Protein Expression and Purification xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Protein Expression and Purification

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

An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability

Junping Fan [a,b], Jie Heng [a], Shuyan Dai [a], Neil Shaw [a], Bei Zhou [a], Bo Huang [a], Zheng He [a], Ya Wang [a], Taijiao Jiang [a], Xuemei Li [a], Zhijie Liu [a], Xianping Wang [a,*], Xuejun C. Zhang [a]

[a] National Laboratory of Macromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, China
[b] Key Laboratory for Cell Proliferation and Regulation Biology of the Ministry of Education, Institute of Cell Biology, Beijing Normal University, 19 Xinjiekouwai Avenue, Beijing 100875, China

ARTICLE INFO

Article history:
Received 26 October 2010
and in revised form 14 February 2011
Available online xxxx

Keywords:
Transmembrane proteins
High throughput
Ligation-independent cloning
Green fluorescence protein
Thermofluor
Protein expression

ABSTRACT

Membrane proteins account for about 30% of the genomes sequenced to date and play important roles in a variety of cellular functions. However, determining the three-dimensional structures of membrane proteins continues to pose a major challenge for structural biologists due to difficulties in recombinant expression and purification. We describe here a high throughput pipeline for *Escherichia coli* based membrane protein expression and purification. A ligation-independent cloning (LIC)-based vector encoding a C-terminal green fluorescence protein (GFP) tag was used for cloning in a high throughput mode. The GFP tag facilitated expression screening in *E. coli* through both cell culture fluorescence measurements and in-gel fluorescence imaging. Positive candidates from the GFP screening were subsequently sub-cloned into a LIC-based, GFP free vector for further expression and purification. The expressed, C-terminal His-tagged membrane proteins were purified *via* membrane enrichment and Ni-affinity chromatography. Thermofluor technique was applied to screen optimal buffers and detergents for the purified membrane proteins. This pipeline has been successfully tested for membrane proteins from *E. coli* and can be potentially expanded to other prokaryotes.

© 2011 Published by Elsevier Inc.

Introduction

Transmembrane proteins (referred to membrane proteins hereafter) constitute 20–30% of prokaryotic and eukaryotic genomes sequenced to date [1,2]. They are important cellular components and play essential roles in many biological processes such as signal transduction, solute and macromolecular transport, and energy metabolism [3]. Owing to their crucial cellular functions, membrane proteins have also become major pharmaceutical targets [4,5]. In contrast to their biological significance, the number of known three-dimensional structures of membrane proteins is currently slim, comprising of only about 1% of the entries in the Protein Data Bank (PDB)[1] [6]. A number of factors are attributed to the difficulties in studying membrane protein structures [7]. First, membrane proteins in general are more difficult to overexpress in recombinant forms than water soluble proteins. Recombinant membrane proteins can be fairly toxic to their host cells by competing for cellular membrane with endogenous membrane proteins, changing physical–chemical properties of the membrane system or directly disrupting biochemical pathways. Secondly, even when overexpressed, a membrane protein is often difficult to be purified in its native form. Detergents are required to extract membrane proteins from cellular membrane [8]. Since different membrane proteins may behave distinctively in a variety of detergents, it is essential to identify optimal detergent type and concentration for extraction, purification, and crystallization of each given recombinant membrane protein. Thirdly, because protein-detergent complexes are often heterogeneous and pack poorly in a lattice, the success rate for crystallizing membrane protein samples is usually lower than water soluble proteins. One strategy to improve the chance of obtaining useable membrane protein crystals for structural study is to increase our search base. This may be achieved either by simultaneously trying numerous candidates or searching in parallel multiple homologs (including mutations) for the protein of interest [9].

Many efforts have been devoted to developing protocols of structural proteomics for membrane proteins [9–16]. Here, we report a potentially high throughput approach to expressing and purifying recombinant *Escherichia coli* membrane proteins in an *E. coli* host as part of our efforts towards membrane protein structural proteomics. This procedure has been shown to be successful

* Corresponding author.
E-mail address: wangxp@moon.ibp.ac.cn (X. Wang).

[1] Abbreviations used: GFP, green fluorescence protein; LIC, ligation-independent cloning; PDB, Protein Data Bank; TEV, tobacco etch virus (protease).

1046-5928/$ - see front matter © 2011 Published by Elsevier Inc.
doi:10.1016/j.pep.2011.02.010

Please cite this article in press as: J. Fan et al., An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.02.010

UNCORRECTED PROOF

in providing purified samples for membrane protein crystallization. It can also be used for expressing and purifying membrane proteins from species other than *E. coli*.

Materials and methods

Target selection protocol

Amino acid sequences of *E. coli* proteins were downloaded from the NCBI Batch Entrez (http://www.ncbi.nlm.nih.gov/sites/batchentrez) by specifying GenBank IDs of *E. coli* proteins collected from the GenProtEC database (http://genprotec.mbl.edu/).

Transmembrane helices as well as the C-terminus locations were predicted with the program TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) [17] which is one of the most commonly used programs in membrane protein prediction [1].

For *E. coli* proteins with more than one transmembrane helix, we searched PDB (http://www.pdb.org/) using a portable version of the Blast program (ftp.ncbi.nlm.nih.gov) [18] to identify possible homologous proteins of known three-dimensional structures.

Sequence based parameters were calculated with the program ProtParam Tool (http://au.expasy.org/tools/protparam.html).

Some functional information was extracted from annotation of the corresponding FASTA sequence files (http://www.ncbi.nlm.nih.gov/BLAST/fasta.shtml).

The results were further assembled into a master spread-sheet file using a local program written in the Perl language.

Vector construction

To facilitate high throughput cloning of candidate genes, we modified an existing pET28 based vector, pWaldo-d [19] which encodes a C-terminal TEV (tobacco etch virus) protease cleavage site, an enhanced green fluorescence protein (referred to GFP hereafter) [20], and an oct-histidine (His_8) tag. Our new vector, referred as pIBP-WX11, contains a ligation-independent cloning (LIC) site [21] and was constructed by inserting a LIC motif between *Nde*I and *BamH*I sites of the multiple cloning site of the pWaldo-d vector. The LIC motif was synthesized as two pieces of DNA, WX11-for 5′-T ATG CTG TAC TTC CAA TCC AAT ATT TGG GTG GGA TAA CCG-3′ and WX11-rev 5′-G ATC CGG TTA TCC CAC CCA AAT ATT GGA TTG GAA GTA CAG CA-3′ (*Ssp*I site is underlined), which were annealed and inserted into pWaldo-d using restriction endonucleases. To prevent accidental expression of GFP from the empty vector, the C-terminal TEV-GFP-His_8 tags were shifted to a different reading frame from the 5′-end ATG starting codon. In addition, another vector, pIBP-WF12 was constructed by deleting the TEV cleavage site and GFP gene from the pIBP-WX11 vector. To make this vector, first the original vector was treated with *BamH*I and *Sma*I, the latter of which is unique in the vector and located downstream of the His_8-tag region. Next, a PCR reaction was performed to amplify the section from the His_8 region to the *Sma*I site from the original vector with primers WF12-for 5′-CAT GGA TCC CAT CAT CAT CAC CAC-3′ (an added *BamH*I site is underlined) and WF12-rev 5'-CTT CCC GGG AAA ACA GCA TTC C-3' (*Sma*I site is underlined). Finally, the PCR product treated with *BamH*I and *Sma*I was ligated back to the original vector digested with the same enzymes.

Ligation-independent cloning

To prepare a LIC ready vector [21], an expression vector as described above was linearized with *Ssp*I, and T4 DNA polymerase was used to create the two single stranded ends in a 40 μl reaction containing 200 ng vector, 5 mM DTT, 2.5 mM dGTP, 1× T4 polymerase buffer, and 2 units of LIC certified T4 DNA polymerase (Novagen, US). The reaction mixture was incubated at room temperature (RT, 25 °C) for 45 min, and the T4 DNA polymerase was inactivated by heating at 75 °C for 20 min. To clone *E. coli* membrane protein open reading frames (ORFs) into the expression vector, each selected ORF was amplified using Ex Taq™ DNA polymerase (Takara, Japan) and *E. coli* genomic DNA as the template in a polymer chain reaction (PCR). To introduce the overlap sequences for LIC cloning, the forward primer contained a sequence of 5′-TAC TTC CAA TCC AAT GCT-ORF and the reverse primer contained a sequence of 5′-T TAT CCC ACC CAA ATG-ORF (reversed). With one base shift the reverse primer should put the downstream tags back into the correct reading frame. The amplified DNA encoding the candidate ORF was purified by electrophoresis and similarly treated with T4 DNA polymerase in a 40 μl reaction each containing 800 ng PCR product, 5 mM DTT, 2.5 mM dCTP, 1× T4 DNA polymerase buffer, and 2 units of T4 DNA polymerase. The T4 DNA polymerase-treated vector (3 μl) and PCR product to be inserted (6 μl each) were then annealed at room temperature for 5 min, followed by addition of 50 μl of chemically treated competent DH5α cells (of efficiency *ca.* 10^8 cfu/μg DNA) and regular heat-shock transformation. Single colonies were selected from kanamycin (25 mg/mL) plates after 10–12 h, and plasmids were extracted. At this point, the identity and integrity of each construct was verified with colony PCR using LIC specific primers and with DNA sequencing.

Small-scale expression screening

In this experiment, we screened for suitable host strains and optimal cell growth conditions for a set of 6 (or multiples of 6) selected candidate membrane proteins following a previously published protocol [12]. Biomek FXP liquid transfer station (Beckman, US) was programmed to perform this experiment with minimal manual intervention.

Transformation

A 96-deep well plate (1–12 x A–H; Greiner, Germany) was used to perform the screening. Competent cells of strains BL21, Tuner, C41, and C43 [22] were thawed on ice, and 20 μl of each competent cells were distributed into rows A–D (columns 1–6 only), respectively. For each of the six membrane protein candidates, 1.5 μl (*ca.* 20 ng/μl) plasmids purified in Section "Ligation-independent cloning" were distributed into columns 1–6 (rows A–D only), respectively. There were 24 reactions in total at this point. The plate was incubated on ice for 30 min. Heat-shock was performed by placing the 96-well plate at 42 °C for 45 s and incubating on ice for 3 min. Sixty microlitres of warm (42 °C) LB media was added to each reaction, and the plate was incubated at 37 °C for 1 h.

The 96-well plate was centrifuged at 4000*g* for 20 min, and the supernatant was discarded. One fifty microlitres LB with kanamycin at 25 mg/L was added to each well. The plate was covered with a gas permeable adhesive seal (AeraSeal Film; Omega) and incubated at 37 °C overnight with 200 rpm shaking.

Expression screen

Four expression media (LB, Luria–Bertani medium; 2xYT, yeast extract and tryptone; M9 + LB, minimal M9 broth + LB; and TB, terrific broth) [23] were prepared. Each medium was tested for growth of the 24 transformants obtained as described in Section "Transformation" One fifty microlitres of selected medium supplemented with kanamycin (25 mg/L) and 30 μl of the overnight culture were transferred into each well of a new plate for each temperature to be tested (*e.g.* 16 °C and 25 °C). Total 192 combinations (6 candidates × 4 strains × 4 media × 2 temperatures) or two 96-well

UNCORRECTED PROOF

Please cite this article in press as: J. Fan et al., An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.02.010

plates (one for each temperature) were analyzed. When only one temperature was to be tested, we usually chose 16 °C.

All plates were incubated at 37 °C with 220 rpm shaking. Cells were induced after 5.5 h (*ca.* 0.6–0.8 $OD_{600\ nm}$) for expression of recombinant proteins by adding 7.5 μl IPTG (isopropyl β-D-thiogalactoside) at 1 mM final concentration. Cell cultures grew for another 18 h at either 25 °C or 16 °C.

Fluorescence-based cell culture screening

For whole cell fluorescence measurements, cells were harvested by centrifugation at 3300*g* for 30 min. Cells from each well were then re-suspended in 100 μl of Tris buffer (20 mM Tris–HCl (pH 7.0), 100 mM NaCl, and 10% (v/v) glycerol). The cells were transferred to a black plate of UV transparent bottom (Greiner, Germany), and its $OD_{600\ nm}$ absorption was recorded. GFP emission from whole cells was recorded from the bottom of the plate using a Varioskan Flash fluorescence reader (Thermo, US) with an excitation wavelength of 484 nm and an emission wavelength of 512 nm.

In-gel fluorescence

In-gel fluorescence was used to investigate integrity and quantity of the GFP-fusion proteins. For example, during high throughput evaluation of cell cultures, 100 μl lysis buffer (B-PER Protein Extraction Reagents, Thermo) was added to cell pellet of each of the 96-wells, and the plate was incubated on ice for 1 h. Twenty microlitres of each cell lysate were mixed with 5 μl 5× SDS–PAGE loading buffer (250 mM Tris–HCl (pH 6.8), 10% (w/v) SDS, 0.5% (w/v) bromophenol blue, 50% (v/v) glycerol, and 5% (v/v) β-mercaptoethanol). 10 μl of the sample mixture was loaded on an SDS–PAGE gel (10% SDS and 15% polyacrylamide) after 10 min incubation at 37 °C. The gel was imaged directly using a fluorescence imaging system, Gel Documentation and Image Analysis System (Sagecreation, China) and subsequently stained with Coomassie blue.

Medium scale membrane protein purification

Cells were lysed using a high pressure homogenizer, EmulsiFlex-C3 (Avestin, Canada). Cell debris was removed by high-speed centrifugation at 17,000*g* for 15 min at 4 °C, and the supernatant was ultracentrifuged at 100,000*g* for 1 h at 4 °C to enrich the membrane. A detergent, *e.g.* DDM, was applied at 1% final concentration to dissolve the membrane on ice. Undissolved components were removed with another round of ultracentrifugation. The supernatant, presumably containing the membrane protein, was subjected to metal chelate chromatography followed by size exclusion chromatography and/or ion-exchange chromatography.

Thermofluor based buffer and detergent screening

Thermofluor experiments [24] were performed with an qPCR instrument, Rotor-Gene 6600 (Corbett Research, Australia) which is equipped with a blue light channel. Thiol-specific fluorochrome N-[4-7-(diethylamino-4-methyl-3-coumarinyl) phenyl] maleimide (CPM) (Invitrogen) was used as the fluorescence probe which binds buried cysteine residues during protein denaturation as the temperature increased. This results in a fluorescence signal with an excitation maximum at 387 nm and an emission maximum at 463 nm [25]. Up to 72 samples could be measured in one experiment within 1.5 h. The volume of each sample was 20 μl, and the final concentration of the protein was 0.1–0.2 μg/μl. The protein samples were typically diluted from an original buffer (*e.g.* 0.1 M Tris–HCl (pH 7.0) and 0.05% DDM) into the buffer to be analyzed at a 1:20 ratio. Buffers being tested included sodium acetate, MES, HEPES, Tris, and bicine among others. The pH range varied from 5 to 9, and salt concentration varied from 0.1 to 1.0 M. In a separate assay, the following detergents were evaluated: $C_{12}E_8$ (octaethylene glycol monododecyl ether), FC-12 (fos-choline-12), LDAO (n-dodecyl-n,n-dimethylamine- n-oxide), DM (n-decyl-β-D-maltopyranoside), DDM (n-dodecyl-β-D-maltopyranoside), chapso, Cymal-6, βOG (n-octyl-β-D-glucopyranoside), and nOG (n-nonyl-β-D-glucopyranoside) typically at 5× CMC concentrations. All detergents were purchased from Anatrace (US) at the highest purities available. The qPCR instrument was programmed to increase temperature from 25 to 95 °C at a rate of 1 °C/min. Fluorescence signals were recorded using the blue channel of the instrument. Melting temperature, *T*m, was estimated to be the temperature corresponding to the minimum of the first derivative of the denaturation curve.

Results

Candidate selection

One goal of our current work is to build a platform for high throughput expression and purification of prokaryotic membrane proteins for structural and functional studies (Fig. 1). *E. coli* was chosen as the expression system, and as a proof-of-principle test we selected membrane proteins from *E. coli* as targets.

ORFs from *E. coli* genome were analyzed for (a) their predicted transmembrane helices; (b) position of the C-terminus relative to the cytosol; (c) presence of homologous structures deposited in PDB; and (d) known biological functions. All candidates should have at least two transmembrane helices. Their C-termini were predicted to be on the cytosolic side of the membrane, because our initial expression screen would be based on a C-terminal GFP tag which is reported to work best when located inside the cell [26]. In addition, we set a cutoff for homologous structures from PDB at 30% sequence identity to avoid duplicating efforts of other groups. While the above three criteria were evaluated by a few computer programs, annotated functions of candidates were analyzed manually based on information collected from literature and databases (*e.g.* PCDB from www.pcdb.org).

From 4300 ORFs of *E. coli* genome, over 400 targets that satisfy the above criteria were selected. As an initial test, 75 candidates were included in the current work. Molecular weights of these candidates range between 17.7 and 62.5 kDa. The number of predicted transmembrane helices for each protein varied between 3 and 14 with their C-termini inside the cell. Table 1 lists some statistics of the target set included in the current work. The candidates included transporters (35 targets), enzymes (29 targets) and proteins with a host of other functions. Furthermore, most candidates had a human homologue, so that a successful structural study may have a greater, more general impact.

Cloning

Based on previous reports using GFP as a reporter for membrane protein expression [27,28], we utilized a GFP-based screening system to evaluate the expression potential of our candidate *E. coli* membrane proteins in an *E. coli* host. There exists a correlation between proper folding, membrane insertion of target proteins, and their GFP fluorescence [29]. However, one potential limitation of such a technique is that the C-terminal GFP moiety may not mature properly if the C-terminus of the target protein is located in the peripheral space [29]. Genes encoding C-terminal inside candidates were cloned into the pIBP-WX11 vector which contained a LIC cloning site and encoded a C-terminal GFP tag (Fig. 2). The LIC method does not require target-specific restriction enzymes for cloning and, therefore, is particularly useful for high throughput applications [21]. Among the 75 candidates, 66 (88%) were cloned

UNCORRECTED PROOF

Please cite this article in press as: J. Fan et al., An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.02.010

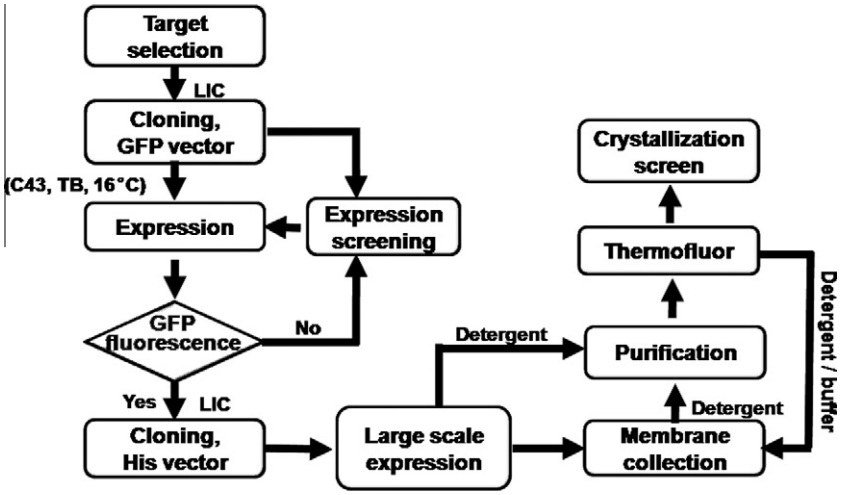

Fig. 1. Flow chart of membrane protein expression screening and purification.

Table 1
Statistics of the target membrane proteins. Seventy five *E. coli* membrane proteins were included in this study. Some statistics of these proteins at different stages of the screening are listed. The transmembrane helix numbers were predicted by TMHMM, and other numbers were calculated by the ProtParam Tool program.

	Min	Max	Average
Target selection: 75			
No. of a.a.	100	560	358
MW (kDa)	17.7	62.5	39.4
No. of helices	2	14	8.3
No. of Cys.	0	8	2.8
pI	5.2	10.6	8.5
Cloning: 66			
No. of a.a.	100	560	345
MW (kDa)	17.7	61.3	37.0
No. of helices	2	14	8.2
No. of Cys.	0	8	2.7
pI	5.2	10.6	8.5
GFP-screening: 30			
No. of a.a.	147	552	351
MW (kDa)	17.7	61.3	33.7
No. of helices	3	14	8.2
No. of Cys.	0	8	2.6
pI	5.5	10.6	8.6
Subcloning: 20			
No. of a.a.	158	549	321
MW (kDa)	17.7	59.2	33.7
No. of helices	3	13	7.8
No. of Cys.	0	7	2.1
pI	5.5	10.6	8.4
Purification: 15			
No. of a.a.	162	549	328
MW (kDa)	18.2	59.2	35.7
No. of helices	4	13	8
No. of Cys.	0	7	2.3
pI	5.5	10.6	8.7

successfully, as judged by growth of the transformants and subsequent DNA sequencing. All cloning failures occurred at the PCR step when target genes were amplified from the *E. coli* genomic DNA. There was no clear preference among our target proteins on what properties might have influenced the cloning (Table 1). For a high throughput process, sequence confirmation could be carried out after expression screening to focus only on those clones that had showed positive green fluorescence.

Small-scale expression screening

To determine whether it is possible to have a default growth condition for a high throughput expression screening in the current system, we performed several small-scale expression studies. In each experiment, we usually chose six candidate membrane proteins, four *E. coli* strains (BL21, Tuner, C41, and C43), four media (LB, 2xYT, M9 + LB, and TB), and two temperatures (16 °C and 25 °C) as variables. Fig. 3 shows results from a representative experiment. Four of the six targets showed significant expression under certain conditions judged by whole cell lysate fluorescence measurements, in-gel fluorescence, and anti-His immunoblot analyses (Fig. 3 and data not shown). In particular, we demonstrated that GFP fluorescence signals are fairly reproducible at the whole cell level (Fig. S1), suggesting that the fluorescence value of whole cell culture may serve as a reliable indicator of expression level of candidate proteins. Additionally, in-gel fluorescence allows assessment of the integrity of membrane protein-GFP fusions and provides a rapid and generic alternative for immunoblotting of membrane proteins. Furthermore, our results showed that a combination of the C43 strain, TB medium with IPTG induction, and 16 °C gave the best expression among combinations we tested (Fig. 3B and data not shown). This observation was consistent with the facts that TB is the most nutrition-rich among the media being compared [30] and that the *E. coli* C43 strain is specifically selected for membrane protein expression and is more tolerant to toxicity of overexpressed recombinant proteins [22].

The combination of these optimal parameters was subsequently used as a default condition for expression screening of more candidate membrane proteins in a 96-well format. Among the 66 successful clones, 30 targets gave positive GFP fluorescence (*i.e.* more than three time as high as the background value). The total ratio was 40% (*i.e.* 30/75). All target proteins of lengths between 100 and 147 amino acid residues resulted in low or no expression and were thus excluded from further studies. Meanwhile, the average molecular weight of positive targets decreased, indicating some larger target proteins were also unfavorable in this expression screening (Table 1).

Overexpression of candidate membrane proteins without a GFP tag

Initial attempts at direct purification of a number of GFP-fusion proteins resulted in several failures during proteolytic removal of the GFP tag. The TEV protease digestion was often incomplete and did not yield sufficient quantities of purified target proteins for further characterization. We speculated that part of the reason for this failure was that the detergent used (*i.e.* DDM. See below) had some inhibitory effects. However, a control experiment to test detergent effects on the TEV protease activity towards a GST-TEV-

UNCORRECTED PROOF

Please cite this article in press as: J. Fan et al., An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.02.010

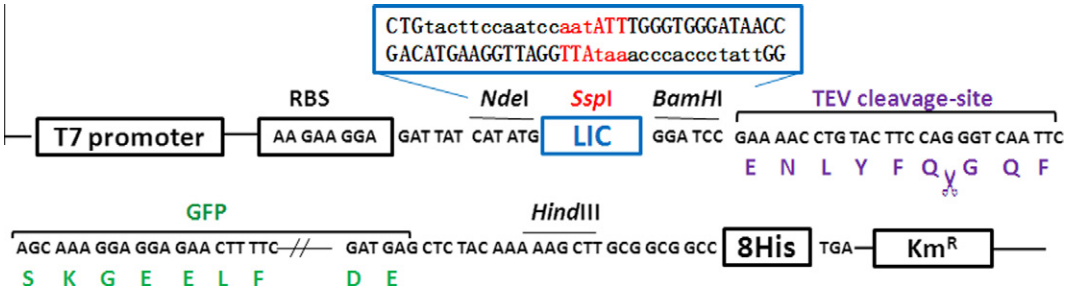

Fig. 2. Schematic diagram of the pIBP-WX11 vector. The LIC motif which is enlarged in the top blue box is inserted using the *Nde*I and *BamH*I restriction sites of the pWaldo-d vector. The *Ssp*I site marked in red is used to linealize the vector. Nucleotides to be removed by T4 DNA polymerase during LIC preparation are shown in lower case letters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

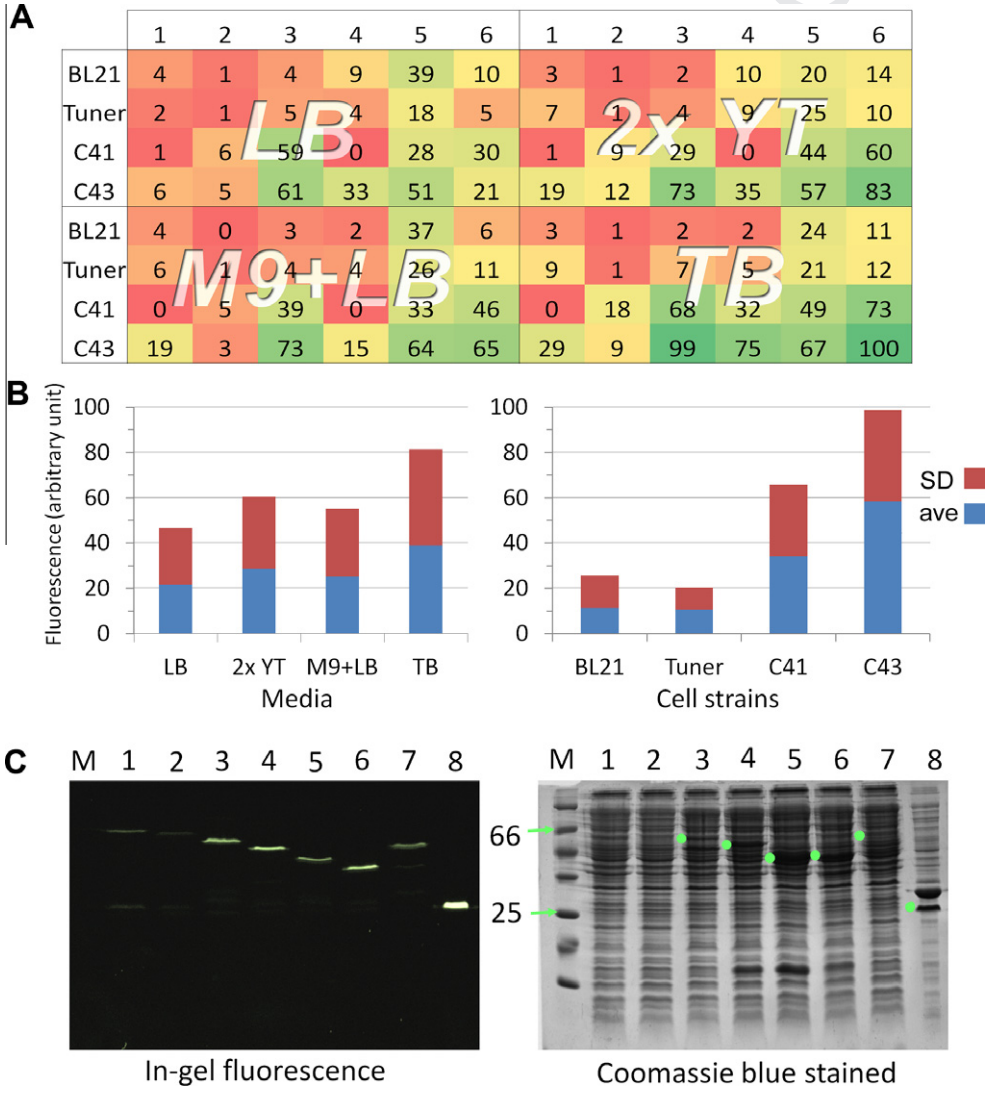

Fig. 3. Membrane protein expression screening. (A) Screening of optimal expression conditions. Six randomly selected membrane proteins (#1–6) from our candidate list were expressed in a 96-deep well plate using a combination of 4 different media (LB, 2xYT, M9 + LB, and TB) and 4 *E. coli* strains (BL21, Tuner, C41, and C43) at 16 °C. The cell cultures were induced with IPTG. Fluorescence values of the cell cultures were recorded from a black-wall plate with a plate reader, and the results are color coded here: the highest expression (arbitrarily normalized to 100) is indicated with dark green color, the lowest expression is colored in red, and others in between. This experiment was repeated several times, and the results were reproducible. (B) Statistics of the fluorescence readings for each of the four media (left) and of the four cell strains (right). The blue bars are the average value of the corresponding row or block in (A), and the magenta parts are the corresponding standard deviation. (C) SDS–PAGE in-gel fluorescence of 6 candidate membrane proteins (#1–6). All proteins were expressed under the condition of TB media, C43 strain, and 16 °C. Cell cultures were induced with IPTG. Total cell lysates were mixed with SDS buffer, and the supernatants were subjected to electrophoresis. Positive controls include a GST fusion protein replacing membrane protein in the same vector (lane 7) and GFP alone (lane 8). The in-gel fluorescence image is shown on the left, and the Coomassie blue stained gel on the right for comparison. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

Please cite this article in press as: J. Fan et al., An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.02.010

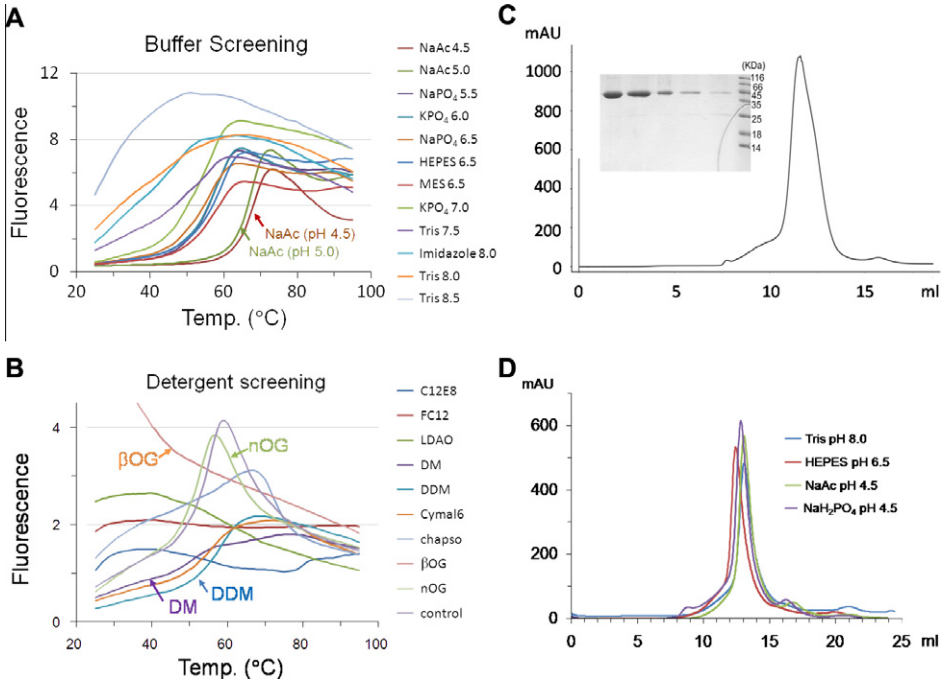

Fig. 4. Buffer-detergent screening using thermofluor. (A) Effects of buffers and variation of pH on the thermal stability of one candidate membrane protein (#11). The protein sample (4 μg/well) was prepared in 100 mM NaCl, 0.025% DDM, and 0.1 M buffer. (B) Effects of detergents on the thermal stability. 20 mM HEPES (pH 6.5), 100 mM NaCl, and a variety of detergents at 5× CMC concentration were tested. All measurements were repeated multiple times with reproducible results. Representative denaturation curves are shown. (C) Quality of the protein sample used in thermofluor assays. The protein sample (#11) was subjected to Ni-affinity chromatography and SEC using Superdex 200, and an SDS–PAGE analysis of the SEC peak is shown in the inset. (D) Variation of buffer conditions may have a less dramatic effect on the SEC profile than on the thermofluor profile of a target protein.

GFP-fusion protein encoded by a pIBP-WX11 based plasmid showed that DDM had little if any effect on the TEV protease mediated cleavage (Fig. S2), suggesting that in our current system other factors such as the fusion protein conformation might be the main reason for the problem [31].

To avoid this difficulty, we sub-cloned the high-yield target proteins from the GFP-based expression screen into a His-tag vector (pIBP-WF12) using the LIC method. Of the 30 positive clones, top 20 clones that gave higher fluorescence were sub-cloned into the pIBP-WF12 vector. The C-terminal His-tag fusion proteins were expressed under a similar condition to that used for the corresponding GFP fusion and purified using Ni-affinity chromatography. The expression level for each candidate was evaluated by SDS–PAGE with Coomassie blue staining and anti-polyhistidine immunoblot (data not shown). Consistent with previous reports [27], most of these candidate proteins maintained their high expression after switching to the His-tag expression vector.

Purification of these recombinant proteins followed standard protocols [27], including harvesting membranes with ultracentrifugation, membrane solubilization with detergents, and Ni-affinity chromatography. Of the above 20 target membrane proteins, 15 showed good yields, *i.e.* more than 0.5 mg recombinant protein per liter of cell culture after initial Ni-affinity chromatography. As in the GFP fluorescence screening, more target proteins of either shorter or longer chains were eliminated (Table 1).

Thermofluor-based buffer-detergent screening

For high throughput membrane protein purification, it is desirable to have a universal technique for sample quality control. Thermofluor technique can be used to determine thermal stability of protein samples without prior knowledge of protein functions [24]; therefore, it can serve as a quality control technique as soon as the protein sample is partially purified. Furthermore, because a thermofluor measurement takes only about 2–4 μg protein sample and it can be performed with a programmable qPCR instrument, this technique can be used to screen optimal buffers, salts, and detergents in a high throughput mode. For membrane proteins, the highly reactive thiol-specific fluorochrome CPM can be used as the fluorescence probe that binds buried cysteine residues during protein denaturation as the temperature increases [25]. In line with the requirement that the target protein contains buried cysteine residues, of our 75 initial candidates 64 contained at least one cysteine residue, and half of them had three or more cysteine residues (also see Table 1).

Since a thermofluor assay would not work for a raw sample of membrane protein, initial purification must be performed to prepare samples for the assay, for example, with a Ni-affinity chromatography and even with an additional sizing exclusion chromatography (SEC). Before an optimal condition had been determined, this preliminary purification was usually done under a default condition (Fig. 4C). Based on previous reports [11], DDM was chosen as our default detergent of high throughput screening for both membrane extraction and purification. Once the candidate membrane protein was purified from a medium scale preparation, a thermofluor experiment could be performed to screen for optimal combinations of buffers and detergents. In the example shown in Fig. 4, the candidate protein (#11) is a 47 kDa transporter, contains seven cysteine residues, and was predicted to have 12 transmembrane helices and a pI value of 9.5. Until now, there has not any report about the protein functional form in vivo or in vitro. Results of the buffer screening showed that, for this candidate, NaAc buffer (pH 4.5 and 5.0) worked the best (Fig. 4A); and buffers of pH values higher than neutral appeared destabilizing the protein. However, we found that switching from the default condition to some low pH buffer did not show further improvement of protein behavior in a SEC chromatography (Fig. 4D), suggesting either that the thermofluor assay is a more sensitive measurement for protein stability or that a correlation between thermal stability and oligomerization status is weak. On the other hand, oligomerization status of a protein sample would

UNCORRECTED

Please cite this article in press as: J. Fan et al., An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.02.010

clearly affect the results of its thermofluor assay. For example, samples of one target protein collected from different fractions of an SEC experiment often behaved distinctively in a thermofluor assay. In our hand, oligomers of the same protein sample would lower the initial fluorescence value in the thermofluor assay compared to its monomer counterpart, but effects of oligomerization on the melting temperature was not significant (data not shown).

Similarly, thermofluor detergent screening showed that DDM and Cymal-6 were among the best detergents in stabilizing this particular membrane protein (#11) and others during thermal denaturation (Fig. 4B and data not shown). It is noted that the maltoside heads and main bodies of alkyl chains of DDM and Cymal-6 are the same, while their tails differ (a methyl *vs.* cyclohexyl group). In addition, DDM and nOG worked better than their shorter cousins DM and βOG, respectively. Moreover, the detergent LDAO appeared destabilizing to many of our protein samples tested in the CPM-thermofluor assay. This is consistent with the notion that LDAO is more destructive to membrane proteins [8]. On the other hand, it was possible that some types of detergents may simply not be compatible with the fluorescence probe CPM for unknown reasons. To test such a possibility, we performed the same thermofluor assay using chicken egg white lysozyme (Sigma) and a variety of detergents. Under the experimental condition (Fig. S3), most detergents gave consistent thermal denaturation curves which are also comparable with data obtained from a thermofluor experiment using a thiol group independent fluorescent dye (Sypro Orange) (data not shown). Results from this control experiment suggested that the CPM fluorescence dye is compatible with most detergents used in such an assay. It is interesting to note that this 129 amino acid residue lysozyme contains 8 cysteine residues, forming four buried disulfide bonds (PDB ID 3IJU). Our results also indicated that thiol groups involved in disulfide bond formation in a folded protein can also be targeted by CPM in a thermofluor assay. This would allow the CPM based thermofluor assay to be applied to more cysteine containing membrane proteins independent of redox status of their cysteine residues, provided that these cysteine residues are buried in the folded state [25].

Discussion

A membrane protein expression, purification pipeline has been developed based on a GFP mediated screening system (Fig. 1). As the starting point of this pipeline, a software package was assembled by integrating a number of on-line resources to facilitate candidate selection from a given species following a set of rules such as the number of predicted transmembrane helices and whether homologous known structures exist in PDB.

Based on LIC [21] and GFP-tag techniques, we developed a high throughput screening scheme to identify candidates of membrane proteins for *E. coli* expression. We have tested this scheme using *E. coli* membrane proteins as a show case and have implemented it with a robotic liquid handling system. Genes encoding candidate membrane proteins are first cloned into a new vector (Fig. 2) which is modified from an existing pET28 based GFP encoding vector, pWaldo-d [19], to make it LIC compatible. Using whole cell fluorescence measurements, dot blot, and in-gel fluorescence, the system can potentially be used to identify optimal expression conditions such as expression strain, medium, and temperature for every candidate proteins.

Our study showed that the fluorescence intensity of a GFP-fusion protein is a good indicator of the protein expression level in a high throughput screening (Fig. 3 and Fig. S1), consistent with previous report [27,28]. Since C-terminal GFP tag is a more reliable reporter of membrane-integrated expression [29], we chose a construct of C-terminal GFP tag for our candidate membrane proteins. Nevertheless, due to low folding efficiency of the GFP tag in *E. coli* periplasmic space, this method would have more reliable fluorescence when the C-terminus is located inside the cell [26]. Since bioinformatics prediction of membrane protein topology is fairly reliable [26], we decided to select our targets to feed into the high throughput pipeline based on such a prediction. Experimental data about the topology would also be helpful in making such selections. For those C-terminal outside proteins which account about 30% of membrane proteins [32], their heterogeneous expression can be screened with alternative high throughput methods for example fluorescence of an N-terminal GFP tag or a dot blot against a polyHis tag. The current pipeline can be used without much change to accommodate an N-terminal GFP tag based screening protocol. Such a method would improve screening efficiency by including N-in-C-out integral membrane proteins and giving additional chances for those membrane proteins that would become unstable if fused with a C-terminal tag [33].

Although a GFP tag has many advantages during expression screening and purification, the fusion protein usually requires removing the tag in order to facilitate further stability-functional studies and crystallization. However removing the tag can be problematic, because of improper folding of the fusion protein in a detergent-protein complex and inhibitory effects of some detergents on the proteolytic enzyme used (Fig. S2) [31,34]. To avoid difficulties associated with proteolytic cleavage of the GFP tag by the TEV protease, we choose to switch those target proteins that give good expression in our GFP-based screening into a GFP free vector, pIBP-WF12, for medium or large scale expression. Alternatively, one may use an in vivo method to remove the tag, although it would require co-expression of the corresponding protease [35]. It has been reported that there is a good correlation between expressions of GFP based high throughput screening and GFP-free expression, which means if one protein can give a high fluorescence in a GFP-fusion form it usually can give good expression without GFP [27]. And our results support such a notion.

Detergent choice has significant implications for protein oligomeric state, stability, homogeneity, and crystallization [8]. Therefore, a fast, effective method to search for optimal detergents is essential for high throughput membrane protein purification. It is shown that DDM is the most commonly used detergent in membrane protein extraction and purification [11]; therefore, it was used as our default choice of detergent for those proteins of high expression during their initial extraction and purification. Once we obtained the protein sample in a purified from, a screening for optimal combinations of detergent(s) and buffer(s) is performed using the thermofluor technique [24].

A higher thermal stability can often be translated into better behavior during protein purification and crystallization [36–38]. Thermofluor is a technique to measure protein thermal stability [24,37]. Since it does not require prior knowledge about functions of the target protein, the thermofluor technique is considered a universal method for protein quality control. It has obvious advantages over classical stability assays based on time- and/or temperature-dependent loss of functions [39], although thermofluor usually requires a (partially) purified protein sample. Because of its high throughput capacity, thermofluor can be used to screen a large number of buffer, salt, and detergent combinations to identify optimal conditions for the protein stability, which makes thermofluor a more attractive quality control assay and screening tool than other stability assays such as circular dichroism. Our results show that the pH, buffer type, and detergents have significant influence on the protein denaturation curves (Fig. 4). Stability-friendly detergents may even suggest specific interactions with the membrane proteins, mimicking favorable lipid-membrane protein interaction in a native state. In cases of our target proteins, we find that detergents of longer alkyl chains in general result in higher melting tem-

UNCORRECTED PROOF

Please cite this article in press as: J. Fan et al., An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability, Protein Expr. Purif. (2011), doi:10.1016/j.pep.2011.02.010

perature than short chain detergent (Fig. 4B), which would be consistent with previous observations that detergents of longer alkyl chains are more effective in membrane protein solubilization [8]. These results provide us with some practical guidance for further protein sample preparation and crystallization. Nevertheless, a potential downside to the use of thermofluor CPM assays is that it does not give information regarding the oligomeric state of the protein preparation, although misfolded proteins often result in a high baseline. In case that a proper oligomeric status is essential for functions of the target protein, thermofluor results may not predict which oligomeric status is the best to be pursuit further; however, it at least allows us to eliminate those clearly unfavorable conditions during sample preparation. With our limited experience, we attempt to conclude that a set of good thermofluor results including higher melting temperatures and steeper denaturation curves are a sufficient condition (not a necessary one) for the target membrane protein to behave well during sample purification. For a high throughput expression–purification pipeline, such a technique would be helpful in prioritizing target proteins.

In summary, membrane protein structural genomics is a natural extension of the one for soluble proteins. Many techniques as well as infrastructure developed for soluble protein structural proteomics will be transplanted to study membrane protein structures. New techniques will be developed to solve membrane protein specific problems. Our work is part of these collective, continuous efforts to integrate existing methods into a high throughput pipeline for structural studies of membrane proteins.

Acknowledgment

The authors are grateful to Dr. Geoffrey S. Waldo of Alamos National Laboratory for the original pWaldo-d vector. We also thank staff members of the Structural Biology Core Facility in the Institute of Biophysics, Chinese Academy of Sciences (CAS), for their excellent technical assistance, especially Mr. Jianhui Li, Ms. Zhenwei Yang, and Mr. Xudong Zhao. This work was supported by the MOST (China) "973" Project grant 2009CB918803.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.pep.2011.02.010.

References

[1] L. Fagerberg, K. Jonasson, M. Von Heijne, L. Uhlen, Berglund, Prediction of the human membrane proteome, Proteomics 10 (2010) 1141–1149.
[2] E. Wallin, G. von Heijne, Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms, Protein Sci. 7 (1998) 1029–1038.
[3] S.H. White, Biophysical dissection of membrane proteins, Nature 459 (2009) 344–346.
[4] R.E. Hubbard, Structure-based drug discovery and protein targets in the CNS, Neuropharmacology 60 (2011) 7–23.
[5] T.L. Blundell, Structure-based drug design, Nature 384 (1996) 23–26.
[6] H. Berman, K. Henrick, H. Nakamura, Announcing the worldwide Protein Data Bank, Nat. Struct. Biol. 10 (2003) 980.
[7] J.J. Lacapere, E. Pebay-Peyroula, J.M. Neumann, C. Etchebest, Determining membrane protein structures: still a challenge!, Trends Biochem Sci. 32 (2007) 259–270.
[8] G.G. Prive, Detergents for the stabilization and crystallization of membrane proteins, Methods 41 (2007) 388–397.
[9] F. Mancia, J. Love, High-throughput expression and purification of membrane proteins, J. Struct. Biol. 172 (2010) 85–93.
[10] S. Eshaghi, High-throughput expression and detergent screening of integral membrane proteins, Methods Mol. Biol. 498 (2009) 265–271.
[11] M.S. Willis, C.M. Koth, Structural proteomics of membrane proteins: a survey of published techniques and design of a rational high throughput strategy, Methods Mol. Biol. 426 (2008) 277–295.
[12] K. McLuskey, M. Gabrielsen, F. Kroner, I. Black, R.J. Cogdell, N.W. Isaacs, A protocol for high throughput methods for the expression and purification of inner membrane proteins, Mol. Membr. Biol. 25 (2008) 599–608.
[13] S. Newstead, H. Kim, G. Von Heijne, S. Iwata, D. Drew, High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*, Proc. Natl. Acad. Sci. USA 104 (2007) 13936–13941.
[14] S. Eshaghi, M. Hedren, M.I. Nasser, T. Hammarberg, A. Thornell, P. Nordlund, An efficient strategy for high-throughput expression screening of recombinant integral membrane proteins, Protein Sci. 14 (2005) 676–683.
[15] J. Love, F. Mancia, L. Shapiro, M. Punta, B. Rost, M. Girvin, D.N. Wang, M. Zhou, J.F. Hunt, T. Szyperski, E. Gouaux, R. MacKinnon, A. McDermott, B. Honig, M. Inouye, G. Montelione, W.A. Hendrickson, The New York Consortium on Membrane Protein Structure (NYCOMPS): a high-throughput platform for structural genomics of integral membrane proteins, J. Struct. Funct. Genomics 11 (2010) 191–199.
[16] F.A. Hays, Z. Roe-Zurz, M. Li, L. Kelly, F. Gruswitz, A. Sali, R.M. Stroud, Ratiocinative screen of eukaryotic integral membrane protein expression and solubilization for structure determination, J. Struct. Funct. Genomics 10 (2009) 9–16.
[17] E.L. Sonnhammer, G. Von Heijne, A. Krogh, A hidden Markov model for predicting transmembrane helices in protein sequences, Proc. Int. Conf. Intell. Syst. Mol. Biol. 6 (1998) 175–182.
[18] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications, BMC Bioinformatics 10 (2009) 421.
[19] S. Cabantous, J.D. Pedelacq, B.L. Mark, C. Naranjo, T.C. Terwilliger, G.S. Waldo, Recent advances in GFP folding reporter and split-GFP solubility reporter technologies Application to improving the folding and solubility of recalcitrant proteins from *Mycobacterium tuberculosis*, J. Struct. Funct. Genomics 6 (2005) 113–119.
[20] G. Zhang, V. Gurtu, S.R. Kain, An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells, Biochem. Biophys. Res. Commun. 227 (1996) 707–711.
[21] N.S. Berrow, D. Alderton, S. Sainsbury, J. Nettleship, R. Assenberg, N. Rahman, D.I. Stuart, R.J. Owens, A versatile ligation-independent cloning method suitable for high-throughput expression screening applications, Nucleic Acids Res. 35 (2007) e45.
[22] B. Miroux, J.E. Walker, Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels, J. Mol. Biol. 260 (1996) 289–298.
[23] A.Z. Shaw, B. Miroux, A general approach for heterologous membrane protein expression in Escherichia coli: the uncoupling protein, UCP1, as an example, Methods Mol. Biol. 228 (2003) 23–35.
[24] M.D. Cummings, M.A. Farnum, M.I. Nelen, Universal screening methods and applications of ThermoFluor, J. Biomol. Screen. 11 (2006) 854–863.
[25] A.I. Alexandrov, M. Mileni, E.Y. Chien, M.A. Hanson, R.C. Stevens, Microscale fluorescent thermal stability assay for membrane proteins, Structure 16 (2008) 351–359.
[26] D. Drew, D. Sjostrand, J. Nilsson, T. Urbig, C.N. Chin, J.W. De Gier, G. von Heijne, Rapid topology mapping of *Escherichia coli* inner-membrane proteins by prediction and PhoA/GFP fusion analysis, Proc. Natl. Acad. Sci. USA 99 (2002) 2690–2695.
[27] J. Hammon, D.V. Palanivelu, J. Chen, C. Patel, D.L. Minor Jr., A green fluorescent protein screen for identification of well-expressed membrane proteins from a cohort of extremophilic organisms, Protein Sci. 18 (2009) 121–133.
[28] D. Drew, S. Newstead, Y. Sonoda, H. Kim, G. Von Heijne, S. Iwata, GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*, Nat. Protoc. 3 (2008) 784–798.
[29] D. Drew, M. Lerch, E. Kunji, D.J. Slotboom, J.W. De Gier, Optimization of membrane protein overexpression and purification using GFP fusions, Nat. Methods 3 (2006) 303–313.
[30] S. Jaganaman, A. Pinto, M. Tarasev, D.P. Ballou, High levels of expression of the iron-sulfur proteins phthalate dioxygenase and phthalate dioxygenase reductase in *Escherichia coli*, Protein Expr. Purif. 52 (2007) 273–279.
[31] A.K. Lundback, S. Van den Berg, H. Hebert, H. Berglund, S. Eshaghi, Exploring the activity of tobacco etch virus protease in detergent solutions, Anal. Biochem. 382 (2008) 69–71.
[32] D.O. Daley, M. Rapp, E. Granseth, K. Melen, D. Drew, G. von Heijne, Global topology analysis of the *Escherichia coli* inner membrane proteome, Science 308 (2005) 1321–1323.
[33] D. Yernool, O. Boudker, Y. Jin, E. Gouaux, Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*, Nature 431 (2004) 811–818.
[34] A.K. Mohanty, C.R. Simmons, M.C. Wiener, Inhibition of tobacco etch virus protease activity by detergents, Protein Expr. Purif. 27 (2003) 109–114.
[35] M.I. Donnelly, M. Zhou, C.S. Millard, S. Clancy, L. Stols, W.H. Eschenfeldt, F.R. Collart, A. Joachimiak, An expression vector tailored for large-scale, high-throughput purification of recombinant proteins, Protein Expr. Purif. 47 (2006) 446–454.
[36] V.L. Postis, S.E. Deacon, P.C. Roach, G.S. Wright, X. Xia, J.C. Ingram, J.M. Hadden, P.J. Henderson, S.E. Phillips, M.J. McPherson, S.A. Baldwin, A high-throughput assay of membrane protein stability, Mol. Membr. Biol. 25 (2008) 617–624.
[37] U.B. Ericsson, B.M. Hallberg, G.T. Detitta, N. Dekker, P. Nordlund, Thermofluor-based high-throughput stability optimization of proteins for structural studies, Anal. Biochem. 357 (2006) 289–298.
[38] F. Forneris, R. Orru, D. Bonivento, L.R. Chiarelli, A. Mattevi, ThermoFAD, a Thermofluor-adapted flavin ad hoc detection system for protein folding and ligand binding, FEBS J. 276 (2009) 2833–2840.
[39] J.U. Bowie, Stabilizing membrane proteins, Curr. Opin. Struct. Biol. 11 (2001) 397–402.

UNCORRECTED PROOF