This is a repository copy of *Microbial expression systems for membrane proteins*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/129554/

Version: Published Version

**Article:**
Dilworth, MV, Piel, MS, Bettaney, KE et al. (9 more authors) (2018) Microbial expression systems for membrane proteins. Methods, 147. pp. 3-39. ISSN 1046-2023

https://doi.org/10.1016/j.ymeth.2018.04.009

---

**Reuse**
See Attached

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
Microbial expression systems for membrane proteins

Marvin V. Dilworth, Mathilde S. Piel, Kim E. Bettaney, Piykee Ma, Ji Luo, David Sharples, David R. Poyner, Stephanie R. Gross, Karine Moncoq, Peter J.F. Henderson, Bruno Miroux, Roslyn M. Bill.

A R T I C L E   I N F O

Keywords:
- Recombinant membrane proteins
- Expression plasmid vector
- Tag
- Promoter
- Detergent

A B S T R A C T

Despite many high-profile successes, recombinant membrane protein production remains a technical challenge; it is still the case that many fewer membrane protein structures have been published than those of soluble proteins. However, progress is being made because empirical methods have been developed to produce the required quantity and quality of these challenging targets. This review focuses on the microbial expression systems that are a key source of recombinant prokaryotic and eukaryotic membrane proteins for structural studies. We provide an overview of the host strains, tags and promoters that, in our experience, are most likely to yield protein suitable for structural and functional characterization. We also catalogue the detergents used for solubilization and crystallization studies of these proteins. Here, we emphasize a combination of practical methods, not necessarily high-throughput, which can be implemented in any laboratory equipped for recombinant DNA technology and microbial cell culture.

1. Recombinant membrane protein production in microbes

Few membrane proteins are naturally abundant in their native membranes; in order to characterize them biophysically and biochemically, recombination of their genes with more efficient promoters and regulators of expression are required [1]. Unsurprisingly, the few naturally-abundant membrane proteins (including mammalian and bacterial rhodopsins, aquaporins and complexes involved in respiration and photosynthesis) were amongst the first to have their crystallographic structures solved: the first high-resolution structure of a membrane protein was that of the photosynthetic reaction centre from Blastochloris viridis published in 1995 [2]. In 1998, the first recombinant membrane protein structures were published: those of the prokaryotic proteins Mscl [3] and KcsA [4], both produced in Escherichia coli. The first structures of recombinant mammalian membrane proteins were solved in 2005 using protein that had been produced in yeast cells: the rabbit Ca\(^{2+}\)-ATPase, SERCA1a, was produced in Saccharomyces cerevisiae [5] and the rat voltage-dependent potassium ion channel, Kv1.2 was produced in Pichia pastoris [6]. These early results established microbes as efficient and effective host systems for synthesizing membrane proteins.

While baculovirus-infected insect cells and mammalian cell-lines have been used very successfully for both prokaryotic and eukaryotic membrane protein production [1], we note that microbes have remained a consistently-popular choice because they are quick, easy and cheap to culture and they can produce high-quality protein suitable for subsequent study. In November 2017, Stephen White’s database (blanco.biomol.ucd.edu/mpstruc/) recorded that almost a third (31%) of all membrane protein coordinate files deposited in the Protein Data Bank (www.rcsb.org/pdb/home/home.do; PDB) were derived from recombinant proteins; notably, 71% of all unique structures were derived from microbial sources. Of these, 64% were produced in E. coli, 4% in P. pastoris and 3% in S. cerevisiae.

This review focuses on current approaches to selecting expression plasmids (especially with respect to their purification tags and promoters), microbial strains and culture conditions to enable the detergent-based purification of functional membrane proteins for biophysical characterization and crystallization trials (for subsequent

https://doi.org/10.1016/j.ymeth.2018.04.009
Received 18 January 2018; Received in revised form 8 March 2018; Accepted 10 April 2018

Please cite this article as: Dilworth, M.V., Methods (2018), https://doi.org/10.1016/j.ymeth.2018.04.009
Table 1
Unique membrane protein structures derived from recombinant E. coli proteins produced in E. coli (homologous expression).

| PDB   | Name                                                                 | TM | Length | Strain(s) | Vector | Tag       | Size | N/C | Solubilization | Structure | Date  |
|-------|----------------------------------------------------------------------|----|--------|-----------|--------|-----------|------|-----|---------------|-----------|-------|
| 1B12  | Signal peptidease (SPase)                                            | 0  | 280    | BL21(DE3) | pET3b  | NT        | 6    | N   | IB-gua        | X-ray-TX100 | 1999  |
| 1J79  | Dihydroorotate dehydrogenase (DdoB)                                  | 0  | 347    | XLI-Blue  | NA     | No vector | 6    | NT  | None         | None      | 2001  |
| 2QC2  | Glycerol-3-phosphate dehydrogenase (GlpD)                            | 0  | 501    | XLI-Blue/ JM109 | pQE30 | His | 6 | N  | OG          | X-ray-OG   | 2008  |

**Alpha-helical**

| 1A91  | Subunit C of the F_{1}F_{o} ATP synthase                            | 2  | 79     | MEG119 NA | pC35   | NS        | NS   | NS  | NS           | InNMR-solvent | 1998  |
| 1FFT  | Electron transport chain complexes                                   | 15 + 2 + 5 + 3 | 663 + 315 + 204 + 109 | GO105 NA | pRHis6 NI | His | 6 | C  | MD           | X-ray-OG   | 2000  |
| 1FX8  | Glyceraldehyde 3-phosphate dehydrogenase (Glyc)                     | 3  | 281    | NS        | NS     | NS        | NS   | NS  | NS           | X-ray-OG   | 2000  |
| 1L0V  | Native fumarate reductase complex (FrdA)                             | 3 + 3 | 602 + 244 + 131 + 119 | DW35 NA | pHis3 NI  | NT   | C   | 12E9 | X-ray-C12E9  | 2002  |
| 1KFQ  | Formate dehydrogenase-N (FdhGHI)                                    | 1 + 4 | 1005 + 294 + 217 | GL101 NA | pET    | His | 10 | N   | LDAO         | X-ray-LDAO | 2002  |
| 1L7V  | Vitamin B12 transporter (BtuCD)                                      | 10  | 326 + 249 | BL21(DE3) | pET    | His | 10 | N   | LDAO         | X-ray-LDAO | 2002  |
| 1NEK  | Glycerol-3-phosphate dehydrogenase (GlpD)                           | 3 + 3 | 115 + 129 + 238 + 588 | DW35 NA | pSDH15 NI | NT   | C   | 12E9 | X-ray-C12E9  | 2003  |
| 1OT5  | Multi-drug efflux transporter (EmrD)                                 | 14  | 473    | BL21(DE3) | pET28b | His | 6 | C   | DM          | X-ray-DDM or DM | 2003  |
| 1PW4  | Glycerol-3-phosphate dehydrogenase (GlpD)                           | 12  | 1049   | DHF5a     | pUC515A | NT   | 6 | C   | DDM         | X-ray-DDM | 2003  |
| 1P77  | Lactose permease (LacY)                                             | 12  | 417    | XLI-Blue  | pT7-5  | His | 10 | C   | DDM         | X-ray-DDM | 2003  |
| 1Q16  | Nitrate reductase A (NarGHI)                                        | 5 + 2 | 1247 + 512 + 225 | LG2048 | pAV700 | NT   | C   | 12E9 | X-ray-C12E9 | 2003  |
| 1RC2  | Aquaporin (AqpZ)                                                    | 8 + 4 | 231    | C43(DE3) | pET28b | His | 6 | N   | OG          | X-ray-OG   | 2003  |
| 1U7G  | Ammonia channel (AmlB)                                              | 11 + 3 | 428   | C43(DE3) | pET28b | His | 6 | C   | DDM         | X-ray-OG   | 2003  |
| 1X0F  | Ammonia channel (AmlB)                                              | 11 + 3 | 428   | C43(DE3) | pET28b | His | 6 | C   | DDM         | X-ray-OG   | 2003  |
| 1ZCD  | Na" / H" + antipporter (NhaA)                                       | 12  | 388    | Rk20 + B834DE3 | pAXH | His | NS | N  | DDM         | X-ray-DDM or Cyma6 | 2005  |
| 2ABB  | Aquaporin (Agg)                                                     | 8 + 4 | 231    | BL21(DE3) | pLyS6 | NS    | His | NS | N         | X-ray-OG   | 2005  |
| 1T9T  | Multi-drug efflux transporter (AcrB)                                 | 12  | 1049   | C43(DE3) | pUC515A, pET24a (2GFP) | NT or His | 6 | C   | DDM         | X-ray-DDM or Cyma6 | 2005  |
| 2GF3  | Multidrug transporter (EmrD)                                        | 12  | 375    | NF        | NF      | NF    | NF | NF | NF          | X-ray-DDM | 2006  |
| 2DH1  | Multi-drug efflux transporter (AcrB)                                 | 12  | 1053   | JM109, W3104 | pACB3 (pUC118) | His | 6 | C   | DDM         | X-ray-DDM or MD | 2006  |
| 2J5B  | Translocon for capsular polysaccharides (Wza)                        | 1 + 8 | 359    | LE392    | pWQ126 (pBAD24) | NT | 6 | SB3-14 | X-ray-DDM | 2006  |
| 2HI7  | Periplasmic oxidase complex (DshB-DshA)                              | 4   | 176 + 208 | M15     | pQE70  | His | 6 | C   | UDM        | X-ray-DDM | 2006  |
| 2OAU  | Voltage-modulated mechanosensitive channel (MscS)                    | 3 + 7 | 286    | BL21(DE3) | pET28b | His | 6 | N   | FC14        | X-ray-FC14 | 2007  |
| 2QC2  | Multi-drug efflux transporter (AcrB)                                 | 12  | 1053   | BL21(DE3) Gold | pSORT1 | His | 6 | C   | DDM         | X-ray-DDM | 2007  |
| 2VBN  | Lactose permease (LacY)                                             | 12  | 417    | XLI-blue | pT7-5  | His | 10 | C   | DDM         | X-ray-DDM | 2007  |
| 2QFI  | Zinc transporter (YizP)                                             | 6 + 2 | 300    | BL21(DE3) | pLyS6 | pET15b | His | 6 | N         | X-ray-DDM | 2007  |
| 2R6G  | Maltose uptake transporter (MaltFGK2, 8 MaLF + 6 MaLG)               | 514 + 296 + 381 + 370 | HM741    | pACYC184 + PFG23 | X-ray-UOM | 6 | C   | DDM         | X-ray-DDM | 2007  |
| 3B5D  | Multi-drug efflux transporter (EmsE)                                 | 4 + 2 | 110    | BL21(DE3) | pET15b | His | 6 | N   | NG          | X-ray-NG  | 2007  |
| 3DHW  | Methionine uptake transporter (MetNI)                                | 5 + 2 | 343 + 217 | BL21(DE3) Gold | pET19b | His | 10 | N   | DDM         | X-ray-DDM | 2008  |
| 3EJU  | Periplasmic oxidase complex (DshB-DshA)                              | 4   | 208 + 176 | HM125   | pQE70  | His | 6 | C   | NM          | X-ray-NM  | 2008  |
| 3FI1  | Na" / H" + antipporter (NhaA)                                       | 12  | 388    | BL21(DE3) | pET    | His | 6 | C   | DDM         | EM-lipids (2D) | 2009  |
| 3ZUO  | Periplasmic oxidase complex (DshB-DshA)                              | 4   | 176    | M15     | pQE70  | His | 6 | C   | UDM         | X-ray-DFPC | 2009  |
| 2WCD  | Cytosolysin A (Cya, HlyE)                                           | 3 + 12 | 309    | Tuner(DE3) | pET11a | His | 6 | N   | None        | X-ray-DDM | 2009  |
| 2KDC  | Diacylglycerol kinase (DAGk), Domain-swapped homotrimer              | 3 + 3 | 122    | BL21(DE3) | pSD005 | His | 6 | N   | Empigen     | InNMR-DPC | 2009  |

(continued on next page)
| PDB  | Name                                      | TM | Length | Strain(s) | Vector   | Tag     | Size | N/C | Solubilization | Structure | Date       |
|------|-------------------------------------------|----|--------|-----------|----------|---------|------|-----|----------------|------------|------------|
| 3JQO | Type IV OM secretion complex              | 1×14 | 396     | B834(DE3) | pASK-IBA3 | Strep   | C    | MD (1) | X-ray-LDAO     | 2009       |
| 3KCU | Aquaporin-like formate transporter (FocA) | 6×5  | 285     | BL21(DE3) | pET21b   | His     | 6    | C   | OG             | X-ray-MD   | 2009       |
| 3R6B | Arginine:agmatine antipporter (AdfC)      | 12  | 445     | BL21(DE3) | pET15b   | His     | 6    | N   | DDM            | X-ray-MD   | 2010       |
| 3HFX | Electron transport chain complexes I (NuoLMNJK) | 14+4+14+3+4 | 613 + 509 + 485 + 100 + 147 + 194 | BL21(DE3) | No vector | NT (1) | DDM | X-ray-MD       | 2010       |
| 3NMO, 4ENE, 4MQX | H⁺ /Cl⁻ exchange transporter | 14+5 | 465     | NS (2) | pASK-IBA2 | His     | 6    | C   | DM             | X-ray-MD   | 2010       |
| 2WSX | Carnitine transporter (CaT)               | 12  | 304     | BL21(DE3) | pET15b   | His     | 6    | N   | DDM            | X-ray-MD   | 2010       |
| 3KQ7 | Metal-ion efflux pump (CusA)              | 12  | 304     | BL21(DE3) | pET15b   | His     | 6    | N   | DDM            | X-ray-MD   | 2010       |
| 3Q67 | Lactose permease (LacY)                   | 12  | 304     | BL21(DE3) | pET15b   | His     | 6    | N   | DDM            | X-ray-MD   | 2010       |
| 2LYS | Peptidoglycan glycosyltransferase         | 1   | 768     | BL21(DE3) | pET15b   | His     | 6    | N   | DDM            | X-ray-MD   | 2010       |
| 4DJK | Glutamate-GABA antipporter (GadC)         | 12  | 304     | NS (2) | pET15b   | His     | 6    | N   | DDM            | X-ray-MD   | 2010       |
| 4E3  | Vitamin B12 transporter (BtuCDF)          | 10  | 304     | BL21(DE3) | CodonPlus | pET     | His  | 10 | NDM            | X-ray-C28B | 2010       |
| 4GBY | Proton:xylose symporter (XylE)            | 12  | 304     | BL21(DE3) | pET15b   | His     | 6    | N   | DDM            | X-ray-MD   | 2010       |
| 4GD3, 3UJE | O₂-tolerant hydrogenase-I in complex | 1+1+4 | 372 + 597 + 235 | FT004 NA (3) | No vector | His    | 6    | C   | TX100          | X-ray-MD   | 2013       |
| 2LTQ | Periplasmic oxidase (DobB)                | 4   | 768     | C43(DE3) | pQE70    | His     | 6    | C   | DDM            | mNMR-E. coli | 2013       |
| 2LZS | Twin arginine translocase (TatA)          | 1×9 | 55      | BL21(DE3) | pLysS    | pET24a  | His  | 6   | C   | E29          | X-ray-MD   | 2013       |
| 4U9U | Nitrile transporter (NarU)                | 12  | 445     | BL21(DE3) | pET15b   | His     | 6    | N   | DDM            | X-ray-MD   | 2010       |
| 4JAR | Proton:xylose symporter (XylE)            | 12  | 445     | BL21(DE3) | pET15b   | His     | 6    | N   | DDM            | X-ray-MD   | 2010       |
| 3Z64 | Dicarboxylglycerol kinase (DAK)           | 3×3 | 122     | BL21(DE3) | pET21b   | His     | 6    | N   | DDM            | X-ray-MD   | 2010       |
| 4A5U | Na⁺ /H⁺ antiporter (NhaA)                 | 12  | 388     | BL21(DE3) | pLysS    | pET derivative | His    | 6 | C   | DDM            | X-ray-MD   | 2013       |
| 3WDO | Drug efflux transporter (YajR)            | 12  | 454     | C43(DE3) | pET28a   | His     | 6    | N   | DDM            | X-ray-MD   | 2010       |
| 4C68 | Multi-drug efflux transporter (AcrB/Z)    | 12+1 | 1049 + 49 | C43(DE3) | pET21a or pRSFduet-1 | His    | 6    | C   | DDM            | X-ray-MD   | 2013       |
| 4PX7 | Phosphatidylglycerophosphate phospotase B (PggB) | 6 | 254     | C43(DE3) | pET28a   | His     | 6    | N   | DDM            | X-ray-MD   | 2013       |
| 4PLO, 5OFR | Antimicrobial peptide transporter (McjD) | 6×2 | 580     | C43(DE3) | pValdoGPPd | His    | 6    | C   | DDM            | X-ray-MD   | 2014       |
| 2MNP | Rhodanese (YgaP)                         | 2×2 | 68      | BL21(DE3) | pET3a    | His     | 6    | N   | DHC/MPG        | X-ray-MD   | 2014       |
| 4QKQ | Proton:xylose symporter (XylE)            | 12  | 474     | C43(DE3) | pET15b   | His     | 6    | N   | DDM            | X-ray-MD   | 2014       |
| 4Q6S | Peptide transporter (YghG)                | 14  | 493     | C43(DE3) | pET28a   | His     | 6    | N   | DDM            | X-ray-MD   | 2014       |
| 4Q02 | Rhodobactin protein (GlpG)               | 6   | 200     | C43(DE3) | pET24a   | His     | 6    | N   | DDM            | X-ray-MD   | 2014       |
| 4UNV, 5JMN | Multi-drug efflux transporter (AcrB) | 12  | 1057    | C43(DE3) | pET24a   | His     | 6    | N   | DDM            | X-ray-MD   | 2014       |
| 3WSP | Insertase (YidC)                         | 5   | 548     | BL21(DE3) | pET28a   | His     | 8    | C   | DDM            | X-ray-MD   | 2014       |
| 4XSM | SemiSWEET transporter                   | 3×2 | 89      | BL21(DE3) | pET21b   | His     | 8    | C   | DDM            | X-ray-MD   | 2014       |
| 4RP9 | Vitamin C transporter (UsaA)              | 10  | 465     | C43(DE3) | pET15b or pET21b | His    | 8    | C   | NM            | X-ray-MD   | 2015       |
| 5AJ | Voltage-modulated mechanosensitive channel (MacS) | 3×7 | 286     | MJFH612 | pRH616   | His     | 6    | C   | DDM            | X-ray-MD   | 2015       |
| 414V | Nitrate/nitrite exchanger (NarK)         | 12  | 463     | C43(DE3) | pET     | His     | 8    | C   | DDM            | X-ray-MD   | 2015       |
| 414V | 40AA                                      | 12  | 417     | C41(DE3) | pETIb    | His     | 6    | C   | DDM            | X-ray-MD   | 2015       |
| 42D1 | Multidrug resistance transporter (MdfA)  | 12  | 410     | C41(DE3) | pET2a    | His     | 6    | C   | DDM            | X-ray-MD   | 2015       |
| 4UXS | Dicarboxylglycerol kinase (DAK)           | 3×3 | 122     | WH1061   | pTRH6s8  | His     | 6    | N   | Empigen        | X-ray-MD   | 2015       |
| 5AZC | Phosphatidylglycerol: phospholipid transferase (Lgl) | 7 | 291     | C43(DE3) | pET28a   | His     | 6    | C   | DDM            | X-ray-MD   | 2015       |
Table 1 (continued)

| PDB   | Name                          | TM | Length | Strain(s)                          | Vector       | Tag | Size | N/C | Solubilization | Structure | Date   |
|-------|-------------------------------|----|--------|------------------------------------|--------------|-----|------|-----|----------------|-----------|--------|
| 5F5B  | Rhomboid protease (GlpG)      | 6  | 211    | C43(DE3)                           | pGEX-6P-1    | GST | N    | DDM | X-ray-CHAPSO   | Nicelles  | 2016   |
| 5KNN  | F-ion channel homologue (Fluc)| 4×2| 126    | BL21(DE3)                          | pASK-IBA2    | His | 6    | C   | X-ray-DM       |           | 2016   |
| 5J4N  | Arginine:agmatine antiporter (AdiC) | 12 | 445    | BL21(DE3) pLysS                    | pZUDE21      | His | 10   | C   | X-ray-NG       |           | 2016   |
| 5SV0, 5SV1 | Lactose permease (LacY)    | 3+1| 244 + 58| BL21(DE3) pET28b(ExbB) + pCDF-1b(ExbD) |               |     |      |     | X-ray-DDM      |           | 2016   |
| 5GXB  | Lactose permease (LacY)      | 12 | 417    | BL21(DE3)                          | pTT7-S       | His | 6    | C   | X-ray-DDM      | LCP       | 2016   |
| 5T4O  | Frα, ATP synthase            | 5+1+2| 513 + 460 + 287 + 177 + 139 + 271 + 155 + 79|               |              |     |      |     | X-ray-CHAPSO   |           | 2016   |
| 5JII  | Sensor histidine kinase (NarQ)| 4 | 230    | SI1                                | pSCodon1.2   | His | 6    | C   | X-ray-DDM      |           | 2016   |
| 5N6H  | Apolipoprotein N-acyltransferase (Lnt) | 8 | 512    | C43(DE3)                          | pET28a       | His | 6   | N   | X-ray-DM (LCP) |           | 2016   |
| 5XHQ  | Apolipoprotein N-acyltransferase (Lnt) | 8 | 512    | C41(DE3) pET19b                   | pET22b       | His | 8   | C   | X-ray-DM       |           | 2016   |
| 5TV4  | Lactose permease (LacY)      | 6  | 582    | BL21(DE3) Gold                    | pET3b        | His | 10  | N   | DDM           |           | 2016   |

Beta-barrel

| PDB   | Name                          | TM | Length | Strain(s)                          | Vector       | Tag | Size | N/C | Solubilization | Structure | Date   |
|-------|-------------------------------|----|--------|------------------------------------|--------------|-----|------|-----|----------------|-----------|--------|
| 1MPF  | Porin from colicin-resistant (OmpF) | 16 | 340    | B2B1107 NA^3                       | Top10        |     | NT^5 | OPOE| X-ray-MD       |           | 1995   |
| 1GFM  | Porin (OmpF)                  | 16 | 340    | Popp6510 NA^3                      | pAC1         |     | NT^5 | OPOE| X-ray-MD       |           | 1996   |
| 1MMP  | Maltooligosaccharide permease  | 16 | 2757   | P1601 NA^3                         | pAC1         |     | NT^5 | OPOE| XRAY-OPOE or MD|           | 1997   |
| 1F79, 50OE | Porin (OmpF)             | 16 | 340    | BL21(DE3)                          | pGEM-SZf (+) lac |     | NT^5 | SDS | X-ray-OPOE      |           | 1999   |
| 1BY3, 1BY5 | Ferrichrome-iron receptor (FhuA) | 22 | 747    | B834(DE3)/21(DE3)                  | pET         |     | NT^5 | SDS | X-ray-OPOE      |           | 1999   |
| 1IFE  | Siderophore transporter (FecA) | 22 | 747    | BL21(DE3) pET17b                   | TX100        |     | NT^5 | SDS | X-ray-OPOE      |           | 1999   |
| 1QJG  | Porin (OmpK)                  | 8  | 148    | BL21(DE3) pET3b                    | IB-gua^6     |     | NT^5 | SDS | X-ray-C8E4      |           | 1999   |
| 1QKC, 1F11, 2FCP, 1FCP | Ferrichrome-iron receptor (FhuA) | 22 | 747    | AW740 [ompF]                       | pHIx405      |     | NT^5 | SDS | X-ray-C8E4      |           | 1999   |

(continued on next page)
| PDB | Name                                  | TM | Length | Strain(s) | Vector       | Tag   | Size | N/C | Solubilization | Structure | Date    |
|-----|---------------------------------------|----|--------|-----------|-------------|-------|------|-----|----------------|-----------|---------|
| 2JQY | Porin (OmpG)                          | 14 | 280    |           | BL21(DE3) pIysS | pT7-SMC | NT   | IB-urea | lsNMR-FC12     | Xray-MD   | 2007    |
| 2QOM | Autotransporter (EspP)                | 12 | 285    |           | BL21(DE3)    | PC6H1  | His  | 6   | C              | Elugent   | Xray-MD | 2008    |
| 2VDE | OM protein (TolC)                     | 4  | 460    |           | C43(DE3)/C41(DE3) | pET14b  | His  | 8   | C              | Tx100     | Xray-DDM | 2008    |
| 2VQI | P pilus usher translocation (PapC)    | 24 | 515    |           | BR34(DE3)   | pDG2   | His  | 6   | C              | DDM       | Xray-MD  | 2008    |
| 3DWN | Long-chain fatty acid transporter (FadL) | 14 | 427    |           | C43(DE3)    | pBAD   | His  | 6   | C              | LDAO      | Xray-C8E4 | 2008    |
| 2WRJ (Mb), 2WJQ | Porin (NanC) | 12 | 214    |           | BL21(DE3) pIysS | pT7 | NT  | OP0E, IB-urea | Xray-LDAO, Xray-FC12 | 2009 |
| 3HW9, 2ZFG | Porin (OmpF) | 16 | 340    |           | Ml225 NA | pR272  | NT  | IB-gua | Xray-OG | Xray-MD | 2009    |
| 3PIK, 4K7R, 4K34 | Hemoglobin protease (Hbp) | 12 | 308    |           | C43(DE3)    | pET22b | His  | 6   | C              | SB3-12 DDM | Xray-MD | 2011    |
| 3RFZ | P pilus (FimD)                        | 24 | 843 + 211 + 279 | B834(DE3) / pAN2 (fimD) + pET31001 (fimC) | Strep + His | 6   | C   | DDM | Xray-MD | 2011    |
| 3OHN | P pilus (FimD)                        | 24 | 558    |           | BR34(DE3)   | pNH297/pET54 | His  | 6   | C              | DDM       | Xray-C8E4 | 2011    |
| 4E1S, 1F02 | Intimin outer membrane β-domain | 12 | 242    |           | BL21(DE3)   | pET9/pET21a | His  | 10  | N              | Elugent   | Xray-LDAO | 2012    |
| 2M06 | Porin (OmpX)                          | 18 | 148    |           | BL21(DE3)   | pET11a  | NT  | IB-gua | LDAO or DDM | C8E4 Cymal6 | 2012 |
| 4J30 | P pilus (FimD)                        | 24 | 843 + 211 + 279 | Tuner(DE3) / pNH237 (fimC) + pAN2 | Strep + His | 6   | C   | DDM | Xray-MD | 2013    |
| 2YNK | OM lectin (Wzi)                       | 18 | 456    |           | Top10/B834(DE3) | pBAD | His  | 6   | C              | SB3-14 Xray-LDAO | 2013 |
| 4C00 | Auto transporter (TamA)               | 16 | 559    |           | BL21(DE3)   | pET22b  | His  | 6   | N              | OG       | Xray-OG | 2013    |
| 4K7K | Heavy metal efflux pump (CusC)        | 4  | 446    |           | BL21(DE3) Star | pET15b | His  | 6   | C              | Cymal6 Xray-cymal6 | 2011 |
| 4C4V | (BamA)                                | 16 | 467 (344-810) | BL21(DE3) | pET30b | His  | 6   | C              | IB-gua | Xray-MD | 2014    |
| 4UV3 | Amyloid secretion channel (CsgG)      | 4 × 9 | 277 |           | BL21(DE3)   | pQlinkN  | Strep | C   | DDM | Xray-MD | 2014    |
| 4Q79 | Amyloid secretion channel (CsgC)      | 4 × 9 | 277 |           | BL21(DE3)   | pQlinkN  | Strep | C   | DDM | Xray-MD | 2014    |
| 4SU | Porin (OmpF)                          | 16 | 340    |           | C43(DE3) NA  | pNH114  | His  | 8   | C              | DDM       | Xray-C8E4 | 2016    |
| 5EOQ | β-barrel assembly machine (BamACDE)   | 16 | 810 + 344 + 245 + 113 | BL21(DE3) | pNH114 | His  | 8   | C              | DDM       | Xray-MD | 2016    |
| 5AYW | β-barrel assembly machine (BamABCDE)  | 16 | 810 + 392 + 344 + 245 + 113 | C43(DE3) | pQlink  | Strep + His | 6   | C   | DDM | Xray-MD | 2016    |
| 5DOO | β-barrel assembly machine (BamABCDE)  | 16 | 810 + 392 + 344 + 245 + 113 | BD8150 | pGY120 | His  | 8   | C   | DDM | Xray-MD | 2016    |
| 5JO | β-barrel assembly machine (BamABCDE)  | 16 | 810 + 392 + 344 + 245 + 113 | BL21(DE3) | pJH114 | His  | 8   | C   | DDM | EM-DDM | 2016    |
| 5WQ7 | Secretin (GspD)                       | 16 | 650    |           | DH5a       | pASK-I8A3c | Strep | C   | MD | EM-MD | 2016    |

1 Number of transmembrane domains.  
2 N- or C-terminal position.  
3 Solubilization detergent.  
4 Detergent used for structure determination (NB: LCP is lipid cubic phase; lsNMR is solution phase NMR; ssNMR is solid state NMR).  
5 No tag.  
6 Inclusion bodies solubilized in 6 M guanidine hydrochloride.  
7 Not specified in PDB or corresponding publication.  
8 E. coli membrane protein produced using its native promoter.  
9 No induction.  
10 Mixed detergent.  
11 Not found; article was not accessible.  
12 Other; the tag was inserted within the protein.
structural studies by X-ray crystallography, as well as the newly-invigorated technique of electron microscopy [7]). We have experience of automated methods using robots that, once commissioned and optimized, can dramatically increase the number of constructs and hosts explored and reduce the time required to reach success. This review is intended for laboratories without access to such facilities, meaning that the approaches discussed here should be widely applicable.

2. An overview of microbial expression hosts, tags and promoters

2.1. An overview of microbial host usage

The expression systems used in generating high-resolution structures of recombinant membrane proteins have been documented by Stephen White in his analysis of the PDB (blanco.biomol.uci.edu/mpstruc/). Biophysical studies of membrane proteins (especially NMR and crystallographic techniques) require large quantities (0.1–10 mM) of homogenous, correctly-folded, purified protein; a focus on data extracted from the PDB has therefore allowed us to identify systems that have the capability of producing the required quantity and quality of these challenging targets. This review updates our previous study [8] of E. coli expression systems and extends that work to include S. cerevisiae and P. pastoris. Together, these three host systems account for the production of the vast majority of recombinant membrane proteins in microbes, although Lactococcus lactis (see PDB entry 4US3), Pseudomonas fluorescens (SKUD) and Schizosaccharomyces pombe (2PNO) have also been used successfully as microbial cell factories in a minority of cases.

In November 2017, of the 729 unique membrane protein structures (uMPS) derived from recombinant proteins and deposited in the PDB, 521 were produced in microbial host cells. E. coli was clearly the cell factory of choice (producing 468 uMPS, Tables 1 and 2), followed by the yeast hosts, P. pastoris (31 uMPS, Table 3) and S. cerevisiae (22 uMPS, Table 4). Table 5 summarizes the use of microbial expression systems and the origin of the target uMPS. With a growing number of uMPS being deposited in the PDB, heterologous membrane protein production is becoming dominant over the production of homologous targets. Higher eukaryotic uMPS, in particular, have more recently been obtained using all three microbial systems (Table 5).

Yeast expression systems have been used almost exclusively in the production of large, eukaryotic membrane proteins; in the case of P. pastoris, the targets were mainly of mammalian and plant origin (Table 3). Yeast hosts have mainly produced α-helical membrane proteins, while E. coli has also been used to produce β-barrel proteins, probably because many such proteins are found natively in the E. coli outer membrane (Fig. 1).

Fig. 2 shows that above 500–600 amino acids (~50–60 kDa), the number of uMPS decreases dramatically, suggesting that E. coli cannot efficiently produce large proteins; this may be because ribosomes drop off very long mRNAs leading to incomplete synthesis products. In contrast, yeast expression systems can cope with larger proteins up to 1400 residues in length (~150 kDa) (Fig. 2). When we interrogated the data for eukaryotic (mammalian, plant, fish, anemone and worm) membrane proteins produced in E. coli (Table 6), we identified 47 uMPS. Of these 47 uMPS, 7 were monomeric membrane proteins and 23 were small peptides or proteins containing only one transmembrane domain. Fifteen uMPS were produced as inclusion bodies and subsequently refolded, 17 were purified in mild detergent and, of those, 3 were membrane proteins with more than 4 transmembrane α-helices that had been crystallized in the presence of detergent (PDB codes: 2Q7M; 4BUO; 4O6Y) and 1 was studied by electron microscopy (3DWW).

2.2. An overview of tag usage in microbial expression systems

Construct design is an integral part of defining an appropriate expression system, with key considerations being the size and predicted secondary structure of the target protein as well as the planned purification strategy. SMART (protein domain identification; http://smart.embl-heidelberg.de/help/smart_about.shtml) or Jpred (secondary structure prediction; http://www.compbio.dundee.ac.uk/jpred/) approaches can be used to describe the protein architecture and may help in deciding where to place any tags.

Addition of a polyhistidine tag is the most popular strategy for large-scale purification of recombinant membrane proteins on nickel-affinity columns. This is especially true for E. coli, where other affinity purification tags have had very little impact: 392 of the 447 tagged proteins produced in E. coli contain a polyhistidine tag; 18 contain a GFP tag and 17 were fused to maltose binding protein (MBP). Other tags such as Strep and Flag account for no more than 13 uMPS (Fig. 3A, Tables 1 and 2). While the overall numbers are lower for uMPS from yeast-derived proteins, GFP is emerging as a useful tag to track the purification of proteins from yeast membranes (Fig. 3B, Tables 3 and 4). GFP can be particularly useful for monitoring production yields or the oligomeric state of a membrane protein-GFP fusion via fluorescent size exclusion chromatography experiments (F-SEC) [9]. Other tags (e.g. Strep and Flag) are also more frequently used in yeast expression systems (Fig. 3B).

Irrespective of the host used, polyhistidine tag placement is approximately equally favoured at the amino- or carboxy-terminus of the target protein (Fig. 4A, Tables 1–4). For constructs with amino-terminal tags, protein synthesis is usually initiated using a sequence of at least three amino acids before that of the tag. For example, in plasmid pRSET (Invitrogen), the sequence is MRGHis6, while the protein used to solve structure 4V3G contained the following amino-terminal tag: ANVRLHI-His6-LE (Table 2). Fig. 4B shows that 35% of uMPS produced in microbial host cells contained polyhistidine tags with more than 6 histidines. An interesting example is the insertion of a tandem array of 6 histidines separated by a glycine (see 5D07, Table 3).

Fluorescent tags are an increasingly popular choice for examination of protein quality [10]. In bacteria, dual Ribosome-Binding-Site (RBS) expression vectors such as pET-Duet (Novagen) enable the cloning of a gene encoding a reporter fluorescent protein downstream of the target gene. This allows the cell population to be assessed by flow cytometry for stability and toxicity of the expression construct and to establish optimal induction conditions. Double RBS vectors from the pET-Duet series have been used to produce nine multi-subunit membrane proteins (see Table 2 for 4HZU, 4HQU, 4HG6, 4NRE, 4N4R, 5AWW, 4YMS, 3DL8 and Table 1 for 4C48), with five being produced in E. coli host strain C43(DE3).

There is no general rule regarding cleavage sequences, but TEV protease, which is easy to produce in-house, is widely used for membrane protein purification (see 4C00, 3WVF, 4X5M and 4JA3 for examples) because it is still active in the presence of the most commonly-used detergents [11]. Thrombin protease is also widely used (see 2VQI, 2ABM and 3B5D for examples).

2.3. Promoter usage for E. coli expression

We analyzed how the 468 membrane proteins in Table 5 had been produced. Some uMPS were produced in more than one expression system and therefore Tables 7 and 8 list a total of 477 combinations of promoter and E. coli host strain. As we previously observed (in our 2015 analysis of 213 uMPS [8]), the T7 RNA polymerase (T7RNAP)-based expression system is the most widely-used followed by the ara, T5 and tet promoter-based expression systems (Fig. 5). The data in Tables 1 and 2 are presented in chronological order, meaning that the later entries reflect the most recent trends in promoter, strain and vector choice.

For the production of non-E. coli (heterologous) membrane proteins in an E. coli host, the T7RNAP-based expression system is predominant and is used together with six bacterial strains, BL21(DE3), C43(DE3), C41(DE3), BL21(DE3) pLyS5, BL21(DE3) CodonPlus or
| PDB   | Organism       | Membrane Protein Name                  | TM | TMV | Strain(s)    | Vector | Tag    | Size | N/C  | Solubilization | Structure | Date       |
|-------|----------------|----------------------------------------|----|-----|--------------|--------|--------|------|------|----------------|-----------|------------|
| 1UUM  | R. rattus      | Flavin dihydroorotate dehydrogenase (DHOD) | 0  | 372 | XL1-Blue     | pASKDr | His    | 6    | N    | None           | Xray-OG   | 2004       |
| 2FNQ, 3FG4 | P. homomalla   | Lipoxigenases (LOXs)                  | 0  | 699 | BL21(DE3)    | NS    | His    | 6    | N    | None           | None      | 2006       |
| 2OQO  | A. aeolicus    | Peptidoglycan glycosyltransferase (PGT) | 0  | 200 | BL21(DE3)    | pET48b | His    | 6    | N    | CHAPS          | CHAPS     | 2007       |
| 2PRM  | H. sapiens     | Dihydroorotate dehydrogenase (DHODH)  | 0  | 367 | BL21(DE3)    | pET19b | His    | 10   | N    | TX100          | Xray-MD   | 2008       |
| 2K5U, 2KSQ | S. cerevisiae  | Adip-ribosylation factors (ARFs)       | 0  | 181 | BL21(DE3)    | pET20b | His    | 6    | C    | None           | InNMR-DMPC/DMPC or Xray-LDAO | 2009       |
| 3L65  | P. falciparum  | Dihydroorotate dehydrogenase (PHDODH) | 0  | 415 | BL21(DE3)    | pET28b | His    | 6    | N    | C12E9          | Xray-LDAO | 2009       |
| 3L7I  | S. epidermidis | Polymerase (TagF)                     | 0  | 729 | BL21(DE3)    | pET28b | His    | 6    | C    | CHAPS          | Xray-CHAPS| 2010       |
| 3OQO  | H. sapiens     | Enzyme 5-lipoxygenase (5LOX)           | 0  | 691 | Rosetta™ (DE3) | pET14b | His    | 6    | N    | None           | None      | 2011       |
| 2OQO  | A. aeolicus    | Glycosyltransferase (WaaA, KdA)       | 0  | 374 | BL21(DE3) CodonPlus | pUM212/216 | His    | 10   | N    | TX100          | Xray-Cymal6| 2009       |
| 1J6W  | B. subtilis    | Signal peptide peptidase A (SppA)     | 0  | 240 | BL21(DE3) Tuner | pET28b | His    | 6    | N    | None           | Xray-DDM  | 2012       |
| 4G9K  | A. aeolicus    | Kinase (LpxK)                          | 0  | 317 | C41(DE3)     | pET21b | GST    | N    | DDN | None           | Xray-DDM  | 2012       |
| 4G9K  | S. cerevisiae  | NADH dehydrogenase (Ndi1)             | 0  | 471 | BL21(DE3) pLexS | pET16b | His    | 10   | N    | DDN           | Xray-DDM or Xray-DDM | 2012       |
| 4GGM  | C. crescentus | Phosphodiester hydrolase (Lpxd)       | 0  | 283 | C41(DE3)     | T7 based | His    | 10   | N    | None           | None      | 2012       |
| 4GG6  | S. cerevisiae  | Dihydrogenase (Ndi1)                  | 0  | 502 | C43(DE3)     | pQ680L | His    | 6    | N    | TX100          | Xray-TX100| 2012       |
| 4HHS  | A. thaliana    | Fatty acid α-oxoxygenase (α-DOX)      | 0  | 652 | M15          | pQ630  | His    | 6    | N    | None           | Xray-NG   | 2013       |
| 3V9V  | T. brucei      | Alternative oxidase (AOX)              | 0  | 329 | BL21(DE3)    | pET15b | His    | 6    | N    | OG             | Xray-MD   | 2013       |
| 2YOC  | K. oxytoca     | Lipoprotein pullulanase (PulA)         | 0  | 1078| pAP5198      | pCHA14486 | His    | 6    | C    | NS            | NS        | 2013       |
| 4NRE  | H. sapiens     | Enzyme 15-lipoxygenase-2 (15LOX-2)    | 0  | 696 | Rosetta™ (DE3) | pETDuet-1 | His    | 6    | C    | None           | None      | 2014       |
| 4NM9  | G. sulfurreducens | Proline utilization A (PutA)          | 0  | 1005| BL21Al       | pNIC28-Ins4 | His    | 6    | N    | None           | None      | 2014       |
| 4NWZ  | C. thermarum  | Non-proton pumping type II NADH dehydrogenase (NDH-2) | 0  | 405 | C41(DE3)     | pTRG68b2 | His    | 6    | C    | OG             | Xray-OG   | 2014       |
| 4PLA  | H. sapiens     | Phthiacyclinins 4-kinase type Ila (P4K Ila) | 0  | 556 | BL21(DE3) Star | pRSFD  | His    | 6    | N    | None           | None      | 2014       |
| 4QN9  | H. sapiens     | Fatty-acid ethanolamides (FAEs)       | 0  | 393 | Rosetta-origamiB (DE3) | pLexS | His    | 6    | C    | TX100          | Xray-DC   | 2015       |
| 4WGH  | S. aureus      | Bacterial type I signal peptidases    | 0  | 542 | BL21(DE3) CodonPlus | pProExHta | His    | 6    | C    | None           | None      | 2015       |
| 5B49  | P. aeruginosa  | UDP-diacylglycerolamine                 | 0  | 248 | BL34(DE3) pLexS | pET26b | His    | 6    | C    | TX100          | None      | 2016       |
| 5K7N  | A. baumannii   | Lipid A acyltransferase LpxM           | 0  | 333 | C41(DE3)     | pRham  | His    | 6    | N    | DDN           | Xray-LMNG | 2016       |
| Alpha-helical | S. lividans | Channel (KcsA)                         | 2 × 4 | 160 | XL1-Blue     | pQE32  | His    | 6    | N    | DM             | EPR-DDM; Xray-DDM or Xray-LDAO | 1998       |

(continued on next page)
| PDB  | Organism       | Name                     | TM | Length | Strain(s)               | Vector | Tag   | N/C | Solubilization   | Structure   | Date     |
|------|----------------|--------------------------|----|--------|-------------------------|--------|-------|-----|-----------------|-------------|----------|
| 1KPL | S. typhimurium  | Transporter (H⁺/Cl⁻⁻)    | 18 | 473    | BL21(DE3)               | pET28b | His   | 6   | C               | DM          | Xray-OM   | 2002 |
| 3RB2 | M. thermotrophicus | Potassium channel (MhK) | 2  | 340    | XL1-Blue, SG1309        | pQE70  | His   | 6   | C               | DM          | Xray-LDAO | 2002 |
| 1ORQ | A. pernix      | Channel (KvAP)           | 6  | 223    | XL1-Blue                | pQE60  | His   | 6   | C               | DM          | Xray-OM   | 2003 |
| 1XFH | P. horikoshii  | Glutamate transporter homol (GlpPh) | 8  | 422    | TOP10 or DH10B          | pBAD   | His   | 8   | N               | DM          | Xray-DM   | 2004 |
| 1WAZ | M. morganii    | Transporter (MerF)       | 2  | 46     | C43(DE3)                | pET31b | His   | 6   | C               | IB-gua + SDS | InNMR-SDS | 2005 |
| 1ZLL | H. sapiens     | Phospholamban homopentamer | 1  | 52     | BL21(DE3)               | pMALc2x | MBP   | N   | NS              | InNMR-PC12 | saNMR-DMPC | 2005 |
| 2B2F | A. fulgidus    | Transporter (Amt-1)      | 11 | 399    | C43(DE3)                | pET21a | His   | 6   | C               | DM          | Xray-DM   | 2005 |
| 2P2B | M. marburgensis | Aqp (AugM)              | 6  | 246    | BL21(DE3) CodonPlus     | NS³    | His   | 10  | N               | OG          | Xray-OM   | 2005 |
| 2BJJ | T. maritima    | Mg²⁺ transporter (CorA)  | 2  | 351    | NS³                    | pET15b | His   | 6   | N               | DM          | Xray-DM   | 2005 |
| 2AH9 | S. lividans    | Channel (KcsA)           | 2  | 155    | C41(DE3)               | pIVEX-2.4d | His   | 6   | N               | FC12        | InNMR-PC12 | 2006 |
| 3E66 | B. cereus      | Channel (NaK)            | 2  | 96-114 | SG1309                 | pQ650  | His   | 6   | C               | DM          | Xray-DM   | 2006 |
| 3B9W | N. europaea    | Rh protein ammonia or CO₂ channel | 11 | 407    | GT1000 D (glnK, amtB)   | pAD7; nitrogen promoter | His   | 6   | C               | OG          | Xray-OM   | 2007 |
| 3BEH | M. loti        | Flippase (MshA)          | 6  | 582    | BL21(DE3)              | pET19b | His   | 10  | N               | UDM         | Xray-UDM  | 2007 |
| 2LJB | M. morganii    | Cyclic nucleotide-regulated K⁺ channel | 6  | 355    | JM83                    | pASK-IBA2 | His   | 6   | C               | DM          | Xray-DM   | 2008 |
| 2LNF | E. coli        | Channel (M2)             | 1  | 35     | BL21(DE3)              | pMM-LR6 | His   | 9   | N               | IB-gua + TX100 | InNMR-SDS | 2006 |
| 2VQF | P. horikoshii  | Aspartate transporter (GlpPh) | 6  | 360    | pBAD24                 | His   | 8    | NS³ | DDM             |             | Xray-DM   | 2007 |
| 2Q7M | H. sapiens     | Lipoygenase protein (FLAP) | 4  | 161    | BL21(DE3)              | pET28a | His   | 6   | C               | DM          | Xray-OM   | 2007 |
| 2QXU | E. coli        | Leucine transporter (LexT) | 12 | 511    | BL21(DE3) pLYsS        | pBAD   | His   | 6   | C               | NS³         | DM        | 2007 |
| 2VXY | T. thermophilus | Transporter (MgtE)       | 5  | 473    | C41(DE3)               | pET28a | His   | 6   | N               | DM          | Xray-DM   | 2007 |
| 3BW9 | N. europaea    | Rh protein ammonia or CO₂ channel | 11 | 407    | GT1000 D (glnK, amtB)   | pAD7; nitrogen promoter | His   | 6   | C               | OG          | Xray-OM   | 2007 |
| 3BEH | M. loti        | Flippase (MshA)          | 6  | 582    | BL21(DE3)              | pET19b | His   | 10  | N               | UDM         | Xray-UDM  | 2007 |
| 2LJB | M. morganii    | Cyclic nucleotide-regulated K⁺ channel | 6  | 355    | JM83                    | pASK-IBA2 | His   | 6   | C               | DM          | Xray-DM   | 2008 |
| 2LNF | E. coli        | Channel (M2)             | 1  | 35     | BL21(DE3)              | pMM-LR6 | His   | 9   | N               | IB-gua + TX100 | InNMR-SDS | 2006 |
| 2VQF | P. horikoshii  | Aspartate transporter (GlpPh) | 6  | 360    | pBAD24                 | His   | 8    | NS³ | DDM             |             | Xray-DM   | 2007 |
| 2Q7M | H. sapiens     | Lipoygenase protein (FLAP) | 4  | 161    | BL21(DE3)              | pET28a | His   | 6   | C               | DM          | Xray-OM   | 2007 |
| 2QXU | E. coli        | Leucine transporter (LexT) | 12 | 511    | BL21(DE3) pLYsS        | pBAD   | His   | 6   | C               | NS³         | DM        | 2007 |
| 2LJB | M. morganii    | Cyclic nucleotide-regulated K⁺ channel | 6  | 355    | JM83                    | pASK-IBA2 | His   | 6   | C               | DM          | Xray-DM   | 2008 |
| 2LNF | E. coli        | Channel (M2)             | 1  | 35     | BL21(DE3)              | pMM-LR6 | His   | 9   | N               | IB-gua + TX100 | InNMR-SDS | 2006 |
| 2VQF | P. horikoshii  | Aspartate transporter (GlpPh) | 6  | 360    | pBAD24                 | His   | 8    | NS³ | DDM             |             | Xray-DM   | 2007 |
| 2Q7M | H. sapiens     | Lipoygenase protein (FLAP) | 4  | 161    | BL21(DE3)              | pET28a | His   | 6   | C               | DM          | Xray-OM   | 2007 |
| 2QXU | E. coli        | Leucine transporter (LexT) | 12 | 511    | BL21(DE3) pLYsS        | pBAD   | His   | 6   | C               | NS³         | DM        | 2007 |

(continued on next page)
| PDB | Organism          | Name                                                                 | TM1 | Length | Strain(s) | Vector          | Tag    | Size | N/C | Solubilization | Structure | Date       |
|-----|------------------|----------------------------------------------------------------------|-----|--------|-----------|----------------|--------|------|-----|----------------|-----------|------------|
| 3DIN| T. maritima      | SecYEG protein in complex with SecA                                | 10  | Y      | 431 + 65 + 76 | BL21(DE3)     | pBAD22, pACYC | No tag | DDM | Xray-Cymal6     | 2008      |            |
| 2ZJS| T. thermophilus  | Translocon (SecYE)                                                  | 10  | 1      | 438 + 60     | AD202         | pTV118N lac   | His    | 6   | C               | DDM       | Xray-DDM   | 2008      |
| 2JLN, 2X79| M. liquefaciens | Benzyl-hydantoin transporter (Mhp1)                                | 12  | 501    |            | BLR            | pTTQ18 | His | 6   | C               | DDM       | Xray-NM    | 2008      |
| 3DL8| A. aeolicus      | Channel (SecYEG)                                                    | 10  | Y      | 429 + 65 + 107 | C43(DE3)     | pET Duet-1, pCDF Duet-1 | His    | 6   | C               | Xray-Cymal6 | 2008      |
| 3FSA, 2QEI, 2A6S, 3GJD, 3MPN, 3TUG, 3QS4, 4MM4| A. aeolicus VF5 | Symporter (LeuT)                                                   | 12  | 2      | 519        | C41(DE3)     | pET16b | His | 8   | C               | Xray-OG   | 2008      |

(continued on next page)
### Table 2 (continued)

| PDB     | Organism                     | Name                          | TM<sup>1</sup> | Length | Strain(s)       | Vector       | Tag      | Size | N/C<sup>2</sup> | Solubilization<sup>3</sup> | Structure<sup>4</sup> | Date         |
|---------|------------------------------|-------------------------------|-----------------|--------|-----------------|--------------|----------|------|----------------|----------------------------|------------------------|---------------|
| 2QKA, 3EHZ, 4HFI |                               |                               |                 |        |                 |              |          |      |                |                            |                        |               |
| 2XQ2    | V. parahaemolyticus          | Na<sup>+</sup>/galactose transporter (ySGLT) | 14              | 593    | TOP10           | pBAD        | His      | 6    | C              | DM                         | Xray-Trp-DM          | 2010          |
| 2XUT    | S. oneidensis                | Transporter (PepTSt)          | 14              | 524    |                 | C43(DE3)     | pWaldo-GPFP | GFP  | + His         | 8 C                         | DM                     | Xray-DMM      | 2010          |
| 4A2N, 3PW, 3EAM, 3IQG, 4QHS, 5HCJ, 5LAE |                               |                               |                 |        |                 |              |          |      |                |                            |                        |               |
| 3P50    | G. violaceus                 | Ligand-gated ion channel (GLIC) | 4 × 5           | 359    |                 | C43(DE3)     | pET20b    | His  | + MBP         | 6 N                         | DM                     | Xray-DMM      | 2011          |
| 3P5Z    | V. parahaemolyticus          | Transporter (TrkH)            | 10 × 2          | 485    | BL21(DE3)       | pET31        | His      | 10   | C              | DM                         | Xray-DMM              | 2011          |
| 3N00    | S. oneidensis                | H<sup>+</sup>/Cl<sup>-</sup> exchange transporter | 18              | 466    |                 | 420399      | pASk     | His  | 6 C           | DM                         | Xray-DMM              | 2011          |
| 3O0J    | H. influenzae                | Peptidase (GlpG)              | 6               | 196    | TOP10           | pBAD        | His      | 6    | C              | DM                         | Xray-C12E8            | 2011          |
| 3QNQ    | B. cereus                    | Transporter (Chbc E1C)        | 10 × 2          | 433    | BL21(DE3)       | pET        | His      | 10   | C              | DM                         | Xray-DMM              | 2011          |
| 2KYV    | H. sapiens                   | Phospholamban homopentamer    | 1 × 5           | 52     | BL21(DE3)       | pET        | MBP      |      | N             | TX100                       | InNMR-FC12             | 2011          |
| 3AQP    | T. thermophilus              | (SecDF)                       | 12              | 735    | BL21(DE3)       | CodonPlus   | pET26b   | His  | 6 C           | DM                         | Xray-DMM              | 2011          |
| 2L9U    | H. sapiens                   | Transmembrane domain (ErbB3)  | 1 × 2           | 40     |                 | pET22b      | His      | 6    | C              | Cell-free expression-pellet | InNMR-FC12             | 2011          |
| 3RFU    | L. pneumophila               | Copper efflux ATPase          | 8               | 736    | C43(DE3)        | pET22b      | His      | 6    | C              | C12E8                       | Xray-MD<sup>5</sup>   | 2011          |
| 3RCE, 3RFU | Campylobacter lari         | OST in complex (PepI3)        | 13              | 724    | BL21(Gold SCM6) | pBAD       | His      | 10   | C              | C12E8                       | Xray-DMM              | 2011          |
| 4BJU, 3RFU | L. pneumophila              | Copper-transporting ATPase (LpCopa) | 8               | 736    | C43(DE3)        | pET22b      | His      | 6 NS | NS NS         | C12E8                       | Xray-DMM              | 2011          |
| 3SIX    | M. martalphilus              | Preflagellin aspartyl protease (FlaK) | 6               | 237    | C43(DE3)        | pET28a/43b | His      | 6    | C              | FC12                        | Xray-Cymal6          | 2011          |
| 2LCX    | M. musculus                  | Mitochondrial uncoupling protein 2 (UCP2) | 6               | 303    |                      | Rosetta™ (2 DE3) | pET21   | His  | 6 C           | FC12                        | FC12                  | 2011          |
| 3QDC, 2KSY, 1H68, 1H2S | N. pharaonis                 | Rhodopsin (SRII)              | 7               | 239    | BL21(DE3)       | and pET27b/28b | His      | 7    | C              | OG or DDM                   | Xray-OG or DDM        | 2011          |
| 3ZUY    | N. meningitidis              | Bacterial ASBT homologues     | 10              | 323    | C43(DE3)        | pWaldo-GPFP  | GFP  + His | 8    | C              | DM                         | Xray-LDAO              | 2011          |
| 32HS, 2WLJ, 4LP8 | M. magneto-tacticum         | Channel (Kirbαc3.1)           | 2 × 4           | 339    | BL21(DE3)       | CodonPlus   | pET30a   | His  | 6    | C              | Xray                    | 2011          |
| 4A2N    | M. acidovorans               | Transferase (ICMT)            | 5               | 194    | C41(DE3)        | pTriEX/pOPIN | GFP  + His | 7    | C              | DDM                        | Xray-DMM              | 2011          |
| 3AYF    | G. stea- thermophilus        | Nitric oxide reductase (qNOR) | 14              | 800    | Rosetta2 (DE3)  | pET22b      | His      | 6    | C              | TX100                       | Xray-O2               | 2011          |
| 3SVU    | M. janauchii                 | Exchanger (Na<sup>+</sup>/Ca<sup>+</sup>) | 9               | 320    | BL21(DE3) pLyS  | pQ660      | His      | 10   | C              | DDM                        | Xray-DMM              | 2011          |
| 3T1J    | V. cholerae                  | Nucp family protein           | 8 × 3           | 424    | C41(DE3)        | pET26       | MBP + His | 10   | N              | DDM                        | Xray-DMM              | 2011          |
| 3RQW    | E. chrysantheri              | Ligand-gated ion channel with acetylcholine (ELIC) | 4 × 5           | 322    | Rosetta™ (2 DE3) pLyS | pET26b   | His  | 10 N          | UDM                        | Xray-DMM              | 2011          |
| 3TDO    | C. difficile                 | Hydrosulfide ion channel (FNT3) | 6 × 5           | 268    | BL21(DE3) pLyS  | pBAD       | His      | 10   | C              | DDM                        | Xray-OG               | 2011          |
| 3QF4    | T. maritima                  | Heterodimeric ABC exporter    | 6 × 2           | 587    | C43(DE3) or MC1061 | pBAD24  | His      | 10   | N              | DDM                        | X-ray-DMM             | 2012          |
| 3VMT    | S. aureus                    | Glycosyltransferase           | 1               | 263    | BL21(DE3) CodonPlus | pET15b   | His      | 6    | N              | DDM                        | Xray-DMM              | 2012          |
| 3VOU    | B. weihen- stephanensis (NaK) and S. pantoea (NaV) | NaK chimera with NaV | 2 × 4       | 148    | KRX             | pET21b     | His      | 4 × 6 | Oth<sup>6</sup> | DM                         | Xray-DMM              | 2012          |
| 2LCX    | H. sapiens                   | ErbB4                         | 1 × 2           | 44     | BL21(DE3) pLyS  | pGEMEXI    | TrxA-His | 6    | N              | TX100                       |                        | 2012          |

(continued on next page)
| PDB  | Organism                          | Name                           | TM | Length | Strain(s) | Vector       | Tag    | Size | N/C | Solubilization | Structure | Date  |
|------|-----------------------------------|--------------------------------|-----|--------|-----------|--------------|--------|------|-----|----------------|-----------|-------|
| 4DXW | *A. proteobacterium himb114*     | Na⁺ channel (NaV)              | 6   | 4      | 228       | BL21(DE3)    | pET21b | His  | 6   | C               | DDM       | Xray-NG| 2012 |
| 4APS | *S. thermophilus*                | Transporter (PepTSo)           | 12 + 2 | 491   | C43(DE3)  | pWaldo-GFPe  | GFP + His | 8   | C   | DDM                | Xray-DDM  | 2012  |
| 4FLH | *G. violaceus*                   | Ligand-gated ion channel with  | 4 × 5 | 359   | Rosetta²(2(DE3) pLysS | pET26b | MRP + His | 10  | N   | DDM                | Xray-TDM  | 2012  |
| 4F4L | *M. marinus*                     | Voltage-gated sodium channel   | 2 × 4 | 112   | C41(DE3)  | pET15b       | His     | 6   | N   | DDM                | Xray-MD² | 2012  |
| 2NL  | *H. sapiens*                     | Receptor (CXCR1)               | 7   | 309   | BL21(DE3) | pGEK2a       | GST + His | 6   | C   | IB-SDS             | isoNMR-DMPC/DMPC | 2012  |
| 4FS5 | *V. cholerae*                    | Symporter                      | 11 × 2 | 462   | BL21(DE3) | pET        | His     | 10  | N   | DDM                | Xray-MD² | 2012  |
| 4EV6 | *M. jannaschii*                  | Transporter (CorA Mg²⁺⁺)       | 2 × 5 | 317   | C43(DE3) | pNIC28(pET28) | His     | 10  | N   | UDM                | Xray-UDM | 2012  |
| 4H33 | *L. monocytogenes*               | Channel (GluK)                 | 2 × 4 | 137   | BL21(DE3) | pQE70       | His     | 6   | C   | DDM                | Xray-MD² | 2012  |
| 4EEB | *T. maritima*                    | Mg²⁺⁺ transportor (CorA)       | 2 × 5 | 351   | BL21(DE3) | pET15b       | His     | 6   | N   | DDM                | Xray-MD² | 2012  |
| 4BA4 | *A. aeroliktus*                  | Twin-arginine translocase (TatC) | 6   | 240   | Lemo56(DE3) | pWaldo-GFPe | GFP + His | 8   | C   | NMR-FC12           | Trilutriarginic acid | 2013  |
| 3UX4 | *H. pylori*                      | Channel (urea)                 | 6   | 195   | C43(DE3) | pET       | His     | 6   | N   | Oth²               | DM        | Xray-MD² | 2012 |
| 4HYG | *M. marinigri*                   | Presenilin (PSH)               | 9 × 4 | 301   | BL21(DE3) | pET21b       | His     | 8   | N   | NM                 | Xray-MD² | 2012  |
| 4HGR | *S. pharaonis*                   | Cellulose synthase (BcsA-BcsB) | 8 + 1 | 802 + 707 | Rosetta²(2(DE3) pETDuet | pETDuet | His     | 12  | N   | TX100               | Xray-LDAO | 2012  |
| 4G1U | *Y. pestis*                      | Transporter (HmuV)             | 10 × 2 | 334   | BL21(DE3) | pET19b       | His     | 10  | N   | DDM                | Xray-MD² | 2012  |
| 4G2X | *H. sapiens*                     | Apelin receptor                | 1   | 64    | BL21(DE3) | pEXP5-CT     | His     | 6   | C   | IB-acetonitrile    | isoNMR-FC12 | 2013  |
| 3ZKR | *E. chrysanthemi*                | Ligand-gated channel (ELIC)    | 4 × 5 | 307   | C43(DE3) | pET11a       | MBP     | N   | UDM   | Xray-UDM           | (LCP)     | 2013  |
| 4HIU | *N. gonorrhoeae*                 | MATE transporter (NorM)        | 12 | 459   | BL21(DE3) | pET15b       | His     | 6   | N   | DDM                | Xray-MD² | 2012  |
| 3VYN | *P. furiosus*                    | Multidrug and toxin compound   | 12 | 461   | C43(DE3) | pET11a       | His     | 6   | C   | DDM                | Xray-Cymal6 | 2013  |
| 4J9L | *P. pastoris*                    | Channel (KtrAB)                | 8 × 2 | 485   | BL21(DE3) | pET24d       | His     | NS² | N   | DDM                | Xray-Cymal6 | 2013  |
| 4HTS | *A. aeroliktus*                  | Twin-arginine translocase (TatC) | 6   | 236   | BL21(DE3) | pET33b       | His     | 6   | N   | DDM                | Xray-DHPC or DDM | 2013  |
| 4JQ6 | *Isoalted from the Mediterranean* | Proteorhodopsin (blue-light   | 7 × 6 | 235   | C43(DE3) | pET28       | His     | NS² | NS² | DM                | Xray-DM (bireucles) | 2013  |
| 4KLY | *G. macropyro-                      | Proteorhodopsin (PRs),        | 7 × 6 | 235   | C43(DE3) | pET28a      | His     | 6   | C   | DM                | Xray-DM   | 2013  |
| 2M6X | *H. pylori*                      | p7 hexamer channels           | 2 × 6 | 63    | BL21(DE3) | pMM- LR6     | His-trpLE | 9   | N   | IB-gua + TX100     | isoNMR-DMPC/DMPC | 2013  |
| 4KPP | *A. fulgidus*                    | Ca²⁺⁺/H⁺ antiporter (CaX)     | 12 | 405   | C41(DE3) | pET        | GFP + His | 8   | C   | DDM                | Xray-LDAO | 2013  |
| 4IKS | *S. violaceus*                   | Ca²⁺⁺/H⁺ antiporter (Yke)     | 11 × 3 | 351   | BL21(DE3) | pET22b       | His     | 6   | Oth² | DDM                | Xray-DM² | 2013  |
| 3W9J | *P. aeruginosa*                  | Multi-drug efflux transporter  | 12 × 3 | 1052  | MG1655   | pUCP20-BHis | His     | 10  | C   | DDM                | Xray-DM² | 2013  |
| 4KVN | *G. kaustophilus*                | Proton-dependent oligopeptide  | 12 × 5 | 507   | C41(DE3) | ΔacrB       | pCGFP-BC | GFP + His | 8   | C   | DDM                | Xray-DM² | 2013  |
| 4HYJ | *E. sibericum*                   | Proteorhodopsin                | 7   | 258   | Rosetta²(2(DE3) pET26a | pET32a | His     | 6   | C   | DDM                | Xray-DM² | 2013  |
| 2M3G | *Anabaena sp.*                   | Sensory rhodopsin              | 7× 3 | 235   | BL21(DE3) | pET21b       | His     | 6   | C   | DDM                | Xray-NM   | 2013  |

(continued on next page)
| PDBe | Organism | Name | TM | Length | Strain(s) | Vector | Tag | Size | N/C | Solubilization | Structure | Date |
|------|----------|------|-----|--------|-----------|--------|-----|------|-----|---------------|-----------|------|
| 2M6B | H. sapiens | Glycine receptor (hGlyR-a1) | 4 × 5 | 150 | BL21 (DE3) pLysS | pET31b | His | 6 | C | IB-NS | inNMR-PPG | 2013 |
| 47J2, 5CKR | A. aerculus | MutNac pentapeptide translocase (MarX) | 10 | 565 | C41 (DE3) | NS | MBP + His | 10 | N | DDM | Xray-DM | 2013 |
| 4K10 | T. kodakarenosis | Aspartate transporter | 8 | 431 | MC1061 | pBAD24 | His | 8 | C | DM | Xray-OG | 2013 |
| 3ZLZ, 4RCB, 5ZWX | M. marinus | Sodium channels | 4 | 149 | C41 (DE3) | pET15b | His | NS | NS | DDM | Xray-HEGAI0 | 2013 |
| 2L2L | H. sapiens | Growth factor receptor 3 (FGFR3) | 1 × 2 | 43 | BL21 (DE3) pLysS | pGEMEX1 | TrxA-His | 6 | N | TX100 | inNMR-FC12 | 2013 |
| 4M1J | P. mirabilis | Carminine transporter (CaiT) | 12 × 3 | 504 | BL21 (DE3) pLysS | pET15b | FLAG-His | 6 | C | Cymal5 | Xray-Cymal5 | 2013 |
| 4LDS | S. epidermidis | Glucose/H+ symporter (GlcP) | 12 | 446 | C41 (DE3) | pET15b | His | NS N | N | DDM | Xray-DM | 2013 |
| 4K9J | C. metallidurans | ZnA Zn(II)/proton antiporter | 12 | 1045 | C41 (DE3) | pET30b | His | 6 | C | DDM | Xray-DM | 2013 |
| 4LTO | A. ehrlichii | Voltage-gated sodium channel (NaVaC1p) | 2 × 4 | 152 | C41 (DE3) | pET24 | MBP + His | 6 | N | DDM | Xray-DM | 2013 |
| 4L2S, 5G0N | B. halodurans | MATE transporter (DinB-FH) | 12 | 446 | BL21 (DE3) | pET15b | His | 6 | C | DDM | Xray-DM | 2013 |
| 3WAJ | A. fulgidus | OST (AgIB) | 13 | 875 | C43 (DE3) | pET52b | His | 10 | C | DDM | Xray-MD | 2013 |
| 4C7R | C. glutamicum | Glycine betaine transporter (BetP) (BetP) | 12 × 3 | 566 | DH5α | pBA7 | Strep | N | DDM | Xray-Cymal5 | 2013 |
| 2MAW | H. sapiens | Neuronal acetylcholine receptor | 4 | 137 | Rosetta® 2 (DE3) pLysS | pET28a | His | 6 | N | NS | inNMR-LDAO, Xray-NEG or C7G | 2013 |
| 2M8R, 3HID | R. norvegicus | Synaptotagmin 1A, TM & syntaxin complex | 1 + 1 | 109 + 91 | Rosetta® 2 (DE3) pLysS | pET28a | His | 6 | N | TX100 or OG | inNMR-FC12 | 2013 |
| 4CAD | M. maripulorum | Protease Rec1 (CAAX) | 8 | 271 | C41 (DE3) | pTrIX | His | 7 | C | UDM | Xray-MD | 2013 |
| 4N7W | Y. frederiksenii | Bile acid symporter | 12 | 307 | BL21 (DE3) Gold | pET | His | 10 | N | DDM | Xray-MD | 2013 |
| 4M64 | S. typhimurium | AcrB/NS-membrane symporter (MelB) | 12 | 476 | DW2 | pK95 native | His | 10 | C | UDM | Xray-MD | 2014 |
| 4BUO | R. norvegicus | Neurotensin receptor (NTS1) | 7 | 335 | BL21 Tuner | pBR322 | MBP + His | 6 | N | MD | Xray-MD | 2014 |
| 4NOY | A. thaliana | Cytochrome B561 (Cyto bfl1) | 6 | 230 | BL21 | pET15b | His | 6 | N | DM | Xray-NG | 2014 |
| 4P9Y | P. luminescens | Tc toxin (TcA) | 2 × 5 | 2516 | BL21 (DE3) CodonPlus | pET28a | His | 6 | N | None | None | 2014 |
| 2MS9 | H. sapiens | Endothelial growth factor receptor 2 | 1 × 2 | 37 | Cell-free expression | pET20b | HA | N | Cell-free expression pellet-trifluoroethanol//H2O/trifluoroacetic acid | inNMR-FC12 | 2014 |
| 4M8M | N. aromaticivorans | ABC transporter (NalA1m1) | 6 × 2 | 614 | BL21 (DE3) | pJL-H6/pET21a | His | 6 | C | MD | Xray-MD | 2014 |
| 4MYC | S. cerevisiae | ABC transporter (Atm1) | 6 × 2 | 598 | NS | pASK-IBA1 | Strep | C | DDM | Xray-DMD | 2014 |
| 2MFR | H. sapiens | Insulin receptor (AAS 940–980) | 1 | 57 | BL21 (DE3) | pET29b | His | 6 | C | IB-urea | inNMR-FC12 | 2014 |
| 2MGG | M. musculus | Translocator protein (TSPO) | 5 | 169 | BL21 (DE3) | pET15b | His | 6 | N | IB-BS | inNMR-FC12 | 2014 |
| 3WO6 | B. halodurans | Insertase (YidC) | 5 | 267 | C41 (DE3) | pET | His or GFP + His | 8 | N/C | DDM | Xray-DMD | 2014 |
| 406M | A. fulgidus | Alcohol phosphotransferase (AF2299) | 6 | 372 | BL21 (DE3) pLysS | pMCSG7 | His | 10 | N | DM | Xray-DLDA or C5O85 | 2014 |
| 4PGR | B. subtilis | pH-sensitive channel (YetU) | 7 | 214 | BL21 (DE3) pLysS | pET or pMCSG | His | 10 | N | DDM | Xray-DLDA | 2014 |
| 4XZS, 4P19 | P. horikoshii | Aspartate transporter (GluPh) | 8 | 422 | DH10B | pBAD24 | His | 8 | C | DDM | Xray-DLDA | 2014 |
| 4MND | A. fulgidus | CDP-alcohol phosphotriphosphatidytransferase | 6 | 479 | C43 (DE3) | pET25b | His | 10 | C | TX100 | Xray-DLDA | 2014 |
| 4TPH | S. oneidensis | Oligopeptide-proton symporter (PepTSo) | 14 | 516 | C41 (DE3) | pNIC-CTHF(pET) | His | 6 | C | DDM | Xray-DLDA | 2014 |
| 4Q4H | T. maritima | ABC exporter | 6 × 6 | 587 + 598 | C43 (DE3) | pBAD24 | His | 10 | N | DDM | Xray-DLDA | 2014 |
| 4TQ3 | A. fulgidus | UbiA homolog | 9 | 303 | BL21 (DE3) | pET | His | NS | DDM | Xray-OX | Xray-OG | 2014 |
| 4MT1 | N. gonorrhoeae | Multidrug efflux pump (MtrD) | 12 | 1056 | C43 (DE3) ΔacrB | pET15bΔacrD | His | 6 | C | Cymal6 | Xray-Cymal6 | 2014 |
| 4QNC, 5UHQ | L. biflexa | SemiSWEET | 3 | 85 | BL21 (DE3) | pJexpress411 | His | 10 | C | DDM | Xray-DMD | 2014 |
| 4QND | Vibrio sp. 418 | | 3 × 2 | 97 | BL21 (DE3) | pJexpress411 | His | 10 | C | DDM | Xray-DMD | 2014 |

(continued on next page)
| PDB | Organism | Name | TM | Length | Strain(s) | Vector | Tag | Size | N/C | Solubilization | Structure | Date |
|-----|----------|------|-----|--------|-----------|--------|-----|------|-----|----------------|-----------|------|
| 4TWD | *D. chrysanthemi* | SemiSWEET in outward-open state | 4 × 5 | 307 | C43(DE3) | pET11a | MBP | N | UDM | Xray-UDM | 2014 |
| 4WD7 | *K. pneumoniae* | Bestrophin homolog (KpBest) | 4 × 5 | 301 | BL21(DE3) pLysS | pMC5G7 | His | 10 | N | DDM | Xray-DDM | 2014 |
| 4Q7N | *N. mucosa* | Vitamin B$_{12}$ transporter (PnuC) | 8 × 3 | 863 | MC1061/BL21(DE3) | pBAD/pET21d (SoMet) | His | 6 | C | DDM | Xray-DDM | 2014 |
| 4QMV | *M. alcaliphilum* | Δ$^{15}$-sterol reductase (MasR1) | 10 | 427 | C43(DE3) | pET21b | His | 8 | N | DDM | Xray-OG | 2014 |
| 4WGV | *S. capitis* | Transition-metal ion transporter (NRAMP) | 11 | 415 | BL21(DE3) pLysS pMCSG7 His 10 | N | DDM | Xray-DDM | 2014 |
| 4RNG | *T. yellowstonii* | SemiSWEET in occluded state | 3 × 2 | 88 | BL21(DE3) | pET | His | 6 | C | DDM | Xray-DDM | 2014 |
| 4QTN | *N. mucosa* | Vitamin B$_{3}$ transporter (PnuC) | 8 × 3 | 863 | MC1061/BL21(DE3) pBAD/pET21d (SeMet) | His 6 | C | DDM | Xray-NG | 2014 |
| 4QUV | *M. alcaliphilum* | Δ$^{14}$-sterol reductase (MaSR1) | 10 | 427 | C43(DE3) | pET21b | His | 8 | N | DDM | Xray-DDM | 2014 |
| 4W7V | *S. capitis* | Transition-metal ion transporter (NRAMP) | 11 | 415 | BL21(DE3) pLysS pMCSG7 His 10 | N | DDM | Xray-DDM | 2014 |
| 4RNG | *T. yellowstonii* | SemiSWEET in occluded state | 3 × 2 | 88 | BL21(DE3) | pET | His | 6 | C | DDM | Xray-DDM | 2014 |
| 4QTN | *N. mucosa* | Vitamin B$_{3}$ transporter (PnuC) | 8 × 3 | 863 | MC1061/BL21(DE3) pBAD/pET21d (SeMet) | His 6 | C | DDM | Xray-NG | 2014 |
| 4QUV | *M. alcaliphilum* | Δ$^{14}$-sterol reductase (MaSR1) | 10 | 427 | C43(DE3) | pET21b | His | 8 | N | DDM | Xray-DDM | 2014 |

(continued on next page)
| PDB | Organism                          | Name                                      | TM | Length | Strain(s) | Vector           | Tag   | Size | N/C | Solubilization                      | Structure | Date |
|-----|----------------------------------|-------------------------------------------|-----|--------|-----------|------------------|-------|------|-----|-------------------------------------|-----------|------|
| 5AWW| *T. thermophilus*                | Transloco (SecYEG)                        | 10 + 1 | 438 + 60 + 117 | BL21(DE3) | pACYC-Duet       | His   | 6    | C   | DDM                                | Xray-DDM  | 2015 |
| 5EH4| *H. sapiens*                     | Glycophorin A (GpA)                      | 1 + 2 | 30     | BL21(DE3) | pMM              | His-trpLE | 9    | N   | IB-gua + TX100                      | Xray-None  | 2015 |
| 5EKP| *Synechocystis sp.*              | Polyisoprenyl-glycosyltransferase (GtrB) | 2 × 4 | 318    | BL21(DE3) | pMSG7            | His   | 6    | N   | DM                                 | Xray-DM    | 2016 |
| 5E23| *T. thermophilus*                | Sodium/proton antiporters (NapA)         | 12   | 385    | Lemo21(DE3) | pWaldo         | GFP + His | 8    | C   | DDM                                | Xray-DM    | 2016 |
| 5AIS| *S. enterica*                    | Citrate symporter CitS                   | 11   | 448    | C41(DE3) | pET21d          | His   | 10   | N   | DM                                 | Xray-DM    | 2016 |
| 5E2M| *C. metallidurans*              | Glycosyltransferase (ArnT)               | 13   | 575    | BL21(DE3) | pNCOYOMPS       | His   | 10   | N   | DDM                                | Xray-DM    | 2016 |
| 5JGF| *T. maritima*                    | Mg²⁺ transporter (CorA)                  | 2 × 5 | 351    | BL21(DE3) | pET15b          | His   | 6    | N   | DDM                                | EM-DDM     | 2016 |
| 5N2A| *H. sapiens*                     | Receptor tyrosine kinases (HER or ErbB)  | 1 × 2 | 58     | C41(DE3) | pGEMEX-1        | NT²   |      |    | Cell-free expression                | InNMR-Fc12 | 2016 |
| 5DIR| *P. aeruginosa*                  | Lipoprotein signal peptidase II          | 4    | 169    | C41(DE3) | pET28a          | His   | 6    | N   | FC12                                | Xray-FC12  | 2016 |
| 5EUL| *G. thermo-dentriticans/*B. subtilis | Transloco (SecYE) with SecAYE             | 1 + 10 | 841 + 430 + 70 | BL21(DE3) | pWaldo         | GFP + Streep | N   |    | C   | DDM                                | Xray-DM    | 2016 |
| 5EC1| *S. lividans*                    | Potassium channel mutant (KcsA)          | 2 × 4 | 160    | BL21(DE3) | NS⁵           | His   | 6    | N   | FC12 or DM                          | Xray-DM    | 2016 |
| 5AZD| *T. thermophilus*                | Thermophilic rhodopsin (TR)              | 7    | 268    | BL21(DE3) | pET21c          | His   | 6    | C   | DDM                                | Xray-DDM   | 2016 |
| 2N7Q| *H. sapiens*                     | Human nicastin                           | 1    | 54     | C41(DE3) | pET29b          | His   | 6    | N   | IB-urea                             | InNMR-SDS  | 2016 |
| 5ID3| *C. elegans*                     | Mitochondrial calcium uniporner          | 2 × 5 | 159    | BL21(DE3) | pET21a          | His   | 6    | C   | FC14                                | InNMR-FC14 | 2016 |
| 5JJZ| *L. delbrueckii*                 | Folate ECF transporter (FolT)            | 6    | 265    | MC1061 | pET28           | His   | 10   | N   | DDM                                | Xray-DM    | 2016 |
| 2N4X| *A. fulgidus*                    | Electron transporter (Ced(A))            | 6    | 208    | C43(DE3)-SEN12 | pET28 | His   | 6    | C   | DDM                                | Xray-DM    | 2016 |
| 5I20| *S. novella*                     | Aromatic amino acid exporter (YddG)      | 10   | 287    | Rosetta²⇐2(DE3) | pET | His   | 8    | C   | DDM                                | Xray-DM    | 2016 |
| 5IJS| *A. aerolicus*                   | Leucine transferase (LeuT)               | 12 × 2 | 519    | C41(DE3) | pET16b          | His   | 6    | C   | DDM                                | Xray-OG    | 2016 |
| 5JYN| *S. lividans*                    | HIV-1 envelope spike (Env)               | 1 × 3 | 40     | BL21(DE3) | pMM-LR6         | His-trpLE | 9    | N   | IB-gua                             | NMR-DMPC/DDHPC | 2016 |
| 5G28| *N. marinus*                     | Chloride-pumping rhodopsin (CIR)         | 7    | 275    | BL21(DE3) | CodonPlus       | pET21b | His   | 6    | C   | DDM                                | Xray-DDM   | 2016 |
| 5RS7| *B. cereus*                      | ABC heme importer (BhuUV)               | 10 × 2 | 385    | C41(DE3) | pET19b          | His   | 8    | N   | DM or NG                            | Xray-NG    | 2016 |
| 5KTE| *D. radiodurans*                 | Transition-metal ion transporter (NRAMP) | 11   | 436    | C41(DE3) | pET21         | His   | 8    | N   | DDM                                | Xray-DM    | 2016 |
| 5T77| *T. africanus*                   | MOP flippase (MurJ)                      | 14   | 475    | C41(DE3) | pET26          | MBP + His and His | 10   | C   | DDM                                | Xray-DM    | 2016 |
| 5MKK| *T. thermophilus*                | ABC transporter (TmrAB)                  | 6 + 6 | 611 + 577 | BL21(DE3) | pET22b          | His   | 10   | C   | DDM                                | Xray-Cymal5 | 2017 |
| 5ITE| *H. walsbyi*                     | Bacteriorhodopsin                      | 7    | 268    | C43(DE3) | pET28b         | His   | 6    | C   | DDM or SMA                          | Xray-OG or SMA | 2017 |
| 5FQ2| *N. meningitidis*                | Lipid A transferase                     | 5    | 544    | BL21(DE3) | pLyxS          | pTrc99A | His   | 6    | C   | DDM                                | Xray-MD² | 2017 |
| 5SKT| *M. musculus*                    | high density lipoprotein (HDL)          | 2    | 73     | BL21(DE3) | pQe30           | His    | 8    | N   | Empigen                             | InNMR-LPPG | 2017 |
| 5LS2| *A. aerolicus*                   | ABC transporter (AaPrtD)                | 6 × 2 | 572    | C43(DE3) | pQe30          | NS²   | His   | 6    | C   | NS²                                | InNMR-MD   | 2017 |
| 5SXY| *P. aeruginosa*                  | ABC transporter (LptB2FG)               | 6 (G) + 247 + 362 + 355 | C43(DE3) | pQe30          | His    | 6    | C   | DDM                                | Xray-MD² | 2017 |
| 5V4S| *L. licerasiae*                   | Cyclic nucleotide-gated channel (LKB)   | 7    | 465    | C43(DE3) | pET | His    | 8    | N   | LMNG                               | EM-LMNG    | 2017 |
| 5UNI| *T. thermophilus*                | Nicotinamide nucleotide transhydrogenase (TH) | 3 + 9 | 94 + 261 | BL21(DE3) | pTV118N (lac) | His    | 8    | C   | DDM                                | Xray-DDM   | 2017 |
| 5XAM| *D. radiodurans*                 | Protein-export enhancer (SecDF)          | 10   | 740    | BL21(DE3) | pET156l2mpN | His    | 6    | C   | DDM                                | Xray-DDM   | 2017 |
| 5KHN| *B. multivorans*                 | Hopanoid transporter (HptN)             | 12   | 877    | BL21(DE3) | pET156l2mpN | His    | 6    | C   | DDM                                | Xray-DDM   | 2017 |

(continued on next page)
| PDB  | Organism   | Name                        | TM1 | Length | Strain(s) | Vector | Tag  | Size | N/C | Solubilization | Structure | Date     |
|------|------------|-----------------------------|-----|--------|-----------|--------|------|------|-----|----------------|-----------|----------|
| 5WUC | S. acidocaldarius | Trimeric intracellular cation (TRIC) | 7 × 3 | 237    | BL21(DE3) pLyS5 | pET   | His  | 10   | C   | DDM            | Xray-DM   | 2017     |
| 5VRE | C. minuta   | Lysosomal K⁺ channel         | 6 × 4 | 203    | BL21(DE3) | pET15b | His  | 6    | N   | DDM            | Xray-DMNG | 2017     |
| 5WUF | C. psych-rerythraea | Trimeric intracellular cation (TRIC) | 7 × 3 | 249    | BL21(DE3) pLyS5 | pET   | His  | 10   | C   | DDM            | Xray-DM   | 2017     |
| 5TOO | C. jejuni    | Multi-drug efflux transporter (CmeB) | 12  | 1040   | BL21(DE3) pET15b | His  | 6    | N   | DDM            | Xray-DM   | 2017     |
| 5SGO | A. baumannii | ABC transporter (MacB)       | 4x2 | 671    | BL21(DE3) | pET22 | His  | 6    | N   | DDM            | Xray-DM   | 2017     |
| 5LIL | A. actinomy-cetemcomitans | ABC transporter (MacB) | 4x2 | 664    | BL21(DE3) | pET28 | His  | 6    | N   | LMNG           | Xray-LMNG | 2017     |

**Beta-barrel**

| 1FPO | C. perfringens | Perfringolysin a | 4  | 500    | JM109 | pRT10 | His  | 6    | N   | None            | None      | 1998     |
| 1H6S, 1BH3 | R. blanca | Porin (E1M/A166K) | 16 × 3 | 307    | BL21(DE3) pLyS5 | pET1b | NT² | 6   | LDAO            | Xray-C8E4 | 1998     |
| 1UYN | H. influenzae | Autotransporter (trimeric Hia) | 12  | 308    | BL21(DE3) | pET1a | NT² | 6   | None            | Xray-C10E5 | 2004     |
| 3O44, 1XZ7 | V. cholerae | Pore-forming toxin (Cytosolin) | 14  | 741    | Origami B | phi-parallel12 | His  | 6    | N   | C10E6, None     | None      | 2005     |
| 2GR8 | H. influenzae | Trimeric autotransporter (Hia) | 4 × 3 | 99     | B834 | pASK-B1A12 | Strep | N   | Elugent | Xray-C8E4 | 2006     |
| 2ODJ | P. aeruginosa | Channel (OprD) | 18 | 443    | C43(DE3) | pBAD22 | His  | 6    | C   | MD⁶            | Xray-C8E4 | 2007     |
| 2VDF | N. meningitidis | OM adherin (OpaA) | 10  | 253    | AR58 | pMG1 | NS⁵ | 6   | LDAO            | Xray-C10E5 | 2005     |
| 2QTK | P. aeruginosa | Benzoate channel (OprK) | 18 | 390    | C43(DE3) | pBAD22 | His  | 8    | C   | Elugent        | Xray-C8E4 | 2008     |
| 2KT4 | H. sapiens | Voltage-dependent anion channel (VDAC) | 19 | 291    | M15 | FDM6 | His  | 6    | C   | IB-gua         | Xray-C8E4 | 2008     |
| 3EMN, 4C69 | M. musculus | Anion channel (VDAC-1) | 19 | 295    | M15 | pQE9 | His  | 6    | N   | IB-gua         | Xray-C8E4 | 2008     |
| 3DWO | P. aeruginosa | Fatty acid transporter (FadLI) | 14 | 451    | C43(DE3) | pBAD22 | His  | 6    | C   | LDAO            | Xray-C10E5 | 2008     |
| 2K0L | K. pneumoniae | OuterMP (OmpA) | 8 | 216    | BL21(DE3) | pET21c | His  | 6    | C   | IB-gua         | Xray-C8E4 | 2008     |
| 3CL0 | S. marscens | Heme receptor complex (HasR) | 12 | 319    | C43(DE3) | pET21a | His  | 8    | C   | LDAO            | Xray-C8E4 | 2008     |
| 3FID | S. typhimurium | Lipid A deacetylase (LpxR) | 22 | 707    | BL21(DE3) | pET20b | His  | 6    | C   | Elugent | Xray-C8E4 | 2008     |
| 3FHH | P. aeruginosa | Outer MP (OprM) | 4 × 3 | 485    | C43(DE3) | pET22b | His  | 8    | C   | LDAO            | Xray-C8E4 | 2009     |
| 3EMN, 4C69 | M. musculus | Anion channel (VDAC-1) | 19 | 295    | M15 | pQE9 | His  | 6    | N   | IB-gua         | Xray-C8E4 | 2008     |
| 3DWO | P. aeruginosa | OuterMP (OmpA) | 8 | 216    | BL21(DE3) | pET21c | His  | 6    | C   | LDAO            | Xray-C10E5 | 2008     |
| 3CL0 | S. marcescens | Heme receptor complex (HasR) | 22 | 865    | MC4100 | pRT2 | His  | 6    | N   | Elugent | Xray-C8E4 | 2008     |
| 3FID | S. typhimurium | Lipid A deacetylase (LpxR) | 12 | 319    | C43(DE3) | pET21a | His  | 8    | C   | LDAO            | Xray-C8E4 | 2009     |
| 3FHH | P. aeruginosa | Outer MP (OprM) | 22 | 640    | C43(DE3) | pET20b | His  | 6    | C   | Lon-gua        | Xray-C8E4 | 2009     |
| 3KVN | P. aeruginosa | Auto transporter (EstA) | 12 | 646    | C43(DE3) + BL21(DE3) | pET22 | His  | 7    | N   | LDAO            | Xray-C8E4 | 2010     |
| 2X55 | Y. pestis | OM protease (ompA) | 10 | 293    | C43(DE3) | pET22 | His  | 6    | C   | LDAO            | Xray-C8E4 | 2010     |
| 4QKY, 3NJT | B. pertussis | Transporter FhaC | 16 | 554    | BL21(DE3) | pET24d | His  | 6    | N   | OG             | Xray-OG   | 2010     |
| 2X27 | P. aeruginosa | Outer MP G (OprG) | 8 | 233    | C43(DE3) | pBAD | His  | 6    | C   | MD⁵            | Xray-C8E4 | 2010     |
| 3ANZ | S. aureus | α-hemolysin | 2 × 2 | 647    | C43(DE3) | pET28b | His  | 6    | C   | None            | None      | 2011     |
| 3SNG | S. typhi | Porin (OmpF) | 16 | 341    | GJ1158 | pET20b | His  | 6    | C   | LDAO            | Xray-C8E4 | 2011     |
| 3RRH | P. aeruginosa | Export protein (AlgE) | 18 | 479    | BL21(DE3) | pET28a | His  | 6    | C   | LDAO            | Xray-OmPOE| 2011     |
| 2LHF | P. aeruginosa | Outer MP H (OprH) | 8 | 179    | BL21(DE3) | pET20a | His  | 6    | C   | LDAO            | Xray-OmPOE| 2011     |
| 3B07, 4P1Y, 4P1X | S. aureus | γ-hemolysin: (LukF) and (Hlg2) | 16 | 309 + 290 | BL21(DE3) | pET26 + pRAREI | His  | 6    | N   | None            | None      | 2011     |
| 3QRA | Y. pestis | Adhesion protein (Aii) | 8 | 157    | BL21(DE3) | pET16b | His  | 6    | N   | MD³            | Xray-C8E4 | 2011     |
| 3SY7, 3SY9, 3SYB | P. aeruginosa | Aromatic hydrocarbon (OscK3, OcpD) | 18 | 401    | BL21(DE3) | pET22 | His  | 6    | N   | MD³            | Xray-C8E4 | 2012     |
| 3SZD | P. aeruginosa | Channel (OcpK2, OcpD) | 18 | 405    | BL21(DE3) | pET22 | His  | 6    | N   | MD³            | Xray-C8E4 | 2012     |

*(continued on next page)*
| PDB  | Organism     | Name                           | TM   | Length | Strain(s)       | Vector | Tag    | Size | N/C  | Solubilization | Structure       | Date   |
|------|--------------|--------------------------------|------|--------|-----------------|--------|--------|------|------|----------------|-----------------|--------|
| 3SY9 | P. aeruginosa | OM carboxylate channels (Occ)  | 18   | 430    | BL21(DE3)       | pB22   | His    | 6    | N    | MD             | X-ray-C8E4      | 2012   |
| 3SYS | P. aeruginosa | OM carboxylate channels (Occ)  | 18   | 430    | BL21(DE3)       | pET22b | His    | 10   | N    | Elugent        | X-ray-C8E4      | 2012   |
| 4ETX | N. meningitida| Transferrin (TbpA)             | 22   | 915    | BL21(DE3)       | pET9   | His    | 10   | N    | Elugent        | X-ray-MD (LCP)  | 2012   |
| 4ETY | Y. pseudo-    | tuberculosis                    | 12   | 245    | BL21(DE3)       | pET9   | His    | 10   | N    | Elugent        | X-ray-MD (LCP)  | 2012   |
| 4EY  | P. putida     | Carbohydrate transporter (OprK) | 16   | 436    | BL21(DE3)       | pBAD22 | His    | 7    | N    | Elugent        | X-ray-C8E4      | 2012   |
| 2LME | Y. enterocolita| Autotransporter (trimeric YadA)| 12   | 105    | BL21(DE3)       | pIBA1adM | NT    |      |      |                |                |        |
| 3VZT | N. meningitida| Export protein (AlgE)          | 16   | 355    | BL21(DE3)       | pET1b  | His    | 6    | C    | IB-urea        | X-ray-LDAO      | 2014   |
| 3VX4 | P. aeruginosa | OM protein (PorB)              | 22   | 915    | BL21(DE3)       | pET200 | NS     |      |      | OG             | X-ray-TDDG      | 2012   |
| 4K3C | H. ducreyi    | β-barrel assembly machinery (RamA) | 16   | 532    | BL21(DE3)       | pET20b | His    | 10   | NS   | Elugent        | X-ray-C8E4      | 2013   |
| 4HSC | S. pyogenes   | Streptolysin O pore-forming toxin | 4    | 571    | XL1-Blue        | NP     | NP     |      |      |                |                |        |
| 4UIM | D. rerio      | Voltage-dependent anion channel 2 | 19   | 289    | M15             | pQE60  | His    | 6    | C    | IB-gua         | X-ray-C8E4      | 2014   |
| 4FR7 | D. damanii    | OM porin (KdgM)                | 12   | 222    | BL21(DE3) omplpLysS | pSKM17 | His    | 6    | C    | OPOE           | X-ray-C8E4      | 2014   |
| 4N4R | S. enterica   | LPS transport proteins (LptD-LptE complex) | 26   | 786(LptD) + 196(LptE) | C43(DE3) | pET28b and pACYCDuet-1 | pBAD22 | His    | 6    | C    | LDAO           | X-ray-C8E4      | 2014   |
| 4Q5S | S. flexneri   | Lipopolysaccharide transport (Lpt) | 30   | 802(LptD) + 175(LptE) | BL21(DE3) | pBAD22 | His    | 6    | C    | LDAO           | X-ray-C8E4      | 2014   |
| 4MT0 | N. gonorrhoeae| OM multidrug efflux pump (MtrE) | 4    | 467    | C43(DE3)       | pBAD22b2mnrE | His    | 6    | C    | DDM           | X-ray-MD     | 2014   |
| 3JPC | B. anthracis  | Protective antigen (PA-63)     | 14   | 562    | BL21(DE3)       | pET22b-PA NT | TrxA-His | 6    | N    | None           | EM-Igpal       | 2015   |
| 4MKO | P. entomophila| β-barrel pore-forming toxins (β-PFT) | 2    | 236    | Rosetta™ 2(DE3) pLysS | pETG-20A | None |      |      |                |                |        |
| 4RL9 | A. baumannii  | Carboxenem-associated OM protein | 8    | 255    | BL21(DE3) and C43(DE3) | pET15, pR22 | His    | 6    | N or C | IB-urea | X-ray-C8E4 | 2015   |
| 4V3G | K. oxytoca    | OM protein (CymA)               | 14   | 339    | C43(DE3) ΔcymABC2 | pET22 | His    | 7    | N    | Elugent        | X-ray-C8E4      | 2015   |
| 2N2M | Y. pestis     | Attachment invasion locus (All) | 8    | 156    | NF             | NF     | NF     | NF    |      |                | IsNMR-DePC      | 2015   |
| 4RL8 | P. putida     | OM protein (Cgo4313)           | 12   | 275    | C43(DE3)       | pET22 | His    | 6    | C    | IB-urea        | X-ray-C8E4 or MD | 2015 |
| 5BUN | S. enterica   | Antigenic OM protein            | 12   | 467    | BL21(DE3)       | pH DST | His    | NS    | NS   | DDM           | X-ray-DMM       | 2015   |
| 2N6L | P. aeruginosa | OM protein (OprG)              | 8    | 215    | BL21(DE3)       | pET30a | His    | 6    | C    | IB-urea        | X-ray-DMM       | 2015   |
| 5DL7 | A. baumannii  | OM carboxylate channel (OccA)   | 18   | 419    | C43(DE3)       | pET22 | His    | 7    | N    | Elugent        | X-ray-C8E4      | 2016   |
| 5DL5 | A. baumannii  | OM carboxylate channel (OccA)   | 18   | 430    | C43(DE3)       | pET22 | His    | 7    | N    | Elugent        | X-ray-C8E4      | 2016   |
| 5DL6 | A. baumannii  | OM carboxylate channel (OccA)   | 18   | 413    | C43(DE3)       | pET22 | His    | 7    | N    | Elugent        | X-ray-C8E4      | 2016   |
| 5DL8 | A. baumannii  | OM carboxylate channel (OccA)   | 18   | 407    | C43(DE3)       | pET22 | His    | 7    | N    | Elugent        | X-ray-C8E4      | 2016   |
| 5GAQ | E. fetida     | β-pore-forming toxins (β-PFTs)  | 2    | 310    | Rosetta™ 2(DE3) | pHis-Parallel1 | His    | 6    | C    | None          | EM-DDM         | 2016   |
| 5JXM | Y. pestis     | Lipopolysaccharide transport (Lpt) | 26   | 577    | BL21(DE3)       | pET9 /pCDFc1b | His    | 10 and 6 | N and C | Elugent | X-ray-C8E4 | 2016   |
| 5IV8 | K. pneumoniae | Lipopolysaccharide transport (Lpt) | 26   | 601 + 182 | BL21(DE3)       | pET9/pCDFc1b | His    | 10 and 6 | N and C | Elugent | X-ray-C8E4 | 2016   |
| 5IVA | P. aeruginosa | Lipopolysaccharide transport (Lpt) | 26   | 646    | BL21(DE3)       | pET9/pCDFc1b | His    | 10 and 6 | N and C | Elugent | X-ray-C8E4 | 2016   |
| 5AZO | P. aeruginosa | Multidrug efflux pump (OprN-J) | 12   | 455    | C43(DE3)       | pET21b | His    | 6    | C    | TX100          | NPOE           | 2016   |
| 5FVN | E. cloacae    | Porin (OmpX)                   | 16   | 342    | BL21(DE3) omplp8 | pBAD24 | NT    |      |      |                |                |        |

(continued on next page)
Table 9 lists the genotypes of the bacterial hosts identified in this analysis.

Expression systems that are not T7RNAP-based do not require λDE3-containing hosts. Despite this, it is noticeable that in the case of the ara expression system, C43(DE3) is used more than any other strain: 16 out of 47 non- E. coli uMPS and 5 out of 12 E. coli uMPS were produced in C43(DE3). Whether the lac super-repressor mutation or another mutation found in this host [12] is advantageous for the regulation of the arabinose promoter remains to be demonstrated.

The T7RNAP-based expression system in combination with C41(DE3) or C43(DE3) has been mostly used to produce α-helical membrane proteins, while BL21(DE3) and other BL21(DE3) derivatives were also used to produce β-barrel membrane proteins. The situation is opposite for the arabinose expression system where C43(DE3) hosts produced mainly β-barrels. In the T7RNAP-based expression system, C41(DE3) and C43(DE3) hosts were more frequently used for uMPS containing more than 7 transmembrane domains, while BL21(DE3) was preferentially used for smaller proteins, typically with 1–2 transmembrane domains (Tables 1 and 2).

It is clear that selecting the optimal combination of promoter, tag and bacterial host is key to achieving suitable recombinant membrane protein yields for biophysical studies. In order to provide some guidance, in our experience the following applies to the T7RNAP-based expression system with the C41(DE3) and C43(DE3) bacterial strains, which were originally derived using high copy number plasmids (200–600 copies/cell, such as those containing the pMB1 origin of replication). A non-exhaustive list of suitable plasmids includes pMW7 and derivatives (pHis and pRun) [13,14], pGEM (Promega), pRSET and pDEST (Invitrogen), pVEX (5prime) and pPR-IBA (IBA). It is important to note that the chosen plasmid should not contain lacI or lacO sequences because further attenuation of the T7 promoter is often not needed for those expression hosts (see also comments on stability testing in Section 3). For BL21(DE3) derivatives, medium copy number vectors (pET series) and those containing lacI or lacO sequences (e.g. pET 3, 9, 14, 17, 20 or 23 from Novagen) are more suitable because they reduce the amount of T7RNAP before induction. Use of the companion plasmid pLyS inhibits T7RNAP after induction. The BL21AI host, which contains the T7RNAP gene under the control of the arabinose promoter or the Lemo21 host [15], which contains a companion plasmid expressing the lysozyme gene under the control of the rhamnose promoter, may also be useful to titrate the amount or activity of T7RNAP.

### 2.4. Promoter usage for yeast expression

Table 10 lists the yeast promoters and strains that are integral components of yeast expression systems. Table 11 lists the corresponding genotypes. Typically, episomal plasmids are used for expression in *S. cerevisiae*, while the expression cassette is integrated into the genome of *P. pastoris*. This situation probably results from the reproduction of early successes with these combinations. Since the *P. pastoris* system depends upon very strong promoters, only a few copies of the gene (as present in stably-integrated strains) are required to

| Table 2 (continued) | PMB | Organism | Name | Linear | PDR-2(DE3) | 885C (DE3) | 650C (DE3) | 305C (DE3) | BL21(DE3) | RP493(DE3) | 450C (DE3) | 540C (DE3) | 545C (DE3) | 750C (DE3) | 850C (DE3) |
|---------------------|-----|----------|------|--------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| 2.4. Promoter usage for yeast expression

Table 10 lists the yeast promoters and strains that are integral components of yeast expression systems. Table 11 lists the corresponding genotypes. Typically, episomal plasmids are used for expression in *S. cerevisiae*, while the expression cassette is integrated into the genome of *P. pastoris*. This situation probably results from the reproduction of early successes with these combinations. Since the *P. pastoris* system depends upon very strong promoters, only a few copies of the gene (as present in stably-integrated strains) are required to
Table 3: Unique membrane protein structures derived from recombinant proteins produced in Pichia pastoris.

| PDB | Organism | Description | TM | Size | Host | Vector | Tag | L | N/C | Solub | Structure | Date |
|-----|----------|-------------|-----|------|------|--------|-----|---|-----|-------|-----------|------|
| 1GOS, 1OJA | H. sapiens | Monoamine oxidase B (MAO) | 0 | 520 | KM71 | pPIC3.5 K | NS | None | Xray-LDAO or SB3-12 | Xray-DM | 2001 |
| 2BXR | H. sapiens | Monoamine oxidase A (MAO) | 0 | 527 | KM71 | pPIC3.5 K | NS | None | Xray-OG | Xray-OG | 2005 |
| Alpha-helical | | | | | | | |
| 2AT9, 3LUT | R. norvegicus | Voltage-dependent potassium ion channel (kV channels) | 6 × 4 | 333 | SMD1163 (HIS +) | pPICZC | His | 8 | N | DDM | Xray-DM | 2005 |
| 1Z98, 3CLL | S. cerevisiae | Plant aquaporin | 7 | 281 | KM71H | pPIC3.5 K | NS | | | | Xray-DM | |
| 2U1H, 4JZC | H. sapiens | LTC4 synthase | 4 × 3 | 156 | KM71H | pPICZB | His | 6 | C | OG | Xray-OG | 2005 |
| 2RSR, 3LMH | R. norvegicus | Voltage-dependent K1 channels (kV) | 6 × 6 | 514 + 333 | SMD1163 (HIS +) | pPICZ-C | His | 10 | N | DDM | Xray-MD | 2007 |
| 3D9S | H. sapiens | Human aquaporin 5 | 6 | 286 | X33 | pPICZ-B | NT | | | | Xray-NG | 2008 |
| 3GSO | H. sapiens | Aquaporin (AQP) | 2 × 4 | 343 | SMD1163 (HIS +) | pPICZB | GFP | C | DM | Xray-DM | 2010 |
| 3RZE | H. sapiens | Histamine-H1 receptor (H1R) | 7 | 452 | SMD1163 | pPICZK | GFP + His | 8 | C | DDM | Xray-DM | 2011 |
| 3SYO | H. sapiens | G protein-gated K+ channel (GIRK) | 2 × 4 | 340 | SMD1163 (HIS +) | pPICZC | GFP + His | 10 | C | DM or Xray-DM | DDM | 2011 |
| 4F4C | C. elegans | P-glycoprotein (P-gp) | 12 | 1321 | SMD1163 | pPICZ and pVL1393 | GFP + His | 10 | C | DDM | Xray-DM | 2012 |
| 4HKR | D. melanogaster | Voltage-dependent calcium channels (Ca2P channels) | 4 × 6 | 214 | SMD1163 (HIS +) | pPICZC | His | 6 | C | DDM | Xray-MD | 2012 |
| 3UME, 4WFF | H. sapiens | Human aquaporin 2 (AQP2) | 7 | 230 | SMD1163 | pPICZB | GFP + His | 10 | C | DDM | Xray-FC12 | 2011 |
| 3UKM | H. sapiens | Human aquaporin 5 | 6 | 286 | X33 | pPICZ-B | NT | | | | Xray-NG | 2011 |
| 4NEF | G. gallus | Inward-rectifier potassium channels | 4 × 6 | 409 | SMD1163 (HIS +) | pPICZ | Anti-tubulin Ab | 5 | C | DDM | Xray-DM | 2012 |
| 3UE7, 4KFM | M. musculus | P-glycoprotein 12 | 12 | 1321 | SMD1163 | pPICZK | GFP + His | 10 | N | DDM | Xray-DM | 2012 |
| 4RDQ, 5T5N | C. elegans | Calcium release-activated calcium channel | 4 × 5 | 409 | SMD1163 (HIS +) | pPICZ | Anti-tubulin Ab | 5 | C | DDM | Xray-DM | 2012 |
| 3WME | H. sapiens | Human aquaporin 5 | 6 | 286 | X33 | pPICZ-B | NT | | | | Xray-NG | 2011 |
| 4FIC | C. elegans | Voltage-dependent calcium channels (Ca2P channels) | 4 × 6 | 214 | SMD1163 (HIS +) | pPICZC | His | 6 | C | DDM | Xray-MD | 2012 |
| 3UE7, 4KFM | M. musculus | P-glycoprotein 12 | 12 | 1321 | SMD1163 | pPICZK | GFP + His | 10 | N | DDM | Xray-DM | 2012 |
| 4RDQ, 5T5N | C. elegans | Voltage-dependent potassium channels (kV) | 6 × 4 | 333 | SMD1163 (HIS +) | pPICZB | GFP + His | 10 | C | DDM | Xray-DM | 2012 |
| 3UE7, 4KFM | M. musculus | P-glycoprotein 12 | 12 | 1321 | SMD1163 | pPICZK | GFP + His | 10 | N | DDM | Xray-DM | 2012 |

1 Number of transmembrane domains.
2 Tag length.
3 N- or C-terminal position.
4 Solubilization detergent.
5 Detergent used for structure determination.
6 Not specified in PDB or corresponding publication.
7 Mixed detergent.
8 No tag.
9 Not found or article was not accessible.
obtain sufficient levels of mRNA, although it is apparent that copy number and protein yield are not linearly correlated. The negative impact of secretory stress and clonal instability are areas on which investigators have focussed attention in the search for productive recombiant *P. pastoris* strains [16]. A recent report has streamlined the ‘time-to-strain pipeline’ in *P. pastoris* [17]. In contrast, in *S. cerevisiae*, the promoter may be 10- to 100-fold weaker, so the use of episomal plasmids with high copy numbers is advantageous; episomal plasmids are available for *P. pastoris* [18], but are not yet widely used in structural biology projects.

The strong *S. cerevisiae* promoter, *P*<sub>GAL1</sub>, is induced with galactose while *P*<sub>AOX1</sub> (a very strong *P. pastoris* promoter) is induced with methanol [19]. In choosing a strong promoter, the idea is that transcription should not be rate limiting. However, high mRNA synthesis rates may be countered by high rates of mRNA degradation [20]. Evidence from bacterial expression systems suggests that lowering promoter efficiency via mutation can lead to improved functional yields of membrane proteins for some, but not all, targets [21]. It has been proposed that the ideal inducible system would completely uncouple cell growth from recombinant synthesis, which requires the host cell to remain metabolically capable of transcription and translation in a growth-arrested state. In this scenario, all metabolic fluxes would be diverted to the production of recombinant protein [22]. While this approach is yet to be demonstrated for membrane protein production in yeast cells, soluble chloramphenicol acetyltransferase was produced to more than 40% of total cell protein in *E. coli* [23] suggesting that this may be a strategy worth exploring in yeast. As for bacteria, yeast growth rates often (but not always) decline dramatically upon induction of yeast cultures, in part achieving this state.

3. Bacterial expression systems for membrane protein production: *P*<sub>lacI</sub>-based expression protocols

### 3.1. Optimization of culture growth conditions for improved membrane protein production

We have previously examined the importance of optimizing growth conditions for improved membrane protein production in bacterial host cells [24,25] and have published an analysis of the T7RNAP-based expression system [8]. Here we combine these insights with our updated analysis of the PDB. For simplicity, we refer only to the T7RNAP-based expression system in this section, but the principles of most of our advice can be applied more widely to other microbial expression hosts.

An important, but simple, test that should be done prior to culturing recombinant strains is to assess whether the selected plasmid/bacterial host combination is stable over time in the medium to be used for large-scale production. We suggest assessing individual cultures from five independent colonies. After overnight growth in the presence of a suitable antibiotic, 10<sup>-6</sup>, 10<sup>-7</sup> and 10<sup>-8</sup> dilutions should be plated on 2×TY agar with and without antibiotic. If the same number of colonies is obtained in the absence or presence of antibiotic, then the plasmid is stable and it is appropriate to proceed to growing large-scale cultures. However, if the number of colonies is higher in the absence of antibiotic, the expression plasmid is unstable (even prior to inducing expression of the target gene), and it is not advisable to prepare a large-scale culture. In this scenario, it would be prudent to change to a better regulated host strain such as C41(DE3), C43(DE3), Lemo21(DE3) or BL21(DE3) pLysS. Alternatively, some investigators do not plate cells after heat shock but use the whole transformation medium as a pre-culture [26]. By doing this, they take advantage of the significant variability in target gene expression level from one colony to another, in the hope of achieving a reasonable recombinant protein yield.

In general, however, it is preferable to start from freshly-transformed bacterial cells. A typical approach is to inoculate 5 ml 2×TY medium with an isolated colony and incubate overnight. The next morning, this should be used to inoculate 500 ml of 2×TY medium in a 2.5 L flask. The culture should reach an optical density of 0.6 in fewer than 5 h; if not, then the basal expression level of the target gene must be impairing cell growth, which usually affects the stability of the expression plasmid. Instead of the typical induction protocol (0.7 mM IPTG at *A*<sub>600</sub> = 0.6), two options are also worth trying. The first is to add IPTG and instead to let the culture grow overnight at 30 °C or 37 °C. This protocol works well for high copy number plasmids that are not regulated (i.e. they lack the T7lac promoter and/or multicopy *lacI* or lysis gene expression); two membrane protein structures were obtained without inducing the culture in this way [27,28]. The second method is to add IPTG at the beginning of the stationary phase (*A*<sub>600</sub> = 1) either in trace amounts (10 μM) following the improved protocol of Alfasi and colleagues [29] or at a high concentration (0.7 mM) in the stationary phase (Table 12). However, adding IPTG in the stationary phase is not recommended when using C41(DE3) or C43(DE3) and will result in decreased expression levels of the target gene.

### 3.2. Selection of mutant T7RNAP-based expression strains for toxic genes

C41(DE3) and C43(DE3) were originally selected as part of a strategy to produce a membrane protein target that was toxic to BL21(DE3) host cells [30]. The protocol summarized here allows the selection of a bacterial strain to produce any given toxic target membrane protein. Having a reporter gene such as GFP makes the experiment faster but is not essential; C41(DE3) and C43(DE3) were selected without the use of a fluorescent reporter.

The expression plasmid containing the gene of interest should be transformed into BL21(DE3) using calcium chloride and with 1–10 ng of plasmid. After incubation of the 1 ml transformation culture for 1 h at 37 °C, 100 μl are spread onto a 2×TY plates with antibiotic and onto 2×TY plates with antibiotic supplemented with either 0.4 mM or 0.7 mM IPTG (this range avoids the non-specific toxicity of IPTG above 0.7 mM). If the vector expressing the target membrane protein does not prevent cell growth on IPTG-containing plates, mutant strains cannot be selected. If there are hundreds of colonies in the absence of IPTG but very few in the presence of IPTG, some mutants may appear at high frequency.

Typically, five selection experiments can be performed in one day: five 250 ml flasks containing 50 ml 2×TY medium with antibiotic are each inoculated with one bacterial colony. Once the culture has reached *A*<sub>600</sub> = 0.4–0.6, IPTG is added at 0.7 mM final concentration to induce gene expression. One to two hours after induction, 1 ml culture is harvested and serial 10<sup>-1</sup> to 10<sup>-4</sup> dilutions are plated onto the IPTG- and antibiotic-containing plates. The frequency of appearance of mutant hosts varies from 10<sup>-4</sup> to 10<sup>-6</sup> [30]. After an overnight incubation at 37 °C, the number of colonies of different sizes is counted. Large colonies have usually lost the ability to express the target gene in contrast to small colonies, which arise at a frequency of 1–20%. Fig. 6 shows selection experiments with the green fluorescent protein (GFP) as a reporter gene. Panel A shows the size difference between mutant hosts under normal light while panel B shows the same plate under UV exposure. Almost all the small colonies are green and therefore express high amounts of GFP. Large colonies exhibit no or weak fluorescence. Panel C shows a selection experiment where all colonies are small. Among them, some exhibit very high fluorescence intensity. Panel D shows another independent experiment where medium colonies are fluorescent, while the very small ones are not. In the case where a GFP reporter is not being used, membrane protein production can be assessed by immuno-detection or by staining an SDS-PAGE gel with Coomassie Brilliant Blue.

To check whether the mutation is within the bacterial genome of the colony or the plasmid DNA, mutant colonies need to be cured of the expression vector. The vector DNA can be isolated using standard ‘miniprep’ protocols, while the colony can be cured of the plasmid through spontaneous loss in the absence of antibiotic (ten days were...
Table 4

Unique membrane protein structures derived from recombinant proteins produced in *Saccharomyces cerevisiae*.

| PDB   | Organism          | Description                                      | TM\(^1\) | Size | Host   | Vector | Tag   | L\(^2\) | N/C\(^3\) | Solub\(^4\) | Struct\(^5\) | Date   |
|-------|-------------------|--------------------------------------------------|----------|------|--------|--------|-------|--------|---------|-----------|-------------|--------|
| 1OSW  | *R. norvegicus*   | Monoamine oxidase (MAO)                          | 1        | 534  | BJ2168 | YEp51  | His   | 6      | N       | FC12      | Xray- MD\(^7\) | 2004   |
| 2ZX5  | *R. sapiens*      | Monoamine oxidase (MAO)                          | 1        | 513  | BJ2168 | YEp51  | His   | 6      | N       | FC12      | Xray- MD\(^7\) | 2008   |
| 4AI1  | *V. radiata*      | H\(^+\) translocating pyrophosphatases (H\(^+\)-Pases) | 16       | 766  | BJ2168 | pYH6   | His   | 6      | C       | DDM       | Xray-DM    | 2012   |
| 4AV3  | *T. maritima*     | Pyrophosphatases (M-Pases)                        | 16       | 735  | RI1991 | pRS1024| His   | 6      | N       | DDM       | Xray-OGNG or Cyma5 | 2012   |
| 4IL3  | *S. mitake*       | CaaX protease Ste24p                             | 7        | 461  | BJ5460 | pGSP46 | His   | 10     | C       | DDM       | Xray-C12E7 | 2013   |
| 4J05  | *P. indica*       | Phosphate-H\(^+\) symporter (PHS)                | 12       | 530  | DSY-5  | p423-GAL1 | His + FLAG | 10    | C       | DDM       | Xray-NG    | 2013   |
| 4K1C  | *S. cerevisiae*   | Ca\(^{2+}\)/H\(^+\) exchanger (VCX1)            | 11       | 421  | DSY-5  | p423-GAL1 | His + FLAG | 10    | C       | DDM       | Xray-DDM   | 2013   |
| 4NAB  | *O. cuniculus*    | Sarco(endo)plasmic reticulum Ca\(^{2+}\)-ATPase (SERCA) | 16       | 1000 | W303.1b pYdP60 | BAD | C | CI2E8 | Xray-C12E8 | 2013   |
| 4C9G  | *S. cerevisiae*   | Mitochondrial ADP/ATP carrier                     | 6        | 318  | WB12   | pYES3  | His   | 8/9    | N       | UDM       | Xray-DM or Cyma5 | 2014   |
| 4LXJ  | *S. cerevisiae*   | Cytochrome P450                                  | 1        | 536  | MMLY41 | NS\(^6\) | His   | 6      | C       | DDM       | Xray-DM    | 2014   |
| 4WIS  | *N. haematococca* | Anoactin family (TMEM16)                         | 10 × 2   | 735  | FGY217 | pYES2  | GFP + His | 10    | N       | DDM       | Xray-EM-DMM | 2014   |
| 3JPT  | *S. cerevisiae*   | H\(^+\)-ATPases (V-ATPases) (subunits ABCDEFGIac-c’d) | 8a + 40 cc’c’ | 616 + 517 + 392 + 256 + 23 + 118F + 478 + 840a + 160 cc’c’ + 345d | JTY002 | NS\(^2\) | FLAG | C | DM | EM-DMM | Xray-DM, or NG or OGNG | 2015   |
| 5AZN  | *A. thaliana*     | Proton-coupled transporters (NRT1/PRK)           | 12       | 590  | FGY217 | pKS426GAL1 | GFP + His | 8    | C       | DDM       | Xray-DDM   | 2015   |
| 5AEK  | *S. cerevisiae*   | (Mep2)                                           | 10       | 505  | W303.1b ppe4A | p38aA | His | 6      | C       | MD\(^7\) | Xray-DMNG  | 2016   |
| 5AEZ  | *C. albicans*     | Mep2 proteins                                    | 11 × 3   | 486  | W303.1b ppe4A | p38aA | His | 6      | C       | MD\(^7\) | Xray-DMNG  | 2016   |
| 5DQQ  | *A. thaliana*     | Two-pore channels (TPCs)                         | 12 × 2   | 723  | DSY-5  | p423-GAL1 | His | 10    | C       | DDM       | Xray-DDM   | 2016   |
| 5H9   | *R. norvegicus*   | Transient receptor potential (TRP)               | 6 × 4    | 770  | BJ5457 | pYepM   | 1D4 epitope | C | LMNG | EM-DMG | Xray-DMNG  | 2016   |
| 5EQI  | *H. sapiens*      | Glucose transporter I (hGLUT1)                   | 12       | 492  | DSY-5  | p423-GAL1 | His | 10    | C       | DDM       | Xray-NG    | 2016   |
| 5I6C  | *A. nidulans*     | Uric acid/xanthine H\(^+\) symporter (UapA)      | 14 × 2   | 574  | FGY217 | pdDGFP | GFP + His | 8    | C       | DDM       | Xray-DDM   | 2016   |
| 5TJ5  | *S. cerevisiae*   | Vacuolar-type ATPases (V-ATPases) (subunits aarc’c’d) | 6a + 40cc’c’ + 2e + 2f | 680a + 150c + 147c + 213c’ + 387d + 57e + 54f | CACY1 | NS\(^2\) | FLAG | C | DM | EM-DMM | Xray-DDM, or OGNG | 2016   |
| 5LZQ  | *T. maritima*     | Pyrophosphatases (M-Pases)                        | 16 × 2   | 735  | RIJ2168 | pYES2  | His | 6      | C       | DDM       | Xray-DM    | 2016   |
| 516P  | *S. cerevisiae*   | (Hrd1-Hrd3)                                      | 8 × 2    | 407  | INVS1  | pKS42X | Strept | C  | DM  | EM-amphipol\(^8\) | 2017   |

\(^1\) Number of transmembrane domains.

\(^2\) Tag length.

\(^3\) N- or C-terminal position.

\(^4\) Solubilization detergent.

\(^5\) Detergent used for structure determination.

\(^6\) Not specified in PDB or corresponding publication.

\(^7\) Mixed detergent.

\(^8\) For an overview of amphipols, see J Membr Biol, 247 (2014) 759–96.
Table 5
Number of unique membrane protein structures derived from recombinant proteins produced in microbial expression systems.

| Origin of membrane protein | Expression system |
|----------------------------|-------------------|
|                            | *Escherichia coli* | *Pichia pastoris* | *Saccharomyces cerevisiae* | Total |
| *E. coli*                  | 144               | 0                | 0                            | 144   |
| Non-*E. coli* bacteria     | 237               | 0                | 2                            | 239   |
| Plant                      | 2                 | 5                | 3                            | 10    |
| Mammal                     | 41                | 18               | 5                            | 64    |
| Bird                       | 0                 | 3                | 0                            | 3     |
| Virus                      | 5                 | 0                | 0                            | 5     |
| Fungus                     | 4                 | 1                | 12                           | 17    |
| Parasite                   | 2                 | 0                | 0                            | 2     |
| Other1                     | 5                 | 4                | 0                            | 9     |
| Total1                     | 468               | 31               | 22                           | 521   |

1 Plexaura homomalla, Danio rerio, Eisenia fetida, Actinia fragacea, Caenorhabditis elegans.
2 The PDB was analyzed in November 2017.

Fig. 1. Secondary structure analysis of recombinant membrane proteins produced in *E. coli* and yeast for which a structure has been resolved. Data were obtained from Tables 1–4. The percentage of unique membrane protein structures is plotted as a function of their amino acid content for recombinant membrane proteins produced either in yeast (grey) or *E. coli* (black).

3.4. Collecting proliferated membranes or inclusion bodies from *E. coli* hosts

Formation of inclusion bodies containing a recombinant membrane protein (IBM) occurs frequently in bacteria especially for non-*E. coli* targets. Inclusion body formation is usually not toxic to the cell, the recombinant protein can be accumulated to very high levels and, in some cases, the protein is in an ‘amyloid’ form which entraps functional protein [4]. Bacterial inclusion bodies have been shown to spontaneously penetrate mammalian cells and can be targeted to specific receptors, opening the way to deliver functional drugs. Due to their natural abundance and the fact that Ni²⁺-affinity chromatography can be performed in denaturing conditions, IBM can be purified in large quantities. One application is their use as an alternative to peptides for raising specific antibodies against eukaryotic proteins [5,6]. As mentioned in Section 2.1, large scale refolding of inclusion bodies has been attempted in the field of structural biology and some progress has been made especially for NMR analysis. For instance, several G protein-coupled receptors (GPCRs) have been produced in a functional form (after refolding of *E. coli*-produced inclusion bodies in amphipols [7,8]).
Table 6
Unique eukaryotic membrane protein structures derived from recombinant proteins produced in E. coli.

| PDB | Organism | Name | TM | Length | Strain(s) | Vector | Tag | Size | N/C | Solub | Struct | Date |
|-----|----------|------|----|--------|-----------|--------|-----|------|-----|-------|--------|------|
| 1UUM | R. rattus | Flavin dihydroorotate dehydrogenase (DHOD) | 0 | 372 | XL1-Blue Tetra | pASKDr | His | 6 | N | None | Xray-OG | 2004 |
| 2PRM | H. sapiens | Dihydroorotate dehydrogenase (DHODH) | 0 | 367 | BL21(DE3) T7 | pET19b | His | 10 | N | TX100 | Xray-MD | 2008 |
| 3O8Y | H. sapiens | Enzyme 5-lipoxygenase (5LOX) | 0 | 691 | Rosetta™ 2(DE3) T7 | pET14b | His | 6 | N | None | None | 2011 |
| 4HIS | A. thaliana | Fatty acid ω-dioxygenase (ω-Dox) | 1 | 43 | BL21(DE3) T7 | pETDuet-1 | His | 6 | N | DM | Xray-NG | 2013 |
| 4NRE | H. sapiens | Enzyme 5-lipoxygenase-2 (5LOX-2) | 1 | 43 | BL21(DE3) T7 | pETDuet-1 | His | 6 | N | None | None | 2014 |
| 4PLA | H. sapiens | Phosphatidylinositol 4-kinase type IIa (PI4K IIa) | 0 | 556 | BL21(DE3) Star T7 | pRSPD | His | 6 | N | None | None | 2014 |
| 4Q9N | H. sapiens | Fatty-acid ethanolamides (FAEs) | 0 | 393 | Rosetta-origamiB (DE3) pET28a | pMAL-C2/8a | His | 6 | N | Sodium cholate | Xray-DC | 2015 |

### Alpha-helical

| PDB | Organism | Name | TM | Length | Strain(s) | Vector | Tag | Size | N/C | Solub | Struct | Date |
|-----|----------|------|----|--------|-----------|--------|-----|------|-----|-------|--------|------|
| 1L2L, 2M3B | H. sapiens | Phospholamban homopentamer | 1 | 52 | BL21(DE3); Other | pMALc2 x | MBP | N | NS | | inNMR-FC12 | 2005 |
| 2HAC | H. sapiens | TCR-CD3, TM dimer complex | 1 | 24 | BL21(DE3) pET28a | pGEX2 | His | 9 | N | IB-gua + TX100 | inNMR-SDX/FC12 | 2006 |
| 2J01 | H. sapiens | Neuronal acetylcholine receptor (ChR2) | 0 | 330 | Rosetta™ 2(DX) | pET22b | His | 6 | C | Cell-free expression-pellet | inNMR-FC12 | 2009 |
| 2KTV | M. musculus | Mitochondrial uncoupling protein 2 | 2 | 303 | Rosetta™ 2(DE3) pET21 | pET21 | His | 6 | C | FC12 | FC12 | 2011 |
| 2L2N | H. sapiens | Receptor (CXC1) | 0 | 239 | BL21(DE3) pET28a | pGEX2 | His | 6 | C | Cell-free expression-pellet | inNMR-FC12 | 2014 |
| 2LOU | M. musculus | Endothelial growth factor receptor 2 | 2 | 330 | Rosetta™ 2(DX) | pET22b | His | 6 | C | FC12 | FC12 | 2014 |
| 2M6B | H. sapiens | Glycine receptor (GlyR-α1) | 4 | 150 | BL21(DE3) pET28a | pET31b | His | 6 | C | IB-Ns | inNMR-SDX/FC12 | 2014 |
| 2L2L | H. sapiens | Growth factor receptor 3 (GFR3) | 1 | 143 | Rosetta™ 2(DE3) pET28a | pGEX2 | His | 6 | C | Cell-free expression-pellet | inNMR-FC12 | 2014 |
| 2MAW | H. sapiens | Neuronal acetylcholine receptor | 4 | 147 | Rosetta™ 2(DE3) pET28a | pET22b | His | 6 | C | Cell-free expression-pellet | inNMR-FC12 | 2014 |
| 2MR, 3HD7 | R. norvegicus | Syntaxin 1A, TM & syntaxin complex | 1 | 209 | Rosetta™ 2(DE3) pET28a | pGEX2 | His | 6 | C | Cell-free expression-pellet | inNMR-FC12 | 2014 |
| 4BUO | R. norvegicus | Neurotensin receptor | 0 | 335 | BL21(DE3) pET22b | pET15b | His | 6 | N | MD | Xray-OG | 2014 |
| 4O6Y | A. thaliana | Cytochrome b561 (Cyto b561/B) | 0 | 239 | Rosetta™ 2(DE3) pET22b | pET15b | His | 6 | N | DM | Xray-OG | 2014 |
| 2MS9 | M. musculus | Translocator protein (TSPO) | 2 | 377 | BL21(DE3) pET22b | pET15b | His | 6 | N | None | Xray-OG | 2014 |
| 2MGO | M. musculus | Translocator protein (TSPO) | 5 | 169 | BL21(DE3) pET22b | pET15b | His | 6 | N | IB-SD | inNMR-FC12 | 2014 |
| 2MFH | H. sapiens | Insulin receptor (IIS) | 0 | 580 | BL21(DE3) pET22b | pET15b | His | 6 | N | DM | Xray-OG | 2014 |
| 2MCO | R. norvegicus | Neurotrophin receptor (p75) | 1 | 24 | BL21(DE3) pET22b | pET15b | His | 6 | N | None | Xray-OG | 2014 |
| 4T58 | A. fragilis | Haemolytic fragaceatoxin C (FrcA) | 0 | 80 | BL21(DE3) pET22b | pET15b | His | 6 | N | None | Xray-DMD | 2015 |
| 4WOL | H. sapiens | Signaling module (DAP12) | 0 | 33 | BL21(DE3) pET22b | pET15b | His | 6 | N | IB-SD | Xray-OG | 2015 |
| 5E4A | H. sapiens | Glycophorin A (Gpa) | 0 | 556 | BL21(DE3) pET22b | pET15b | His | 6 | N | None | Xray-OG | 2015 |
| 2N2A | H. sapiens | Receptor tyrosine kinases (HER or ErbB) | 0 | 24 | BL21(DE3) pET22b | pGEX2 | His | 6 | N | None | Xray-OG | 2016 |

(continued on next page)
### Table 2 (continued)

| PB2 | Organism | Name | N/C | Solub | Date   | TM | Length | Strategy | Insertion | Tag | Solub |
|-----|-----------|------|-----|-------|--------|-----|--------|----------|-----------|-----|-------|
|     |           |      |     |       |        | 2014|        |          |           |     |       |
| 2JK4 | D. rerio | D. rerio | C. elegans | 19 | 289 | 2016 |     | C       | None     | 6    | C     |       |
|     |           |      |     |       |        | 2014|        |          |           |     |       |
| 2DE3 | H. sapiens | Anion channel (VDAC-1) | H. sapiens | 19 | 294 | 2014 |     | C       | None     | 6    | C     |       |

- **Number of transmembrane domains.**
- **Solubilization detergent.**
- **Vector.**
- **Strategy.**
- **Insertion.**
- **Solub.**
- **TM.**
- **Length.**
- **N/C.**
- **Date.**
- **Organism.**
- **Name.**
- **Number of transmembrane domains.**
- **Solubilization detergent.**
- **Vector.**
- **Strategy.**
- **Insertion.**
- **Solub.**
- **TM.**
- **Length.**
- **N/C.**
- **Date.**
- **Organism.**
- **Name.**

**Note:**
- **Not found in PDB or corresponding publication.**
- **Not found or article was not accessible.**
- **Inclusion bodies solubilised in 6 M guanidine hydrochloride.**
- **Mixed detergents.**
- **Not specified in PDB or corresponding publication.**
- **Inclusion bodies solubilised in 6 M guanidine hydrochloride.**

---

**4. Methods**

As shown in Table 5, *E. coli* has been engineered and optimized for use as an expression host to produce proteins from both prokaryotic and eukaryotic organisms. This section is concerned exclusively with the overexpression of genes encoding prokaryotic membrane proteins, for which *E. coli* is usually an ideal expression host. The strain of *E. coli* illustrated here, BL21(DE3), was selected for its lack of both the lon and ompT proteases, and as a consequence of the previous successes achieved for high-level expression of membrane transport proteins [44-51]. Overexpression of all target genes is initially examined and verified by the culture of *E. coli* BL21(DE3) host cells, harbouring the

---

**and successfully studied by high resolution NMR [9]. However, refolding of IBMP is challenging because some surfactants maintain misfolded membrane proteins in solution, as exemplified by the mitochondrial uncoupling protein structure (2LCK, Table 2) which is not physiologically relevant [10]. Consequently, although producing IBMP for structural studies could be considered, we have focused on targeting heterologous membrane protein targets to bacterial membranes, ideally in proliferating membranes.**

Intracellular formation of membranes in *E. coli* has been observed upon the overproduction of several classes of proteins: 1. integral membrane proteins including the whole ATP-synthase [32] or AtpF, its membrane bound subunit b [33], the chemotaxis receptor Tsr [34], the sn-glycerol-3-phosphate acyltransferase [35] and the fumarate reductase [36]; 2. monotopic membrane proteins including the glycosyltransferase MurG [37]; the monoglycosyl diacylglycerol synthase (MGS) from *Acholeplasma laidlawii* [38,39] and the N-methyltransferase PmtA from *Agrobacterium tumefaciens* [40]; and 3. Amphipatic protein oligomers made of caveolin [41,42] or deriving from elastin-like peptide repeats (ELP, [43]).

AtpF is a good example of an *E. coli* membrane protein that can be produced either as inclusion bodies in C41(DE3) or in a folded state within internal proliferating membranes in C43(DE3). Accumulation of *atpF* mRNA is similar 3 h after induction in both expression hosts but the time course of expression is delayed by 30 min in C43(DE3) [3]. Optimized expression conditions were 16 h of induction with 0.7 mM IPTG at 25 °C. In these conditions, the viability of the cells was restored and overproduction of AtpF did not trigger toxicity.

Despite this example, inclusion body formation is frequent and difficult to avoid with eukaryotic membrane protein targets. When producing a membrane protein in bacteria, it is therefore important to check for the presence of inclusion bodies and to prepare carefully cellular or internal bacterial membranes. Inclusion bodies can be isolated following two centrifugation steps: 600g for 10 min to collect unbroken cells and cell debris in the pellet, followed by 10,000g for 15 min at 4 °C to collect inclusion bodies from the supernatant. **Intracellular formation of membranes in eukaryotic organisms.**

**This section is concerned exclusively with the use of plasmid pTTQ18**

As shown in Table 5, *E. coli* has been engineered and optimized for use as an expression host to produce proteins from both prokaryotic and eukaryotic organisms. This section is concerned exclusively with the overexpression of genes encoding prokaryotic membrane proteins, for which *E. coli* is usually an ideal expression host. The strain of *E. coli* illustrated here, BL21(DE3), was selected for its lack of both the lon and ompT proteases, and as a consequence of the previous successes achieved for high-level expression of membrane transport proteins [44-51]. Overexpression of all target genes is initially examined and verified by the culture of *E. coli* BL21(DE3) host cells, harbouring the
plasmid pTTQ18 [45–47] containing the gene of interest, in 50 ml LB medium and inducing with 0.5 mM IPTG at mid-log phase (A_{680} ∼ 0.4–0.6). The cells are harvested 3 h after induction of the tac promoter and total membranes are prepared from spheroplasts by the water lysis method (Fig. 7). The total membrane proteins are separated by SDS-PAGE and stained with Coomassie brilliant blue and/or analysed by Western blotting with an anti-His antibody. If a protein is found to overexpress well, scaling-up of bacterial culture volumes is undertaken [44]. This is often performed with 30 or 100 L fermenters [52] and inner membranes containing the protein of interest are prepared from the cells using sucrose density gradients (Fig. 7). Note that whole cell lysates can be used for this screening step, but there is a danger of missing successful expression, because the protein is located only in the membrane fraction comprising less than 10% of total cell protein, potentially leading to false-negative results.

![Fig. 3](image1.png)

**Fig. 3.** Tag usage for recombinant membrane protein production in microbes. Data were obtained from Tables 1–4. The number of unique membrane protein structures is plotted as a function of the purification tag present in the corresponding recombinant membrane protein following production in (A) bacteria (black) or (B) yeast (grey).

In our extensive experience of using plasmid pTTQ18 for the heterologous production of bacterial membrane proteins in *E. coli* BL21(DE3), inclusion bodies did not appear. Rather the recombinant protein appeared in the membrane fraction of the disrupted host cell, where it was functionally active in all cases tested.

### 4.1. General choices and considerations for cloning into plasmid pTTQ18

Our preferred cloning strategy is based on the traditional restriction enzyme method, which involves the digestion of both vector and amplified DNA fragments with the relevant restriction enzymes to enable DNA ligation. This method is not high-throughput but is reliable. The pUC-based plasmid pTTQ18 [45] is a high copy number vector, which has been used successfully for the overexpression of diverse membrane transport proteins of the Major Facilitator Superfamily (MFS)

![Fig. 4](image2.png)

**Fig. 4.** Polyhistidine tag usage in microbial expression systems. Data were obtained from Tables 1–4. (A) The position of the polyhistidine tag within a recombinant membrane protein and (B) the number of histidine residues it contains are shown.
Table 7
Promoter and E. coli strain combinations used for the production of recombinant non-E. coli (heterologous) membrane proteins.

| Bacterial strain | Promoter used in the expression plasmid | Total |
|------------------|------------------------------------------|-------|
|                  | T7 | ara | T5 | tet | trp | lac | rh | Total |
| BL21(DE3)        | 88 | 7  | 2  | 3   | 6  | 1   | 1  | 107  |
| BL21(DE3) ΔacrB  | 1  | 1  |    |     |    |     |    | 1    |
| BL21(DE3) omp5   | 1  | 2  |    |     |    |     |    | 2    |
| BL21(DE3) ΔacrBΔBmacABlysS    | 1  | 1  |    |     |    |     |    | 1    |
| C43(DE3)         | 34 | 16 | 2  |      |    |     |    | 52   |
| C43(DE3) ΔcyoABCD| 1  | 1  |    |     |    |     |    | 1    |
| C43(DE3) ΔacrB   | 1  | 1  |    |     |    |     |    | 1    |
| C43(DE3)         | 27 | 2  | 2  |      |    |     |    | 29   |
| C41(DE3) ΔacrB   | 1  | 1  |    |     |    |     |    | 1    |
| BL21(DE3) pLysS  | 2  | 2  | 1  |      |    |     |    | 7    |
| BL21(DE3) omp8 pLysS | 1 | 1 | | | | | 1 |
| BL21(DE3) CodonPlus1 | 14| 1| 1| 3 | 19 |
| Rosetta™ 2(DE3) T7 | 8 | 8 |      |    |     |    | 8    |
| Rosetta™ 2(DE3) pLysS T7 | 5 | 5 |      |    |     |    | 5    |
| BL21(DE3) Tuner | 2  | 1  | 3  |      |    |     |    | 6    |
| BL21(DE3) Star   | 2  | 2  |    |      |    |     |    | 4    |
| SEI T7           | 1  | 1  |    |      |    |     |    | 1    |
| BL21(DE3) Gold   | 3  | 1  | 4  |      |    |     |    | 8    |
| BL21-Gold pLysS  | 1  | 1  |    |      |    |     |    | 1    |
| GJ1158           | 1  | 1  |    |      |    |     |    | 1    |
| Lemo21(DE3)      | 2  | 2  |    |      |    |     |    | 4    |
| Lemo56c(DE3)     | 1  | 1  |    |      |    |     |    | 1    |
| Origami B        | 1  | 1  |    |      |    |     |    | 1    |
| Rosetta- Origami B pLysS | 1 | 1 | | | | | 1 |
| B834             | 1  | 1  | 3  |      |    |     |    | 5    |
| B834 pLysS       | 1  | 1  |    |      |    |     |    | 1    |
| PA (LagF)        | 1  | 1  |    |      |    |     |    | 1    |
| DH10B/TOP10      | 7  | 7  |    |      |    |     |    | 14   |
| XL1-Blue         | 1  | 5  | 1  | 1    |    |     |    | 8    |
| BL21(DE3) AI     | 1  | 1  | 2  |      |    |     |    | 4    |
| DW2              | 1  |     |    |      |    |     |    | 1    |
| DHSta            | 2  | 2  | 2  |      |    |     |    | 6    |
| SG1309           | 2  | 2  | 1  |      |    |     |    | 5    |
| MC4100           | 1  | 1  |    |      |    |     |    | 1    |
| SCM6             | 1  | 1  |    |      |    |     |    | 1    |
| MC1061           | 5  | 5  |    |      |    |     |    | 10   |
| JM83             | 2  | 2  |    |      |    |     |    | 4    |
| M15              | 1  | 1  | 2  |      |    |     |    | 4    |
| EP51             | 1  | 1  | 2  |      |    |     |    | 4    |
| RKR              | 1  | 1  |    |      |    |     |    | 1    |
| MG1655           | 1  | 1  |    |      |    |     |    | 1    |
| JM109            | 1  | 1  |    |      |    |     |    | 1    |
| 420399           | 1  | 1  |    |      |    |     |    | 1    |
| AR58             | 1  | 1  |    |      |    |     |    | 1    |
| BLR              | 1  | 1  |    |      |    |     |    | 1    |
| GT1000 Al(lgKnKamB) | 1 | 1 | | | | | 1 |
| PAP5198          | 1  | 1  |    |      |    |     |    | 1    |
| K83             | 1  | 1  |    |      |    |     |    | 1    |
| AD202            | 1  | 1  |    |      |    |     |    | 1    |
| Not found        | 2  | 2  |    |      |    |     |    | 4    |
| Not specified    | 4  | 1  | 4  |      |    |     |    | 12   |
| Cell-free expression | 1 | 1 | | | | | 1 |
| Total            | 229| 47| 15| 24 | 5 | 333 |    |      |

1 Cell-free expression of genes encoding membrane proteins using E. coli lysates;
2 The number of expression hosts is higher than the number of uMPS because some recombinant proteins were produced in several expression hosts.

[46,47,50,53–55], the S-Helix Inverted Repeat Transporter superfamily (‘S-HIRT’, commonly known as the ‘LeuT’ superfamily [54,55], two-component system (TCS) membrane regulatory proteins [56,57], Proteobacterial Acinetobacter Chlorhexidine Efflux (PACE) family efflux proteins [58] and soluble proteins (e.g. [59]). The efficacy of pTTQ18 as an expression vector for different classes of membrane proteins has been tested and compared with other types of plasmid construct [46,47,54,60]. Also, the desirability of placing a tag (usually (His)₆) at either the carboxyl-terminus or the amino-terminus of the cloned gene has been discussed [61]. Plasmid pTTQ18 contains a polyclinker/lac2α region flanked by a hybrid trp-lac (tac) promoter. The tac promoter consists of the -35 region of the trp promoter fused with the lacUV5 -10 region of the lac promoter (Fig. 8). Basal expression of the tac promoter is minimized by binding of the LacI repressor, encoded by the lacI gene, to the lac operator downstream of the promoter. Also downstream of the tac promoter is the multicloning site, which permits the use of either EcoRI or NdeI restriction enzyme sites at the 5’ end of the amplified gene, and Pet or HindIII enzyme sites at the 3’ end of the gene, for successful ligation. The pTTQ18 plasmid also contains the bla gene for the expression of β-lactamase, conferring ampicillin or carbenicillin resistance (Fig. 8).

The affinity tag of choice in this strategy is the RGSHis₆ motif, which is present on a modified pTTQ18 between the PetI and the HindIII restriction sites, so incorporating the tag onto the carboxyl-terminus of the protein (Fig. 8). The orientation of the carboxyl-terminus of the protein is important since previous experience has shown that, if the carboxyl-terminus is periplasmic, the use of the hexahistidine tag will be unsuccessful (Saidjiamo, M., Baldwin, S.A., personal communications). A possible cause for this is the inability of the hydrophilic histidine tag to traverse the hydrophobic membrane domain. It is therefore necessary before cloning to assess the predicted topology of the protein,
and investigate whether the location of the carboxyl-terminus of the protein is expected to be cytoplasmic or periplasmic, using topology prediction programmes such as TMHMM (www.cbs.dtu.dk/services/TMHMM/). The location of the RGSHis tag on the plasmid also dictates the use of the Pstl restriction site to enable correct fusion of the protein with the tag. However, if Pstl cannot be used because of an internal Pstl site within the gene of interest, then the RGSHis tag can instead be added at the primer level, and the HindIII restriction site used for restriction/ligation cloning.

4.2. Cloning of genes encoding membrane transport proteins

The PCR primers designed for use in many of our studies introduced EcoRI or NdeI and PstI or HindIII restriction sites at the 5’ and 3’ ends of the gene respectively (Fig. 8). The reaction itself was conducted using a set of different melting/annealing/extension temperatures that varied depending on either the melting temperature of the primers or the GC content of the DNA to be amplified.

Following successful amplification, the resulting DNA fragment is purified and digested with the relevant restriction enzymes. We always designed sticky-ended ligation of the DNA fragment with pTTQ18/RGSHis6, using the DNA ligase enzyme. The freshly-ligated DNA is used to transform E. coli XL1-Blue cells for propagation of the plasmid and the resulting carbenicillin-resistant colonies are selected and screened using PCR. Size estimation of the amplified gene can be performed by agarose gel electrophoresis and used to confirm the presence of the

![Fig. 5. Promoter usage for recombinant membrane protein production in E. coli. Data were obtained from Tables 1–3. The graph shows the promoters used for the heterologous (black) and homologous (grey) production of membrane proteins in E. coli. NS: not specified; Native: the native promoter of the gene encoding the target membrane protein.]

Table 9

| Strains for T7RNAP-based expression | Genotype |
|------------------------------------|----------|
| BL21(DE3)                          | E. coli str. F’ ompT gál dcm lon hsdS(rK- mK+) λDE3 [lacI lacUV5 T7 + rpsL80 ind1 ampr) [malB + lacI574 lacI574] |
| C41(DE3)                           | BL21(DE3) F' proY438 melB653 yvgQ1103 yhwA290 ydcZ71 ycfB815 ropC3023 yehU679 rbsD4(53) λDE3 [lacI lacUV5] |
| C43(DE3)                           | BL21(DE3) F' Δ(lacUV5)866-870 fur:Val lona(l84y) yibB90 yjcO665 cydA::IS1 Δ(cmcF-ompC) Δ(yjIV-yjI) rbsD4(53) λDE3 [lacI lacUV5] |
| BL21(DE3)                          | BL21(DE3) pLYS5 (CamR) |
| BL21(DE3) CodonPlus                | BL21 dcm + TetR λDE3 endA Hte [argU proyl CamR] |
| BL21(DE3) Star                     | BL21 rE313 λ(DE3) |
| BL21(DE3) Rosetta pLYS5            | BL21 λ(DE3) pLYS5 (CamR) |
| BL21(AI)                           | BL21 lacY1 λ(DE3) |
| Other expression strains           | Genotype |
| BL21Rosome                         | BL21 RARE (CamR) |
| BL21-Gold                          | BL21 dcm + TetR endA Hte |
| BL21-T1R                          | fpuA2 [lon] ompT gál dcm ΔhSD |
| Origami B                          | BL21 lacY1 aphC gat522::Tn10 trvX (KanR TetR) |
| BB34                               | F’ ompT endA (rB- mb-) gál dcm met |
| BLR                               | F’ ompT endA (rB- mb-) gál dcm Δ(ser-recA)306::Tn10 (TetR) |
| DH10B-TOP10                        | F’ mcrA Δ(mtr-hsdRMS-mrBC) 8780lacZAM15 ΔlacX74 supG recA1 endA139 Δ(ara-lac)7697 galK16 galR16 endA1 λ ΔH10B rpsL(StrR) |
| KRR                               | [F’ tralX6, ΔompP proA + B lacI Δ(lacZ)15] ΔompT endA1 recA1 endA1 galR96 (Nalr) thi-1 hisR17 (R B a + mK + +) λDE3 (recA-Δ lacI574 Δlac-proAB) Δ(lacZ::T7) RNA polymerase |
| XL10-Gold-XL1-Blue                 | F’ [proAB lacIqZAM15 Tn10(TetR Amy CmR)] recA1 endA1 galV44 thi-1 grrA96 recA1 endA1 Hte ΔlacX183 Δ(mcrB::hsdRMS-mrBC)1573 TetR |
| F’ [proAB, lacI qZAM15 Tn10(TetR)] | F’ [proAB, lacIqZAM15 Tn10(TetR)] recA1 endA1 galV44 thi-1 grrA96 recA1 supI sup44 hisR17 (K r – mK + +) l |
| DH5a                              | F’ [proAB lacIqZAM15 Δ(lacI2Y-argU)]T169 deor recA1 endA1 hisR17 (K r – mK + +) pbsA supE44 λDE3 thi-1 grrA96 recA1 |
| SG13009                           | NaI(s) Str(s) Rif(s) ThI- lacI- Ara-1 Gal-1 M13(+) F’ RecA[+] Uvr[+] Lon-1 |
| LS8164                            | ΔaddR ΔaddL |
| MC4110                            | F’ [araL139B r Δ(argF-lac169)] & lambda- e14-fhDS301 Δ(rsak-rst-R)725 (fras255) recA1 endA1570(strR) rbsR22 Δ(rmod-fmE3622::cIs1 |
| MCM6                               | ΔaddC |
| MCI061                            | F’ Δ(ara-lac)7697 Δ(ara-lac169)1573 galK16 galR16 λ Δe14-merCR recA1 endA1570(strR) supT1 merR1 hisR2(R- m+ +) |
| JMB18                             | rpsL ara Δ(lac-proAB) 4800dacZAM15 |
| Other                              | PA[Δ(qppR)] |
incorporating the appropriate restriction sites for cloning, and also pTTQ18 plasmid containing the sequenced gene is used to transform BL21(DE3) cells for expression studies.

DNA sequencing to confirm the expected integrity of the cloned gene is more certainly established by single restriction digestion analysis with the relevant restriction gene of interest. The resulting positive colonies are cultured and the structural determination.

Genotypes of yeast strains used to produce recombinant membrane proteins for Table 11

Table 11

| Yeast strain | Genotype |
|--------------|----------|
| S. cerevisiae | Genotype/phenotype |
| ADΔ | MATa URA3::HIS3/MATα leu2 trp1 ura3-52 prb1-1122 pep4-3 prc1-407 gal2 |
| BJ1991 | MATa +/+ X33 wild-type/Mut Δ |
| BJ2168 | MATa pep4-3/araA112 ura3-52 leu2-3 spor1-1 pep4-3 prc1-407 gal2 |
| BJ5457 | MATa pep4-4/His3 pep1-1 trp1 ura3-52 leu2Δ his3Δ lys2-801 can1 |
| BJ5460 | MATa ura3-52 trp1 lys2-801 leu2Δ his3Δ 200 pep4-4/His3 pep1-16 |
| CACY1 | MATa ura3-52 trp1 lys2-801 leu2Δ his3Δ 200 pep4-4/His3 pep1-16 |
| DSY-5 | MATa leu2Δ trp1 ura3-52 his3Δ pep4-4/Prb1 |
| FGY217 | MATa ura3-52 lys2Δ201 pep4Δ |
| INV51 | MATa his3Δ111111 leu2Δ trp1-289 ura3-52 |
| JTY002 | MATa leu2Δ trp1-1 can1-100 ura3-3 ade2-1 his3Δ11,15 pep4Δ |
| WB12 | MATa ade2-1 leu2-3, 112 his3Δ22, 15 trp1-1 ura3-1 can1-100 aac1::E::UA::LacA::hisG |
| P. pastoris | Genotype/phenotype |
| GS115 | his4/Mut+ His+ |
| KMT1 | his4 arg4 aux1::ARG4/Mut+ His+ |
| SMD1163 | his4 Δpep4 ΔApr1/Mut+ His+ |
| X33 | wild-type/Mut+ |

gene of interest. The resulting positive colonies are cultured and the plasmid DNA extracted. The plasmid DNA can be subjected to double and single restriction digestion analysis with the relevant restriction enzymes to check that the size of the DNA insert is approximately as expected. Integrity of the cloned gene is more certainly established by DNA sequencing to confirm that the gene has been cloned without mutation and is inserted into pTTQ18 with the correct orientation. The pTTQ18 plasmid containing the sequenced gene is used to transform E. coli BL21(DE3) cells for expression studies.

An important alternative strategy is to synthesize the gene de novo incorporating the appropriate restriction sites for cloning, and also modifying the codon usage of heterologous genes so they fit better to the codon usage of E. coli.

4.3. Optimising the production of recombinant membrane transport proteins from plasmid pTTQ18

4.3.1. Production and characterization of the protein YwtG as an exemplar

One example that we can consider in detail is the gene ywtG from Bacillus subtilis. The translated amino acid sequence indicates YwtG is a putative membrane transport protein, which from BLAST similarity searches is predicted to be a MFS sugar transporter. It shares 46% sequence identity with a D-xylose:proton symporter from L. brevis (XyIT), 39% with an arabino:proton symporter from B. subtilis (AraE), 38% with a major myo-inositol:proton transporter from B. subtilis (IoT) and 38% with a D-galactose:proton transporter from E. coli (GalP). Wild type YwtG consists of 457 amino acids with a calculated Mr of 49,192.49 and is predicted by TMHMM to consist of 12 transmembrane helices with both the amino- and carboxyl-terminal located in the cytoplasm. YwtG contains many of the characteristic elements of the sugar porter sub-family of the MFS including a long, central cytoplasmic loop, and the RGRXX sequence motif found between helices 2 and 3. As the topology of the protein allows for the addition of the carboxyl-terminal hexahistidine tag, the recombinant YwtG protein will contain 17 additional residues, increasing the Mr, to 51,041.44.

Analysis of the ywtG gene revealed that there are no inherent EcoRI or Psfl sites within the gene enabling the use of primers designed to introduce an EcoRI and Psfl restriction site at the 5’ and 3’ ends of the gene respectively. The ywtG gene was successfully amplified from B. subtilis genomic DNA using an annealing temperature of 60°C. This fragment was digested with EcoRI and Psfl, yielding a DNA fragment of 1.5 kb (actual – 1.371 kb, Fig. 9A). This was ligated into pTTQ18 and used to transform E. coli XLI-Blue cells. The resulting carbenicillin resistant E. coli XLI-Blue colonies were PCR screened, which revealed six positives (Fig. 9B). These were cultured and the plasmid DNA extracted. Double restriction digestion analysis of pTTQ18/ywtG with EcoRI and Psfl yielded two DNA fragments at 1.3 kb and 4.6 kb, which are similar in size to the gene ywtG (1.371 kb) and pTTQ18/RGSHis6 (4.59 kb; Fig. 9C). DNA sequencing was performed on pTTQ18/ywtG, which revealed that the full length ywtG gene had been cloned successfully but one mutation was present – base number 1000 was changed from guanine (G) to adenine (A), resulting in the YwtG(His) mutant D334N. However, it is not known if this residue is important to structure or function.

E. coli BL21(DE3) host cells harbouring the plasmid pTTQ18/ywtG were grown in LB medium and induced with 0.5 mM IPTG. Total membranes were prepared and separated by SDS-PAGE. The Coomassie Brilliant Blue stained gel revealed a protein band in the membranes from the induced cells with an apparent mass of ~ 31 kDa (well below the predicted Mr) that was absent in the uninduced cell membranes, which constituted 16% of the total membrane protein (Fig. 9D). A positive signal was observed on the Western blot that confirmed the identity of the YwtG(His) protein (Fig. 9D). A minor signal was also observed in the uninduced cells, which is possibly due to ‘leaky’ derepression of the tac promoter on pTTQ18.

4.3.2. Anomalous migration of recombinant YwtG and other membrane transport proteins in SDS-PAGE gels

The relative molecular masses of protein bands observed in Coomassie-stained SDS-PAGE gels and Western blotting were determined by comparing their migrating distances with those of standard protein molecular weight markers. There is a linear relationship between the log10Mr of YwtG, BC0935, BC5418 and YhjI and the distance they migrate on the SDS-PAGE gel.

For many membrane proteins, and especially membrane transport proteins, boiling to solubilize in SDS before running the gel leads to irreversible aggregation and insolubility. Instead we routinely...
solubilize the protein in SDS at temperatures of 30–60 °C for 10–60 min. Such solubilized samples may contain partially-unfolded protein and/or sub-optimal SDS:protein ratios that lead to anomalous migration in the gel. Generally the observed molecular weight is less than predicted (Table 13), though there are often higher molecular weight bands that may represent completely unfolded protein, oligomers, or aggregates. In our experience, the anomalous lower molecular weight bands, as well as the higher ones, are an idiosyncrasy of SDS-protein behaviour – see e.g. [50,53,55–58] and their presence does not imply anything is wrong with the protein.

4.3.3. Dependence of recombinant protein yields on growth and induction conditions

The extent of growth before and after induction can vary, meaning that it is advisable to conduct trials that aim to maximize the amount of cells without compromising the level of the desired recombinant protein in the membrane. Usually, recombinant protein yields in amounts greater than 2–5% of the total inner membrane protein composition trigger cell toxicity, compromise growth and reduce the biomass yield of cells and membranes, whereas it is necessary to achieve levels of 10–50% in order to facilitate later purification and the minimization of contaminating proteins. Generally, 10% is regarded as satisfactory, 20–30% is desirable and often achieved, and 50% was achieved in only one case out of over 100 recombinant proteins produced. These higher levels seriously compromise cell growth and net production. Thus, a compromise needs to be arrived at where induction is left late to maximize the yield of cells, but not so late that the level of induction is reduced.

For each protein that is taken forward for characterization and purification, we try first to determine whether expression is best in rich or minimal medium. We then try a dose–response curve measuring the level of expression achieved in membranes exposed to zero and Fig. 6. Selection of bacterial strains for improved recombinant membrane protein production using GFP as a gene reporter. Isolation of bacterial mutant hosts was performed as described in Section 3.2 of this review and previously [30]. Briefly, the pMW7-GFP-Xa expression plasmid was transformed into BL21(DE3) cells and a single colony was inoculated in 50 ml 2*TY medium. At A600 = 0.4, cells were diluted in water and 100 μl of the 10⁻¹ dilution were plated on an IPTG-containing plate. Plates were illuminated under (A) normal light (two small colonies that did not emit fluorescence are encircled) or (B) UV light (arrows indicate four large colonies that emitted a diffuse fluorescence). Panels (C) and (D) show two other independent experiments with petri dishes illuminated under UV light.

Table 12

| Size of colonies on IPTG plate¹ | Inoculation | Induction | IPTG concentration | Temperature after induction |
|---------------------------------|-------------|-----------|--------------------|-----------------------------|
| No colony                       | no preculture² | No induction¹ | None               | 30 °C or below               |
| Small (>10% reduction)          | preculture³  | A600 = 1   | 10 μM, 0.1 μM      | 37 °C or 25 °C              |
| Minor reduction (<10%)          | preculture⁴  | A600 < 0.6 | 0.4 or 0.7 mM      | 37 °C or 25 °C              |
|                                 | preculture⁵  | A600 < 0.4 | 0.7 mM            | 37 °C or 25 °C              |

¹ Try 0.7 mM and 0.4 mM of IPTG and check the phenotype on plates at 37 °C and room temperature.
² If you need a preculture to grow large volumes or to inoculate a fermenter, then check plasmid stability.
³ See [27,28].
⁴ See [29] for a complete description of the procedure.
⁵ Pre-warm the medium and use the pre-culture at 10⁻² dilution; when the plasmid is stable, antibiotic is no longer required in the large-scale culture.
increments within 0.01–2 mM IPTG. In some cases we have also explored ‘autoinduction’ using lactose/glucose mixtures [44]. This economizes on expensive IPTG, but can take some time before achieving good results. An optimal situation is arrived at where larger scale (30–100 L) cell growth is conducted in fermenters [32] and extended to $A_{600} = 0.6–1.0$, when inducer is added for 1–3 h before harvesting, cooling and freezing the concentrated cell suspension at −80 °C for storage. In the great majority of expression studies using constructs in the pTTQ18 plasmid we have simply used a growth temperature of 37 °C and maintained it during induction, but there are indications that lowering the temperature at the time of induction is beneficial.

Provided the concentrated cells are kept frozen at −80 °C, the recombinant proteins in the inner membrane that we have studied appear to be immortal. Aliquots of cells can be thawed, membranes prepared and the proteins purified any time later (yes, years), though their stability is not necessarily guaranteed during and after purification, of course.

5. Yeast expression systems for membrane protein production

Yeast is both microbial and eukaryotic, meaning it is quick, cheap and easy to culture, whilst having the post-translational pathways present in higher eukaryotic host cells that are absent in bacteria [62]. The two yeast species most widely used for recombinant membrane protein production are *S. cerevisiae* and *P. pastoris* [63,64] (Table 5). Both grow quickly in a range of complex and defined media (doubling times are typically 2.5 h when glucose is the carbon source) in vessels ranging from multi-well plates to shake flasks and bioreactors [64].

*P. pastoris* is notable for being able to grow to very high cell densities under controlled conditions where oxygenation rates are high (>100 g/L dry cell weight; >500 $A_{600}$ units/mL [19]) and therefore has the potential to produce large amounts of recombinant membrane protein for structural analysis. High-resolution crystal structures of the adenosine $A_{2A}$ [65] and the histamine $H_1$ [66] GPCRs have been solved using recombinant protein derived from *P. pastoris*. More recently, a 2.9 Å resolution crystal structure was published of the first plant multidrug and toxic compound extrusion (MATE) transporter to be structurally characterized; the crystals were formed using recombinant protein synthesized in *P. pastoris* [67].

*S. cerevisiae* is notable for being supported by a more extensive literature than *P. pastoris*. Its genetics are also better understood (http://www.yeastgenome.org/). This means that there is a much wider range of tools and strains for improved membrane protein production in this yeast. Recent examples of its use include the generation of the 4.4 Å cryo-EM structure of the rat TRPV2 channel [68] and the 3.0 Å crystal structure of the wild-type human GLUT1 glucose transporter in complex with cytochalasin [69].

The experimental strategy for obtaining the structure of the histamine $H_1$ receptor provides an example of making best use of the two yeast species’ strengths: crystals were obtained from protein produced in *P. pastoris*, while initial screening to define the best expression construct was performed in *S. cerevisiae* [70]. In principle, many of the tools established for *S. cerevisiae* could be transferred to *P. pastoris* (for which a genome sequence was published in 2009 [71]) combining the strengths of both yeast species, although such work would be time-consuming. In our laboratory, we often start with *P. pastoris* and, if the production is not straightforward, use *S. cerevisiae* to troubleshoot [64]. In the following sections, we include the production of the human GPCR, adenosine $A_{2A}$ receptor (h$A_{2A}$R), in both species as an exemplar (Figs. 10 and 11).

5.1. *Saccharomyces cerevisiae*

In studies examining the host response to recombinant membrane protein production, the unfolded protein response (UPR) and altered ribosomal biogenesis [73] have been identified as major determinants of high yields in yeast, although the precise mechanistic reasons for this remain unclear.
5.1. Selection of yeast expression strains for improved membrane protein production

A comprehensive strain collection exists from which potential expression hosts can be selected, supported by information in the Saccharomyces Genome Database (http://www.yeastgenome.org/). The yeast deletion collections comprise over 21,000 mutant strains with precise start-to-stop deletions of approximately 6000 S. cerevisiae ORFs [74]. The collections include heterozygous and homozygous diploids as well as haploids of both MATa and MATα mating types. Individual strains or the complete collection can be obtained from Euroscarf (http://web.uni-frankfurt.de/b15/mikro/euroscarf/) or the American Type Culture Collection (http://www.atcc.org/). Dharmacco sells the Yeast Tet-Promoters Hughes Collection (yTHC) with 800 essential yeast genes under control of a tetracycline-regulated promoter that permits experimental regulation of essential genes. A number of specifically-engineered S. cerevisiae strains also exists including those with ‘humanized’ sterol and glycosylation pathways [75]. Protease-deficient strains are a consistently-popular choice in membrane protein structural biology projects (Table 10 and 11). Often, the standard BY4741 laboratory strain (MATα, ura3Δ, leu2Δ0, met15Δ0, his3Δ1) is a good start, but it is not always the most successful, as shown for the expression of hAαAR (Fig. 10A).

We previously selected four strains of S. cerevisiae for their ability to produce the aquaporin Fps1 in sufficient yield for further study [73]. Yields from the yeast strains spt3Δ, srb5Δ, gcn5Δ and yTHC8MS1 (supplemented with 0.5 µg/mL doxycline) that had been transformed with an expression plasmid containing 249 base pairs of 5’ untranslated region (UTR) in addition to the primary FPS1 open reading frame (ORF) were 10–80 times higher than yields from wild-type cells expressing the same plasmid. One of the strains increased recombinant yields of hAαAR and soluble green fluorescent protein (GFP); all but gcn5Δ were found to exhibit a block in translation initiation. Expression of the eukaryotic transcriptional activator GCN4 was increased in these strains and they also exhibited constitutive phosphorylation of the eukaryotic initiation factor, eIF2α. Both responses are indicative of a constitutively-stressed phenotype.

Investigation of the 5’UTR of FPS1 in the expression construct revealed two untranslated ORFs (uORF1 and uORF2) upstream of the primary ORF. Deletion of either uORF1 or uORF1 and uORF2 further improved recombinant yields in our four strains; the highest yields of the uORF deletions were obtained from wild-type cells. Frame-shifting the stop codon of the native uORF (uORF2) so that it extended into the FPS1 ORF did not substantially alter Fps1 yields in spt3Δ wild-type cells, suggesting that high-yielding strains are able to bypass 5’uORFs in the FPS1 gene via leaky scanning, which is a known stress-response mechanism. Yields of recombinant hAαAR, GFP and horseradish peroxidase could be improved in one or more of the yeast strains suggesting that a stressed phenotype may also be important in high-yielding cell factories [76].

From these studies we concluded that regulation of FPS1 levels in yeast by translational control might be functionally important and the presence of a native uORF (uORF2) may be required to maintain low levels of Fps1 under normal conditions, but higher levels as part of a stress response. We also concluded that constitutively-stressed yeast strains may be useful high-yielding microbial cell factories for recombinant membrane protein production [76].

5.1.2. Using selective advantage to improve membrane protein yields in S. cerevisiae

Making the production of a target recombinant protein a condition for yeast cell survival should give producers a selective advantage over non-producers. This principle has been examined previously for the production of membrane proteins in prokaryotic hosts [77–79]. However, functional yields were not assessed; instead total yields were quantified by immunoblot [77–79]. We therefore investigated whether yeast cells could be given a selective advantage to produce high yields of hAαAR by fusing it with the orotidine-5-monophosphate decarboxylase polypeptide (Ura3p). Ura3p catalyzes the sixth step in the de novo biosynthesis of uridine monophosphate in yeast and is required by ura3 deletion strains when they are cultured in uracil-deficient growth medium [80].

Transformation of S. cerevisiae strain BY4741 (MATα, ura3Δ, leu2Δ0, met15Δ0, his3Δ1) with a control plasmid (pYXX22-hA2AR) generated colonies on solid histidine-deficient growth medium that were designated A2 (the control, producing hAαAR). BY4741 transformants expressing hAαAR-Ura3p (following transformation with pYXX22-hA2AR-URA3) were selected using either a 1- or 2-step process. In the 1-step process, yeast cells were grown on solid uracil-deficient medium immediately following transformation; this generated a single colony designated A2U1 (A2SU1 was similarly generated following transformation of the yeast deletion mutant strain, spo3Δ). In the 2-step process, yeast cells were cultured on solid histidine-deficient medium following transformation and colonies were spotted onto solid uracil-deficient medium (generating A2H1; Table 14 and Fig. 10).

The total yield of hAαAR-Ura3p fusion proteins was analysed by immunoblot following transformation and selection on nutrient-deficient medium. Table 14 shows that the yield of hAαAR-Ura3p from A2H1 was almost 7-fold higher than the yield of hAαAR from the A2 control. No hAαAR-Ura3p was detected from the A2U1 transformant. The yield of hAαAR-Ura3p from A2SU1 was just over half that of A2, with sp33:hAαAR, showing no signal. Changes in the expression levels could also be determined using immunofluorescence staining (Fig. 10) where increased levels of the newly-synthesized A2 receptors could be seen along with the Ura3p-fusions in A2H1 and A2SU1 (Fig. 10). This suggests that making the production of a recombinant GPCR a condition for cell survival through nutrient selection is an effective method to increase total yield. To determine whether the hAαAR-Ura3p we had produced was correctly folded and thereby estimate the functional yield of the hAαAR moiety, a radio-ligand binding assay was performed [81]. Radio-ligand binding analysis was done using the well-characterized antagonist [3H]ZM241385 (82) on 100 µg of total membrane extract from A2, A2H1 and A2SU1. Table 14 shows that A2H1 produced only a minimal increase (1.6 ± 0.1 pmol mg⁻¹) of correctly-folded hAαAR-Ura3p compared to the A2 control (1.1 ± 0.1 pmol mg⁻¹). The yield of hAαAR-Ura3p from A2U1 was negligible (0.2 ± 0.01 pmol mg⁻¹). These findings suggested that the protein produced using this strategy was a heterologous mixture of correctly folded (binding-competent) and misfolded (binding incompetent) protein. In contrast the functional yield of hAαAR-Ura3p from A2SU1 was increased almost 3-fold (3.0 ± 0.2 pmol mg⁻¹) over the A2 control and 6-fold over the mutant strain control sp33:hAαAR (Table 14). Notably, ligand binding activity could be recovered from A2H1, but not A2 or A2SU1, by solubilising the hAαAR-Ura3p in n-dodecyl β-D-maltopyranoside (DDM; Table 14). We validated single-point binding with full saturation curves for A2, A2H1 and A2SU1 in membranes. The Bmax values from these experiments validate the single-point saturation values. Affinity was determined using competition binding experiments, with the pKd being 8.3–8.6 for A2, A2H1 and A2SU1.

In order to rationalize why the functional yield was lower than the total yield, we examined the localization of hAαAR-Ura3p using confocal microscopy following staining of yeast spheroplasts with a mouse anti-hexahistidine antibody followed by an Alexa488-conjugated goat anti-mouse antibody. Fig. 10A shows confocal images for BY4741 expressing no recombinant protein (panel i), the control plasmid pYXX22-A2AR (panel ii) and A2H1 (panel iii); in the latter image, a vacuolar localization of the recombinant protein is observed. Fig. 10A, panel iv shows that vacuolar accumulation of the hAαAR -Ura3p fusion could be reduced by using the BY4741 sp33Δ strain. Homologous competition radio-ligand binding with [3H]ZM241385 (Fig. 10B) demonstrated that hAαAR and hAαAR-Ura3p had comparable pKd values (8.3–8.6) as reported in the literature [83].

When nutrient selection was used as a strategy to increase the yield
of another recombinant GPCR, the human β2 adrenergic receptor (hβ2AR), high yielding transformants were found to have been generated using the 2-step method. B2U1 and B2US gave a 3.2- and 2.5-fold increase in total yield and a 5- and 7-fold increase in functional yield (when solubilized with DDM) of hβ2AR-Ura3p, respectively, compared to the B2 control. The 1-step method (leading to transformant B2H3) did not result in an increased functional yield following DDM-solubilization, either the receptor was not correctly-folded in yeast or solubilization using 2.5% styrene maleic acid copolymer (SMA; 2:1 styrene to maleic acid ratio), increased the functional yield of the control B2 and B2H3 compared to DDM solubilization.

The constructs used in these experiments were not truncated thermostable constructs that had been optimized for recombinant expression. Rather, they were the full wild-type sequences (that had been codon optimized for yeast). The data in Table 14 suggest that prior to solubilization, either the receptor was not correctly-folded in yeast membranes or that it was expressed below the limit of detection when we assayed 100 µg membranes. Extraction by surfactant and subsequent concentration resulted in function being detected suggesting that this process had recovered correctly-folded recombinant protein. This approach demonstrates the power of applying a selective advantage strategy to recombinant GPCR production and provides insight into the role of targeting and quality control.

5.2. Pichia pastoris

One notable and highly-beneficial feature of producing recombinant membrane proteins in P. pastoris (and that has been reviewed extensively elsewhere) is that exceptionally high yields of correctly-folded protein can be obtained, especially under the tightly-controlled conditions achieved in bioreactor cultures. For example, yields of both human aquaporin 1 (hAQP1) and hAβ2AR in bioreactors were more than double those achieved in equivalent shake flask cultures [19]. Moreover, the bioreactors produced higher quality membrane protein as determined by functional assay (more than 150 pmol/mg is reported in several studies [84]). Isolation of hAQP1 was possible at 90 mg/L, while yields of 13 mg/100 g cells were reported for a codon-optimized P-glycoprotein construct [19]. P. pastoris is therefore a highly attractive system for the production of folded, eukaryotic membrane proteins although yields remain protein dependent.

P. pastoris expression plasmids are usually integrated into the yeast genome to produce a stable production strain. Since it is not possible to control precisely the number of copies that integrate, or indeed where they integrate within the genome, the optimal clone must be selected although yields remain protein dependent.
strains are available in which to integrate the expression plasmid for the generation of a recombinant production strain (Table 10 and 11). The wild-type strain, X33, the histidine auxotroph GS115, and the slow-methanol-utilization strain KM71H, have all been used to produce membrane proteins for structural studies [19]. Protease-deficient strains such as SMD1163, which lacks proteinase A and proteinase B, are also available (Table 10 and 11).

In all these strains, *P. pastoris* (like *S. cerevisiae*) post-translationally glycosylates membrane proteins by adding core (Man)8-GlcNAc2 groups, but not the higher-order structures found in humans and other mammals; compared to *S. cerevisiae*, the mannose chains also tend to be shorter. However, the effects of these non-native modifications are not necessarily detrimental and need to be assessed on a case-by-case basis [84]. The high-resolution structure of a glycosylated form of the *Caenorhabditis elegans* P-glycoprotein (using recombinant protein produced in *P. pastoris*) demonstrates that yeast glycosylation does not necessarily hinder crystal formation [87]. Nonetheless, in order to overcome potential bottlenecks in producing, purifying, characterizing and crystallizing human proteins in yeast, engineered strains have been developed including strains with ‘humanized’ glycosylation [88,89] and sterol pathways.

Most proteins produced in *P. pastoris* for structural biology use variations of the standard methanol induction protocol. Fig. 11 shows an example of the recombinant production and purification of hA2AR following the ‘Pichia Fermentation Process Guidelines’ (Invitrogen). The hA2AR protein in this study was tagged with an amino-terminal decahistidine-tag and incorporated an N154Q mutation to prevent glycosylation (the corresponding gene was expressed from the pPICZ αA expression plasmid). Cells were cultured in a bioreactor and depletion of glycerol in the initial glycerol batch phase was indicated by a spike in the dissolved oxygen (DO) reading. This was followed by a fed-batch phase with a 50% (w/v) glycerol solution and a 3 h starvation phase to achieve complete glycerol consumption. During the final hour of starvation, the temperature was reduced from 30 °C to 22 °C and allowed to stabilize. Theophylline, a non-selective hA2AR antagonist (10 mM) was then added to the culture to increase stabilization during expression.

---

**Table 13**

Predicted and calculated sizes of membrane proteins. Protein sizes were calculated from both Coomassie-stained and immunoblotted SDS-PAGE gels. YwtG from *Bacillus subtilis* is predicted to be a sugar transport protein, BC0935 from *Bacillus cereus* a dicarboxylate/α-ketoglutarate transporter, BC5418 from *B. cereus* a sugar/metabolite transporter and Yhj1 from *B. subtilis* a glucose transporter. The data for YwtG are from a different experiment from that shown in Fig. 9, explaining the minor discrepancies in apparent molecular masses.

| Protein | Predicted molecular mass (kDa) | Determined by Coomassie-stained SDS-PAGE | Determined by Western blot detecting the His6 tag | Ratio of observed size/actual size |
|---------|-------------------------------|------------------------------------------|-----------------------------------------------|----------------------------------|
|         | Migrated distance (cm)       | Calculated size (kDa)                    | Migrated distance (cm)                        | Calculated size (kDa)            |                                  |
| YwtG    | 51,042                        | 5.5                                      | 40                                             | 4.3                             | 0.78                             |
| BC0935  | 49,906                        | 6.2                                      | 34                                             | 6.75                            | 0.68                             |
| BC5418  | 45,541                        | 6.6                                      | 31                                             | 5.0                             | 0.68                             |
| Yhj1    | 49,477                        | 6.5                                      | 32                                             | 4.85                            | 0.65                             |
The cells were induced with 100% methanol at an initial feed rate of 1.92 ml/h for 17 h to allow adaptation to methanol. When a steady DO rate and fast DO spike time were obtained, the feed rate was increased to 3.96 ml/h for the remainder of the culture duration. The entire methanol fed-batch phase lasted approximately 40 h with a total of ~125 ml of methanol fed per litre of initial volume. The cells were then harvested by centrifugation.

In a study of the regulation of carbon substrate utilization, we cultured wild-type *P. pastoris* cells in methanol and found that a higher proportion of the total mRNA pool was associated with two or more ribosomes (and therefore judged to be highly translated) compared to t>ured wild-type S. cerevisiae cells in any other non-inducing growth condition.

### Table 14

Characterization of recombinant hA<sub>2A</sub>R/hA<sub>2A</sub>R-Ura3p- and β<sub>2</sub>AR/β<sub>2</sub>AR-Ura3p-producing transformants. Confocal microscopy visualization of recombinant hA<sub>2A</sub>R/ hA<sub>2A</sub>R-Ura3p in transformed S. cerevisiae using AlexaFluor488 antibodies was done to assess whether hA<sub>2A</sub>R/hA<sub>2A</sub>R-Ura3p was localized in the membrane or had been internalized to the vacuole. Immunoblots (50 µg total membrane protein loaded per well, as determined by BCA assay, and probed with Clontech anti-His<sub>6</sub> antibody) were quantified using ImageJ. This allowed comparison of recombinant protein yield from A2H1, A2U1 and A2SU1 (BY4741 or B2 control). The functional yield for hA<sub>2A</sub>R/hA<sub>2A</sub>R-Ura3p-producing transformants. Confocal microscopy visualization of (i) control BY4241 cells, (ii) recombinant hA<sub>2A</sub>R produced from transformant A2, (iii) recombinant hA<sub>2A</sub>R-Ura3p expressed after a 1-step selection from transformant A2H2 and (iv) recombinant hA<sub>2A</sub>R-Ura3p expressed after a 2-step selection from transformant A2SU1. Cells were grown to A<sub>600</sub> = 4–5 and were visualized using rabbit anti-His<sub>6</sub> (Clontech) as the primary antibody and an Alexa-Fluor488-conjugated anti-rabbit secondary antibody. (B) Homologous competition binding experiments were performed for hA<sub>2A</sub>R/hA<sub>2A</sub>R-Ura3p produced from A2, A2H2 and A2SU1 using labelled ([<sup>3</sup>H]) and unlabelled ZM241385. Experiments were done on 100 µg total membrane protein, with A2 acting as the control. Error bars represent the standard deviation (n = 3).

| Transformant name | Expression strain | Vacular Internalization | Total yield from immunoblot (Arbitrary units; relative to control) | Functional yield (pmol mg<sup>−1</sup>) | Functional yield following DDM solubilization (pmol mg<sup>−1</sup>) | Functional yield following SMA solubilization (pmol mg<sup>−1</sup>) |
|-------------------|-------------------|------------------------|---------------------------------------------------------------|---------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|
| hA<sub>2A</sub>R | A2 (control)      | BY4741                 | No                                                           | 1.0                                   | 1.1 (0.2)                                                      | 1.2 (0.3)                                                      |
|                  | A2U1              | BY4741                 | Yes                                                          | 6.8 (0.8)                             | 1.6 (0.1)                                                      | 5.8 (1.6)                                                      |
|                  | A2SU1             | BY4741                 | n/a                                                          | 0.0                                   | 0.2 (0.01)                                                     | n/a                                                           |
|                  | spt3Δ               | BY4741                 | No                                                           | 0.6 (0.2)                             | 3.0 (0.2)                                                      | 3.2 (0.7)                                                      |
|                  | spt3Δ/hA<sub>2A</sub>R | BY4741             | n/a                                                          | n/a                                  | 0.5 (0.1)                                                      | n/a                                                           |
| β<sub>2</sub>AR | B2 (control)      | BY4741                 | n/a                                                          | 1.0                                   | 0.0                                                           | 0.3 (0.02)                                                     |
|                  | B2H3              | BY4741                 | n/a                                                          | 2.4 (1.5)                             | 1.6 (0.3)                                                      | 1.43 (0.4)                                                     |
|                  | B2U1              | BY4741                 | n/a                                                          | 3.2 (0.3)                             | 0.0                                                           | 0.3 (0.1)                                                      |
|                  | B2SU5             | BY4741                 | n/a                                                          | 2.5 (0.2)                             | 2.2 (1.0)                                                      | 1.12 (0.6)                                                     |
This observation suggests that high recombinant protein yields in methanol-grown cells are due not just to promoter strength, but also to the global response of *P. pastoris* to growth on methanol [90]. We have also demonstrated pre-induction expression under the control of *P*<sub>AOX1</sub> [91], suggesting that the uncoupling of growth and protein synthesis in *P. pastoris* cells has not yet been achieved and may provide opportunities for future optimization studies.

6. An overview of detergent usage for microbially-produced recombinant membrane proteins

For membrane protein investigations, the choice of the detergent is crucial, as a suitable one is needed to prepare a pure, stable and monodispersed protein in solution but also to grow well-ordered crystals without preventing crystal contacts. As a consequence, the best detergent for solubilization is often not the best for crystallization and a detergent exchange procedure during purification is a common approach (Tables 1–4). Notably, more than 50% of the membrane proteins in the PDB have been crystallized in a detergent or a detergent mixture that is different from the detergent used for membrane protein solubilization (Fig. 12).

Folded membrane proteins in their native membranes can usually be solubilized with detergent. However after production in homologous membranes, it is frequently found that recombinant membrane proteins are difficult to solubilize. Our simple solubilization screen compares the solubility of the target membrane protein at 1 mg/mL in three different detergents ten times above their critical micellar concentration (cmc): DDM (1% final concentration), FC12 (1%) and SDS (2%). After 1 h incubation at 4 °C on a stirring wheel, insoluble material is removed by ultracentrifugation at 100,000 g for 30 min. The pellet is resuspended with TEP buffer (0.25 mM EDTA, 0.1 mM phenylmethylsulfonyl fluoride and 10 mM Tris-Cl pH 7.8; same volume as that of the supernatant) and both solubilized and non-solubilized fractions are loaded onto an SDS-PAGE gel. If the target membrane protein is solubilized only by SDS, it is likely to be in inclusion bodies (these structures are most frequently associated with bacterial expression systems). If it is solubilized by all three detergents, then it is likely to be well-folded. If DDM cannot solubilize the target membrane protein, experience tells us that it is likely to be misfolded. However, there are many combinations of detergent that can be used to potentially overcome this problem.

The data presented in Fig. 12 show significant differences between detergents used in solubilization and crystallization of β-barrel (Fig. 12A, Supplementary Table 1) and α-helical (Fig. 12B–D, Supplementary Table 1) membrane proteins. For monotopic and α-helical membrane proteins, the most common detergents used for solubilization are by far the maltosides (78%, 72% and 59% for *E. coli*, *S. cerevisiae* and *P. pastoris* expressions systems, respectively) and with DDM contributing more than 80% to this detergent family. DDM is a mild, low cmc, long alkyl chain detergent and has been found to be very stabilising explaining its success in maintaining dynamic membrane proteins in solution. However its large micelles are not well adapted to form ordered, well-diffracting crystals because they limit essential crystal contacts. Therefore, while maltosides are the major detergents used for crystallization of α-helical membrane proteins produced in *E. coli* (48%), an important contribution is made by the smaller glucosides (15%) and especially OG, followed by detergent mixtures (15%) and Cymal detergents (6%). For membrane proteins produced in *S. cerevisiae* and *P. pastoris*, the profile of detergent usage is similar with the exception of neopentylglycol detergents that are used more often (7% and 14%, respectively).

Glucosides (OG, NG) have a high cmc and form small micelles allowing better packing in crystal lattices and resulting in better diffracting crystals [92,93]. OG in particular has been used successfully for channel proteins [92] such as the five OG-crystallized aquaporins derived from *P. pastoris*. The success of detergent mixtures also suggests that combinations of detergents (mostly with small micelle-sized detergents) have a useful contribution to make.

For β-barrels, the three most successful detergents for solubilization are detergent mixtures (24%), the zwitterionic amine oxide detergents (17%) and the maltosides DDM/DM (12%). Strikingly, C8E4, C10E5 or C10E6 are the best detergents for crystallization accounting for half of the structures (42%) followed by detergent mixtures (18%), and amine oxide detergents (16%), in agreement with an earlier study [94]. The high stability of the β-barrel fold supports the use of smaller-micelle-size detergents and more destabilizing detergents.

7. Conclusions

Microbes have an important role to play in membrane protein structural biology projects. *E. coli*, *P. pastoris* and *S. cerevisiae* have together been used to produce 71% of all unique structures in the PDB that...
were derived from recombinant sources. In this review we have focused on an analysis of the host strains, tags and promoters that, in our experience, are most likely to yield protein suitable for structural and functional characterization. We have also exemplified some of our preferred protocols. There are, of course, many other factors that could be considered including codon optimization, mutagenesis, the use of other microbes, engineering of the membrane lipid composition and an in-depth analysis of the culture medium composition. We note, however, that in many cases the approaches we have catalogued provide the requisite quantity and quality of protein for further study.

One of the major challenges in the forthcoming years will be to overcome the barrier of producing complex eukaryotic membrane proteins in microbial systems. There are numerous reports of misfolded recombinant proteins being produced in human cells and tuning human promoters to favour efficient folding is still in its infancy. In contrast there are several initiatives to ‘humanize’ microorganisms. For instance S. cerevisiae has been engineered to synthetize cholesterol instead of ergosterol in order to favour the activity of human GPCRs in yeast membranes. Some T7RNAP-based E. coli strains are fully devoid of lipopolysaccharides and are now recognized as being as safe as Lactobacilllas. Finally, the genetic diversity of microorganisms is now a source of inspiration. For instance some groups are developing semi-synthetic hosts based on magnetotactic bacteria that contain sophisticated intracellular organelles [95]. We therefore anticipate that microbes will continue to make important contributions to the production of recombinant membrane proteins from a range of prokaryotic and eukaryotic organisms.

Acknowledgements

We acknowledge funding from the Biotechnology and Biological Sciences Research Council (BBSRC; via grants BB/N007417/1, BB/P025927/1 and BB/P022685/1 to RMB) and the Innovative Medicines Joint Undertaking under Grant Agreement number 115583 to the ND4BB ENABLE Consortium to RMB. Parts of the research for this review were funded through the EC Membrane Protein Consortium (E-MeP, FP6 LSHG-CT-2004-504601), the European Drug Initiative for Channels and Transporters (EDICT, FP7 Health-F4-2007-201924 and BBBSRC grant BB/D524832/1 to RMB and PJFH. MVD’s PhD student-ship was funded by BBSRC iCASE BB/H016244/1 with Glycoform Ltd. PJFH is grateful to the Leverhulme Trust for an Emeritus Research Fellowship (Grant number EM-2014-045). MP’s PhD fellowship is funded by the French Ministry of Research and Education.
[79] E. Massey-Gendel, A. Zhao, G. Boulting, H.Y. Kim, M.A. Balamotis, L.M. Seligman, R.K. Nakamoto, J.U. Bowie, Protein Sci. 18 (2009) 372–383.

[80] P.J. Flynn, R.J. Reece, Mol. Cell. Biol. 19 (1999) 882–888.

[81] S. Singh, M. Zhang, N. Bertheleme, E. Karu, P.G. Strange, B. Byrne, Curr. Protoc. Protein Sci. Chapter 29 (2012) Unit 29 23.

[82] V.P. Jaakola, M.T. Griffith, M.A. Hanson, V. Cherezov, E.Y. Chien, J.R. Lane, A.P. Ijzerman, R.C. Stevens, Science 322 (2008) 1211–1217.

[83] S. Singh, D. Hedley, E. Karu, A. Gras, S. Iwata, J. Ruprecht, P.G. Strange, B. Byrne, Protein Expr. Purif. 74 (2010) 80–87.

[84] T. Yurugi-Kobayashi, H. Asada, M. Shiroishi, T. Shimamura, S. Funamoto, N. Katsuki, K. Ito, T. Sugawara, N. Tokuda, H. Tsujimoto, T. Murata, N. Nomura, K. Haga, T. Haga, S. Iwata, T. Kobayashi, Biochem. Biophys. Res. Commun. 380 (2009) 271–276.

[85] T. Vogl, L. Gebbie, R.W. Palfreyman, R. Speight, Appl. Environ. Microbiol. (2018).

[86] M. Guerfal, S. Ryckaert, P.P. Jacobs, P. Ameloot, K. Van Craenenbroeck, R. Derycke, N. Callewaert, Microb. Cell Fact. 9 (2010) 49.

[87] M.S. Jin, M.L. Oldham, Q. Zhang, J. Chen, Nature 490 (2012) 566–569.

[88] B. Laukens, C. De Wachter, N. Callewaert, Methods Mol. Biol. 1321 (2015) 103–122.

[89] S.R. Hamilton, P. Bobrowicz, B. Bobrowicz, R.C. Davidson, H. Li, T. Mitchell, J.H. Nett, S. Rauch, T.A. Stadheim, H. Wischnewski, S. Wildt, T.U. Gerngross, Science 301 (2003) 1244–1246.

[90] R. Priellhofer, S.P. Cartwright, A.B. Graf, M. Valli, R.M. Bill, D. Mattanovich, B. Gasser, BMC Genomics 16 (2015) 167.

[91] Z. Basa, S.J. Routledge, M. Jamshad, M. Clare, D. Sarkar, I. Dickerson, M. Ganzlin, D.R. Poyner, R.M. Bill, Microb. Cell Fact. 13 (2014) 127.

[92] S. Newstead, S. Ferrandon, S. Iwata, Protein Sci. 17 (2008) 466–472.

[93] J.L. Parker, S. Newstead, Adv. Exp. Med. Biol. 922 (2016) 61–72.

[94] S. Newstead, J. Hobbs, D. Jordan, E.P. Carpenter, S. Iwata, Mol. Membr. Biol. 25 (2008) 631–638.

[95] I. Kolinko, A. Lohuse, S. Borg, O. Raschdorf, C. Jogleur, Q. Tu, M. Postai, E. Tompa, J.M. Plitzko, A. Brachmann, G. Wanner, R. Muller, Y. Zhang, D. Schuler, Nat. Nanotechnol. 9 (2014) 193–197.