

SCIENTIFIC REPORTS

OPEN

Escherichia coli as host for membrane protein structure determination: a global analysis

Georges Hattab, Dror E. Warschawski, Karine Moncoq & Bruno Miroux

Received: 17 March 2015

Accepted: 11 June 2015

Published: 10 July 2015

The structural biology of membrane proteins (MP) is hampered by the difficulty in producing and purifying them. A comprehensive analysis of protein databases revealed that 213 unique membrane protein structures have been obtained after production of the target protein in *E. coli*. The primary expression system used was the one based on the T7 RNA polymerase, followed by the arabinose and T5 promoter based expression systems. The C41λ(DE3) and C43λ(DE3) bacterial mutant hosts have contributed to 28% of non *E. coli* membrane protein structures. A large scale analysis of expression protocols demonstrated a preference for a combination of bacterial host-vector together with a bimodal distribution of induction temperature and of inducer concentration. Altogether our analysis provides a set of rules for the optimal use of bacterial expression systems in membrane protein production.

Membrane proteins (MP) play a central role in several biological processes, which includes cell signaling, ion and metabolites transport, and energy conversion. Since the first MP structure was determined in 1986¹, over 450 unique MP structures have been obtained (see crystal structure list from White², and NMR structure list from Warschawski³). They provide molecular details explaining how MP work. Despite numerous breakthroughs in X-ray diffraction^{4,5}, NMR⁶ and electron microscopy⁷, as well as in MP production^{8–11} and stabilization¹², the structural biology of MP is hampered by the production of the recombinant protein and its purification in a functional state. In 1986, Studier and colleagues¹³ set up a powerful bacterial expression system, in which the RNA polymerase from the bacteriophage T7 specifically drove the transcription of the target gene inserted in the expression plasmid, under the control of the T7 promoter. However, one of the main drawbacks of the T7 based expression system is that the rate of transcription of the target gene is rather fast because the T7 RNA polymerase (T7 RNAPol) transcription activity is over ten times faster than *E. coli* RNA polymerase. Moreover, the expression system is further enhanced by the copy number of the expression plasmid. Consequently, upon expression of the T7 RNAPol, an excess of target RNA is produced, which is often toxic to the cell, and triggers uncoupling between transcription and translation and growth arrest^{14,15}. Therefore, several strategies have been developed to attenuate and better regulate the T7 expression system: 1. Introducing a T7/*lac* hybrid promoter within the expression plasmid; 2. Over-expressing the T7 lysozyme, a natural inhibitor of the T7 RNAPol¹⁶; 3. Expressing the T7 RNAPol under the control of the tightly regulated arabinose promoter (BL21-AI, Invitrogen).

In 1996, two spontaneous mutants of the BL21λ(DE3) bacterial host, namely C41λ(DE3) and C43λ(DE3), were isolated by exploiting the toxicity of the over-expression of the oxoglutarate mitochondrial carrier gene and of *atpF* encoding the b subunit of the *E. coli* ATP synthase, respectively¹⁷. We found that the level of accumulation of the target gene mRNA is ten times lower in C41λ(DE3) and is delayed by one hour in the C41λ(DE3)-derived bacterial host, C43λ(DE3). More recently, Wagner *et al.* have shown that the level of T7 RNAPol is strongly reduced in both mutants, thereby allowing the bacterial cell to mediate cell growth with protein production¹⁸. Ensuring viability allowed

Laboratoire de Biologie Physico-Chimique des Protéines Membranaires, Institut de Biologie Physico-Chimique, CNRS, Univ Paris Diderot, Sorbonne Paris Cité, PSL research university, Paris, France. Correspondence and requests for materials should be addressed to K.M. (email: Karine.Moncoq@ibpc.fr) or B.M. (email: Bruno.Miroux@ibpc.fr)

metabolic adaptation, as illustrated by the production of the b subunit of the ATP synthase. In the C41 λ (DE3) host, the b subunit was found in a partially unfolded state whereas in the C43 λ (DE3) host, the production of the protein was accompanied by intense membrane proliferation with the b subunit in the correctly folded state¹⁹. In parallel to developments in the T7 based expression system, alternative expression systems have been established by employing arabinose²⁰, lactose, tetracycline²¹, or T5 promoters²². Today, a profusion of expression plasmids and bacterial hosts are available for protein over-production^{23,24}, however there is no clear rationale for choosing the appropriate bacterial expression system in each individual case.

Our objective here was to perform a global analysis of existing expression systems in the frame of MP production and structure determination. In a first step, entry codes referring to membrane protein structures obtained from *E. coli* were extracted from the Protein Data Bank (PDB) and the two major expression systems, T7 and arabinose promoter based, were identified. In a second step, a bibliographic database was constructed to perform an extensive analysis of expression protocols. The results we have obtained thus provide a systematic set of rules for the successful production of membrane proteins in *E. coli*.

Results

Analysis of bacterial expression systems for MP structure determination. At the time of the analysis, June 2014, 213 unique MP structures (including 72 *Escherichia coli* MP, see supplementary Tables 1 and 2) were retrieved from the crystallographic² and NMR³ databases on the basis of having been produced in *E. coli*. First, we focused our analysis on the heterologous production of MP in *E. coli*. Table 1 summarizes the distribution of the 163 expression vector/bacterial hosts used for obtaining the 141 unique non *E. coli* MP structures. The first remarkable observation is that the T7 promoter based expression system is dominant (63%) followed by the arabinose, *tac* and T5 promoter based expression systems (17%, 9% and 7%, respectively). The pASK tetracycline induced expression vector also shows a detectable impact (5%), which is notable given that it is exclusively marketed by a small biotech company (IBA, Goettingen Germany). Within the T7 based expression system, five bacterial hosts, namely BL21 λ (DE3), C41 λ (DE3), C43 λ (DE3), BL21 λ (DE3)-pLysS, and BL21 λ (DE3)-CodonPlus are extensively used (91%). The bacterial host BL21(DE3) is first (40 MP structures), followed by the two mutant hosts, C41 λ (DE3) and C43 λ (DE3) (16 and 18 MP structures respectively), and then the combination of BL21 λ (DE3) with a companion plasmid expressing either lysozyme or a rare tRNA (12 and 7 MP structures respectively). Bacterial hosts other than those mentioned above have only had a marginal impact in the field (1 to 2 MP structures).

Next we asked whether some bacterial hosts had more success with integral membrane proteins (IMP), which are the most difficult class of membrane proteins to express. Non *E. coli* MP structures obtained within the T7 expression system were classified according to their secondary structures and topologies. As shown in Fig. 1, half the α -helical integral membrane protein structures obtained so far are produced either in C41 λ (DE3) or C43 λ (DE3) (14 and 17 IMP, respectively). Within this IMP group, distribution of the number of MP transmembrane spans is independent of the bacterial host (supplementary Figure 1). In contrast, no β -barrelMP are produced in the T7 based expression system with those mutant hosts (Fig. 1). In the arabinose promoter based expression system (Table 1), the bacterial host C43 λ (DE3) surprisingly appears as the best choice (10 MP structures, including 7 β -barrel MP), followed by the BL21-*T1*^R and TOP10/DHB10 hosts (4 and 3 MP structures respectively). The bacterial strain XL1-Blue is preferred when using the T5 promoter system (4 MP structures). Next, we determined whether the success of the C41 λ (DE3) and C43 λ (DE3) mutants is specific to heterologous production of MP, or if it also impacted the homologous production of MP.

Table 2 shows the distribution of expression hosts and promoter, for 72 unique MP structures found in the PDB. The greater number of unique expression systems, 111, for producing only 72 unique MP structures, suggests that, in contrast to the heterologous production of MP, the choice of the promoter and the expression host is more flexible. For instance, the crystal structure of OmpG was obtained after production of the protein in C43 λ (DE3) with an arabinose promoter expression plasmid (2F1C²⁵), or in C41 λ (DE3) (2IWW²⁶) and in BL21 λ (DE3)pLysE (2JQY²⁷) with a T7 based expression plasmid. Similarly, the lactose permease was expressed under the control of its native promoter in the XL1Blue host grown in a 150l fermenter (1PV7, 2CFQ^{28,29}), and more recently using either C41 λ (DE3) or C43 λ (DE3) with a T7 based expression plasmid^{30,31} (see supplementary Table 2). Table 2 also shows the impact of the T7 system (60%) and of native promoters (18%). For example, OmpC (2J1N³²), LamB (1MPM³³), AcrB (2RDD³⁴), and the electron-transfer chain complex 1 (3M9C³⁵), were produced under the control of their own promoter in multi-copy plasmids (Supplementary table 2). Within the T7 based expression system, the leading host is BL21 λ (DE3) (28 MP structures) followed by C43 λ (DE3), C41 λ (DE3) and BL21 λ (DE3)pLysS (10, 7, and 6 MP structures, respectively).

Overall, both C43 λ (DE3) and C41 λ (DE3) mutant hosts contributed to 28% of non *E. coli* MP structures and 19% of *E. coli* MP structures deposited into the PDB.

Analysis of T7 and arabinose based expression protocols. A database was constructed, containing 2817 articles citing either Miroux and Walker¹⁷, Studier and Moffatt¹³, or Guzman *et al.*²⁰, for the use of C41 λ (DE3)/C43 λ (DE3) (group 1), BL21 λ (DE3) (group 2) bacterial hosts in the T7 based expression

Bacterial host	Promoter				
	T7	<i>ara</i>	T5	<i>tet</i>	<i>trp/tac</i>
BL21λ(DE3)	40	1	1	2	2
C43λ(DE3)	18	10	1		–
C41λ(DE3)	16				
BL21λ(DE3) pLysS	12	2	1		
BL21λ(DE3) CodonPlus ¹	7			1	1
BL21 Star λ(DE3)	1	1			
BL21λ(DE3) Rosetta pLysS	1				
BL21λ(DE3) Tuner	1				1
BL21Rosetta	2				
BL21(AI)		1			
BL21-Gold	1				
BL21-T1 ^R		4			
Lemo21	1				
Origami B			1		
B834	1			1	1
BLR					1
DH10B/ TOP10		3			
XL1-Blue		1	4		1
DH5a				1	3
SG13009			2		
MC4100		1			
SCM6		1			
MC1061		2			
JM83				2	
M15			1		
KRX					1
JM109					1
Not specified	1				2
Other		1 ³		1	
Total	102	28	11	8	14
Total (%)	63	17	7	5	8

Table 1. Bacterial expression hosts for non *E. coli* MP structure determination.

¹BL21-CodonPlus-λ(DE3)-RIL(4), BL21-CodonPlus λ(DE3)-RP(4), BL21-CodonPlus λ(DE3)-RIPL(1), BL21-CodonPlus λ(DE3)(1), BL21-CodonPlus λ(DE3)-RIL-X(1). ²An *E. coli* K strain that contains a chromosomal copy of the T7 RNA polymerase gene under rhamnose promoter (Promega). ³*Pseudomonas aeruginosa* used as expression system.

system, or any bacterial host in the arabinose inducible promoter system (group 3), respectively. All groups were first parsed with the regular expression “membrane protein”, which was found in 77% of the articles in group 1, 25% of articles in group 2 and 45% of articles in group 3 (Table 3). Next, we investigated expression protocols, focusing on inducer concentration and growth temperature (see Materials and Methods section). Explicit values for IPTG concentration or growth temperature were unstated in half of the articles (Table 3). Fig. 2 shows a bimodal distribution for both parameters. As expected, the majority of articles refer to 1mM IPTG concentration as the induction condition and 37°C growth temperature. However, in 50% of the cases, IPTG concentration is below 0.5 mM, and in 38% of the cases growth temperature is equal to, or below, 30°C (Fig. 2). Of note, a significant number of publications, 94 altogether from groups 1 and 2, refer to IPTG concentration below 10 μM, which is consistent with the recently published improved induction protocol at 8 μM IPTG³⁶. Accordingly, we found two MP structures where the recombinant protein was obtained without any addition of IPTG in BL21λ(DE3) cell cultures (2PRM³⁷, 4EIT³⁸).

Next, the distribution of plasmids in all three groups of articles was investigated (Table 3). Frequency citation of T7 based high copy number plasmids is greater in group 1 (19% versus 13% in group 2),

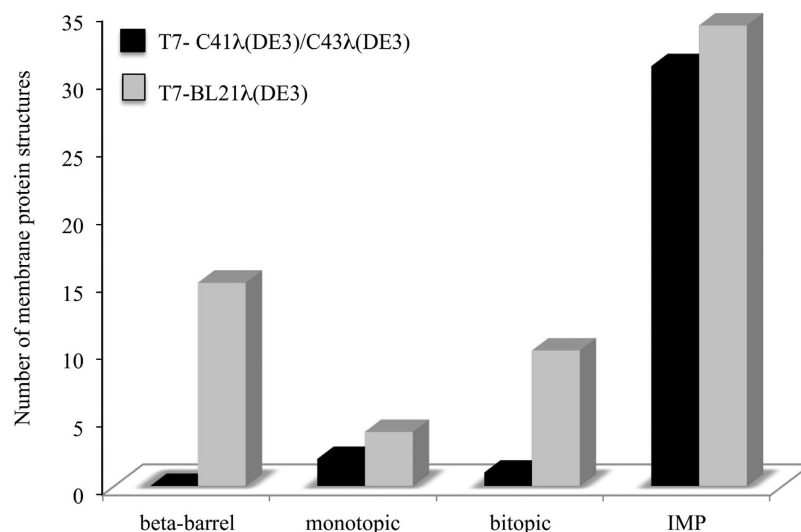


Figure 1. Distribution of secondary structures in MP structures within the T7 expression system.

Membrane protein structures obtained from overexpression in the T7 system (102 see Table 1) were classified according to their secondary structure and topologies. For α -helical membrane proteins the number of transmembrane spans was represented as follow: monotopic (without transmembrane span), bitopic (1 transmembrane span) and integral membrane proteins (IMP, more than 1 transmembrane α -helices).

whereas attenuation plasmids are less cited (10% versus 16% in group 2). Arabinose based promoter pBAD plasmids increased in group 1 (9% versus 3% in group 2), which correlates well with the significant number of MP structures obtained with the C43 λ (DE3)/pBAD host-expression vector system combination (10 and 4 MP structures in Tables 1 and 2 respectively). Citation of pBAD plasmids were found in 99% of group 3 articles. The use of pRARE companion plasmids supplementing rare tRNA is marginal in all three groups, which is consistent with the small number of MP structures that are produced with the BL21 λ (DE3) CodonPlus bacterial host (7, see Table 1). To confirm the link between high copy number plasmids and C41 λ (DE3) and C43 λ (DE3) bacterial hosts, a subset of the bibliographic database comprising frequent users of those mutant hosts, which was identified by the number of citations in group 1, was further analysed. Table 4 summarizes plasmid usage from eight laboratories representing 124 citations. In this subset, the use of the C43 λ (DE3) bacterial host was low (13%) and exclusively associated with low copy-number vectors. In more than half the studies (53%), high copy number plasmids (pRSET from Invitrogen or pHis/pMW7 expression vectors^{39,40}) were used in combination with C41 λ (DE3).

Discussion

One of the primary objectives of this study was to assess the impact of bacterial expression hosts for membrane protein structure determination. We found that 28% of all non *E. coli* MP structures have been resolved from MP produced from the C41 λ (DE3) and C43 λ (DE3) bacterial hosts, which has been distributed since 2000 by Lucigen. Thus these hosts, together with the parental host BL21 λ (DE3), have significantly contributed to the success of bacterial expression systems in structural biology. In contrast, other expression systems have had moderate or little impact on the field, most likely because they may have failed to provide sufficient amount of protein for structural studies. Significantly, both mutant hosts have been used for MP difficult to produce in large amounts. For instance, they have been used mainly to produce α -helical MP, which comprise 50% of non *E. coli* MP within the T7 based expression system, rather than being used to produce beta-barrel proteins, which can be produced by almost any expression system (see OmpG, supplementary Table 2). Regarding bitopic MP, essentially produced in the BL21 λ (DE3) bacterial host, it should be noted that in all cases but one (3VMT⁴¹, a glycosyltransferase from *Staphylococcus aureus*), a small truncated form of the protein (usually 30–60 amino-acids), excluding the soluble domains, was produced in *E. coli* for NMR studies.

One explanation for the success of the C41 λ (DE3) and C43 λ (DE3) bacterial hosts is that there is improved regulation in the expression of the target gene. Before induction, the basal level of expression of the target gene mRNA is undetectable, but the strength of the expression system is not compromised after induction¹⁷. Secondly, the decrease in accumulation of the target mRNA in those mutants improves the coupling between transcription and translation⁴², a critical issue for high-level production of some MP⁴³. Wagner *et al.* have demonstrated that in the C41 λ (DE3) and C43 λ (DE3) mutant hosts the level of T7 RNA polymerase is decreased tenfold¹⁸, which is a reasonable explanation for the

Bacterial host	Promoter						
	T7	Native	<i>ara</i>	<i>tac</i>	T5	<i>tet</i>	Other
BL21λ(DE3)	28	2		3			
C43λ(DE3)	10		4				
C41λ(DE3)	7						
BL21λ(DE3) pLysS	6						
BL21λ(DE3) pRIL	1						
BL21Starλ(DE3)	1						
BL21λ(DE3)Star pLysS	1						
BL21λ(DE3) Tuner	2						
BL21-Gold	3						
B834 (DE3)	3		1			1	
XL1-Blue		2		1	1		
DH5-α		2					1
JM109		1			1		
TOP10		1		1			
BZB1007		2					
HN705		1					
LS6164			1				
LE392			1				
UT5600							1
WH1061				1			
MEG119		1					
LMG194			1				
HN741				1			
LCB2048				1			
AW740		1					
RK20				1			
MH225		1					
TNE012		1					
FT004		1					
DW35		1					
MC4100		1					
GO105		1					
Not found							
Total	62	19	8	9	2	1	2
Total (%)	60	18	8	9	2	1	2

Table 2. Bacterial expression hosts for *E. coli* MP structures determination.

improved regulation in this system. Similarly, co-production of the lysozyme has been shown to inhibit the activity of the T7 RNA polymerase¹⁶, which thereby reduces the basal and induced amount of enzyme available for the transcription of the target gene. Recently, the Cole group isolated new bacterial derivatives of BL21λ(DE3)⁴⁴. The authors used the chemotaxis protein CheY fused to GFP as a model. The level of the CheY-GFP fusion protein was increased by 25% in their most potent BL21λ(DE3) derivative, P2-BL21λ(DE3), as compared to the C41λ(DE3) host. At the chromosomal level however, the situation is still unclear. Wagner *et al.* have shown that in C41λ(DE3) and C43λ(DE3) the *lacUV5* promoter, which drives the expression of the T7 RNAPol gene within the lambda DE3 insertion, has been replaced by the natural *lac* promoter, likely due to homologous recombination with the genomic copy of the *lac* operon¹⁸. Cole *et al.* found the same mutation in the T7 promoter but they postulated that additional mutations account for the production levels of the CheY-GFP fusion protein between P2-BL21λ(DE3) and C41λ(DE3). The secondary mutation in C43λ(DE3), which derives from C41λ(DE3), is still unknown.

Group of articles	1 (T7)	2 (T7)	3 (<i>ara</i>)
Number of articles, Web of Knowledge ¹	876	4626	2310
Unique articles converted and analysed	756	1056 ²	1005 ²
Citation of membrane protein (%) ³	77	25	45
Articles with explicit amount of inducer	393	539	633
Article with explicit growth temperature values	256	493	526
Frequency of expression vector citation (%)			
pET ^{3,4} vectors	40	34	19
pET vectors $\Delta(lacI/lacO)^{3,5}$	13	13	4
High copy number vectors $\Delta(lacI/lacO)^{3,6}$	19	13	8
T7 attenuation vectors ^{3,7}	10	16	4
pRARE codon vectors ³	2	0.4	0.5
Vectors other than T7 based ^{3,8}	25	13	99 ⁹

Table 3. Analysis of the bibliographic database. ¹Number of citing articles at time of study (July 2012).

²Free access articles only. ³Frequency of word pattern count within the group of articles, limited to 1 match per article. ⁴pET (Novagen) are medium copy number plasmids (20–50/cell). ⁵pET(3, 9, 14, 17, 20, 23). ⁶pMW7 and derivatives (pHis and pRun), pGEM (Promega), pRSETand pDEST (Invitrogen), pIVEX (5prime), and pPR-IBA (IBA) are all high copy number plasmids (200–600/cell). ⁷*placI*, *plysS*, *plysE* (Novagen). ⁸pGEX (GE Healthcare), pASK (IBA), pQE (Quiagen), pMAL (New England Biolabs) and pBAD (Invitrogen-life technologies). ⁹pBAD exclusively.

Despite the difficulty in analyzing a large dataset of expression protocols, some general experimental rules have emerged. Our analysis revealed that IPTG concentration and growth temperature are important parameters that are complementary to the choice of a bacterial host. Plasmid usage analysis revealed that high copy number plasmids were preferentially used with C41 λ (DE3), consistent with the fact that this mutant host was selected using a high copy number plasmid (pMW7⁴⁰ encoding the oxoglutarate mitochondrial carrier⁴⁵). Altogether, our data highlight that most of the strategies developed to improve expression systems have focused on limiting the toxicity for the bacterial host. Dong and co-workers were the first to show that in both *tac* and T7 based expression systems, and in presence of high copy number expression plasmids, gratuitous over-production of proteins leads to 16S ribosomal RNA destruction and loss of protein translation capacity, which is inversely correlated with the production level of the target protein. Restoring the fitness of the bacteria not only increases the yield of the over-produced protein but also impacts the folding and targeting of the over-produced protein. For instance, fine tuning of the target gene mRNA accumulation impacted the folding efficiency of the protein, which is exemplified by the production of AtpF in C43 λ (DE3) membranes¹⁹. This initial findings with a simple bitopic *E. coli* membrane protein is now re-enforced by the success of C41 λ (DE3) and C43 λ (DE3) bacterial hosts' ability to produce more complex α -helical MP. The choice of the appropriate bacterial host should thus rely on the viability of the cells⁴⁶. In practice, a high copy number vector should be used in combination with the C41 λ (DE3) host to take advantage of the strength of the T7 based expression system whereas, for more difficult MP, the C43 λ (DE3) host, especially in combination with low copy number plasmids, offers the possibility to strongly attenuate the transcription of the target gene.

Our analysis of the PDB shows that they are very few mammalian MP produced in *E. coli* for structural studies. Two approaches have been developed to achieve eukaryotic MP production in *E. coli*. The first one involves engineering the bacterial host to improve its fitness during MP overproduction. Skretas and co-workers used GFP monitoring⁴⁷ and cell sorting to identify genes (*nagD*, *nlpD*, *ptsN-rapZ-npr*), which, when co-expressed in multi-copy vector, enhances GPCR heterologous production in bacteria. They helped maintain periplasm and cell-envelope integrity, which in turn increased the folding efficiency of the newly synthesized GPCR⁴⁸. The second approach is based on increasing the amount of correctly folded protein either by random or site directed mutagenesis. Sarkar *et al.* have elegantly addressed this challenge with the neurotensin receptor (NTR1)⁴⁹. They have expressed a library of plasmids encoding random variants of the neurotensin receptor in *E. coli*. Using fluorescent ligands and cell sorting, they could identify mutants of NTR1 exhibiting a higher level of production, not only in *E. coli* but in yeast and mammalian cells as well. These mutants were also thermoresistant, which points to a common requirement for *in vivo* folding, membrane insertion and thermostability. Here we propose a third approach that could be developed together with the two strategies mentioned above. Pechman and Frydman⁵⁰ proposed that codon optimality regulates the rhythm of translation elongation to ensure the efficiency of cotranslational folding of the peptide. In the case of MP, this principle could be applied not only to the cotranslational folding of the peptide but also to its interaction with the membrane targeting

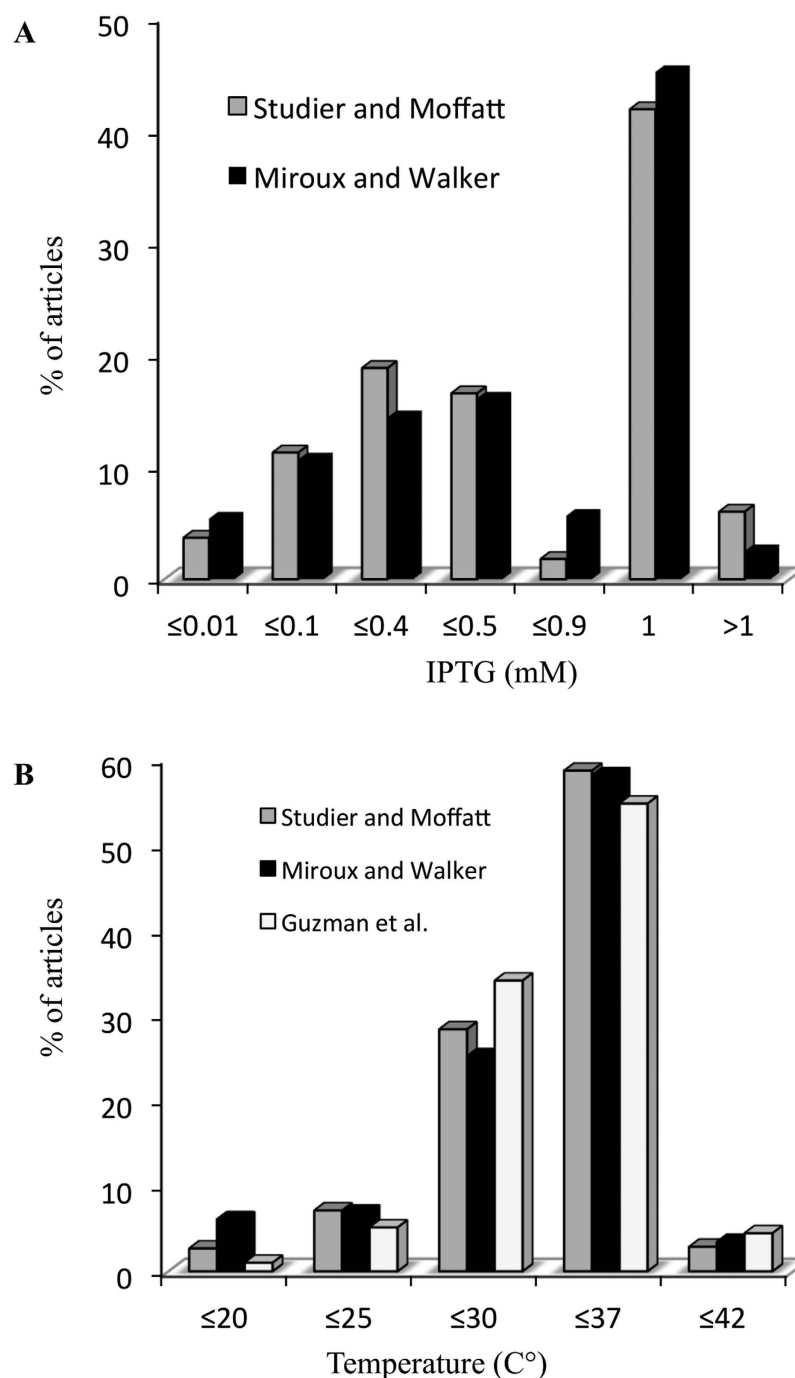


Figure 2. Expression protocol parameters in T7 and arabinose based expression systems. Inducer concentrations and temperatures of growth were extracted using regular expression patterns in articles citing either Miroux and Walker, Studier and Moffatt or Guzman *et al.* for the recombinant expression of proteins in *E. coli* (see Materials and Methods). **A.** IPTG concentration in the T7 based expression system; **B.** temperature of growth in both T7 and arabinose based expression systems. Data are expressed as percentages of the total number of articles where an explicit value was found (See Table 2).

machinery⁵¹. The fact that eukaryotic MP do not follow the same translation rhythms as prokaryotic ones, could also explain why eukaryotic MP often aggregate as inclusion bodies when overproduced in *E. coli*. Adapting codon optimality of eukaryotic MP genes to the *E. coli* translation dynamic could help prevent inclusion body formation, and this, together with the practical rules that have emerged from this study, could reveal the way in which the sequence space coverage of membrane proteome production in *E. coli* could be extended to eukaryotic sequences.

Laboratory	Number of articles in group 1 ¹⁷	Bacterial host vector combination			
		C41λ(DE3)			C43λ(DE3)
		pRSET	pHis/pMW7	pET	pET
Fersht A.R.	27	20		2	
Lowe J.	20	3	15	2	
Clarke J.	18	8	7		
Bycroft M.	17	13			
De la Cruz F.	12			6	3
Winkler H.H.	10			4	2
Dimroth P.	10				7
Suh S.W.	10				4
Total	124	44	22	14	16
Distribution (%)		35	18	11	13

Table 4. Distribution of plasmids in C41λ(DE3) and C43λ(DE3) hosts by most frequent users. ¹Origin or name of the vector not specified. ²pIVEX, pGEM, pMAL-C41λ(DE3). ³pUC8. ⁴pMal/C41λ(DE3).

Materials and Methods

Analysis of membrane protein structure databases. Since NMR or crystal structure determination of proteins usually require milligram amounts of pure protein, this was used as a criterion to assess the success of expression systems. MP structures were extracted from the Protein Data Bank, by using the crystal structure database, interface and search engine developed by Steve White² and the NMR structure database developed by Dror Warschawski³. Only accession codes referring to MP produced from *E. coli* were kept. Secondary accession numbers (structure obtained in presence of inhibitors or ligands, single mutations, changes in crystallisation conditions) were merged into one single entry. Note that when a MP structure was obtained using two different expression systems or bacterial hosts, both accession codes were kept. All related articles were downloaded and screened for expression hosts and vectors.

Construction of the literature database. Articles citing Miroux and Walker¹⁷ for the use of the C41λ(DE3) and C43λ(DE3) hosts (group 1) Moffatt and Studier¹³ for the use BL21λ(DE3) host (group 2) and Guzman *et al.*²⁰ for the use of the arabinose expression system (group 3) were downloaded from PubMed. The PMID list of each control group was converted to a PMCID (unique reference number for PubMed) list using the online PubMed PMCID/PMID/NIHMSID Converter. After conversion into text files of the downloaded articles, group 1 contained 756 articles out of 876 listed through the Web of Knowledge (WoK, using the INIST-CNRS access gate). Groups 2 and 3 consisted of 1005 and 1056 free access articles, respectively (Table 2). The text files for all three groups can be downloaded at https://www.dropbox.com/sh/cf6z7bj3k1sxegg/AADU1JLQ4fW0aeG_-HgltQsCa?dl=0.

Parsing of the literature database. The literature database was parsed for singular keywords or a combination of regular expressions. One positive hit per article was sufficient to select the articles of interest. Temperature and IPTG queries required manual annotation to avoid the counting of misleading hits.

Temperature, inducer and plasmid search. In the global temperature search, all temperatures were recovered: growth temperature, as well as storage, centrifugation and denaturation temperatures, searching for a regular expression of the form of one, two or three digits preceded by a minus or a space and followed by a Celsius or a space then a c; transcribed into: '(\\s|-)([0-9]*)'(\\s°)c. Common and repetitive expressions were transcribed into a regular expression of the form '(harvested or cultured or grown or cultivated) at' preceding the explicit temperature value. The specific temperature search targeted only temperatures of growth. Regarding IPTG induction, the line containing the term 'iptg' was recovered, and the explicit amounts of IPTG was recovered by a manual annotation. Distributions of IPTG concentrations and of temperatures of induction between groups were compared using ANOVA tests. P value < 0.05 were considered as statistically significant. Plasmids were counted and classified in several groups depending on the origin of replication and on the presence of inhibiting sequences from the *lac* operon (*lacO*, *lacI* sequences) either within the expression plasmid or in companion plasmids. Plasmid nomenclature was not always consistent and plasmid names having lost their original nomenclature (pET, pRSET, pGEM, pMal etc...) could not be retrieved. Consequently, pET vectors are explicitly cited in only 40% and 33% of group 1 and 2, respectively. Plasmids encoding either lysozyme (pLysS/E), which has been shown to inhibit the T7 RNA polymerase, or encoding rare tRNA sequences, were also included

in the search. T7 vectors were divided into three categories: pET based vectors, which are medium copy number plasmids (colE1 origin of replication); high copy number T7 based vectors (pMB1 origin of replication) and T7 attenuation vectors encoding either the *lacI* repressor or the lysozyme. Expression plasmids other than T7 based ones, such as pBAD, pcDNA and pRARE codon, were counted separately. The pcDNA eukaryotic expression plasmid served as a negative control for the regular expression search.

References

- Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3 Å resolution. *Nature* **318**, 618–624 (1985).
- White, S. Membrane proteins of known 3D structure determined by X-ray crystallography. at <<http://blanco.biomol.uci.edu/mpstruc/>>
- Warschawski, D. E. Membrane proteins of known structure determined by NMR. at <<http://www.drorlist.com/nmr/MPNMR.html>>
- Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat Protoc* **4**, 706–731 (2009).
- Hendrickson, W. A. Anomalous diffraction in crystallographic phase evaluation. *Q. Rev. Biophys.* **47**, 49–93 (2014).
- Catoire, L. J. *et al.* Structure of a GPCR Ligand in Its Receptor-Bound State: Leukotriene B4 Adopts a Highly Constrained Conformation When Associated to Human BLT2. *J Am Chem Soc* (2010). doi: 10.1021/ja101868c
- Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107–112 (2013).
- Andréll, J. & Tate, C. G. Overexpression of membrane proteins in mammalian cells for structural studies. *Mol. Membr. Biol.* **30**, 52–63 (2013).
- Bornert, O., Alkhaloui, F., Logez, C. & Wagner, R. Overexpression of membrane proteins using *Pichia pastoris*. *Curr Protoc Protein Sci* **Chapter 29**, Unit 29.2 (2012).
- Mancia, F. & Love, J. High-throughput expression and purification of membrane proteins. *J. Struct. Biol.* **172**, 85–93 (2010).
- Bruni, R. & Kloss, B. High-throughput cloning and expression of integral membrane proteins in *Escherichia coli*. *Curr Protoc Protein Sci* **74**, Unit 29.6. (2013).
- Popot, J.-L. Amphipols, nanodiscs, and fluorinated surfactants: three nonconventional approaches to studying membrane proteins in aqueous solutions. *Annu. Rev. Biochem.* **79**, 737–775 (2010).
- Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130 (1986).
- Dong, H., Nilsson, L. & Kurland, C. G. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J. Bacteriol.* **177**, 1497–1504 (1995).
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Meth. Enzymol.* **185**, 60–89 (1990).
- Moffatt, B. A. & Studier, F. W. T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell* **49**, 221–227 (1987).
- Miroux, B. & Walker, J. E. 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 (1996).
- Wagner, S. *et al.* Tuning *Escherichia coli* for membrane protein overexpression. *Proceedings of the National Academy of Sciences* **105**, 14371–14376 (2008).
- Arechaga, I. *et al.* Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale over-production of the b subunit of F(1)F(o) ATP synthase. *FEBS Lett* **482**, 215–219 (2000).
- Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
- Skerra, A. Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. *Gene* **151**, 131–135 (1994).
- Bujard, H. *et al.* A T5 promoter-based transcription-translation system for the analysis of proteins *in vitro* and *in vivo*. *Meth. Enzymol.* **155**, 416–433 (1987).
- Terpe, K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol* **72**, 211–222 (2006).
- Terpe, K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol* **60**, 523–533 (2003).
- Subbarao, G. V. & van den Berg, B. Crystal structure of the monomeric porin OmpG. *J. Mol. Biol.* **360**, 750–759 (2006).
- Yildiz, Ö., Vinothkumar, K. R., Goswami, P. & Kühlbrandt, W. Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation. *The EMBO Journal* **25**, 3702–3713 (2006).
- Liang, B. & Tamm, L. K. Structure of outer membrane protein G by solution NMR spectroscopy. *PNAS* **104**, 16140–16145 (2007).
- Abramson, J. *et al.* Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* **301**, 610–615 (2003).
- Mirza, O., Guan, L., Verner, G., Iwata, S. & Kaback, H. R. Structural evidence for induced fit and a mechanism for sugar/H⁺ symport in LacY. *The EMBO Journal* **25**, 1177–1183 (2006).
- Chaptal, V. *et al.* Crystal structure of lactose permease in complex with an affinity inactivator yields unique insight into sugar recognition. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9361–9366 (2011).
- Kumar, H. *et al.* Structure of sugar-bound LacY. *PNAS* **111**, 1784–1788 (2014).
- Baslé, A., Rummel, G., Storici, P., Rosenbusch, J. P. & Schirmer, T. Crystal Structure of Osmoporin OmpC from *E. coli* at 2.0 Å. *Journal of Molecular Biology* **362**, 933–942 (2006).
- Dutzler, R., Wang, Y.-F., Rizkallah, P. J., Rosenbusch, J. P. & Schirmer, T. Crystal structures of various maltooligosaccharides bound to maltoporin reveal a specific sugar translocation pathway. *Structure* **4**, 127–134 (1996).
- Törnroth-Horsefield, S. *et al.* Crystal Structure of AcrB in Complex with a Single Transmembrane Subunit Reveals Another Twist. *Structure* **15**, 1663–1673 (2007).
- Efremov, R. G., Baradaran, R. & Sazanov, L. A. The architecture of respiratory complex I. *Nature* **465**, 441–445 (2010).
- Sevastyanovich, Y. *et al.* Exploitation of GFP fusion proteins and stress avoidance as a generic strategy for the production of high-quality recombinant proteins. *FEMS Microbiol. Lett.* **299**, 86–94 (2009).
- Walse, B. *et al.* The structures of human dihydroorotate dehydrogenase with and without inhibitor reveal conformational flexibility in the inhibitor and substrate binding sites. *Biochemistry* **47**, 8929–8936 (2008).
- Fairman, J. W. *et al.* Crystal structures of the outer membrane domain of intimin and invasins from enterohemorrhagic *E. coli* and enteropathogenic *Y. pseudotuberculosis*. *Structure* **20**, 1233–1243 (2012).
- Orriss, G. L. *et al.* The delta- and epsilon-subunits of bovine F1-ATPase interact to form a heterodimeric subcomplex. *Biochem. J.* **314** (Pt 2), 695–700 (1996).

40. Way, M., Pope, B., Gooch, J., Hawkins, M. & Weeds, A. G. Identification of a region in segment 1 of gelsolin critical for actin binding. *EMBO J.* **9**, 4103–4109 (1990).
41. Huang, C.-Y. *et al.* Crystal structure of Staphylococcus aureus transglycosylase in complex with a lipid II analog and elucidation of peptidoglycan synthesis mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6496–6501 (2012).
42. Walker, J. & Miroux, B. in *Manual of Industrial Microbiology and Biotechnology*, 2nd edition (*MIMB2*). (A. L. Demain & J. E. Davies, 1999).
43. Arechaga, I., Miroux, B., Runswick, M. J. & Walker, J. E. Over-expression of Escherichia coli F1F(o)-ATPase subunit a is inhibited by instability of the uncB gene transcript. *FEBS Lett.* **547**, 97–100 (2003).
44. Alfasi, S. *et al.* Use of GFP fusions for the isolation of Escherichia coli strains for improved production of different target recombinant proteins. *J. Biotechnol.* **156**, 11–21 (2011).
45. Fiermonte, G., Walker, J. E. & Palmieri, F. Abundant bacterial expression and reconstitution of an intrinsic membrane-transport protein from bovine mitochondria. *Biochem. J.* **294** (Pt 1), 293–299 (1993).
46. Shaw, A. Z. & Miroux, B. A general approach for heterologous membrane protein expression in Escherichia coli: the uncoupling protein, UCP1, as an example. *Methods Mol. Biol.* **228**, 23–35 (2003).
47. Drew, D., Lerch, M., Kunji, E., Slotboom, D.-J. & de Gier, J.-W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nat. Methods* **3**, 303–313 (2006).
48. Skretas, G., Makino, T., Varadarajan, N., Pogson, M. & Georgiou, G. Multi-copy genes that enhance the yield of mammalian G protein-coupled receptors in Escherichia coli. *Metab. Eng.* **14**, 591–602 (2012).
49. Sarkar, C. A. *et al.* From the Cover: Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity. *Proceedings of the National Academy of Sciences* **105**, 14808–14813 (2008).
50. Pechmann, S. & Frydman, J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.* **20**, 237–243 (2013).
51. Pechmann, S., Chartron, J. W. & Frydman, J. Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP *in vivo*. *Nat. Struct. Mol. Biol.* (2014). doi: 10.1038/nsmb.2919

Acknowledgments

We thank Philippe Delepelaire and Lavanya Premvardhan for critical reading of the manuscript. This work was supported by the Agence National de La Recherche (ANR MIT-2M, 2010 BLAN1518), the Centre National de la Recherche Scientifique, and by the “Initiative d’Excellence” program from the French State (Grant “DYNAMO”, ANR-11-LABEX-0011-01).

Author Contributions

G.H., D.W., K.M. and B.M. contributed to the analysis. G.H. and B.M. prepared figures and tables. B.M. wrote the main manuscript text. All authors reviewed the manuscript. B.M. and K.M. have jointly supervised the work.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Hattab, G. *et al.* *Escherichia coli* as host for membrane protein structure determination: a global analysis. *Sci. Rep.* **5**, 12097; doi: 10.1038/srep12097 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

***Escherichia coli* as host for membrane protein structure determination: a global analysis.**

Georges Hattab, Dror E. Warschawski, Karine Moncoq and Bruno Miroux

Laboratoire de Biologie Physico-Chimique des Protéines Membranaires, Institut de Biologie Physico-Chimique, CNRS, Univ Paris Diderot, Sorbonne Paris Cité, PSL research university, Paris, France

Correspondance address :

Bruno.Miroux@ibpc.fr and Karine.Moncoq@ibpc.fr

Laboratoire de Biologie Physico-Chimique des Protéines Membranaires, UMR7099 CNRS IBPC, 13 rue Pierre et Marie Curie 75005 Paris, France

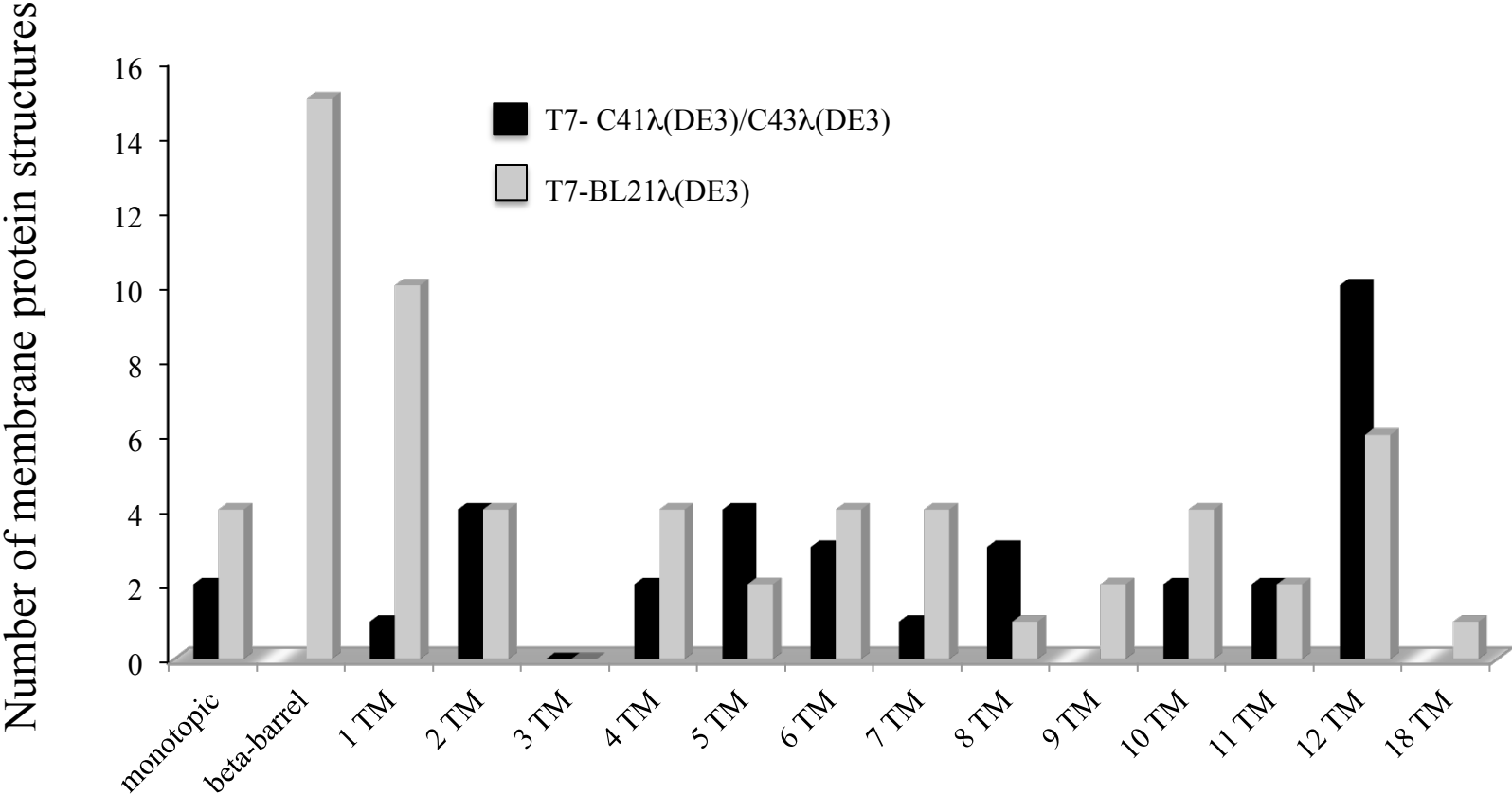
Telephone: 33 1 58 41 52 25

Fax: 33 1 58 41 52 24

Keywords:

structural biology, membrane protein, production of recombinant proteins, *Escherichia coli*, T7 RNA polymerase, bibliographic analysis.

Supplementary Figure-1 Miroux



Supplementary Figure legend

Supplementary Fig. 1. Distribution of secondary structures in MP structures within the T7 expression system.

Membrane protein structures obtained from overexpression in the T7 system (102 see Table 1) were classified according to their secondary structure and topologies. For α -helical membrane proteins the number of transmembrane spans (TM) was represented from 1 to 18.

Supplementary Table 1-Miroux

List of non *E. coli* MP produced in *E. coli* host for structural studies.

T7 expression system: C41(DE3) and C43(DE3) hosts			
PDB via Mpstruc (S. White)	Organism	Description	Name of bacterial host
4EHW	<i>Aquifex aeolicus</i> VF5	Kinase (LpxK)	C41λ(DE3)
4GGM	<i>Bacillus thuringiensis</i> serovar <i>konkukian</i>	Phosphodiester hydrolase (LpxI)	C41λ(DE3)
2J01	<i>Homo sapiens</i>	Phospholemmann (FXD1)	C43λ(DE3)
4F4L	<i>Magnetococcus marinus</i> MC-1	Bacterial voltage-gated sodium channel pore	C41λ(DE3)
2A9H	<i>Streptomyces lividans</i>	Channel (KcsA)	C41λ(DE3)
1WAZ	<i>Morganella morganii</i>	Transporter (MerF)	C43λ(DE3)
4LTO	<i>Alkalilimnicola ehrlichii</i>	Pore-only volt-gated sodium channel (NaVAe1p)	C41λ(DE3)
3P50, 3P4W, 3EAM, 3IGQ	<i>Gloeobacter violaceus</i>	Pentameric ligand-gated ion channel (GLIC)	C43λ(DE3)
3ZKR	<i>Erwinia chrysanthemi</i>	Prokaryotic pentameric ligand-gated ion channel (ELIC) in complex with bromoform:	C43λ(DE3)
4HZU	<i>Lactobacillus brevis</i>	ECF transporter complex	C43λ(DE3)
4A2N	<i>Methanosarcina acetivorans</i>	Transferase (ICMT)	C41λ(DE3)
3KP9	<i>Synechococcus</i> sp.	Thioredoxin domain protein (VKORC1)	C43λ(DE3)
2YVX, 2ZY9	<i>Thermus thermophilus</i> HB8	Transporter (MgtE)	C41λ(DE3)
3UX4	<i>Helicobacter pylori</i> J99	Channel (urea)	C43λ(DE3)
3S0X	<i>Methanococcus maripaludis</i>	Preflagellin aspartyl protease (FlaK)	C43λ(DE3)
3ODJ	<i>Haemophilus influenzae</i>	Peptidase (GlpG)	C43λ(DE3)
4JQ6	Isolated from the Mediterranean Sea	Proteorhodopsin (blue-light absorbing), Med12BPR.	C43λ(DE3)
3TIJ	<i>Vibrio cholerae</i>	Nupc family protein	C41λ(DE3)
4CAD	<i>Methanococcus maripaludis</i>	Protease Rce1 (CAAX)	C41λ(DE3)
3RFU	<i>Legionella pneumophila</i>	Copper efflux ATPase	C43λ(DE3)
4J72	<i>Aquifex aeolicus</i>	Phospho-MurNAc-pentapeptide translocase (MraY)	C41λ(DE3)
3ZUY	<i>Neisseria meningitidis</i>	Bacterial ASBT homologues	C43λ(DE3)
3WAJ	<i>Archaeoglobus fulgidus</i>	OST with bound zinc & sulfate (AglB)	C43λ(DE3)
2B2F	<i>Archaeoglobus fulgidus</i>	Transporter (Amt-1)	C43λ(DE3)
3F3A, 2QE1, 2A65, 3GJD, 3MPN, 3TT1, 3USG 3QS4, 4MM4	<i>Aquifex aeolicus</i> VF5	Symporter (LeuT)	C41λ(DE3)
4IKV	<i>Geobacillus kaustophilus</i>	Proton-dependent oligopeptide transporter (POT)	C41λ(DE3)ΔacrB
4LDS	<i>Staphylococcus epidermidis</i>	Glucose/H ⁺ symporter (GlcP)	C41λ(DE3)
3VVN	<i>Pyrococcus furiosus</i>	Multidrug and Toxin Compound Extrusion (MATE)	C41λ(DE3)ΔacrB
4KPP	<i>Archaeoglobus fulgidus</i>	Ca ²⁺ /H ⁺ antiporter (CaX)	C41λ(DE3)
4K0J	<i>Cupriavidus metallidurans</i>	ZneA Zn(II)/proton antiporter	C43λ(DE3)
3GIA	<i>Methanocaldococcus jannaschii</i>	protein MJ0609 (ApcT)	C41λ(DE3)
4APS	<i>Streptococcus thermophilus</i>	Transporter (PepTSo)	C43λ(DE3)
2XUT	<i>Shewanella oneidensis</i>	Transporter (PepTSt)	C43λ(DE3)
2V50	<i>Pseudomonas aeruginosa</i>	Transporter (MexB)	C43λ(DE3)
T7 expression system: BL21(DE3) derivatives			
PDB via Mpstruc (S. White)	Organism	Type	Strain(s)
3L7I	<i>Staphylococcus epidermidis</i>	Polymerase (TagF)	BL21λ(DE3)
4G9K	<i>Saccharomyces cerevisiae</i>	NADH dehydrogenase (Ndi1)	BL21λ(DE3)pLysS
3VV9	<i>Trypanosoma brucei</i> brucei	Alternative oxidase (AOX)	BL21λ(DE3)
2XCI	<i>Aquifex aeolicus</i>	Glycosyltransferase (WaaA, KdsB)	BL21CodonPlusλ(DE3)
3RBH	<i>Pseudomonas aeruginosa</i>	Export Protein (AlgE)	BL21CodonPlusλ(DE3)
1H6S, 1BH3	<i>Rhodopseudomonas blastica</i>	Porin (E1M/A116K)	BL21λ(DE3)pLysS
3FID	<i>Salmonella typhimurium</i>	lipid A deacylase (LpxR)	BL21Starλ(DE3)
4E1T	<i>Yersinia pseudotuberculosis</i>	Invasin beta-domain	BL21λ(DE3)
3V8X	<i>Neisseria meningitidis</i>	Transferrin (TbpA)	BL21λ(DE3)
3QRA	<i>Yersinia pestis</i>	Adhesion Protein (Ail)	BL21λ(DE3)
3A2R	<i>Neisseria meningitidis</i>	OuterMP (PorB)	BL21λ(DE3)
2K0L (lsNMR)	<i>Klebsiella pneumoniae</i>	OuterMP (OmpA)	BL21λ(DE3)
2K4T (lsNMR)	<i>Homo sapiens</i>	Anion channel (VDAC-1)	BL21λ(DE3)
3EMN	<i>Mus musculus</i>	Anion channel (VDAC-1)	BL21λ(DE3)
2QDZ	<i>Bordetella pertussis</i>	Transporter (FhaC)	BL21λ(DE3)
1UYN	<i>Haemophilus influenzae</i>	Autotransporter (Trimeric Hia)	BL21λ(DE3)
3B07	<i>Staphylococcus aureus</i>	γ-hemolysin composed of LukF and Hlg2	B834λ(DE3)
4K3B	<i>Neisseria gonorrhoeae</i>	BamA with POTRA domains 1 - 5	BL21λ(DE3)
4K3C	<i>Haemophilus ducreyi</i>	BamA with POTRA domains 4 & 5	BL21λ(DE3)

Sup-Table1-Miroux

2LOU	<i>Homo sapiens</i>	Apelin receptor	BL21λ(DE3)
2MFR	<i>Homo sapiens</i>	Insulin receptor TM domain (AAs 940-980)	BL21λ(DE3)
3VMT	<i>Staphylococcus aureus</i>	Glycosyltransferase	BL21CodonPlusλ(DE3)
2KS1, 2JWA (lsNMR)	<i>Homo sapiens</i>	ErbB1/ErbB2, TM heterodimer	BL21λ(DE3)pLysS
2LZL (lsNMR)	<i>Homo sapiens</i>	Fibroblast growth factor receptor 3 (FGFR3, TM)	BL21λ(DE3)pLysS
2LCX (lsNMR)	<i>Homo sapiens</i>	ErbB4, TM dimer	BL21λ(DE3)pLysS
2LOJ, 2LY0, 2KIX	<i>Influenza A</i>	Channel (AM2)	BL21CodonPlusλ(DE3) and BL21λ(DE3)pLysS
2KYV	<i>Homo sapiens</i>	Phospholamban Homopentamer	BL21λ(DE3)
2HAC (lsNMR)	<i>Homo sapiens</i>	TCR-CD3, TM dimer complex	BL21λ(DE3)
2M8R, 3HD7	<i>Rattus norvegicus</i>	Syntaxin 1A, TM & Syntaxin complex	BL21λ(DE3)
2HN2	<i>Thermotoga maritima</i>	Transporter (CorA Mg2+)	BL21CodonPlusλ(DE3)
2H30	<i>Morganella morganii</i>	Transporter (MerF HgII)	BL21λ(DE3)pLysS
4EV6	<i>Methanocaldococcus jannaschii</i>	Transporter (CorA Mg2+)	BL21λ(DE3)Rosetta
3ZRS, 2WLJ, 4LP8	<i>Magnetospirillum magnetotacticum</i>	Channel (KirBac3.1)	BL21CodonPlusλ(DE3)
3DWW (electron microscopy)	<i>Homo sapiens</i>	human microsomal prostaglandin E synthase 1	BL21λ(DE3)pLysS
2XQA, 3EHZ, 4HFI	<i>Gloeobacter violaceus</i>	Channel (pentameric GLIC)	BL21λ(DE3)
2VL0	<i>Herpes simplex</i>	Channel (pentameric ELIC)	BL21λ(DE3)
2Q7M	<i>Homo sapiens</i>	Lipoxygenase protein (FLAP)	BL21λ(DE3)
2M6B	<i>Homo sapiens</i>	Human glycine receptor (hGlyR-α1), TM monomer)	BL21λ(DE3)pLysS
4HTS	<i>Aquifex aeolicus</i>	Twin-arginine translocase receptor (TatC)	BL21λ(DE3)
4HUQ	<i>Lactobacillus brevis</i>	Folate ECF transporter complex	BL21λ(DE3)
3K3F	<i>Desulfovibrio vulgaris</i>	Transporter (Urea)	BL21λ(DE3)
3B60	<i>Salmonella typhimurium</i>	Flippase (MsbA)	BL21λ(DE3)
4DXW	<i>Alpha proteobacterium himb114</i>	Na channel (NaV)	BL21λ(DE3)
3P5N	<i>Staphylococcus aureus</i>	Transporter (RibU)	BL21λ(DE3)
4J9U	<i>Vibrio parahaemolyticus</i>	Potassium ion transporter (TrkH) in complex with TrkA	BL21λ(DE3)
2LNL (ssNMR)	<i>Homo sapiens</i>	Receptor (CXCR1)	BL21λ(DE3)
3QDC, 2KSY, 1H68, 1H2S	<i>Natronomonas pharaonis</i>	Rhodopsin (SRII)	BL21λ(DE3) and BL21λ(DE3) Tuner
4HYJ	<i>Exiguobacterium sibiricum</i>	Proteorhodopsin: Lysine is the proton donor in this novel proteorhodopsin	Rosettaλ(DE3)pLysS
4J7C	<i>Bacillus subtilis</i>	Potassium ion transporter (KtrAB)	BL21λ(DE3)
4HYG	<i>Methanoculleus marisnigri</i>	Presenilin(PSH)	BL21λ(DE3)
4HG6	<i>Rhodobacter sphaeroides</i>	Cellulose synthase/cellulose translocation intermediate (BcsA-BcsB)	Rosetta 2
3QNQ	<i>Bacillus cereus</i>	Transporter (ChbC EIIc)	BL21λ(DE3)
3PJZ	<i>Vibrio parahemolyticus</i>	Transporter (TrkH)	BL21λ(DE3)
4G1U	<i>Yersinia pestis</i>	Transporter (HmuUV)	BL21Goldλ(DE3)
3M71	<i>Haemophilus influenzae</i>	Anion channel (SLAC1)	BL21λ(DE3)pLysS
4KJS	<i>Bacillus subtilis</i>	Ca2+/H+ antiporter (YfkE)	BL21λ(DE3)
4F35	<i>Vibrio cholerae</i>	Symporter	BL21λ(DE3)
3AQP	<i>Thermus thermophilus</i>	(SecDF)	BL21CodonPlusλ(DE3)
4LZ6	<i>Bacillus halodurans</i>	Multidrug and Toxin Compound Extrusion (MATE) transporter (DinF-BH)	BL21λ(DE3)
4BWZ	<i>Thermus thermophilus</i>	Na+/H+ antiporter (NapA)	ΔacrABΔmacABΔyojHI Lemo21λ(DE3)
4M8J	<i>Proteus mirabilis</i>	Carnitine transporter (CaiT)	BL21λ(DE3)pLysS
4HUK	<i>Neisseria gonorrhoeae</i>	Multidrug and Toxin Compound Extrusion (MATE) transporter (NorM)	BL21λ(DE3)
3MKT	<i>Vibrio cholerae</i>	Transporter (NorM & MATE)	BL21λ(DE3)
1KPL	<i>Salmonella typhimurium</i>	Transporter (H+/Cl-)	BL21λ(DE3)

Arabinose promoter based expression system

PDB via Mpstruc (S. White)	Organism	Type	Strain(s)
3SY7, 2SY9, 3SYB	<i>Pseudomonas aeruginosa</i>	Channel (OccD1, OprD)	BL21λ(DE3)T1phage resistant
3SZD	<i>Pseudomonas aeruginosa</i>	Channel (OccK2, OpdF)	BL21λ(DE3)T1phage resistant
3SZV, 3T0S, 3T20, 3T24	<i>Pseudomonas aeruginosa</i>	Aromatic Hydrocarbon (OccK3, OpdO)	BL21λ(DE3)T1phage resistant
4GEY	<i>Pseudomonas putida</i>	Carbohydrate-specific transporter (OprB)	BL21λ(DE3)T1phage
2LHF	<i>Pseudomonas aeruginosa</i>	Outer MP (OprH)	PA(ΔoprH)
2X27	<i>Pseudomonas aeruginosa</i>	Outer MP (OprG)	C43λ(DE3)
2X55	<i>Yersinia pestis</i>	Plasminogen activator (Pla)	C43λ(DE3)
3D5K	<i>Pseudomonas aeruginosa</i>	Outer MP (OprM)	C43λ(DE3)
2QTK	<i>Pseudomonas aeruginosa</i>	Benzoate channel (OpdK)	C43λ(DE3)
2ODJ	<i>Pseudomonas aeruginosa</i>	Channel (OprD)	C43λ(DE3)

Sup-Table1-Miroux

3KVN	<i>Pseudomonas aeruginosa</i>	Autotransporter (EstA)	C43λ(DE3) & BL21Starλ(DE3)
3CSL	<i>Serratia marcescens</i>	Heme receptor complex (HasR)	MC4100 derivative
3DWO	<i>Pseudomonas aeruginosa</i>	Fatty acid transporter (FadL)	C43λ(DE3)
3TDO	<i>Clostridium difficile</i>	Hydrosulfide Ion Channel (FNT3)	BL21λ(DE3)pLysS
3ODJ, 2NR9	<i>Haemophilus influenzae</i>	IntraMembrane peptidase (GlpG)	TOP10
3KLY	<i>Vibrio cholerae</i>	Formate transporter (FocA)	C43λ(DE3)
3QF4	<i>Thermotoga maritima</i>	Heterodimeric ABC exporter	C43λ(DE3) or MC1061
1XFH, 3KBC, 3V8F	<i>Pyrococcus horikoshii</i>	Glutamate Transporter Homologue (GltPh)	TOP10 or DH10B
4KY0	<i>Thermococcus kodakarensis</i>	Aspartate Transporter	MC1061
3DL8	<i>Aquifex aeolicus</i>	Channel (SecYEG)	C43λ(DE3)
3MP7	<i>Pyrococcus furiosus</i>	Primed channel (SecYeb)	BL21λ(DE3)AI
3DIN	<i>Thermotoga maritima</i> MSB8	SecYEG protein in complex SecA	BL21λ(DE3)
2QJU	<i>Aquifex aeolicus</i>	Leucine Transporter (LeuT)	BL21λ(DE3)pLysS
3RCE	<i>Campylobacter lari</i>	OST in complex (PglB)	BL21-Gold SCM6
3DH4	<i>Vibrio parahaemolyticus</i>	NaGalactose Transporter (ySGLT)	XL1-blue
2XQ2	<i>Vibrio parahaemolyticus</i>	NaGalactose Transporter (ySGLT)	TOP10
T5 promoter based expression system			
PDB via Mpstruc (S. White)	Organism	Type	Strain(s)
4HHS	<i>Arabidopsis thaliana</i>	Fatty acid α-dioxygenase (α-DOX)	M15
3O44	<i>Vibrio cholerae</i>	Pore-forming toxin (Cytosolin)	Origami B
4G6G	<i>Saccharomyces cerevisiae</i>	Dehydrogenase (Ndi1)	C43λ(DE3)
4GX0	<i>Geobacter sulfurreducens</i>	Channel (GsuK)	BL21λ(DE3)
3RBZ, 4EI2 4HY0, 4L73, 1LNQ,	<i>Methanothermobacter</i>	Potassium channel and mutants (MthK)	XL1-Blue, SG1309
3LDC,	<i>thermautotrophicus</i>		
4H33	<i>Listeria monocytogenes</i>	Channel (KyLm)	XL1-Blue
3STL, 3OR7, 3EFF, 1BL8	<i>Streptomyces lividans</i>	Channel (KcsA)	XL1-Blue
3E86, 3KOD, 3OUF, 3T1C, 3E86,	<i>Bacillus cereus</i>	Channel (NaK)	SG13009
2Q67, 2AHY			
1ORQ	<i>Aeropyrum pernix</i>	Channel (KyAP)	XL1-Blue
3V5U	<i>Methanocaldococcus janaschii</i>	Exchanger (NaCa)	BL21λ(DE3)pLysS
Tetracyclin promoter based expression system			
PDB via Mpstruc (S. White)	Organism	Type	Strain(s)
2LME	<i>Yersinia enterocolitica</i>	Autotransporter (trimeric YadA)	BL21λ(DE3)
2GR8	<i>Haemophilus influenzae</i>	Trimeric autotransporter (Hia)	B834
3IGA	<i>Streptomyces lividans</i>	Potassium channel (KcsA)	JM-83
3BEH	<i>Mesorhizobium loti</i> Mloti	K1 cyclic nucleotide-regulated K+-channel	JM83
4DOJ, 2WIT	<i>Corynebacterium glutamicum</i>	Glycine betaine transporter (BetP)	DH5a
			BL21 RIL-Xλ(DE3)
3NCY	<i>Salmonella enterica</i>	Antiporter (AdiC)	BL21λ(DE3)
3NDO	<i>Synechocystis</i> sp. pcc 6803	H+/Cl- Eukaryotic Exchange Transporter	variant strain from Stratagene lot # 0420399
Other promoter (Tac, Trp, Rham) based expression system			
PDB via Mpstruc (S. White)	Organism	Type	Strain(s)
1LKF	<i>Staphylococcus aureus</i>	Component of γ-hemolysin (LukF)	B834 and DH5a
4HSC	<i>Streptococcus pyogenes</i>	Streptolysin O pore-forming toxin	XL-1 Blue
1CWV	<i>Yersinia pseudotuberculosis</i>	Invasin C-terminal passenger domain	not found
2M6X (1sNMR)	<i>Hepatitis C virus</i>	p7 hexamer channels (isolate EUH1480)	DH5α
1ZLL, 2M3B	<i>Homo sapiens</i>	Phospholamban homopentamer	BL21λ(DE3)
2KNC (1sNMR)	<i>Homo sapiens</i>	Human Integrin αIIbβ3 transmembrane-cytoplasmic heterodimer	BL21λ(DE3)
3VOU	<i>Bacillus weihenstephanensis</i> (NaK) and <i>Sulfitobacter pontiacus</i> (NaV)	NaK channel chimera with grafted C-terminal region of a NaV channel	KRX strain (Promega)
2F2B	<i>Methanothermobacter marburgensis</i>	Aquaporin water channel (AqpM)	not found
4BUO	<i>Rattus norvegicus</i>	Neurotensin receptor produced by direct evolution (NTS1)	BL21 TUNER
2M3G (1s NMR)	<i>Anabaena (Nostoc)</i> sp. PCC7120	Sensory Rhodopsin	BL21-Codonplus-RIL
2JLN	<i>Microbacterium liquefaciens</i>	Benzyl-hydantoin transporter (Mhp1)	BLR (Novagen)
4C7R	<i>Corynebacterium glutamicum</i>	Glycine betaine transporter (BetP)	DH5α
3W9J	<i>Pseudomonas aeruginosa</i>	Bacterial multi-drug efflux transporter (MexB)	JM109

Supplementary Table 2-Miroux

List of *E. coli* MP produced in *E. coli* hosts for structural studies

PDB via Mpstruc (S. White)	Type of membrane proteins	Name of bacterial host
MONOTOPIC		
1B12	Signal Peptidase (SPase) in complex with a β -lactam inhibitor	BL21 λ (DE3)
2QCU	Glycerol-3-phosphate dehydrogenase (GlpD, native).	XL1blue/JM109
1J79, 1XGE	Dihydroorotate Dehydrogenase.	XL1blue
BETA BARRELS		
1MPF	OmpF Porin from colicin-resistant.	BZB1107
1GFM	OmpF Porin, D113G mutant.	Top10
1BT9, 300E	OmpF Porin, D74A mutant.	BL21 λ (DE3)
1HXX	OmpF Porin, Y106F Mutant.	BL21 λ (DE3)Domp8
3HW9, 2ZFG	OmpF Porin	MH225
2J1N	OmpC Osmoporin.	BZB1107
2XE1	OmpC Osmoporin clinical variant OmpC06.	HN705 Domp8
2F1C	OmpG *monomeric* porin.	C43 λ (DE3)
2IWW	OmpG *monomeric* porin.	C41 λ (DE3)
2JQY	OmpG by solution NMR spectroscopy.	BL21 λ (DE3)pLysS
1MPM, 1MAL, 1AF6	LamB Maltoporin in complex with maltose.	K12 pop65I0
1EK9, 1TQQ	Outer membrane protein TolC central to multidrug efflux and protein export.	BL21 λ (DE3)
2VDE	TolC outer membrane protein (Y362F, R367E), partially open state.	C43 λ (DE3) and C41 λ (DE3)
3PIK, 4K7R, 4K7K, 4K34	CusC, the Outer Membrane Component of a Heavy Metal Efflux Pump.	C43 λ (DE3)/BL21Star λ (DE3)
1NQE/F/G/H	BtuB cobalamin transporter.	not found
1UJW	BtuB with bound colicin E3 R-domain.	TNE012 (K12 tsx-ompA- ompB-)
2GSK	BtuB:TonB.	BL21Star λ (DE3)pLysS
2YSU	Complex of the Colicin E2 R-domain and Its BtuB Receptor.	TNE012 (K12 tsx-ompA- ompB-)
2HDI	Colicin I receptor Cir in complex with Colicin Ia binding domain.	BL21 λ (DE3)
1QJP/1BXW, 1G90	OmpA.	BL21 λ (DE3)
2JMM	OmpA with four shortened loops: NMR Structure DHPC micelles.	BL21 λ (DE3)Gold
1I78	OmpT outer membrane protease.	BL21 λ (DE3), DH5 α and B884
2F1V	OmpW outer membrane protein.	C43 λ (DE3)
1ORM, 1Q9F	OmpX: , NMR (DHPC micelles).	BL21 λ (DE3)pLysS
2MO6	OmpX in optimized nanodiscs: NMR Structure In DPC micelles.	BL21 λ (DE3)
1FW2	OmpLA (PldA) outer membrane phospholipase A monomer with Ca ⁺⁺ .	BL21 λ (DE3)
1ILZ	OmpLA (PldA) active-site mutant (N156A).	BL21 λ (DE3) Δ pldA
2WJR	NanC Porin, model for KdgM porin family.	BL21 λ (DE3)pLysS
1MM4, 1MM5, 1THQ, 3GP6	PagP outer membrane palmitoyl transferase.	BL21 λ (DE3)
1T16, 3PGR	FadL long-chain fatty acid transporter.	C43 λ (DE3)
3DWN	FadL long-chain fatty acid transporter A77E/S100R mutant.	Ls6164, C43 λ (DE3)
1BY3	FhuA, Ferrichrome-iron receptor without ligand. With ligand: 1BY5.	B834(DE3)/BL21 λ (DE3)
1QKC, 1FI1, 2FCP	FhuA in complex with albomycin.	AW740 [DompF zch:TnlO DotnpCftuA31]
1FCP	SeMet-FhuA.	DL41
2GRX	FhuA in complex with TonB.	AW740
1FEP, 1PNZ	FepA, Ferric enterobactin receptor.	BL21 λ (DE3)
1KMO	FecA, siderophore transporter.	UT5600
2VQI	P pilus usher translocation domain.	B834 λ (DE3)
3RFZ	P pilus FimD usher bound to FimC:FimH substrate.	B834 λ (DE3)
4J30	P pilus FimD usher in complex with FimC:FimF:FimG:FimH.	Tuner λ (DE3)
2YNK	Wzi outer-membrane lectin.	Top10/B834 λ (DE3)
2QOM, 3SLJ, 3SZE	EspP autotransporter.	BL21 λ (DE3)
3AEH	Hbp (hemoglobin protease) self-cleaving autotransporter with truncated passenger.	C43 λ (DE3)
1WXR	Hbp (hemoglobin protease) full-length passenger domain.	DH5 α
4-E1S, 1F02	Intimin outer membrane β -domain.	BL21 λ (DE3)
4C00	TamA Autotransporter, full length.	BL21 λ (DE3)
TRANSMEMBRANE PROTEINS: ALPHA-HELICAL		
3FWM	Peptidoglycan Glycosyltransferase penicillin-binding protein 1b.	BL21 λ (DE3)
2WCD	Cytolysin A (ClyA, aka HlyE).	Tuner λ (DE3)
2J58	Wza translocon for capsular polysaccharides.	LE392
3JQO	Type IV outer membrane secretion complex.	B834 λ (DE3)
2OAU, 2VV5	MscS voltage-modulated mechanosensitive channel.	BL21 λ (DE3)
1RC2, 209D, 3NK5	AqpZ aquaporin water channel.	C43 λ (DE3)
2ABM	AqpZ aquaporin showing two conformations of Arg-189.	BL21 λ (DE3)pLysS
1FX8, 1LDF	GlpF glycerol facilitator channel.	not found
3KCU	FocA, pentameric aquaporin-like formate transporter.	BL21 λ (DE3)
1U7G	AmtB ammonia channel (mutant).	C41 λ (DE3)
1XQF, 2NMR, 2NUU, 2NS1	AmtB ammonia channel.	C43 λ (DE3)

3TXT, 3UBB, 2LEP, 4HDD	GlpG rhomboid-family intramembrane protease.	BL21λ(DE3)
2LZS	TatA, Twin arginine translocase.	BL21λ(DE3)pLysS
1OTS, 2EXW, 4FG6	H ⁺ /Cl ⁻ Exchange Transporter: Formerly ClC Chloride Channel.	BL21λ(DE3)
3NMO	Monomeric H ⁺ /Cl ⁻ Exchange Transporter.	stratagene
4ENE	H ⁺ /Cl ⁻ Exchange Transporter (truncated). Truncation: Residues 2-16 at N-terminal and 461-464 at C-terminal.	stratagene
1OY6	AcrB bacterial multi-drug efflux transporter.	DH5α
1T9T, 2GIF	AcrB bacterial multi-drug efflux transporter.	C43λ(DE3)
2HQC	AcrB bacterial multi-drug efflux transporter.	BL21λ(DE3)Gold
2DHH, 2RDD, 2W1B, 3AOB, 3W9H, 1IWG	AcrB bacterial multi-drug efflux transporter.	JM109
3K07, 3NE5	CusA metal-ion efflux pump.	BL21λ(DE3)
3OOC	CusB membrane fusion protein (apo protein).	DH5α
3B5D	EmrE bacterial multi-drug efflux transporter with bound TPP substrate.	BL21λ(DE3)
1PV7, 2CFQ	LacY Lactose Permease Transporter.	not found
2V8N	LacY Lactose Permease (wild-type) with TDG.	XL1blue
2Y5Y	LacY Lactose Permease with covalently bound MTS-gal.	C43λ(DE3)
4OAA	LacY Lactose Permease Transporter (G46W/G262W mutant) with bound lactose analog.	C41λ(DE3)
3O7Q	FucP Fucose Transporter in outward-facing conformation.	BL21λ(DE3)
1PW4	GlpT Glycerol-3-Phosphate Transporter.	LMG194
2GFP	EmrD Multidrug Transporter.	not found
4GBY	XylE proton:xylose symporter with bound D-xylose.	BL21λ(DE3)
4AJ3	XylE proton:xylose symporter in partially occluded inward-open state.	C41λ(DE3)
4IU9	NarU nitrate transporter.	not found
4JR9	NarK nitrate/nitrite exchanger. A member of the nitrate/nitrite porter family (NNP).	C41λ(DE3)
3WDO	YajR drug efflux transporter.	C43λ(DE3)
3QE7	Nucleobase/ascorbate transporter (NAT) UraA uracil/H ⁺ symporter.	BL21λ(DE3)
3HFX	CaIT carnitine transporter.	C41λ(DE3)
2WSX	CaIT carnitine transporter Fully-open inward-facing conformation.	BL21λ(DE3)
3LRB, 3L1L, 3OB6	AdiC Arginine:Agmatine Antiporter.	BL21λ(DE3)
4DJK	GadC glutamate-GABA antiporter.	not found
2QFI, 3H90	YiiP Zinc Transporter.	BL21λ(DE3)
1ZCD	NhaA Na ⁺ /H ⁺ antiporte.	rk20
3F11	NhaA Na ⁺ /H ⁺ antiporter Difference maps show structural changes with changes in pH.	BL21λ(DE3)
1L7V, 2QI9	BtuCD Vitamin B12 Transporter.	BL21λ(DE3)
4F13	BtuCD-F Vitamin B12 Transporter with bound AMP-PNP.	BL21λ(DE3)CodonPlusRIPL
2R6G, 4JBW, 3PV0	MalFGK2-MBP Maltose uptake transporter complex.	HN741
3RLF	MalFGK2-MBP Maltose uptake transporter complex with bound MgAMPPNP.	HN597 (DuncB-C i1ur:TnlO araD lac rpsLI/F' laca lacZ::Tn5, proA ⁺ proB ⁺)
4KHZ, 4KI0	MalFGK ₂ -MBP Maltose uptake transporter complex; pre-translocation conformation bound to maltoheptaose.	AD126, Top10F', BL21I(DE3)
3DHW	MetNI Methionine uptake transporter complex MetN-C2 domain.	BL21λ(DE3)Gold
3OAA	F1-ATPase in an autoinhibited conformation.	not found
1A91	Subunit C of the F1Fo ATP synthase.	MEG119
2KDC	Diacylglycerol kinase (DAGK). Domain-swapped homotrimer.	BL21λ(DE3) WH1061
3ZE4	Diacylglycerol kinase (DAGK).	not found
2QCU	Glycerol-3-phosphate dehydrogenase (GlpD, native).	XL1Blue
1Q16, 1SIW, 1Y4Z	NarGHI Nitrate Reductase A.	LCB2048
2LTQ	Dsbb in POPE lipid bilayer: Cys41Ser mutant. Solid-state NMR used to refine the X-ray structure 2ZUQ.	C43λ(DE3)
4GD3, 3USE	O2-tolerant Hydrogenase-1 in complex with cytochrome b: Structure includes transmembrane helices.	FT004
3M9C, 3RKO	Electron Transport Chain Complexes I.	BL21λ(DE3)
1FUM	Electron Transport Chain Complexes: Complex II Fumarate Reductase Complex	not found
1LOV	Native Fumarate Reductase Complex: +HQN0.	DW35
1KQF	Formate dehydrogenase-N: HQNO complex.	GL101
1NEK, 2ACZ, 2WDQ, 2WP9	Succinate:quinone oxidoreductase (SQR, Complex II).	MC4100
1FFT	Electron Transport Chain Complexes: Complex IV (Cytochrome C Oxidase).	G0105 lacking terminal oxydases