High-throughput expression and purification of membrane proteins

Filippo Mancia a, James Love b, * 

a Department of Physiology and Cellular Biophysics, Columbia University, New York, NY, USA
b New York Structural Biology Center, New York, NY, USA

1. Introduction

Every biological process requiring exchange of information between cells or intracellular compartments or detection of and response to either extracellular stimuli or cues or nutrients, relies for success on one or more proteins embedded in the cell membranes and endowed with the capability of accurately carrying out such a task. Membrane proteins comprise approximately 30% of all proteins in both prokaryotic and eukaryotic organisms (Wallin and von Heijne, 1998), and, not surprisingly, represent by far the most thought-after pharmacological targets. However, despite their importance, they are but a fraction of a percent of those with a known structure. As of December 2009, 214 unique entries in the protein data bank (PDB; Berman et al., 2003), only 0.5% of the total, were accounted for by membrane-spanning proteins. Indeed, integral membrane proteins present formidable, albeit not insurmountable, challenges for structural analysis. The field has progressed tremendously since the first result in three dimensions, by electron crystallography at 7 Å, on bacteriorhodopsin (Henderson and Unwin, 1975) and the first high-resolution structure, at 3 Å by X-ray crystallography on a photosynthetic reaction center (Deisenhofer et al., 1984), and membrane protein structures have been determined at an accelerated pace in recent years. Nevertheless, structural output for this class of macromolecules remains negligible in comparison to their soluble counterparts. The initial objective of the New York Consortium of Membrane Protein Structure (NYCOMPS) was to construct an automated HT pipeline for integral membrane protein production and preliminary characterization that would achieve a comparable level of productivity to that observed for soluble proteins in the initial phases of the Protein Structure Initiative (PSI). To fulfill this goal, we designed and implemented a centralized Protein Production Facility at the New York Structural Biology Center (NYSBC) to implement cloning and screening activities. The methodologies underpinning this pipeline, presented in detail below and in schematic form in Fig. 1, were developed over a period of years through iterative rounds of testing and optimization and have so far resulted in ~7000 cloned and screened targets, yielding crystals from ~30 of these, and 24 structures from six unique novel membrane protein.

1.1. Overview of the field

The natural association of membrane proteins with lipid bilayers, and the resulting need of detergents for extraction and purification complicate their structural analysis. Once disferring membrane-protein crystals are obtained, for example, the diffraction analysis is typically as straightforward as it is for aqueous soluble macromolecules. Problems that arise in the recombinant expression of membrane proteins are even more limiting than difficulties in purification and crystallization. Furthermore, the establishment of a suitable recombinant expression system is typically a prerequisite for successful structure determination, and represents the first bottleneck of the entire process (Grisshammer and Tate, 1995). There are only few examples present in high-abundance from natural sources, and membrane proteins are most often intolerant to their aqueous environment.
Fig. 1. Schematic representation of the NYCOMPS HT pipeline for the identification of prokaryotic membrane proteins suitable for structural studies.
the transcript. LIC has been successfully used for large-scale membrane protein cloning efforts (Clark et al., xxxx).

For HT expression, the consensus points to E. coli being the preferred host for prokaryotic proteins, although L. lactis (Kunji et al., 2003) and cell-free systems and have also been proposed (Liguori et al., 2007; Schwarz et al., 2008). Economical small-scale incubators have been developed that are useful for high-density growth of E. coli in deep-well blocks. Optical densities measured at 600 nm approaching 10 units are possible with just 600 mL of culture (Page and R., 2004). Cultures grown under these conditions appear also to be scalable with growth in both specialized 96-tube airlift fermenters (Lesley et al., 2002) and ultrafilters (Brodsky and Cronin, 2006).

Determination of optimal expression, extraction, and purification parameters has been the focus of several studies on varying numbers of membrane protein targets, ranging from several tens to few hundreds. In two manuscripts, Dobrovetsky et al. outline a HT process for membrane protein expression utilizing a single affinity tag and promoter system, one extraction and purification detergent followed by ion-exchange chromatography or SEC as a final purification step (Dobrovetsky et al., 2005; Dobrovetsky et al., 2007). Using these techniques they were able to screen 280 E. coli and Thermotoga maritima integral membrane proteins. The authors conclude that, in a manner analogous to the techniques involved in HT platforms for soluble proteins, similar approaches for membrane proteins will succeed, but with higher attrition rates. Lewinson et al. investigate many of the necessary parameters needed for a HT approach to the production of membrane proteins (Lewinson et al., 2008), focusing on the prokaryotic P-type transporter family (Yatime et al., 2009). The authors set out to express and purify multiple homologues, in multiple strains, temperatures, affinity tags, promoters and extraction and purification detergents. They find that many factors appear to impact the final outcome, but conclude that (i) the closer phylum of the target gene to the expression host, the better the possible outcome; (ii) the promoter may affect the amount of protein produced and the membrane incorporation levels; (iii) the position and nature of the affinity tag may have a profound impact on expression levels; (iv) that only a small subset of extraction detergents results in a high percentage of success in solubilizing a majority of the membrane proteins, with n-dodecyl-b-d-maltopyranoside (DDM) being the most favored. In a similar study by Eshaghi et al. the expression profile of 49 E. coli membrane proteins was analyzed in this host varying, tag position, expression strain and extraction detergent (Eshaghi et al., 2005). Gateway (Hartley et al., 2000) was utilized for the cloning, which may not be ideal for structural studies due to the introduction of additional amino acids at the protein termini. These parameters were assessed for success by dot blot and gel filtration analysis of some of the targets. The authors show that many assays can be conducted in a multwell plate assay, and hence a high throughput may be achieved. The authors also settled on Fos-choline 12 (FC-12) as their standard extraction detergent, due to its good solubilization properties. However, care should be used before considering the Fos-choline detergents as the panacea for membrane protein studies, because it's suitability as a crystallization detergent has yet to be widely accepted (Klock et al., 2008). Indeed, a review of all membrane protein structures deposited in the PDB suggests that a majority of targets are solubilized in just a small subset of detergents, with DDM being the most commonly used for extraction and purification (Willis and Koth, 2008). The same review also concludes that for recombinant proteins poly-histidines are the most commonly used engineered-affinity tags, and that in ~50% of the cases, SEC is employed as the final purification step.

Whilst a single, monodisperse elution peak from SEC does not guarantee success, this technique is a very informative, predictive method for likelihood of crystallization of membrane proteins (Wang et al., 2003), as it is for soluble proteins (Klock et al., 2008). The elution profile of a membrane protein from a SEC column equilibrated in a given detergent can provide a reliable estimate on their aggregation state, and, in general, on their "well-being" in that surfactant. SEC can be readily adapted to HT methods with micro-volume HPLCs fitted with autoloaders and appropriately sized columns. It can also be given added value by being coupled to static-light scattering and refractive index detectors, allowing quantitative evaluation of excess detergent, micelle size and aggregation state (Veeleser et al., 2009). SEC analysis can be streamlined further by eliminating other time-consuming purification steps, and using an in-line fluorescence detector to monitor the elution profiles of GFP-fusions directly from miniscule amounts of detergent-solubilized cell extracts (Kawate and Gouaux, 2006). Fluorescence can also be used to estimate expression levels of GFP-membrane protein fusions, as shown by Hannon et al. on of ~300 proteins from 18 bacterial and archeal extremophiles (Hammon et al., 2009). The authors find that after 'benchmarking' levels of fluorescence from the fusion tag to levels of target protein expression, they can easily detect membrane proteins produced in amounts suitable for structural studies. The only requirements are that the terminus of the protein that the GFP is attached to remains in the reducing environment of the cytoplasm, so that the GFP fluorescence develops properly (Daley et al., 2005). This may render a percentage of membrane proteins not amenable to this technique. In addition, it may still be necessary to perform standard SDS-PAGE analysis to verify that the fluorescence signal recorded is from the intact, full-length fusion, rather than from truncation products.

The studies reviewed above and our own experience, allowed us to construct an expression and screening platform for prokaryotic membrane proteins, presented here.

2. Materials and methods

2.1. Target selection protocol

In-depth discussion and experimental details for the target selection process can be found in Punta et al. (2009). The initial NYCOMP target set comprised more than 300,000 annotated sequences from the RefSeq collection (Pruitt et al., 2007) belonging to 96 fully sequenced prokaryotic genomes. Since most targets have no experimental annotation linking them to the membrane, the prediction program TMHMM2 (Krogh et al., 2001) was used to predict transmembrane helices (TMHs). Although prediction methods are estimated to be very accurate, they will inevitably make helix prediction mistakes. Therefore, we retained only proteins with ≥2 predicted TMHs. Following this step, redundancy was reduced by filtering out targets with exceedingly similar sequences, guaranteeing that no two proteins in our dataset shared more than 98% pairwise sequence identity (CD-hit; Li and Godzik, 2006). Furthermore, in order to minimize the probability of introducing water-soluble non-integral membrane proteins into our pipeline, sequences with two predicted TMHs, for which the position of the most N-terminal TMH overlapped with a predicted signal peptide sequence where excluded. Indeed, the most common mistake of TMH prediction programs is to predict an N-terminal TMH in place of a signal peptide (Moller et al., 2001). Finally, target sequences were excluded in proteins that were predicted to have more than 15 consecutive disordered residues and hence might be problematic for crystallization (Enzoul et al., 2006). The remaining 39,037 sequences constitute the "NYCOMP98" dataset.

The targets to be cloned are selected from the NYCOMP98 dataset of membrane proteins. Targets are selected following a
two-step process. (1) Valid targets that either constitute promising candidates for structure determination and/or are of utmost biological interest are identified. These are referred to as "seeds". (2) The seeds are expanded into families of typically homologous proteins that have a predicted membrane region structures similar to that of the seed. Only sequences that are part of the NYCOMPS98 set of valid targets were used for the expansion of the seeds.

NYCOMPS seeds were chosen according to two distinct tracks that are referred to as "central selection" and "nomination". Central selected seeds were selected from a list of proteins that have previously been successfully expressed in E. coli (Daley et al., 2005). On the other hand, nominated seeds were "hand-picked" by laboratories participating in the NYCOMPS consortium and adjunct members from the community. Most of these seeds are well-characterized proteins of known function. One key difference between centrally selected and nominated seeds is that novelty is not enforced on the latter set. Instead, the observed similarities to proteins in the PDB was checked, and reported to the nominating group, which was ultimately responsible for the final decision on whether or not to pursue that specific target.

The seed expansion procedure is the same for centrally selected and nominated targets and is based on reciprocal sequence similarity in the predicted transmembrane (TM) region between the seed and the NYCOMPS98 set of target proteins. In particular, given a PSI-BLAST (Altschul et al., 1997) alignment between the seed and a NYCOMPS98 protein with E-value of <10\(^{-3}\), the requirement is that more than 50% of the residues predicted to be in "DMHs" in both proteins are aligned. After a seed is expanded into a family of proteins predicted to have similar TM cores, all family members are subject to additional filters. Firstly, any centrally selected target with significant similarity in the predicted TM region to proteins in the PDB (PSI-BLAST E-value <1 and alignment covering >25% of the target TM region) is filtered out. This filter is not applied to nominated targets. Secondly, all candidates for which there is evidence that they might constitute individual subunits of hetero-oligomeric complexes (using information extracted from EcoCyc; (Keseler et al., 2009) are removed from the list. Finally, there is a correction phase dedicated to "inconsistencies" within the families. Proteins for which the number of predicted TMHs differs greatly from that of the seed are typically discarded, as well as proteins that differ significantly in their length with respect to the seed. Additionally, proteins that align well with the consensus N-terminus of the family (when any such consensus can be identified) but that have additional N-terminal amino acids, are typically excluded, because they often constitute cases of proteins that may have been erroneously annotated. In-depth discussion and experimental details for the target selection process can be found in Punta et al. (2009).

2.2. Cloning procedures

Amplification primers are designed by "Primer prime'er", a program for automated oligonucleotide construction developed by the North-Eastern Structural Genomics Consortium (NESG; Everett et al., 2004). Forward and reverse primers are supplied at a 5 μM concentration in a pre-mixed form in 384-well blocks (IDT, Inc.).

In order to simplify our cloning process, we developed standard pNYCOMPS vectors for prokaryotic expression. These reagents and their sequence information are publicly available at the PSI Materials Repository (http://psimr.asu.edu/). Each plasmid carries a cassette harboring kanamycin resistance and is based on the IPTG inducible T7 promoter pET vector system (Novagen, Inc.). pNYCOMPS plasmids encodes a T7 protease (Kapust et al., 2002) cleavable composite FLAG/deca-HIS affinity tag at either the N- or C-terminus. We have also introduced a ccdB "death gene" (Bernard and Couturier, 1992) within the plasmid to negatively select uncut or self-ligating vector thus abolishing background in the bacterial transformation following annealing.

Ninety-six genomic DNA templates were purchased from ATCC and arrayed in a single 96-well plate at working concentration (10 ng/μL). A list of organisms from which the genomic DNAs were obtained can be found in Punta et al. (2009). A Biomek FX liquid handling robot (Beckman Coulter, Inc.) is used to assemble the amplification reaction in a 384-well plate by mixing a PCR master mix, primers and required matching template DNA in a final volume of 15 μL. KOD Hot Start DNA Polymerase (Novagen, Inc.) is used in the amplification reaction with one standard set of cycling parameters. The high processivity and proofreading ability of this polymerase ensures a success rate of greater than 95% without the need for further optimization. Standard conditions for PCR are as follows: 1 x KOD Hot Start Buffer, 1.5 mM MgSO\(_4\), 0.2 mM (each) dNTPs, 0.3 μM (each) primers, 30 ng genomic template DNA, 10% DMSO, 0.02 U/μL KOD Hot Start DNA Polymerase, and PCR Grade water to 15 μL. PCR reactions are carried out in 384-well plates using standard cycling conditions: initial denaturation and activation, 2 min at 95 °C, is followed by 40 cycles of denaturation for 15 s at 95 °C, annealing for 30 s at 59 °C, and elongation for 30 s at 72 °C. After PCR amplification, samples are removed from the thermocycler, and they are assayed for the presence or absence of products of approximately the correct size by loading 2 μL of each reaction onto a pre-cast, 96-well, ethidium bromide containing 2% agarose gel (E-Gen\(^{TM}\); Invitrogen, Inc.). Inserts are purified using Agencourt Solid Phase Reversible Immobilization (SPRI) technology (Beckman Coulter, Inc.) in a 384-well format. SPRI uses selective capture of nucleic acids onto magnetic micro-particles, which can readily be washed prior to an elution step containing purified DNA (DeAngelis et al., 1995). Purified inserts are eluted in 15 μL of Agencourt RE buffer.

Constructs are generated by ligation-independent cloning (LIC; Aslanidis and de Jong, 1990), a technique which is both simple to use and highly cost effective. T4 DNA Polymerase is used to generate complimentary single-stranded overhangs on both insert and linearized vector. Following treatment, insert and vector are mixed and incubated at 22 °C for 1 h to allow spontaneous annealing of complimentary single-stranded overhangs.

In preparation for T4 DNA Polymerase treatment, cloning vectors pNYCOMPS-LIC-ccdB-TFH10+(C-term.) and pNYCOMPS-LIC-ccdB-FH10+(N-term.) are digested with restriction enzymes BsuRI or SnaBl, respectively. Following digestion, cut plasmids are run on preparative 1% agarose TBE gels and the linearized plasmid is excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen, Inc.), eluted in low-EDTA TE buffer and diluted to a working concentration of 50 ng/μL. LIC treatment reactions are generally performed in large batches, in 96-well thermocycler plates. Standard LIC treatment conditions for the vectors are as follows: 1 x New England Biolabs, Inc. (NEB) Buffer 2, 1 x NEB BSA, 25 ng/μL vector, 2.5 mM dGTP (N-term vector) or dCTP (C-term vector), 0.0375 U/μL T4 DNA Polymerase, and PCR Grade water to final volume. Reactions are incubated at 22 °C for 1 h followed by heat inactivation at 75 °C for 20 min. Inserts are prepared in a manner similar to that of the cloning vectors. The only difference being that in order to make complimentary single-stranded overhangs, the purified PCR products are LIC treated in the presence of the complimentary nucleotide, dCTP for N-terminal inserts and dGTP for C-terminal inserts. The LIC insert master mix is as follows: 1 x NEB Buffer 2, 1 x NEB BSA, 2.5 mM dCTP (N-terminal inserts) or dGTP (C-terminal inserts), 0.0375 U/μL T4 DNA polymerase, and PCR Grade water to final volume. Eight microlitre of the standard LIC insert master mix is combined with 2 μL of purified PCR product and incubated at 22 °C for 1 h followed by heat inactivation at 75 °C for 20 min.
After the vector and inserts have been LIC treated, 2 μL of vector and 4 μL of insert are combined and incubated at 22 °C for 1 h. About 2 μL of 25 mM EDTA pH 8.0 is added to each annealing reaction and incubated at 22 °C for 5 min. Annealing reactions are typically performed in either 96- or 384-well plates according to scale of the experiment.

Following this step, 1–2 μL of circular plasmid is transformed into 20 μL of phage-resistant competent cells (DH10B-TIR-1), and plated on selective agar containing 25 μg/mL kanamycin in 24-well blocks. Blocks are incubated overnight at 37 °C. The following day, 1 colony per target is picked manually and grown in 800 μL of 2 × YT containing 25 μg/mL kanamycin in a 96-well deep-well block at 37 °C in HT growth incubators (Vertiga; Thomson Instrument Company, Inc.) for 16 h to amplify plasmid DNA. Agencourt CosmIdPrep magnetic SPIR technology (Beckman Coulter, Inc.) is used to purify the DNA in an automated manner at a rate of six 96-well blocks every 3 h. At this point, the identity and integrity of each construct is verified by sequencing by Agencourt Biosciences (Beckman Coulter, Inc.).

In order to test for expression and purification potential of each constructed target, expression screens are performed into a phage resistant expression strain (BL21 (DE3) pLysS–T1R; Sigma–Aldrich, Inc.). About 0.6 mL cultures of each transformant are grown to high density in a GNFfermenter (GNF Systems, Inc.) and plated on selective agar containing 25 g/mL kanamycin in a 96-well deep-well block at 37 °C in HT growth incubators (Vertiga; Thomson Instrument Company, Inc.) for 16 h to amplify plasmid DNA. Agencourt CosmIdPrep magnetic SPIR technology (Beckman Coulter, Inc.) is used to purify the DNA in an automated manner at a rate of six 96-well blocks every 3 h. At this point, the identity and integrity of each construct is verified by sequencing by Agencourt Biosciences (Beckman Coulter, Inc.).

In order to test for expression and purification potential of each membrane protein target, expression plasmids are transformed into a phage resistant expression strain (BL21 (DE3) pLysS–T1R; Sigma–Aldrich, Inc.). About 0.6 mL cultures of each transformant are grown in 96-well deep-well block at 37 °C in a HT growth incubators (Vertiga; Thomson Instrument Company, Inc.) for 16 h to amplify plasmid DNA. Agencourt CosmIdPrep magnetic SPIR technology (Beckman Coulter, Inc.) is used to purify the DNA in an automated manner at a rate of six 96-well blocks every 3 h. At this point, the identity and integrity of each construct is verified by sequencing by Agencourt Biosciences (Beckman Coulter, Inc.).

2.3. Small-scale expression screens

In order to test for expression and purification potential of each membrane protein target, expression plasmids are transformed into a phage resistant expression strain (BL21 (DE3) pLysS–T1R; Sigma–Aldrich, Inc.). About 0.6 mL cultures of each transformant are grown in 96-well deep-well block at 37 °C in a HT growth incubators (Vertiga; Thomson Instrument Company, Inc.) for 16 h to amplify plasmid DNA. Agencourt CosmIdPrep magnetic SPIR technology (Beckman Coulter, Inc.) is used to purify the DNA in an automated manner at a rate of six 96-well blocks every 3 h. At this point, the identity and integrity of each construct is verified by sequencing by Agencourt Biosciences (Beckman Coulter, Inc.).

2.4. Medium-scale expression tests

Plasmids harboring genes coding for proteins expressing at suitable levels and purity as shown by the small-scale assay are re-arrayed in a 96-well format, transformed into expression strains, and grown to high density in a GNFfermenter (GNF Systems, Inc.). Cell pellets harvested from this fermenter are treated via the same processes as previously described for the small-scale expression screens, but the larger cell masses necessitate the use of 50 mL tubes and 10 mL drip columns, rather than 96-well blocks. These tubes and columns are conveniently arrayed in custom-made 96-well racks for processing. Purified proteins are eluted from the metal-affinity resin, and ~1/40th of the volume is run on an SDS-PAGE gel and proteins are detected, and yields estimated by staining with coomassie blue.

2.5. Detergent selection and stability assay

The eluted proteins from the previous step are each divided into 4, 40 μl aliquots in a 96-well assay plate. The concentration of DDM in the elution buffer is 0.02% w/v. The chosen detergents, β-octyl-glucopyranoside (β-OG), n-dodecyl-N,N-dimethylamine-N-oxide (LDAO), tetraethylene glycol monooctyl ether (C8E4) and DDM, are added, as 4 μL stock solutions, to 0.2% over the respective amount required to achieve a concentration of twice the CMC (i.e. the percentage for a 2 × CMC solution for β-OG is 0.5%, hence this detergent is added to a final concentration of 0.7%). This represents an excess of the second detergent and is meant to allow for at least partial replacement of the associated DDM molecules, carried from the metal-affinity chromatography elution step. After detergent addition and mixing, the plate is incubated at 25 °C for 2 h. Subsequently, the samples are centrifuged at 3000g for 20 min to pellet any precipitate that has formed and the clear supernatant is transferred to the HPLC system. Samples are transferred to a thermostated holder, which is held at 4 °C, in an Agilent 1200 HPLC equipped with a 4.6 mm × 30 cm TSK-Gel SuperSW3000 column (Tosoh Biosciences, Inc.). The column is equilibrated with gel filtration buffer containing 2 × CMC DDM. Using the autosampler, 5 μL of each sample is loaded onto the column and the elution from the column monitored at 280 nm. The column has an approximate bed volume of 5 mL and a flow rate of 0.25 mL/min is used. This process is repeated automatically until all of the samples have been processed.

Alternatively, the protein of choice can be assayed in a 12 detergent screen where iterative injections (5 μL) of the eluted protein (in DDM buffer) are made onto a SEC column (4.6 mm × 30 cm TSK-Gel SuperSW3000 column) pre-equilibrated in the target detergent at 2 × CMC. The detergents used are DDM, n-decyl-β-maltopyranosanide (DM), n-nonyl-β-maltopyranosanide (NM), n-octyl-β-maltopyranosanide (OM), OG, n-nonyl-β-glucopyranosanide (NG), 5-cyclohexyl-1-pentyl-β-o-maltopyranosanide (Cymal-5), 4-cyclohexyl-1-buty1-β-o-maltopyranosanide (Cymal-4), pentaethylene glycol monooctyl ether (C8E5), C8E4, Fos-choline 10 (FC-10) and FC-12. Elution profiles are optionally monitored by UV absorbance at 280nm, three angle static-light scattering and refractive index detection using a miniDAWN™ TREOS system and Optilab® rEX (Wyatt Instruments, Inc.).

3. Results and discussion

3.1. Target selection

Approximately 10,000 targets have been bio-informatically processed and selected from the “NYCOMPS98” dataset of ~39,000 integral membrane proteins, to date. Details can be found at the public databases TargetDB (http://targetdb.pdb.org/) and Pepceb (http://pepcdb.pdb.org/). The average number of membrane proteins we predicted in all 96 genomes is 24%, which is in-line with other predictions. This number varies from 19% (Methanocaldococcus jannaschii) to 30% (Clostridium perfringens) (Knight et al., 2004; Liu and Rost, 2001). Most of the targets selected are between 100 and 500 amino acids in length and have between 2 and 12 predicted TMHs. 81% of all targets align to one or more Pfam-A families (HMMER E-value <10⁻³), with these families collectively covering more than half of the target predicted TM region. Full details of these experiments are published elsewhere (see Punta et al., 2009).
3.2. Automated cloning

As of winter 2009, 6667 cloning attempts have been conducted resulting in 5113 sequence-verified constructs (77% success). Failures in cloning may result from any step in the process (PCR amplification, purification of inserts, LIC treatments, transformation or sequencing failure). Failure may also at times result from bioinformatics selection of inexistent targets due to incorrect annotation of sequences in public databases. Periodic re-processing of failed cloning efforts are made when sufficient numbers have built up, but these seem to recover only ~5% of targets at most.

3.3. Small and medium expression and purification trails

The goal of the NYCOMPS pipeline for expression and purification of membrane proteins is to identify samples that not only can be over-expressed, but that are also suitable for structural studies. Therefore, the small-scale expression tests have been designed and are conducted not only to monitor expression levels under a fixed and optimized set of parameters, but also to identify targets that can be purified to sufficient yields for the process to be cost-effective. The threshold for selection at this stage is that the yields of the membrane protein extracted from a 0.6 mL culture in DDM and purified by metal-affinity chromatography be sufficient for detection on an SDS–PAGE stained with coomassie blue. Following this initial screening protocol, the lower expression limit set is approximately 0.5 mg/L, a value deemed acceptable for cost-effective scale-up. DDM was selected as the ‘standard’ detergent due to its positive properties as determined by others (Lewinson et al., 2008) and by in-house conducted optimization experiments (data not shown), and cost effectiveness. All these small-scale steps mirror the processes that are subsequently reused at a production scale for structural studies. An SDS–PAGE gel from a representative experiment conducted at this scale is shown in Fig. 2.

A prerequisite for these experiments was to identify a method for lysing a large number of membrane protein samples in parallel, without causing denaturation or aggregation. Chemical lysis solutions are not ideal for membrane proteins as they frequently contain unsuitably harsh detergents. Lysis by French-pressure cell cannot be adapted to 96, 0.6 mL samples, and iterative rounds of freezing and thawing were deemed to be somewhat inefficient. Sonication was chosen as the most viable option. Plate sonicators or multi-head sonicators are commercially available, but the energy distribution has the tendency of being uneven, often leading to over-heating of some samples and no lysis of others. We therefore designed a robotic sonicator that consists of an electronically controlled sonicator probe mounted on a robot arm that can enter sequentially each well of a 96-well deep-well block, energize, and move to the next well in turn (Fig. 3). The sonicator head is kept energized momentarily after leaving the liquid surface but still within the confines of the block well to remove liquid attached to the probe. As a result, no cross-contamination could be detected by western blot (data not shown). The block is held on a metal platform at 2 °C to minimize heating and multiple, typically three at most, rounds of short bursts are used with a 13 min well-to-well delay time to minimize over-heating. Processing of 96 samples is achieved in 13–39 min according to the number of rounds used. This robotic sonicator can also be used in conjunction with a robot-addressable 96-well ultra centrifuge to isolate membrane fractions in HT format, although this additional purification step was most often deemed unnecessary.

Protein stability and aggregation state in different, “crystallization-friendly” detergents are assayed by SEC following an HT procedure. However, the amount of material necessary for these experiments requires a mid-scale growth culture. To this end, we have made use of an airlift fermenter (GNFermenter, GNF systems, Inc.) that allows 96 independent cultures to be simultaneously grown to extremely high cell densities. The cell mass yield from ~70 mLs of culture is typically equivalent to that obtained with ~500 mL grown in a 2 L shake flask. Moreover, expression yields for cells grown with the fermenter are comparable to those achieved in a 96-well deep-well block, thus minimizing issues of scalability.

Typically, ~23% of membrane protein targets pass the small and mid-scale selection steps. This success rate is in agreement with that of other reports (Dobrovetsky et al., 2005). A representative gel of the material eluted from the metal-affinity chromatography step for the mid-scale expression and purification step is shown in Fig. 4. These samples are then passed onto the stability screens discussed below.

3.4. Detergent selection and stability assays

The selection of purification detergent, and possibly more importantly the choice of crystallization detergent, is a critical decision in the entire process of membrane protein structure determination.

Fig. 2. Representative results from small-scale expression tests. A 24-well SDS–PAGE gel stained with coomassie blue. Membrane proteins are purified from 0.6 mL culture volumes. An omni-present contaminant is marked.

The capability of a given protein to withstand a ‘harsh’ detergent correlates with its stability, and stable proteins are more amenable to structural investigation. Therefore, we wished to select proteins that are stable at moderately elevated temperatures in large excesses of short-chain detergents. Proteins yielding elution profiles from SEC that are unperturbed by this treatment are classified as stable, and are likely to perform well at production scale-up and/or crystallization. Proteins that are detergent-selective are classified as workable. Proteins that do not behave well in any detergent are either discarded or redirected to a rescue pathway, depending on the importance of the target. Two examples derived from this assay are shown in Fig. 5. In one case, the protein appears to withstand all conditions—albeit with a somewhat diminished yield in C8E4 and to a lesser extent, in β-OG (Fig. 5A). In contrast, a second protein appears to only tolerate DDM, as sample treated with other detergents elutes from SEC as an aggregate, in the void volume of the column (Fig. 5B). We typically perform this stability assay with three short-chain detergents (see legend to Fig. 5), while we include only DDM in the mobile phase. This reduces running costs and time, thus increasing the throughput. Unfolding or aggregation of a membrane protein in a given short-chain detergent is typically an irreversible process, unlikely to be rescued by removal of the agent in a DDM-only containing mobile phase.

Fig. 3. Picture of custom-made sonicator robot. Key features are indicated with arrows. Sample holder, pictured here, can be interchanged with a 96-well deep-well block or with a custom-made holder for 6, 50 mL centrifuge tubes.

Fig. 4. Representative results from mid-scale expression and purification tests. Commassie-stained SDS–PAGE gel of 24 purified membrane proteins from the medium-scale pipeline. About 2.5% of the total volume eluted from the metal-affinity chromatography step was loaded in each lane. Two contaminant bands are marked. Interestingly, these contaminants differ from those present in the small-scale expression screens. This difference could be due to different growth conditions of the bacteria.

Fig. 5. Stability in short-chain detergents. SEC elution profiles of two different proteins treated with large excess of DDM (blue), C8E4 (green), LDAO (red) and β-OG (pink). The SEC experiments are performed with DDM included in the mobile phase at twice its CMC. Stability of the protein shown in (A) is apparent when compared to the one in (B), as indicated by presents of a large percentage of aggregated material which elutes after approximately 7 min, in the void volume of the column.
A second detergent assay consists of iteratively injecting the partially purified protein onto a SEC column pre-equilibrated in any one of 12 different detergent-containing mobile phases (see methods section for list of detergents). Arrayed UV traces for one protein are shown in Fig. 6. Elution profiles are monitored by UV absorbance at 280nM, and optionally also by three angle static-light scattering and refractive index detection. These data can be used to formulate an estimate of molecular weight, aggregation status and to determine the size of the protein/detergent micelle and build up a profile of detergent preference for a given protein. Adequate detergent exchange on the column often leads to a shift in the retention time of the elution peak and this be confirmed by calculations from the light scattering and refractive index detection (Hayashi et al., 1989; Slotboom et al., 2008). These data also allow detection and quantification of multimeric species.

Approximately 40% of samples that enter the gel filtration assays give a ‘positive’ elution profile in at least one detergent. These proteins are distributed for scale-up and crystallization experiments.

4. Conclusions

We present here a structural genomics pipeline for the HT targeting, cloning, expression, purification and biophysical characterization in different detergent of integral membrane proteins to select those most suitable for structural investigation. This pipeline has been used to target and process thousands of integral membrane proteins and produces many suitable targets for scale up. This highly-automated pre-selection approach serves to concentrate the resources for the labor-intensive downstream steps (crystallization and structure determination) on those proteins with highest probability of success.

Retrospective data analysis of proteins that successfully emerge from the screening processes provides precious feedback into our selection procedures for continuous rounds of improvement. Furthermore, the technologies and processes described herein, could, if deemed necessary, readily be modified to re-process targets under a different set of conditions.

The pipeline presented here has led to biologically-interesting structures such as a bacterial homologue of the kidney urea transporter (Levin et al., 2009) and a pentameric formate channel (Waight and A.B., 2009), with several others on the horizon. Finally, it may be worth noting that the NYCOMPS facilities are available to the general community via the Protein Structure Initiative community-nominated targets proposal system (http://cnt.psi-structuralgenomics.org/CNT/targetlogin.jsp).

Acknowledgments

The authors wish to thank the NYCOMPS team of researchers who generated a substantial amount of the data presented in this manuscript. We are particularly grateful to Marco Punta (Columbia University and University of Munich) for the bioinformatics analysis, and to the following researchers at the NYCOMPS core laboratory at the New York Structural Biology Center, for the cloning, expression and purification of the targets: Brian Kloss, Patricia Rodriguez, Arianne Morrison, Renato Bruni and Brandan Hillerich. We also wish to acknowledge Larry Shapiro and Wayne Hendrickson for their invaluable contributions to the design and implementation of this platform. This work was funded by NIGMS GM075026 (Wayne Hendrickson, P.I.).

References


