Methods 147 (2018) 3–39

Contents lists available at ScienceDirect

Methods

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

Microbial expression systems for membrane proteins

Marvin V. Dilworth[a,1,2], Mathilde S. Piel[b,1], Kim E. Bettaney[c,1], Pikyee Ma[c,3], Ji Luo[c], David Sharples[c], David R. Poyner[a], Stephane R. Gross[a], Karine Moncoq[b], Peter J.F. Henderson[c,*], Bruno Miroux[b,*], Roslyn M. Bill[a,*]

[a] School of Life & Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK
[b] Laboratoire de Biologie Physico-Chimique des Protéines Membranaires, UMR 7099, CNRS, Université Paris Diderot, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France
[c] Astbury Centre for Structural Molecular Biology and School of Biomedical Sciences, University of Leeds, Leeds LS2 9JT, UK

ARTICLE INFO

Keywords:
Recombinant membrane proteins
Expression plasmid vector
Tag
Promoter
Detergent

ABSTRACT

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.uci.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

* Corresponding authors.
E-mail addresses: p.j.f.henderson@leeds.ac.uk (P.J.F. Henderson), bruno.miroux@ibpc.fr (B. Miroux), r.m.bill@aston.ac.uk (R.M. Bill).
[1] Equal contributions by: Marvin V. Dilworth, Mathilde Piel and Kim E. Bettaney.
[2] Current address: Department of Chemistry, King's College London, Britannia House, 7 Trinity Street, London SE1 1DB, UK.
[3] Current address: Paul Scherrer Institute, CH-5232 Villigen PSI, Switzerland.

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
Available online 12 April 2018

Table 1
Unique membrane protein structures derived from recombinant *E. coli* proteins produced in *E. coli* (homologous expression).

PDB	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
Monotopic											
1B12	Signal peptidase (SPase)	0	280	BL21(DE3)	pET3b	NT[5]			IB-gua[6]	Xray-TX100	1999
1J79, 1XGE	Dihydroorotate dehydrogenase	0	347	XL1-Blue NA[8]	No vector	NT[5]			None	None	2001
2QCU	Glycerol-3-phosphate dehydrogenase (GlpD)	0	501	XL1-Blue/JM109	pQE30	His	6	N	OG	Xray-OG	2008
Alpha-helical											
1A91	Subunit C of the F_1F_o ATP synthase	2	79	MEG119 NA[8]	pC35	NS[7]	NS[7]	NS[7]	NS[7]	lsNMR-solvent	1998
1FFT	Electron transport chain complexes	15 + 2 + 5 + 3	663 + 315 + 204 + 109	GO105 NA[8]	pJRhisB NI[9]	His	6	C	MD[10]	Xray-OG	2000
1FX8, 1LDF	Glycerol facilitator channel (GlpF)	8	281	NS[7]	NS[7]	His	6	N	OG	Xray-OG	2000
1L0V, 5VPN	Native fumarate reductase complex (FrdABCD)	3 + 3	602 + 244 + 131 + 119	DW35 NA[8]	pH3 NI[9]	NT[5]			C12E9	Xray-C12E9	2002
1KQF	Formate dehydrogenase-N (FdnGHI)	1 + 4	1015 + 294 + 217	GL101 NA[8]	NF[11]	NF[11]			NF[11]	NF[11]	2002
1L7V, 2QI9	Vitamin B12 transporter (BtuCD)	10	326 + 249	BL21(DE3)	pET	His	10	N	LDAO	Xray-LDAO	2002
1NEK, 2ACZ, 2WDQ, 2WP9	Succinate:quinone oxidoreductases (dhCDAB)	3 + 3	115 + 129 + 238 + 588	DW35 NA[8]	PSDH15 NI[9]	NT[5]			C12E9	Xray-C12E9 or DM or MD[10]	2003
1OTS, 2EXW, 4FG6	H^+/Cl^- exchange transporter	14	473	BL21(DE3)	pET28b	His	6	C	DM	Xray-DM or OM	2003
1OY6	Multi-drug efflux transporter (AcrB)	12	1049	DH5α	pUC515A	NT[5]			DDM	Xray-DDM	2003
1PW4	Glycerol-3-phosphate transporter (GlpT)	12	451	LMG194	pBAD	His	NS[7]	C	DDM	Xray-MD[10]	2003
1PV7, 2CFQ	Lactose permease (LacY)	12	417	XL1-Blue	pT7-5	His	10	C	DDM	Xray-MD[10]	2003
1Q16, 1SIW, 1Y4Z	Nitrate reductase A (NarGHI)	5 × 2	1247 + 512 + 225	LCB2048	pVA700	NT[5]			C12E9	Xray-C12E9	2003
1RC2, 2O9D, 3NK5	Aquaporin (AqpZ)	8 × 4	231	C43(DE3)	pET28b	His	6	N	OG	Xray-OG	2003
1U7G	Ammonia channel (AmtB)	11 × 3	428	C41(DE3)	pET29b	His	6	C	OG	Xray-OG	2004
1XQF, 2NMR, 2NUU, 2NS1	Ammonia channel (AmtB)	11 × 3	428	C43(DE3)	pET22b	His	6	C	DDM	Xray-LDAO	2004
1ZCD	Na^+/H^+ antiporter (NhaA)	12	388	Rk20 + B834DE3	pAXH	His	NS[7]	NS[7]	DDM	Xray-MD[10]	2005
2ABM	Aquaporin (AqpZ)	8 × 4	231	BL21(DE3) pLysS	NS[7]	His	NS[7]	N	OG	Xray-OG	2005
1T9T, 2GIF	Multi-drug efflux transporter (AcrB)	12	1049	C43(DE3)	pUC515A, pET24a (2GIF)	NT[5]or His	6	C	DDM	Xray-DDM or Cymal6	2005
2GFP	Multidrug transporter (EmrD)	12	375	NF[11]	NF[11]	NF[11]	NF[11]	NF[11]	NF[11]	Xray-DDM	2006
2DHH, 2RDD, 2W1B, 3AOB, 3W9H, 1IWG	Multi-drug efflux transporter (AcrB)	12	1053	JM109, W3104 (ΔAcrA/B)	pACBH (pUC118)	His	6	C	DDM	Xray-DDM or MD[10]	2006
2J58	Translocon for capsular polysaccharides (Wza)	1 × 8	359	LE392	pWQ126 (pBAD24)	NT[5]			SB3-14	Xray-DDM	2006
2HI7	Periplasmic oxidase complex (DsbB-DsbA)	4	176 + 208	M15	pQE70	His	6	C	UDM	Xray-MD[10]	2006
2OAU, 2VV5	Voltage-modulated mechanosensitive channel (MscS)	3 × 7	286	BL21(DE3)	pET28b	His	6	N	FC14	Xray-FC14	2007
2HQC	Multi-drug efflux transporter (AcrB)	12	1053	BL21(DE3) Gold	pSORT1	His	6	C	DDM	Xray-DDM	2007
2V8N	Lactose permease (LacY)	12	417	XL1-Blue	pT7-5	His	10	C	DDM	Xray-MD[10]	2007
2QFI, 3H90	Zinc transporter (YiiP)	6 × 2	300	BL21(DE3) pLysS	pET15b	His	6	N	DDM	Xray-MD[10]	2007
2R6G, 4JBW, 3PV0, 3RLF, 4KHZ, 4KI0	Maltose uptake transporter (MalFGK2)	8 (MalF) + 6 (MalG)	514 + 296 + 381 + 370	HN741	pACYC184 + PFG23	His	6	C	DDM	Xray-UDM	2007
3B5D	Multi-drug efflux transporter (EmrE)	4 × 2	110	BL21(DE3)	pET15b	His	6	N	NG	Xray-NG	2007
3DHW	Methionine uptake transporter (MetNI)	5 × 2	343 + 217	BL21(DE3) Gold	pET19b	His	10	N	DDM	Xray-DDM	2008
3E9J	Periplasmic oxidase complex (DsbB-DsbA)	4	208 + 176	HM125	pQE70	His	6	C	NM	Xray-NM	2008
3FI1	Na^+/H^+ antiporter (NhaA)	12	388	BL21(DE3)	pET	His	6	C	DDM	EM-lipids (2D)	2009
2ZUQ	Periplasmic oxidase (DsbB)	4	176	M15	pQE70	His	6	C	UDM	Xray-DHPC	2009
2WCD	Cytolysin A (ClyA, HlyE)	3 × 12	309	Tuner(DE3)	pET11a	His	6	N	None	Xray-DDM	2009
2KDC	Diacylglycerol kinase (DAGK). Domain-swapped homotrimer	3 × 3	122	BL21(DE3) WH1061	pSD005	His	6	N	Empigen	lsNMR-DPC	2009

(*continued on next page*)

Table 1 (*continued*)

PDB	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
3JQO	Type IV OM secretion complex	1 × 14	396	B834(DE3)	pASK-IBA3	Strep		C	MD[10]	Xray-LDAO	2009
3KCU	Aquaporin-like formate transporter (FocA)	6 × 5	285	BL21(DE3)	pET21b	His	6	C	OG	Xray-MD[10]	2009
3LRB, 3L1L, 3OB6	Arginine:agmatine antiporter (AdiC)	12	445	BL21(DE3)	pET15b	His	6	N	DDM	Xray-NG	2010
3HFX	Carnitine transporter (CaiT)	12	504	C41(DE3)	pET28b	GFP + His	6	C	DDM	Xray-MD[10]	2010
3M9C, 3RKO	Electron transport chain complexes I (NuoLMNAJK)	14 + 14 + 14 + 3 + 3 + 5	613 + 509 + 485 + 100 + 147 + 184	BL21(DE3) NA[8]	No vector	NT[5]			DDM	Xray-MD[10]	2010
3NMO, 4ENE, 4MQX	H^+/Cl^- exchange transporter	14	465	NS[7]	pASK-IBA2	His	6	C	DM	Xray-DM, Cymal4 or DMNG	2010
2WSX	Carnitine transporter (CaiT)	12	504	BL21(DE3)	pET15b	His	6	N	Cymal5	Xray-Cymal5	2010
3O7Q	Fucose transporter (FucP)	12	438	BL21(DE3)	pET21b	His	6	C	DDM	Xray-NG	2010
3K07, 3NE5	Metal-ion efflux pump (CusA)	12	1055	BL21(DE3)	pET15b	His	6	N	Cymal6	Xray-Cymal6	2010
3QE7	Nucleobase/ascorbate transporter	12	429	Bl21(DE3)	pET21b	His	6	C	DDM	Xray-NG	2011
2Y5Y	Lactose permease (LacY)	12	417	C43(DE3)	pET28	His	6	C	DDM	Xray-MD[10]	2011
3TXT, 3UBB	Rhomboid protease (GlpG)	6	185	BL21(DE3)	pET41b	His	8	C	DM	Xray-NG	2011
3VMA	Peptidoglycan glycosyltransferase penicillin-binding protein 1b	1	768	BL21(DE3)	pET15b	His	6	N	DDM	Xray-LDAO	2012
4DJK	Glutamate-GABA antiporter (GadC)	12	511	NS[7]	pET15b	His	6	N	OG	Xray-MD[10]	2012
4FI3	Vitamin B12 transporter (BtuCDF)	10	326 + 249	BL21(DE3) CodonPlus	pET	His	10	N	LDAO	Xray-C12E8	2012
4GBY	Proton:xylose symporter (XylE)	12	491	BL21(DE3)	pET15b	His	6	N	DDM	Xray-MD[10]	2012
4GD3, 3USE	O_2-tolerant hydrogenase-1 in complex with cytochrome *b*	1 + 1 + 4	372 + 597 + 235	FT004 NA[8]	No vector	His	6	C	TX100	Xray-DDM	2013
2LTQ	Periplasmic oxidase (DsbB)	4	176	C43(DE3)	pQE70	His	6	C	DDM	ssNMR-*E. coli* lipids	2013
2LZS	Twin arginine translocase (TatA)	1 × 9	55	BL21(DE3) pLysS	pET24a	His	6	C	C12E9	lsNMR-DPC	2013
4IU9	Nitrate transporter (NarU)	12	462	NS[7]	pET21b	GST		NS[7]	DDM	Xray-MD[10]	2013
4JA3	Proton:xylose symporter (XylE)	12	491	C41(DE3)	pTH27	His	6	N	DM	Xray-DM	2013
4JR9	Nitrate/nitrite exchanger (NarK)	12	463	C41(DE3)	pET15b	His	8	N	DM	Xray-DM	2013
3ZE4	Diacylglycerol kinase (DAGK)	3 × 3	122	BL21(DE3)	pSD005	His	6	N	Empigen	Xray-DM (LCP)	2013
4AU5	Na^+/H^+ antiporter (NhaA)	12	388	BL21(DE3) pLysS	pET derivative	GFP + His	8	C	DDM	Xray-MD[10]	2013
3WDO	Drug efflux transporter (YajR)	12	454	C43(DE3)	pET28a	His	6	C	DDM	Xray-NG	2013
4C48, 5V78	Multi-drug efflux transporter (AcrB/Z)	12 + 1	1049 + 49	C43(DE3)	pET21a/pRSFduet-1	His	6	C	DDM	Xray-MD[10]	2014
4PX7	Phosphatidylglycerophosphate phosphatase B (PgpB)	6	254	C43(DE3)	pET28a	His	6	C	DDM	Xray-MD[10]	2014
4PL0, 5OFR	Antimicrobial peptide transporter (McjD)	6 × 2	580	C43(DE3)	pWaldoGFPd	GFP + His	8	C	DDM	Xray-NG	2014
2MPN	Rhodanese (YgaP)	2 × 2	68	BL21(DE3) Star pLysS	pET3a	His	6	N	DHPC/LMPG	lsNMR-DHPC/LMPG	2014
4QIQ	Proton:xylose symporter (XylE)	12	474	C43(DE3)	pET15b	His	6	N	DM	Xray-MD[10]	2014
4Q65	Peptide transporter (YbgH)	14	493	C43(DE3)	pET28a	His	8	C	DDM	Xray-MD[10]	2014
4QO2	Rhomboid protease (GlpG)	6	200	C43(DE3)	pET15b	His	6	C	DM	Xray-NG	2014
4U8V, 5JMN	Multi-drug efflux transporter (AcrB)	12	1057	C43(DE3)	pET24a	His	6	C	DDM	Xray-DDM	2014
3WVF	Insertase (YidC)	5	548	BL21(DE3)	pTV118N	His	8	C	DM	Xray-DM (LCP)	2014
4X5M	SemiSWEET transporter	3 × 2	89	Rosetta™ 2(DE3)	pET	His	8	C	DDM	Xray-DDM	2015
4RP9	Vitamin C transporter (UlaA)	10	465	C43(DE3)	pET15b or pET21b	His	8	C	NM	Xray-MD[10]	2015
5AJI	Voltage-modulated mechanosensitive channel (MscS)	3 × 7	286	MJF612	pTRcYH6	His	6	C	DDM	Xray-DDM	2015
4U4V	Nitrate/nitrite exchanger (NarK)	12	463	C41(DE3)	pET	GFP + His	8	C	DDM	Xray-DDM (LCP)	2015
4ZYR, 4OAA	Lactose permease (LacY)	12	417	C41(DE3)	PT7-5	His	6	C	DDM	Xray-NG	2015
4ZP0	Multidrug resistance transporter (MdfA)	12	410	C43(DE3)	pET28a	His	6	C	DM	Xray-MD[10]	2015
4UXX	Diacylglycerol kinase (DAGK)	3 × 3	122	WH1061	pTrcHisB	His	6	N	Empigen	Xray-DM (LCP)	2015
5AZC	Phosphatidylglycerol: prolipoprotein diacylglyceryl transferase (Lgt)	7	291	C43(DE3)	pET28a	His	6	C	DM	Xray-OG	2016

(*continued on next page*)

Table 1 (*continued*)

PDB	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
5F5B	Rhomboid protease (GlpG)	6	211	C43(DE3)	pGEX-6P-1	GST		N	DDM	Xray-CHAPSO (bicelles)	2016
5KBN	F-ion channel homologue (Fluc)	4 × 2	126	BL21(DE3)	pASK-IBA2	His	6	C	DM	Xray-DM	2016
5J4N	Arginine:agmatine antiporter (AdiC)	12	445	BL21(DE3) pLysS	pZUDF21	His	10	C	NG	Xray-NG	2016
5SV0, 5SV1	(ExbB/ExbD)	3 + 1	244 + 58	BL21(DE3)	pET28b(ExbB) + pCDF-1b(ExbD)	His	10	C	DDM	Xray-DDM	2016
5GXB	Lactose permease (LacY)	12	417	BL21(DE3)	PT7-5	His	6	C	DDM	Xray-DDM (LCP)	2016
5T4O	F_1F_o ATP synthase	5 + 1 + 2	513 + 460 + 287 + 177 + 139 + 271 + 155 + 79	DK8 NA[8]	pFV2 NI[9]	His	6	N	Digitonin	EM-Digitonin	2016
5IJI	Sensor histidine kinase (NarQ)	4	230	SE1	pSCodon1.2	His	6	C	DDM	Xray-DM (LCP)	2017
5N6H	Apoplipoprotein N-acyltransferase (Lnt)	8	512	C43(DE3)	pET28a	His	6	N	LMNG	Xray-LMNG (LCP)	2017
5XHQ	Apoplipoprotein N-acyltransferase (Lnt)	8	512	C41(DE3)	pET22b	His	8	C	DM	Xray-MD[10]	2017
5TV4	Lipid "flippase" (MsbA)	6	582	BL21(DE3) Star pLysS	pET19b	His	10	N	DDM	EM-POPG (nanodisc)	2017
Beta-barrel											
1MPF	Porin from colicin-resistant (OmpF)	16	340	BZB1107 NA[8]	Native	NT[5]			OPOE	Xray-MD[10]	1995
1GFM	Porin (OmpF)	16	340	Top10	pGEM-7Zf (+) lac	NT[5]			OPOE	Xray-MD[10]	1996
1MPM, 1MAL, 1AF6	Maltoporin (LamB)	18	421	Pop6510 NA[8]	pAC1	NT[5]			OPOE	XRAY-OPOE or MD[10]	1997
1BT9, 3O0E	Porin (OmpF)	16	340	BL21(DE3)	pGEM-5Zf (+)	NT[5]			SDS	Xray-OPOE	1999
1BY3, 1BY5	Ferrichrome-iron receptor (FhuA)	22	747	B834(DE3)/BL21(DE3)	pET	NT[5]			OPOE	Xray-OPOE	1999
1FEP	Siderophore transporter (FecA)	22	747	BL21(DE3)	pET17b	NT[5]			TX100	Xray-LDAO	1999
1QJ8	Porin (OmpX)	8	148	BL21(DE3) pLysS	pET3b	NT[5]			IB-gua[6]	Xray-C8E4	1999
1QKC, 1FI1, 2FCP, 1FCP	Ferrichrome-iron receptor (FhuA)	22	747	AW740 [ΔompF zch:TnlO ΔotnpCfltuA31] NA[8]	pHX405	His	6	Oth[12]	LDAO	Xray-DDAO	2000
1EK9, 1TQQ	OM protein (TolC)	3 × 4	428	BL21(DE3) NA[8]	pAX629 pACYC184	NT[5]			TX100	Xray-MD[10]	2000
1QJP, 1BXW, 1G90	Porin (OmpA)	8	171	BL21(DE3)	pET3b, pET14b	NT[5]			IB-gua[6] or urea	Xray-C8E4, lsNMR-DPC	2000
1FW2	OM phospholipase A (PldA)	12	275	BL21(DE3)	pT7.7	NT[5]			IB-urea	Xray-OG	2001
1HXX	Porin (OmpF)	16	340	BL21(DE3)Δomp8	pGEM-5Zf (+)	NT[5]			SDS	Xray-OPOE	2001
1I78	OM protease (OmpT)	10	297	BL21(DE3)	pET13a	NT[5]			IB-urea	Xray-OG	2001
1ILZ	OM phospholipase A (PldA)	12	275	BL21(DE3) Δ pldA	pND1/pRK3	NT[5]			IB-urea	Xray-OG	2001
1KMO	Siderophore transporter (FecA)	22	774	U600 NA[8]	pSV66	NT[5]			TX100	Xray-LDAO	2002
1MM4, 1MM5, 1THQ, 3GP6	Palimitoyl transferease (PagP)	8	170	BL21(DE3)	pET	His	6	C	IB-gua[6] or SDS	lsNMR-OG or DPC; Xray-LDAO or SDS	2002
1ORM, 1Q9F	Porin (OmpX)	8	148	BL21(DE3) pLysS	pET3b	NT[5]			IB-gua[6]	lsNMR-DHPC	2003
1NQ (E-H), 2GSK	Cobalamin transporter (BtuB)	22	594	BL21(DE3) Star pLysS	pET22b	NT[5]			LDAO	Xray-MD[10]	2003
1PNZ	Siderophore transporter (FecA)	22	774	BL21(DE3)	pET20b	His	10	N	Elugent	Xray-LDAO	2003
1UJW, 2YSU	Cobalamin transporter (BtuB)	22	594	TNE012 ($tsx^-ompA^-ompB^-$) NA[8]	pJC3	NT[5]			OG	Xray-LDAO	2003
1T16, 3PGR	Long-chain fatty acid transporter (FadL)	14	427	C43(DE3)	pBAD	His	6	C	MD[10]	Xray-C8E4	2004
2F1V	Outer membrane protein (OmpW)	8	197	C43(DE3)	pBAD	His	6	C	LDAO	Xray-LDAO or C8E4	2006
2GRX	Ferrichrome-iron receptor (FhuA)	22	747	AW740 NA[8]	pHX405	His	6	OT[11]	LDAO	Xray-C8E4	2006
2F1C	Porin (OmpG)	14	381	C43(DE3)	pBAD	His	6	C	LDAO	Xray-OG	2006
2IWW	Porin (OmpG)	14	281	C41(DE3)	pET26b	NT[5]			IB-urea	Xray-LDAO or OG	2006
2J1N	Osmoporin (OmpC)	16	346	BZB1107 NA[8]	No vector	NT			OPOE	Xray-MD[10]	2006
2HDI	Colicin I receptor (Cir)	22	639	BL21(DE3)	pET20b	His	6	C	Elugent	Xray-MD[10]	2007
2JMM	Porin (OmpA)	8	156	BL21(DE3) Gold	pET3b	NT[5]			IB-gua[6]	lsNMR-DHPC	2007

(*continued on next page*)

Table 1 (*continued*)

PDB	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
2JQY	Porin (OmpG)	14	280	BL21(DE3) pLysS	pT7-SMC	NT[5]			IB-urea	lsNMR-FC12	2007
2QOM	Autotransporter (EspP)	12	285	BL21(DE3)	PC6H1	His	6	C	Elugent	Xray-MD[10]	2007
2VDE	OM protein (TolC)	4 × 3	460	C43(DE3)/C41(DE3)	pET14b	His	8	C	Tx100	Xray-DDM	2008
2VQI	P pilus usher translocation (PapC)	24	515	B834(DE3)	pDG2	His	6	C	DDM	Xray-MD[10]	2008
3DWN	Long-chain fatty acid transporter (FadL)	14	427	C43(DE3)	pBAD	His	6	C	LDAO	Xray-C8E4	2008
2WJR (Mb), 2WJQ (IB)	Porin (NanC)	12	214	BL21(DE3) pLysS	pT7	NT[5]			OPOE, IB-urea	Xray-LDAO, Xray-FC-12	2009
3HW9, 2ZFG	Porin (OmpF)	16	340	MH225 NA[8]	pR272	NT[5]			OPOE	Xray-MD[10]	2009
2XE1	Osmoporin (OmpC)	16	346	HN705Δomp8 NA[8]	PHSG575	NT[5]			SB3-14	Xray-MD[10] or OG	2010
3AEH	Hemoglobin protease (Hbp)	12	308	C43(DE3)	pET22b	His	6	C	SB3-12	Xray-OG	2010
3PIK, 4K7R, 4K34	Heavy metal efflux pump (CusC)	4 × 3	446	C43(DE3)	pBAD22	His	6	C	LDAO or DDM or Cymal6	Xray-DDM or Cymal6	2011
3RFZ	P pilus (FimD)	24	843 + 211 + 279	B834(DE3) /	pAN2 (fimD) + pETS1001 (fimC)	Strep + His	6	C	DDM	Xray-MD[10]	2011
3OHN	P pilus (FimD)	24	558	B834(DE3)	pNH297/pETS4	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)	8	148	BL21(DE3)	pET11a	NT[5]			IB-gua[6]	lsNMR-DMPC/DMPG (nanodisc)	2012
4J3O	P pilus (FimD)	24	843 + 211 + 154 + 279 + 144	Tuner(DE3) /	pNH237 (fimC) + pAN2 (fimD)	Strep + His	6	C	DDM	Xray-MD[10]	2013
2YNK	OM lectin (Wzi)	18	456	Top10/B834(DE3)	pBAD	His	6	C	SB3-14	Xray-LDAO	2013
4C00	Autotransporter (TamA)	16	559	BL21(DE3)	pET22b	His	6	N	OG	Xray-OG	2013
4K7K	Heavy metal efflux pump (CusC)	4 × 3	446	BL21(DE3) Star	pET15b	His	6	C	Cymal6	Xray-cymal6	2013
4C4V	(BamA)	16	467 (344–810)	BL21(DE3)	pET30b	His	6	C	IB-gua[6]	Xray-LDAO	2014
4UV3	Amyloid secretion channel (CsgG)	4 × 9	277	BL21(DE3)	pQLinkN	Strep		C	DDM	Xray-MD[10]	2014
4Q79	Amyloid secretion channel (CsgG)	4 × 9	277	BL21(DE3)	pQLinkN	His	6	N	LDAO	Xray-LDAO	2014
4D5U	Porin (OmpF)	16	340	C41(DE3) NA[8]	No vector	NT[5]			FC12	Xray-FC12	2015
5EKQ	β-barrel assembly machine (BamACDE)	16	810 + 344 + 245 + 113	BL21(DE3)	pJH114	His	8	C	DDM	Xray-C8E4	2016
5AYW	β-barrel assembly machine (BamABCDE)	16	810 + 392 + 344 + 245 + 113	C43(DE3)	pQLink	Strep + His	6	C	DDM	Xray-MD[10]	2016
5D0O	β-barrel assembly machine (BamABCDE)	16	810 + 392 + 344 + 245 + 113	HDB150	pYG120	His	8	C	DDM	Xray-MD[10]	2016
5LJO	β-barrel assembly machine (BamABCDE)	16	810 + 392 + 344 + 245 + 113	BL21(DE3)	pJH114	His	8	C	DDM	EM-DDM	2016
5WQ7	Secretin (GspD)	4 × 15	650	DH5a	pASK-IBA3c	Strep		C	MD[10]	EM-MD[10]	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* (5KUD) 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 monotopic 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 carboxyl-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 MRGSHis_6, while the protein used to solve structure 4V3G contained the following amino-terminal tag: AN-VRLQHis_7LE (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 5DO7, 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, 4HUQ, 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 predominantly used together with six bacterial strains, BL21(DE3), C43(DE3), C41(DE3), BL21(DE3) pLysS, BL21(DE3) CodonPlus or

Table 2
Unique membrane protein structures derived from recombinant non-*E. coli* proteins produced in *E. coli* (heterologous expression).

PDB	Organism	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
Monotopic												
1UUM	*R. rattus*	Flavin dihydroorotate dehydrogenase (DHOD)	0	372	XL1-Blue	pASKDr	His	6	N	None	Xray-OG	2004
2FNQ, 3FG4	*P. homomalla*	Lipoxygenases (LOXs)	0	699	BL21(DE3)	NS[5]	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[6]	2008
2K5U, 2KSQ	*S. cerevisiae*	ADP-ribosylation factors (ARFs)	0	181	BL21(DE3)	pET20b	His	6	C	None	lsNMR-DMPC/DHPC	2009
3I65	*P. falciparum*	Dihydroorotate dehydrogenase (PfDHODH)	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
3O8Y	*H. sapiens*	Enzyme 5-lipoxygenase (5LOX)	0	691	Rosetta™ 2(DE3)	pET14b	His	6	N	None	None	2011
2XCI	*A. aeolicus*	Glycosyltransferase (WaaA, KdsB)	0	374	BL21(DE3) CodonPlus	pUM212/216	His	10	N	TX100	Xray-Cymal6	2011
3RST, 4KWB	*B. subtilis*	Signal peptide peptidase A (SppA)	0	240	BL21(DE3) Tuner	pET28b	His	6	N	None	Xray-DDM	2012
4EHW	*A. aeolicus*	Kinase (LpxK)	0	317	C41(DE3)	pET21b	GST		N	DDM	Xray-DDM	2012
4G9K	*S. cerevisiae*	NADH dehydrogenase (Ndi1)	0	471	BL21(DE3) pLysS	pET16b	His	10	N	DDM	Xray-DDM or DM	2012
4GGM	*C. crescentus*	Phosphodiester hydrolase (LpxI)	0	283	C41(DE3)	T7 based	His	10	N	None	None	2012
4G6G	*S. cerevisiae*	Dehydrogenase (Ndi1)	0	502	C43(DE3)	pQE80L	His	6	N	TX100	Xray-TX100	2012
4HHS	*A. thaliana*	Fatty acid α-dioxygenase (α-DOX)	0	652	M15	pQE30	His	6	N	DM	Xray-NG	2013
3VV9	*T. brucei*	Alternative oxidase (AOX)	0	329	BL21(DE3)	pET15b	His	6	N	OG	Xray-MD[6]	2013
2YOC	*K. oxytoca*	Lipoprotein pullulanase (PulA)	0	1078	pAP5198	pCHAP4486	His	6	C	NS[5]	NS[5]	2013
4NRE	*H. sapiens*	Enzyme 15-lipoxygenase-2 (15LOX-2)	0	696	Rosetta™ 2(DE3)	pETDuet-1	His	6	N	None	None	2014
4NM9	*G. sulfurreducens*	Proline utilization A (PutA)	0	1005	BL21AI	pNIC28-Bsa4	His	6	N	None	None	2014
4NWZ	*C. thermarum*	Non-proton pumping type II NADH dehydrogenase (NDH-2)	0	405	C41(DE3)	pTRCndh2	His	6	C	OG	Xray-OG	2014
4PLA	*H. sapiens*	Phatidylinositol 4-kinase type IIa (PI4K IIa)	0	556	BL21(DE3) Star	pRSFD	His	6	N	None	None	2014
4QN9	*H. sapiens*	Fatty-acid ethanolamides (FAEs)	0	393	Rosetta-origamiB (DE3) pLysS	pMAL	MBP + His	6	C	TX100	Xray-DC	2015
4WVG	*S. aureus*	Bacterial type I signal peptidases	0	542	BL21(DE3) CodonPlus	pProExHta	MBP		N	None	None	2015
5B49	*P. aeruginosa*	UDP-diacylglucosamine pyrophosphohydrase (LpxH)	0	248	B834(DE3) pLysS	pET26b	His	6	C	TX100	None	2016
5KN7	*A. baumannii*	Lipid A acyltransferase LpxM	0	333	C41(DE3)	pRham	His	6	N	DDM	Xray-LMNG	2016
Alpha-helical												
3STL, 3OR7, 3EFF, 1BL8, 1K4C, 2BOB, 2ITC, 3PJS, 4MSW, 2JK5	*S. lividans*	Channel (KcsA)	2 × 4	160	XL1-Blue	pQE32	His	6	N	DM	EPR-DDM; Xray-DM or LDAO	1998
2CPB	*Phage M13*	Major coat protein of M13	1	50	K38	*NF*[7]	NF[7]			NF[7]	lsNMR-FC12 or SDS	1998
1F6G	*S. lividans*	Potassium channel (KcsA)	2 × 4	160	XL1-Blue	pQE32	His	4	N or C	DDM		2001

(*continued on next page*)

Table 2 (*continued*)

PDB	Organism	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
											EPR-asolectin liposome	
1KPL	*S. typhimurium*	Transporter (H^+/Cl^-)	18 × 2	473	BL21(DE3)	pET28b	His	6	C	DM	Xray-OM	2002
3RBZ, 4EI2, 4HYO, 4L73, 1LNQ, 3LDC,	*M. thermauto-trophicus*	Potassium channel (MthK)	2 × 4	340	XL1-Blue, SG1309	pQE70	His	6	C	DM	Xray-LDAO	2002
1ORQ, 2A0L, 2KYH	*A. pernix*	Channel (KyAP)	6	223	XL1-Blue	pQE60	His	6	C	DM	Xray-DM	2003
1XFH, 3KBC, 3V8F	*P. horikoshii*	Glutamate transporter homol (GltPh)	8 × 3	422	TOP10 or DH10B	pBAD	His	8	N	DM	Xray-DM	2004
1WAZ, 2LJ2	*M. morganii*	Transporter (MerF)	2	46	C43(DE3)	pET31b	His	6	C	IB-gua + SDS	lsNMR-SDS ssNMR-DMPC	2005
1ZLL, 2M3B	*H. sapiens*	Phospholamban homopentamer	1 × 5	52	BL21(DE3)	pMALc2x	MBP		N	NS[5]	lsNMR-FC12 ssNMR-DOPC/DOPE	2005
2B2F	*A. fulgidus*	Transporter (Amt-1)	11 × 3	399	C43(DE3)	pET21a	His	6	C	DDM	Xray-LDAO	2005
2F2B	*M. marburgensis*	Aquaporin (AqpM)	6 × 4	246	BL21(DE3) CodonPlus	NS[5]	His	10	N	OG	Xray-OG	2005
2BBJ	*T. maritima*	Mg^{2+} transporter (CorA)	2 × 5	351	NS[5]	pET15b	His	6	N	DDM	Xray-DDM	2005
2A9H	*S. lividans*	Channel (KcsA)	2 × 4	155	C41(DE3)	pIVEX-2.4d	His	6	N	FC12	lsNMR-FC12	2006
3E86, 3K0D, 3OUF, 3T1C, 3E86, 2Q67, 2AHY	*B. cereus*	Channel (NaK)	2 × 4	'96–114	SG1309	pQE60	His	6	C	DM	Xray-DM	2006
2HN2, 2IUB	*T. maritima*	Transporter (CorA Mg^{2+})	2 × 5	351	BL21(DE3) CodonPlus	pET15b	His	6	N	DDM	Xray-DDM	2006
2H3O	*M. morganii*	Transporter (MerF HgII)	2	61	BL21(DE3) pLysS	pET31	His	6	C	IB-gua	ssNMR-14-O-PC/6-O-PC	2006
2HAC	*H. sapiens*	TCR-CD3, TM dimer complex	1 × 2	33	BL21(DE3)	pMM-LR6	His	9	N	IB-gua + TX100	lsNMR-SDS/FC12	2006
2NR9	*H. influenzae*	Intramembrane peptidase (GlpG)	6 × 2	196	TOP10	pBAD	His	6	C	DDM	Xray-C12E8	2006
2NWL	*P. horikoshii*	Aspartate transporter (GltPh)	8	422	Top10	pBAD24	His	8	NS[5]	DDM	Xray-DM	2007
2JO1	*H. sapiens*	Phospholemman (FXYD1)	1	72	C43(DE3)	pETBcl-XL	Bcl-XL + His	6	N	IB-gua	lsNMR-SDS	2007
2Q7M	*H. sapiens*	Lipoxygenase protein (FLAP)	4 × 3	161	BL21(DE3)	pET28a	His	6	C	DDM	Xray-MD[6]	2007
2QJU	*A. aeolicus*	Leucine transporter (LeuT)	12	511	BL21(DE3) pLysS	pBAD	His	NS[5]	NS[5]	DDM	Xray-OG	2007
2YVX, 2ZY9	*T. thermophilus*	Transporter (MgtE)	5 × 2	473	C41(DE3)	pET28a	His	6	N	DDM	Xray-DDM	2007
3B9W	*N. europaea*	Rh protein ammonia or CO_2 channel	11	407	GT1000 D (glnK, amtB)	pAD7; nitrogen promoter	His	6	C	OG	Xray-OG	2007
3B60	*S. typhimurium*	Flippase (MsbA)	6 × 2	582	BL21(DE3)	pET19b	His	10	N	UDM	Xray-UDM	2007
3BEH	*M. loti*	Cyclic nucleotide-regulated K^+ channel	6 × 4	355	JM83	pASK-IBA2	His	6	C	DM	Xray-DM	2008
2LJB, 2RLF	*Influenza A, B*	Channel (M2)	1 × 4	35	BL21(DE3)	pMM-LR6	His-trpLE	9	N	IB-gua + TX100	lsNMR-DHPC	2008
2VL0	*E. chrysanthemi*	Channel (pentameric ELIC)	4 × 5	321	BL21(DE3)	pET26b	MBP + His	10	N	UDM	Xray-UDM	2008
2VQG	*C. glutamicum*	Porin B (PorB)	1 × 5	99	BL21(DE3)	pGEX-3X	GST		N	None	Xray-C10E9, OG, NG, DM	2008
3DH4	*V. parahaemolyticus*	Na^+/galactose transporter (ySGLT)	14	530	XL1-Blue	pBAD	His	6	C	DM	Xray-DM	2008
2K1L	*H. sapiens*	Receptor tyrosine kinase (EphA1)	1	38	BL21(DE3) pLysS	pGEMEX1	TrxA-His	NS[5]	N	TX100	lsNMR-DMPC/DHPC	2008

(*continued on next page*)

Table 2 (*continued*)

PDB	Organism	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
3DIN	*T. maritima*	SecYEG protein in complex with SecA	10 (Y) + (E) + 2 (G)	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) + 1 (E) + 2 (G)	429 + 65 + 107	C43(DE3)	pET Duet-1, pCDF Duet-1	His	6	C	DM	Xray-Cymal6	2008
3F3A, 2QEI, 2A65, 3GJD, 3MPN, 3TT1, 3USG 3QS4, 4MM4	*A. aeolicus VF5*	Symporter (LeuT)	12 × 2	519	C41(DE3)	pET16b	His	8	C	DDM	Xray-OG	2008
2V50	*P. aeruginosa*	Transporter (MexB)	12 × 3	1052	C43(DE3)	pET28	His	6	C	DDM	Xray-DDM	2009
4DOJ, 2WIT	*C. glutamicum*	Glycine betaine transporter (BetP)	12 × 3	566	DH5a, BL21(DE3) CodonPlus	IBA7	Strep		N	DDM	Xray-Cymal5	2009
3DWW	*H. sapiens*	Prostaglandin E synthase 1	4 × 3	158	BL21(DE3) pLysS	pSP19T7LT	His	6	N	TX100	EM-TX100	2009
2K9Y	*H. sapiens*	Receptor tyrosine kinase (Eph2)	1	41	BL21(DE3) pLysS	pGEMEX1	TrxA-His	NS[5]	N	TX100	NMR-MeOH/ $CHCl_3$/water	2009
3GIA	*M. jannaschii*	Protein MJ0609 (ApcT)	12	444	C41(DE3)	pET3a-GFP	GFP + His	8	C	DDM	Xray-OTG	2009
2KNC	*H. sapiens*	Integrin αIIbβ3	1 + 1	54 and 79	BL21(DE3)	pMAL-C2	MBP-His	6	N	TX100	lsNMR- CD_3CN/H2O	2009
3K3F	*D. vulgaris*	Transporter (urea)	10 × 3	533	BL21(DE3)	pET-SUMO	SUMO + His	NS[5]	N	DDM	Xray-OM	2009
3IGA	*S. lividans*	Potassium channel (KcsA)	2 × 4	124	JM83	pASK90	His	6	N	DM	Xray-DM	2009
2KOG	*R. norvegicus*	Synaptobrevin	1	119	BL21 (DE3)/BL21(DE3) pRil	pET15b/28a	His	6	N	Sodium cholate	lsNMR-FC12	2009
3KLY	*V. cholerae*	Formate transporter (FocA)	6 × 5	280	C43(DE3)	pBAD	His	10	C	DDM	Xray-OG	2009
3KP9	*Synechococcus sp.*	Thioredoxin domain protein (VKORC1)	5	291	C43(DE3)	pET20b	His	6	C	DDM	Xray-DDM	2010
4NV5, 3KP9	*Synechococcus sp.*	Vitamin K epoxide reductase (VKOR)	5	291	BL21 (DE3)	NS[5]	His	6	C	DDM	Xray-DDM	2010
3M71	*H. influenzae*	Anion channel (SLAC1)	10 × 3	328	BL21(DE3) pLysS	pET	His	10	C	DDM	Xray-LDAO or OG	2010
2KS1, 2JWA	*H. sapiens*	ErbB1/ErbB2	1 × 2	44	BL21(DE3) pLysS	pGEM-EX1	TrxA-His	6	N	TX100	lsNMR-DHPC/DMPC	2010
3NCY	*S. enterica*	Antiporter (AdiC)	12 × 2	445	BL21(DE3)	pASK-IBA2	His	6	N	DM	Xray-DM	2010
2KPF	*H. sapiens*	Glycophorin A (GpA)	1 × 2	38	NF[7]	NF[7]	NF[7]				lsNMR-DHPC/DMPC	2010
3MKT	*V. cholerae*	MATE transporter (NorM)	12	461	BL21(DE3)	pET19b	His	10	N	DDM	Xray-DDM	2010
3MP7	*P. furiosus*	Primed channel (SecYEb)	10 (Y) + 1 (E)	482 + 61	BL21AI	pBAD	His	6	C	DDM + OG	Xray-OG	2010
3P5N	*S. aureus*	Transporter (RibU)	6 × 2	189	BL21(DE3)	pET15b	His	6	N	NG	Xray-NG	2010
2L35	*H. sapiens*	Signaling module (DAP12)	2 + 1	63 + 32	BL21(DE3)	pMM-LR6	His-trpLE	9	N	IB-Gua + TX100	lsNMR-FC14 + SDS	2010
2L0J, 2LY0	*Influenza A*	Channel (AM2)	1 × 4	44	BL21(DE3) pLysS	PET30-23d pMALc2x	*MBP or* His	*NS*[5]	C	DDM OG	*ssNMR-DOPC/DOPCE* lsNMR-FC12	*2010*
	G. violaceus	Channel (pentameric GLIC)	4 × 5	317	BL21(DE3)	pET26b	MBP + His	10	N	DDM	Xray-DDM	2010

(*continued on next page*)

Table 2 (*continued*)

PDB	Organism	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
2XQA, 3EHZ, 4HFI												
2XQ2	*V. parahaemolyticus*	Na^+/galactose transporter (ySGLT)	14	593	TOP10	pBAD	His	6	C	DM	Xray-Tri-DM	2010
2XUT	*S. oneidensis*	Transporter (PepTSt)	14	524	C43(DE3)	pWaldo-GFPe	GFP + His	8	C	DDM	Xray-DDM	2010
4A2N	*M. acetivorans*	Transferase (ICMT)	5	194	C41(DE3)	pTriEX/pOPIN	GFP + His	7	C	DDM	Xray-DDM	2012–01–11
3P50, 3P4W, 3EAM, 3IGQ, 4QH5, 5HCJ, 5L4E	*G. violaceus*	Ligand-gated ion channel (GLIC)	4 × 5	359	C43(DE3)	pET20b	MBP + His	6	N	DDM	Xray-DDM	2011
3PJZ	*V. para-hemolyticus*	Transporter (TrkH)	10 × 2	485	BL21(DE3)	pET31	His	10	C	DM	Xray-DM	2011
3ND0	*Synechocystis sp. pcc 6803*	H^+/Cl^- exchange transporter	18	466	420399	pASK	His	6	C	DM	Xray-DM	2011
3ODJ	*H. influenzae*	Peptidase (GlpG)	6	196	TOP10	pBAD	His	6	C	DDM	Xray-C12E8	2011
3QNQ	*B. cereus*	Transporter (ChbC EIIC)	10 × 2	433	BL21(DE3)	pET	His	10	C	DDM	Xray-NM	2011
2KYV	*H. sapiens*	Phospholamban homopentamer	1 × 5	52	BL21(DE3)	pET	MBP		N	TX100	lsNMR-FC12 ssNMR-DOPC/DOPE	2011
3AQP	*T. thermophilus*	(SecDF)	12	735	BL21(DE3) CodonPlus	pET26b	His	6	C	DDM	Xray-DDM	2011
2L9U	*H. sapiens*	Transmembrane domain (ErbB3)	1 × 2	40	Cell-free expression	pET22b	His	6	C	Cell-free expression-pellet	lsNMR-FC12	2011
3RFU	*L. pneumophila*	Copper efflux ATPase	8	736	C43(DE3)	pET22b	His	6	C	C12E8	Xray-MD[6]	2011
3RCE	*Campylobacter lari*	OST in complex (PglB)	13	724	BL21-Gold SCM6	pBAD	His	10	C	DDM	Xray-DDM	2011
4BBJ, 3RFU	*L. pneumophila*	Copper-transporting ATPase (LpCopA)	8	736	C43(DE3)	pET22b	His	NS[5]	NS[5]	C12E8	Xray-C12E8	2011
3S0X	*M. maripaludis*	Preflagellin aspartyl protease (FlaK)	6	237	C43(DE3)	pET28a/43b	His	6	C	FC12	Xray-Cymal6	2011
2LCK	*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 BL21(DE3) Tuner	pET27b/28b	His	7	C	OG or DDM	Xray-OG or DDM (LCP) lsNMR-DHPC	2011
3ZUY	*N. meningitidis*	Bacterial ASBT homologues	10	323	C43(DE3)	pWaldo-GFPe	GFP + His	8	C	DDM	Xray-LDAO	2011
3ZRS, 2WLJ, 4LP8	*M. magneto-tacticum*	Channel (KirBac3.1)	2 × 4	339	BL21(DE3) CodonPlus	pET30a	His	6	C	DM	Xray-HEGA10	2012
4A2N	*M. acetivorans*	Transferase (ICMT)	5	194	C41(DE3)	pTriEX/pOPIN	GFP + His	7	C	DDM	Xray-DDM	2012
3AYF	*G. stearo-thermophilus*	Nitric oxide reductase (qNOR)	14	800	Rosetta2 (DE3) BL21(DE3) CodonPlus	pET22b	His	6	C	TX100	Xray-OG	2012
3V5U	*M. janaschii*	Exchanger (Na^+/Ca^+)	9	320	BL21(DE3) pLysS	pQE60	His	10	C	DDM	Xray-DDM (LCP)	2012
3TIJ	*V. cholerae*	Nupc family protein	8 × 3	424	C41(DE3)	pET26	MBP + His	10	N	DDM	Xray-DM	2012
3RQW	*E. chrysanthemi*	Ligand-gated ion channel with acetylcholine (ELIC)	4 × 5	322	Rosetta™ 2(DE3) pLysS	pET26b	His	10	N	UDM	Xray-DDM	2012
3TDO	*C. difficile*	Hydrosulfide ion channel (FNT3)	6 × 5	268	BL21(DE3) pLysS	pBAD	His	10	C	DDM	Xray-OG	2012
3QF4	*T. maritima*	Heterodimeric ABC exporter	6 × 2	587	C43(DE3) or MC1061	pBAD24	His	10	N	DDM	X-ray-DDM	2012
3VMT	*S. aureus*	Glycosyltransferase	1	263	BL21(DE3) CodonPlus	pET15b	His	6	N	DM	Xray-DM	2012
3VOU	*B. weihen-stephanensis (NaK) and S. pontiacus (NaV)*	NaK chimera with NaV	2 × 4	148	KRX	pET21b	His	4 × 6	Oth[8]	DM	Xray-DM	2012
2LCX	*H. sapiens*	ErbB4	1 × 2	44	BL21(DE3) pLysS	pGEMEX1	TrxA-His	6	N	TX100		2012

(*continued on next page*)

Table 2 (*continued*)

PDB	Organism	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
											lsNMR-DHPC/DMPC	
4DXW	*A. proteo-bacterium 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
4F8H	*G. violaceus*	Ligand-gated ion channel with ketamine (GLIC)	4 × 5	359	Rosetta™ 2(DE3) pLysS	pET26b	MBP + His	10	N	DDM	Xray-TDM	2012
4F4L	*M. marinus*	Voltage-gated sodium channel pore	2 × 4	112	C41(DE3)	pET15b	His	6	N	DDM	Xray-MD[6]	2012
2LNL	*H. sapiens*	Receptor (CXCR1)	7	309	BL21(DE3)	pGEX2a	GST + His	6	C	IB-SDS	ssNMR-DMPC	2012
4F35	*V. cholerae*	Symporter	11 × 2	462	BL21(DE3)	pET	His	10	N	DM	Xray-MD[6]	2012
4EV6	*M. jannaschii*	Transporter (CorA Mg^{2+})	2 × 5	317	Rosetta™ 2(DE3)	pNIC28 (pET28)	His	6	N	UDM	Xray-UDM	2012
4H33	*L. monocytogenes*	Channel (KyLm)	2 × 4	137	XL1-Blue	pQE70	His	6	C	DM	Xray-DM (LCP)	2012
4EEB	*T. maritima*	Mg^{2+} transporter (CorA)	2 × 5	351	BL21(DE3) CodonPlus	pET15b	His	6	N	DDM	Xray-DDM	2012
4B4A	*A. aeolicus*	Twin-arginine translocase (TatC)	6	240	Lemo56(DE3)	pWaldo-GFPe	GFP + His	8	C	LMNG	Xray-LMNG	2012
3UX4	*H. pylori*	Channel (urea)	6 × 6	195	C43(DE3)	pET	His	6	Oth[8]	DM	Xray-MD[6]	2012
4HYG	*M. marisnigri*	Presenilin (PSH)	9 × 4	301	BL21(DE3)	pET21b	His	8	N	NM	Xray-MD[6]	2012
4HG6	*R. sphaeroides*	Cellulose synthase (BcsA-BcsB)	8 + 1	802 + 707	Rosetta™ 2(DE3)	pETDuet	His	12	N	TX100	Xray-LDAO	2012
4G1U	*Y. pestis*	Transporter (HmuUV)	10 × 2	334	BL21(DE3) Gold	pET19b	His	10	N	DDM	Xray-DDM	2012
4GX0	*G. sulfurreducens*	Channel (GsuK)	2 × 4	565	BL21(DE3)	pQE70	His	NS[5]	C	DM	Xray-DM	2012
2LOU	*H. sapiens*	Apelin receptor	1	64	BL21(DE3)	pEXP5-CT	His	6	C	IB- acetonitrile /trifluoroacetic acid	lsNMR-FC12	2013
3ZKR	*E. chrysanthemi*	Ligand-gated ion channel (ELIC)	4 × 5	307	C43(DE3)	pET11a	MBP		N	UDM	Xray-UDM	2013
4HUK	*N. gonorrhoeae*	MATE transporter (NorM)	12	459	BL21(DE3)	pET15b	His	6	N	DDM	Xray-DDM	2013
3VVN	*P. furiosus*	Multidrug and toxin compound extrusion (MATE)	12	461	C41(DE3) ΔacrB	pET11a	His	6	C	DDM	Xray-Cymal6 (LCP)	2013
4J9U	*V. parahaemolyticus*	Potassium ion transporter (TrkH)	10 × 2	485	BL21(DE3)	pET31b	His	NS[5]	C	DM	Xray-DM	2013
4HZU	*L. brevis*	ECF transporter complex	2 + 5	166 + 266	C43(DE3)	pRSF-Duet-1	His	12	N	DDM	Xray-DM	2013
4HUQ	*L. brevis*	Folate ECF transporter (FolT + EcfT)	2 + 5	174/280	BL21(DE3)	PET-Duet	His	6	N	DDM	Xray-DM	2013
4J7C	*B. subtilis*	Potassium ion transporter (KtrAB)	8 × 2	465 + 222	BL21(DE3)	pET24d	His	NS[5]	N	DDM	Xray-Cymal6	2013
4HTS	*A. aeolicus*	Twin-arginine translocase (TatC)	6	236	BL21(DE3)	pET33b	His	6	N	DDM	Xray-DHPC or DDM	2013
4JQ6	Isolated from the Mediterranean Sea	Proteorhodopsin (blue-light absorbing)	7 × 6	235	C43(DE3)	pET28	His	NS[5]	NS[5]	DM	Xray-DM (bicelles)	2013
4KLY	*Gammaproteo-bacterium*	Proteorhodopsins (PRs),	7	259	C43(DE3)	pET28a	His	6	C	DM	Xray-DM	2013
2M6X	Hepatitis C virus	p7 hexamer channels	2 × 6	63	BL21(DE3)	pMM- LR6	His-trpLE	9	N	IB-gua + TX100	lsNMR-FC12	2013
4KPP	*A. fulgidus*	Ca^{2+}/H^+ antiporter (CaX)	12	405	C41(DE3)	pET	GFP + His	8	C	DDM	Xray-DMNG (LCP)	2013
4KJS	*B. subtilis*	Ca^{2+}/H^+ antiporter (YfkE)	11 × 3	351	BL21(DE3)	pET22b	His	6	Oth[8]	DDM	Xray-DDM	2013
3W9J	*P. aeruginosa*	Multi-drug efflux transporter (MexB)	12 × 3	1052	MG1655	pUCP20-BHis	His	10	C	DDM	Xray-DDM	2013
4IKV	*G. kaustophilus*	Proton-dependent oligopeptide transporter (POT)	12	507	C41(DE3) ΔacrB	pCGFP-BC	GFP + His	8	C	DDM	Xray-DDM (LCP)	2013
4HYJ	*E. sibiricum*	Proteorhodopsin	7	258	Rosetta™ 2(DE3) pLysS	pET32a	His	6	C	DDM	Xray-DDM (LCP)	2013
2M3G	*Anabaena sp. PCC7120*	Sensory rhodopsin	7x3	235	BL21(DE3) CodonPlus		His	6	C	DDM	ssNMR-DMPC/DMPA	2013
4BWZ	*T. thermophilus*	Na^+/H^+ antiporter (NapA)	12 × 2	394	Lemo21(DE3)	pWaldo-GFPe	GFP + His	8	C	DDM	Xray-NM	2013

(*continued on next page*)

Table 2 (*continued*)

PDB	Organism	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
2M6B	*H. sapiens*	Glycine receptor (hGlyR-α1)	4 × 5	150	BL21(DE3) pLysS	pET31b	His	6	C	IB-NS[5]	lsNMR-LPPG	2013
4J72, 5CKR	*A. aeolicus*	MurNAc-pentapeptide translocase (MraY)	10	365	C41(DE3)	NS[5]	MBP + His	10	N	DDM	Xray-DM	2013
4KY0	*T. kodakarensis*	Aspartate transporter	8	431	MC1061	pBAD24	His	8	C	DM	Xray-OG	2013
3ZJZ, 4CBC, 5BZB	*M. marinus*	Sodium channels	4	149	C41(DE3)	pET15b	His	NS[5]	NS[5]	DDM	Xray-HEGA10	2013
2LZL	*H. sapiens*	Growth factor receptor 3 (FGFR3)	1 × 2	43	BL21(DE3) pLysS	pGEMEX1	TrxA-His	6	N	TX100	lsNMR-FC12/SDS	2013
4M8J	*P. mirabilis*	Carnitine 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[5]	N	DDM	Xray-DDM	2013
4K0J	*C. metallidurans*	ZneA Zn(II)/proton antiporter	12	1045	C43(DE3)	pET30b	His	6	C	DDM	Xray-DDM	2013
4LTO	*A. ehrlichii*	Voltage-gated sodium channel (NaVAe1p)	2 × 4	152	C41(DE3)	pET24	MBP + His	6	N	DDM	Xray-DDM	2013
4LZ6, 5C6N	*B. halodurans*	MATE transporter (DinF-BH)	12	446	BL21(DE3) ΔacrABΔmacABΔyojHI	pET15b	His	6	C	DDM	Xray-DDM	2013
3WAJ	*A. fulgidus*	OST (AglB)	13	875	C43(DE3)	pET52b	His	10	C	DDM	Xray-MD[6]	2013
4C7R	*C. glutamicum*	Glycine betaine transporter (BetP)	12 × 3	566	DH5α	pIBA7	Strep		N	DDM	Xray-Cymal5	2013
2MAW	*H. sapiens*	Neuronal acetylcholine receptor	4	137	Rosetta™ 2(DE3) pLysS	pMCSG7	His	6	N	NS[5]	lsNMR-LDAO	2013
2M8R, 3HD7	*R. norvegicus*	Syntaxin 1A, TM & syntaxin complex	1 + 1	109 + 91	BL21(DE3)	pET28a	His	6	N	TX100 or OG	lsNMR-FC12, Xray-NG or C7G	2013
4CAD	*M. maripaludis*	Protease Rce1 (CAAX)	8	271	C41(DE3)	pTriEX	His	7	C	UDM	Xray-MD[6]	2013
4N7W	*Y. frederiksenii*	Bile acid symporter	10	307	BL21(DE3) Gold	pET	His	10	N	DDM	DM	2013
4M64	*S. typhimurium*	Na^+/melibiose symporter (MelB)	12	476	DW2	pK95 native	His	10	C	UDM	Xray-MD[6]	2014
4BUO	*R. norvegicus*	Neurotensin receptor (NTS1)	7	335	BL21 Tuner	pBR322	MBP + His	6	N	MD[6]	Xray-MD[6]	2014
4O6Y	*A. thaliana*	Cytochrome B561 (Cyt b_{561}-B)	6	230	BL21(DE3)	pET15b	His	6	NS[5]	DM	Xray-NG	2014
4O9Y	*P. luminescens*	Tc toxin (TcA)	2 × 5	2516	BL21(DE3) CodonPlus	pET28a	His	6	N	None	None	2014
2M59	*H. sapiens*	Endothelial growth factor receptor 2	1 × 2	37	Cell-free expression	pET20b	HA		N	Cell-free expression pellet-trifluoroethanol// H_2O/trifluoroacetic acid	lsNMR-FC12	2014
4MRN	*N. aromaticivorans*	ABC transporter (NaAtm1)	6 × 2	614	BL21(DE3)	pJL-H6/pET21a	His	6	C	MD[6]	Xray-MD[6]	2014
4MYC	*S. cerevisiae*	ABC transporter (Atm1)	6 × 2	598	NS[5]	pASK-IBA1	Strep		C	DDM	Xray-DDM	2014
2MFR	*H. sapiens*	Insulin receptor (AAs 940–980)	1	57	BL21(DE3)	pET29b	His	6	C	IB-urea	lsNMR-FC12	2014
2MGY	*M. musculus*	Translocator protein (TSPO)	5	169	BL21(DE3)	pET15b	His	6	N	IB-SDS	lsNMR-FC12	2014
3WO6	*B. halodurans*	Insertase (YidC)	5	267	C41(DE3)	pET	His or GFP + His	8	N/C	DDM	Xray-DDM (LCP)	2014
4O6M	*A. fulgidus*	Alcohol phosphotransferase (AF2299)	6	372	BL21(DE3) pLysS	pMCSG7	His	10	N	DM	Xray-DM (LCP)	2014
4PGR	*B. subtilis*	pH-sensitive channel (YetJ)	7	214	BL21(DE3) pLysS	pET or pMCSG	His	10	N	DDM	Xray-LDAO or C10E5	2014
4X2S, 4P19	*P. horikoshii*	Aspartate transporter (GltPh)	8	422	DH10B	pBAD24	His	8	C	DDM	Xray-DM	2014
4MND	*A. fulgidus*	CDP-alcohol phosphatidyltransferase	6	479	C43(DE3)	pET52b	His	10	C	TX100	Xray-TX100 (LCP)	2014
4TPH	*S. oneidensis*	Oligopeptide-proton symporter (PepTSo)	14	516	C41(DE3)	pNIC-CTHF(pET)	His	6	C	DDM	Xray-DDM	2014
4Q4H	*T. maritima*	ABC exporter	6 + 6	587 + 598	C43(DE3) MC1061	*pBAD24*	His	10	N	DDM	Xray-DDM	2014
4TQ3	*A. fulgidus*	UbiA homolog	9	303	BL21(DE3)	pET	His	NS[5]	N	DM	Xray-OG (LCP)	2014
4MT1	*N. gonorrhoeae*	Multidrug efflux pump (MtrD)	12	1056	C43(DE3) ΔacrB	pET15bΩ*mtrD*	His	6	C	Cymal6	Xray-Cymal6	2014
4QNC, 5UHQ	*L. biflexa*	SemiSWEET	3	85	BL21(DE3)	pJexpress411	His	10	C	DDM	Xray-DDM	2014
4QND	*Vibrio sp. n418*		3 × 2	97	BL21(DE3)	pJexpress411	His	10	C	DDM	Xray-DDM	2014

(*continued on next page*)

Table 2 (*continued*)

PDB	Organism	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
		SemiSWEET in outward-open state										
4TWD	*D. chrysanthemi*	Ligand-gated ion channel with Br-memantine (ELIC)	4 × 5	307	C43(DE3)	pET11a	MBP		N	UDM	Xray-UDM	2014
4WD7	*K. pneumoniae*	Bestrophin homolog (KpBest)	4 × 5	301	BL21(DE3) pLysS	pMCSG7	His	10	N	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-OG	2014
4QUV	*M. alcaliphilum*	Δ^{14}-sterol reductase (MaSR1)	10	427	C43(DE3)	pET21b	His	8	N	DDM	Xray-NG	2014
4WGV	*S. capitis*	Transition-metal ion transporter (NRAMP)	11	415	MC1061	pBXC3H	His	10	C	DM	Xray-NM	2014
4RNG	*T. yellowstonii*	SemiSWEET in occluded state	3 × 2	88	BL21(DE3)	pET21b	His	6	C	DDM	Xray-DDM	2014
4U9L	*T. thermophilus*	Mg^{2+} transporter (MgtE)	5 × 2	179–450	C41(DE3)	pET	His	6	N	DDM	Xray-DDM	2014
4CZB	*M. jannaschii*	Sodium proton antiporter (Mjnhap1)	13	426	BL21(DE3)	pET21a	None		C	FC12	Xray-Cymal5	2014
2MMU	*M. tuberculosis*	Cell division protein (CrgA)	2	101	BL21(DE3) CodonPlus	pET29b	His	6	C	Empigen	ssNMR-POPG/POPC	2014
2MIC	*R. norvegicus*	Neurotrophin receptor (p75)	1 × 2	41	Cell-free expression	NS[5]	NT[9]			Cell-free expression pellet-sarkosyl	lsNMR-FC12	2014
4O93	*T. thermophilus*	Nicotinamide nucleotide transhydrogenase	3 + 9	94 + 270	BL21(DE3)	pET21a	His	6	N or C	DDM	Xray-DDM (LCP)	2014
4RYQ	*B. cereus*	Translocator protein (TSPO)	5	181	BL21(DE3)	pMCSG7	His	10	N	DDM	Xray-DDM (LCP)	2015
4UC1	*R. sphaeroides*	Translocator protein (TSPO)	5	158	BL21(DE3)	pRK415 (lac)	His	NS[5]	NS[5]	DDM	Xray-DM (LCP)	2015
4TSY	*A. fragacea*	Haemolytic fragaceatoxin C (FraC)	1 × 8	179	BL21(DE3)	pBAT-4	NT[9]			None	Xray-DDM	2015
4XTL	*D. eikasta*	Rhodopsin 2 (KR2)	7	288	SE1	pSCodon1.2	His	6	C	DDM	Xray-DDM	2015
3X3B	*D. eikasta*	Rhodopsin 2 (KR2)	7	290	C41(DE3)	pET	His	6	C	DDM	Xray-DDM	2015
4YMS	*C. subterraneus*	Amino acid ABC transporter (Art (QN)2)	5 × 2	240 + 220	C43(DE3)	pACYCDuet	His	6	C	DDM	Xray-CHAPSO	2015
4R0C	*A. borkumensis*	Transporter (YdaH)	9	492	BL21(DE3) ΔacrB	pET15bΩ*ydaH*	His	6	N	DDM	Xray-MD[6]	2015
4WOL	*H. sapiens*	Signaling module (DAP12)	1 × 3	33	BL21(DE3)	pMM	His-trpLE	9	N	IB-gua + TX100	Xray-None (LCP)	2015
4YCR	*H. influenzae*	Anion channel (SLAC1)	10	328	C43(DE3)	pWaldoGFPe	His	8	C	NF[7]	Xray-OG	2015
4W6V	*Y. enterocolitica*	Dipeptide transporter (PepT)	14	519	BL21(DE3) pLysS	pZUDF21	His	10	C	DDM	Xray-DDM	2015
4RY2	*R. thermocellum*	ABC transporter (PCAT1)	6 × 2	730	BL21(DE3) CodonPlus	pMCSG7	GST		N	DDM	Xray-CHAPSO	2015
4TQU	*Sphingomonas sp.*	ABC transporter	6 (M1) + 6 (M2)	301 + 305 + 2x363 + 516	BL21(DE3) Gold pLysS	pET21b	His	10	C	DDM	Xray-MD[6] or Cymal 6	2015
4QI1	*H. walsbyi*	Bacteriorhodopsin-I	7	262	C43(DE3)	pET21d	His	6	C	DDM	Xray-DM	2015
5C78	*C. jejuni*	Flippase (PglK)	6 × 2	564	BL21(DE3) Gold	pET19b	His	10	N	DDM	Xray-DDM	2015
4Y7K	*M. acetivorans*	Mechanosensitive channel of large conductance (MscL)	2 × 5	275	C41(DE3)	pET15b	His	6	N	TX100	Xray-MD[6]	2015
5A40	*B. pertussis*	F^- ion channel homologue (Fluc)	4 × 2	128	BL21(DE3)	pASK-IBA2	His	6	C	DM	Xray-DM (LCP)	2015
5C8J	*M. jannaschii*	YidC-like protein (MJ0480)	4	198	BL21(DE3)	pET28a	His	6	C	DDM	Xray-DDM	2015
5DA0	*D. geothermalis*	Proton-coupled fumarate symporter (SLC26)	14	499	MC1061	pBXC3GH	GFP + His	10	C	DM	Xray-DM or NM	2015
5C6P	*N. gonorrhoeae*	MATE transporter (NorM)	12	459	BL21(DE3)	pET15b	His	10	C	DDM	Xray-DDM	2015
4XU4	*M. vanbaalenii*	Insig homologue (MvINS)	6 × 3	210	BL21(DE3)	pET21b	His	6	C	LDAO	Xray-OG	2015
5AYN	*B. bacteriovarus*	Fe^{2+} transporter (SLC40A1)	12	432	C41(DE3)	pWaldo-GFPe	GFP + His	8	C	DDM	Xray-LMNG (LCP)	2015
5D91	*R. Salmoninarum*	Phosphatidylinositolphosphate synthase	6	336 + Af2299	BL21(DE3) pLysS	pMCSG7	His	10	N	DDM	Xray-DDM	2015

(*continued on next page*)

Table 2 (*continued*)

PDB	Organism	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
5AWW	*T. thermophilus*	Translocon (SecYEG)	10 + 1 + 2	438 + 60 + 117	BL21(DE3)	pACYC-Duet	His	6	C	DDM	Xray-DDM (LCP)	2015
5EH4	*H. sapiens*	Glycophorin A (GpA)	1 × 2	30	BL21(DE3)	pMM	His-trpLE	9	N	IB-gua + TX100	Xray-None (LCP)	2015
5EKP	*Synechocystis sp. PCC6803*	Polyisoprenyl-glycosyltransferase (GtrB)	2 × 4	318	BL21(DE3) pLysS	pMCSG7	His	6	N	DM	Xray-DM	2016
5BZ3	*T. thermophilus*	Sodium/proton antiporters (NapA)	12	385	Lemo21 (DE3)	pWaldo	GFP + His	8	C	DDM	Xray-NM (LCP)	2016
5A1S	*S. enterica*	Citrate symporter CitS	11	448	C41(DE3)	pET21d	His	10	N	DM	Xray-MD[6]	2016
5EZM	*C. metallidurans*	Glycosyltransferase (ArnT)	13	575	BL21(DE3) pLysS	pNYCOMPS	His	10	N	DDM	Xray-DDM (LCP)	2016
3JCF	*T. maritima*	Mg^{2+} transporter (CorA)	2 × 5	351	BL21(DE3) pLysS	pET15b	His	6	N	DDM	EM-DDM	2016
2N2A	*H. sapiens*	Receptor tyrosine kinases (HER or ErbB)	1 × 2	58	Cell-free expression	pGEMEX-1	NT[9]			Cell-free expression pellet-sarkosyl	lsNMR-FC12	2016
5DIR	*P. aeruginosa*	Lipoprotein signal peptidase II	4	169	C41(DE3)	pET28a	His	6	N	FC12	Xray-FC12 (LCP)	2016
5EUL	*G. thermo-denitrificans/B. subtilis*	Translocon (SecYE) with SecAYE	1 + 10 + 1	841 + 430 + 70	EP51	pTet SecYE, pBAD SecA	His GFP + Strep	*8*	C	DDM	Xray-DDM	2016
5EC1	*S. lividans*	Potassium channel mutant (KcsA)	2 × 4	160	BL21(DE3)	NS[5]	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 nicastrin	1	54	NS[5]	pET29b	His	6	N	IB-urea	lsNMR-SDS or FC12	2016
5ID3	*C. elegans*	Mitochondrial calcium uniporter	2 × 5	159	BL21(DE3)	pET21a	His	6	C	FC14	lsNMR-FC14	2016
5JSZ	*L. delbrueckii*	Folate ECF transporter (FolT)	6	265	MC1061	p2BAD	His	10	N	DDM	Xray-DDM	2016
2N4X	*A. fulgidus*	Electron transporter (Ccd(A))	6	208	C43(DE3)-SEN212	pET28	His	6	C	FC12	lsNMR-FC12	2016
5IWS	*B. cereus*	EIIC maltose transporter MalT	10	545	NS[5]	pMCSG28	His	6	C	DDM	Xray-DDM	2016
5I20	*S. novella*	Aromatic amino acid exporter (YddG)	10	287	Rosetta™ 2(DE3)	pET	His	8	C	DDM	Xray-DDM (LCP)	2016
5JAE	*A. aeolicus*	(LeuT)	12 × 2	519	C41(DE3)	pET16b	His	6	C	DDM	Xray-OG	2016
5JYN	Human immunodeficiency virus 1	HIV-1 envelope spike (Env)	1 × 3	40	BL21(DE3)	pMM-LR6	His-trpLE	9	N	IB-gua	NMR-DMPC/DHPC	2016
5G28	*N. marinus*	Chloride-pumping rhodopsin (ClR)	7	275	BL21(DE3) CodonPlus	pET21b	His	6	C	DDM	Xray-DDM	2016
5B57	*B. cenocepacia*	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-MD[6]	2016
5T77	*T. africanus*	MOP flippase (MurJ)	14	475	C41(DE3)	pET26	MBP + His and His	10	C	DDM	Xray-DMNG (LCP)	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 (LCP)	2017
5FGN	*N. meningitidis*	Lipid A transferase	5	544	BL21(DE3) pLysS	pTrc99A	His	6	C	DDM	Xray-MD[6]	2017
5KTF	*M. musculus*	high density lipoprotein (HDL)	2	73	BL21(DE3) CodonPlus	pQE30	His	8	N	Empigen	lsNMR-LPPG	2017
5L22	*A. aeolicus*	ABC transporter (AaPrtD)	6 × 2	572	NS[5]	NS[5]	His	6	C	NS[5]	NS[5]	2017
5X5Y	*P. aeruginosa*	ABC transporter (LptB2FG)	6 (F) + 6 (G)	2 × 247 + 362 + 355	C43(DE3)	pQLink	His	6	C	DDM	Xray-MD[6]	2017
5V4S	*L. licerasiae*	Cyclic nucleotide-gated channel (LliK)	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) NF[7]	NF[7]	NF[7]			NF[7]	NF[7]	2017
5XAM	*D. radiodurans*	Protein-export enhancer (SecDF)	10	740	BL21(DE3)	pTV118N (lac)	His	8	C	DDM	Xray-DDM	2017
5KHN	*B. multivorans*	Hopanoid transporter (HpnN)	12	877	BL21(DE3)	pET15bΩ*hpnN*	His	6	C	DDM	Xray-DDM	2017

(*continued on next page*)

Table 2 (*continued*)

PDB	Organism	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
5WUC	*S. acidocaldarius*	Trimeric intracellular cation (TRIC)	7 × 3	237	BL21(DE3) pLysS	pET	His	10	C	DDM	Xray-DM	2017
5VRE	*C. minutus*	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) pLysS	pET	His	10	C	DDM	Xray-DM	2017
5T0O	*C. jejuni*	Multi-drug efflux transporter (CmeB)	12	1040	BL21(DE3)	pET15bΩcmeB	His	6	N	Cymal6	Xray-Cymal6	2017
5GKO	*A. baumannii*	ABC transporter (MacB)	4x2	671	BL21(DE3)	pET22	His	6	N	UDM	Xray-UDM	2017
5LIL	*A. actinomy-cetemcomitans*	ABC transporter (MacB)	4x2	664	C43(DE3)	pET28	His	6	N	LMNG	Xray-LMNG	2017
Beta-barrel												
1PFO	*C. perfringens*	Perfringolysin-O	4	500	JM109	pRT10	His	6	N	None	None	1998
1H6S, 1BH3	*R. blastica*	Porin (E1M/A116K)	16 × 3	307	BL21(DE3) pLysS	pET3b	NT[9]			LDAO	Xray-C8E4	1998
1UYN	*H. influenzae*	Autotransporter (trimeric Hia)	12	308	BL21(DE3)	pET11a	NT[9]			IB-urea	Xray-C10E5	2004
3O44, 1XEZ	*V. cholerae*	Pore-forming toxin (Cytosolin)	14	741	Origami B	pHis-parallel2	His	6	N	C10E6, None	Xray-C10E6 or OG	2005
2GR8	*H. influenzae*	Trimeric autotransporter (Hia)	4 × 3	99	B834	pASK-IBA12	Strep		N	Elugent	Xray-C8E4	2006
2ODJ	*P. aeruginosa*	Channel (OprD)	18	443	C43(DE3)	pBAD22	His	6	C	MD[6]	Xray-C8E4	2007
2VDF	*N. meningitidis*	OM adhesin (OpcA)	10	253	AR58	pMG1	NS[5]			LDAO	Xray-C10E5	2007
2QTK	*P. aeruginosa*	Benzoate channel (OpdK)	18	390	C43(DE3)	pBAD22	His	8	C	LDAO	Xray-C8E4	2008
2K4T	*H. sapiens*	Anion channel (VDAC-1)	19	291	BL21(DE3)	pET21a	His	6	C	IB-urea	lsNMR-LDAO	2008
2JK4	*H. sapiens*	Voltage-dependent anion channel (VDAC)	19	294	M15 NS[5]	PDS56	His	6	C	IB-gua	Xray-Cymal5	2008
3EMN,4C69	*M. musculus*	Anion channel (VDAC-1)	19	295	M15	pQE9	His	6	N	IB-gua	Xray-LDAO (bicelles)	2008
3DWO	*P. aeruginosa*	Fatty acid transporter (FadL)	14	451	C43(DE3)	pBAD22	His	6	C	LDAO	Xray-C8E4	2008
2K0L	*K. pneumoniae*	OuterMP (OmpA)	8	216	BL21(DE3)	pET21c	His	6	C	IB-gua	lsNMR-DHPC	2008
3CSL	*S. marcescens*	Heme receptor complex (HasR)	22	865	MC4100	pFR2	NT[9]			SB3-14	Xray-C8E4	2009
3FID	*S. typhimurium*	Lipid A deacylase (LpxR)	12	319	BL21(DE3) Star	pET21a	NT[9]			IB-urea	Xray-C10E5	2009
3EFM	*B. pertussis*	OM transporter (FauA)	22	707	BL21(DE3)	pET20b	His	6	Oth[8]	OPOE	Xray-C8E4 or C8E5	2009
3D5K	*P. aeruginosa*	Outer MP (OprM)	4 × 3	485	C43(DE3)	pB22	His	8	C	LDAO	Xray-C8E4	2009
3FHH	*S. dysenteriae*	Heme/hemoglobin OM receptor (ShuA)	22	640	BL21(DE3)	pET20b	His	6	Oth[8]	OPOE	Xray-C8E4 or C8E5	2009
3KVN	*P. aeruginosa*	Autotransporter (EstA)	12	646	C43(DE3) + BL21(DE3) Star	pB22	His	7	N	LDAO	Xray-C8E4	2010
2X55	*Y. pestis*	OM protease (omptin)	10	293	C43(DE3)	pB22	His	6	C	LDAO	Xray-C8E4	2010
4QKY, 3NJT	*B. pertussis*	Transporter FhaC	16	554	BL21(DE3) omp5	pET24d	His	6	N	OG	Xray-OG	2010
2X27	*P. aeruginosa*	Outer MP G (OprG)	8	233	C43(DE3)	pBAD	His	6	C	MD[6]	Xray-C8E4	2010
3ANZ	*S. aureus*	α-hemolysin	2 × 7	302	B834(DE3)	pET28b	His	6	C	None	None	2011
3NSG	*S. typhi*	Porin (OmpF)	16	341	GJ1158	pET20b	NT[9]			IB-urea	Xray-LDAO	2011
3RBH	*P. aeruginosa*	Export protein (AlgE)	18	479	BL21(DE3) CodonPlus	pET28a	His	6	N	DM	Xray-OTOE	2011
2LHF	*P. aeruginosa*	Outer MP H (OprH)	8	179	BL21(DE3)	pET30a	His	6	C	IB-urea	lsNMR-DHPC	2011
3B07, 4P1Y, 4P1X	*S. aureus*	γ-hemolysin: (LukF) and (Hlg2)	16	309 + 290	B834(DE3)	pET26 + pRAREII	His	6	N	None	None	2011
3QRA	*Y. pestis*	Adhesion protein (Ail)	8	157	BL21(DE3)	pET16b	NT[9]			IB-gua	Xray-C8E4	2011
3SY7, 3SY9, 3SYB	*P. aeruginosa*	Channel (OccD1, OprD)	18	443	BL21(DE3)	pB22	His	6	N	MD[6]	Xray-C8E4	2012
3SZV, 3T0S, 3T20, 3T24	*P. aeruginosa*	Aromatic hydrocarbon (OccK3, OpdO)	18	401	BL21(DE3)	pB22	His	6	N	MD[6]	Xray-C8E4	2012
3SZD	*P. aeruginosa*	Channel (OccK2, OpdF)	18	405	BL21(DE3)	pB22	His	6	N	MD[6]	Xray-C8E4	2012

(*continued on next page*)

Table 2 (*continued*)

PDB	Organism	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
3SY9, 3SYS	*P. aeruginosa*	OM carboxylate channels (Occ)	18	430	BL21(DE3)	pB22	His	6	N	MD[6]	Xray-C8E4	2012
3V8X	*N. meningitidis*	Transferrin (TbpA)	22	915	BL21(DE3)	pET20b	His	10	N	Elugent	Xray-C8E4	2012
4E1T	*Y. pseudo-tuberculosis*	Invasin beta-domain	12	245	BL21(DE3)	pET9	His	10	N	Elugent	Xray-MD[6] (LCP)	2012
4GEY	*P. putida*	Carbohydrate transporter (OprB)	16	436	BL21(DE3)	pBAD22	His	7	N	Elugent	Xray-C8E4	2012
2LME	*Y. entero-colitica*	Autotransporter (trimeric YadA)	12	105	BL21(DE3)	pIBAYadM	NT[9]			OPOE	ssNMR-OPOE (micro-crystals)	*2012*
3VZT	*N. meningitidis*	OM protein (PorB)	16	355	BL21(DE3)	pET21b	His	6	C	IB-urea	Xray-TDDG	2012
4AFK, 4XNL	*P. aeruginosa*	Export protein (AlgE)	18	458	BL21(DE3)	pET200	NS[5]			OG	Xray-LDAO (LCP)	2013
4K3C	*H. ducreyi*	β-barrel assembly machinery (BamA)	16	532	BL21(DE3)	pET20b	His	10	NS[5]	Elugent	Xray-C8E4 (bicelles)	2013
4HSC	*S. pyogenes*	Streptolysin O pore-forming toxin	4	571	XL1-Blue	NF[7]	NF[7]			NF[7]	NF[7]	2013
4BUM	*D. rerio*	Voltage-dependent anion channel 2	19	289	M15	pQE60	His	6	C	IB-gua	Xray-LDAO	2014
4PR7	*D. dadantii*	OM porin (KdgM)	12	222	BL21(DE3) omp8 pLysS	pKSM717	His	6	C	OPOE	Xray-C8E4	2014
4N4R	*S. enterica*	LPS transport proteins (LptD–LptE complex)	26	786(LptD) + 196 (LptE)	C43(DE3)	pET28b and pACYCDuet-1	His	6	C	SB3-14	Xray-OG	2014
4Q35	*S. flexneri*	Lipopolysaccharide transport (Lpt)	26	802 (LptD) + 175(LptE)	BL21(DE3)	pBAD22	His	6	C	LDAO	Xray-C8E4	2014
4MT0	*N. gonorrhoeae*	OM multidrug efflux pump (MtrE)	4	467	C43(DE3)	pBAD22bΩ*mtrE*	His	6	C	DDM	Xray-MD[6]	2014
3J9C	*B. anthracis*	Protective antigen (PA-63)	14	562	BL21(DE3)	pET22b-PA	NT[9]			None	EM-Igepal	2015
4MKO	*P. entomophila*	β-barrel pore-forming toxins (β-PFT),	2	236	Rosetta™ 2(DE3) pLysS	pETG-20A	TrxA-His	6	N	None	None	2015
4RL9	*A. baumannii*	Carbapenem-associated OM protein	8	255	BL21(DE3) and C43(DE3)	pET15, pB22	His	6	N or C	IB-urea	Xray-C8E4	2015
4V3G	*K. oxytoca*	OM protein (CymA)	14	339	C43(DE3) ΔcyoABCD	pB22	His	7	N	Elugent	Xray-C8E4	2015
2N2M	*Y pestis*	Attachment invasion locus (Ail)	8	156	NF[7]	NF[7]	NF[7]	NF[7]	NF[7]	NS[5]	lsNMR-DePC	2015
4RL8	*P. putida*	OM channel (COG4313)	12	275	C43(DE3)	pB22	His	6	C	IB-urea	Xray-C8E4 or MD[6]	2015
5BUN	*S. enterica*	Antigenic OM protein	12	467	BL21(DE3)	pHDST	His	NS[5]	NS[5]	DDM	Xray-DDM	2015
2N6L	*P. aeruginosa*	OM protein (OprG)	8	215	BL21(DE3)	pET30a	His	6	C	IB-urea	ssNMR-DHPC (micelles)	2015
5DL7	*A. baumannii*	OM carboxylate channel (OccAB3)	18	419	C43(DE3)	pB22	His	7	N	Elugent	Xray-C8E4	2016
5DL5	*A. baumannii*	OM carboxylate channel (OccAB1)	18	430	C43(DE3)	pB22	His	7	N	Elugent	Xray-C8E4	2016
5DL6	*A. baumannii*	OM carboxylate channel (OccAB2)	18	413	C43(DE3)	pB22	His	7	N	Elugent	Xray-C8E4	2016
5DL8	*A. baumannii*	OM carboxylate channel (OccAB4)	18	407	C43(DE3)	pB22	His	7	N	Elugent	Xray-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
5IXM	*Y. pestis*	Lipopolysaccharide transport (Lpt)	26	577 (LptD) + 198 (LptE)	BL21(DE3)	pET9 /pCDFc1b	His	10 and 6	N and C	Elugent	Xray-C8E4	2016
5IV8	*K. pneumoniae*	Lipopolysaccharide transport (Lpt)	26	601 + 182 (LptE)	BL21(DE3)	pET9/pCDFc1b	His	10 and 6	N and C	Elugent	Xray-C8E4	2016
5IVA	*P. aeruginosa*	Lipopolysaccharide transport (Lpt)	26	646 (LptD) + 192 (LptE)	BL21(DE3)	pET9/pCDFc1b	His	10 and 6	N and C	Elugent	Xray-C8E4	2016
5AZO, 5AZS	*P. aeruginosa*	Multidrug efflux pump (OprN-J)	12	455	C43(DE3)	pET21b	His	6	C	TX100	NF[7]	2016
5FVN	*E. cloacae*	Porin (OMPs)	16	342	BL21(DE3) omp8	pBAD24	NT[9]			LDAO	Xray-C8E4	2016

(*continued on next page*)

Table 2 (*continued*)

PDB	Organism	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solubilization[3]	Structure[4]	Date
5IMW	*S. intermedius*	Cholesterol-dependent cytolysins (CDCs)	4	471	BL21(DE3)	pMCSg7	His	6	N	None	None	2016
5IMY	*G. vaginalis*	Cholesterol-dependent cytolysins (CDCs)	4	490	BL21(DE3)	pMCSg7	His	6	N	None	None	2016
4ZGV	*P. atrosepticum*	Plant-ferredoxin receptor (FusA)	22	868	BL21(DE3)	pET28a	His	6	N	IB-urea	Xray-MD[6]	2016
5LDV	*C. jejuni*	Major OM protein (MOMP)	18	408	C43(DE3)	pTAMA	His	6	N	Elugent	Xray-C8E4	2016
5WQ8	*V. cholerae*	Secretin (GspD)	60	650	DH5α	pASK-IBA3c	Strep		C	MD[6]	EM-LDAO	2016
5LY6	*S. pneumoniae*	Pneumolysin	4 × 42	471	BL21(DE3)	pET15b	His	6	N	None	None	2017
5VJ8	*Yersinia pestis*	Attachment invasion locus (Ail)	8	157	BL21(DE3)	pET30b	His or NT[9]	6	C	IB-urea	lsNMR-DMPC/DMPG (nanodisc)	2017

[1] Number of transmembrane domains.
[2] N- or C-terminal position.
[3] Solubilization detergent.
[4] Detergent used for structure determination.
[5] Mixed detergents.
[6] Not specified in PDB or corresponding publication.
[7] Not found or article was not accessible.
[8] Inclusion bodies solubilised in 6 M guanidine hydrochloride.
[9] No tag.

Rosetta™ 2(DE3) (Table 7). The bacterial host BL21(DE3) (including mutant derivatives) is most used (111 uMPS), followed by the two mutant hosts, C43(DE3) and C41(DE3) (54 and 30 uMPS, respectively). The BL21(DE3) host together with plasmids expressing either lysozyme or a rare tRNA (28 and 19 uMPS) follows in fourth and fifth position, respectively. Rosetta™ 2(DE3) was used for a total of 13 uMPS. Bacterial hosts other than those mentioned above have only had a marginal impact in the field (1 to 8 uMPS each). Production of homologous membrane proteins in *E. coli* shows a similar pattern (Table 8), although native *E. coli* promoters were more frequently used (Fig. 5). As expected, rare tRNA plasmids were not typically used for the production of homologous membrane proteins. The two most used bacterial hosts were BL21(DE3) and C43(DE3) yielding 41 and 26 uMPS, respectively (Table 8). 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 *lacI* 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), pIVEX (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* and *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 3
Unique membrane protein structures derived from recombinant proteins produced in *Pichia pastoris*.

PDB	Organism	Description	TM[1]	Size	Host	Vector	Tag	L[2]	N/C[3]	Solub[4]	Struct[5]	Date
Monotopic												
1GOS, 1OJA	*H. sapiens*	Monoamine oxidase B (MAO)	0	520	KM71	pPIC3.5 K	NS[6]			None	Xray-LDAO or SB3-12	2001
2BXR	*H. sapiens*	Monoamine oxidase A (MAO)	0	527	KM71	pPIC3.5 K	NS[6]			None	Xray-OG	2005
Alpha-helical												
2A79, 3LUT	*R. norvegicus*	Voltage-dependent potassium ion channels (kV channels)	6 × 4	333	SMD1163 (HIS^+)	pPICZC	His	8	N	DDM	Xray-DM	2005
1Z98, 3CLL	*S. oleracea*	Plant aquaporin	7	281	X33	pPICZB	His	6	C	OG	Xray-OG	2005
2UUH, 4JCZ	*H. sapiens*	LTC4 synthase	4 × 3	156	KM71H	pPICZA	His	6	N	MD[7]	Xray-DDM	2007
2R9R, 3LNM	*R. norvegicus*	Voltage-dependent K1 channels (kV)	6 × 4	514 + 333	SMD1163 (HIS^+)	pPICZ-C	His	10	N	DDM	Xray- MD[7]	2007
3D9S	*H. sapiens*	Human aquaporin 5	6	266	X33	pPICZ-B	NT[8]			NG	Xray-NG	2008
3G5U, 4M1M, 4Q9H	*M. musculus*	P-glycoprotein (Pgp)	12	1284	GS115	pHIL-D2	His	6	C	TX100	Xray-DDM	2009
3GD8	*H. sapiens*	Aquaporin (AQP) 4	6	223	X33	pPICZ	His + FLAG	8	N	OG	Xray-OG	2009
2W2E, 3ZOJ	*P. pastoris*	Yeast aquaporin	6	279	GS115-his4	pPICZαB	His	6	C	OG	Xray-OG	2009
3JYC, 3SPI	*G. gallus*	Inward-rectifier potassium channels	2 × 4	343	SMD1163 (HIS^+)	pPICZB	GFP		C	DM	Xray-DM	2010
3RZE	*H. sapiens*	Histamine-H1 receptor (H1R)	7	452	SMD1163	pPIC9K	GFP + His	8	C	DDM	Xray-DDM	2011
3SYO, 4KFM	*M. musculus*	G- protein-gated K^+ channel (GIRK)	2 × 4	340	SMD1163 (HIS^+)	pPICZ	GFP + His	10	C	DM or DDM	Xray-DM or DDM	2011
3VG9	*H. sapiens*	A_{2A} adenosine receptor (A2AAR)	7	326	SMD1163	pPIC9K	His + FLAG	10	C	DDM	Xray-DDM	2012
3UM7, 4I9W, 4WFF	*H. sapiens*	Potassium channel $K_{2P}4.1$ (TRAAK)	4 × 2	309	SMD1163	pPICZB	GFP + His	10	C	DDM	Xray-FC12	2012
3UKM	*H. sapiens*	Two–pore domain potassium (K^+) channels (K2P channels)	4 × 2	280	SMD1163 (HIS^+)	pPICZC	GFP + His	10	N	DDM	Xray- MD[7]	2012
4F4C	*C. elegans*	P-glycoprotein (P-gp)	12	1321	SMD1163	pPICZ and pVL1393	GFP + His	10	C	DDM	Xray-UDTM	2012
4HKR	*D. melano-gaster*	Calcium release- activated calcium channel Orai	4 × 6	214	SMD1163 (HIS^+)	pPICZC	His	6	C	DDM	Xray- MD[7]	2012
3WME	*C. merolae*	P-glycoprotein	6 × 2	612	SMD1163	pPICZA	His	10	C	C12E9	Xray-DM	2014
4NEF	*H. sapiens*	Human aquaporin 2 (AQP2)	6	242	GS115 aqy1Δ	pPICZB	His	8	N	NG	Xray-OGNG	2014
4RDQ, 5T5N	*G. gallus*	Bestrophin calcium-activated chloride channels (CaCCs)	4 × 5	409	SMD1163 (HIS^+)	pPICZ	Anti-tubulin Ab	5	C	DDM	Xray-DDM	2014
5CTG	*O. sativa*	SWEET transporters	7	224		pPICZC	GFP + His		C	DDM	Xray-NG	2015
5E1J, 5TUA	*A. thaliana*	Two-pore channels (TPCs)	12 × 2	741	SMD1163	pPICZ	GFP + His	8	C	DDM	Xray-LMNG	2015
5I32	*A. thaliana*	Aquaporins of the TIP subfamily	6	275	X33	pPICZB	His	10	N	OG	Xray-OG	2016
5DO7	*H. sapiens*	ABC transporters	6 + 6	666 + 685	KM71H	pSGP18 and pLIC (pPICZB)	Calmodulin $His_6GlyHis_6$	12	C	DDM	Xray- MD[7]	2016
5KUK	*G. gallus*	Inward rectifier potassium (Kir) channel	2 × 4	343	SMD1163 (HIS^+)	pPICZB	Flag		C	DM	Xray-DM	2016
5EGI	*C. elegans*	Trimeric intracellular cation (TRIC) channel family	7 × 3	257	GS115	pPICZA/C	His	6	C	TX100	Xray-DM	2016
5U1D	*H. sapiens*	ABC transporter	6 + 6	748 + 686 + 88	SMD1163 (HIS^+)	pPICZ	Protein A		C		EM- MD[7]	2017
5VK5	*M. musculus*	Potassium (K2P) channels of the TREK subfamily	4	312	SMD1163H	pPICZ	GFP + His	10	C	MD[7]	Xray- MD[7]	2017
5WIE	*R. norvegicus*	Voltage-gated K^+ channels (kV)	6 × 4	(532) + β-subunit (333)	SMD1163 (HIS^+)	pPICZB	Strep	× 2	N	DDM	Xray- MD[7]	2017
5XJJ	*C. sativa*	Multidrug and toxic compound extrusion (MATE)	12	455	NF[9]	NF[9]	NF[9]	NF[9]	NF[9]	NF[9]	NF[9]	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] 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 recombinant *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_{GAL1}, is induced with galactose while P_{AOX1} (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_{T7}-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^{-6}, 10^{-7} and 10^{-8} 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 preculture [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_{600} = 0.6), two options are also worth trying. The first is not 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 T7*lac* promoter and/or multicopy *lacI* or lysozyme 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_{600} = 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_{600} = 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^{-1} to 10^{-4} dilutions are plated onto the IPTG- and antibiotic-containing plates. The frequency of appearance of mutant hosts varies from 10^{-4} to 10^{-6} [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
Alpha-helical												
1O5W	*R. norvegicus*	Monoamine oxidase (MAO)	1	534	BJ2168	YEp51	His	6	N	FC12	Xray-MD[7]	2004
2Z5X	*R. sapiens*	Monoamine oxidase (MAO)	1	513	BJ2168	YEp51	His	6	N	FC12	Xray- MD[7]	2008
4A01	*V. radiata*	H^+ translocating pyrophosphatases (H^+ -PPases)	16	766	BJ2168	pYVH6	His	6	C	DDM	Xray-DM	2012
4AV3	*T. maritima*	Pyrophosphatases (M-PPases)	16	735	BJ1991	pRS1024	His	6	N	DDM	Xray-OGNG or Cymal5	2012
4IL3	*S. mikatae*	CaaX protease Ste24p	7	461	BJ5460	pSGP46	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)	10	1000	W303.1b Gal4-2	pYeDP60	BAD		C	C12E8	Xray-C12E8	2013
4C9G	*S. cerevisiae*	Mitochondrial ADP/ATP carrier	6	318	WB12	pYES3	His	8/9	N	UDM	Xray-DM or Cymal5	2014
4LXJ	*S. cerevisiae*	Cytochrome P450	1	536	MMLY941	NS[6]	His	6	C	DM	Xray-DM	2014
4WIS	*N. haematococca*	Anoctamin family (TMEM16)	10 × 2	735	FGY217 (Δura)	pYES2	GFP + His	10	N	DDM	Xray- MD[7]	2014
3J9T	*S. cerevisiae*	H^+-ATPases (V-ATPases) (subunits ABCDEFGHacc'c"d)	8a + 40 cc'c"	616 + 517 + 392 + 256 + 23 + 118F + 478 + 840a + 160 cc'c" + 345d	JTY002	NS[6]	FLAG		C	DM	EM-DDM	2015
5A2N	*A. thaliana*	Proton-coupled transporters (NRT1/PTR)	12	590	FGY217 (pep4Δ)	pRS426GAL1	GFP + His	8	C	DDM	Xray-DDM	2015
5AEX	*S. cerevisiae*	(Mep2)	10	505	W303 (pep4Δ)	p83νΔ	His	6	C	MD[7]	Xray-DMNG	2016
5AEZ	*C. albicans*	Mep2 proteins	11 × 3	486	W303 (pep4Δ)	p83νΔ	His	6	C	MD[7]	Xray-DM or NG or OGNG	2016
5DQQ	*A. thaliana*	Two-pore channels (TPCs)	12 × 2	723	DSY-5	p423-GAL1	His	10	C	DDM	Xray-DDM	2016
5HI9	*R. norvegicus*	Transient receptor potential (TRP)	6 × 4	770	BJ5457	pYepM	1D4 epitope		C	LMNG	EM-DMNG	2016
5EQI	*H. sapiens*	Glucose transporter 1 (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 ac8c'c"def)	6a + 40c8c'c" + 2e + 2f	680a + 150c + 147c' + 213c" + 297d + 57e + 54f	CACY1	NS[6]	FLAG		C	DM	EM-DDM	2016
5LZQ	*T. maritima*	Pyrophosphatases (M-PPases)	16 × 2	735	BJ2168	pYES2	His	6	C	DDM	Xray-DM	2016
5V6P	*S. cerevisiae*	(Hrd1-Hrd3)	8 × 2	407	INVSc1	pRS42X	Strep		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
Archaea	28	0	0	28
Parasite	2	0	0	2
Other[1]	5	4	0	9
Total[2]	468	31	22	521

[1] *Plexaura homomalla*, *Danio rerio*, *Eisenia fetida*, *Actinia fragacea*, *Caenorhabditis elegans*.

[2] The PDB was analyzed in November 2017.

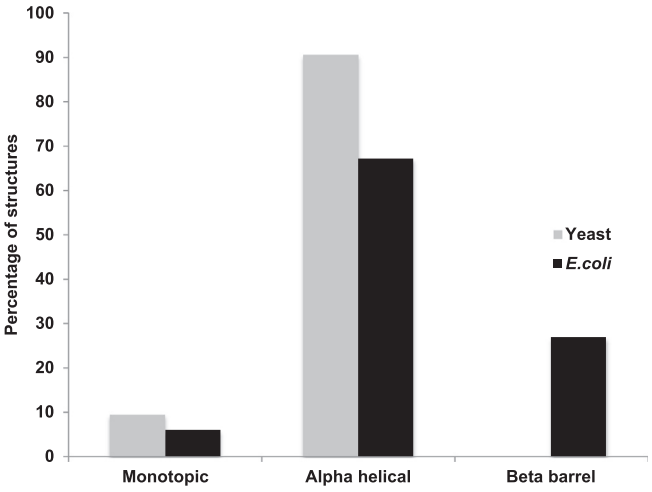

Fig. 1. Secondary structure analysis of recombinant membrane proteins produced in *E. coli* and yeast for which a structure has been resolved. Data were obtained from Tables 1–4. The percentage of unique membrane protein structures is plotted as a function of their secondary structure for recombinant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

required to cure C41(DE3) from the pOGCP expression plasmid [30]). If the mutation is in the expression vector, transformation of the isolated plasmid into BL21(DE3) cells should give colonies on IPTG-containing plates; if there are no colonies, then the isolated colony carries the mutation.

3.3. Expression of non-toxic or moderately-toxic target genes

Expression of genes encoding non-toxic or moderately-toxic membrane proteins cloned in T7 expression plasmids lead to colony formation on IPTG-containing plates. Toxicity is inversely proportional to the size of colonies on these plates. We have observed that antibiotic use is not required in large-scale cultures, providing that antibiotic has been added to the preculture [24]. The induction protocol must be adjusted depending on the size of the colonies on IPTG plates (Table 12). If the size reduction is marginal compared to plates lacking IPTG (< 10%), this may suggest that the production yield of the target membrane protein is very low. To maximize the chance of obtaining high yields, 0.7 mM IPTG should be added at the early exponential phase ($A_{600} \leq 0.4$). If the size of the colonies is decreased by 10% or more, then IPTG should be added at $A_{600} = 0.6$ at the two

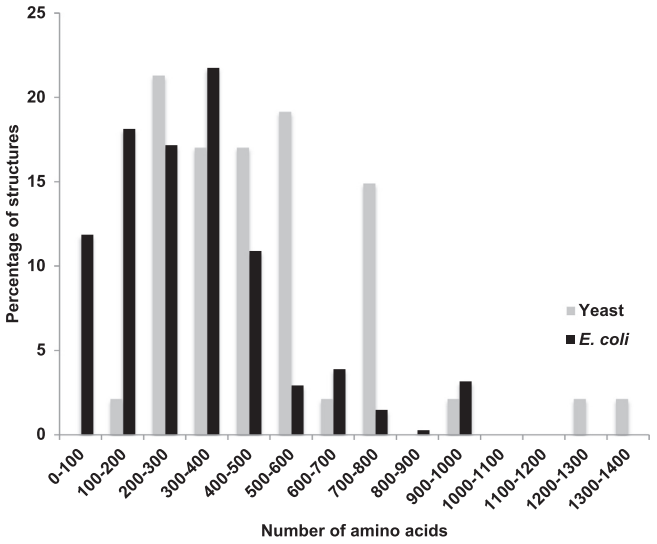

Fig. 2. Size distribution of recombinant membrane proteins produced in *E. coli* and yeast for which a structure has been resolved. Data were obtained from Tables 1–4. The percentage of unique membrane protein structures is plotted as a function of their amino acid content for recombinant membrane proteins produced either in yeast (grey) or *E. coli* (black).

concentrations that are most frequently used [8]: 0.4 mM and 0.7 mM (Table 12). Autoinduction has been used with the T7 and arabinose expression systems ([31] and Table 2 in the cases of 4HYJ, 4KJS and 3FID). In *E. coli*, glucose is a catabolic repressor that is catabolized before any other carbon source. Autoinduction media take advantage of this; they contain glucose to allow the bacterial cells to grow to high densities, but when the glucose has been exhausted, cells switch on operons involved in the catabolism of the other carbon sources present. Autoinduction media contain a defined amount of lactose that can bind to *lacI* and stimulate the expression of T7RNAP. Commercial autoinduction media are not cheap, but are a useful option when leaky expression is toxic and prevents cell growth prior to IPTG addition. Another option to circumvent toxicity is to decrease the temperature of the culture 30 min before IPTG addition. In a previous study [8], we demonstrated that in approximately 50% of studies using T7 expression systems, lowering the temperature (i) prevented the formation of inclusion bodies, (ii) improved the solubility of the recombinant membrane protein, (iii) reduced toxicity or (iv) prevented overgrowth of the culture by cells that had lost the expression plasmid [30].

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

Formation of inclusion bodies containing a recombinant membrane protein (IBMP) 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 Ni^{2+}-affinity chromatography can be performed in denaturing conditions, IBMP 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[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solub[3]	Struct[4]	Date
Monotopic												
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[5]	2008
3O8Y	*H. sapiens*	Enzyme 5-lipoxygenase (5LOX)	0	691	Rosetta™ 2(DE3) T7	pET14b	His	6	N	None	None	2011
4HHS	*A. thaliana*	Fatty acid α-dioxygenase (α-DOX)	0	652	M15 T5	pQE30	His	6	N	DM	Xray-NG	2013
4NRE	*H. sapiens*	Enzyme 15-lipoxygenase-2 (15LOX-2)	0	696	Rosetta™ 2(DE3) T7	pETDuet-1	His	6	N	None	None	2014
4PLA	*H. sapiens*	Phatidylinositol 4-kinase type IIa (PI4K IIa)	0	556	BL21(DE3) Star T7	pRSFD	His	6	N	None	None	2014
4QN9	*H. sapiens*	Fatty-acid ethanolamides (FAEs)	0	393	Rosetta-origamiB (DE3) pLysS; Other	pMAL	MBP + His	6	C	TX100	Xray-DC	2015
Alpha-helical												
1ZLL, 2M3B	*H. sapiens*	Phospholamban homopentamer	1 × 5	52	BL21(DE3); Other	pMALc2 x	MBP		N	NS[6]	lsNMR-FC12 ssNMR-DOPC/DOPE	2005
2HAC	*H. sapiens*	TCR-CD3, TM dimer complex	1 × 2	33	BL21(DE3) T7	pMM-LR6	His	9	N	IB-gua[8] + TX100	lsNMR-SDS/FC12	2006
2JO1	*H. sapiens*	Phospholemman (FXYD1)	1	72	C43(DE3) T7	pETBcl-XL	Bcl-XL + His	6	N	IB-gua[8]	lsNMR-SDS	2007
2Q7M	*H. sapiens*	Lipoxygenase protein (FLAP)	4 × 3	161	BL21(DE3) T7	pET28a	His	6	C	DDM	Xray-MD[5]	2007
2K1L	*H. sapiens*	Receptor tyrosine kinase (EphA1)	1	38	BL21(DE3) pLysS T7	pGEMEX1	TrxA-His	NS[6]	N	TX100	lsNMR-DMPC/DHPC	2008
3DWW	*H. sapiens*	Prostaglandin E synthase 1	4 × 3	158	BL21(DE3) pLysS T7	pSP19T7LT	His	6	N	TX100	EM-TX100	2009
2K9Y	*H. sapiens*	Receptor tyrosine kinases (Eph2)	1	41	BL21(DE3) pLysS T7	pGEMEX1	TrxA-His	NS[6]	N	TX100	lsNMR-MeOH/$CHCl_3$/H_2O	2009
2KNC	*H. sapiens*	Integrin αIIbβ3	1 + 1	54 and 79	BL21(DE3) Other	pMAL-C2	MBP-His	6	N	TX100	lsNMR-CD_3CN/H_2O	2009
2KOG	*R. norvegicus*	Synaptobrevin	1	119	BL21(DE3)/BL21(DE3) pRil T7	pET15b/28a	His	6	N	Sodium cholate	lsNMR-FC12	2009
2KS1, 2JWA	*H. sapiens*	ErbB1/ErbB2	1 × 2	44	BL21(DE3) pLysS T7	pGEMEX1	TrxA-His	6	N	TX100	lsNMR-DHPC/DMPC	2010
2KPF	*H. sapiens*	Glycophorin A (GpA)	1 × 2	38	NF[7]	NF[7]	NF[7]				lsNMR-DHPC/DMPC	2010
2L35	*H. sapiens*	Signaling module (DAP12)	2 + 1	63 + 32	BL21(DE3) T7	pMM-LR6	His-trpLE	9	N	IB-gua[8] + TX100	lsNMR-FC14 + SDS	2010
2KYV	*H. sapiens*	Phospholamban homopentamer	1 × 5	52	BL21(DE3) T7	pET	MBP		N	TX100	lsNMR-FC12 ssNMR-DOPE/DOPE	2011
2L9U	*H. sapiens*	Transmembrane domain (ErbB3)	1 × 2	40	Cell-free expression	pET22b	His	6	C	Cell-free expression-pellet	lsNMR-FC12	2011
2LCK	*M. musculus*	Mitochondrial uncoupling protein 2 (UCP2)	6	303	Rosetta™ 2(DE3)	pET21	His	6	C	FC12	FC12	2011
2LCX	*H. sapiens*	ErbB4	1 × 2	44	BL21(DE3) pLysS T7	pGEMEX1	TrxA-His	6	N	TX100	lsNMR-DHPC/DMPC	2012
2LNL	*H. sapiens*	Receptor (CXCR1)	7	309	BL21(DE3) T7	pGEX2a	GST + His	6	C	IB-SDS	ssNMR-DMPC	2012
2LOU	*H. sapiens*	Apelin receptor	1	64	BL21(DE3) T7	pEXP5-CT	His	6	C	IB- acetonitrile /trifluoracetic acid	lsNMR-FC12	2013
2M6B	*H. sapiens*	Glycine receptor (hGlyR-α1)	4 × 5	150	BL21(DE3) pLysS T7	pET31b	His	6	C	IB-NS[6]	lsNMR-LPPG	2013
2LZL	*H. sapiens*	Growth factor receptor 3 (FGFR3)	1 × 2	43	BL21(DE3) pLysS T7	pGEMEX1	TrxA-His	6	N	TX100	lsNMR-FC12/SDS	2013
2MAW	*H. sapiens*	Neuronal acetylcholine receptor	4	137	Rosetta™ 2(DE3) pLysS T7	pMCSG7	His	6	N	NS[6]	lsNMR-LDAO	2013
2M8R, 3HD7	*R. norvegicus*	Syntaxin 1A, TM & syntaxin complex	1 + 1	109 + 91	BL21(DE3) T7	pET28a	His	6	N	TX100 or OG	lsNMR-FC12, Xray-NG or C7G	2013
4BUO	*R. norvegicus*	Neurotensin receptor (NTS1)	7	335	BL21 Tuner Other	pBR322	MBP + His	6	N	MD[5]	Xray-MD[5]	2014
4O6Y	*A. thaliana*	Cytochrome B561 (Cyt b_{561}-B)	6	230	BL21(DE3) T7	pET15b	His	6	NS[6]	DM	Xray-NG	2014
2M59	*H. sapiens*	Endothelial growth factor receptor 2	1 x 2	37	Cell-free expression	pET20b	HA		N	Cell-free expression pellet-TFE/H2O/TFA	lsNMR-FC12	2014
2MGY	*M. musculus*	Translocator protein (TSPO)	5	169	BL21(DE3)	pET15b	His	6	N	IB-SDS	lsNMR-FC12	2014
2MFR	*H. sapiens*	Insulin receptor (AAs 940–980)	1	57	BL21(DE3) T7	pET29b	His	6	C	IB-urea	lsNMR-FC12	2014
2MIC	*R. norvegicus*	Neurotrophin receptor (p75)	1 × 2	41	Cell-free expression	NS[6]	NT[9]			Cell-free expression pellet-sarkosyl	lsNMR-FC12	2014
4TSY	*A. fragacea*	Haemolytic fragaceatoxin C (FraC)	1 × 8	179	BL21(DE3) T7	pBAT-4	NT[9]			None	Xray-DDM	2015
4WOL	*H. sapiens*	Signaling module (DAP12)	1 × 3	33	BL21(DE3) T7	pMM	His-trpLE	9	N	IB-gua[8] + TX100	Xray-None (LCP)	2015
5EH4	*H. sapiens*	Glycophorin A (GpA)	1 × 2	30	BL21(DE3) T7	pMM	His-trpLE	9	N	IB-gua[8] + TX100	Xray-None (LCP)	2015
2N2A	*H. sapiens*	Receptor tyrosine kinases (HER or ErbB)	1 × 2	58	Cell-free expression	pGEMEX-1	NT[9]			Cell-free expression pellet-sarkosyl	lsNMR-FC12	2016

(*continued on next page*)

Table 6 (*continued*)

PDB	Organism	Name	TM[1]	Length	Strain(s)	Vector	Tag	Size	N/C[2]	Solub[3]	Struct[4]	Date
2N7Q	*H. sapiens*	Human nicastrin	1	54	NS[6]; T7	pET29b	His	6	N	IB-urea	lsNMR-SDS or FC12	2016
5ID3	*C. elegans*	Mitochondrial calcium uniporter	2 × 5	159	BL21(DE3) T7	pET21a	His	6	C	FC14	lsNMR-FC14	2016
5KTF	*M. musculus*	High density lipoprotein (HDL)	2	73	BL21(DE3) CodonPlus	pQE30	His	8	N	Empigen	lsNMR-LPPG	2017
2K4T	*H. sapiens*	Anion channel (VDAC-1)	19	291	BL21(DE3) T7	pET21a	His	6	C	IB-urea	lsNMR-LDAO	2008
Beta-barrel												
3EMN, 4C69	*M. musculus*	Anion channel (VDAC-1)	19	295	M15	pQE9	His	6	N	IB-gua[8]	Xray-LDAO (bicelles)	2008
2JK4	*H. sapiens*	Voltage-dependent anion channel (VDAC)	19	294	M15 ; NS[6]	PDS56	His	6	C	IB-gua[8]	Xray-Cymal5	2008
4BUM	*D. rerio*	Voltage-dependent anion channel 2	19	289	M15T5	pQE60	His	6	C	IB-gua[8]	Xray-LDAO	2014
5GAQ	*E. fetida*	β-pore-forming toxins (β-PFTs)	2	310	Rosetta™ 2(DE3) T7	pHis-Parallel1	His	6	C	None	EM-DDM	2016

[1] Number of transmembrane domains.
[2] N- or C-terminal position.
[3] Solubilization detergent.
[4] Detergent used for structure determination.
[5] Mixed detergents.
[6] Not specified in PDB or corresponding publication.
[7] Not found or article was not accessible.
[8] Inclusion bodies solubilised in 6 M guanidine hydrochloride.
[9] No tag.

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 monoglycosyldiacylglycerol 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: 600*g* for 10 min to collect unbroken cells and cell debris in the pellet, followed by 10,000*g* for 15 min at 4 °C to collect inclusion bodies from the supernatant. Bacterial membranes remain in the 10,000*g* supernatant; they can be pelleted after high speed centrifugation, usually 100,000*g* for 1 h. To collect bacterial membranes in the absence of inclusion bodies, disrupt the bacteria (at least 1 L of culture) by passing the suspension twice through a French Press or cell disruptor. If a recombinant membrane protein triggers internal membrane proliferation, such as AtpF-induced intracellular membranes [11], those membranes can be immediately collected following low speed centrifugation: 2500*g* for 10 min (P1 pellet). The pellet contains internal membranes but also unbroken cells and debris that need to be washed away. The supernatant (S1) contains inner and outer membranes, which are collected by centrifugation of S1 at 100,000*g* for 1 h at 4 °C. Proliferated membranes within P1 are then washed and unbroken cells are removed after centrifugation at 2500*g* for 10 min at 4 °C. The supernatant (S2) contains the washed internal membranes, which are collected after 1 h centrifugation at 100,000*g*. The next step is to separate membrane vesicles according to their specific density on a sucrose gradient. For high purity requirements, continuous gradients are used.

4. P_{tac}-based protocols: 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

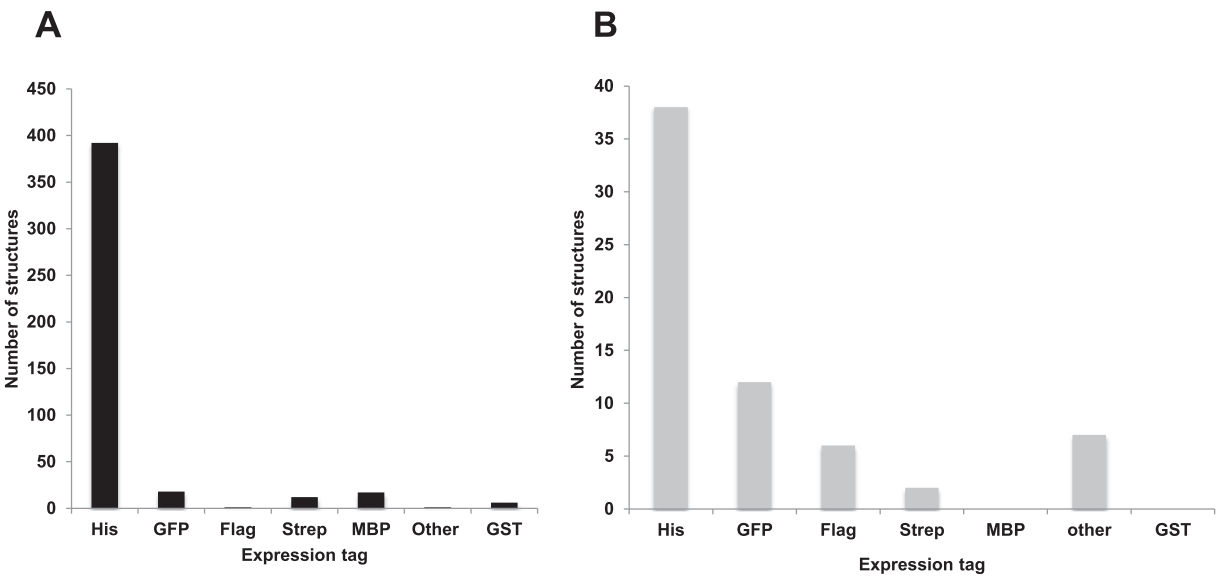

Fig. 3. Tag usage for recombinant membrane protein production in microbes. Data were obtained from Tables 1–4. The number of unique membrane protein structures is plotted as a function of the purification tag present in the corresponding recombinant membrane protein following production in (A) bacteria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

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.

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)

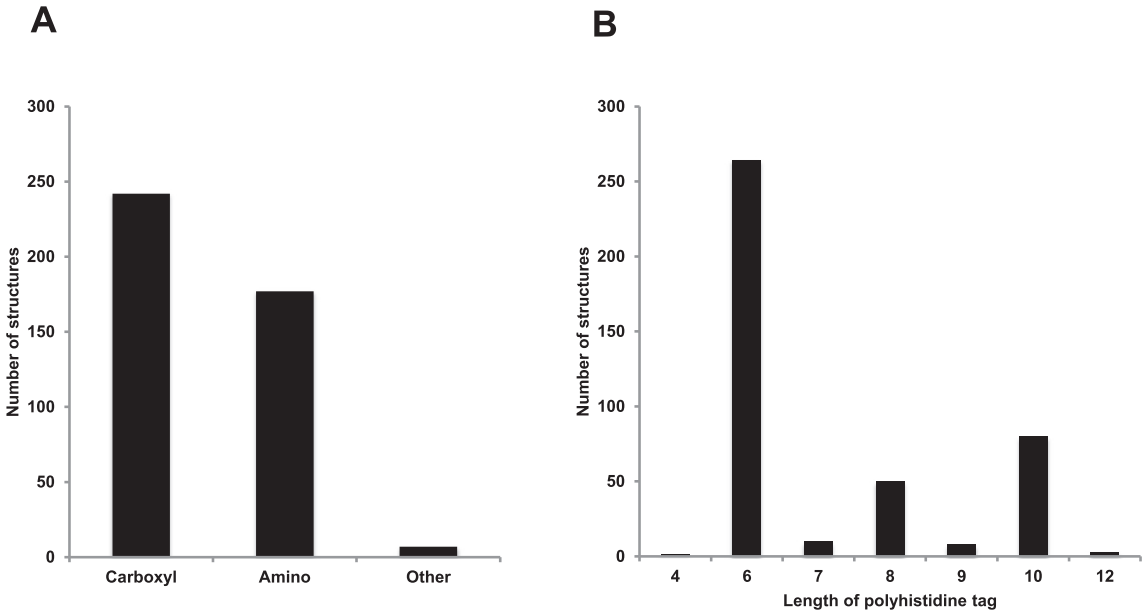

Fig. 4. Polyhistidine tag usage in microbial expression systems. Data were obtained from Tables 1–4. (A) The position of the polyhistidine tag within a recombinant membrane protein and (B) the number of histidine residues it contains are shown.

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, tac, lac, rham*	Not specified	
BL21(DE3)	88	7	2	3	6	1	107
BL21(DE3) ΔacrB	1						1
BL21(DE3) omp5	1	1					2
BL21(DE3) ΔacrABΔmacABΔyojHI	1						1
C43(DE3)	34	16	2				52
C43(DE3) ΔcyoABCD		1					1
C43(DE3) ΔacrB		1					1
C41(DE3)	27				2		29
C41(DE3) ΔacrB	1						1
BL21(DE3) pLysS	24	2	1				27
BL21(DE3) omp8 pLysS	1						1
BL21(DE3) CodonPlus1	14		1	1	3		19
Rosetta™ 2(DE3) T7	8						8
Rosetta™ 2(DE3) pLysS T7	5						5
BL21(DE3) Tuner	2				1		3
BL21(DE3) Star	2						2
SE1 T7	1						1
BL21(DE3) Gold	3	1					4
BL21-Gold pLysS	1						1
GJ1158	1						1
Lemo21(DE3)	2						2
Lemo56(DE3)	1						1
Origami B			1				1
Rosetta- Origami B pLysS					1		1
B834	2			1			3
B834 pLysS	1						1
PA (ΔoprH)		1					1
DH10B/TOP10		7					7
XL1-Blue		1	5	1	1		8
BL21(DE3) AI	1	1					2
DW2					1		1
DH5α				2	1		3
SG1309			2				2
MC4100		1					1
SCM6		1					1
MC1061		5					5
JM83				2			2
M15			1			1	2
EP51		1		1			2
KRX					1		1
MG1655					1		1
JM109					1		1
420399				1			1
AR58					1		1
BLR					1		1
GT1000 Δ(glnK,amtB)					1		1
PAP5198					1		1
K38						1	1
AD202					1		1
Not found						2	2
Not specified	3			1			4
Cell-free expression[1]	4						4
Total	229	47	15	13	24	5	333[2]

[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 5-Helix Inverted Repeat Transporter superfamily ('5-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

Table 8
Promoter and *E. coli* strain combinations used for the production of recombinant *E. coli* (homologous) membrane proteins.

Bacterial strain	Promoter used in the expression plasmid						
	T7	*ara*	*T5*	*tet*	*trp, tac, lac, rham*	Native	Total
BL21(DE3)	31		2	1	5	2	41
C43(DE3)	18	5	2		1		26
C41(DE3)	9					1	10
BL21(DE3) pLysS	7						7
BL21(DE3) CodonPlus	1						1
BL21(DE3) Star	1						1
BL21(DE3) Star pLysS	3						3
Rosetta™ 2(DE3)	1						1
BL21(DE3) Tuner	1	1					2
SE1 T7	1						1
BL21(DE3) Gold	3						3
LE392		1					1
LMG194		1					1
TNE012						1	1
HM125			1				1
XL1-Blue			1			1	2
DW35						2	2
BZB1107						2	2
AW740						1	1
MH225						1	1
HN705Δomp8						1	1
MEG119						1	1
GO105						1	1
GL101						1	1
FT004						1	1
DK8						1	1
pop6510						1	1
B834	2	3		1	1	1	8
WH1061					1		1
HN741					1		1
MJF612					1		1
DH5α				1	1		2
TOP10		1			1		2
HDB150					1		1
M15			2				2
LCB2048					1		1
UT5600						1	1
JM109			1		1		2
Not specified	2			1	1	3	7
Total	80	12	9	4	16	23	144

[46,47,54,60]. Also, the desirability of placing a tag (usually $(His)_n$) at either the carboxyl-terminus or the amino-terminus of the cloned gene has been discussed [61]. Plasmid pTTQ18 contains a polylinker/*lacZ*α 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*^{q} 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 *Eco*RI or *Nde*I restriction enzyme sites at the 5′ end of the amplified gene, and *Pst*I or *Hin*dIII 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 $RGSHis_6$ motif, which is present on a modified pTTQ18 between the *Pst*I and the *Hin*dIII 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 (Saidijam, 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,

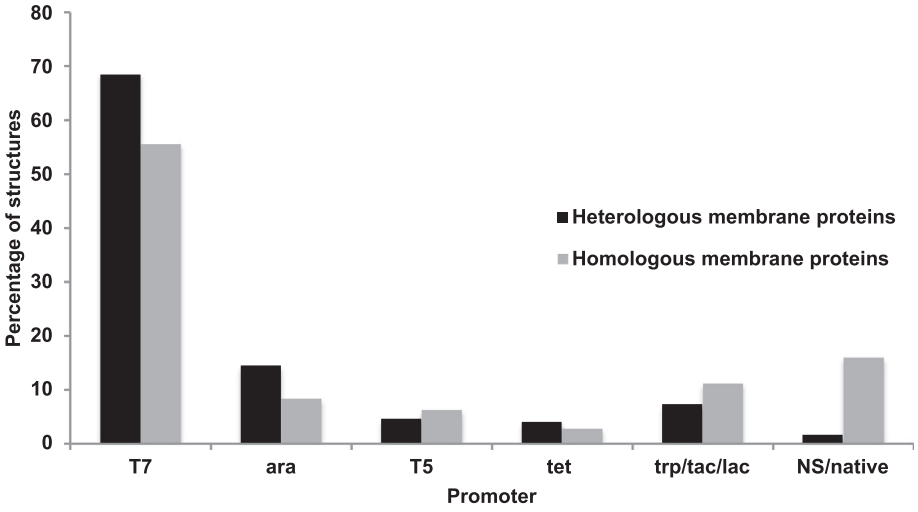

Fig. 5. Promoter usage for recombinant membrane protein production in *E. coli*. Data were obtained from Tables 1–3. The graph shows the promoters used for the heterologous (black) and homologous (grey) production of membrane proteins in *E. coli*. NS: not specified; Native: the native promoter of the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

Table 9
Genotypes of *E. coli* strains used to produce recombinant membrane proteins for structural determination.

Strains for T7RNAP-based expression	Genotype
BL21(DE3)	*E. coli* str. B F^- *ompT gal dcm lon* $hsdS_B(r_B^-m_B^-)$ λ(DE3 [*lacI lacUV5 T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K\text{-}12}$(λ^S)
C41(DE3)	BL21λ(DE3) F- *proY438 melB653 ycgO1103 yhhA290 ydcD71 zwf-815 rpoC3023 yehU679 rbsDΔ(IS3) λ(DE3 [lacUV5])*
C43(DE3)	C41λ(DE3) *F- Δ(dcuS)866–870 fur::Val lonΔ(IS4)* yibJ90 yjcO665 cydA::IS1 *Δ(ccmF-ompC) Δ(yjiV-yjjN) rbsDΔ(IS3) λ(DE3 [lacI574 lacUV5])*
BL21(DE3) pLysS	BL21λ(DE3) pLysS (CamR)
BL21(DE3) CodonPlus	BL21 *dcm* + TetR λ(DE3) *end*A Hte [*arg*U *pro*L CamR]
BL21(DE3) Star	BL21 *rne*131 λ(DE3)
BL21(DE3) Rosetta pLysS	BL21 λ(DE3) pLysSRARE (CamR)
Tuner^{TM}(DE3)	BL21 *lacZ*Y1 λ(DE3)
BL21(AI)	BL21 lon *ara*B::T7RNAP-tetA
Other expression strains	Genotype
BL21Rosetta	BL21 RARE (CamR)
BL21-Gold	BL21 *dcm* + TetR *end*A Hte
BL21-T1^R	*fhu*A2 [lon] *ompT gal* [*dcm*] Δ*hsd*S
Origami B	BL21 *lac*Y1 *aph*C *gor*522::Tn10 *trx*B (KanR TetR)
B834	F- *ompT hsd*SB (rB- mB-) *gal dcm met*
BLR	F- *ompT hsd*SB (rB- mB-) *gal dcm* Δ(srl-recA)306::Tn10 (TetR)
DH10B TOP10	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ*lac*X74 *nup*G *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*E15 *gal*K16 *rps*L *end*A1 λ- DH10B *rps*L(StrR)
KRX	[F′, *tra*D36, Δ*ompP pro*A + B+ *lac*Iq Δ(*lacZ*)M15] Δ*ompT end*A1 *rec*A1 *gyr*A96 (Nalr) *thi-1 hsd*R17 (rk– mk+) e14– (McrA–) *rel*A1 *sup*E44 Δ(*lac-pro*AB) Δ(*rha*BAD)::T7 RNA polymerase
XL10-Gold XL1-Blue	F' [*pro*AB *lac*IqZΔM15 Tn10(TetR Amy CmR)] *rec*A1 *end*A1 *gln*V44 *thi-1 gyr*A96 *rel*A1 *lac* Hte Δ(*mcr*A)183 Δ(*mcr*CB-*hsd*SMR-*mrr*)173 TetR F' [*pro*AB, *lac*Iq ZΔM15 Tn10(TetR)] *rec*A1 *end*A1 *gyr*A96 *thi-1 rel*A1 *sup*E44 *hsd*R17(rK- mK +) l-
DH5α	F- ø80*dlacZ*ΔM15 Δ(*lacZYA-arg*F)U169 *deo*R *rec*A1 *end*A1 *hsd*R17(rK – mK+) *pho*A *sup*E44 λ– *thi-1 gyr*A96 *rel*A1
SG13009	Nal[s] Str[s] Rif[s] Thi[-] lac[-] Ara[+] Gal[+] Mtl[-] F[-] RecA[+] Uvr[+] Lon[+]
LS6164	Δ*fad*R Δ*fad*L
MC4100	F- [*ara*D139]B/r Δ(*arg*F-*lac*)169* &lambda- *e14- flh*D5301 Δ(fruK-yeiR)725 (fruA25) *rel*A1 *rps*L150(strR) *rbs*R22 Δ(*fim*B-*fim*E)632(::IS1) deoC1
SCM6	NS (Patented)
MC1061	F- Δ(*ara-leu*)7697 [*ara*D139]B/r Δ(*cod*B-*lac*I)3 *gal*K16 *gal*E15 λ- e14- *mcr*A0 *rel*A1 *rps*L150(strR) *spo*T1 *mcr*B1 *hsd*R2(r- m+)
JM83	*rps*L *ara* Δ(*lac-pro*AB) Φ80*dlacZ*ΔM15
Other	PA(*ΔoprH)*

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$_6$ tag on the plasmid also dictates the use of the *Pst*I restriction site to enable correct fusion of the protein with the tag. However, if *Pst*I cannot be used because of an internal *Pst*I site within the gene of interest, then the RGSHis$_6$ tag can instead be added at the primer level, and the *Hin*dIII 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 *Eco*RI or *Nde*I and *Pst*I or *Hin*dIII 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/RGSHis$_6$, 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

Table 10
Promoter and yeast strain combinations used for the production of homologous and heterologous recombinant membrane proteins.

Yeast strain	Promoter used in the expression plasmid				
	Inducible *GAL*	Inducible *AOX1*	Constitutive *PMA1*	Not stated	Total
S. cerevisiae					
ADΔ				1	1
BJ1991			1		1
BJ2168	3			1	4
BJ5457	1				1
BJ5460	1				1
CACY1				1	1
DSY-5	4				4
FGY217	2				2
INVSc1	1				1
JTY002				1	1
W303 pep4Δ	1			1	2
WB12	1				1
Not stated	1			1	2
Total	15		1	6	22
P. pastoris					
GS115		4			4
KM71		4			4
SMD1163		17			17
X33		4			4
Not stated		1		1	2
Total		30		1	31

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

S. cerevisiae	Genotype
ADΔ	MATα PDR1-3 Δyor1::hisG Δsnq2::hisG Δpdr3::hisG Δpdr10::hisG Δpdr11::hisG Δycf1::hisG Δpdr5::hisG Δpdr15::hisG Δura3 ΔhisAD124567 Δpdr5::hisG Δpdr15::hisG, Δura3
BJ1991	MATα pep4-3 prbl-1122 ura3-52 leu2 trpl
BJ2168	MATa leu2 trp1 ura3-52 prb1-1122 pep4-3 prc1-407 gal2
BJ5457	MATα pep4::HIS3 prb1-Δ trp1 ura3-52 leu2-Δ his3-Δ lys2-801 can1
BJ5460	MATa ura3-52 trp1 lys2-801 leu2Δ1 his 3Δ200 pep4::HIS3 prb1Δ16
CACY1	
DSY-5	MATa leu2 trp1 ura3-52 his3 pep4 prb1
FGY217	MATa ura3-52 lys2Δ201 pep4Δ
INVSc1	MATa his3Δ1 leu2 trp1-289 ura3-52
JTY002	
W303 pep4Δ	MATa leu2-3112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 pep4Δ
WB12	MATα ade2-1 leu2-3, 112 his3-22, 15 trp1-1 ura3-1 can1-100 aac1::LEU2 aac2::HIS3
P. pastoris	Genotype/phenotype
GS115	his4/Mut$^+$ His$^-$
KM71	his4 arg4 aox1::ARG4/Mut^S His$^-$
SMD1163	his4 Δpep4 Δprb1/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* (XylT), 39% with an arabinose:proton symporter from *B. subtilis* (AraE), 38% with a major myo-inositol:proton transporter from *B. subtilis* (IolT) and 38% with a D-galactose:proton transporter from *E. coli* (GalP). Wild type YwtG consists of 457 amino acids with a calculated M_r of 49,192.49 and is predicted by TMHMM to consist of 12 transmembrane helices with both the amino- and carboxyl-termini 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 RGXRR 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 M_r to 51,041.44.

Analysis of the *ywtG* gene revealed that there are no inherent *Eco*RI or *Pst*I sites within the gene enabling the use of primers designed to introduce an *Eco*RI and *Pst*I 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 *Eco*RI and *Pst*I, yielding a DNA fragment of 1.5 kbp (actual – 1.371 kbp; Fig. 9A). This was ligated into pTTQ18 and used to transform *E. coli* XL1-Blue cells. The resulting carbenicillin resistant *E. coli* XL1-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 *Eco*RI and *Pst*I yielded two DNA fragments at 1.3 kbp and 4.6 kbp, which are similar in size to the gene *ywtG* (1.371 kbp) and pTTQ18/RGSHis$_6$ (4.59 kbp; 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)$_6$ 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)$_6$ 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 film were determined by comparing their migrating distances with those of standard protein molecular weight markers. There is a linear relationship between the $\log_{10}M_r$ 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

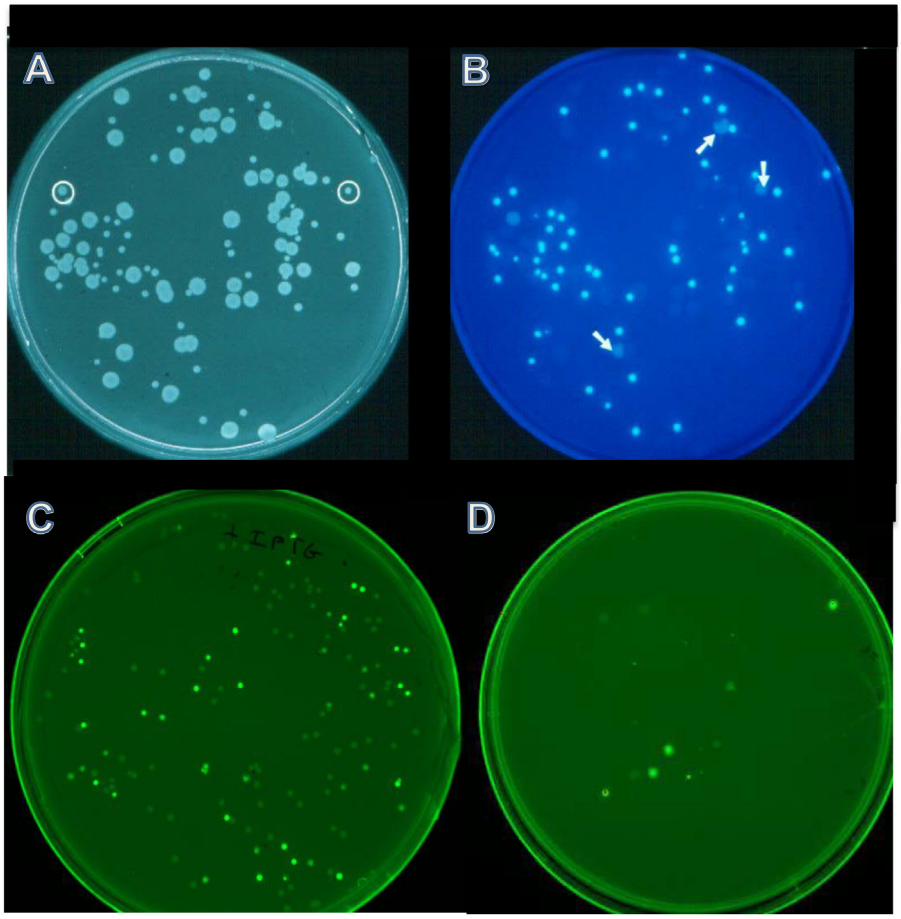

Fig. 6. Selection of bacterial strains for improved recombinant membrane protein production using GFP as a gene reporter. Isolation of bacterial mutant hosts was performed as described in Section 3.2 of this review and previously [30]. Briefly, the pMW7-GFP-Xa expression plasmid was transformed into BL21(DE3) cells and a single colony was inoculated in 50 ml 2*TY medium. At A_{600} = 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.

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

Table 12
Optimization of growth conditions in the IPTG-inducible T7RNAP expression system.

Size of colonies on IPTG plate[1]	Inoculation	Induction	IPTG concentration	Temperature after induction
No colony	no preculture[2]	No induction[3]	None	30 °C or below
		A_{600} = 1	10 µM[4], 0.1 µM	
Small (> 10% reduction)	preculture[5]	A_{600} < 0.6	0.4 or 0.7 mM	37 °C or 25 °C
Minor reduction (< 10%)	preculture	A_{600} < 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.

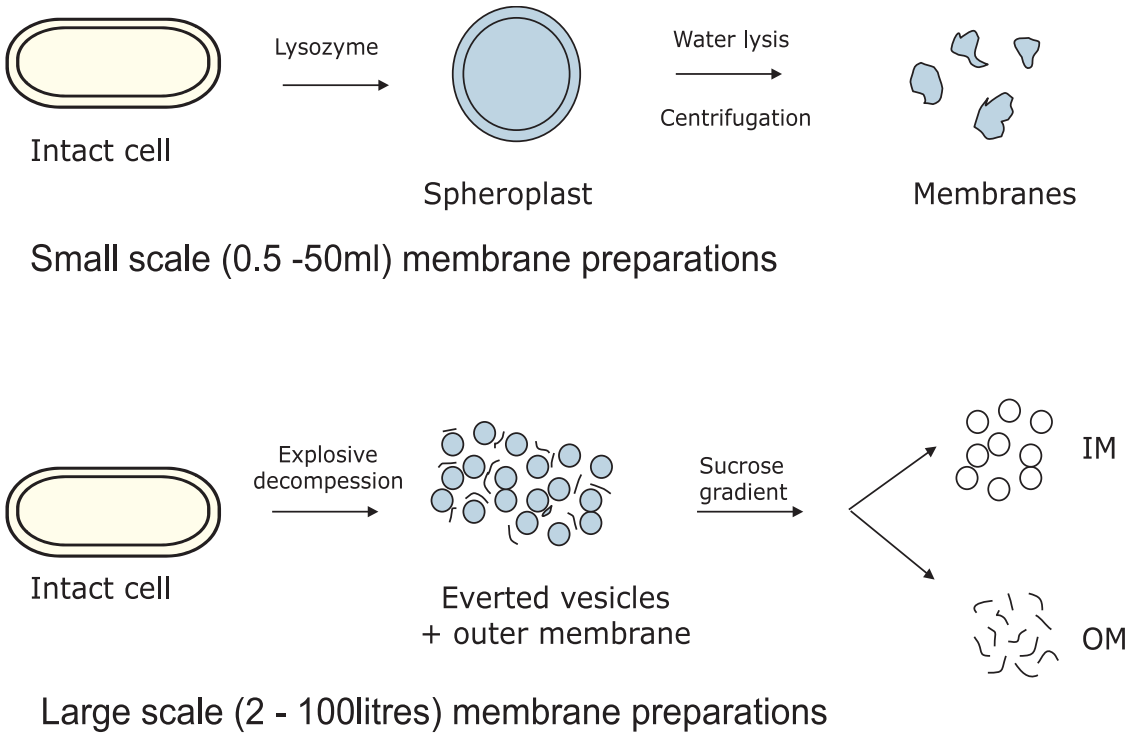

Fig. 7. Preparation of membranes from *E. coli* cultures. Schemes for the (A) small-scale preparation of mixed inner and outer membranes or (B) large-scale preparation of separated inner or outer membranes from cultures of *E. coli*.

increments within 0.01–2 mM IPTG. In some cases we have also explored 'autoinduction' using lactose/glucose mixtures [44]. This economises 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 [52] and extended to A_{680} = 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 ($hA_{2A}R$), 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 [72] 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.

5.1.1. 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 *MAT*a and *MAT*α mating types. Individual strains or the complete collection can be obtained from Euroscarf (http://web.uni-frankfurt.de/fb15/mikro/euroscarf/) or the American Type Culture Collection (http://www.atcc.org/). Dharmacon 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, *ura3Δ0, leu2Δ0, met15Δ0, his3Δ1*) is a good start, but it is not always the most successful, as shown for the expression of $hA_{2A}R$ (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 yTHC*BMS1* (supplemented with 0.5 μg/mL doxycycline) 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 $hA_{2A}R$ 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* or 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 $hA_{2A}R$, 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 $hA_{2A}R$ 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, *ura3Δ0, leu2Δ0, met15Δ0, his3Δ1*) with a control plasmid (pYX222-h*A2AR*) generated colonies on solid histidine-deficient growth medium that were designated A2 (the control, producing $hA_{2A}R$). BY4741 transformants expressing $hA_{2A}R$-Ura3p (following transformation with pYX222-h*A2AR-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, *spt3Δ*). 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 $hA_{2A}R$-Ura3p fusion proteins was analysed by immunoblot following transformation and selection on nutrient-deficient medium. Table 14 shows that the yield of $hA_{2A}R$-Ura3p from A2H1 was almost 7-fold higher than the yield of $hA_{2A}R$ from the A2 control. No $hA_{2A}R$-Ura3p was detected from the A2U1 transformant. The yield of $hA_{2A}R$-Ura3p from A2SU1 was just over half that of A2, with spt3Δ:$hA_{2A}R$, 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 $hA_{2A}R$-Ura3p we had produced was correctly folded and thereby estimate the functional yield of the $hA_{2A}R$ moiety, a radio-ligand binding assay was performed [81]. Radio-ligand binding analysis was done using the well-characterized antagonist [^{3}H]ZM241385 [82] on 100 μg of total membrane extract from A2, A2H1 and A2U1. Table 14 shows that A2H1 produced only a minimal increase (1.6 ± 0.1 pmol mg^{-1}) of correctly-folded $hA_{2A}R$-Ura3p compared to the A2 control (1.1 ± 0.1 pmol mg^{-1}). The yield of $hA_{2A}R$-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 $hA_{2A}R$-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 spt3Δ:$hA_{2A}R$ (Table 14). Notably, ligand binding activity could be recovered from A2H1, but not A2 or A2SU1, by solubilising the $hA_{2A}R$-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 B_{max} values from these experiments validate the single-point saturation values. Affinity was determined using competition binding experiments, with the pK_d 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 $hA_{2A}R$-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 pYX222-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 $hA_{2A}R$ -Ura3p fusion could be reduced by using the BY4741 *spt3Δ* strain. Homologous competition radio-ligand binding with [^{3}H]ZM241385 (Fig. 10B) demonstrated that $hA_{2A}R$ and $hA_{2A}R$-Ura3p had comparable pK_d values (8.3–8.6) as reported in the literature [83].

When nutrient selection was used as a strategy to increase the yield

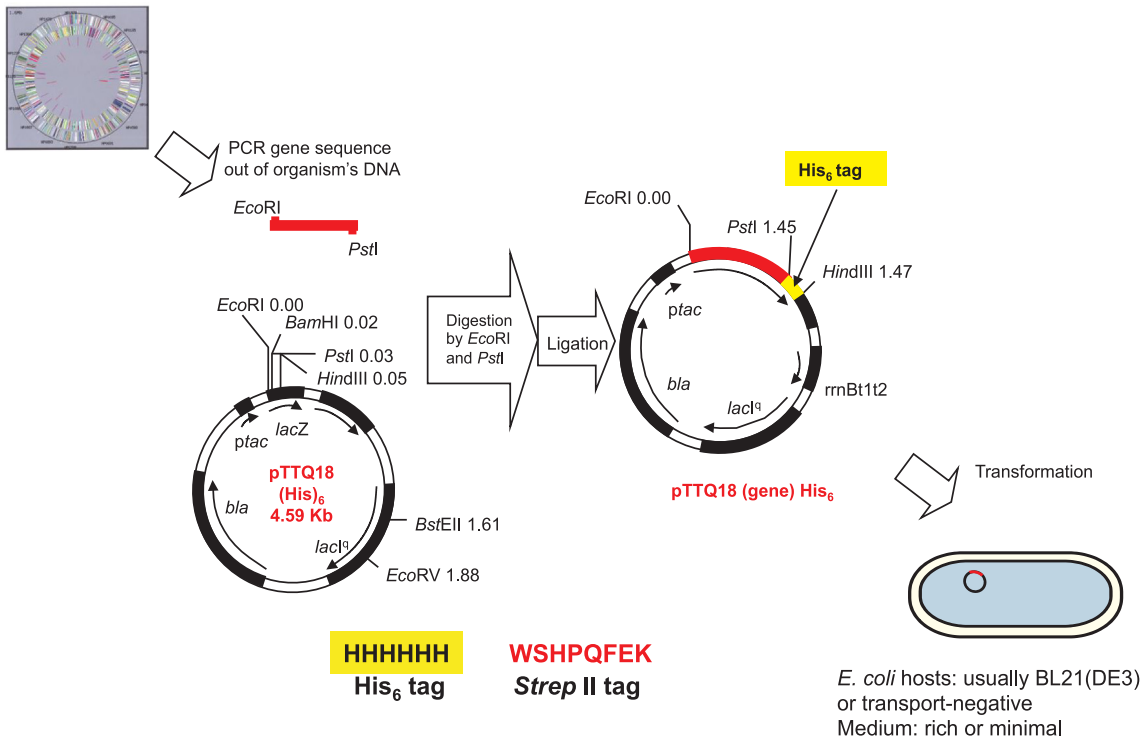

Fig. 8. Strategy for cloning and expressing genes of bacterial membrane proteins using plasmid pTTQ18-His_6. Each target gene was inserted into the multiple cloning site (MCS) downstream of the *tac* promoter in the plasmid pTTQ18-His_6 in order to amplify gene expression. Two different restriction enzymes, *Eco*RI and *Pst*I were used to ensure correct orientation of the gene on ligation into plasmid pTTQ18-His_6, as well as to prevent re-ligation of the plasmid. First, the membrane protein gene was amplified by PCR using bacterial genomic DNA as template and introducing *Eco*RI at the 5′-end and *Pst*I at the 3′-end, followed by digestion with these two enzymes and ligation with *Eco*RI-*Pst*I – digested pTTQ18-His_6. The resulting plasmid construct with the gene inserted was then transformed into *E. coli* XL10-Gold cells, followed by colony PCR to identify positive clones. The sequence encoding the hexahistidine tag (yellow) is incorporated into the pTTQ18 plasmid so it is in frame with the ligated gene. This works well when the carboxyl terminus of the recombinant protein is finally located inside the cell membrane. However, if the carboxyl terminus is destined to be outside the cell membrane, translocation of the fused positively-charged histidines appears to compromise expression. In this latter case, fusion a Strep II tag (red) can be used instead. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of another recombinant GPCR, the human β2 adrenergic receptor ($h\beta_{2A}R$), high yielding transformants were found to have been generated using the 2-step method. B2U1 and B2U5 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 $h\beta_{2A}R$-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. However solubilization using 2.5% styrene maleic acid co-polymer (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 $hA_{2A}R$ 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 experimentally [85]. One approach is to screen on increasing concentrations of antibiotic (usually zeocin) to obtain so-called 'jackpot' clones. However, the correlation between the copy number of the integrated expression cassette (as determined by resistance to increasing zeocin concentrations) and the final yield of recombinant protein is not always positive [16]. Sometimes clones with lower copy numbers are more productive, suggesting that the cellular machinery is overwhelmed in jackpot clones (resulting in misfolded or degraded protein). Consistent with this idea, $hA_{2A}R$ yields were increased 1.8-fold when the corresponding gene was co-expressed in *P. pastoris* with the stress-response gene *HAC1* [86]; Hac1 drives transcription of UPR genes.

In contrast to the situation in *S. cerevisiae*, far fewer *P. pastoris*

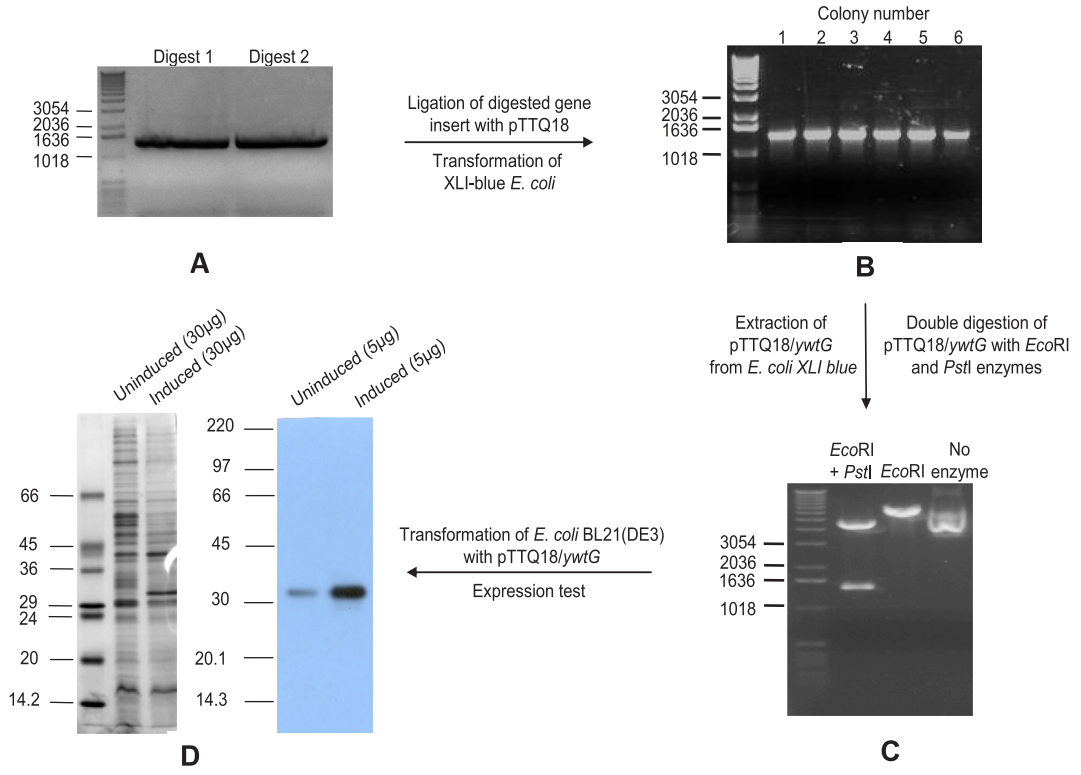

Fig. 9. Production of the YwtGHis$_6$ protein from *B. subtilis* in *E. coli.* (A) Amplification and digestion of the *ywtG* gene. (B) PCR screening of carbenicillin-resistant *E. coli* XL1-Blue colonies. (C) Identification of the *ywtG* gene using double restriction digestion of pTTQ18/*ywtG* with *Eco*RI and *Pst*I. (D) Coomassie Brilliant Blue stained gel (left panel) of membrane preparations made from induced and uninduced cells (as indicated) and Western blot analysis (right panel) of the overexpressed YwtGHis$_6$ protein. There is some expression in the uninduced cells, indicative of 'leaky' expression. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 $(Man)_8$-$(GlcNAc)_2$ 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 $hA_{2A}R$ following the 'Pichia Fermentation Process Guidelines' (Invitrogen). The $hA_{2A}R$ 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 pPICZαA 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 $hA_{2A}R$ 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 His$_6$ 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	41	0.78
BC0935	49,906	6.2	34	4.75	35	0.68
BC5418	45,541	6.6	31	5.0	31	0.68
YhjI	49,477	6.5	32	4.85	35	0.65

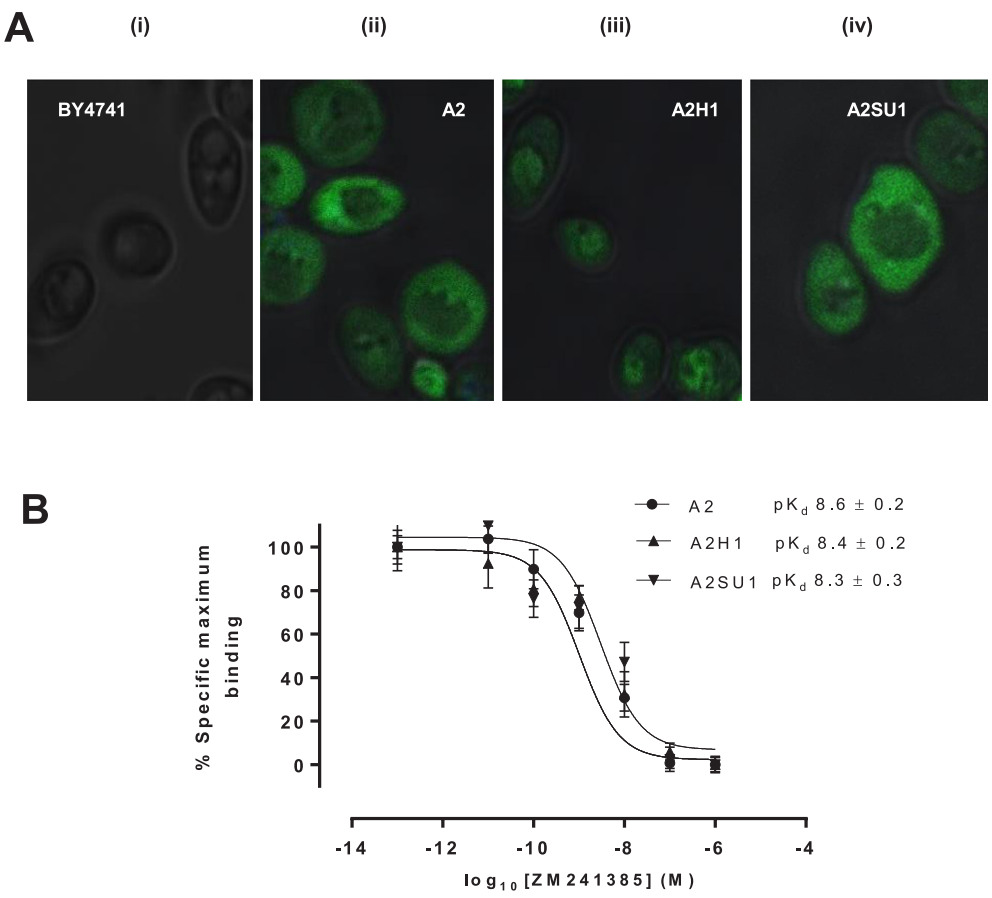

Fig. 10. Localization and pharmacological analysis of recombinant $hA_{2A}R/hA_{2A}R$-Ura3p expression in *S. cerevisiae*. (A) Confocal microscopy visualization of (i) control BY4241 cells, (ii) recombinant $hA_{2A}R$ produced from transformant A2, (iii) recombinant $hA_{2A}R$-Ura3p expressed after a 1-step selection from transformant A2H2 and (iv) recombinant $hA_{2A}R$-Ura3p expressed after a 2-step selection from transformant A2SU1. Cells were grown to $A_{600} = 4–5$ and were visualized using rabbit anti-His_6 (Clontech) as the primary antibody and an Alexa-Fluor488-conjugated anti-rabbit secondary antibody. (B) Homologous competition binding experiments were performed for $hA_{2A}R/hA_{2A}R$-Ura3p produced from A2, A2H2 and A2SU1 using labelled ($[^3H]$) 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).

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 the same cells cultured in any other non-inducing growth condition

Table 14
Characterization of recombinant $hA_{2A}R/hA_{2A}R$-Ura3p- and β_2AR/β_2AR-Ura3p-producing transformants. Confocal microscopy visualization of recombinant $hA_{2A}R/hA_{2A}R$-Ura3p in transformed *S. cerevisiae* using AlexFluor488 antibodies was done to assess whether $hA_{2A}R/hA_{2A}R$-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_6 antibody) were quantified using ImageJ. This allowed comparison of recombinant protein yield from A2H1, A2U1 and A2SU1 (BY4741 or *spt3Δ* were transformed with pYX222-h*A2AR-URA3* and grown under conditions of nutrient selection) compared with the A2 control (BY4741 transformed with pYX222-h*A2AR*) or *spt3Δ*:$hA_{2A}R$ (*spt3Δ* transformed with pYX222-h*A2AR*). For $h\beta_{2A}R$, B2H3, B2U1 and B2U5 were compared with the B2 control. The functional yield for $hA_{2A}R/hA_{2A}R$-Ura3p or β_2AR/β_2AR-Ura3p in yeast cell membranes or following solubilization with 2.5% DDM, 0.5% CHS was determined by single-point saturation binding using the antagonists $[^3H]$ZM241385 or $[^3H]$CGP 12 177, respectively and 100 µg total membrane protein per experiment, as described in [73]. A 1-way ANOVA with a Holm-Sidak's multiple comparison test gave p = 0.001 (***) for A2H1 (without DDM treatment) versus A2H1(solubilized with DDM). Additionally, β_2AR/β_2AR-Ura3p was solubilized using 2.5% (w/v) styrene maleic acid (SMA) polymer with a 2:1 ratio of styrene to maleic acid. All data are derived from at least 3 independent biological replicates, with error ± SEM in parenthesis, where applicable.

Transformant name	Expression strain	Vacuolar Internalization	Total yield from immunoblot (Arbitrary units; relative to control)	Functional yield (pmol mg^{-1})	Functional yield following DDM solubilization (pmol mg^{-1})	Functional yield following SMA solubilization (pmol mg^{-1})
$hA_{2A}R$						
A2 (control)	BY4741	No	1.0	1.1 (0.2)	1.2 (0.3)	n/a
A2H1	BY4741	Yes	6.8 (0.8)	1.6 (0.1)	5.8 (1.6)	n/a
A2U1	BY4741	n/a	0.0	0.2 (0.01)	n/a	n/a
A2SU1	*spt3Δ*	No	0.6 (0.2)	3.0 (0.2)	3.2 (0.7)	n/a
spt3Δ:$hA_{2A}R$	*spt3Δ*	No	n/a	0.5 (0.1)	n/a	n/a
β_2AR						
B2 (control)	BY4741	n/a	1.0	0.0	0.3 (0.02)	1.26 (0.2)
B2H3	BY4741	n/a	2.4 (1.5)	0.0	0.3 (0.1)	0.93 (0.3)
B2U1	BY4741	n/a	3.2 (0.3)	0.0	1.6 (0.3)	1.43 (0.4)
B2U5	BY4741	n/a	2.5 (0.2)	0.0	2.2 (1.0)	1.12 (0.6)

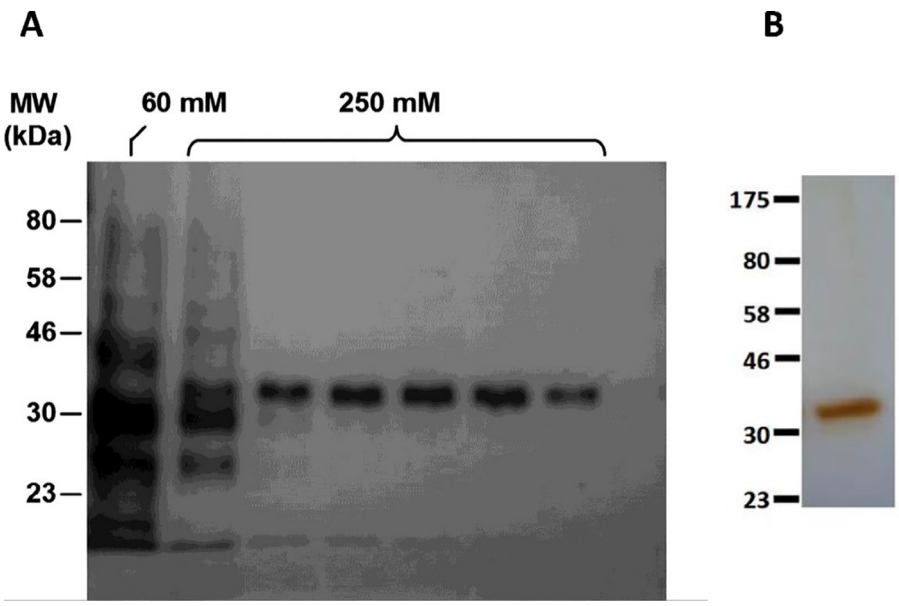

Fig. 11. Purification of the human adenosine A_{2A} receptor in *P. pastoris* under the control of P_{AOX1}. (A) Human A_{2A}R eluted from Ni^{2+}–NTA linked agarose as a single band in Coomassie-stained fractions with 250 mM imidazole. (B) Silver-stained band of the 250 mM imidazole fraction with an anti-histidine antibody.

[90]. 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_{AOX1} [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 heterologous 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* expression 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

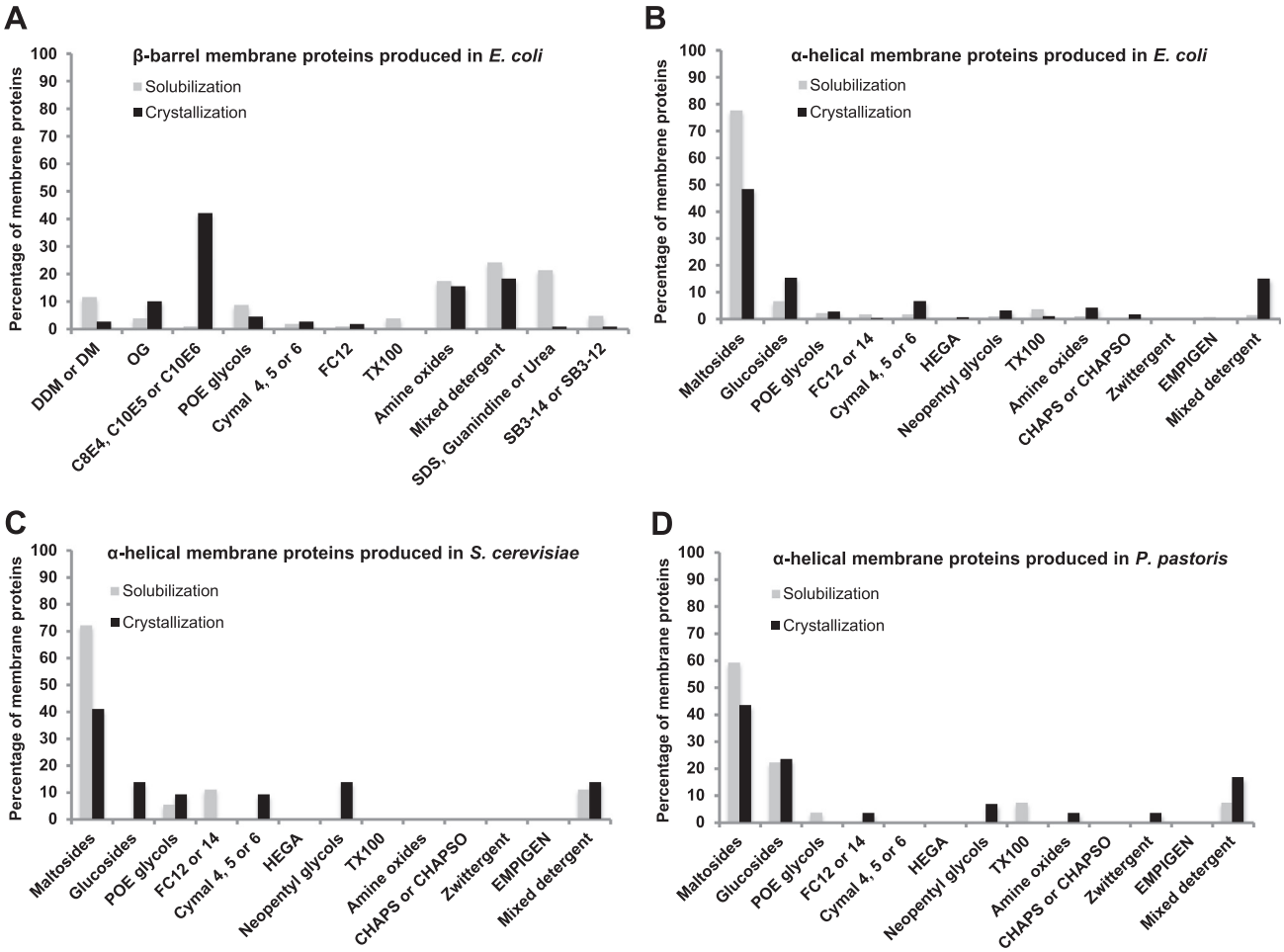

Fig. 12. Detergent usage for the solubilization and crystallization of membrane proteins. Data were obtained from Tables 1–4. Detergent usage for the solubilization (grey) and crystallization (black) of (A) β-barrel membrane proteins produced in *E. coli* and of α-helical membrane proteins produced in (B) *E. coli*, (C) *S. cerevisiae* or (D) *P. pastoris*. Detergents are classified as follows: Maltosides (DDM, UDM, Tri-DM, DM, NM, OM and TDM); Glucosides (OG, OTG, NG, C7G); Polyoxyethyleneglycols (C12E9, C12E8, C10E5, C8E4, OPOE and C12E7); Fos cholines (FC12 and FC14); Cymals (Cymal 4, Cymal 5 and Cymal 6); Neopentylglycols (LMNG, DMNG and OGNG); Amine oxides (LDAO and DDAO); Zwittergent (SB-3 14 and SB-3 12); Mixed (Mixed detergents).

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, may 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 *Lactobacillus*. 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 (EMeP, 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 studentship 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). Fermenters in PJFH's laboratory and allied equipment were funded by the BBSRC (MPSI BBS/B/14418), the Wellcome Trust (JIF 062164/Z/00/Z) and the University of Leeds. We acknowledge the 'Initiative d'Excellence' program from the French State (Grant 'DYNAMO', ANR-11-LABEX-0011-01). MP's PhD fellowship is funded by the French Ministry of Research and Education.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ymeth.2018.04.009.

References

[1] R.M. Bill, P.J. Henderson, S. Iwata, E.R. Kunji, H. Michel, R. Neutze, S. Newstead, B. Poolman, C.G. Tate, H. Vogel, Nat. Biotechnol. 29 (2011) 335–340.
[2] J. Deisenhofer, O. Epp, I. Sinning, H. Michel, J. Mol. Biol. 246 (1995) 429–457.
[3] G. Chang, R.H. Spencer, A.T. Lee, M.T. Barclay, D.C. Rees, Science 282 (1998) 2220–2226.
[4] D.A. Doyle, J. Morais Cabral, R.A. Pfuetzner, A. Kuo, J.M. Gulbis, S.L. Cohen, B.T. Chait, R. MacKinnon, Science 280 (1998) 69–77.
[5] M. Jidenko, R.C. Nielsen, T.L. Sorensen, J.V. Moller, M. le Maire, P. Nissen, C. Jaxel, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 11687–11691.
[6] S.B. Long, E.B. Campbell, R. Mackinnon, Science 309 (2005) 897–903.
[7] D. Elmlund, S.N. Le, H. Elmlund, Curr. Opin. Struct. Biol. 46 (2017) 1–6.
[8] G. Hattab, D.E. Warschawski, K. Moncoq, B. Miroux, Sci. Rep. 5 (2015) 12097.
[9] M. Hattori, R.E. Hibbs, E. Gouaux, Structure 20 (2012) 1293–1299.
[10] D. Drew, M. Lerch, E. Kunji, D.J. Slotboom, J.W. de Gier, Nat. Methods 3 (2006) 303–313.
[11] Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, S. Graslund, P. Nordlund, J. Weigelt, B.M. Hallberg, J. Bray, O. Gileadi, S. Knapp, U. Oppermann, C. Arrowsmith, R. Hui, J. Ming, S. dhe-Paganon, H.W. Park, A. Savchenko, A. Yee, A. Edwards, R. Vincentelli, C. Cambillau, R. Kim, S.H. Kim, Z. Rao, Y. Shi, T.C. Terwilliger, C.Y. Kim, L.W. Hung, G.S. Waldo, Y. Peleg, S. Albeck, T. Unger, O. Dym, J. Prilusky, J.L. Sussman, R.C. Stevens, S.A. Lesley, I.A. Wilson, A. Joachimiak, F. Collart, I. Dementieva, M.I. Donnelly, W.H. Eschenfeldt, Y. Kim, L. Stols, R. Wu, M. Zhou, S.K. Burley, J.S. Emtage, J.M. Sauder, D. Thompson, K. Bain, J. Luz, T. Gheyi, F. Zhang, S. Atwell, S.C. Almo, J.B. Bonanno, A. Fiser, S. Swaminathan, F.W. Studier, M.R. Chance, A. Sali, T.B. Acton, R. Xiao, L. Zhao, L.C. Ma, J.F. Hunt, L. Tong, K. Cunningham, M. Inouye, S. Anderson, H. Janjua, R. Shastry, C.K. Ho, D. Wang, H. Wang, M. Jiang, G.T. Montelione, D.I. Stuart, R.J. Owens, S. Daenke, A. Schutz, U. Heinemann, S. Yokoyama, K. Bussow, K.C. Gunsalus, Nat Methods 5 (2008) 135–146.
[12] S.K. Kwon, S.K. Kim, D.H. Lee, J.F. Kim, Sci. Rep. 5 (2015) 16076.
[13] M. Way, B. Pope, J. Gooch, M. Hawkins, A.G. Weeds, EMBO J. 9 (1990) 4103–4109.
[14] G.L. Orriss, M.J. Runswick, I.R. Collinson, B. Miroux, I.M. Fearnley, J.M. Skehel, J.E. Walker, Biochem. J. 314 (Pt. 2) (1996) 695–700.
[15] S. Wagner, M.M. Klepsch, S. Schlegel, A. Appel, R. Draheim, M. Tarry, M. Hogbom, K.J. van Wijk, D.J. Slotboom, J.O. Persson, J.W. de Gier, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 14371–14376.
[16] R. Aw, K.M. Polizzi, Microb. Cell Fact. 12 (2013) 128.
[17] K.E. Royle, K. Polizzi, Sci. Rep. 7 (2017) 15817.
[18] T. Sasagawa, M. Matsui, Y. Kobayashi, M. Otagiri, S. Moriya, Y. Sakamoto, Y. Ito, C.C. Lee, K. Kitamoto, M. Arioka, Plasmid 65 (2011) 65–69.
[19] B. Byrne, Curr. Opin. Struct. Biol. 32c (2015) 9–17.
[20] A. Siwaszek, M. Ukleja, A. Dziembowski, RNA Biol. 11 (2014) 1122–1136.
[21] S. Schlegel, J. Lofblom, C. Lee, A. Hjelm, M. Klepsch, M. Strous, D. Drew, D.J. Slotboom, J.W. de Gier, J. Mol. Biol. 423 (2012) 648–659.
[22] S. Mahalik, A.K. Sharma, K.J. Mukherjee, Microb. Cell Fact. 13 (2014) 177.
[23] D.C. Rowe, D.K. Summers, Appl. Environ. Microbiol. 65 (1999) 2710–2715.
[24] A.Z. Shaw, B. Miroux, Methods Mol. Biol. 228 (2003) 23–35.
[25] M. Zoonens, B. Miroux, Methods Mol. Biol. 601 (2010) 49–66.
[26] J. Roge, J.M. Betton, Microb. Cell Fact. 4 (2005) 18.
[27] J.W. Fairman, N. Dautin, D. Wojtowicz, W. Liu, N. Noinaj, T.J. Barnard, E. Udho, T.M. Przytycka, V. Cherezov, S.K. Buchanan, Structure 20 (2012) 1233–1243.
[28] B. Walse, V.T. Dufe, B. Svensson, I. Fritzson, L. Dahlberg, A. Khairoullina, U. Wellmar, S. Al-Karadaghi, Biochemistry 47 (2008) 8929–8936.
[29] S. Alfasi, Y. Sevastsyanovich, L. Zaffaroni, L. Griffiths, R. Hall, J. Cole, J. Biotechnol. 156 (2011) 11–21.
[30] B. Miroux, J.E. Walker, J. Mol. Biol. 260 (1996) 289–298.
[31] F.W. Studier, Protein Expr. Purif. 41 (2005) 207–234.
[32] K. von Meyenburg, B.B. Jorgensen, B. van Deurs, EMBO J. 3 (1984) 1791–1797.
[33] I. Arechaga, B. Miroux, S. Karrasch, R. Huijbregts, B. de Kruijff, M.J. Runswick, J.E. Walker, FEBS Lett. 482 (2000) 215–219.
[34] J. Lefman, P. Zhang, T. Hirai, R.M. Weis, J. Juliani, D. Bliss, M. Kessel, E. Bos, P.J. Peters, S. Subramaniam, J. Bacteriol. 186 (2004) 5052–5061.
[35] W.O. Wilkison, J.P. Walsh, J.M. Corless, R.M. Bell, J. Biol. Chem. 261 (1986) 9951–9958.
[36] J.H. Weiner, B.D. Lemire, M.L. Elmes, R.D. Bradley, D.G. Scraba, J. Bacteriol. 158 (1984) 590–596.
[37] E. van den Brink-van, J.W. der Laan, R.E. Boots, G.M. Spelbrink, E. Kool, J.A. Breukink, B. de Killian, Kruijff, J. Bacteriol. 185 (2003) 3773–3779.
[38] H.M. Eriksson, P. Wessman, C. Ge, K. Edwards, A. Wieslander, J. Biol. Chem. 284 (2009) 33904–33914.
[39] C. Ge, J. Gomez-Llobregat, M.J. Skwark, J.M. Ruysschaert, A. Wieslander, M. Linden, FEBS J. 281 (2014) 3667–3684.
[40] L. Danne, M. Aktas, A. Unger, W.A. Linke, R. Erdmann, F. Narberhaus, MBio 8 (2017).
[41] J. Shin, Y.H. Jung, D.H. Cho, M. Park, K.E. Lee, Y. Yang, C. Jeong, B.H. Sung, J.H. Sohn, J.B. Park, D.H. Kweon, Enzyme Microb. Technol. 79–80 (2015) 55–62.
[42] P.J. Walser, N. Ariotti, M. Howes, C. Ferguson, R. Webb, D. Schwudke, N. Leneva, K.J. Cho, L. Cooper, J. Rae, M. Floetenmeyer, V.M. Oorschot, U. Skoglund, K. Simons, J.F. Hancock, R.G. Parton, Cell 150 (2012) 752–763.
[43] M.C. Huber, A. Schreiber, P. von Olshausen, B.R. Varga, O. Kretz, B. Joch, S. Barnert, R. Schubert, S. Eimer, P. Kele, S.M. Schiller, Nat. Mater. 14 (2015) 125–132.
[44] S.E. Deacon, P.C. Roach, V.L. Postis, G.S. Wright, X. Xia, S.E. Phillips, J.P. Knox, P.J. Henderson, M.J. McPherson, S.A. Baldwin, Mol. Membr. Biol. 25 (2008) 588–598.
[45] M.J. Stark, Gene 51 (1987) 255–267.
[46] S. Surade, M. Klein, P.C. Stolt-Bergner, C. Muenke, A. Roy, H. Michel, Protein Sci. 15 (2006) 2178–2189.
[47] D.N. Wang, M. Safferling, M.J. Lemieux, H. Griffith, Y. Chen, X.D. Li, BBA 1610 (2003) 23–36.
[48] M. Saidijam, G. Benedetti, Q. Ren, Z. Xu, C.J. Hoyle, S.L. Palmer, A. Ward, K.E. Bettaney, G. Szakonyi, J. Meuller, S. Morrison, M.K. Pos, P. Butaye, K. Walravens, K. Langton, R.B. Herbert, R.A. Skurray, I.T. Paulsen, J. O'Reilly, N.G. Rutherford, M.H. Brown, R.M. Bill, P.J. Henderson, Curr. Drug Targets 7 (2006) 793–811.
[49] M. Saidijam, K.E. Bettaney, G. Szakonyi, G. Psakis, K. Shibayama, S. Suzuki, J.L. Clough, V. Blessie, A. Abu-Bakr, S. Baumberg, J. Meuller, C.K. Hoyle, S.L. Palmer, P. Butaye, K. Walravens, S.G. Patching, J. O'Reilly, N.G. Rutherford, R.M. Bill, D.I. Roper, M.K. Phillips-Jones, P.J. Henderson, Biochem. Soc. Trans. 33 (2005) 867–872.
[50] M. Saidijam, G. Psakis, J.L. Clough, J. Meuller, S. Suzuki, C.J. Hoyle, S.L. Palmer, S.M. Morrison, M.K. Pos, R.C. Essenberg, M.C. Maiden, A. Abu-bakr, S.G. Baumberg, A.A. Neyfakh, J.K. Griffith, M.J. Stark, A. Ward, J. O'Reilly, N.G. Rutherford, M.K. Phillips-Jones, P.J. Henderson, FEBS Lett. 555 (2003) 170–175.
[51] G. Szakonyi, D. Leng, P. Ma, K.E. Bettaney, M. Saidijam, A. Ward, S. Zibaei, A.T. Gardiner, R.J. Cogdell, P. Butaye, A.B. Kolsto, J. O'Reilly, R.J. Hope, N.G. Rutherford, C.J. Hoyle, P.J. Henderson, J. Antimicrob. Chemother. 59 (2007) 1265–1270.
[52] P.C.J. Roach, J. O'Reilly, H. Norbertczak, R.A. Hope, H. Venter, S.G. Patching, M. Jamshad, P.G Stockley, S.A. Baldwin, R.B. Herbert, N.G. Rutherford, R.M. Bill, P.J. F. Henderson in Practical Fermentation Technology (B. McNeil and L.M. Harvey) Chapter 3 (2008) pp. 37–67. J. Wiley, Chichester, UK.
[53] W.J. Liang, K.J. Wilson, H. Xie, J. Knol, S. Suzuki, N.G. Rutherford, P.J. Henderson, R.A. Jefferson, J. Bacteriol. 187 (2005) 2377–2385.
[54] P. Ma, F. Varela, M. Magoch, A.R. Silva, A.L. Rosario, J. Brito, T.F. Oliveira, P. Nogly, M. Pessanha, M. Stelter, A. Kletzin, P.J. Henderson, M. Archer, PLoS One 8 (2013) e76913.
[55] S. Suzuki, P.J. Henderson, J. Bacteriol. 188 (2006) 3329–3336.
[56] P. Ma, H.M. Yuille, V. Blessie, N. Gohring, Z. Igloi, K. Nishiguchi, J. Nakayama, P.J. Henderson, M.K. Phillips-Jones, Mol. Membr. Biol. 25 (2008) 449–473.
[57] C.A. Potter, A. Ward, C. Laguri, M.P. Williamson, P.J. Henderson, M.K. Phillips-Jones, J. Mol. Biol. 320 (2002) 201–213.
[58] K.A. Hassan, Q. Liu, P.J. Henderson, I.T. Paulsen, MBio 6 (2015).
[59] R. Rani, A. Mannan, M.A. Yameen, B. Mirza, Miner. Biotecnol. 25 (2013) 95–100.
[60] K.A. Hassan, Z. Xu, R.E. Watkins, R.G. Brennan, R.A. Skurray, M.H. Brown, Protein Expr. Purif. 64 (2009) 118–124.
[61] M. Rahman, F. Ismat, M.J. McPherson, S.A. Baldwin, Mol. Membr. Biol. 24 (2007) 407–418.
[62] Z. Bawa, C.E. Bland, N. Bonander, N. Bora, S.P. Cartwright, M. Clare, M.T. Conner, R.A. Darby, M.V. Dilworth, W.J. Holmes, M. Jamshad, S.J. Routledge, S.R. Gross, R.M. Bill, Biochem. Soc. Trans. 39 (2011) 719–723.
[63] S.J. Routledge, L. Mikaliunaite, A. Patel, M. Clare, S.P. Cartwright, Z. Bawa, M.D. Wilks, F. Low, D. Hardy, A.J. Rothnie, R.M. Bill, Methods 95 (2016) 26–37.
[64] R.A. Darby, S.P. Cartwright, M.V. Dilworth, R.M. Bill, Methods Mol. Biol. 866 (2012) 11–23.
[65] T. Hino, T. Arakawa, H. Iwanari, T. Yurugi-Kobayashi, C. Ikeda-Suno, Y. Nakada-Nakura, O. Kusano-Arai, S. Weyand, T. Shimamura, N. Nomura, A.D. Cameron, T. Kobayashi, T. Hamakubo, S. Iwata, T. Murata, Nature 482 (2012) 237–240.
[66] T. Shimamura, M. Shiroishi, S. Weyand, H. Tsujimoto, G. Winter, V. Katritch, R. Abagyan, V. Cherezov, W. Liu, G.W. Han, T. Kobayashi, R.C. Stevens, S. Iwata, Nature 475 (2011) 65–70.
[67] Y. Tanaka, S. Iwaki, T. Tsukazaki, Structure 25 (2017) 1455–1460.
[68] K.W. Huynh, M.R. Cohen, J. Jiang, A. Samanta, D.T. Lodowski, Z.H. Zhou, V.Y. Moiseenkova-Bell, Nat. Commun. 7 (2016) 11130.
[69] K. Kapoor, J.S. Finer-Moore, B.P. Pedersen, L. Caboni, A. Waight, R.C. Hillig, P. Bringmann, I. Heisler, T. Muller, H. Siebeneicher, R.M. Stroud, Proc. Natl. Acad. Sci. U.S.A. 113 (2016) 4711–4716.
[70] M. Shiroishi, H. Tsujimoto, H. Makyio, H. Asada, T. Yurugi-Kobayashi, T. Shimamura, T. Murata, N. Nomura, T. Haga, S. Iwata, T. Kobayashi, Microb. Cell Fact. 11 (2012) 78.
[71] K. De Schutter, Y.C. Lin, P. Tiels, A. Van Hecke, S. Glinka, J. Weber-Lehmann, P. Rouze, Y. Van de Peer, N. Callewaert, Nat. Biotechnol. 27 (2009) 561–566.
[72] B. Gasser, M. Maurer, J. Rautio, M. Sauer, A. Bhattacharyya, M. Saloheimo, M. Penttila, D. Mattanovich, BMC Genomics 8 (2007) 179.
[73] N. Bonander, R.A. Darby, L. Grgic, N. Bora, J. Wen, S. Brogna, D.R. Poyner, M.A. O'Neill, R.M. Bill, Microb. Cell Fact. 8 (2009) 10.
[74] G. Giaever, C. Nislow, Genetics 197 (2014) 451–465.
[75] C. Arico, C. Bonnet, C. Javaud, Methods Mol. Biol. 988 (2013) 45–57.
[76] S.P. Cartwright, R.A. Darby, D. Sarkar, N. Bonander, S.R. Gross, M.P. Ashe, R.M. Bill, Microb. Cell Fact. 16 (2017) 41.
[77] N. Gul, D.M. Linares, F.Y. Ho, B. Poolman, J. Mol. Biol. 426 (2014) 136–149.
[78] D.M. Linares, E.R. Geertsma, B. Poolman, J. Mol. Biol. 401 (2010) 45–55.

[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. Kara, 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. Kara, 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. Katsuta, 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. Rausch, T.A. Stadheim, H. Wischnewski, S. Wildt, T.U. Gerngross, Science 301 (2003) 1244–1246.
[90] R. Prielhofer, S.P. Cartwright, A.B. Graf, M. Valli, R.M. Bill, D. Mattanovich, B. Gasser, BMC Genomics 16 (2015) 167.
[91] Z. Bawa, 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. Lohsse, S. Borg, O. Raschdorf, C. Jogler, Q. Tu, M. Posfai, E. Tompa, J.M. Plitzko, A. Brachmann, G. Wanner, R. Muller, Y. Zhang, D. Schuler, Nat. Nanotechnol. 9 (2014) 193–197.