
30 April 2011
ISSN 1871-6784, Volume 28, Issue 3
SPECIAL ISSUE: ADVANCES IN CELL-FREE PROTEIN EXPRESSION
EDITED BY: MICHAEL OVERDUIN & MICHAEL TAUSSIG
New
BIOTECHNOLOGY
• Original research & reviews
• News & views
• Opinion papers & discussion
• Industry news
Official Journal of the EFB
EUROPEAN FEDERATION OF BIOTECHNOLOGY

New Biotechnology • Volume 28, Number 3 • April 2011

Research Paper

Cell-free expression and labeling strategies for a new decade in solid-state NMR

Alaa Abdine, Michiel A. Verhoeven and Dror E. Warschawski

UMR 7099, CNRS and Université Paris Diderot, IBPC, 13 rue Pierre et Marie Curie, F-75005 Paris, France

Although solid-state NMR and cell-free expression have recently become standard methods in biology, the combination of the two is still at a very early stage of development. In this article, we will explore several approaches by which cell-free expression could help solid-state NMR in its quest for biomolecular structure and mechanism elucidation. Far from being just another structure determination technique, this quest is motivated by the unique possibility of using solid-state NMR to determine the high resolution structure of a membrane protein *within its native environment*, the lipid membrane. We will examine the specific sample preparation requirements that such a goal imposes and how cell-free expression can play a key role in such a protocol.

Introduction

After 50 years of development, solid-state nuclear magnetic resonance (SSNMR) has finally entered, in 2002, the small club of techniques providing the RCSB Protein Structure Database with high resolution atomic structures of biomolecules. About 20 SSNMR protein structures have been deposited in the past decade, 12 of them in the past two years (http://www.drorlist.com/nmr/SPNMR.html). Following the lesson learned from solution-state NMR, SSNMR will benefit immensely from developments in biochemistry, protein overproduction, sample preparation and isotopic labeling strategies. At the same time, Cell-Free Expression (CFE) is becoming an affordable technique for a protein biochemistry laboratory. Although SSNMR and CFE are entering the standard toolbox of the structural biologist, both techniques are still confronted with specific bottlenecks when tackling *membrane proteins*. We believe that CFE and SSNMR, together, could join forces to try and increase the throughput of membrane protein structure determination. In this article, we will consider several approaches by which CFE can help to design proteins that are isotopically labeled, specifically for the various steps of molecular structure determination by SSNMR: resonance assignments, distance measurements and structure–function relationship determination.

Corresponding author: Warschawski, D.E. (Dror.Warschawski@ibpc.fr)

Solid-state NMR

Nuclear magnetic resonance comes in different flavors and can be applied to a variety of samples, from individual molecules to the entire human body. As an example, soluble proteins under 30 kDa tumble in water with a correlation time of under 20 ns, which is fast enough to obtain high resolution ^{1}H NMR spectra in a high magnetic field and allow for the determination of a three-dimensional molecular structure. With such an approach, solution-state NMR has provided the structures of some 6000 soluble proteins in the past 20 years [1].

When the molecules do not tumble fast enough, spectral resolution is lost. The field of SSNMR applies not only to solid molecules, but also when the molecules are too large, the solution is too viscous or the temperature is too low. In that case, other approaches have to be followed, to compensate for nature's fate and to regain high spectral resolution. One such approach is magic-angle spinning (MAS), where the solid sample is spun very fast, almost mimicking the fast molecular tumbling in solution. The spinning rate is limited by current technology to around 70 kHz, or even 30 kHz on standard equipment, which is hardly enough to compete with nature in ^{1}H NMR spectroscopy, but largely enough for most ^{13}C or ^{15}N biomolecular NMR experiments. SSNMR is a fast growing technique that benefits immensely from constant technological developments such as high magnetic fields or fast spinning NMR probes [2].

1871-6784/$ - see front matter doi:10.1016/j.nbt.2010.07.014

Many intriguing biological molecules are 'solid' in the NMR sense, or can become solid (dried, frozen or crystallized) for the purpose of the experiment. When a crystal is large enough, the molecular structure can be determined by X-ray diffraction. When it is too small (also known as micro-crystalline), it can still be studied by SSNMR [3]. Some biological macromolecules are solid by nature. Amyloid peptides are soluble, but they polymerize to form solid fibers that are unsuitable for solution-state NMR or X-ray crystallography, but not for SSNMR [4]. Last, but not least, membrane proteins embedded in their hydrated lipid bilayers are not strictly solid, but they are solid in the NMR sense: they tumble too slowly to be studied by solution-state NMR (or by X-ray crystallography, for that matter). SSNMR has the immense advantage of being suited to the study of membrane proteins *within* their native environment [5]. All these approaches, combined with major technological developments, have allowed for the elucidation of about 20 protein structures by solid-state NMR since 2002 (http://www.drorlist.com/nmr/SPNMR.html).

NMR, whether in solution or in the solid state, is not a very sensitive technique. A typical sample contains 1 mM of protein, which requires the production of mg quantities and the development of protein overexpression systems. In addition, NMR often requires isotopic labels in the protein (mostly ^{13}C, ^{15}N and sometimes ^{2}H), incorporated uniformly or at specific places. As mentioned above, standard solution-state NMR is limited to relatively small molecules (under 100 kDa), otherwise the spectral resolution is lost. Because SSNMR does not rely on molecular tumbling, resolution (the inverse of the linewidth of a spectral line) is independent of molecular size, but strongly depends on sample homogeneity. For a given resolution, nuclei with similar magnetic environments will resonate at similar frequencies and may not be distinguishable one from another (spectral crowding). This limit can be overcome by designing a molecule (whatever the size) where only a *subset* of the atoms is 'NMR-visible', while the remaining atoms are invisible. Although a significant technological breakthrough was achieved in the past decade, that has allowed for structure determination of biomolecules by SSNMR, the next breakthrough will come from biotechnological developments that will help produce, in large quantities, proteins that will be engineered to incorporate isotopic labels at specific places. In this context, cell-free synthesis could well play a major role in the next decade.

Cell-free expression for solid-state NMR

Cell-free expression is a major development in structural biology, for both NMR and X-ray crystallography, because it allows for fast and easy overproduction of proteins, both wild-type and mutant, with or without non-natural amino acids or additional cofactors. For NMR, CFE offers the additional possibility of incorporating a selected subset of isotopically labeled amino acids, with very little metabolic scrambling. In addition, the rapid development of cell-free lysates now allows for a wide range of post-translational modifications [6].

For SSNMR, CFE offers several methods of preparing a membrane protein sample [7]. Membrane proteins are notoriously difficult to overproduce in cells, due to the availability of limited membrane surface. In addition, membrane proteins often require a complex targeting mechanism to transport them inside their host membrane. In many cases, they end up as inclusion bodies in the cell, largely unfolded, sometimes irreversibly [8]. By contrast, CFE can produce a membrane protein sample in two days, directly into lipid liposomes or nanodiscs, without the addition of detergents and with a purification method that does not require a tag. Should this method fail, several other options exist, such as using CFE in the presence of detergents or other surfactants. In that case, a tag is required for purification and the protein needs to be reconstituted in a lipid membrane for another two days and a slightly lower yield [9,10]. With CFE, the choice of lipid or lipid mixture, concentration, pH and ionic strength in the final sample is very flexible.

Labeling strategies

Cellular versus cell-free expression

Creative biochemists have come up with a variety of methods to produce molecules tailor made for NMR, by incorporating or removing isotopes at specific places in the molecule. The most popular method is uniform labeling in *E. coli*, using $^{15}NH_4Cl$ as a nitrogen source, ^{13}C-labeled glucose as a carbon source and/or $^{2}H_2O$ as a deuterium source [11]. When uniform labeling results in an overcrowded spectrum, one can use other sources of carbon, such as $^{13}C_2$-glycerol or $^{13}C_{1,3}$-glycerol, to produce partially labeled proteins [12]. Another way is to feed the bacteria directly with amino acids that are expected to end up in the protein, either with labeled amino acids (specific labeling), or with labeled nutriments together with 'cold' amino acids (reversed labeling [13]). The resulting proteins are never isotopically pure, because of metabolic scrambling, but they are of great help in the first steps of a protein NMR study. All these approaches have their advantages and disadvantages, including a cost that depends on the desired labeling scheme (see below).

CFE has opened up new ways to produce those samples, and new types of samples are available. Isotopic labels are introduced into the protein via the labeled amino acids in the CFE system. Amino acids are available with various patterns, uniformly or specifically ^{13}C, ^{15}N and/or ^{2}H labeled [14,15]. If a uniformly ^{13}C,^{15}N,^{2}H-labeled protein is required, all 20 uniformly labeled amino acids have to be incorporated, which can be prohibitively expensive, although affordable commercial mixtures are already available [16]. By contrast, for specific labeling, CFE is a very competitive method that is soon to become routine in an NMR laboratory. The efficiency of this approach in SSNMR has been shown in the context of heavily crowded membrane protein spectra [17,18], as can be seen in Fig. 1 with the 2D ^{13}C NMR spectra of the mechano-sensitive ion channel MscL, a 75-kDa pentameric α-helical protein reconstituted in a hydrated DOPC bilayer. Spectrum 1a shows a uniformly labeled protein, while spectrum 1b shows a specifically labeled protein on isoleucines and threonines only. Crowding in the specifically labeled spectrum is reduced, allowing the characterization of 10 out of 16 labeled isoleucines, and of all three threonine spin systems [18].

After obtaining a well-resolved set of NMR spectra, and before measuring distances between atoms that will contribute to the protein structure determination, one needs to assign each resonance to an atom on a specific amino acid. A simplistic approach would consist of making 20 different samples, each containing *one* labeled amino acid type, easily assigned by NMR. The cost of each

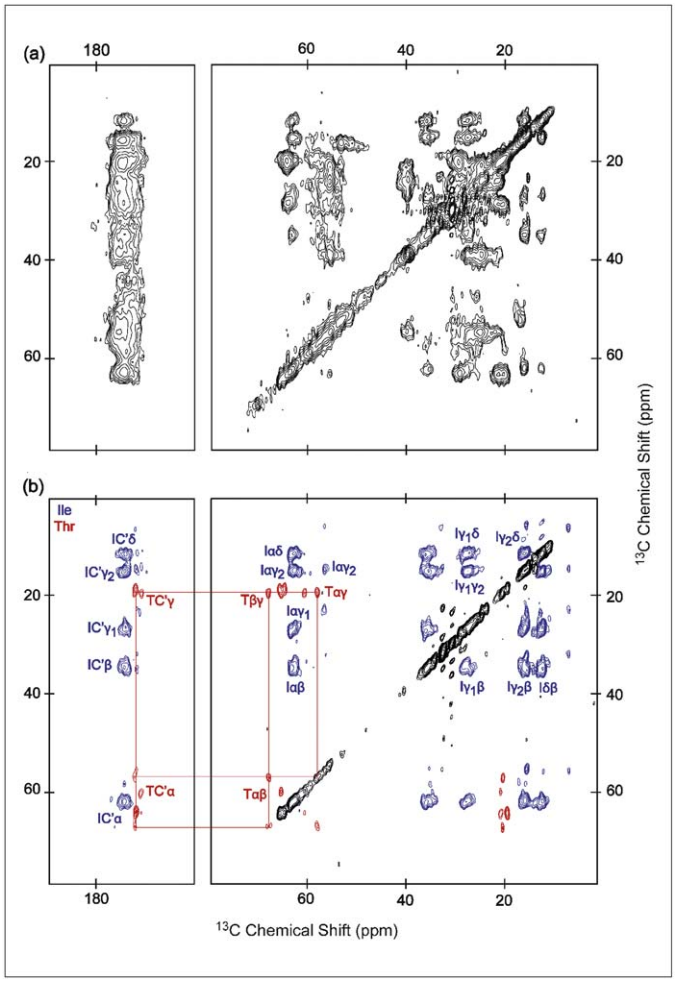

FIGURE 1

2D ^{13}C–^{13}C DARR NMR (100 ms mixing time) spectra of MscL/DOPC: **(a)** U-^{13}C,^{15}N-labeled protein produced in *E. coli* and **(b)** (Ile_{16}, Thr_3)-^{13}C,^{15}N-labeled protein produced *in vitro*. Identified isoleucine cross-peaks are colored in blue, while threonine cross-peaks are in red. The assignment pattern for one of the three threonines is indicated on the spectrum.

sample and the amount of NMR data that would be required point toward searching for better approaches. The main purpose of this article is to explore the various CFE strategies that one could adopt when studying a biomolecule by SSNMR.

Combinatorial approaches

For the initial and crucial assignment step, several combinatorial approaches have been suggested, tested and refined for solution-state NMR, based on the sole knowledge of the protein amino acid sequence. Rather than labeling all amino acids at once, which would often result in an overcrowded NMR spectrum, the idea is to make a reasonable number of samples, each with a different labeling scheme and a reasonable number of labels, covering all amino acids with all samples. The Otting group [19] suggests making five different samples, each comprising seven labeled amino acids, including only *one* of the five most abundant residues. The less abundant amino acids appear in several samples and can be identified by comparison between the five sets of NMR spectra. The 'dual combinatorial' approach of the Dötsch group [20] aims at identifying unique *pairs* of neighboring residue types, labeling these types and assigning those residues by observing unique cross-peaks. While the first approach allows the assignment of residue *types* (all glycines, all valines, among others), the second approach is less general, but allows *specific* residue assignment of glycine 22 and valine 23 (among others), as long as the (glycine 22–valine 23) pair is the only glycine–valine neighboring pair in the protein sequence. A preliminary example of this method is shown in Fig. 2 where a sample of MscL was ^{13}C-labeled on all phenylalanines, isoleucines, methionines, prolines and arginines using CFE, resulting in 9 unique pairs of labeled neighboring residue types along the protein sequence. Although the unique cross-peaks have not yet been localized on this spectrum with a short mixing time, many intra-residue cross-peaks have been identified.

Complete assignment of a protein is a difficult task, and sometimes 45% assignment is enough to continue and determine a protein structure [5]. Nevertheless, in some cases, a particularly important residue needs to be assigned unambiguously, despite spectral overlap. Site-directed mutagenesis offers the possibility of growing mutant proteins, both in cells and *in vitro*, that would replace all but one residue of a certain type by another type. The remaining residue could then be isotopically labeled and identified with certainty by NMR. In such a case, CFE can be preferred over biosynthesis, as exemplified by Lehner *et al*. [17] when they managed to identify the key residue for transport in the EmrE membrane protein by SSNMR.

These approaches have been designed for solution-state NMR, where ^{1}H–^{15}N spectra are acquired, amide protons play an essential role, prolines are excluded, and spectral overlap is worse in α-helices than in β-sheets. In SSNMR, ^{13}C–^{13}C and ^{13}C–^{15}N spectra

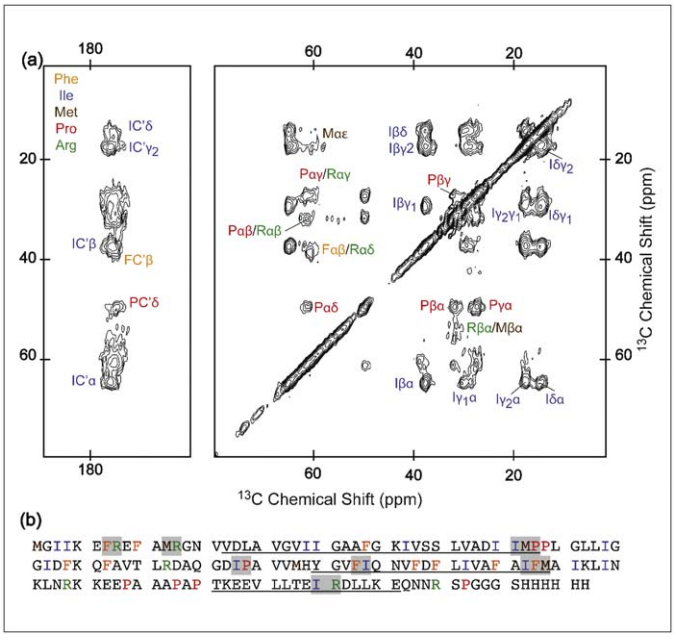

FIGURE 2

(a) 2D ^{13}C–^{13}C DARR NMR (50 ms mixing time) spectrum of (Phe_{10}, Ile_{16}, Met_5, Pro_7, Arg_6)-^{13}C,^{15}N-labeled MscL/DOPC produced *in vitro*. Preliminary identified intra-residue cross-peaks are indicated on the spectrum. **(b)** MscL amino acid sequence with the corresponding color coding. The nine unique pairs of labeled neighboring residue types are highlighted in gray. The three α-helices, of which the first two are transmembrane, are underlined.

are acquired, prolines are welcome, and spectral overlap is comparable within each secondary structure. Resolution is generally worse in SSNMR than in its solution-state counterpart, because magnetic interactions are stronger and because magnetic and conformational inhomogeneities are not entirely averaged out by molecular motion or packing. Another difference stems from the nature of polarization transfers. In solution-state NMR, polarization is transferred from one nucleus to its neighbor through chemical bonds, via the scalar coupling. In SSNMR, transfer is accomplished through space, via the dipolar coupling. Although dipolar transfer can be voluntarily limited to very short distances so as to resemble scalar transfers, it can also be extended over longer distances, and probe the environment of a nucleus beyond its neighbor. In other words, while this approach has been designed, in solution-state NMR, mostly for helping in the crucial residue assignment step, it can also be used in SSNMR and extended for long distance measurements and structure determination.

Predictive approaches

One major goal of specific labeling is to reduce the spectral crowding on NMR spectra. Although combinatorial approaches only require prior knowledge of the protein amino acid sequence, they do not take into account the regions of the spectra where those amino acids will appear and, potentially, overlap. Because a structural model, or the structure of an analog protein, is often available, the expected NMR spectra can be predicted and help choose the labeling scheme that will minimize spectral overlap. A variety of (usually free) software packages have been designed for this purpose, such as SPARTA [21], and this approach has been refined by Sweredoski *et al.* [22], again in the case of solution-state NMR, ^{1}H and ^{15}N detection.

The limit of such an approach is that of the NMR spectral resolution. The position and crowding of cross-peaks can be predicted, but not their linewidth. In SSNMR, ^{13}C linewidth can go from around 0.5 ppm in a quasi-crystalline membrane protein [23] to 0.8 ppm in a membrane with a high lipid and water content [18]. As an example, using a computer model for the membrane channel MscL [24], we have predicted that alanine, lysine, valine and tyrosine ^{13}C signals would not overlap. Through CFE, we have synthesized a protein with those four amino acid types ^{13}C-labeled. Indeed, as can be seen in Fig. 3, the various amino acid types do not overlap. Unfortunately, due to low spectral resolution of this membrane preparation, all amino acids of the *same* type overlap with one another and could not be assigned. For example, 13 valine carbon lines (of 0.8 ppm width each) are overlapped within a 3 ppm wide square. In addition, with this particular labeling scheme, no unique pair of neighboring amino acids was found either. By contrast, although most observed cross-peaks are intra-residues, a couple of alanine–valine *inter-residue* cross-signals are observed in Fig. 3, confirming that distances could be extracted with this approach, for future structure determination.

Structure–function approaches

Structure determination is most useful for understanding biological functions and molecular mechanisms. In this context, protein function can be understood by knowing a rough overall structure (secondary or low resolution structure) and a precise structure of only the crucial parts of the protein (active site, dimerization site, among others). Regardless of prior assignment or structural determination, another labeling strategy could therefore focus solely on regions that are expected to be important for the protein function, or whose structure needs to be refined. Once those crucial amino acids are identified, they can be isotopically labeled, observed by NMR and distances can be measured. Upon ligand binding, channel opening or conformational change, for example, these distances can be used as probes of molecular motion, and help visualizing the protein mechanism.

In the context of a structure–function study, the possibility of studying the structure of a mutant with a different functional phenotype is of great importance. Unfortunately, such mutants are sometimes lethal for bacterial growth. For example, a membrane channel that would be permanently open would cause a lethal cell leakage. In such a case, CFE is a very attractive alternative production method for structural biology, because the produced mutant can consequently be studied by electron microscopy [25] or by solid-state NMR.

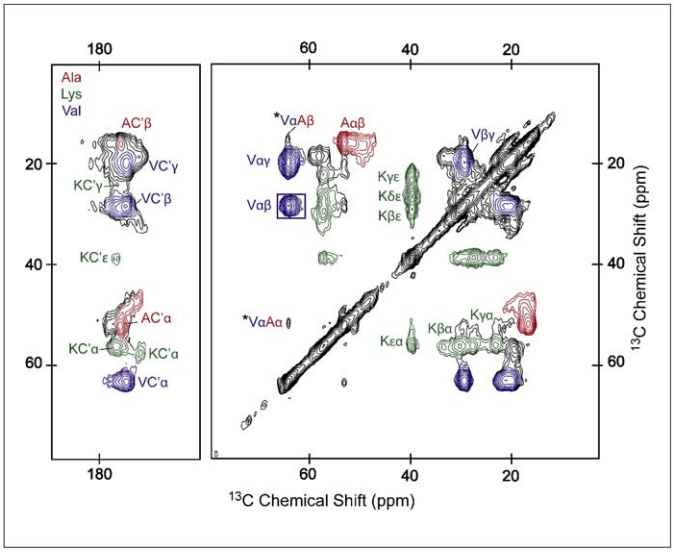

FIGURE 3

2D ^{13}C–^{13}C DARR NMR (100 ms mixing time) spectrum of (Ala_{15}, Lys_9, Val_{13}, Tyr_1)-^{13}C,^{15}N-labeled MscL/DOPC produced *in vitro*. Each identified set of intra-residue cross-peaks is colored differently. The blue square contains all 13 valine Cα to Cβ cross-peaks. Cross-peaks indicated with a star correspond to inter-residue peaks between valine Cα and alanine Cα and Cβ.

Cost

Although costs are highly variable, the cost difference between expression methods is an argument that is often put forward. Besides the labor cost, which is lower for CFE because the protocol is faster, it is therefore interesting to compare the approximate protein production cost for each approach, as of today. The price of 1 mg of uniformly ^{13}C,^{15}N-labeled protein expressed in *E. coli* is mostly the price of 3 g of uniformly ^{13}C-labeled glucose, in the order of 600 €. The price of 1 mg of a specifically labeled protein expressed *in vitro* through the Roche system is a combination of the price of the Roche RTS kit (300 €) and that of the ^{13}C,^{15}N-labeled amino acids (10 mg, around 50 €, of each). The price of 1 mg of a specifically labeled protein expressed in *E. coli* would be

that of the labeled amino acids, which would have to be added in larger quantities (100 mg approximately). For six ^{13}C,^{15}N-labeled residues, CFE would produce 1 mg of protein for 600 €, whereas biosynthesis would cost 3000 € per mg. Using a commercial kit such as RTS is advantageous for its convenience and reliability, for saving the time (and manpower) to make the lysate and also for managing the stocks. However, there are many protocols available for making lysates [26], which would reduce the cost of CFE by an additional significant factor.

Conclusion

In this article, we have provided an overview of several approaches that could be followed for protein studies by solid-state NMR. In addition to the methods developed for solution-state NMR, a similar effort has to be devoted to adapt them for their solid-state counterpart. Among those, we have described several labeling schemes developed in our laboratory on the membrane channel MscL, using cell-free expression or expression in *E. coli*, that will hopefully lead to a protein structure determination by SSNMR.

The rate of protein structure deposition at the RCSB Protein Structure Database, and determined by solid-state NMR, is not going to slow down soon. We hope that membrane protein structures will be a significant number of those in the next decade. While NMR still benefits from technological development (dynamic nuclear polarization, cryo-NMR, among others), progress in biomolecular NMR will also originate from the wet laboratory, by optimizing sample preparation, for which we believed cell-free expression of membrane proteins may well play a major role.

Acknowledgements

This work was supported by a fellowship from the Ministère de l'Enseignement Supérieur et de la Recherche (to AA), by the CNRS (UMR 7099), the ANR (ANR-06-JCJC0014) and the Université Paris Diderot. We thank Carole Gardiennet and Cathy Etchebest for help with the simulations, Carine Van Heijenoort for the NMR, Anders Pedersen, Catherine Berrier, Kyu-Ho Park and Francesca Zito for various tricks during sample preparations, and for helpful discussions.

References

1 Billeter, M. *et al.* (2008) Solution NMR structure determination of proteins revisited. *J. Biomol. NMR* 42, 155–158
2 Lesage, A. (2009) Recent advances in solid-state NMR spectroscopy of spin $I = 1/2$ nuclei. *Phys. Chem. Chem. Phys.* 11, 6876–6891
3 Bertini, I. *et al.* (2010) High-resolution solid-state NMR structure of a 17.6 kDa protein. *J. Am. Chem. Soc.* 132, 1032–1040
4 Wasmer, C. *et al.* (2008) Amyloid fibrils of the HET-s(218–289) prion form a beta solenoid with a triangular hydrophobic core. *Science* 319, 1523–1526
5 Cady, S.D. *et al.* (2010) Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. *Nature* 463, 689–692
6 He, M. (2008) Cell-free protein synthesis: applications in proteomics and biotechnology. *Nat. Biotechnol.* 25, 126–132
7 Abdine, A. et al. (2011) Cell-free membrane protein expression for solid-state NMR. In Protein NMR Techniques, 3rd Edition (Shekhtman A. and Burz D.S. eds), chapter 5, Methods in Molecular Biology series, Humana Press.
8 Wagner, S. *et al.* (2006) Rationalizing membrane protein overexpression. *Trends Biotechnol.* 24, 364–371
9 Park, K.H. *et al.* (2007) Fluorinated and hemifluorinated surfactants as alternatives to detergents for membrane protein cell-free synthesis. *Biochem. J.* 403, 183–187
10 Park, K.H. *et al.* In the cauldron of cell-free synthesis of membrane proteins: playing with new surfactants. *Nat. Biotechnol.* (this issue)
11 Goto, N.K. and Kay, L.E. (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr. Opin. Struct. Biol.* 10, 585–592
12 Castellani, F. *et al.* (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* 420, 98–102
13 Heise, H. *et al.* (2005) Molecular-level secondary structure, polymorphism, and dynamics of full-length alpha-synuclein fibrils studied by solid-state NMR. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15871–15876
14 Kainosho, M. *et al.* (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature* 440, 52–57
15 Hiller, M. *et al.* (2008) [2,3-(13)C]-labeling of aromatic residues-getting a head start in the magic-angle-spinning NMR assignment of membrane proteins. *J. Am. Chem. Soc.* 130, 408–409
16 Etezady-Esfarjani, T. *et al.* (2007) Cell-free protein synthesis of perdeuterated proteins for NMR studies. *J. Biomol. NMR* 39, 229–238
17 Lehner, I. *et al.* (2008) The key residue for substrate transport (Glu14) in the EmrE dimer is asymmetric. *J. Biol. Chem.* 283, 3281–3288
18 Abdine, A. *et al.* (2010) Structural study of the membrane protein MscL using cell-free expression and solid-state NMR. *J. Magn. Reson.* 204, 155–159
19 Ozawa, K. *et al.* (2006) ^{15}N-labelled proteins by cell-free protein synthesis. Strategies for high-throughput NMR studies of proteins and protein–ligand complexes. *FEBS J.* 273, 4154–4159
20 Sobhanifar, S. *et al.* (2010) Cell-free expression and stable isotope labelling strategies for membrane proteins. *J. Biomol. NMR* 46, 33–43
21 Shen, Y. and Bax, A. (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J. Biomol. NMR* 38, 289–302
22 Sweredoski, M.J. *et al.* (2007) Minimizing the overlap problem in protein NMR: a computational framework for precision amino acid labeling. *Bioinformatics* 23, 2829–2835
23 Shi, L. *et al.* (2009) Three-dimensional solid-state NMR study of a seven-helical integral membrane proton pump – structural insights. *J. Mol. Biol.* 386, 1078–1093
24 Valadié, H. *et al.* (2003) Dynamical properties of the MscL of *Escherichia coli*: a normal mode analysis. *J. Mol. Biol.* 332, 657–674
25 Yoshimura, K. *et al.* (2008) Gating-associated conformational changes in the mechanosensitive channel MscL. *Proc. Natl. Acad. Sci. U. S. A.* 105, 4033–4038
26 Schneider, B. *et al.* (2010) Membrane protein expression in cell-free systems. *Methods Mol. Biol.* 601, 165–186