A green fluorescent protein screen for identification of well-expressed membrane proteins from a cohort of extremophilic organisms

Justus Hammon,[†] Dinesh V. Palanivelu,[†] Joy Chen, Chintan Patel, and Daniel L. Minor Jr.*

Cardiovascular Research Institute, Department of Biochemistry and Biophysics, Department of Cellular and Molecular Pharmacology, California Institute for Quantitative Biosciences, University of California, San Francisco, California 94158-2330

Received 3 September 2008; Revised 10 October 2008; Accepted 15 October 2008
DOI: 10.1002/pro.18
Published online 2 December 2008 proteinscience.org

Abstract: Green fluorescent protein (GFP) fusion proteins provide a potentially facile tool for identification of well expressed, properly behaved membrane proteins for biochemical and structural study. Here, we present a GFP-expression survey of >300 membrane proteins from 18 bacterial and archaeal extremophiles, organisms expected to be rich sources of membrane proteins having robust biophysical properties. We find that GFP-fusion fluorescence intensity is an excellent indicator of over-expression potential. By employing a follow-up optimization protocol using a suite of non-GFP constructs and different expression temperatures, we obtain 0.5–15 mg L^{-1} expression levels for 90% of the tested candidate proteins that pass the GFP screen. Evaluation of the results suggests that certain organisms may serve as better sources of well-expressed membrane proteins than others, that the degree to which codon usage matches the expression host is uncorrelated with success rate, and that the combination of GFP screening and expression optimization is essential for producing biochemically tractable quantities of material.

Keywords: membrane protein expression; GFP-screen; extremophile

Introduction

Membrane proteins are the gateways for signals that pass between a cell and its surroundings and are central to ionic signaling mechanisms and the material transport required for nutrient uptake and toxin disposal. Understanding membrane protein molecular architecture remains one of the major challenges of modern structural biology efforts. One of the main obstacles is the difficulty associated with obtaining sufficient quantities of pure material for biochemical and structural analysis.[1,2] Because most membrane proteins are not found in readily accessible over-enriched native sources, it is necessary to develop ways to overproduce membrane protein constructs at will. The ability to overproduce proteins in heterologous systems has revolutionized structural biology studies of soluble proteins but similar scale successes have not yet been had for membrane proteins. Thus, pursuit of robust methods to express and characterize membrane proteins remains an area of vigorous research efforts.

The most rigorous test of whether a protein can be crystallized is to overexpress, purify, and place the

Additional Supporting Information may be found in the online version of this article.

[†]Justus Hammon and Dinesh V. Palanivelu contributed equally to this work.

Grant sponsors: NIH-NIGMS (Membrane Protein Expression Center and Center for the Structures of Membrane Proteins); American Heart Association.

*Correspondence to: Daniel L. Minor, Jr., Cardiovascular Research Institute, Department of Biochemistry and Biophysics, Department of Cellular and Molecular Pharmacology, California Institute for Quantitative Biosciences, University of California, San Francisco, California 94158-2330.
E-mail: daniel.minor@ucsf.edu

protein of interest in crystallization trials. However, taking any given protein through this procedure requires a large investment of time and resources and has a high probability of a negative outcome for any given candidate. There have been ongoing efforts to develop methods that serve as readily accessible proxies to identify well-expressed, well-behaved protein samples that should have a higher probability of leading to crystallizable material without the need for an initial complete characterization of each candidate. Such approaches attempt to leverage the power of molecular biology to provide a rapid way to screen large numbers of constructs in parallel and identify the most promising for further, more intensive characterization. A leading tactic in this regard has been the use of fusion proteins bearing Green Fluorescent Protein (GFP).[3–9] Because of its spectral characteristics, GFP fluorescence provides a unique and sensitive means for detecting tagged proteins within a cellular environment. Maturation of the GFP chromophore and development of the unique GFP spectral signal are directly coupled to proper GFP folding.[10,11] Because of this property, there should be a correlation between the fate of the target protein and the fate of the GFP chromophore.[3] Thus, the level of fluorescence signal from a fused GFP molecule is expected to provide a reasonable measure of the ability of the candidate protein to be expressed at high levels in a properly folded form.

Here, we report the implementation of a *E. coli*-based whole cell, GFP fluorescence assay to survey the expression properties of a large cohort of membrane proteins (314 candidates) from extremophilic organisms. We show that measurement of the fluorescence produced by a C-terminal GFP fusion of the candidate protein relative to a set of benchmark GFP-fusion proteins provides a reliable way to identify membrane proteins that can be expressed in the 0.5–15 mg L^{-1} range in the membrane fraction following expression optimization. The production of such quantities facilitates further characterization by detergent extraction screens and size exclusion chromatography. Importantly, our results demonstrate that the use of the whole-cell fluorescence detection method provides a rapid way to assess the overexpression potential of many membrane protein with a limited investment of labor. This stands in sharp contrast to traditional, labor-intensive methods for expression screening using SDS-PAGE and Western blot. Analysis of candidate protein expression behavior with respect to protein properties suggests that although there appear to be no particular limits on protein size or number of transmembrane segments, proteins in the 20–50 kDa range and having <7 transmembrane segments have a higher success rate. Interestingly, we find no correlation between expression level and the degree to which the codon usage of the target gene matches the expression host.

Results

Fluorescence-based screening for identifying membrane proteins having overexpression potential

Following previous reports using GFP as a reporter for membrane protein expression,[4,7–9,12] we established an *E. coli* GFP-based expression system (Supporting Fig. 1) to evaluate the expression potential of a large cohort of membrane proteins from organisms found in extreme and unusual environments. To allow rapid clone generation, we constructed a ligase independent cloning vector[13] to facilitate direct cloning of PCR products of the gene of interest (Supporting Fig. 1). Target protein fusions contained a C-terminal tag that bore in sequence, a Tobacco Etch Virus (TEV) protease site,[14] eGFP,[15] and an octahistidine (His_8) tag [Fig. 1(a)]. On the basis of previous studies,[4,7–9,12] we anticipated that there was likely to be a correlation between proper folding and membrane insertion of the target protein and GFP fluorescence. One limit to this assay is that the C-terminal GFP moiety will only mature if the C-terminus of the target protein is intracellular.[7] As a consequence, any well-expressed membrane proteins having a topology that places the C-terminus extracellularly will be missed. An estimated 70% of integral membrane proteins have their C-terminus in the cytoplasm.[16] Given the expected high proportion of candidates with cytosolic C-termini, and the possibility that a single error in transmembrane protein prediction would eliminate potential candidates by inadvertently switching the predicted topology, we did not use topology prediction algorithms to prescreen candidates.

We chose candidate membrane proteins from organisms that live in extreme pH, high salt, the presence of high concentrations of heavy metals, high temperatures, or some combination of these extremes, based on the rationale that proteins found in such organisms, and in particular those in the membrane, which should have direct contact with the harsh environment, are likely to be very stable and robust subjects for biochemical and crystallographic studies.[17] The criteria for selection of candidate organisms were that in addition to living in an unusual environment the selected organisms had to have a completely sequenced genome and have genomic DNA that could be obtained from the ATCC collection to be used as template for PCR.

Membrane proteins were selected for the expression survey using a combination of annotation searches for ion channel, transporter, exchanger, antiporter, carrier, and major facilitator superfamily members, and by searching the genome of each candidate organism with a generic transmembrane helix profile. From an initial list of ~400 selected candidates, we succeeded in making 314 GFP fusion constructs and tested their expression levels by measuring in cell GFP

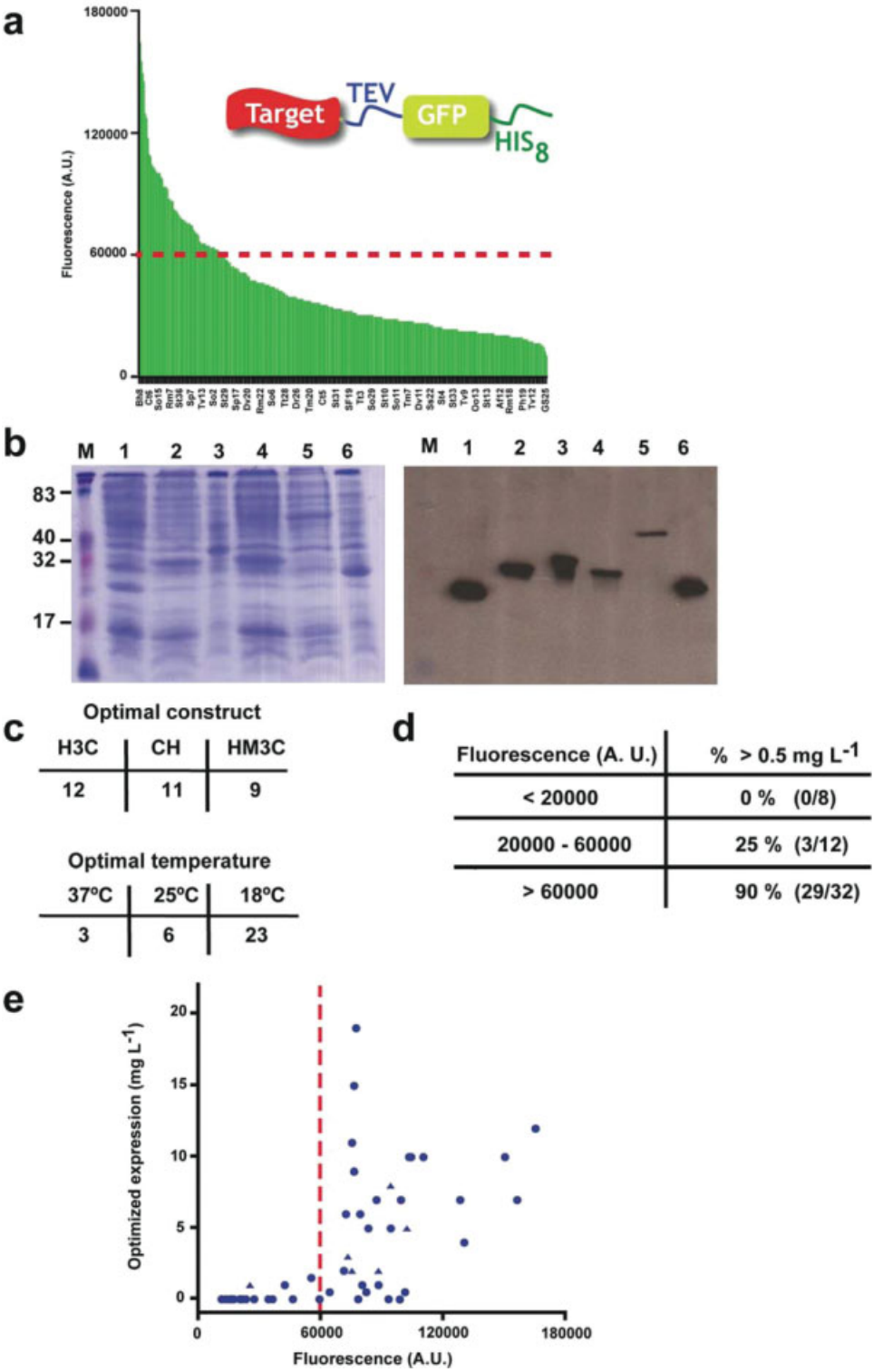


Figure 1. GFP-based integral membrane protein expression screen. (**a**) Fluorescence measurements of 314 GFP fusions assayed in *E. coli*. Dashed line indicates the cutoff threshold level based on the values obtained three benchmark proteins, KvAP, AQPz, and YhjX. Inset shows schematic of GFP constructs. (**b**) SDS-PAGE analysis of the C41 *E. coli* membrane fractions from several extremophile membrane protein candidates following expression optimization analyzed by coomassie-blue staining (left) and anti-polyhistidine tag western blot (right). Lanes are: M, Marker; 1, Af27 (22 kDa); 2, Rm29 (30 kDa); 3, Rm8 (30 kDa); 4, Ph11 (29 kDa); 5, Sp2 (49 kDa); 6, Rm4 (25 kDa). (**c**) Tables showing the numbers of proteins having optimal expression as H3C, CH, and HM3C affinity tagged constructs and showing the numbers of proteins having the optimal expression at different expression temperatures. (**d**) Percentage of optimized candidates with expression >0.5 mg L^{-1} of culture, grouped according to their fluorescence values from the GFP screen. (**e**) Comparison of GFP fluorescence versus final amount of protein following optimization. Expression levels were determined by following single-step metal affinity purification in FC-12 and measurement of the amount of protein purified by absorbance (circles) or by comparison with a protein standard by western blot (triangles).

Table I. *List of Extremophilic Organisms Selected, Number of Candidates Screened, Number Positive, and Hit Rates in Each Organism*

Organism	Environment	Candidates screened	>60,000 F.A.U.	
			# Positive	% Positive
Aeropyrum pernix	Hydrothermal vent	8	1	13
Archaeoglobus fulgidus	Hydrothermal vent, oil deposits, hot springs	21	7	33
Bacillus halodurans	pH >9.5, high salt, Dead Sea	5	1	20
Chlorobium tepidium	New Zealand hot spring	6	3	50
Deinococcus radiodurans	Radiation resistant	14	3	21
Desulfovibrio vulgaris	Tolerates heavy metals	24	3	13
Geobacter sulfurreducens	Hydrocarbon contaminated soils, tolerates heavy metals	13	2	15
Oenococcus oeni	Low pH, high alcohol	17	1	6
Pyrococcus furiosus	Geothermal marine sediments	6	1	17
Pyrococcus horikoshii	Hydrothermal vent	21	3	14
Ralstonia metallidurans	Heavy metal containing industrial wastes	25	11	44
Shewanella oneidensis	Heavy metal reducer	26	8	31
Silicibacter pomeroyi	Sea water, sulfur reducer	12	6	50
Sulfolobus solfataricus	Volcanic hot springs	19	0	0
Streptococcus thermophilus	Yoghurt, cheese, low pH	30	7	23
Thermotoga maritima	Geothermal marine sediments	20	3	15
Thermus thermophilus	Yellowstone hot spring	18	3	17
Thermoplasma volcanium	Hot springs, low pH	29	1	3
Total		314	64	20

fluorescence following induction and expression at 25°C (Supporting Table I).

Comparison of expression using a simple fluorescence assay that measured the GFP fluorescence signal from a standardized density of cells of the 314 GFP fusions yielded a wide distribution of fluorescence values, 20,000–180,000 F.A.U. (Fluorescence Arbitrary Units) [Fig. 1(a), Supporting Table I]. Uninduced cells grown to similar density yielded a background value of ~20,000 F.A.U. To set a criterion for the level of expression that would indicate the most promising candidates, we examined the performance of three benchmark membrane protein-GFP fusions. Two of the benchmark proteins were chosen because they have had X-ray crystallographic structures determined from *E. coli* expressed material, AQPz[18,19] and KvAP.[20,21] The third benchmark protein, YhjX, was the best expressed membrane protein in a GFP-based expression survey of *E coli* membrane proteins.[6] Our rationale was that expression of any test protein-GFP fusion at levels equivalent to or better than these benchmarks should indicate the production of acceptable levels of properly folded material for further characterization. On the basis of the benchmark protein-GFP fusion values, AQPz (73,000 F.A.U.), KvAP (118,000 F.A.U.), and YhjX (97,000 F.A.U.), we defined a positive threshold of 60,000 F.A.U. This value was chosen to allow a generous lower bound for candidates that should be as good or better than the benchmark proteins. Sixty-four of the screened candidates (20%) yielded fluorescence values at or above the threshold (60,000 A.U.). Interestingly, we obtained vast differences in success among the extremophilic organisms (Table I). *C. tepidium*, *R. metallidurans*, and *S. pomeroyi* were good sources of well-expressed candidates, having success rates as high as 40–50%, whereas *O. oeni*, *S. solfataricus*, and *T. volcanium* proved to be very poor sources having very low success rates, 0–6%. Further analysis revealed that certain membrane protein classes express better than others (Table II). For example, candidates from the mechanosensitive ion channel and from the

Table II. *Expression Success Analyzed by Family*

Membrane protein class	Candidates tested	% of total candidates	>60,000 F.A.U.	% of candidates >60,000 F.A.U.
Antiporter	25	8.0	3	12.0
Exchanger	12	3.8	2	16.6
Ion channel	24	7.6	6	25.0
MATE transporter	6	1.9	2	33.3
MotA proton channel	7	2.2	5	71.4
Mechanosensitive ion channel	28	8.9	15	53.5
Transporter	177	56.4	27	15.2
Other	35	11.1	4	11.4
Total	314		64	20.3

Analysis of success rate by membrane protein class.

flagellar motor MotA proton channel families had high success rates (53.5% and 71.4%, respectively), whereas candidates from the antiporter, exchanger, and transporter families were less likely to express well (12.0%, 16.6%, and 15.2%, respectively). The origins of the differences are unclear but may relate to a combination of protein molecular weight, total number of transmembrane segments, toxic consequences of protein activity from the overexpressed protein, and the potential requirement for auxiliary subunits that were not included in the expression study.

Overexpression optimization without GFP

Initial attempts at direct purification of a number of GFP constructs having high F.A.U. scores (Af11, Bh8, Ct6, and Dr35) proved problematic and did not directly yield sufficient quantities of purified material for further characterization. In consideration of the possibility that the GFP itself was proving a burden to production of the candidate protein, we cloned 52 candidates from three different pools; 32 high-expressing proteins (>60,000 F.A.U.), 12 medium to low expressing proteins (60,000–20,000 F.A.U.), and 8 low expressing proteins (<20,000 F.A.U.) into three affinity tag formats for protein expression optimization experiments. These 156 constructs were subjected to overexpression optimization trials in which each construct was grown as a small scale culture (7 mL) and tested for expression at three different induction temperatures (18°C, 25°C, and 37°C). Expression levels for each sample were assayed evaluating SDS-PAGE by coomassie blue staining and western blot using an anti-polyhistidine antibody.

The optimization experiments with non-GFP constructs caused clear improvements compared to the temperature-optimized GFP fusion construct of the same target (Supporting Fig. 2). Figure 1(b) shows examples of membrane fractions for six of the candidate proteins following expression optimization. Thirty-two candidates of the 52 showed expression by both coomassie-staining and Western blot analysis. Both the affinity tag and expression temperature were important variables for expression optimization [Fig. 1(c)]. No particular tag predominated as the best choice. In general, a low induction temperature, 18°C, proved most beneficial. Comparison of the amount of protein produced under optimized expression conditions by using single step using metal affinity purification in Fos-choline-12 (FC-12), which was found as the most robust detergent for membrane protein extraction (see below), followed by quantification of the amount of protein expressed in the membrane fraction showed clear differences between the ability to convert proteins from different GFP score categories into expressed material [Fig. 1(d,e)]. Strikingly, we were able to produce 90% of the candidates from the high-expressing category at levels of ≥ 0.5 mg L^{-1} [Fig. 1(d)] with 50% of these made at ≥ 5 mg L^{-1}. The success rate for producing ≥ 0.5 mg L^{-1} of protein from the medium to low expressing category was also good, 25%. None of the candidates from the category having GFP values near the background fluorescence value gave expression levels >0.5 mg L^{-1} following optimization. Comparison of the amounts of expressed proteins following optimization with the GFP screen fluorescence values shows that the fluorescence value served as a reliable indicator of candidate proteins that could be expressed in the membrane fraction at a level that is useful for further biochemical characterization [Fig. 1(e)]. On the basis of these success rates, we estimate that 90% of the candidates that passed the GFP screen could be produced at ≥ 0.5 mg L^{-1} culture if subjected to optimization experiments. These results highlight the value of including benchmark standards in execution of GFP-based protein expression surveys and indicate that a simple GFP screen should be sufficient for initial identification of proteins with a high-likelihood of overexpression potential.

Detergent screening by size exclusion chromatography

Detergents are a necessary component of any membrane protein purification effort and constitute a complex variable that requires extensive investigation.[22,23] Detergent choice has significant implications for protein oligomeric state, stability, homogeneity, and crystallization. Unfortunately, it is impossible to predict which detergent will be suitable for extraction, purification, and crystallization of a given membrane protein. Having sufficient expressed material for a variety of candidate proteins from different membrane protein classes following our expression optimization experiments, we screened candidates for detergent solubilization and size exclusion profile behavior using a panel of several different detergents. These experiments were aimed at assessing whether the overexpressed material could yield protein that would be of suitable quality for further biophysical characterization and crystallization trails. We chose 24 well-expressed membrane proteins and used a panel of six commonly used detergents decyl-β-D-maltoside (DM), octyl-β-D-glucopyranoside (OG), dodecyl-β-D-maltoside (DDM), Fos-choline-12 (FC-12), *n*-dodecyl-*N*,*N*-dimethylamine-*N*-oxide (LDAO), and 3-[(3-cholamidopropyl) dimethylammonio]-2-hydroxy-1-propanesulfonate (CHAPSO). Following extraction from the membrane fraction and a single metal affinity purification step, 21 of the 24 candidates gave an included peak in the gel filtration profiles in at least one detergent (Fig. 2, Table III). Ten of the 24 were found to be monodisperse based on the presence of a single, symmetrical peak (Fig. 2, Supporting Table II). The presence of a monodisperse size exclusion profile has been shown to be a good indicator of membrane protein samples that are of suitable quality for further biochemical, biophysical, and crystallization studies.[8,9,24–27] It is difficult to

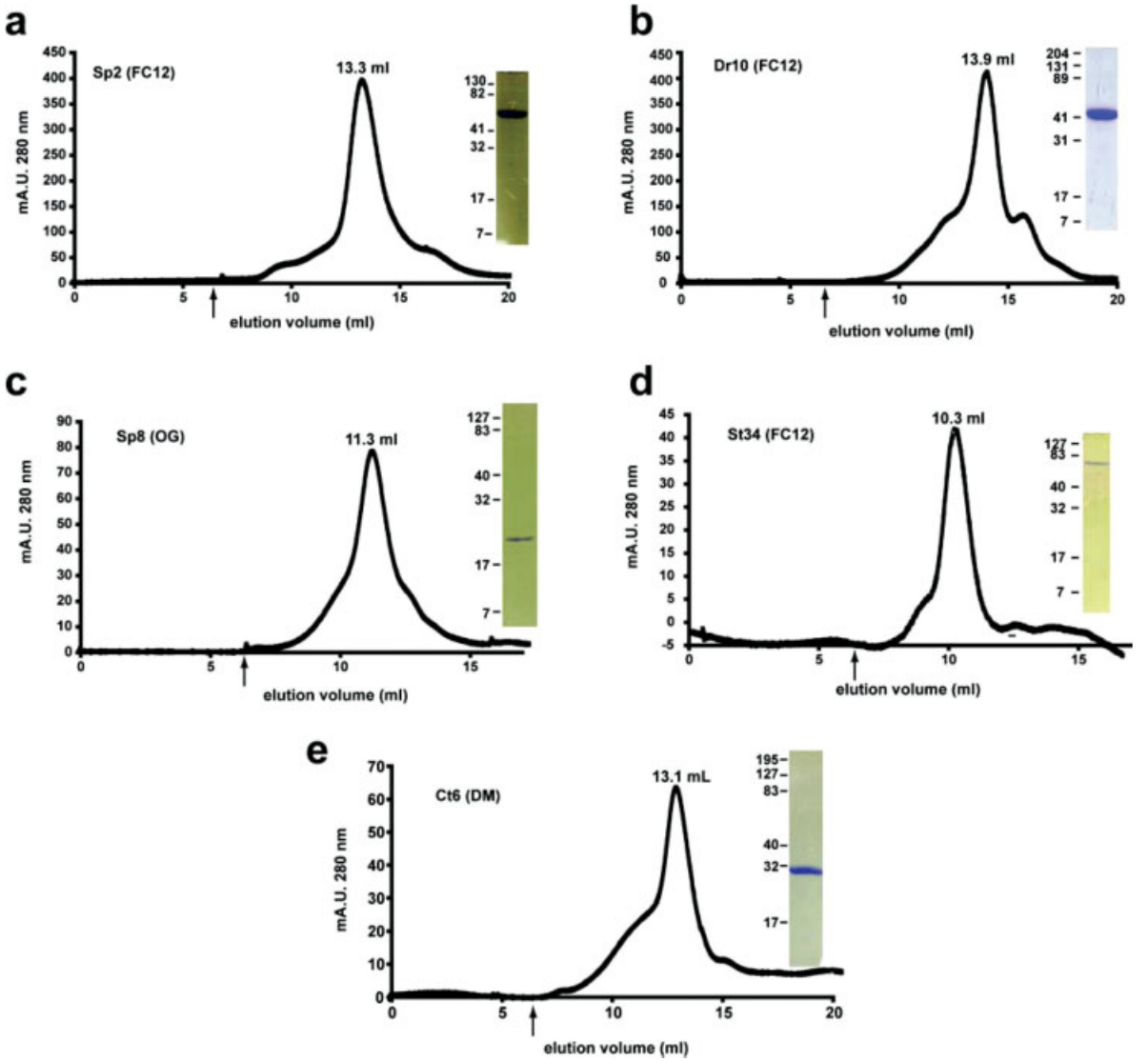


Figure 2. Superdex200 gel filtration profiles and SDS-PAGE analysis of exemplar candidates following batch Ni-NTA purification. The detergent used is indicated in parentheses. Elution volume and coomassie-blue stained SDS-PAGE of the main gel filtration peak fraction are shown. Arrow indicates the void volume. Sample identity, annotation, monomer molecular weight, and approximate size of the observed protein detergent complex size are: (**a**) Sp2, MscS channel, 49 kDa, ~120 kDa; (**b**) Dr10, MscS channel, 40 kDa, ~95 kDa; (**c**) Sp8, MscL channel, 15 kDa, ~180 kDa; (**d**) St34, ABC transporter, 65 kDa, ~320 kDa; (**e**) Ct6, proton channel, 27 kDa, ~110 kDa (also see Supporting Table II).

estimate the oligomeric state of a membrane protein from size exclusion results, particularly as the standards are soluble, globular proteins, and the exact biological assembly of the particular proteins we examined is unknown. Nevertheless, we found that all of the monodisperse samples ran with apparent molecular weights that are larger than that expected for monomers. This property was maintained regardless of the detergent type in the cases where more than one detergent yielded a monodisperse peak (Fig. 2, Supporting Table II).

Interestingly, we found that FC-12 was the most successful detergent in terms of both solubilization and behavior in gel filtration. DDM was the second most effective followed by LDAO, OG, and DM. CHAPSO proved to be least effective at both solubilization and in yielding protein samples having good size exclusion chromatography behavior. The rank order of effectiveness in terms of the ability of specific detergents to produce samples having monodisperse profiles is FC12, DDM, LDAO and is similar to that reported by Drew and colleagues for a recent survey of eukaryotic membrane proteins.[9]

Fos-choline detergents have been used successfully to purify, crystallize, and determine membrane protein structures,[28,29] for membrane protein NMR studies,[30] and for electron microscopy studies of transporter oligomers.[31] FC12 has further been shown to be

Table III. *Detergent Extraction and Gel Filtration Behavior of Overexpressed Extremophile Membrane Proteins*

Target	Construct	Induction temp. (°C)	Yield (mg/L)	Detergents						Annotation
				DM	OG	DDM	FC12	CHAPSO	LDAO	
Ct6	HM3C	37	~10	++	−	++	++	−	++	MotA/TolQ/ExbB proton channel
Rm3	H3C	18	6	n.p.	n.p.	n.p.	+	n.p.	n.p.	MotA/TolQ/ExbB proton channel
Gs26	CH	18	1.5	n.a.	n.a.	n.a.	−	n.a.	n.a.	MotA/TolQ/ExbB proton channel
Rm4	CH	18	10	−	n.p.	+	++	−	−	MotA/TolQ/ExbB proton channel
Dr10	H3C	18	9	−	n.p.	++	++	−	−	Mechanosensitive ion channel, MscS
Rm8	CH	18	7	n.p.	n.p.	n.p.	n.p.	n.p.	n.p.	Mechanosensitive ion channel, MscS
Sp7	H3C	18	15	−	−	−	++	−	−	Mechanosensitive ion channel, MscS
Rm11	H3C	18	7	−	n.p.	−	+	−	+	Mechanosensitive ion channel, MscS
Sp2	H3C	18	11	−	n.p.	−	++	−	−	Mechanosensitive ion channel, MscS
Sp8	H3C	25	2	−	++	++	++	−	++	Mechanosensitive ion channel, MscL
Dr35	CH	18	0.5	n.p	n.a	n.a	+	n.a	n.a	Multidrug efflux transporter
Rm29	H3C	18	7	n.p.	n.p.	n.p.	++	n.p.	−	High-affinity Nickel transporter
Sp1	H3C	25	20	−	−	+	+	−	−	Voltage-gated sodium channel subunit
So15	H3C	25	>10	−	n.p	+	+	n.p	n.p	Iron compound, ABC transporter, permease
So22	H3C	18	<1	n.a	n.a	n.a	+	n.a	n.a	Sulfate ABC transporter, permease
Gs6	HM3C	25	~5	+	+	+	−	n.p	+	Heavy metal efflux pump
St34	CH	18	1	−	n.a	n.a	++	n.a	n.a	Lipid/multidrug/protein-type ABC exporter
Tm23	H3C	18	~0.5	n.p	n.a	n.a	+	n.a	n.a	Na^+/H^+ antiporter
Af11	HM3C	25	6	n.p	n.p	+	+	+	n.p	Potassium channel
Bh8	H3C	37	12	−	−	−	−	−	−	Potassium channel
Af27	CH	18	10	−	+	−	−	−	−	No significant homologues
Ph11	CH	18	10	+	+	+	+	−	++	Na^+/solute symporter
So25	HM3C	18	8	+	n.p	+	+	−	+	Multidrug resistance
Dr18	CH	18	1	n.a	++	++	++	n.a	n.a	Multidrug resistance

List of purified candidates showing several parameters including type of construct, induction temperature, yield/L, and gel-filtration behavior in selected detergents. ++, mondisperse peak; +, good gel-filtration profile; −, poor gel-filtration profile; n.p., no gel-filtration profile; n.a, not tested.

efficient at membrane protein solubilization and at yielding monodisperse samples as judged by size exclusion chromatography.[7] Nevertheless, concerns have been raised that even though Fos-choline detergents are efficient at extracting proteins from membrane, they may be harsh enough to need to be treated with caution[27] and may not always yield a stable sample.[32] It is notable that 11 of the 19 proteins that gave good gel filtration profiles in FC12 also gave similar gel filtration profiles in at least one other more mild detergent. This result suggests that, at least in these cases, the samples were not likely to be simply solubilized and denatured by FC12.

Our detergent extraction and gel filtration experiments identify many candidates that were able to produce protein that gave good gel filtration behavior in at least one detergent following a very simple purification protocol (21/24 candidates) (see Fig. 4). This observation provides further support for the idea that the original GFP screen was a good proxy for identifying membrane proteins that have the potential to yield material of suitable quality for biophysical experiments.

Discussion

Although there are now a number of reports in which GFP-based screens have been used to identify proteins or constructs that have appropriate biophysical characteristics for structural biology experiments,[3,5–9] there have remained some questions regarding the fidelity of the assay. Our results establish that there is an excellent correspondence between the behavior of membrane protein GFP-fusions and the potential of a given candidate to be overproduced. Of the candidates that pass the screen and were subjected to further characterization, 90% can be produced at $\geq$0.5 mg L^{-1} of culture, with half of these made at $\geq$5 mg L^{-1}. These expression scales are suitable for investing further effort into purification, biophysical characterization, and crystallization trials. An important feature of this work is that the screening can be done manually and requires no specialized robotic equipment. Thus, such screens should be applicable in a wide range of settings.

Membrane protein expression: influence of molecular weight, transmembrane segment number, and codon usage

Our survey of the expression behavior of a large number of membrane proteins from diverse sources provides a dataset that yields insight into the factors that influence the ability to overproduce a candidate membrane protein [Fig. 3(a,b)]. There are some clear general lessons regarding the likelihood of success for a candidate based on a few simple properties. In general, the total molecular weight of the candidate appears to be an important factor. We found many well expressed membrane proteins in the 20–50 kDa

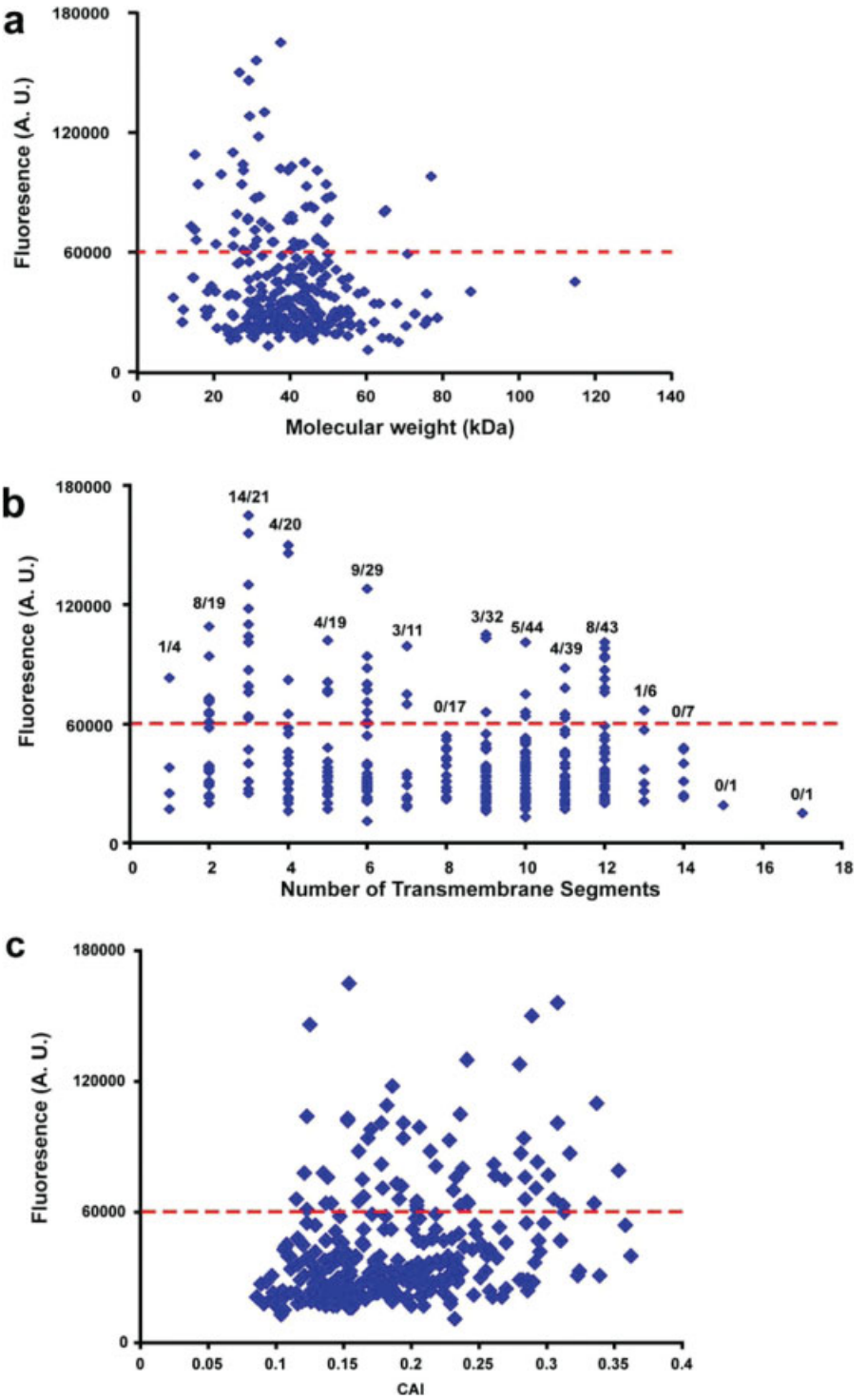


Figure 3. Distribution of MW, TMs, and codon usage among the GFP-screened candidates with respect to their fluorescence scores. (**a**) Molecular weight, (**b**) Predicted number of transmembrane helices. Fraction expressed per class: 1 TM (25%), 2 TM (42%), 3 TM (67%), 4 TM (20%), 5 TM (21%), 6 TM (31%), 7TM (27%), 8 TM (0%), 9 TM (9%), 10 TM (11%), 11 TM (10%), 12 TM (18%), 13 TM (17%), 14 TM (0%), 15 TM (0%), 17 TM (0%). (**c**) CAI for each gene, calculated using RSCU values from the class II highly-expressed *E. coli* gene reference set.[33] Higher CAI values correspond to better matches with the host codon usage.

range, (59/262, 22.5%), but very few in the >50 kDa range (5/52, 9.6%). Similarly, there is a clear difference in the success rates with respect to the number of predicted transmembrane segments; ~35% of the 123 candidate proteins having 1–7 transmembrane express better than the cutoff whereas only 11% of the 191

proteins having 8–17 transmembrane segments expressed well. It is particularly striking that 67% of the proteins having three predicted transmembrane segments expressed as well as or better than the benchmark proteins.

There is a general notion that one of the parameters that may have an impact on the ability to overexpress a candidate protein in *E. coli* is how suited the codon usage of the candidate gene is to well-expressed proteins from host organism.[34,35] However, data supporting this belief are equivocal. There are anecdotal reports of successes,[36,37] failures,[38] and variable outcomes[39] for particular cases of codon optimization. Large scale study of a set of 1000 soluble proteins from *Plasmodium falciparum,* an organism with proteins that are known to be particularly challenging to express in *E. coli*, found no significant correlation with codon usage.[40] As we had a diverse set of proteins from many organisms that have a range codon usages, we asked whether there was any correlation between how well the candidate proteins expressed and the similarity of their codon usage to that of well-expressed *E. coli* genes as measured by the "codon adaptation index," CAI. This parameter gives a length independent assessment of how well each test gene matches the codon usage of well-expressed proteins from the host.[33] Strikingly, for those candidates that pass our threshold GFP value, we find no correlation between the CAI for a given gene and expression level [Fig. 3(c)]. Indeed, a number of the proteins having the highest GFP values can be produced at very high levels following optimization in spite of having very low CAI scores (Ph11, CAI = 0.123, 10 mg L^{-1}; So15, CAI = 0.153, >10 mg L^{-1}; Bh8 CAI = 0.154, 12 mg L^{-1}). Even though individual rare codons may certainly cause expression difficulties,[34,35] our data support the conclusion that codon optimization is not a key parameter for heterologous membrane protein expression in *E. coli*.

Comparisons other GFP-membrane protein screens

Given the challenges associated with membrane protein biochemistry, there have been a number of reports describing the implementation of GFP-based screens as a means to identify potentially tractable candidate proteins. Although large scale GFP screens in *E. coli*,[6] *S. cereviseae*[9] have reported success in identification of well-expressed candidates, multi-milligram expression levels of non-host derived proteins per liter of culture have not yet been generally demonstrated. A key advance reported here is the use of benchmark proteins to set a threshold level for well-expressed candidates paired with construct and expression temperature optimization protocol. Together, this approach identifies a large number of candidate proteins that can be made at $\geq$0.5 mg L^{-1} with some candidates expressing better than 10 mg L^{-1}. Similar to the recently reported *S. cerevisiae* survey,[9] we find that Fos-choline 12 (FC-12) has the best success at both solubilizing membrane expressed material and producing material having a good gel filtration profile (Table III). The high success with fos-choline detergents in membrane protein solubilization has raised some of the concern regarding the potential destabilizing effects of this detergent class.[27,32] Importantly, a large number of the expressed extremophile membrane proteins that can be extracted from the membrane using FC-12 are also extracted from the membrane and give good gel filtration profiles in other, more conventionally used detergents.

Methods

Target selection and sequence analysis

Annotation searches for membrane proteins denoted as ion channels, transporters, exchangers, antiporters, carriers and major facilitator superfamily in the genomes of *Aeropyrum pernix* (Ap), *Archaeoglobus fulgidus* (Af), *Bacillus halodurans* (Bh), *Chlorobium tepidium* (Ct), *Deinococcus radiodurans* (Dr), *Desulfovibrio vulgaris* (Dv), *Geobacter sulfurreducens* (Gs), *Oenococcus oeni* (Oo), *Pyrococcus furiosus* (Pf), *Pyrococcus horikoshii* (Ph), *Ralstonia metallidurans* (Rm), *Shewanella oeneidensis* (So), *Silicibacter pomeroyi* (Sp), *Streptococcus thermophilus* (St), *Sulfolobus solfataricus* (Ss), *Thermotoga maritima* (Tm), *Thermus thermophilus* (Tt), and *Thermoplasma volcanium* (Tv) were performed using the Integr8 protein database (http://www.ebi.ac.uk/integr8). To find unannotated the membrane proteins, each genome was searched with a sequence representing a generic transmembrane helix using BLASTP[41] at the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov). This search yielded a list of membrane proteins in the chosen organism from which we selected several proteins with no homologues of known structure. All transmembrane helix predictions were performed with TMHMM 2.0[16] (http://www.cbs.dtu.dk/services/TMHMM-2.0). A complete list of screened proteins and values from the GFP screen can be found in Supporting Table I.

Construct generation

All genes were cloned using ligation-independent cloning (LIC)[13] of PCR products containing LIC overhangs and directly cloned into vector containing a complementary LIC cassette (Supporting Fig. 1). For the GFP screening, genes encoding candidate membrane proteins were cloned into a pET-based LIC vector that harbors a Tobacco etch virus protease cleavage site followed by a C-terminally His_8 tagged GFP. The three vectors used in overexpression optimization experiments were H3C (N-terminal His_8 tag-3C protease cleavage site[42] followed by the gene of interest), HM3C (N-terminal His_6 tag-maltose binding protein-3C protease cleavage site followed by the gene of

interest), and CH (gene of interest followed by a TEV protease cleavage site and a His_8 tag). The GFP and CH vectors share the same LIC cassette, whereas, the H3C and HM3C vectors share a different LIC cassette (Supporting Fig. 1). PCR reactions were done using Phusion high-fidelity DNA polymerase (Finnzymes) and genomic DNA for each parent organism. Genomic DNA was obtained from ATCC (www.atcc.org).

Expression of GFP-fusion and non-GFP fusion membrane proteins

Two different cell types of *E. coli* were used. BL21 (DE3) pLysS were used for screening of GFP-fusion candidates and C41 cells[43] were used for expression optimization and protein production. BL21 (DE3) pLysS cells were used for the GFP screen owing to the observation that they are reported to be less sensitive to variations in $OD_{600\ nm}$ of induction than the C41 strain.[7]

GFP screening

Individual *E. coli* BL21 (DE3) pLysS cultures for each of the candidate GFP-fusions were grown in 7 mL Luria-Bertani (LB) media at 37°C to $OD_{600\ nm}$ ~0.5. Induction was done at 25°C using 0.4 m*M* IPTG. After 20 h of growth post-induction, the cell pellet was recovered from 1 mL of culture (2 min at 13,000 rpm in a bench top 5415D centrifuge (Eppendorf)) and resuspended in 200 μL of 200 m*M* KCl, 20 m*M* Tris pH 8.0. Samples were subjected to the whole-cell fluorescence assay (465–485 nm excitation, 515–575 nm emission) using a $20/20^n$ luminometer with a Blue fluorescent module (Turner Biosystems, U.S.A.). In each experiment uninduced cells and cells expressing the GFP fusion of BH8, a well-expressed candidate protein, were run as controls. These controls have reproducible fluorescence measurements over a number of trials, F.A.U. of 23,667 ± 1021, $n = 6$, and 174,964 ± 5397, $n = 6$, for uninduced and BH8 expressing cells, respectively.

Expression optimization using non-GFP constructs

Proteins for overexpression optimization were selected based on the fluorescence values from the GFP screen. Genes for the selected proteins were cloned into H3C, CH, and HM3C tag formats and transformed into C41 cells.[43] Cultures were grown in 7 mL LB media at 37°C. Induction was performed at $OD_{600\ nm}$ ~0.5 using 0.4 m*M* IPTG. Immediately following induction, cultures were moved to one of three different expression temperatures, 18°, 25°, or 37°C, and were grown for 48, 20, and 3 h, respectively. Following expression, 500 μL of cells were harvested by centrifugation (2 min at 13,000 rpm in a bench top 5415D centrifuge (Eppendorf), resuspended in 50 μL of 200 m*M* KCl, 20 m*M* Tris, pH 8.0, and lysed by the addition of 50 μL of 2% SDS, 20% glycerol, 0.05% bromophenol blue, 125 m*M* Tris pH 6.8. Expression was assayed by SDS-PAGE analysis with Coomassie staining and Western blot using an anti-His_5 HRP conjugate antibody (Qiagen). The best-optimized condition from this assay was used for a scale-up.

Membrane preparations

C41 *E. coli* cells harboring the expression construct of interest were grown in 1 L 2xYT (5 g NaCl, 16 g tryptone, 10 g yeast extract) under conditions that were determined to give the best expression for each individual construct under investigation. Cell pellets obtained from the 1 L culture were suspended in 90 mL lysis buffer (200 m*M* KCl, 20 m*M* Tris pH 8.0, 8% glycerol, 2 m*M* EDTA, 1 m*M* PMSF, 1 m*M* DTT) and disrupted by sonication for 4 min on ice using a Sonic Dismembrator model 500 (Fisher). Cell lysate was centrifuged in a JA 25.50 rotor (Beckman Coulter) at 7000*g* for 35 min at 4°C to remove unbroken cells and debris. The supernatant was then centrifuged in a Ti45 rotor (Beckman Coulter) at 40,000 rpm for 2 h at 4°C to separate the membrane pellet. Pellets were homogenized in a storage buffer (200 m*M* KCl, 20 m*M* Tris pH 8.0, 8% glycerol) with a Dounce Tissue Grinder (Kimble Kontes LLC) and frozen at −80°C.

Membrane protein extraction, purification, and characterization

All extraction, purification, and characterization were done at 4°C. Membrane extractions and purifications were performed with six detergents (Anatrace, USA): DM (40 m*M*), OG (200 m*M*), LDAO (2% v/v), DDM (20 m*M*), CHAPSO (160 m*M*), and FC12 (40 m*M*), added and incubated for 2 h at 4°C. Detergent-solubilized membrane proteins were separated from insoluble material by ultracentrifugation for 45 min at 55,000 rpm at 4°C in a TLA 110 rotor on a bench-top ultracentrifuge (Beckman Coulter). Following recovery, the supernatant was incubated with a bed volume of 0.5 mL of pre-equilibrated Ni-NTA superflow beads (Qiagen) at 4°C. Protein-bound beads were washed with the appropriate detergent buffer consisting of 200 m*M* KCl, 30 m*M* imidazole, 8% glycerol (v/v), detergent (3× critical micelle concentration (CMC), except for LDAO which was used at 10× CMC), 20 m*M* Tris, pH 8.0 and eluted with the same detergent buffer containing 300 m*M* imidazole. The eluted protein samples were >80–90% pure based on SDS-PAGE analysis. Protein concentration was determined by absorbance at 280 nm.[44] Purified samples were subjected to size exclusion chromatography on a Superdex200 column (GE Healthcare) to assess the homogeneity and stability in a running buffer of 200 m*M* KCl, 8% glycerol (v/v), detergent (3× CMC, except for LDAO which was used at 10× CMC), 20 m*M* Tris, pH 8.0.

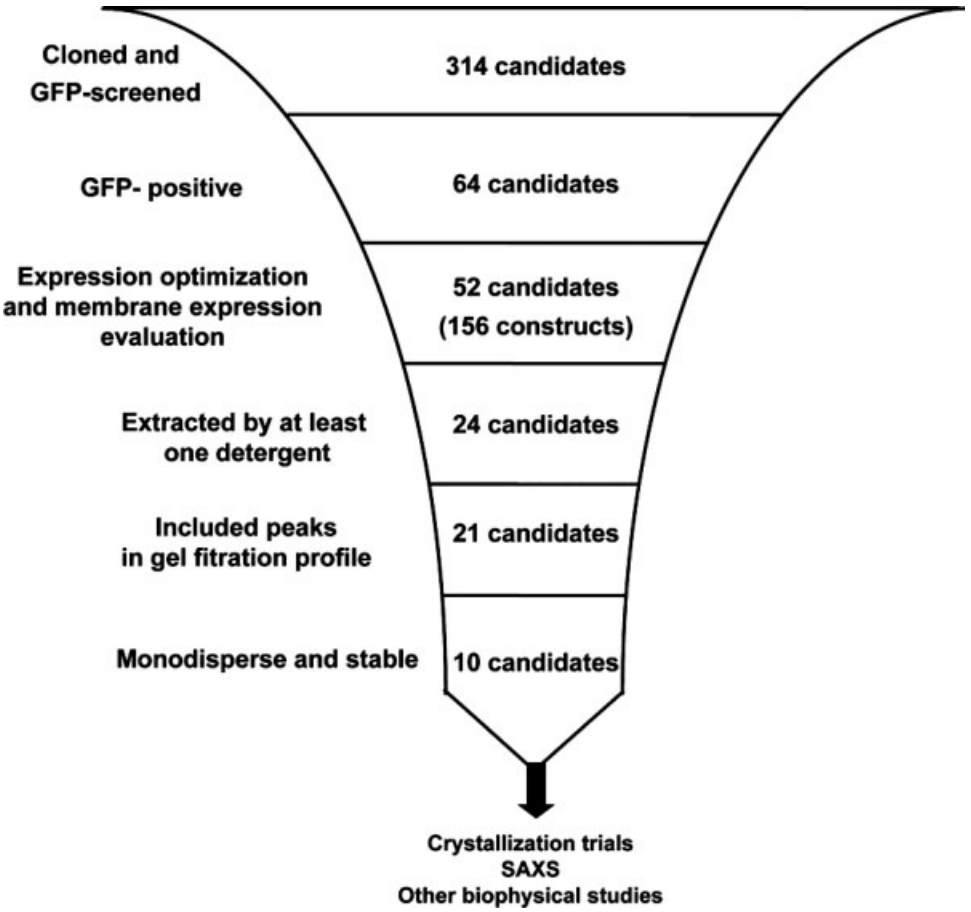


Figure 4. Summary of extremophile membrane protein GFP screen. Results are presented as a funnel. At each stage some candidates are eliminated due to a failure to pass progressively more stringent tests. The exception is the step from GFP-positive to expression optimization in which 32 GFP-positive proteins along with 20 representatives from categories having lower GFP signals were chosen for expression optimization and further testing. Candidates passing the final stage of being monodisperse and stable are suitable for immediate evaluation by crystallization trials, small-angle X-ray scattering (SAXS), or other biophysical and functional analysis.

Codon adaptation index calculation

Codon adaptation index (CAI) values for each candidate gene were calculated according to Sharp and Li[33] using the webserver (http://www.evolvingcode.net/codon/cai/cais.php).[45] CAI $= \mathrm{CAI}_{obs}/\mathrm{CAI}_{max}$ where the observed CAI for a given gene sequence $\mathrm{CAI}_{obs} = (\prod_{k=1}^{L} RSCU_k)^{1/L}$ and the maximal CAI score is $\mathrm{CAI}_{max} = (\prod_{k=1}^{L} RSCU_{k\,max})^{1/L}$. RSCU_k is the "relative synonymous codon usage" value for the kth codon in the gene, RSCU_{kmax} is the maximum RSCU value for the amino acid encoded by the kth codon in the gene, and L is the number of codons in the gene. $\mathrm{RSCU}_{ij} = x_{ij} \Big/ \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ where x_{ij} is the number of occurrences of the jth codon for the ith amino acid and n_i is the number (from 1–6) of alternative codons for the ith amino acid. RSCU_{kmax} values were calculated from the class II highly-expressed *E. coli* gene reference set.[33]

Conclusions

Producing membrane proteins and membrane protein complexes for biochemical and structural study remains a major challenge. The development of methods that provide a rapid way to identify membrane proteins that can be expressed in sufficient amounts and of sufficient quality is an area of intensive effort.[4,6–9,46] This report provides strong evidence that screening GFP-fusions provides a fruitful approach for identifying well-expressed candidate membrane proteins that display favorable biophysical behaviors suitable for further functional and structural investigation, and that a key factor in success is the investment in optimization of construct tags and induction temperature.

Structural and biochemical studies of integral membrane proteins remains challenging because identifying a candidate protein that has good expression, that can be readily purified, and that has suitable behavior in detergent solution can require a

substantial investment in time and resources. By combining a GFP-screen with a simple optimization protocol, we have been able to achieve a high success rate identifying candidate proteins that can be well expressed and that display desirable biophysical characteristics. A key implementation was the use of benchmark proteins to set a level of GFP fluorescence that will identify a candidate with a high likelihood of success. We anticipate that the large number of diverse membrane proteins we identify here as having good expression profiles will provide excellent starting points for both functional and structural studies. A summary of the outcome of the screen is shown in Figure 4. As with other expression surveys of panels of membrane proteins in which multiple levels of behavior are evaluated,[27] the approach we report here functions as a funnel in which the there is an attrition of candidates as the stringency imposed at each step increases. It may be possible to salvage some failed candidates by additional exploration of alternate constructs that are designed to eliminate regions of predicted polypeptide disorder. Further, just as combination of mutagenesis and GFP screens has proved useful for defining appropriate constructs of soluble proteins for crystallization and structure determination,[47,48] it is likely similar combination mutagenesis/GFP screens should prove fruitful for enhancing the production of crystal grade membrane protein samples or for rescuing failed candidates. Similarly, GFP-based studies using other expression hosts and as a means to identify well-expressed complexes through co-expression strategies[49–51] should enable the structural investigation of an array of membrane proteins and membrane protein complexes.

Acknowledgment

We thank K. Brejc, J.M. Berger, and M. Pioletti for comments on the manuscript and members of the Minor lab for input at all stages of this work. DLM is an AHA Established Investigator.

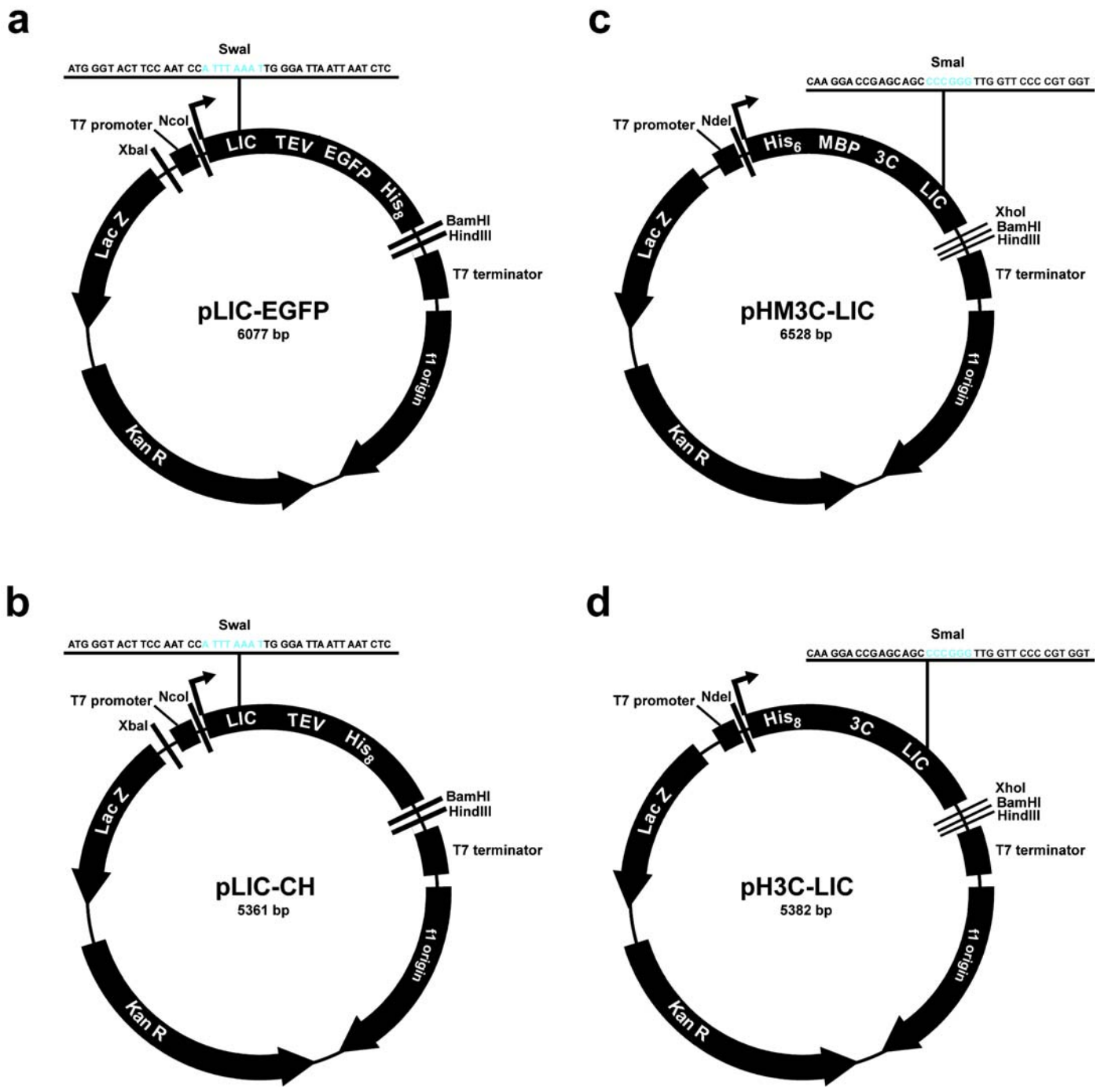
References

1. Grisshammer R, Tate CG (1995) Overexpression of integral membrane proteins for structural studies. Q Rev Biophys 28:315–422.
2. Lacapere JJ, Pebay-Peyroula E, Neumann JM, Etchebest C (2007) Determining membrane protein structures: still a challenge! Trends Biochem Sci32:259–270.
3. Waldo GS, Standish BM, Berendzen J, Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol 17:691–695.
4. Drew DE, von Heijne G, Nordlund P, de Gier JW (2001) Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*. FEBS Lett 507:220–224.
5. Cabantous S, Pedelacq JD, Mark BL, Naranjo C, Terwilliger TC, Waldo GS (2005) 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 6:113–119.
6. Drew D, Slotboom DJ, Friso G, Reda T, Genevaux P, Rapp M, Meindl-Beinker NM, Lambert W, Lerch M, Daley DO, Van Wijk KJ, Hirst J, Kunji E, De Gier JW (2005) A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. Protein Sci 14:2011–2017.
7. Drew D, Lerch M, Kunji E, Slotboom DJ, de Gier JW (2006) Optimization of membrane protein overexpression and purification using GFP fusions. Nat Methods 3: 303–313.
8. Kawate T, Gouaux E (2006) Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure 14:673–681.
9. Newstead S, Kim H, von Heijne G, Iwata S, Drew D (2007) High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 104:13936–13941.
10. Reid BG, Flynn GC (1997) Chromophore formation in green fluorescent protein. Biochemistry 36:6786–6791.
11. Wachter RM (2007) Chromogenic cross-link formation in green fluorescent protein. Acc Chem Res 40:120–127.
12. Drew D, Sjostrand D, Nilsson J, Urbig T, Chin CN, de Gier JW, von Heijne G (2002) Rapid topology mapping of Escherichia coli inner-membrane proteins by prediction and PhoA/GFP fusion analysis. Proc Natl Acad Sci USA 99:2690–2695.
13. Aslanidis C, de Jong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res 18:6069–6074.
14. Kapust RB, Waugh DS (2000) Controlled intracellular processing of fusion proteins by TEV protease. Protein Expr Purif 19:312–318.
15. Zhang G, Gurtu V, Kain SR (1996) An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. Biochem Biophys Res Commun 227:707–711.
16. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580.
17. Jenney FE, Jr, Adams MW (2008) The impact of extremophiles on structural genomics (and vice versa). Extremophiles 12:39–50.
18. Savage DF, Egea PF, Robles-Colmenares Y, O'Connell JD, III, Stroud RM (2003) Architecture and selectivity in aquaporins: 2.5 a X-ray structure of aquaporin Z. PLoS Biol 1:E72.
19. Jiang J, Daniels BV, Fu D (2006) Crystal structure of AqpZ tetramer reveals two distinct Arg-189 conformations associated with water permeation through the narrowest constriction of the water-conducting channel. J Biol Chem 281:454–460.
20. Jiang Y, Lee A, Chen J, Ruta V, Cadene M, Chait BT, MacKinnon R (2003) X-ray structure of a voltage-dependent K+ channel. Nature 423:33–41.
21. Lee SY, Lee A, Chen J, MacKinnon R (2005) Structure of the KvAP voltage-dependent K+ channel and its dependence on the lipid membrane. Proc Natl Acad Sci USA 102:15441–15446.
22. Wiener MC (2004) A pedestrian guide to membrane protein crystallization. Methods 34:364–372.
23. Prive GG (2007) Detergents for the stabilization and crystallization of membrane proteins. Methods 41: 388–397.
24. Michel H (1982) Three-dimensional crystals of a membrane protein complex. The photosynthetic reaction centre from Rhodopseudomonas viridis. J Mol Biol 158: 567–572.

25. Borths EL, Locher KP, Lee AT, Rees DC (2002) The structure of *Escherichia coli* BtuF and binding to its cognate ATP binding cassette transporter. Proc Natl Acad Sci USA 99:16642–16647.
26. Daniels BV, Jiang JS, Fu D (2004) Crystallization and preliminary crystallographic analysis of the *Escherichia coli* water channel AqpZ. Acta Crystallogr D Biol Crystallogr 60:561–563.
27. Lewinson O, Lee AT, Rees DC (2008) The funnel approach to the precrystallization production of membrane proteins. J Mol Biol 377:62–73.
28. Bass RB, Strop P, Barclay M, Rees DC (2002) Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. Science 298:1582–1587.
29. Wang W, Black SS, Edwards MD, Miller S, Morrison EL, Bartlett W, Dong C, Naismith JH, Booth IR (2008) The structure of an open form of an *E. coli* mechanosensitive channel at 3.45 Å resolution. Science 321:1179–1183.
30. Baker KA, Tzitzilonis C, Kwiatkowski W, Choe S, Riek R (2007) Conformational dynamics of the KcsA potassium channel governs gating properties. Nat Struct Mol Biol 14:1089–1095.
31. McDevitt CA, Collins RF, Conway M, Modok S, Storm J, Kerr ID, Ford RC, Callaghan R (2006) Purification and 3D structural analysis of oligomeric human multidrug transporter ABCG2. Structure 14:1623–1632.
32. Carpenter EP, Beis K, Cameron AD, Iwata S (2008) Overcoming the challenges of membrane protein crystallography. Curr Opin Struct Biol 18:581–586.
33. Sharp PM, Li WH (1987) The codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15: 1281–1295.
34. Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. Trends Biotechnol 22:346–353.
35. Jana S, Deb JK (2005) Strategies for efficient production of heterologous proteins in *Escherichia coli*. Appl Microbiol Biotechnol 67:289–298.
36. Hernan RA, Hui HL, Andracki ME, Noble RW, Sligar SG, Walder JA, Walder RY (1992) Human hemoglobin expression in Escherichia coli: importance of optimal codon usage. Biochemistry 31:8619–8628.
37. Li A, Kato Z, Ohnishi H, Hashimoto K, Matsukuma E, Omoya K, Yamamoto Y, Kondo N (2003) Optimized gene synthesis and high expression of human interleukin-18. Protein Expr Purif 32:110–118.
38. Hedfalk K, Pettersson N, Oberg F, Hohmann S, Gordon E (2008) Production, characterization and crystallization of the *Plasmodium falciparum* aquaporin. Protein Expr Purif 59:69–78.
39. Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O (2008) Codon optimization can improve expression of human genes in *Escherichia coli*: a multi-gene study. Protein Expr Purif 59:94–102.
40. Mehlin C, Boni E, Buckner FS, Engel L, Feist T, Gelb MH, Haji L, Kim D, Liu C, Mueller N, Myler PJ, Reddy JT, Sampson JN, Subramanian E, Van Voorhis WC, Worthey E, Zucker F, Hol WG (2006) Heterologous expression of proteins from *Plasmodium falciparum*: results from 1000 genes. Mol Biochem Parasitol 148: 144–160.
41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 1215: 403–410.
42. Cordingley MG, Callahan PL, Sardana VV, Garsky VM, Colonno RJ (1990) Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. J Biol Chem 265:9062–9065.
43. Miroux B, Walker JE (1996) 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 260:289–298.
44. Edelhoch H (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry 6: 1948–1954.
45. Wu G, Culley DE, Zhang W (2005) Predicted highly expressed genes in the genomes of *Streptomyces coelicolor* and *Streptomyces avermitilis* and the implications for their metabolism. Microbiology 151:2175–2187.
46. Molina DM, Cornvik T, Eshaghi S, Haeggstrom JZ, Nordlund P, Sabet MI (2008) Engineering membrane protein overproduction in *Escherichia coli*. Protein Sci 17: 673–680.
47. Jacobs SA, Podell ER, Cech TR (2006) Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. Nat Struct Mol Biol 13:218–225.
48. Jacobs SA, Podell ER, Wuttke DS, Cech TR (2005) Soluble domains of telomerase reverse transcriptase identified by high-throughput screening. Protein Sci 14: 2051–2058.
49. Fitzgerald DJ, Berger P, Schaffitzel C, Yamada K, Richmond TJ, Berger I (2006) Protein complex expression by using multigene baculoviral vectors. Nat Methods 3: 1021–1032.
50. Fitzgerald DJ, Schaffitzel C, Berger P, Wellinger R, Bieniossek C, Richmond TJ, Berger I (2007) Multiprotein expression strategy for structural biology of eukaryotic complexes. Structure 15:275–279.
51. Bieniossek C, Richmond TJ, Berger I (2008) MultiBac: multigene baculovirus-based eukaryotic protein complex production. Curr Protoc Protein Sci 5:Unit 5 20.

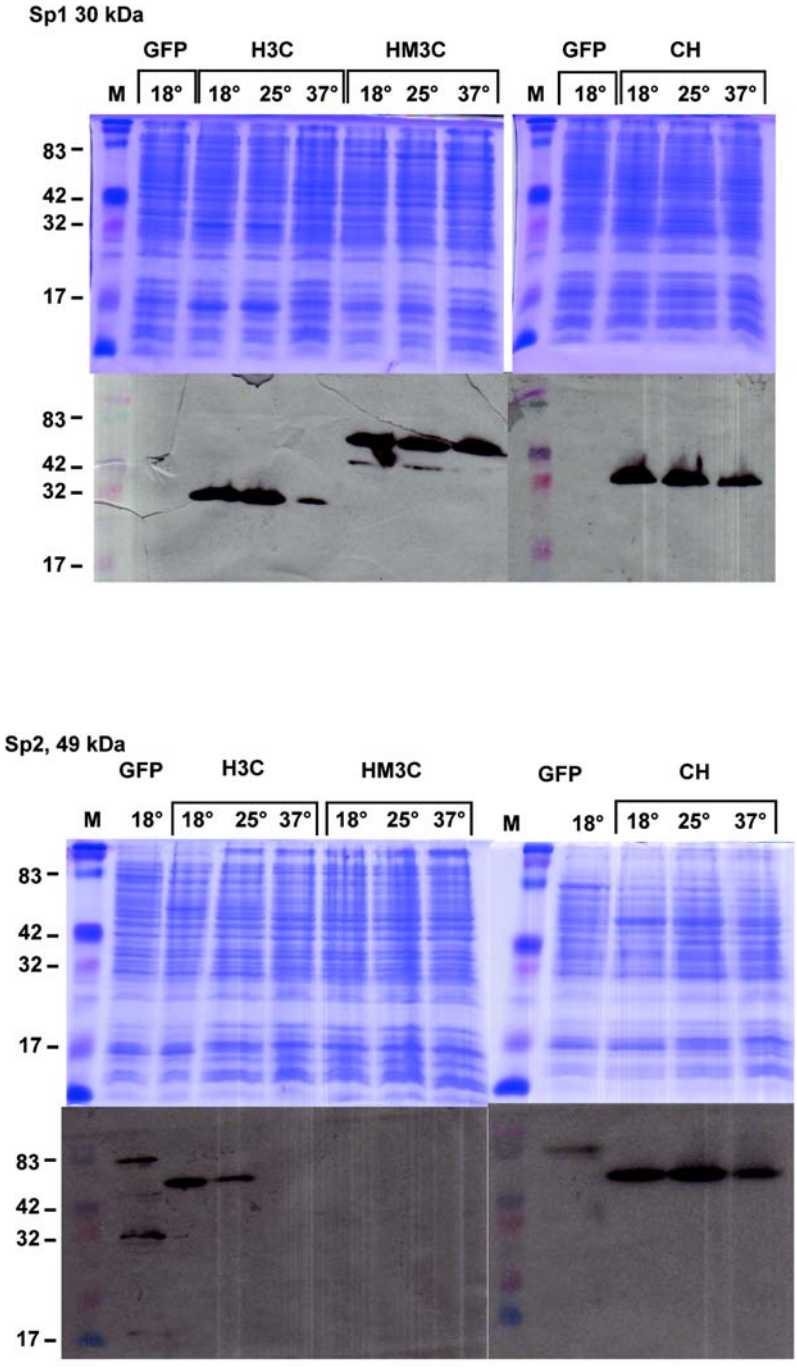
Supplementary Figure 1

Hammon *et al.*



Supplementary Figure 1 Diagram of the LIC vectors used in this study. In each case the restriction site used to prepare the LIC site in the vector is shown in light blue.

Supplementary Figure 2



Supplementary Figure 2 Examples of construct and temperature optimization for two candidates, Sp1 and Sp2. GFP indicates the GFP-fusion construct for each protein. H3C, HM3C, and CH indicate N-terminal His_8, N-terminal His_6-maltose binding protein, and C-terminal His_8 versions, respectively. Western blots are probed with an anti-polyhistidine antibody. Expression temperatures are indicated.

Supplementary Table 1 – Candidate membrane proteins

Candidate	ID	FSU	Annotation	#TM	MW (Daltons)
Aeropyrum pernix					
Ap4	Q9YC10	31000	MscS, Mechanosensitive ion channel	4	31155
Ap6	Q9YFR5	54000	MFS transporter	8	43464
Ap7	Q9YEN5	25000	Hypothetical protein	10	39202
Ap9	Q9YAM9	45000	Putative MFS transporter	11	36327
Ap10	Q9Y9L9	22000	Putative MFS transporter	11	39213
Ap11	Q9Y918	47000	Putative MFS transporter	8	41811
Ap12	Q9Y8L7	101000	Hypothetical protein	10	39483
Archaeoglobus fulgidus					
Af1	O30022	31000	Na^{+}/H^{+} antiporter (NapA-1)	12	41026
Af3	O29412	46000	Na^{+}/H^{+} antiporter (Nhe2)	12	53710
Af4	O29023	34000	Na^{+}/H^{+} antiporter (NapA-2)	9	37387
Af8	O29022	25000	Hypothetical protein	1	11774
Af9	O28857	34000	Chloride channel, putative	10	63443
Af10	O28726	118000	Uncharacterized MscS family protein	3	31734
Af11	O28600	72000	Potassium channel, putative	2	34510
Af12	O28169	22000	UPF0056 membrane protein AF_2111	6	23062
Af15	O29993	23000	Iron (II) transporter (FeoB-1)	8	70367
Af16	O29472	61000	Magnesium and cobalt transporter (CorA)	2	40272
Af17	O29285	65000	Ammonium transporter (Amt-1)	11	40992
Af19	O30277	34000	Iron (II) transporter (FeoB-2)	10	62039
Af20	NP_071083.1	31000	Multidrug resistance protein	14	55068
Af21	NP_069832.1	39000	Hypothetical protein	2	24641
Af22	NP_071101.1	23000	Hypothetical protein	7	32164
Af23	NP_070962.1	66000	Hypothetical protein	6	31291
Af24	NP_070708.1	61000	Daunorubicin resistance membrane protein (drrB)	6	27683
Af26	NP_070381.1	41000	Hypothetical protein	10	29579
Af27	NP_070263.1	99000	Hypothetical protein	7	21891
Af28	NP_069332.1	43000	Hypothetical protein	4	19371
Af29	NP_069540.1	23000	Hypothetical protein	9	43060
Bacillus halodurans					
Bh4	Q9KBF6	31000	Potassium channel	2	12013
Bh7	Q9K7W8	46000	Flagellar motor apparatus	4	29219
Bh8	Q9K7M1	165000	Potassium channel	3	37552
Bh9	Q9K5Q2	26000	Cation-transporting ATPase	6	75920
Bh10	Q9Z9W7	33000	CLC channel	9	46107
Chlorobium tepidium					
Ct4	Q8KEM8	82000	Mechanosensitive channel	4	46213
Ct5	Q8KDP0	37000	Hypothetical protein	2	9434
Ct6	Q8KC40	150000	MotA/TolQ/ExbB proton channel	4	26700
Ct7	Q8KB92	33000	MscS, Mechanosensitive ion channel	5	38591

Ct8	Q8KAY4	130000	MscS, Mechanosensitive ion channel	3	33300
Ct10	Q8KDB0	40000	Na^{+}/H^{+} antiporter	11	87355
Deinococcus radiodurans					
Dr1	Q9RXE0	29000	Cation exchanger, putative	9	42790
Dr2	Q9RWL4	28000	Na^{+}/H^{+} antiporter, putative	9	46772
Dr5	Q9RRU0	46000	Na^{+}/H^{+} antiporter, putative	11	48180
Dr10	Q9RXU5	76000	MscS, Mechanosensitive ion channel	5	40602
Dr17	Q9RVH3	44000	MFS transporter	11	48100
Dr18	Q9RV07	43000	Multidrug resistance protein-related	10	43914
Dr21	Q9RUQ8	50000	Transport protein, putative	9	45904
Dr24	Q9RSZ8	45000	MFS transporter	12	39382
Dr26	Q9RSM6	40000	MFS transporter	14	47095
Dr27	Q9RSF5	33000	Drug transport protein	12	45547
Dr28	Q9RS23	36000	MFS transporter	10	39230
Dr29	Q9RS20	29000	Multidrug efflux transporter, putative	12	42530
Dr34	Q9RZS1	66000	Integral membrane protein, LmrP	10	46970
Dr35	Q9RZF0	82461	Multidrug-efflux transporter, putative	12	44153
Desulfovibrio vulgaris					
Dv1	Q728I6	93000	Na^{+}/Ca^{2+}-exchanger family	12	44333
Dv2	Q729P8	59000	Glutathione-regulated potassium efflux, KefB	12	70823
Dv3	Q72WJ3	57000	MFS transporter	11	41595
Dv5	Q72EP3	48000	Drug resistance transporter, putative	14	49436
Dv6	Q72E39	47000	MFS transporter	9	42758
Dv7	Q72CX0	33000	MFS transporter	12	45938
Dv8	Q72C49	18000	MFS transporter	9	41256
Dv9	Q72AU5	52000	MFS transporter	10	41000
Dv11	Q72C83	28000	Na^{+}/H^{+} antiporter	8	33510
Dv12	Q72AL4	57000	Na^{+}/H^{+} antiporter family protein	13	45282
Dv14	Q72F34	52000	Na^{+}/H^{+} antiporter, NhaC	12	48203
Dv15	Q726J9	23000	Na^{+}/H^{+} antiporter, NhaC	10	49455
Dv16	Q72WT3	25000	Magnesium transporter MgtE, putative	5	36903
Dv18	Q72D72	23000	Cobalt ABC transporter, permease protein	4	28216
Dv19	Q72D38	29000	Transporter	11	47929
Dv20	Q72CX3	50000	Urea transporter, putative	10	35757
Dv21	Q729P4	22000	Phosphate transporter family protein	8	35162
Dv22	Q729B6	36000	Transporter, CorA family	2	39017
Dv23	Q72CQ2	63000	Ammonium transporter	11	42167
Dv24	Q72WF3	34000	Zinc transporter	8	29595
Dv25	YP_011767.1	88000	MATE efflux family protein	11	50711
Dv26	YP_010408.1	47000	Permease	8	43166
Dv27	YP_010853.1	29000	Hypothetical protein	7	24838
Dv28	YP_011850.1	58000	Hypothetical protein	4	37786
Geobacter sulfurreducens					

Gs2	Q74DW2	27000	Sodium/hydrogen exchanger family, TrkA domain	11	52033
Gs3	Q74DW1	24000	Sodium/hydrogen exchanger family, TrkA domain	11	75301
Gs5	Q74GI1	20000	Hypothetical protein	9	31527
Gs6	Q74BA5	83000	Heavy metal efflux pump, CzcB family	1	45404
Gs7	Q74GR0	26000	MFS transporter	8	44477
Gs10	Q74F57	27000	Transporter	10	47181
Gs15	Q749J8	22000	MFS transporter	8	41219
Gs20	Q74CP0	48000	MS channel superfamily	5	40428
Gs21	Q74CF1	11000	MS channel superfamily	6	60400
Gs22	Q74B94	39000	Cation-transporting ATPase	6	75732
Gs25	Q74AJ1	17000	Chloride channel family protein	10	66028
Gs26	Q748H5	55000	Chemotaxis MotA protein	4	26957
Gs27	Q749E9	66000	MscL	2	15430
Oenococcus oeni					
Oo1	Q04FY2	21000	Chloride channel protein EriC	11	45175
Oo2	Q04FK1	73000	MscL	2	14068
Oo3	Q04DD4	40000	MscS, Mechanosensitive ion channel	3	32721
Oo5	Q04HA9	29000	Mg^{2+} and Co^{2+} transporter	2	34259
Oo6	Q04HA7	28000	ABC-type cobalt transport system, permease, CbiQ	5	25155
Oo7	Q04H42	21000	Mn^{2+} and Fe^{2+} transporter, NRAMP family	11	48555
Oo9	Q04GU3	20000	Mg^{2+} and Co^{2+} transporter	2	36748
Oo10	Q04GR9	30000	Amino acid transporter	13	51746
Oo11	Q04GP4	22000	Putative citrate transporter	10	35446
Oo12	Q04G55	20000	ABC-type cobalt transport system, permease, CbiQ	4	30199
Oo13	Q04G49	23000	Spermidine/putrescine ABC transporter permease	6	30596
Oo14	Q04G48	30000	Spermidine/putrescine ABC transporter permease	6	30014
Oo17	Q04FM3	16000	Mn^{2+} and Fe^{2+} transporter, NRAMP family	9	46064
Oo18	Q04FM2	22000	ABC-type cobalt transport system, permease, CbiQ	4	28639
Oo23	Q04E65	35000	Nicotinamide mononucleotide transporter	7	28963
Oo24	Q04DX6	24000	Amino acid transporter	10	48378
Oo25	Q04DX5	16000	ABC-type cobalt transport system, permease, CbiQ	4	24364
Pyrococcus furiosus					
Pf5	Q8U3L6	43000	Auxin efflux carrier	8	32356
Pf6	Q8U427	46000	$Na^{+/}H^{+}$ antiporter	10	41154
Pf7	Q8U422	78000	$Na^{+/}H^{+}$ antiporter	12	40600
Pf8	Q8U3V5	55000	Putative cation antiporter	9	29331
Pf9	Q8U1P1	44000	$Na^{+/}H^{+}$ antiporter	11	42922
Pf10	Q8U0I6	34000	Hypothetical sodium antiporter	12	46136
Pyrococcus horikoshii					
Ph1	O57747	78000	CIC channel	11	40162

Ph2	O58074	102000	MscS, Mechanosensitive ion channel	5	37487
Ph3	O59323	23000	Potassium channel	2	55869
Ph7	O58682	28000	Na^{+}/H^{+} antiporter	11	42680
Ph8	O59255	31000	Na^{+}/H^{+} antiporter	12	46107
Ph10	NP_142156.1	22000	MFS/permease	10	44056
Ph11	NP_143811.1	104000	Na^{+}/solute symporter (limited C-terminal homology)	3	27623
Ph12	NP_142608.1	30000	Type II secretion system protein, GspF	4	41078
Ph13	NP_143637.1	21000	LrgB-like protein	6	24013
Ph14	NP_143430.1	24000	Hypothetical protein	9	37835
Ph15	NP_877871.1	23000	Hypothetical protein	8	34828
Ph16	NP_142647.1	24000	MFS transporter	12	48362
Ph18	NP_142106.1	20000	Hypothetical protein	9	52934
Ph19	NP_142251.1	20000	Hypothetical protein	5	25697
Ph21	NP_143648.1	29000	TRK system potassium uptake protein	12	55063
Ph23	NP_143207.1	40000	Hypothetical protein	4	20422
Ph25	NP_142088.1	28000	Permease, drug/metabolite transporter	10	30675
Ph26	NP_142317.1	41000	MscS, Mechanosensitive ion channel	5	37487
Ph27	NP_142356.1	24000	Hypothetical protein	5	37409
Ph29	NP_143043.1	24000	Hypothetical protein	9	58340

Ralstonia metallidurans

Rm1	Q1LR85	39000	Voltage-gated chloride channel	8	57798
Rm2	Q1LMC6	77000	MscS, Mechanosensitive ion channel	5	50055
Rm3	Q1LL19	79000	MotA/TolQ/ExbB proton channel	3	26112
Rm4	Q1LJX5	110000	MotA/TolQ/ExbB proton channel	3	25044
Rm6	Q1LIE1	42000	Voltage-gated chloride channel	8	54767
Rm7	Q1LIC3	94000	MscL, Mechanosensitive ion channel	2	15869
Rm8	Q1LE13	87000	Mechanosensitive ion channel	3	30819
Rm9	Q1LD70	36000	MscS, Mechanosensitive ion channel	5	39873
Rm10	Q1LBM8	63000	MscS, Mechanosensitive ion channel	3	30634
Rm11	Q1LB14	156000	Mechanosensitive ion channel	3	31167
Rm12	Q58AF6	37000	MFS transporter family protein	9	44644
Rm13	Q1LSG7	40000	MFS transporter family protein	10	43827
Rm14	Q1LR44	101000	MFS transporter family protein	12	47161
Rm16	Q1LMW3	44000	Sodium/hydrogen exchanger	11	42774
Rm17	Q1LHL8	53000	Sodium/hydrogen exchanger	10	42958
Rm18	Q1LDP0	21000	Sodium/calcium exchanger	10	40585
Rm19	Q1LDC9	30000	Sodium/hydrogen exchanger	11	45624
Rm22	Q1LSH3	47000	Drug resistance transporter EmrB/QacA subfamily	14	55393
Rm25	Q1LQB7	28000	Phosphonate uptake transporter	6	31871
Rm26	Q1LQB3	21000	Drug resistance transporter Bcr/CflA subfamily	10	43882
Rm29	Q1LN60	128000	High-affinity nickel-transporter	6	29393
Rm30	Q1LLZ8	54000	General substrate transporter	12	46684
Rm31	Q1LLX4	31000	Phosphate transporter	8	35203
Rm32	Q1LLM4	87000	Nitrate transporter	12	49438
Rm33	Q1LKD2	28000	Chromate transporter	5	18096
Rm34	Q1LKD1	64000	Chromate transporter	3	20533

Shewanella oneidensis

So1	Q8EHC6	25000	Chloride channel	11	62047
So2	Q8EFY9	64000	MotA/TolQ/ExbB proton channel	3	48770
So3	Q8EED5	19000	Channel protein, hemolysin III family subfamily	7	23884
So4	Q8EDJ5	63000	MotA/TolQ/ExbB proton channel family	3	25103
So5	Q8E842	109000	Potassium/ion transporter. Voltage-gated ion channel superfamily	2	15022
So6	Q8E808	45000	Heavy metal efflux two-channel pump, HME family	12	114694
So7	Q8EJE5	47000	MscL, Mechanosensitive ion channel	3	14475
So8	Q8EJV2	27000	Sodium/calcium exchanger	10	32767
So9	Q8EH64	67000	Sodium/hydrogen exchanger family protein	13	47159
So10	Q8EE53	29000	Sodium/hydrogen exchanger	12	72718
So11	NP_715670.1	29000	Potassium uptake protein TrkH	11	52592
So12	Q8EAX3	29000	Ion transporter	6	30890
So13	Q8EJE2	21000	Drug resistance transporter, EmrB/QacA family	13	58647
So15	Q8EI19	103000	Iron-compound ABC transporter, permease protein	9	40456
So16	Q8EGN5	38000	Magnesium transporter, putative	5	50475
So17	Q8E9W6	40000	Toxin secretion ABC transporter protein, HlyB	3	59422
So19	Q8E802	25000	Multidrug efflux transporter, DMT superfamily	3	11653
So20	Q8EBC5	71000	Sulfate ABC transporter, permease	6	30779
So21	Q8E8L0	32000	Sulfate ABC transporter, permease	6	31850
So22	Q8EBC4	88000	Sulfate ABC transporter, permease	6	32013
So23	Q8E8K9	22000	Sulfate ABC transporter, permease	6	30983
So24	Q8EHS2	20000	Magnesium transporter	5	49444
So25	Q8EES3	94000	Multidrug resistance protein, norM	12	49441
So27	NP_718490.1	52000	Permease	8	39941
So29	NP_718628.1	31000	Sodium:alanine symporter family	11	53290
So30	NP_716923.1	17000	Hypothetical protein	1	30527
Silicibacter pomeroyi					
Sp1	Q6TMY8	77000	Voltage-gated sodium channel subunit	6	28969
Sp4	Q5LTJ3	60000	Cation channel family protein	6	28726
Sp7	Q5LMR6	76000	Mechanosensitive ion channel family	3	28930
Sp8	Q5LMR7	71000	MscL, Mechanosensitive ion channel	2	15050
Sp12	Q5LX43	37000	Auxin efflux carrier family	10	32884
Sp13	Q5LMH2	24000	Auxin efflux carrier family	9	31885
Sp17	Q5LW76	54000	Transporter, formate/nitrate family	6	26247
Sp2	Q5LX04	75000	Mechanosensitive ion channel family	7	49412
Sp21	Q5LTV8	31000	His/Glu/Gln/Arg/opine ABC transporter	5	30850
Sp22	Q5LTL8	105000	MFS transporter	9	43814
Sp24	Q5LT36	55000	Ammonium transporter family	11	49898
Sp25	Q5LSI9	37000	MFS transporter	12	44161
Sulfolobus solfataricus					
Ss1	Q97U23	17000	Chloride channel	11	64147
Ss2	Q97XX3	13000	Auxin transporter	10	34295
Ss4	Q97Z08	17000	Na^+/H^+ antiporter	11	41628

Ss6	Q97UN8	15000	Na^+/H^+ antiporter	17	68431
Ss7	Q97U31	21000	Na^+/Ca^{2+} exchanger	9	34129
Ss8	Q97XI3	31000	Potassium channel protein	3	37250
Ss9	Q97XW8	17000	Proton symporter	5	26116
Ss10	Q97Y43	32000	MFS transporter	12	46426
Ss11	Q97Y15	19000	Multi-drug resistance transporter	11	52278
Ss12	Q97XG8	18000	Multidrug-efflux transporter	7	29929
Ss13	Q97X92	20000	Multidrug-efflux transporter	10	42744
Ss14	Q97UL0	19000	Manganese transporter	11	43182
Ss15	Q97UD8	27000	MFS transporter	12	46520
Ss16	Q97U91	23000	MFS transporter	14	53392
Ss17	NP_342075.1	22000	Multi-drug resistance transporter	10	44180
Ss18	NP_343625.1	34000	Cation-efflux system membrane protein czcD	5	30276
Ss20	NP_343223.1	18000	Hypothetical protein	9	55174
Ss21	NP_343113.1	20000	Amino acid transporter-like	12	49427
Ss22	Q97ZU1	27000	Iron (III) ABC transporter, permease (HEMU1)	9	35682

Streptococcus thermophilus

St1	YP_139963.1	34000	Amino acid transporter	11	67947
St2	Q5M6C3	30000	MscS, Mechanosensitive ion channel	2	31356
St3	Q5M426	38000	Chloride channel (CIC) family protein	9	43711
St4	Q5M424	25000	Chloride channel	9	55405
St5	Q5M610	31000	Urea channel	6	18983
St6	Q5M682	29000	Auxin efflux carrier, malate permease	9	34519
St7	Q5M683	33000	Zinc ABC uptake transporter	7	30517
St8	Q5M3H4	29000	Oligopeptide ABC uptake transporter	6	55752
St9	Q5M3H5	33000	Oligopeptide ABC uptake transporter	6	34695
St10	Q5M5I1	30000	Ammonium uptake transporter	11	45964
St11	Q5M369	65000	CorA cation transporter (MIT) family	2	35301
St12	Q5M2Q8	24000	CorA cation transporter (MIT) family	2	36345
St13	Q5M4D5	22000	Iron compound ABC uptake transporter	9	37472
St14	Q5M4D4	23000	Iron compound ABC uptake transporter	7	36264
St15	Q5M460	65000	Cell division ABC transporter, permease	4	35683
St16	Q5M5U0	48000	Branched-chain amino acid ABC uptake transporter	8	31433
St17	Q5M5T9	48000	Branched-chain amino acid ABC uptake transporter	9	33857
St21	Q5M3F1	39000	Polysaccharide biosynthesis protein	10	48773
St23	Q5M601	146000	ABC transporter	4	29179
St24	Q5M315	75000	Drug/metabolite transporter (DMT) superfamily	10	32562
St25	Q5M245	35000	ABC transporter	4	30024
St26	Q5M618	28000	Potassium transporter (Trk)	10	52892
St28	Q5M2Z8	98000	Na^+/H^+ antiporter	12	76976
St29	Q5M3N9	59000	Multi Antimicrobial Extrusion (MATE) family,drug:Na^+ antiporter	12	49893
St30	Q5M2E7	31000	Multi Antimicrobial Extrusion (MATE) family,drug:Na^+ antiporter	12	46034
St31	Q5M3H0	35000	Oxalate:formate antiporter	12	43533
St33	Q5M5Z6	24000	Drug:H^+ antiporter-3 (DHA3) family protein	11	44740
St34	YP_139060.1	80000	lipid/multidrug/protein-type ABC exporter	6	64690

St35	YP_140291.1	33000	Multi Antimicrobial Extrusion (MATE) family drug:Na+ antiporter	12	46033
St36	YP_138963.1	81000	Multidrug ABC exporter	5	65120
Thermotoga maritima					
Tm4	Q9WXN7	35000	Oligopeptide ABC transporter, permease	6	37909
Tm5	Q9WXT4	36000	Iron(III) ABC transporter, permease	9	34972
Tm6	Q9WXV9	39000	Sugar ABC transporter, permease	9	35925
Tm7	Q9WY33	28000	Iron(III) ABC transporter, permease	9	36982
Tm8	Q9WY46	94000	ABC transporter, permease, cysTW family	6	27379
Tm9	Q9WYN3	52000	Ammonium transporter	10	46427
Tm10	Q9WZ61	38000	Amino acid ABC transporter, permease	2	23740
Tm11	Q9X0P4	27000	Mg^{2+} transporter MgtE, putative	4	50662
Tm12	Q9X1C4	27000	Antibiotic ABC transporter	6	28665
Tm13	Q9WZS2	28000	Multidrug resistance protein, norM	11	51054
Tm14	NP_228062.1	22000	Na^+-translocating NADH-quinone reductase, Nqr5 subunit	6	20865
Tm15	NP_227902.1	28000	Virulence factor MviN-related protein	12	51850
Tm16	NP_229456.1	70000	Hypothetical protein	7	25363
Tm18	NP_229615.1	31000	Chromate transport protein	4	17718
Tm20	NP_228770.1	38000	Hypothetical protein	1	25797
Tm21	NP_228234.1	30000	Auxin efflux carrier	10	32389
Tm22	NP_229098.1	27000	RND exporter	12	78574
Tm23	NP_228915.1	101000	N-terminal similarity to Na^+/H^+ antiporter	3	27806
Tm24	NP_228307.1	27000	Hypothetical protein	5	32050
Tm25	NP_229521.1	37000	Transporter	10	31785
Thermus thermophilus					
Tt1	Q72K84	38000	Na^+/H^+ antiporter	11	42713
Tt3	Q72IM0	32000	Na^+/H^+ antiporter	10	40012
Tt5	Q72G86	37000	Trk system potassium uptake, trkA	13	41656
Tt6	Q72KL1	25000	Transporter	10	32411
Tt7	Q72HX9	58000	Mechanosensitive ion channel	2	32679
Tt8	Q72HB7	22000	Zinc uptake transporter/gufA protein	7	27223
Tt11	Q72KC2	28000	MFS Transporter	10	41707
Tt14	Q72J32	38000	Transporter	10	42757
Tt15	Q72I99	64000	Permease	10	42370
Tt17	Q72HL7	45000	MFS Transporter	10	37086
Tt18	Q72H63	33000	Fosmidomycin resistance protein	11	40120
Tt21	Q72GR5	76000	MFS Transporter	12	39465
Tt22	Q72GQ5	64000	Permease	10	43497
Tt23	Q72GF5	48000	Permease	12	40475
Tt25	Q72G81	40000	Hypothetical protein	6	18390
Tt26	Q72G80	17000	Macrolide-efflux transmembrane protein	9	37196
Tt27	Q745T8	31000	Sugar transporter	9	45260
Tt28	Q745T4	42000	Sugar transporter	12	45728
Thermoplasma volcanium					
Tv3	Q97AV5	23000	MscS, Mechanosensitive ion channel	4	31515
Tv4	Q97AH5	23000	MscS, Mechanosensitive ion channel	4	32887

Tv5	Q979Z2	27000	Potassium channel	3	38605
Tv7	Q97BP9	23000	Na^+/Ca^{2+} NHX type exchanger	8	29467
Tv8	Q979W5	21000	Na^{+}/H^{+} antiporter	11	45453
Tv9	Q97CN2	23000	Metabolite transporter	12	48124
Tv10	Q97CG2	20000	Multidrug-efflux transporter	12	43334
Tv11	Q97CC5	51000	Metabolite transporter	10	52078
Tv12	Q97C87	18000	MFS permease	10	45743
Tv13	Q97C83	66000	MFS, probably peptide permease	9	47011
Tv14	Q97C76	21000	MFS permease	12	47118
Tv16	Q97C55	21000	MFS	12	41457
Tv18	Q97BG6	20000	Multidrug resistance protein	11	46294
Tv19	Q97BD9	19000	Multidrug-efflux transporter	15	51634
Tv21	Q97B97	22000	Glycerol-3-phosphate transporter	12	45332
Tv25	Q97AJ5	24000	Multidrug resistance protein	14	50523
Tv27	Q97A80	37000	Metabolite transporter	12	51689
Tv29	Q979S3	27000	MFS permease	12	41524
Tv30	Q979Q7	24000	Multidrug resistance protein	11	37029
Tv33	Q979G7	36000	Transporter	12	44526
Tv34	Q978Z4	21000	Sugar transporter	11	44246
Tv35	Q978X8	21000	MFS transporter	11	40690
Tv36	Q978M7	47000	MFS transporter	3	14711
Tv37	Q978M5	23000	MFS transporter	10	40052
Tv38	Q978M4	26000	Multidrug resistance protein	13	51115
Tv39	Q978J3	29000	MFS transporter	11	40942
Tv40	Q978I2	27000	MFS transporter	12	43556
Tv41	Q978G9	36000	Multidrug efflux permease	10	42317
Tv42	Q978G3	24000	Multidrug resistance protein	14	49844

Supplementary Table 1 Candidate extremeophile membrane proteins tested. Uniprot code/gene ID, molecular weight, annotation information, number of transmembrane segments and FSU score in the GFP-screen are shown.

Supplementary Table 2 – Size exclusion apparent molecular weights for monodisperse samples

Candidate	Construct	Monomer (kDa)	Protein-Detergent complex (kDa)						Annotation
			DM	OG	DDM	FC12	CHAPSO	LDAO	
Ct6	HM3C	27	108	-	160	164*	-	98	MotA/TolQ/ExbB proton channel
Rm4	CH	25	-	n.p.	n.p.	132	-	-	MotA/TolQ/ExbB proton channel
Dr10	H3C	40	-	n.p.	144	95	-	-	Mechanosensitive ion channel, MscS
Sp7	H3C	29	-	-	-	109	-	-	Mechanosensitive ion channel, MscS
Sp2	H3C	49	-	n.p.	-	120	-	-	Mechanosensitive ion channel, MscS
Sp8	H3C	15	-	180	133	240	-	180	Mechanosensitive ion channel, MscL
Rm29	H3C	30	n.p.	n.p.	n.p.	102	n.p.	-	High-affinity Nickel transporter
St34	CH	65	-	n.a	n.a	320	n.a	n.a	Lipid/multidrug/protein-type ABC exporter
Ph11	CH	28	+	+	+	+	-	110	Na^+/solute symporter
Dr18	CH	44	n.a	165	153	130	n.a	n.a	Multidrug resistance

Supplementary Table 2 Apparent molecular weight of the protein detergent complex for the ten monodisperse extremophile membrane proteins. * indicates measurement of the Ct6-HM3C uncleaved construct (monomer 72 kDa). Other Ct6 values are for samples after HM3C tag removal.