

Review

## Structural genomics of membrane proteins

Peter Walian\*, Timothy A Cross<sup>†</sup> and Bing K Jap\*

Addresses: \*Life Sciences Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720, USA. <sup>†</sup>National High Magnetic Field Lab and Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306, USA.

Correspondence: Bing K Jap. E-mail: bkjap@lbl.gov

Published: 15 March 2004

*Genome Biology* 2004, **5**:215

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/4/215>

© 2004 BioMed Central Ltd

### Abstract

Improvements in the fields of membrane-protein molecular biology and biochemistry, technical advances in structural data collection and processing, and the availability of numerous sequenced genomes have paved the way for membrane-protein structural genomics efforts. There has been significant recent progress, but various issues essential for high-throughput membrane-protein structure determination remain to be resolved.

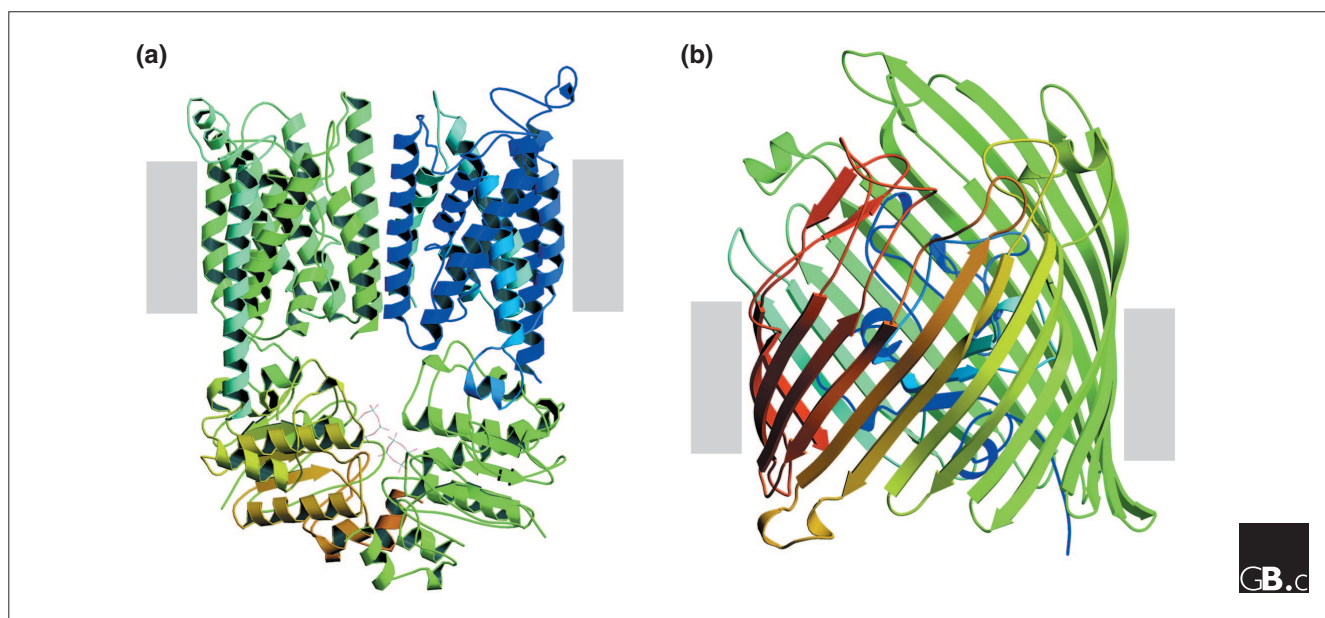
The goal of determining the structure of membrane proteins continues to define a substantial region of the structural biology horizon. While significant progress has been made over the past five years, the ratio of structures solved for membrane proteins to those solved for soluble proteins remains small, such that membrane proteins comprise less than 1 in 100 of the structures deposited in the Protein Data Bank (PDB) [1].

Although the total number of membrane-protein structures determined to date is but a small fraction of all protein structures determined, the more than 100 structures on deposit in public databases represent a substantial start. A burgeoning database [2] already contains examples of proteins with seven transmembrane helices, ion and water channels, transporters, ATPases, porins, toxins, and an array of proteins involved in energy production. While substantial architectural diversity can be found among the membrane-protein structures determined to date, they clearly fall into one of two general categories, those containing  $\alpha$ -helical transmembrane regions and those with transmembrane regions composed of  $\beta$  strands arranged to form barrel-like structures, the latter being primarily represented by the porins of bacterial outer membranes. Examples of structures from these two categories are shown in Figure 1.

The relatively small number of membrane-protein structures determined to date stems primarily from the requirement

for solubilization of membrane proteins before crystallization, while preserving the structural integrity of the solubilized protein. Despite this challenge, the need to increase the number of known membrane-protein structures is clear and is further emphasized by the estimate that more than 30% of a typical cell's proteins are membrane proteins [3] and that more than half of all membrane proteins are predicted to be pharmaceutical targets [4]. The recent modest increase in the rate of determining membrane-protein structures has been facilitated by improvements in the areas of membrane-protein molecular biology and biochemistry, and through technical advances in synchrotron X-ray beamlines for crystallography, high-field nuclear magnetic resonance (NMR) and high-resolution electron microscopy. The availability of sequenced genomes spanning a broad range of species has vastly improved searches for structural homologs and the prediction of previously unknown membrane proteins. These factors have converged to help set the stage for the determination of membrane-protein structures rapidly and on a large scale.

In recent years a number of consortia, bringing together researchers from a variety of academic and research institutions [5-7], have been established to address and execute the goals of structural genomics - that is, to dramatically increase the database of known protein structures by developing and applying methodologies to determine them as rapidly and cost-effectively as possible. To date, however, only one group

**Figure 1**

Membrane-protein models representing the two general categories of transmembrane-region structure. **(a)** Side-view of BtuCD Vitamin B12 Transporter [51] containing an  $\alpha$ -helical transmembrane region (PDB accession code 1L7V); and **(b)** side-view of FecA Ferric Citrate Uptake Receptor [52] featuring a  $\beta$ -barrel transmembrane region (PDB accession code 1KMO). Light gray blocks on the sides of each model depict the approximate limits of the lipid bilayer. The figure was produced using MOLSCRIPT [53] and Raster3D [54].

(*Mycobacterium tuberculosis* Membrane Protein Structural Genomics [8]) has taken on as its primary mission the high-throughput determination of membrane-protein structures. While the efforts of this group are ongoing, substantial progress has already been made in the construction of expression vectors on a large scale.

This article provides an overview of the factors essential for the determination of membrane-protein structures in high-throughput fashion and the progress that has been made so far in these areas. The key issues that arise for a researcher who wishes to determine the structure of membrane proteins at the atomic level are: how to produce sufficient protein, and once produced how to solubilize and purify the protein; then, how to crystallize the protein, or whether instead to study it in solution; and finally, how to scale up such methods for high-throughput structure determination.

## Protein overexpression

### Target selection

High-resolution structure-determination efforts typically require milligram quantities of proteins. Overexpression of prokaryotic genes in bacterial vectors currently provides the most direct and productive route to fulfilling this need [9,10]. Studies on genes with introns will require full-length cDNAs derived from mRNA libraries, and this represents another degree of complexity. Groups such as the Mammalian Gene Collection (MGC) [11], for example, have

created resources for the production and distribution of full-length human genes [12].

Prokaryotic genomes are also logical choices as target genomes for membrane-protein structural genomics efforts [13]. The initial goals of these efforts will be to clone, overexpress and purify the known and putative membrane proteins of their selected genomes. Potential membrane-protein targets can be identified from functional studies or on the basis of knowledge from previously characterized homologous genes. In many instances, however, homology-based predictions of protein type and function will not be possible. For these proteins, assignment of putative membrane-protein status will have to be based on predictions of transmembrane segments using the many bioinformatic tools now available (for example, see [14]).

Ideally, all successfully expressed and solubilized target membrane proteins should be distributed to X-ray and electron crystallography groups, and appropriate protein samples to NMR spectroscopy teams, for simultaneous efforts at structure determination and maximization of the likelihood that rapid progress will be made.

### Expression constructs

Once a membrane protein has been prioritized, by whatever means, for structure determination, the next step must be to overexpress the protein in a way that allows significant quantities to be isolated and purified for further study. The

majority of structural genomics consortia are pursuing high-throughput protein expression through constructs expressed in *Escherichia coli*. Expression in *E. coli* has numerous attributes that make it such a strong choice. It has clear advantages currently with respect to cost per gene expressed, the variety of specialized expression vectors available and the well-developed methods for labeling target proteins for NMR and X-ray diffraction studies [9,10].

Expression vectors based on promoters used by T7 RNA polymerase are in widespread use for the overexpression of soluble proteins among the various consortia [9,10]. It also appears that for the immediate future this class of vectors will be favored by research groups overexpressing membrane proteins. One concern, especially with respect to the overexpression of membrane proteins, is the effect of target-protein expression levels in the uninduced state ('leakiness') on host-cell health, and subsequently on the ability to overexpress properly folded proteins at high levels. Vectors with promoters less prone to leakage expression may have to be sought for the successful overexpression of certain membrane proteins.

Purification of overexpressed protein is greatly simplified and idealized for high-throughput studies through the use of constructs in which the target gene is fused to an affinity tag, whereby the tag can be placed at either the amino- or the carboxy-terminal end of the target protein, with a number of options in construct design. Examples of tags include glutathione *S*-transferase, maltose-binding protein and polyhistidine. By virtue of their ease of use, polyhistidine tags have seen the broadest application [9,10]. Although there are indications that amino-terminal polyhistidine-tag fusion proteins may have a better expression record with respect to membrane proteins [8], in our view the performance of both amino- and carboxy-terminal tagged constructs should be evaluated on a case-by-case basis with respect to target-protein overexpression, solubility, and crystallizability. To facilitate the subsequent removal of affinity tags, protease recognition sites can also be incorporated into the constructs; and in the case of membrane proteins it is desirable that these sites support the use of detergent-resistant proteases, so as to be compatible with detergent-based purification procedures.

Structure-determination efforts on human gene products have been limited, because of difficulties in obtaining high expression levels of protein. Many human genes will probably require some form of eukaryotic expression vector for successful overexpression. Numerous yeast, insect and mammalian cell lines could potentially serve in this capacity [13]; the development of eukaryotic expression methodologies tailored for high-throughput applications, however, is still in the nascent stages.

#### Host cells

The choice of host cells for overexpression of a given protein will depend on various factors, such as the source of the origi-

nal gene, the protein's fold complexity and the potential need for folding partners, and requirements for post-translational modification. As discussed earlier, the use of *E. coli* expression vectors runs strongly across the various structural genomics consortia and, in turn, dictates that some strain of *E. coli* will serve as host cell. Although there are a number of strains that have been used to express membrane proteins, BL21 (DE3) and a derivative of BL21 optimized for membrane-protein expression (designated C43) appear to be best suited for the task. Strain variant C43 grows more slowly than BL21, and in doing so may provide more time for the host cell to deal properly with higher than normal levels of membrane-protein expression [15]. Expression of both soluble and membrane proteins in a given bacterial strain can be quite sensitive to post-induction incubation temperature. The amount of overexpressed target membrane protein localized within lipid bilayers may be increased, and the occurrence of inclusion bodies containing aggregated protein reduced, by lowering incubation temperatures following induction.

Although not as practical for high-throughput purposes as bacterial expression systems, certain eukaryotic target proteins, either single polypeptides or those of multiple subunit complexes, may require 'higher' cell types to achieve adequate expression. Potential drawbacks to the use of eukaryotic cell types can derive from difficulties in protein isolation and yield, longer doubling times, and cost. Dealing with post-translational modifications, such as the removal of glycosylation usually required for success in crystallization, can be particularly challenging and may require inhibition of the host cell's glycosylation pathway during expression, or modification of the construct sequence, or treatment of expressed proteins with glycosidases [16].

#### Cell-free systems

Although still an evolving technology, cell-free expression systems offer an alternative methodology for the overexpression of proteins [17-19]. In such a system, essential protein expression machinery is obtained from cell lysates, which can be isolated and prepared in-house or obtained from commercial sources. Commercial systems are presently available with an advertised capability of producing 150 mg of protein from 30 ml of reaction mixture over a 24-hour time period [20]. Clearly this level of expression is well suited for high-throughput structural studies. For membrane proteins, however, expression away from lipid bilayer environments can, understandably, result in problems with protein folding and solubility. Supplementation of the reaction mixtures with detergents and lipids may provide a means of extending the utility of this approach to membrane proteins.

#### Solubilization and purification

##### Selection of solubilization detergents

Once sufficiently high level expression of the target membrane protein has been established, preferably in the host-cell

membrane, the next step is to determine the detergents best-suited for solubilization and subsequent purification. A wide variety of detergents suitable for membrane-protein solubilization are currently available (see, for example, [21]). Some of the most popular detergent families include the alkyl glucosides and maltosides, polyoxyethylenes, alkyldimethylamine oxides, and cholate derivatives. Experience has shown that the detergent selected for membrane extraction may not be the detergent of choice for crystallization. Broadly speaking, both the length of the detergent's hydrocarbon chain (hydrophobic domain) and the size of its polar head group (hydrophilic domain) are major factors affecting the stability of the solubilized protein - longer chain lengths and larger head groups are generally more favorable for the stability of the protein. When necessary, solubilizing detergents can be exchanged for other detergents through dialysis, or while the target protein is bound to chromatographic media. Some factors that need to be evaluated in choosing a detergent at the solubilization stage may be extraction yield, stability of the solubilized protein, and cost.

A particularly important criterion in selecting a detergent is its effect on a protein's structure and function. Certain detergents, particularly ionic ones, can denature membrane proteins, even when used at relatively low concentrations. Undesirable outcomes can involve varying degrees of denaturation, separation of subunits from multimeric or multi-subunit complexes, and aggregation [13]. Such potential results should be avoided prior to attempting crystallization and in collecting solution NMR data, for which samples should be monodisperse and stable, often at concentrations up to 10 mg/ml [10,22].

Evaluation of a detergent's effect on target-protein stability can provide a relatively quick means of assessing a detergent's suitability. One simple but effective test we have used involves solubilizing target proteins in candidate detergents and storing the mixtures overnight at room temperature. The various preparations can then be quickly evaluated to determine whether or not the protein has precipitated. Those solubilized proteins appearing stable can be examined more closely to determine the extent of homogeneity. Molecular-sieve chromatography, which separates molecules primarily by size, can reveal protein aggregation and/or oligomerization, as well as provide a means of improving sample homogeneity. Dynamic light scattering is another approach that can provide much of the same information about particle size as molecular-sieve chromatography, but more rapidly [23]. A form of NMR spectroscopy, heteronuclear single quantum correlation (HSQC), can also be used as a screening tool for the rapid assessment of target-protein quality [22]. If the function of a target protein is known or confidently predicted, functional assays should ideally be used to ensure that the protein is fully active in the candidate detergent.

## Purification

As in the case of high-throughput structure-determination efforts for soluble proteins, the process of purifying overexpressed membrane proteins has been substantially streamlined through the use of affinity tags. When coupled to the output of optimized host-cell systems, milligram quantities of relatively pure protein can be obtained following a single chromatographic step [13]. In many instances the target protein will already be sufficiently pure at this stage to begin structure-determination efforts. During this phase of purification it is also appropriate to address the possibility of whether the solubilizing detergent used is suitable for crystallization and for maintaining a monodisperse solution when the protein is highly concentrated. The effects of alternative detergents can be investigated by exchanging detergents while the target protein is bound to the affinity column. Should the affinity-column-purified sample require further cleanup, use of molecular-sieve chromatography is usually sufficient to remove minor contaminants and aggregates. On those occasions when residual contaminants cannot be isolated from the target protein on the basis of molecular weight differences, an alternative additional chromatographic step, such as ion-exchange chromatography, may be necessary [10].

## Structure determination

There are currently several approaches for determining the structure of membrane proteins, notably X-ray crystallography, electron crystallography and NMR spectroscopy. Given its history of demonstrated success, X-ray crystallography is regarded as the most widely proven tool for structure-determination efforts. But target-protein characteristics, such as molecular weight, solubility and crystallizability, may dictate that other methodologies are better suited for a particular gene product. For example, small detergent-solubilized membrane proteins or peptides with a very large hydrophobic surface area to volume ratio, which may not have good solubility properties at high concentration, may be excellent candidates for NMR spectroscopy.

## X-ray crystallography

X-ray crystallography provides an established means for obtaining high-resolution structural data from membrane proteins. With this approach, molecular weight seldom limits the choice of target protein, and determination of structures at atomic-level resolution is a very realistic goal. Often diffraction from a single crystal is sufficient for high-resolution structure determination. In many cases, the difficulties that were in the past associated with interpreting X-ray diffraction amplitudes in terms of how they reflect the underlying crystal structure - known as the 'phase problem' - have been dramatically reduced through the use of multi-wavelength anomalous diffraction techniques that rely on the use of X rays of multiple wavelengths and externally provided anomalously scattering atoms that yield reference points within the crystal

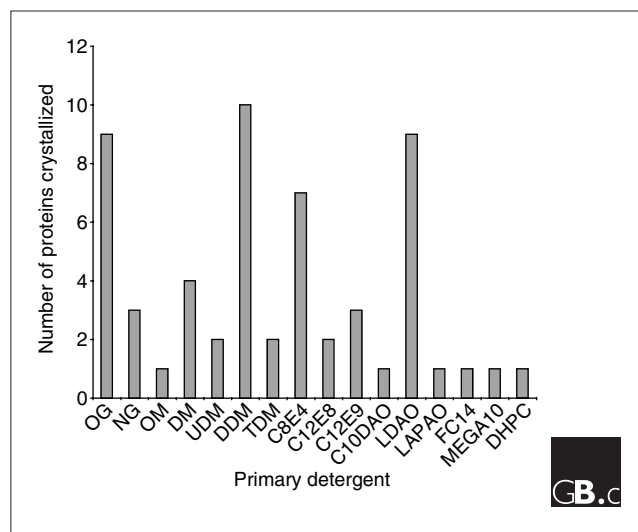
structure [24]. For example, tunable synchrotron X-ray sources facilitate the rapid phasing of diffraction data obtained from selenomethionine-derivatized target proteins, prepared through the metabolic labeling of proteins expressed in *E. coli*. Synchrotron X-ray sources also make it possible to obtain high-resolution datasets from microcrystals [25], which are often no larger than 50  $\mu\text{m}$  in their longest dimension, reducing potential bottlenecks associated with the need to optimize crystallization conditions in an effort to obtain large crystals. Further gains in sample throughput rates can be realized through automation-assisted screening of sample wells for the presence of crystals, and automated crystal handling and data collection.

The major challenge of the X-ray diffraction structure-determination approach lies in obtaining suitable three-dimensional crystals. As with soluble proteins, homogeneity and stability of the purified protein at high concentration is often critical for obtaining crystals. The strategies for crystallizing membrane proteins are similarly centered on reducing the solubility of the target protein under conditions that allow for the establishment of crystal-forming contacts between neighboring molecules [26]. Protein solubility is typically lowered through the use of precipitating agents, such as ammonium sulfate and polyethylene glycols. Experimentally variable parameters affecting the degree and nature of molecule-to-molecule contact include pH, ionic strength and temperature. A factor unique to the crystallization of membrane proteins is the presence of the substantial concentrations of detergent required to maintain solubility of the target protein. To ensure solubility of a target protein the concentration of detergent must be kept above the critical micelle concentration (CMC) which, depending on the detergent in question, could be well into the millimolar range.

As mentioned above, a detergent that is well-suited for protein solubilization may not be the detergent of choice for crystallization, and detergents can be exchanged through dialysis or during purification while the target protein is bound to a chromatographic column. Just as with the screening of detergents for the optimization of solubilization, a variety of detergents should be screened during the course of crystallization trials. An impressive range of detergents has been used to obtain crystals yielding high-resolution diffraction (Figure 2). From crystal-packing considerations, detergents with the potential to yield the smallest possible micelle region on the solubilized protein should best support the formation of the protein-to-protein contacts needed for crystallization. A potential downside is that smaller detergents tend to have higher CMCs, requiring higher concentrations to maintain protein solubility, and are more likely to destabilize native structure. The goal should therefore be to identify the smallest detergents that maintain homogeneous and monodisperse solutions of structurally sound target proteins. Alternatively, the use of secondary detergents or amphiphiles as additives to alter the properties of mixed

micelles has also yielded high-quality crystals [27,28]. The location, or mere presence, of an affinity tag may play a role in determining whether a protein will crystallize, and it may be advisable to conduct crystallization trials on target proteins in which the polyhistidine affinity tags have been removed. Cleavage sites engineered into the expressed protein can be used to remove these tags using a detergent-resistant protease.

Membrane proteins have been crystallized using vapor diffusion (in which hanging and sitting drops of a solution containing the target protein are allowed to equilibrate with a reservoir solution containing a higher concentration of precipitant), and less frequently by dialysis and batch methods (where protein, precipitant and buffer are mixed to be at or very near the final concentrations required for crystallization). Sparse matrix screens (relatively small sets of crystallization conditions that survey a broad range of parameter space in coarse intervals) allow for rapid sampling of a diverse range of precipitant, pH and ionic strength conditions [29-31] have been successfully applied and are even available commercially (see, for example, [32]). These crystallization and screening methods lend themselves well to high-throughput robotics-based automation [33,34]. Recently



**Figure 2**  
Primary detergents used to obtain crystals of membrane proteins suitable for high-resolution structural studies [1]. The proportion of proteins solved with each detergent is indicated. Abbreviations: OG, n-octyl- $\beta$ -D-glucopyranoside; NG, n-nonyl- $\beta$ -D-glucopyranoside; OM, n-octyl- $\beta$ -D-maltopyranoside; DM, n-decyl- $\beta$ -D-maltopyranoside; UDM, n-undecyl- $\beta$ -D-maltopyranoside; DDM, n-dodecyl- $\beta$ -D-maltopyranoside; TDM, n-tridecyl- $\beta$ -D-maltopyranoside; C8E4, polyoxyethylene (4) octyl ether; C12E8, polyoxyethylene (8) dodecyl ether; C12E9, polyoxyethylene (9) dodecyl ether; C10DAO, n-decyl-N,N-dimethylamine-N-oxide; LDAO, n-dodecyl-N,N-dimethylamine-N-oxide; LAPAO, 3-laurylamido-N,N-dimethylpropylaminoxide; FC14, n-tetradecylphosphocholine; MEGA10, decanoyl-N-methylglucamide; DHPC, 1,2-diheptanoyl-sn-glycero-3-phosphocholine.

developed microfluidics devices can also support the rapid setup and evaluation of extensive crystallization screens using extremely small amounts of sample [35]. For example, in one commercially available system it is possible to survey up to 144 conditions from a total of 3  $\mu$ l of sample [36,37]. This method mixes reagent and sample through the process of free interface diffusion, whereby the protein and reagent are free to move throughout the system, and may allow for novel high-throughput surveys of crystallization space.

Techniques directly targeting the unique concerns of membrane-protein crystallization have also been developed; these include methods involving the use of lipidic cubic phases [38] and bicelles [39]. The rationale behind these methods is the notion that placing the solubilized protein back into a native-like environment will improve the chances of crystallization. Both of these approaches involve crystallization of the membrane protein within the context of lipid bilayers and have been used to produce well-ordered crystals of bacterial rhodopsins; but it remains to be determined to what extent the same approaches will apply to other membrane-protein families.

### **NMR spectroscopy**

Several different NMR technologies utilize a wide variety of membrane-mimetic environments. Solution NMR requires isotropic motions of the protein, and hence membrane proteins must be solubilized within detergent micelles. In homogeneous monodisperse samples, membrane proteins typically maintain not only their secondary and tertiary structure, but also their quaternary structure within micelles. In solid-state NMR, membrane proteins can be characterized in aligned planar bilayers by using orientational restraints that relate each atomic site of the protein to a reference axis perpendicular to the plane of the lipid bilayer. Solid-state NMR can also be used with samples that are not uniformly aligned, such as multilamellar liposomes, by using distance and torsional restraints that, respectively, constrain the structure by interatomic distances or define the relative orientation of adjacent atomic groups. Moreover, it may be possible to characterize membrane protein structure by solid-state NMR using micro- and nano-crystals of membrane proteins again through distance and torsional restraints. While NMR does not require diffraction-quality crystallization of membrane proteins, sample preparation is still a bottleneck, whether it is at the stage of detergent solubilization of the protein at high concentration, the reconstitution of protein into liposomes, or the uniform alignment of bilayer samples.

Solution NMR methodology has advanced with new procedures, such as transverse relaxation optimized spectroscopy (TROSY), that aid in data collection of samples that tumble slowly on the NMR frequency scale (500 to 900 MHz). NMR spectra of proteins require the use of  $^{15}\text{N}$  and  $^{13}\text{C}$  isotopic labeling to achieve sensitivity and resolution in the spectra.

The collection of structural restraints using this methodology is primarily from residual dipolar couplings (RDCs) that are derived from samples that have a slight degree of alignment with respect to the magnetic field and nuclear Overhauser effect (NOE)-derived inter-proton distances from samples that are extensively deuterated [40]. Such deuteration is required to improve the resolution in  $^1\text{H}$  spectra. Excellent progress has been made recently in the development of partial alignment of these proteins by using stretched polyacryamide gels that generate an anisotropic environment for the protein [41]. In other words, the protein in these gels has a slight preference for one orientation over other possible orientations. For  $\alpha$  helices, a pattern with 3.6 resonances per cycle is observed, in a phenomenon known as a dipolar wave [42]. Because  $\alpha$  helices have 3.6 residues per turn about the helical axis, a pattern in the RDCs of the backbone amide  $^{15}\text{N}$  resonances is observed with the same periodicity. The amplitude of the waves represented on plots of RDCs versus residue number is characteristic of the tilt angle of the helix with respect to the alignment axis and the magnetic field axis. Recent success with this approach has resulted in submissions of structures to the PDB [43] and progress with polytopic oligomeric proteins is progressing in additional laboratories (see [44], for example).

In solid-state NMR two technologies are utilized [45], one requiring aligned planar bilayer samples and the other using magic angle spinning (MAS) samples, in which samples of liposomes or micro- or nano-crystals are rotated about an axis inclined at  $54.7^\circ$  with respect to the magnetic field. In this way the anisotropic properties of the spectra are removed and a solution-like spectrum is observed. Uniformly aligned bilayers yield anisotropic observables, such as dipolar and quadrupolar couplings, as well as anisotropic chemical shifts. In other words, these NMR spin interactions display an orientation dependence with respect to the axis of the magnetic field of the NMR spectrometer. In this way the observed couplings and chemical shifts can be related to the orientation of the atomic sites with respect to the bilayer normal, which is aligned parallel to the magnetic field. As for the dipolar waves described above, the spectra of uniformly aligned samples in which the  $^{15}\text{N}$ - $^1\text{H}$  dipolar interaction is correlated with the  $^{15}\text{N}$  chemical shift, result in circular patterns of resonances for  $\alpha$ -helical segments with 3.6 resonances per turn, reminiscent of helical wheels. Here the patterns, known as PISA wheels [46,47], represent an opportunity to assess helix tilt angles and orientations without the need for, or with minimal, resonance assignments, respectively. This methodology represents an excellent screening tool for low-resolution structural information. Spectral sensitivity and resolution have been dramatically improved in the past decade, such that backbone structures of small proteins are now possible, and complete structures of peptides have been demonstrated. MAS experiments lead to distance and torsion restraints that are highly complementary to orientational restraints. Resolution is dramatically improved in MAS experiments by using

crystalline samples in which the conformation and environment of each protein is nearly identical.

While the first solid-state NMR structure was deposited in the PDB in 1997, since then another nine structures have been deposited in the data bank, and these nine were determined using either uniformly aligned samples or MAS samples. In addition, the structures of three  $\beta$ -barrel membrane proteins have been solved by solution NMR.

### Electron crystallography

Two-dimensional crystals, which are sheet-like crystals one unit-cell thick, are ideally suited for electron-crystallographic structure-determination methods. Membrane proteins crystallized within the context of lipid bilayers represent one excellent example; numerous electron crystallography structural studies of membrane proteins have been performed using such specimens. There is no 'phase problem' in electron crystallography as there is in X-ray crystallography, since electron micrographs of crystalline samples yield images from which phases can be determined directly. To produce a structural dataset consisting of diffraction amplitudes and phases, electron diffraction patterns (to obtain more accurate amplitudes) and electron micrographs (to obtain phases) are collected from tilted and untilted two-dimensional crystals. These data are subsequently processed and merged into three-dimensional sets of structure factors. Three-dimensional density maps of the target molecules obtained from their structure factors are then modeled and interpreted in much the same manner as electron density maps derived from X-ray diffraction data. Recent improvements in electron microscope automation have led to increased data-collection rates and reduced processing times.

Naturally occurring two-dimensional crystals of bacteriorhodopsin (found in the purple membrane of halobacteria) have yielded the best quality electron crystallographic data from a membrane protein to date, allowing an atomic-level model of this membrane protein to be obtained [48,49]. A substantial number of detergent-solubilized membrane proteins have been reconstituted to form two-dimensional crystals, some very well ordered and over 1  $\mu$ m in size. Because of limits in sample tilting, however, the distribution of resolution in density maps produced from two-dimensional crystals is anisotropic. The quality of diffraction, in the best direction, from the most useful of these crystals has ranged from about 7 Å resolution, which is sufficient to reveal the presence of transmembrane helices, up to the 4 to 3 Å resolution range, where the main chain of the polypeptide can be modeled and the larger side chains assigned.

Several methodologies have evolved for obtaining two-dimensional crystals from solubilized membrane proteins; these include reconstitution of membrane proteins into lipid bilayers, and crystallization along lipid monolayers at air-water interfaces or on preformed lipid tubes [50]. The

approach based on lipid-bilayer reconstitution is the only method that to date has yielded high-resolution structural data. The reconstitution procedures involve mixing the detergent-solubilized target protein and lipid at relatively low lipid-to-protein ratios, followed by removal, or reduction in the concentration, of detergent. This may be done by dialysis, by adsorption of detergent to polystyrene beads, or by dilution of the sample. Upon removal of detergent, protein and lipid can associate to form membranes with a high density of proteins; under appropriate conditions, these are organized into crystalline arrays. The formation and quality of the resulting crystals depend on parameters such as the choice of lipid, protein concentration, protein-to-lipid ratio, detergents, rates of detergent removal, temperature and other factors, such as pH and ionic strength that are often found useful in three-dimensional protein crystallization. As with the other structure-determination techniques described above, the target membrane protein to be studied should be in a pure, homogeneous and stable state. The protein concentration required for these experiments is, however, substantially lower than for X-ray and NMR methods, at about 1 mg/ml.

### Frontiers of membrane-protein structural determination

Dramatic improvements in a range of technologies associated with membrane-protein structure determination have been realized over the past ten years, particularly in the areas of protein solubilization, crystallization and NMR sample preparation, as well as data collection and processing. Automation of processes in some of these areas is expected to further accelerate progress. The availability of a broad spectrum of fully sequenced genomes, coupled with advanced molecular biology techniques, means that literally thousands of membrane proteins can be made available for study. Clearly the time is right for membrane-protein structural genomics efforts to move into full swing.

### Acknowledgements

The authors gratefully acknowledge Robert Nakamoto for helpful discussions and Young Do Kwon for figure preparation. This work was supported by funding from the National Institutes of Health (P01-GM64676) and by the US Department of Energy.

### References

1. **Protein Data Bank** [<http://www.rcsb.org/pdb>]
2. **Membrane Proteins of Known 3D Structure** [[http://blanco.biomol.uci.edu/Membrane\\_Proteins\\_xtal.html](http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html)]
3. Wallin E, von Heijne G: **Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms.** *Protein Sci* 1998, **7**:1029-1038.
4. Russell RB, Eggleston DS: **New roles for structure in biology and drug discovery.** *Nat Struct Biol* 2000, **7 Suppl**:928-930.
5. Heinemann U: **Structural genomics in Europe: slow start, strong finish?** *Nat Struct Biol* 2000, **7 Suppl**:940-942.
6. Terwilliger T: **Structural genomics in North America.** *Nat Struct Biol* 2000, **7 Suppl**:935-939.
7. Yokoyama S, Hirota H, Kigawa T, Yabuki T, Shirouzu M, Terada T, Ito Y, Matsuo Y, Kuroda Y, Nishimura Y, et al.: **Structural**

- genomics projects in Japan. *Nat Struct Biol* 2000, **7 Suppl**:943-945.
8. **Mycobacterium tuberculosis Membrane Protein Collaboratory** [<http://magnet.fsu.edu/~rhudy/grant/tb-start.html>]
  9. Edwards AM, Arrowsmith CH, Christendat D, Dharamsi A, Friesen JD, Greenblatt JF, Vedadi M: **Protein production: feeding the crystallographers and NMR spectroscopists.** *Nat Struct Biol* 2000, **7 Suppl**:970-972.
  10. Goulding CW, Perry LJ: **Protein production in *Escherichia coli* for structural studies by X-ray crystallography.** *J Struct Biol* 2003, **142**:133-143.
  11. **Mammalian Gene Collection** [<http://mgc.nci.nih.gov/>]
  12. **ATCC** [<http://www.atcc.org/>]
  13. Loll PJ: **Membrane protein structural biology: the high throughput challenge.** *J Struct Biol* 2003, **142**:144-153.
  14. **Expert Protein Analysis System ExPASy Molecular Biology Server** [<http://www.expasy.ch/>]
  15. Miroux B, Walker JE: **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* 1996, **260**:289-298.
  16. Sui H, Han B-G, Lee JK, Walian P, Jap BK: **Structural basis of water-specific transport through the AQPI water channel.** *Nature* 2001, **414**:872-878.
  17. Jermutus L, Ryabova LA, Pluckthun A: **Recent advances in producing and selecting functional proteins by using cell-free translation.** *Curr Opin Biotechnol* 1998, **9**:534-548.
  18. Kigawa T, Yabuki T, Yoshida Y, Tsutsui M, Ito Y, Shibata T, Yokoyama S: **Cell-free production and stable-isotope labeling of milligram quantities of proteins.** *FEBS Lett* 1999, **442**:15-19.
  19. Sawasaki T, Ogasawara T, Morishita R, Endo Y: **A cell-free protein synthesis system for high-throughput proteomics.** *Proc Natl Acad Sci USA* 2002, **99**:14652-14657.
  20. **Rapid Translation System** [<http://www.roche-applied-science.com/sis/proteinexpression/>]
  21. **Anatrace** [<http://www.anatrace.com/>]
  22. Montellione GT, Zheng D, Huang YJ, Gunsalus KC, Szyperski T: **Protein NMR spectroscopy in structural genomics.** *Nat Struct Biol* 2000, **7 Suppl**:982-985.
  23. Ferre-D'Amare AR, Burley SK: **Dynamic light scattering in evaluating crystallizability of macromolecules.** *Methods Enzymol* 1997, **276**:157-166.
  24. Hendrickson WA, Horton JR, LeMaster DM: **Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure.** *EMBO J* 1990, **9**:1665-1672.
  25. Pebay-Peyroula E, Rummel G, Rosenbusch JP, Landau EM: **X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases.** *Science* 1997, **277**:1676-1681.
  26. Ostermeier C, Michel H: **Crystallization of membrane proteins.** *Curr Opin Struct Biol* 1997, **7**:697-701.
  27. Deisenhofer J, Epp O, Miki K, Huber R, Michel H: **Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3 Å resolution.** *Nature* 1985, **318**:618-624.
  28. Iwata S, Lee JW, Okada K, Lee JK, Iwata M, Rasmussen B, Link TA, Ramaswamy S, Jap BK: **Complete structure of the II-subunit bovine mitochondrial cytochrome bcl complex.** *Science* 1998, **281**:64-71.
  29. Carter CW Jr, Carter CW: **Protein crystallization using incomplete factorial experiments.** *J Biol Chem* 1979, **254**:12219-12223.
  30. Cudney R, Patel S, Weisgraber K, Newhouse Y, McPherson A: **Screening and optimization strategies for macromolecular crystal growth.** *Acta Crystallogr D* 1994, **50**:414-423.
  31. Jancarik J, Kim S-H: **Sparse matrix sampling: a screening method for crystallization of proteins.** *J Appl Crystallogr* 1991, **24**:409-411.
  32. **Hampton Research** [<http://www.hamptonresearch.com/>]
  33. Hui R, Edwards A: **High-throughput protein crystallization.** *J Struct Biol* 2003, **142**:154-161.
  34. Stevens RC: **High throughput protein crystallization.** *Curr Opin Struct Biol* 2000, **10**:558-563.
  35. van der Woerd M, Ferree D, Pusey M: **The promise of macromolecular crystallization in microfluidic chips.** *J Struct Biol* 2003, **142**:180-187.
  36. Hansen CL, Skordalakes E, Berger JM, Quake SR: **A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion.** *Proc Natl Acad Sci USA* 2002, **99**:16531-16536.
  37. **Fluidigm** [<http://www.fluidigm.com/>]
  38. Landau EM, Rosenbusch JP: **Lipidic cubic phases: a novel concept for the crystallization of membrane proteins.** *Proc Natl Acad Sci USA* 1996, **93**:14532-14535.
  39. Faham S, Bowie JU: **Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure.** *J Mol Biol* 2002, **316**:1-6.
  40. Zwahlen C, Vincent SJF, Gardner KH, Kay LE: **Significantly improved resolution for NOE correlations from valine and isoleucine (C<sup>γ2</sup>) ethyl groups in <sup>15</sup>N, <sup>13</sup>C- and <sup>15</sup>N, <sup>13</sup>C,<sup>2</sup>H-labeled proteins.** *J Am Chem Soc* 1998, **120**:4825-4831.
  41. Tycko R, Blanco FJ, Ishii Y: **Alignment of biopolymers in strained gels: a new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR.** *J Am Chem Soc* 2000, **122**:9340-9341.
  42. Melesh MF, Opella SJ: **Dipolar waves as NMR maps of helices in proteins.** *J Magn Reson* 2003, **163**:288-299.
  43. Marassi FM, Opella SJ: **Simultaneous assignment and structure determination of a membrane protein from NMR orientational restraints.** *Protein Sci* 2003, **12**:403-411.
  44. Oxenoid K, Sonnichsen FD, Sanders CR: **Topology and secondary structure of the N-terminal domain of diacylglycerol kinase.** *Biochemistry* 2002, **41**:12876-12882.
  45. Fu R, Cross TA: **Solid-state nuclear magnetic resonance investigation of protein and polypeptide structure.** *Annu Rev Biophys Biomol Struct* 1999, **28**:235-268.
  46. Marassi F, Opella SJ: **A solid-state NMR index of helical membrane protein structure and topology.** *J Magn Reson* 2000, **144**:150-155.
  47. Wang J, Denny J, Tian C, Kim S, Mo Y, Kovacs F, Song Z, Nishimura K, Gan Z, Fu R, et al.: **Imaging membrane protein helical wheels.** *J Magn Reson* 2000, **144**:162-167.
  48. Grigorieff N, Ceska TA, Downing KH, Baldwin JM, Henderson R: **Electron-crystallographic refinement of the structure of bacteriorhodopsin.** *J Mol Biol* 1996, **259**:393-421.
  49. Kimura Y, Vassilyev DG, Miyazawa A, Kidera A, Matsushima M, Mitsuoka K, Murata K, Hirai T, Fujiyoshi Y: **Surface of bacteriorhodopsin revealed by high-resolution electron crystallography.** *Nature* 1997, **389**:206-211.
  50. Jap BK, Walian PJ: **Crystallization of proteins: two-dimensional.** In *Nature Encyclopedia of Life Sciences* London: Nature Publishing Group, April 2001. (doi:10.1038/ngp.els.0003041)
  51. Locher KP, Lee AT, Rees DC: **The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism.** *Science* 2002, **296**:1091-1098.
  52. Ferguson AD, Chakraborty R, Smith BS, Esser L, van der Helm D, Deisenhofer J: **Structural basis of gating by the outer membrane transporter FecA.** *Science* 2002, **295**:1715-1719.
  53. Kraulis PJ: **MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures.** *J Appl Crystallogr* 1991, **24**:946-950.
  54. Merritt EA, Bacon DJ: **Raster3D: photorealistic molecular graphics.** *Methods Enzymol* 1997, **277**:505-524.