# Designer proteins: applications of genetic code expansion in cell biology

*Lloyd Davis and Jason W. Chin*

Abstract | Designer amino acids, beyond the canonical 20 that are normally used by cells, can now be site-specifically encoded into proteins in cells and organisms. This is achieved using 'orthogonal' aminoacyl-tRNA synthetase–tRNA pairs that direct amino acid incorporation in response to an amber stop codon (UAG) placed in a gene of interest. Using this approach, it is now possible to study biology *in vitro* and *in vivo* with an increased level of molecular precision. This has allowed new biological insights into protein conformational changes, protein interactions, elementary processes in signal transduction and the role of post-translational modifications.

Selective pressure incorporation
Bacteria that are auxotrophic for a natural amino acid are used in conjunction with a related unnatural amino acid, leading to the incorporation of the unnatural amino acid throughout the cell's proteome.

*Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK. Correspondence to J.W.C. e-mail: chin@mrc-lmb.cam.ac.uk*
doi:10.1038/nrm3286
Published online
15 February 2012

In the past few years, the ability to incorporate unnatural amino acids into proteins at defined sites has begun to have a direct impact on the ability of scientists to study biological processes that are difficult or impossible to address by more classical methods. One of the most powerful approaches for incorporating unnatural amino acids site-specifically into proteins expressed in cells is genetic code expansion. In this approach, an aminoacyl-tRNA synthetase and a tRNA are used to specifically insert the unnatural amino acid during mRNA translation, in response to an amber stop codon (UAG) placed at a user-defined site in a gene of interest (FIG. 1a). Another robust, and often complementary, approach for site-specifically incorporating unnatural amino acids into recombinant proteins is native chemical ligation (BOX 1). Unnatural amino acids can also be incorporated non-specifically throughout the proteome by an approach known as selective pressure incorporation[1]. However, this leads to a proteome bearing unnatural amino acids in the place of every occurrence of a particular natural amino acid, rather than the quantitative installation of an unnatural amino acid at a defined site in a protein that is achieved by genetic code expansion.

Methodological advances in genetic code expansion now allow the site-specific incorporation of unnatural amino acids into proteins in bacteria, yeast, mammalian cells and *Caenorhabditis elegans*[2–7]. These advances have been rapidly translated into new insights in diverse areas of biology. Encoded photocrosslinkers have allowed interactions that are weak, transient, or pH sensitive to be determined and assigned to distinct functional states *in vivo*, especially in membranes. The incorporation of post-translational modifications has allowed the synthesis of homogenously and site-specifically modified proteins in cases in which the modifying enzymes cannot be used or in which the modification has been identified by mass spectrometry but its regulators are unknown. In several cases, this has revealed the function of these modifications and/or the enzymes that regulate them. The ability to incorporate photocaged amino acids allows a specific function of a protein to be activated in cells with a pulse of light, forming the basis of 'photochemical genetic' strategies for time-resolved studies of signalling and transport processes. The ability to incorporate biophysical probes — for example, ones that show altered infrared spectra in response to a protein conformational change — is also providing exquisite time-resolved insight into how proteins respond to stimuli. These methods for incorporating unnatural amino acids have been combined with imaging, single-molecule studies, biophysical approaches, structural biology and mass spectrometry to provide a powerful new set of approaches for studying biology.

In this Review, we describe the basis of methods for site-specifically incorporating unnatural amino acids into proteins and we summarize the range of amino acids that can be incorporated and their diverse applications in cell biology. We highlight key examples in which the methods have been applied to address a particular biological question and have provided previously unattainable insight. For the distinct applications of other incorporation techniques, such as selective pressure incorporation, readers are referred to another review[1].



© 2012 Macmillan Publishers Limited. All rights reserved

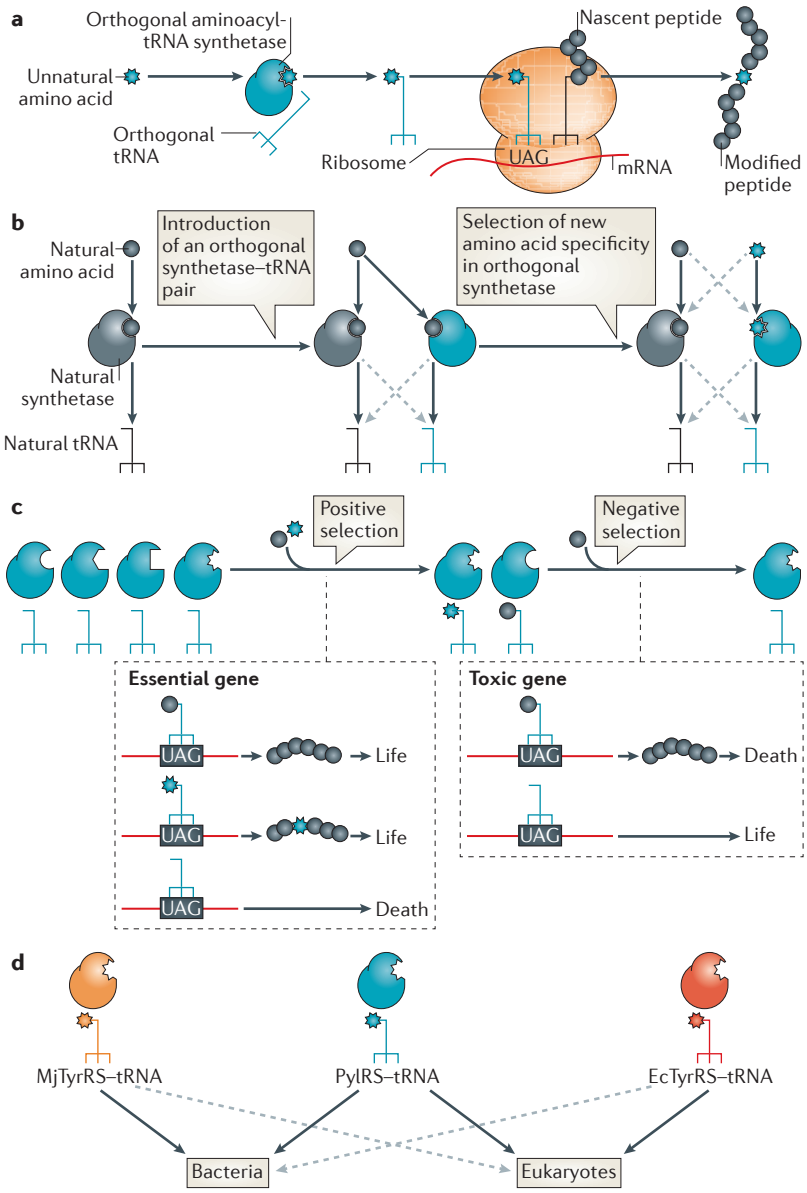



Figure 1 | **Expanding the genetic code. a** | To incorporate an unnatural amino acid into a protein, the amino acid is normally added to the cell growth medium and is then taken up by the cell. It is specifically recognized by an 'orthogonal' aminoacyl-tRNA synthetase and attached to the orthogonal amber suppressor tRNA, which is decoded on the ribosome during translation in response to an introduced amber codon (UAG), allowing its incorporation into the peptide that otherwise consists of natural amino acids. **b** | Orthogonal synthetase tRNA pairs are generated in two steps: import of a heterologous aminoacyl-tRNA synthetase–tRNA pair into a host containing a set of natural synthetases that use natural amino acids; and the subsequent selection of a mutated active site in the orthogonal synthetase that recognizes an unnatural amino acid. **c** | To generate a synthetase with this altered specificity, a large library of active-site variants of the synthetase is subject to positive selection for activity with either natural or unnatural amino acids, by virtue of their ability to suppress an introduced stop codon and so allow complete translation of a gene that is essential for survival. The synthetases that use natural amino acids are subsequently removed by a negative-selection step, in which they use natural amino acids to suppress a stop codon introduced in a toxic gene, which leads to cell death. **d** | The pyrrolysyl-tRNA synthetase (PylRS)–tRNA pair is orthogonal and can be used in both bacteria and eukaryotic cells. By contrast, *Methanococcus jannaschii* Tyrosyl-tRNA synthetase (*Mj*TyrRS)–tRNA is orthogonal only in bacteria and *Escherichia coli* Tyrosyl-tRNA synthetase (*Ec*TyrRS)–tRNA is orthogonal only in eukaryotic cells; the dashed arrows connect synthetase–tRNA pairs to organisms in which they cannot be used for genetic code expansion because they cross-react with endogenous synthetases or tRNAs.

## Expanding the genetic code

During translation, each codon in an mRNA is recognized by a specific tRNA anticodon that is aminoacylated with the appropriate amino acid by the enzymatic action of an aminoacyl tRNA synthetase. As the ribosome moves along the mRNA, a chain of amino acids is formed until a stop codon, which is not recognized by any tRNA, is encountered, allowing peptide release. Genetic code expansion uses an 'orthogonal' aminoacyl tRNA synthetase–tRNA pair (that is, a synthetase that does not aminoacylate the normal tRNA range of a cell and a tRNA that is not a substrate for the normal synthetase range of the cell) to direct the incorporation of an unnatural amino acid in response to an amber stop codon in a gene of interest[2,8]. The unnatural amino acid is generally added to the medium in which the cells are grown. As the set of endogenous tRNAs and synthetases are distinct in different cells and organisms, synthetases and tRNAs are defined as orthogonal (or not) with respect to particular hosts.

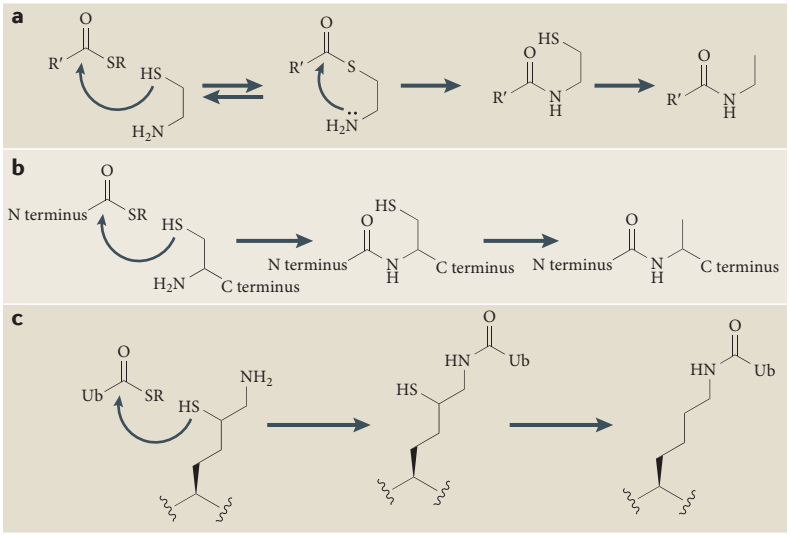
A two-step procedure is now well established by which orthogonal synthetase–tRNA pairs for incorporating unnatural amino acids can be isolated (FIG. 1b,c). In the first step, a synthetase–tRNA pair is selected from an organism that is evolutionarily divergent from the host of interest. When expressed in the host of interest, the orthogonal synthetase and tRNA should not cross-react with the host synthetases and tRNAs because their sequences and structures have diverged through natural evolutionary processes. They do, in general, still recognize a natural amino acid in the host cell; so, a second step is required to alter the specificity of the synthetase enzyme so that it exclusively recognizes an unnatural amino acid and not natural amino acids. This can be achieved by a two-step genetic selection in which large libraries of mutations in the synthetase active site are created and then variant synthetases that uniquely use the unnatural amino acid are selected[3,8,9].

Nearly all genetic code expansion approaches use one of four synthetase–tRNA pairs, each of which can be used in particular model organisms. The *Methanococcus jannaschii* Tyrosyl-tRNA synthetase (*Mj*TyrRS)–$tRNA_{CUA}$ pair was the first pair to be used to incorporate unnatural amino acids, and it is orthogonal to the synthetases and tRNAs in *Escherichia coli* but not to those in eukaryotic cells[8]. The *E. coli* Tyrosyl-tRNA synthetase (*Ec*TyrRS)–$tRNA_{CUA}$ and *E. coli* Leucyl-tRNA synthetase (*Ec*LeuRS)–$tRNA_{CUA}$ pairs are orthogonal to the synthetases and tRNAs in yeast and mammalian cells but not to those in bacteria[3,10]. And the pyrrolysyl-tRNA synthetase (PylRS)–$tRNA_{CUA}$ pair from *Methanosarcina* species is orthogonal to synthetases and tRNAs in *E. coli*, yeast, mammalian cells and *C. elegans*[4–6,11] (FIG. 1d). The PylRS–$tRNA_{CUA}$ pair has two advantages over the other pairs. First, this synthetase does not use one of the 20 canonical amino acids and there is, therefore, no need to destroy the natural synthetase activity before creating specificity for a new amino acid. Indeed, several useful unnatural amino acids can now be specifically incorporated using the unmodified PylRS enzyme. Second, whereas the use of other synthetase–tRNA pairs for



© 2012 Macmillan Publishers Limited. All rights reserved

Box 1 | **Chemical ligation**




Chemical ligation approaches can be used to assemble modified proteins from unprotected polypeptide fragments, and can also be used to generate ubiquitylated proteins and peptides. These approaches are complementary to genetic code expansion for the synthesis of recombinant proteins bearing site-specific modification. Reviews of this method have been published elsewhere[141,142].

Native chemical ligation uses *S*-to-*N* acyl shift chemistry for peptide bond formation (see the figure, part **a**). This occurs through the reaction of a thioester with a 1,2-amino thiol. The initial product that is formed can subsequently be desulphurized. A protein can also be assembled by native chemical ligation of two fragments: one containing a carboxy-terminal thioester (which may be generated by protein expression and intein fusion thiolysis) and another bearing an amino-terminal 1,2-amino thiol, as found on Cys (see the figure, part **b**). This allows the introduction of multiple modifications into the N-terminal portion via solid phase peptide synthesis (SPPS). The ligation of multiple fragments is possible. Following ligation, the Cys may be desulphurized to yield Ala. Using δ-thiol Lys (or other Lys derivatives) and ubiquitin (Ub) it is possible to generate site-specifically ubiquitylated proteins (see the figure, part **c**). This was first demonstrated on synthetic peptides[87,90–94], but it is now possible to also genetically encode δ-thiol-Lys derivatives[46], allowing the ubiquitylation of recombinant proteins that are larger than those accessible by SPPS.

incorporating unnatural amino acids is limited to cells from particular organisms, PylRS variants that recognize new amino acids can be selected in *E. coli*, in which selections are the most straightforward, and then used to incorporate unnatural amino acids in yeast, mammalian cells and *C. elegans*[4–6,11].

## Mapping protein interactions

A challenge in almost any biological process is to understand how proteins interact, and this can be particularly difficult when interactions are transient, weak or at particular subcellular locations. The site-specific incorporation of amino acids that crosslink with nearby molecules in response to light, including benzophenones (such as *p*-benzoyl-L-phenylalanine, (Bpa))[3,9], azido derivatives of Phe[3,12] and diazirines[13–16], provide powerful tools to gain information about the interactions of a specific protein with other molecules both *in vitro* and *in vivo*. In a typical experiment, the unnatural amino acid that contains a photocrosslinker is genetically installed at a single site in a protein and the protein is then irradiated with ultraviolet (UV) light in the presence of potential binding partners. Binding partners in proximity to the photocrosslinker become covalently trapped, allowing the subsequent identification of factors that interact with this protein.

Site-specific photocrosslinking has been used to gain information on protein interactions that are difficult to study by other methods, particularly for proteins in membranes[17–21] and weak, transient or pH-dependent interactions that may be systematically lost in non-covalent methods such as TAP tagging (tandem affinity purification tagging). Crosslinking using genetically encoded Bpa has been used to: define the interactions of chaperones, including trigger factor[22,23], ClpB[24,25] and GroEL[26], with substrates; identify protein interactions that are important in cell cycle regulation[27]; define conformational changes in RNA polymerases[28]; and determine the interactions and topology of large transcriptional initiation complexes *in vitro* and *in vivo*[29,30]. It has also been used to define protein interactions at the inner[20,21] and outer membrane[19] of *E. coli*, the mitochondrial[31] and endoplasmic reticulum (ER) membranes in yeast[32], and the plasma membrane in mammalian cells[33].

***In vivo* photocrosslinking of membrane proteins.** A study examining the interaction of the bacterial SecY–SecE–SecG (SecYEG) translocon with SecA *in vivo* exemplifies how new biological information about dynamic protein interactions of macromolecular complexes may be obtained using genetically encoded site-specific photocrosslinkers[20]. SecYEG (the eukaryotic homologue of which is SEC61α–SEC61β–SEC61γ) is a channel-like transmembrane complex that mediates both the co-translational integration of proteins at the *E. coli* inner membrane and the SecA- and SecB-dependent post-translational translocation of periplasmic and outer membrane proteins. Several aspects of SecA-mediated translocation have been unclear, including where the binding interface between SecA and SecYEG is and how binding of SecA to SecYEG changes during the SecA ATPase cycle to drive protein translocation. Introduction of the photocrosslinker Bpa at specific sites in SecY, a central component of SecYEG, has allowed these questions to be addressed (FIG. 2a).

To investigate the interface between SecY and SecA, 53 different amino acid positions across the six cytoplasmic domains (C1–C6) of SecY tagged with $His_6$–MYC were mutated to Bpa by creating the corresponding amber mutants in a *secY* gene. By irradiating the mutant cells to induce crosslinking and then purifying SecY–$His_6$–MYC complexes and western blotting with antibodies against SecA, it was possible to identify the sites of SecY crosslinking to SecA. This approach demonstrated that SecA interacts with C2, C4, C5 and C6 of SecY in SecYEG, which is consistent with a subsequent crystal structure[34] (FIG. 2b).

This approach also provided insights into the conformational changes that take place during SecA-mediated ATP hydrolysis and pre-protein translocation. In this case, crosslinking was carried out in the presence or absence of sodium azide, which arrests SecA ATPase activity and stabilizes SecA in the 'membrane-inserted' state.

TAP tagging
(Tandem affinity purification tagging). A process in which a protein is carboxy-terminally tagged with a peptide containing a calmodulin-binding peptide, a TEV protease cleavage site and protein A. The protein is first purified using Immunoglobulin G-coated beads that bind protein A. The protein fusion is then cleaved from the gene of interest by the TEV protease and purified.

Chaperones
Proteins that assist other macromolecules in folding and/or unfolding and assembly and/or disassembly.



© 2012 Macmillan Publishers Limited. All rights reserved

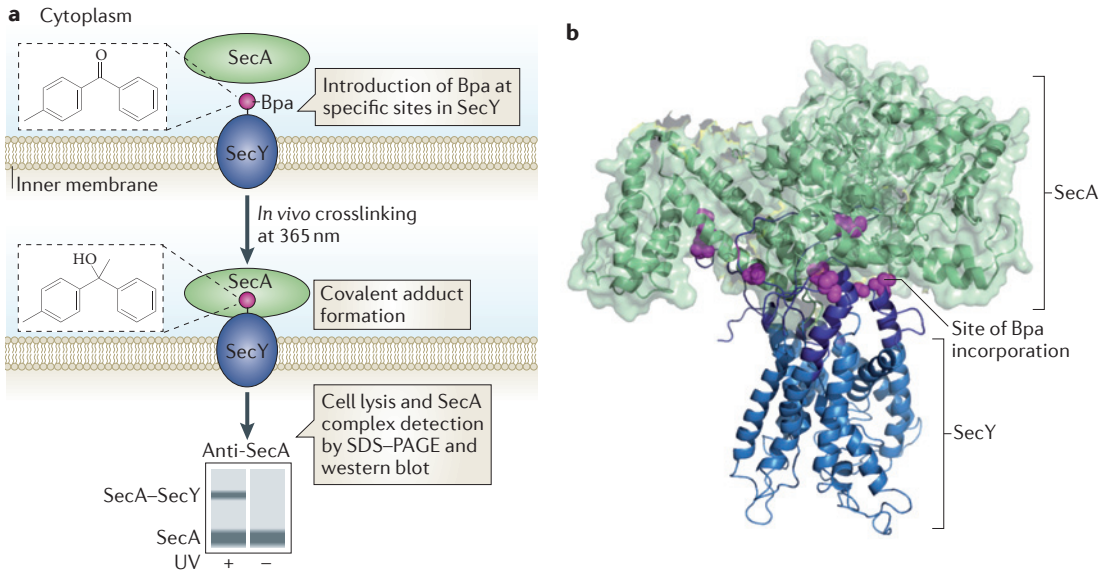



Figure 2 | ***In vivo* photocrosslinking of membrane proteins and study of protein interactions.** The molecular surfaces involved in the interaction between SecA and the SecY component of the SecY–SecE–SecG (SecYEG) translocon at the inner membrane in *Escherichia coli* can be revealed via *in vivo* protein crosslinking with genetically encoded *p*-benzoyl-L-phenylalanine (Bpa). **a** | To analyse the basis of the interaction between SecA and SecY, Bpa was introduced at 53 positions in the cytoplasmic loops of SecY. The protein was then crosslinked *in vivo*, forming covalent adducts with SecA. Covalently linked SecA–SecY complexes can then be isolated through cell lysis, and the complexes can be detected by SDS–PAGE and western blot analysis with antibodies against SecA. **b** | The crystal structure of SecA–SecYEG complex from *Thermotoga maritima* (Protein Databank ID: 3DIN[34]) allows a homology model of the *E. coli* complex to be built (note that SecE and SecG are omitted here). The residues where *in vivo* crosslinking has been observed are shown in magenta, and all lie at the SecA–SecY interface. UV, ultraviolet light.

The presence of sodium azide increased crosslinking efficiency for residues within C6 of SecY but had little effect on crosslinking to other regions. Further photocrosslinking experiments were carried out *in vitro* on inverted membrane vesicles, and showed that photocrosslinking of SecA to C6 of SecY, but not to C5, depended on the presence of both ATP and a translocation substrate protein, outer membrane protein A precursor (proOmpA). These experiments suggest that C5 of SecY in SecYEG constitutively interacts with SecA but that C6 of SecY engages with SecA only when a protein substrate is being actively translocated. Further crosslinking experiments using truncated forms of SecA suggested that its amino-terminal region is sufficient to mediate the constitutive interaction with C5 of SecY. Thus, this example illustrates how the Bpa photocrosslinker can be used to determine the regions of contact between a membrane protein and its interacting partner, as well as how these interactions are affected by the functional state of its interacting partner.

## Installing post-translational modifications

Post-translational modifications of proteins can dynamically alter their molecular function, localization and stability, and they regulate almost all aspects of biology. There is extensive crosstalk between modifications, making it challenging to address the function of any particular modification by genetic approaches. Mass spectrometry-based approaches have made spectacular progress in defining the modified proteome[35] but provide little information about how modifications alter protein structure, function or activity. Moreover, the identification of modification sites using these approaches does not provide any information about the enzymes that add or remove the modifications, making it challenging to understand the relationship between an observed modification and its biological function.

It can be difficult to produce homogeneously and site-specifically modified proteins on which to perform biochemistry, enzymology or structural biology and thereby assess the role of post-translational modifications. For example, the enzymes that carry out the modification *in vivo* may be unknown or may not modify the protein site specifically or completely *in vitro*. Genetic code expansion methods have been reported for introducing Tyr nitration[36], Tyr sulphation[37], analogues of phosphorylated Tyr[38,39], phosphorylated Ser[40], Lys monomethylation and dimethylation[41–45], Lys acetylation[6] and Lys ubiquitylation[46–48] into recombinant proteins, and these approaches are beginning to help define the function and regulation of several post-translational modifications.

***Installing phosphorylation, sulphation and nitration.*** Protein phosphorylation on Ser, Thr and Tyr residues, which is promoted by protein kinases and removed by phosphatases, is perhaps the most well-studied post-translational modification. Several analogues that mimic Tyr phosphorylation have been introduced into proteins using genetic code expansion[38,39].



© 2012 Macmillan Publishers Limited. All rights reserved

For example, a carboxylic acid derivative of Tyr that can be incorporated into proteins in *E. coli* may mimic some features of phosphorylation[39]. A clever strategy for converting *p*-azido Phe, which can be incorporated into proteins in bacteria and eukaryotic cells, to the phosphoamidate analogue of phosphorylated Tyr has also been reported[38]. Phosphorylated Ser has also been incorporated into proteins, albeit with low efficiency; this is achieved by importing an archeal synthetase and tRNA pair (in which the synthetase naturally aminoacylates the tRNA with phosphoserine as an intermediate to Cysteinyl-tRNA) and introducing alterations to elongation factor-Tu (EF-Tu) so that it accepts the tRNA which is aminoacylated with phosphoserine[40]. Sulphonated Tyr, a modification that is found naturally on cell surface receptors, has been incorporated into proteins[37], and this has allowed antibodies with higher affinity for the ligands of sulphated receptors to be identified[49]. Last, addition of 3-nitrotyrosine, a modification of Tyr resulting from oxidative damage and correlated with numerous disease states, has been used to demonstrate that site-specific nitration on manganese superoxide dismutase — a modification that is also observed *in vivo* — leads to its complete inactivation[36].

***Installing Lys methylation.*** Lys residues in proteins can be monomethylated, dimethylated and trimethylated by methyltransferases, and methylation can be removed by demethylases[50]. Both the degree and site of methylation may control biological outcomes, and the role of methylation on histones in regulating chromatin function has been extensively studied. For example, methylation of Lys9 in histone H3 (H3K9) is associated with heterochromatin formation[51,52]. Monomethylated and dimethylated Lys can be installed into recombinant proteins using the PylRS–tRNA pair[44]. Using the unmodified PylRS–$tRNA_{CUA}$ pair, a 'protected' version of monomethylated Lys ($N^{\varepsilon}$-tert-butyloxycarbonyl-$N^{\varepsilon}$-methyl-L-Lys) was site-specifically introduced into recombinant histone H3 produced in *E. coli*. Removal of the tert-butyloxycarbonyl group from the protein under mild conditions then revealed monomethylated Lys at the genetically programmed site. This two-step approach was necessary because $N^{\varepsilon}$-methyl-L-Lys is so similar to Lys that it is difficult to create an aminoacyl-tRNA synthetase to specifically incorporate methylated Lys into proteins in the presence of the high Lys concentration that is constitutively present in the cell. The two-step approach for installing Lys methylation has been extended using different protecting groups that can be removed with light or metal complexes[41–43]. A technique for genetically directing Lys dimethylation in recombinant histones, thereby generating site-specifically dimethylated histones in large quantities, has also been described[45]. So far, these approaches for installing monomethyl and dimethyl Lys in recombinant H3 have been used in proof-of-principle experiments to immunoprecipitate a known interaction partner (heterochromatin protein 1 (HP1)) of monomethylated and dimethylated H3K9 (REFS 44,45). But it seems likely that these approaches will facilitate new biological discoveries.

***Installing Lys acetylation.*** Lys residues in proteins may be acetylated by acetyltransferases and deacetylated by histone deacetylases (HDACs) or sirtuins. Lys acetylation is a key post-translational modification in regulating chromatin structure and function and epigenetic phenomena[53,54]. Moreover, mass spectrometry and a range of other methods have now revealed that diverse biological processes, from metabolism to signalling and immunity, are also regulated by acetylation[55,56]. Acetyl-lysine can be installed into recombinant proteins produced in *E. coli* using a modified PylRS–$tRNA_{CUA}$ pair[6]. Using this approach, several recombinant proteins bearing homogeneous acetylation at a defined site have been produced. This has provided new insights into the role of acetylation that would be challenging or impossible to achieve by other methods, including insights into the role of acetylation in regulating chromatin structure and function, enzyme activity, immunosuppression, viral infection and tumour suppressor and transcription factor activity[6,57–61].

***The role of H3K56 acetylation.*** Acetylation of H3K56 has been implicated in diverse biological processes, including the definition of epigenetic status and the regulation of transcription, DNA replication and DNA repair[62–77]. Recombinant H3 bearing quantitative acetylation at Lys56 has been produced in *E. coli* and then reconstituted into histone octamers and assembled into nucleosome core particles and nucleosomal arrays. This has allowed testing of how H3K56 acetylation may control diverse phenomena by altering higher order chromatin structure; remodelling or repositioning nucleosomes; allowing nucleosome core particle 'breathing' or opening on DNA; or recruiting factors to this modification[57].

Contrary to the prevailing 'dogma', H3K56 acetylation did not measurably affect the chromatin compaction of nucleosomal arrays in this analysis. Remodelling assays of single nucleosome movements on DNA revealed that acetylation of H3K56 caused only modest increases in repositioning of nucleosomes by SWI/SNF and RSC remodelling enzymes.

The crystal structure of the nucleosome core particle shows that H3K56 is near the entry and exit site of DNA and that Lys56 in H3 makes a water-mediated contact with the DNA phosphate backbone[78]. Therefore, it was proposed that neutralization of the Lys charge by acetylation might alter DNA breathing through the local disengagement of regions of the DNA from the protein surface in the protein–DNA complex. This was experimentally demonstrated and quantified by single-nucleosome fluorescence resonance energy transfer (FRET) experiments[57,79] (FIG. 3a). FRET probes on the DNA revealed an increase in breathing of DNA on the nucleosome core particles that contained acetylated H3K56 compared with those that contained wild-type H3. Moreover, by changing the position of the FRET probes on the DNA, it was possible to show that the acetylation-dependent increase in breathing localized to one turn of the DNA helix. This histone acetylation-dependent increase in DNA breathing may increase the accessibility of the DNA and the modified nucleosome to other factors that alter chromatin, providing a possible biophysical basis

Elongation factor-Tu
(EF-Tu). A protein that binds to aminoacylated tRNAs and delivers them to the ribosome for protein synthesis.

Nucleosome
A unit of DNA packaging in eukaryotes in which a length of DNA is wrapped around an octamer of histone proteins.

SWI/SNF
(Switch/sucrose nonfermentable). An ATP-dependent multiprotein nucleosome-remodelling complex that is found in yeast.

RSC
(Remodels the structure of chromatin). An ATP-dependent multiprotein nucleosome-remodelling complex.



© 2012 Macmillan Publishers Limited. All rights reserved

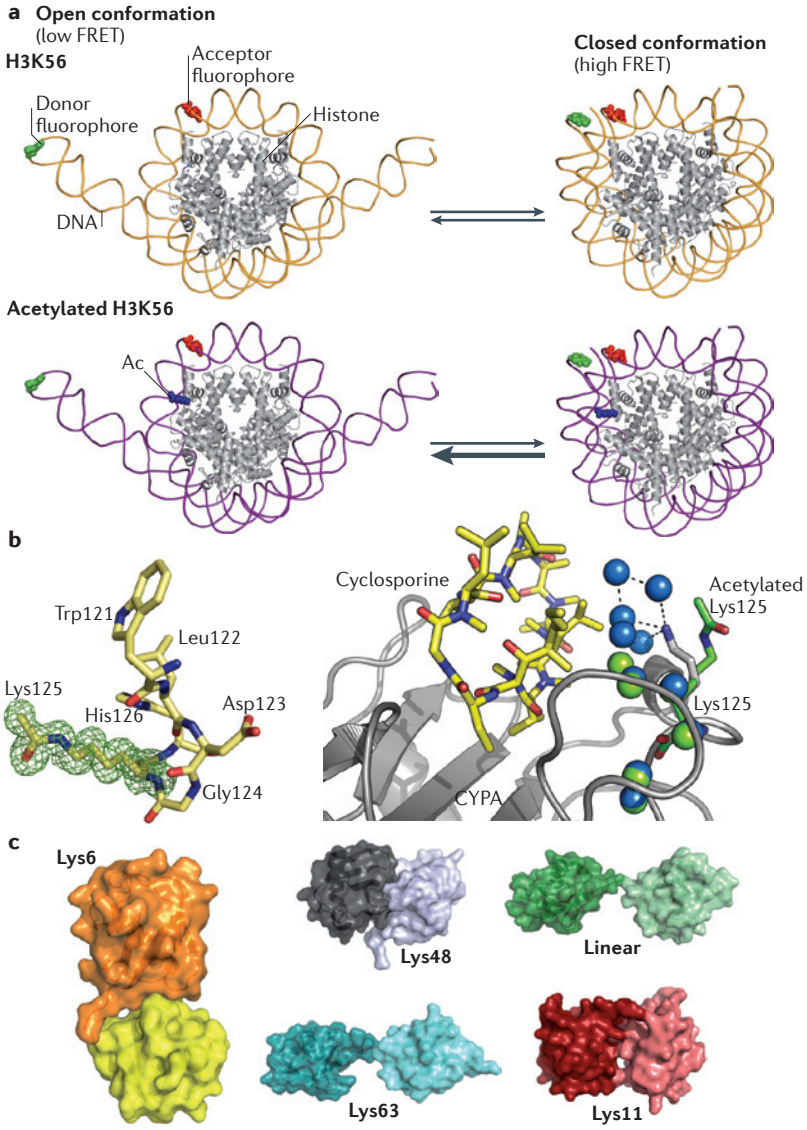



Figure 3 | **Applications of genetically encoded post-translational modifications. a** | Genetically encoded $N^{\varepsilon}$-Lys acetylation allows the role of acetylation in DNA 'breathing' to be assessed by single-molecule fluorescence resonance energy transfer (FRET). When FRET is measured between a donor and an acceptor fluorophore on DNA, a larger fraction of nucleosomes containing acetylated Lys56 on histone H3 (H3K56) are found with low FRET efficiency, suggesting that Lys56 acetylation favours DNA breathing and a more open conformation. The models are based on the X-ray structure of the nucleosome core particle (Protein Databank (PDB) ID: 1KX5)[78]. **b** | Genetically encoded $N^{\varepsilon}$-Lys acetylation allows the high-resolution X-ray structures of acetylated proteins and their complexes to be solved. The high-resolution structure of acetylated Lys125 from acetylated cyclophilin A (CYPA) is shown on the left and includes the experimental electron density map (green mesh). A comparison between unmodified CYPA and acetylated CYPA, both in complex with cyclosporine, is shown on the right (PDB ID: 2CPL[176]). Five water molecules (blue spheres) that are ordered at the interface between CYPA and the small molecule cyclosporine in the unacetylated complex rearrange in the acetylated complex, reducing the affinity of this interaction. **c** | Genetically directed ubiquitylation has allowed the determination of an X-ray structure of atypical Lys6-linked diubiquitin (PDB ID: 2XK5 (REF. 47)) (yellow and orange), which shows that it forms a distinct asymmetric conformation compared with other ubiquitin linkages, such as Lys48 (PDB ID: 1AAR[177]), Lys63 (PDB ID: 2JF5 (REF. 178)), linear chains (in which the amino terminus of one ubiquitin is linked to the carboxyl terminus of another ubiquitin (PDB ID: 2W9N[178])) and Lys11 (PDB ID: 2XEW[179]). Images in part **a** are reproduced, with permission, from REF. 57 © (2009) Elsevier. Images in part **b** are reproduced, with permission, from REF. 58 © (2010) Macmillan Publishers Ltd. All rights reserved. Images in part **c** are modified, with permission, from REF. 47 © (2010) Macmillan Publishers Ltd. All rights reserved.

for observed cellular phenomena. It is important to note that the ability to produce homogeneously and quantitatively modified protein at Lys56 of histone H3 was essential to making interpretable measurements in these single-molecule experiments. Thus, these data demonstrate how access to defined modified proteins allows mechanistic hypotheses to be tested in a unique way.

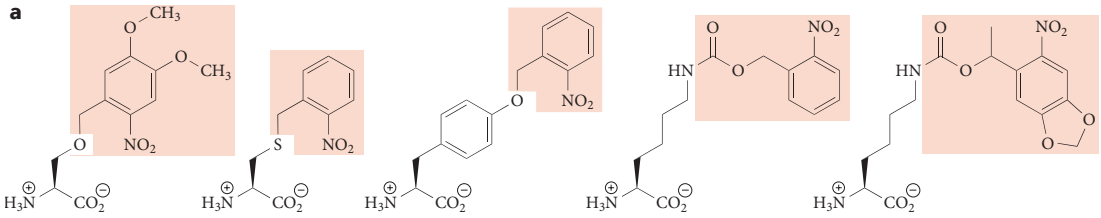
***CYPA acetylation during viral infection and immunity.*** Cyclophilin A (CYPA; also known as PPIA) is a *cis–trans* prolyl isomerase, and two of the most important roles of CYPA are in controlling immunosuppression and viral infection. CYPA is the target of the widely used immunosuppressive molecule cyclosporine: the complex formed by CYPA and cyclosporine binds and allosterically inhibits calcineurin, leading to suppression of the T cell-mediated immune response[80]. CYPA is also an essential host protein for the efficient replication of several viruses, including HIV-1 (REFS 81,82). During HIV-1 infection, CYPA interacts with the capsid protein GAG and is packaged into budding HIV-1 virions; and inhibition of CYPA by cyclosporine substantially decreases viral titre in human cells. It has been proposed that disassembly of the HIV capsid, which is an essential step in HIV infection, may be catalysed by CYPA-mediated isomerization of the Gly89–Pro90 bond in the HIV capsid.

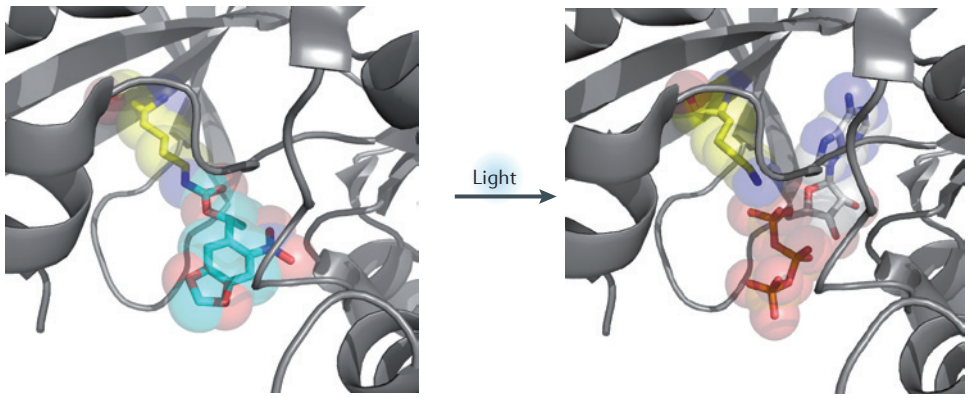
The acetylation of a cyclophilin at Lys125 was identified in a proteomics screen[56]. It was subsequently demonstrated that a substantial proportion of CYPA is acetylated in HeLa cells and human T cell lines, suggesting that the acetylated form of the protein is biologically relevant[58]. To test this, recombinant CYPA bearing homogenous acetylation at Lys125 was prepared by overexpression in *E. coli* using an acetyllysyl-tRNA synthetase–$tRNA_{CUA}$ pair, allowing detailed biophysical and enzymatic characterization of acetylated CYPA and associated complexes, as well as the first high-resolution structures of an acetylated protein and associated complexes[58] (FIG. 3b). The acetylation of CYPA at

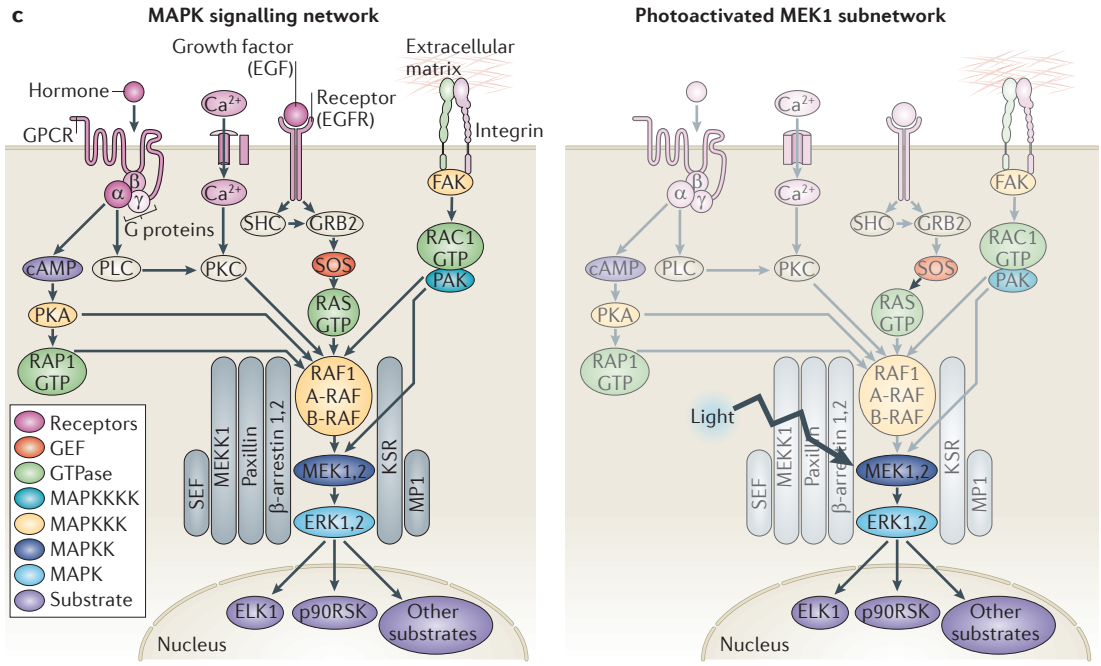
Prolyl isomerase
An enzyme that is responsible for the *cis–trans* isomerization of peptide bonds that are on the amino-terminal side of Pro residues.

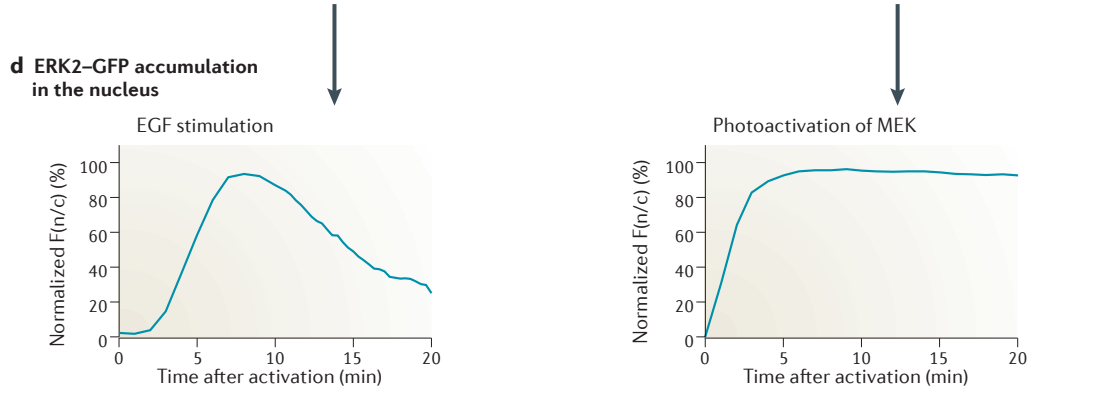


© 2012 Macmillan Publishers Limited. All rights reserved




**Enthalpy**
A measure (in thermodynamics) of the total internal energy of a system plus its pressure multiplied by its volume.

**Entropy**
A thermodynamic property that can be used to determine the energy that is not available for work.

**E1, E2 and E3**
The enzymes by which ubiquitin is added to cellular proteins. The E1 activates ubiquitin by attaching the molecule to its active site Cys residue. The E2 then binds the ubiquitin molecule, also by a Cys residue. The E2 then binds an E3 ligase which binds the target protein and catalyses the transfer of the ubiquitin to a Lys residue of the target protein.

**Atypical ubiquitin chains**
Ubiquitin chains that are not linked by the common Lys48 or Lys63 linkages but by one of the other Lys residues of ubiquitin.

**Sumoylation**
The covalent attachment of small ubiquitin-like modifier (SUMO) to a substrate protein.

**Neddylation**
A process, analogous to ubiquitylation, in which ubiquitin-like protein NEDD8 is conjugated to a protein substrate.



© 2012 Macmillan Publishers Limited. All rights reserved

◂ Figure 4 | **Genetically encoded photocaged amino acids allow insights into cell biology in real time. a** | Photocaging natural amino acids. The natural amino acids are shown with the attached photocages (in pink). From left to right: Ser caged with a 4,5-dimethoxy-2-nitrobenzyl group; Cys caged with a 2-nitrobenzyl group; Tyr caged with a 2-nirobenzyl group; Lys caged with a 2-nitrobenzyloxycarbonyl group; and Lys caged with a 6-nitropiperonyloxycarbonyl group. **b** | Photocaging a near-universally conserved key residue in the active site of MAPK/ERK kinase (MEK), along with the introduction of activating mutations elsewhere in the kinase, allows rapid activation of kinase function in response to light. The models are based on a previously obtained MEK1 structure (Protein Databank ID: 1S9I[180]). **c** | The mitogen-activated protein kinase (MAPK) signalling pathway can be activated *in vivo* on binding of growth factors, such as epidermal growth factor (EGF), to receptors, such as EGF receptor (EGFR), and also shows intricate crosstalk with other pathways, such as those downstream of integrins and G protein-coupled receptors (GPCRs). Activation of MEK1 and MEK2 allows phosphorylation and nuclear import of ERK1 and ERK2, allowing substrate activation. Photoactivation of the MEK kinase allows activation of a specific 'subnetwork' in MAPK signalling. **d** | How photoactivation of MEK, compared with whole-pathway stimulation with EGF, affects the kinetics of ERK2–green fluorescent protein (GFP) accumulation in the nucleus has provided insights into the elementary steps in signalling. The y-axis represents the percentage normalized fluorescence expressed as a fraction of nuclear GFP divided by cytoplasmic GFP fluorescence (F(n/c)). cAMP, cyclic AMP; FAK, focal adhesion kinase; GEF, guanine nucleotide exchange factor; GRB2, growth factor receptor-bound 2; KSR, kinase suppressor of RAS; MAPKK, MAPK kinase; MAPKKK, MAPKK kinase; MAPKKKK, MAPKKK kinase; MEKK1, MEK kinase 1; MP1, MAPK scaffold protein 1; p90RSK, p90 ribosomal protein S6 kinase; PAK, p21-activated kinase; PK, protein kinase; PLC, phospholipase C; SHC, SH2 domain-containing transforming protein. Images in parts **b**–**d** are modified, with permission, from REF. 98 © (2011) American Chemical Society.

Lys125 decreased its affinity for cyclosporine 20-fold. Comparison of the crystal structure of acetylated CYPA in complex with cyclosporine and the same complex for non-acetylated CYPA revealed that the acetylation of Lys125 disrupts a network of five water molecules that form an interface between unacetylated CYPA and cyclosporine. This acetylation-mediated solvent remodelling illustrates how acetylation may control protein–small molecule interactions without eliciting rearrangement of the structure or inducing allosteric changes. CYPA acetylation also decreased the affinity of the CYPA–cyclosporine complex for calcineurin twofold. As cyclosporine-mediated immunosuppression requires its binding to CYPA and the CYPA–cyclosporine complex binding to calcineurin, both of which are decreased by acetylation, the effect of CYPA acetylation on immunosuppression may be greater than the effect measured for each step individually.

Measurements on the CYPA–HIV-1 capsid interaction revealed that neither acetylation nor a Lys-to-Gln mutation at position 125 changes the affinity of CYPA for the HIV-1 capsid, but that the enthalpy and entropy contributions to this affinity do change[58]. Therefore, the data are consistent with acetylated CYPA, Lys125Gln CYPA and wild-type CYPA binding the capsid through different modes, demonstrating that, although Gln is often used by researchers as a mimic of acetylation, it is not a good mimic for the acetylation of CYPA. The crystal structure of acetylated CYPA in complex with the HIV-1 M group capsid N-terminal domain also showed that the Gly89–Pro90 peptide bond of the HIV-1 capsid was in the *cis* form, rather than the *trans* form observed in the unacetylated CYPA complex, and acetylation decreased the catalytic efficiency for *cis–trans* prolyl isomerization of a model substrate 35-fold. These data are consistent with a model in which acetylation of CYPA decreases capsid isomerization and, therefore, potentially has an impact on capsid disassembly and HIV-1 infection. Together, these studies also demonstrate how insight into the consequences of a modification on biological function can be obtained through the synthesis and study of the modified protein.

***Installing Lys ubiquitylation.*** The post-translational modification of target proteins with the 76-amino-acid protein ubiquitin regulates many aspects of eukaryotic cell biology[47,83]. Ubiquitylation, in which the carboxy-terminal carboxylate of ubiquitin is attached to a Lys residue to form an isopeptide bond, is naturally mediated by E1, E2 and E3 enzymes and removed by deubiquitinases (isopeptidases). However, the specific combinations of enzymes that are used to ubiquitylate many proteins are unknown, making it challenging to understand the roles of protein ubiquitylation and to decipher deubiquitinase specificity. As ubiquitin is itself a protein, this post-translational modification cannot be genetically encoded directly using an evolved synthetase and a tRNA, but genetic code expansion methods have been developed to site-specifically install Lys derivatives that allow ubiquitin to be attached to proteins, in some cases through a native isopeptide bond.

For example, $N^{\varepsilon}$-cysteinyl-L-Lys has been incorporated into proteins in response to the amber codon using evolved PylRS–$tRNA_{CUA}$ pairs[48,84]. Native chemical ligation at this amino acid with a ubiquitin (residues 1–75) thioester generates an isopeptide bond[48,85]. However, this approach generates a non-native linkage in which the C-terminal Gly of ubiquitin, which is universally conserved, is mutated to Cys (or Ala following desulphurization). This linkage is not a substrate for deubiquitinases and this may be an advantage in approaches aiming to identify binding partners for ubiquitylated proteins in complex mixtures that contain deubiquitinases.

Two methods have used genetic code expansion to create native isopeptide bonds between recombinant proteins and ubiquitin[46,47]. In the first approach, termed GOPAL (genetically encoded orthogonal protection and activated ligation)[47], a protected version of Lys is installed in response to the amber stop codon and defines the ultimate site of isopeptide bond formation. A series of chemical protection and deprotection steps are then used to direct the formation of an isopeptide bond with ubiquitin. Although GOPAL, and its variants, can be used to generate atypical ubiquitin chains to study their function[47,86], it does require protein unfolding and may, therefore, be challenging to apply to the ubiquitylation of large multidomain proteins that cannot be refolded. In a second approach, a PylRS–tRNA pair has been evolved to incorporate a δ-thiol-Lys derivative[46], which is known, from work on synthetic peptides[87,88], to react with C-terminal thioesters of ubiquitin to yield an isopeptide bond (BOX 1). Desulphurization of the δ-thiol following isopeptide bond formation renders the approach entirely traceless. This approach was used



© 2012 Macmillan Publishers Limited. All rights reserved

to make ubiquitylated small ubiquitin-like modifier (SUMO) for the first time[46], and simple extensions of the approach should allow access to large proteins bearing defined ubiquitylation, as well as to proteins bearing sumoylation or neddylation.

***Atypical ubiquitin chain structure and deubiquitinase specificity.*** Atypical ubiquitin chains make up a large fraction of polyubiquitin in cells[89]. However, the biological function, assembly and disassembly of atypical chains are less well understood than they are for Lys48 and Lys63 chains, which have established roles in proteasomal degradation and signalling, respectively. It was not previously possible to make these linkages so that their function and regulation could be investigated, as the specific ligases responsible for their synthesis were unknown. Using GOPAL, Lys6- and Lys29-linked diubiquitins were synthesized that contained the minimal ubiquitin–ubiquitin linkage of atypical ubiquitin chains[47]. Atypical dimers have also been made in other studies by peptide synthesis and native chemical ligation[87,88,90–94].

The crystal structure of Lys6-linked diubiquitin, produced using GOPAL[47], reveals an asymmetric compact conformation (FIG. 3c). The structure is distinct from previously observed diubiquitin structures for other isopeptide linkages and is also distinct from a computational model of Lys6-linked diubiquitin. Interestingly, iterative modelling of a longer Lys6-linked ubiquitin chain, on the basis of the diubiquitin structure, suggests the possibility that Lys6-linked ubiquitin may form a helical filament. Thus, the structure of Lys6-linked diubiquitin reveals molecular features that may subsequently explain how Lys6-linked ubiquitin polymers may be uniquely recognized in the cell to elicit particular biological outcomes.

The ability to synthesize atypical ubiquitin chains has allowed 10% of human deubiquitinases, which may each disassemble chains, to be profiled for their isopeptide linkage specificity[47]. These experiments revealed that atypical linkages are the preferred substrates of certain deubiquitinases; notably, the OTU domain deubiquitinase TRABID (also known as ZRANB1) cleaves Lys29-linked diubiquitin 40 times more rapidly than Lys63-linked ubiquitin[95]. As TRABID has been implicated as a positive regulator of WNT signalling, these experiments reveal a potential new link between WNT signalling and atypical ubiquitin chains and highlight the potential of these approaches for assessing the regulation and biological roles of both classic and atypical ubiquitin chains.

## Photocaged amino acids

Protein function may be abrogated by site-specifically replacing a natural amino acid in the protein with an unnatural photocaged version of the same amino acid, the presence of which blocks a particular function of the natural protein. A millisecond pulse of light may then be used to de-cage the unnatural amino acid and restore this function. A key advantage of this approach is that it allows a defined function of a specific protein to be rapidly and specifically activated inside living cells and organisms, and so facilitates the study of the resulting adaptive changes.

Photocaged Tyr[96], Cys[10], Ser[97], and Lys[7] have all been genetically encoded using evolved aminoacyl-tRNA synthetase–$tRNA_{CUA}$ pairs (FIG. 4a), and some of these amino acids have been used for biological studies. For example, photocaged Ser has been used to examine the contribution of two phosphorylation sites in the yeast transcription factor Pho4 to its export from the nucleus[97]. Photocaged Lys derivatives have been incorporated in response to the amber codon in *E. coli*, yeast and mammalian cells[4,7,98,99]. The use of these approaches for mammalian cells was demonstrated by genetically encoding a caged Lys in place of a key Lys within a nuclear localization sequence fused to green fluorescent protein (GFP) and/or in place of a key Lys in p53 that has been reported to be important for nuclear localization[7]. This mislocalized both proteins to the cytosol, allowing the kinetics of their nuclear import to be measured reproducibly in single cells[7] following a millisecond-to-second pulse of light.

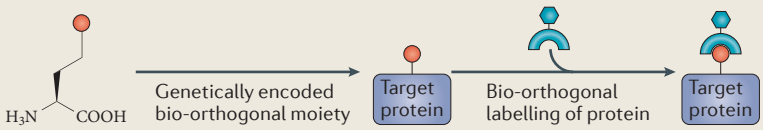
***Insights into MEK signalling dynamics.*** The RAF–MEK–ERK (RAF–MAPK/ERK kinase (also known as MAPKK)–extracellular signal-regulated kinase) mitogen-activated protein kinase (MAPK) signalling pathway, which is crucial for regulating diverse cellular phenomena, including growth and proliferation, is typically activated at the cell surface by, for example, the binding of epidermal growth factor (EGF) to the EGF receptor (EGFR). EGF stimulation leads to relocalization of ERK2 from the cytoplasm to the nucleus, and this can be followed by fluorescence microscopy using an ERK2–GFP fusion[98,100] (FIG. 4). Following EGF stimulation, there is a lag phase before ERK2–GFP accumulates in the nucleus, and ERK2–GFP then dissipates from the nucleus despite persistent EGF stimulation; this latter phenomenon — described as exact adaptation[100] — suggests there may be negative feedback in the pathway. To begin to understand how the kinetics of these steps contribute to information processing, and to dissect the architecture of feedback control mechanisms, it would be very helpful to be able to rapidly turn on a single kinase in this pathway to activate a signalling 'subnetwork' by a process that is independent of extracellular stimuli.

By introducing mutations to make MEK constitutively active and simultaneously replacing the key Lys residue in the active site of MEK with a photocaged Lys, a photoactivatable version of MEK was created in mammalian cells[98] (FIG. 4b). When cells were subjected to a pulse of light, active MEK was produced that phosphorylated ERK2–GFP, leading to its nuclear import. The kinetics of this process were followed by fluorescence microscopy, and additional experiments demonstrated that activation of MEK leads to phosphorylation of ERK2, as well as to the phosphorylation of nuclear targets of ERK2 (REF. 98). This demonstrated that a subnetwork of MAPK signalling had been specifically photoactivated by a process that was independent of receptor stimulation (FIG. 4c).



© 2012 Macmillan Publishers Limited. All rights reserved

Box 2 | **Bio-orthogonal chemistry and protein labelling**




Bio-orthogonal chemical reactions allow labelling of an incorporated unnatural amino acid containing a bio-orthogonal functional group (orange circle) with a chemical reporter (blue hexagon; see the figure). These reactions proceed under biologically compatible conditions, in which the reactants form a product with each other but do not react with chemical groups found in other biological molecules or in cells and organisms[143]. Amino acids bearing bio-orthogonal functional groups, including azides, alkynes, ketones, alkenes, tetrazoles, anilines, 1,2-amino thiols, aryl halides and boronic acids, can now be site-specifically incorporated into proteins and labelled through bio-orthogonal reactions[84,144–172] (see Supplementary information S1 (table)). The site-specific introduction of probes into a protein via the genetic installation of a bio-orthogonal functional group and subsequent chemical labelling allows almost any probe to be conjugated to the protein, and it has been possible to site-specifically label proteins *in vitro* with fluorophores, biotin and polyethylene glycol (PEG) in model studies.

The bio-orthogonal chemistries that can currently be encoded, with notable exceptions[84,152,172], have rate constants of $10^{-4}$ $M^{-1}s^{-1}$ to $10^{-2}$ $M^{-1}s^{-1}$ (REF. 173), which are six to eight orders of magnitude slower than methods that are used to successfully label proteins in or on cells. However, recent advances in the development of very rapid bio-orthogonal reactions, including the incredibly rapid reaction between tetrazines and strained alkenes[174,175], and the encoding of their reaction partners[84,152,172], have begun to allow site-specific labelling of cellular proteins[172]. This is an area with tremendous potential, which may be fully realized as faster bio-orthogonal reactions are developed and their components are encoded.

Comparison of the kinetics of ERK2–GFP nuclear accumulation following MEK photoactivation or endogenous receptor stimulation provided some interesting insights (FIG. 4d). Whereas there is a lag phase of minutes before the nuclear accumulation of ERK2–GFP following receptor stimulation, this long lag phase is not present in the photoactivated subnetwork, suggesting that the lag may result from steps upstream of MEK. Once initiated, the rates of ERK2–GFP nuclear accumulation are comparable, suggesting that the rate-determining step for ERK2 nuclear import may lie between MEK and ERK2. In addition, although exact adaptation is observed in the receptor-stimulated case, this is not observed in the photoactivated subnetwork. This suggests that the feedback mechanism that controls exact adaptation may operate upstream of MEK in the MAPK pathway. Finally, following the import of ERK2–GFP with high time resolution showed that it accumulated sigmoidally, which is consistent with its distributive dual phosphorylation, previously observed *in vitro*[101], being rate-determining for nuclear import. Because the Lys caged in these experiments is almost universally conserved in protein kinases[102], it should be possible to apply variations of this approach to understanding the elementary steps in many kinase pathways. Extensions of this approach may provide insights that are not readily accessible by classical genetics or by small interfering RNA-based approaches.

Distributive
A mode of activation in which two or more events occur independently. In the case of ERK2, which is diphosphorylated when activated, the MEK1–ERK2 complex dissociates after each phosphorylation event.

## Introducing biophysical probes and labels

Several amino acids bearing biophysical probes can now be directly incorporated into proteins. These include fluorescent probes, NMR probes and infrared probes that can report on the location of proteins within cells or the chemical environment of the region of a protein in which they are installed. Proteins may also be site-specifically labelled with heavy atoms to facilitate X-ray crystallography.

A range of methods is also available for labelling proteins with a much wider range of other molecules, including cytotoxins, biotin and fluorophores. These methods include the use of fluorescent protein fusions[103–106], self-labelling proteins (for example, SnapTag, HaloTag and ClipTag)[107–110], ligases (for example, biotin ligase, lipoic acid ligase, sortase and phosphopantetheinyl transferase)[111–118] and self-labelling tags (for example, tetracysteine and tetraserine)[119,120]. Some of these approaches allow rapid labelling and have had a substantial impact on biological studies. However, they require the use of protein fusions and/or the introduction of additional sequences into the protein of interest. This may disturb the structure and function of the protein and can make it challenging to place probes at any position in a protein. Moreover, the range of probes that can be incorporated by some of these methods is limited[105,106,121]. The site-specific installation of bio-orthogonal chemical groups into a protein by genetic code expansion and the labelling of proteins via specific bio-orthogonal reactions (BOX 2; see Supplementary information S1 (table)), provides an emerging route to site-specific protein labelling with almost any probe.

***Probing protein structure by X-ray crystallography and NMR.*** Labels and probes can be incorporated that may allow new structural insight into proteins. Heavy atoms may be incorporated into proteins to assist with phasing X-ray data for structure determination. For example, *p*-Iodo-L-Phe has been incorporated into T4 lysozyme, and use of the iodine to phase the X-ray data allowed the crystal structure of T4 lysozyme to be obtained with less data than would be required using more traditional methods[122]. However, this approach has not yet yielded information that could not have been obtained using established methods, such as selenomethionine incorporation[123] for phasing. NMR-active versions of unnatural amino acids (bearing $^{19}F$, $^{13}C$ or $^{15}N$ isotopes) have also been site-specifically incorporated into proteins[124–129], allowing a single site in a protein to be observed by NMR. Proof-of-principle experiments have demonstrated that site-specific $^{19}F$ labelling of specific proteins can be detected by NMR in *E. coli*[129]. In principle, 'in-cell NMR' has the potential to yield new insights into the conformational changes that proteins undergo inside cells.

***Optical probes: fluorescent and infrared probes.*** Four fluorescent amino acids have been incorporated into proteins[130–133]. An amino acid containing a coumarin fluorophore has been encoded in *E. coli*[132] and has been used to label the chaperonin protein GroEL. Fusions of GroEL to GFP are problematic for imaging GroEL localization because they interfere with GroEL function. By contrast, a coumarin-labelled GroEL was shown to be functional and allowed the localization of GroEL during stress response and heat shock to be observed, providing evidence that GroEL does not relocalize upon exposure to heat shock[134]. An environmentally sensitive dansyl amino



© 2012 Macmillan Publishers Limited. All rights reserved

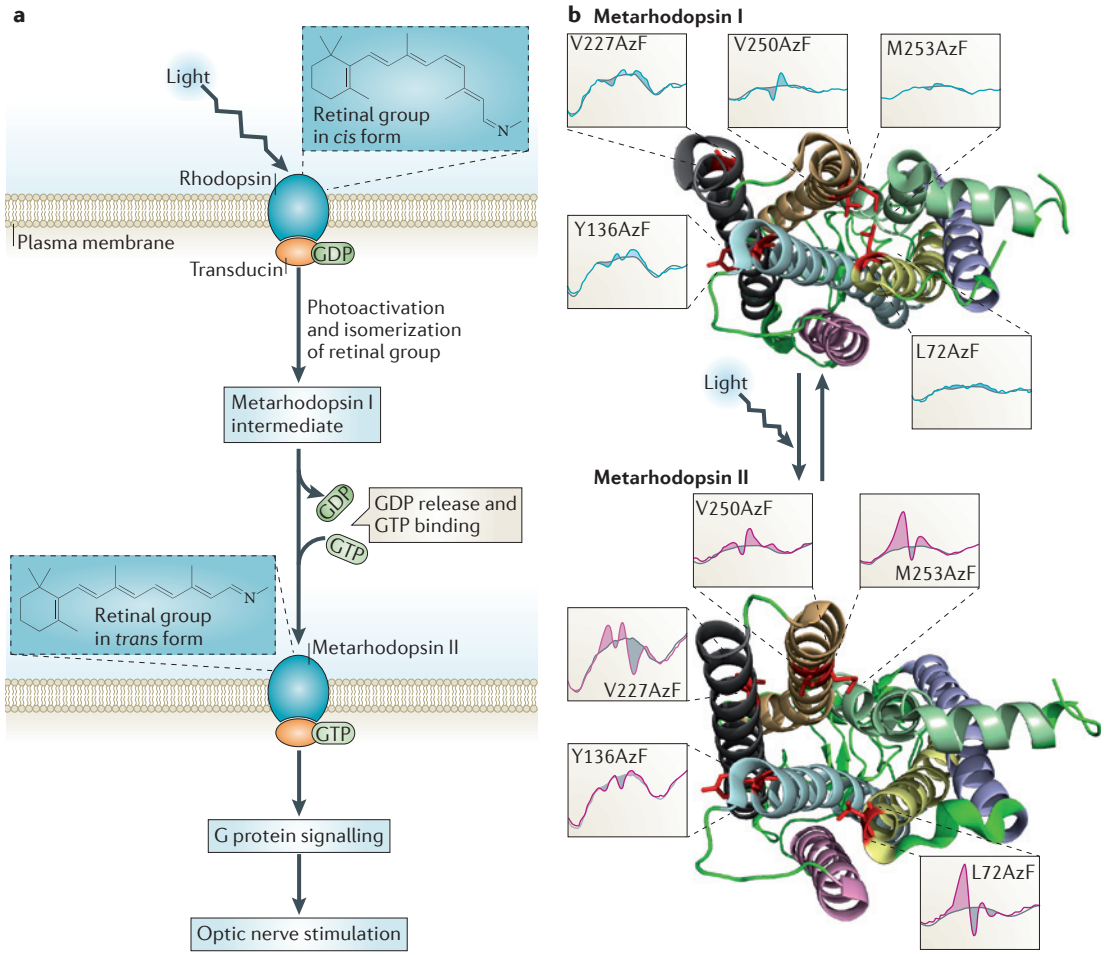



Figure 5 | **Use of genetically encoded infrared probes to dissect local GPCR conformational changes during activation. a** | Inactive rhodopsin, which is located at the cell membrane, is activated by light, which induces conformational changes in a retinal group of the chromophore from a *cis* to a *trans* form. These trigger further changes in the structure of the rhodopsin protein, resulting in conversion of rhodopsin to metarhodopsin II, via the structural intermediate state of metarhodopsin I. The conversion to metarhodopsin II is sensed by the associated G protein transducin, which undergoes a structural change causing GDP to be released from the pocket and allowing GTP to bind. GTP-bound transducin is then able to induce downstream events, leading to G protein signalling and optic nerve stimulation. **b** | Incorporation of *p*-azido-L-phenylalanine (AzF) at several positions (red) in helices 5 and 6 of rhodopsin has provided insights into the conformational changes that occur during the formation of metarhodopsin II via the structural intermediate metarhodopsin I upon exposure to light. Fourier transform infrared (FTIR) spectra between 2,200 and 2,000 $cm^{-1}$ report the altered local chemical environment experienced by the AzF probe. Data for the metarhodopsin I state are shown here mapped onto an inactive bovine rhodopsin structure (Protein Databank (PDB) ID: 1GZM[181]). The spectral data for metarhodopsin II (red) are shown mapped on to the structure of an active bovine opsin (PDB ID: 3DQB[182]). The grey spectra show the equivalent region of a control protein that does not contain the azido-Phe. Spectra for positions V227 and Y136 are shown at x 3 magnification. Residues 136 and 250 already show downshifts during the transition to metarhodopsin I and only show small downshifts from metarhodopsin I to metarhodopsin II, indicating that they move to a less polar environment. Residues 253 and 72 show very large upshifts from metarhodopsin I to metarhodopsin II, indicating that they move to a more polar environment in this transition. Additionally, residue 227 shows a downshift in transition to metarhodopsin I, followed by an upshift from metarhodopsin I to metarhodopsin II. Taken together, the data suggest that the rotation of helix 5 (black) occurs by the metarhodopsin I state, but rotation of helices 2 (yellow) and 5 (orange) do not occur until the metarhodopsin II transition. The shaded areas show deviation from the wild-type rhodopsin spectra (containing no AzF). GPCR, G protein-coupled receptor. Images in part **b** are reproduced, with permission, from REF. 138 © (2010) Macmillan Publishers Ltd. All rights reserved.

acid that changes its fluorescence intensity and its wavelength of fluorescence emission as functions of the hydrophobicity of its local environment has been incorporated in *Saccharomyces cerevisiae* and used to follow protein denaturation *in vitro*[133].

Compared with established methods for fluorescently labelling proteins, one advantage with this technique is that the introduction of these fluorescent unnatural amino acids introduces minimal perturbation to protein structure and is unlikely to interfere with the function



© 2012 Macmillan Publishers Limited. All rights reserved

and localization of proteins. However, the amino acids that can currently be encoded are excited with blue light, making them less than ideal for cellular imaging. In addition, for applications in cells, it may be challenging in some cases to wash out the background of free fluorescent amino acid to give a high signal to noise ratio, a problem that is not encountered when using fluorescent protein fusions.

Phe derivatives bearing nitriles[135] or azides[3,12] have been site-specifically incorporated into proteins. Nitriles and azides have unique vibrational signatures in the spectral range between 1900 $cm^{-1}$ and 2300 $cm^{-1}$, which is well separated from intrinsic protein vibrations. And azide and nitrile vibrations are sensitive to polarity and electric fields, making them good environmentally sensitive probes for Fourier transform infrared (FTIR) spectroscopy. As nitriles and azides are small, they are considered to be less likely to introduce perturbations into protein structure than environmentally sensitive fluorophore labels. However, FTIR is generally less sensitive than fluorescence and requires larger amounts of material. And, although the labelling of proteins using fluorescent amino acids has been demonstrated in cells[136], the use of genetically encoded infrared probes has only been demonstrated *in vitro*.

***Insight into GPCR signalling via infrared probes.*** Incorporation of an azide derivative of Phe in combination with FTIR has been used to examine the conformational changes that take place when rhodopsin, the G protein-coupled receptor (GPCR) that is responsible for vision in dim light, becomes activated by light and binds a cytoplasmic G protein[137,138]. Upon photoactivation, an endogenous retinal group in rhodopsin is isomerized from the *cis* to the *trans* form. This isomerization is coupled to a series of rearrangements in the transmembrane helices, although the precise temporal sequence and the mechanism of the resulting protein conformational changes are unclear. To address this, *p*-azido-L-Phe was incorporated at specific sites in helices 5 and 6 of rhodopsin in mammalian cells, and purified rhodopsin, reconstituted into liposomes, was examined by time-resolved FTIR for changes in the local environment that the azido group experiences as rhodopsin changes from an inactive to an active conformation[137,138] (FIG. 5). In these experiments, comparison of the time-resolved FTIR spectra for the azide-containing mutants and the unmodified protein provides a level of internal control for the effects of introducing the unnatural amino acid on protein function. The FTIR experiments provide evidence that some helix rearrangement occurs early on (in the intermediate metarhodopsin I state), that helices 5 and 6 of rhodopsin are altered before the active state of the receptor is acquired and that receptor activation is defined by a final movement of helix 6. Thus, the introduction of these Phe derivatives into proteins can allow dissection of the structural intermediates that a protein transits through during the activation of signalling.

## Conclusions

It is now possible to site-specifically encode the incorporation of a range of unnatural amino acids into proteins. The functionality that can be directed into proteins has been further expanded by encoding amino acids that can undergo subsequent, specific chemical reactions. Genetically encoded unnatural amino acids now allow protein interactions to be defined *in vivo*; the role of post-translational modifications to be defined; protein function to be activated in milliseconds inside live cells to dissect biological roles; and conformational changes of proteins to be followed in real time.

Emerging approaches allow the incorporation of multiple distinct unnatural amino acids into proteins[139,140], and these should facilitate the development of a whole new set of applications to biological problems through, for example, the encoding of FRET pairs for imaging applications. In addition, the incorporation of unnatural amino acids in animals[11] may also allow the real-time dissection of molecular processes, including development and neural processing, *in vivo*.


1. Johnson, J. A., Lu, Y. Y., Van Deventer, J. A. & Tirrell, D. A. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr. Opin. Chem. Biol.* **14**, 774–780 (2010).
2. Xie, J. & Schultz, P. G. A chemical toolkit for proteins — an expanded genetic code. *Nature Rev. Mol. Cell Biol.* **7**, 775–782 (2006).
3. Chin, J. W. *et al.* An expanded eukaryotic genetic code. *Science* **301**, 964–967 (2003).
4. Hancock, S. M., Uprety, R., Deiters, A. & Chin, J. W. Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *J. Am. Chem. Soc.* **132**, 14819–14824 (2010).
5. Mukai, T. *et al.* Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* **371**, 818–822 (2008).
6. Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding $N^{\varepsilon}$-acetyllysine in recombinant proteins. *Nature Chem. Biol.* **4**, 232–234 (2008). **Demonstrates, for the first time, that Lys acetylation can be site-specifically encoded in proteins.**
7. Gautier, A. *et al.* Genetically encoded photocontrol of protein localization in mammalian cells. *J. Am. Chem. Soc.* **132**, 4086–4088 (2010).
8. Xie, J. M. & Schultz, P. G. An expanding genetic code. *Methods* **36**, 227–238 (2005).
9. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **99**, 11020–11024 (2002). **Describes the first genetically encoded photocrosslinking between the monomers of a glutathione *S*-transferase dimer in *E. coli*.**
10. Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **126**, 14306–14307 (2004).
11. Greiss, S. & Chin, J. W. Expanding the genetic code of an animal. *J. Am. Chem. Soc.* **133**, 14196–14199 (2011). **Shows that it is possible to use genetic code expansion to specifically modify a protein in an animal.**
12. Chin, J. W. *et al.* Addition of *p*-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J. Am. Chem. Soc.* **124**, 9026–9027 (2002).
13. Ai, H.-W., Shen, W., Sagi, A., Chen, P. R. & Schultz, P. G. Probing protein–protein interactions with a genetically encoded photo-crosslinking amino acid. *Chembiochem* **12**, 1854–1857 (2011).
14. Chou, C., Uprety, R., Davis, L., Chin, J. W. & Deiters, A. Genetically encoding an aliphatic diazirine for protein photocrosslinking. *Chem. Sci.* **2**, 480–483 (2011).
15. Tippmann, E. M., Liu, W., Summerer, D., Mack, A. V. & Schultz, P. G. A genetically encoded diazirine photocrosslinker in *Escherichia coli*. *Chembiochem* **8**, 2210–2214 (2007).
16. Zhang, M. *et al.* A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. *Naure Chem. Biol.* **7**, 671–677 (2011).
17. Braig, D., Bar, C., Thumfart, J. O. & Koch, H. G. Two cooperating helices constitute the lipid-binding domain of the bacterial SRP receptor. *J. Mol. Biol.* **390**, 401–413 (2009).
18. Hino, N. *et al.* Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nature Methods* **2**, 201–206 (2005).
19. Ieva, R. & Bernstein, H. D. Interaction of an autotransporter passenger domain with BamA during its translocation across the bacterial outer membrane. *Proc. Natl Acad. Sci. USA* **106**, 19120–19125 (2009).
20. Mori, H. & Ito, K. Different modes of SecY–SecA interactions revealed by site-directed *in vivo* photo-cross-linking. *Proc. Natl Acad. Sci. USA* **103**, 16159–16164 (2006).
21. Okuda, S. & Tokuda, H. Model of mouth-to-mouth transfer of bacterial lipoproteins through inner membrane LolC, periplasmic LolA, and outer membrane LolB. *Proc. Natl Acad. Sci. USA* **106**, 5877–5882 (2009).




© 2012 Macmillan Publishers Limited. All rights reserved

22. Kaiser, C. M. *et al.* Real-time observation of trigger factor function on translating ribosomes. *Nature* **444**, 455–460 (2006).
23. Lakshmipathy, S. K. *et al.* Identification of nascent chain interaction sites on trigger factor. *J. Biol. Chem.* **282**, 12186–12193 (2007).
24. Haslberger, T. *et al.* M domains couple the ClpB threading motor with the DnaK chaperone activity. *Mol. Cell* **25**, 247–260 (2007).
25. Schlieker, C. *et al.* Substrate recognition by the AAA$^+$ chaperone ClpB. *Nature Struct. Mol. Biol.* **11**, 607–615 (2004).
26. Liu, C. *et al.* Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463**, 197–202 (2010).
27. Boos, D., Kuffer, C., Lenobel, R., Korner, R. & Stemmann, O. Phosphorylation-dependent binding of cyclin B1 to a Cdc6-like domain of human separase. *J. Biol. Chem.* **283**, 816–823 (2008).
28. Tagami, S. *et al.* Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. *Nature* **468**, 978–982 (2010).
29. Mohibullah, N. & Hahn, S. Site-specific cross-linking of TBP *in vivo* and *in vitro* reveals a direct functional interaction with the SAGA subunit Spt3. *Genes Dev.* **22**, 2994–3006 (2008).
30. Chen, H. T., Warfield, L. & Hahn, S. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nature Struct. Mol. Biol.* **14**, 696–703 (2007).
31. Yamano, K., Tanaka-Yamano, S. & Endo, T. Tom7 regulates Mdm10-mediated assembly of the mitochondrial import channel protein Tom40. *J. Biol. Chem.* **285**, 41222–41231 (2010).
32. Carvalho, P., Stanley, A. M. & Rapoport, T. A. Retrotranslocation of a misfolded luminal ER protein by the ubiquitin-ligase Hrd1p. *Cell* **143**, 579–591 (2010).
33. Hino, N. *et al.* Genetic incorporation of a photo-crosslinkable amino acid reveals novel protein complexes with GRB2 in mammalian cells. *J. Mol. Biol.* **406**, 343–353 (2011).
34. Zimmer, J., Nam, Y. & Rapoport, T. A. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455**, 936–943 (2008).
35. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nature Biotech.* **21**, 255–261 (2003).
36. Neumann, H., Hazen, J. L., Weinstein, J., Mehl, R. A. & Chin, J. W. Genetically encoding protein oxidative damage. *J. Am. Chem. Soc.* **130**, 4028–4033 (2008).
37. Liu, C. C. & Schultz, P. G. Recombinant expression of selectively sulfated proteins in *Escherichia coli*. *Nature Biotech.* **24**, 1436–1440 (2006).
38. Serwa, R. *et al.* Chemoselective staudinger-phosphite reaction of azides for the phosphorylation of proteins. *Angew. Chem. Int. Ed. Engl.* **48**, 8234–8239 (2009).
39. Xie, J. M., Supekova, L. & Schultz, P. G. A genetically encoded metabolically stable analogue of phosphotyrosine in *Escherichia coli*. *ACS Chem. Biol.* **2**, 474–478 (2007).
40. Park, H.-S. *et al.* Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science* **333**, 1151–1154 (2011).
41. Ai, H. W., Lee, J. W. & Schultz, P. G. A method to site-specifically introduce methyllysine into proteins in *E. coli*. *Chem. Commun. (Camb.)* **46**, 5506–5508 (2010).
42. Groff, D., Chen, P. R., Peters, F. B. & Schultz, P. G. A genetically encoded epsilon-*N*-methyl lysine in mammalian cells. *Chembiochem* **11**, 1066–1068 (2010).
43. Wang, Y. S. *et al.* A genetically encoded photocaged $N^{\varepsilon}$-methyl-L-lysine. *Mol. Biosyst.* **6**, 1557–1560 (2010).
44. Nguyen, D. P., Garcia Alai, M. M., Kapadnis, P. B., Neumann, H. & Chin, J. W. Genetically encoding $N^{\varepsilon}$-methyl-L-lysine in recombinant histones. *J. Am. Chem. Soc.* **131**, 14194–14195 (2009).
45. Nguyen, D. P., Garcia Alai, M. M., Virdee, S. & Chin, J. W. Genetically directing ε-*N*, *N*-dimethyl-L-lysine in recombinant histones. *Chem. Biol.* **17**, 1072–1076 (2010).
46. Virdee, S. *et al.* Traceless and site-specific ubiquitination of recombinant proteins. *J. Am. Chem. Soc.* **133**, 10708–10711 (2011).
47. Virdee, S., Ye, Y., Nguyen, D. P., Komander, D. & Chin, J. W. Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. *Nature Chem. Biol.* **6**, 750–757 (2010).
**References 46 and 47 demonstrate the site-specific ubiquitylation of recombinant proteins. Reference 47 also demonstrates the potential of this approach for elucidating the specificity of deubiquitinases and solving the structures of ubiquitylated proteins.**
48. Li, X., Fekner, T., Ottesen, J. J. & Chan, M. K. A pyrrolysine analogue for site-specific protein ubiquitination. *Angew. Chem. Int. Ed. Engl.* **48**, 9184–9187 (2009).
49. Liu, C. C. *et al.* Protein evolution with an expanded genetic code. *Proc. Natl Acad. Sci. USA* **105**, 17688–17693 (2008).
50. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
51. Kim, J. *et al.* Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep.* **7**, 397–403 (2006).
52. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870–1881 (2003).
53. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
54. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
55. Zhao, S. M. *et al.* Regulation of cellular metabolism by protein lysine acetylation. *Science* **327**, 1000–1004 (2010).
56. Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834–840 (2009).
57. Neumann, H. *et al.* A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol. Cell.* **36**, 153–163 (2009).
**Uses site-specific histone acetylation and single-molecule FRET studies to examine how acetylation of H3K56 can influence the conformation of DNA.**
58. Lammers, M., Neumann, H., Chin, J. W. & James, L. C. Acetylation regulates cyclophilin A catalysis, immunosuppression and HIV isomerization. *Nature Chem. Biol.* **6**, 331–337 (2010).
**Reports site-specific acetylation of cyclophilin A and structural insights into the interactions between cyclophilin A and cyclosporine, and cyclophilin A and HIV capsid, using X-ray crystallography.**
59. Arbely, E. *et al.* Acetylation of lysine 120 of p53 endows DNA-binding specificity at effective physiological salt concentration. *Proc. Natl Acad. Sci. USA* **108**, 8251–8256 (2011).
60. Oppikofer, M. *et al.* A dual role of H4K16 acetylation in the establishment of yeast silent chromatin. *EMBO J.* **30**, 2610–2621 (2011).
61. Thao, S., Chen, C.-S., Zhu, H. & Escalante-Semerena, J. C. $N^{\varepsilon}$-lysine acetylation of a bacterial transcription factor inhibits its DNA-binding activity. *PLoS ONE* **5**, e15123 (2010).
62. Celic, I. *et al.* The sirtuins Hst3 and Hst4p preserve genome integrity by controlling histone H3 lysine 56 deacetylation. *Curr. Biol.* **16**, 1280–1289 (2006).
63. Chen, C.-C. *et al.* Acetylated lysine 56 on histone H3 drives chromatin assembly after repair and signals for the completion of repair. *Cell* **134**, 231–243 (2008).
64. Das, C., Lucia, M. S., Hansen, K. C. & Tyler, J. K. CBP/p300-mediated acetylation of histone H3 on lysine 56. *Nature* **459**, 113–117 (2009).
65. Driscoll, R., Hudson, A. & Jackson, S. P. Yeast Rtt109 promotes genome stability by acetylating histone H3 on lysine 56. *Science* **315**, 649–652 (2007).
66. Garcia, B. A. *et al.* Organismal differences in post-translational modifications in histones H3 and H4. *J. Biol. Chem.* **282**, 7641–7655 (2007).
67. Han, J. *et al.* Rtt109 acetylates histone H3 Lysine 56 and functions in DNA replication. *Science* **315**, 653–655 (2007).
68. Hyland, E. M. *et al.* Insights into the role of histone H3 and histone H4 core modifiable residues in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **25**, 10060–10070 (2005).
69. Li, Q. *et al.* Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. *Cell* **134**, 244–255 (2008).
70. Masumoto, H., Hawke, D., Kobayashi, R. & Verreault, A. A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature* **436**, 294–298 (2005).
71. Ozdemir, A. *et al.* Characterization of lysine 56 of histone H3 as an acetylation site in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280**, 25949–25952 (2005).
72. Rufiange, A., Jacques, P.-E., Bhat, W., Robert, F. & Nourani, A. Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. *Mol. Cell* **27**, 393–405 (2007).
73. Williams, S. K., Truong, D. & Tyler, J. K. Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. *Proc. Natl Acad. Sci. USA* **105**, 9000–9005 (2008).
74. Xie, W. *et al.* Histone H3 lysine 56 acetylation is linked to the core transcriptional network in human embryonic stem cells. *Mol. Cell* **33**, 417–427 (2009).
75. Xu, F., Zhang, K. & Grunstein, M. Acetylation in histone h3 globular domain regulates gene expression in yeast. *Cell* **121**, 375–385 (2005).
76. Xu, F., Zhang, Q., Zhang, K., Xie, W. & Grunstein, M. Sir2 deacetylates histone H3 lysine 56 to regulate telomeric heterochromatin structure in yeast. *Mol. Cell* **27**, 890–900 (2007).
77. Yang, B., Miller, A. & Kirchmaier, A. L. HST3/HST4-dependent deacetylation of lysine 56 of histone H3 in silent chromatin. *Mol. Biol. Cell* **19**, 4993–5005 (2008).
78. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).
79. Koopmans, W. J. A., Buning, R., Schmidt, T. & van Noort, J. spFRET using alternating excitation and FCS reveals progressive DNA unwrapping in nucleosomes. *Biophys. J.* **97**, 195–204 (2009).
80. Roehrl, M. H. A. *et al.* Selective inhibition of calcineurin–NFAT signaling by blocking protein–protein interaction with small organic molecules. *Proc. Natl Acad. Sci. USA* **101**, 7554–7559 (2004).
81. Franke, E. K., Yuan, H. E. H. & Luban, J. Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* **372**, 359–362 (1994).
82. Thali, M. *et al.* Functional association of cyclophilin A with HIV-1 virions. *Nature* **372**, 363–365 (1994).
83. Chen, Z. J. & Sun, L. J. Nonproteolytic functions of ubiquitin in cell signaling. *Mol. Cell* **33**, 275–286 (2009).
84. Nguyen, D. P., Elliott, T., Holt, M., Muir, T. W. & Chin, J. W. Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. *J. Am. Chem. Soc.* **133**, 11418–11421 (2011).
85. McGinty, R. K. *et al.* Structure-activity analysis of semisynthetic nucleosomes: mechanistic insights into the stimulation of Dot1L by ubiquitylated histone H2B. *ACS Chem. Biol.* **4**, 958–968 (2009).
86. Castaneda, C. *et al.* Nonenzymatic assembly of natural polyubiquitin chains of any linkage composition and isotopic labeling scheme. *J. Am. Chem. Soc.* **133**, 17855–17868 (2011).
87. Kumar, K. S. A., Spasser, L., Erlich, L. A., Bavikar, S. N. & Brik, A. Total chemical synthesis of di-ubiquitin chains. *Angew. Chem. Int. Ed. Engl.* **49**, 9126–9131 (2010).
88. Ajish Kumar, K. S., Haj-Yahya, M., Olschewski, D., Lashuel, H. A. & Brik, A. Highly efficient and chemoselective peptide ubiquitylation. *Angew. Chem. Int. Ed. Engl.* **48**, 8090–8094 (2009).
89. Xu, P. *et al.* Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell* **137**, 133–145 (2009).
90. Chatterjee, C., McGinty, R. K., Pellois, J.-P. & Muir, T. W. Auxiliary-mediated site-specific peptide ubiquitylation. *Angew. Chem. Int. Ed. Engl.* **46**, 2814–2818 (2007).
91. El Oualid, F. *et al.* Chemical synthesis of ubiquitin, ubiquitin-based probes, and diubiquitin. *Angew. Chem. Int. Ed. Engl.* **49**, 10149–10153 (2010).
92. McGinty, R. K., Kim, J., Chatterjee, C., Roeder, R. G. & Muir, T. W. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **453**, 812–816 (2008).
93. Yang, R., Pasunooti, K. K., Li, F., Liu, X.-W. & Liu, C.-F. Dual native chemical ligation at lysine. *J. Am. Chem. Soc.* **131**, 13592–13593 (2009).
94. Yang, R. L., Pasunooti, K. K., Li, F. P., Liu, X. W. & Liu, C. F. Synthesis of K48-linked diubiquitin using dual native chemical ligation at lysine. *Chem. Commun.* **46**, 7199–7201 (2010).
95. Tran, H., Hamada, F., Schwarz-Romond, T. & Bienz, M. Trabid, a new positive regulator of Wnt-induced transcription with preference for binding and cleaving K63-linked ubiquitin chains. *Genes Dev.* **22**, 528–542 (2008).




© 2012 Macmillan Publishers Limited. All rights reserved

96. Deiters, A., Groff, D., Ryu, Y., Xie, J. & Schultz, P. G. A genetically encoded photocaged tyrosine. *Angew. Chem. Int. Ed. Engl.* **45**, 2728–2731 (2006).
97. Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M. & Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nature Chem. Biol.* **3**, 769–772 (2007).
98. Gautier, A., Deiters, A. & Chin, J. W. Light-activated kinases enable temporal dissection of signaling networks in living cells. *J. Am. Chem. Soc.* **133**, 2124–2127 (2011).
**Uses a photocaged amino acid in MEK to allow specific photoactivation of a subnetwork in MAPK signalling pathway and dissection of key regulatory steps in this pathway.**
99. Chen, P. R. *et al.* A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chem. Int. Ed. Engl.* **48**, 4052–4055 (2009).
100. Cohen-Saidon, C., Cohen, A. A., Sigal, A., Liron, Y. & Alon, U. Dynamics and variability of ERK2 response to EGF in individual living cells. *Mol. Cell* **36**, 885–893 (2009).
101. Burack, W. R. & Sturgill, T. W. The activating dual phosphorylation of MAPK by MEK is nonprocessive. *Biochemistry* **36**, 5929–5933 (1997).
102. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
103. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. & Prasher, D. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–805 (1994).
104. Heim, R., Prasher, D. C. & Tsien, R. Y. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl Acad. Sci. USA* **91**, 12501–12504 (1994).
105. Giepmans, B. N. G., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217–224 (2006).
106. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nature Methods* **2**, 905–909 (2005).
107. Los, G. V. *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).
108. Keppler, A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nature Biotech.* **21**, 86–89 (2003).
109. Kosaka, N. *et al. In vivo* stable tumor-specific painting in various colors using dehalogenase-based protein-tag fluorescent ligands. *Bioconjug. Chem.* **20**, 1367–1374 (2009).
110. Gautier, A. *et al.* An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.* **15**, 128–136 (2008).
111. George, N., Pick, H., Vogel, H., Johnsson, N. & Johnsson, K. Specific labeling of cell surface proteins with chemically diverse compounds. *J. Am. Chem. Soc.* **126**, 8896–8897 (2004).
112. Zhou, Z., Koglin, A., Wang, Y., McMahon, A. P. & Walsh, C. T. An eight residue fragment of an acyl carrier protein suffices for post-translational introduction of fluorescent pantetheinyl arms in protein modification *in vitro* and *in vivo*. *J. Am. Chem. Soc.* **130**, 9925–9930 (2008).
113. Yin, J. *et al.* Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc. Natl Acad. Sci. USA* **102**, 15815–15820 (2005).
114. Fernandez-Suarez, M. *et al.* Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nature Biotech.* **25**, 1483–1487 (2007).
115. Uttamapinant, C. *et al.* A fluorophore ligase for site-specific protein labeling inside living cells. *Proc. Natl Acad. Sci. USA* **107**, 10914–10919 (2010).
116. Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E. & Ploegh, H. L. Sortagging: a versatile method for protein labeling. *Nature Chem. Biol.* **3**, 707–708 (2007).
117. Antos, J. M. *et al.* Site-specific *N*- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J. Am. Chem. Soc.* **131**, 10800–10801 (2009).
118. Liu, D. S. *et al.* Diels–Alder cycloaddition for fluorophore targeting to specific proteins inside living cells. *J. Am. Chem. Soc.* **134**, 792–795 (2011).
119. Griffin, B. A., Adams, S. R. & Tsien, R. Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281**, 269–272 (1998).
120. Halo, T. L., Appelbaum, J., Hobert, E. M., Balkin, D. M. & Schepartz, A. Selective recognition of protein tetraserine motifs with a cell-permeable, pro-fluorescent bis-boronic acid. *J. Am. Chem. Soc.* **131**, 438–439 (2008).
121. Hinner, M. J. & Johnsson, K. How to obtain labeled proteins and what to do with them. *Curr. Opin. Biotechnol.* **21**, 766–776 (2010).
122. Xie, J. M. *et al.* The site-specific incorporation of *p*-iodo-L-phenylalanine into proteins for structure determination. *Nature Biotech.* **22**, 1297–1301 (2004).
123. Walden, H. Selenium incorporation using recombinant techniques. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 352–357 (2010).
124. Cellitti, S. E. *et al. 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**, 9268–9281 (2008).
125. Deiters, A., Geierstanger, B. H. & Schultz, P. G. Site-specific *in vivo* labeling of proteins for NMR studies. *Chembiochem* **6**, 55–58 (2005).
126. Hammill, J. T., Miyake-Stoner, S., Hazen, J. L., Jackson, J. C. & Mehl, R. A. Preparation of site-specifically labeled fluorinated proteins for $^{19}$F-NMR structural characterization. *Nature Protoc.* **2**, 2601–2607 (2007).
127. Jackson, J. C., Hammill, J. T. & Mehl, R. A. 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**, 1160–1166 (2007).
128. Lampe, J. N., Brandman, R., Sivaramakrishnan, S. & de Montellano, P. R. O. Two-dimensional NMR and all-atom molecular dynamics of cytochrome P450 CYP119 reveal hidden conformational substates. *J. Biol. Chem.* **285**, 9594–9603 (2010).
129. Li, C. G. *et al.* Protein $^{19}$F NMR in *Escherichia coli*. *J. Am. Chem. Soc.* **132**, 321–327 (2010).
130. Wang, L., Brock, A. & Schultz, P. G. Adding L-3-(2-naphthyl)alanine to the genetic code of *E. coli*. *J. Am. Chem. Soc.* **124**, 1836–1837 (2002).
131. Lee, H. S., Guo, J., Lemke, E. A., Dimla, R. D. & Schultz, P. G. Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in *Saccharomyces cerevisiae*. *J. Am. Chem. Soc.* **131**, 12921–12923 (2009).
132. Wang, J., Xie, J. & Schultz, P. G. A genetically encoded fluorescent amino acid. *J. Am. Chem. Soc.* **128**, 8738–8739 (2006).
133. Summerer, D. *et al.* A genetically encoded fluorescent amino acid. *Proc. Natl Acad. Sci. USA* **103**, 9785–9789 (2006).
134. Chapman, E., Farr, G. W., Furtak, K. & Horwich, A. L. A small molecule inhibitor selective for a variant ATP-binding site of the chaperonin GroEL. *Bioorg. Med. Chem. Lett.* **19**, 811–813 (2009).
135. Schultz, K. C. *et al.* A genetically encoded infrared probe. *J. Am. Chem. Soc.* **128**, 13984–13985 (2006).
136. Charbon, G. *et al.* Localization of GroEL determined by *in vivo* incorporation of a fluorescent amino acid. *Bioorg. Med. Chem. Lett.* **21**, 6067–6070 (2011).
137. Ye, S. X., Huber, T., Vogel, R. & Sakmar, T. P. FTIR analysis of GPCR activation using azido probes. *Nature Chem. Biol.* **5**, 397–399 (2009).
138. Ye, S. X. *et al.* Tracking G-protein-coupled receptor activation using genetically encoded infrared probes. *Nature* **464**, 1386–1389 (2010).
**Describes the use of infrared probes to follow the conformational changes that occur in structural intermediates of the GPCR rhodopsin during activation in response to light.**
139. Neumann, H., Slusarczyk, A. L. & Chin, J. W. *De novo* generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs. *J. Am. Chem. Soc.* **132**, 2142–2144 (2010).
140. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441–444 (2010).
**References 139 and 140 detail approaches for encoding more than one unnatural amino acid through engineering the ribosome and generating new synthetase tRNA pairs.**
141. Muralidharan, V. & Muir, T. W. Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nature Methods* **3**, 429–438 (2006).
142. Vila-Perelló, M. & Muir, T. W. Biological applications of protein splicing. *Cell* **143**, 191–200 (2010).
143. Boyce, M. & Bertozzi, C. R. Bringing chemistry to life. *Nature Methods* **8**, 638–642 (2011).
144. Zhang, Z. *et al.* A new strategy for the site-specific modification of proteins *in vivo*. *Biochemistry* **42**, 6735–6746 (2003).
145. Ye, S. *et al.* Site-specific incorporation of keto amino acids into functional G protein-coupled receptors using unnatural amino acid mutagenesis. *J. Biol. Chem.* **283**, 1525–1533 (2008).
146. Tsao, M. L., Tian, F. & Schultz, P. G. Selective Staudinger modification of proteins containing *p*-azidophenylalanine. *Chembiochem* **6**, 2147–2149 (2005).
147. Tian, F., Tsao, M.-L. & Schultz, P. G. A phage display system with unnatural amino acids. *J. Am. Chem. Soc.* **126**, 15962–15963 (2004).
148. Brustad, E. M., Lemke, E. A., Schultz, P. G. & Deniz, A. A. A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *J. Am. Chem. Soc.* **130**, 17664–17665 (2008).
149. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted 3 + 2 azide-alkyne cycloaddition for covalent modification of blomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046–15047 (2004).
150. Yanagisawa, T. *et al.* Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(varepsilon)-(*o*-Azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol.* **15**, 1187–1197 (2008).
151. Wang, J. *et al.* A biosynthetic route to photoclick chemistry on proteins. *J. Am. Chem. Soc.* **132**, 14812–14818 (2010).
152. Wang, Y., Song, W., Hu, W. J. & Lin, Q. Fast alkene functionalization *in vivo* by photoclick chemistry: HOMO lifting of nitrile imine dipoles. *Angew. Chem. Int. Ed. Engl.* **48**, 5330–5333 (2009).
153. Song, W., Wang, Y., Qu, J. & Lin, Q. Selective functionalization of a genetically encoded alkene-containing protein via "photoclick chemistry" in bacterial cells. *J. Am. Chem. Soc.* **130**, 9654–9655 (2008).
154. Song, W., Wang, Y., Qu, J., Madden, M. M. & Lin, Q. A photoinducible 1,3-dipolar cycloaddition reaction for rapid, selective modification of tetrazole-containing proteins. *Angew. Chem. Int. Ed. Engl.* **47**, 2832–2835 (2008).
155. Nguyen, D. P. *et al.* Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNA$_{CUA}$ pair and click chemistry. *J. Am. Chem. Soc.* **131**, 8720–8721 (2009).
156. Fekner, T., Li, X., Lee, M. M. & Chan, M. K. A pyrrolysine analogue for protein click chemistry. *Angew. Chem. Int. Ed. Engl.* **48**, 1633–1635 (2009).
157. Deiters, A., Cropp, T. A., Summerer, D., Mukherji, M. & Schultz, P. G. Site-specific PEGylation of proteins containing unnatural amino acids. *Bioorg. Med. Chem. Lett.* **14**, 5743–5745 (2004).
158. Carrico, Z. M., Romanini, D. W., Mehl, R. A. & Francis, M. B. Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem. Commun. (Camb.)* 1205–1207 (2008).
159. Wang, L., Zhang, Z., Brock, A. & Schultz, P. G. Addition of the keto functional group to the genetic code of *Escherichiacoli*. *Proc. Natl Acad. Sci. USA* **100**, 56–61 (2003).
160. Mehl, R. A. *et al.* Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **125**, 935–939 (2003).
161. Fleissner, M. R. *et al.* Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proc. Natl Acad. Sci. USA* **106**, 21637–21642 (2009).
162. Zeng, H., Xie, J. & Schultz, P. G. Genetic introduction of a diketone-containing amino acid into proteins. *Bioorg. Med. Chem. Lett.* **16**, 5356–5359 (2006).
163. Huang, Y. *et al.* Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications. *Bioorg. Med. Chem. Lett.* **20**, 878–880 (2010).
164. Deiters, A. *et al.* Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*. *J. Am. Chem. Soc.* **125**, 11782–11783 (2003).
165. Deiters, A. & Schultz, P. G. *In vivo* incorporation of an alkyne into proteins in *Escherichia coli*. *Bioorg. Med. Chem. Lett.* **15**, 1521–1524 (2005).




© 2012 Macmillan Publishers Limited. All rights reserved

166. Plass, T., Milles, S., Koehler, C., Schultz, C. & Lemke, E. A. Genetically encoded copper-free click chemistry. *Angew. Chem. Int. Ed. Engl.* **50**, 3878–3881 (2011).
167. Tong, G. J., Hsiao, S. C., Carrico, Z. M. & Francis, M. B. Viral capsid DNA aptamer conjugates as multivalent cell-targeting vehicles. *J. Am. Chem. Soc.* **131**, 11174–11178 (2009).
168. Kodama, K. *et al.* Regioselective carbon–carbon bond formation in proteins with palladium catalysis; new protein chemistry by organometallic chemistry. *Chembiochem* **7**, 134–139 (2006).
169. Kodama, K. *et al.* Site-specific functionalization of proteins by organopalladium reactions. *Chembiochem* **8**, 232–238 (2007).
170. Spicer, C. D. & Davis, B. G. Palladium-mediated site-selective Suzuki–Miyaura protein modification at genetically encoded aryl halides. *Chem. Commun.* **47**, 1698–1700 (2011).
171. Brustad, E. *et al.* A genetically encoded boronate-containing amino acid. *Angew. Chem. Int. Ed. Engl.* **47**, 8220–8223 (2008).
172. Lang, K. *et al.* Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction *Nature Chem.* 5 Feb 2012 (doi:10.1038/nchem.1250).
**Describes rapid and site-specific fluorescent labelling of a protein on mammalian cells through genetic code expansion.**
173. Agard, N. J., Baskin, J. M., Prescher, J. A., Lo, A. & Bertozzi, C. R. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **1**, 644–648 (2006).
174. Devaraj, N. K., Weissleder, R. & Hilderbrand, S. A. Tetrazine-based cycloadditions: application to pretargeted live cell imaging. *Bioconjug. Chem.* **19**, 2297–2299 (2008).
175. Blackman, M. L., Royzen, M. & Fox, J. M. Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels–Alder reactivity. *J. Am. Chem. Soc.* **130**, 13518–13519 (2008).
176. Hengming, K. Similarities and differences between human cyclophilin A and other β-barrel structures: structural refinement at 1.63 Å resolution. *J. Mol. Biol.* **228**, 539–550 (1992).
177. Cook, W. J., Jeffrey, L. C., Carson, M., Chen, Z. & Pickart, C. M. Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2). *J. Biol. Chem.* **267**, 16467–16471 (1992).
178. Komander, D. *et al.* Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. *EMBO Rep.* **10**, 466–473 (2009).
179. Bremm, A., Freund, S. M. V. & Komander, D. Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase Cezanne. *Nature Struct. Mol. Biol.* **17**, 939–947 (2010).
180. Ohren, J. F. *et al.* Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. *Nature Struct. Mol. Biol.* **11**, 1192–1197 (2004).
181. Li, J., Edwards, P. C., Burghammer, M., Villa, C. & Schertler, G. F. X. Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol.* **343**, 1409–1438 (2004).
182. Scheerer, P. *et al.* Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497–502 (2008).

**Acknowledgements**
We are grateful to the UK Medical Research Council for funding(U105181009 and MC_UP_A024_1008) and to K. Lang and S. Hancock for help with the manuscript.

**Competing interests statement**
The authors declare no competing financial interests.

## DATABASES
**Protein Databank:** http://www.rcsb.org/pdb/home/home.do
1AAR | 1KX5 | 1GZM | 1S9J | 2CPL | 2JF5 | 2W9N | 2XEW | 2XK5 | 3DIN | 3DQB

## FURTHER INFORMATION
**Jason W. Chin's homepage:** http://www2.mrc-lmb.cam.ac.uk/group-leaders/a-to-g/j-chin

## SUPPLEMENTARY INFORMATION
**See online article:** S1 (table)

**ALL LINKS ARE ACTIVE IN THE ONLINE PDF**



© 2012 Macmillan Publishers Limited. All rights reserved