1.18 Small and Wide Angle X-ray Scattering from Biological Macromolecules and their Complexes in Solution

S Doniach, Stanford University and SLAC National Accelerator Laboratory, Stanford, CA, USA
J Lipfert, Delft University of Technology, Delft, The Netherlands

1.18.1	**Introduction**	376
1.18.2	**Scattering Basics**	377
1.18.2.1	Background Subtraction	377
1.18.2.2	Determination of the Radius of Gyration R_g and the Pair Distribution Function $P(r)$	378
1.18.2.3	Effects of Interparticle Interference	379
1.18.2.4	Measurement of Interparticle Potentials Using the SAXS Interparticle Interference	380
1.18.3	**Effects of Hydration**	380
1.18.3.1	*Ab Initio* Methods for Estimating the Hydration Shell	381
1.18.4	**Three-Dimensional Reconstruction of Low Resolution Density Maps from SAXS Data**	381
1.18.4.1	Estimating the Resolution of 3-D-Reconstructed Density Maps	382
1.18.4.2	Example: Optimal Number of Beads Justified by Cytochrome c Experimental Data	382
1.18.4.3	Use of SAXS Data to Help Determine Protein Structure from Sequence; High-Throughput Analysis	383
1.18.5	**Disordered Peptides and Proteins in Solution: Free Energy Landscape for Denaturation**	383
1.18.5.1	Importance of Solvation Free Energy and Hydration for Disordered Peptides and Proteins	384
1.18.6	**Calculation of SAXS Profiles for Detergent Micelles and Protein–Detergent Complexes**	385
1.18.6.1	Micelle Mixtures and Protein–Detergent Complexes	386
1.18.6.2	Atomic Scale Modeling of SAXS from Detergent Micelles and Protein–Detergent Complexes (PDCs)	386
1.18.7	**Combination of SAXS and NMR for Structure Determination of Multi-Domain Proteins**	387
1.18.8	**Solution Structure of Small Functional RNAs**	388
1.18.8.1	SAXS as a Filter for Candidate RNA Structures on the Pathway to a Fully Folded State	389
1.18.8.2	Using Low Resolution Bead Models to Estimate Free Energies as a Function of Salt Concentration	391
1.18.9	**Use of Nanogold Markers for SAXS Determination of Intramolecular Distance Distributions**	393
1.18.9.1	Possible Use of Anomalous SAXS to Extract Distance Distributions from Gold-Labeled Biomolecules	395
Acknowledgments		395
References		395

Abbreviations

FRET	Förster Resonance Energy Transfer
MD	molecular dynamics
MSG	Malate Synthase G
NLPB	numerical solution of the non-linear Poisson-Boltzmann equation
NOE	nuclear Overhauser effect
PB	Poisson-Boltzmann
PDC	protein–detergent complex
RISM	Reference Interaction Site Model
SANS	small-angle neutron scattering
SAXS	small-angle X-ray scattering
SDS	sodium dodecyl sulfate
SVD	Singular Value Decomposition
TPP	thiamine pyrophosphate

1.18.1 Introduction

The use of X-ray scattering to study the structure of biomolecules in solution dates back to the first half of the twentieth century.[1] However, because of the low scattering contrast between low-Z elements and background water, the ratio of signal-to-background is quite low, so that the usefulness of the technique with conventional X-ray sources is somewhat limited. Although the use of X-ray tube sources is still practical today, obtaining a useable SAXS profile takes a matter of hours.

With the advent of high intensity synchrotron radiation sources in the 1970s, X-ray scattering, particularly at small angles where the scattered intensity is highest, obtaining useful data from biomolecules under a large range of solvent conditions, becomes a matter of minutes, or more recently seconds. Time-resolved SAXS becomes feasible under these conditions when suitably averaged.

An important advance in relating scattering data to molecular structure came about with the development of increasingly powerful software for application of non-linear refinement methods to obtain shape information about biomolecules in solution. When coupled with the increasing power of computation, this has led to the significant impact of X-ray scattering in the field of structural molecular biology over the last decade or so.

A number of general reviews of the use of SAXS for biomolecules in solution are available which cover a wide range of aspects of the technique.[2–6] In this chapter, we focus on recent developments in the use of SAXS and WAXS for a variety of problems in structural molecular biology. We make no attempt at a comprehensive overview of the field, but aim to point out promising directions in the use of X-ray scattering, some of which are still under development.

Of course, the main source of information on biomolecular structure remains X-ray crystallography. However, X-ray scattering works for molecules in solution under physiological conditions, therefore, it can give complementary information, albeit at lower spatial resolution, to that obtainable by crystallography. It is important to note that X-ray scattering also complements a number of other structural techniques which do not depend on the formation of crystals, including ultracentrifugation, light scattering, NMR, EPR, electron microscopy and FRET. So, understanding a given structural problem should make use of information from as many techniques as are available.

1.18.2 Scattering Basics

SAXS measurements are in principle simple to carry out: the solution containing the biomolecule in a sample cell or a capillary is placed in the X-ray beam (see the reviews cited in the Introduction). The scattered X-rays are recorded by an area CCD detector, or, more recently, by a 2-D pixel detector[109] **Figure 1** illustrates some of the basic concepts by which the model-free parameters, radius of gyration R_g and the forward scattering intensity $I(0)$ may be extracted from the data in a very simple way. In the section below, we discuss how the observed SAXS profile $I(q)$ may be related to the crystal structure, for molecules of known structure.

1.18.2.1 Background Subtraction

The amplitude for scattering of X-rays of wavelength λ through an angle θ from a set of identical molecules at positions $\vec{R}_j$ in a

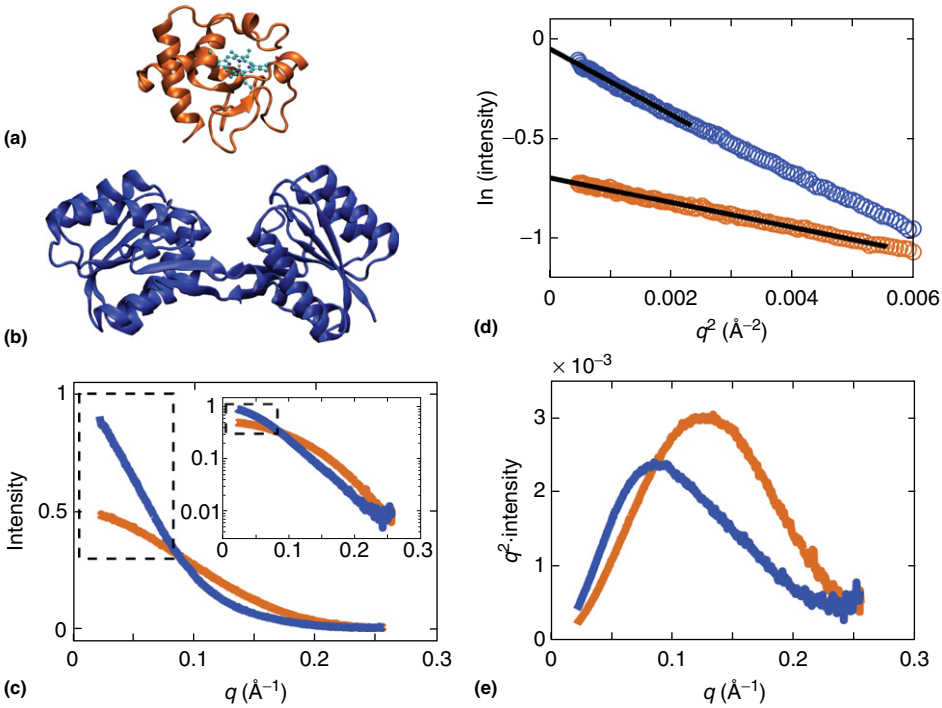

Figure 1 SAXS scattering profiles and basic analyses for two differently-sized proteins. (a) Crystal structure of cytochrome *c* (PDB accession code 1CRC). The protein is shown in cartoon representation and the heme is rendered in ball-and-stick representation. Cytochrome *c* has a molecular weight of 11.8 kDa. (b) Crystal structure of the *apo* form of ribose binding protein (i.e., in the absence of ribose) (PDB accession code 1URP). Ribose binding protein has a molecular weight of 28.5 kDa. (c) SAXS profiles of cytochrome *c* (orange) and ribose binding protein (blue) shown as a linear plot of I vs. q and a logarithmic plot ln I vs. q in the inset. The dashed rectangle indicates the low scattering angle range that is used in the Guinier analysis (panel (d)). It is apparent that the scattering intensity falls much more rapidly with increasing q for the larger ribose binding protein, compared to the smaller cytochrome *c*. (d) Guiner plot $\ln(I(q))$ vs. q^2 of the low q region for the data shown in panel (c). The black lines indicate the fitting ranges, chosen such that $q_{max}\,R_g \approx 1$ for both proteins. This criterion implies a considerably smaller q_{max} for the larger ribose binding protein. The fitted R_g values are 13.8 ± 0.5 Å for cytochrome *c* and 21.6 ± 0.5 Å for ribose binding protein, in good agreement with the values of 13.7 Å and 21.5 Å, respectively, computed from the crystal structures using the software crysol. Reproduced from Svergun, D. I.; Barberato, C.; Koch, M. H. J. Evaluation of x-ray solution scattering curves from atomic models. *J. Appl. Crystallogr.* **1995**, *28*, 768–773. (e) Kratky representation of q^2I vs. q of the same two scattering profiles. The Kratky representation is particularly useful for monitoring conformational changes occurring in protein or RNA folding. As expected for well-folded proteins, both scattering profiles show a clear peak in the Kratky representation. The peak is relatively symmetric for the approximately spherical cytochrome *c* and more asymmetric for the more elongated ribose binding protein.

vacuum is given by[7]

$$F(\vec{q}) = \sum_{R_j, r_i} f_i e^{i\vec{q}\cdot(\vec{R}_j + \vec{r}_i)} \quad [1]$$

where the scattering vector $\vec{q}$ is oriented along the scattering direction and $|q| = 4\pi \sin(\theta/2)/\lambda$. $\vec{r}_i$ are the atomic positions within the molecule measured relative to $\vec{R}_j$ and f_i are the atomic scattering amplitudes. In solution, the scattering amplitude is given by the difference of eqn [1] and the scattering by an identically shaped volume of the solvent averaged over the molecular positions of the solvent (usually water), or 'ghost' scattering:

$$F(\vec{q}) = \sum_{R_j, r_i} f_i e^{i\vec{q}\cdot(\vec{R}_j + \vec{r}_i)} - F_{\text{ghost}}(\vec{q}) \quad [2]$$

In reality, once the molecule is solvated, it also acquires a solvation shell of water molecules (roughly 3 Å thick) which should be included in the atomic coordinates $\vec{r}_I$ of the molecule itself. This will be further discussed below.

At sufficiently low concentrations, the interparticle scattering interference may be neglected and the scattering amplitude per molecule reduces to

$$F(\vec{q}) = \sum_{i} f_i e^{i\vec{q}\cdot\vec{r}_i} - F_{\text{ghost}}(\vec{q}) \quad [3]$$

with resulting scattered intensity per molecule given by the integral over all orientations Ω of the molecular axis relative to the scattering vector direction by

$$I(q) = \frac{1}{4\pi} \int d\Omega \left| \sum_i f_i e^{i\vec{q}\cdot\vec{r}_i} - F_{\text{ghost}}(\vec{q}) \right|^2 \quad [4]$$

For small angle ($q < 0.25$ Å^{-1}) X-ray scattering (SAXS), the q-dependence of the f_i may be neglected, and the f_i are given by the number of electrons in the ith atom. However, at large angles (WAXS) the q-dependence of the atomic form factors (as used in X-ray crystallography) needs to be included.

A simple way to take into account the effect of the solvent background subtraction is to replace the atomic form factors f_i by the number of electrons/atom(i) minus the equivalent number of solvent electrons in the volume occupied by the type of atom at position $\vec{r}_i$. This is done in the program CRYSOL by Svergun and collaborators.[8]

There have historically been a number of methods published for numerical evaluation of the angular integral in $I(q)$. These will be discussed below and include explicit sampling of the unit sphere and the spherical harmonic expansion method used in CRYSOL.

Traditionally, the analysis of SAXS for simple geometrical models has been analyzed in terms of the Debye integral:

$$I(q) = \frac{1}{2} \int d\cos\theta_{\vec{q}} \sum_{i,j} f_i f_j e^{i\vec{q}\cdot(\vec{r}_i - \vec{r}_j)} = \sum_{i,j} f_i f_j \frac{\sin(qr_{ij})}{qr_{ij}} \quad [5]$$

where we now denote the f_i to include the background subtraction as defined above. This may then be written in terms of the distribution of a weighted histogram of interatomic distances within the molecule, $P(r) = \sum_{ij} f_i f_j \delta(r - r_{ij})$

$$I(q) = \int_0^\infty dr P(r) \frac{\sin qr}{qr} \quad [6]$$

However, use of this formula for randomly placed particles (such as water molecules or bead models) can lead to problems as will be discussed below. Nevertheless, the Debye formula is useful for understanding the problems of extracting the radius of gyration, R_g and the pair distribution $P(r)$, so we will continue to use it in the following discussion.

1.18.2.2 Determination of the Radius of Gyration R_g and the Pair Distribution Function $P(r)$

Guinier's method[9] for calculation of R_g involves a linear least square fit of the (non-linear) log of $I(q)$ based on the approximation valid at low q:

$$I(q) \approx \left(\sum_i f_i \right)^2 e^{-R_g^2 q^2/3} \quad [7]$$

from which one may extract R_g^2 and the $q = 0$ intercept, $I(0)$, by a linear fit of ln $I(q)$ vs q^2:

$$\ln(I(q)) \approx \ln(I(0)) - \frac{R_g^2 q^2}{3} \quad [8]$$

Note that $I(0)$ is proportional to the square of the contrast between the number of electrons in the molecule and the background solvent: $I(0) = \left(\sum_i f_i\right)^2$. For proteins in water, the contrast reduces the effective number of electrons in the molecule by approximately a factor of 10.

Viewed from the point of view of Fourier transforms and Shannon's sampling theorem, one might expect the accuracy of R_g to be limited by the maximum scattered wave vector q_{max}.

For $q_{max} \approx 0.25$ Å^{-1}, this is given by $2\pi/0.25 = 25$ Å. However, Guinier's method is an example of super-resolution, recently exploited as a powerful tool in optical fluorescence studies of molecules:[10–12] by non-linear fitting of the expected optical signal, a parameter – in this case, R_g – may be evaluated at much greater accuracy than expected from linear diffraction theory. The accuracy of the evaluated R_g from the linear regression fit to the log is generally found to be much higher than found experimentally. That is, repeated measurements tend to show a larger variance in the measured R_g. This is due to systematic errors resulting from a number of additional factors. These include uncertain sample concentrations, effects of interparticle interference, possible storage ring beam motions (SAXS is generally more sensitive than the beam position monitors in many storage rings), etc. So in the end, the actual error in R_g needs to be determined empirically.

For highly anisotropic or disordered molecules, the range of validity of eqn [8] is much less than $q \approx \pi/R_g$ expected for a sphere (see discussion in the review by Doniach[2]). Svergun and Koch[6] give an empirical fitting range criterion valid for most macromolecules $q_{max} R_g \leq 1.3$, better yet ≤ 1.0. Another limitation is for data where there are large error bars for $I(q)$ at

low q from dilute samples, samples with particularly low contrast, or short measurement time in the case of time-resolved measurements. In this case, we have found that bootstrap sampling[13] allows for an accurate estimation of the expected errors of R_g. In this method, a limited range of measured points in the Guinier region, $\{I(q_j)\}$ are randomly picked and the linear regression is applied. For example, if 20 sample points are measured, an ensemble of, say, 50 points is picked at random from this set of 20. Some points are repeated a number of times and others dropped out. This procedure is repeated a large number of times (of order 100–1000 s) resulting in a histogram of R_g and $I(0)$ values from which a mean and variance are easily estimated.

Extraction of the pair distribution function $P(r)$ from SAXS data may be expressed as a linear least squares minimization problem of the function χ^2:

$$\chi^2 = \sum_j \left| \frac{I(q_i)}{\sigma_i} - \sum_j \left\{ \frac{\sin(q_i r_j)}{q_i r_j} \right\} P(r_j) \right|^2 \qquad [9]$$

where the q_i are a set of measurement points with experimental errors σ_i and the r_j the desired set of distances, leading to a matrix inversion $P(r_i) = \mathbf{K}^{-1}|_{ij} I(q_j)$ with kernel $\mathbf{K}_{ij} = \sin(q_j r_i)/q_j r_i$. However, the matrix $\mathbf{K}_{ij}$ is highly ill-conditioned and, in practice, errors in $I(q)$ are too amplified to allow for a simple linear inversion to be possible.

This type of problem may be attacked using prior knowledge of the expected form of $P(r)$ (see discussion in 'Numerical Recipes'[14]). This prior knowledge is applied in the form of a non-linear constraint on the form of the solution of the minimization problem. This was applied by Svergun and colleagues in their GNOM algorithm[15] by the method of Tikhonov.[16] This leads to a smoothness constraint in the kernel of the integral equation

$$\chi^2 = \sum_j \left| \frac{I(q_i)}{\sigma_i} - \sum_j \left\{ \frac{\sin(q_i r_j)}{q_i r_j} \right\} P(r_j) \right|^2 + \alpha\Phi \qquad [10]$$

where

$$\Phi = \delta r \sum_j P(r_j)^2 + \sum_j \{P(r_j) - P(r_{j-1})\}^2. \qquad [11]$$

This may be solved by linear least squares for the set of $\{P(r_i)\}$ unknowns, but the results depend in a highly non-linear way on the constraint parameter α. At this point, the procedure for extracting $P(r)$ contains a number of assumptions leading to a plausible solution. This leads to a method of fixing a desirable value of α which includes several criteria: smoothness, stability, absence of systematic deviations, etc. Some of these criteria are improved by increasing α, others are improved by decreasing α. For ways in which the current version of GNOM selects a compromise value for α, the reader should see the notes in the current release of GNOM.

An important user-input is the expected maximum pair distance in the evaluation of $P(r)$. This needs to be evaluated by trial and error. The usual criterion applied is that $P(r)$ should go smoothly to zero as $r \to r_{max}$. This may be determined by running GNOM for a series of r_{max} incremented by, say 0.5 Å, in successive runs and picking the value which gives the smoothest decrease at r_{max}.[15,17]

1.18.2.3 Effects of Interparticle Interference

As the molecular concentration is increased, the approximation of treating the SAXS as coming from isolated molecules starts to break down. Starting from the general expression

$$F(\vec{q}) = \sum_{R_j, r_i} f_i e^{i\vec{q}\cdot(\vec{R}_j + \vec{r}_i)} \qquad [12]$$

(where background subtraction is implied) we can approximate the resulting scattering as

$$I_{total} = |F(\vec{q})|^2 = \left| \sum_{R_j} e^{i\vec{q}\cdot\vec{R}_j} \sum_{r_i} f_i e^{i\vec{q}\cdot\vec{r}_i} \right|^2. \qquad [13]$$

The thermodynamic average of the scattering may then be written as

$$\langle I_{total}(q) \rangle = \int d^3R_i d^3R_j \sum_{R_i, R_j} \langle e^{i\vec{q}\cdot(\vec{R}_i - \vec{R}_j)} F_i^*(\vec{q}) F_j(\vec{q}) \rangle \qquad [14]$$

where $F_i(\vec{q})$ is the scattering amplitude from the oriented isolated molecule at position R_i. In the simplest approximation, neglecting angular dependence of inter-molecular interactions, this may be approximated in terms of spherically isotropic scattering from isolated molecules as

$$\langle I_{total}(q) \rangle \approx \int d^3R e^{i\vec{q}\cdot\vec{R}} \rho(\vec{R}) I(q) = S(q) I(q) \qquad [15]$$

where $S(q)$ is the Fourier transform of the pair distribution function of the molecules in the solution

$$\begin{aligned} S(q) &= \left\{ \int d^3R e^{i\vec{q}\cdot\vec{R}} \delta\left(\vec{R}\right) \right. \\ &\quad \left. + \int d^3R_i d^3R_j \sum_{i \neq j} e^{i\vec{q}\cdot(\vec{R}_i - \vec{R}_j)} \left(1 + e^{-\beta V(\vec{R}_i - \vec{R}_j)}\right) \right\} \\ &= \left(1 + \frac{N}{\mathrm{Vol}} \int d^3R_{ij} \sum_{i,j} e^{i\vec{q}\cdot\vec{R}_{ij}} e^{-\beta V(R_{ij})}\right) \end{aligned} \qquad [16]$$

Here, the first term is from the self-scattering ($i = j$) terms. Vol_{nbor} is a normalization volume for the pair distribution integral. For $V =$ hard core repulsion at radius R_0, the interference term is

$$\begin{aligned} \frac{S(q)}{N} &= \left(1 + c \int_{R_0}^{\infty} R dR \frac{\sin(qR)}{qR}\right) \\ &= 1 + c\left(\frac{R_0}{q^2}\cos(qR_0) - \frac{\sin(qR_0)}{q^3}\right) \end{aligned} \qquad [17]$$

where c is the concentration relative to the volume swept out by the nearest neighbor $= N/\mathrm{Vol}_{nbor}$. In the limit $q \to 0$, we then have

$$S(q)/N - 1 \propto -cR_0^3 \qquad [18]$$

so in the case of hard core repulsion, the interparticle interference reduces the forward scattering by an amount corresponding to the number of neighboring molecules in the

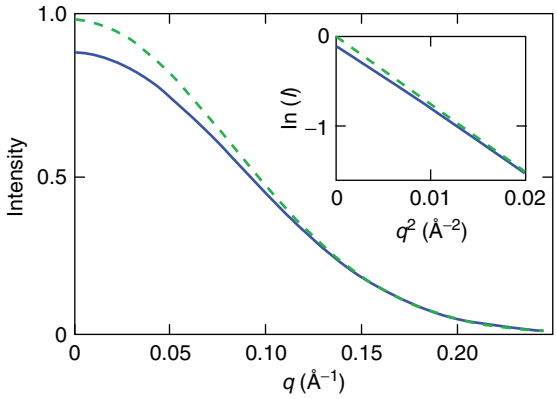

Figure 2 Illustration of the effects of interparticle interference on a SAXS profile. $I(q)$ for the Guinier expression $\exp(-q^2R_g^2/3)$ with $R_g = 15$ Å (green dashed curve) and the same profile with interparticle interference added (using $R_0 = 25$ Å; solid blue curve). The inset shows a Guinier plot $\ln(I(q))$ vs q^2 of the same data. Repulsive interactions lead to a reduction of the forward scattering and apparent radius of gyration compared to the dilute limit.

normalization volume $\mathrm{Vol}_{\mathrm{nbor}} = 4\pi/3\{(2R_0)^3 - R_0^3\}$. As q increases, the interference scattering factor $S(q)$ decreases (see Figure 2), leading to an apparent reduction of the radius of gyration. For this reason, it is important to measure R_g as a function of concentration and extrapolate towards $c \to 0$ to obtain a robust estimate of R_g.

In the case of an attractive intermolecular interaction, the probability of finding a neighbor within $2R_0$ is increased rather than decreased so that the apparent R_g is increased. This leads to a plot of R_g vs c as increasing as $c \to 0$ rather than decreasing as in the case of a repulsive potential.

1.18.2.4 Measurement of Interparticle Potentials Using the SAXS Interparticle Interference

As discussed above, the SAXS profile for a concentrated solution includes scattering from pairs of molecules. Within the spherical approximation, we can write

$$S(c,q) = 1 + \rho \int dr 4\pi^2 (g(r) - 1)(\sin qr/qr) \qquad [19]$$

where $\rho = cN_a/M$ is the number density of particles, M the macromolecular molecular weight (Da), N_a is Avogadro's number and $g(r)$ is the pair distribution function of the molecules. By expansion in powers of the concentration, $S(c, q)$ may then be related to the second virial coefficient resulting from the interparticle interaction via

$$1/S(c,0) = 1 + 2cMA_2 \qquad [20]$$

where A_2 is the second virial coefficient given by

$$A_2 = 2\pi N_a/M^2 \int r^2 dr(1 - \exp(-V(r)/k_B T)) \qquad [21]$$

where k_B is Boltzman's constant. Therefore, A_2 is positive for repulsive interactions and negative for attractive ones (see discussion by Tardieu et al.[18]).

By fitting the q-dependence of the scattering function $S(q)$ to a DLVO potential model (Derjaguin, Landau, Verwey, Overbeek[19]), Tardieu, Belloni and collaborators[20] have been able to study the phase diagram of proteins in solution, and in particular, the effects of salt and PEG (polyethylene glycol) on protein crystallization. The DLVO model is usually a good model for colloidal systems such as soluble proteins at low ionic strength (i.e., <0.2 M ionic strength). The model includes hard sphere, van der Waals and Coulombic interactions. With small compact proteins, the van der Waals forces were shown to be equivalent to a short range, about 3 Å, attractive Yukawa potential of 2–3 k_BT.[20,21]

1.18.3 Effects of Hydration

For X-ray scattering from biological macromolecules in solution, the effects on the scattering profile of the hydration shell which results from the interaction of the biomolecules with the surrounding water need to be taken into account. Although any individual water molecule is probably exchanging rapidly from a position near the molecular surface with the bulk water, the accumulation of waters around the molecule leads to a local change of electron density relative to that of bulk water (0.33 e^- Å^{-3}) which extends a distance of 3 Å or so out from the molecular surface and varies as one goes around the molecule depending on how the surface residues of the molecule interact with the water.[22,23]

These effects were included in the CRYSOL program of Svergun and collaborators[8] by assuming a shell around the surface of the molecule of thickness 3 Å and a constant value of the density relative to the bulk which is included as an adjustable parameter. This was tested out by comparing the resulting predictions for the X-ray scattering profile of lysozyme with experiments. The predictions were also made for small-angle neutron scattering (SANS) measurements where the contrast relative to the solvent may be varied by altering the concentration of D_2O relative to H_2O since scattering of neutrons from deuterium is much different from that from hydrogen. By adjusting the density of the solvation shell, pretty good agreement with experiment could be obtained, showing that for most proteins, CRYSOL does a reasonably good job of predicting the SAXS profile.[24]

However, since the introduction of CRYSOL in the 1990s, new situations have arisen where the need to replace CRYSOL with a more detailed description of hydration is becoming evident. These include the measurement of wide-angle scattering (WAXS) where the thickness of the water shell starts to be comparable to the Shannon resolution $2\pi/q_{\max}$, molecules such as GRO-EL with a topologically complex molecular surface (i.e., with cavities[25]), and molecules with highly variable electron density such as membrane protein–detergent complexes.[4] Another area where improvements to CRYSOL are needed is that of disordered peptides or proteins, including denatured proteins. Here hydrophobic surfaces are exposed to solvent (sometimes in the presence of a denaturant such as urea or guanidine hydrochloride). In such cases the nature of the water shell needs to be estimated for each member of the ensemble of random conformations, and the application of CRYSOL needs to be superseded by a more flexible way of

estimating the scattering profile. This is further discussed in the following paragraph.

A straightforward, if theory-dependent, way to estimate the solvation of a biomolecule is to run molecular dynamics (MD) with an atomic resolution model of the protein in explicit water.[26] This procedure generates time-resolved snapshots of the distribution of water (or other solvent components such as ions or urea) around the protein. However, generating SAXS profiles for a given snapshot presents a technical problem since use of the Debye formula for a random array of scatterers can lead to speckle-type singularities where the random phases of scattering from molecules in the array can be accidentally coherent leading to a large scattering in a particular (random) $\vec{q}$ direction.[27] These singularities can be avoided if the angular integral

$$I(q) = \int d\cos\theta_q \left|F(\vec{q})\right|^2 = \int d\cos\theta_q \left|\sum_i e^{i\vec{q}\cdot\vec{r}_i}\right|^2 \qquad [22]$$

leading to the Debye formula has a smoothing filter applied. A simple way to do this is by replacing the integral as a sum over a set of sample points on the unit sphere:

$$I(q) \approx \sum_j w_j \left|F(q,\Omega_j)\right|^2 \qquad [23]$$

The sample points Ω_j covering the unit sphere and corresponding weights w_j may be determined by the method of cubature.[28] Here the coarse graining, resulting from finite sampling of the continuous integral, smooths out the singularities which lead to speckle in the Debye formula.

Another way to do this smoothing is incorporated into the spherical harmonic expansion of $F(\vec{q})$ used in the CRYSOL algorithm[8]

$$F(\vec{q}) = \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} F_{l,m}(q) Y_{lm}(\Omega_q) \qquad [24]$$

Then by limiting the maximum order $l_{\max}$ of spherical harmonics used in the integral over Ω_q of $\sum_{lm}\left|F_{l,m}(q)Y_{lm}(\Omega_q)\right|^2$, a similar smoothing is obtained. This may then be applied to a pdb file containing the molecule of interest together with a sample of the surrounding MD waters. This would then need to be repeated a number of times to average over different snapshots of the MD water. However, for this to work, the contrast correction to the scattering factor f_{oxygen} needs to be defined for a water molecule in the solvent (as opposed to oxygens which are part of the protein). To our knowledge, this has not been tried to date.

For an example of application of the cubature method to estimate the WAXS profile for lysozyme in water, see the recent paper by Makowski and collaborators.[29] A further example is given in the membrane protein section (below).

1.18.3.1 *Ab Initio* Methods for Estimating the Hydration Shell

An alternative to using MD methods to estimate the hydration shell, which can be quite time consuming and rather impractical when an ensemble of molecular conformations needs to be examined, is to use methods which estimate hydration *ab initio*. The most developed of these is the Reference Interaction Site Model (RISM) method in which the pair correlation functions for water in the presence of the solute are estimated by using approximations to the hypernetted chain approach. This approach, which has been extensively developed for bulk fluids has also been extended to include interactions with molecules such as proteins in solution.[30] However, applications to SAXS and WAXS profile calculations have not reached the publication stage at the present time.

A related, but more phenomenological method for estimating hydration is the Langevin dipole method, an extension of the Poisson-Boltzmann approach to computing ionic shielding of molecules in solution.[31] In this approach, the sample volume is mapped on to a Cartesian grid and the chemical potential $\mu_{\text{water}}(\vec{x})$ of the solvating water is allowed to vary over the grid. The solvation free energy of the system is then estimated based on the dipolar contribution of the water molecules to the local electrostatic potential. The NLPB (numerical solution of the non-linear Poisson-Boltzmann equations) is then applied to reach a self-consistent solution for the local density of water at each point of the grid, given by $\rho_{\text{water}}(\vec{x}) = \exp(\mu_{\text{water}}(\vec{x})/k_{\text{B}}T)$. This method has the advantage that the effects of ions in the solution are simply included in the calculation of the electrostatic potential. NLPB software under development[32] can execute in the order of minutes of CPU time, so it is realistic to apply to ensembles of conformers. Application of this method to compute SAXS and WAXS profiles is currently under development.

1.18.4 Three-Dimensional Reconstruction of Low Resolution Density Maps from SAXS Data

Since the SAXS profile is a function of only one variable, q, it was believed for many years that only one-dimensional models of the molecular density could be extracted from the data. That is the $P(r)$ function (Section 1.18.1, eqn [6]) which measures the histogram of interatomic distances in the molecule. However, by including prior physical constraints on the electron density of the molecule in a non-linear reconstruction procedure, analysis of the SAXS profile may be extended to three dimensions. The key idea, which may be traced back to reconstruction of density maps from small molecule crystallography[33] is that the density of a biological molecule is relatively constant over the molecular volume and quite a bit larger (of order $0.42\,e^-\,Å^{-3}$ compared to $0.33\,e^-\,Å^{-3}$) than that of bulk water. Application of this constraint to SAXS data was first formulated in terms of the spherical harmonic expansion of the scattering form factors $F_{\text{lm}}(q)$ by Svergun and Stührman.[34]

Then the discovery by Chacon and collaborators[35] that 3-D density maps could be reconstructed from 1-D SAXS data by an algorithm in which a low resolution model is built from 'dummy atoms' or 'beads', showed that this approach was surprisingly robust. (This is analogous to solving the inverse scattering problem of reconstructions of electron microscope data.) Since then, a number of algorithms have been formulated[36–38] for fairly efficient 3-D-reconstruction from SAXS data. They are based on the principle that minimizing a

fitting function of the form

$$\chi^2 = \sum_i \{I_{\text{obs}}(q_i) - I_{\text{calc}}(\{\vec{r}_1 \ldots \vec{r}_m\}, q_i)\}^2 \qquad [25]$$

with respect to the bead positions $\{\vec{r}_1 \ldots \vec{r}_m\}$ results in a 3-D-array of beads whose scattering profile reproduces the experimental SAXS profile $I_{\text{obs}}(q_i)$. The physical constraint of constant density within the bead model is imposed by placing beads on a grid and adding or taking away beads in an iterative manner to converge to a minimum of χ^2. (Note that the scattering from the individual beads is generally calculated by treating the beads as point particles, leading to the Debye's formula for the scattering from the beads.)

How does the algorithm manage to produce a three-dimensional model from one-dimensional data? Note that not only is the anisotropy of the molecule unknown, but also the X-ray scattering phases are unknown. The way the symmetry of the data is broken comes about from the reconstruction algorithm where beads are selectively added or taken away. The first steps of the procedure break rotational symmetry. Subsequent steps then optimize the density in the already symmetry-broken frame defined by the initial steps. Extensive tests by Walther et al.[38] on a variety of proteins of very different shapes showed that the procedure is surprisingly robust and the likelihood surface defined by χ^2 in the space of bead positions is relatively smooth, so that the algorithm is unlikely to get stuck in local minima. The algorithm even managed to model internal cavities in Gro-El![25] This approach was then extended by Svergun[39] by inclusion of simulated annealing steps which help prevent the computation getting stuck in local minima.

Note that the loss of phase information in the SAXS data means that multiple solutions are in principle possible by changing the unknown phase factors. In practice, this problem only seems to arise when applied to very symmetrical geometric models.[40] Usually when doing 3-D-reconstructions, the algorithm is run a number of times using different random number starting seeds. Then the resulting models may be combined using the software DAMAVER[40] to produce an average density map.

1.18.4.1 Estimating the Resolution of 3-D-Reconstructed Density Maps

The statement of the linear sampling theorem of Fourier analysis (Nyquist, Shannon and others) states that a function defined as a Fourier transform with maximum period $q_{\max}/2\pi$ (in our case $P(r)$) is completely determined by giving its ordinates at a series of points spaced $\pi/q_{\max}$ Å apart. So for $q_{\max} = 0.25$ Å^{-1}, this limits the number of values of $P(r)$ to a series of r values spaced by around 12 Å. Clearly, bead models which define a set of 10's or 100's of 3-vector positions do not fulfill this criterion. The 3-D-reconstructions are derived by a highly non-linear procedure which means they are super-resolution models in the sense discussed in connection with Guinier's formula (Section 1.18.1). So how can we assess the resolution in position space of a bead model? The kinds of linear resolution criteria applied in crystallography do not apply.

A solution to this problem proposed by Kretchetov[41] lies in the concept of 'model noise'. If we pick a certain number N_{model} of beads to use in a 3-D-reconstruction, then one finds that each model generated with a different starting seed gives a slightly different pair distribution function $P(r)$ with a corresponding variation in $I(q)$ profiles computed using the Debye formula (eqn [6]). The value of N_{model} in turn defines an average distance L_{avg} between the beads, depending on the effective volume of the molecule. For a given bead model, we can generate any number of random variations by defining regions Γ_i as spheres around each initial bead position $\vec{r}_i$, with radius δl around each model bead, and assigning new bead positions $\vec{r}'_i$ randomly within the sphere of each $\{\Gamma_i\}$. When the resulting ensemble of $I(q)$ profiles is calculated, there will be a mean and variance at each wave vector q which define the width of the 'model noise' associated with the parameters L_{avg} and δl. As the 'smoothness parameter' $\delta l/L_{\text{avg}}$ is increased, the Γ_i regions will overlap and the *average* over the $I(q)$ curves will get smoother and smoother and eventually all of the details will be lost.

What determines the optimum value of the smoothness parameter? As δl is increased, the calculated profiles will get smoother and smoother so the fit to the data will get worse, while as δl is decreased, the model noise will increase. Hence, for a given set of data, there is an optimal value for δl at a given N_{model} which is seen by plotting χ^2 (or an improved version of χ^2) as a function of δl. δl_{opt} is then determined as the value of δl which minimizes χ^2.

If we compare the ensemble of calculated profiles $\{I_{\text{calc}}(q)\}$ for models with given N_{model}, and hence given L_{avg}, using δl_{opt} with the measured profile, it is reasonable to adjust N_{model} so that the model noise matches the experimental (counting statistics) noise in a q-region of interest. On trying various values of N_{model}, δl and L_{avg} are found not to scale linearly. Consideration of a number of different proteins leads to an empirical relation between δl and L_{avg} as[41]

$$\delta l\big|_{\text{opt}} = 3^{1\pm0.2}(L_{\text{avg}} - 2.5^{1\pm0.3}) \qquad [26]$$

which defines the resolution of the model as a function of L_{avg} and hence of the number of beads. This in turn leads to a criterion for the number of beads: N_{model} must be chosen such that the model noise overlaps the experimental noise in the q-range of interest.

As a result of this matching of the model noise to the experimental noise, we can determine the optimum number of beads (and hence L_{avg}) for a given data set. How to turn this into a 'resolution' of the resulting model depends on how the model will be applied. The simplest use of such a bead model is to define a molecular surface. From the above definition, one can assert semi-quantitatively that the precision of determination of points on the surface (in fact of distances between pairs of surface points) is of order L_{avg}.

1.18.4.2 Example: Optimal Number of Beads Justified by Cytochrome c Experimental Data

As an example, we apply these criteria to a bead model of cytochrome *c*. The experimental data measured for cytochrome *c* at SLAC are quite noisy, but available up to high

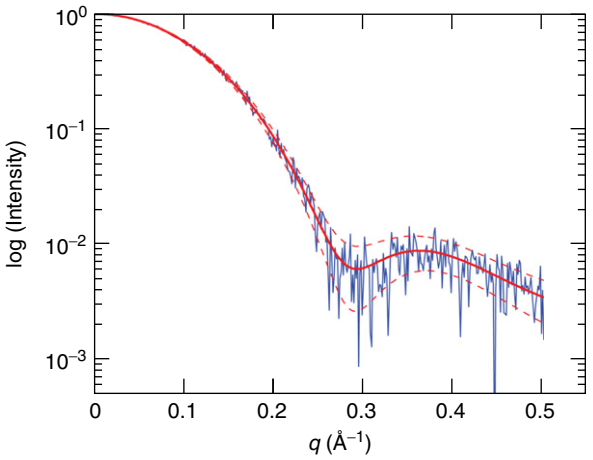

Figure 3 Model noise in reconstructions of cytochrome *c* scattering. The SAXS profiles of reconstructed 3-D-models of cytochrome *c* with $n = 170$ beads and $\delta l = 1.6$ Å (red line) gives model noise (shown as the upper and lower dashed red lines) comparable with the experimental data (blue). Reconstruction with more than 170 beads is not justified by the experimental data. Figure adapted from Kretchetov, A. *Reconstruction of 3-D Electron Density Models from Small Angle X-Ray Scattering Data for Macromolecules: A Space Averaging Approach*, Ph.D. thesis, Stanford University, Dept. of Applied Physics, Stanford, CA, 2005.

$q_{max} = 0.5$ Å^{-1}. A number of reconstructions were computed, with a selection of different numbers of beads. The model noise was computed for each reconstruction using $\delta l|_{opt}$ as determined in eqn [26]. Applying the algorithm we found that a model with 170 beads ($L_{avg} = 3.0$ Å, $\delta l = 1.6$ Å) makes the model noise comparable with the experimental noise. The number of beads needed is approximately twice that which would be expected assuming the protein volume is $1.5\ 10^4$ Å^3 and the average bead spacing $\pi/q_{max} = 6$ Å. **Figure 3** demonstrates the experimental data and profile of the reconstructed model. The dashed line shows the model noise corridor.

The resulting model is a set of spherical beads with radius $\delta l = 1.6$ Å. The average spacing between centers of neighboring beads is $L_{avg} = 3$ Å which defines the resolution of the model in the sense of accuracy of distances between pairs of points on the surface. Note that the Fourier resolution determined by π/q_{max} would be ~ 6 Å for $q_{max} = 0.5$, so that the non-linear fitting procedure leads to around a factor two improvement in distance resolution as well as giving a three-dimensional model hence improving the resolution in solid angle from 4π steradians (for $P(r)$) to something like $4\pi/N_{model} \sim 0.07$ steradians.

1.18.4.3 Use of SAXS Data to Help Determine Protein Structure from Sequence; High-Throughput Analysis

The generation of 3-D bead models from SAXS data leads to the question how to define the intra-protein structural resolution of these 'low resolution' density maps beyond that of the protein shape. This is probably best defined by asking how good is the fit of the bead model representation of the electron density inside the boundary of the protein compared to the density derived from crystallography. As far as known to the authors, this question has not been studied in any detail at the present time.

A more practical question is whether the resolution is good enough to help identify the 'protein fold' (in the sense of tertiary arrangement of the secondary structure helices and sheets) of a protein of known sequence but of unknown structure. This can be asked at a number of levels. If sequence homology of the unknown protein is sufficient to identify with a protein of known structure, then work on high throughput techniques to identify unknown proteins generated in proteomics studies has shown that SAXS data can provide solution structural information at resolutions often sufficient for functional insights into how these proteins work in the context of their pathways and networks.[42]

For proteins for which there is insufficient sequence homology to a protein of known structure, SAXS data may be used as a filter for *ab initio* algorithms which generate candidate structures based on the sequence of a given protein. This has been studied for a large set of proteins of varying fold types.[43] The probability of filtering out the good candidates was found to depend strongly on the type of fold involved.[43,44]

1.18.5 Disordered Peptides and Proteins in Solution: Free Energy Landscape for Denaturation

As opposed to SAXS from fully folded proteins or polynucleotides, interpretation of SAXS data from denatured proteins, natively disordered proteins, or solutions of peptides, involves calculation of a SAXS profile for an ensemble of conformers. A key point here is that such ensembles are built of a Boltzmann weighted aggregate of conformers. The SAXS data measure the thermodynamic average of such an ensemble, but contain no information about the probability distribution of the ensemble of conformers. Furthermore, the fact that the molecules have multiple conformers indicates that they are in a complex free energy landscape which may have multiple minima, so that treating the data in terms of a single predominant conformation may be an oversimplification. Finally, the denatured state of the protein, or multiple conformations in the case of a peptide, indicates that the molecule has lost some of its hydrophobicity and tends to expose more accessible surface area than a fully folded protein.

A basic question for denatured proteins is the character of the free energy landscape of the protein–denaturant mixture. By measuring the effects of varying solvent conditions such as concentration of denaturant, some information may be gained by determination of R_g as a function of denaturant concentration. This can be fit against a thermodynamic model of denaturation using a Singular Value Decomposition (SVD) analysis to determine the possible occupancies of a number of denatured local minima in the free energy surface (see Doniach review[2]). As an example, a study of the thermodynamics of cytochrome *c* denaturation as a function of guanidine hydrochloride (GdnHCl) revealed the existence of more than one denatured state at intermediate GdnHCl concentrations.[45]

Similar methods may be used to fit kinetic models of folding to time-resolved SAXS measurements of protein or RNA folding. Comparison of time-resolved SAXS data with optical fluorescence measurements using Förster Resonance Energy Transfer (FRET) of protein folding have raised questions about the initial kinetics of protein folding: is there a very fast polymer-collapse taking a denatured protein to a more compact local minimum unfolded ensemble before a '2-state' folding event takes place? (see work by Eaton and collaborators[46]).

Because the denatured or natively unfolded state tends to expose more accessible surface area than when fully folded, the scattering contributions of the hydration shells of the conformers in the ensemble will vary depending on the degree of surface exposure of the individual conformers.

1.18.5.1 Importance of Solvation Free Energy and Hydration for Disordered Peptides and Proteins

Calculation of a SAXS profile for an ensemble of conformers requires a method for generating ensembles of possible conformers for the disordered molecule. Published results have depended either on molecular dynamics simulations or on *ad hoc* prescriptions for combining fragments of peptide sequence taken from the pdb using a variety of stochastic dihedral angles.[47–49] SAXS measurements of an equilibrium ensemble of the artificial peptide XAO have revealed fundamental problems in computer simulation of the conformer ensemble.[50] As shown in **Figure 4**, the Kratky plot of the SAXS data for the peptide solution rises linearly with q, a behavior characteristic of unfolded proteins and peptides.[2] However, the Kratky profile of a typical simulated PPII helix used for comparison curves over as q increases (**Figure 4**). In addition, the R_g computed from conformations of the XAO peptide sampled in extensive molecular dynamics simulations was consistently larger than the experimental value.[50]

In subsequent work, Scheraga and coworkers demonstrated that an ensemble of XAO conformations generated by an MD method that takes into account experimental constraints from NMR measurements gave an R_g in close agreement with the experimental results.[51,52] However, reproducing the linear increase at high q in the Kratky plot remains a challenge (**Figure 5**). While the interpretation for the observed discrepancies is still under study, the treatment of the hydration shell appears to play an important role. **Figure 5** shows a set of profiles including the effects of hydration on the SAXS profiles. Profiles were computed for the same ensemble of structures using different methods to take into account the hydration shell. Treating hydration in the Langevin-dipole approximation[31,32,53,54] appears to give an overall reasonable agreement with the experimental data (**Figure 4**, main graph). In contrast, using the constant hydration shell density implemented in the software CRYSOL[8] appears to give less favorable results (**Figure 4**, inset). We note that these preliminary results do not yet address the issue of how the hydration free energy changes the relative Boltzmann weights of different conformations in the ensemble. However, our results indicate that the treatment of hydration plays an important role even in the calculation of

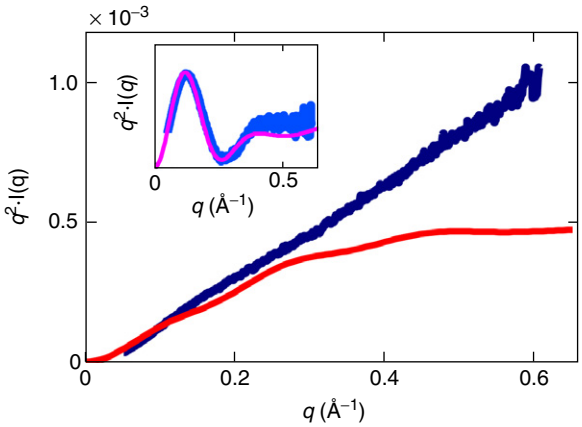

Figure 4 SAXS data for the XAO peptide. Scattering profile of the XAO peptide at a concentration of 5 mg ml^{-1} in 100 mM acetate (blue line) and the theoretical prediction for the ideal PPII helix (red solid line). The data are plotted in Kratky representation of q^2 $I(q)$ as a function of q. For comparison, the inset shows data for horse heart cytochrome c (1.25 mg ml^{-1} in 100 mM acetate) measured on the same set-up and used as a molecular weight standard (light blue line) and a theoretical profile using the known crystal structure with Protein Data Bank accession code 1CRC and CRYSOL (magneta line). Adapted from Zagrovic, B.; Lipfert, J.; Sorin, E.; Millettt, I.; van Gunsteren, W.; Doniach, S.; Pande, V. Unusual compactness of a polyproline type II structure. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11698–11703. Copyright by PNAS.

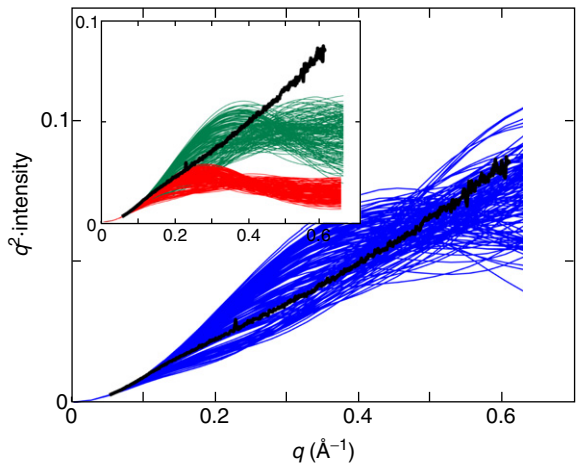

Figure 5 The importance of hydration for the calculation of scattering profiles from unfolded ensembles of peptides. Experimental scattering data for the XAO peptide are shown in Kratky representation of q^2 $I(q)$ as a function of q (thick black line). The experimental profile is compared to SAXS profiles computed for an ensemble of conformers of the XAO peptide that was generated using an MD method with experimental NMR constraints.[52] Scattering profiles were computed for the same ensemble of structures using the aquasaxs method that treats hydration in a Langevin-dipole formalism[53] (main graph, blue lines) and using the program CRYSOL[8] with default setting for the hydration (inset, red lines) and without the hydration shell (inset, green lines).

scattering profiles for an unweighted ensemble of unfolded structures.

These preliminary results indicate that it is essential to include estimates of solvation shell contributions to the SAXS expected for the Boltzmann ensemble of conformers. SAXS measurements taken on their own contain no information about the distribution of conformers in an ensemble. Consequently, methods which do not take into account solvation are very likely to give false results for SAXS averages over an ensemble which is not Boltzmann weighted, so may lead to misleading interpretations of SAXS data from disordered peptides, proteins or RNA.[48] Clearly this consideration will also apply to proteins or RNA consisting of folded domains linked by flexible linkers.[47]

1.18.6 Calculation of SAXS Profiles for Detergent Micelles and Protein–Detergent Complexes

Integral membrane proteins comprise ≈30% of most proteomes[55] and facilitate transport and signaling across cell membranes. Despite their importance, less than 1% of known protein structures are of membrane proteins.[56] One major obstacle to membrane protein structure determination is the selection of a detergent that mimics the native lipid bilayer and stabilizes the protein fold. Detergents are small amphipathic molecules that are used to solubilize membrane proteins for structural and functional investigations. However, unlike phospholipid bilayers, detergents form micelles, which are spheroidal and have a core composed of the detergent hydrophobic tails. Micelles have different shapes and sizes depending on the detergent chemical structure. For structural investigations, a multitude of detergents is screened until a condition that provides high-quality crystals or NMR samples. An understanding of detergent properties is consequently desirable for determination of optimal detergent conditions for extraction, purification, and structural and functional characterization of membrane proteins.

SAXS is found to be a powerful probe of the size and shape of detergent micelles. Both SANS and SAXS have been used to study the solution structure and interactions of detergent micelles.[57] It is also found to be a useful tool for analysis of the structure of protein–detergent complexes (PDCs).[4,58–60]

SAXS profiles for typical micelles are found to have a 'double hump' feature (i.e., a pronounced second maximum (counting the maximum at $I(q=0)$ as the first maximum)). A characteristic profile is shown in **Figure 6**, where we take as an example the detergent FC-10 (n-decylphosphocholine). This pronounced second maximum is a consequence of the micelle internal structure with the detergent headgroups forming the outer surface of the micelle and the aliphatic tails forming the inside of the micelle (**Figure 7**). As the tails are predominantly hydrocarbon chains of form $(CH_2)_n$, they have an electron density lower than that of the surrounding water. This leads to a negative scattering amplitude when the background is subtracted which results in interference in the X-ray scattering of the headgroup contribution relative to the tail contribution, resulting in the additional maximum in the 'double hump' form of the SAXS profile.[61]

From SAXS studies of a number of different detergents, it has been observed that the position of the second maximum, q_{max}, of the SAXS intensity observed for detergents with alkyl tail groups is a direct measurement of the head group–head group spacing, L, across the micelle:[61,62] $L = 2\pi/q_{max}$. Determination of the characteristic head group–head group distance from the position of the second maximum in the scattering intensity is straightforward and robust, and the results are in good agreement with the parameters derived from the two-component ellipsoid models.[61,62] This characteristic distance across the micelle likely constrains detergent–protein interactions and appears to be a determinant of homogeneity of the conformation of proteins solubilized by detergent.[62]

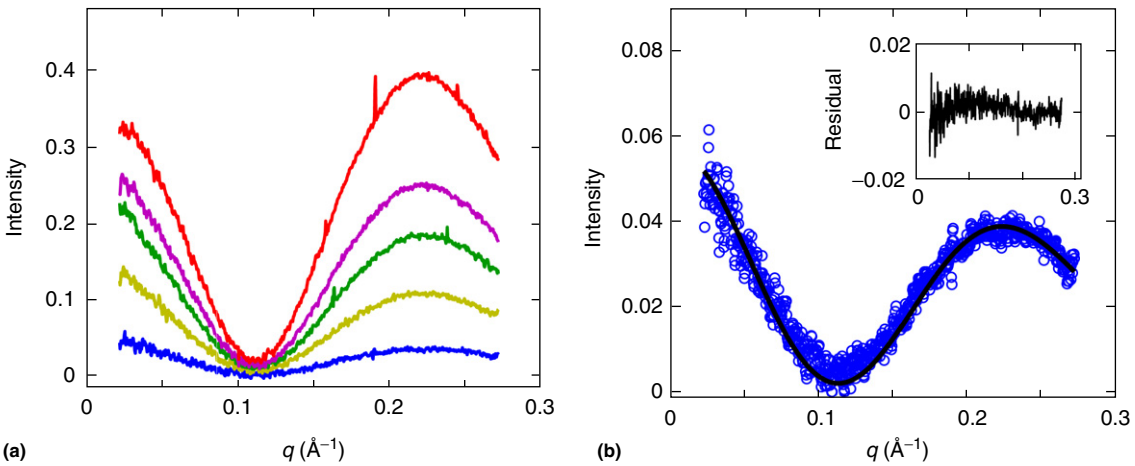

Figure 6 SAXS analysis of *n*-decylphosphocholine (FC-10) detergent micelles. (a) SAXS profiles $I(q)$ of FC-10 at detergent concentrations of 25 (blue), 50 (light brown), 75 (green), 100 (purple), and 150 (red) mM. (b) Two-component ellipsoid fit (black solid line) and scattering intensity recorded at 25 mM detergent concentration (blue circles). The residuals of the fit are shown in the inset. The fit for FC-10 has parameters $\rho_1 = 0.273$ (e Å^{-3}), $\rho_2 = 0.490$ (e Å^{-3}), $a = 20.7 - 21.2$ Å, $b = 13.4 - 13.6$ Å, $t_a \approx t_b = 2.7 - 3.0$ Å, $R_g = 25$ Å. Adapted from Figures 2(a) and 2(d) in Lipfert, J.; Columbus, L.; Chu, V. B.; Lesley, S. A.; Doniach, S. Size and shape of detergent micelles determined by small-angle x-ray scattering. *J. Phys. Chem. B* **2007**, *111*, 12427–12438. Copyright by American Chemical Society.

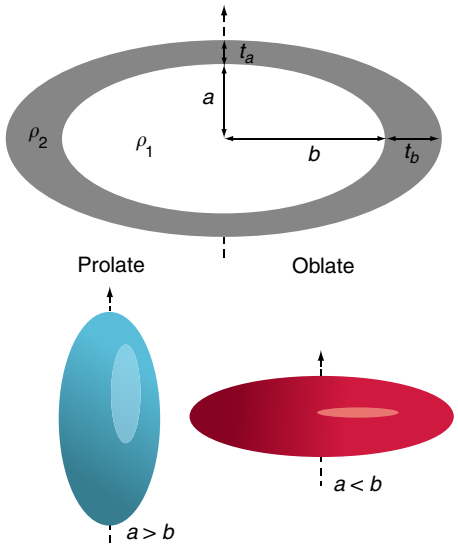

Figure 7 Schematic diagram of the two-component ellipsoid model. The rotation symmetry axis is shown as a dashed line. The model consists of a hydrophobic core with dimensions a and b and electron density ρ_1. The hydrophobic core is surrounded by a head group region with thickness t_a and t_b and the electron density ρ_2. For $a > b$, the ellipsoid is prolate (lower left), for the case $a < b$, the ellipsoid is oblate (lower right).

1.18.6.1 Micelle Mixtures and Protein–Detergent Complexes

In a recent study on the model polytopic α-helical membrane protein TM0026 using a combination of NMR, EPR and SAXS, a correlation was found between protein conformations, micelle size and thickness, and homogeneity of nuclear magnetic resonance (NMR) and electron paramagnetic resonance (EPR) spectra.[62] TM0026 is a membrane protein of unknown function from the thermophile *Thermotoga maritima* and was initially characterized as part of the high throughput structure determination pipeline of the Joint Center for Structural Genomics.[58] This led to the exploration of whether engineering mixed micelles by mixing detergents at different ratios might be a way to systematically change the size and shape of detergent micelles.

The dependence of L, the distance between the head group centers across the short diameter of the spheroid, determined from the position of the second peak in the scattering intensity, on the mixed micelle composition for two detergents A and B was fit by the relationship

$$L(\chi_A) = L_B + (L_A - L_B)\chi_A \qquad [27]$$

with the mixing ratio $\chi_A = ([A])/([A] + [B])$. $[A]$ and $[B]$ are the detergent concentrations that have been corrected for the monomeric detergent concentration using the relationship $X_{i,\text{monomer}} \approx (X_i \cdot cmc_i)$, where X_i and cmc_i are the mol fraction and critical micelle concentration of detergent species. The linear dependence of L on the mixing ratio appears to hold for a wide range of detergent mixtures. The NMR and EPR spectra found for TM0026 protein–detergent complexes suggest that the protein conformations are strongly influenced by the size and thickness of the detergent micelle. The characteristic headgroup–headgroup dimension L measured for micelles in the absence of proteins appears to reflect an intrinsic packing preference of the detergent and we observe a similar thickness in TM0026 complexes that yield well-resolved NMR spectra.

Taken together, the results indicate that well-resolved NMR spectra are observed when the hydrophobic component of L (which is equal to L minus the thickness of the headgroup) $\approx$29–31 Å matches the length (30–33 Å) of the hydrophobic part of the α-helices of TM0026. We note that this hydrophobic dimension matches the average hydrophobic thickness of the *T. maritima* bilayer ($\approx$30 Å). All in all, these data indicate that, rather than exhaustively screening a multitude of detergents, it should be possible to rationally engineer appropriate mixed micelles for NMR structure determination, or possibly for crystal formation, following simple principles and from a limited set of detergents.

1.18.6.2 Atomic Scale Modeling of SAXS from Detergent Micelles and Protein–Detergent Complexes (PDCs)

In order to interpret SAXS profiles for solubilized membrane protein PDCs, we need an approach which can work with atomic scale models, generated by MD or other means, of the combined protein and accompanying 'belt' of detergent molecules represented in a pdb file. Such calculations need to take into account the fact that the detergent hydrocarbon chains have a lower density than water, so contribute a negative scattering amplitude to the scattering profile due to the negative contrast with the solvent. One approach to taking into account this negative scattering amplitude is to perform the background subtraction explicitly on an atomic scale model of the micelle and protein and integrate over angles numerically by the method of cubature.

We illustrate this approach with a rough molecular model of a micelle built from individual molecules of SDS (sodium dodecyl sulfate) detergent. The latter was built by taking a pdb representation of a detergent monomer, then assembling the heads on a cylinder with rounded end caps, and tails pointing in to the axis of symmetry. However, further energy minimization has not been carried out at the present time. We construct a 'ghost' scatterer for the solvent by creating an array of dummy atoms with lattice spacing δx which fill the same spatial shape as does the micelle. This is done by constructing a periodic integer array and writing the coordinates of all points within a distance ξ of any atom in the micelle model to an output pdb file which fills the space occupied by the target molecule (see **Figure 8(a)**, for which δx was set at 2.7 Å, and ξ was set at 3.4 Å). The scattering form factors for both the molecule and the ghost array of dummy atoms are computed at an array of X-ray wavevectors $\vec{q}$ where the $\vec{q}$ directions are selected from a non-uniform spiral on the unit sphere taken from the 'cubature' algorithm of Fliege and Maier[28] (this algorithm is a unit sphere analog of the Gauss quadrature integrator in 1-D which picks sampling points non-uniformly in a way which guarantees exact results for integrating a polynomial of given order).

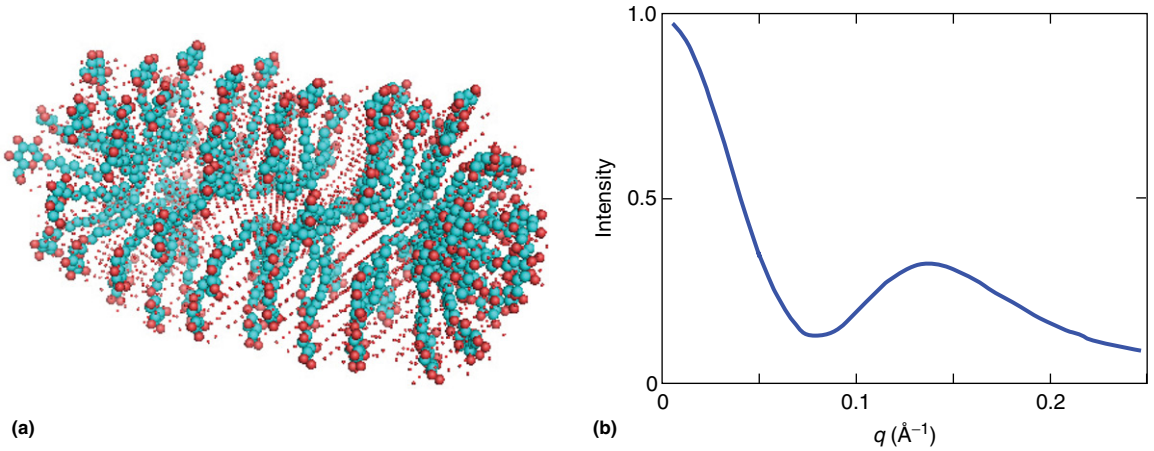

Figure 8 Molecular model of a SDS micelle. (a) Micelle model generated by placing pdb monomers and ghost dummy atoms representing water scattering (small red spheres). (b) SAXS profile for the model of a SDS micelle with contrast determined by subtracting a form factor for its ghost adjusted for water volume (see text).

For given $\vec{q}$, the form factors are given by

$$F(\vec{q}) = \sum_i f_i \exp(i\vec{q} \cdot \vec{r}_i) \qquad [28]$$

where the $\vec{r}_i$ are the positions of the atoms in the micelle or of the dummy atoms, and the f_i are the electron numbers for the detergent atoms or an equivalent electron number for the water, represented by the dummy atoms. The latter are calculated as $f_{\text{water}} = 0.334\ (\delta x)^3$. Finally, the scattering profile is computed as

$$I(q) = \sum_{q_j} w_j \left| F_{\text{micelle}}(\vec{q}_j) - F_{\text{ghost}}(\vec{q}_j) \right|^2 \qquad [29]$$

where the directions $\vec{q}_j$ and weights w_j are taken from Fliege's tables (available at: http://www.personal.soton.ac.uk/jf1w07/nodes/nodes.html).

When evaluating the resulting profile for a model micelle of SDS, an additional factor was needed to adjust the water density to take account of the non-space-filling character of the model micelle. Since the micelle model leaves spaces in between the monomers which were not optimized by energy minimization, we need an additional adjustment factor in front of F_{ghost} which leads to a variable height of the second diffraction peak, and was adjusted by hand to give a profile similar to that found experimentally. In the case shown in **Figure 8(b)**, the factor was found to lie in the range 0.53–0.57. This result indicates that fitting the measured scattering profile will place quite strong bounds on the number of monomers and the level of packing optimization needed in modeling a given micelle.

1.18.7 Combination of SAXS and NMR for Structure Determination of Multi-Domain Proteins

NMR has proven to be a highly effective method for determining protein structure at atomic resolution (provided the protein is not too large to give effective motional narrowing by rotational diffusion). However, for multidomain proteins, even though the conformations and relative orientations of the individual domains can be determined accurately by using backbone-backbone nuclear Overhauser effects (NOEs) and extensive sets of residual dipolar couplings (RDCs), the need for data determining relative positioning of the individual domains becomes critical.

Based on this need, Bax and collaborators have established systematic protocols to combine SAXS data into macromolecular structure refinement so as to compensate for the deficiency of the translational information derived from interdomain NOEs.[63] To do this, a model of the protein at atomic resolution obtained from NMR data was refined against the SAXS data by minimizing a cost function of the form

$$\chi^2 = \frac{1}{N_q - 1} \sum_k^{N_q} \left\{ \frac{I_{\text{expt}}(q_k) - c_k I_{\text{calc}}(q_k)}{\sigma(q_k)} \right\}^2 \qquad [30]$$

where the SAXS profile $I_{\text{calc}}(q_k)$ evaluated at measurement angles q_k is calculated using the Debye formula. Differentiation of χ^2 with respect to the current atomic coordinates yields a set of atomic forces that add SAXS constraints to those from an empirical force field used in the MD-based structure refinement. However, the refinement against SAXS data in combination with other data sources (in this case, a set of NMR-generated restraints) leads to a considerable computational overhead.

A solution to this problem implemented by the Bax group[63] is to lump together groups of proximal atoms into 'globs' and replace their scattering amplitude by spherically averaged scattering form factors which they term a 'globic' approximation. Since the refinement scales as $N_q N^2$, where N is the number of degrees of freedom (atomic coordinates) used in the NMR refinement, a reduction of N input heavy atoms into approximately $N/3$ globs, leads to a reduction of the required CPU time by about an order of magnitude. As a result of the globbic approximation, the calculated scattering

intensity curves show small but systematic differences with respect to the exact ones, obtained from all-atom calculations. These differences may be rectified to some extent by multiplying the calculated scattering curves by coefficients $c(q_k)$ determined from the atomic coordinates of each glob to compensate for these differences. Another speed up was obtained by limiting the number of sampling points q_k, resulting in an overall speed up of the refinement by a factor of 80.

The method was applied for the two-domain 177-amino-acid protein, γS crystallin, using an experimental SAXS data set fitted over a q-range from 0.03 to 0.2 $Å^{-1}$. Inclusion of these data during structure refinement decreased the backbone coordinate root-mean-square difference between the derived model and the high-resolution crystal structure of a 54% homologous γB crystallin from 1.96 ± 0.07 to 1.31 ± 0.04 Å.

In a subsequent study, Bax and collaborators[64] applied this method to a structural study of the 82 kDa enzyme Malate Synthase G (MSG), currently the largest single chain protein solved by solution NMR[64]. The protein is composed of four domains that include a central TIM barrel β_8/α_8 core, an N-terminal α-helical clasp, an α/β domain, and a C-terminal 5-helix bundle that forms a plug for one of the ends of the β-barrel of the core domain and is connected to it by a flexible linker. A principal conclusion from this work is that, even in the absence of changes in such simple parameters as R_g or R_{max}, SAXS data can improve the structural accuracy by influencing the relative positioning of the individual domains within the overall model. The bulk of this improvement does not result from either the lowest angle part of the scattering profile, which is sensitive to the overall particle dimension, or from the higher-angle data, which are influenced by variation of the protein's internal density. Instead, the intermediate angle data, which reflect the low-resolution particle shape, are the driving force behind the rearrangement of the MSG domains. The joint analysis of NMR and scattering data indicates that the magnitude of the differences between the MSG domain orientations in solution and in the crystal is only slightly higher than the experimental uncertainty of the relative domain orientations.

1.18.8 Solution Structure of Small Functional RNAs

Advances in the biochemistry of RNA[65] and in genome sequencing have led to the recognition that RNA functions not just as a messenger as in mRNA or as a nucleotide transporter as in tRNA but also in many other roles including enzymatic activity (ribozymes) and control of gene expression (as in riboswitches). The biochemistry and biophysics of functional RNAs are complicated by difficulties of crystallization of polynucleotides, in part because of the high negative charge of the phosphate backbone and in part because many states of RNA occupy ensembles of conformers in the absence of stabilizing substrates.

SAXS is starting to play a useful role in determination of the 3-D structure of small functional RNAs.[66–69] Building of low resolution density maps by 3-D-reconstruction allows for generating bead models of the variety of thermodynamic states occupied by a given RNA molecule as a function of the counterion concentration of ions of different valencies (monovalent as for Na^+, divalent as for Mg^{2+}, and higher valency molecular ions such as spermidine and spermine), and of substrate binding.[70–72] When coupled with other structural techniques such as hydroxy radical determination of unprotected backbone regions and other biochemical methods, FRET applied to flexible junctions and some NMR studies,[66,73–75] it has been possible to generate structures of close to atomic resolution for a variety of thermodynamic states.

As an example of how complementary information can be used in the construction of RNA structural models, we show a figure (**Figure 9**) of a 3-D-model of the VS ribozyme built by combining FRET studies with bead models generated by SAXS.[73] The Varkud satellite (VS) ribozyme[76,77] is the largest

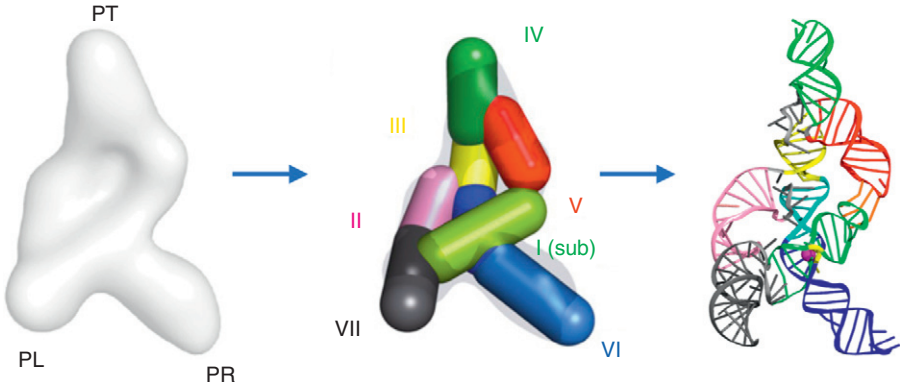

Figure 9 Modeling of the complete VS ribozyme from SAXS data. Left: low-resolution 3-D-model obtained from *ab initio* reconstructions using SAXS data using the software DAMMIN.[36] Middle: intermediate model where the double-stranded regions of the ribozyme are replaced by cylinders that have the length and diameter of the corresponding RNA helices. The cylinders are placed into the *ab initio* model by taking into account SAXS data of several RNA constructs and prior information from FRET. Right: atomic model of the VS ribozyme. The model was obtained by replacing the cylinders with standard A-form helices and homology modeling of the junctions followed by rigid body refinement. Figure adapted from Lipfert, J.; Ouellet, J.; Norman, D.; Doniach, S.; Lilley, D. The complete VS ribozyme in solution studied by small-angle X-ray scattering. *Structure* **2008**, *16*, 1357–1367.

of the known nucleolytic ribozymes, and the only one for which there is no crystal structure. Analysis of the secondary structure showed that the full ribozyme comprises seven helical sections. The 5′ stem-loop contains the cleavage site and can be considered the substrate. The representation of the density map shown in **Figure 9** is derived from the bead model using the software Situs.[78]

The model was built in a stepwise procedure. First, SAXS data of both the full VS ribozyme and of several sub-components were used to build *ab initio* low-resolution density maps of all constructs. In a second step, double-stranded regions were represented with cylinders of the appropriate dimensions. By taking into account the low-resolution SAXS models of several constructs as well as prior information from biochemical and FRET studies,[79,80] the cylinders could be placed into the low-resolution densities. In the next step, the cylinders were replaced by standard A-form RNA helices. In addition, models of the junction regions were built from homology. Finally, energy-minimization refinement against the standard stereochemical restraints was used to regularize and refine the structure. The path of the backbone is indicated by the ribbon, and the positions of base pairs are indicated by the bars (**Figure 9**). The color coding of helices matches that of the secondary structure.[73] The scissile phosphate is shown by the purple ball, and the probable active site components A756 and G638 are highlighted in yellow (**Figure 9**).

The modeling enables some ideas on the relation between the structure and function of the ribozyme to be verified by physical measurement. The SAXS data provide physical evidence for the location of the substrate within the ribozyme, showing that close interaction between the substrate loop and the A730 loop is very achievable in the context of this structure. This conforms very well to current mechanistic ideas, in which a significant element of the catalysis of phosphoryl transfer arises by general acid-base catalysis by the nucleobases of A756 and G638. In addition, the new structure provides some clues to the function of the additional helix VII in the formation of the active complex.

1.18.8.1 SAXS as a Filter for Candidate RNA Structures on the Pathway to a Fully Folded State

Low resolution bead models derived from SAXS data are useful for characterizing changes in states of a functional RNA as solvent conditions are altered. Nevertheless, they do not provide any direct link between the three-dimensional structure of the RNA and its primary sequence; the placing of structural elements of the RNA within the envelope cannot be derived from the bead models alone.

Development of a number of algorithms[81–84] for generating candidate 3-D-structures of RNA from sequence via the secondary structures derived by M-fold and related algorithms, makes available the possibility of generating a large number of candidate tertiary folds for different thermodynamic states of small RNAs based on sequence alone. As discussed above, plausible physical tertiary folded structures can be selected by putting together a number of biochemical and biophysical characterizations. Here, using the thiamine pyrophosphate (TPP) riboswitch as an example, we demonstrate that the SAXS data can also be used to filter out the most physically realistic of the tertiary structures generated by these algorithms.

The TPP riboswitch regulates genes that code for proteins involved in the thiamine biosynthetic pathway.[85,86] The overall changes in structure on going from no Mg^{2+} or ligand, through the Mg^{2+} induced partially folded state, to the fully folded state with Mg^{2+} and ligand are illustrated in **Figure 10**.[87] Crystal structures have been obtained for the TPP riboswitch *thiM* from *Escherichia coli* and for the TPP riboswitch *thiC* from *Arabidopsis thaliana* bound to TPP.[88,89] The folded, ligand stabilized structures of the riboswitch for the different organisms show a high degree of similarity, folded into the same overall tuning-fork conformation, with the

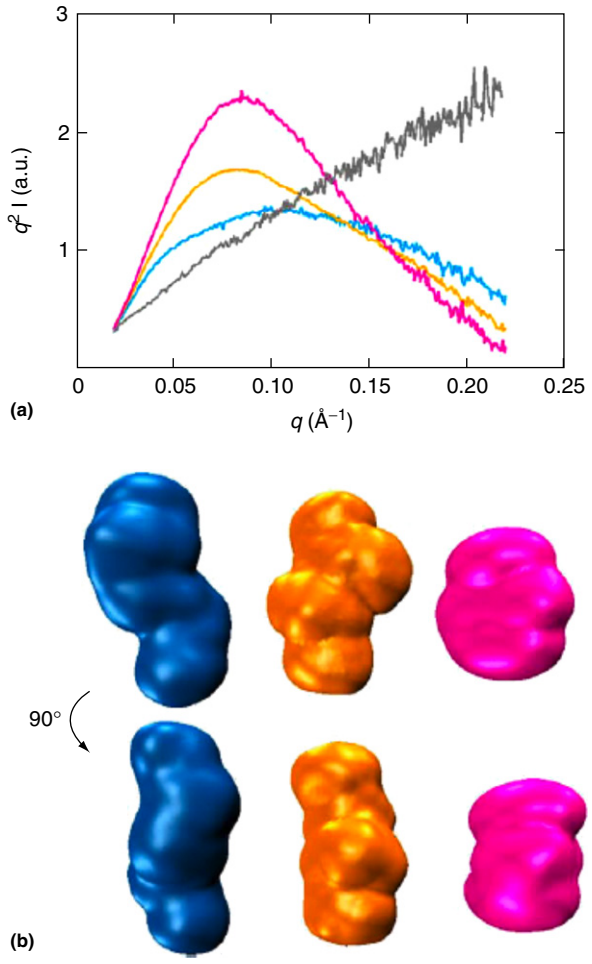

Figure 10 Conformation of the TPP riboswitch aptamer as a function of solution conditions determined by SAXS. (a) Scattering profiles in Kratky representation ($q^2I(q)$ versus q) of the TPP aptamer in the absence of Mg^{2+} and TPP (blue), defined as the unfolded state; in 10 mM Mg^{2+} but no TPP (orange), defined as the intermediate state; and in 10 mM Mg^{2+} and 10 mM TPP (pink), defined as the folded state. Data are also shown for the denatured riboswitch in 7 M urea (gray). (b) Low-resolution bead models for the TPP riboswitch in its unfolded, intermediate, and folded states, obtained from the SAXS profiles (same color code as in (a)). Figure adapted from Ali, M.; Lipfert, J.; Seifert, S.; Herschlag, D.; Doniach, S. The ligand-free state of the TPP riboswitch: A partially folded RNA structure. *J. Mol. Biol.* **2010**, *396*, 153–165.

ligand bound between the two prongs in an extended conformation. However, the intermediate states on the folding pathway of the riboswitch cannot be characterized by crystallization.

The secondary structure of the riboswitch, illustrated in **Figure 11(a)**, consists of five basepaired helical stems (P1P5) connected by non-basepaired junction regions. The architecture of the riboswitch is built around a three-way junction, with helix P1 branching into two stacked helices in each branch (P2 and P3 on one side, and P4 and P5 on the other side) and with the two branches arranged in a parallel fashion. The P2/P3 stack forms the pyrimidine sensor helix, where the thiamine moiety of TPP is inserted into a pocket formed by the J2/3 junction, and the P4/P5 stack is identified as the pyrophosphate sensor helix, where junction J4/5 forms a hydrogen-bonding network with the pyrophosphate end of the ligand.

As in the case of the VS ribozyme, SAXS data, combined with electrophoretic gel mobility and chemical probing approaches were used to characterize the global conformation of the TPP riboswitch. Models were generated using MC-Sym.[84] Drawing on the hierarchical folding characteristic of RNA, such that regions of local secondary structure are typically stably formed under conditions not yet conducive to tertiary structure formation, the structure of the TPP riboswitch was modeled as rigid basepaired helices connected by unpaired junction regions. These junction regions govern the overall arrangement of the helices in the global fold observed for the RNA. A total of 40 000 candidate structures were generated for the riboswitch. SAXS data provided a straightforward way to filter out candidate structures on the folding pathway shown in **Figure 10** for the case of the TPP riboswitch.

SAXS profiles were computed for all of the 40 000 models of the whole riboswitch using CRYSOL, and these profiles were filtered against experimental SAXS data for the RNA. The results are shown in **Figure 11** for the candidates which best fit the SAXS data. The 38 highest-ranking models from the MC-sym structures gave p-values of >0.95 when tested against experimental data for the ligand-free state of the riboswitch. There were considerable variations between these 38 models, with an average RMSD of ~ 12 Å between the different models. The differences arise from the twist of helices with respect to one another and from small changes in interhelical angles, which cannot be differentiated using SAXS. However, the models all have the same overall conformation, with helical branches P2/P3 and P4/P5 separated by an average angle of $120 \pm 15^{\circ}$. One representative model is shown in **Figure 11(b)**: the two arms of the tuning fork are now lined up in opposite directions as opposed to being parallel in the ligand bound state.

For the ligand-bound state of the riboswitch, the 27 top-ranked models produced good fits to the experimental data, as determined by a chi-squared goodness-of-fit test. There was more variation between these models than between the set of 38 models isolated for the ligand-free state, as indicated by a higher average RMSD of ~ 17 Å between the models (see **Figure 11(c)** for an example). SAXS profiles generated from the crystal structure and from the model give similar fits to the experimental SAXS data. The overall conformation of the model is similar to the crystal structure,[89] in which helices P2/P3 and P4/P5 come together in an almost parallel fashion, however, higher-resolution details, such as the orientation of the individual residues, helical twists, and structures of the non-basepaired regions, differ between the two structures. The TPP ligand has an important stabilizing effect on the folded state of this riboswitch and, since MC-Sym does not account

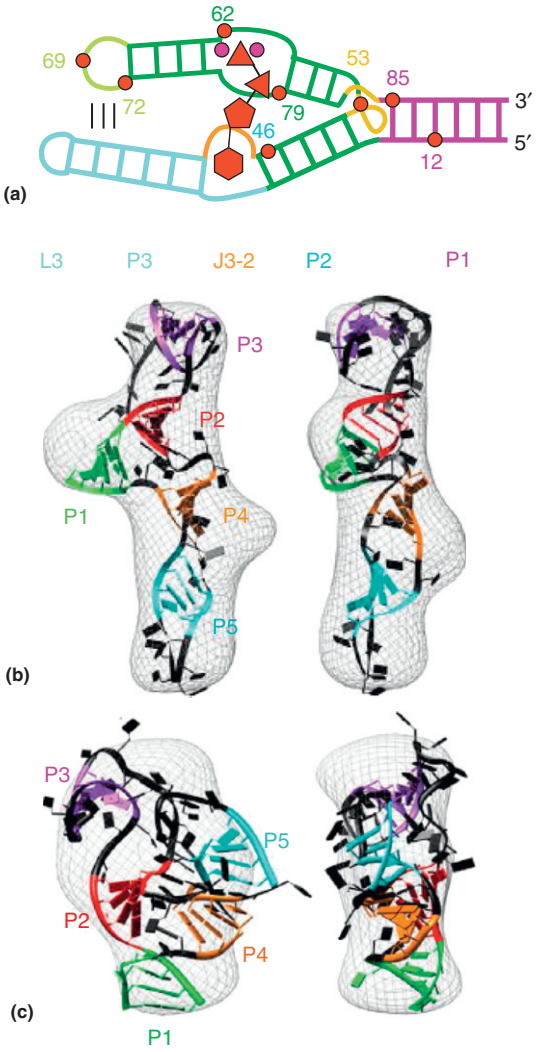

Figure 11 Structure modeling of the TPP riboswitch aptamer in the ligand-free and ligand-bound conformation.[87] (a) The secondary structure of the ligand bound state.[90,91] (b) Atomic scale tertiary structure model of the TPP riboswitch aptamer in the ligand-free state, superimposed on the SAXS-derived electron density (shown as a gray mesh). The color code matches that used in the secondary structure panel. (c) Atomic scale tertiary structure model and SAXS-derived electron density of the TPP riboswitch aptamer in the ligand-bound state. Same color code as in panel (b). Models were generated using MC-Sym.[84] (a) Adapted from Lang, K.; Rieder, R.; Micura, R. Ligand-induced folding of the thiM TPP riboswitch investigated by a structure-based fluorescence spectroscopic approach. *Nucleic Acids Res.* **2007**, *35*, 5370–5378. Copyright by Oxford University Press. (b)–(c) Adapted from Ali, M.; Lipfert, J.; Seifert, S.; Herschlag, D.; Doniach, S. The ligand-free state of the TPP riboswitch: A partially folded RNA structure. *J. Mol. Biol.* **2010**, *396*, 153–165. Copyright by Oxford University Press.

for the effects of ligand binding on RNA structure, MC-Sym cannot be expected to capture the bound state of the RNA. This limitation of MC-Sym will lead to undersampling of the bound conformation in the presence of TPP. The use of the SAXS data in selecting models of the bound state therefore helps in identifying the correct global conformation from all of the other models.

A recent paper by Yang et al.[92] develops the above approach by incorporating the SAXS data as a filter in the generation of a large ensemble of 3-D-models using MC-Sym. Their method, termed Fast-SAXS-RNA, is built on a reduced two-particle representation for each nucleotide. Their results indicate that a two-particle representation is sufficient to achieve reasonable accuracy in the calculation of SAXS patterns for the case of tRNA structures. They report that, because of the significant reduction in computational cost for testing a model, the method enables them to rank, order, and filter a large ensemble of 3-D-models extremely efficiently. Usage of SAXS data as a constraint succeeds in filtering out most non-native folds and identifying well-fit models at a low-resolution level.

1.18.8.2 Using Low Resolution Bead Models to Estimate Free Energies as a Function of Salt Concentration

Breaker and coworkers reported a tandem aptamer riboswitch (VCI-II) that binds glycine cooperatively.[93] In subsequent work, a combination of hydroxyl radical footprinting and SAXS was used to study the conformations of this tandem aptamer as a function of Mg^{2+} and glycine concentration.[72] SAXS data identified three thermodynamic states: In low salt with no magnesium present, the VCI-II construct has an extended overall conformation (termed U), presumably representing unfolded structures.

Addition of millimolar concentrations of Mg^{2+} in the absence of glycine leads to a significant compaction and partial folding (termed the M state), as judged by the observed changes in R_g and in hydroxyl radical protections. In the presence of millimolar Mg^{2+} concentrations, the tandem aptamer binds glycine cooperatively. The glycine bound state is termed B. 3-D-reconstructions of the structural conformations of each of the states has allowed for visualization of the shapes adopted by the three states.[72] However, the sequence of the VCI-II glycine riboswitch is sufficiently complex that an atomic scale model has not yet been constructed at the time of writing. The glycine binding transition involves a further compaction, additional tertiary packing interactions and further uptake of magnesium ions relative to the M state in high Mg^{2+} and in the absence of glycine.

To probe the role of ion interactions in folding and ligand-binding of the VCI-II tandem aptamer, the global solution structure of the VCI-II RNA was assayed by SAXS under a variety of buffer conditions. Scattering data were obtained in the absence and presence of glycine for a range of monovalent and divalent ions. The measurements employed either 20 mM divalent, which is >100-fold above the Mg^{2+} midpoint concentration obtained previously for transition from the U to M state (~0.1 mM Mg^{2+}),[59] or 2 M monovalent salt. Similarly, measurements in the presence of glycine used a concentration of 20 mM glycine, >100-fold above the previously observed glycine midpoint of ~90 μM in 10 mM Mg^{2+}.[72]

In the absence of glycine, the scattering profiles for 2 M monovalent (Na^+ and K^+) and 20 mM of any of the tested divalent ions (Mg^{2+}, Ca^{2+}, Ba^{2+}, Sr^{2+}, Zn^{2+}) were similar (see **Figure 12**, green lines, for examples). For ease of comparison, the scattering profiles in the presence of 20 mM Mg^{2+} without (**Figure 12**, dashed green lines) and with 20 mM glycine (**Figure 12**, dashed red lines) are shown in each of the panels. This result indicates that the VCI-II tandem aptamer adopts similar global conformations, termed the M state, in the absence of glycine and in the presence of sufficient concentrations of the tested mono- and divalent ions.

In contrast, the scattering profiles in the presence of 20 mM glycine are markedly different with the different ion species (**Figure 12**, red lines).[94] Whereas the profiles for Ca^{2+} and Mn^{2+} are indistinguishable from the scattering pattern observed in the presence of Mg^{2+} and glycine, the profiles with added Zn^{2+}, Sr^{2+}, Ba^{2+}, Na^+ or K^+ do not change appreciably from those in the absence of glycine (see **Figure 12** for examples, green lines in the presence of glycine and red lines in the absence of glycine). The results indicate that Mg^{2+}, Ca^{2+} and Mn^{2+} can support glycine binding but that

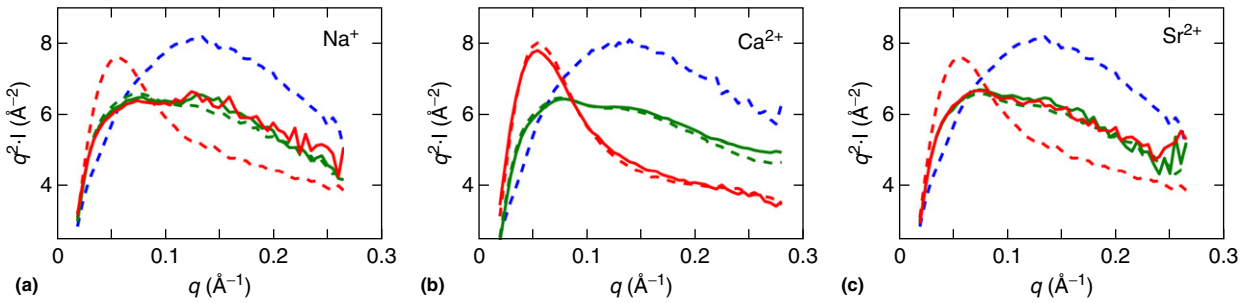

Figure 12 SAXS data for the VCI-II aptamer of the glycine riboswitch in the presence of different ions in the absence and presence of glycine. Kratky plots of SAXS profiles are plotted in the presence of 2 M Na^+ (panel a), 20 mM Ca^{2+} (panel b), and 20 mM Sr^{2+} (panel c). Profiles are shown in the absence (green, solid lines) and presence (red, solid lines) of 20 mM glycine.[94] For comparison, profiles are shown in 50 mM Na-MOPS buffer only (blue, dashed lines), in the presence of 20 mM Mg^{2+} (green, dashed lines), and in the presence of 20 mM Mg^{2+} and 20 mM glycine (red, dashed lines). Data adapted from Figures 2(a), 2(c) and 2(e) in Lipfert, J.; Sim, A. Y. L.; Herschlag, D.; Doniach, S. Dissecting electrostatic screening, specific ion binding, and ligand binding in an energetic model for glycine riboswitch folding. *RNA* **2010**, *16*, 708–719. Copyright by RNA Society.

the monovalent ions and other divalent ions cannot. This observation strongly suggests the presence of at least one specific divalent metal ion binding site that is formed in the final folded structure.

To analyze further the observed changes in R_g as a function of $[Mg^{2+}]$, a simple three-state thermodynamic model that describes the energetic coupling between magnesium-induced folding and glycine binding in terms of the dependence of ΔG on $[Mg^{2+}]$ and [glycine] concentrations was fit to the SAXS data. Then use of the bead-model arising from the 3-D-reconstruction was found to allow for a further understanding of the thermodynamics of the Mg^{2+} induced transition: using the Poisson-Boltzmann (PB) equation and a simplified representation of the electrostatics in which the backbone phosphate charges are distributed uniformly among the beads of the low-resolution model, an estimate of the change in electrostatic contribution to ΔG as a function of the Mg^{2+} concentration could be calculated.[94]

To allow for a simple comparison of the data characterizing the U to M transition, an empirical fit was applied to the Hill equation.[95] According to the Hill equation, the fractional occupancy of the M state (i.e., the amount of VCI-II RNA in the M conformation relative to the U state) is given by:

$$f_M = \frac{[\text{counterion}]^m}{K^m + [\text{counterion}]^m} \quad [31]$$

where [counterion] represents the concentration of the monovalent ($[Na^+]$) or divalent ($[Mg^{2+}]$ or $[Sr^{2+}]$) counterions. K is the concentration of Na^+ or Sr^{2+} required for both the U and M states to be equally populated at equilibrium; m is the Hill coefficient. We emphasize that the Hill model is used for mathematical convenience, as an empirical equation used to describe the data. Unlike in its usual interpretation, m does not correspond to the number of specifically bound ligands or ions but is rather a fitting parameter loosely describing the observed cooperativity of the transition.[96–99] From the fit to the SAXS data, $m = 2.7 \pm 0.5$ and the midpoint is in the range of ~ 100 mM Na^+.

The relative population of the U and M states is determined by their free energy difference, $\Delta G_{MU} = \Delta G_M - \Delta G_U$. The free energy difference ΔG_{MU} has several contributions, including electrostatic interactions, base stacking, hydrogen bonding, and conformational entropy of the U and M states. Many of these contributions are known to be difficult to quantify. However, as the primary interest of this analysis is in the salt-dependent contribution to the free energy difference, a distinction may be made between salt-independent contributions, ΔG^0_{MU} which includes many terms, and a salt-dependent term, $\Delta G_{MU}([X])$ with ion species X. The free energy difference between the U and M states at a given concentration of $[X]$ is then

$$\begin{aligned}\Delta G_{MU} &= \Delta G^0_{MU} + \Delta G_M([X]) - \Delta G_U([X]) \\ &= \Delta G^0_{MU} + \Delta\Delta G_{MU}([X])\end{aligned} \quad [32]$$

The salt-dependent term

$$\Delta\Delta G_{MU}([X]) = \Delta G_M([X]) - \Delta G_U([X]) \quad [33]$$

summarizes all of the components of the free energy that are dependent on the ionic conditions, such as the electrostatic interactions of ions in solution with the charged RNA molecule and possible effects due to specific ion binding. The low-resolution structures consist of dummy residues or beads. As there is no direct correspondence between beads and nucleic acid residues, the total charge of the molecule ($Q_{tot} = -226$ e, where e is the elementary charge) was divided equally amongst all N_{bead} beads, such that $Q_{bead} = Q_{tot}/N_{bead}$. PB energies were computed for these models as a function of salt concentrations using the software APBS.[100] The PB predictions were compared to the experimentally observed fractional occupancies of the M state for Na^+ and Mg^{2+} titrations in the absence of glycine using the relationship

$$f_M([X]) = 1 - \frac{1}{1 + \exp(-\Delta G^0_{MU} + \Delta\Delta G_{MU}([X]))/k_B T} \quad [34]$$

where $\Delta\Delta G_{MU}([X])$ is computed from PB theory and ΔG^0_{MU} is the fitting constant. The PB prediction of the salt dependence for the U to M transitions was found to agree well with the experimental observations for both the Na^+ ($\chi^2 = 0.58$, probability that random fluctuation accounts for the observed discrepancies, $p = 0.99$) and Mg^{2+} ($\chi^2 = 5.91$, $p = 0.83$) titrations (**Figure 13**).

For the glycine bound state, however, PB modeling of the electrostatic energy is insufficient. From the SAXS studies illustrated in **Figure 12**, it was found that Mg^{2+}, Mn^{2+}, and Ca^{2+}, but not Sr^{2+}, Ba^{2+}, and Zn^{2+}, support glycine binding and population of the glycine-bound B state. The discrimination of at least 100-fold between different species of divalent ions cannot be accounted for by simple electrostatic theories[101,102] and strongly suggests that formation of the B state involves specific ion site binding. Divalent ion titrations in a background of high concentrations of monovalent ions have been shown to isolate the effects of any specific site-bound divalent ions, as the monovalent background competes away nonspecifically interacting divalent ions.[97,103,104]

Specific ion binding will involve chemical binding with covalent contributions which are not taken into account in the PB theory. In this situation, the fitted Hill coefficient provides a lower bound on the number of specifically site-bound ions (and is equal to the number of specific binding sites if binding is perfectly cooperative). Fitting the Mg^{2+} titration data in the presence of glycine and 2 M NaCl to a two-state Hill model for the M to B transition gave a Mg^{2+} Hill coefficient of 2.8 ± 1.3 and a midpoint of 2.0 ± 1.1 mM Mg^{2+}. Previously published data and model predict a midpoint for the M to B transition of ~ 0.3 mM Mg^{2+} in the presence of negligible amounts of monovalent salt.[72] This value is much lower than the value of ~ 2.8 mM Mg^{2+} determined from the titrations in 2 M NaCl background. This difference is qualitatively expected if the M to B transition involves specific ion binding in the absence of a substantial overall compaction, as the presence of a high concentration of monovalent ions depletes the local ion atmosphere around the RNA of Mg^{2+} ions, making it more difficult to occupy specific divalent ion binding sites.[97,101,104]

Thus, the conclusion from these studies is that glycine-binding to the VCI-II tandem aptamer requires specific divalent

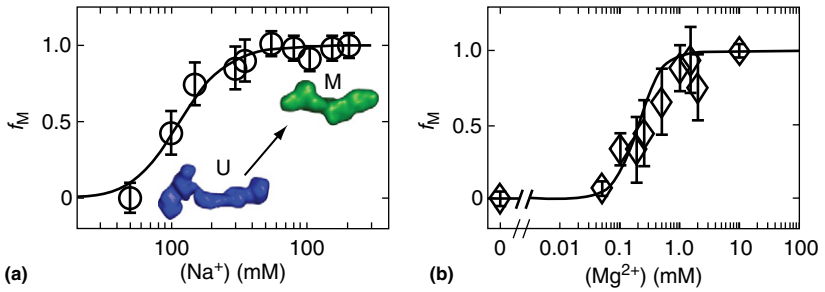

Figure 13 Ion-induced compaction of the glycine riboswitch. The U to M state transition as a function of Na^+ and Mg^{2+} concentrations monitored by SAXS compared with Poisson-Boltzmann (PB) predictions of the electrostatic effects from the ion atmosphere. (a) Fractional occupancy of the M state as a function of Na^+ concentration from SAXS (from R_g fitting, open circle symbols) and prediction from PB theory (solid line). Low resolution reconstructions of the U and M state from SAXS data are shown in the inset. (b) Fractional occupancy of the M state as a function of Mg^{2+} concentration from SAXS (from R_g fitting, open diamond symbols) and prediction from PB theory (solid line). Data adapted from Lipfert, J.; Das, R.; Chu, V. B.; Kudaravalli, M.; Boyd, N.; Herschlag, D.; Doniach, S. Structural transitions and thermodynamics of a glycine-dependent riboswitch from *Vibrio cholerae*. *J. Mol. Biol.* **2007**, *365*, 1393–1406 and Lipfert, J.; Sim, A. Y. L.; Herschlag, D.; Doniach, S. Dissecting electrostatic screening, specific ion binding, and ligand binding in an energetic model for glycine riboswitch folding. *RNA* **2010**, *16*, 708–719. Copyright by RNA Society.

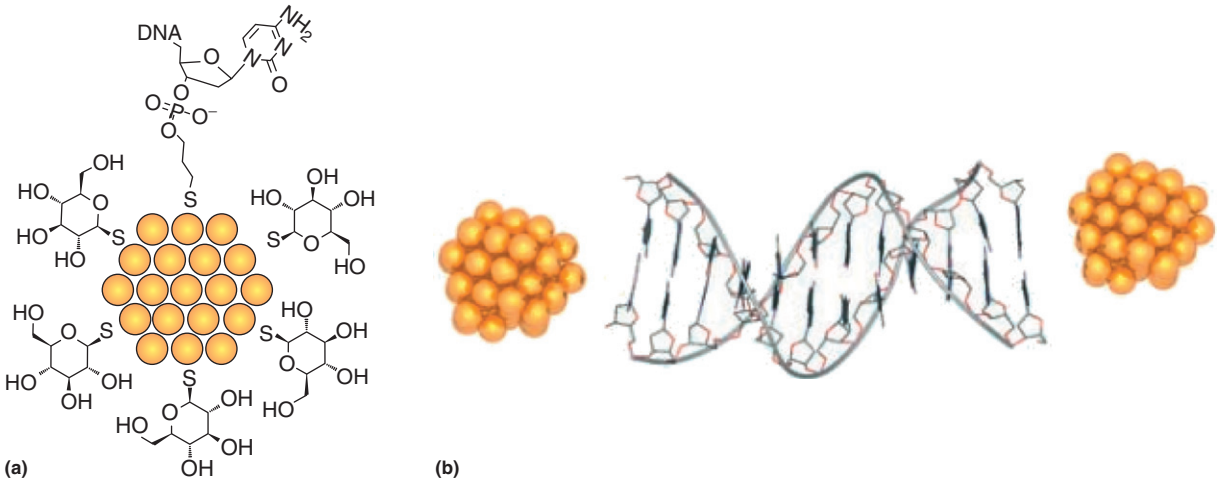

Figure 14 Gold-labeled DNA for SAXS measurements. (a) Schematic diagram of a thioglucose passivated gold nanocrystal coupled to a DNA oligonucleotide bearing a 3′ sulfhydryl group. The sulfhydryl group forms a bond directly to the nanocrystal core. (b) Model coordinates of a DNA duplex with a gold nanocrystal at either end (cluster ligands and propyl linkers are not shown). Adapted from Figure 1 in Mathew-Fenn, R. S.; Das, R.; Harbury, P. A. B. Remeasuring the double helix. *Science* **2008**, *322*, 446–449. Copyright by AAAS and Mathew-Fenn, R. S.; Das, R.; Silverman, J. A.; Walker, P. A.; Harbury, P. A. A molecular ruler for measuring quantitative distance distributions. *PLoS ONE* **2008**, *3*, e3229.

ion binding and, based on the observed Hill coefficient, suggests the presence of at least two specific ion binding sites.[94]

1.18.9 Use of Nanogold Markers for SAXS Determination of Intramolecular Distance Distributions

\The availability of gold nanoparticles covalently coated with surface thiol layers has engendered a wide range of applications in chemistry, biology and other fields.[105] From the point of view of X-ray scattering, gold is an ideal molecular label because of its extremely high electron density. The ability to attach nanogold particles to specific chemical groups in macromolecules such as DNA or proteins[106] then leads to a way to determine structural details involving the placement of the gold particles relative to an electron density map of the macromolecule, or as a molecular ruler to measure distances between specific groups in the molecule labeled by nanogold particles.

Recent work by Harbury and collaborators[107,108] has applied SAXS to study end-to-end fluctuations in small DNA oligonucleotides labeled with nanogold particles. In these experiments, gold nanoparticles are coupled to the ends of DNA oligomers via a sulfhydryl-gold bond (**Figure 14**).

The X-ray scattering amplitude from the labeled DNA may be written

$$F(\vec{q}) = \sum_i f_i^{\mathrm{DNA}} e^{i\vec{q}\cdot\vec{r}_i^{\mathrm{DNA}}} + \sum_j f_j^{Au} e^{i\vec{q}\cdot\vec{r}_j^{Au}} \qquad [35]$$

where $\vec{r}_i^{\,\mathrm{DNA}}$ are the positions of the DNA atoms, scattering factors f_i^{DNA}, and the $\vec{r}_j^{\,\mathrm{Au}}$ are the positions of the gold atoms, scattering factors f_j^{Au}. So it may be seen that the scattered intensity

$$I(q) = \int d\Omega_q \left|F(\vec{q})\right|^2 \qquad [36]$$

is made up of three terms: a DNA scattering term, a gold–gold scattering term, and a DNA–gold interference term:

$$I(q) = I^{\mathrm{DNA-DNA}}(q) + I^{\mathrm{Au-Au}}(q) + I^{\mathrm{DNA-Au}}(q) \qquad [37]$$

with

$$I^{\mathrm{DNA-DNA}}(q) = \int d\Omega_q \sum_{ij} f_i^{\mathrm{DNA}} f_j^{*\mathrm{DNA}} e^{i\vec{q}\cdot(\vec{r}_i^{\,\mathrm{DNA}} - \vec{r}_j^{\,\mathrm{DNA}})} \qquad [38]$$

$$I^{\mathrm{Au-Au}}(q) = \int d\Omega_q \sum_{ij} f_i^{\mathrm{Au}} f_j^{*\mathrm{Au}} e^{i\vec{q}\cdot(\vec{r}_i^{\,\mathrm{Au}} - \vec{r}_j^{\,\mathrm{Au}})} \qquad [39]$$

$$I^{\mathrm{DNA-Au}}(q) = \int d\Omega_q \sum_{ij} f_i^{\mathrm{DNA}} f_j^{*\mathrm{Au}} e^{i\vec{q}\cdot(\vec{r}_i^{\,\mathrm{DNA}} - \vec{r}_j^{\,\mathrm{Au}})} + \text{complex conjugate} \qquad [40]$$

For two gold particles, the gold–gold term in the total scattering further breaks up into a gold-self scattering term and an interference term between the two gold particles. Within a spherical approximation for the gold particles

$$I^{\mathrm{Au-Au}}(q)/2I^{\mathrm{Au}}(q) \simeq I^{\mathrm{Au,self}}(q) + I^{\mathrm{Au,interf}}(q) = I^{\mathrm{Au}}(q)\left(1 + \frac{\sin(qR_{12})}{qR_{12}}\right) \qquad [41]$$

where $R_{12} = \left|\vec{R}_1 - \vec{R}_2\right|$ and we have treated the two particles, centered at R_1 and R_2 as identical. $I^{\mathrm{Au}}(q)$ is the scattering profile for a single gold particle

$$I^{\mathrm{Au}}(q) = \sum_{ij} f_i^{\mathrm{Au}} f_j^{\mathrm{Au}} \frac{\sin(q|r_{ij}|)}{q|r_{ij}|}. \qquad [42]$$

Equation (41) leads to a direct measurement for the distance between the gold particles. In the case that there is an ensemble of labeled DNA molecules with probability distribution $P^{\mathrm{Au}}(R)$ of the gold–gold distances, then, within the spherical approximation, the pair distance distribution may be extracted by inverting the relation

$$I^{\mathrm{Au-Au}}(q)/2I^{Au}(q) = \int dR P^{\mathrm{Au}}(R)\left(1 + \frac{\sin(qR)}{qR}\right) \qquad [43]$$

As discussed in Section 1.18.2.2, this is an ill-conditioned problem which needs the application of prior constraints via a non-linear optimization method. To extract the gold–gold term $I^{\mathrm{Au-Au}}(q)$ from SAXS data, separate determinations of the individual gold particle scattering term $I^{\mathrm{Au}}(q)$, the interference term $I^{\mathrm{DNA-Au}}(q)$ and the unlabeled DNA scattering term $I^{\mathrm{DNA-DNA}}(q)$ are needed. Mathew-Fenn et al.[107,108] determined these terms by chemical preparation of a singly labeled DNA, and by direct measurement of the SAXS from the unbound gold particles, the DNA on its own and the singly labeled DNA in addition to the doubly labeled sample. They inverted the gold–gold scattering term by application of a non-negative least squares algorithm cross-checked by a maximum entropy algorithm.

Mathew-Fenn et al. used this method (**Figure 15**) to measure the mean and variance of end-to-end length for a series of DNA double helices in solution.[107,108] They found that the variance in end-to-end length follows a quadratic dependence on the number of base pairs rather than the linear dependence expected from a simple elastic model, suggesting

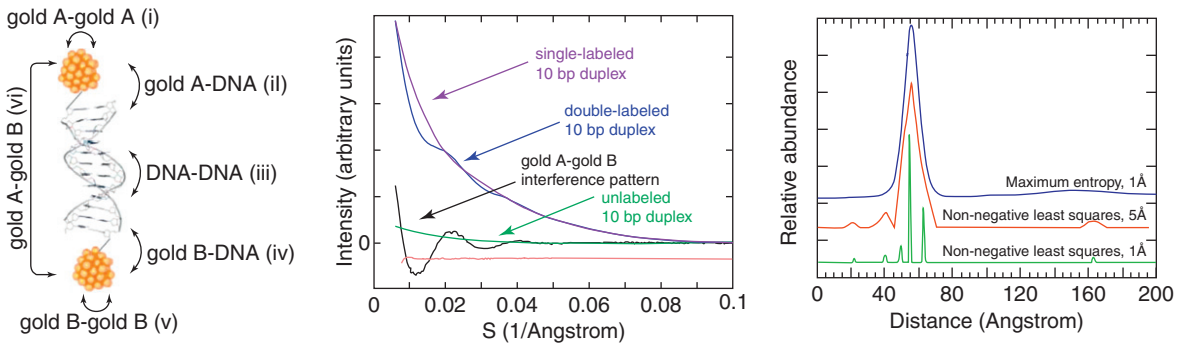

Figure 15 Gold nanolabels as distance markers in a SAXS experiment. Left panel: Model coordinates of a 12 base-pair DNA duplex bearing a gold nanocrystal at each end. Various types of scattering interference between gold nanocrystals and DNA are illustrated with labeled arrows. Center panel: The scattering profiles for 10 base-pair double-labeled (blue), single-labeled (purple, magenta; indistinguishable), and unlabeled (green) DNA duplexes. The probe-probe scattering interference pattern (black) is obtained by adding the double-labeled and unlabeled profiles and subtracting off the single-labeled profiles. The residual difference between this interference pattern and the transform of the probability distribution in the panel on the right is plotted in red, and offset downward. Right panel: Distance distributions obtained by decomposing the scattering interference pattern in the middle panel into a linear combination of the basis profiles. Three different transformation methods are illustrated. They are offset vertically from one another for clarity. Figure adapted from Mathew-Fenn, R. S.; Das, R.; Silverman, J. A.; Walker, P. A.; Harbury, P. A. A molecular ruler for measuring quantitative distance distributions. *PLoS ONE* **2008**, *3*, e3229.

that DNA stretching is cooperative over more than two turns of the DNA double helix.

1.18.9.1 Possible Use of Anomalous SAXS to Extract Distance Distributions from Gold-Labeled Biomolecules

An interesting possibility, which arises when using SAXS for gold-labeled molecules, is to measure the scattering as a function of X-ray energy (e.g., by tuning the incoming X-rays through the L3 gold resonance close to 11.9 keV). In this case, the gold scattering factors $f^{\mathrm{Au}}(\omega)$ become complex, with the real and imaginary parts varying markedly as a function of the X-ray frequency ω. Inserting $f^{\mathrm{Au}}(\omega) = f'(\omega) + if''(\omega)$ into eqns [39] and [40], we get

$$\begin{aligned} I^{\mathrm{DNA-Au}}(q) &= \int d\Omega_q \sum_{ij} f_i^{\mathrm{DNA}} f_j'(\omega) e^{i\vec{q}\cdot(\vec{r}_i^{\,\mathrm{DNA}} - \vec{r}_j^{\,\mathrm{Au}})} \\ &= f'(\omega) \sum_i f_i \underbrace{\frac{\sin(q|\vec{r}_i - \vec{R}|)}{q|\vec{r}_i - \vec{R}|}}_{I^{\mathrm{Au-Au}}(q)} \\ &= |f^{\mathrm{Au}}(\omega)|^2 \sum_{ij} \frac{\sin(q|r_{ij}^{\mathrm{Au}}|)}{q|r_{ij}^{\mathrm{Au}}|} \end{aligned} \quad [44]$$

So by tuning through the gold L3 edge, one can, in principle, extract the gold–DNA interference term without having to prepare a separate singly labeled sample. In practice, however, this has proved to be a difficult task, possibly associated with problems in mechanical reproducibility of the monochromator setting of the X-ray frequency as sweeps through the gold edge are repeated.

Acknowledgments

This work was supported by the Netherlands Organisation for Scientific Research (NWO) via a VENI grant to J.L., by the National Institutes of Health grant PO1 GM0066275, and by the Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, under contract DE-AC02-76SF00515. Use of the Advanced Photon Source, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the U.S. DOE under Contract No. DE-AC02-06CH11357.

References

[1] Guinier, A.; Fournet, G. Small-Angle Scattering of X-rays. Wiley: New York, 1955.
[2] Doniach, S. Changes in biomolecular conformations seen by small angle x-ray scattering. *Chem. Rev.* **2001**, *101*, 1763–1778.
[3] Koch, M. H. J.; Vachette, P.; Svergun, D. I. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys.* **2003**, *36*, 147–227.
[4] Lipfert, J.; Doniach, S. Small-angle x-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 307–327.
[5] Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **2007**, *40*, 191–285.
[6] Svergun, D. I.; Koch, M. H. J. Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* **2003**, *66*, 1735–1782.
[7] Als-Nielsen, A.; McMorrow, D. *Elements of Modern X-Ray Physics*. John Wiley & Sons, Ltd: West Sussex, UK, 2001.
[8] Svergun, D. I.; Barberato, C.; Koch, M. H. J. Evaluation of x-ray solution scattering curves from atomic models. *J. Appl. Crystallogr.* **1995**, *28*, 768–773.
[9] Guinier, A. La diffraction des rayons X aux très petits angles: Application à l'étude de phénomènes ultramicroscopiques. *Ann. Phys. (Paris)* **1939**, *12*, 161–237.
[10] Betzig, E.; Patterson, G. H.; Sougrat, R.; Lindwasser, O. W.; Olenych, S.; Bonifacino, J. S.; Davidson, M. W.; Lippincott-Schwartz, J.; Hess, H. F. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **2006**, *313*, 1642–1645.
[11] Hell, S. W. Far-field optical nanoscopy. *Science* **2007**, *316*, 1153–1158.
[12] Rust, M. J.; Bates, M.; Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **2006**, *3*, 793–795.
[13] Efron, B. *The Bootstrap and Modern Statistics*. Stanford University, Division of Biostatistics: Stanford, CA, 1999.
[14] Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*. Cambridge University Press: Cambridge, UK, 1992.
[15] Svergun, D. I. Determination of the regularization parameter in indirect transform methods using perceptual criteria. *J. Appl. Crystallogr* **1992**, *25*, 495–503.
[16] Tikhonov, A.; Arsenin, V. *Solution of Ill-Posed Problems*. Wiley: New York, 1977.
[17] Petoukhov, M. V.; Konarev, P. V.; Kikhney, A. G.; Svergun, D. I. Atsas 2.1 – towards automated and web-supported small-angle scattering data analysis. *J. Appl. Crystallogr.* **2007**, *40*, S223–S228.
[18] Tardieu, A.; Bonneté, F.; Finet, S.; Vivarés, D. Understanding salt or peg induced attractive interactions to crystallize biological macromolecules. *Acta Crystallogr.* **2002**, *D58*, 1549–1553.
[19] Verwey, E. J. W.; Overbeek, J. T. G. *Theory of the Stability of Lyophobic Colloids*. Elsevier: Amsterdam, 1948.
[20] Malfois, M.; Bonnete, F.; Belloni, L.; Tardieu, A. A model of attractive interactions to account for fluid-fluid phase separation of protein solutions. *J. Chem. Phys.* **1996**, *105*, 3290–3300.
[21] Ducruix, A.; Guilloteau, J. P.; Riès-Kautt, M.; Tardieu, A. Protein interactions as seen by solution x-ray scattering prior to crystallogenesis. *J. Cryst. Growth* **1996**, *168*, 28–39.
[22] Hubbard, S.; Doniach, S. A. Monte-Carlo calculation of the interparticle interference in small-angle X-ray-scattering. *J. Appl. Crystallogr.* **1988**, *21*, 953–959.
[23] Levitt, M.; Sharon, R. Accurate simulation of protein dynamics in solution. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 7557–7561.
[24] Svergun, D.; Richard, S.; Koch, M.; Sayers, Z.; Kuprin, S.; Zaccai, G. Protein hydration in solution: Experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 2267–2272.
[25] Meyer, A. S.; Gillespie, J. R.; Walther, D.; Millet, I. S.; Doniach, S.; Frydman, J. Closing the folding chamber of the eukaryotic chaperonin requires the transition state of ATP hydrolysis. *Cell* **2003**, *113*, 369–381.
[26] Merzel, F.; Smith, J. C. SASSIM: a method for calculating small-angle X-ray and neutron scattering and the associated molecular envelope from explicit-atom models of solvated proteins. *Acta Crystallogr. D Biol. Crystallogr.* **2002**, *58*, 242–249.
[27] Shapiro, H.J. *Simulations of the Hydration of Folded and Unfolded Proteins, with Comparison to Small-Angle X-ray Scattering*, Ph.D. thesis, Stanford University, Dept. of Physics, Stanford, CA, 2002.
[28] Fliege, J.; Maier, U. The distribution of points on the sphere and corresponding cubature formulae. *IMA J. Numer. Anal.* **1999**, *19*, 317–334.
[29] Park, S.; Bardhan, J. P.; Roux, B.; Makowski, L. Simulated x-ray scattering of protein solutions using explicit-solvent models. *J. Chem. Phys.* **2009**, *130*, 134114.
[30] Miyata, T.; Hirata, F. Combination of molecular dynamics method and 3-D-RISM theory for conformational sampling of large flexible molecules in solution. *J. Comp. Chem.* **2008**, *29*, 871–882.
[31] Koehl, P.; Orland, H.; Delarue, M. Computing ion solvation free energies using the dipolar Poisson model. *J. Phys. Chem. B.* **2009**, *113*, 5694–5697.

[32] Koehl, P.; Delarue, M. AQUASOL: An efficient solver for the dipolar Poisson-Boltzmann-Langevin equation. *J. Chem. Phys.* **2010**, *132*, 064101.

[33] Frängsmyr, T.; Malmström, B. G. *Jerome Karle: Nobel Lectures in Chemistry 1981–1990.* World Scientific Publishing Co.: Singapore, 1992.

[34] Svergun, D. I.; Stuhrmann, H. B. New developments in direct shape determination from small-angle scattering. 1. Theory and model-calculations. *Acta Crystallogr.* **1991**, *A47*, 736–744.

[35] Chacon, P.; Moran, F.; Diaz, J. F.; Pantos, E.; Andreu, J. M. Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophys. J.* **1998**, *74*, 2760–2775.

[36] Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **1999**, *76*, 2879–2886.

[37] Svergun, D. I.; Petoukhov, M. V.; Koch, M. H. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **2001**, *80*, 2946–2953.

[38] Walther, D.; Cohen, F. E.; Doniach, S. Reconstruction of low resolution three-dimensional density maps from one-dimensional small angle x-ray scattering data for biomolecules in solution. *J. Appl. Crystallogr.* **2000**, *33*, 350–363.

[39] Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **1999**, *76*, 2879–2886.

[40] Volkov, V. V.; Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **2003**, *36*, 860–864.

[41] Kretchetov, A. *Reconstruction of 3-D Electron Density Models from Small Angle X-Ray Scattering Data for Macromolecules: A Space Averaging Approach*, Ph.D. thesis, Stanford University, Dept. of Applied Physics, Stanford, CA, 2005.

[42] Hura, G. L.; Menon, A. L.; Hammel, M.; Rambo, R. P.; Poole, II F. L.; Tsutakawa, S. E.; Jenney, Jr. F. E.; Classen, S.; Frankel, K. A.; Hopkins, R. C.; Yang, S.-J.; Scott, J. W.; Dillard, B. D.; Adams, M. W. W.; Tainer, J. A. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods.* **2009**, *6*, 606–U83.

[43] Zheng, W.; Doniach, S. Protein structure prediction constrained by solution X-ray scattering data and structural homology identification. *J. Mol. Biol.* **2002**, *316*, 173–187.

[44] Wu, Y.; Tian, X.; Lu, M.; Chen, M.; Qinghua, W.; Ma, J. Folding of small helical proteins assisted by small-angle x-ray scattering profiles. *Structure* **2005**, *13*, 1587–1597.

[45] Segel, D. J.; Fink, A. L.; Hodgson, K. O.; Doniach, S. Protein denaturation: A small-angle x-ray scattering study of the ensemble of unfolded states of cytochrome c. *Biochemistry* **1998**, *37*, 12443–12451.

[46] Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1528–1533.

[47] Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17002–17007.

[48] Bernado, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **2007**, *129*, 5656–5664.

[49] Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13099–13104.

[50] Zagrovic, B.; Lipfert, J.; Sorin, E.; Millettt, I.; van Gunsteren, W.; Doniach, S.; Pande, V. Unusual compactness of a polyproline type II structure. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11698–11703.

[51] Makowska, J.; Rodziewicz-Motowidlo, S.; Baginska, K.; Makowski, M.; Vila, J. A.; Liwo, A.; Chmurzynski, L.; Scheraga, H. A. Further evidence for the absence of polyproline II stretch in the XAO peptide. *Biophys. J.* **2007**, *92*, 2904–2917.

[52] Makowska, J.; Rodziewicz-Motowidlo, S.; Baginska, K.; Vila, J.; Liwo, A.; Chmurzynski, L.; Scheraga, H. Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1744–1749.

[53] Poitevin, F.; Orland, H.; Doniach, S.; Koehl, P.; Delarue, M. AquaSAXS: A web server for computation and fitting of SAXS profiles with non-uniformly hydrated atomic models. *Nucleic Acids Res.* **2011**, *39*, W184–W189.

[54] Azuara, C.; Orland, H.; Bon, M.; Koehl, P.; Delarue, M. Incorporating dipolar solvents with variable density in Poisson-Boltzmann electrostatics. *Biophys. J.* **2008**, *95*, 5587–5605.

[55] Wallin, E.; von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **1998**, *7*, 1029–1038.

[56] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

[57] Chen, S. H. Small angle neutron scattering studies of the structure and interactions in micellar and microemulsion systems. *Annu. Rev. Phys. Chem.* **1986**, *37*, 351–399.

[58] Columbus, L.; Lipfert, J.; Klock, H.; Millett, I. S.; Doniach, S.; Lesley, S. Expression, purification and characterization of Thermotoga maritima α-helical membrane proteins for structure determination. *Protein Sci.* **2006**, *15*, 961–975.

[59] Lipfert, J.; Columbus, L.; Chu, V. B.; Doniach, S. Analysis of small-angle x-ray scattering data of protein-detergent complexes by singular value decomposition. *J. Appl. Crystallogr.* **2007**, *40*, S235–S239.

[60] Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Sligar, S. G.; Schulten, K. Assembly of lipids and proteins into lipoprotein particles. *J. Phys. Chem. B* **2007**, *111*, 11095–11104.

[61] Lipfert, J.; Columbus, L.; Chu, V. B.; Lesley, S. A.; Doniach, S. Size and shape of detergent micelles determined by small-angle x-ray scattering. *J. Phys. Chem. B* **2007**, *111*, 12427–12438.

[62] Columbus, L.; Lipfert, J.; Jambunathan, K.; Fox, D. A.; Sim, A. Y. L.; Doniach, S.; Lesley, S. A. Mixing and matching detergents for membrane protein nmr structure determination. *J. Am. Chem. Soc.* **2009**, *131*, 7320–7326.

[63] Grishaev, A.; Wu, J.; Trewhella, J.; Bax, A. Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *J. Am. Chem. Soc.* **2005**, *127*, 16621–16628.

[64] Grishaev, A.; Tugarinov, V.; Kay, L. E.; Trewhella, J.; Bax, A. Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. *J. Biomol. NMR* **2008**, *40*, 95–106.

[65] Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **1982**, *31*, 147–157.

[66] Doniach, S.; Lipfert, J. Use of small angle x-ray scattering (SAXS) to characterize conformational states of functional RNAs. *Methods Enzymol* **2009**, *469*, 237–251.

[67] Lipfert, J.; Chu, V. B.; Bai, Y.; Herschlag, D.; Doniach, S. Low resolution models for nucleic acids from small-angle x-ray scattering with applications to electrostatic modeling. *J. Appl. Crystallogr* **2007**, *40*, S229–S234.

[68] Lipfert, J.; Herschlag, D.; Doniach, S. Riboswitch conformations revealed by small-angle X-ray scattering. *Methods Mol. Biol.* **2009**, *540*, 141–159.

[69] Rambo, R. P.; Tainer, J. A. Improving small-angle X-ray scattering data for structural analyses of the RNA world. *RNA* **2010**, *16*, 638–646.

[70] Baird, N. J.; Ferre-D'Amare, A. R. Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis. *RNA* **2010**, *16*, 598–609.

[71] Kulshina, N.; Baird, N. J.; Ferre-D'Amare, A. R. Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch. *Nat. Struct. Mol. Biol.* **2009**, *16*, 1212–1217.

[72] Lipfert, J.; Das, R.; Chu, V. B.; Kudaravalli, M.; Boyd, N.; Herschlag, D.; Doniach, S. Structural transitions and thermodynamics of a glycine-dependent riboswitch from Vibrio cholerae. *J. Mol. Biol.* **2007**, *365*, 1393–1406.

[73] Lipfert, J.; Ouellet, J.; Norman, D.; Doniach, S.; Lilley, D. The complete VS ribozyme in solution studied by small-angle X-ray scattering. *Structure* **2008**, *16*, 1357–1367.

[74] Wang, J.; Zuo, X.; Yu, P.; Xu, H.; Starich, M. R.; Tiede, D. M.; Shapiro, B. A.; Schwieters, C. D.; Wang, Y. X. A method for helical RNA global structure determination in solution using small-angle x-ray scattering and NMR measurements. *J. Mol. Biol.* **2009**, *393*, 717–734.

[75] Zuo, X.; Wang, J.; Foster, T. R.; Schwieters, C. D.; Tiede, D. M.; Butcher, S. E.; Wang, Y. X. Global molecular structure and interfaces: refining an RNA:RNA complex structure using solution X-ray scattering data. *J. Am. Chem. Soc.* **2008**, *130*, 3292–3293.

[76] Lilley, D. M. J. The varkud satellite ribozyme. *RNA* **2004**, *10*, 151–158.

[77] Saville, B. J.; Collins, R. A. A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell* **1990**, *61*, 685–696.

[78] Wriggers, W.; Milligan, R. A.; McCammon, J. A. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **1999**, *125*, 185–195.
[79] Lafontaine, D. A.; Norman, D. G.; Lilley, D. M. Structure, folding and activity of the VS ribozyme: importance of the 2-3-6 helical junction. *EMBO J.* **2001**, *20*, 1415–1424.
[80] Lafontaine, D. A.; Norman, D. G.; Lilley, D. M. J. The global structure of the VS ribozyme. *EMBO J.* **2002**, *21*, 2461–2471.
[81] Das, R.; Baker, D. Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 14664–14669.
[82] Das, R.; Karanicolas, J.; Baker, D. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods* **2010**, *7*, 291–294.
[83] Flores, S. C.; Altman, R. B. Turning limited experimental information into 3-D models of RNA. *RNA* **2010**, *16*, 1769–1778.
[84] Parisien, M.; Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **2008**, *452*, 51–55.
[85] Kubodera, T.; Watanabe, M.; Yoshiuchi, K.; Yamashita, N.; Nishimura, A.; Nakai, S.; Gomi, K.; Hanamoto, H. Thiamine-regulated gene expression of Aspergillus oryzae thiA requires splicing of the intron containing a riboswitch-like domain in the 5′-UTR. *FEBS Lett.* **2003**, *555*, 516–520.
[86] Sudarsan, N.; Wickiser, J.; Nakamura, S.; Ebert, M.; Breaker, R. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* **2003**, *17*, 2688–2697.
[87] Ali, M.; Lipfert, J.; Seifert, S.; Herschlag, D.; Doniach, S. The ligand-free state of the TPP riboswitch: A partially folded RNA structure. *J. Mol. Biol.* **2010**, *396*, 153–165.
[88] Serganov, A.; Polonskaia, A.; Phan, A.; Breaker, R.; Patel, D. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* **2006**, *441*, 1167–1171.
[89] Thore, S.; Leibundgut, M.; Ban, N. Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science* **2006**, *312*, 1208–1211.
[90] Mayer, G.; Raddatz, M.-S. L.; Grunwald, J. D.; Famulok, M. RNA ligands that distinguish metabolite-induced conformations in the TPP riboswitch. *Angew. Chem. Int. Ed.* **2007**, *46*, 557–560.
[91] Lang, K.; Rieder, R.; Micura, R. Ligand-induced folding of the thiM TPP riboswitch investigated by a structure-based fluorescence spectroscopic approach. *Nucleic Acids Res.* **2007**, *35*, 5370–5378.
[92] Yang, S.; Parisien, M.; Major, F.; Roux, B. RNA Structure determination using SAXS data. *J. Phys. Chem. B* **2010**, *114*, 10039–10048.
[93] Mandal, M.; Lee, M.; Barrick, J. E.; Weinberg, Z.; Emilsson, G. M.; Ruzzo, W. L.; Breaker, R. R. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **2004**, *306*, 275–279.
[94] Lipfert, J.; Sim, A. Y. L.; Herschlag, D.; Doniach, S. Dissecting electrostatic screening, specific ion binding, and ligand binding in an energetic model for glycine riboswitch folding. *RNA* **2010**, *16*, 708–719.
[95] Hill, A. V. The possible effects of the aggregation of the molecules of haemoglobin on its oxygen dissociation curve. *J. Physiol. (Lond.)* **1910**, *40*, 4–7.
[96] Bai, Y.; Chu, V. B.; Lipfert, J.; Pande, V. S.; Herschlag, D.; Doniach, S. Critical assessment of nucleic acid electrostatics via experimental and computational investigation of an unfolded state ensemble. *J. Am. Chem. Soc.* **2008**, *130*, 12334–12341.
[97] Das, R.; Travers, K. J.; Bai, Y.; Herschlag, D. Determining the Mg^{2+} stoichiometry for folding an RNA metal ion core. *J. Am. Chem. Soc.* **2005**, *127*, 8272–8273.
[98] Draper, D. E. RNA folding: Thermodynamic and molecular descriptions of the roles of ions. *Biophys. J.* **2008**, *95*, 5489–5495.
[99] Misra, V. K.; Draper, D. E. The interpretation of Mg^{2+} binding isotherms for nucleic acids using Poisson-Boltzmann theory. *J. Mol. Biol.* **1999**, *294*, 1135–1147.
[100] Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10037–10041.
[101] Bai, Y.; Greenfeld, M.; Travers, K. J.; Chu, V. B.; Lipfert, J.; Doniach, S.; Herschlag, D. Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J. Am. Chem. Soc.* **2007**, *129*, 14981–14988.
[102] Chu, V. B.; Bai, Y.; Lipfert, J.; Herschlag, D.; Doniach, S. Evaluation of ion binding to DNA duplexes using a size-modified Poisson-Boltzmann theory. *Biophys J.* **2007**, *93*, 3202–3209.
[103] Bukhman, Y.; Draper, D. Affinities and selectivities of divalent cation binding sites within an RNA tertiary structure. *J. Mol. Biol.* **1997**, *273*, 1020–1031.
[104] Travers, K. J.; Boyd, N.; Herschlag, D. Low specificity of metal ion binding in the metal ion core of a folded RNA. *RNA* **2007**, *13*, 1205–1213.
[105] Daniel, M.; Astruc, D. Gold nanoparticles: Assembly, supramolecular chemistry, quantum-size-related properties, and applications toward biology, catalysis, and nanotechnology. *Chem. Rev.* **2004**, *104*, 293–346.
[106] Jadzinsky, P. D.; Calero, G.; Ackerson, C. J.; Bushnell, D. A.; Kornberg, R. D. Structure of a thiol monolayer-protected gold nanoparticle at 1.1 angstrom resolution. *Science* **2007**, *318*, 430–433.
[107] Mathew-Fenn, R. S.; Das, R.; Harbury, P. A. B. Remeasuring the double helix. *Science* **2008**, *322*, 446–449.
[108] Mathew-Fenn, R. S.; Das, R.; Silverman, J. A.; Walker, P. A.; Harbury, P. A. A molecular ruler for measuring quantitative distance distributions. *PLoS ONE* **2008**, *3*, e3229.
[109] Broennimann, C.; Eikenberry, E. F.; Henrich, B.; Horisberger, R.; Huelsen, G.; Pohl, E.; Schmitt, B.; Schulze-Briese, C.; Suzuki, M.; Tomizaki, T.; Toyokawa, H.; Wagner, A. The PILATUS 1M detector. *J. Synchrotron Radiat.* **2006**, *13*, 120–130.