



HAL
open science

The impact of protein characterization in structural proteomics

Arie Geerlof, J. Brown, B. Coutard, M.-P. Egloff, F. Enguita, M. Fogg, R. Gilbert, M. Groves, A. Haouz, J. Nettleship, et al.

► **To cite this version:**

Arie Geerlof, J. Brown, B. Coutard, M.-P. Egloff, F. Enguita, et al.. The impact of protein characterization in structural proteomics. *Acta crystallographica Section D: Structural biology* [1993-..], International Union of Crystallography, 2006, 62 (10), pp.1125 - 1136. 10.1107/s0907444906030307 . pasteur-03525330

HAL Id: pasteur-03525330

<https://hal-pasteur.archives-ouvertes.fr/pasteur-03525330>

Submitted on 13 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution| 4.0 International License

The impact of protein characterization in structural proteomics

Arie Geerlof,^{a*} J. Brown,^b
B. Coutard,^c M.-P. Egloff,^c
F. J. Enguita,^d M. J. Fogg,^e
R. J. C. Gilbert,^b M. R. Groves,^a
A. Haouz,^f J. E. Nettleship,^b
P. Nordlund,^g R. J. Owens,^b
M. Ruff,^h S. Sainsbury,^b
D. I. Svergun^a and
Matthias Wilmanns^a

^aEMBL Hamburg Outstation, Notkestrasse 85, D-22603 Hamburg, Germany, ^bDivision of Structural Biology, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, England, ^cArchitecture et Fonction des Macromolécules Biologiques, UMR 6098, CNRS/Universités de Provence/ Université de la Méditerranée, Parc Scientifique et Technologique de Luminy, Case 932163, Avenue de Luminy 13288, Marseille CEDEX 09, France, ^dMacromolecular Crystallography Laboratory, Instituto Tecnológica Química e Biológica, Universidade Nova de Lisboa, 2781-901 Oeiras, Portugal, ^eYork Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, England, ^fUnité de Biochimie Structurale, Institut Pasteur, 25–28 Rue du Dr Roux, 75724 Paris CEDEX 15, France, ^gDepartment of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-109 51 Stockholm, Sweden, and ^hInstitut de Génétique et de Biologie Moléculaire et Cellulaire, 1 Rue Laurent Fries, BP 163, 67404 Illkirch CEDEX, France

Correspondence e-mail:
arie.geerlof@embl-hamburg.de

Protein characterization plays a role in two key aspects of structural proteomics. The first is the quality assessment of the produced protein preparations. Obtaining well diffracting crystals is one of the major bottlenecks in the structure-determination pipeline. Often, this is caused by the poor quality of the protein preparation used for crystallization trials. Hence, it is essential to perform an extensive quality assessment of the protein preparations prior to crystallization and to use the results in the evaluation of the process. Here, a protein-production and crystallization strategy is proposed with threshold values for protein purity (95%) and monodispersity (85%) below which a further optimization of the protein-production process is strongly recommended. The second aspect is the determination of protein characteristics such as domains, oligomeric state, post-translational modifications and protein–protein and protein–ligand interactions. In this paper, applications and new developments of protein-characterization methods using MS, fluorescence spectroscopy, static light scattering, analytical ultracentrifugation and small-angle X-ray scattering within the EC Structural Proteomics in Europe contract are described. Examples of the application of the various methods are given.

1. Introduction

If protein structure determination is viewed in terms of a pipeline, there are generally considered to be two major bottlenecks: protein production and crystallization (or formation of a suitable sample for NMR). The first of these is particularly severe for the laboratories involved in the Structural Proteomics In Europe (SPINE) consortium, since many of the proteins targeted for structure determination by these groups are eukaryotic and the major ‘workhorse’ expression system is *Escherichia coli*. Indeed, this is reflected in the overall statistics for the SPINE project, which show some 30% of constructs yield soluble protein (Banci *et al.*, 2006), rising to 89% for the expression of carefully chosen bacterial proteins (Alzari *et al.*, 2006). Similarly, less than 30% of the total set of SPINE targets which express as soluble protein either crystallize or are suitable for NMR analysis (Banci *et al.*, 2006), although success rates for subsets of favourable targets rise to 60% (for example, anthrax proteins; see Au *et al.*, 2006). Given this attrition rate, it is clearly vital to do everything possible to optimize the sample entered into crystallization (or NMR) screening. Empirical observations suggest that there is a correlation between the biophysical properties of a protein preparation and the probability of obtaining crystals, *e.g.* monodispersity measured by dynamic light scattering appears to correlate well with the ability to crystallize (D’Arcy, 1994;

Received 30 November 2005

Accepted 2 August 2006

Table 1

Important biophysical properties for quality assessment of a protein preparation and the most frequently used methods for their determination.

Biophysical property	Methods used for its determination
Purity	SDS-PAGE, size-exclusion chromatography, mass spectrometry
Monodispersity, aggregation state	Dynamic light scattering, size-exclusion chromatography, analytical ultracentrifugation
Conformational state	Static light scattering, analytical ultracentrifugation
Folding state	Nuclear magnetic resonance spectroscopy, circular dichroism, Fourier-transformed infrared spectroscopy
Stability	Differential scanning calorimetry, circular dichroism, thermal shift assay
Identity, modifications	Mass spectrometry

Ferré-D'Amaré & Burley, 1994, 1997). Hence, it is important to carry out quality assessment (QA) of purified proteins if initial crystal screening is unsuccessful or indeed in parallel with these trials. The most important biophysical properties and the most used methods for their determination are summarized in Table 1.

SPINE is the largest European structural proteomics consortium, with over 20 laboratories spread over the continent and in Israel. During the 3 y course of the project, the different laboratories have developed their own protocols for protein production (Alzari *et al.*, 2006; Aricescu *et al.*, 2006) and QA of the produced proteins. To investigate which QA methods are used routinely, a survey was carried among the SPINE partners. In almost all purification protocols size-exclusion chromatography (SEC) is used as a polishing step and the SEC elution profiles provide the first indication of protein purity, homogeneity and oligomeric state. In all laboratories the purity is also checked by SDS-PAGE, which yields the approximate subunit weight of the purified protein, thereby confirming the identity of the protein. In some laboratories QA ends here and the proteins are concentrated and used for crystallization trials. However, in most the proteins are subjected to other QA methods to determine additional biophysical properties. Dynamic light scattering (DLS) is routinely used by most laboratories to determine the monodispersity of the protein. Mass spectrometry (MS) can be used to check the quality of the preparation, to confirm the identity of the purified protein by accurate determination of its molecular mass and to obtain information about post-translational modifications such as phosphorylation and glycosylation. Roughly half of the laboratories routinely use MS. Methods to determine the folding state of the sample, such as circular dichroism (CD) and NMR, are less routinely employed. This is mainly a consequence of the lower availability of the necessary equipment, although this in itself presumably reflects the relatively low information content of a technique such as CD. The same is true for methods to determine the activity or functionality which often cannot be performed because of the lack of a suitable assay. Other QA methods that were mentioned in the survey but are not (or not yet) routinely used are analytical SEC, native PAGE and static light scattering (SLS). SLS is a potentially very powerful technique as discussed in §4 below. The fluorescence-based

thermal shift assay, a method to monitor protein stability, is also emerging as a QA tool of considerable value (see §3).

The development of protein characterization methods was part of the Protein Technologies workpackage 3 of the SPINE project. The tasks included, among others, the development of methods for routine QA of protein sequence composition, including post-translational modifications, by MS and the characterization of proteins by fluorescence spectroscopy, but also of less routinely used methods such as

small-angle X-ray scattering (SAXS) for protein association and shape determination. In this article, we describe the achievements within SPINE regarding the application and development of the various methods for indicating the likelihood of obtaining crystals from a given protein preparation in crystallization trials within the context of structural proteomics. Data were collected for 256 proteins from bacterial sources and used to evaluate the predictive power of biophysical properties for protein crystallizability. Such information can guide a protein production and crystallization strategy.

2. Use of mass spectrometry in structural proteomics

2.1. Background

Electrospray ionization MS (ESI-MS; Fenn *et al.*, 1989) and matrix-assisted laser-desorption ionization MS (MALDI-MS; Hillenkamp *et al.*, 1991), together with high-performance liquid chromatography (HPLC), have allowed the characterization of samples containing multiple biological and non-biological components (Bakhtiar & Tse, 2000). Single-stage MS is used to provide accurate mass measurements of proteins and peptides, whereas tandem MS gives additional information regarding the sequence of a chosen peptide.

Within structural proteomics, MS can be used for a variety of tasks such as construct verification, domain determination, crystal content verification, analysis of selenomethionine incorporation and post-translational modifications. In addition, native MS is able to probe the oligomeric state of proteins as well as protein-protein and protein-ligand interactions. Such use of MS can facilitate structure determination by X-ray diffraction or NMR (Cohen & Chait, 2001). MS is being used for quality assurance by structural genomics/proteomics consortia in the USA (Pantazatos *et al.*, 2004; Jeon *et al.*, 2005), Japan (Yokoyama *et al.*, 2000) and Europe (Jawhari *et al.*, 2004; Potier *et al.*, 2003; Nettleship *et al.*, 2005).

2.2. Protein analysis

2.2.1. QA of intact proteins. Both ESI-MS and MALDI-MS are used by several SPINE partners (Oxford, Hamburg, Marseille, Strasbourg, York) to measure the total mass of a purified protein sample (Cohen & Chait, 2001). Such

measurements can verify a construct and give information concerning any degradation products, contaminants and covalently bound partners as well as the purity of the protein sample. ESI-MS is preferred as it is accurate to ~ 1 Da for intact proteins. Several SPINE partners have developed automated methods for standard MS measurements.

For example, in Oxford, protein masses from around 8 to 100 kDa and selenomethionine incorporation in a sample are successfully measured automatically. A further application is the assessment of the level of methylation of lysine residues where the protein has undergone reductive methylation prior to crystallization. Each protein that comes through the Oxford pipeline is analysed by LC-ESI-MS to determine its accurate mass and selenomethionine incorporation (if appropriate). Around 78% of samples submitted for analysis give the expected accurate mass. In Strasbourg a high-throughput platform has been established consisting of ESI-MS combined with an orthogonal acceleration time-of-flight (TOF) mass analyzer. This platform is routinely used for protein and protein-complex characterization as well as for the screening of protein-ligand complexes (see §2.4).

2.2.2. Domain identification. Limited proteolysis combined with automated ESI-MS can be used for the identification of suitable domain boundaries or to identify structural units within a multi-domain protein (Jawhari *et al.*, 2004; Gervais *et al.*, 2004). The method has been routinely used by some SPINE partners in cases where the structure determination of the full-length protein has failed. The resultant data are used to design an optimized construct for crystallization trials.

2.2.3. Tryptic digest analysis. Proteins which are not amenable to intact protein analysis, for example if they are large and do not ionize in the mass spectrometer, can be analysed by tryptic digest followed by LC-MS or LC-MS/MS. In Oxford this technique has been used successfully for determination of selenomethionine incorporation in RNA-dependent RNA polymerase from *Neurospora crassa* (approximate molecular weight 120 kDa) produced in yeast (Laurila *et al.*, 2005). The isotope pattern of selenium was used in comparison to that of sulfur to quantify the level of selenomethionine present.

2.2.4. Glycosylated protein analysis. One of the most common post-translational modifications of protein is glycosylation (Gahmberg & Tolvanen, 1996). The presence of N- and O-linked glycans has important consequences for structural studies since they can hinder crystallization owing to their flexibility, heterogeneity and ability to obscure potential sites for crystal contacts (Butters *et al.*, 1999). Heterogeneity in glycosylation is caused by both variable site occupancy and variation in glycan-chain branching and composition. In Oxford, a method for the analysis of glycosylation-site occupancy has been developed, involving the removal of potential N-glycosylation sites by PNGase F, which converts the asparagine residue to aspartic acid. The sample is then subjected to tryptic digest followed by LC-ESI-MS analysis, which can detect the resulting 1 Da mass shift. Separation of glycosylated and non-glycosylated peptides may be performed using zwitterion chromatography-hydrophilic interaction

chromatography (ZIC-HILIC). This method is employed routinely for proteins produced *via* secretion in eukaryotic cell lines. Results can be used either to guide mutation of partially occupied glycosylation sites or to determine a strategy employing various enzymatic digests to modify the number of sugars attached to a protein.

Glycosylation-site analysis was used in Oxford to show that the human receptor-type tyrosine protein phosphatase mu expressed in HEK293 cells contains three asparagines that are glycosylated, four that are not glycosylated and two that are partially glycosylated. Similarly, the glycosylation-site analysis on the vaccinia viral replication protein A28, also produced in HEK293, cells showed that only two of the three predicted glycosylated asparagine residues are in fact glycosylated.

2.3. Analysis of crystal content

Measurement of the accurate mass of proteins present in crystal samples is very useful to determine if the crystal has grown after degradation of the protein or, in the case of cocrystallization, if all the expected proteins and/or ligands are present in the crystal (Potier *et al.*, 2000). One of the major obstacles to the use of crystal-derived protein samples for MS is the use of PEG in many crystallization conditions, which swamps the spectrum obtained. The Oxford laboratories have developed a protocol for the successful removal of PEG consisting of the extensive washing of the crystal with acetonitrile followed by dissolving it in a small amount of buffer containing 8 M urea (Nettlehip *et al.*, 2005). For example, this protocol was used to determine which form of a nitrogen regulatory protein from *Neisseria meningitidis* had crystallized (SPINE workpackage 9; see Alzari *et al.*, 2006). LC-ESI-MS had shown that the preparation contained both the phosphorylated and the non-phosphorylated protein. After crystallization, MS revealed that the crystals contained only the non-phosphorylated form (Fig. 1). In addition, a preparation of human lymphotoxin β -receptor produced *via* secretion from HEK293 cells was seen to contain both full-length and truncated protein prior to crystallization. LC-ESI-MS analysis, carried out on a cluster of crystals after sample preparation and PNGase F treatment to remove the glycans (Aricescu *et al.*, 2006), showed that only the truncated form was present in the crystals, which led to a redesign of the construct. For very small crystals, the number of protein molecules in the crystal sample (which is dependent on both crystal size and the molecular mass of the protein) may be too low for detection by ESI-MS.

Another example of the usefulness of the determination of the crystal content by MS comes from Marseille, where the expression of the dengue virus NS5 protein gave such low yields that crystallographic study was not feasible. Therefore, a 297-amino-acid N-terminal fragment was designed using bioinformatics, cloned, highly expressed in a soluble form, purified and crystallized. In the resulting crystal structure the last 33 residues of this domain were not visible in the electron-density maps (Egloff *et al.*, 2002). However, MS analysis on dissolved crystals verified the presence of the expected 297

amino acids, thus demonstrating that the last 33 residues were present but disordered. Subsequently, the full-length NS5 proteins from two other flaviviruses (tick-borne encephalitis virus, strains V61 and V67) were expressed. After purification by immobilized metal ion-affinity chromatography (IMAC), the analysis of the protein profile by SDS-PAGE revealed that the dominant species for both proteins was the result of proteolytic cleavage, the site of which was determined by

MALDI-MS to map perfectly onto the point where the dengue virus crystal structure became disordered.

2.4. Determination of non-covalent interactions by native MS

Native MS is a powerful technique for the determination of the oligomeric state of a protein and for the investigation of non-covalent interactions such as protein-protein, protein-DNA and protein-ligand interactions. For example, the Strasbourg laboratory has developed a general two-step strategy to study interactions involving nuclear receptors and ligands (antagonists and/or agonists). Firstly, the purity, homogeneity and exact mass of the protein are determined by ESI-MS under denaturing conditions. Secondly, non-covalent complexes are detected by ESI-MS using non-denaturing conditions. The latter step allows new interaction partners to be found for selected proteins (ligand screening), the determination of the binding stoichiometry of each species involved in the formation of the complex, the monitoring of changes induced by the addition of a given ligand in the association/dissociation pattern and the evaluation of the binding affinities of the ligands. This new approach has generated interesting results such as the detection and identification of fortuitous ligands for retinoic acid-related orphan receptor (ROR) and for the ultraspiracle protein (USP) (Potier *et al.*, 2003) and the characterization of all-*trans* retinoic acid as a potential ligand for the orphan nuclear receptor ROR (Stehlin-Gaon *et al.*, 2003).

2.5. Summary of MS techniques

Experience of the SPINE partners has highlighted the usefulness of MS for routine QA of proteins as discussed in §2.2, which can determine the suitability of a sample for crystallization trials. However, MS also has a key analytical role in gaining more information regarding a target protein, such as (i) investigating incorporation of selenomethionine or another modification (§2.2.1 and §2.2.3), (ii) determining whether the optimal domain has been expressed (§2.2.2), (iii) determining the occupation of potential glycosylation sites (§2.2.4), (iv) determining whether crystals contain full-length protein with the expected modifications (§2.3) and (v) determining the interaction partners of the protein (§2.4). These data can be useful in both the crystallization of a sample and in structure determination.

3. Analysis of protein stability by the thermal shift (ThermoFluor) method

3.1. Background

Fluorescence-based thermal shift (ThermoFluor) assays, using environmentally sensitive dyes to monitor protein unfolding with respect to temperature, are becoming widely used as a general biophysical method for assessing conditions that can enhance protein stability. The midpoint of the unfolding transition, as determined from analysis of the melting curve, can be taken as an approximation of the melting temperature (T_m) of a protein (Pantoliano *et al.*,

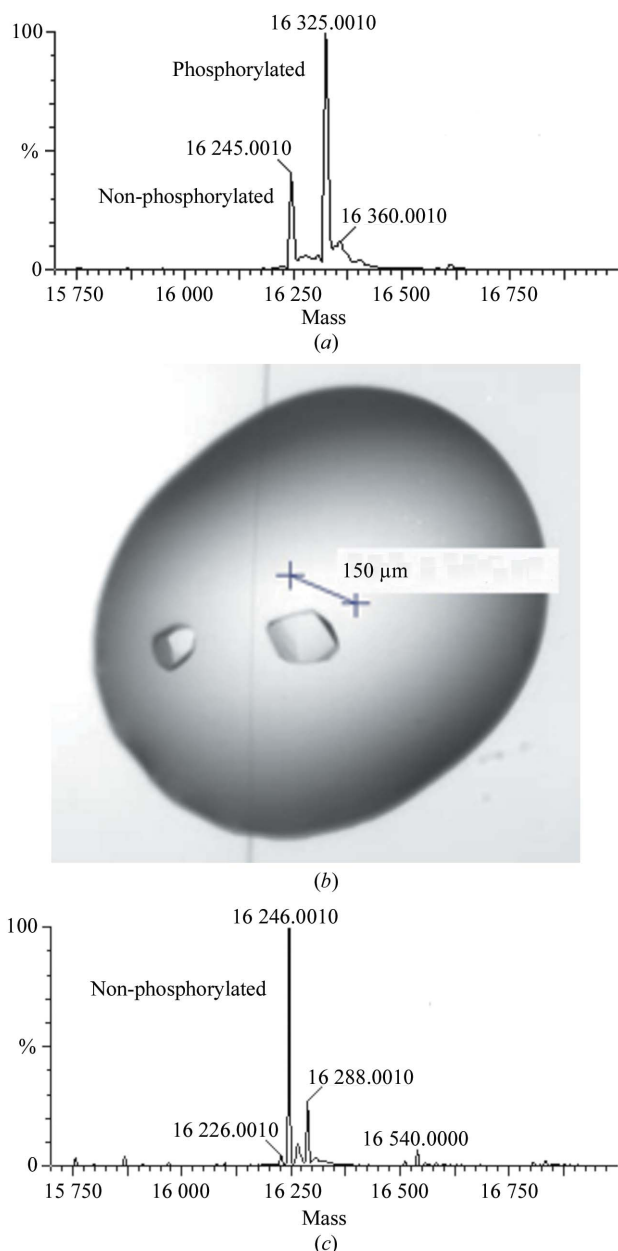


Figure 1 Analysis by MS of a nitrogen regulatory protein from *N. meningitides* before and after crystallization. (a) Mass spectrum of the purified protein prior to crystallization. The preparation contained both the phosphorylated and the non-phosphorylated protein form. (b) Image of the crystal taken for MS analysis. The protein was crystallized in 100 mM citrate pH 5 and 3.2 M ammonium sulfate. (c) Mass spectrum of the crystallized protein sample prepared using the protocol described in Nettleship *et al.* (2005). The crystal only contained the non-phosphorylated protein form.

2001). Any shift in T_m under different conditions is indicative of a change in protein stability. A positive shift in T_m indicates a stabilization of the protein by an increase in structural order and a reduction in conformational flexibility, while a negative shift in T_m indicates a destabilization. Comparison of the T_m in the presence of a particular ligand with that in its absence can give an estimate of the binding affinity of that ligand. A miniaturized version using custom-designed machinery has been described (Pantoliano *et al.*, 2001) and researchers have since adapted the technique to use commercially available real-time PCR machines, of which there are several models available (Lo *et al.*, 2004), and a number of SPINE laboratories have now implemented this technique (Oxford, Marseille, Stockholm). Such machines can perform simple thermal shift assays using small amounts of protein (1–15 μg per experiment). The fluorophore Sypro Orange (Molecular Probes) is commonly used. Supplied in DMSO, Sypro Orange is environmentally sensitive and fluoresces strongly when located in regions of low dielectric constant such as the hydrophobic regions exposed in unfolded proteins.

3.2. Methodology and case studies

The thermal shift method involves stepwise heating of a buffered solution of protein and dye, with and without ligands and other additives, from around 293 K to approximately 368 K. The use of a commercially available real-time PCR instrument allows for the HTP screening of the thermal melting curves (96 samples in 1 h). The reaction volume is typically 15–25 μl and the overall heating rate is normally about 1 K min^{-1} , with a hold step for fluorescence reading every 0.2–0.5 K. Only relatively small quantities of protein are needed (1–15 μg) per assay. As the reaction mixture is heated, the protein unfolds and there is a consequent increase in fluorescence.

El Omari and colleagues (Oxford) screened several potential cofactor ligands for the pro-apoptotic CC3 protein by thermal shift. Their finding that NADPH gave a significant T_m compared with NADH suggested that NADPH is the preferred ligand, a conclusion further supported when the only crystals obtained for CC3 grew in the presence of NADPH (El Omari *et al.*, 2005). Studies of the innate murine immune receptor dectin-1 using thermal shift confirmed that the protein regained bioactivity after refolding from *E. coli* inclusion bodies and that dectin-1 binds divalent cations, an unexpected attribute initially implied by the X-ray crystal structure of the protein (Brown *et al.*, 2006). A destabilizing effect was observed when studying mouse MICAL, a multi-domain flavoenzyme-signalling molecule involved in axon guidance (Siebold *et al.*, 2005). The melting temperature of the FAD-containing catalytic domain alone (mMICAL₄₈₉) decreased significantly with increasing NADPH concentration. This suggested that addition of NADPH lowered the stability of mMICAL₄₈₉, most likely as a consequence of a conformational change in the protein induced by reduction of the flavin ring by NADPH.

In Stockholm the ThermoFluor assay has been used to identify optimal conditions for protein crystallization (Ericsson *et al.*, 2006), the rationale being that stable proteins will crystallize more readily. To evaluate the method, the thermal melting curves of 25 proteins were measured in different buffers with or without additives. Next, the proteins were purified in optimal buffers and crystallized in the presence of a stabilizing additive. A twofold increase in the number of crystallization leads was observed compared with crystallization trials performed without these additives.

In Marseille, a combined approach including both solubility and thermostability determination was used to establish optimal conditions for subsequent protein purification and storage prior to crystallization screening. After IMAC purification, the protein was micro-dialyzed in 24 different condi-

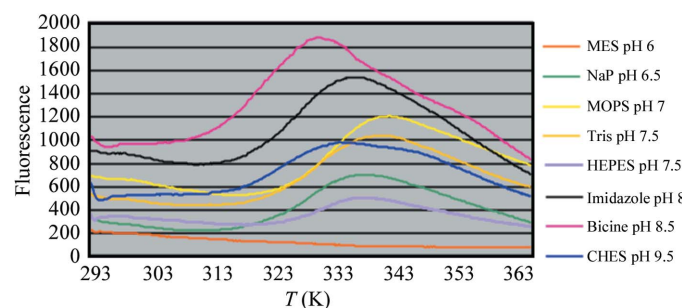


Figure 2

Thermal shift assay results for a protein measured at constant ionic strength in different buffers with varying pH values. The melting temperature (T_m) and therefore the stability of the protein decreases with increasing pH. Below pH 6 the protein precipitated completely and no fluorescence signal was recorded.

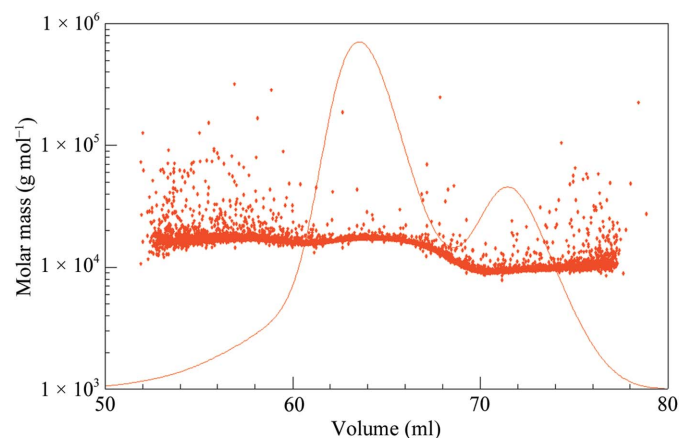


Figure 3

Analysis of the oligomeric state of a protein by the combination of size-exclusion chromatography and static light scattering. The different oligomers of a small single-domain protein (7.0 kDa) covalently attached to a ligand peptide were separated by SEC. The purification was performed on a HiLoad Superdex 75 16/60 column (GE Healthcare) in 50 mM Tris-HCl pH 8.5 containing 200 mM NaCl. The absorbance at 280 nm (solid line) was measured in series with the static light-scattering signal (dots) using a Wyatt Technologies miniDAWN. The molecular weights of the two peaks in the elution profile were calculated to be 7.3 and 13.7 kDa, respectively, which correlates well with the monomeric and dimeric protein forms.

tions, varying the pH, ionic strength and buffer composition, and its solubility was recorded by measuring the optical density at 340 nm (the optical density increases proportionally to the amount of precipitate produced; Trésaugues *et al.*, 2004; Vincentelli *et al.*, 2004). After filtration, the soluble part was subjected to analysis by thermal shift assay (Fig. 2). In most applications both assays led to convergent results. However, sometimes solubility and thermostability data diverge, particularly concerning the ionic strength of a buffer. For example, one of the proteins of interest requires a salt concentration of 100 mM or more for solubility at a given pH, whereas it is more thermostable in the absence of salt. In this case, the only condition leading to successful crystallization used a storage buffer containing 50 mM salt (B. Coutard, unpublished results).

3.3. Evaluation of the method

The experience gained in SPINE laboratories indicates that thermal shift assays are well suited to high-throughput proteomics since they offer a simple and rapid method of acquiring a large amount of information about the stability and biophysical attributes of a protein. Not all additives have a positive effect and negative shifts can be suggestive of the protein becoming destabilized. The sensitivity of commercially available real-time PCR machines means that low amounts of protein are required. A major strength of the thermal shift technique comes from comparing melting curves in one condition with those in another; for example, in different buffer conditions or in the presence or absence of ligands. Possible functions can thus be assigned to proteins where no other characterization has been performed and sequence and structural homology gives only limited clues (Carver *et al.*, 2005; Constantine *et al.*, 2005). Potential crystallization and storage conditions, cofactor and other additive preferences can be swiftly assessed, allowing subsequent experiments and trials to be optimized to maximize potential success. Furthermore, drug screening based on T_m values can identify positive hits and allow calculation of binding affinities which compare favourably with those obtained by calorimetry, a much more expensive, time-consuming and material-hungry technique (Lo *et al.*, 2004; Matulis *et al.*, 2005). Overall the technique has proven value, which is reflected in its rapid incorporation into the battery of routine characterization tools used by several SPINE partners.

4. Protein characterization by static light scattering

4.1. Background

The analysis of purified proteins by static light scattering (SLS) is a powerful technique in the quality control of samples prepared for crystallization. It is a non-invasive light-scattering method (measuring at 680 nm) that can be used in-line with standard purification equipment and does not lead to any loss of sample. The minimal mass required to produce a measurable signal is dependent upon the protein molecular weight (larger proteins scatter more light, so less is required).

A basic setup, such as the Wyatt Technologies MiniDAWN, can measure accurate signals from amounts as low as 20 µg of lysozyme (14.3 kDa). Accurate determination of *in vitro* solution oligomeric state can also aid the analysis of crystal structures, providing strong evidence that a crystal 'oligomer' is indeed a solution state.

SLS is especially powerful when used in combination with SEC. The equilibration of an SEC column in a single buffer allows the determination of accurate scattering baselines. SEC separates proteins based upon their average solution radius of gyration (K_{av}) and relies upon the use of calibration or standard curves to determine molecular mass. Thus, SEC can lead to inaccurate mass determination if the sample under analysis has shape properties that are distinct from those used in the standard curve or if the molecule under examination undergoes significant conformational rearrangement. In contrast, since the SLS data are based upon the scattering of light from proteins, they are independent of molecular conformation and dependent only upon the absolute molecular mass of the sample, resulting in mass determinations that are typically accurate to within 2–5%.

Besides accurate mass determination, SLS can be used to find solution conditions that are favourable for crystallization, the so-called 'crystallization slot'. This is performed by the determination of the second virial coefficient (B), which can be measured directly by SLS (Wilson, 2003). It is observed that crystals form more readily if the B value lies within a certain narrow range, while outside this range the proteins have a higher tendency to precipitate or remain soluble.

4.2. Conformational state analysis of a protein by SLS: case study

The following example is based upon the purification of a single protein domain (7 kDa) performed in Hamburg. A construct was generated in which the ligand peptide is covalently attached to the C-terminus of the domain under examination. The protein was purified by immobilized metal-affinity chromatography (IMAC) prior to application of the sample to SEC. Examination of the elution profile clearly indicates that two distinct species are identifiable (Fig. 3). However, comparison of the elution peaks with a standard curve results in estimated molecular weights of the proteins that do not correspond well to defined oligomeric states. In contrast, analysis by SLS clearly demonstrates that the measured weights of the two peaks correspond well to monomeric and dimeric forms of the protein (7.3 and 13.7 kDa, respectively). SLS data measured from constructs that do not contain the ligand peptide show no such dimer formation (data not shown). Thus, these data indicate that the presence of the ligand peptide at the C-terminus of the construct is sufficient to promote dimerization of the protein. Further, the weights of the purified protein are experimentally measured to an accuracy of better than 2% and the *in vitro* solution oligomeric state is unambiguously defined.

Analysis of the SLS data can also be used to define more properly which regions of the elution profile are best

combined for crystallization purposes. In the example given, it can be clearly seen that in both monomer and dimer peaks there are regions of flat (horizontal) and sloped weight values. As SLS measures the average molecular weight of the sample, the use of SLS data allows selection of elution fractions that contain only a single species, rather than a mixture of monomer and dimer. Conversely, regions of the elution profile corresponding to the 'sloped' weight values contain both oligomeric states. The separation of proteins into homogeneous weight states is often an important step in producing protein of sufficient quality for crystallization.

4.3. Limitations of SLS

As SLS data are dependent upon the scattered light at a specific wavelength, the absorption properties of the protein are critical. If the sample under examination shows significant absorption at the wavelength used for analysis (680 nm), a significant error will be introduced. While SLS is highly sensitive, the scattering signal is also dependent upon the amounts of each sample available. Thus, a highly dilute sample will not produce high-quality data. Additionally, the accurate determination of molecular weight from SLS data is absolutely dependent on the purity of the sample under analysis. Should the elution from a separation media contain more than a single weight species, the resultant scattering will be from the 'weight average' of the sample. If SEC is used as a final polishing step in the purification of proteins for crystallization, the samples are typically of sufficient purity for accurate weight measurement using SLS, whereas samples from an initial capture step such as IMAC will often be of insufficient purity. Finally, as is the case for all scattering techniques, the presence of large particles must be avoided as these will typically scatter so strongly that they mask the signals from the sample. These particles are often generated in small amounts by the pumps of modern chromatography systems and are not filtered out by the separation media. Thus, it is advisable to use an in-line filter (100–200 μm) to remove such particles.

5. Analysis of protein conformation by analytical ultracentrifugation

5.1. Background

Analytical ultracentrifugation (AUC) is a very useful accessory technique in the QA of samples for crystallization, validation of the oligomerization states captured in crystals and interrogation of protein function, both in the equilibrium and velocity modes of operation. In the equilibrium mode, the molecular weight of proteins and other species in solution can be determined directly and in a manner unaffected by sample shape considerations. In the velocity mode, the sedimentation coefficient of a species can be determined, which is sensitive to molecular shape in a way that can be modelled and which is highly sensitive to polydispersity. Modern analytical ultracentrifuges have absorbance and interference optics, allowing the design of experiments making use of differential absorbances displayed by different classes of molecules or

chromophoric labels and the analysis of concentration ranges from 0.1 to 100 mg ml^{-1} and more.

5.2. Case studies

The dimerization of proteins such as B7-1, Mona/Gads and SARS coronavirus (CoV) nsp9 within their crystal forms has been shown using AUC to be relevant to their molecular function (Oxford). In the case of B7-1, the crystallographic dimer was shown to form in solution (Ikemizu *et al.*, 2000). This permitted the development of models of interaction between B7-1 and CTLA-4 involving avidity enhancement of cell–cell interaction through the formation of extended zipper-like regions of attachment (Ikemizu *et al.*, 2000; Stamper *et al.*, 2001). AUC was used to show in addition that the homologous B7-2 is instead monomeric, thus being weakly stimulatory of a T-cell response whereas B7-1 is strongly inhibitory (Collins *et al.*, 2002). In the case of Mona/Gads, dimerization of this intracellular signal-mediating module was shown to be *via* mutual chelation of an unidentified ion, which AUC demonstrated to be zinc (Harkiolaki *et al.*, 2003). Finally, the dimer of SARS-CoV nsp9 was validated by AUC (Sutton *et al.*, 2004).

More complex mixtures of aggregation states in proteins can be easily observed using velocity approaches, particularly if the $g(S)$, $c(S)$ or $c(M)$ methods of data analysis are employed (Schuck, 2003). These analytical tools are model-independent and provide a profile of the differently sedimenting species found in solution. The multi-levelled assembly of large complexes can be analysed in this manner (Solovyova *et al.*, 2004), as well as the effects proteins have on each other's behaviour even if a stable hetero-interaction cannot be identified. Thus, nsp9 of SARS-CoV showed both monomer and dimer peaks in a $g(S)$ analysis of velocity data (as it had shown dimeric protein by equilibrium methods, discussed above) and nsp8 displayed a complex and poly-disperse profile of interaction (Sutton *et al.*, 2004). The combination of nsp9 and nsp8, however, produced a marked change in the sedimentation behaviour of the proteins and thus suggested an effect on each other's oligomerization state even though a stable hetero-oligomer could not be identified.

Sedimentation coefficient values derived from velocity profiles of proteins can also be used to support certain models of the overall shape of a molecule or complex over others, as for example with CD8 (Merry *et al.*, 2003) and ribosomes (RJCG, unpublished results). More ordinarily, sedimentation-velocity analysis is a highly sensitive way to demonstrate monodispersity in a protein preparation.

5.3. Evaluation of the method

The key strength of AUC for our strategies has proved to be the contributions it has made to our understanding the functions of proteins for which we have solved structures and in confirming that contacts observed within oligomeric assemblies in crystals also occur in solution. A special benefit of AUC is the way in which very wide ranges of concentration can be analysed (0.1–100 mg ml^{-1} ; Mavaddat *et al.*, 2000; Shapiro *et al.*, 1996), easily bracketing those encountered in

crystallography and NMR experiments and in other structural techniques such as cryo-EM. Additionally, this concentration range also includes those used in other biophysical approaches reviewed here, such as light scattering, mass spectrometry and small-angle scattering. AUC is therefore an especially complementary technique. By the use of a combination of sedimentation-equilibrium experiments (for systems of relatively well characterized assembly state, such as monomer–dimer equilibria) and sedimentation-velocity experiments (where the samples are more heterogeneous) with modelling of sedimentation coefficients from crystal structures, we have been able to make use of AUC to draw together functional and structural insights.

6. Assessment of protein association and shape determination by small-angle X-ray scattering

6.1. Background

Small-angle X-ray scattering (SAXS) allows the study of the low-resolution structure of native particles in solution under nearly physiological conditions (Feigin & Svergun, 1987). The potential of SAXS has been boosted by the advent of high-brilliance synchrotron-radiation sources accompanied by the development of novel data-analysis methods (Svergun & Koch, 2002, 2003) which significantly improved the resolution and reliability of the models constructed. SAXS offers several options for the structural characterization of biological macromolecules in solution.

(i) If no structural information about the macromolecule is available, its shape can be reconstructed *ab initio* at about 1–2 nm resolution (Svergun, 1999; Svergun *et al.*, 2001).

(ii) If high-resolution models of subunits or domains in a complex are available, the quaternary structure can be determined by rigid-body modelling against the SAXS data (Konarev *et al.*, 2001; Petoukhov & Svergun, 2005). This option is especially useful for the analysis of macromolecular complexes.

(iii) If high-resolution methods yield an incomplete structure of the macromolecule, overall configurations of missing (flexible) loops and domains in solution can be reconstructed using the SAXS data (Petoukhov *et al.*, 2002; Marquez *et al.*, 2003).

(iv) SAXS can rapidly validate in solution models for oligomeric states proposed on the basis of crystal-packing contacts (Svergun *et al.*, 2000; Morth *et al.*, 2004).

(v) SAXS allows the quantification of the components of equilibrium and non-equilibrium mixtures [*e.g.* oligomerization, (dis)assembly or (un)folding processes] and changes in time or upon changing the conditions (Sokolova *et al.*, 2001; Kozielski *et al.*, 2001; van den Heuvel *et al.*, 2003).

A synchrotron SAXS experiment typically requires a few milligrams of protein at sub-millimolar concentrations and measurement times of seconds to minutes. Monodisperse solutions are required for *ab initio* or rigid-body analysis. Most notably, SAXS has practically no size limitations, allowing individual macromolecules of a few kDa and macromolecular

machines of dozens of MDa to be studied. The technique is easily coupled with other structural, biophysical and biochemical methods. In particular, information on inter-residue contacts from mutagenesis studies or distances from FRET is readily included to constrain rigid-body refinement of complexes (Petoukhov & Svergun, 2005). Joint use of SAXS

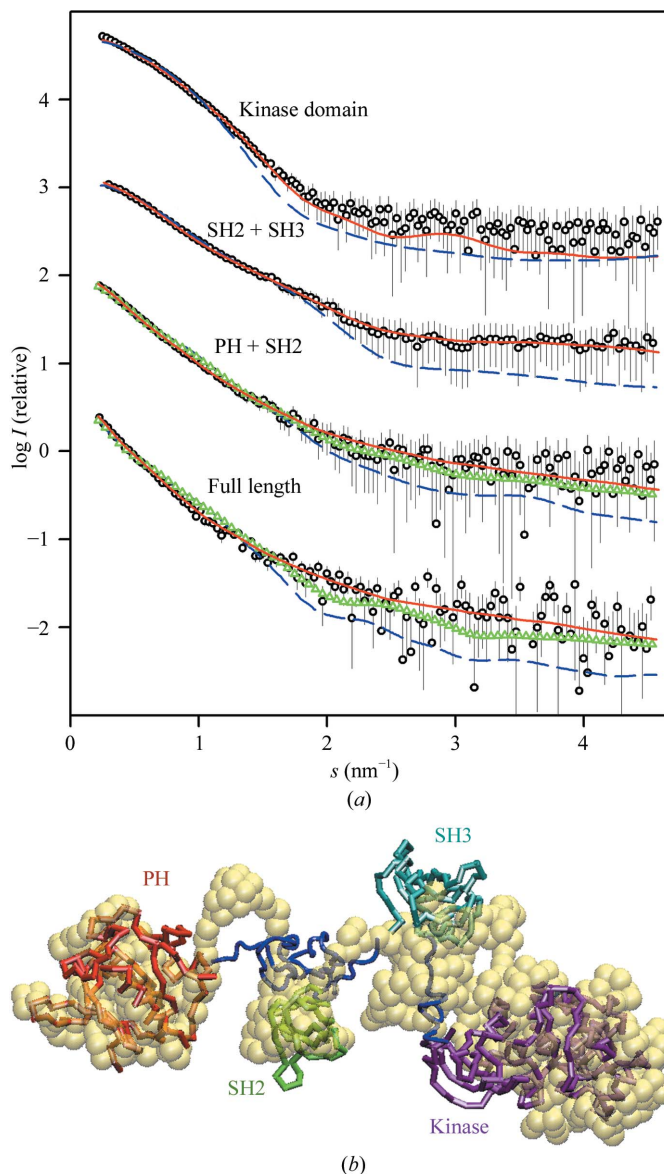


Figure 4 X-ray scattering patterns and the model of Btk. (a) Experimental scattering from the Btk constructs (dots with error bars), computed scattering from *ab initio* models (red solid lines) and scattering from the models obtained by rigid-body refinement (blue dashed lines; for the kinase domain, computed scattering from its atomic structure). For PH-SH2 and full-length protein, scattering from models is displayed as green triangles. The plot displays the logarithm of the scattering intensity as a function of momentum transfer $s = 4\pi\sin(\theta)/\lambda$, where θ is the scattering angle and λ is the X-ray wavelength. The scattering patterns are appropriately displaced in the logarithmic scale for better visualization. (b) Superposition of a typical *ab initio* low-resolution model of the full-length Btk construct (semitransparent beads) with the Btk model obtained by rigid-body refinement of the positions of individual domains (colour-coded C^α chains). Typical conformations of interdomain linkers constructed by the program CREDO are displayed as blue C^α chains.

Table 2

Summary of a quality-assessment survey carried out within the SPINE laboratories.

Data were collected on the quality assessment of the preparations of 256 bacterial target proteins expressed in *E. coli*. The set contained proteins from *M. tuberculosis*, *B. anthracis*, *C. jejuni*, *N. meningitides* and *N. gonorrhoea*. The purity of the preparations was determined by SDS-PAGE, the monodispersity by dynamic light scattering, the folding state by circular dichroism, the foldedness by NMR and the identity by MS. The threshold values were determined after the evaluation of the quality-assessment data from this survey. These were compared with the number of protein preparations that yielded crystals in crystallization trials.

Protein property	Threshold value	Proteins	Crystals	%
Purity	>95%	229	134	59
	<95%	27	10	37
Monodispersity	>85%	94	58	62
	<85%	24	4	17
Folding state	Not determined	138	82	59
	<60% random coil	18	13	72
	>60% random coil	4	2	50
Folded	Not determined	234	129	55
	Yes	1	1	100
Identity	Not determined	255	143	56
	Determined	102	65	64
Total	Not determined	154	79	51
		256	144	56

and NMR data allows the construction of more reliable models of multi-domain and multi-subunit proteins (Mattinen *et al.*, 2002; Grishaev *et al.*, 2005; Petoukhov & Svergun, 2005). In many cases, functional questions can be addressed directly, as SAXS data can be correlated with results such as kinetics, spectroscopy or interaction studies in solution under similar experimental conditions (protein concentration, temperature, buffer composition, presence of ligands). In recent years, biological SAXS has experienced a real resurgence (reviewed by Svergun & Koch, 2002, 2003; Vachette *et al.*, 2003; Koch *et al.*, 2003).

6.2. Case study

An example of a SPINE collaboration between Hamburg and Grenoble using SAXS is provided by a study of the domain structure of Bruton's tyrosine kinase (Btk; Marquez *et al.*, 2003). Btk is a non-receptor protein tyrosine kinase (nrPTK) essential for the development of B lymphocytes in humans and mice. Like Src and Abl PTKs, Btk contains a conserved cassette formed by SH3, SH2 and protein kinase domains, but differs by the presence of an N-terminal PH domain and the Tec homology region. SAXS data were collected independently from the full-length protein and from several deletion mutants (Fig. 4a). The programs *DAMMIN* (Svergun, 1999) and *GASBOR* (Svergun *et al.*, 2001) were employed to determine low-resolution shapes of these constructs *ab initio* from the scattering data and these *ab initio* models indicated a linear arrangement of domains. This arrangement was further confirmed by interactive rigid-body modelling using known high-resolution structures of individual domains and the probable conformations of missing linkers between the domains were subsequently constructed. The final model of Btk (Fig. 4b) displays an extended

conformation with little or no interdomain interactions. In agreement with this model, deletion of non-catalytic domains failed to enhance the activity of Btk. These results indicate that, in contrast to Src and Abl, Btk might not require an assembled conformation for the regulation of its activity. The rigid-body model was constructed by a multi-step procedure involving interactive modelling with the program *MASSHA* (Konarev *et al.*, 2001) with missing linkers added by the program *CREDO* (Petoukhov *et al.*, 2002). The recently developed global modelling programs *SASREF* and *BUNCH* (Petoukhov & Svergun, 2005) allow the construction of models automatically (also by fitting simultaneously multiple scattering curves) and gave results similar to those in Fig. 4 with little user intervention.

7. Correlation between biophysical properties and successful protein crystallization

7.1. Introduction

Studies on individual proteins have suggested that there is a correlation between the biophysical properties of a protein preparation, such as purity, homogeneity, solubility and stability, and the probability of that protein forming well diffracting crystals. For instance, proteins that are monodisperse in undersaturated solutions crystallize more readily than proteins that are (partially) aggregated or polydisperse (D'Arcy, 1994; Ferré-D'Amaré & Burley, 1994, 1997). In the largest DLS survey carried out so far (D'Arcy, 1994), 44 proteins gave narrow unimodal distributions, of which 34 (77%) crystallized and 60% of the ten that gave broad unimodal distributions crystallized. In contrast, 12 proteins had multimodal (or polydisperse) distributions and only one of these crystallized (8%). However, few systematic studies have been published on the correlation between other biophysical properties and protein crystallizability (however, see §3.2 above).

7.2. Quality-assessment survey

To investigate the effect of the biophysical properties of purified proteins on their crystallizability, a survey was undertaken within SPINE. Partner laboratories supplied data on the purity, homogeneity, monodispersity, folding state and identity of these proteins and whether crystals were obtained in crystallization trials set up with these preparations. For practical reasons, the survey was limited to bacterial targets. Data were collected for 256 proteins from different organisms such as *Mycobacterium tuberculosis*, *Bacillus anthracis*, *Neisseria meningitides*, *N. gonorrhoea* and *Campylobacter jejuni*. The results are summarized in Table 2.

The purity of all protein preparations was checked by SDS-PAGE. The vast majority of the proteins (229) had a purity of more than 95% and more than half of these yielded crystals in the crystallization screens used. Of the proteins with a lower purity, crystals were obtained in little more than one third of the cases, indicating that protein purity has a significant effect on crystallizability. The purification protocols of all proteins in

this survey contained SEC as the final polishing step. This was included to remove aggregates and to obtain homogeneous preparations. Next, the proteins were highly concentrated for crystallization. During this process the conformational state of a protein may change and new aggregates may form. Therefore, it is recommended to check the conformational/aggregation state of a protein by DLS, for example, before crystallization trials are set up. This was performed for approximately half of the proteins and the results were compared with the number of crystals obtained. Initially, a cutoff of 90% monodispersity was chosen, but the data showed a stronger correlation when this was changed to 85%. Most of the 94 tested proteins had a monodispersity of more than 85% and crystals were obtained for 58 of these (62%). Of the 24 proteins that were less than 85% monodisperse, only four yielded crystals (17%). This confirms earlier reports (see §7.1) that there is a strong correlation between the monodispersity of a protein preparation and the probability of obtaining crystals. In the laboratories that contributed data to this survey, methods for the determination of the folding state such as CD and NMR are not routinely used. Hence, insufficient data were available to allow for a proper correlation between the folding state of proteins and the crystallizability. Nevertheless, if a protein is not properly folded, has large unstructured regions or is even naturally unfolded, the likelihood that it will crystallize is small. Therefore, it is noteworthy that two proteins that showed more than 60% random coil in their CD spectra still yielded crystals. MS was performed on 102 of the 256 proteins to check the identity and to determine the quality of the protein preparation. Crystals were obtained with 65 of these (64%), which was only marginally better than was achieved with the proteins that were not tested (51%). This result indicates that there is no clear correlation between MS data and protein crystallizability.

7.3. The role of quality assessment in the protein-production and crystallization pipeline

The introduction of automated high-throughput production of proteins has led to the development of standardized protocols (Lesley, 2001; Stevens, 2000). These usually consist of an IMAC step, a protease-cleavage step to remove tags and a SEC polishing step and are performed in a standard buffer system. A major drawback of this approach is that it excludes the possibility of optimizing the purification and storage conditions for each individual protein, which may lead to protein preparations of insufficient quality for crystallization. Hence, it becomes necessary to establish a protein-production and crystallization strategy in which protein-quality assessment plays an important role. The strategy shown in Fig. 5 is based on our experiences within SPINE and the results of the QA survey (Table 2).

The first serious bottleneck in the protein-production pipeline is to obtain sufficient amounts of soluble protein. If all attempts to improve the soluble expression have failed (see Alzari *et al.*, 2006; Aricescu *et al.*, 2006), the next step should

be to optimize the gene constructs by using bioinformatics tools for the identification of individual domains and regions of low complexity (Albeck *et al.*, 2006; Esnouf *et al.*, 2006). This information can then be used to design new more stable expression constructs. Experimentally, these constructs can be identified by performing limited proteolysis on the full-length protein or by exopeptidase treatment (see §2.4 and MRG, unpublished results) followed by MS analysis (see §2.4). However, once a potential route to producing soluble protein has been established, the next major bottleneck is to obtain protein of sufficient quality for crystallization.

Our QA survey shows a clear correlation between the biophysical properties of a protein preparation and the probability of obtaining well diffracting crystals. Therefore, every purified protein that comes through the protein-production pipeline should undergo QA and the outcome of this assessment should play a decisive role in determining the next step. In addition, the database accumulated through recording QA data will provide an increasingly valuable resource for future data-mining exercises.

Protein purity and monodispersity are relatively easily determined. Both show a strong correlation with protein crystallizability and threshold values have been determined (Table 2). A protein purity of more than 95% is almost always

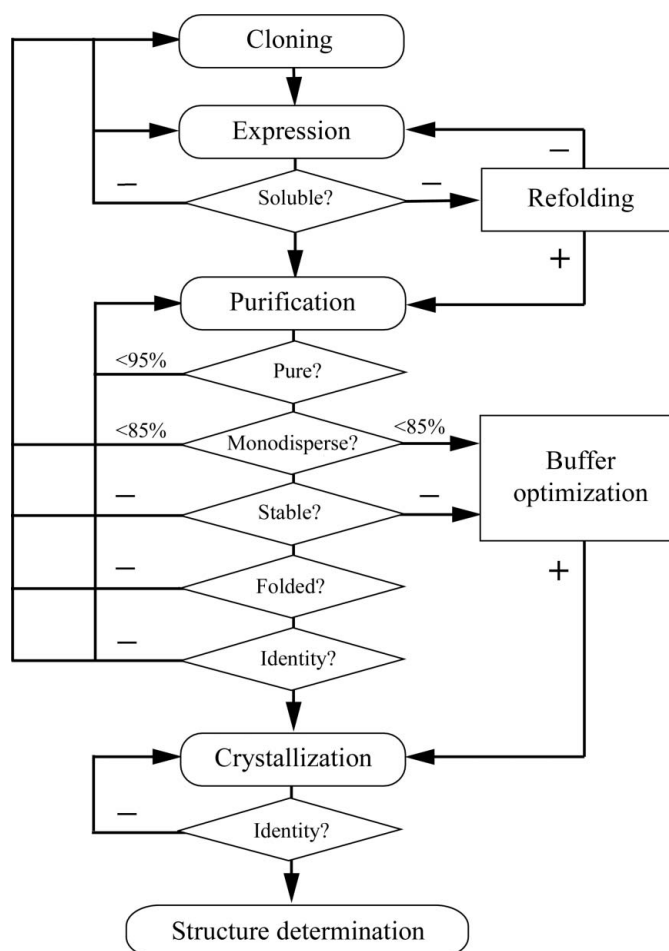


Figure 5 Protein-production and crystallization strategy.

obtained with the purification protocols used. In the exceptional case that the purity is below this value the purification process could be adapted; for instance, by the addition of other chromatographic steps. An even stronger correlation was found with the aggregation state of the protein. To have a good chance of obtaining crystals, the monodispersity of the protein preparation should be more than 85%. Below this threshold value it is advisable to improve the quality of the protein preparations by optimizing the purification protocol and/or the buffer composition. For the latter, several methods are available including solubility screening with or without DLS analysis (Jancarik *et al.*, 2004; Collins *et al.*, 2004) and thermal shift (ThermoFluor) assay (Ericsson *et al.*, 2006). The result is a greatly increased chance of obtaining crystals.

For other protein properties such as stability and folding state, insufficient data were collected in our QA survey to determine threshold values. However, CD or NMR experiments could be used to identify optimal expression and purification conditions for protein folding. The stability of a protein can be improved by optimizing the buffer composition and/or by the addition of a stabilizing ligand (§3.2). If all of these approaches fail to improve the properties of a protein preparation, what remains is to optimize the gene constructs (see above) and check the identity of a protein by MS. In the case of protein–protein or protein–ligand complexes, it is particularly important to know if all components of the complex are present in the preparation before continuing with the crystallization trials. After crystals are obtained, MS can be used to check if they contain the intact or partially degraded protein or if all the proteins and/or ligands are present in the case of cocrystallization.

7.4. Evaluation of the strategy

The data obtained in our QA survey have shown a strong correlation between certain biophysical parameters (especially purity and monodispersity) and protein crystallizability. For other parameters, no clear correlation was found or insufficient data were collected to draw any conclusions. These results suggest a protein-production and crystallization strategy (Fig. 5) which can be applied to provide protein preparations with sufficiently high quality for crystallization. As described in §7.3, this can be achieved by incorporating QA in the protein-production pipeline and directly using the results in the determination of the next step in the process. On the other hand, in many cases it will be more practical to use QA to improve protein preparations in cases where initial crystallization trials have failed. This is especially true for proteins that are of high value or are particularly difficult to produce.

This work was funded by the European Commission as SPINE (Structural Proteomics In Europe) Contract No. QL2-CT-2002-00988 under the RTD programme 'Quality of Life and Management of Living Resources'. AH is supported by a grant from RNG (Réseau National des Génopoles).

References

- Albeck, S. *et al.* (2006). *Acta Cryst.* **D62**, 1184–1195.
- Alzari, P. M. *et al.* (2006). *Acta Cryst.* **D62**, 1103–1113.
- Aricescu, A. R. *et al.* (2006). *Acta Cryst.* **D62**, 1114–112.
- Au, K. *et al.* (2006). *Acta Cryst.* **D62**, 1267–1275.
- Bakhtiar, R. & Tse, F. L. (2000). *Mutagenesis*, **15**, 415–430.
- Banci, L. *et al.* (2006). *Acta Cryst.* **D62**, 1208–1217.
- Brown, J., O'Callaghan, C. A., Marshall, A. S. J., Gilbert, R. J., Siebold, C., Gordon, S., Brown, G. D. & Jones, E. Y. (2006). Submitted.
- Butters, T. D., Sparks, L. M., Harlos, K., Ikemizu, S., Stuart, D. I., Jones, E. Y. & Davis, S. J. (1999). *Protein Sci.* **8**, 1696–1701.
- Carver, T. E. *et al.* (2005). *J. Biol. Chem.* **280**, 11704–11712.
- Cohen, S. L. & Chait, B. T. (2001). *Annu. Rev. Biophys. Biomol. Struct.* **30**, 67–85.
- Collins, A. V., Brodie, D. W., Gilbert, R. J., Iaboni, A., Manso-Sancho, R., Walse, B., Stuart, D. I., van der Merwe, P. A. & Davis, S. J. (2002). *Immunity*, **17**, 201–210.
- Collins, B. K., Tomanicek, S. J., Lyamicheva, N., Kaiser, M. W. & Mueser, T. C. (2004). *Acta Cryst.* **D60**, 1674–1678.
- Constantine, K. L., Krystek, S. R., Healy, M. D., Doyle, M. L., Siemers, N. O., Thanassi, J., Yan, N., Xie, D., Goldfarb, V., Yanchunas, J., Tao, L., Dougherty, B. A. & Farmer, B. T. II (2005). *Protein Sci.* **14**, 1472–1484.
- D'Arcy, A. (1994). *Acta Cryst.* **D50**, 469–471.
- Egloff, M. P., Benarroch, D., Selisko, B., Romette, J. L. & Canard, B. (2002). *EMBO J.* **21**, 2757–2768.
- El Omari, K., Bird, L. E., Nichols, C. E., Ren, J. & Stammers, D. K. (2005). *J. Biol. Chem.* **280**, 18229–18236.
- Ericsson, U., Hallberg, M., DeTitta, G., Dekker, N. & Nordlund, P. (2006). Submitted.
- Esnouf, R. M., Hamer, R., Sussman, J. L., Silman, I., Trudgian, D., Yang, Z.-R. & Prilusky, J. (2006). *Acta Cryst.* **D62**, 1260–1266.
- Feigin, L. A. & Svergun, D. I. (1987). *Structure Analysis by Small-Angle X-ray and Neutron Scattering*. New York: Plenum Press.
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. (1989). *Science*, **246**, 64–71.
- Ferré-D'Amaré, A. R. & Burley, S. K. (1994). *Structure*, **2**, 357–359.
- Ferré-D'Amaré, A. R. & Burley, S. K. (1997). *Methods Enzymol.* **276**, 157–166.
- Gahmberg, C. G. & Tolvanen, M. (1996). *Trends Biochem. Sci.* **21**, 308–311.
- Gervais, V., Lamour, V., Jawhari, A., Frindel, F., Wasielewski, E., Dubaele, S., Egly, J. M., Thierry, J. C., Kieffer, B. & Poterszman, A. (2004). *Nature Struct. Mol. Biol.* **11**, 616–622.
- Grishaev, A., Wu, J., Trehwella, J. & Bax, A. (2005). *J. Am. Chem. Soc.* **127**, 16621–16628.
- Harkiolaki, M., Lewitzky, M., Gilbert, R. J., Jones, E. Y., Bourette, R. P., Mouchiroud, G., Sondermann, H., Moarefi, I. & Feller, S. M. (2003). *EMBO J.* **22**, 2571–2582.
- Heuvel, R. H. van den, Svergun, D. I., Petoukhov, M. V., Coda, A., Curti, B., Ravasio, S., Vanoni, M. A. & Mattevi, A. (2003). *J. Mol. Biol.* **330**, 113–128.
- Hillenkamp, F., Karas, M., Beavis, R. C. & Chait, B. T. (1991). *Anal. Chem.* **63**, 1193A–1203A.
- Ikemizu, S., Gilbert, R. J., Fennelly, J. A., Collins, A. V., Harlos, K., Jones, E. Y., Stuart, D. I. & Davis, S. J. (2000). *Immunity*, **12**, 51–60.
- Jancarik, J., Pufan, R., Hong, C., Kim, S.-H. & Kim, R. (2004). *Acta Cryst.* **D60**, 1670–1673.
- Jawhari, A., Boussert, S., Lamour, V., Atkinson, R. A., Kieffer, B., Poch, O., Potier, N., van Dorselaer, A., Moras, D. & Poterszman, A. (2004). *Biochemistry*, **43**, 14420–14430.
- Jeon, W. B., Aceti, D. J., Bingman, C. A., Vojtik, F. C., Olson, A. C., Ellefson, J. M., McCombs, J. E., Sreenath, H. K., Blommel, P. G., Seder, K. D., Burns, B. T., Geetha, H. V., Harms, A. C., Sabat, G., Sussman, M. R., Fox, B. G. & Phillips, G. N. Jr (2005). *J. Struct. Funct. Genomics*, **6**, 143–147.

- Koch, M. H. J., Vachette, P. & Svergun, D. I. (2003). *Quart. Rev. Biophys.* **36**, 147–227.
- Konarev, P. V., Petoukhov, M. V. & Svergun, D. I. (2001). *J. Appl. Cryst.* **34**, 527–532.
- Kozielski, F., Svergun, D., Zaccari, G., Wade, R. H. & Koch, M. H. J. (2001). *J. Biol. Chem.* **276**, 1267–1275.
- Laurila, M. R., Salgado, P. S., Makeyev, E. V., Nettleship, J., Stuart, D. I., Grimes, J. M. & Bamford, D. H. (2005). *J. Struct. Biol.* **149**, 111–115.
- Lesley, S. A. (2001). *Protein Expr. Purif.* **22**, 159–164.
- Lo, M. C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M. & Ellestad, G. (2004). *Anal. Biochem.* **332**, 153–159.
- Marquez, J. A., Smith, C. I., Petoukhov, M. V., Lo Surdo, P., Mattsson, P. T., Knekt, M., Westlund, A., Scheffzek, K., Saraste, M. & Svergun, D. I. (2003). *EMBO J.* **22**, 4616–4624.
- Mattinen, M. L., Paakkonen, K., Ikonen, T., Craven, J., Drakenberg, T., Serimaa, R., Waltho, J. & Annala, A. (2002). *Biophys. J.* **83**, 1177–1183.
- Matulis, D., Kranz, J. K., Salemme, F. R. & Todd, M. J. (2005). *Biochemistry*, **44**, 5258–5266.
- Mavaddat, N., Mason, D. W., Atkinson, P. D., Evans, E. J., Gilbert, R. J., Stuart, D. I., Fennelly, J. A., Barclay, A. N., Davis, S. J. & Brown, M. H. (2000). *J. Biol. Chem.* **275**, 28100–28109.
- Merry, A. H., Gilbert, R. J., Shore, D. A., Royle, L., Miroshnychenko, O., Vuong, M., Wormald, M. R., Harvey, D. J., Dwek, R. A., Classon, B. J., Rudd, P. M. & Davis, S. J. (2003). *J. Biol. Chem.* **278**, 27119–27128.
- Morth, J. P., Feng, V., Perry, L. J., Svergun, D. I. & Tucker, P. A. (2004). *Structure*, **12**, 1595–1605.
- Nettleship, J. E., Walter, T. S., Aplin, R., Stammers, D. K. & Owens, R. J. (2005). *Acta Cryst.* **D61**, 643–645.
- Pantazatos, D., Kim, J. S., Klock, H. E., Stevens, R. C., Wilson, I. A., Lesley, S. A. & Woods, V. L. Jr (2004). *Proc. Natl Acad. Sci. USA*, **101**, 751–756.
- Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P. & Salemme, F. R. (2001). *J. Biomol. Screen.* **6**, 429–440.
- Petoukhov, M. V., Eady, N. A., Brown, K. A. & Svergun, D. I. (2002). *Biophys. J.* **83**, 3113–3125.
- Petoukhov, M. V. & Svergun, D. I. (2005). *Biophys. J.* **89**, 1237–1250.
- Potier, N., Billas, I. M., Steinmetz, A., Schaeffer, C., van Dorsselaer, A., Moras, D. & Renaud, J. P. (2003). *Protein Sci.* **12**, 725–733.
- Potier, N., Lamour, V., Poterszman, A., Thierry, J. C., Moras, D. & Van Dorsselaer, A. (2000). *Acta Cryst.* **D56**, 1583–1590.
- Schuck, P. (2003). *Anal. Biochem.* **320**, 104–124.
- Shapiro, L., Doyle, J. P., Hensley, P., Colman, D. R. & Hendrickson, W. A. (1996). *Neuron*, **17**, 435–449.
- Siebold, C., Berrow, N., Walter, T. S., Harlos, K., Owens, R. J., Stuart, D. I., Terman, J. R., Kolodkin, A. L., Pasterkamp, R. J. & Jones, E. Y. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 16836–16841.
- Sokolova, A., Malfois, M., Caldentey, J., Svergun, D. I., Koch, M. H. J., Bamford, D. H. & Tuma, R. (2001). *J. Biol. Chem.* **276**, 46187–46195.
- Solovyova, A. S., Nollmann, M., Mitchell, T. J. & Byron, O. (2004). *Biophys. J.* **87**, 540–552.
- Stamper, C. C., Zhang, Y., Tobin, J. F., Erbe, D. V., Ikemizu, S., Davis, S. J., Stahl, M. L., Seehra, J., Somers, W. S. & Mosyak, L. (2001). *Nature (London)*, **410**, 608–611.
- Stehlin-Gaon, C., Willmann, D., Zeyer, D., Sanglier, S., Van Dorsselaer, A., Renaud, J. P., Moras, D. & Schule, R. (2003). *Nature Struct. Biol.* **10**, 820–825.
- Stevens, R. C. (2000). *Structure Fold. Des.* **8**, R177–R185.
- Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., Berrow, N., Owens, R., Gilbert, R., Davidson, A., Siddell, S., Poon, L. L., Diprose, J., Alderton, D., Walsh, M., Grimes, J. M. & Stuart, D. I. (2004). *Structure*, **12**, 341–353.
- Svergun, D. I. (1999). *Biophys. J.* **76**, 2879–2886.
- Svergun, D. I. & Koch, M. H. J. (2002). *Curr. Opin. Struct. Biol.* **12**, 654–660.
- Svergun, D. I. & Koch, M. H. J. (2003). *Rep. Prog. Phys.* **66**, 1735–1782.
- Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). *Biophys. J.* **80**, 2946–2953.
- Svergun, D. I., Petoukhov, M. V., Koch, M. H. J. & Koenig, S. (2000). *J. Biol. Chem.* **275**, 297–302.
- Trésaugues, L., Collinet, B., Minard, P., Henckes, G., Aufrere, R., Blondeau, K., Liger, D., Zhou, C. Z., Janin, J., Van Tilbeurgh, H. & Quevillon-Cheruel, S. (2004). *J. Struct. Funct. Genomics*, **5**, 195–204.
- Vachette, P., Koch, M. H. & Svergun, D. I. (2003). *Methods Enzymol.* **374**, 584–615.
- Vincentelli, R., Canaan, S., Campanacci, V., Valencia, C., Maurin, D., Frassinetti, F., Scappucini-Calvo, L., Bourne, Y., Cambillau, C. & Bignon, C. (2004). *Protein Sci.* **13**, 2782–2792.
- Wilson, W. W. (2003). *J. Struct. Biol.* **142**, 56–65.
- Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T., Ito, Y., Matsuo, Y., Kuroda, Y., Nishimura, Y., Kyogoku, Y., Miki, K., Masui, R. & Kuramitsu, S. (2000). *Nature Struct. Biol.* **7**, Suppl., 943–945.