

Site-specific labeling of proteins with NMR-active unnatural amino acids

David H. Jones · Susan E. Cellitti · Xueshi Hao · Qiong Zhang ·
Michael Jahnz · Daniel Summerer · Peter G. Schultz · Tetsuo Uno ·
Bernhard H. Geierstanger

Received: 26 May 2009 / Accepted: 17 July 2009 / Published online: 9 August 2009
© Springer Science+Business Media B.V. 2009

Abstract A large number of amino acids other than the canonical amino acids can now be easily incorporated in vivo into proteins at genetically encoded positions. The technology requires an orthogonal tRNA/aminoacyl-tRNA synthetase pair specific for the unnatural amino acid that is added to the media while a TAG amber or frame shift codon specifies the incorporation site in the protein to be studied. These unnatural amino acids can be isotopically labeled and provide unique opportunities for site-specific labeling of proteins for NMR studies. In this perspective, we discuss these opportunities including new photocaged unnatural amino acids, outline usage of metal chelating and spin-labeled unnatural amino acids and expand the approach to in-cell NMR experiments.

Keywords Site-specific labeling ·
Unnatural amino acids · Spin label · Metal chelator ·
In-cell NMR

Electronic supplementary material The online version of this article (doi:10.1007/s10858-009-9365-4) contains supplementary material, which is available to authorized users.

D. H. Jones · S. E. Cellitti · X. Hao · Q. Zhang ·
P. G. Schultz · T. Uno · B. H. Geierstanger (✉)
Genomics Institute of the Novartis Research Foundation, 10675
John Jay Hopkins Drive, San Diego, CA 92121-1125, USA
e-mail: bgeierst@gnf.org

M. Jahnz · D. Summerer · P. G. Schultz
Department of Chemistry and the Skaggs Institute for Chemical
Biology, The Scripps Research Institute, 10550 North Torrey
Pines Road, La Jolla, CA 92037, USA

Introduction

Nuclear magnetic resonance spectroscopy (NMR) is currently capable of studying the structure, ligand interactions and conformational changes of proteins and complexes approaching a megadalton in size (Wüthrich 1998; Tugarinov et al. 2004; Mittermaier and Kay 2006; Luy 2007). The ability to extend NMR to such large systems has been fueled by improvements in instrumentation, experiments tailored for large proteins (Pervushin et al. 1997; Riek et al. 1999; Tugarinov et al. 2003; Khaneja et al. 2004; Bromek et al. 2005) and isotope labeling approaches (Gardner et al. 1997; Gardner and Kay 1998; Goto and Kay 2000; Gross et al. 2003; Tugarinov and Kay 2005; Kainosho et al. 2006). However, the increased number of NMR correlations and broad resonance lines associated with larger proteins will generally result in more congested spectra that may become impossible to assign. A simple approach to reduce signal complexity is to employ amino acid type labeling methods (Markley et al. 1968; Muchmore et al. 1989). Unfortunately these methods are limited when applied to larger systems because assigning a large number of identical residues may still prove challenging. Single-site labeling with unnatural amino acids represents a powerful approach as incorporation instantaneously provides signal assignments. Here, we discuss the unique opportunities and applications of single-site labeling with NMR-active unnatural amino acids.

In vivo incorporation of unnatural amino acids

In vivo methods for site-specific incorporation of unnatural amino acids (Wang et al. 2001) provide unprecedented opportunities for protein labeling. The label can be directly

introduced or conjugated through the introduction of chemical functionalities not present on the canonical natural amino acids (reviewed in ref. (Xie and Schultz 2005b, 2005a, 2006; Wang et al. 2006b)). This approach requires an orthogonal tRNA/aminoacyl-tRNA synthetase pair that specifically incorporates the unnatural amino acid into the target protein in response to a TAG amber or frame shift codon. The unnatural amino acid is typically added to the culture media and is presumably transferred into the cells through endogenous transport mechanisms. Methods for selecting such unnatural amino acid-specific tRNA/aminoacyl-tRNA synthetase pairs from mutant libraries are well established. Depending on the amino acid structure, different host organisms including *E. coli* (Wang et al. 2001), mammalian cells (Liu et al. 2007), *Saccharomyces cerevisiae* and *Pichia pastoris* (Young et al. 2009), incorporate unnatural amino acids with high fidelity and high protein yield. Over 50 unnatural amino acids can now be incorporated into proteins (Xie and Schultz 2005b, a, 2006; Wang et al. 2006b).

Unnatural amino acids provide unique opportunities for structural biologists. The ability to site-specifically introduce post-translational modifications or analogous chemical structures is particularly important for studying signaling pathways. For example, a stable analog of phospho-tyrosine (Xie et al. 2007a) and sulfo-tyrosine (Liu and Schultz 2006; Liu et al. 2009) can be incorporated. The latter is an important post-translational modification for chemokine signaling (Veldkamp et al. 2006, 2008). Methylation or acetylation of lysines can be mimicked via chemical derivatization of phenylselenocysteine (Guo et al. 2008) or through direct incorporation of a modified lysine derivative (Neumann et al. 2008). Unnatural amino acids such as *p*-iodo-phenylalanine and *p*-bromo-phenylalanine can be used to obtain phases for X-ray crystallography (Xie et al. 2004). IR and fluorescent amino acids have also been introduced into proteins to provide spectroscopic probes at specific sites (Wang et al. 2003, 2006a; Schultz et al. 2006; Tsao et al. 2006; Groff et al. 2009). Here, we will focus on applications utilizing NMR-active unnatural amino acids (Fig. 1). These include isotopically labeled unnatural amino acids and photocaged amino acids. Spin-labeled and metal chelating unnatural amino acids will also be discussed. Finally, preliminary in-cell NMR experiments will be described.

Incorporation of NMR-active unnatural amino acids into proteins

The first NMR study using site-specific incorporation of an unnatural amino acid at TAG sites was performed 17 years ago by Wemmer, Schultz and coworkers (Ellman et al.

1992). Incorporation of ^{13}C -labeled alanine into T4 lysozyme was accomplished in an in vitro reaction using chemically aminoacylated suppressor tRNA. Subsequent technological advances have enabled the in vivo incorporation of many different unnatural amino acids and in a proof-of-concept study, ^{15}N -labeled *p*-methoxy-phenylalanine (*p*-OMePhe, Fig. 1a) was incorporated at modest yields into myoglobin (Deiters et al. 2005). More recently, an improved *E. coli* incorporation system was developed that readily enabled the production of sufficient quantities of protein for NMR using minimal or rich media (Cellitti et al. 2008). As is illustrated below, it is now possible to produce fully deuterated NMR samples containing isotopically labeled unnatural amino acids using minimal D_2O media.

Unnatural amino acid incorporation in *E. coli* currently requires two plasmids: one for the expression of the target TAG mutant protein and a second plasmid for the production of the orthogonal tRNA and aminoacyl-tRNA synthetase pair. The unnatural amino acid is typically added to the culture media 30–60 min prior to induction at a final concentration between 0.5 and 5 mM. For protein expression in rich media, we found that the activity of the engineered aminoacyl-tRNA synthetase limited expression of full-length protein. Because TAG naturally acts as a stop codon, an insufficient supply of charged tRNA led to the production of large amounts of truncated protein. Inducing over-expression of the aminoacyl-tRNA synthetase along with the protein of interest resulted in high levels of full-length protein and prevented truncation (Cellitti et al. 2008). In fact, for the thioesterase domain of human fatty acid synthase (FAS-TE; 33 kDa) sufficient protein for a 0.1 mM 500 μl NMR sample was typically obtained from a 50 ml culture. This was confirmed for three different unnatural amino acids: *p*-trifluoromethoxy-phenylalanine (*p*-OCF₃-Phe, Fig. 1a), $^{13}\text{C}/^{15}\text{N}$ -labeled *p*-OMePhe (Fig. 1a), and ^{15}N -labeled *o*-nitrobenzyl-tyrosine (*o*-NBTyr, Fig. 1b) that were individually incorporated at 11 different sites around the active site using only 8–25 mg of the unnatural amino acid per 50 ml culture (Cellitti et al. 2008).

For optimal protein expression yields, a series of plasmids with inducible expression of the aminoacyl-tRNA synthetases are now available. These expression systems are capable of producing NMR samples with an unnatural amino acid incorporated at two different sites within a protein using a double TAG mutant (unpublished results). High level expression can result in truncated material, and it is advantageous to employ a C-terminal affinity tag to simplify purification of the full-length protein. In the absence of a functional assay it is useful to confirm that the unnatural amino acid does not significantly alter the protein's structure. By preparing uniformly or selectively ^{15}N -labeled protein with unlabeled unnatural amino acid incorporated, the HSQC spectrum of the mutant can be

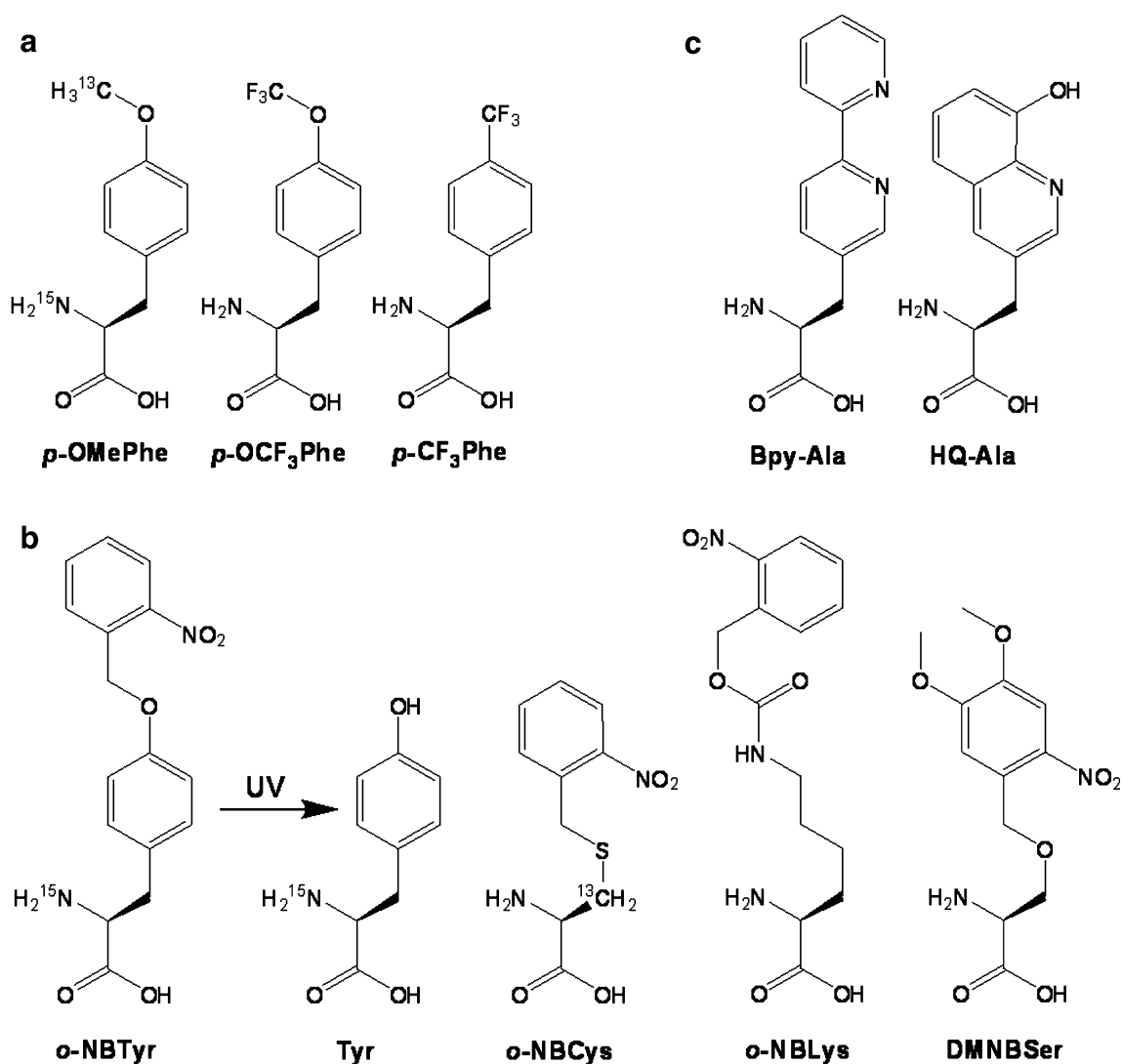


Fig. 1 NMR-active unnatural amino acids. **a.** Isotopically labeled *p*-methoxy-phenylalanine (*p*-OMePhe) (Wang et al. 2001), its trifluorinated analog *p*-OCF₃Phe (Cellitti et al. 2008) and *p*-trifluoromethyl-phenylalanine (CF₃Phe) (Jackson et al. 2007). **b.** Photocaged unnatural amino acids for the site-specific labeling of tyrosine (*o*-NBTyr) (Deiters et al. 2006), cysteine (*o*-NBCys) (Wu et al. 2004), lysine (*o*-NBLys) (Chen et al. 2009) and serine (DMNBSer) (Lemke

et al. 2007). The *o*-nitrobenzyl (*o*-NB) photocaging group is readily removed by UV illumination while the 4,5-dimethoxy-2-nitrobenzyl (DMNB) moiety can be removed by blue visible light. **c.** Metal chelating unnatural amino acids (2,2'-bipyridin-5-yl)-alanine (Bpy-Ala) (Xie et al. 2007b; Lee and Schultz 2008) and 2-amino-3-(8-hydroxyquinolin-3-yl)-propanoic acid (HQ-Ala) (Lee et al. 2009)

compared to that of the wild-type protein (Cellitti et al. 2008). Labeling with two different unnatural amino acids is in principle possible using a TAG codon for the first, and a frame-shifted codon for a second unnatural amino acid. Such an experiment has however not been performed to date.

¹⁵N/¹³C-labeled unnatural amino acids

The ability of NMR active unnatural amino acids to instantaneously provide signal assignment is extremely useful for the study of ligand binding to large proteins. For

example, ¹³C-labeled *p*-OMePhe (Fig. 1a) has recently been used to study local ligand-induced conformational heterogeneity in cytochrome P450 CYP119 (Lampe et al. 2008). We have incorporated ¹⁵N/¹³C-labeled and fluorinated unnatural amino acids at 11 sites in FAS-TE and monitored the binding of a tool compound using chemical shift perturbation and changes in conformational exchange behavior of certain residues (Cellitti et al. 2008). However, one caveat with only using chemical shift perturbations is that they can result from direct contacts with the compound or from ligand-induced conformational changes. Developing a clear understanding of the binding mode generally requires detailed atomic information regarding

the complex. Isotope filtered experiments, where only one of the binding partners is labeled, can provide intermolecular NOE correlations that can subsequently be used to determine the complex structure (reviewed in (Breeze 2000)). For complexes with an existing structure or known binding mode it is unnecessary to attempt a complete de novo structure determination. Instead, it is sufficient to use a small number of key intermolecular distances to guide docking simulations (Fesik et al. 1992; Morgan et al. 1995). Unnatural amino acids provide the ability to selectively incorporate an isotope label into a protein near a known binding site and measure these distances using NOE restraints.

To validate this approach, fully deuterated FAS-TE mutants were produced with a single $^{15}\text{N}/^{13}\text{C}$ -labeled *p*-OMePhe residue incorporated (Fig. 2). Samples were deuterated in order to remove protein signals and to minimize spin-diffusion. Deuteration allowed the use of longer mix times and avoided relayed NOE artifacts. As expected, the yield of deuterated protein was approximately 25% of the level obtained in rich media. 2D ^1H - ^{13}C HSQC-NOESY spectra (Fesik and Zuiderweg 1988) were recorded for selected mutants before and after the addition of a tool compound (Fig. 2).

In the absence of the compound the spectra show three correlations corresponding to the ^1H - ^{13}C methoxy group and intra-residue NOE peaks to the δ and ϵ side-chain protons. The magnitude of the chemical shift change for the single methoxy ^1H - ^{13}C cross peak observed upon compound addition was identical to previous studies (Cellitti et al. 2008). For the mutant Tyr-2343-*p*-OMePhe measurements were only made after tool compound addition because no peak was observed in the apo-protein HSQC. Presumably this region of the protein is undergoing conformational exchange in the absence of compound. Additional correlations (see arrows in Fig. 2c) for the Leu-2222, Tyr-2343, and Tyr-2347-*p*-OMePhe mutants appear upon compound binding and represent inter-molecular NOE peaks from the unnatural amino acid to the tool compound. The spectrum for Tyr-2343-*p*-OMePhe protein shows strong NOE peaks to aliphatic and aromatic resonances, whereas Leu-2222-*p*-OMePhe has only aromatic peaks, and Tyr-2347-*p*-OMePhe shows only weak aliphatic correlations. The weak aromatic correlations in Tyr-2347-*p*-OMePhe reflect the presence of a small amount of apo-FAS-TE due to the limited solubility of the compound and has been observed in previous studies (Cellitti et al. 2008). Comparing the intensity of the NOE peaks indicates that the aliphatic chain of the tool compound is closest to Tyr-2343-*p*-OMePhe. The chemical shifts of the NOE peaks to the compound's aromatic protons are very different for Tyr-2343-*p*-OMePhe and Leu-2222-*p*-OMePhe, suggesting that the two aromatic rings of the compound (the difluorobenzene and benzodioxole moieties)

form distinct contacts with the two residues. The mutant Phe-2375-*p*-OMePhe only shows intra-residue correlations, indicating that the compound is located beyond the $\sim 6 \text{ \AA}$ upper distance limit for NOE measurement. Although the resonances of the compound bound to FAS-TE have not yet been assigned, the data still provide qualitative information regarding the binding mode. More importantly, this example demonstrates how isotopically labeled unnatural amino acids in combination with deuterated protein can in principle be used to characterize ligand-protein complexes without obtaining protein signal assignments. Deuteration not only removes protein resonances not relevant for this technique but allows it to be extended to even larger proteins and complexes.

Fluorinated unnatural amino acids

Fluorine represents an attractive nucleus with high intrinsic sensitivity, 100% natural abundance of the NMR active isotope, and the absence of any natural background (Gerig 1994). Incorporation of 5-Fluoro-tryptophan, 3-fluoro-tyrosine, 4-fluoro-phenylalanine, and other fluorinated amino acids in place of their natural amino acid counterparts is an established protein labeling technique for ^{19}F -NMR studies (Gerig 1994; Danielson and Falke 1996; Gakh et al. 2000; Frieden et al. 2004). Site-specific labeling is not possible and signal assignment requires production of single-site mutant proteins. Two different fluorinated unnatural amino acids (Fig. 1a) have recently been introduced site-specifically into proteins and have been used to monitor ligand binding and structural transitions by ^{19}F -NMR (Hammill et al. 2007; Jackson et al. 2007; Cellitti et al. 2008). For *p*-OCF₃Phe, we observed significant variations in the width of the protein fluorine resonance line (Cellitti et al. 2008). These observations likely reflect the high sensitivity of fluorine resonances to conformational exchange and fluctuations in the proteins. Line-broadening by dynamic processes may therefore limit to some extent the application of fluorinated unnatural amino acids to protein NMR studies. Never-the-less, fluorinated unnatural amino acids are attractive probes for ligand binding and protein folding studies in vitro and in vivo (Jackson et al. 2007; Cellitti et al. 2008).

Photocaged unnatural amino acids

Just like any other mutation, an unnatural amino acid has the potential to modify the structure and function of the target protein. In fact, in FAS-TE, substituting Tyr-2307 (adjacent to the active site residue Ser-2308) with *p*-OCF₃Phe or *p*-OMePhe inhibited tool compound binding

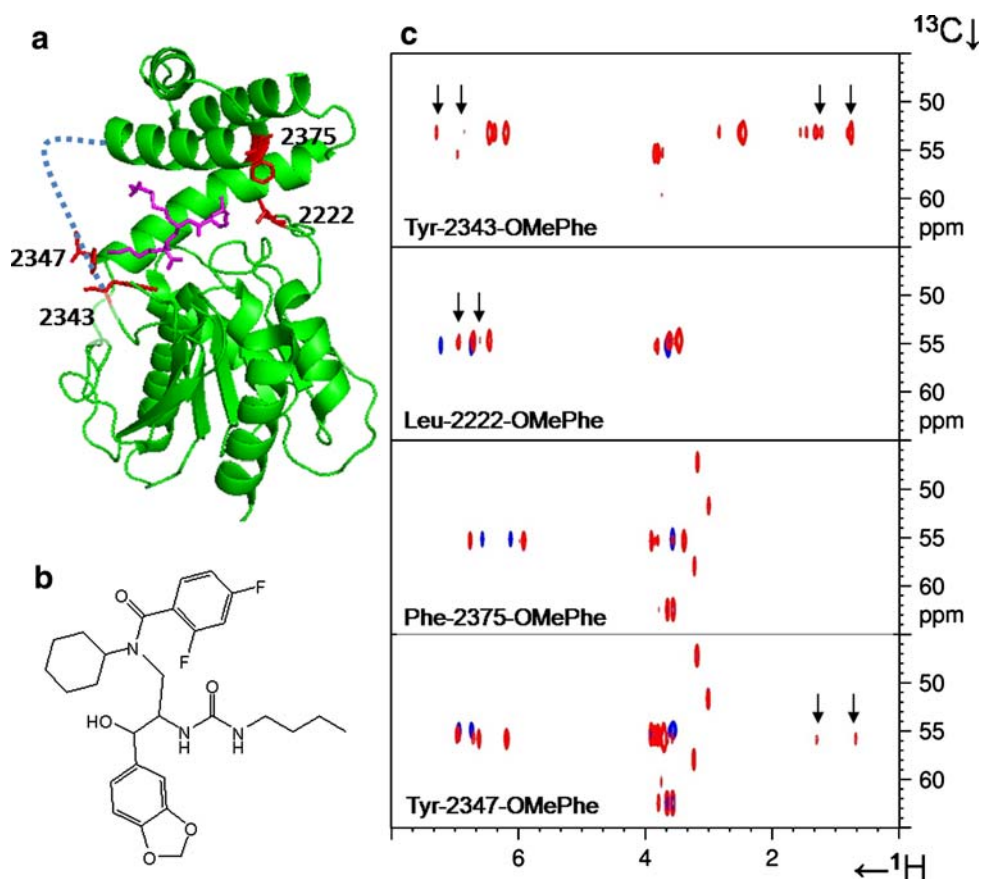


Fig. 2 2D ^1H - ^{13}C HSQC-NOESY experiments using deuterated FAS-TE with $^{15}\text{N}/^{13}\text{C}$ -labeled *p*-OMePhe incorporated at selected positions. **a**. Structure of the orlistat-FAS-TE complex (2PX6.pdb) (Pemble et al. 2007). Missing loops are indicated with a *dashed line* and orlistat is shown in *magenta lines*. The sidechain positions of the different single mutants are shown in *red*. **b**. Tool compound used in binding studies. **c**. 2D HSQC-NOESY spectra (Fesik and Zuiderweg 1988) for the indicated mutants. The spectra for each mutant before (*blue*) and after compound addition (*red*) are overlaid. The only exception is Tyr-2343 which did not have an HSQC peak in the

absence of compound. The intense aromatic peaks correspond to intramolecular peaks between the ^{13}C methoxy group and the sidechain. *Arrows mark* intermolecular correlations from *p*-OMePhe to the tool compound. The additional cross peaks were also found in HSQC spectra (see Supplemental Fig. 1) and correspond to natural abundance ^{13}C peaks from contaminants (glycerol, tris, TCEP) in the buffer. In the case of the Tyr-2347 mutant protein, the limited solubility of the tool compound resulted in the presence of some residual apo protein

(Cellitti et al. 2008). The bulky CF_3 or CH_3 group apparently blocked binding based on the lack of chemical shift perturbations in the protein upon compound addition. To confirm this, the study was repeated *o*-nitrobenzyl-tyrosine (*o*-NBTyr, Fig. 1b) incorporated at the same site: The photocaged form of *o*-NBTyr also prevented binding of the tool compound binding. UV cleavage of the photocage regenerated the natural tyrosine residue and re-established compound binding (Cellitti et al. 2008).

This example illustrates the unique ability of photocaged unnatural amino acids to introduce a single NMR label without perturbing the protein sequence (Cellitti et al. 2008). The photocage provides sufficient dissimilarity to natural side chains to enable the selection of a specific orthogonal tRNA/aminoacyl-tRNA synthetase and allows regeneration of the natural amino acid side chain via light-

induced decaging. Similarly designed photocaged unnatural amino acids could be developed for the site-specific NMR labeling of serine, cysteine, lysine, threonine, aspartic acid, asparagine, glutamic acid and glutamine. Mutating the naturally occurring orthogonal tRNA/RS pair that encodes pyrrolysine at TAG codons in *Methanosarcinae* (Hao et al. 2002; Srinivasan et al. 2002) is a particularly attractive approach toward developing such an array of photocaged unnatural amino acids because the resulting synthetases will enable incorporation in *E. coli* and in mammalian cells (Chen et al. 2009). In a first application, the system was used to evolve an aminoacyl-tRNA synthetase for photocaged lysine, *o*-nitrobenzyl-oxy-carbonyl- N^ϵ -lysine (*o*-NBLys, Fig. 1b) (Chen et al. 2009). The photocaged amino acids *o*-nitrobenzyl-cysteine (*o*-NBCys) (Wu et al. 2004) and 4,5-dimethoxy-2-

nitrobenzyl-serine (DMNBSer) (Lemke et al. 2007) (Fig. 1b) are currently available for incorporation in yeast and mammalian cells.

To explore its utility for functional studies by NMR, we incorporated ^{13}C -labeled *o*-NBCys into human superoxide dismutase (hSOD). The hSOD dimer catalyzes the conversion of superoxide radical to hydrogen peroxide and oxygen (Fridovich 1986) and requires one copper and one zinc cation per monomer for activity. Cys57 was replaced with *o*-NBCys because this residue has been implicated in copper binding (Banci et al. 2006). After the addition of excess ZnCl_2 , CuSO_4 , and isoascorbate, the HSQC spectra of the photo-cleaved mutant and wild-type protein are indistinguishable indicating that the mutant has assumed wild-type structure (Supplemental Fig. 2). The ^{13}C -labeled methylene group of Cys57 can readily be observed (Fig. 3). The observations are complicated because the dimer-monomer equilibrium of hSOD is influenced by metal binding and oxidation states (Arnesano et al. 2004). Based on comparison with published spectra (Assfalg et al. 2003), the data suggest that wild-type and photocaged hSOD are each partially complexed with metal after purification. To date we have been unable to completely remove the metal in order to accurately study the role of Cys57 in the copper binding process.

Photocaged amino acids not only provide a mechanism for site-specifically incorporating an NMR probe without changing the protein sequence but they can function as a “switch” to modulate protein activity. For example, it has been shown that a pro-apoptotic cysteine protease caspase-3 was inactivated by the incorporation of *o*-NBCys at the active site and regained 40% of its activity after photocleavage (Wu et al. 2004). Similarly, photocaged serine (DMNBSer, Fig. 1b) has recently been used in cell culture experiments to spatially and temporally control phosphorylation of a transcription factor (Lemke et al. 2007). For many applications, the DMNB group represents a more attractive photocaging group than the nitrobenzyl group because it can be efficiently cleaved using blue light in traditional microscopes. Time-resolved NMR experiments to study ligand induced conformational changes or enzymatic reactions may develop into a particularly interesting application of photocaged unnatural amino acids in the future.

Spin-labeled and metal chelating unnatural amino acids

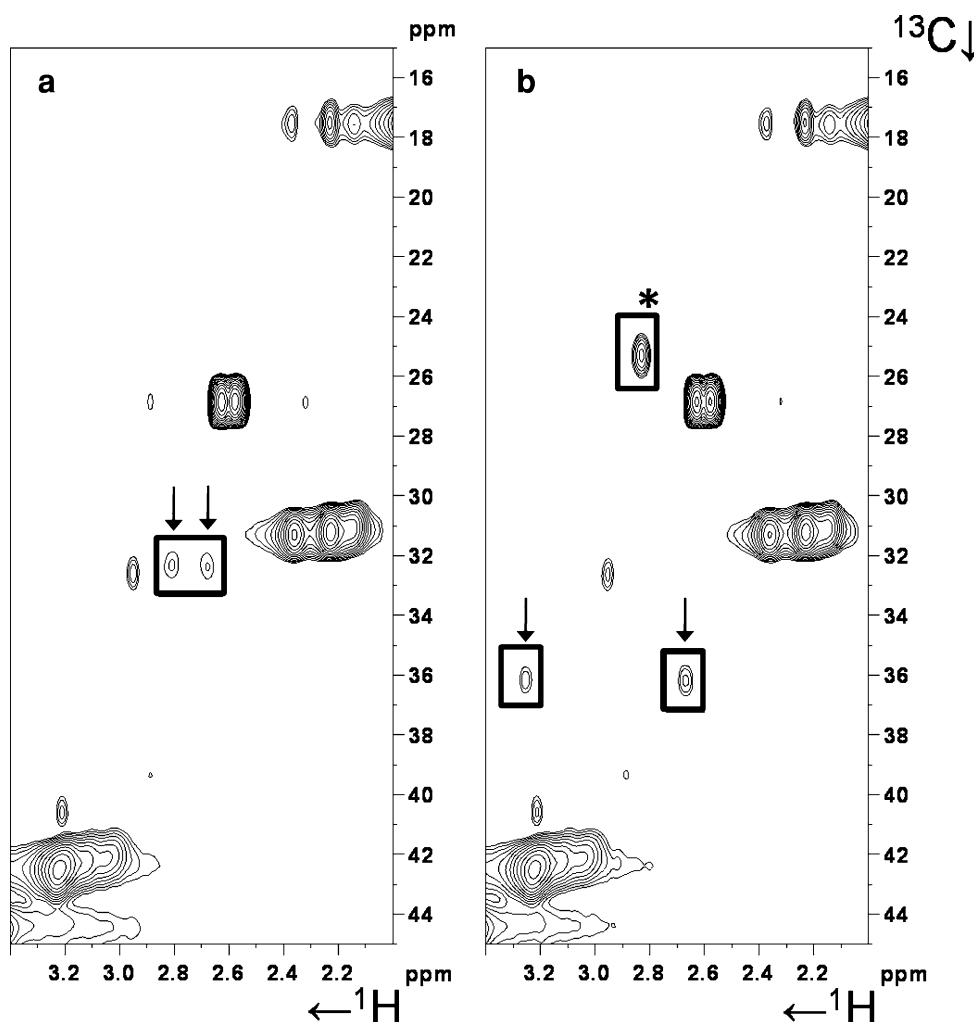
Long-range distance restraints can significantly improve the precision and accuracy of NMR structures. They are particularly important for systems with limited NOE's-such as RNA, and membrane proteins (Clare et al. 1999, 2007; Liang et al. 2006). Distance restraints out to $\sim 35 \text{ \AA}$ can

be obtained from paramagnetic relaxation enhancement (PRE) that arises from dipolar interactions between a nucleus and the unpaired electron of a spin-labeled group (Bloembergen and Morgan 1961; Kosen 1989) or a paramagnetic metal ion (Otting 2008). Site-directed spin labeling of isotopically labeled protein can be used to obtain resonance assignments (Pintacuda et al. 2004), to determine the global fold and structure of proteins (Battiste and Wagner 2000; Gaponenko et al. 2000; Liang et al. 2006), as well as for ligand screening (Jahnke et al. 2001). PRE measurements have the potential to characterize the binding mode of weakly bound ligands (Constantine 2001). In addition to PRE, paramagnetic metals can induce pseudocontact shifts (PCS's) (Allegrozzi et al. 2000; Gaponenko et al. 2004) and weak alignment resulting in residual dipolar couplings (RDC's) (Tolman et al. 1995). PCS's can provide accurate distance restraints out to 40 \AA from the metal ion (Allegrozzi et al. 2000; Gaponenko et al. 2004) and RDC's provide relative angular information for the entire molecule (Tolman et al. 1995). Lanthanides are particularly suited for these studies because their magnetic properties differ, making it possible to modulate the effects by incorporating different ions into the same metal chelating site.

A number of metal chelating tags have been developed and typically involve creating a fusion protein with a terminal metal binding motif or covalently attaching a synthetic chelating agent/peptide via a disulfide bond (Keizers et al. 2007; Otting 2008). Similarly, spin labeling is traditionally facilitated by chemical coupling to a reactive amino acid such as cysteine. In order to obtain site-specific incorporation, a new cysteine must be introduced and all solvent accessible cysteines must be mutated. The important role cysteines play in the folding, stability, and structure makes this process potentially difficult. In vivo incorporation of an unnatural amino acid featuring a spin-labeled or a metal chelating group would minimize or avoid these problems and extend the range of NMR (and EPR) applications.

Ideally, the spin label or metal ion should be rigidly positioned on the molecule because relative mobility associated with these tags can represent a significant impediment to PRE, PCS and RDC measurement (Su et al. 2008). Metal binding sites must also be enantiomerically pure to prevent peak doubling from the formation of diastereomeric complexes (Ikegami et al. 2004). Metal chelators should also be suitable for incorporation at different sites within the protein to help overcome the degeneracy inherent to PCS and RDC measurements. Unnatural amino acids, in general, can readily be introduced at different sites and offer the potential to introduce a paramagnetic metal or spin label at several rigid and well defined positions.

Fig. 3 Incorporation of ^{13}C -labeled *o*-NBCys into hSOD. ^1H - ^{13}C HSQC spectra before **a** and after **b** photo-cleavage (20 mM sodium phosphate buffer, 20 mM DTT, pH 5.0, 300 K). Resonances of the labeled CH_2 group are highlighted with *solid boxes* and *arrows*. The resonance at 25 ppm (*star* and *extra box*) in **b**, is a buffer component introduced in the process of re-purification after cleavage. Additional peaks are natural abundance peaks of DDT and other buffer components



So far two metal chelating unnatural amino acids have been developed (Fig. 1c). The Cu^{2+} -loaded form of a bipyridyl derivative of alanine (Bpy-Ala) has been incorporated into a DNA binding protein and resulted in site-specific DNA cleavage (Lee and Schultz 2008). Zn-coordination to a hydroxyquinoline derivative of alanine (HQ-Ala) has been observed in a recent crystal structure (Lee et al. 2009) but NMR measurements have been complicated by metal mediated protein oligomerization (unpublished results). Thus far several attempts to evolve an aminoacyl-tRNA synthetase for a spin-labeled unnatural amino acid have proven unsuccessful (unpublished results). New aminoacyl-tRNA synthetase libraries are being designed to accommodate the bulky side chains typical of tetramethyl-piperidinyloxyl or tetramethyl-pyrroline-oxyl spin labeling group (Langen et al. 2000; Columbus and Hubbell 2002), hopefully resulting in a spin-labeled unnatural amino acid in the near future. However, site-specific spin-labeling has been accomplished via the selective modification of the unnatural amino acid *p*-acetyl-

phenylalanine with alkoxyamine-derivatized spin labels (Schultz and Hubbell et al. unpublished results).

Unnatural amino acids and in-cell NMR measurements

Intracellular proteins typically function in an extremely viscous environment packed with other biological molecules. Consequently, interactions with other cellular proteins, intracellular structures, countless low molecular weight compounds as well as the general physical properties of the cytosol could affect the structure and properties of proteins. NMR is uniquely capable of providing atomic level structural information of proteins inside of living cells because it is non-destructive (Serber et al. 2005; Selenko and Wagner 2006). In-cell NMR experiments have been used to study protein-protein interactions (Burz et al. 2006b, a), protein dynamics (Dedmon et al. 2002; McNulty et al. 2006), protein structure (Sakakibara et al. 2009), and compound binding (Hubbard et al. 2003;

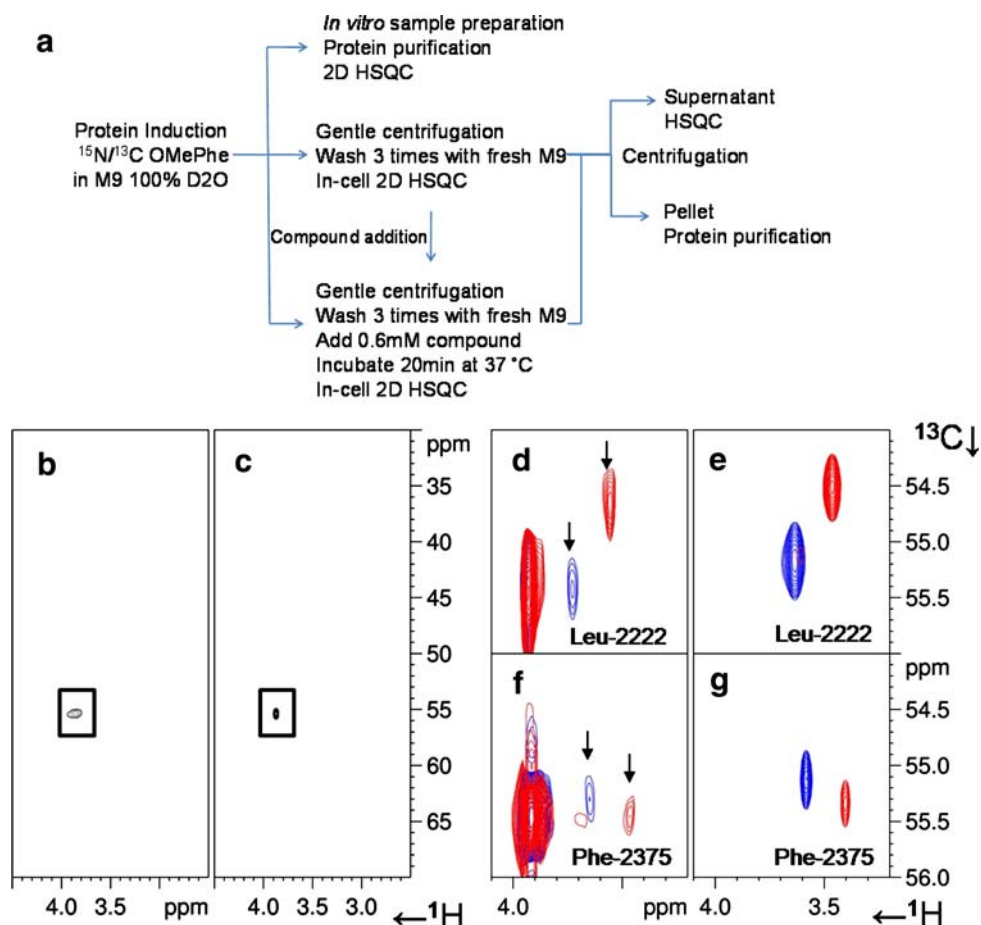


Fig. 4 In-cell NMR experiments using deuterated FAS-TE with $^{15}\text{N}/^{13}\text{C}$ -labeled *p*-OMePhe incorporated at selected positions. **a** In-cell sample preparation scheme illustrating the protocols used for preparing deuterated samples, adding compound, and controls. The latter were used to confirm protein expression and intracellular binding. **b** In-cell ^1H - ^{13}C HSQC spectrum of Tyr-2347-*p*-OMePhe FAS-TE. **c** ^1H - ^{13}C HSQC spectrum of the supernatant after centrifugation to remove the bacteria from the sample used in **b**. The signal in both samples (boxes) is detected at the same chemical shift and corresponds to extracellular, “free” *p*-OMePhe amino acid found in the media; an in-cell peak for the labeled protein is not detectable. The difference in linewidth for the two spectra probably reflects the higher viscosity and the difficulty of shimming with the cell suspension sample in (**b**). **d** In-cell ^1H - ^{13}C HSQC spectrum of

Leu-2222-*p*-OMePhe protein. **e** ^1H - ^{13}C HSQC spectra of purified Leu-2222-*p*-OMePhe protein. **f** In-cell ^1H - ^{13}C HSQC spectrum of Phe-2375-*p*-OMePhe protein. **g** ^1H - ^{13}C HSQC spectra of purified Phe-2375-*p*-OMePhe protein. The apo spectra are shown in blue, and shown in red are the spectra obtained in the presence of tool compound. The spectra in **d** and **f** show the appearance of a new peak (indicated with an arrow) that corresponds to the intracellular methoxy resonances of *p*-OMePhe incorporated into FAS-TE proteins. This peak shifts upon compound addition. The magnitude of the chemical shift change upon compound binding was very similar to in vitro samples prepared using purified protein (panel **e** and **g**, respectively). Peaks corresponding to *p*-OMePhe-labeled protein were not detected in the supernatant after cell removal by centrifugation (data not shown)

Serber et al. 2004). The latter studies, when performed in the appropriate cell type, can verify that a compound is cell permeable and binds directly to the target protein, ruling out off-target effects that can often not be distinguished by cellular screening assays.

In-cell NMR experiments generally require specific labeling to distinguish the protein of interest from the native molecules and to overcome signal overlap associated with the greater linewidths (Serber et al. 2005). This is particularly important for methyl groups which are close to ligands (Stockman and Dalvit 2002) and have excellent sensitivity but also very high background signals

(Wieruszski et al. 2001; Serber et al. 2004). Amino acid specific and uniform labeling is relatively easy to achieve in prokaryotic systems using appropriate media and a strong promoter/polymerase expression system to increase the concentration of the desired protein relative to that of cellular proteins (Serber et al. 2005; Selenko and Wagner 2006). However, studies employing other cell types typically involve obtaining labeled protein by expression in *E. coli* and delivering the protein inside the cell either by injection (Selenko et al. 2006; Serber et al. 2006) or using cell penetrating peptides (Inomata et al. 2009). Unnatural amino acids are uniquely suited for in-cell NMR

applications because of the ability to selectively label a protein without any signals from native proteins in a variety of different cell types.

The feasibility of acquiring in-cell protein NMR spectra was tested using deuterated FAS-TE mutants containing $^{15}\text{N}/^{13}\text{C}$ -labeled *p*-OMePhe at single sites and expressed in *E. coli*. Similar to other protocols (Sakakibara et al. 2009), intact *E. coli* cells were used for in-cell tool compound binding experiments as outlined in Fig. 4a. ^1H - ^{13}C HSQC peak signals that corresponded to *p*-OMePhe labeled protein inside the cells could be detected for only two of the four FAS-TE mutants tested (Table 1). For the Tyr-2347-*p*-OMePhe mutant (Fig. 4b, c) and for the Tyr-2343-*p*-OMePhe mutant (data not shown), the only detectable signals in the presence or absence of the tool compound correspond to those of $^{15}\text{N}/^{13}\text{C}$ -labeled *p*-OMePhe amino acid in the extracellular space. This conclusion is based on a comparison of the ^1H - ^{13}C HSQC spectra of the cell suspension of the Tyr-2347-*p*-OMePhe mutant (Fig. 4b) and that of the supernatant after the cells were removed by centrifugation (Fig. 4c). Except for differences in line-width, presumably because of the higher viscosity and the difficulty of shimming for the cell suspension sample, the two spectra are identical and the peaks at 3.87 and 55.35 ppm (solid boxes) correspond to “free” *p*-OMePhe amino acid in both samples as verified by a spectrum of the pre-induction media (data not shown).

Previous studies have noted the presence of strong peaks associated with free labeled amino acids in in-cell NMR experiments (Hubbard et al. 2003; Serber et al. 2004). However, for FAS-TE mutants Leu-2222-*p*-OMePhe and Phe-2375-*p*-OMePhe an additional methoxy cross peak is detected (arrows in Fig. 4d, f, respectively). These cross peaks shifted in the presence of the compound (red spectra) and therefore correspond to the ^{13}C -labeled methoxy group of *p*-OMePhe incorporated into FAS-TE at these two positions. The chemical shift changes upon compound binding are very similar in magnitude for in-cell and

in vitro samples prepared separately (Fig. 4e, g). The peaks detected in the in-cell experiments are associated with intracellular proteins because the spectra of the supernatant acquired after cell removal are identical to Fig. 4c and only show the resonance of the free labeled amino acid. The latter observation also verifies that the bacteria did not undergo lysis to any significant extent during the experiment.

Because two of the mutants failed to show evidence of an in-cell protein peak, protein expression was confirmed by purification of the FAS-TE proteins from *E. coli* cells recovered from the in-cell NMR samples (Fig. 4a and Table 1). Interestingly, samples with the highest expression levels and therefore highest cytoplasmic concentrations did not show a protein HSQC peak (note that Tyr-2347-*p*-OMePhe did not yield an HSQC peak at two different concentrations). In vitro samples demonstrated a tendency to precipitate at high concentrations and the in-cell results may be a property of FAS-TE aggregating at higher concentration or simply a function of the high cytoplasmic protein concentration causing increased viscosity and line broadening. While previous studies have examined the minimum levels of expression required for in-cell NMR studies (Serber et al. 2001), these findings suggest that in-cell samples are similar to in vitro samples in the sense that total concentration must fall within a certain range for optimal sensitivity. Consequently, for future in-cell studies, we will attempt to vary the cytoplasmic concentration of the protein of interest by altering the expression times and/or the promoter.

Most importantly, these preliminary in-cell NMR studies recapitulated the results of our previous in vitro binding studies and illustrate the use of unnatural amino acid labeling to detect compound binding to a protein inside of intact cells. The combination of labeled unnatural amino acid and deuteration resulted in spectra that were virtually free from background signals, making it trivial to identify the in-cell protein peak. To our knowledge, the 33 kDa

Table 1 Amounts of FAS-TE *p*-OMePhe mutant protein and corresponding cytoplasmic concentration recovered from in-cell NMR samples

<i>p</i> -OMe-Phe FAS-TE mutant	In-cell HSQC peak	Protein (mg) ^a	OD ₆₀₀ ^b	Culture ^c (ml)	Cytoplasmic concentration ^d (uM)
Leu-2222	Yes	0.024	1.38	50	29
Phe-2375	Yes	0.035	1.00	22	69
Tyr-2347	No	0.032	0.76	20	108
Tyr-2343	No	0.161	1.34	40	128
Tyr-2347	No	0.378	1.10	25	537

^a Total amount of protein purified after acquiring NMR data

^b Optical density (OD) at 600 nm of culture when harvested

^c Total amount of culture used in^a

^d Concentration calculated assuming that an OD₆₀₀ of 1 corresponds to 1×10^9 cells/ml (Sambrook and Russell 2001) and a volume of 7×10^{-16} l/bacteria (CyberCell Database)

FAS-TE represents the largest protein used for in-cell binding studies.

Concluding remarks

In vivo incorporation of NMR-active unnatural amino acids is readily accomplishable, and detailed protocols and materials are available (through P.G.S; Schultz@scripps.edu). A number of isotopically labeled unnatural amino acids have been tested, and their utility for characterizing the binding mode of a small molecule ligand to a 33 kDa protein of pharmaceutical interest has been illustrated. Photocaged unnatural amino acids facilitate “traceless” labeling of proteins at tyrosine, cysteine, serine and lysine residues and provide a powerful method to control protein function. In combination with deuteration, unnatural amino acids can likely be used to study larger system without the need for assigning protein resonances. Last but not least, we use unnatural amino acids to detect compound binding to a target protein inside of intact cells. The above discussions and these examples hopefully illustrate the potential role unnatural amino acids can play in expanding NMR structural biology.

Acknowledgments We thank Huiyong Hu for the synthesis of ^{15}N -labeled *o*-NBTyr, and Hyun Soo Lee for samples of HQ-Ala.

References

- Allegrozzi M, Bertini I, Janik MBL, Lee Y-M, Liu G et al (2000) Lanthanide-induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 Å; from the metal ion. *J Am Chem Soc* 122(17):4154–4161
- Amesano F, Banci L, Bertini I, Martinelli M, Furukawa Y et al (2004) The unusually stable quaternary structure of human Cu, Zn-superoxide dismutase I is controlled by both metal occupancy and disulfide status. *J Biol Chem* 279(46):47998–48003
- Assfalg M, Banci L, Bertini I, Turano P, Vasos PR (2003) Superoxide dismutase folding/unfolding pathway: role of the metal ions in modulating structural and dynamical features. *J Mol Biol* 330:145–158
- Banci L, Bertini I, Cantini F, D’Amelio N, Gaggelli E (2006) Human SOD1 before harboring the catalytic metal: solution structure of copper-depleted, disulfide-reduced form. *J Biol Chem* 281(4):2333–2337
- Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* 39(18):5355–5365
- Bloembergen N, Morgan LO (1961) Proton relaxation times in paramagnetic solutions effects of electron spin relaxation. *J Chem Phys* 34(3):842–850
- Breeze AL (2000) Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog NMR Spectrosc* 36(4):323–372
- Bromek K, Lee D, Hauhart R, Krych-Goldberg M, Atkinson JP et al (2005) Polychromatic selective population inversion for TROSY experiments with large proteins. *J Am Chem Soc* 127(1):405–411
- Burz DS, Dutta K, Cowburn D, Shekhtman A (2006a) Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). *Nat Methods* 3(2):91–93
- Burz DS, Dutta K, Cowburn D, Shekhtman A (2006b) In-cell NMR for protein-protein interactions (STINT-NMR). *Nat Protoc* 1(1):146–152
- Cellitti SE, Jones DH, Lagpacan L, Hao XS, Zhang Q et al (2008) In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J Am Chem Soc* 130(29):9268–9281
- Chen PR, Groff D, Guo J, Ou W, Cellitti S et al (2009) A facile system for encoding unnatural amino acids in mammalian cells. *Angew Chem Int Ed Engl* 48:4052–4055
- Clore GM, Starich MR, Bewley CA, Cai ML, Kuszewski J (1999) Impact of residual dipolar couplings on the accuracy of NMR structures determined from a minimal number of NOE restraints. *J Am Chem Soc* 121(27):6513–6514
- Clore GM, Tang C, Iwahara J (2007) Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Curr Opin Struct Biol* 17(5):603–616
- Columbus L, Hubbell WL (2002) A new spin on protein dynamics. *Trends Biochem Sci* 27(6):288–295
- Constantine KL (2001) Evaluation of site-directed spin labeling for characterizing protein-ligand complexes using simulated restraints. *Biophys J* 81(3):1275–1284
- Danielson MA, Falke JJ (1996) Use of F-19 NMR to probe protein structure and conformational changes. *Annu Rev Biophys Biomol Struct* 25:163–195
- Dedmon MM, Patel CN, Young GB, Pielak GJ (2002) FlgM gains structure in living cells. *Proc Natl Acad Sci USA* 99(20):12681–12684
- Deiters A, Geierstanger BH, Schultz PG (2005) Site-specific in vivo labeling of proteins for NMR studies. *ChemBioChem* 6(1):55–58
- Deiters A, Groff D, Ryu Y, Xie J, Schultz PG (2006) A genetically encoded photocaged tyrosine. *Angew Chem Int Ed Engl* 45:2728–2731
- Ellman J, Volkman BF, Mendel D, Schultz PG, Wemmer DE (1992) Site-specific isotopic labeling of proteins for NMR studies. *J Am Chem Soc* 114(20):7959–7961
- Fesik SW, Züderweg ERP (1988) Heteronuclear three-dimensional NMR spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra. *J Magn Reson* 78:588–593
- Fesik SW, Neri P, Meadows R, Olejniczak ET, Gemmecker G (1992) A model of the cyclophilin/cyclosporin A (CSA) complex from NMR and X-ray data suggests that CSA binds as a transition-state analog. *J Am Chem Soc* 114(8):3165–3166
- Fridovich I (1986) Superoxide dismutases. *Adv Enzymol Relat Areas Mol Biol* 58:61–97
- Frieden C, Hoeltzli SD, Bann JG (2004) The preparation of ^{19}F -labeled proteins for NMR studies. *Methods Enzymol* 380:400–415
- Gakh YG, Gakh AA, Gronenborn AM (2000) Fluorine as an NMR probe for structural studies of chemical and biological systems. *Magn Reson Chem* 38(7):551–558
- Gaponenko V, Howarth JW, Columbus L, Gasmi-Seabrook G, Yuan J et al (2000) Protein global fold determination using site-directed spin and isotope labeling. *Protein Sci* 9:302–309
- Gaponenko V, Sarma SP, Altieri AS, Horita DA, Li J et al (2004) Improving the accuracy of NMR structures of large proteins

- using pseudocontact shifts as long-range restraints. *J Biomol NMR* 28(3):205–212
- Gardner KH, Kay LE (1998) The use of ^2H , ^{13}C , ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* 27:357–406
- Gardner KH, Rosen MK, Kay LE (1997) Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry* 36(6):1389–1401
- Gerig JT (1994) Fluorine NMR of proteins. *Prog Nucl Magn Reson Spectroscopy* 26(4):293–370
- Goto NK, Kay LE (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr Opin Struct Biol* 10(5):585–592
- Groff D, Thielges MC, Cellitti S, Schultz PG, Romesberg FE (2009) Efforts toward the direct experimental characterization of enzyme microenvironments: tyrosine 100 in dihydrofolate reductase. *Angew Chem Int Ed Engl* 48:3478–3481
- Gross JD, Gelev VM, Wagner G (2003) A sensitive and robust method for obtaining intermolecular NOEs between side chains in large protein complexes. *J Biomol NMR* 25(3):235–242
- Guo J, Wang J, Lee JS, Schultz PG (2008) Site-specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. *Angewandte Chemie International Ed Engl* 47(34):6399–6401
- Hammill JT, Miyake-Stoner S, Hazen JL, Jackson JC, Mehl RA (2007) Preparation of site-specifically labeled fluorinated proteins for F-19-NMR structural characterization. *Nat Protoc* 2(10):2601–2607
- Hao B, Gong W, Ferguson TK, James CM, Krzycki JA et al (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase. [see comment]. *Science* 296(5572):1462–1466
- Hubbard JA, MacLachlan LK, King GW, Jones JJ, Fosberry AP (2003) Nuclear magnetic resonance spectroscopy reveals the functional state of the signalling protein CheY in vivo in *Escherichia coli*. *Mol Microbiol* 49(5):1191–1200
- Ikegami T, Verdier L, Sakahai P, Grimme S, Pescatore B et al (2004) Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions. *J Biomol NMR* 29(3):339–349
- Inomata K, Ohno A, Tochio H, Isogai S, Tenno T et al (2009) High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* 458(7234):106–109
- Jackson JC, Hammill JT, Mehl RA (2007) Site-specific incorporation of a (^{19}F) -amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J Am Chem Soc* 129(5):1160–1166
- Jahnke W, Rudisser S, Zurini M (2001) Spin label enhanced NMR screening. *J Am Chem Soc* 123(13):3149–3150
- Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono A et al (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature* 440(7080):52–57
- Keizers PH, Desreux JF, Overhand M, Ubbink M (2007) Increased paramagnetic effect of a lanthanide protein probe by two-point attachment. *J Am Chem Soc* 129(30):9292–9293
- Khaneja N, Li JS, Kehlet C, Luy B, Glaser SJ (2004) Broadband relaxation-optimized polarization transfer in magnetic resonance. *Proc Natl Acad Sci USA* 101(41):14742–14747
- Kosen PA (1989) Spin labeling of proteins. *Methods Enzymol* 177:86–121
- Lampe JN, Floor SN, Gross JD, Nishida CR, Jiang Y et al (2008) Ligand-induced conformational heterogeneity of cytochrome P450 CYP119 identified by 2D NMR spectroscopy with the unnatural amino acid (^{13}C) -p-methoxyphenylalanine. *J Am Chem Soc* 130(48):16168–16169
- Langen R, Oh KJ, Cascio D, Hubbell WL (2000) Crystal structures of spin labeled T4 lysozyme mutants: implications for the interpretation of EPR spectra in terms of structure. *Biochemistry* 39(29):8396–8405
- Lee HS, Schultz PG (2008) Biosynthesis of a site-specific DNA cleaving protein. *J Am Chem Soc* 130(40):13194–13195
- Lee HS, Spraggon G, Schultz PG, Wang F (2009) Genetic incorporation of a metal-ion chelating amino acid into proteins as a biophysical probe. *J Am Chem Soc* 131(7):2481–2483
- Lemke EA, Summerer D, Geierstanger BH, Brittain SM, Schultz PG (2007) Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat Chem Biol* 3(12):769–772
- Liang B, Bushweller JH, Tamm LK (2006) Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. *J Am Chem Soc* 128(13):4389–4397
- Liu CC, Schultz PG (2006) Recombinant expression of selectively sulfated proteins in *Escherichia coli*. *Nat Biotechnol* 24(11):1436–1440
- Liu W, Brock A, Chen S, Chen S, Schultz PG (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat Methods* 4(3):239–244
- Liu CC, Cellitti S, Geierstanger BH, Schultz PG (2009) Efficient expression of tyrosine-sulfated proteins in *E. coli* using an expanded code. (submitted)
- Luy B (2007) Approaching the megadalton: NMR spectroscopy of protein complexes. *Angew Chem Int Ed Engl* 46(23):4214–4216
- Markley JL, Putter I, Jardetzky O (1968) High-resolution nuclear magnetic resonance spectra of selectively deuterated staphylococcal nuclease. *Science* 161(3847):1249
- McNulty BC, Young GB, Pielak GJ (2006) Macromolecular crowding in the *Escherichia coli* periplasm maintains alpha-synuclein disorder. *J Mol Biol* 355(5):893–897
- Mittermaier A, Kay LE (2006) New tools provide new insights in NMR studies of protein dynamics. *Science* 312(5771):224–228
- Morgan WD, Birdsall B, Polshakov VI, Sali D, Kompis I et al (1995) Solution structure of a brodimoprim analogue in its complex with *Lactobacillus casei* dihydrofolate reductase. *Biochemistry* 34(37):11690–11702
- Muchmore DC, McIntosh LP, Russell CB, Anderson DE, Dahlquist FW (1989) Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods Enzymol* 177:44–73
- Neumann H, Peak-Chew SY, Chin JW (2008) Genetically encoding Nε-acetyllysine in recombinant proteins. *Nat Chem Biol* epublished.
- Otting G (2008) Prospects for lanthanides in structural biology by NMR. *J Biomol NMR* 42(1):1–9
- Pemble CW, Johnson LC, Kridel SJ, Lowther WT (2007) Crystal structure of the thioesterase domain of human fatty acid synthase inhibited by orlistat. *Nat Struct Mol Biol* 14(8):704–709
- Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* 94(23):12366–12371
- Pintacuda G, Keniry MA, Huber T, Park AY, Dixon NE et al (2004) Fast structure-based assignment of ^{15}N HSQC spectra of selectively ^{15}N -labeled paramagnetic proteins. *J Am Chem Soc* 126(9):2963–2970
- Riek R, Wider G, Pervushin K, Wüthrich K (1999) Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proc Natl Acad Sci USA* 96(9):4918–4923
- Sakakibara D, Sasaki A, Ikeya T, Hamatsu J, Hanashima T et al (2009) Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* 458(7234):102–105
- Sambrook J, Russell DW (2001) Molecular cloning, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA

- Schultz KC, Supekova L, Ryu YH, Xie JM, Perera R et al (2006) A genetically encoded infrared probe. *J Am Chem Soc* 128(43):13984–13985
- Selenko P, Wagner G (2006) NMR mapping of protein interactions in living cells. *Nat Methods* 3(2):80–81
- Selenko P, Serber Z, Gadea B, Ruderman J, Wagner G (2006) Quantitative NMR analysis of the protein G B1 domain in *Xenopus laevis* egg extracts and intact oocytes. *Proc Natl Acad Sci USA* 103(32):11904–11909
- Serber Z, Ledwidge R, Miller SM, Dotsch V (2001) Evaluation of parameters critical to observing proteins inside living *Escherichia coli* by in-cell NMR spectroscopy. *J Am Chem Soc* 123(37):8895–8901
- Serber Z, Straub W, Corsini L, Nomura AM, Shimba N et al (2004) Methyl groups as probes for proteins and complexes in in-cell NMR experiments. *J Am Chem Soc* 126(22):7119–7125
- Serber Z, Corsini L, Durst F, Dotsch V (2005) In-cell NMR spectroscopy. *Methods Enzymol* 394:17–41
- Serber Z, Selenko P, Hansel R, Reckel S, Lohr F et al (2006) Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nat Protoc* 1(6):2701–2709
- Srinivasan G, James CM, Krzycki JA (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* 296(5572):1459–1462
- Stockman BJ, Dalvit C (2002) NMR screening techniques in drug discovery and drug design. *Prog NMR Spectrosc* 41(3–4):187–231
- Su XC, Man B, Beeren S, Liang H, Simonsen S et al (2008) A dipicolinic acid tag for rigid lanthanide tagging of proteins and paramagnetic NMR spectroscopy. *J Am Chem Soc* 130(32):10486–10487
- Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH (1995) Nuclear magnetic dipole interactions in field-oriented proteins—information for structure determination in solution. *Proc Natl Acad Sci USA* 92(20):9279–9283
- Tsao ML, Summner D, Ryu YH, Schultz PG (2006) The genetic incorporation of a distance probe into proteins in *Escherichia coli*. *J Am Chem Soc* 128(14):4572–4573
- Tugarinov V, Kay LE (2005) Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. *ChemBioChem* 6(9):1567–1577
- Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced $^1\text{H}[^1\text{J}]\text{C}$ NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* 125(34):10420–10428
- Tugarinov V, Hwang PM, Kay LE (2004) Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annu Rev Biochem* 73:107–146
- Veldkamp CT, Seibert C, Peterson FC, Sakmar TP, Volkman BF (2006) Recognition of a CXCR4 sulfotyrosine by the chemokine stromal cell-derived factor-1alpha (SDF-1alpha/CXCL12). *J Mol Biol* 359(5):1400–1409
- Veldkamp CT, Seibert C, Peterson FC, De La Cruz NB, Haugner JCIII et al (2008) Structural basis of CXCR4 sulfotyrosine recognition by the chemokine SDF-1/CXCL12. *Sci Signal* 1(37):ra4
- Wang L, Brock A, Herberich B, Schultz PG (2001) Expanding the genetic code of *Escherichia coli*. *Science* 292(5516):498–500
- Wang L, Xie J, Deniz AA, Schultz PG (2003) Unnatural amino acid mutagenesis of green fluorescent protein. *J Org Chem* 68(1):174–176
- Wang JY, Xie JM, Schultz PG (2006a) A genetically encoded fluorescent amino acid. *J Am Chem Soc* 128(27):8738–8739
- Wang L, Xie J, Schultz PG (2006b) Expanding the genetic code. *Annu Rev Biophys Biomol Struct* 35:225–249
- Wieruszkeski JM, Bohin A, Bohin JP, Lippens G (2001) In vivo detection of the cyclic osmoregulated periplasmic glucan of *Ralstonia solanacearum* by high-resolution magic angle spinning NMR. *J Magn Reson* 151(1):118–123
- Wu N, Deiters A, Cropp TA, King D, Schultz PG (2004) A genetically encoded photocaged amino acid. *J Am Chem Soc* 126(44):14306–14307
- Wüthrich K (1998) The second decade—into the third millennium. *Nat Struct Biol* 5(Suppl S):492–495
- Xie J, Schultz PG (2005a) An expanding genetic code. *Methods* 36(3):227–238
- Xie J, Schultz PG (2005b) Adding amino acids to the genetic repertoire. *Curr Opin Chem Biol* 9(6):548–554
- Xie J, Schultz PG (2006) A chemical toolkit for proteins—an expanded genetic code. *Nat Rev Mol Cell Biol* 7(10):775–782
- Xie J, Wang L, Wu N, Brock A, Spraggon G et al (2004) The site-specific incorporation of *p*-iodo-*L*-phenylalanine into proteins for structure determination. *Nat Biotechnol* 22(10):1297–1301
- Xie JM, Supekova L, Schultz PG (2007a) A genetically encoded metabolically stable analogue of phosphotyrosine in *Escherichia coli*. *Acs Chem Biol* 2(7):474–478
- Xie JM, Liu WS, Schultz PG (2007b) A genetically encoded bidentate, metal-binding amino acid. *Angew Chem Int Ed Engl* 46(48):9239–9242
- Young TS, Ahmad I, Brock A, Schultz PG (2009) Expanding the genetic repertoire of the methylotrophic Yeast *Pichia pastoris*. *Biochemistry* 48(12):2643–2653