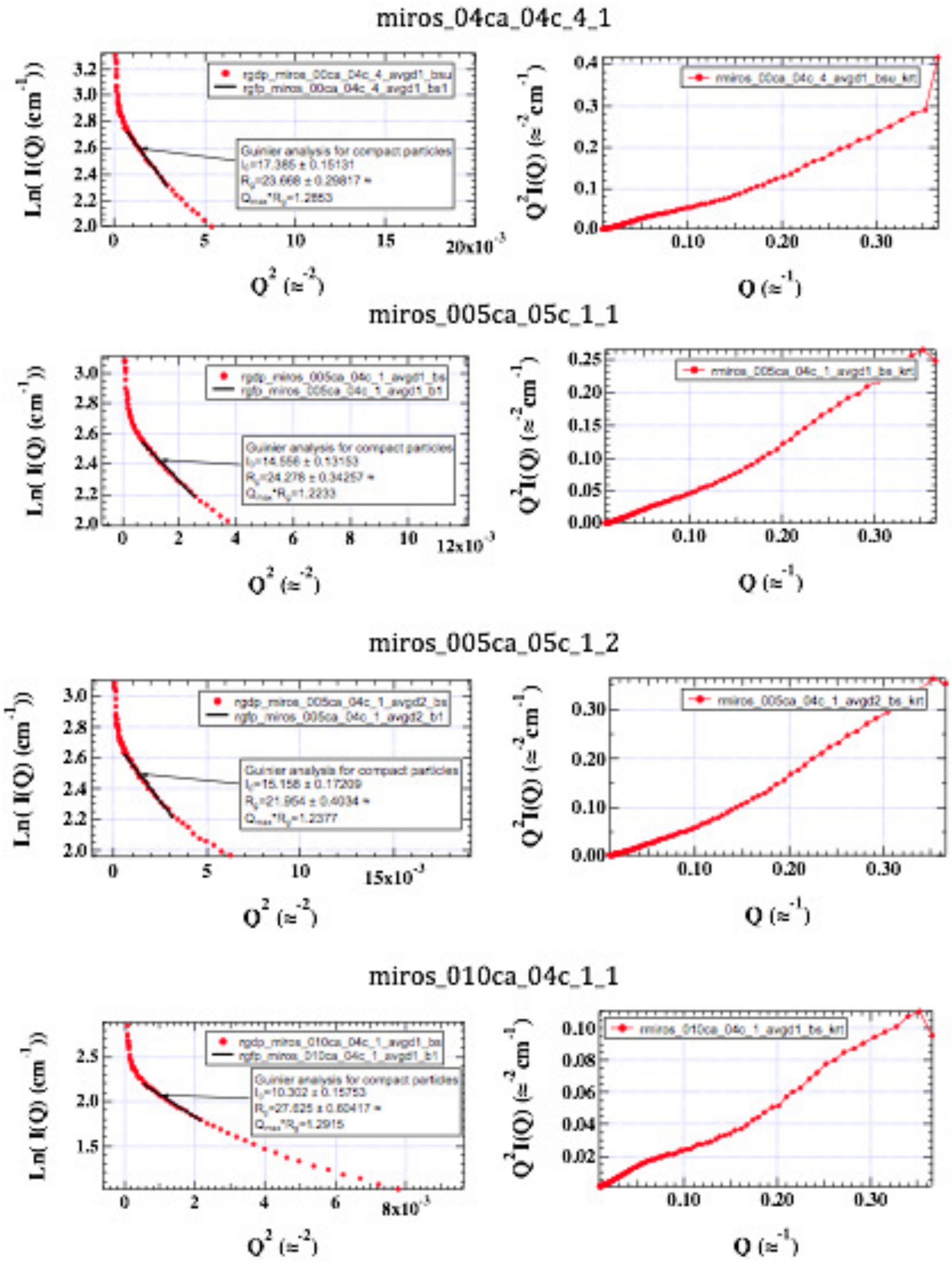
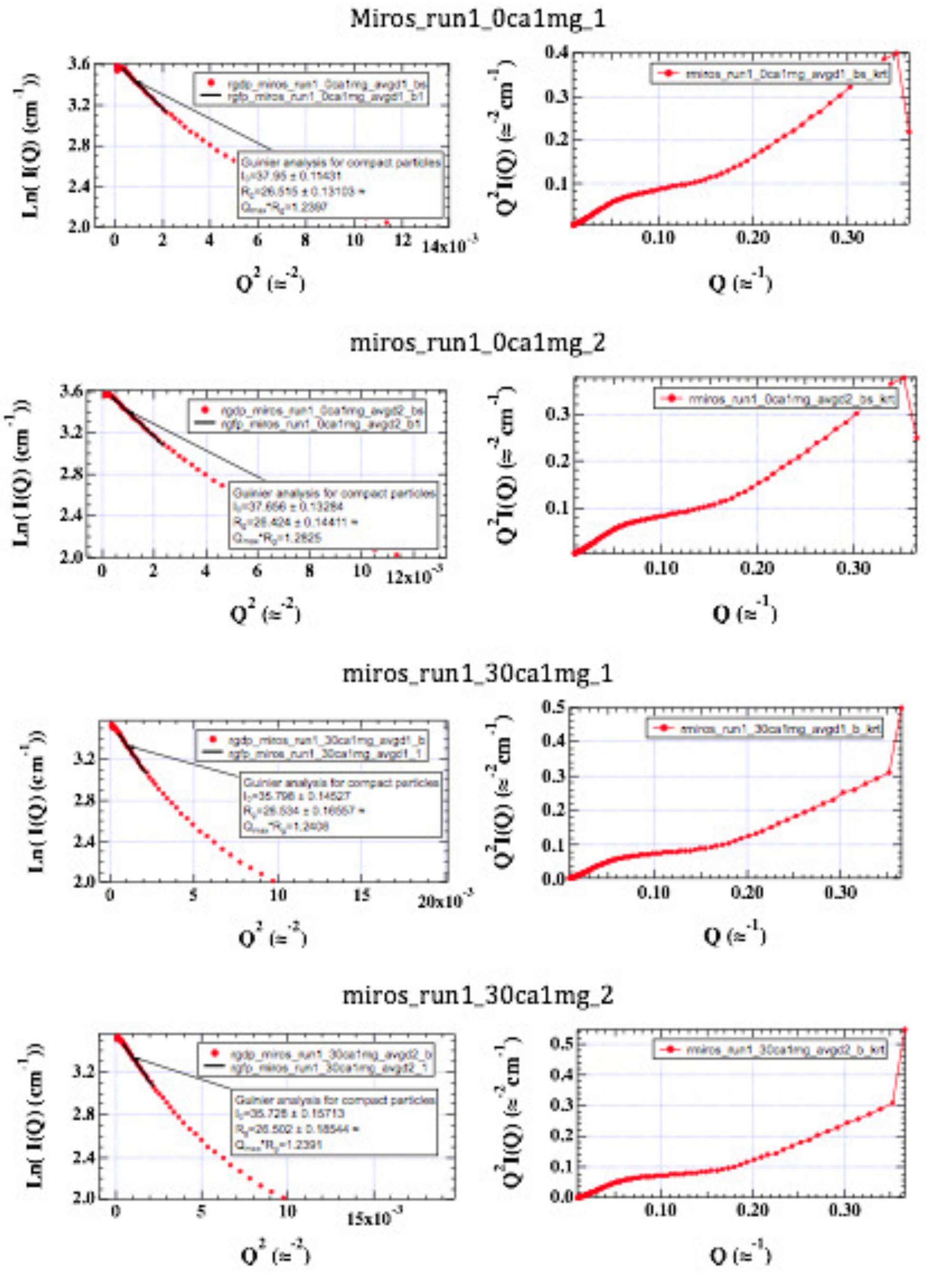


Figure 7: Guinier and Kratky plots for miro-S

.



Miros_run1_0ca1mg_1

miros_run1_0ca1mg_2

miros_run1_30ca1mg_1

miros_run1_30ca1mg_2

Figure 8: Guinier and Kratky plots for miro-S

.

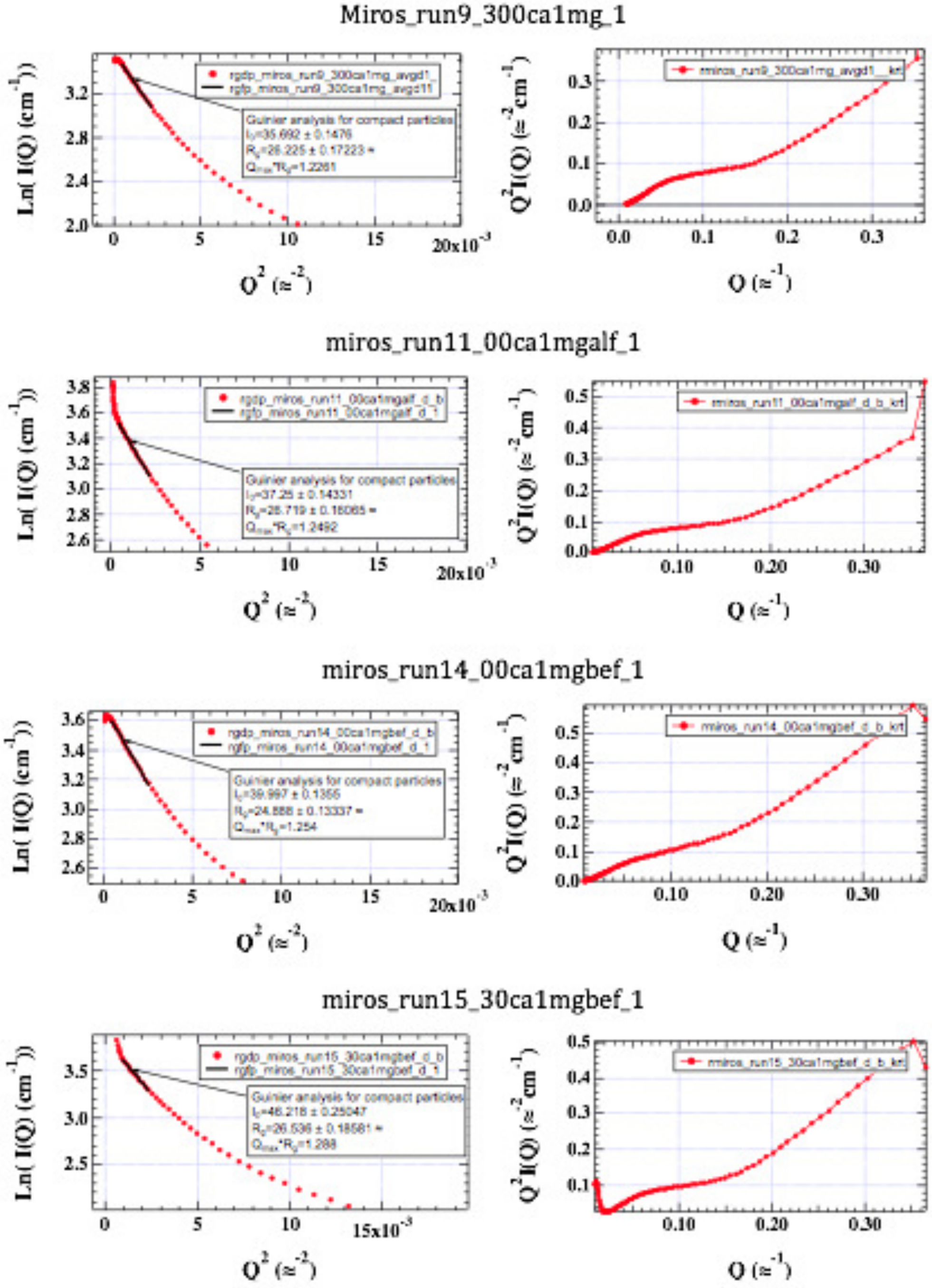


Figure 9: Guinier and Kratky plots for miro-S

.

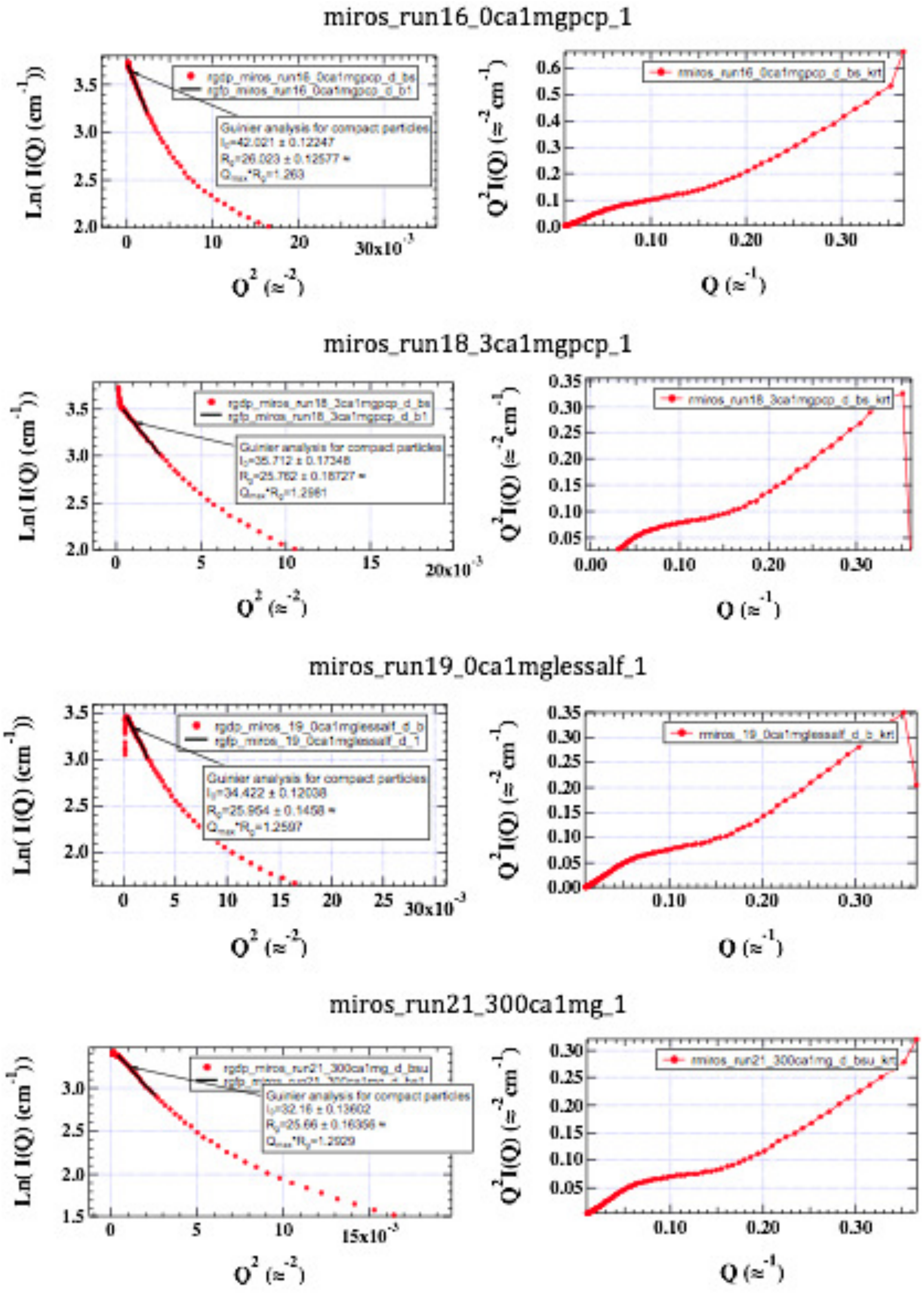


Figure 10: Guinier and Kratky plots for miro-S

.

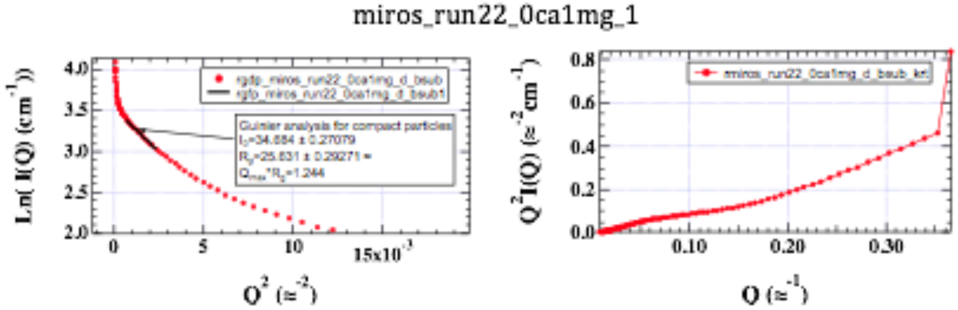
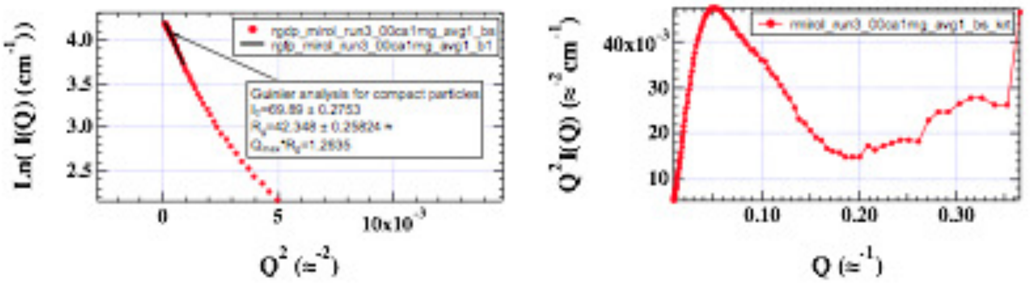


Figure 11: Guinier and Kratky plots for miro-S

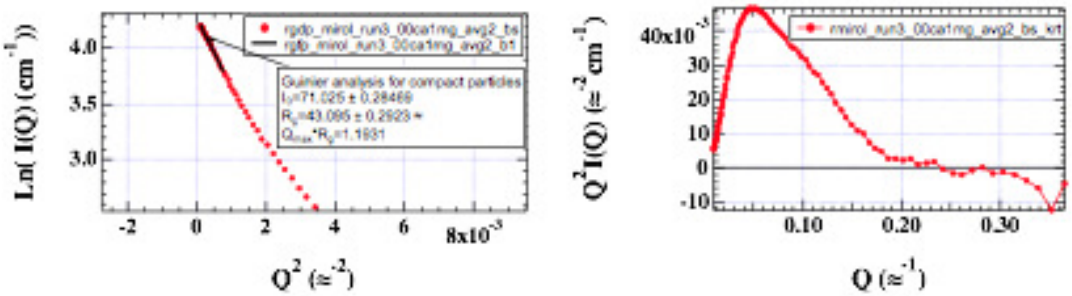
.

MIRO-L GUNIER AND KRATKY PLOTS

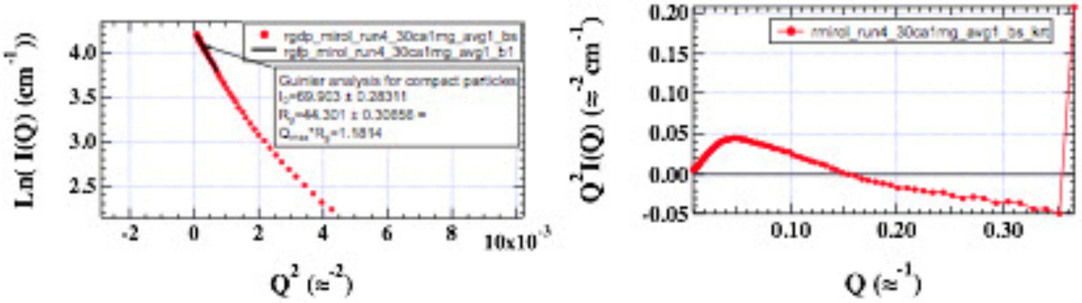
File: mirol_run3_00ca1mg_1



File: mirol_run3_00ca1mg_2



File: mirol_run4_30ca1mg_1



File: mirol_run4_30ca1mg_2

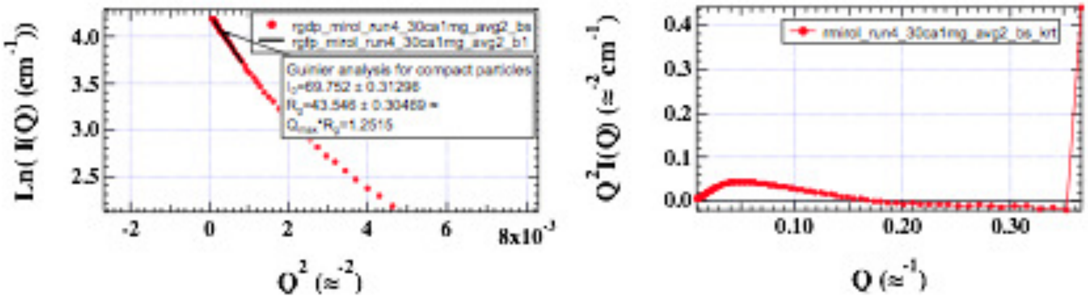
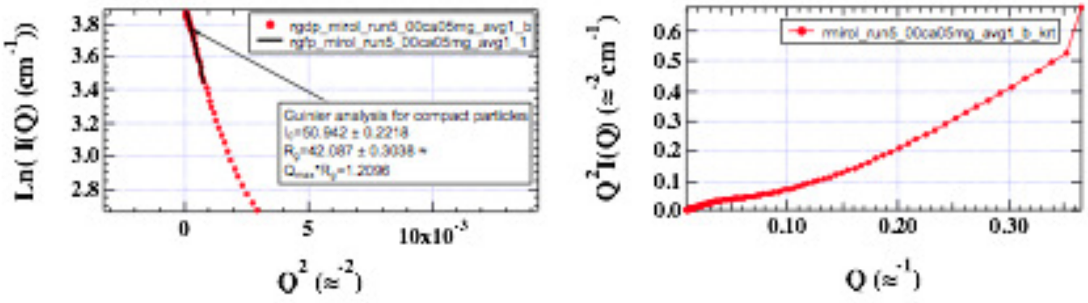


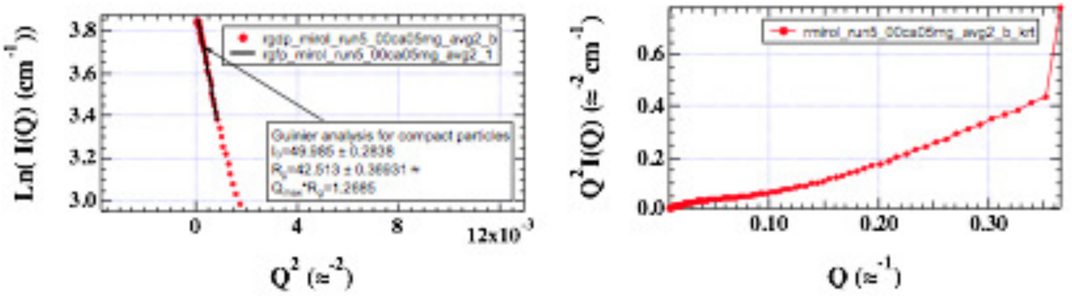
Figure 12: Guinier and Kratky plots for miro-L

.

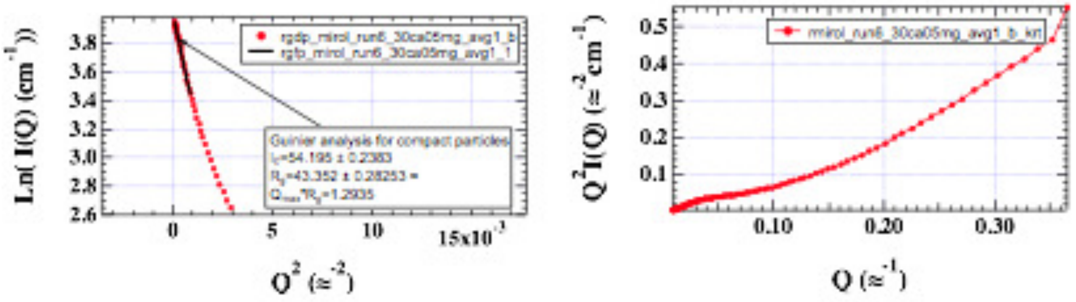
File: mirol_run5_00ca05mg_1



File: mirol_run5_00ca05mg_2



File: mirol_run6_30ca05mg_1



File: mirol_run7_00ca05mgcpp_1

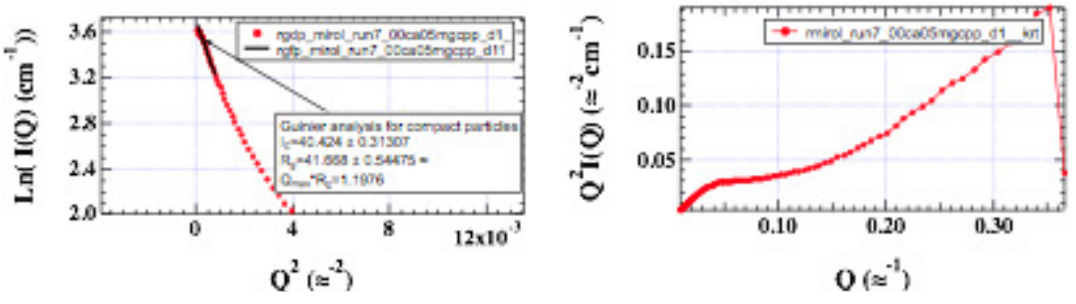
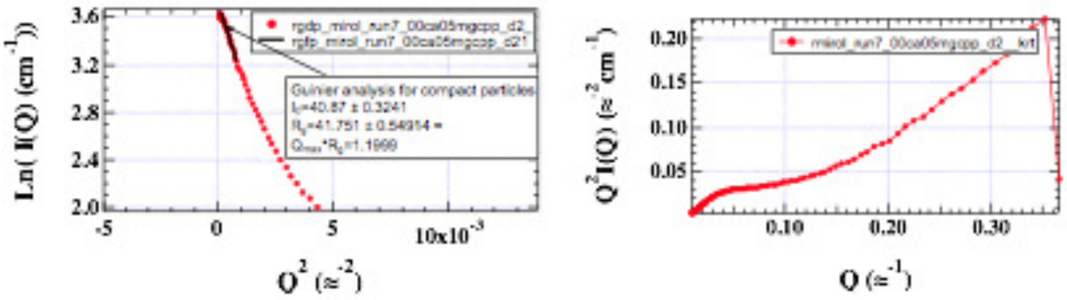


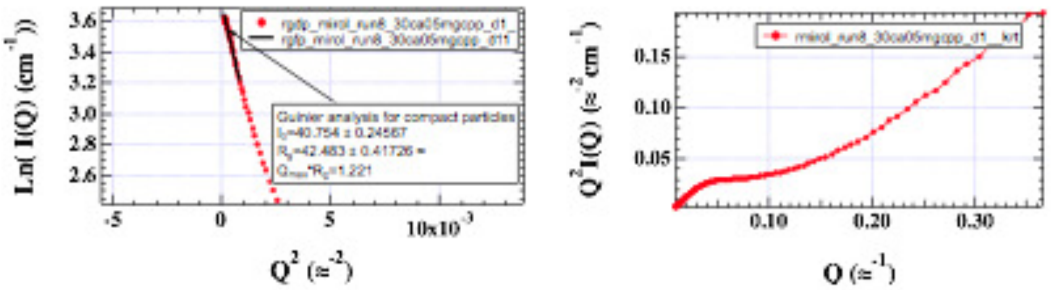
Figure 13: Guinier and Kratky plots for miro-L

.

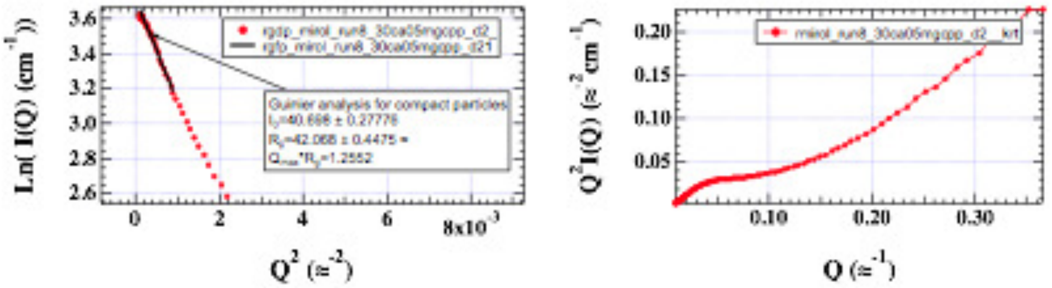
File: mirol_run7_00ca05mgcpp_2



File: mirol_run8_30ca05mgcpp_1



File: mirol_run8_30ca05mgcpp_2



File: mirol_run10_300ca05mg_1

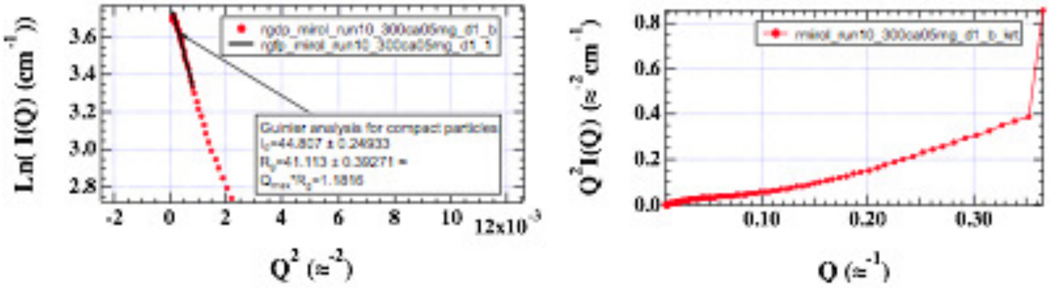
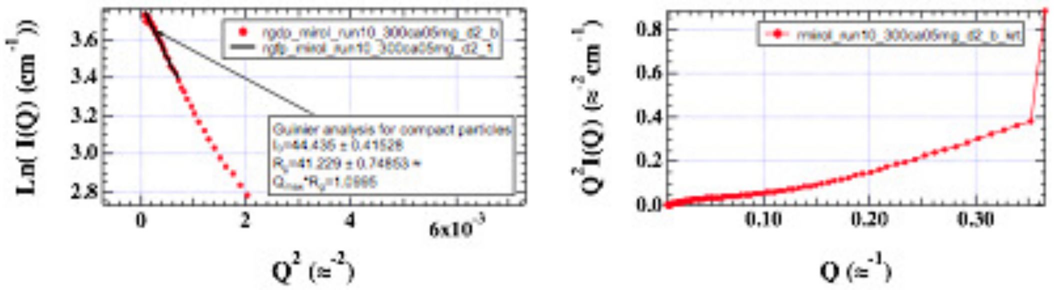


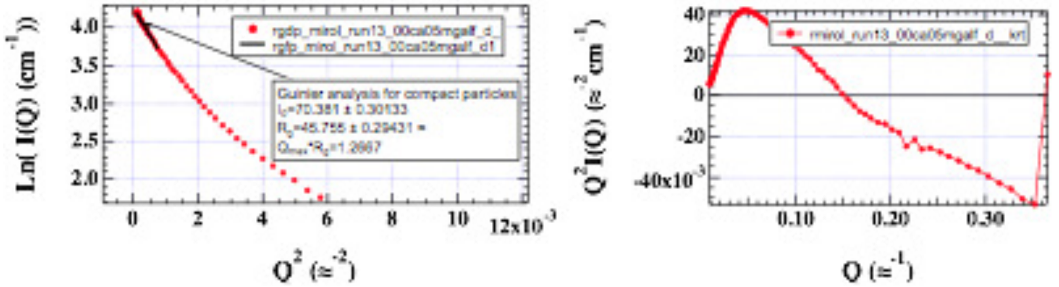
Figure 14: Guinier and Kratky plots for miro-L

.

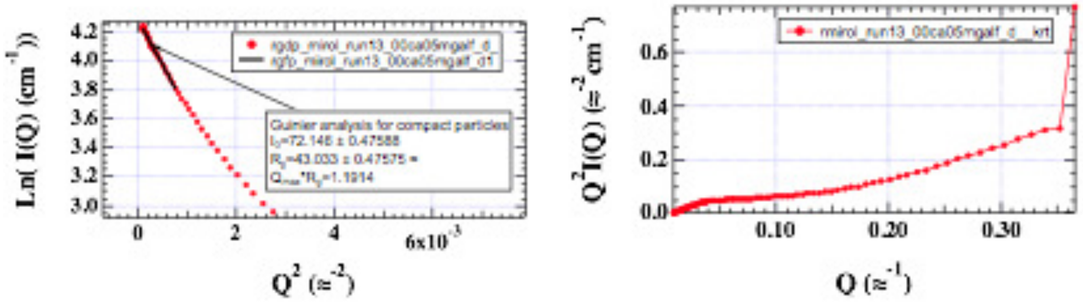
File: mirol_run10_300ca05mg_2



File: mirol_run13_00ca05mgalf_1



File: mirol_run13_00ca05mgalf_2



File: mirol_run23_0ca1mg_1

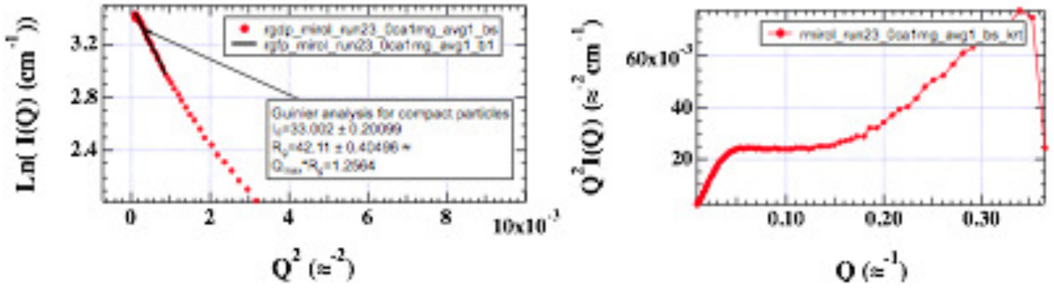
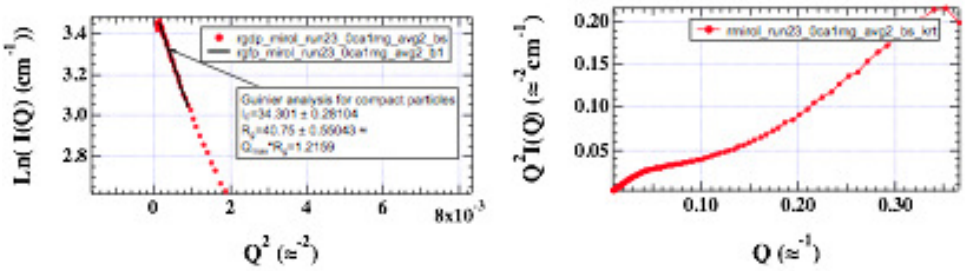


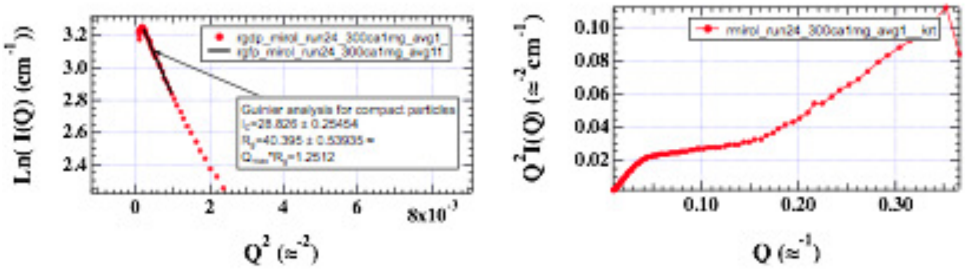
Figure 15: Guinier and Kratky plots for miro-L

.

File: mirol_run23_0ca1mg_2



File: mirol_run24_300ca1mg_1



File: mirol_run24_300ca1mg_2

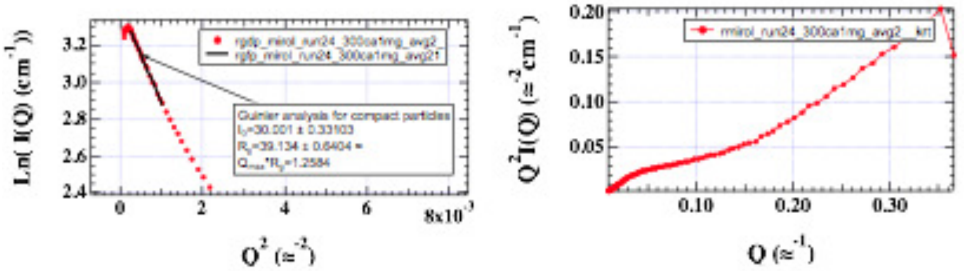


Figure 16: Guinier and Kratky plots for miro-L

.

REFERENCES

Akiyama, Shuji, and Satoshi Takahashi, *Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering*, PNAS, 99.3, (2002), 1329-1334.

Bilsel, Osman, and C Robert Matthews, *Molecular dimensions and their distributions in early folding intermediates*, Current Opinion in Structural Biology, 16, (2006), 86-93.

Brulet A, Boue F, Cotton JP, *About the Experimental Determination of the Persistence Length of Wormlike Chains of Polystyrene*, J. Phys. II France, (1996) 885-891.

Bushnell GW, Louie GV, Brayer GD, *High-resolution three-dimensional structure of horse heart cytochrome c*, J Mol Biol., (1990), Jul 20, 214(2), 585-95.

Charbonnier JB, Renaud E, Miron S, Le Du MH, Blouquit Y, Duchambon P, Christova P, Shosheva A, Rose T, Angulo JF, Craescu CT., *Structural, thermodynamic, and cellular characterization of human centrin 2 interaction with xeroderma pigmentosum group C protein*, J Mol Biol., (2007), Nov 2, 373(4), 1032-46, Epub 2007 Aug 25.

Debye, P., Zerstreuung von rontgenstrahlen, *Scattering from non-crystalline substances*, Ann. Phys., (1915), 46, 809-823.

Dill, Ken, and Sabrina Bromberg, *Molecular Driving Forces*, New York, Garland Science, (2003), 142-144.

Dill, Ken, and Ghosh, Kingshuk, *Computing protein stabilities from their chain lengths*, PNAS, Vol. 106, No. 26, June 30 (2009), 10649-10654.

Dobson, C.M., Sali, A., and Karplus, M., *Protein folding: a perspective from theory and experiment*, Angew. Chem. Int. Ed. Eng., (1998), 37, 868.

Elmer, Margaret, *Direct Visualization of Cold Denaturation of Cytochrome-C Using Small Angle X-ray Scattering*, MS Thesis, DePaul University Physics Dept. (2010).

Fischetti, R, S Stepanov, G Rosenbaum, R Barrea, and E Black, *The BIOCAT undulator beamline 18ID: a facility for biological non-crystalline diffraction and x-ray absorption spectroscopy at the Advanced Photon Source*, Journal of Synchrotron Radiation 11, (2004), 399-405.

Glatter, O., *A new method for the evaluation of small-angle scattering data*, Journal of Applied Crystallography, (1977), 10, 415421.

Glater, E.E., L.J. Megeath, R.S. Stowers, and T.L. Schwarz, *Axonal transport of mitochondria requires milton to recruit kinesin heavy chain and is light chain independent*, J. Cell Biol, (2006), 173, 545557.

Guinier, A., *X-ray diffraction in cyrstals, imperfect crystals, and amorphous bodies*, Dover Publications, (1994), 323.

Hirshberg M, Stockley RW, Dodson G, Webb MR., *The crystal structure of human rac1, a member of the rho-family complexed with a GTP analogue*, Nat Struct Biol., (1997) Feb, 4(2), 147-52.

Ho YP, Yang MW, Chen LT, Yang YC., *Relative calcium-binding strengths of amino acids determined using the kinetic method*, Rapid Commun Mass Spectrom., (2007), 21(6), 1083-9.

Hsu, IJ, and YJ Shiu, *A solution study on the local and global structure changes of cytochrome c: an unfolding process induced by urea*, J Phys Chem A 111.38, (2007), 9286-90.

Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P., *Optimization by Simulated Annealing*, Science, New Series, Vol. 220, No. 4598, May 13 (1983), pp. 671-680.

Landahl, Asta, Banks, Elmer, GrandPre, *Low resolution structures of cold, warm, and chemically denatured cytochrome-c via SAXS*, Poster presentation at the 2013 Biophysics Society Annual Meeting

MacAskill, A.F., and Kittler, J.T., *Control of mitochondrial transport and localization in neurons*, Trends in Cell Biology, Volume 20, Issue 2, December 18 (2009), 102-112.

Margoliash E, Smith EL, Kreil G, Tuppy H, *Amino-acid sequence of horse heart cytochrome c*, Nature, Dec 23 (1961), 192, 11251127.

Nelson, David L., and Michael M. Cox, *Principles of Biochemistry*, 4th ed. New York, W.H. Freeman and Company, (2005), 146-154.

Pain, Roger, *Mechanisms of protein folding*, 2nd ed., Avon, United Kingdom, Oxford University Press, USA, (2000), 1-144.

Pollack, Lois, and Mark W. Tate, *Compactness of a protein measured by submillisecond smallangle x-ray scattering*, Proc. Natl. Acad. Sci. Usa. 96, (1999), 10115-10117

Rice, Sarah E., and Gelfand, Vladimir I., *Paradigm lost: milton connects kinesin heavy chain to miro on mitochondria*, Journal of Cell Biology, Vol. 173, No. 4, (2006).

Robertson, Andrew D, and Kenneth P Murphy, *Protein Structure and the Energetics of Protein Stability*, American Chemical Society, 96, 383-4, (1997)

Schnablegger, Heimo, and Singh, Yashveer, *The SAXS guide, getting acquainted with the principles*, 2nd ed., Anton Paar GmbH, (2011).

Schneidman-Duhovny, D., Hammel, M., Sali, A., *FoXS: A Web server for Rapid Computation and Fitting of SAXS Profiles*, NAR (2010), 38 Suppl: W540-4.

Svergun, D.I., *Determination of the regularization parameter in indirect-transform methods using perceptual criteria*, J. Appl. Crystallogr. 25, (1992), 495-503.

Svergun, D.I., *Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing*, Biophys J., (1999), 2879-2886.

Vachette, P., *Scattering of X-rays, EMBO Practical Course on Solution Scattering from Biological Macromolecules*, Hamburg, October 25th November 1st (2010).

Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R.F., Sykes, B.D., Wishart, D.D., *VADAR: a web server for quantitative evaluation of protein structure quality*, Nucleic Acids Res., July 1 (2003), 31 (13): 3316.3319.