WIDENING BOTTLENECKS IN MEMBRANE PROTEIN STRUCTURE PIPELINES

Roslyn M Bill[1], Peter JF Henderson[2], So Iwata[3], Edmund RS Kunji[4], Hartmut Michel[5], Richard Neutze[6], Simon Newstead[7], Bert Poolman[8], Christopher G Tate[9], Horst Vogel[10]

[1]School of Life and Health Sciences, Aston University, Birmingham B4 7ET, UK
[2] Astbury Centre for Structural Molecular Biology, Institute of Membrane and Systems Biology, University of Leeds, Leeds LS2 9JT, UK
[3]Membrane Protein Crystallography Group, Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, UK
[4]MRC Mitochondrial Biology Unit, Wellcome Trust / MRC Building, Hills Road, Cambridge CB2 0XY, UK
[5]Department of Molecular Membrane Biology, Max Planck Institute of Biophysics, D-60438 Frankfurt am Main, Germany
[6]Department of Chemistry, Biochemistry and Biophysics, University of Gothenburg, SE-405 30 Göteborg, Sweden
[7]Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, UK
[8]Department of Biochemistry, University of Groningen, 9747 AG Groningen, The Netherlands
[9]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK
[10]Laboratory of Physical Chemistry of Polymers and Membranes, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

Correspondence should be addressed to R.M.B. (r.m.bill@aston.ac.uk)

Advances in technology are energizing the field of membrane protein structure.

After decades of slow progress, the pace of research on membrane protein structures is beginning to quicken thanks to various improvements in technology, including protein engineering and micro-focus X-ray diffraction. Recent breakthroughs include structures of GPCR's[1-6], P-ATPases[7, 8], secondary active transporters[9-17], ABC transporters[18, 19] and ion channels[20]. Here we review these developments and, where possible, highlight generic new approaches to solving membrane protein structures based on the technological advances of the past few years. Rational approaches to overcoming the bottlenecks in structure determination are urgently required as membrane proteins, which typically comprise ~30% of the proteomes of organisms, are still dramatically underrepresented in the structural database of the Protein Data Bank.

Electron crystallography is currently the only technique that can solve structures of membrane proteins in their native environment, as exemplified by the seminal structure of bacteriorhodopsin solved by electron diffraction using naturally occurring two-dimensional crystals[21]. But this approach is not widely used because the production of two-dimensional crystals that diffract to high-resolution is far from simple, with only about 8 membrane protein structures determined to atomic resolution, although another 30-40 structures have been determined at an intermediate resolution sufficient to delineate transmembrane α-helices. In addition, the methodology to determine structures from two-dimensional crystals has not been developed into user-friendly software as is the case for determining structures from X-ray diffraction data. Therefore, the predominant technique in membrane protein structural projects is X-ray crystallography of three-dimensional crystals.

To produce crystals that diffract to high resolution, sufficient amounts of the membrane protein are required in a form that is stable and compatible with well-ordered packing. Few membrane proteins are naturally abundant in their native membranes, with notable exceptions such as mammalian and bacterial rhodopsins, aquaporins, respiratory complexes, ATPases, photosynthetic complexes, reaction centers and light harvesting proteins. Inevitably, these proteins were among the first to have their structures solved. For the vast majority, however, recombinant production is the first bottleneck that must be tackled to secure the hundreds-of-milligram quantities necessary for a successful structural biology project.

Once sufficient expression has been achieved, the next barrier is purification of the protein in stable form. The native membrane environment imparts considerable stability to membrane proteins through its lipid composition and physicochemical properties. When solubilized in detergents, many membrane proteins cannot be purified as they rapidly denature and often aggregate. This second bottleneck is particularly acute for membrane

proteins from higher eukaryotes, and most membrane protein structures determined to date are from bacteria or archaea, often focusing on homologs of mammalian proteins. These structures have been solved mostly using X-ray diffraction in combination with detergent crystallization protocols.

Crystallization trials have benefited from major strides in automation and miniaturization in recent years. However, the success rate in advancing from purified protein to high-resolution structure is still disappointingly low, and this third bottleneck is exacerbated by the challenges of data collection from microcrystals. Nevertheless, the rate of progress is accelerating, with structures of recombinant membrane proteins[22] becoming increasingly significant within the pipeline of membrane protein structural biology (**Fig. 1a**). Moreover, as the database of membrane protein structures has grown, the use of molecular replacement to sidestep the time-consuming task of experimental phasing has become more common (**Fig. 1b**).

Rationalizing production of recombinant membrane proteins

Understanding the host organism better is an emerging strategy for achieving high yields of recombinant membrane proteins[23, 24] through improvements of the host cell. In contrast, the conventional approach of repeated rounds of trial-and-error 'optimization' simply varies external parameters (e.g., promoter and fusion tag combinations or culture process parameters such as pH, temperature and aeration) and cannot provide insight into the biology of recombinant protein production. More targeted approaches, such as deletions in protease or secretion pathways based on speculation about where bottlenecks lie, can be successful on a case-by-case basis, but also do not reveal the relevant mechanisms of a high-yielding cell.

Knowledge of how membrane proteins are synthesized in the cell is still very poor. For example, although each host cell appears to have a number of unique accessory factors required for membrane protein biogenesis, their precise roles are unclear. This means that, although any membrane protein can in principle be produced in any system, because of subtle differences between signal-recognition particles, translocon components, cellular chaperones and foldases, the efficiency of heterologous overproduction may be very low. An extensive comparison of membrane proteins produced in various heterologous host systems[25] concluded, perhaps unsurprisingly, that the best host was the one most closely related in evolution to the source of the target membrane protein. Hence, production of mammalian membrane proteins in *E. coli*[26] usually requires considerable time and effort to achieve functional levels suitable for subsequent purification. Successful strategies, which rely on using low-copy-number plasmids, weak promoters and low temperatures during induction[27], are thought to allow sufficient time for folding of the membrane protein while

keeping the amount of mRNA encoding the mammalian membrane protein to a minimum. This prevents the cellular ribosomes from synthesizing the heterologous protein at the expense of host-cell proteins and prevents accumulation of misfolded protein.

These conditions seem to parallel the prolonged expression profiles for mammalian membrane proteins produced in mammalian cells. For example, the mu-opioid receptor has a half-time of appearance at the cell surface of 135 min, with 120 min required for the nascent polypeptide chain to fold and exit the endoplasmic reticulum[28]. The cystic fibrosis transmembrane conductance regulator (CFTR), an ABC transporter, shows a similarly slow maturation period to reach the cell surface[29]. In both cases, a considerable proportion of the nascent polypeptide chain is misfolded and degraded, so only ~40% and 25% of the mu-opioid receptor and CFTR nascent polypeptide chains, respectively, actually make it to the cell surface in a functional form. Similarly, overproduction of the mu-opioid receptor in *Pichia pastoris* results in only 22% of the receptor being functional[30]. The role of molecular chaperones in this folding process is largely unexplored, although calnexin is likely to play a part in the folding of N-glycosylated membrane proteins, such as the serotonin transporter[31], where it can also recruit other molecular chaperones to form a folding complex. Only a few attempts have been made to improve membrane protein production by co-expression of molecular chaperones, and they have mostly met with only a modest success, with a 2–3 fold improvement in yields[31, 32]. Presumably, the levels of multiple molecular chaperones have to be carefully controlled before the full folding pathway can be accelerated to enhance significantly membrane protein overproduction.

Understanding the protein biogenesis machinery and the physiological response of host cells to membrane protein production is crucial for identifying the bottlenecks in expression and designing strategies to improve yields. The application of ‘omics’ technologies has already contributed to our understanding of membrane protein production in bacteria and yeast and provided rationales for the forward engineering of these cells[23, 24, 33]. Interestingly, all these studies have shown that tuning the transcript levels of identified genes (either up or down) is crucial for successful production trials. In addition, much is to be gained from the optimization of the downstream steps of membrane protein biogenesis, but again finetuning is critical as the pathway components may otherwise heighten the production hurdle.

One advance in this area arose from an early analysis of membrane protein production in *Escherichia coli*[34]. Increases in levels of chaperones and proteases were associated with increased membrane protein production, and it was speculated that low yields were due to limited Sec translocon capacity. Previously, a systems biotechnology approach to recombinant membrane protein production in the eukaryotic microbe *Saccharomyces cerevisiae* had identified 39 host-cell genes whose production was significantly altered when the aquaporin Fps1 was produced under high-yielding conditions (20°C, pH5) compared to

low-yielding standard growth conditions (30°C, pH5) [35]. In particular, an essential gene, *BMS1*, with a role in ribosome biogenesis, was identified as always up-regulated in high-yielding host cells[23, 36]. Subsequent overproduction of *BMS1* in a doxycycline-titratable manner revealed that maximal Fps1 yield was significantly correlated with an optimum level of *BMS1* transcript. By further titrating the overproduction of *BMS1*, the functional yields of a range of membrane proteins could be improved by a factor of up to 70. In the future it will be possible to apply this approach to a yeast species that has been used widely in membrane protein projects, given the recent publication of a curated *P. pastoris* genome[37], as well as to other host cells and protein targets.

A second advance has been in the development of alternative host-cell factories. Unexpectedly, the accumulation of host-cell biomass does not necessarily lead to a correlated increase in membrane protein yield, and in the case of G-protein-coupled receptors (GPCRs) produced in yeast, specific activity is often lower[38]. Indeed, it has been noted that higher cell densities can generate cellular stresses leading to modifications in membrane composition[39], and that this modified environment influences the activity of recombinant proteins. Consequently, medium cell density fermentation procedures for GPCR production have been suggested to be preferable to those that maximize biomass yields[38]. In a recent example of host development, a respiratory *S. cerevisiae* strain was reported that has improved biomass properties, leading to increased functional yields without the need to resort to complex cultivation schemes[40]. The yield of functional human adenosine A_{2A} receptor was quadrupled in this new strain compared to that from wild-type cells.

The Gram-positive bacterium *Lactococcus lactis* has also been used as an alternative host to produce a wide range of eukaryotic and prokaryotic membrane proteins[41], enabling a comparison of the production potential of *L. lactis* and *E. coli*[42, 43]. Although a large fraction of proteins could be produced in both hosts, some could only be produced in one or the other. Notably, for about half of the proteins produced in *E. coli*, additional bands of lower molecular weight were observed, indicative of breakdown products, whereas only 10% of the proteins produced in *L. lactis* were degraded. The ability to incorporate selenomethionine efficiently into proteins (>90%) produced in *L. lactis*[44, 45] now greatly extends the usefulness of this production host for X-ray crystallography projects[46].

The establishment of green fluorescent protein (GFP) as a reporter to assess quantitatively the functional yields of membrane proteins has also moved the field forward[47-50]. When a membrane protein that is fused N-terminally to GFP becomes misfolded during biosynthesis, it drags GFP into a misfolded, SDS-sensitive state. If, however, the membrane protein is properly folded, the GFP barrel will be synthesized as a fluorescent, SDS-resistant moiety; the SDS-sensitive and SDS-resistant conformations can be readily discriminated on SDS-PAGE and immunoblots. Thus, one can simultaneously quantify the levels of folded and

aggregated membrane protein. The method requires only standard equipment, small culture samples, is not labor-intensive and can greatly facilitate the optimization of membrane protein production and crystallization experiments in *E. coli, L. lactis* and yeast[47-50]. GFP fusions have also been widely used in pre-crystallization strategies using fluorescence-detection size-exclusion chromatography[51]. Only nanogram quantities of impure protein are needed to evaluate the localization and yield, the degree of monodispersity, and the approximate molecular mass of the recombinant protein. Additionally, using a directed-evolution approach and combining GFP-fusions with an antibiotic resistance marker, it has been possible to select for host strains that produce more functional membrane protein[33].

Fluorescent labeling of recombinant membrane proteins on the extracellular side of the cell's plasma membrane offers an important advantage as only functional proteins correctly inserted into the membrane are visualized, whereas incorrectly folded proteins in the cytoplasm remain invisible. Post-translational labeling with small fluorescent probes such as fused acyl carrier protein tags can deliver novel information about the functional state of GPCRs both in live cells and in detergent-solublilized forms[52, 53]. In the case of transporters and ionotropic receptors, functional activity is classically assayed by measuring the transport of charged molecules or ions across native or reconstituted membranes. This requires the very time-consuming preparation of planar bilayer membranes or patch-clamp experiments. In a recent development, electrophysiological tests on chips have been used to substantially reduce the time and material required for testing function during purification of some membrane proteins[54].

Improving the stability of membrane proteins

The stability of a membrane protein in detergent solution is crucial for producing well-diffracting crystals[55]. Successful conditions have been found through extensive screens using rapid assays such as GFP-tagging coupled to size exclusion chromatography or dot-blotting techniques. Although this approach identifies the few membrane proteins naturally stable in detergent, it is not useful for determining the structure of particular mammalian membrane proteins of interest. In the last couple of years, several new approaches have been developed to improve the stability of membrane proteins, particularly GPCRs[56].

Among GPCRs, only rhodopsin is present in native tissues at sufficiently high levels to allow purification of milligrams of protein. Rhodopsin is also extremely stable in detergent, which has allowed structure determination of bovine rhodopsin[57] and squid rhodopsin[2] both in the dark-adapted inactive state and in an active-like state[3, 6]. The stability of dark-adapted rhodopsin is partly due to the fact that it remains in a single conformation until a photon of light activates its covalently bound chromophore, retinal. In contrast, hormone-binding GPCRs have long resisted crystallization because in detergent solution they are in

equilibrium between two basic conformations: one that bind G proteins (R*) and another that cannot (R). A mixture of multiple conformations makes crystal formation less likely. In addition, R* is itself often unstable, leading to rapid inactivation of all the receptor molecules in solution as R* denatures and more R is converted to R*. A combination of ligand and large lipid-detergent micelles can stabilize many GPCRs, but the resultant species are so large that the occluded hydrophilic surfaces effectively prevent crystallization. This problem has been addressed by binding a F_{ab} antibody fragment to the intracellular part of the receptor, dramatically increasing the potential surfaces for making crystal contacts. The strategy was used to determine the structure of the β_2-adrenergic receptor (β_2AR) at 3.4 Å resolution[4] (**Fig. 2a**). Despite this breakthrough, the resolution of the crystals was insufficient to model side-chains accurately or to delineate the bound inverse agonist, carazolol.

A second strategy for increasing the hydrophilic surface is to engineer fusions with T4 lysozyme (**Fig. 2a**). Insertion of this protein in the third intracellular loop of β_2AR gave better crystals that diffracted to higher resolution[1, 5] compared with F_{ab} fragment co-crystallization. In this case, however, crystals could not be obtained by vapor diffusion and were instead formed in lipid cubic phase[58]. As the major constituent of lipid cubic phase is the single-chain lipid mono-olein, which is denaturing, the key to crystallizing the β_2AR-T4 lysozyme fusion was to add cholesteryl hemisuccinate, which dramatically improved the stability of the receptor. In fact, cholesteryl hemisuccinate had long been known to stabilize GPCRs solublilized in dodecylmaltoside and is essential for purifying the neurotensin and adenosine A_{2A} receptors in functional form[27, 59]. As cholesteryl hemisuccinate dramatically increases the size of the dodecylmaltoside micelle, no crystals have yet been grown by vapor diffusion from GPCRs purified in dodecylmaltoside / cholesteryl hemisuccinate . Presumably in lipid cubic phase, excess detergent and cholesteryl hemisuccinate diffuses into the mono-olein, allowing crystallization to occur. This T4 lysozyme strategy has also been successfully applied in the crystallization and structure determination of the adenosine A_{2A} receptor [60].

A third strategy, which does not rely on fusion proteins or binding partners, is based on the observation that short-chain detergents form small micelles around membrane proteins. Compared with micelles generated with longer-chain detergents, these micelles leave larger hydrophilic areas exposed to form crystal contacts, as can be inferred from the systematic size comparisons of detergent micelles containing the mitochondrial ADP/ATP carrier[61] (**Fig. 2b**). However, short-chain detergents are far more denaturing than the long-chain detergents normally used to purify GPCRs in functional form[55]. Therefore, their use generally requires protein thermostabilization. Because bacterial proteins can be thermostabilized by single point mutations[62, 63], a strategy was developed in which GPCRs were systematically mutated by alanine scanning and each mutant tested for thermostability using a radioligand binding assay coupled to a heating step[64-67]. The thermostabilizing point mutations were then

combined to make an optimally stable mutant containing 4–6 point mutations. Using this approach it was possible to stabilize GPCRs in both agonist- and antagonist-binding conformations that were also more stable in short-chain detergents, allowing the structure of the thermostabilized β_1AR-m23 to be determined to 2.7 Å resolution upon crystallization in octylthioglucoside[68] (**Fig. 2a**).

The introduction of mutations, deletions or insertions to obtain crystals inevitably alters a protein's characteristics and may affect conformational dynamics and ligand affinities. Wherever possible, retention of native functionality must be monitored, for example, with binding studies[65]. Clearly, this issue must be carefully considered when using structures of engineered proteins to draw conclusions about function. For the β_1 and β_2 AR structures determined to date[1, 69-71], it is gratifying that they are all consistent with one another with respect to the ligand-binding pocket and that they explain a wealth of earlier biochemical and pharmacological data. The presence of T4 lysozyme in the β_2 AR fusion does perturb the structure of cytoplasmic loops 2 and 3, limiting the utility of the structure for understanding the binding of intracellular effectors. However, the fusion strategy has been successfully applied to other GPCRs, such as the CXCR4 chemokine receptor[72] and the dopamine D3 receptor[73]. Much more work is clearly required to understand how membrane proteins like GPCRs function in the cell, and any method for increasing the probability of obtaining crystals is therefore valuable.

Increasing success rates of crystal optimization and structure solution

Whether or not a membrane protein has been engineered, its structure after detergent solubilization and crystallization (or in the conditions required for nuclear magnetic resonance studies) may diverge from its native structure. Unfortunately, there are very few cases in which high-resolution structures have been solved by more than one technique, but for bacteriorhodopsin[74] and AQP0[75] (electron and X-ray crystallography) and for sensory rhodopsin[76] (NMR and X-ray crystallography), no large differences in the structural folds were observed. However, a lipidic environment can facilitate conformational changes, as demonstrated for the transport cycle of bacteriorhodopsin[74, 77] and the voltage sensors of the voltage-dependent potassium channel[78]. Thus, it is likely that lipids will be increasingly used in future crystallization trials, whether as sponge phase, lipid cubic phase or detergent-lipid micelles.

Crystallization robot technologies, which can dispense nanoliter-scale drops in 96-well plates, have substantially increased the number of crystallization conditions that can be explored with limited amounts of sample. Robotics is also having an enormous impact on the collection of X-ray diffraction data because sample-exchange robots allow crystals to be replaced without the need to enter the experimental hutch. In combination with rapid crystal

alignment tools, these recent technical advances have increased the number of protein crystals screened at the European Synchrotron Radiation Facility by more than an order of magnitude in less than a decade[79]. More crystals being screened for diffraction translates into collection of higher-quality X-ray diffraction data, accelerating the rate of progress.

Another recent innovation at synchrotrons is micro-focus beamlines[79]. Smaller X-ray beams allow useful diffraction data to be extracted from smaller crystals, reducing the time required for crystal optimization. Although tighter focus comes at the cost of greater radiation damage to the crystal, this problem can be solved by merging data from several crystals, as was done for the structure of a human β_2AR[4]. Microfocus beams also make it possible to examine the diffraction quality at different regions of the same crystal, allowing data to be collected from the best-diffracting regions[79]. High-quality electron density maps to 1.5 Å resolution have been recovered from crystals of the soluble protein xylanase II using a microfocus beam of 1 μm^2 without any significant radiation damage[80]. The combination of microfocus X-ray beams with rapid-readout pixel- based detectors[81] reduces background further using 'fine slicing' (very small oscillations between each frame) methods of data collection, improving the resolution of data that can be extracted from a crystal. This push toward increasingly focused X-ray beams will continue as data are recorded from sub-micron scale crystals at emerging X-ray-free electron sources; the short X-ray pulse characteristics (~100 fs) of this source should enable the traditional radiation barrier of structural biology to be superseded[82].

As the database of membrane protein structures has grown, the use of molecular replacement for phasing is increasing (**Fig. 1b**). The combination of molecular replacement and co-crystallization with antibody fragments[83, 84] or large insertions of known structure[1, 5] (**Fig. 2a**) is also very powerful, since these additions can aid phasing by molecular replacement. Nevertheless, a recent series of very similar structures from transporters with no significant sequence homology[9-17] highlights the continued importance of experimental phasing methods (**Fig. 1b**). A useful innovation in this respect is the development of a convenient method for identifying heavy-atom derivatives covalently bound to cysteine residues of solubilized membrane proteins[85]. In combination with cysteine mutation scanning, this approach can facilitate reliable incorporation of heavy atoms for phasing before crystallization. Labeling phospholipids specifically with heavy-atom derivatives was successful for identifying lipid binding sites but lacked sufficient order to facilitate phasing[86].

Because membrane protein crystals frequently diffract to lower resolution than do soluble proteins, innovations that enable structures to be built and refined more reliably at low to medium resolution could have a strong impact on the field. A recently proposed approach to structural refinement that exploits higher-resolution structural information from homologous structures but allows global and local deformations[87] may enable membrane

protein structures to be refined in the grey zone around 4 Å resolution. Although it is too early to judge the impact of this approach on the field, more conservative approaches, such as applying H-bond restraints in transmembrane helices, can certainly improve the results of structural refinement, as was demonstrated for the sodium-hydantoin transporter, Mhp1[14, 17].

New X-ray diffraction and scattering methods are also emerging that move beyond the study of a static resting conformation and observe conformational changes within membrane proteins in real time at room temperature. Time-resolved Laue diffraction has been used to observe light-induced electron density changes in a photosynthetic reaction centre[88], and time-resolved wide-angle X-ray scattering has provided low-resolution overviews of light-induced helical movements with time in bacteriorhodopsin and proteorhodopsin[89]. Although Laue diffraction will always be limited by the need to probe reversible reactions in highly ordered and robust crystals, time-resolved wide-angle X-ray scattering could develop into a generic technique for visualizing the time-scales and nature of global conformational changes in membrane proteins.

Conclusion

Membrane protein families are defined by similarities in their amino acid sequences, yet individual proteins in a family can behave very differently with regard to production, stability, crystallization and other biophysical and biochemical properties. Conversely, proteins unrelated by amino acid sequence may have very similar crystal structures[15, 16]. Several new technologies have recently emerged to help identify and control the biological pathways underpinning recombinant membrane protein production, to understand why membrane proteins become inactivated upon detergent solubilization and to identify the critical parameters in obtaining high-resolution diffraction data. In combination, these diverse approaches provide a technical platform for overcoming the major bottlenecks in membrane protein structural biology. This potential to build on recent successes is creating an atmosphere of confidence that is contagious, triggering a growth in the number of scientists forming collaborations, like ours, committed to addressing the major challenges in membrane protein structural biology.

Acknowledgements

We thank Dr A. D. Cameron for access to information in advance of its publication. This work was supported by the European Membrane Protein Consortium via contract LSHG-CT-2004-504601 (E-MeP) and by the European Drug Initiative on Channels and Transporters via contract HEALTH-F4-2007-201924 (EDICT).

References

1. Cherezov, V. et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258-1265 (2007).
2. Murakami, M. & Kouyama, T. Crystal structure of squid rhodopsin. *Nature* **453**, 363-367 (2008).
3. Park, J.H., Scheerer, P., Hofmann, K.P., Choe, H.W. & Ernst, O.P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183-187 (2008).
4. Rasmussen, S.G. et al. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **450**, 383-387 (2007).
5. Rosenbaum, D.M. et al. GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* **318**, 1266-1273 (2007).
6. Scheerer, P. et al. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497-502 (2008).
7. Olesen, C. et al. The structural basis of calcium transport by the calcium pump. *Nature* **450**, 1036-1042 (2007).
8. Shinoda, T., Ogawa, H., Cornelius, F. & Toyoshima, C. Crystal structure of the sodium-potassium pump at 2.4 A resolution. *Nature* **459**, 446-450 (2009).
9. Faham, S. et al. The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na+/sugar symport. *Science* **321**, 810-814 (2008).
10. Fang, Y. et al. Structure of a prokaryotic virtual proton pump at 3.2 A resolution. *Nature* **460**, 1040-1043 (2009).
11. Gao, X. et al. Structure and mechanism of an amino acid antiporter. *Science* **324**, 1565-1568 (2009).
12. Ressl, S., Terwisscha van Scheltinga, A.C., Vonrhein, C., Ott, V. & Ziegler, C. Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP. *Nature* **458**, 47-52 (2009).
13. Shaffer, P.L., Goehring, A., Shankaranarayanan, A. & Gouaux, E. Structure and mechanism of a Na+-independent amino acid transporter. *Science* **325**, 1010-1014 (2009).
14. Shimamura, T. et al. Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science* **328**, 470-473 (2010).
15. Weyand, S. et al. in Handbook of Metalloproteins. (ed. A. Messerschmidt) (John Wiley, Chichester; 2010).
16. Weyand, S. et al. The alternating access mechanism of transport as observed in the sodium-hydantoin transporter Mhp1. *J Synchrotron Radiat* **18** 20-23 (2011).
17. Weyand, S. et al. Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* **322**, 709-713 (2008).
18. Aller, S.G. et al. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* **323**, 1718-1722 (2009).
19. Gerber, S., Comellas-Bigler, M., Goetz, B.A. & Locher, K.P. Structural basis of trans-inhibition in a molybdate/tungstate ABC transporter. *Science* **321**, 246-250 (2008).
20. Hilf, R.J. & Dutzler, R. Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel. *Nature* **457**, 115-118 (2009).
21. Henderson, R. et al. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J Mol Biol* **213**, 899-929 (1990).
22. Baker, M. Making membrane proteins for structures: a trillion tiny tweaks. *Nat Methods* **7**, 429-434 (2010).
23. Bonander, N. et al. Altering the ribosomal subunit ratio in yeast maximizes recombinant protein yield. *Microb Cell Fact* **8**, 10 (2009).
24. Wagner, S. et al. Tuning *Escherichia coli* for membrane protein overexpression. *Proc Natl Acad Sci U S A* **105**, 14371-14376 (2008).
25. Grisshammer, R. & Tate, C. Overexpression of integral membrane proteins for structural studies. *Q Rev Biophys* **3**, 315-422 (1995).
26. Tate, C.G. Overexpression of mammalian integral membrane proteins for structural studies. *FEBS Lett* **504**, 94-98 (2001).

27. Tucker, J. & Grisshammer, R. Purification of a rat neurotensin receptor expressed in *Escherichia coli*. *Biochem J* **317 (Pt 3)**, 891-899 (1996).
28. Petaja-Repo, U.E., Hogue, M., Laperriere, A., Walker, P. & Bouvier, M. Export from the endoplasmic reticulum represents the limiting step in the maturation and cell surface expression of the human delta opioid receptor. *J Biol Chem* **275**, 13727-13736 (2000).
29. Ward, C.L. & Kopito, R.R. Intracellular turnover of cystic fibrosis transmembrane conductance regulator. Inefficient processing and rapid degradation of wild-type and mutant proteins. *J Biol Chem* **269**, 25710-25718 (1994).
30. Sarramegna, V., Talmont, F., Demange, P. & Milon, A. Heterologous expression of G-protein-coupled receptors: comparison of expression systems from the standpoint of large-scale production and purification. *Cell Mol Life Sci* **60**, 1529-1546 (2003).
31. Tate, C.G., Whiteley, E. & Betenbaugh, M.J. Molecular chaperones stimulate the functional expression of the cocaine-sensitive serotonin transporter. *J Biol Chem* **274**, 17551-17558 (1999).
32. Higgins, M.K., Demir, M. & Tate, C.G. Calnexin co-expression and the use of weaker promoters increase the expression of correctly assembled Shaker potassium channel in insect cells. *Biochim Biophys Acta* **1610**, 124-132 (2003).
33. Linares, D.M., Geertsma, E.R. & Poolman, B. Evolved *Lactococcus lactis* strains for enhanced expression of recombinant membrane proteins. *J Mol Biol* **401**, 45-55 (2010).
34. Wagner, S. et al. Consequences of membrane protein overexpression in *Escherichia coli*. *Mol Cell Proteomics* **6**, 1527-1550 (2007).
35. Bonander, N. et al. Design of improved membrane protein production experiments: quantitation of the host response. *Protein Sci* **14**, 1729-1740 (2005).
36. Bonander, N. & Bill, R.M. Relieving the first bottleneck in the drug discovery pipeline: using array technologies to rationalise membrane protein production. *Expert Rev Proteomics* **6**, 501-505 (2009).
37. De Schutter, K. et al. Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nat Biotechnol* **27**, 561-566 (2009).
38. Singh, S. et al. Large-scale functional expression of WT and truncated human adenosine A2A receptor in *Pichia pastoris* bioreactor cultures. *Microb Cell Fact* **7**, 28 (2008).
39. Mattanovich, D., Gasser, B., Hohenblum, H. & Sauer, M. Stress in recombinant protein producing yeasts. *J Biotechnol* **113**, 121-135 (2004).
40. Ferndahl, C. et al. Increasing cell biomass in *Saccharomyces cerevisiae* increases recombinant protein yield: the use of a respiratory strain as a microbial cell factory. *Microb Cell Fact* **9**, 47 (2010).
41. Kunji, E.R., Slotboom, D.J. & Poolman, B. *Lactococcus lactis* as host for overproduction of functional membrane proteins. *Biochim Biophys Acta* **1610**, 97-108 (2003).
42. Surade, S. et al. Comparative analysis and "expression space" coverage of the production of prokaryotic membrane proteins for structural genomics. *Protein Sci* **15**, 2178-2189 (2006).
43. Geertsma, E.R. & Poolman, B. Production of membrane proteins in *Escherichia coli* and *Lactococcus lactis*. *Methods Mol Biol* **601**, 17-38 (2010).
44. Berntsson, R.P. et al. Selenomethionine incorporation in proteins expressed in *Lactococcus lactis*. *Protein Sci* **18**, 1121-1127 (2009).
45. Marreddy, R.K. et al. Amino acid accumulation limits the overexpression of proteins in *Lactococcus lactis*. *PLoS One* **5**, e10317 (2010).
46. Berntsson, R.P. et al. The structural basis for peptide selection by the transport receptor OppA. *EMBO J* **28**, 1332-1340 (2009).
47. Drew, D., Lerch, M., Kunji, E., Slotboom, D.J. & de Gier, J.W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nat Methods* **3**, 303-313 (2006).

48. Drew, D. et al. A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci* **14**, 2011-2017 (2005).
49. Geertsma, E.R., Groeneveld, M., Slotboom, D.J. & Poolman, B. Quality control of overexpressed membrane proteins. *Proc Natl Acad Sci U S A* **105**, 5722-5727 (2008).
50. Newstead, S., Kim, H., von Heijne, G., Iwata, S. & Drew, D. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **104**, 13936-13941 (2007).
51. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673-681 (2006).
52. Jacquier, V., Prummer, M., Segura, J.M., Pick, H. & Vogel, H. Visualizing odorant receptor trafficking in living cells down to the single-molecule level. *Proc Natl Acad Sci U S A* **103**, 14325-14330 (2006).
53. Meyer, B.H. et al. FRET imaging reveals that functional neurokinin-1 receptors are monomeric and reside in membrane microdomains of live cells. *Proc Natl Acad Sci U S A* **103**, 2138-2143 (2006).
54. Danelon, C., Terrettaz, S., Guenat, O., Koudelka, M. & Vogel, H. Probing the function of ionotropic and G protein-coupled receptors in surface-confined membranes. *Methods* **46**, 104-115 (2008).
55. Tate, C.G. Practical considerations of membrane protein instability during purification and crystallisation. *Methods Mol Biol* **601**, 187-203 (2010).
56. Tate, C.G. & Schertler, G.F. Engineering G protein-coupled receptors to facilitate their structure determination. *Curr Opin Struct Biol* **19**, 386-395 (2009).
57. Palczewski, K. et al. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-745 (2000).
58. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat Protoc* **4**, 706-731 (2009).
59. Weiss, H.M. & Grisshammer, R. Purification and characterization of the human adenosine A(2a) receptor functionally expressed in *Escherichia coli*. *Eur J Biochem* **269**, 82-92 (2002).
60. Jaakola, V.P. et al. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **322**, 1211-1217 (2008).
61. Kunji, E.R., Harding, M., Butler, P.J. & Akamine, P. Determination of the molecular mass and dimensions of membrane proteins by size exclusion chromatography. *Methods* **46**, 62-72 (2008).
62. Lau, F.W., Nauli, S., Zhou, Y. & Bowie, J.U. Changing single side-chains can greatly enhance the resistance of a membrane protein to irreversible inactivation. *J Mol Biol* **290**, 559-564 (1999).
63. Zhou, Y. & Bowie, J.U. Building a thermostable membrane protein. *J Biol Chem* **275**, 6975-6979 (2000).
64. Magnani, F., Shibata, Y., Serrano-Vega, M.J. & Tate, C.G. Co-evolving stability and conformational homogeneity of the human adenosine A2a receptor. *Proc Natl Acad Sci U S A* **105**, 10744-10749 (2008).
65. Serrano-Vega, M.J., Magnani, F., Shibata, Y. & Tate, C.G. Conformational thermostabilization of the beta1-adrenergic receptor in a detergent-resistant form. *Proc Natl Acad Sci U S A* **105**, 877-882 (2008).
66. Serrano-Vega, M.J. & Tate, C.G. Transferability of thermostabilizing mutations between beta-adrenergic receptors. *Mol Membr Biol* **26**, 385-396 (2009).
67. Shibata, Y. et al. Thermostabilization of the neurotensin receptor NTS1. *J Mol Biol* **390**, 262-277 (2009).
68. Warne, T., Serrano-Vega, M.J., Tate, C.G. & Schertler, G.F. Development and crystallization of a minimal thermostabilised G protein-coupled receptor. *Protein Expr Purif* **65**, 204-213 (2009).

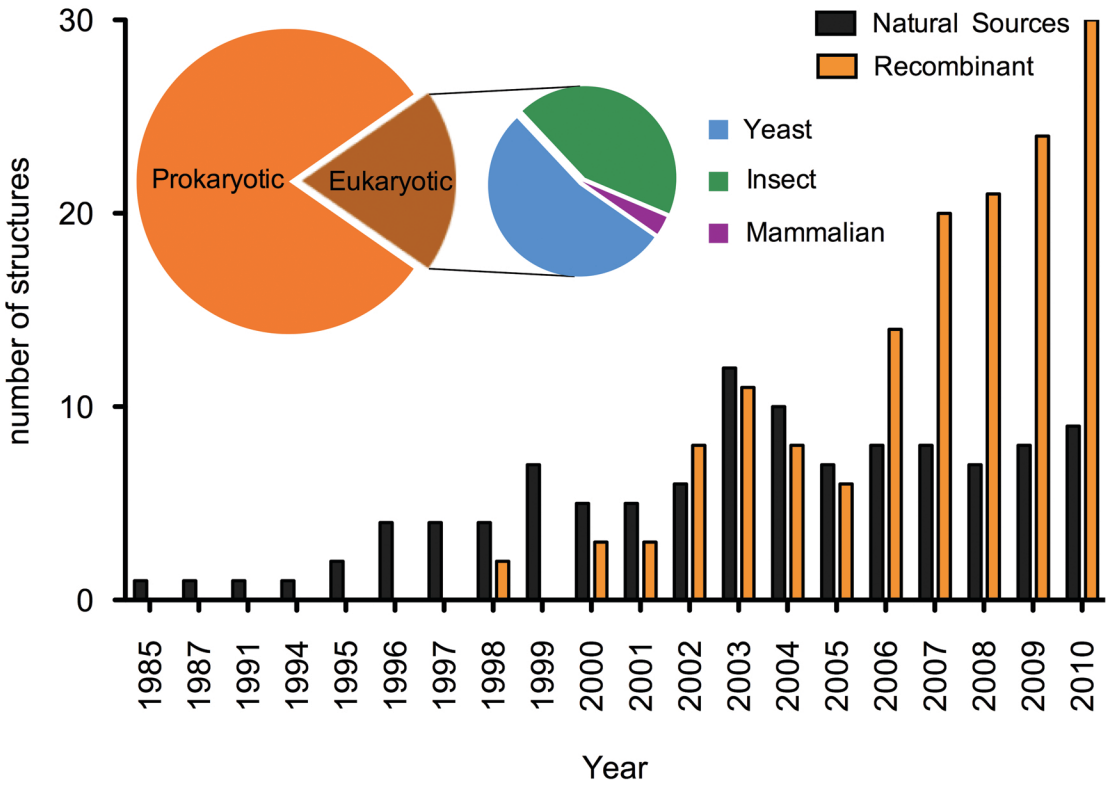
69. Hanson, M.A. et al. A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor. *Structure* **16**, 897-905 (2008).
70. Wacker, D. et al. Conserved binding mode of human beta2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *J Am Chem Soc* **132**, 11443-11445 (2010).
71. Warne, T. et al. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486-491 (2008).
72. Wu, B. et al. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330**, 1066-1071 (2010).
73. Chien, E.Y. et al. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **330**, 1091-1095 (2010).
74. Hirai, T. & Subramaniam, S. Protein conformational changes in the bacteriorhodopsin photocycle: comparison of findings from electron and X-ray crystallographic analyses. *PLoS One* **4**, e5769 (2009).
75. de Groot, B.L., Engel, A. & Grubmuller, H. The structure of the aquaporin-1 water channel: a comparison between cryo-electron microscopy and X-ray crystallography. *J Mol Biol* **325**, 485-493 (2003).
76. Gautier, A., Mott, H.R., Bostock, M.J., Kirkpatrick, J.P. & Nietlispach, D. Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nat Struct Mol Biol* **17**, 768-774 (2010).
77. Neutze, R. et al. Bacteriorhodopsin: a high-resolution structural view of vectorial proton transport. *Biochim Biophys Acta* **1565**, 144-167 (2002).
78. Long, S.B., Campbell, E.B. & Mackinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K+ channel. *Science* **309**, 897-903 (2005).
79. Bowler, M.W. et al. Diffraction Cartography: applying microbeams to macromolecular crystallography sample evaluation and data collection. *Acta Crystallogr. D* **66**, 855-864 (2010).
80. Moukhametzianov, R. et al. Protein crystallography with a micrometre-sized synchrotron-radiation beam. *Acta Crystallogr D Biol Crystallogr* **64**, 158-166 (2008).
81. Henrich, B. et al. PILATUS: A single photon counting pixel detector for X-ray applications. *Nuclear Instruments & Methods in Physics Research Section a-Accelerators Spectrometers Detectors and Associated Equipment* **607**, 247-249 (2009).
82. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752-757 (2000).
83. Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. Structure at 2.8 A resolution of cytochrome c oxidase from *Paracoccus denitrificans*. *Nature* **376**, 660-669 (1995).
84. Ostermeier, C., Iwata, S., Ludwig, B. & Michel, H. Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase. *Nat Struct Biol* **2**, 842-846 (1995).
85. Chaptal, V. et al. Fluorescence Detection of Heavy Atom Labeling (FD-HAL): A rapid method for identifying covalently modified cysteine residues by phasing atoms. *J Struct Biol* (2010).
86. Roszak, A.W., Gardiner, A.T., Isaacs, N.W. & Cogdell, R.J. Brominated lipids identify lipid binding sites on the surface of the reaction center from *Rhodobacter sphaeroides*. *Biochemistry* **46**, 2909-2916 (2007).
87. Schroder, G.F., Levitt, M. & Brunger, A.T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218-1222 (2010).
88. Wohri, A.B. et al. Light-induced structural changes in a photosynthetic reaction center caught by Laue diffraction. *Science* **328**, 630-633 (2010).
89. Andersson, M. et al. Structural dynamics of light-driven proton pumps. *Structure* **17**, 1265-1275 (2009).
90. Bamber, L., Harding, M., Butler, P.J. & Kunji, E.R. Yeast mitochondrial ADP/ATP carriers are monomeric in detergents. *Proc Natl Acad Sci U S A* **103**, 16224-16229 (2006).

Figure legends

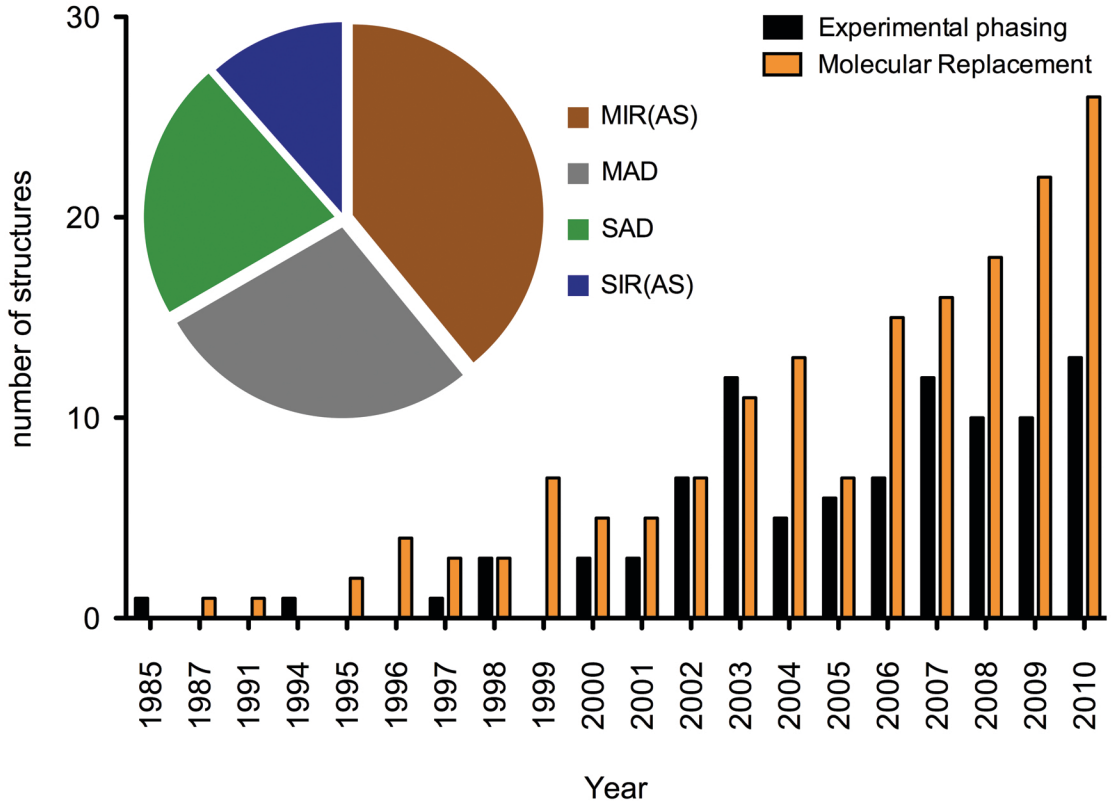
Figure 1. Progress in solving prokaryotic and eukaryotic membrane protein structures. (**a**) Trends in the use of host cells for the production of recombinant membrane proteins used in structural studies. The number of unique α-helical integral membrane protein structures deposited each year since 1985 is broken down according to whether the structure was derived from natural (black) or recombinant (orange) sources. Inset: a pie chart showing the breakdown of various recombinant host sources. (**b**) Trends in phasing methods for new membrane protein structures. The number of unique structures solved using either experimental (black) or molecular replacement (orange) methods is shown. Inset: a pie chart showing the breakdown of various experimental phasing methods. Unique structures are defined according to http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html.

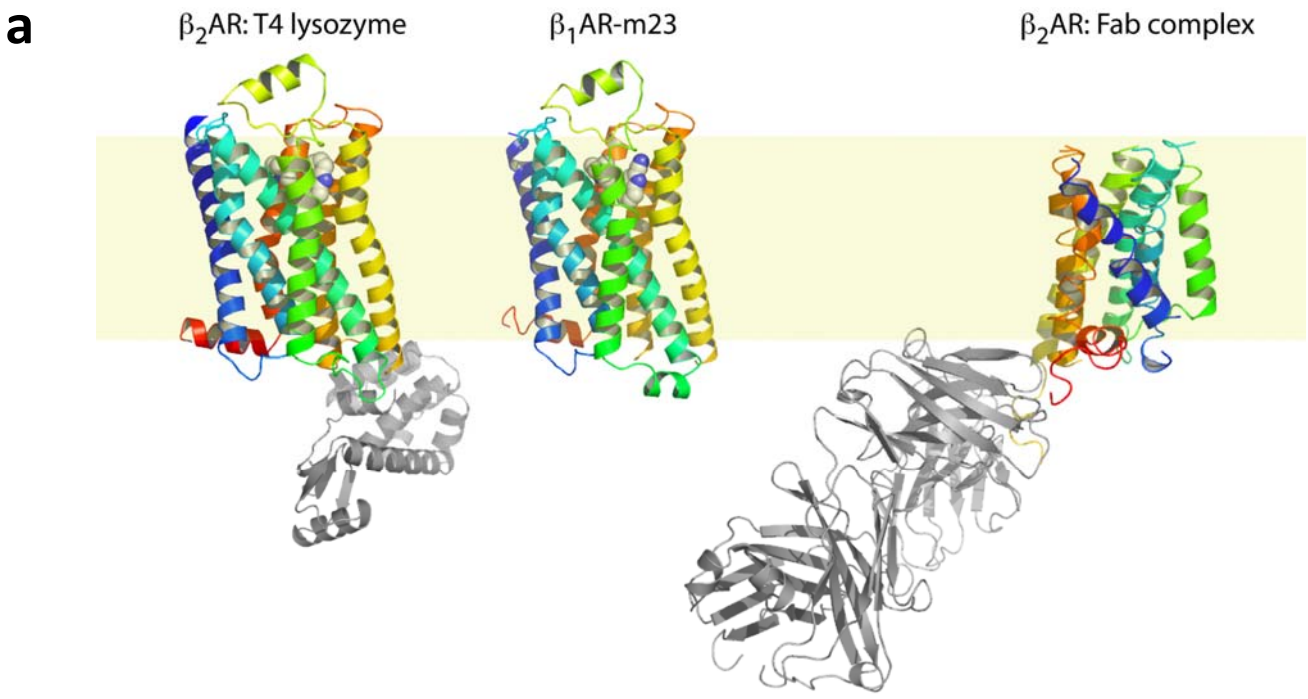
Figure 2. Strategies for crystallizing membrane proteins (**a**) Structures of β-AR achieved using different approaches: β_2AR engineered with T4 lysozyme inserted into intracellular loop 3 was crystallized in lipid cubic phase with cholesteryl hemisuccinate added to stabilize the receptor (left); the structure of thermostabilized β_1AR-m23 purified in octythioglucoside was determined by vapor diffusion crystallization (center); β_2AR stabilized in bicelles and bound to a F_{ab} antibody fragment was crystallized by vapor diffusion (right)[55]. The receptors are shown in rainbow coloration, and T4 lysozyme and F_{ab} antibody fragment are shown in grey. (**b**) The relative sizes of the detergent micelles surrounding a small membrane protein. The mitochondrial ATP/ADP carrier (30 kDa) was purified in detergents of the alkyl-maltoside series with decreasing hydrocarbon chain length from tridecylmaltoside (purple) to octylmaltoside (red). The dimensions of the detergent micelles were inferred from the Stokes radii of the free and protein-detergent micelles determined by size exclusion chromatography[90].

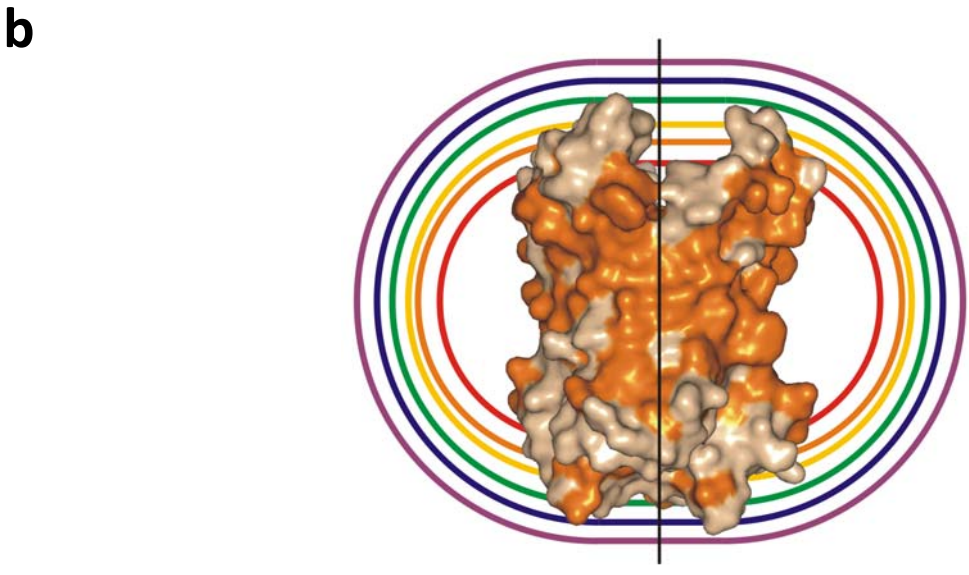
a


30
20
10
0
number of structures
Natural Sources
Recombinant
Prokaryotic
Eukaryotic
Yeast
Insect
Mammalian
1985
1987
1991
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
Year

b


30
20
10
0
number of structures
Experimental phasing
Molecular Replacement
MIR(AS)
MAD
SAD
SIR(AS)
1985
1987
1991
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
Year


a
β2AR: T4 lysozyme
β1AR-m23
β2AR: Fab complex


b