Journal of Structural Biology 162 (2008) 451–459

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

An approach to quality management in structural biology: Biophysical selection of proteins for successful crystallization

Frank H. Niesen [a,1,*], Anja Koch [a,2], Ulf Lenski [b,c,3], Ulrich Harttig [c,3,6], Yvette Roske [b,4], Udo Heinemann [b,d,5], Klaus Peter Hofmann [a,5]

[a] Institut für Medizinische Physik und Biophysik, Charité Universitätsmedizin Berlin, Ziegel straße 5/9, D-10117 Berlin, Germany
[b] Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Straße 10, D-13125 Berlin, Germany
[c] RZPD German Resource Center for Genome Research GmbH, Heubnerweg 6, 14059 Berlin, Germany
[d] Institut für Chemie und Biochemie, Freie Universität Berlin, Takustraße 6, D-14195 Berlin, Germany

ARTICLE INFO

Article history:
Received 1 November 2007
Received in revised form 5 March 2008
Accepted 6 March 2008
Available online 25 March 2008

Keywords:
Protein crystallization
Protein fold
Structural genomics
Dynamic light scattering (DLS)
Differential scanning calorimetry (DSC)
Fourier-transform infrared spectroscopy (FTIR)

ABSTRACT

Aggregation, incorrect folding and low stability are common obstacles for protein structure determination, and are often discovered at a very late state of protein production. In many cases, however, the reasons for failure to obtain diffracting crystals remain entirely unknown. We report on the contribution of systematic biophysical characterization to the success in structural determination of human proteins of unknown fold. Routine analysis using dynamic light scattering (DLS), differential scanning calorimetry (DSC) and Fourier-transform infrared spectroscopy (FTIR) was employed to evaluate fold and stability of 263 purified protein samples (98 different human proteins). We found that FTIR-monitored temperature scanning may be used to detect incorrect folding and discovered a positive correlation between unfolding enthalpy measured with DSC and the size of small, globular proteins that may be used to estimate the quality of protein preparations. Furthermore, our work establishes that the risk of aggregation during concentration of proteins may be reduced through DLS monitoring. In summary, our study demonstrates that biophysical characterization provides an ideal tool to facilitate quality management for structural biology and many other areas of biological research.

1. Introduction

Structural biology provides information about proteins at an atomic level, thus contributing major insight into protein function, enzymatic activities and interactions with other macromolecules in the cell. The international effort of high-throughput genome sequencing projects results in the identification of all genes of more and more species, and structural genomics (SG) projects aim at the fast and efficient 3D structure determination of the encoded gene products with primary focus on proteins. Since SG was initiated at many locations in North America, Japan, Israel and Europe, the number of structures submitted to the Protein Data Bank, PDB (http://pdb.rcsb.org/pdb; Berman et al., 2000; Berman et al., 2002) has increased exponentially. Furthermore, SG contributes about 50% of the first structures identified in a protein family, providing important inference of the fold of homologous proteins in other species including humans (Chandonia and Brenner, 2006).

The main principle of the SG projects is to apply high-throughput methods on all levels from clone to structure on hundreds of protein targets in parallel (Gileadi et al., 2007; Todd et al., 2005; Zhang and Kim, 2003). Many methods have been developed to accomplish high-throughput in cloning, expression and purification of recombinantly produced proteins from bacteria, yeast and eukaryotic cells. Consequently, for the majority of cytoplasmic and extracellular proteins at least one construct can be purified in soluble form. The proteins then enter highly sophisticated, fast crystallization screens complemented by semi-automated methods for acquisition of diffraction data and efficient algorithms for structure determination (reviewed in: McPherson, 2004; Pusey et al., 2005). A close look at the success rates of SG projects, however, reveals that for little more than 10% of the purified proteins crystals with sufficient diffraction can be generated (Bourne et al., 2004; Pusey et al., 2005). Protein aggregation—the formation of irregular assemblies of protein molecules in non-native conformation whose re-dissolvation requires specific conditions and

* Corresponding author. Address: Structural Genomics Consortium, University of Oxford, ORCRB, Roosevelt Drive, Oxford, OX3 7DQ, UK. Fax: +44 1865 617575.
E-mail address: frank.niesen@sgc.ox.ac.uk (F.H. Niesen).

[1] Project management, database development, data visualization, experimental design, FTIR, DLS, DSC, manuscript preparation.
[2] FTIR, DLS, DSC.
[3] Database development, data mining tools and data visualization.
[4] Crystallization, DLS.
[5] Project design and management, manuscript preparation.
[6] Present address: DIFE, Arthur-Scheunert-Allee 114-116, 14558 Nuthetal, Germany.

1047-8477/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2008.03.007

almost always a long time scale (Fink, 1998)—represents the obvious reason for failure in a large number of these cases. However, for most of the other cases where proteins fail to crystallize no reason at all is usually revealed.

The now common 'pipeline' approach, which was pioneered by the protein structure factory (PSF, www.proteinstrukturfabrik.de, Heinemann et al., 2000; Heinemann et al., 2003) and various similar initiatives (Terwilliger, 2000; Yokoyama et al., 2000), applies successive steps for cloning, expression, purification, crystallization, X-ray diffraction and structural determination to many proteins in parallel. At the PSF, a total set of 890 targets, excluding proteins containing transmembrane regions, coiled-coil structures, or extended low-complexity sequences were selected for the first operational phase (Büssow et al., 2004; Büssow et al., 2005). One of the distinctive features of the PSF was a routine implementation of biophysical methods to characterize all proteins that were subjected to crystallization. Our aim was to test the value of these procedures for effective quality management at the nexus between purification and crystallography. The formation of protein crystals requires a nucleation process in a supersaturated solution that can only be induced for proteins at sufficient concentration. Supersaturation defines a 'metastable' phase, followed by the entry into the stable phase where the rate of re-dissolvation of particles is equal or lower than the rate of new protein molecules joining the growing crystal. For highly concentrated proteins, however, precipitates may become energetically favored, and in the absence of precipitants like ammonium sulfate or polyethylene glycol it is unlikely that the native fold, hence the active conformation, is retained (Schein, 1990). Instead, the disturbance of the protein's water sheath, due to displacement of water molecules with neighboring protein molecules, is likely to result in partial unfolding of the protein, and normally internal hydrophobic patches become exposed at the surface. To diminish their contact to water molecules, intermolecular clusters of protein molecules form at the energy minimum (Franks, 2002).

During their way through the structure analysis pipeline of an SG project the value of the proteins increases exponentially: while their purity and concentration increases, substantial loss of material during each involved procedure decreases the protein amount strongly. Since degradation of protein is likely to occur, the time between breakage of the cells and the start of the crystallization experiment should be as short as possible. Consequently, the biophysical characterization needs to comprise fast measurements that acquire data that can be unambiguously interpreted regarding fold and stability. Of special importance is that all information that is needed for downstream procedures, e.g., concentration of the protein, has to become available with minimal delay.

Here we describe powerful tools for data management and visualization of data acquired by dynamic light scattering (DLS), differential scanning calorimetry (DSC), fluorescence spectroscopy and Fourier-transform infrared spectroscopy (FTIR) that were created for this purpose. The setup enabled us to recognize common properties among 263 protein preparations (98 unique proteins) that were used to evaluate the propensity of the protein to crystallize.

2. Materials and methods

2.1. Expression and purification of proteins

Proteins were expressed and purified from cells of *Escherichia coli*, *Saccharomyces cerevisiae* or *Pichia pastoris* comprising affinity tags either on their N- and C-terminus (His_6 and Strep-II, respectively) or an N-terminal His_6 tag that was cleaved during purification as described (Büssow et al., 2004; Holz et al., 2003; Boettner et al., 2002). Typically, proteins were subjected to gel filtration on a Superose 12 16/50 column as the last step of purification, transferring them into standard buffer (15 mM Tris–HCl, pH 7.4, 50 mM NaCl, 2 mM DTT, 0.1 mM EDTA, 0.02% NaN_3) in the process. Protein concentrations were determined from UV absorption at 280 nm (Cary 50, Varian) using extinction coefficients (ε_{280}) calculated from the protein sequence (Pace et al., 1995). All purification steps were accompanied by SDS–PAGE analysis to confirm purity of the proteins. The proteins were concentrated by ultrafiltration (4 °C) through polyethersulfone membranes of appropriate cut-off size (VivaSpin, Sartorius). The final preparations were stored at 4 °C for crystallization screening.

2.2. Fourier-transform infrared spectroscopy

Prior to the measurement, proteins were deuterated to reduce superimposition in the Amide I region of the infrared spectrum by water. Concentrated protein samples were transferred into 20 mM sodium/potassium phosphate buffer, pD 7.0, containing 150 mM NaCl by 2× ultrafiltration on nitrocellulose membranes (VivaSpin 500, Sartorius). The buffer was freshly re-diluted into D_2O after freeze-drying of buffer prepared with H_2O. About 50 μl of buffer was injected into a freshly assembled temperature-controlled 30-mm diameter transmission cell with two BaF_2 windows and a 50 μl polytetrafluorethylene gasket. A reference UV absorption spectrum was taken (Cary 50, Varian; equipped with a home-made sample holder), and reference single beam spectra recorded with a Bruker ifs 113v spectrometer, equipped with a LN_2-cooled bolometer, were recorded after 30 min under continuous flow of dry air. Then the buffer was removed and the concentrated protein injected. The protein concentration was determined from absorption at 280 nm (see above). Then the transmission cell was placed into the sample chamber of the spectrometer and subjected to a continuous flow of dry air for 30 min before the first spectrum was taken. During a linear, manually controlled increase of the temperature (ΔT = 1 °C/min; water bath) spectra were recorded every 5 °C. The second derivatives of the difference spectra were calculated using the instrument software (Opus 4.2, Bruker) and normalized to the protein concentration.

2.3. Automated differential scanning calorimetry

Experiments were performed using a capillary microcalorimeter equipped with an auto sampler to exchange samples from Peltier-cooled microplates (capDSC, MicroCal, LLC (Plotnikov et al., 2002)). Samples were scanned between 10 and 90 °C, at a rate of 1 °C/min. Six buffer reference samples were measured prior to each set of samples. All protein samples were diluted 10–20-fold into 20 mM of buffer (sodium citrate, pH 6, or Na_2HPO_4/KH_2PO_4, pH 7.0 or 8.0, respectively) containing 150 mM NaCl in deep-well microplates (2 ml). Of each sample 300 μl were transferred into a 96-well clear-bottom black microplate (Corning) for the determination of the protein concentrations from absorbance at 280 nm (MikroTek DS, Bio-Tek Instruments). The enthalpy of transition from the native (n) to the denatured (d) state, ΔH_{n-d}, was calculated from the molar heat capacity difference, $\Delta C_P/T$, between the sample cell filled with protein solution and a reference cell filled with buffer alone, using the instrument software (Origin, Originlab). However, for the majority of the samples a non-two state mechanism was observed, and van't Hoff enthalpies were calculated. All enthalpies were normalized to the protein concentration.

2.4. Dynamic light scattering (DLS) and precipitation test

Samples were generated by stepwise (3–4 steps) concentration of an aliquot (2 ml) of a protein preparation (from 0.1 to 1.0 mg/ml; ultrafiltration with MWCO = 10 kDa). One sample, of