SUPPLEMENTARY METHODS

Links to on-line protocols:
http://strgen.org/protocols/
http://www.nesg.org/nesg.swf?page=sops
http://www.thesgc.com/SGC-WebPages/toronto-technology.php
http://www.sgc.ox.ac.uk/emethods

Detailed protocols:
1. Designing the cDNA constructs
2. Cloning method (LIC)
3. Expression screening in 96-well format
4. Large-scale protein expression
5. Large-scale purification
6. Protein characterization

Additional protocols:
A. Chemical induction of chaperones
B. Co-expression with plasmid-encoded chaperones
C. Co-expression of interacting proteins

1. Designing the cDNA constructs

Full-length proteins often contain regions that inhibit the recovery of soluble recombinant protein. These may include flexible or disordered segments, as well as exposed hydrophobic patches. Expression of these proteins is greatly enhanced by removing such regions by recombinant construct design. The construct boundaries are hard to predict precisely, so a mixture of domain boundary prediction, disordered area prediction, and empirical testing is often used. While no precise protocol can be defined, data are usually integrated from the following general sources, together with literature on the protein of interest or on related proteins. The first stage is to collect data on related sequences and structures, and then attempt to predict structural features.

1. Conserved domain search: (for rough boundaries of conserved domains)
   http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi (incorporated in NCBI-blast searches)
   http://www.sanger.ac.uk/Software/Pfam/
   http://smart.embl-heidelberg.de/
2. 3D structure prediction from structures of homologous proteins in the PDB:
   http://www.rcsb.org/pdb/
   http://bioinf.cs.ucl.ac.uk/psipred/ (mGenTHREADER)
3. Secondary structure prediction (if there is no structural homologue):
   http://bioinf.cs.ucl.ac.uk/psipred/ (PSIPRED)
   http://www.compbio.dundee.ac.uk/~www-jpred/submit.html
4. Transmembrane helix prediction:
   http://www.ch.embnet.org/software/TMPRED_form.html
5. Predicting unfolded regions:
   http://bioinf.cs.ucl.ac.uk/disopred/
   http://bioportal.weizmann.ac.il/fldbin/findex
   http://www.strubi.ox.ac.uk/RONN

http://sbcweb.pdc.kth.se/cgi-bin/maccallr/disorder/submit.pl
http://globplot.embl.de/

For proteins with close structural homologues, structural templates can be found using a simple BLAST search of the Protein DataBank (PDB). However, BLAST tends to trim the alignment (it is a local alignment tool not a global alignment tool) and does not provide any secondary structure prediction. mGenthreader, on the other hand, provides the top 10 alignments with representative PDB templates/entries and the results show the actual secondary structure from the PDB aligned with the predicted secondary structure of the target sequence. In our hands, the confidence score of the mGenThreader results gives a good indicator of alignment quality. CERT and HIGH confidence alignments are generally trustworthy while MEDIUM, LOW and GUESS alignments are generally not trustworthy.

When inheriting the suggested start and stop positions from a PDB template, it is important to assess the quality of the alignment between the two sequence entities. The query sequence must be aligned to a protein that has the same overall fold or topology based on the arrangement of predicted secondary structures. The confidence in the predicted secondary structure can be increased if the secondary structure elements at least approximately match the elements in the PDB template.

The mGenThreader results enable the user to visually assess the quality of the alignment; the user must exercise some judgment in cases of marginal quality.

If there are no suitable PDB templates available then the designing the approximate N- and C-terminal boundaries becomes more difficult. The secondary structure prediction from PSI-PRED must be examined and 'structured regions' identified, i.e. regions that do not contain long predicted coil regions. Predicted unfolded regions of the protein should be avoided, as well as low-complexity sequences. However, none of these methods is sufficient; a more extensive series of deletion constructs may be needed to identify soluble fragments experimentally. Valuable information can also be obtained from experimental domain boundary analysis (e.g. partial proteolysis) of the full-length protein expressed in eukaryotic systems or extracted from natural sources.

It is typical to select 3 to 5 start points and 3 to 5 stop points based on the query-template alignment. These start and stop points should almost always be made within "coil" regions or at the very ends of predicted secondary structure elements, unless there is compelling evidence to support a non-coil start/stop point. Such evidence could be direct experience of successful constructs that are soluble and form crystals from such a non-orthodox approach. Try to avoid including too many bulky hydrophobic residues (F, W or Y) near the N or C-terminal regions as they may cause aggregation. The start/stop points should also be outside of the 'core' secondary structure elements. This strategy generates between 9 and 25 constructs for each experiment.

2. Cloning method (LIC)

Primer design for LIC:

Once the construct boundaries have been selected, standard algorithms can be used to design oligonucleotide primers to amplify the selected fragments from a cDNA clone. In primer selection, it is important to avoid self- or cross-hybridization of the primer 3' ends, and it is desirable to have similar annealing temperatures for both primers. Other parameters, such as length or GC content, are less important.

Ligation-independent cloning (LIC) of PCR fragments relies on addition of short 5' extensions to the PCR primers that match sequences in the vector. When using InFusion cloning, these extensions must be at least 15 nucleotides long, matching the last 15 nucleotides next to the cleavage/cloning site in the vector (see also http://www.thesgc.com/SGC-WebPages/toronto-technology.php).

For T4 DNA polymerase-based LIC, the 5' extensions on each primer (of length 12-15 bases) are designed so that the sequence is composed of only three nucleotides, lacking the fourth. For example, when using the plasmid pNIC28-Bsa4 (SGC, GenBank accession EF198106), the following extensions are used in the primers:
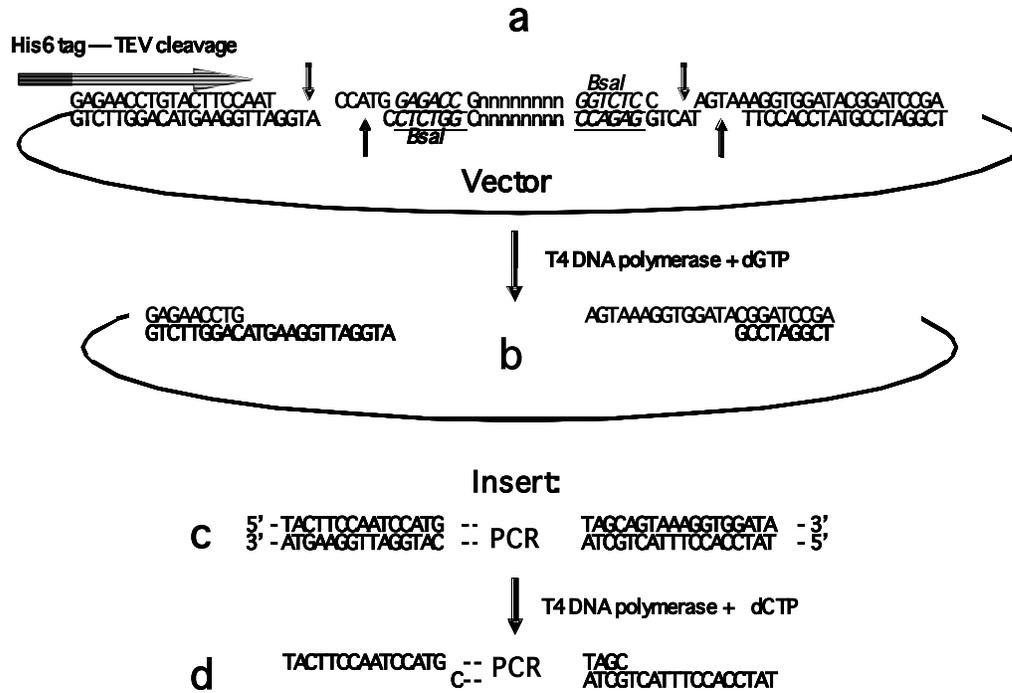Upstream primer: TACTTCCAATCC**ATG** (the ATG encodes methionine in-frame)
Downstream primer: TATCCACCTTTACT**G**TCA (the TCA is a stop codon in reverse).
In both cases, treating the purified PCR fragments with T4 DNA polymerase in presence of dCTP will result in 3'-5' exonuclease activity, which will degrade the opposing strand down to the first C, which is opposite the first G (bold) in the primer sequence. The resulting single-stranded extensions precisely match such single stranded extensions in the vector. The 3'-exonuclease activity of T4 DNA polymerase is thus used to expose complementary single-stranded stretches on the vector and inserts. The treated vector and inserts are annealed and introduced directly into bacteria.

The generic cloning vector pNIC28-Bsa4 is based on a lac-regulated T7 promoter, driving expression of an amino-terminal fusion peptide including a 6-histidine tag and a TEV protease cleavage site. The cloning process consists of the following steps (more details can be found in Stols *et al*[1]):

1. Amplification of inserts by PCR, using a proofreading DNA polymerase (Platinum Pfx). The primers include the following extensions: upstream, TACTTCCAATCC<u>ATG</u> and downstream, TATCCACCTTTACTG<u>TCA</u>.
2. PCR fragments are analyzed by gel electrophoresis and purified using a PCR purification kit (primer and nucleotide removal). Failed PCRs are repeated under alternative conditions if necessary (using PCR enhancers like DMSO; varying annealing temperature, template and Mg^{2+}-concentration, etc.).
3. Treat PCR fragments with T4 DNA polymerase and dCTP.
4. Cut 5 μg of the vector with *Bsa*I, purify, then treat with T4 DNA polymerase and dGTP. This amount of vector is sufficient to create expression vectors for at least 96 constructs.
5. Combine the treated vector and inserts, anneal for 10 minutes at room temperature, then transform bacteria.
6. The next day, pick 2 colonies from each transformation, analyze by PCR with vector-specific primers, and check carefully the size of the resultant fragments.
7. Make minipreps of the positive clones. At least some of the clones should be sequenced.

a

His6 tag — TEV cleavage

GAGAACCTGTACTTCCAAT CCATG *GAGACC* Gnnnnnnnn *GGTCTC* C AGTAAAGGTGGATACGGATCCGA
GTCTTGGACATGAAGGTTAGGTA *CCTCTGG* Cnnnnnnnn *CCAGAG* GTCAT TTCCACCTATGCCTAGGCT

BsaI

BsaI

Vector

T4 DNA polymerase + dGTP

b

GAGAACCTG   AGTAAAGGTGGATACGGATCCGA
GTCTTGGACATGAAGGTTAGGTA   GCCTAGGCT

Insert

c 5'-TACTTCCAATCCATG -- PCR TAGCAGTAAAGGTGGATA -3'
3'-ATGAAGGTTAGGTAC --   ATCGTCATTTCCACCTAT -5'

T4 DNA polymerase + dCTP

d TACTTCCAATCCATG -- PCR TAGC
C--   ATCGTCATTTCCACCTAT

Schematic view of LIC cloning.

(a) The cloning vector pNIC28-Bsa4 is cleaved by the non-palindromic enzyme BsaI. (b) The resulting ends are trimmed by T4 DNA polymerase and dGTP. (c) PCR fragments are generated with the indicated extensions. (d) The PCR fragments are trimmed by T4 DNA polymerase and dCTP, generating single-stranded tails complementary to the vector tails.

Supplementary Table 1 – Comparison of different cloning methods

	LIC (T4 DNA pol)	LIC (InFusion®)	Gateway®	Restriction cloning
Single procedure for any target	Yes	Yes	Yes	No
Sequences added to target	Flexible (restriction on nucleotide composition of LIC area)	Completely flexible	Not flexible (*att* sequences added at both ends)	At most two amino acids at each end
Transfer to different vectors	Independent PCR cloning into separate vectors	Independent PCR cloning into separate vectors. (Transfer to recipient vectors using cre recombinase is possible with the appropriate vectors)	Easy, non-mutagenic transfer from initial PCR clone to acceptor vectors (lambda integrase)	Restriction-based transfer to compatible acceptor vectors
Cost	Lowest	Highest	Highest	Medium

3. Expression screening in 96-well format

Many constructs are screened in parallel to identify the one(s) that produce the most soluble protein, time in order to select candidates for scale-up production. The protocol is designed for a 96-well format using a culture volume of 1 mL in a deep-well plate.

Small-scale expression in 96-well format
1. Transform the DNA constructs to the appropriate expression strain and plate on small (60 mm) agar plates.
2. Start cultures. Prepare a 96-well deep-well (2 mL) plate with 1 mL medium in each well. Pick 5-10 colonies and inoculate the corresponding plate position in the deep well plate. Seal the culture plate with an air-porous plate seal and place it in the shaker at 37ºC, >500 rpm. Let the cultures grow for approximately 3 hours (not to saturation).
3. Expression cultures. Measure the OD_{600} of your starter cultures in a plate reader and make dilutions with fresh TB into a new deep well plate so that all cultures have an OD_{600} of approximately 0.2 (in the range 0.1 to 0.4).
4. Save at least 100 μL of the starter culture and add sterile glycerol to a final conc. of 25% (v/v) to make glycerol stocks.
5. Incubate the expression cultures at 37ºC >500 rpm for an additional 1.5 hours, then reduce temperature to 18ºC.
6. Induce expression by adding IPTG to a final concentration of 0.5 mM and incubate at 18 ºC overnight with shaking.
7. Take out aliquot of the cells to measure total expression. Mix 6 μl of the over-night cultures with 2 μl of a 4xSDS-PAGE sample buffer. Seal and heat at 99 ºC for 10-15 min.
8. Harvest the rest of the cells by centrifugation at 4000 x *g* for 15 min, pour off the culture medium and freeze the cell pellets at -80ºC.

Purification in 96-well format
1. Prepare lysis buffer in a 50 mL Falcon tube:

25 mL	Lysis Buffer (100 mM HEPES, 10 mM imidazole, 500 mM NaCl, and 10% (v/v) glycerol – pH 8.0)
2.5 mL	Lysozyme (10 mg/mL)
1 mL	Protease Inhibitor mix (Complete EDTA-free, 5 tablets dissolved in 10 mL ddH$_2$O)
250 μL	DDM, *n*-Dodecyl ß-D-maltoside (10% (v/v) solution)
125 μL	MgSO$_4$ (200 mM)
12.5 μL	Benzonase (250 U/μL)
12.5 μL	TCEP (1 M)

2. Add 250 μL Lysis buffer to each frozen cell pellet and thaw for 5 min in a room temperature water bath. Vortex or shake until all cell pellets are resuspended.
3. Freeze at -80ºC until frozen, usually about 15-20 min if using a 96-well plate.
4. Thaw the samples as in step 2, re-freeze and thaw again.
5. Centrifuge the samples at 4000 x *g* for 10 min and transfer the supernatants to a 0.65 μm filter plate (Millipore Multiscreen HTS) on top of a sterile collection plate. Centrifuge at 2000 x *g* for 20 min.
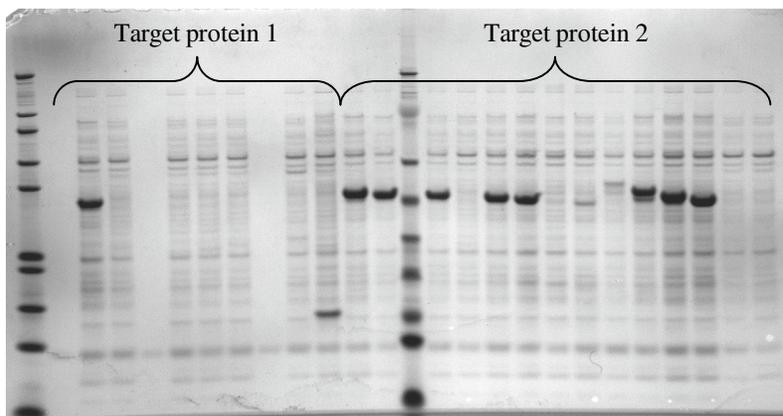
6. Dispense 40 μL of a 50% (v/v) slurry of Ni^{2+}-loaded chelating Sepharose Fast Flow resin into each well in a 1.2 μm filter plate (Millipore Multiscreen HTS) on top of a waste collection plate.
7. Add the clarified lysate and incubate the plate with gentle mixing for 15-20 min. Centrifuge at 100 x g for 1 min.
8. Discard the flow-through. Add 600 μL of wash buffer (20 mM HEPES, 500 mM NaCl, 25 mM imidazole, 10% (v/v) glycerol and 0.5 mM TCEP – pH 8.0). Mix gently for 2 min, then centrifuge 100 x g for 1 min. Repeat the wash step twice.
9. Place the filter plate on top of a sterile collection plate. Add 40 μL of elution buffer (20 mM HEPES, 500 mM NaCl, 500 mM imidazole, 10% (v/v) glycerol and 0.5 mM TCEP). Incubate with gentle mixing for 20 min and then centrifuge at 300 x g for 5 min to collect the eluate.

SDS-PAGE analysis (Novex NuPAGE 4-12% Bis-Tris, 26 well Invitrogen)
1. Load 3 μL of the gel samples with total expression on gel.
2. Prepare gel samples of purified soluble protein by mixing 4 μL of 4x SDS-PAGE sample buffer with 12 μL of the collected eluates. Load 16 μL of the gel samples with purified protein on gel.
.
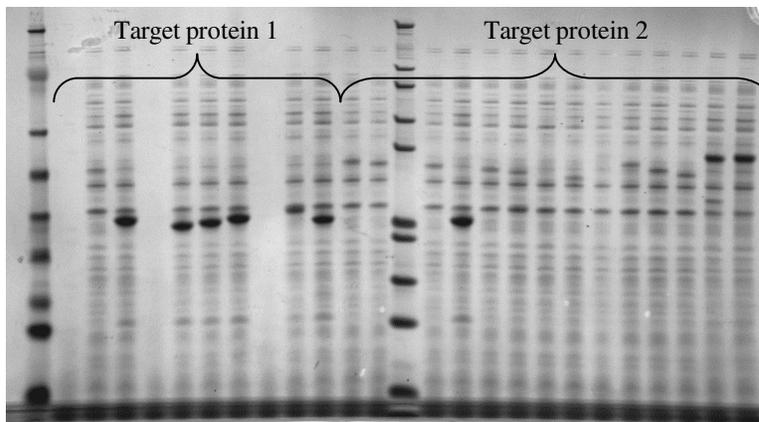
Supplementary Figure 1. Small-scale total expression
Samples from total expression of 24 constructs from two target proteins on a Novex BisTris 4-12% NuPAGE gel, 26 wells (Invitrogen). Lane 1 contains size marker SeeBlue plus2 (Invitrogen) and lane 14 contains Mark 12 (Invitrogen).

4. Large-scale protein expression

1. Inoculate starter cultures of 20 ml (~1/100 of the final cultivation volume) Terrific Broth supplemented with 8 g/L glycerol and appropriate antibiotics, with several fresh colonies from an agar plate or with cells from a frozen glycerol stock.
2. Incubate overnight at 30°C at 175 rpm.
3. The following morning, inoculate 10 ml of the starter culture into 750 ml of Terrific Broth in a 2.5 L TunAir shake flask.
4. Incubate at 37°C at 175 rpm (Infors Unitron Incubator) until an optical density at 600 nm (OD_{600}) of 1-1.5 is reached.
5. Lower the temperature of the cultures to 18°C for 1h.
6. At OD_{600} approximately 3, induce expression by adding 375 μl 1 M IPTG (final conc. of 0.5 mM).
7. Continue expression overnight at 18°C, giving a total expression time of approx. 20 h and final OD_{600} values of around 10.
8. Harvest by centrifugation at 4500 x *g* for 10 min.

5. Large-scale purification

A. Materials and stock solutions.
0.5 M Sodium phosphate, pH 8.0
5.0 M NaCl
60% Glycerol
500 mM HEPES, pH 7.5
3.0 M Imidazole, pH 7.5
0.5 M TCEP
50% (w/v) Polyethyleneimine (Sigma)
50 mM $NiSO_4$, 50 mM Tris-HCl, pH 7.5
100 mM EDTA, pH 8.0
0.5 M NaOH
Protease inhibitor cocktail (Calbiochem)
Benzonase nuclease (250 units/ul, Merck)
NiNTA resin (Qiagen) OR HisTrap FF crude column, 1-ml OR 5-ml (GE Healthcare).
TEV protease, (recombinant, hexahistidine-tagged).

B. Working solutions:
1. **2x lysis buffer:** 0.1 M sodium phosphate, pH 8.0, 1.0 M NaCl, 20 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP, 2 x protease inhibitors (500x dilution of stock), and 20 units/ml Benzonase™.
2. **5% (v/v) Polyethyleneimine (PEI)**, diluted from 50 stock and adjusted to pH 7.5 with HCl.
3. **Affinity loading buffer:** 0.05 M sodium phosphate, pH 8.0, 0.5 M NaCl, 10 mM imidazole, 5% (v/v) glycerol, 0.5 mM TCEP.

4. **Affinity wash buffer:** 0.05 M sodium phosphate, pH 8.0, 0.5 M NaCl, 30 mM imidazole, 5% (v/v) glycerol, 0.5 mM TCEP.
5. **Affinity elution buffer:** 0.05 M sodium phosphate, pH 8.0, 0.5 M NaCl, 0.25 M imidazole, 5% (v/v) glycerol, 0.5 mM TCEP.
6. **Gel filtration buffer:** 10 mM HEPES, pH 7.5, 0.5 M NaCl, 5% glycerol, 1 mM TCEP.

C. Procedure.
1. Thaw the cell pellets, if frozen. Weigh the cells, and add 1 volume (v/w) of 2x lysis buffer. All purification steps are done at 4-7ºC.
2. Cell lysis can proceed by one of the following:
   ∞ High-pressure homogenization (French press, Fluidizer): Dilute the cell lysate with 2-3 volumes of affinity loading buffer, then disrupt the cells according to the manufacturer's instructions.
   ∞ Freeze-thaw: Add to the cell suspension 1 mg/ml lysozyme from a freshly-prepared solution of 10 mg/ml in affinity loading buffer. Freeze and thaw the sample one or more times. Dilute the lysate with 2-3 volumes of affinity loading buffer.
   ∞ Sonication: add lysozyme as above, and after incubation for 15 min on ice, sonicate intermittently for 1-2 minutes. Dilute the lysate with 2-3 volumes of affinity loading buffer.
3. Removal of cell debris, insoluble material and nucleic acids: Measure the volume of the diluted lysate, and add PEI to 0.15% (v/v) from a 5% stock (v/v) (adjusted to pH 7.5) while mixing. After 15 min incubation on ice, distribute in centrifuge tubes and spin for 30 min at 25,000 x *g*. Decant the clarified supernatant to a clean vessel. If the pellets are slimy and contaminate the supernatant, add 0.05% (v/v) PEI to the supernatant and centrifuge again, to remove residual DNA. It is advisable to remove or degrade high molecular weight DNA, which can cause trouble in purification and destroy sensitive columns. This can be done by precipitation with polycations such as PEI, as described. Alternatively, the lysate may be passed through a DEAE-cellulose column in 0.5 M NaCl. The high salt concentration allows DNA to bind to the column, while preventing the binding of most proteins. Another approach is to include nucleases, such as Benzonase™ (2-5 units/ml), in the lysis buffer.
4. Load the clarified lysate on an IMAC column. In general, the capacity of the column should match the expected amount of recombinant protein. For example, 1 ml of Ni-NTA or HisTrap medium should be suitable to extract moderately expressed proteins (2-10 mg) from 1L cultures. If large volumes of lysate are processed (e.g. from several litres of culture), components of the lysate seem to reduce binding in a time-dependent manner. Consequently, column loading should be completed in no more than an hour. Use either gravity flow or batch binding from loose resin, or 5 ml packed columns.
5. After column loading is completed, wash with affinity loading buffer followed by affinity wash buffer, until no more protein comes out (10-20 column volumes of each).
6. Elute the bound protein with affinity elution buffer. Identify and pool the peak fractions by measuring A_{280} or using a protein assay reagent (e.g. Bradford).
7. Equilibrate a gel filtration column with gel filtration buffer. The choice of column depends on the expected size of the native protein (considering possible oligomerization), and on the amount of protein. Superdex 75 and Superdex 200

prep grade, XK 16/60 columns are most useful for 1-30 mg of protein in a sample volume of up to 7.5 ml. XK 26/60 columns can be used for larger sample volumes if needed.

8. Analyze protein-containing fractions by SDS-PAGE. Well-expressed recombinant proteins should be identified as the most prominent band; further assays such as immunoblotting, mass spectrometry or enzyme activity assays may be used for verification.

9. Tag cleavage (optional): Measure the concentration of protein, then add 1/20 (mol/mol) histidine-tagged TEV protease. Incubate overnight at 4-8°C. Pass the protein through a bed of Ni-affinity resin (0.1-0.5 ml) equilibrated in gel filtration buffer. This should remove contaminating proteins that spuriously bind to the Ni column but are unaffected by TEV cleavage. If the digestion is incomplete after overnight incubation, concentrate the protein sample the next morning. This usually helps to speed up digestion as substrate concentration and reaction rate are increased. If this still does not work, proceed with IMAC anyway.

10. Concentrate the protein by centrifugal ultrafiltration, using cartridges with the appropriate molecular weight cutoff. Initially, try to concentrate a small portion of the pooled protein fraction; if the desired concentration can be reached in a reasonable time, proceed to concentrate the remainder of the protein. It is advisable to periodically pipette the solution up and down to mix it as to avoid large concentration gradients to build up during the concentration procedure. If the protein fails to be concentrated, or the concentration takes an inordinate amount of time, the protein may be aggregating. One should test different buffers (pH, ionic strength, composition and additives) and find conditions that stabilize the protein. An option is to add 2.5% (v/v) of 1,2-propanediol to the sample prior to concentration to avoid that the protein sticks to the membrane[2].

6. Protein characterization

Mass spectrometric analysis
A mass spectrum of the intact protein is extremely informative. If available, we advocate the use of a coupled liquid chromatography mass spectrometer (LC-MS) system with ESI-TOF in the ionization and mass analysis steps, primarily due to the ease of sample preparation and mass accuracy for macromolecules. The mass spectrum provides an accurate molecular weight (within a few Daltons) estimate that can be used to verify protein identity, and can reveal whether proteolysis has occurred and whether there are co-purifying proteins. The spectrum will also reveal additional protein species whose identifies can be deduced by comparing the observed mass differences with a database of known posttranslational and other chemical modifications (http://www.abrf.org/index.cfm/dm.home).

Static or Dynamic Light Scattering
Proteins that are stable in solution are often monodisperse[3]. If a purified protein tends to precipitate out of solution, or if the gel filtration profile reveals multiple peaks, it may be useful to identify solution conditions in which the protein is monodisperse. The degree of polydispersity can be characterized using static light scattering. Using this measurement, which is quite rapid, buffer conditions can be sampled for their ability to favour monodispersity. If this study is carried out on IMAC-purified proteins, the buffer

promoting the highest degree of monodispersity can be chosen for the next step of preparative gel filtration purification.

Measuring protein stability to identify optimal solution conditions
The behaviour of a protein can be improved significantly by adjusting the solution conditions, which might include salt and pH conditions, as well as the presence of a specific ligand. It is possible to screen for more optimal solution conditions by measuring the resistance of the protein to thermal denaturation. The existing methods require only tens of micrograms of protein, and can be implemented on common laboratory equipment[4].

Additional protocols

A. Chemical induction of chaperones.
1. Prepare an inoculation culture by standard methods.
2. When the inoculation culture is still in log-phase add it to the medium in large scale flasks at 1/100 dilution.
3. When large cultures have reached OD_{600} 0.5, reduce the temperature to 20ºC and simultaneously add ethanol to 4% (v/v) or benzyl alcohol to 10 mM.
4. When cultures have reached 20ºC, induce expression of the target protein by adding IPTG to 0.1-1 mM.
5. Induce overnight (16 hrs).
6. Harvest.

B. Co–expresssion of plasmid- encoded chaperones
1. Prepare a chaperone strain by introducing a suitable chaperone vector (3; commercialized by Takara Bio Inc.) into an expression strain such as BL21 (DE3).
2. Transform the chaperone strain with the required expression plasmid.
3. Pick positive clones and prepare an inoculation culture by standard methods.
4. When the inoculation culture is still in log-phase growth, inoculate large scale flasks with a 1/100 dilution.
5. When large scale flasks have reached OD_{600} 0.5, reduce the temperature to 20ºC and simultaneously induce chaperone expression by adding tetracycline to 2 µg/L. If simultaneous induction of heat-shock proteins is wanted, reduce the tetracycline added to 1 µg/L and add ethanol to 2% (v/v) or benzyl alcohol to 5 mM.
6. When cultures have reached 20ºC, induce expression of target protein by adding IPTG to 0.1-1 mM.
7. Induce further overnight (16 hrs).
8. Harvest.

C. Co-expression of interacting proteins
Recent estimates predict that about a third of all eukaryotic proteins and 80% of all cancer associated signaling molecules contain unstructured domains, and at least 20% of eukaryotic proteins are part of larger protein assemblies. Many intrinsically disordered protein domains and components of protein assemblies mediate protein interactions[5] and these proteins often require an interacting protein for correct folding and stability[6,7]. Such proteins, as well as proteins with unstructured domains, often cannot be expressed in *E. coli* in soluble form. It has proven possible to improve the properties of these proteins by co-expressing the interacting protein[8,9.] In *E. coli*, three co-expression strategies have been

exploited successfully: i) multiple expression plasmids carrying different selectable markers, ii) expression of multiple recombinant proteins from a polycistronic RNA on a single plasmid[10] and iii) expression of multiple recombinant proteins from independent promoters on a single plasmid[9].

There have been no systematic large-scale comparisons of the successes and failures of the different experimental approaches for co-expression, and thus it is impossible to categorically recommend one approach over the others. However, based on our collective experiences, the most pragmatic strategy for the expression of two different proteins would be to express each from a different plasmid. This strategy facilitates cloning and also the combinatorial expression of different versions/derivatives of each protein. The setup of co-expression systems is particularly valuable when entire families of interacting proteins are being studied. In cases in which three or more proteins are to be expressed, it is possible to combine the various strategies; for example to express two proteins from one plasmid and one from another. The decision as to which combination of proteins to express from the same plasmid is specific to the system.

As an example of the considerations in designing such an experiment, the SGC (Structural Genomics Consortium) recently determined the structure of two ternary complexes of SOCS (Suppressors of cytokine signaling) family members. SOCS proteins are found as parts of ternary complexes with two other proteins, Elongin A and Elongin B. In human, 48 proteins contain a SOCS-box which forms a ~40 residue interaction motif. Expression of SOCS proteins, as well as any N- or C-terminal truncation, in isolation resulted either in the expression of an insoluble protein or in protein samples that aggregated during purification. The problem was overcome by expressing a ternary SOCS complex in *E. coli*: SOCS proteins and derivative constructs were expressed from one plasmid (p11, a T7 expression vector with an ampicillin resistance marker); and Elongin B and Elongin C were expressed from a dicistronic RNA from a different plasmid (pACYCDUET-ElCB). This system has already lead to the successful purification of two SOCS ternary complexes and to the successful soluble expression all SOCS-box containing proteins attempted to date[11]. Expression vectors have been developed which allow for co-expression of up to eight recombinant proteins using four different plasmids. In this system double cistronic expression vectors that carry different replication origins and selection markers are used[8] (Duet vectors are shown in the Table below, http://www.emdbiosciences.com/html/NVG/home.html).

Characteristics of duet vectors

Plasmid	Replicon	Copy number	Resistance marker
pETDuet-1	ColE1	~40	Ampicillin
pACYCDuet-1	P15A (pBR322)	10-12	Chloramphenicol
pCDFDuet-1	CloDF13	~20-40	Streptomycin
pRSFDuet-1	RSF1030	>100	Kanamycin
pCOLADuet-1	ColA	~20-40	Kanamycin

Adapted from product information found on http://www.emdbiosciences.com/html/NVG/home.html

References

1.   Stols, L. *et al*. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Prot. Expr. Purif.* **25**, 8-15 (2002).

2.   Banipal, T.S. and Singh, G. Thermodynamic study of solvation of some amino acids, diglycine and lysozyme in aqueous and mixed aqueous solutions. *Thermochim. Acta* **412**, 63-83 (2003).

3.   Ferre-D'Amare, A.R. & Burley, S.K. Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies. *Structure* **2**, 357-359 (1994).

4.   Niesen, F.H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212-2221 (2007).

5.   Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M. & Uversky, V.N. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **272**, 5129-5148 (2005).

6.   Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F. & Jones, D.T. The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20**, 2138-2139 (2004).

7.   Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradovic, Z. & Dunker, A.K. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **323**, 573-584 (2002).

8.   Tolia, N.H. & Joshua-Tor, L. Strategies for protein coexpression in *Escherichia coli*. *Nat. Methods* **3**, 55-64 (2006).

9.   Romier, C. et al. Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. *Acta Crystallogr.* **62**, 1232-1242 (2006).

10.   Scheich, C., Kummel, D., Soumailakakis, D., Heinemann, U. & Büssow, K. Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res.* **35**, e43 (2007).

11.   Bullock, A.N., Debreczeni, J.E., Edwards, A.M., Sundström, M. & Knapp, S. Crystal structure of the SOCS2-elongin C-elongin B complex defines a prototypical SOCS box ubiquitin ligase. *Proc. Natl. Acad. Sci. USA* **103**, 7637-7642 (2006).