

REVIEW

The recombinant expression systems for structure determination of eukaryotic membrane proteins

Yuan He^{1,2}, Kan Wang³, Nieng Yan^{1,2}✉¹ State Key Laboratory of Bio-membrane and Membrane Biotechnology, Tsinghua university, Beijing 100084, China² Center for Structural Biology, School of Medicine, Tsinghua university, Beijing 100084, China³ China-Japan Friendship Hospital, Beijing 100029, China✉ Correspondence: nyan@tsinghua.edu.cn (N. Yan)

Received May 16, 2014 Accepted June 16, 2014

ABSTRACT

Eukaryotic membrane proteins, many of which are key players in various biological processes, constitute more than half of the drug targets and represent important candidates for structural studies. In contrast to their physiological significance, only very limited number of eukaryotic membrane protein structures have been obtained due to the technical challenges in the generation of recombinant proteins. In this review, we examine the major recombinant expression systems for eukaryotic membrane proteins and compare their relative advantages and disadvantages. We also attempted to summarize the recent technical strategies in the advancement of eukaryotic membrane protein purification and crystallization.

KEYWORDS eukaryotic membrane proteins, recombinant expression, structural biology, integral membrane proteins (IMPs), fluorescence detected size exclusion chromatography (FSEC), protein purification and crystallization

INTRODUCTION

It is estimated that approximately 30% of the protein-coding genes are for integral membrane proteins (IMPs) in human (Overington et al., 2006; Murray et al., 2012). IMPs are critical players for many important physiological processes including metabolism, signal transduction, and energy conversion and utilization (Krogh et al., 2001). Aberrant expressions and activities of IMPs are associated with a variety of diseases such as cancer, Alzheimer's disease,

and metabolic diseases (Ishikawa et al., 2004; Sanders and Myers, 2004; Overington et al., 2006; Aisenbrey et al., 2008; Bkaily and Al-Khoury, 2014). IMPs constitute more than 50% of the US Food and Drug Administration (FDA)-approved drug targets (Russell and Eggleston, 2000; Yildirim et al., 2007). Structures of eukaryotic membrane proteins are actively pursued for structure-based drug development.

In contrast to their physiological and pathophysiological significance, the progress on the structure biology of IMPs, particularly eukaryotic IMPs, has been relatively slow. By the end of March 2014, in total 466 unique membrane protein structures have been reported (Snider and Stephen, 2014), the majority of which are of prokaryotic origins. With respect to eukaryotic IMPs, more than half of the determined structures are for proteins obtained from endogenous sources (Bill et al., 2011). These proteins, exemplified by the electron transport chain complexes (Tsukahara et al., 1996; Xia et al., 1997; Sun et al., 2005), ATP synthases (Abrahams et al., 1994; Liu et al., 2004; Amunts et al., 2007), and photosystems (Kurusu et al., 2003; Liu et al., 2004; Amunts et al., 2007), usually exist in abundance and are biochemically stable, hence representing ideal candidates for structural analysis. However, the total types of endogenously abundant eukaryotic IMPs are limited. The majority of IMPs exist in low copies in the host species. Therefore, structural determination of most eukaryotic IMPs requires recombinant expression of the target proteins. The first atomic-resolution structure of a eukaryotic IMP obtained through recombinant expression, Kv1.2, was reported in 2005 (Long et al., 2005). Ever since, less than seventy structures have been obtained for eukaryotic IMPs generated through recombinant expression systems (Fig. 1).

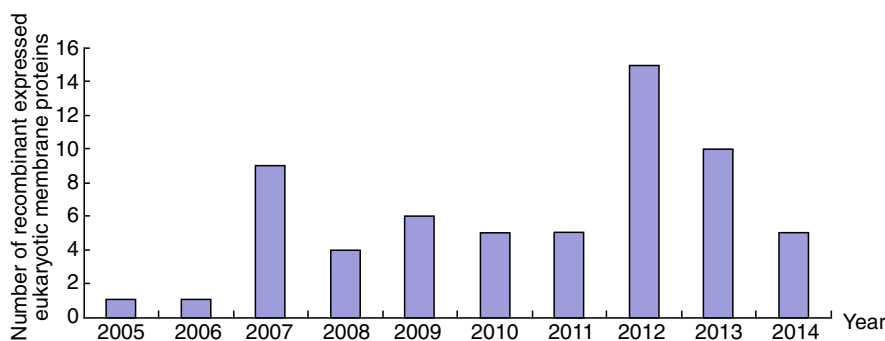


Figure 1. The development trends in recombinant expression eukaryotic membrane proteins. The structure number of eukaryotic membrane protein is limited by some obstacles such as low yield and instability in detergents. Since the first eukaryotic membrane protein structure was determined in 2005, over sixty structures have been emerged until now.

Out of the many challenges facing structural study of eukaryotic IMPs, production of sufficient quantities of well-behaved recombinant proteins represents the real technical bottleneck. Embedded in lipid bilayers, the structural integrity and proper functions of IMPs rely on the interactions with surrounding lipids (Phillips et al., 2009), which stabilize membrane proteins, provide lattice contacts, and in some occasions function as indispensable co-factors (van Meer et al., 2008). Recombinant expression of membrane proteins therefore requires a proper membrane environment. Whereas *Escherichia coli* proved to be the best host for most of prokaryotic IMPs of known structures, eukaryotic IMPs, with very few exceptions, requires eukaryotic expression systems including yeast, baculovirus-infected insect cells, and mammalian cells (Bill et al., 2011; Snider and Stephen, 2014).

In this review, in the hope of extracting some general principles on the expression and crystallization of eukaryotic membrane proteins, we examine the expression systems for the eukaryotic IMPs whose structures are obtained, attempt to summarize and compare the advantages and disadvantages of the representative recombinant expression systems, and delineate the detailed information in eukaryotic membrane protein purification and crystallization (Table 1).

RECOMBINANT EXPRESSION SYSTEMS FOR EUKARYOTIC MEMBRANE PROTEINS

The recombinantly expressed eukaryotic IMPs of known structures were obtained from four systems: *E. coli*, yeasts (*Pichia Pastoris* and *Saccharomyces cerevisiae*), insect cells, and mammalian cells. These expression systems have their respective advantages and disadvantages. The choice of an appropriate expression system remains empirical, largely depending on the biochemical and biological properties of the target proteins (Bernaudat et al., 2011). Among the recombinantly expressed eukaryotic IMPs whose structures have been solved, 4 were expressed in *E. coli*, 20 in yeast,

35 in insect cells, and 3 in mammalian cells. Below we will discuss these four expression systems.

E. coli

As the most frequently exploited recombinant expression system, *E. coli* BL21 (DE3) has the obvious advantage of rapid replication, time-saving operation, inexpensive cost, and mature and easy genetic manipulations (Sahdev et al., 2008). *E. coli* C43 (DE3) and C41 (DE3) strains were developed for over-expression of membrane proteins (Mironov and Walker, 1996; Dumon-Seignovert et al., 2004). Indeed, these *E. coli* strains were employed to over-express the large majority of prokaryotic IMPs whose structures were finally obtained. However, as the prokaryotic expression systems, they may lack the essential lipids, molecular chaperons, and post-translational modifications that are required for the correct membrane insertion, folding, and function of eukaryotic IMPs (Sahdev et al., 2008). As a result, only 4 structures were obtained for eukaryotic IMPs expressed in *E. coli* (Table 2). Despite the challenge to express eukaryotic membrane proteins in *E. coli*, researchers attempted to overcome these hurdles with codon-optimization (Burgess-Brown et al., 2008) and protein fusion with Mistic or GlpF tag to promote protein expression (AegeanSoftware, 2005; Drew et al., 2006; Neophytou et al., 2007), and co-expression of post-translational machineries to facilitate protein folding (Mironova et al., 2005; Mijakovic et al., 2006). Regardless of the effort, *E. coli* may not be an ideal system for eukaryotic IMP expression.

Yeast

Among the many yeast species, *Pichia Pastoris* (*Pichia*) and *Saccharomyces cerevisiae* (*S. cerevisiae*), which have been genetically well characterized, are the major systems to overexpress eukaryotic IMPs (Strausberg and Strausberg, 2001; Bornert et al., 2012). *Schizosaccharomyces pombe* is also employed for overexpression of IMPs, but not as widely

Table 1. Expression, purification, and crystallization information for eukaryotic membrane proteins*

No.	Expression systems	Protein name	Family	Extraction detergent	Purification detergent	Gel filtration detergent	Final concentration (mg/mL)	Temperature (°C)	Methods
1	<i>E. coli</i>	FLAP	MAPEG	DDM	DDM	C ₁₂ E ₈ + C ₆ E ₄			Sitting drop
2	C43(DE3)	PfAQP	Water channel	OG	OG	OG	6	18	Hanging drop
3	BI21(DE3)	Kir3.1-prokaryotic Kir channel chimera	Potassium channel	DDM	DDM	NG	8	20	Sitting drop
4	BI21(DE3)	Cytochrome b ₅₆₁	Electron transport chain complexes	DM	NM	NG		18	Hanging drop
5	Yeast <i>Pichia Pastoris</i>	Kv1.2 with β subunit	Potassium channel	DDM	DDM	DM	10	20	Hanging drop
6		Kv1.2-Kv2.1paddle	Potassium channel	DDM	DDM	CYMAL6 + CYMAL7	10	20	Hanging drop
7		Kv2.1paddle-Kv1.2 (F233 W)	Potassium channel	DDM	DDM	DDM	10	20	Hanging drop
8		Kir2.2 inward-rectifier	Potassium channel	DM	DM	DM	8	20	Hanging drop
9		GIRK2 (Kir3.2) K ⁺ channel	Potassium channel	DM	DM	DM	6-7	20	Hanging drop
10		K _{2P} 1.1(KWIK-1)	Potassium channel	DDM	DDM	DDM	10	20	Hanging drop
11		K _{2P} 4.1(TRAAK)	Potassium channel	DDM	DDM	DM	5	4	Hanging drop
12	<i>Pichia Pastoris</i>	Calcium release-activated calcium channel	Calcium channel	DDM	DDM	NM + NG	16	17	Hanging drop
13		SoPIP2;1	Water channel	OG	OG	OG	10	4	Hanging drop
14		HsAQP5	Water channel	NG	NG	NG	10	8	Hanging drop
15	Yeast	HsAQP4	Water channel	OG	OG	OG	30	25	Hanging drop
16		P-Glycoprotein	ABC transporter	Triton	DDM	DDM	10	4	Sitting drop
17		P-Glycoprotein	ABC transporter	DDM	DDM	UDM		4	Hanging drop
18		LTC ₄ S	MAPEG	DDM	DDM	DDM	6.5	4	Sitting drop
19		Histamine H ₁ receptor	GPCR	DDM	DDM	DDM	30-40	20	LCP
20	<i>S. cerevisiae</i>	AHA2 (H ⁺ pump)	Pump	DDM	DDM	C12E8 + CYMAL5	20-30	4	Hanging drop
21		VtH ⁺ -Ppase	M-PPase	DDM	DM	DM	10	20	Hanging drop

Table 1. continued

No.	Expression systems	Protein name	Family	Extraction detergent	Purification detergent	Gel filtration detergent	Final concentration (mg/mL)	Temperature (°C)	Methods
22		NRT1.1	MFS transporter	DDM	DDM	DDM	10	4	Hanging drop
23		CAAX protease Ste24p	Intramembrane protease	DDM	DDM/C ₁₂ E ₇	C ₁₂ E ₇	7.35	4/17	Hanging drop
24		PIPT	MFS transporter	DDM	DDM	NG	10–15	20	Hanging drop
25	Insect cell	<i>S. frugiperda</i> β ₂ AR (Fab)	GPCR	DDM	DDM	DDM	8–12	22	Bicelle
26		β ₂ AR (T4L)	GPCR	DDM	DDM	DDM	Concentrated	22	LCP
27		β ₂ AR-agonist complex	GPCR	MNG	MNG	MNG (0.1%)	50	20	LCP
28		β ₂ AR-GS complex	GPCR	MNG	MNG	MNG (0.1%)	90	20	LCP
29		A _{2A} adenosine receptor	GPCR	DDM	DDM	DDM	70	20	LCP
30		CXCR4	GPCR	DDM	DDM	DDM	60–70	20	LCP
31		Dopamine D ₃ receptor	GPCR	DDM	DDM	DDM	20–30	20	LCP
32		Sphingosine 1-phosphate receptor	GPCR	DDM	DDM	DDM	100	20	LCP
33		M ₂ muscarinic acetylcholine receptor	GPCR	Digitonin + Na-choleate	DM	MNG	30	20	LCP
34		M ₃ muscarinic acetylcholine receptor	GPCR	DDM	DDM	MNG	60	20	LCP
35		κ-Opioid receptor	GPCR	DDM	DDM	DDM	40	20	LCP
36		μ-Opioid receptor	GPCR	DDM + CHAPS + CHS	DDM + CHAPS + CHS	MNG + CHS	30	20	LCP
37		δ-Opioid receptor	GPCR	MNG + CHAPS + CHS	MNG + CHAPS + CHS	MNG + CHAPS + CHS	50	20	LCP
38	Insect cell	<i>S. frugiperda</i> N/OFQ receptor	GPCR	DDM + CHS	DDM + CHS	DDM + CHS	40	20	LCP
39		CCR5	GPCR	DDM + CHS	DDM + CHS	DDM + CHS	40–50	20	LCP
40		PAR1	GPCR	DDM + CHS	DDM + CHS	DDM + CHS	40–50	20	LCP
41		5-HT _{1B/2A} serotonin receptor	GPCR	DDM + CHS	DDM + CHS	DDM + CHS	50–80	20	LCP
42		Smoothed receptor	GPCR	DDM + CHS	DDM + CHS	DDM + CHS	50–60	20	LCP
43		Glucagon receptor	GPCR	DDM + CHS	DDM + CHS	DDM + CHS	80	20	LCP
44		Metabotropic Glutamate Receptor 1	GPCR	DDM + CHS	DDM + CHS	DDM + CHS	50–80	20	LCP
45		P2X ₄	Channel	DDM	DDM	DDM	2	4	Hanging drop

Table 1. continued

No.	Expression systems	Protein name	Family	Extraction detergent	Purification detergent	Gel filtration detergent	Final concentration (mg/mL)	Temperature (°C)	Methods
46		ASIC1	Channel	DDM	DDM	DDM	5	4	Hanging drop
47		GluA2	Channel	DDM	DDM	C ₁₁ Thio + lipids	2	4	Hanging drop
48		GLuCl α	Cys-loop receptor	DDM	DDM	DDM	2	4	Hanging drop
49		CX26	Gap junction	DDM	DDM	UDM	30	4	Hanging drop
50		UT-B	Urea Transporter	DM	DM	OG	8	4	Sitting drop
51		ZMPSTE24	Intramembrane protease	DDM + CHS or OGNG + CHS	DDM + CHS or OGNG + CHS	DDM + CHS or OGNG + CHS	9–11	20	Sitting drop
52	Insect cell	ABCB10	ABC transporter	DDM	DDM	DDM		20	Sitting drop
53		Claudin-15	Tight junction	DDM	LMNG	LMNG	7	20	LCP
54		NRT1.1	MFS transporter	DDM	DDM	DDM	10	4	Hanging drop
55	<i>Trichoplusia ni</i>	β 1 adrenergic receptor	GPCR	DM	Octylthioglucoiside	Octylthioglucoiside	6	18	Hanging drop
56		NTS1 Neurotensin Receptor	GPCR	MNG + CHS	MNG + CHS	MNG + CHS	20–25	20	LCP
57		CmCIC	H ⁺ /Cl ⁻ exchange transporter	DM	DM	DM	10	20	Hanging drop
58		Corticotropin-releasing factor receptor	GPCR	DM	DM	DM	20–30	22.5	LCP
59		GLUT1	MFS transporter	DDM	DDM	NG	10	4	Hanging drop
60	COS-1 cells	Rhodopsin	Rhodopsin	DDM	DDM	C ₈ E ₄			
61	Mammalian cell	RhCG	Channel	OG	OG	OG	5	20	Hanging drop
62		Dopamine transporter	Solute carrier transporter	DDM	DDM CHS POPC: POPE:POPG = 3:1:1	DM CHS POPE	3	4	Hanging drop

* The blank in the table is due to the details in the reported method has not been mentioned.

Table 2. *E. coli* as an expression system for eukaryotic membrane protein

Expression systems	No.	Protein	Species	PDB code	Reference	
<i>E. coli</i>	BI21(DE3)	1	FLAP	<i>Homo sapiens</i>	2Q7 M 2Q7R	Ferguson et al., 2007
	C43 (DE3)	2	PfAQP	<i>Plasmodium falciparum</i>	3C02	Newby et al., 2008
	BI21(DE3)	3	Kir3.1-prokaryotic Kir channel chimera	<i>Streptomyces lividans</i>	2QKS	Nishida et al., 2007
	BI21(DE3)	4	Cytochrome b ₅₆₁	<i>Arabidopsis thaliana</i>	4O6Y, 4O79, 4O7G	Lu et al., 2014

Table 3. Yeast as an expression system for eukaryotic membrane protein*

Expression systems	No.	Protein	Species	PDB code	Reference	
Yeast	<i>Pichia Pastoris</i>	1	Kv1.2 with β subunit	<i>Drosophila melanogaster</i>	2A79	Long et al., 2005
		2	Kv1.2-Kv2.1 paddle	<i>Rattus norvegicus</i>	2R9R	Long et al., 2007
		3	Kv2.1paddle-Kv1.2 (F233 W)	<i>Rattus norvegicus</i>	3LNM	Tao et al., 2010
		4	Kir2.2 Inward-Rectifier	<i>Gallus gallus</i>	3JYC	Tao et al., 2009
		5	GIRK2 (Kir3.2) channel	<i>Mus musculus</i>	3SYO	Whorton and MacKinnon, 2011
		6	K _{2P} 1.1 (KWIK-1)	<i>Homo sapiens</i>	3UKM	Miller and Long, 2012
		7	K _{2P} 4.1 (TRAAK)	<i>Homo sapiens</i>	3UM7	Brohawn et al., 2012
		8	Calcium release-activated calcium channel	<i>Drosophila melanogaster</i>	4HKR	Xiaowei Hou, 2012
		9	SoPIP2;1	<i>Spinacia oleracea</i>	1Z98 2B5F	Tomroth-Horsefield et al., 2006
		10	HsAQP5	<i>Homo sapiens</i>	3D9S	Horsefield et al., 2008
		11	HsAQP4	<i>Homo sapiens</i>	3GD8	Ho et al., 2009
		12	P-Glycoprotein	<i>M. musculus</i>	3G5U, 3G60, 3G61	Aller et al., 2009
		13	P-Glycoprotein	<i>Caenorhabditis elegans</i>	4F4C	Jin et al., 2012
		14	LTC4S	<i>Homo sapiens</i>	2PNO	Ago et al., 2007
		15	Histamine H ₁ receptor	<i>Homo sapiens</i>	3RZE	Shimamura et al., 2011
	<i>S. cerevisiae</i>	16	AHA2 (H ⁺ pump)	<i>Arabidopsis thaliana</i>	3B8C	Pedersen et al., 2007
		17	VrH ⁺ -Ppase	<i>Vigna radiata</i>	4A01	Lin et al., 2012
		18	NRT1.1	<i>Arabidopsis thaliana</i>	4CL4	Parker and Newstead, 2014
		19	CAAX protease Ste24p	<i>Saccharomyces mikatae</i>	4IL3	Pryor et al., 2013
		20	PiPT	<i>Piriformospora indica</i>	4J05	Pedersen et al., 2013

* For some proteins like GPCR and potassium channel, only the representative ones are listed.

Table 4. Insect cell as an expression system for eukaryotic membrane protein*

Expression systems	No.	Protein	Species	PDB code	Reference
Insect cell <i>S. frugiperda</i>	1	β_2 AR (Fab)	<i>Homo sapiens</i>	2R4R 2R4S	Rasmussen et al., 2007
	2	β_2 AR (T4L)	<i>Homo sapiens</i>	2RH1	Cherezov et al., 2007
	3	β_2 AR-agonist complex	<i>Homo sapiens</i>	3PDS	Rosenbaum et al., 2011
	4	β_2 AR-GS complex	<i>Homo sapiens</i>	3SN6	Rasmussen et al., 2011a, b
	5	A _{2A} adenosine receptor	<i>Homo sapiens</i>	3EML	Jaakola et al., 2008
	6	CXCR4	<i>Homo sapiens</i>	3ODU 3OE8	Wu et al., 2010
	7	Dopamine D3 receptor	<i>Homo sapiens</i>	3PBL	Chien et al., 2010
	8	Sphingosine 1-phosphate receptor subtype 1	<i>Homo sapiens</i>	3V2 W 3V3Y	Hanson et al., 2012
	9	M2 muscarinic acetylcholine receptor	<i>Homo sapiens</i>	3UON	Haga et al., 2012
	10	M3 muscarinic acetylcholine receptor	<i>Rattus norvegicus</i>	4DAJ	Kruse et al., 2012
	11	κ -Opioid receptor	<i>Homo sapiens</i>	4DJH	Wu et al., 2012
	12	μ -Opioid receptor	<i>Mus musculus</i>	4DKL	Manglik et al., 2012
	13	δ -Opioid receptor	<i>Mus musculus</i>	4EJ4	Granier et al., 2012
	14	N/OFQ receptor	<i>Homo sapiens</i>	4EA3	Thompson et al., 2012
	15	CCR5	<i>Homo sapiens</i>	4MBS	Tan et al., 2013
	16	PAR1	<i>Homo sapiens</i>	3VW7	Zhang et al., 2012
	17	5-HT _{1B/2B} serotonin receptor	<i>Homo sapiens</i>	4IAR 4IB4	Wang et al., 2013a, b; Wacker et al., 2013
	18	Smoothed receptor	<i>Homo sapiens</i>	4JKV	Wang et al., 2013a, b
	19	Glucagon receptor	<i>Homo sapiens</i>	4L6R	Siu et al., 2013
	20	Metabotropic glutamate receptor1	<i>Homo sapiens</i>	4OR2	Wu et al., 2014
	21	P2X ₄	<i>Danio rerio</i> (Zebra fish)	3I5D 3H9 V 4DW1	Kawate et al., 2009; Hattori and Gouaux, 2012
	22	ASIC1	<i>Gallus gallus</i>	2QTS 3HGC	Jasti et al., 2007; Gonzales et al., 2009
	23	GluA2	<i>Rat</i>	3KG2 3KGC	Sobolevsky et al., 2009
	24	GLuCl α	<i>Caenorhabditis elegans</i>	3RHW, 3RIF, 3RI5 3RIA	Hibbs and Gouaux, 2011
	25	CX26	<i>Homo sapiens</i>	2ZW3	Maeda et al., 2009
	26	UT-B	<i>Bos taurus</i>	4EZC 4EZD	Levin et al., 2012
	27	ZMPSTE24	<i>Homo sapiens</i>	4AW6	Quigley et al., 2013
	28	ABCB10	<i>Homo sapiens</i>	4AYT	Shintre et al., 2013
	29	Caludin-15	<i>Mus Musculus</i>	4P79	Suzuki et al., 2014
	30	NRT1.1	<i>Arabidopsis thaliana</i>	4OH3	Sun et al., 2014
<i>Trichoplusia ni</i>	31	β 1 adrenergic receptor	<i>Meleagris gallopavo</i>	2VT4	Warne et al., 2008
	32	NTS1 Neurotensin Receptor	<i>Rattus norvegicus</i>	4GRV	White et al., 2012
	33	CmCIC	<i>Cyanidios-chydonmerolae</i>	3ORG	Feng et al., 2010
	34	Corticotropin-releasing factor receptor	<i>Homo sapiens</i>	4K5Y	Hollenstein et al., 2013
	35	GLUT1	<i>Homo sapiens</i>	4PYP	Deng et al., 2014

* For some proteins like GPCR and potassium channel, only the representative ones are listed.

Table 5. Mammalian cell as an expression system for eukaryotic membrane protein

Expression system	No.	Protein	Species	PDB code	Reference
Mammalian HEK293	1	Rodopsin	<i>Homo sapiens</i>	2J4Y	Standfuss et al., 2007
	2	RhCG	<i>Homo sapiens</i>	3HD6	Gruswitz et al., 2010
	3	Dopamine transporter	<i>Homo sapiens</i>	4M48	Penmatsa et al., 2013

Table 6. Comparison among four expression system

	<i>E. coli</i>	Yeast (<i>Pichia</i>)	Insect cell (Sf9)	Mammalian cell (HEK293)
Duration time before cell cultivation (Days)	3–5	6–8	25–30	Transient: 3–5 Stable: at least 30
Cell cultivation time for 1L test (Days)	1–2	3–7	2–4	2–4
Cost for 1L test (\$) in China	15–20	20–25	200–250	200–250
Number of available eukaryotic IMP structures	4	20	34	3

as *Pichia* and *S. cerevisiae* (Yang et al., 2009). During the past thirty years, yeast has proved to be a useful expression system: 15 eukaryotic IMP structures have been determined for proteins expressed in *Pichia* expression system and 5 by *S. cerevisiae*. Most of the structurally available eukaryotic channels such as potassium channels and water channels were expressed in yeast, as listed in Table 3.

Pichia is considered the best expression system among yeast species (Cereghino and Cregg, 2000). Several elements contribute to its increasing applications, including the simplicity of genetic manipulation, the high yield of heterologous protein, the cost-effective chemical reagents, as well as the ability of post-translational modifications (Macauley-Patrick et al., 2005). For these reasons, *Pichia* is a more suitable expression system for producing eukaryotic IMP than *E. coli*. *Pichia* shares the advantage of the molecular and genetic manipulation with *S. cerevisiae*, yet it adds extra advantage of 10- to 100- fold biomass out of the same cultural volume comparing with *S. cerevisiae* (Macauley-Patrick et al., 2005).

The improved techniques and the commercial availability together promote the development of *Pichia* (Cereghino and Cregg, 2000). *Pichia* is a methylotrophic yeast, capable of utilizing methanol as its sole carbon source (Yurimoto and Sakai, 2009). A promoter derived from the alcohol oxidase I (AOX1), which is the first-step enzyme in the methanol metabolism, strictly controls the expression of the foreign proteins (Macauley-Patrick et al., 2005). The commercial vector pPICZ (or pPICZ α) takes advantage of the AOX1 promoter, being induced by methanol (Li et al., 2007). AOX1 promoter is prevailing than other promoters like *PMA1* and *GPD1* for its strong and highly inducible ability (Cereghino and Cregg, 2000). After the vector is readily prepared and transformed into the competent cells, the target gene can be inserted into the *Pichia* genome in high efficiency via homologous recombination to generate stable cell lines, and then the colonies with multiple copies that exhibit the highest protein expression level will be screened out through zeocin-spread plates (Daly and Hearn, 2005). This zeocin selective

marker for transformation selection is important regarding to the convenience of genetic manipulation in yeast. All the procedure typically takes about 10–15 days for a complete procedure from subcloning to protein expression. A potential disadvantage of the yeast culture concerns the difficulty in cell disruption due to the thick and hard cell walls.

Insect cell

The baculovirus infected insect cell system is undoubtedly the dominant heterologous expression system for obtaining eukaryotic IMPs (Contreras-Gomez et al., 2014). The most common method for generating recombinant baculovirus is based on the site-specific transposition of an expression cassette into a baculovirus shuttle vector (bacmid) that is amplified in *E. coli* (Ciccarone et al., 1998). The process is very convenient: clone the target gene into the pFastBac vector which uses the strong AcMNPV polyhedron (PH) as the promoter for high level protein expression, then transform the pFastBac vector into DH10Bac *E. coli* competent cells. DH10Bac cells possess a baculovirus shuttle vector (bacmid) with a transposon site and a helper plasmid, thus can help the pFastBac vector to have a transposition on the bacmid. Once the transposition occurs and the recombinant bacmid is generated, the bacmid could be isolated and purified for transfection. After the insect cells are cultured into a desired confluence, they are transfected by the purified bacmid DNA to generate a recombinant baculovirus that used for preliminary expression test (Contreras-Gomez et al., 2014). The pFastBac is ampicillin resistance and Bacmid is kanamycin resistance, and these selective markers provide expedience for this baculovirus expression system. It takes approximately 3–4 weeks to complete these procedures for initial protein expression test.

There are two most popular insect cell lines used for IMP expression, *Spodoptera frugiperda* (Sf9) and *Trichoplusia ni* (Hi5). Heterologous proteins have disparate performances on the yield and behavior when expressed in these two cell

lines (Unger and Peleg, 2012). Till now, 30 structures were obtained for eukaryotic IMPs from Sf9 expression system and 5 from Hi5 (Table 4).

After the protein IL-2 was first expressed in large scale with the baculovirus-infected insect cells in 1985, this system has been quickly accepted and widely used (Smith et al., 1983; Maeda et al., 1985). Owing to the convenience of scale up, safety and accuracy (Kost et al., 2005), the baculoviral insect cell system has yielded the largest number of eukaryotic IMPs up to date (Table 4). Notably, among the 35 eukaryotic IMP structures, 23 are of G-protein coupled receptors (GPCR) (Table 4). The insect cell system has been the prevailing expression system for eukaryotic IMP. However, the cost for the cultural medium may represent a serious roadblock for most laboratories.

Mammalian cell

Mammalian expression system has become one of the popular recombinant protein production systems for its proper post-translational modification and human protein-like structure assembly (Khan, 2013). HEK (human embryo kidney) and CHO (Chinese hamster ovary) are two broadly used cell lines for recombinant expression. These two cell lines are extensively applied by researchers to do functional assay such as the electrophysiological assay (Kawate et al., 2009). Both these two cell lines can be applied for transient and stable transfections (Zhu, 2012). For the transient transfection approach, it is relative easier to reach to a reasonable protein expression level, but this expression level may vary from batch to batch. On the other hand, although the proteins have higher productivity and less variation in the stable transfection method, it is very time consuming (one month at least) (Condreay et al., 1999; Baldwin et al., 2003). Consequently, it is a balance for scientists to choose between these two transfection methods.

HEK293 is a specific cell line originally derived from HEK cells, while the number "293" comes from Graham's habit of numbering his experiments (Louis et al., 1997). Large scale, transient transfection of HEK293 in suspension culture is a reliable way to generate milligram quantities of recombinant eukaryotic IMPs. When the gene of interest is ligated into the vector pcDNA3 or pCMV5, the complete plasmid is then transfected into the HEK293 cells and the cells are harvested after 48 h (Thomas and Smart, 2005). The whole procedure is more or less similar to that of the insect cell system, only with a couple of exceptions. For example, 5%–10% CO₂ is required for maintaining the HEK293 cells, and the culture temperature is 37°C for HEK293 but not 27°C as for insect cells. The overall process usually requires one to two weeks from initial cloning to small scale test for the transient expression. However, ascribe to the low yield, slow growth rate and higher cost of complex media (Sunley and Butler, 2010), the number of eukaryotic IMP structures generated based on the mammalian cells is very limited. So far, only three eukaryotic IMP structures are from this system, and two of them are obtained from HEK293 cells (Table 5).

The BacMam system has to be mentioned for its safety, reproducibility and efficiency (Dukkipati et al., 2008). The baculoviruses are engineered by inserting a mammalian expression cassette for delivering foreign genes in mammalian cells. Their non-replicating property makes they are safe and well-tolerated by mammalian cells. BacMam system gains widespread use for its safety and rapid manipulation (Reeves et al., 2002; Bacongus and Gouaux, 2012). Depending on the cell type, cell division rate and transduction efficiency, it lasts 5–14 days to detect the gene expression (Dukkipati et al., 2008). The dopamine transporter structure was determined by the BacMam system (Penmatsa et al., 2013).

From the foregoing discussion, it is concluded that every expression system has their distinctive properties for protein expression. We compare their relative merits for an intuitive understanding of each system which can help researchers to make the best choice for their proteins expression (Table 6).

HOMOLOGUE SCREEN

Eukaryotic membrane proteins are very difficult to yield in large quantities, and most of them tend to be unstable in the presence of detergents. As a result, identification of well-expressed proteins is very essential. Homologue screen is widely applied for researchers to discover well-behaved proteins (Kawate et al., 2009; Xiaowei Hou, 2012).

Fluorescence detected size exclusion chromatography (FSEC) is a powerful method for homologue screen (Drew et al., 2006; Newstead et al., 2007). Compared with the common protocols, GFP fusion membrane proteins can be detected by measuring fluorescence in whole cells during the over-expression process. It saves time to help people preclude proteins that have no expression or low expression level. Also, it is much easier to assess the integrity of proteins by detecting the fluorescence in SDS polyacrylamide gels. Moreover, FSEC could be employed to figure out the most stable detergents in initial detergent screen. Considering these benefits, this technology is very widely applied (Jasti et al., 2007; Gonzales et al., 2009; Kawate et al., 2009; Sobolevsky et al., 2009). Taking P2X receptor as an example (Kawate et al., 2009), because of its aggregation and instability problems, researchers applied this method to screen 35 orthologs and finally got one species which was fit for crystallization. FSEC is proven to be one of the most robust methods to facilitate the identification of appropriate candidates for solving the structures of eukaryotic membrane proteins.

OPTIMAL CONSTRUCTS DESIGN

Optimizing constructs is very beneficial for getting the well-packed crystals. One way for optimizing constructs is to "cut off". Limited proteolysis is a conventional method to find the optimal constructs. Besides, it is worth noting that either N-terminal tag or C-terminal tag is removed before crystallization in most crystallization cases (Long et al., 2005; Long

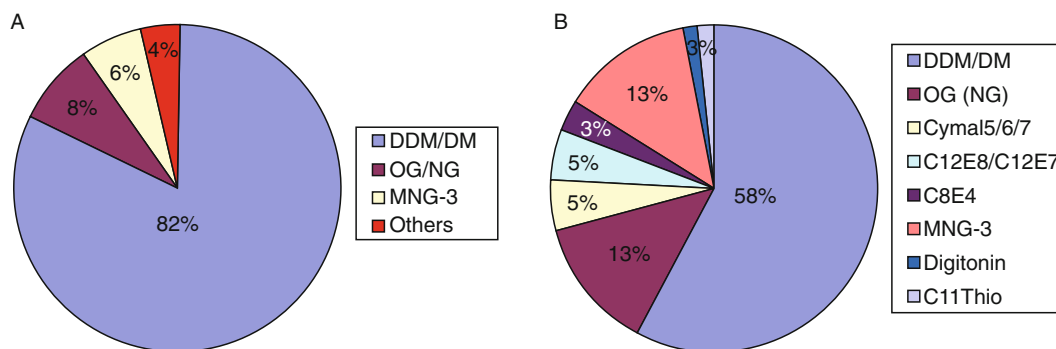


Figure 2. Detergents used for extraction and crystallization of eukaryotic membrane proteins. (A) Detergents for protein extraction and purification. DDM/DM can be applied for most eukaryotic membrane proteins in extraction step. (B) Detergents for protein crystallization. DDM/DM is the major detergent for the crystallization of eukaryotic membrane proteins.

et al., 2007; Gonzales et al., 2009; Maeda et al., 2009; Sobolevsky et al., 2009; Tao et al., 2009). For instance, the desensitized ASIC1 was crystallized by removal of 25 N-terminal and 64 C-terminal residues (Jasti et al., 2007).

The contrary way for optimizing constructs is to “add up”. T4 lysozyme (T4L) insertion and Fab/nanobody replacement are applied to produce stable proteins. The T4L fragment is soluble enough to effectively increase the solvent-exposed area, thereby facilitating protein-protein interactions and generating novel crystal packing interfaces (Cherezov et al., 2007). Fab/nanobody, which are generated from monoclonal antibodies, can reduce the protein flexibility and improve the conformational homogeneity (Zhou et al., 2001; Rasmussen et al., 2007). GPCR is one of the most successful cases employing T4L and Fab/nanobody to the ultimate structure determination (Rasmussen et al., 2007; Rasmussen et al., 2011a, b).

Mutagenesis is an alternative way for constructs design. In order to improve the crystallization behavior and stabilize the tetrameric state of the glutamate receptor GluA2, point mutations were introduced, preventing non-specific aggregation and disulphide bond formation (Sobolevsky et al., 2009). And E329Q was introduced in order to stabilize GLUT1 in a certain conformation (Deng et al., 2014). Plus, glycosylation is the most common post-translational modification of eukaryotic membrane proteins and leads to heterogeneity of proteins. Thus, mutating of glycosylation sites or deglycosylation by enzymes is an essential step for crystallization (Deng et al., 2014).

DETERGENTS, LIPIDS AND CRYSTALLIZATION

We have summarized the detergents used for protein purification and crystallization from Table 1. 51 eukaryotic membrane proteins can be extracted from DDM or DM (Fig. 2A), suggesting DDM/DM are the detergents suitable for the extraction process of the majority of eukaryotic membrane proteins. Collaterally, nearly half of the eukaryotic membrane protein crystals are obtained from DDM/DM, indicating DDM/DM are worthy of a trial for crystallization in

the first place (Fig. 2B and Table 1). Apart from these conventionally applied detergents, new detergents have also been developed to meet the new requirements. For example, when purifying β_2 adrenergic receptor-Gs protein, the authors stabilized protein complex by exchanging DDM with a newly developed maltose neopentyl glycol detergent MNG-3 (NG310, Anatrace) to prevent the complex dissociated from original detergent DDM (Chae et al., 2010; Rasmussen et al., 2011a, b).

It is worth noting that additional lipids are able to help crystal packing. There are three ways of lipid combinations. The first is mixing lipids with detergent(s) in hanging or sitting drop during crystallization. Take mammalian voltage-dependent *shaker* family potassium channel as an example, the author utilized 0.1 mg/mL 3:1:1 POPC: POPE: POPG throughout purification and crystallization to obtain crystals (Long et al., 2005). The second approach is lipid cubic phase (LCP) method. The lipid cubic phase is a dynamic structure consisting of a highly organized single lipid bilayer pervaded by an inter-connected aqueous channel (Landau and Rosenbusch, 1996). Martin has an elaborate discussion about LCP method which we will not go into details in this review (Caffrey and Cherezov, 2009). The crystal structure of β_2 AR-GS complex was determined by the use of 7.7 MAG as the host lipid for crystallization (Rasmussen et al., 2011a, b). The third way is bicelle method, which is regarded as an intermediate approach between the traditional detergent crystallization method and the rigid LCP method. Bicelle can be considered as a lipid bilayer disc that formed by a long chain lipid and a short chain lipid or detergent (Agah and Faham, 2012). The general composition is 3:1 DMPC: CHAPSO. Several protein structures were determined utilizing bicelle method (Rasmussen et al., 2007; Payandeh et al., 2011).

Last but not the least, we will elaborate a few messages for the crystallization of eukaryotic membrane protein drawn from Table 1: (a) Protein concentration: almost all the protein concentration for crystallization is above 5 mg/mL. (b) Crystallization temperature: if we expel the LCP method that is routinely crystallized at $20 \pm 2^\circ\text{C}$, nearly half of the eukaryotic membrane

proteins are crystallized at low temperature, especially on 4°C. At cold temperature, for protein with “normal” solubility, protein will be more soluble in high salt and precipitate from lower concentration of the precipitant reagents, and also the equilibrium diffusion rate occurs more slowly. These manifest that crystallization at lower temperature is absolutely an indispensable trial. (c) Crystallization methods: hanging drop or sitting drop crystallization method is the main and conventional approach for most eukaryotic membrane protein. LCP method is an up-rising star which is extensively applied in determining the GPCR's structures which we have mentioned before. Remarkably, LCP method is not only propitious to GPCR, but also is able to be applied for none-GPCR protein structures determination (Suzuki et al., 2014).

CONCLUSION

In this review, we discuss the benefits and drawbacks of different expression systems for eukaryotic membrane protein, and illustrate some general methods of recent advances for eukaryotic membrane protein purification and crystallization. We hope our work can provide help to those who are interested and work on eukaryotic membrane proteins. Although the discussion of eukaryotic membrane protein structure determined by Cryo-EM or NMR is beyond the scope of this review, the general methodologies and technical strategies summarized here also come to an aid in protein yield augment and sample homogeneity improvement for Cryo-EM and NMR. They are very powerful tools to solve structures, for instance, the Cryo-EM was applied to determine TrpV1 structures (Cao et al., 2013; Liao et al., 2013). With the development of advanced technologies, more and more eukaryotic membrane protein structures will emerge to answer the most significant questions in life sciences and provide the novel pharmaceutical targets in drug design.

ACKNOWLEDGEMENTS

We apologize to colleagues whose work could not be cited due to the scope of this review. We would like to thank members in Yan laboratory for discussions. We thank Brendan Lehnert, Xinlei Sheng, Quanxiu Li, Dan Ma and Xinhui Zhou for critical reading. This work was supported by funds from the National Basic Research Program (973 Program) (No. 2011CB910501), the National Natural Science Foundation of China (Grant Nos. 31321062-20131319400, 31125009, and 91017011), and funds from Tsinghua-Peking Center for Life Sciences. The research of N.Y. was supported in part by an International Early Career Scientist grant from the Howard Hughes Medical Institute.

ABBREVIATIONS

β₂AR, human β₂ adrenergic G-protein-coupled receptor; ABCB10, ATP-binding cassette (ABC) transporters; AHA2, *Arabidopsis thaliana* auto-inhibited H1-ATPase 2; ASIC1, acid-sensing ion channel 1; CAAX protease Ste24p, C is cysteine residue, A is an aliphatic residue and X is any residue. It is a zinc metalloprotease catalyzing two proteolytic steps in the maturation of yeast mating

pheromone a-factor; C₈E₄, tetraethyleneglycol mono-octyl ether; C₁₂E₇, dodecylheptaglycol; C₁₂E₈, polyoxyethylene dodecyl ether; CHS, cholesteryl hemisuccinate; CmClC, cyanidioschyzon merolae chloride (Cl⁻) ions transporter; C₁₁Thio, n-undecyl-β-D-thiomaltopyranoside; CXCR4, human chemokine receptors; CX26, connexin 26 gap junction; CYMAL5, 5-cyclohexyl-1-pentyl-β-D-maltoside; CYMAL6, 6-cyclohexyl-1-hexyl-β-D-maltoside; CYMAL7, 7-cyclohexyl-1-heptyl-β-D-maltoside; DDM, n-dodecyl-β-D-maltoside; DM, n-decyl-β-D-maltoside; FLAP, 5-lipoxygenase-activating protein; GIRK2 (Kir3.2), K⁺ channel: G protein-gated K⁺ channels; GluA2, α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)-sensitive ionotropic glutamate receptor; GLuCl_α, *Caenorhabditis elegans* glutamate-gated chloride channel α (GluCl), an inhibitory anion-selective Cys-loop receptor; HsAQP4, human aquaporin 4; HsAQP5, human aquaporin 5; Human BK channel, high-conductance voltage- and Ca²⁺-activated K⁺ channels; K2P1, two-pore domain potassium (K⁺) channels; Kv1.2, voltage-dependent shaker family potassium channel; Kv1.2-Kv2.1 paddle, ‘paddle-chimaera channel’, voltage-sensor paddle has been transferred from Kv2.1 to Kv1.2; LTC4S, cysteinyl leukotrienes; M-Ppase, membrane-integral pyrophosphatases; MAPEG, membrane-associated proteins in eicosanoid and glutathione metabolism; MNG, maltose-neopentyl glycol; NG, n-nonyl-β-D-glucopyranoside; NM, n-nonyl-β-D-maltoside; N/OFQ receptor, nociceptin/orphanin FQ receptor; OG, n-octyl-β-D-glucoside; OGN, octyl glucose neopentyl glycol; P2X₄, cation-selective ion channels gated by extracellular ATP; PAR1, protease-activated receptor 1; PfAQP, *Plasmodium falciparum* aquaglyceroporin; PiPT, a Fungal (*Piriformospora indica*) high-affinity phosphate transporter; POPC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; POPE, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine; POPG, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol; RhCG, rhesus C glycoprotein; SoPIP2, 1, spinach plant plasma membrane aquaporin; TRAAK, TWIK-related arachidonic acid-stimulated K⁺ channel; UDTM, n-undecyl-β-maltoside; UT-B, urea transporters-B; Vrh⁺-Ppase, *Vigna radiata* H1-translocating pyrophosphatases; ZMPSTE24, zinc metalloprotease STE24.

COMPLIANCE WITH ETHICS GUIDELINES

Yuan He, Kan Wang, and Nieng Yan declare that we have no conflict of interest.

This review does not contain any studies with human or animal subjects performed by the any of the authors.

OPEN ACCESS

This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

REFERENCES

- Abrahams JP, Leslie AG, Lutter R, Walker JE (1994) Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* 370:621–628

- AegeanSoftware (2005) NoteExpress, 2.0 edn. (NoteExpress is a perfect assistant and information manager for researchers, scholars, students, and librarians. NoteExpress is designed to help you organize research notes and bibliographic references, generate bibliographies automatically, search and capture bibliographic data from Internet with efficiency and ease. NoteExpress is well integrated with Microsoft Word. It can format bibliographies in many popular styles)
- Agah S, Faham S (2012) Crystallization of membrane proteins in bicelles. *Methods Mol Biol* 914:3–16
- Ago H, Kanaoka Y, Irikura D, Lam BK, Shimamura T, Austen KF, Miyano M (2007) Crystal structure of a human membrane protein involved in cysteinyl leukotriene biosynthesis. *Nature* 448:609–612
- Aisenbrey C, Borowik T, Bystrom R, Bokvist M, Lindstrom F, Misiak H, Sani MA, Grobner G (2008) How is protein aggregation in amyloidogenic diseases modulated by biological membranes? *Eur Biophys J* 37:247–255
- Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, Harrell PM, Trinh YT, Zhang Q, Urbatsch IL et al (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* 323:1718–1722
- Amunts A, Drory O, Nelson N (2007) The structure of a plant photosystem I supercomplex at 3.4 Å resolution. *Nature* 447:58–63
- Baconguis I, Gouaux E (2012) Structural plasticity and dynamic selectivity of acid-sensing ion channel–spider toxin complexes. *Nature* 489:400–405
- Baldwin SL, Powell TD, Wonderling RS, Keiser KC, Morales T, Hunter S, McDermott M, Radecki SV, Milhausen MJ (2003) Transient and stable transfection of Chinese hamster ovary cells with the recombinant feline erythropoietin gene and expression, purification, and biological activity of feline erythropoietin protein. *Am J Vet Res* 64:1465–1471
- Bernaudeau F, Frelet-Barrand A, Pochon N, Dementin S, Hivin P, Boutigny S, Rioux JB, Salvi D, Seigneurin-Berny D, Richaud P et al (2011) Heterologous expression of membrane proteins: choosing the appropriate host. *PLoS One* 6:e29191
- Bill RM, Henderson PJ, Iwata S, Kunji ER, Michel H, Neutze R, Newstead S, Poolman B, Tate CG, Vogel H (2011) Overcoming barriers to membrane protein structure determination. *Nat Biotechnol* 29:335–340
- Bkaily G, Al-Khoury J, Jacques D (2014) Nuclear membranes GPCRs: implication in cardiovascular health and diseases. *Curr Vasc Pharmacol* 12(2):215–222
- Bornert O, Alkhalifoui F, Logez C, Wagner R (2012) Overexpression of membrane proteins using *Pichia pastoris*. *Curr Protoc Protein Sci* Chapter 29:22–29
- Brohawn SG, Del MJ, MacKinnon R (2012) Crystal structure of the human K2P TRAAK, a lipid- and mechano-sensitive K⁺ ion channel. *Science* 335:436–441
- Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O (2008) Codon optimization can improve expression of human genes in *Escherichia coli*: a multi-gene study. *Protein Exp Purif* 59:94–102
- Caffrey M, Cherezov V (2009) Crystallizing membrane proteins using lipidic mesophases. *Nat Protoc* 4:706–731
- Cao E, Liao M, Cheng Y, Julius D (2013) TRPV1 structures in distinct conformations reveal activation mechanisms. *Nature* 504:113–118
- Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *Fems Microbiol Rev* 24:45–66
- Chae PS, Rasmussen SG, Rana RR, Gotfryd K, Chandra R, Goren MA, Kruse AC, Nurva S, Loland CJ, Pierre Y et al (2010) Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nat Methods* 7:1003–1008
- Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK et al (2007) High-resolution crystal structure of an engineered human beta₂-adrenergic G protein-coupled receptor. *Science* 318:1258–1265
- Chien EY, Liu W, Zhao Q, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V et al (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* 330:1091–1095
- Ciccarone VC, Polayes DA, Luckow VA (1998) Generation of recombinant baculovirus DNA in *E. coli* using a baculovirus shuttle vector. *Methods Mol Med* 13:213–235
- Condreay JP, Witherspoon SM, Clay WC, Kost TA (1999) Transient and stable gene expression in mammalian cells transduced with a recombinant baculovirus vector. *Proc Natl Acad Sci USA* 96:127–132
- Contreras-Gomez A, Sanchez-Miron A, Garcia-Camacho F, Molina-Grima E, Chisti Y (2014) Protein production using the baculovirus-insect cell expression system. *Biotechnol Prog* 30:1–18
- Daly R, Hearn MT (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J Mol Recognit* 18:119–138
- Deng D, Xu C, Sun P, Wu J, Yan C, Hu M, Yan N (2014) Crystal structure of the human glucose transporter GLUT1. *Nature* 510:121–125
- Drew D, Lerch M, Kunji E, Slotboom DJ, de Gier JW (2006) Optimization of membrane protein overexpression and purification using GFP fusions. *Nat Methods* 3:303–313
- Dukkipati A, Park HH, Waghay D, Fischer S, Garcia KC (2008) BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies. *Protein Exp Purif* 62:160–170
- Dumon-Seignover L, Cariot G, Vuillard L (2004) The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr Purif* 37:203–206
- Feng L, Campbell EB, Hsiung Y, MacKinnon R (2010) Structure of a eukaryotic CLC transporter defines an intermediate state in the transport cycle. *Science* 330:635–641
- Ferguson AD, McKeever BM, Xu S, Wisniewski D, Miller DK, Yamin TT, Spencer RH, Chu L, Ujjainwalla F, Cunningham BR et al (2007) Crystal structure of inhibitor-bound human 5-lipoxygenase-activating protein. *Science* 317:510–512
- Gonzales EB, Kawate T, Gouaux E (2009) Pore architecture and ion sites in acid-sensing ion channels and P2X receptors. *Nature* 460:599–604

- Granier S, Manglik A, Kruse AC, Kobilka TS, Thian FS, Weis WI, Kobilka BK (2012) Structure of the delta-opioid receptor bound to naltrindole. *Nature* 485:400–404
- Gruswitz F, Chaudhary S, Ho JD, Schlessinger A, Pezeshki B, Ho CM, Sali A, Westhoff CM, Stroud RM (2010) Function of human Rh based on structure of RhCG at 2.1 Å. *Proc Natl Acad Sci USA* 107:9638–9643
- Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T et al (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* 482:547–551
- Hanson MA, Roth CB, Jo E, Griffith MT, Scott FL, Reinhart G, Desale H, Clemons B, Cahalan SM, Schuerer SC et al (2012) Crystal structure of a lipid G protein-coupled receptor. *Science* 335:851–855
- Hattori M, Gouaux E (2012) Molecular mechanism of ATP binding and ion channel activation in P2X receptors. *Nature* 485:207–212
- Hibbs RE, Gouaux E (2011) Principles of activation and permeation in an anion-selective Cys-loop receptor. *Nature* 474:54–60
- Ho JD, Yeh R, Sandstrom A, Chorny I, Harries WE, Robbins RA, Miercke LJ, Stroud RM (2009) Crystal structure of human aquaporin 4 at 1.8 Å and its mechanism of conductance. *Proc Natl Acad Sci USA* 106:7437–7442
- Hollenstein K, Kean J, Bortolato A, Cheng RK, Dore AS, Jazayeri A, Cooke RM, Weir M, Marshall FH (2013) Structure of class B GPCR corticotropin-releasing factor receptor 1. *Nature* 499:438–443
- Horsefield R, Norden K, Fellert M, Backmark A, Tomroth-Horsefield S, Terwisscha VSA, Kvassman J, Kjellbom P, Johanson U, Neutze R (2008) High-resolution x-ray structure of human aquaporin 5. *Proc Natl Acad Sci USA* 105:13327–13332
- Ishikawa Y, Inoue N, Zhenfang Y, Nakae Y (2004) Molecular mechanisms and drug development in aquaporin water channel diseases: the translocation of aquaporin-5 from lipid rafts to the apical plasma membranes of parotid glands of normal rats and the impairment of it in diabetic or aged rats. *J Pharmacol Sci* 96:271–275
- Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, Ijzerman AP, Stevens RC (2008) The 2.6 Å crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* 322:1211–1217
- Jasti J, Furukawa H, Gonzales EB, Gouaux E (2007) Structure of acid-sensing ion channel 1 at 1.9 Å resolution and low pH. *Nature* 449:316–323
- Jin MS, Oldham ML, Zhang Q, Chen J (2012) Crystal structure of the multidrug transporter P-glycoprotein from *Caenorhabditis elegans*. *Nature* 490:566–569
- Kawate T, Michel JC, Birdsong WT, Gouaux E (2009) Crystal structure of the ATP-gated P2X(4) ion channel in the closed state. *Nature* 460:592–598
- Khan KH (2013) Gene expression in mammalian cells and its applications. *Adv Pharm Bull* 3:257–263
- Kost TA, Condreay JP, Jarvis DL (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotechnol* 23:567–575
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580
- Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, Green HF, Liu T, Chae PS, Dror RO et al (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* 482:552–556
- Kurusu G, Zhang H, Smith JL, Cramer WA (2003) Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity. *Science* 302:1009–1014
- Landau EM, Rosenbusch JP (1996) Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc Natl Acad Sci USA* 93:14532–14535
- Levin EJ, Cao Y, Enkavi G, Quick M, Pan Y, Tajkhorshid E, Zhou M (2012) Structure and permeation mechanism of a mammalian urea transporter. *Proc Natl Acad Sci USA* 109:11194–11199
- Li P, Anumanthan A, Gao XG, Ilangoan K, Suzara VV, Duzgunes N, Renugopalakrishnan V (2007) Expression of recombinant proteins in *Pichia pastoris*. *Appl Biochem Biotechnol* 142:105–124
- Liao M, Cao E, Julius D, Cheng Y (2013) Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* 504:107–112
- Lin SM, Tsai JY, Hsiao CD, Huang YT, Chiu CL, Liu MH, Tung JY, Liu TH, Pan RL, Sun YJ (2012) Crystal structure of a membrane-embedded H⁺-translocating pyrophosphatase. *Nature* 484:399–403
- Liu Z, Yan H, Wang K, Kuang T, Zhang J, Gui L, An X, Chang W (2004) Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature* 428:287–292
- Long SB, Campbell EB, Mackinnon R (2005) Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel. *Science* 309:897–903
- Long SB, Tao X, Campbell EB, MacKinnon R (2007) Atomic structure of a voltage-dependent K⁺ channel in a lipid membrane-like environment. *Nature* 450:376–382
- Louis N, Eveleigh C, Graham FL (1997) Cloning and sequencing of the cellular-viral junctions from the human adenovirus type 5 transformed 293 cell line. *Virology* 233:423–429
- Lu P, Ma D, Yan C, Gong X, Du M, Shi Y (2014) Structure and mechanism of a eukaryotic transmembrane ascorbate-dependent oxidoreductase. *Proc Natl Acad Sci USA* 111(5):1813–1818
- Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the *Pichia pastoris* expression system. *Yeast* 22:249–270
- Maeda S, Kawai T, Obinata M, Fujiwara H, Horiuchi T, Saeki Y, Sato Y, Furusawa M (1985) Production of human alpha-interferon in silkworm using a baculovirus vector. *Nature* 315:592–594
- Maeda S, Nakagawa S, Suga M, Yamashita E, Oshima A, Fujiyoshi Y, Tsukihara T (2009) Structure of the connexin 26 gap junction channel at 3.5 Å resolution. *Nature* 458:597–602
- Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, Pardo L, Weis WI, Kobilka BK, Granier S (2012) Crystal structure of the micro-opioid receptor bound to a morphinan antagonist. *Nature* 485:321–326
- Mijakovic I, Petranovic D, Macek B, Cepo T, Mann M, Davies J, Jensen PR, Vujaklija D (2006) Bacterial single-stranded DNA-binding proteins are phosphorylated on tyrosine. *Nucleic Acids Res* 34:1588–1596
- Miller AN, Long SB (2012) Crystal structure of the human two-pore domain potassium channel K2P1. *Science* 335:432–436

- Mironova R, Niwa T, Handzhiyski Y, Sredovska A, Ivanov I (2005) Evidence for non-enzymatic glycosylation of *Escherichia coli* chromosomal DNA. *Mol Microbiol* 55:1801–1811
- Miroux B, Walker JE (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* 260:289–298
- Murray CW, Verdonk ML, Rees DC (2012) Experiences in fragment-based drug discovery. *Trends Pharmacol Sci* 33:224–232
- Neophytou I, Harvey R, Lawrence J, Marsh P, Panaretou B, Barlow D (2007) Eukaryotic integral membrane protein expression utilizing the *Escherichia coli* glycerol-conducting channel protein (GlpF). *Appl Microbiol Biotechnol* 77:375–381
- Newby ZE, O'Connell JR, Robles-Colmenares Y, Khademi S, Miercke LJ, Stroud RM (2008) Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite *Plasmodium falciparum*. *Nat Struct Mol Biol* 15:619–625
- Newstead S, Kim H, von Heijne G, Iwata S, Drew D (2007) High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 104:13936–13941
- Nishida M, Cadene M, Chait BT, MacKinnon R (2007) Crystal structure of a Kir3.1-prokaryotic Kir channel chimera. *Embo J* 26:4005–4015
- Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? *Nat Rev Drug Discov* 5:993–996
- Parker JL, Newstead S (2014) Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1. *Nature* 507:68–72
- Payandeh J, Scheuer T, Zheng N, Catterall WA (2011) The crystal structure of a voltage-gated sodium channel. *Nature* 475:353–358
- Pedersen BP, Buch-Pedersen MJ, Morth JP, Palmgren MG, Nissen P (2007) Crystal structure of the plasma membrane proton pump. *Nature* 450:1111–1114
- Pedersen BP, Kumar H, Waight AB, Risenmay AJ, Roe-Zurz Z, Chau BH, Schlessinger A, Bonomi M, Harries W, Sali A et al (2013) Crystal structure of a eukaryotic phosphate transporter. *Nature* 496:533–536
- Penmatsa A, Wang KH, Gouaux E (2013) X-ray structure of dopamine transporter elucidates antidepressant mechanism. *Nature* 503:85–90
- Phillips R, Ursell T, Wiggins P, Sens P (2009) Emerging roles for lipids in shaping membrane-protein function. *Nature* 459:379–385
- Pryor EJ, Horanyi PS, Clark KM, Fedoriw N, Connelly SM, Koszelak-Rosenblum M, Zhu G, Malkowski MG, Wiener MC, Dumont ME (2013) Structure of the integral membrane protein CAAX protease Ste24p. *Science* 339:1600–1604
- Quigley A, Dong YY, Pike AC, Dong L, Shrestha L, Berridge G, Stansfeld PJ, Sansom MS, Edwards AM, Bountra C et al (2013) The structural basis of ZMPSTE24-dependent laminopathies. *Science* 339:1604–1607
- Rasmussen SG, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VR, Sanishvili R, Fischetti RF et al (2007) Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* 450:383–387
- Rasmussen SG, Choi HJ, Fung JJ, Pardon E, Casarosa P, Chae PS, Devree BT, Rosenbaum DM, Thian FS, Kobilka TS et al (2011a) Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature* 469:175–180
- Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D et al (2011b) Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* 477:549–555
- Reeves PJ, Callewaert N, Contreras R, Khorana HG (2002) Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc Natl Acad Sci USA* 99:13419–13424
- Rosenbaum DM, Zhang C, Lyons JA, Holl R, Aragao D, Arlow DH, Rasmussen SG, Choi HJ, Devree BT, Sunahara RK et al (2011) Structure and function of an irreversible agonist-beta(2) adrenoceptor complex. *Nature* 469:236–240
- Russell RB, Eggleston DS (2000) New roles for structure in biology and drug discovery. *Nat Struct Biol* 7(Suppl):928–930
- Sahdev S, Khattar SK, Saini KS (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem* 307:249–264
- Sanders CR, Myers JK (2004) Disease-related misassembly of membrane proteins. *Annu Rev Biophys Biomol Struct* 33:25–51
- Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW et al (2011) Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475:65–70
- Shintre CA, Pike AC, Li Q, Kim JI, Barr AJ, Goubin S, Shrestha L, Yang J, Berridge G, Ross J et al (2013) Structures of ABCB10, a human ATP-binding cassette transporter in apo- and nucleotide-bound states. *Proc Natl Acad Sci USA* 110:9710–9715
- Siu FY, He M, de Graaf C, Han GW, Yang D, Zhang Z, Zhou C, Xu Q, Wacker D, Joseph JS et al (2013) Structure of the human glucagon class B G-protein-coupled receptor. *Nature* 499:444–449
- Smith GE, Summers MD, Fraser MJ (1983) Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol Cell Biol* 3:2156–2165
- Snider C, Stephen HW (2014) Membrane proteins of known 3D structure. <http://blanco.biomol.uci.edu/mpstruc/>
- Sobolevsky AI, Rosconi MP, Gouaux E (2009) X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* 462:745–756
- Standfuss J, Xie G, Edwards PC, Burghammer M, Oprian DD, Schertler GF (2007) Crystal structure of a thermally stable rhodopsin mutant. *J Mol Biol* 372:1179–1188
- Strausberg RL, Strausberg SL (2001) Overview of protein expression in *saccharomyces cerevisiae*. *Curr Protoc Protein Sci* Chapter 5:t5–t6
- Sun F, Huo X, Zhai Y, Wang A, Xu J, Su D, Bartlam M, Rao Z (2005) Crystal structure of mitochondrial respiratory membrane protein complex II. *Cell* 121:1043–1057
- Sun J, Bankston JR, Payandeh J, Hinds TR, Zagotta WN, Zheng N (2014) Crystal structure of the plant dual-affinity nitrate transporter NRT1.1. *Nature* 507:73–77
- Sunley K, Butler M (2010) Strategies for the enhancement of recombinant protein production from mammalian cells by growth arrest. *Biotechnol Adv* 28:385–394
- Suzuki H, Nishizawa T, Tani K, Yamazaki Y, Tamura A, Ishitani R, Dohmae N, Tsukita S, Nureki O, Fujiyoshi Y (2014) Crystal

- structure of a claudin provides insight into the architecture of tight junctions. *Science* 344:304–307
- Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, Li T, Ma L, Fenalti G, Li J et al (2013) Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science* 341:1387–1390
- Tao X, Avalos JL, Chen J, MacKinnon R (2009) Crystal structure of the eukaryotic strong inward-rectifier K⁺ channel Kir2.2 at 3.1 Å resolution. *Science* 326:1668–1674
- Tao X, Lee A, Limapichat W, Dougherty DA, MacKinnon R (2010) A gating charge transfer center in voltage sensors. *Science* 328:67–73
- Thomas P, Smart TG (2005) HEK293 cell line: a vehicle for the expression of recombinant proteins. *J Pharmacol Toxicol Methods* 51:187–200
- Thompson AA, Liu W, Chun E, Katritch V, Wu H, Vardy E, Huang XP, Trapella C, Guerrini R, Calo G et al (2012) Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* 485:395–399
- Tomroth-Horsefield S, Wang Y, Hedfalk K, Johanson U, Karlsson M, Tajkhorshid E, Neutze R, Kjellbom P (2006) Structural mechanism of plant aquaporin gating. *Nature* 439:688–694
- Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S (1996) The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. *Science* 272:1136–1144
- Unger T, Peleg Y (2012) Recombinant protein expression in the baculovirus-infected insect cell system. *Methods Mol Biol* 800:187–199
- van Meer G, Voelker DR, Feigenson GW (2008) Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* 9:112–124
- Wacker D, Wang C, Katritch V, Han GW, Huang XP, Vardy E, McCorvy JD, Jiang Y, Chu M, Siu FY et al (2013) Structural features for functional selectivity at serotonin receptors. *Science* 340:615–619
- Wang C, Jiang Y, Ma J, Wu H, Wacker D, Katritch V, Han GW, Liu W, Huang XP, Vardy E et al (2013a) Structural basis for molecular recognition at serotonin receptors. *Science* 340:610–614
- Wang C, Wu H, Katritch V, Han GW, Huang XP, Liu W, Siu FY, Roth BL, Cherezov V, Stevens RC (2013b) Structure of the human smoothed receptor bound to an antitumour agent. *Nature* 497:338–343
- Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AG, Tate CG, Schertler GF (2008) Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* 454:486–491
- White JF, Noinaj N, Shibata Y, Love J, Kloss B, Xu F, Gvozdenovic-Jeremic J, Shah P, Shiloach J, Tate CG et al (2012) Structure of the agonist-bound neurotensin receptor. *Nature* 490:508–513
- Whorton MR, MacKinnon R (2011) Crystal structure of the mammalian GIRK2 K⁺ channel and gating regulation by G proteins, PIP₂, and sodium. *Cell* 147:199–208
- Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC et al (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330:1066–1071
- Wu H, Wacker D, Mileni M, Katritch V, Han GW, Vardy E, Liu W, Thompson AA, Huang XP, Carroll FI et al (2012) Structure of the human kappa-opioid receptor in complex with JDTic. *Nature* 485:327–332
- Wu H, Wang C, Gregory KJ, Han GW, Cho HP, Xia Y, Niswender CM, Katritch V, Meiler J, Cherezov V et al (2014) Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. *Science* 344:58–64
- Xia D, Yu CA, Kim H, Xia JZ, Kachurin AM, Zhang L, Yu L, Deisenhofer J (1997) Crystal structure of the cytochrome bc₁ complex from bovine heart mitochondria. *Science* 277:60–66
- Hou X, Pedi L, Diver MM, Long SB (2012) Crystal structure of the calcium release-activated calcium channel orai. *Science* 338(6112):1308–1313
- Yang Y, Hu Z, Liu Z, Wang Y, Chen X, Chen G (2009) High human GLUT1, GLUT2, and GLUT3 expression in *Schizosaccharomyces pombe*. *Biochemistry (Mosc)* 74:75–80
- Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M (2007) Drug-target network. *Nat Biotechnol*. 25:1119–1126
- Yurimoto H, Sakai Y (2009) Methanol-inducible gene expression and heterologous protein production in the methylotrophic yeast *Candida boidinii*. *Biotechnol Appl Biochem* 53:85–92
- Zhang C, Srinivasan Y, Arlow DH, Fung JJ, Palmer D, Zheng Y, Green HF, Pandey A, Dror RO, Shaw DE et al (2012) High-resolution crystal structure of human protease-activated receptor 1. *Nature* 492:387–392
- Zhou Y, Morais-Cabral JH, Kaufman A, MacKinnon R (2001) Chemistry of ion coordination and hydration revealed by a K⁺ channel-Fab complex at 2.0 Å resolution. *Nature* 414:43–48
- Zhu J (2012) Mammalian cell protein expression for biopharmaceutical production. *Biotechnol Adv* 30:1158–1170