

Quality control of purified protein

Best practice recommendations

The Association of Resources for Biophysical Research in Europe (ARBRE-MOBIEU) and the Production and Purification Partnership in Europe (P4EU) have produced a joint initiative on recombinant protein quality. We aim to develop a minimal reporting standard/best practice for the quality control of recombinant proteins to ensure that the input material used in biological, biophysical and biochemical research is of high quality. This condition, in turn, will result in more reliable and reproducible final data. The prescribed tests must be simple to perform using standard laboratory equipment, while still producing data acceptable as admission criteria for biophysical or structural biology labs.

Below you find some recommendations for what we propose as (i) minimal information, (ii) minimal quality control parameters and (iii) extended quality control parameters/standards. Example of results with the most used methods are presented.

(i) Minimal information to provide in publications

The aim is to provide sufficient information on the **protein identity, expression and purification parameters** such that the experiment can be replicated reliably in any laboratory.

- **Protein name and full primary structure**, by providing a NCBI (or UniProt) accession number, cloning strategies, and the source of the DNA (species).
- **Expression vector and host strain**, including the tags and cleavage sites used, accompanied by the full amino acid sequence of the final protein, or sufficient details to derive the full amino acid sequence of the final protein.
- **Expression and Purification protocol**, namely the detailed description of all the protein production steps.
- **Protein concentration** (specifying the method used for quantification and the molar extinction coefficient at 280nm, if applicable. Extinction coefficient can be obtained for any sequence by using open access resources -<https://web.expasy.org/protparam/>).
- **Storage conditions**, i.e. final buffer composition (pH, buffers, salts and additives), storage temperature and, where applicable, freezing or lyophilization conditions.

Example of description

eSpCas9 (including both N and C-terminal NLS sequences) was amplified from plasmid eSpCas9(1.1) (Addgene plasmid 71814) using the high fidelity KOD polymerase (Merck) using the following primers:

fwd-AGGAGATATACCATGGCCCCAAAGAAGAAGCGGAAGG

rev-CCACGAAAAAGGCCGGCCAGGCCAAAAAGAAAAAGAAACATCACCATCAC

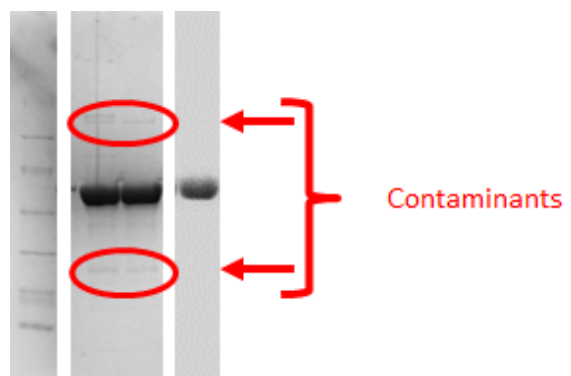
The resulting 4227 b.p. PCR product was DpnI treated to remove the template and purified using Ampure (Beckman Coulter) before In-Fusion (Clontech) into NcoI-PmeI –cut pOPINE (Addgene plasmid 26043). The resulting plasmid will express eSpCas9 with an N-terminal NLS and a C-terminal NLS-His-tag). The plasmid was fully sequenced before *E. coli* Rosetta(DE3) pLysS cells were transformed with it, and transformants were selected on LB agar plates supplemented with 1% glucose, Carbenicillin (50µg/ml) and Chloramphenicol (35µg/ml). A starter culture was grown overnight at 37°C with shaking at 250rpm in LB supplemented with 1% glucose, Carbenicillin (50µg/ml) and Chloramphenicol (35µg/ml). The culture was then diluted 1:50 in TB Overnight Express media auto-induction media (Merck), grown for a further 3 hours at 37°C until the OD₆₀₀ reached approximately 0.5, the temperature was then reduced to 25°C and the culture maintained for a further 24 hours. Cells were collected by centrifugation at 5000g for 10 minutes and the pellets (15 g/L) stored at -80°C before processing. Pellets were lysed in 5 volumes of **Lysis/Loading Buffer** (20mM Tris pH8, 500mM NaCl, 30mM Imidazole supplemented with 1 complete EDTA free tablet [Roche] and 1000Kunits of DNase [Merck]/10g of pellet) at 20Kpsi using a cell disruptor (Constant Systems Ltd. UK). The lysate was cleared by centrifugation for 20 minutes at 30,000xg at 4°C and filtration (cut-off 0.4 µm) before loading onto a 5ml HisTrap HP column (GE Healthcare). After sample loading the column was washed with 10CV of loading buffer before the eSpCas9 was eluted using **Elution Buffer** (20mM Tris pH8, 500mM NaCl, 500mM Imidazole, 2mM DTT, 10% glycerol). Peak fractions were analyzed by SDS-PAGE before pooling and buffer exchange into **DESALT/Storage Buffer** (20mM Tris pH8, 200mM KCl, 10mM MgCl₂) using a HiPrep Desalting column (GE Healthcare). eSpCas9 was then concentrated to a maximum of 3mg/ml using a ultrafiltration device with a 30kDa MW cut-off, concentration was assayed using a Nanodrop instrument (Thermo Scientific) and a calculated A₂₈₀ of 0.69 A_{280nm}/mg/ml). eSpCas9 fraction was then separated into single-use aliquots that were flash-frozen in liquid nitrogen and stored at -80°C until further use.

(ii) Minimal quality control parameters that should be tested on protein sample

- **Purity:** checked by SDS-PAGE, Capillary Electrophoresis (CE) or Reversed-Phase-HPLC (RP-HPLC). The objective here is to assess the presence (and level) of contaminants using techniques that are available to most laboratories. The assessment of sample purity by SDS-PAGE is illustrated below.

Example: SDS-PAGE

If SDS-PAGE is the sole method of assessment of sample purity then ideally no other band other than the expected one for your protein of interest should be detected, as illustrated in the example figure. The staining should be chosen according to the amount loaded on the gel in order to be able to detect contaminants of 1% or less of the total protein load. N.B.: the detection limit of Coomassie blue staining is approximately 100ng per band, for reverse zinc



staining approximately 10ng per band, while fluorescent or silver stains have a detection limit of approximately 1ng of protein per band. In other words, if you load 10 μ L of a solution at 1mg/ml, you will load in total 10 μ g of protein meaning. In order to detect contaminants you will need sensitivity lower than 100ng per band and you should therefore use reverse zinc, fluorescent or silver stain to be able to assess contamination. It is also possible to perform TCA or DOC/TCA or acetone precipitation to concentrate ten or more times the sample.

Common staining protocols can be found in:

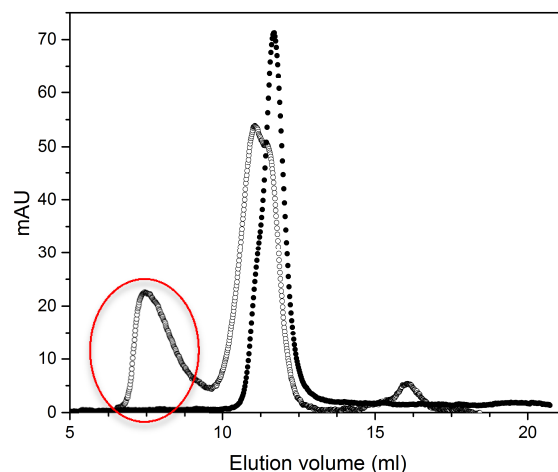
- ⊙ *Quality assessment and optimization of purified protein samples: why and how? Bertrand Raynal, Pascal Lenormand, Bruno Baron, Sylviane Hoos and Patrick England Microbial Cell Factories 2014 ; 13:180.*

More detail about other techniques can be found in:

- ⊙ *Applications of capillary electrophoresis in characterizing recombinant protein therapeutics. Zhao SS, Chen DDY Electrophoresis 2014; 35:96–108.*
- ⊙ *Hydrophobic interaction chromatography for the characterization of monoclonal antibodies and related products. Fekete S, Veuthey JL, Beck A, Guillaume D. J Pharm Biomed Anal. 2016; 130:3-18.*
- **Homogeneity (aggregation state):** checked preferably by Size Exclusion Chromatography (SEC) and/or Dynamic Light Scattering (DLS) or by Size Exclusion Chromatography in combination with Multi Angle Light Scattering (SEC-MALS), Field Flow Fractionation (FFF) or Field Flow Fractionation in combination with Multi Angle Light Scattering (FFF-MALS) or Analytical Ultracentrifugation (AUC). The objective is to assess if the sample has a tendency to form aggregates in the condition used to purify it, and to assess the potential for oligomerization of the protein sample.

Example: SEC

If you are using analytical size exclusion chromatography to assess homogeneity, only regular peaks corresponding to the prevalent monomeric or oligomeric species specific for that protein should be detected, and no aggregates should be detected in the void volume of the column situated in a position corresponding to 1/3 of the total volume of the size exclusion column used (e.g. for column with a total volume of 21ml the void volume is situated at an elution volume of 7ml). The figure below shows a good sample (filled dots) with a single, well-defined, symmetric peak, representing a homogeneous species, and a heterogeneous sample containing aggregates (open dots), note the 'aggregate' peak at the



note the 'aggregate' peak at the

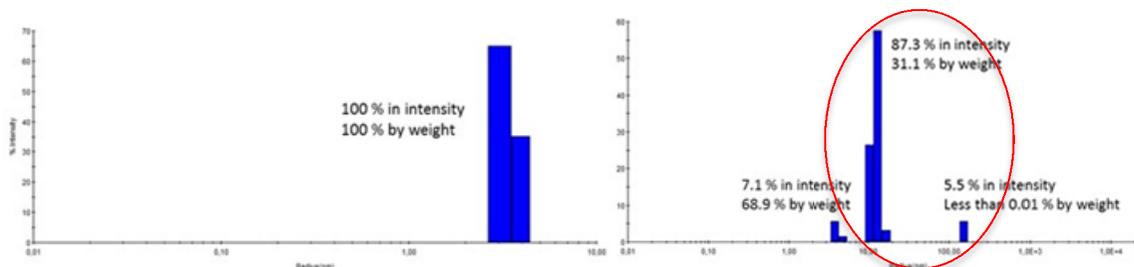
void volume of the column (circled) and the asymmetric main peak. SEC can, of course, also be used preparatively at larger scales to select specific oligomeric states e.g. if a dimer is functional in an experiment or assay but a monomer is not the peak corresponding to the size of the dimer may be selected. Please note that after pooling a specifically 'sized' peak any further processes e.g. concentration or buffer exchange may alter the oligomeric state. Pooled fractions from a single peak should be preferentially checked again by a further SEC round.

More detail about Size exclusion Chromatography can be found in:

- ⊙ Theory and practice of size exclusion chromatography for the analysis of protein aggregates. Fekete S, Beck A, Veuthey J-L, Guillaume D: *J Pharm Biomed Anal* **2014**; **101:161–173**.
- ⊙ Useful practical information can be accessed in: <https://www.gelifesciences.com/en/de/solutions/protein-research/knowledge-center/protein-handbooks> and <http://wolfson.huji.ac.il/purification/>

Example: DLS

If DLS is used to assess the homogeneity of your sample ideally only one species with a low polydispersity (with less than 20% dispersity of the peak) should be detected, and no aggregates should be detectable (as greater than 1-2 percent of your sample). The figure below shows a good sample (Left Figure) with a single peak and a heterogeneous sample containing aggregates (Right Figure).



One should remember that the signal is dependent on the size of the detected particle. For example in the right figure the aggregate of 150nm (right circle) represents more than 5% of signal, however less than 1% in weight, it is therefore negligible. On the contrary, the aggregates with an approximate size of 15nm (central peak) represent nearly 90% of the intensity signal but only 31% of the sample weight. This sample can clearly be defined as non-homogeneous.

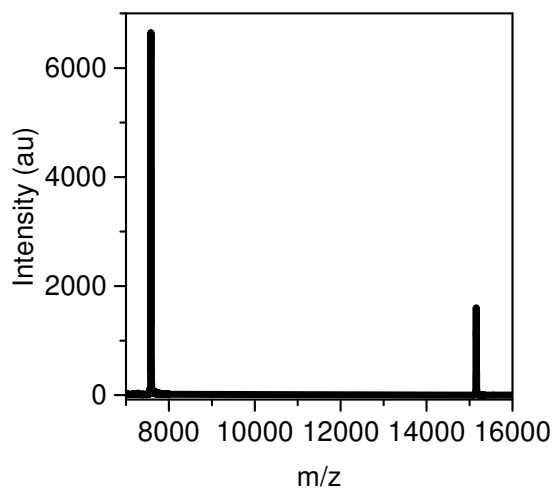
More detail about DLS can be found in:

- ⊙ Dynamic light scattering: a practical guide and applications in biomedical sciences. Stetefeld J, McKenna SA, Patel TR. *Biophys Rev.* **2016**; **8:409-427**.
- ⊙ Nobbmann U, Connah M, Fish B, Varley P, Gee C, Mulot S, Chen J, Zhou L, Lu Y, Shen F, Yi J, Harding SE: Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies. *Biotechnol Genet Eng Rev* **2007**; **24:117–128**.
- ⊙ Philo JS: Is any measurement method optimal for all aggregate sizes and types? *AAPS J* **2006**; **8:E564–571**.

- **Identity:** checked preferably by intact protein mass, peptide mass fingerprinting or Edman sequencing.

Example: Intact mass by MALDI-TOF

For confirmation of sample identity mass spectrometry is a technique that is now widely available, both in academic institutions and commercially. Mass spectrometry will identify species according to their mass over charge ratio. In the intact mass (also known as ‘top-down’ MS) example presented here (see figure), the purified protein has an expected mass of 15154.8 Da. Two peaks can be detected: one with a double charge and a ratio of 7578.34 m/z and one with a single charge and a ratio of 15155.56 m/z. Since these two peaks are in good agreement with the expected molecular mass, the produced protein is the expected one and no trace of degradation or contamination can be detected.



Example: tryptic digest/MS

You may also use ‘bottom-up’ mass spectrometry (e.g. tryptic digest/MS) to confirm the identity of our protein. This method will also detect (and identify) any contaminating proteins but will not provide information on the integrity of your protein. It can be conveniently performed on samples excised from SDS-PAGE gels.

More detail about mass spectrometry can be found in:

- ⊙ Analysis of intact protein isoforms by mass spectrometry. Tipton JD, Tran JC, Catherman AD, Ahlf DR, Durbin KR, Kelleher NL: J Biol Chem **2011**, **286**:25451–25458.
- ⊙ Overview of peptide and protein analysis by mass spectrometry. Zhang G, Annan RS, Carr S, Neubert T Curr Protoc Protein Sci **2010**, **62**:16.1.1–16.1.30

(iii) Extended quality control parameters

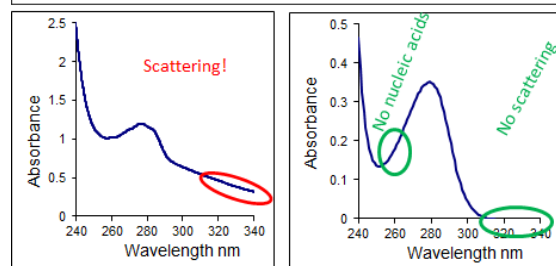
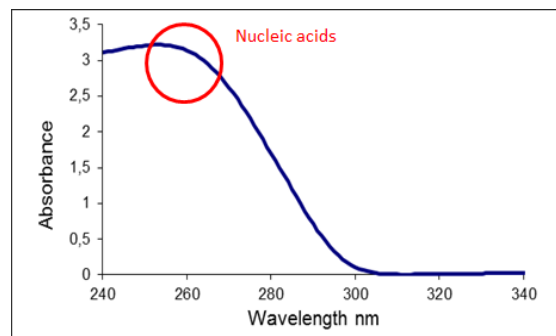
Depending on the intended use and in addition to the methods listed above, the following methods may also be applicable and are highly recommended, although they are not considered to be as essential as those in section (ii)

The first of these tests can be performed simultaneously when determining protein concentration spectrophotometrically (using A_{280nm}) as most instruments will also allow you

to collect data over a wide spectrum of wavelengths. This additional data can be very informative.

- **General quality test by UV spectroscopy** between 200nm and 340nm to check nucleic acid content and general protein fitness/quality: **Mandatory if protein binds nucleic acids.**

As general quality test the simplest qualitative test is the measure of UV absorbance between 240nm and 340nm. One of the main advantages is the availability of such equipment in all the biological laboratories. Apart from measuring the concentration by means of the protein extinction coefficient, the full spectrum will inform on the general quality of the preparation. It should be stressed that the exact protein buffer as to be used as a blank in order As general quality control test the simplest qualitative test is the measurement of UV absorbance between 240 nm and 340 nm. One of the main advantages is the availability of such equipment in all the biological laboratories. Apart from measuring the concentration by means of the protein



extinction coefficient, a full spectrum will inform on the general quality of the preparation as well. It should be stressed that the exact protein buffer needs to be used as a blank in order to avoid errors in the interpretation. A strong signal at 260 nm is usually a sign of nucleic acid contamination. As a general rule the ratio between the absorbance at 260 nm and 280 nm ($Abs\ 260/Abs\ 280$) should give a value close to 0.6 for a good quality protein preparation. Furthermore, a regularly rising absorbance between 340 nm and 300 nm is usually the sign of scattering due to aggregation (see figure). One simple way to confirm scattering is to determine an aggregation index by calculating $100 \times Abs\ 340/[Abs\ 280 - Abs\ 340]$. As a rule of thumb, an index lower than 2 would be acceptable for a non-aggregated protein.

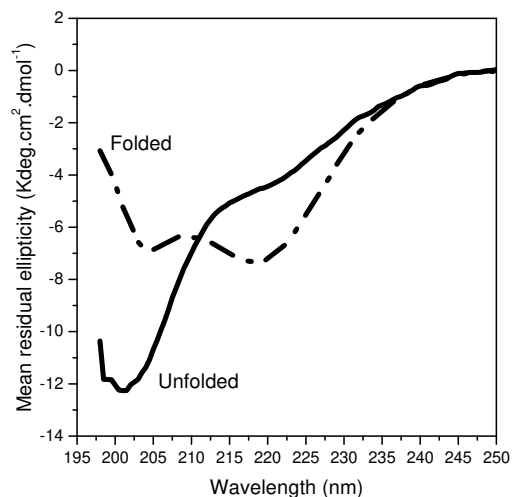
More detail about UV spectrometry can be found in:

- ⊙ Leach SJ, Scheraga HA: Effect of Light Scattering on Ultraviolet Difference Spectra. *J Am Chem Soc* **1960**, **82:4790–4792**.
- ⊙ Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. Glasel JA. *Biotechniques*. **1995**; **18:62-3**.
- ⊙ Dual effects of Tween 80 on protein stability. Wang W1, Wang YJ, Wang DQ. *International Journal of Pharmaceutics* **2008**; **347:31–38**
- ⊙ Monodispersity of recombinant Cre recombinase correlates with its effectiveness in vivo. Capasso P, Aliprandi M, Ossolengo G, Edenhofer F, de Marco A *BMC Biotechnol* **2009**; **9:80**

- **Homogeneity:** The following techniques are complementary to the ones previously described: analytical Ion Exchange Chromatography (IEX), analytical Hydrophobic Interaction Chromatography (HIC) or Isoelectric Focusing (IEF). The objective is to get extra information on the homogeneity of the sample. If you need a full description of the techniques please see:
 - ⊙ Separation techniques: Chromatography Ozlem Coskun North Clin Istanbul. **2016; 3: 156–160.**
 - ⊙ Isoelectric Point Separations of Peptides and Proteins. Pergande MR, Cologna SM. Proteomes. **2017; 5: E4**
- **Conformational stability/folding state:** Circular Dichroism (CD), Differential Scanning Calorimetry (DSC), NMR, Fourier Transform InfraRed (FTIR). The objective of the measurement is to verify that the same folding signature can be seen. One of the classical techniques to perform this type of measurement is to use circular dichroism.

Example: Folding state by circular dichroism

Circular dichroism (CD) looks at the difference of absorption of left and right handed light. In the figure, typical curves of both an unfolded protein (plain line) and of a folded protein (dotted line) are presented.



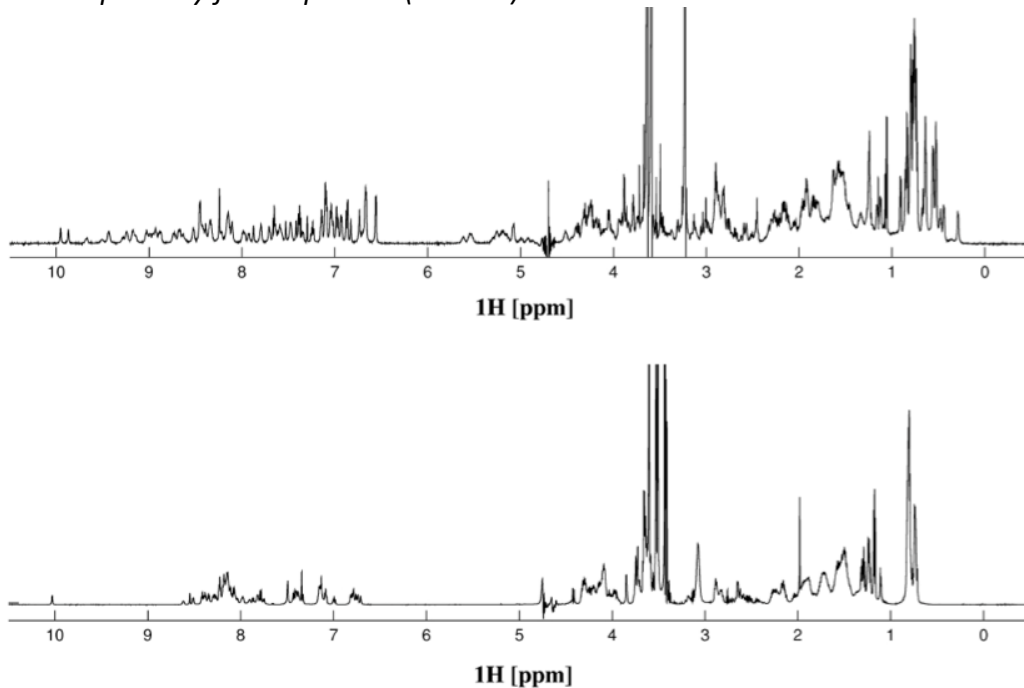
More detail about CD can be found in:

- ⊙ How to study proteins by circular dichroism. Kelly SM1, Jess TJ, Price NC. Biochim Biophys Acta. **2005; 1751:119-39.**
- ⊙ Woody WR. Methods in Enzymology **1995; 246:**
- ⊙ <http://www.proteinchemist.com/cd/cdspec.html>
- ⊙ Circular Dichroism and the conformational analysis of biomolecules. Fasman GD, Plenum Press, **1997** New York and London
- ⊙ Circular dichroism and its application to the study of biomolecules. Martin SR, Schilstra MJ. Methods Cell Biol. **2008; 84:263-93.**

Example: Folding state by NMR spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a spectroscopic method to study magnetically active nuclei in an external magnetic field. Measuring a one dimensional (1D) ¹H NMR spectrum of a protein requires only little material, is pretty fast and the spectrum clearly indicates if a protein is folded or not. In the 1D ¹H spectrum of a well-folded protein (top) the peaks are narrow and sharp and distributed over a large range of chemical shifts (good signal dispersion); signals can be especially found at ppm values < 0.5 (corresponding to high field-

shifted methyl group protons) or > 8.5 (corresponding to down field-shifted amide protons) are observable. In contrast, the peaks are broader and not as widely dispersed in the spectrum of an unfolded or partially folded protein (bottom).



More details about NMR spectroscopy can be found in:

⊙ *Applied NMR Spectroscopy for Chemists and Life Scientists*. Zerbe O, Jurt S. Wiley-VCH Verlag GmbH & Co. KGaA, **2014** Weinheim, Germany

- **Protein competent fraction**, i.e. the relative amount of functionally active protein, measured as specific activity, by active site titration or other suitable methods. The objective of this test is to be able to determine the amount of active molecule in the protein preparation. When the specificity test is based on interaction between molecules this measure can be realized by a surface plasmon resonance (SPR) technique using the “calibration-free concentration analysis” (CFCA) method, which has been implemented in different SPR instruments available commercially. More detailed can be found in:
- ⊙ *Pol E: The importance of correct protein concentration for kinetics and affinity determination in structure-function analysis*. *J Vis Exp* **2010**, **37:2–8**.
- **Endotoxin content**, this analysis is mandatory for protein samples used in combination with cell cultures.
- **Optimization of storage conditions**: To minimize the formation of protein aggregates and to improve solubility, adjustment of several parameters of the sample buffer composition (pH, salinity, the presence of detergents, cryo-protectants or other additives, co-factors or ligands etc.) can be made to increase homogeneity and long term

stability. These conditions can be assessed using automated technique such as DLS (described above) or thermal shift assay.

Example: optimization of buffer condition by automated DLS

This DLS conformation that allows processing a large number of samples in plate format, has simplified buffer condition screening. Buffer matrices for multi-parametric screening of pH, salinity, buffer nature, additives and co-factors can be generated by hand or using simple robotics. One approach is to dilute samples in the different buffer, classically ten times with a final concentration of 1 mg/ml for a 10 kDa protein or 0.1 mg/ml for a 100 kDa protein. The homogeneity of the sample and the presence of aggregates can be assessed for each condition, and the optimal buffer composition can be selected, according to solubility parameters and the downstream application.

More detail about this approach can be found in:

- ⊙ Solubility at the molecular level: development of a critical aggregation concentration (CAC) assay for estimating compound monomer solubility. Wang J, Matayoshi E **Pharm Res** 2012, 29:1745–1754.

Example: optimization of buffer condition by thermal shift assay

Modern Thermofluor and differential scanning fluorimeter that allow processing a large number of samples are methods for screening buffer conditions. Similarly to automated DLS, buffer matrices can be generated.

In a thermal shift assay, protein unfolding is monitored by heating up the samples in a linear temperature gradient in the presence of an environmentally sensitive dye. One of the most commonly used dyes is SYPRO orange, which fluoresces upon interaction with hydrophobic parts of the protein. As the temperature increases, the protein starts to unfold and exposes its hydrophobic core. The dye interacts with these exposed hydrophobic areas and the fluorescence increases. Ideally, the samples will display a sharp increase in fluorescence over a short temperature interval. The inflection point of this sigmoidal curve then represents the protein melting temperature T_m , at which 50% of the protein is unfolded. Most thermal shift assays are performed by using 96-well plates containing various buffer conditions in a real-time PCR machine. Comparing the T_m 's between all these different conditions provides hints to optimize the buffer conditions for protein purification and storage.

More detail about this approach can be found in:

- ⊙ Optimization of protein purification and characterization using Thermofluor screens. Boivin S, Kozak S, Meijers R: **Protein Expr Purif** 2013, 91:192–206.
- ⊙ A thermal stability assay can help to estimate the crystallization likelihood of biological samples. Dupeux F, Röwer M, Seroul G, Blot D, Márquez JA. *Acta Crystallogr D Biol Crystallogr.* 2011; 67:915-9.
- ⊙ Leung S-M, Senisterra G, Ritchie KP, Sadis SE, Lepock JR, Hightower LE. Thermal activation of the bovine Hsc70 molecular chaperone at physiological temperatures: physical evidence of a molecular thermometer. *Cell Stress & Chaperones.* 1996;1(1):78-89.

- **Batch-to-batch consistency:** Mandatory if more than one batch is used-many factors can affect the quality of your protein and you should never assume that all preps are of equal quality. Use some of the methods listed above, e.g. spectroscopic technique such as UV spectroscopy and circular dichroism are rapid and effective methods of quality assessment.