Amino Acids (2014) 46:1565–1582
DOI 10.1007/s00726-014-1699-x

AMINO ACIDS PROTOCOLS

The Center for Optimized Structural Studies (COSS) platform for automation in cloning, expression, and purification of single proteins and protein–protein complexes

Georg Mlynek · Anita Lehner · Jana Neuhold · Sarah Leeb · Julius Kostan · Alexej Charnagalov · Peggy Stolt-Bergner · Kristina Djinović-Carugo · Nikos Pinotsis

Received: 21 September 2013 / Accepted: 14 February 2014 / Published online: 20 March 2014

Abstract Expression in *Escherichia coli* represents the simplest and most cost effective means for the production of recombinant proteins. This is a routine task in structural biology and biochemistry where milligrams of the target protein are required in high purity and monodispersity. To achieve these criteria, the user often needs to screen several constructs in different expression and purification conditions in parallel. We describe a pipeline, implemented in the Center for Optimized Structural Studies, that enables the systematic screening of expression and purification conditions for recombinant proteins and relies on a series of logical decisions. We first use bioinformatics tools to design a series of protein fragments, which we clone in parallel, and subsequently screen in small scale for optimal expression and purification conditions. Based on a scoring system that assesses soluble expression, we then select the top ranking targets for large-scale purification. In the establishment of our pipeline, emphasis was put on streamlining the processes such that it can be easily but not necessarily automatized. In a typical run of about 2 weeks, we are able to prepare and perform small-scale expression screens for 20–100 different constructs followed by large-scale purification of at least 4–6 proteins. The major advantage of our approach is its flexibility, which allows for easy adoption, either partially or entirely, by any average hypothesis driven laboratory in a manual or robot-assisted manner.

Keywords Construct optimization · Ligation-indepedent cloning · Expression screening · Optimized protein production · Structural biology

Introduction

Systematic biochemical and biophysical studies investigating functions of individual proteins in the cell, interactions with their binding partners, or their three-dimensional structure typically require milligram quantities of highly pure and homogeneous samples. As a result, the establishment of rapid and efficient expression and purification pro-

Electronic supplementary material The online version of this article (doi:10.1007/s00726-014-1699-x) contains supplementary material, which is available to authorized users.

G. Mlynek · A. Lehner · J. Neuhold · S. Leeb · J. Kostan · A. Charnagalov · K. Djinović-Carugo (✉) · N. Pinotsis
Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Campus Vienna Biocenter 5, 1030 Vienna, Austria
e-mail: kristina.djinovic@univie.ac.at

Present Address:
A. Lehner · J. Neuhold · P. Stolt-Bergner
Protein Technologies Facility, Campus Science Support Facilities GmbH, Dr. Bohrgasse 3, 1030 Vienna, Austria

P. Stolt-Bergner
Research Institute of Molecular Pathology, Dr. Bohr-gasse 7, 1030 Vienna, Austria

K. Djinović-Carugo
Department of Biochemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Aškerčeva 5, SI-1000 Ljubljana, Slovenia

Present Address:
N. Pinotsis (✉)
Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK
e-mail: nikos.pinotsis@univie.ac.at

cedures for recombinant proteins represents an important asset for any laboratory that studies proteins at a molecular level.

The fastest and most cost-effective approach to produce proteins is recombinant cloning and heterologous expression in *Escherichia coli* (*E. coli*). However, successful production of soluble protein(s) in *E. coli* can be challenging when larger and more complex proteins or eukaryotic proteins, which often require molecular chaperones and/or various post-translation modifications for their stability and function, need to be expressed.

The conventional approaches to overcome low expression and solubility levels of proteins in *E. coli* focus on optimization of expression conditions such as temperature, type of cultivation media and the type of *E. coli* strain (Vincentelli et al. 2003; Graslund et al. 2008b) as well as on utilization of various solubility tags fused to the N- or C-terminus of the protein (Vincentelli et al. 2011; Graslund et al. 2008a; Festa et al. 2013). Often the protein is divided into smaller units representing functionally and structurally defined domains. Here the information derived from bioinformatics tools for analysis of both DNA and amino-acid sequences is essential (Radivojac et al. 2013). However, even for proteins with well-defined domain boundaries, it is still not possible to predict which specific N- and C-terminal boundaries are most compatible with the expression of a soluble protein. Therefore, the use of the nested cloning approach, a combination of different starting and ending points that leads to several constructs per protein target, increases the probability for soluble expression (Graslund et al. 2008a, b).

To address the aforementioned obstacles, expression screenings on small scale were introduced, where optimal cultivation and expression conditions can be transferred to large-scale protein production (Vincentelli et al. 2003; Graslund et al. 2008a; Dahlroth et al. 2006). In the past decade several structural genomics consortia and individual research groups have driven the development and optimization of protocols for protein production in *E. coli* (Graslund et al. 2008a, b). The major benefit of these initiatives for the scientific community was the optimization of routine procedures towards streamlined "consensus" conditions used for initial protein production, thus rendering them affordable and suitable for laboratories without pronounced needs for automation (Graslund et al. 2008a, b). The optimized protocols rely on the use of a single bacterial strain for initial expression screening and auto-induction media, reducing in this way the complexity of the initial screening approach (Vincentelli et al. 2011; Studier 2005). Moreover, as modern bacterial expression screens do not focus on protein expression but rather on the protein yield and purity, new benchmarks for evaluation of soluble protein expression have recently been defined (Vincentelli et al. 2011; Graslund et al. 2008a), in which the ultimate criterion is the potential to proceed with the large scale purification of the target. Hence, failures such as poor expression, formation of inclusion bodies or missfolded products are all considered as negative results.

One of the major tasks of the Center of Optimized Structural Studies (COSS) at the University of Vienna was to adopt and combine approaches and methodologies used for protein expression in *E. coli* and to establish an optimized pipeline for fast and cost-effective protein production. The established platform is capable of dealing with large numbers of constructs, is flexible enough to adopt customized approaches if required, and is suitable for automation.

The COSS protein production platform consists of five consecutive experimental modules focused on (i) cataloging target information, bioinformatics analysis and construct design, (ii) cloning of multiple constructs, (iii) expression screening and evaluation, (iv) large-scale protein expression and purification and (v) biophysical characterization (Fig. 1). Here we present a combination of experimental modules for high-throughput expression in *E. coli* cells and purification of soluble proteins suitable for downstream applications.

Experimental design

Bioinformatics analysis and target curation

The COSS principal aim is to prepare samples amenable to structural and functional studies. Careful design of protein constructs is a critical factor in successful generation of protein crystals. For this reason we selected a series of bioinformatics tools that we routinely use for definition of domain boundaries of the targets as well as for detection of intrinsically disordered amino acid sequences.

The initial analysis is performed with the assistance of two meta-servers, the protein CCD server at the NKI (http://xtal.nki.nl/ccd/Welcome.html) (Mooij et al. 2009) and the meta Protein DisOrder prediction System—PrDOS (http://prdos.hgc.jp/cgi-bin/meta/top.cgi) (Ishida and Kinoshita 2008). The protein CCD meta-server employs as input the DNA sequence of the target and generates various amino acid sequence analyses and predictions, such as secondary structure, intrinsically disordered regions and location of trans-membrane segments, by combining output from several predefined software packages. The PrDOS meta-server gathers information from eight predictors of intrinsically disordered regions and generates a score of disorder for each amino acid in the sequence. It, therefore, provides a more comprehensive analysis of a protein's disordered regions and generates a user-friendly graphical representation of the results. The output from both meta-servers is evaluated and used to design optimal constructs for structural or biochemical studies.

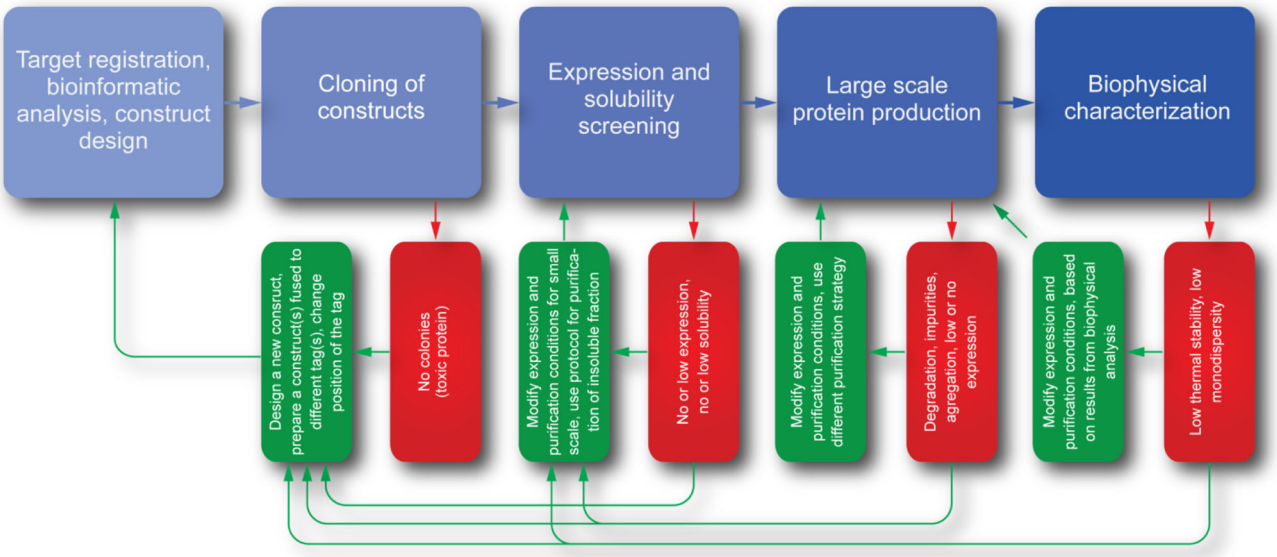

Fig. 1 Workflow and subroutines chart of the COSS platform. The individual experimental modules are shown in *blue boxes*. General problems that might occur within individual/particular module(s) are indicated in *red boxes*, while general solutions to these problems are shown in *green boxes*

To deal with the large number of nested constructs generated per target, it is beneficial to introduce a laboratory information management system (LIMS). We chose to manage our data through the Protein Information Management System (PIMS—http://www.pims-lims.org). PIMS is freely available for academic users and can be hosted on a web server, thereby enabling access from any machine with internet access. Apart from cataloging the basic information for each target, PIMS provides the possibility to track the complete progress for each target from construct design up to the final status of the sample. In parallel PIMS significantly simplified the process of primer design. The only requirement from the user is to define the starting and ending point of each target as well as the expression vector.

Cloning

The design of the COSS expression vectors is inspired by the ACEMBL vector system (Bieniossek et al. 2009). It is based on our needs for automation of the cloning steps for single proteins and protein complexes. The COSS vectors fall into two categories based on the presence or absence of the fusion tag(s) and its type. The first category comprises vectors for generation of proteins with non-cleavable C-terminal His_6-tag or untagged proteins (Fig. 2; Table 1). The second category comprises vectors for generation of proteins with N-terminal His_6- or Strep-tag either alone, or followed by a fusion tag that enhances solubility and expression levels (Fig. 2; Table 1). In this category of

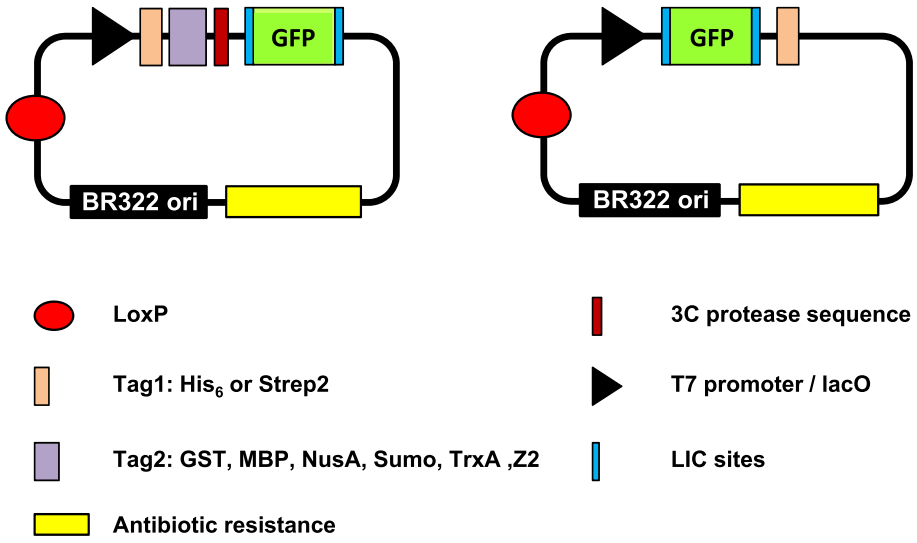

Fig. 2 General representation of our modified pACE vectors with an N-terminal tag (*left panel*) or C-terminal tag (*right panel*). Important modifications include the insertion of a specific LIC site (different for the two categories of vectors), the tags and the 3C protease cleavage site. LoxP, imperfect inverted repeats (LoxP sites), are used for Cre-mediated assembly and disassembly of donor and acceptor vectors

Table 1 Vectors

Vector	Promoter	Selection	Tag	Protease cleavage site	Origin	Forward primer for amplification	Reverse primer for amplification
C-terminal tagged and untagged vectors							
pACE3-CH[a]	T7/lac	Kan	C-His or untagged	N/A	BR322	pace_uncl_FW	pace_uncl_RV
N-terminal tagged vectors							
pACE1-NH	T7/lac	Amp	N-His	3C	BR322	pace_cl_FW	pace_cl_RV
pACE1-NS	T7/lac	Amp	N-Strep	3C	BR322	pace_cl_FW	pace_cl_RV
pACE1-GST	T7/lac	Amp	N-Strep-GST	3C	BR322	pace_cl_FW	pace_cl_RV
pACE1-MBP	T7/lac	Amp	N-Strep-MBP	3C	BR322	pace_cl_FW	pace_cl_RV
pACE1-NusA	T7/lac	Amp	N-Strep-NusA	3C	BR322	pace_cl_FW	pace_cl_RV
pACE1-Sumo	T7/lac	Amp	N-Strep-Sumo	3C	BR322	pace_cl_FW	pace_cl_RV
pACE1-TrxA	T7/lac	Amp	N-Strep-TrxA	3C	BR322	pace_cl_FW	pace_cl_RV
pACE1-Z2	T7/lac	Amp	N-His-Z2	3C	BR322	pace_cl_FW	pace_cl_RV
pACE2-NH	T7/lac	Tet	N-His	3C	BR322	pace_cl_FW	pace_cl_RV
pACE3-NH	T7/lac	Kan	N-His	3C	BR322	pace_cl_FW	pace_cl_RV
pACE3-GST	T7/lac	Kan	N-His-GST	3C	BR322	pace_cl_FW	pace_cl_RV
pACE3-MBP	T7/lac	Kan	N-His-MBP	3C	BR322	pace_cl_FW	pace_cl_RV
pACE3-NusA	T7/lac	Kan	N-His-NusA	3C	BR322	pace_cl_FW	pace_cl_RV
pACE3-Sumo	T7/lac	Kan	N-His-Sumo	3C	BR322	pace_cl_FW	pace_cl_RV
pACE3-TrxA	T7/lac	Kan	N-His-TrxA	3C	BR322	pace_cl_FW	pace_cl_RV
pACE3-Z2	T7/lac	Kan	N-His-Z2	3C	BR322	pace_cl_FW	pace_cl_RV

[a] The untagged construct can be generated by insertion of a stop codon with the optional primer for the insert as specified in Table 2

vectors the tags can be removed through cleavage by the 3C protease from human Rhinovirus type 14 (Cordingley et al. 1990). The vectors cloning sites were optimized for Ligation-Independent Cloning (LIC) (Aslanidis and de Jong 1990) so that PCR products can be simultaneously cloned into vectors containing different solubility tags (Table 1). In addition, the inclusion of a stuffer fragment, i.e. the coding sequence of green fluorescence protein that is disrupted during cloning, allows for negative selection of colonies and, thus, helps to distinguish between "empty" vector and plasmid containing insert. The COSS vectors can also be used for generation of co-expression plasmids by a loxP site-specific recombination using the Cre recombinase (Fitzgerald et al. 2006; Bieniossek et al. 2009).

For our automated PCR and plasmid purifications we utilize commercially available DNA purification kits on the STARlet module from Hamilton Robotics (Djinović-Carugo et al. 2013).

Protein expression screen and large-scale purification

The default vector used for our expression screening comprises an N-terminal His_6-tag followed by a 3C-protease cleavage sequence. We selected the Rosetta™ 2 (DE3) pLysS *E. coli* strain as the default strain for bacterial expression as it supplies tRNAs for 7 rare codons and also contains the pLysS plasmid that further suppresses the basal expression of T7 RNA polymerase prior to induction.

The expression screens are based on a 24- or 96-well plate format suitable for automation using the same STARlet module, where PCR and plasmid purifications take place without altering the deck layout (Djinović-Carugo and Pinotsis 2013).

For large-scale protein expression we use auto-induction media (Studier 2005), thereby reducing the volume of the liquid culture 2–3 times while maintaining the same level of biomass. For large-scale protein purification we utilize automated protocols on ÄKTAxpress™ modules. The protocols were developed for commercially available pre-packed columns and designed to eliminate user and system variations, maximize run reproducibility and standardize yield and purity between batches. Since ÄKTAxpress™ modules are essentially identical in terms of flow paths, delay volumes, gradient delay volumes, predefined column types, etc., our established protocols can be used on any ÄKTAxpress™ module without the requirement for specific system optimization. The corresponding method file can be downloaded from the Supplementary files.

Biophysical characterization

In our COSS platform the quality and suitability of the protein sample for structural or/and functional studies are

assessed by various biophysical assays. The sample monodispersity is routinely evaluated by the inspection of SEC elution profiles (Graslund et al. 2008a) obtained during large-scale purification or by dynamic light scattering (DLS). When required, analytical gel filtration linked to a multi-angle static light scattering (MALS) is performed. This is an ideal approach to directly measure the molar mass distribution of a protein in solution (Striegel and Brewer 2012). The thermal stability of samples under different buffer conditions is assessed by Fluorescence Thermal Shift Assay (Niesen et al. 2007).

Ligation independent cloning (LIC)

Materials

Reagents

- Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific, http://www.thermoscientificbio.com, cat. no. F-548S) or any other high-fidelity polymerase without non-template dependent terminal transferase activity.
- T4 DNA Polymerase (Roche, http://www.roche-applied-science.com, cat. no. 11004786001).
- Primers for amplification of insert and linearized vector, as listed in Table 2.

NOTE: All primer sequences are given as 5′ to 3′, 100 µM solutions are used throughout this protocol.

- DpnI (New England Biolabs, https://www.neb.com, cat. no. R0176S).
- dNTP set (Thermo Scientific, http://www.thermoscientificbio.com, cat. no. R0181).
- Betaine solution 5 M (Sigma, http://www.sigmaaldrich.com, cat. No. B0300).
- Chemically competent *E. coli* DH5α cells (Transformation Efficiency $>1 \times 10^6$ cfu/µg).
- SOC medium.
- LB-agar plates containing proper antibiotics.

Equipment

- 1.5 ml microfuge tubes.
- VWR Skirted 96-well PCR plate (VWR, http://www.vwr.com, cat. no. 211-0297)
- 96-well PCR plate (4titude, http://www.4ti.co.uk, cat. no. 4ti-0960/C).
- QPCR Tube Strips 0.2 ml (peqlab, http://www.peqlab.de, cat. no. 82-1502-A).
- GeneJET Gel Extraction and DNA Cleanup Micro Kit (Thermo Scientific, http://www.thermoscientificbio.com; cat. no. K0831) or any other gel extraction kit.
- PCR Clean-up NucleoSpin 96 (Macherey–Nagel, http://www.mn-net.com/, cat. no. 740658.1).
- GeneJET Plasmid Miniprep Kit (Thermo Scientific, http://www.thermoscientificbio.com cat. no. K0503).
- QIAprep 96 Plus BioRobot Kit (Qiagen, http://www.qiagen.com cat. no. 962241).
- Eppendorf Mastercycler gradient or any other gradient thermal cycler.
- Eppendorf Centrifuge 5810R or any other refrigerating centrifuge with adapters for 96-well PCR plates.
- Microfuge centrifuge (any model which accepts 1.5 ml tubes).
- Thermoblock or thermal cycler accepting 96-well PCR plates.
- NanoDrop 2000c UV–Vis Spectrophotometer (Thermo Scientific, http://www.thermoscientificbio.com) or any Spectrophotometer for low volume samples.
- Multichannel pipettes.
- Agarose Gel Electrophoresis unit including electrophoresis chamber and power supply, electrophoresis buffer, DNA ladder, reagents to prepare agarose gel; or optionally: an automated electrophoresis device, e.g. QIAxcel Advanced System (Qiagen, http://www.qiagen.com cat. no. 9001941).
- Optional: liquid handling workstations for laboratory automation solutions, which is able to process 96-PCR plates and DWPs, e.g. Microlab STARlet from Hamilton Robotics (http://www.hamiltonrobotics.com).

Procedure

LIC proceeds in three stages.

(i) PCR amplification of both vector and insert fragments with a proofreading polymerase.
(ii) T4 DNA-polymerase treatment.
(iii) Mixing and annealing of prepared fragments.

Vector preparation

- To linearize the vector for LIC, mix in a PCR tube or in a 96-well PCR plate:
  - 1 µl FW primer (see Table 2)
  - 1 µl RW primer (see Table 2)
  - 25 µl Phusion Flash Master mix
  - 22.5 µl sterile distilled H_2O
  - 0.5 µl template vector (20–200 ng μl^{-1})
- Use the following PCR cycling program for amplifying any vector referred to in Table 1:
  - Initial denaturation: 98 °C, 10 s
  - Denaturation: 98 °C, 2 s
  - Extension: 72 °C, 90 s
  - Repeat denaturation-extension steps 30 times
  - Final extension: 72 °C, 2 min
  - Hold at 4 °C

NOTE: For our LIC vectors specified in Table 1 no annealing step is required. Please refer to the manual of the

Table 2 Primers

Template to be amplified	Description	Forward (FW) or reverse (RV)	Sequence
Insert	N-terminal tagged insert without ATG at start	FW	CCAGGGGCCCGCCATG
	N-terminal tagged insert with ATG at start	FW	CCAGGGGCCCGCC
	Reverse primer for N-terminal tagged insert	RV	GACCCGACGCGGTTA
	C-terminal tagged insert without ATG at start	FW	AAGAAGGAGAACAACCATG
	C-terminal tagged insert with ATG at start	FW	AAGAAGGAGAACAACC
	Reverse primer for C-terminal tagged insert	RV	GACCCGACGCGGT
	Optional: Reverse primer for untagged insert	RV	GACCCGACGCGGTTTA
Vector	pace_cl_FW	FW	CCGCGTCGGGTCAACGCGTCTCGAGAGATCCGGC
	pace_cl_RV	RV	TGGCGGGCCCCTGGAACAGAACTTC
	pace_uncl_FW	FW	CCGCGTCGGGTCACCAC
	pace_uncl_RV	RV	TGGTTGTTCTCCTTCTTAAAGTTAAATC

DNA polymerase for further information on the protocol especially if you are using a different polymerase.

- Run the PCR products on an agarose gel to remove dNTPs and the template vector.

CRITICAL STEP: Be aware that super-coiled plasmid migrates at ~70 % of the size of a full-length linear species.

- Cut the band containing the linearized vector from the gel and extract the DNA by using a commercially available kit. Elute the purified PCR-product with 50 µl elution buffer or ultrapure water.
- Determine the concentration by measuring the absorbance at 260 nm (1 OD_{260} Unit corresponds to 50 ng μl^{-1} of DNA), with the NanoDrop 2000c UV–Vis Spectrophotometer.

TIMING: 5 h.
PAUSE POINT: The eluted linearized vector DNA can be frozen and stored indefinitely for future use at −20 °C.

- Treat the linearized vector with T4 DNA polymerase in the presence of dTTP at room temperature or 22 °C. The T4 DNA polymerase is active from the 3′ towards the 5′ end of each strand; therefore, the bases are removed from both 3′ ends until the first thymidine (T) residue is reached.
- Set up the reaction in a 96-well PCR plate on ice. Add the polymerase last:
  - 600 ng linearized vector
  - 4 µl incubation buffer, 5× concentrated
  - 0.5 µl dTTP (100 mM)
  - 1.5 µl T4 DNA polymerase (3 U/µl)
  - Add sterile distilled H_2O up to 20 µl
  - Mix by short centrifugation of the plate

CRITICAL STEP: Thaw the incubation buffer once, prepare single-use aliquots, and store them at −20 °C. Do not subject the incubation buffer and dTTP to multiple freeze–thaw cycles.

- Incubate the plate in a thermal cycler for 30 min at 22 °C and then for 20 min at 75 °C to inactivate the T4 DNA polymerase.
- Spin down the plate 10 s at 4000 rpm.

TIMING: 2 h.
PAUSE POINT: The T4 DNA polymerase treated linearized vector can be frozen and stored indefinitely for future use at −20 °C.
TROUBLE SHOOTING: Typical problems comprising both cloning and purification sections of protocol are summarized in Table 3.

Insert preparation

- To amplify the insert for LIC, mix in a 96-well PCR plate:
  - 1 µl FW primer (see Table 2)
  - 1 µl RW primer (see Table 2)
  - 12.5 µl Phusion Flash Mastermix
  - 10 µl sterile distilled H_2O
  - 0.5 µl template DNA
- Use the following thermal cycler program for amplifying inserts. Adjust the extension time as recommended in the polymerase's manual:
  - Initial denaturation: 98 °C, 10 s
  - Denaturation: 98 °C, 2 s
  - Extension: 72 °C, 20 s/1 kb
  - Repeat the denaturation–extension steps 29 times
  - Final extension: 72 °C, 2 min
  - Hold at 4 °C

Table 3 Troubleshooting

Step	Problem	Possible reason	Solution
Vector preparation	No product or low product yield	Not optimal cycling protocol	Use cycling protocol as recommended
		Template DNA may be damaged	Check your template (measure DNA concentration, run agarose gel)
PCR amplification of the insert	No product at all or low yield	High primer Tm	Use betaine or DMSO
		Not optimal cycling protocol	Use cycling protocol as recommended
			Try annealing step, optimally gradient
		Template DNA may be damaged	Check your template
Transformation	No colonies, positive transformation control negative	Wrong antibiotics	Use proper antibiotic
		Low efficiency of competent cells	Increase recovery time after transformation to 2 h. Use competent cells with higher cfu
	Positive transformation control positive, no colonies with T4-DNA polymerase treated inserts	No or low number of annealed constructs recovered	Prepare new T4-DNA Polymerase treated insert
		Colonies contain template vector	Run gel longer, use DpnI treatment, wait till GFP is expressed to distinguish right colonies (white) from wrong ones (yellow)
Large scale purification	Incomplete cleavage of the N-terminal tag	3C protease partially inactive	Confirm that the 3C protease is active
		Tag is partially buried in the structure	Change vector (including a bulky tag, or a vector with C-terminal tag)
			Use slightly denaturating conditions
	Cleaved protein strongly bound to Ni–NTA column	Unspecific interaction with matrix	Increase salt concentration, re-clone to a vector with different tag
	Protein aggregates		Buffer optimization (pH, ionic strength)

NOTE: Increasing the time of the extension step even by 100 % or more, will not affect the result of the reaction. For longer constructs (>2,000 bp) and increased time/kb may be required.

- Purify the DNA by using a commercially available PCR-purification kit. Determine the DNA concentration of purified insert.

TIMING: 4 h.

PAUSE POINT: The purified PCR product can be frozen and stored indefinitely for future use at −20 °C.

- T4 DNA polymerase treatment of the PCR amplified insert. In the next step, the purified insert is incubated with T4 DNA polymerase in the presence of dATP. Because of the $3' \rightarrow 5'$ activity of the T4 DNA polymerase the bases are removed from both $3'$ ends until the first adenine (A) residue is reached.
- Mix in a 96-well PCR plate on ice, add polymerase last.
  - 0.2 pmol insert
  - 4 µl Incubation Buffer, 5× concentrated
  - 0.5 µl dATP (100 mM)
  - 1.5 µl T4 DNA polymerase
  - Fill up with sterile distilled H_2O to 20 µl
  - Mix shortly by centrifugation of the plate.

CRITICAL STEP: Thaw the incubation buffer once, prepare single-use aliquots and store the aliquots at −20 °C. Do not subject the incubation buffer and dATP to multiple freeze–thaw cycles.

- Incubate the reaction mixture in a PCR machine for 30 min at 22 °C and then for 20 min at 75 °C to inactivate the polymerase.
- Spin 10 s at 4,000 rpm in a microfuge centrifuge.

OPTIONAL STEP: DpnI treatment (In case that PCR template and the LIC vector carry the same antibiotic resistance):

- Add 0.5 µl of DpnI per 20 µl T4-DNA polymerase treated sample
- Incubate for 2 h at 37 °C
- Deactivate DpnI for 20 min at 80 °C

TIMING: 2 or 5 h if the treatment with DpnI is required.

PAUSE POINT: The T4 polymerase treated insert can be frozen and stored indefinitely for future use at −20 °C.

Annealing of the insert and the LIC vector

- The complementary overhangs that are created in the vector and insert are long enough for a very specific, enzyme free annealing of the two DNA fragments. Mix in a 96-well PCR plate:
  - 1 µl vector (~0.007 pmol)
  - 2 µl insert DNA (0.02 pmol)
  - Incubate the annealing mixture for 5 min at 22 °C (or room temperature)

■ Mix gently by stirring the solution with the tip
■ Incubate for a further 5 min at 22 °C (or room temperature)

Transformation of E. coli DH5α competent cells with the annealed products

- Thaw the appropriate amount of *DH5α* competent cells on ice.
- Add 33 µl of competent cells to each annealing mixture.
- Incubate plate for 15 min on ice.
- Heat shock *E. coli* cells for 45 s at 42 °C in a thermoblock (or any thermal cycler accepting 96-well plates).
- Transfer the plate immediately on ice and incubate for at least 2 min.
- Add 100 µl SOC medium to each tube and incubate for 1 h at 37 °C in a shaker/incubator.
- Plate out the cell suspension on a LB agar plate containing proper antibiotic.
- Incubate plates overnight at 37 °C.

NOTE: Include as a positive control the template vector and as an optional negative control any vector with resistance different from kanamycin.

TIMING: 15 h with overnight incubation.

Identification of positive constructs

Constructs containing the desired insert can be identified by colony PCR or by preparing of small volume overnight cultures followed by isolation of plasmids by any of standard mini-prep protocols and sequencing or restriction analysis of the plasmids.

NOTE: Colonies containing vector with no insert will turn yellow after some days, when the LB agar plate is stored at 4 °C. Colonies containing vector without insert will express GFP and fluorescence can then be observed under UV light (Fig. 2).

Isolate plasmids from positive colonies and confirm the construct identity by PCR, restriction analysis or sequencing the insert.

NOTE: The plasmid purification can be performed by a liquid handling workstation, e.g. Microlab STARlet (Djinović-Carugo et al. 2013).

TIMING: 4 h for colony PCR or 20 h with overnight cultivation, conventional minipreps and PCR check of insert.

Expression and purification screening

Materials

Reagents

- Trizma® base (Tris base) (Sigma, http://www.sigmaaldrich.com, cat. no. T1503).
- Sodium chloride (Sigma, http://www.sigmaaldrich.com, cat. no. 71376).
- Glycerol (Roth, http://www.carlroth.com; cat. no. 3783.2)
- Imidazole (Sigma, http://www.sigmaaldrich.com, cat. no. 56750).
- Urea (Sigma, http://www.sigmaaldrich.com, cat. no. U5378).
- Chemically competent Rosetta 2 pLysS cells (Merck Millipore, http://www.merckmillipore.com, cat. no. 71401-3).
- SOC medium (Invitrogen, http://www.invitrogen.com; cat. no. 15544-034).
- BugBuster® 10× Protein Extraction Reagent (Merck Millipore, http://www.merckmillipore.com, cat. no. 70921).
- Benzonase® Nuclease (Merck Millipore, http://www.merckmillipore.com, cat. no. 70746).
- Ni Sepharose 6 Fast Flow (GE Healthcare, http://www.gelifesciences.com; cat. no. 17-5318-02).
- Kanamycin (Sigma, http://www.sigmaaldrich.com, cat. no. K0254).
- Chloramphenicol (Sigma, http://www.sigmaaldrich.com, cat. no. C7795).
- Triton™ X-100 (Sigma, http://www.sigmaaldrich.com, cat. no. T7878).
- NuPAGE® LDS Sample Buffer (4×) (Invitrogen, http://www.invitrogen.com; cat. no. NP0008).
- Luria Broth (LB) medium.
- Auto-induction medium ZYP5052, prepared as described in (Studier 2005).
- Antifoam 204 (Sigma, http://www.sigmaaldrich.com, cat. no. A8311).

Equipment

- Multichannel electronic pipette, matrix equalizer pipette or any equivalent multichannel pipette (Thermo Scientific, http://www.matrixtechcorp.com, cat. no. 2230 & 2034).
- Eppendorf Mastercycler gradient or any other thermal cycler.
- A 2 ml 96-deepwell plate (DWP) (Ritter, http://www.ritter-online.de, cat. no. 43001-0020).
- Adhesive film for culture plates, porous (VWR, http://www.vwr.com, cat. no. 391-1261).
- Tape taps (Qiagen, http://www.qiagen.com; cat. no. 1018104).
- Filter plate (Receiver Plate) (Macherey–Nagel, http://www.mn-net.com, cat. no. 740686.4).
- VWR skirted 96-well PCR plate (VWR, http://www.vwr.com, cat. no. 211-0297).
- Microplate 96-well, clear, tissue-culture treated, flat-bottom with lid (BD Bioscience, http://www.bdbiosciences.com, cat. no. 353072).

- Axygen 10 ml 24-deepwell plate (DWP) (Corning, http://www.corning.com; cat. no. P-DW-10ML-24-C).
- Thermoblock or thermal cycler accepting 96-PCR plates.
- Protein Gel Electrophoresis system including electrophoresis chamber and power supply, electrophoresis buffer, commercial precast gels or reagents to prepare custom SDS-PAA gels, gel staining and destaining solutions, protein ladder; or optionally an automated electrophoresis device, e.g. QIAxcel Advanced system (Qiagen, http://www.qiagen.com, cat. no. 9001941).
- Vacuum aperture that fits 96-well plates.
- Refrigerating centrifuge supporting centrifugation of plates (Eppendorf, http://www.eppendorf.at, cat. no. 5810 R) or any equivalent.
- Incubator-Shaker (Sartorius, http://www.sartorius.com, cat. no. BBI-8865221).
- Optional: liquid handling workstations for laboratory automation solutions, which is able to process DWPs, e.g. Microlab Starlet from Hamilton Robotics (http://www.hamiltonrobotics.com).

Reagent setup

- 50 % Glycerol solution: Prepare the 50 % (v/v) glycerol solution by diluting 100 % glycerol in H_2O and then autoclave.
- Optional: Antifoam 204 1 % (v/v) solution in water, store at 4 °C. *Note* it becomes cloudy at room temperature.
- Wash buffer 1: 100 mM Tris/HCl pH 8.0, 300 mM NaCl, 5 % (v/v) glycerol, 20 mM imidazole.
- Lysis buffer: 13.5 ml wash buffer 1, 1.5 ml 10× BugBuster, 7.5 µl Benzonase (0.5 µl/ml final concentration). The final volume is sufficient for 24 samples.
- Wash buffer 2: 100 mM Tris/HCl pH 8.0, 300 mM NaCl, 5 % (v/v) glycerol, 40 mM imidazole. *Note* Antifoam 204 is recommended to add to wash buffers 1 and 2 in 0.005–0.01 % concentration to prevent foaming.
- Elution buffer: 25 mM Tris/HCl, pH 8.0, 200 mM NaCl, 5 % (v/v) glycerol, 500 mM imidazole.
- Denaturation lysis buffer: 50 mM Tris/HCl, pH 8.0, 300 mM NaCl, 8.0 M urea, 20 mM imidazole, 0.5 % (v/v) Triton X-100.
- Denaturation wash buffer: 50 mM Tris/HCl pH 8.0, 300 mM NaCl, 8.0 M urea, 40 mM imidazole
- Denaturation elution buffer: 50 mM Tris/HCl pH 8.0, 8.0 M urea, 500 mM imidazole
- Ni-Sepharose 4×: Ni-Sepharose 6 FF equilibrated with wash buffer 1: To prepare 5 ml of Ni-Sepharose 4× slurry, transfer 1.25 ml settled beads in a 15 ml falcon tube. Fill up to 15 ml with wash buffer and invert for mixing. Centrifuge for 1 min at 700*g*. Discard the buffer. Repeat once and fill up to 5 ml. This volume is enough for 24 samples.

Procedure

Transformation

- Place the VWR skirted 96-well PCR plate on ice and aliquot 33 µl Rosetta 2 pLysS competent cells per well.
- Add 1 µl of plasmid (10–100 ng μl^{-1}) into each well with a multichannel pipette.
- Incubate 15 min on ice.
- Heat shock for 1 min at 42 °C on a thermoblock or thermal cycler.
- Immediately place the PCR plate on ice and incubate for 2 min.
- Add 100 µl SOC medium in each well and seal with adhesive porous film. Incubate at 37 °C for 1 h with shaking at 170–250 rpm.
- During the incubation time, fill a 2 ml-96 DWP plate with 1 ml LB medium per well supplemented with the proper antibiotics for the plasmid containing the gene of interest and for the pLysS plasmid (chloramphenicol).
- Transfer the recovered cells into the 2 ml-96 DWP plate filled with LB. Seal with adhesive porous film. Incubate overnight at 37 °C while shaking at 170–250 rpm.

Creation of glycerol stocks

- Aliquot 50 µl sterile 50 % glycerol solution in a 96-well sterile microplate. Add 50 µl of overnight culture with multichannel pipette. Mix by aspirating and dispensing. Cover the plate with adhesive seal.

TIMING: Approx. 18 h with overnight incubation.
PAUSE POINT: Freeze glycerol stocks and store at −80 °C or continue.

Preparation of small-scale expression cultures

This method is designed for a 24-DWP. It is possible to apply the protocol for a 96-DWP format provided that there is enough aeration (high speed shaker) during expression (Studier 2005). The method provides the flexibility to use a centrifuge instead if a vacuum unit is not available.

- Fill a 10 ml 24-DWP with 5 ml per well auto-induction media supplemented with the required antibiotics.

CRITICAL STEP: Note that due to high phosphate concentrations in auto-induction media the final concentration of kanamycin should be 100 µg ml^{-1} (Studier 2005).

- Inoculate with 50 µl overnight pre-culture. Seal with adhesive porous film. Incubate at 37 °C shaking at 250 rpm until OD_{600} reaches ~0.8. Then reduce temperature to 18 °C and leave shaking overnight.

NOTE: For routine work the overexpression point is typically reached after about 4 h, it is therefore possible to program the shaker to automatically decrease the temperature after that time without further checks.

- Harvest the cells by centrifugation at 3,200*g* for 3 min. Discard medium by carefully inverting the plate.

CRITICAL STEP: Do not increase speed and time during centrifugation. The short time improves significantly the solubilization of the pellet in down-stream applications while the amount of cells remaining in the supernatant after 3 min is negligible.
TIMING: Approx. 20 h including overnight cultivation.
PAUSE POINT: Freeze the pellets at −80 °C or continue.

Purification of soluble fraction

NOTE: The protocol below is developed for purification of multiple samples after small-scale expression and it can be performed either in manual mode or with help of a liquid handling workstation, e.g. Hamilton Microlab STARlet (Djinović-Carugo and Pinotsis 2013).

- Add 600 µl lysis buffer to each well in a 24-DWP and incubate on a shaker at 20 °C for 20 min (or 2 h at 4 °C) 250 rpm.

CRITICAL STEP: The cultures in the end should become viscous.

- Collect the lysates by centrifugation for 15 min at 4 °C at 4,000*g* and transfer supernatants into a 2-ml 96-DWP.

RECOMMENDED STEP: To analyze the insoluble fraction, either store the cell debris after centrifugation at −80 °C for later analyses, or continue following the protocol placed below for purification of insoluble fraction.

- Add 200 µl of Ni-Sepharose 4× slurry to the lysates and incubate for 15 min at 20 °C with shaking at 250 rpm.

CRITICAL STEP: To reproducibly transfer the Ni-Sepharose 4× resin to the lysates use pre-cut or wide-bore pipet tips and re-suspend the slurry before each pipetting step.
NOTE: The purification protocol below is recommended for work at RT.

- Transfer the lysate Ni-Sepharose mixture onto the filter plate placed on the vacuum unit.
- Optionally add 1 % antifoam 204 to a final concentration 0.01 % in each well to prevent foaming when applying vacuum.
- Apply vacuum −50 mbar for 1 min or centrifuge for 1 min at 100*g* if no vacuum unit is available.
- Add 1 ml of wash buffer, apply vacuum −50 mbar for 1 min or centrifuge for 1 min at 100*g* if no vacuum unit is available.
- Add 1 ml of extra wash buffer, apply vacuum −50 mbar for 1 min or centrifuge for 1 min at 100*g* if no vacuum unit is available. Repeat once.
- Add 150 µl of elution buffer, incubate for 2 min and elute by applying vacuum −200 mbar for 10 s or spin 2 min at 100*g* if no vacuum unit is available.
- Prepare samples for SDS-PAGE and run electrophoresis. Transfer 10 µl of each eluent from the previous step into a new 96-well PCR plate, adjust the volume by 4× SDS-PAGE sample buffer.
- Optional: prepare samples for automated electrophoresis, e.g. QIAxcel instrument, according to manufacturer's procedure and run the electrophoresis.

Purification of insoluble fraction

- Add 600 µl of denaturation lysis buffer to each well containing the remained cell debris after the purification of soluble fraction.
- Shake for 1 h at room temperature at 250 rpm.
- Add 200 µl of Ni-Sepharose 4× to the lysates and incubate for 15 min at 20 °C while shaking at 250 rpm.
- Transfer the lysate Ni-Sepharose mixture onto the filter plate already placed on the vacuum apparatus.
- Optionally add 1 % antifoam 204 to a final concentration 0.01 % in each well to prevent foaming when applying vacuum.
- Apply vacuum −50 mbar for 1 min or centrifuge for 1 min at 100*g* if no vacuum unit is available.
- Add 1 ml of denaturation lysis buffer; apply vacuum −50 mbar for 1 min or centrifuge for 1 min at 100*g* if no vacuum unit is available.
- Add 1 ml of denaturation wash buffer, apply vacuum −50 mbar for 1 min or centrifuge for 1 min at 100*g* if no vacuum unit is available. Repeat once
- Add 150 µl of denaturation elution buffer, incubate for 2 min and elute by applying vacuum −200 mbar for 10 s or centrifuge for 2 min at 100*g* if no vacuum unit is available.
- Prepare samples for SDS-PAGE. Transfer 10 µl of each eluent from the previous step into a new 96-well PCR plate and adjust the volume by 4× SDS-PAGE sample buffer.
- Optional: prepare samples for QIAxcel instrument according to manufacturer's procedure and run the electrophoresis.

TIMING: 4 h for purification of either soluble or insoluble fraction and for electrophoresis analysis.

Large scale expression

Materials

Reagents

- Auto-induction media ZYP5052, prepared as described in (Studier 2005).
- Kanamycin (Sigma, http://www.sigmaaldrich.com, cat. no. K0254).
- Chloramphenicol (Sigma, http://www.sigmaaldrich.com, cat. no. C7795).

Equipment

- 2-L baffled shaking flasks.
- Incubator-Shaker (Sartorius, http://www.sartorius.com; cat. no. BBI-8865221).
- Centrifuge Sorvall RC 6 + (Thermo Scientific, http://www.thermoscientificbio.com).

Reagent setup

- Prepare the auto-induction media according to literature (Studier 2005).
- Add 0.5 L of auto-induction medium in each 2 L baffled shaking flask.
- **NOTE:** The purification protocol described below is adjusted to approximately 2 L of expression culture, e.g. set up at least four 2 L shaking flasks with 0.5 L medium in each. This set up provides enough culture aeration required for auto-induction expression protocols.
- Add proper antibiotics. The final kanamycin concentration in auto-induction media should be 100 μg mL^{-1} (Studier 2005).

Procedure

Large scale expression

- Inoculate each flask with 1:100 overnight pre-culture.
- Grow at 37 °C until the cell growth reaches its logarithmic phase (OD_{600} about 0.8) in a shaking incubator at 170 rpm.
- Cool down the incubator at 18 °C and continue with the over-expression stage overnight.
- Harvest the cells by centrifugation at 5,000*g* at 4 °C for 10 min. Store at −20 °C (optionally at −80 °C) or continue with purification. On average 1 L of auto-induction media yields 10–25 g of wet cell pellet.

TIMING: 18 h.

Large scale purification using ÄKTAxpress™

Materials

Reagents

- Trizma® base (Sigma, http://www.sigmaaldrich.com; cat. no. 93363).
- Sodium chloride (Sigma, http://www.sigmaaldrich.com; cat. no. 71376).
- Glycerol (Roth, http://www.carlroth.com; cat. no. 3783.2).
- Imidazole (Sigma, http://www.sigmaaldrich.com; cat. no. 56750).
- Protease inhibitor complete cocktail tablets, EDTA-free (Roche, http://www.roche-applied-science.com, cat. no. 04693132001).
- Deoxyribonuclease I (DNase I, Sigma, http://www.sigmaaldrich.com, cat. no. DN25).
- GST-3C protease. Prepared according to (Walker et al. 1994).

Equipment

- Cell disruptor Sonifier® W-450 D with a flat Microtip (Branson, http://www.emerson.com, Microtip cat. no. 101-148-013).
- Centrifuge Sorvall RC 6 + (Thermo Scientific, http://www.thermoscientificbio.com).
- ÄKTAxpress chromatography system (GE Healthcare, http://www.gelifesciences.com).
- HisTrap FF crude columns 1 ml (GE Healthcare, http://www.gelifesciences.com).
- HiLoad 16/600 Superdex 75 prep grade or HiLoad 16/600 Superdex 200 prep grade column (GE Healthcare, http://www.gelifesciences.com).
- Superloop 10 ml (GE Healthcare, http://www.gelifesciences.com).
- Concentrators Vivaspin of suitable molecular weight cutoffs (Sartorius, http://www.sartorius.com/).

Reagent setup

- Lysis buffer: 100 mM Tris/HCl pH 8.0, 500 mM NaCl, 5 % (v/v) glycerol, 20 mM imidazole. Optionally add freshly each time reducing agent, 5 mM β-mercaptoethanol or 1 mM TCEP.
- Wash buffer: 20 mM Tris/HCl pH 8.0, 300 mM NaCl, 5 % (v/v) glycerol, 40 mM imidazole. Optionally add freshly each time reducing agent, 5 mM β-mercaptoethanol or 1 mM TCEP.
- Size exclusion chromatography (SEC) buffer: 20 mM Tris/HCl pH 8.0, 300 mM NaCl, 5 % (v/v) glycerol. Optionally add freshly each time reducing agent, 2 mM DTT and/or 1 mM EDTA.
- DNase I: 2 mg ml^{-1}. Dissolve 20 mg DNAseI in 10 ml of a solution containing 20 % glycerol and 75 mM NaCl.
- GST-3C protease: prepare 2 ml at a final concentration 0.6 mg ml^{-1} in a SEC buffer, assuming that a 1:50 tagged-protein to protease ratio leads to 100 % cleavage after 12 h and 60 mg of protein can be bound on two 1 ml HisTrap columns.

Equipment setup

- Attach two 1 ml HisTrap™ FF crude columns on position 1 of the designated positions for affinity purification on the ÄKTAxpress™
- Attach the HiLoad Superdex 16/600 column on column position 5 of ÄKTAxpress™ (depending on the size of the purified protein use the Superdex 75 or the Superdex 200 according to manufactures specifications).
- Attach a 10 ml Superloop™ on loop position 5 on the ÄKTAxpress™.
- Set the following buffers in the inlets:
  - A1: Load buffer
  - A3: Extra wash buffer
  - A4: SEC buffer
  - A7: Tag cleavage buffer
  - B1: Elution buffer

Procedure

Large scale purification

Large scale purification of tagged proteins was optimized for ÄKTAxpress™ chromatography modules. The purification protocol is fully automated and comprises of 3 steps: affinity chromatography, on column cleavage and size exclusion chromatography (SEC). Ideally, the protocol results a final sample in <15 h, including sample processing and column equilibration steps.

- Re-suspend the harvested cells with lysis buffer (2 ml for each 1 gram), protease inhibitor complete tablet, 5 mM $MgCl_2$ and 2 µg ml^{-1} DNase I (final concentration).
- Disrupt the cells by sonication (3 times of 2 min intervals, 50 % amplitude, 1 s Pulse On, 1 s Pulse Off).
- Separate cell debris by centrifugation at 4 °C for 20 min at 45,000*g* and collect the supernatant.
- Process the sample on an ÄKTAXpress™ module using the program method downloaded from Supplementary files.
- Collect and analyze eluted fractions by conventional SDS-PAGE or automated electrophoresis.
- Combine fractions containing the protein of interest, measure the concentration and volume, calculate the yield, and store or immediately proceed with downstreaming biophysical characterization.

TIMING: Approx. 20 h including system setup, chromatography and fraction analysis.

Biophysical characterization

Materials

Reagents

Size-exclusion chromatography coupled to multi-angle static light scattering (SEC-MALS)

- SEC-buffer (see reagent in large scale purification section)

Fluorescence thermal shift assays

- SYPRO® Orange protein gel stain 5,000× concentrate in DMSO (Invitrogen, http://www.invitrogen.com; cat. no. S-6650).
- Multiplate™ PCR Plates 96-well, white (Bio Rad, http://www.bio-rad.com; cat. no. MLP9651).
- Microseal (Bio Rad, http://www.bio-rad.com; cat. no. MSB1001).
- Buffer/additive screens; for our purposes we use our in-house screens which composition is available upon request. As substitute, commercial screens, e.g. Hampton research, Slice pH, Solubility & Stability screen, Additive screen, or customized screens can be used (Reinhard et al. 2013; Boivin et al. 2013)

Equipment

Size-exclusion chromatography combined with multi-angle static light scattering (SEC-MALS)

- Liquid chromatography system, Agilent 1260 Infinity (Agilent Technologies, http://www.chem.agilent.com/).
- Refractive index detector, Shodex RI-101 (Shodex, http://www.shodex.net/).
- Multi-angle light scattering detector, miniDawn Treos (Wyatt Technology, http://www.wyatt.eu/).
- HiLoad 10/300 Superdex 200 GL (GE Healthcare, http://www.gelifesciences.com)
- Dynamic light scattering (DLS).
- DynaPro NanoStar (Wyatt Technology, http://www.wyatt.eu/).
- Fluorescence thermal shift assays/differential scanning fluorimetry (DSF).
- Real-Time PCR (RT-PCR) Detection SystemIQ5 (Bio Rad, http://www.bio-rad.com).

Reagent setup

SEC-MALS

- Size exclusion chromatography (SEC) Buffer: 20 mM Tris/HCl pH 8.0, 300 mM NaCl, 5 % (v/v) glycerol. Optionally add freshly prepared reducing agent, 2 mM DTT and/or 1 mM EDTA.

Equipment setup

SEC-MALS

- Equilibrate the system with freshly prepared and filtered SEC-buffer over night with a flow rate of 0.5 ml min^{-1}.
- Use standard operation conditions recommended by manufacturer.

DLS

- Use standard operation conditions recommended by manufacturer.

Procedure

SEC-MALS

- Inject 50–100 µl of sample with a concentration of 2 mg ml^{-1} using autosampler.
- Run chromatography at room temperature with a flow rate of 0.5 ml min^{-1}.
- Analyze the data according to literature (Gerard et al. 2007).

DLS

- To eliminate dust and large aggregates centrifuge protein samples (0.5–2 mg ml^{-1}) for 10 min at 13,000*g* and/or filter them through a 0.1 µm cutoff membrane filter.
- Analyze the data according to literature (Gerard et al. 2007).

Fluorescence thermal shift assays:
Buffer screen:

- Prepare a protein/dye solution by mixing 420 µl protein of 0.5–2 mg ml^{-1} with 2.1 µl 5,000× SYPRO Orange and fill up with SEC buffer to 525 µl.
- Transfer 20 µl of buffer screen from the 96 deep-well block using a multichannel pipette to the PCR-plate.
- Centrifuge for 30 s at 1,500*g*. Add 5 µl of the protein/dye solution to each well.
- Seal the plate and tap to mix.
- Centrifuge for 30 s at 1,500*g*.
- Put the plate in the RT-PCR machine. The temperature is increased from 15 to 95 °C in 0.5 °C increments taking a fluorescence reading every 0.5 °C. Data analysis is performed by the CFX Manager software included with the RT-PCR machine.

Anticipated results

To establish and validate the complete workflow and test the integrity of protocols, we applied them to various fragments of the protein filamin-C. This is a large actin-binding protein expressed predominantly in muscle cells where it stabilizes the actin filament ultrastructure (Stossel et al. 2001; Razinia et al. 2012). It is a functional dimer with each subunit comprising an actin-binding domain followed by 24 immunoglobulin-like (Ig) domains.

Using the platform described above, a battery of soluble fragments encompassing the C-terminal region of filamin-C (Ig domains 16–24) was generated for biochemical and structural studies. For this purpose, 88 constructs were designed and all PCR reactions were run simultaneously, under conditions described in the protocol section. The purity and size of the generated PCR products were analyzed with the QIAxcel Advanced electrophoresis instrument (Figure S1). With exception of two PCR reactions that revealed several bands of different sizes indicating non-specificity of primer binding, all other PCR products showed the expected molecular size (Fig. 3) and were successfully cloned into the pACE3-NH vector (Table 1).

The remaining 86 positive constructs were verified by sequencing, and 22 of them were selected for small-scale expression screening. Rosetta 2 pLysS competent cells

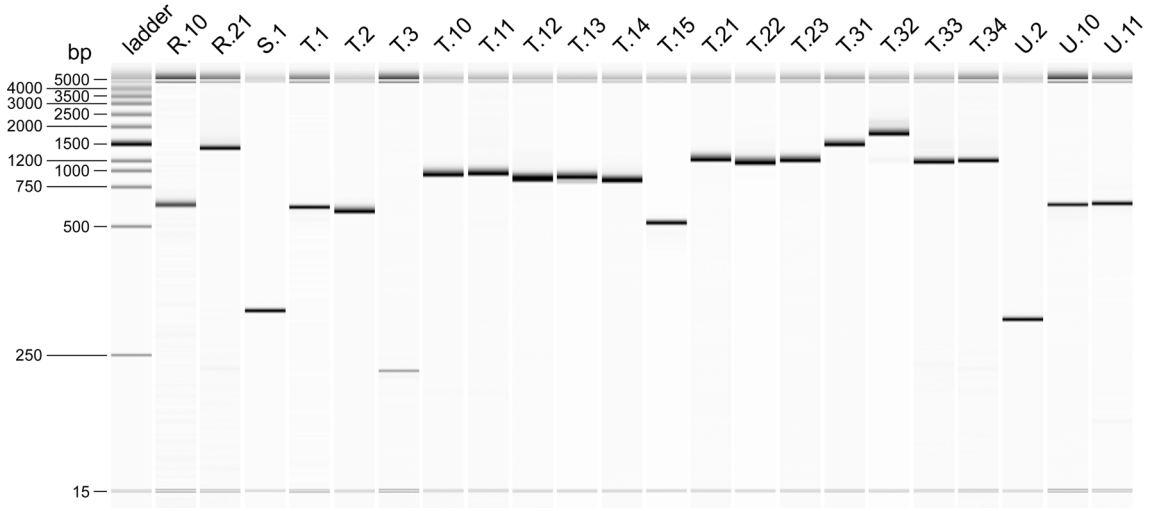

Fig. 3 Generated gel image of selected PCR products. PCR products used for cloning of the 22 selected constructs selected for soluble expression screening. Purity and size of the PCR products were assessed using the ScreenGel software. Expected sizes of PCR products are listed in Table 4

Table 4 PCR products selected for small scale expression screening of filamin-C proteins (Amino acids numbering follows the one for human filamin-C (accession number AB371585.1)

Construct name	Encoded ORF (filamin C, Ig domains)	N-terminal amino acid number	C-terminal amino acid number	Expected size of PCR product (bp)	Expected MW of expressed protein (kDa)
R.10	18–19	1952	2130	565	21.46
R.21	18–21	1973	2402	1321	48.65
S.1	19	2033	2130	325	12.77
T.1	20	2126	2306	574	22.35
T.2	20	2132	2308	562	21.89
T.3	20	2244	2306	220	9.13
T.10	20–21	2132	2402	844	31.58
T.11	20–21	2134	2406	850	31.77
T.12	20–21	2137	2401	826	30.90
T.13	20–21	2144	2402	808	30.08
T.14	20–21	2148	2401	793	29.64
T.15	20–21	2244	2406	520	19.37
T.21	20–22	2132	2497	1129	41.82
T.22	20–22	2137	2497	1114	41.22
T.23	20–22	2144	2497	1093	40.32
T.31	20–23	2132	2596	1426	51.88
T.32	20–23	2144	2596	1390	50.38
T.33	20–23	2244	2596	1090	39.27
T.34	20–23	2246	2596	1084	39.07
U.2	21	2313	2406	313	12.12
U.10	21–22	2313	2497	586	21.97
U.11	21–22	2313	2500	595	22.31

were transformed with the selected plasmids and the resulting *E. coli* cells were grown in 5 ml of auto-induction ZYP5052 media in 24-DWP plates. Cell pellets were then collected and treated according to the protocol for soluble expression screening (Fig. 3; Table 4).

The purification of the expressed proteins was performed with the Microlab STARlet liquid workstation (Djinović-Carugo and Pinotsis 2013). Ready-to-load aliquots for electrophoresis were transferred to 96-well PCR plates at the end of the purification cycle. Soluble proteins were analyzed by SDS-PAGE or by the QIAxcel Advanced electrophoresis instrument (Fig. 4).

Results of the electrophoretic analysis show that the yields of soluble proteins and their purity vary significantly between different constructs. To identify constructs suitable for large-scale production we established a scoring system for soluble expression levels. The scores are based on visual inspection of protein bands in gels of individual protein samples resulting from small-scale expression and purification steps. Briefly, we defined a rank scale from 0 to 3, where 0 stands for no-expression (or no visible product of expected size), 3 indicates a prominent major protein band of corresponding molecular weight, while 1 and 2 relate to two intermediate detectable expression levels. From the 22 filamin-C protein fragments one did not express and scored as zero (0), 17 were scored as two (2) or three (3) and 4 were scored as one (1) mainly due to low purity (Fig. 4).

Proteins with a score of 2 or 3 proceeded to large-scale expression while constructs with a score of 1 were processed with alternative protocols such as use of a C-terminal tag, or other N-terminal tags (Trx, MBP, Sumo, etc.) (Table 1) and re-screened. Taken together, 21 out of 22 of the selected targets showed variable but sufficient expression and solubility levels that are protein dependent. The used scoring system is gauged for production of samples for structural biology studies, but can naturally be customised for the requirements of a specific research question and research approach.

We routinely assessed the expression levels and purity by a combination of affinity media (e.g. Ni Sepharose 6 Fast Flow beads) used to concentrate protein and standard SDS-PAGE to control the yield and homogeneity of the protein. An automated electrophoretic system, such as the QIAxcel Advanced applied here, can reduce the analysis time, facilitate identification of the protein of interest and contribute to the evaluation of the expression level.

To test the large-scale expression and purification module, we chose 4 constructs that scored a 2 or a 3 in the

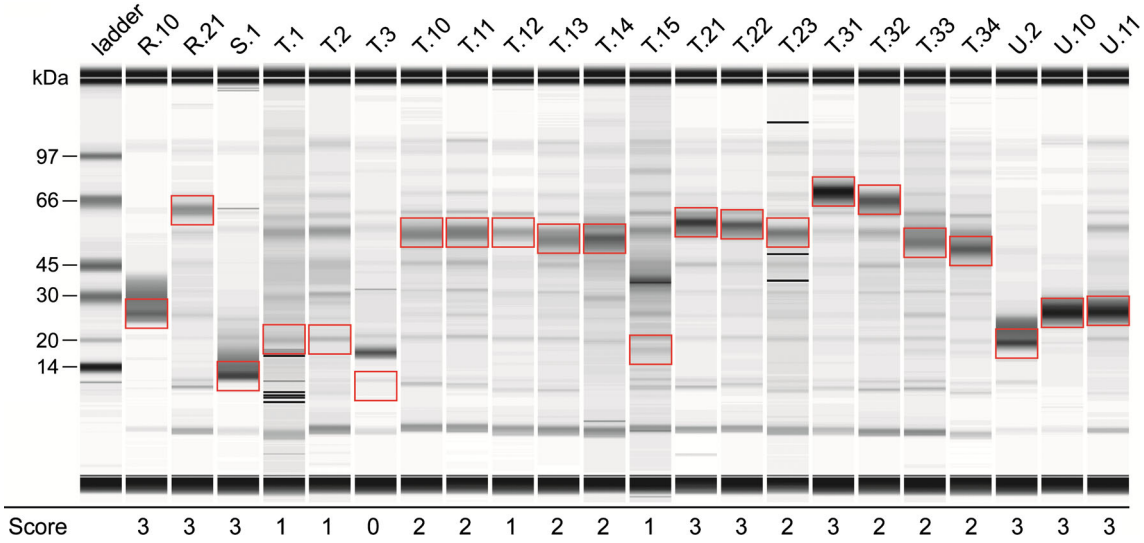

Fig. 4 Generated gel image of proteins purified after small-scale expression screening. Analysis of proteins eluted from Ni–NTA Sepharose 6 Fast Flow beads was performed with the QIAxcel Advanced instrument following the manufacturer's protocol. The gel picture was generated as described in the legend to Fig. 3. *Red boxes* indicate the expected position of the respective protein band on the gel. For further details on the expressed proteins see Table 4. Scores for soluble expression shown here were assigned as described in the results. *Note sharp thin lines* represent spikes, which are typical for capillary electrophoresis technology (for details please refer to QIAxcel System Handbook)

small-scale expression screen and subjected them to large-scale expression according to the aforementioned protocol. All proteins were well expressed and purified to over 90 % homogeneity (Fig. 5).

In our large scale protocol we accelerated the purification process by employing the AKTAxpress system as this substantially reduces the time for the entire procedure. Similar purification results can also be achieved with any other liquid chromatography system.

To check the quality of the generated protein samples and their suitability for downstream applications, including functional and structural studies, additional steps such as assays for protein activity, stability and monodispersity need to be performed (Fig. 1). Monodispersity of samples is deemed sufficient when polydispersity measured by dynamic light scattering (DLS) is below 15 %, and when the sample displays a symmetric SEC elution profile (Graslund et al. 2008a), with expected molecular weight as measured by size exclusion chromatography coupled to multi-angle static light scattering (SEC-MALS). In the fluorescence thermal shift assays different buffer compositions were screened to identify conditions in which the sample was most stable (Dupeux et al. 2011; Reinhard et al. 2013; Boivin et al. 2013). These conditions can be subsequently used in large-scale protein production before further structural and/or biochemical studies (Fig. 1).

The filamin-C constructs purified at large scale were further characterized by a combination of biophysical methods, including DLS, SEC-MALS and differential

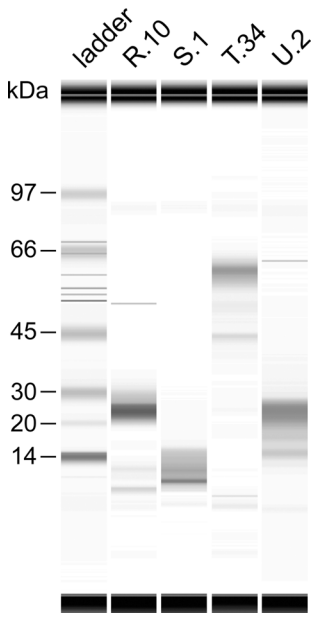

Fig. 5 Generated gel image of proteins purified according to the large-scale expression and purification protocol. Figure generation and analysis of peak fractions eluted from the gel filtration column was performed as described in legends to Fig. 3. For further details, see Table 4

scanning fluorimetry (DSF). Results of these analyses for each of four filamin-C constructs are summarized in Table 5. As shown in Fig. 6, all proteins displayed symmetric SEC elution profiles when analyzed by SEC-MALS.

Table 5 Outlook of large-scale purification and biophysical characterization of filamin-C constructs

Expressed filamin C domain (Construct)	Yield of purified protein from 2 liters of ZYP5052 medium (mg)	Expected MW of purified protein (kDa)	Multi-angle static light scattering		Dynamic light scattering Dispersity (%)	Differential scanning fluorimetry	
			MW (kDa)	Polydispersity index[a]		In standard buffer (Tm, °C)	In optimized conditions (Tm, °C)
Domains 18–19 (R.10)	42	19.43	18.83 ± 0.04	1.055	21.3 ± 1.2	67.5 ± 0.4	73.3 ± 0.4 Sodium phosphate, pH 6.0
Domain 19 (S.1)	31	10.74	11.04 ± 0.02	1.020	43.8 ± 2.2	69.2 ± 0.2	79.5 ± 0.1 Hepes, pH 7.0
Domains 20–23 (T.34)	7	37.04	38.68 ± 0.15	1.099	25.8 ± 2.3	52.8 ± 1.2	59.0 ± 0.4 Glycine, pH 10.5
Domain 21 (U.2)	28	10.10	10.16 ± 0.01	1.002	10.9 ± 0.4	49.2 ± 2.5	70.7 ± 1.2 Sodium phosphate, pH 6.5

M_n the number average molecular weight is the statistical average molecular weight of all the polymer chains in the sample and is defined by: $M_n = \sum N_i M_i / \sum N_i$, where M_i is the molecular weight of a chain and N_i is the number of chains of that molecular weight. M_w is the weight average molecular weight, defined by: $M_w = \sum N_i M_i^2 / \sum N_i M_i$, and it is directly measured by SLS. A monodisperse polymer where all the chain lengths are equal (such as a protein) has an $M_w/M_n = 1$. For the proteins routinely tested in the COSS we assumed that the PD value which is lower than 1.020 indicates a monodisperse sample

[a] Polydispersity index, PD = M_w/M_n

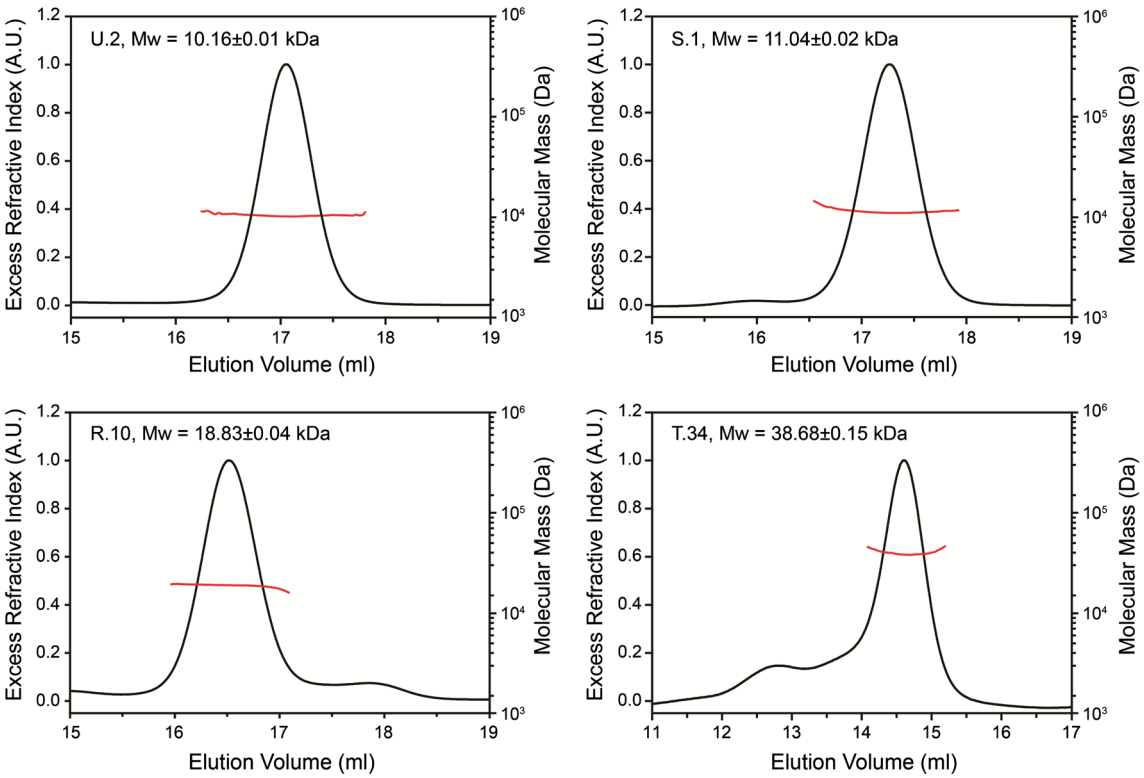

Fig. 6 SEC-MALS elution profiles of filamin-C fragments after large-scale purification. The protein samples were separated on a size exclusion chromatography (SEC) column (HiLoad 10/300 Superdex 200 GL) equilibrated with standard SEC-buffer. The elution profiles are normalized by sample concentration. The *red lines* crossing the SEC profile indicate molecular masses of the proteins determined by connected in-line multi-angle light scattering (MALS) and refractive index (RI) detectors

Furthermore, the molecular masses of these proteins as determined by the MALS analysis (for the protein eluted in the main peak) were in a very good agreement with predicted molecular weight (Table 5; Fig. 6). Examination of the protein polydispersity by both MALS and DLS showed that sample encompassing domain 19 is truly monodisperse, based on aforementioned criteria. The increased polydispersity as observed in other purified domains typically indicates aggregation or lower solubility of the protein. Albeit only one sample (domain 19) displayed polydispersity indicating it is likely to crystallized (Graslund et al. 2008a), other samples have polydispersity in moderate range (measured by DLS < 30 %), that renders them suitable to biochemical and biophysical studies. Additionally, the DSF assay with buffer screen identified buffering molecule, as well as pH, which notably improve thermals stability of samples (from 5.8 to 21.5 °C). These conditions could be employed in further optimization of purification/production protocols (Table 5).

In conclusion, our protein production pipeline allows for the preparation and small-scale expression screening of 20–100 different constructs within 1 week. One additional week is required to process 4–6 selected constructs at large scale. Overall, we successfully implemented the pipeline in which we cloned 30 different targets, which resulted in 383 cloned constructs. All cloned constructs were subjected to small-scale expression screening. 47 selected constructs with confirmed soluble expression were subsequently purified on a large scale and characterized by a set of biophysical methods. Typical yield for most of our targets was 20–30 mg of pure, untagged protein prepared from 2 L of auto-induction medium, the amount being dependent on the resin used and the specific target (Table 5). We used ligation independent cloning, which is less expensive and faster compared to conventional restriction/ligation methods. Robot-assisted small-scale expression screening proved to be an efficient way of identifying promising leads for large scale production. The protocols for large-scale purification developed for the ÄKTAxpress allow one to obtain pure, untagged protein sample via a three-step, completely automated and integrated purification run. Finally, biophysical characterisation of generated samples gives valuable information on their suitability for subsequent functional or/and structural studies. This pipeline was shown to be effective and flexible, and it can be easily adopted partially or entirely in different research laboratories in a manual or robot-assisted manner.

Acknowledgments This project was supported by the Federal Ministry of Economy, Family and Youth through the initiative "Laura Bassi Centres of Expertise", funding Center of Optimized Structural Studies, project Number 253275; JK was also supported by the Austrian Science Fund (FWF) Project P22276. FWF Doctoral program BioToP-Molecular Technology of Proteins (W1224) is acknowledged to providing training for GM. We thank Dr. Nathalie Landstetter at QIAGEN for generously providing the use of the instrument and valuable suggestions and help with experiments.

Conflict of interest The authors declare that they have no conflict of interest.

References

Aslanidis C, de Jong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res 18(20):6069–6074

Bieniossek C, Nie Y, Frey D, Olieric N, Schaffitzel C, Collinson I, Romier C, Berger P, Richmond TJ, Steinmetz MO, Berger I (2009) Automated unrestricted multigene recombineering for multiprotein complex production. Nat Methods 6(6):447–450

Boivin S, Kozak S, Meijers R (2013) Optimization of protein purification and characterization using Thermofluor screens. Protein Expr Purif 91(2):192–206

Cordingley MG, Callahan PL, Sardana VV, Garsky VM, Colonno RJ (1990) Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. J Biol Chem 265(16):9062–9065

Dahlroth SL, Nordlund P, Cornvik T (2006) Colony filtration blotting for screening soluble expression in *Escherichia coli*. Nat Protoc 1(1):253–258

Djinović-Carugo K, Pinotsis N (2013) Automated protein expression and subsequent protein purification screen. http://www.mynewsdesk.com/de/pressroom/hamilton-robotics/pressrelease/view/automatisiertes-screening-von-small-scale-proteinen-906051

Djinović-Carugo K, Pinotsis N, Charnagalov A, Kostan J, Mlynek G (2013) Automation of the Macherey-Nagel NucleoFast PCR Cleanup Kit and the QIAprep® 96Plus Biorobot® on a Hamilton Microlab STARlet Liquid Handler. http://www.mynewsdesk.com/de/pressroom/hamilton-robotics/pressrelease/view/automation-des-macherey-nagel-nucleofast-pcr-cleanup-kits-und-des-qiaprep-r-96plus-biorobot-r-911245

Dupeux F, Rower M, Seroul G, Blot D, Marquez JA (2011) A thermal stability assay can help to estimate the crystallization likelihood of biological samples. Acta Crystallogr D Biol Crystallogr 67(Pt 11):915–919

Festa F, Steel J, Bian X, Labaer J (2013) High-throughput cloning and expression library creation for functional proteomics. Proteomics 13(9):1381–1399

Fitzgerald DJ, Berger P, Schaffitzel C, Yamada K, Richmond TJ, Berger I (2006) Protein complex expression by using multigene baculoviral vectors. Nat Methods 3(12):1021–1032

Gerard FC, Ribeiro Ede A Jr, Albertini AA, Gutsche I, Zaccai G, Ruigrok RW, Jamin M (2007) Unphosphorylated rhabdoviridae phosphoproteins form elongated dimers in solution. Biochemistry 46(36):10328–10338

Graslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park HW, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim SH, Rao Z, Shi Y, Terwilliger TC, Kim CY, Hung LW, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, Prilusky J, Sussman JL, Stevens RC, Lesley SA, Wilson IA, Joachimiak A, Collart F, Dementieva I, Donnelly MI, Eschenfeldt WH, Kim Y, Stols L, Wu R, Zhou M, Burley SK, Emtage JS, Sauder JM, Thompson D, Bain K, Luz J, Gheyi T, Zhang F, Atwell S, Almo SC, Bonanno JB, Fiser A, Swaminathan S, Studier FW, Chance MR, Sali A, Acton TB, Xiao R, Zhao L, Ma LC, Hunt JF, Tong L, Cunningham K, Inouye M, Anderson S, Janjua H, Shastry R, Ho CK, Wang D, Wang H, Jiang M, Montelione GT, Stuart DI, Owens RJ, Daenke

S, Schutz A, Heinemann U, Yokoyama S, Bussow K, Gunsalus KC (2008a) Protein production and purification. Nat Methods 5(2):135–146

Graslund S, Sagemark J, Berglund H, Dahlgren LG, Flores A, Hammarstrom M, Johansson I, Kotenyova T, Nilsson M, Nordlund P, Weigelt J (2008b) The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. Protein Expr Purif 58(2):210–221

Ishida T, Kinoshita K (2008) Prediction of disordered regions in proteins based on the meta approach. Bioinformatics 24(11):1344–1348

Mooij WT, Mitsiki E, Perrakis A (2009) ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments. Nucleic Acids Res 37((Web Server issue)):W402–W405

Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2(9):2212–2221

Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, Sokolov A, Graim K, Funk C, Verspoor K, Ben-Hur A, Pandey G, Yunes JM, Talwalkar AS, Repo S, Souza ML, Piovesan D, Casadio R, Wang Z, Cheng J, Fang H, Gough J, Koskinen P, Toronen P, Nokso-Koivisto J, Holm L, Cozzetto D, Buchan DW, Bryson K, Jones DT, Limaye B, Inamdar H, Datta A, Manjari SK, Joshi R, Chitale M, Kihara D, Lisewski AM, Erdin S, Venner E, Lichtarge O, Rentzsch R, Yang H, Romero AE, Bhat P, Paccanaro A, Hamp T, Kassner R, Seemayer S, Vicedo E, Schaefer C, Achten D, Auer F, Boehm A, Braun T, Hecht M, Heron M, Honigschmid P, Hopf TA, Kaufmann S, Kiening M, Krompass D, Landerer C, Mahlich Y, Roos M, Bjorne J, Salakoski T, Wong A, Shatkay H, Gatzmann F, Sommer I, Wass MN, Sternberg MJ, Skunca N, Supek F, Bosnjak M, Panov P, Dzeroski S, Smuc T, Kourmpetis YA, van Dijk AD, ter Braak CJ, Zhou Y, Gong Q, Dong X, Tian W, Falda M, Fontana P, Lavezzo E, Di Camillo B, Toppo S, Lan L, Djuric N, Guo Y, Vucetic S, Bairoch A, Linial M, Babbitt PC, Brenner SE, Orengo C, Rost B, Mooney SD, Friedberg I (2013) A large-scale evaluation of computational protein function prediction. Nat Methods 10(3):221–227

Razinia Z, Makela T, Ylanne J, Calderwood DA (2012) Filamins in mechanosensing and signaling. Ann Rev Biophys 41:227–246

Reinhard L, Mayerhofer H, Geerlof A, Mueller-Dieckmann J, Weiss MS (2013) Optimization of protein buffer cocktails using Thermofluor. Acta Crystallogr Sect F Struct Biol Cryst Commun 69(Pt 2):209–214

Stossel TP, Condeelis J, Cooley L, Hartwig JH, Noegel A, Schleicher M, Shapiro SS (2001) Filamins as integrators of cell mechanics and signalling. Nat Rev Mol Cell Biol 2(2):138–145

Striegel AM, Brewer AK (2012) Hydrodynamic chromatography. Ann Rev Anal Chem 5:15–34

Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41(1):207–234

Vincentelli R, Bignon C, Gruez A, Canaan S, Sulzenbacher G, Tegoni M, Campanacci V, Cambillau C (2003) Medium-scale structural genomics: strategies for protein expression and crystallization. Acc Chem Res 36(3):165–172

Vincentelli R, Cimino A, Geerlof A, Kubo A, Satou Y, Cambillau C (2011) High-throughput protein expression screening and purification in *Escherichia coli*. Methods 55(1):65–72

Walker PA, Leong LE, Ng PW, Tan SH, Waller S, Murphy D, Porter AG (1994) Efficient and rapid affinity purification of proteins using recombinant fusion proteases. Biotechnology 12(6):601–605