Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

So how do you know you have a macromolecular complex?

Timothy R. Dafforn

Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, England

Correspondence e-mail: t.r.dafforn@bham.ac.uk

Received 30 June 2006
Accepted 7 November 2006

Protein in crystal form is at an extremely high concentration and yet retains the complex secondary structure that defines an active protein. The protein crystal itself is made up of a repeating lattice of protein–protein and protein–solvent interactions. The problem that confronts any crystallographer is to identify those interactions that represent physiological interactions and those that do not. This review explores the tools that are available to provide such information using the original crystal liquor as a sample. The review is aimed at postgraduate and postdoctoral researchers who may well be coming up against this problem for the first time. Techniques are discussed that will provide information on the stoichiometry of complexes as well as low-resolution information on complex structure. Together, these data will help to identify the physiological complex.

1. Introduction: why do we need to know about complexes?

'No man is an island' (John Donne, 1573–1631). Biology thrives through interactions, from the interactions between organisms that make up the biosphere to the interactions between molecules and atoms in the cell. A complete knowledge of these complex associations has the potential to allow us to understand nature. It is the central aim of biology to attain that knowledge.

Of all these biological interactions, perhaps the hardest for the 'man on the street' to understand are those he cannot see. These interactions (between cells, molecules and atoms) have been the objective of biological research for only a hundred years and already much progress has been made. One of the most revolutionary developments of the past fifty years has been the development of techniques that allow us to 'look' directly at these interactions by peering into the very workings of life itself. The trailblazer in this study has been protein X-ray crystallography. Since Max Perutz determined the structure of haemoglobin (Perutz, 1954) and John Kendrew that of myoglobin, the structures produced by X-ray crystallography have intrigued scientists across disciplines. X-ray crystal structures of proteins have not only shown us the beautiful convoluted shape of the peptide backbone, but have also provided information on their interactions.

In the early days, the number of monomers in the complex was generally already well established by biochemical and biophysical studies, making the interpretation of the associations in the crystal a trivial exercise. As more structures were solved, more complexes were determined, but the technically

© 2007 International Union of Crystallography
Printed in Denmark – all rights reserved

Novartis Exhibit 1018, Page 1 of 9

difficult nature of X-ray crystallography meant that again most of these were well studied in solution and hence the physiological relevance of the complex was easily determined. However, during the later part of the last century the technology and protocols used for X-ray crystallography improved and the number of proteins crystallized increased rapidly. It has now become the case that proteins are being crystallized using high-throughput techniques (Terwilliger *et al.*, 2003; Pusey *et al.*, 2005) with only limited biophysical and biochemical characterization of the protein sample. This has led

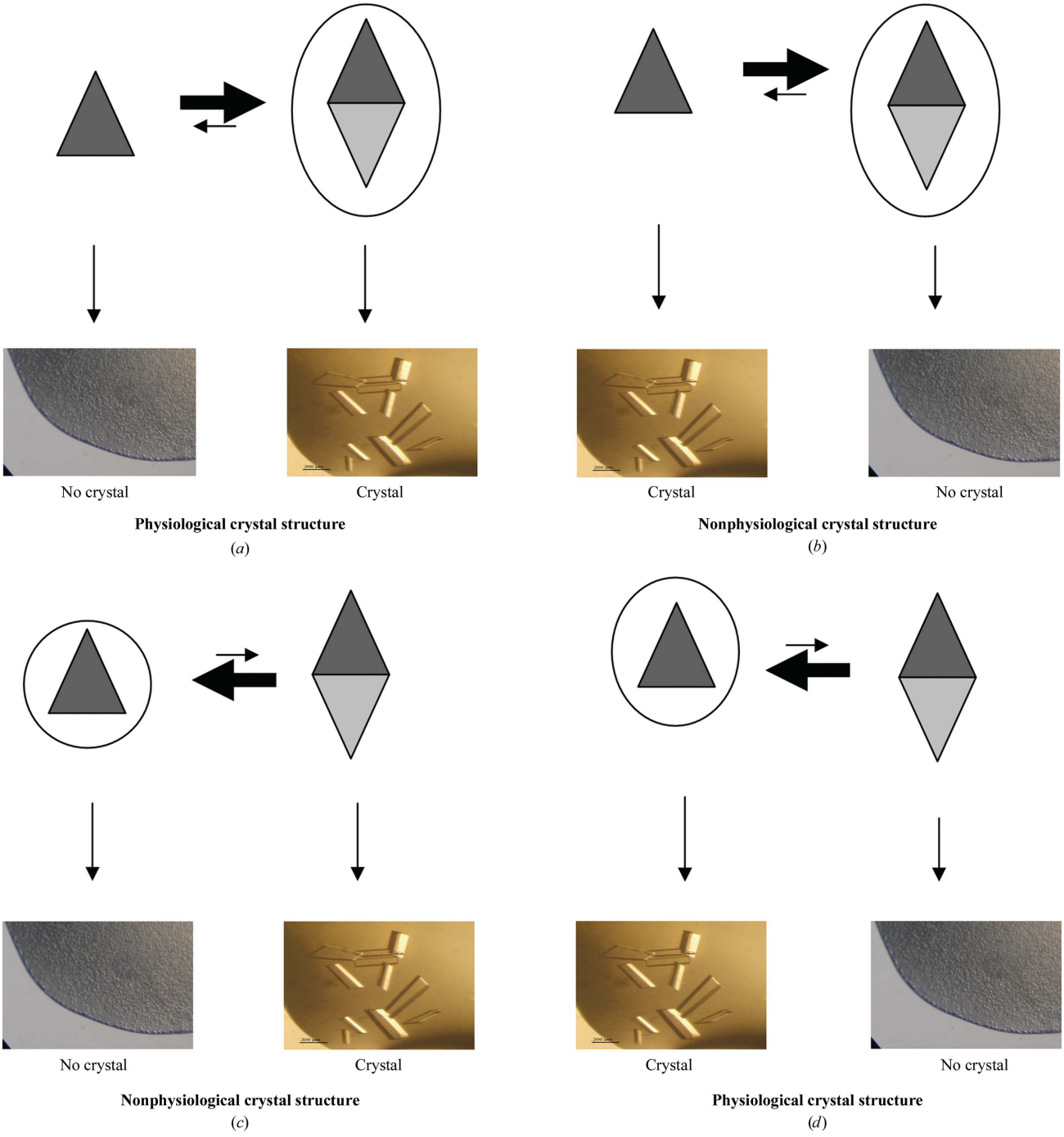

Figure 1
The process of crystallization may select nonphysiological protein associations. (*a*) The physiological state of the protein is a dimer and the dimer can be crystallized to provide a structure. (*b*) The physiological state of the protein is a dimer. The dimer cannot pack into a lattice to produce a crystal, but the monomer alone can. Therefore, the crystal structure contains the nonphysiological state. (*c*) and (*d*) demonstrate an analogous case where the physiological state is a monomer. For clarity, the physiological oligomerization state is circled.

Novartis Exhibit 1018, Page 2 of 9

to the current situation where, in order to determine the biologically relevant complex in a crystal, the scientist has had to return to the techniques of biophysics (Perugini *et al.*, 2005). One approach to determining the 'real' oligomerization state of a protein in a crystal structure has been through computational analysis. Computational biologists have developed a number of algorithms that have the potential to differentiate between physiological and nonphysiological interactions in a crystal (Janin *et al.*, 1988; Wang & Janin, 1993; Janin & Rodier, 1995; Henrick & Thornton, 1998; Robert & Janin, 1998; Bahadur *et al.*, 2004). These algorithms, although important as an indicator, are still not completely reliable. Thus, biophysical and biochemical characterization of protein is essential for determination of protein association states. This review aims to summarize these techniques.

The review provides an overview of the techniques that are available to examine protein–protein associations. Although many techniques exist for such studies, I have concentrated on those that can be applied to the samples used in crystal trials. Hence I have not included the most sensitive techniques as, in general, significant quantities of protein are available.

To begin, I will discuss why complexes in protein crystals are not always those that are relevant in physiology. I will then address the outwardly simple task of determining just how many monomers make up the physiological complex. Thirdly, I will look at a situation where the overall order of the association is unimportant, but where the crystal presents us with a number of possible monomer–monomer orientations which must be distinguished.

2. Why do complexes in crystals not match complexes in biology?

As has been discussed in the previous section, understanding the formation of protein complexes has two direct implications on our understanding of biological systems. So why, if an X-ray crystal structure of a protein provides the coordinates of all the non-H atoms in a protein, can we not always determine the stoichiometry of a protein complex? Surely it should be as simple as counting how many monomeric units are in close contact with one another?

If we take a step back and think about the crystal and crystallization process, then the answer is clear. The conditions for crystallizations are designed to induce protein–protein interactions which will result in a crystal, which after all is the 'mother of all protein complexes'. This is immediately going to cause us problems, as a crystal structure is likely to contain protein–protein interactions that are not physiological but that are stabilized through crystal packing. A limited study of the PDB by Bahadur and colleagues has shown more than 100 protein dimers in the database for proteins that are monomers in solution (Bahadur *et al.*, 2004).

If we examine the process that leads to the production of a protein crystal, then another potential flaw in the process can be appreciated. If we think of a dimeric protein complex, what we must remember is that this is an equilibrium between the monomer state and the dimer state (Fig. 1). The equilibrium position is determined by the affinities of the monomeric units for each other.

$$M + M \Leftrightarrow MM$$

If the dimer is able interact with other dimers in the crystallization liquor to form an ordered three-dimensional association, then a crystal will form containing the dimer. However, it is possible that the dimer in the liquor cannot propagate to form a crystal. In the simplest case this leads to no crystals, a disappointed crystallographer and no structure. However, as the process is in equilibrium, it is possible that some free monomer exists. This is particularly possible given the nonphysiological solution conditions in most crystallization screens. The free monomer could associate with other monomers in a manner that does not form a physiological dimer. This association could propagate to form a crystal, in this case without a physiological association.

$$\begin{array}{c} M + M \Leftrightarrow MM \\ \downarrow \\ \text{Crystal} \end{array}$$

3. What changes upon complex formation?

If it is the aim of a study to examine the potential oligomeric state of a new protein, then it is worth, for a moment, considering the consequences of protein oligomerization. Such consideration will allow potential signals of complex formation to be identified. The most obvious of all physical changes that accompany the formation of a complex is an increase in the molecular weight of the particles in solution. Theoretically, all that is required to characterize an associating system is to measure this molecular weight over a range of particle concentrations. Analysis of these data will provide the order of the oligomerization mechanism (*e.g.* monomer–dimer or monomer–dimer–tetramer *etc.*) as well as affinities for each step. Such information is obtainable (notably using analytical ultracentrifugation) and provides the greatest opportunity for the complete characterization of a system. However, a number of other physical characteristics can also be inherently linked to the formation of oligomers. The binding of one monomer to a second will lead to a reduction in solvent accessibility of the monomer–monomer binding site. On occasions, this change can be exploited to measure complex formation. For example, if the formation of dimers results in burial of hydrophobic surfaces, then a dye (such as ANSA) which changes its spectroscopic character when in contact with a hydrophobic surface can be used to monitor association (Dafforn *et al.*, 1999). The docking of one monomer to another can also alter the environment of amino acids on the common surface. This disturbance can be detected using fluorescence, near-UV circular dichroism (CD; Zsila *et al.*, 2004; Patel *et al.*, 2006) or nuclear magnetic resonance (NMR; Hewitt *et al.*, 1999; Lucas *et al.*, 2003; Zartler *et al.*, 2003). If those residues are aromatic residues such as tryptophan, phenylalanine or tyrosine, then

Novartis Exhibit 1018, Page 3 of 9

changes in fluorescence can be used as a signal for the formation of complex (Owen *et al.*, 1999; Lakowitz, 2006). Formation of a complex can also induce larger changes in the monomer architecture, leading to changes in backbone conformation. These types of changes can be measured using far-UV CD (Kelly & Price, 2000; Misenheimer *et al.*, 2003), Fourier transform infrared (FTIR) (Cooper & Knutson, 1995; Jackson & Mantsch, 1995) spectroscopy or NMR. In some cases, these changes have functional implications; for instance, altering the activity of an enzyme. In these cases, simple enzyme assays can be employed to provide information on oligomerization.

4. How many monomeric units are in the physiological complex?

As mentioned earlier, in the context of crystallography it is often important to determine the true stoichiometry of a complex in solution in order to understand the structure present in the crystal structure. Solving this problem seems like a relatively trivial exercise and many crystallographers maintain that 'careful' examination of the crystal structure will yield the physiologically relevant oligomer. However, there is now a groundswell of opinion that in a significant number of cases the declared oligomeric structures in the PDB database are nonphysiological (Bahadur *et al.*, 2004). In the event that a researcher does set out to determine the solution composition of a protein complex, the methods available are relatively limited. The most popular approach to this problem is invariably size-exclusion chromatography (SEC). SEC utilizes a porous chromatographic matrix which allows particles smaller than the pore size to partition into a larger space than particles larger than the pore size (for an excellent review of the details of SEC, see Winzor, 2003). This means that large particles traverse the column bed more rapidly than small particles, leading to a separation by size. Size-exclusion chromatography has the main advantage that it is relatively cheap and is easy to carry out. However, as is often the case, apparent simplicity in fact belies a very complex process with many factors that can lead to erroneous results. An idealized SEC matrix is utterly inert, allowing no interaction between the particles in solution and itself. Why does this make an ideal matrix? If a particle is able to interact with the matrix, its flow through the column will be retarded (Fig. 2). This retardation will then be erroneously interpreted as a lower relative molecular weight than the true one. Manufacturers have worked hard to reduce these interactions by reducing the charge density of the column to a minimum *etc*. However, the highly variant chemical nature of protein surfaces makes them very effective at adhering to a range of materials. In many cases, the buffer conditions used during the SEC experiment can be altered to reduce interactions with the column. A common approach is to increase the ionic strength as this reduces charge–charge interactions with the column matrix. However, it must always be borne in mind that increasing the ionic strength also has the potential to alter the interactions between the monomers of any complexes. Indeed, if the interaction is charge–charge-based then the complex may dissociate completely. Fortunately, most protein–protein interactions have a significant involvement of hydrophobic interactions, reducing the effect of changes in ionic strength. If a rigorous analysis of the effect of matrix interaction is required, then the experiment should be run at a range of ionic strengths. A plot of apparent weight *versus* ionic strength should then indicate the reliability of weight determined. If the weight is unchanged by ionic strength, then it is likely to be

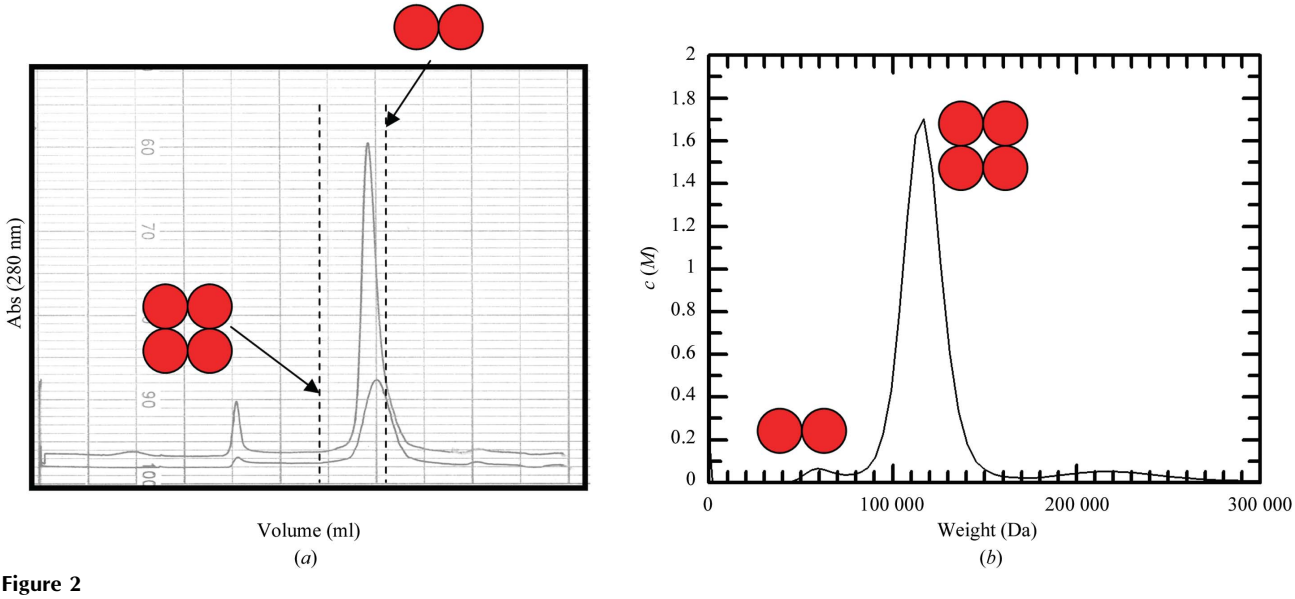

Figure 2
A comparison of data from SEC (*a*) and AUC (*b*) on the same protein. SEC provides a weight that is close to that expected for a dimer, whereas AUC shows a peak for the weight of a tetramer. It is likely that the result using SEC indicates that the protein (which is membrane-associated) interacts with the column matrix, leading to retardation and an erroneously low estimation of weight.

Novartis Exhibit 1018, Page 4 of 9

correct. If the weight increases, then it is likely that the protein is interacting with the column (or the increase in ionic strength is stabilizing a higher order association). If the weight decreases, then it is likely that the protein oligomer is held together by ionic interactions (and is unlikely to be observed in the high ionic strength solutions used for crystallography). In the two preceding cases, if the plots of apparent weight *versus* ionic strength plateau (at high ionic strength in the former and lower in the latter), then the weight value at the plateau will be closer to the correct value.

Even if interactions with the chromatographic matrix are not an issue, the experimentalist also has to take into account other issues which may lead to incorrect weights from SEC. It is common to erroneously view a protein complex as a 'solid' unchanging entity. It must be remembered that the monomers within the complex are in fact in a state of continuous exchange with free monomers in solution. This exchange rate is different for different complexes and is related to the affinity monomers have for each other in a complex. This exchange can have large effects on the observed weight as measured by SEC. Complexes where the exchange is slow compared with the time taken to perform an SEC experiment (and the monomer–monomer affinity is high) will provide a weight that is consistent with the weight of the complex. However, as the exchange increases (and the affinity drops) the apparent weight determined by the SEC begins to reduce towards that of the monomer. This can lead to an underestimation of the number of monomers in the complex. To negate this effect, SEC should be performed on a range of protein concentrations. If the exchange is slow (and the monomer–monomer affinity high) then the weight should not change considerably with concentration. However, if the exchange is fast (and the affinity low) the weight will decrease with concentration. As with the effect of ionic strength, if the plot forms a plateau at high concentrations, this weight may be taken as that of the complex.

The final issue with determination of weight by SEC is that of molecular shape. All SEC measurements are made with reference to measurements made using a 'standard' set of proteins. In most cases, these are commercial samples and are chosen to have negligible interactions with the matrix and to be close to an ideal spherical shape. Use of these references is adequate if the protein (and the complex) you are studying is also close to spherical; however, as the protein structure deviates from this idea, the apparent weight becomes less reliable. This effect can become extreme where monomers and oligomers of a protein are rod-like (Millard *et al.*, 2005). In these cases, results from SEC are usually untrustworthy.

As can be seen, SEC, although simple in concept, suffers from a number of fundamental problems when it comes to determining oligomerization states. It is not the case of a single run using SEC providing a definitive answer. Such studies should as a minimum involve a number of experiments at a range of protein concentrations. Ideally, a plot of ionic strength *versus* apparent molecular weight should also be undertaken. This is a particularly lengthy process as the set of reference proteins also has to be run at each of the ionic strengths. However, taking all these issues into consideration, SEC often provides accurate assessments of protein oligomerization and should not be discounted as a very useful technique.

The inadequacies of SEC discussed in the previous section leads a researcher to ask the question: what other methods are there? In this section, I will discuss some of the other techniques that exist for determination of solution molecular weight. Unlike SEC, the techniques described below measure the molecular weight of a protein in a solution in a sample chamber where the molecular-weight measurement is being made by an instrument or device that is able to 'interrogate' the sample. The requirement for complex instrumentation makes these techniques more costly than a simple SEC setup, but in many cases the quality and reliability of the data produced matches the cost.

To keep within the size limitations of this review, I will limit my discussion to the two techniques most commonly encountered in bioscience, dynamic light scattering (DLS) and analytical ultracentrifugation (AUC).

Dynamic light scattering (also called quasi-elastic light scattering or photon correlation spectroscopy) relies on the observation that the scattering observed from particles in a fluid fluctuates (or flickers) with time (for reviews of the experimental and theoretical details, see Schmitz, 1990; Brown, 1993; Johnson & Gabriel, 1994). This phenomenon can be observed in real life by observing the flickering caused by dust particles in a beam of sunlight. DLS uses a combination of a monochromatic laser light source and a high-speed detector to measure the scattering fluctuations in a sample solution with time. These data are then deconvoluted to produce a weight distribution. The deconvolution relies on the observation that particles in solution are constantly moving owing to random impacts with the particles that make up the fluid. Einstein and Stokes were able to show that the motion of these particles is dependent on a relatively simple relationship

$$R_{\rm H} = kT/6\pi\eta D$$

where T is temperature, k is the Boltzmann constant, D is the diffusion coefficient and η is the solution viscosity and the radius of the particle is $R_{\rm H}$. In a typical DLS experiment, T and η are known. This allows the solution molecular weight to be calculated from a measurement of D using DLS. DLS can also provide an indication as to which solution conditions will allow crystallization (Mikol *et al.*, 1990; Skouri *et al.*, 1991; Wilson, 2003).

Actually making a DLS measurement requires a few practical issues to be taken into account. In general, the sensitivity of DLS is such that at least a 0.25 mg ml^{-1} solution of a typical 50 kDa protein is required to provide a good signal. The concentration required is directly related to the weights of the protein, with lower molecular-weight molecules requiring higher concentration and higher molecular-weight molecules requiring a lower concentration. In the case of samples that have been used for crystallographic studies, this is not usually a problem. Perhaps the greatest limiting factor when it comes to using DLS is the purity of the sample. The deconvolution of

Novartis Exhibit 1018, Page 5 of 9

DLS data uses mathematical procedures that in general can only detect the presence of two or fewer species in solution. Any more than this and deconvolution of the data becomes more difficult and gaining more meaningful results less likely. Samples used for crystallography are usually of a high enough quality that this is not a problem. However, care should be taken to filter the sample before use to remove the large particulates often found in laboratories such as dust, miscellaneous fluff and hairs. We have had most success with 0.3 µm pore-size filters and this simple step can make the difference between a measurable and an unmeasurable sample. The data from the DLS usually comes in the form of a table that contains the molecular weight and radius of gyration of the species and its relative abundance in solution. One factor that has to be taken into account when using DLS is that, like SEC, it relies on the assumption that the shape of proteins approximates to a sphere. If this is not the case, then the mathematical model that is used in the calculation is incorrect. Unlike SEC, it is possible in many of the manufacturers' software packages to alter the model to take into account other shapes, *e.g.* rod, ellipse *etc.*

Analytical ultracentrifugation (AUC) is probably the 'gold standard' when it comes to determination of biomolecular oligomerization but comes at a considerable cost. However, the information gained from AUC can stand alone and in many cases an AUC study yields a plethora of other data that tell us more than just the oligomerization state. Analytical ultracentrifugation determines the solution molecular weight of particles by measuring their motion within a centrifugal field (for detailed reviews of the technical aspects, see Schuster & Toedt, 1996; Minton, 2000; Lebowitz *et al.*, 2002). The field is induced by spinning the sample and the motion of the particle is measured either by relying on the absorbance of light by chromophores within the particle or by using laser interferometry. AUC allows the motion of the particles to be examined in two ways. A sedimentation-velocity (SV) experiment measures the velocity with which particles move out from the centre of the rotor, eventually sedimenting at the bottom of the rotor. A sedimentation-equilibrium (SE) experiment is carried out at a lower speed that does not cause complete sedimentation. Instead, the particles distribute themselves as a gradient within the cells. This equilibrium state is reached when the centrifugal force is balanced by a reverse force induced by the concentration gradient within the cell.

A combination of the two experiments is extremely useful in the study of self-association as each provides subtly different information. A sedimentation-velocity (SV) experiment is a more rapid experiment than an SE experiment, taking approximately 8 h compared with days. Analysis of sedimentation velocity provides information on the size distribution of particles in a sample. The data from an SV experiment looks similar to an SEC trace, the only difference being that units for the axis of an SV distribution plot are generally quoted in terms of the sedimentation coefficient. In cases where the solution contains a relatively small number of species, this distribution plot can be transformed so that the x axis is represented in terms of weight. However, like all the previous techniques, results from SV can be distorted if the particles diverge from a spherical shape. Unlike the other techniques, SV analysis also returns an estimation of the spherical nature of the sample in terms of the frictional ratio. This ratio ranges from 1 (sphere) upwards as the particle becomes more elongated. With this in mind, it is still possible to obtain a good estimation of weight from SV and in a number of cases we have achieved results within 1% of the sequence weight (Fig. 2). In common with the other techniques discussed above, if the particle is in a complex the weight that is measured by the AUC in SV mode is determined by the exchange rate and affinity of the monomeric units for one another. Like the other techniques, this effect can be checked for by using a range of concentrations. In many cases, each AUC experiment can accommodate eight samples, allowing seven concentrations to be analysed simultaneously (the eighth sample is a reference cell). When this is combined with an absorbance-based detection system, data on a wide range of concentrations can be collected (typically 0.1–100 µ*M*).

If an associating system is suspected and a clear answer is not gained from an SV experiment, then an SE experiment is probably required. These experiments are quite lengthy and require that the protein is stable over a number of days at 278 K. As mentioned earlier in this article, an SE experiment produces a continuous concentration gradient of the particle in the sample chamber. For a non-associating system, the shape of the concentration gradient can be analysed to provide a surprisingly accurate solution weight (typically within 0.1%). For an associating system, the situation is more complex. If we consider a single AUC cell where the concentration is low (the end near the axis of the rotor), the law of mass action dictates that solution will tend to contain a higher concentration of monomeric material. Where the concentration is at its highest (the end furthest from the rotor), association is favoured, a decreased proportion of monomeric material will be found and the complex will be populated instead. The entire cell as a whole contains a continuum between and including these two extremes. These distributions can be analysed successfully to yield both the oligomeric weight (and hence the number of monomers in the oligomer) and often the equilibrium constants for the oligomerization reaction. However, analysis of this type of data is complex and requires some prior knowledge. Firstly, an accurate weight is needed for the monomer (not usually a problem if the sequence is known, but post-translational modifications can be an issue). An idea is also needed of what the order of the resulting oligomer is (dimer, trimer, tetramer *etc.*). This piece of information causes something of a dilemma, because if we knew this then we would not be doing AUC. To some extent, this logical impasse can be circumvented by analysing the data using a range of models for different possible oligomerization states. In general, one of the models will fit much better than any of the others, indicating the correct answer. It should be noted that such computational fitting approaches are often improved by increasing the amount of data available to be fitted. For SE AUC, it is

Novartis Exhibit 1018, Page 6 of 9

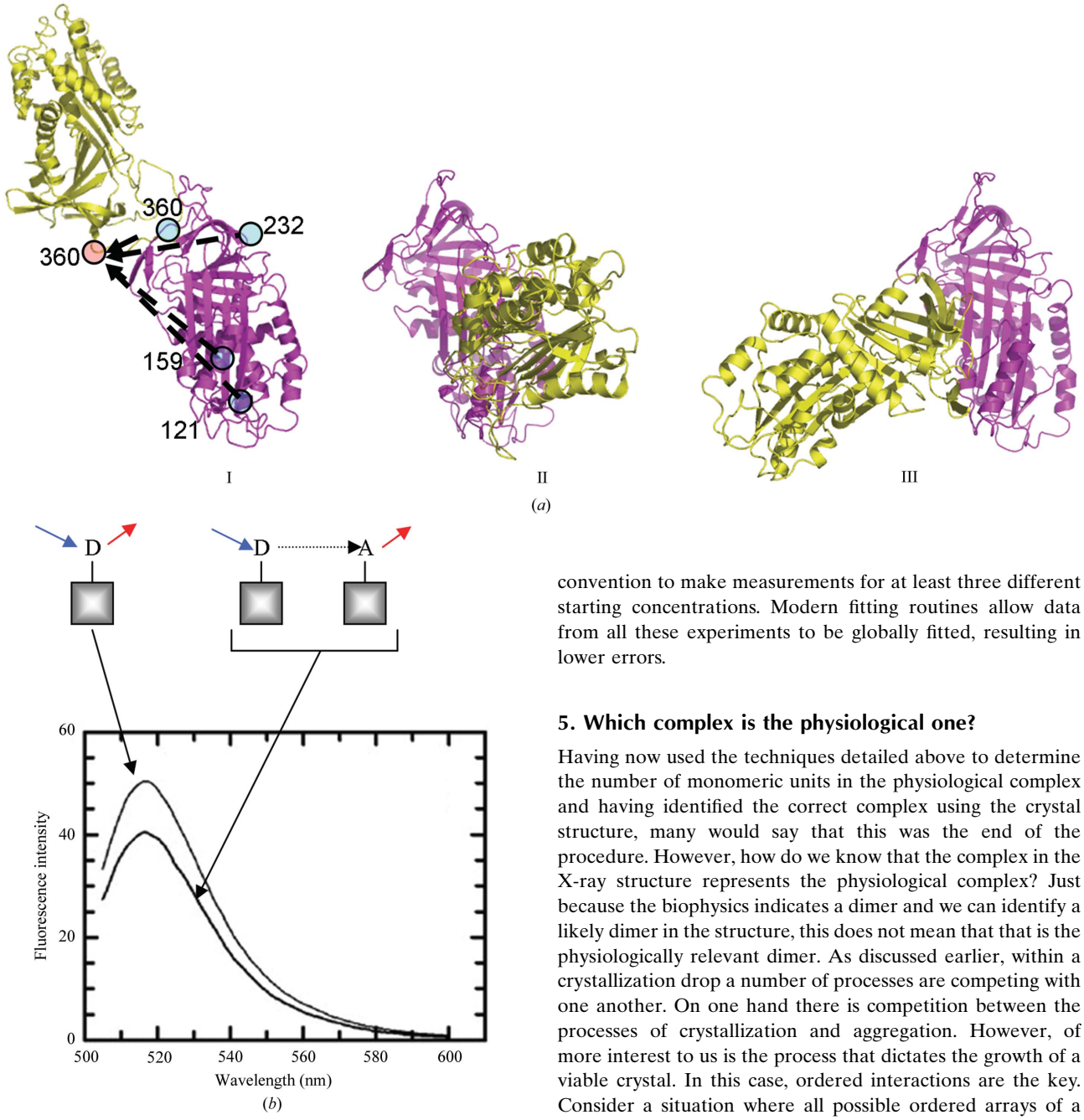

Figure 3
The use of FRET to determine the correct dimer structure for α_1-antitrypsin (Sivasothy *et al.*, 2000). (*a*) Models showing three possible dimer structures: I, II and III. The residues upon which fluorophores are attached are shown in dimer I. (*b*) The fluorescence of a donor fluorophore in the presence of an acceptor fluorophore is measured under conditions that promote and disrupt polymerization. FRET results in a decrease in the fluorescence from the donor fluorophore. A range of FRET signals are measured for proteins with fluorophores at different positions on the surface of α_1-antitrypsin. These are then used to model the structure of the dimer. The correct dimer is dimer III.

convention to make measurements for at least three different starting concentrations. Modern fitting routines allow data from all these experiments to be globally fitted, resulting in lower errors.

5. Which complex is the physiological one?

Having now used the techniques detailed above to determine the number of monomeric units in the physiological complex and having identified the correct complex using the crystal structure, many would say that this was the end of the procedure. However, how do we know that the complex in the X-ray structure represents the physiological complex? Just because the biophysics indicates a dimer and we can identify a likely dimer in the structure, this does not mean that that is the physiologically relevant dimer. As discussed earlier, within a crystallization drop a number of processes are competing with one another. On one hand there is competition between the processes of crystallization and aggregation. However, of more interest to us is the process that dictates the growth of a viable crystal. In this case, ordered interactions are the key. Consider a situation where all possible ordered arrays of a dimer produce structures that cannot propagate to form a large crystal. If, however, as we previously considered, a small proportion of monomer is present in solution (this proportion can be enhanced by solution conditions in the drop, see earlier), then in this case the monomer may associate in an ordered fashion, leading to a crystal. The important point is that this crystal does not contain the physiological complex. However, it is quite possible that in the crystal lattice contacts between monomers could lead a researcher to conclude that a dimer (the wrong dimer) does exist. So the question is: how do we know which is correct? Thankfully, there are often biochemical reasons that indicate whether a complex is the

Novartis Exhibit 1018, Page 7 of 9

correct one; for example, if the complex contains a ligand in an active site that is known to be composed of two monomers. In other cases, the structure will agree with other structures of similar proteins that have been confirmed to be physiological. Alternatively, if the interface between the two monomers is very hydrophobic, it indicates that it is likely to be unstable if exposed. In some cases, however, none of this evidence exists. In this case, we have to resort to biochemical or biophysical measurement (for a review of the use of fluorescence in such studies, see Yan & Marriott, 2003).

Unfortunately, unlike determining the number of monomers in a complex, there are no universal methods for determining which complex is the correct one. Often, examination of the structure will suggest an experiment. For example, if a protein contains a single tryptophan that is on the interface between monomers, then it would be expected that this tryptophan will change intensity and emission maximum upon complex formation. Another method is to chemically cross-link the monomers in the complex and then to determine the cross-link position using mass spectrometry. There are many other techniques; however, I would like to detail one technique that has been successfully used on a number of occasions: fluorescence resonance energy transfer (FRET).

Fluorescence resonance energy transfer is a physical phenomenon which can be used to measure the distances between points on a nanometer scale (for a detailed discussion of FRET and many other fluorescence techniques, see Lakowitz, 2006). The FRET reaction requires the use of two fluorescent probes that have overlapping spectra (the emission spectrum of one overlaps the excitation of the other; Giepmans *et al.*, 2006). When these two fluorescent probes are close to each other (typically <100 Å), FRET can occur. The FRET process involves the absorption of a photon by one of the probes: that with the lower wavelength absorbance maximum. An electron in the probe is promoted to a higher energy state, which then collapses. In the absence of the second probe, the collapse leads to the emission of a photon of lower energy (fluorescence). However, with a second fluorophore nearby, the energy can be transferred nonradiatively to the second fluorophore. An electron in the second probe is then promoted to a higher level which then collapses, giving rise to a photon with an energy (emission maximum) consistent with the properties of the second probe. The key to this process is that the efficiency of the transfer process is extremely sensitive to the distance between the two probes. Therefore, if we measure the efficiency, we can calculate the distance. If a number of these measurements can be made between monomers in a complex, then these can be combined with the information in the crystal structure to produce a structure of the complex. We have used such a system to determine the structure of an oligomer of a member of the serpin superfamily of proteins, α_1-antitrypsin (Sivasothy *et al.*, 2000).

Our study of α_1-antrypsin was initiated by the publication of three structures that showed serpin dimers (Fig. 3*a*), each with quite different monomer–monomer interfaces. Each structure seemed to have merits with regard to what was known about the physiological dimer and there seemed to be no simple biochemical test that would prove that one structure was correct in comparison with the others. To address this problem, we constructed four α_1-antrypsin mutations, introducing a single cysteine in place of a surface serine in each case (we had previously deleted the only natural cysteine in α_1-antrypsin). These mutants were purified and labelled with tetramethylrhodamine iodoacetamide (TMRIA) or fluorescein iodoacetamide (IAF). Together, these two dyes can participate in FRET, with IAF acting as a donor fluorophore and TMRIA as an acceptor (Fig. 3*b*). After labelling, we were left with eight labelled proteins: four mutants each with either of the two labels. These were used in a FRET experiment. To begin with, a donor- and an acceptor-labelled pair of proteins were mixed and the fluorescence spectra of the donor and acceptor were measured. Serpin polymerization requires incubation at elevated temperature, which meant that this experiment could act as a baseline measurement as no interaction would exist between the two species of α_1-antrypsin. The mixture was then left to polymerize for 24 h at 318 K and a second set of spectra were measured. Changes in donor and acceptor fluorescence were then calculated and used to determine the distance between the two probes. FRET data is related to distance data by

$$R = (1/E - 1)^{1/6} R_0, \qquad (1)$$

where R is the distance between the fluorophores, E is the efficiency of FRET (ranging from 1 to 0) and R_0 is the Förster radius for the fluorophore pairs used and is defined by

$$R_0 = [(8.79 \times 10^{-5})\kappa^2 n^{-4} \varphi_{\rm D} J_{\rm DA}]^{1/6}, \qquad (2)$$

where κ^2 is the orientation factor which is assumed to be 2/3 for a freely rotating fluorophore, n is the refractive index (1.4 for water) and $\varphi_{\rm D}$ is the quantum yield of the donor fluorophore. $\varphi_{\rm D}$ can be obtained by measuring the fluorescence intensity of the donor fluorophore compared with a standard solution of sodium fluorescein (10^{-6} *M* in 0.01 *M* NaOH pH 12, $\varphi = 0.79$). $J_{\rm DA}$ is the spectral overlap for the two fluorophores and is 0.5 M^{-1} cm^{-1} nm^4 for TMRIA and IAF.

The procedure was repeated with eight other pairs. The probe–probe distances could then be compared with similar distances calculated from the available crystal structures, taking into account the added length of the probe. The result of this study was a structure that showed that a surface loop on α_1-antrypsin inserted into a vacant β-strand position in a large β-sheet in a second α_1-antrypsin molecule. This structure agreed well with one of the crystal structures of α_1-antrypsin, suggesting that the other structures were the result of the crystallization process.

The approach detailed in the previous section can potentially be applied to any protein complex as long as a fluorescent probe can be inserted into the structure. I have detailed the use of cysteine-linked probes, but probes can also be attached to the protein N-terminus or to lysine residues with relative ease. It is also possible to use the presence of tryptophan residues if a single one exists in one of the pairs. Tryptophan can act as a donor, with a dansyl group acting as a

Novartis Exhibit 1018, Page 8 of 9

partner (Stratikos & Gettins, 1997, 1998; Gettins & Olson, 2004). It should be noted that I have also shown the use of fluorescent intensities for providing a measure of FRET efficiencies. However, this method can be error-prone if a control of non-interacting monomers is not available. FRET efficiencies can also be determined by measurements of the rate of fluorescence decay. This method requires more complex spectrometers that can measure fluorescence decays over a period of nanoseconds, but is becoming more popular.

6. Summary

I hope that within this review I have highlighted some of the biochemical/biophysical techniques that can be used to understand oligomerization states in X-ray crystal structures. I also hope that I have highlighted the need to undertake such studies. It is not just a case of using these techniques when the X-ray structure is in some way ambiguous. In an ideal world, the determination of the oligomerization state in solution would be carried out routinely.

References

Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J. (2004). *J. Mol. Biol.* **336**, 943–955.

Brown, W. (1993). *Dynamic Light Scattering: The Method and Some Applications.* Oxford: Clarendon Press.

Cooper, E. A. & Knutson, K. (1995). *Pharm. Biotechnol.* **7**, 101–143.

Dafforn, T. R., Mahadeva, R., Elliott, P. R., Sivasothy, P. & Lomas, D. A. (1999). *J. Biol. Chem.* **274**, 9548–9555.

Gettins, P. G. & Olson, S. T. (2004). *Methods*, **32**, 110–119.

Giepmans, B. N., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. (2006). *Science*, **312**, 217–224.

Henrick, K. & Thornton, J. M. (1998). *Trends Biochem. Sci.* **23**, 358–361.

Hewitt, C. O., Eszes, C. M., Sessions, R. B., Moreton, K. M., Dafforn, T. R., Takei, J., Dempsey, C. E., Clarke, A. R. & Holbrook, J. J. (1999). *Protein Eng.* **12**, 491–496.

Jackson, M. & Mantsch, H. H. (1995). *Crit. Rev. Biochem. Mol. Biol.* **30**, 95–120.

Janin, J., Miller, S. & Chothia, C. (1988). *J. Mol. Biol.* **204**, 155–164.

Janin, J. & Rodier, F. (1995). *Proteins*, **23**, 580–587.

Johnson, C. S. & Gabriel, D. A. (1994). *Laser Light Scattering.* New York: Dover.

Kelly, S. M. & Price, N. C. (2000). *Curr. Protein Pept. Sci.* **1**, 349–384.

Lakowitz, J. R. (2006). *Principles of Fluorescence Spectroscopy.* Berlin: Springer.

Lebowitz, J., Lewis, M. S. & Schuck, P. (2002). *Protein Sci.* **11**, 2067–2079.

Lucas, L. H., Yan, J., Larive, C. K., Zartler, E. R. & Shapiro, M. J. (2003). *Anal. Chem.* **75**, 627–634.

Mikol, V., Hirsch, E. & Giegé, R. (1990). *J. Mol. Biol.* **213**, 187–195.

Millard, T. H., Bompard, G., Heung, M. Y., Dafforn, T. R., Scott, D. J., Machesky, L. M. & Fütterer, K. (2005). *EMBO J.* **24**, 240–250.

Minton, A. P. (2000). *Exp. Mol. Med.* **32**, 1–5.

Misenheimer, T. M., Hannah, B. L., Annis, D. S. & Mosher, D. F. (2003). *Biochemistry*, **42**, 5125–5132.

Owen, D. J., Vallis, Y., Noble, M. E., Hunter, J. B., Dafforn, T. R., Evans, P. R. & McMahon, H. T. (1999). *Cell*, **97**, 805–815.

Patel, H. V., Vyas, K. A., Savtchenko, R. & Roseman, S. (2006). *J. Biol. Chem.* **281**, 17570–17578.

Perugini, M. A., Griffin, M. D., Smith, B. J., Webb, L. E., Davis, A. J., Handman, E. & Gerrard, J. A. (2005). *Eur. Biophys. J.* **34**, 469–476.

Perutz, M. F. (1954). *Proc. R. Soc. A*, **225**, 264–286.

Pusey, M. L., Liu, Z.-J., Tempel, W., Praissman, J., Lin, D., Wang, B.-C., Gavira, J. A. & Ng, J. D. (2005). *Prog. Biophys. Mol. Biol.* **88**, 359–386.

Robert, C. H. & Janin, J. (1998). *J. Mol. Biol.* **283**, 1037–1047.

Schmitz, K. (1990). *An Introduction to Dynamic Light Scattering by Macromolecules.* San Diego: Academic Press.

Schuster, T. M. & Toedt, J. M. (1996). *Curr. Opin. Struct. Biol.* **6**, 650–658.

Sivasothy, P., Dafforn, T. R., Gettins, P. G. & Lomas, D. A. (2000). *J. Biol. Chem.* **275**, 33663–33668.

Skouri, M., Delsanti, M., Munch, J. P., Lorber, B. & Giegé, R. (1991). *FEBS Lett.* **295**, 84–88.

Stratikos, E. & Gettins, P. G. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 453–458.

Stratikos, E. & Gettins, P. G. (1998). *J. Biol. Chem.* **273**, 15582–15589.

Terwilliger, T. C. *et al.* (2003). *Tuberculosis*, **83**, 223–249.

Wang, X. & Janin, J. (1993). *Acta Cryst.* D**49**, 505–512.

Wilson, W. W. (2003). *J. Struct. Biol.* **142**, 56–65.

Winzor, D. J. (2003). *J. Biochem. Biophys. Methods*, **56**, 15–52.

Yan, Y. & Marriott, G. (2003). *Curr. Opin. Chem. Biol.* **7**, 635–640.

Zartler, E. R., Yan, J., Mo, H., Kline, A. D. & Shapiro, M. J. (2003). *Curr. Top. Med. Chem.* **3**, 25–37.

Zsila, F., Bikadi, Z., Fitos, I. & Simonyi, M. (2004). *Curr. Drug Discov. Technol.* **1**, 133–153.

Novartis Exhibit 1018, Page 9 of 9