CHAPTER 6

Modulating Enzyme Activity via *Incorporation of Non-canonical Amino Acids*

JULES KIRK, THOMAS HUBER AND COLIN J. JACKSON*

Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia
*Email: colin.jackson@anu.edu.au

6.1 Introduction

Naturally occurring proteins are linear polypeptides that are generally comprised of the twenty canonical amino acids. There are two notable examples of additional genetically-encoded amino acids, selenocysteine[1] and pyrrolysine,[2] as well as a large array of post-translational modifications of amino acids, such as phosphorylation,[3] sulfonation[4] and glycosylation.[5,6] Proteins can adopt a variety of structures and spontaneously fold in a multitude of shapes and conformations that allow them to carry out their diverse physiological roles. These roles range from being purely structural proteins, to receptors, catalysts (enzymes), and regulatory proteins, such as transcription factors.

Enzymes, which are the focus of this chapter, have been studied for over a century owing to their remarkable ability to catalyse chemical reactions. They increase the rates of chemical reactions essential to sustain life and allow them to take place on biologically relevant timescales. Over many years, our understanding of enzymatic catalysis has become increasingly sophisticated, with concepts such as transition state theory, catalytic

Catalysis Series No. 32
Modern Biocatalysis: Advances Towards Synthetic Biological Systems
Edited by Gavin Williams and Mélanie Hall
© The Royal Society of Chemistry 2018
Published by the Royal Society of Chemistry, www.rsc.org

preorganization, conformational change and even quantum tunnelling, all contributing to our now relatively mature understanding of these extraordinary molecular machines.[7,8] That being said, the exact biophysical and biochemical basis of enzymatic catalysis is not without controversy. Rational engineering and design, which is the ultimate test of our understanding of protein structure and function, still has a failure rate that is too high for us to truly say that we understand their function. The use of protein mutagenesis is one of the best available approaches to expand our understanding, allowing us to probe the function of proteins by altering their structure.

Owing to their remarkable catalytic properties, enzymes rapidly became the focus of applied research and are frequently used in industrial and medical applications. For instance, enzymes have been used in the production of cheese and as digestive aids since the 19th century. Their use greatly increased through the development of recombinant gene technology, which allowed the facile production of large amounts of specific enzymes through fermentation. Large-scale production led to their widespread use in industry, for example as detergents such as subtilisin,[9] or in medicine, for example tissue plasminogen activator, which is produced recombinantly to break down blood clots.[10] The development of enzymes in biocatalysis has been greatly facilitated through the advent of engineering techniques, from rational design,[11] through to random approaches like directed evolution,[12] and onto computational protein design (see Chapters 4, 5 and 7).[13,14] This has resulted in an explosion in the use of designer enzymes, particularly in the industrial production of valuable chemical precursors.[15]

In parallel to intensive efforts to genetically engineer enzymes, in recent years there has also been increasing focus on new methodologies to chemically modify proteins. The lure of chemical modification has always been the ability to introduce functionalities that go beyond what is available among the canonical amino acids. Traditionally, the most commonly applied form of chemical modification utilizes thiol groups in a protein reacting with maleimides to form covalent bonds with other molecules.[16] This has been used to label proteins with dyes,[17] metal chelators that can act as paramagnetic lanthanide tags,[18] and metal chelators that allow the formation of entirely novel metalloenzymes.[19] A second frequently used avenue to chemically modify proteins takes advantage of reactive amine groups to allow covalent attachment of bulk chemicals. For instance, protein modification with polyethylene glycol (PEGylation) is widely used in the pharmaceutical industry,[20] because it has been shown to significantly increase the circulatory lifetime of biotherapeutics. Finally, the isoelectric point (pI) of proteins can be drastically modified by chemical modification of amino acids,[21] which can facilitate the electrostatic attachment of polymers to create hybrid proteins.[22]

Despite the advances in our understanding of enzyme function and the technical progression of our abilities to genetically engineer improved enzymes and modify them with new chemical functionalities, these areas of research have their limitations. In terms of advancing our understanding of

enzymes, the ubiquitous nature of the canonical amino acids makes it difficult to use particular amino acids as mechanistic probes, or ''handles'' in biophysical characterization; in genetic engineering we are fundamentally limited by the functionalities and reactivity inherent in the canonical amino acids; in chemical modification, we are limited by the reactivities present in the canonical amino acids and since thiol and amine groups are present in most/all proteins, it is impossible to make specific chemical modifications in complex mixtures.

Non-proteogenic or ''unnatural'' amino acid (UAA) mutagenesis was first developed as a method to allow global, or residue-specific replacement of amino acids either through natural metabolic incorporation of structural analogues,[23] or chemical misacylation of tRNA,[24] to yield proteins with UAAs.[25] This approach was improved by taking advantage of suppressor read-through, allowing site-specific incorporation of amino acids using bioorthogonal machinery derived from distantly related organisms.[26–28] UAA mutagenesis is a potential solution to many of the challenges one faces when engineering enzymes, because it allows the introduction of new bioorthogonal reactive groups or biophysically unique probes at specific positions. In the following sections, we provide a brief outline to the residue-specific and site-specific approaches, before highlighting recent prominent examples of the use of UAAs for the study and engineering of enzymes. Finally, we provide some perspectives for the field in coming years, focusing in particular on the challenges to be addressed and the opportunities that are available.

6.2 Residue-specific *vs.* Site-specific UAA Incorporation

6.2.1 Residue-specific UAA Incorporation

The purpose of residue-specific UAA incorporation is to replace a given endogenous amino acid with a specific UAA analogue. Pauling[29] proposed that compounds isosteric to a given amino acid may be bound and acylated to a cognate tRNA by the respective aaRS.[30,31] This phenomenon is observed where structurally similar aliphatic amino acids isoleucine, methionine and valine often compete with the cognate leucine amino acid for activation and acylation by leucyl-tRNA synthetase (LeuRS).[32] Further, mistranslation can be an adaptive response for organism survival: in the hyperthermophilic archaea *Aeropyrum pernix*, MetRS misacylates $tRNA^{Leu}$ with methionine at low temperatures because substituting leucine with methionine in the proteome increases the flexibility of proteins, which is advantageous at low temperatures.[33] Thus, by replacing a given amino acid with an isosteric UAA analogue, the UAA can often be successfully incorporated by the endogenous aminoacyl-tRNA synthetase (aaRS):tRNA pair for the particular amino acid.[34] Alternatively, chemical misacylation of tRNA can allow incorporation of UAAs that are unable to be charged to tRNA using native tRNA synthetases (aaRSs).[24] Because most amino acids are present at multiple positions

throughout a protein's primary sequence (and throughout the proteome), replacement of an amino acid in the cell or in cell-free translation media with an isosteric analogue results in global substitution of that amino acid throughout the protein.[35]

Selective Pressure Incorporation (SPI) is the most commonly used methodology for *in vivo* residue specific UAA incorporation.[36] The procedure is founded upon work in the late fifties, which successfully replaced methionine with the isosteric analogue selenomethionine throughout the *E. coli* proteome.[37] For selenomethionine incorporation, auxotrophic cells that have been genetically engineered to lack the biosynthetic pathways required for the synthesis of methionine are most commonly used, making protein translation dependent on the availability of methionine (or an isosteric analogue) in the growth medium. SPI has been further refined by recombinant DNA technology, which can amplify the expression of a target gene under the control of an artificial promoter.[38] This minimizes any detrimental effects to cell growth and is more economical because the isosteric amino acid analogue is supplied only when cells have reached the appropriate growth stage and overexpression of the recombinant protein has been initiated.[39] The combination of these metabolic engineering advances has allowed for almost any gene to be expressed in conditions where a canonical amino acid has been replaced with its isosteric analogue to allow residue-specific incorporation. It is important to recognize that the SPI methodology depends on an endogenous aaRS misacylating tRNA with a UAA. If this process is inefficient, low UAA incorporation is the result. This can be addressed by simple overexpression of specific aaRSs or through directed evolution or rational design of aaRS for improved UAA affinity.[40]

The primary advantage of residue specific incorporation is the ability to incorporate multiple amino acid substitutions throughout the protein sequence.[41] This multiplicity of substitutions can result in cumulative, or synergistic effects of the UAAs, yielding much larger improvements to certain properties than can be achieved with single amino acid substitutions. For example, this has been exploited to increase the thermal stability of coiled-coil proteins *via* introducing hydrophobic side chains at the helix interface.[42] However, multiple mutations in a protein can also be problematic because a single deleterious substitution may cause the enzyme to become structurally unstable or lose catalytic activity, and thereby masking the effects of other neutral or beneficial substitutions. In general, the probability of the protein misfolding and exhibiting reduced activity increases with the difference between the physical and chemical properties between the canonical amino acids and the UAA replacement.[35]

6.2.2 Site-specific UAA Incorporation

Owing to the lack of control over which amino acids are substituted with UAAs using residue-specific incorporation, site-specific UAA incorporation is a valuable alternative methodology. Site-specific UAA incorporation is

comparatively more involved than residue-specific incorporation in terms of the underlying methodology. It requires a ''bio-orthogonal pair'' of an UAA-specific aminoacyl-tRNA synthetase (uaaRS) and a cognate tRNA that recognises a nonsense or frameshift codon at which the UAA is inserted (Figure 6.1).[43] For example, a bio-orthogonal suppressor tRNA that is complementary to an amber stop codon will allow read-through of the TAG stop codon,[27,44] with UAA incorporation occurring due to the bioorthogonal tRNA having been aminoacylated with an UAA by a cognate bio-orthogonal aaRS.[45,46] tRNA–aaRS pairs from different domains of life are often bio-orthogonal and provide the platform to expand the genetic code of an organism with additional UAAs. The most commonly used approaches for the genetic encoding of UAAs in bacterial cells utilize mutants of the tyrosyl-RS and $tRNA^{Tyr}$ pair from the archaeon *Methanococcus jannaschii*, which do not cross-react in bacterial cells due to a unique recognition of the tRNA C1-G72 base-pair by *aa*RS.[47]

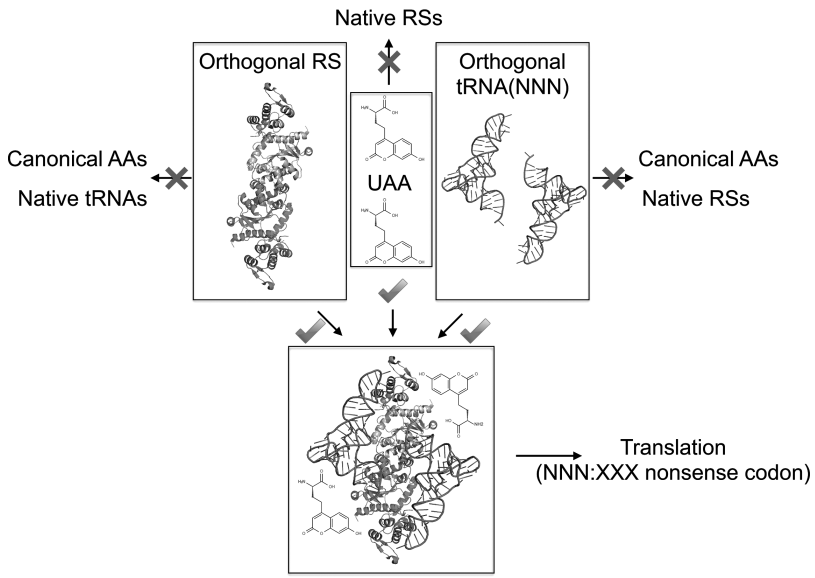

Figure 6.1 Orthogonal translational machinery for *in vivo* UAA incorporation. A tRNA synthetase from a distantly related organism is introduced to the cell and is bio-orthogonal because it does not interact with endogenous tRNA molecules, and is engineered to interact with UAAs in place of canonical/endogenous amino acids. An orthogonal tRNA from a distant related organism is introduced that is recognised by the orthogonal tRNA synthetase, but not recognised by endogenous tRNA synthetases. The tRNA is modified to possess a (typically) nonsense anticodon that will recognise a complementary nonsense codon. This orthogonal tRNA synthetase:tRNA pair then charges the orthogonal tRNA with the UAA; the nonsense codon (usually stop codon, such as UAG) results in suppression of translation termination, resulting in incorporation of the UAA.

6.2.3 *In Vitro* and *In Vivo* UAA Incorporation

Most of the work described above focuses on UAA incorporation *via* natural systems, such as mammalian cells,[48,49] yeast,[50] and prokaryotes.[27] *In vitro* site specific incorporation by cell-free protein synthesis is an alternative and increasingly popular avenue for mutagenesis.[51] The method is generally considered to be less time consuming and more economical, while simultaneously providing more control over components in the reaction mixture.[52] To this end, cell-free protein synthesis can be even applied in cases when the UAA is toxic to cells,[53] and is well suited when chemically misacylated tRNAs are used.[24] While exact details of the protocol may vary to satisfy individual goals, the essential steps are obtaining and adding all the components necessary for RNA transcription and protein translation to a reaction mixture containing an energy source and amino acids.[53]

6.3 Engineering tRNA Synthetases for UAA Incorporation

No discussion of the role of UAAs in enzymology would be complete without first covering how the enzymes that are responsible for incorporating UAAs have themselves been modified to carry out this role. Amino acids are 'charged' to their cognate tRNAs *via* a two-step process catalysed by aaRSs. The amino acid is first adenylated by an ATP molecule in the active site of the aaRS before the amino acid moiety of the resulting amino acid–AMP complex is transferred to the 2′ or 3′ hydroxyl group of the ribose moiety on the 3′ end of the terminal adenine of the tRNA molecule.[54] The various aaRSs and tRNAs form cognate pairs that each correspond to one of the twenty canonical amino acids. Specific and selective binding of amino acids in the aaRSs combined with an elaborate proof-reading process facilitates accurate tRNA aminoacylation and ensures translational fidelity.

One of the greatest barriers to UAA-incorporation is the requirement for the bio-orthogonal aaRS to recognise the UAA and aminoacylate the amber-suppressor tRNA. Recognition is not a major problem for isosteric structural analogues (although this can result in low specificity), but becomes problematic when the structure of the UAA diverges significantly from the natural amino acid. In these instances, it is necessary to use protein engineering to alter the substrate specificity of the aaRS. Directed evolution has been one of the most successful approaches, and begins with the generation of a large library of mutants that have been generated from a bio-orthogonal wild-type protein, for instance through the use of error-prone PCR.[55,56] Alternatively, computational design is initiated with docking and energy minimization programs, such as Rosetta.[57] Here, substrates can be superimposed on a known aaRS structure, as obtained through X-ray crystallography. Atomic interactions between residues around the active site and the substrate are then scored. Favourable interactions are hypothesized to increase binding

affinity to the UAA without negatively impacting on the protein, and mutations with low energy scores are then chosen for experimental testing.[58,59] These methods are thought to be complementary when designing aaRSs that are specific for UAA, because rational design can efficiently explore larger variations in sequence whereas directed evolution is very apt to improve on basal function.[60]

In the variant screening procedure developed by Schultz *et al.*,[61] mutant sequences go through a double sieve selection process to determine their function and translation fidelity *in vivo*. Initially, the antibiotic resistance of the bacteria assesses positive selection; resistance is achieved if the UAA-charged tRNA, aminoacylated by the mutant aaRS, effectively suppresses a nonsense or frameshift codon in the chloramphenicol acetyl transferase gene. Remaining sequences subsequently undergo negative selection. In a growth medium that contains only the 20 proteogenic amino acids, the uaaRS/$tRNA^{UAA}$ pair will be co-expressed with a toxic barnase gene containing nonsense or frameshift codons at permissive sites. If the engineered aaRS misacylates an endogenous amino acid to the $tRNA^{UAA}$, and the amino acid successfully fills the permissive sites in the toxic gene, cell death ensues.[46]

6.4 Enzyme Engineering with UAAs

6.4.1 UAAs for Increased Protein Thermostability

Since the first years of protein engineering, researchers have been seeking ways to increase the thermostability of enzymes, particularly for industrial applications and developing more robust and longer-lived enzymes for use in biocatalysis.[62] As with traditional protein mutagenesis and chemical modification, UAA incorporation has been used to increase protein stability (Figure 6.2).

The fluorous effect has been widely exploited in medicinal chemistry and industry to yield molecules with altered physicochemical properties, with one of the most prominent examples being Teflon.[63] Similarly, one of the most common applications of UAAs in enzyme engineering is the incorporation of fluorinated analogues.[64] Incorporation of fluorinated amino acids is particularly suited to residue-specific approaches because of the minimal structural changes that fluorination causes, which makes the isosteric amino acid analogues compatible with existing translation machinery, while the novel physicochemical properties of fluorinated UAAs can result in marked changes to the target enzymes. Several proteins containing one, two, or three different simultaneously fluorinated amino acids have been produced. For example, Budisa *et al.* have produced a lipase from *Thermoanaerobacter thermohydrosulfuricus* with (4*S*)-fluoroproline, 4-fluorophenylalanine, and 6-fluorotryptophan analogues in place of proline, phenylalanine and tryptophan, respectively.[65] They successfully incorporated monofluorinated amino acids at 24 different positions in a single

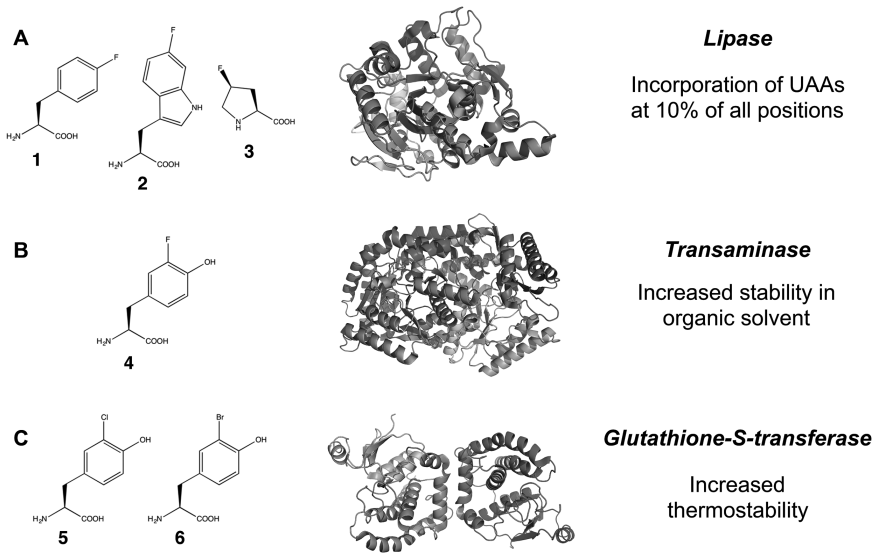

Figure 6.2 The use of UAA in engineering enzymes with enhanced stability. (A) Phenylalanine, tryptophan and proline were replaced by 4-fluorophenylalanine (**1**), 6-fluorotryptophan (**2**), (4*S*)-fluoroproline (**3**), respectively, in a lipase from *T. thermohydrosulfuricus* (represented by the structurally similar lipase from *Geobacillus* sp. SBS-4S, 3AUK),[115] resulting is incorporation of fluorinated UAAs at almost 10% of the sequence positions in the enzyme with no loss of stability. (B) Tyrosine was replaced by the UAA 3-fluorotyrosine (**4**) in a residue-specific manner in the *Vibrio fluvialis* ω-transaminase (4E3Q),[116] resulting in significantly enhanced stability in organic solvents. (C) 3-Chloro- (**5**) and 3-bromo-L-tyrosine (**6**) UAAs were used to replace native tyrosine residues in glutathione-*S*-transferase from *Schistosoma japonicum* (4WR4),[68] leading to significant increases in the thermostability of the enzyme.

expression, corresponding to approximately 10% of the amino acid sequence. Notably, the enzyme retained native structure and significant activity. Yun *et al.* have shown that, by non-specifically incorporating 3-fluorotyrosine into ω-transaminase in place of tyrosine, the fluorinated UAA was able to enhance the thermostability of the enzyme.[66] They also demonstrated that the effects of fluorotyrosine led to increased stability in organic solvent, for instance, increasing the half-life of the enzyme in 20% (v/v) DMSO from less than 10 hours to approximately 50 hours. However, in contrast to these positive reports of the effects of fluorinated UAAs on stability, Holzberger and Marx have shown that multifluorination of DNA polymerase resulted in very little change to the catalytic activity, but did cause loss of thermostability.[67]

Recently, Sakamoto *et al.* have shown that the thermostability of enzymes can also be enhanced by incorporating bulky halogenated UAAs, such as 3-chloro- and 3-bromo-L-tyrosines, at certain sites in glutathione *S*-transferase.[68] The level of stabilization that was achieved was significant

(up to 5.6 kcal mol^{-1}). In this study, the author explored the biophysical basis for the stabilizing effect through the use of X-ray crystallography, which showed that the additional steric bulk from the halogen moieties provided a stabilizing effect by filling internal cavities and forming additional van der Waals interactions. Filling internal cavities in proteins is a well-established mechanism of enzyme stabilization.[69] This mechanistic understanding of the stabilizing effect of these UAAs allowed the authors to extend this work to include the facile stabilization of an industrially relevant azoreductase in the same study.

Altogether, these examples reinforce two observations: (i) at certain positions, mutations can be highly destabilizing, and (ii) UAAs can also provide stabilizing effects in excess of those that can be conferred by canonical amino acids. This suggests that, as our knowledge of the fluoro-effect develops alongside the development of our understanding of the epistatic, context dependent, effects of mutations,[70,71] we are likely to see fluorinated amino acids incorporated more widely, particularly in industrial biocatalysts, owing to their ability to increase stability and resilience to harsh industrial environments.

6.4.2 UAAs for Increased Catalytic Efficiency

When it comes to enzyme engineering, one of the foremost motivations is to increase the activity of these catalysts. As such, there are thousands of reports of better combinations of canonical amino acids within and around the active site of an enzyme, which led to increased activity. A number of approaches have been utilized to achieve this, from rational design,[72] through directed evolution approaches,[73] and most recently, computational design (see Chapters 4 and 7).[13,14] However, these approaches are often limited by the intrinsic reactivity of canonical amino acids; for instance, serine, one of the most common nucleophiles in enzymes, is a relatively poor nucleophile at physiological pH unless it is part of a catalytic triad including histidine and aspartic or glutamic acid. When additional reactivity is required, enzymes have evolved the ability to incorporate non-covalently bound organic or inorganic cofactors, which complicate the development of designer enzymes due to the need to retain, synthesise or recycle these cofactors. UAA mutagenesis is a potential route to producing cofactor-free enzymes with unique chemical properties that can extend beyond those found in nature (Figure 6.3).

Using a residue-specific approach, Antranikian and colleagues have investigated UAA incorporation in a lipase from *Thermoanaerobacter thermohydrosulfiricus.*[74] They showed that global replacement of methionine residues throughout the enzyme with the UAA norleucine resulted in an ~10-fold increase in the enzymatic activity without thermal activation. This was suggested to be due to the significantly greater hydrophobicity of norleucine than that of methionine, which could enhance the hydrophobicity of the lid domain and interactions with hydrophobic lipids.

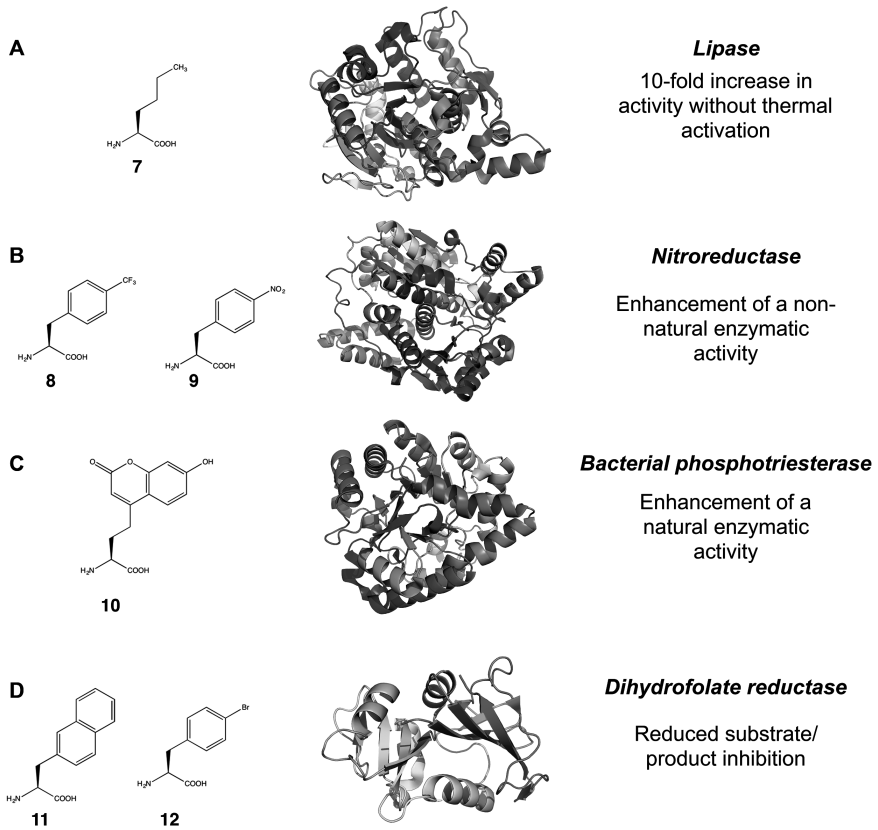

Figure 6.3 The use of UAA in engineering enzymes with improved catalytic activity. (A) Residue-specific replacement of methionine residues with the UAA norleucine (**7**) in a lipase from *T. thermohydrosulfuricus* (represented by the structurally similar lipase from *Geobacillus* sp. SBS-4S, 3AUK),[115] resulted in a 10-fold increase in catalytic activity without enzyme thermal activation. (B) In investigating whether UAA mutagenesis could lead to better catalytic activity than mutagenesis with naturally occurring amino acids, it was found that replacement of a phenylalanine amino acid by either *p*-trifluoromethylphenylalanine (**8**) or *p*-nitrophenylalanine (**9**) in a nitroreductase from *E. coli* (1IDT)[117] was significantly superior than mutation to Tyr, Lys or Asn. (C) When a tyrosine residue within the substrate binding site of a bacterial phosphotriesterase (2R1N)[118] was mutated to a hydroxycoumarinyl amino acid (**10**), the activity was increased to a larger extent that could be achieved by screening thousands of different randomly generated natural mutants of the enzyme. (D) A phenylalanine at position 31 was substituted by 2-napthylphenylalanine (**11**) and *para*-bromo-phenylalanine (**12**) in murine dihydrofolate reductase (1U70),[119] resulting in significantly lower substrate and product inhibition.

Site specific mutagenesis has also been used in enzyme engineering: Mehl and coworkers have shown that the use of UAAs can be advantageous in enzyme engineering and allows rapid improvement of enzyme activity.[75]

In that work, they studied a nitroreductase from *Escherichia coli* that can activate a prodrug used in cancer therapy, CB1954,[76] showing that improvement of the enzyme activity was not only possible (by >30-fold) using UAAs, but that UAAs also out-performed native amino acids at this position. Substitution of phenylalanine at position 124 with *p*-trifluoromethylphenylalanine or *p*-nitrophenylalanine was significantly superior than mutation to Tyr, Lys or Asn. The work also demonstrated that UAAs can lead to the development of enzymes that have greater catalytic efficiency than those that can be generated using the naturally occurring amino acids. Similarly, Fasan and colleagues have investigated the effect of UAAs on the regio-selectivity of cytochromes P450 using a site-specific approach.[77] The addition of *p*-aminophenylalanine, in particular, resulted in a 5-fold increase in the oxidation rate of (+)-nootkatone. Importantly, the functional changes that were observed could not be replicated through the use of any of the 20 canonical amino acids, reinforcing the idea that the UAAs provided new and otherwise inaccessible structure and function to these modified enzymes.

In contrast to the previous studies, which investigated enzymatic catalysis of synthetic substrates or substrates for which the enzyme had not evolved under strong selective pressure to turn over, recent work examined a naturally evolved and highly efficient ($k_{cat}/K_M \approx 10^7\ M^{-1}\,s^{-1}$) phosphotriesterase[78,79] with the ability to detoxify and protect against neurotoxic organophosphates.[80,81] This enzyme was engineered using UAAs to test whether highly efficient natural enzymes can be improved further through UAAs, *i.e.* whether the catalytic potential of the enzyme can be advanced beyond what is possible through natural evolution. Several efforts to increase the activity of the phosphotriesterase through rational design and directed evolution with canonical amino acids, effectively sampling hundreds of thousands of different mutations, yielded only a modest improvement in activity (4-fold) with the substrate paraoxon.[82] In contrast, a single tyrosine > hydroxycoumarinyl amino acid replacement at the substrate binding site yielded an order of magnitude greater activity, establishing that the catalytic limits observed in nature can indeed be exceeded through the use of UAA mutagenesis.[83]

Another area of interest in the context of increasing enzyme efficiency and usefulness involves reducing the level of substrate or product inhibition that enzymes undergo, as substrate/product inhibition has the potential to lessen substrate loading and/or reduce total substrate turnover in biocatalytic applications.[84] The unique attributes of UAAs also have potential in this area. Zheng and Kwon have used site-specific UAA incorporation into murine dihydrofolate reductase (mDHFR) as a model system to control the level of inhibition by methotrexate.[85] Structural analysis highlighted the role that phenylalanine 31 played in substrate and inhibitor binding. This residue was then mutated to *p*-bromophenylalanine and 2-napthylphenylalanine. These unnatural variants of mDHFR displayed interesting kinetic characteristics, with the *p*-bromophenylalanine and 2-napthylphenylalanine

variants exhibiting ~2-fold and ~4-fold higher dissociation constants (K_d) for the inhibitor methotrexate. Surprisingly, both variants also exhibited higher affinity as measured by the Michaelis constant (K_M) for the substrate dihydrofolate (6.5 μM for the wild-type *vs.* 3.4 and 4.8 μM for the respective variants). Although this was accompanied by a trade-off in turnover number (k_{cat}) in the case of the 2-napthylphenylalanine variant, the *p*-bromophenylalanine variant actually displayed enhanced substrate turnover leading to a 4-fold increase in the ratio of inhibitor affinity to substrate affinity (K_d/K_M).

6.4.3 UAAs to Alter Specificity and Selectivity

The desire to increase the catalytic efficiency of enzymes is partly academic—to understand how we can improve enzymes in a general sense—but also derives from a need to make more efficient catalysts for medical and industrial applications. However, any efforts to increase catalytic efficiency assume that the enzyme already has the necessary substrate specificity. In many medical and industrial applications, the target substrate is synthetic and no enzymes have naturally evolved to recognise it. Thus, there is significant interest in engineering enzymes for altered substrate specificity (Figure 6.4). A particular subfield within this topic is the intense research into modulating enzymatic stereo/regioselectivity, which is of particular relevance to the chemical and pharmaceutical industry, where production of the correct stereoisomer can be of particular importance.[86]

Using a residue-specific approach, Fasan and colleagues have investigated the effect of UAAs on the regioselectivity of cytochromes P450.[77] In this work, a number of different aromatic UAAs were introduced at the substrate-binding site and these variants were screened against (*S*)-ibuprofen and (+)-nootkatone. The addition of the UAAs in place of Ala82 dramatically increased the rate of oxidation of (+)-nootkatone to form an allylic alcohol (from 4% oxidation to 62% oxidation in the case of a *p*-acetylphenylalanine) and replacement of Ala78 with *p*-acetylphenylalanine catalysed the oxidation of (+)-nootkatone to (9*R*)-hydroxynootkatone with absolute stereoselectivity, when this reaction did not proceed at detectable rates with the native enzyme. Similarly, replacement of Leu75 with *p*-aminophenylalanine resulted in a significant increase in the oxidative conversion of (*S*)-ibuprofen to produce a tertiary alcohol (from 38% to 86%). Additional residue-specific work to alter enzyme specificity includes the global incorporation of *m*-fluorophenylalanine, replacing phenylalanine, in a lipase from *Thermoanaerobacter thermohydrosulfiricus*, which also resulted in increased substrate range,[74] with the lipase displaying greater activity with a range of triglyceride chains, both greater and shorter than the wild-type enzyme, which is relatively specific for C6–C8 substrates.

A notable example of the use of site-specific UAA mutagenesis to enhance the enantioselectivity of diketoreductase has been recently reported by Chen and colleagues.[87] Diketoreductase is an important biocatalyst that is capable of stereoselectively reducing β,δ-diketo esters to dihydroxy products.[88]

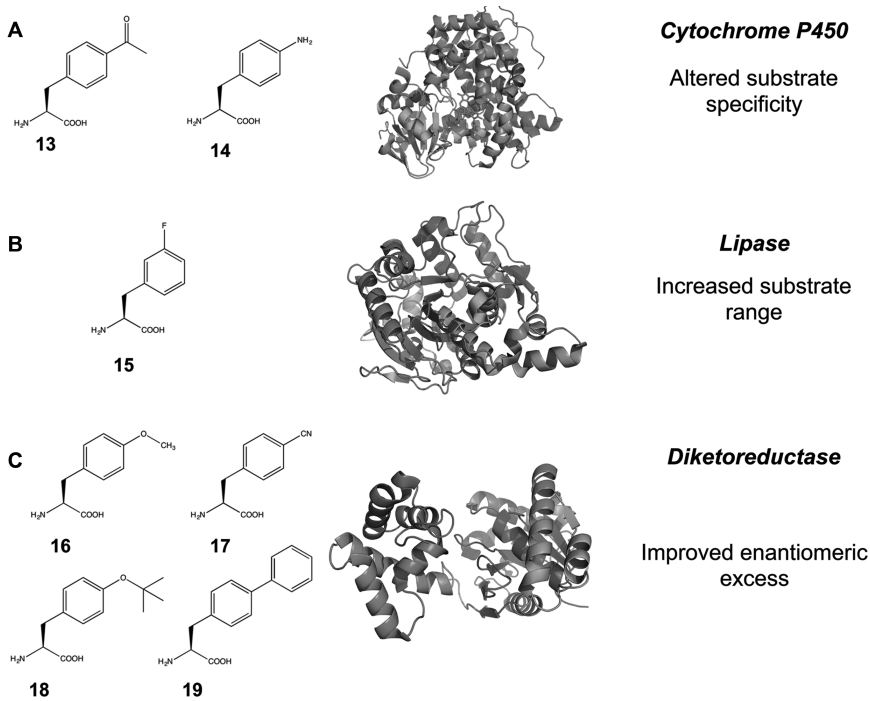

Figure 6.4 The use of UAA in engineering enzymes with altered substrate specificity. (A) The substrate specificity of a cytochrome P450 (2IJ2)[120] was altered by introducing the UAAs *para*-acetyl-phenylalanine (**13**) and *para*-amino-phenylalanine (**14**) into the active site of the enzyme. (B) The lipase from *T. thermohydrosulfuricus* (represented by the structurally similar lipase from *Geobacillus* sp. SBS-4S, 3AUK)[115] was engineered by the residue-specific substitution of phenylalanine with *m*-fluorophenylalanine (**15**), leading to an increase in the substrate range. (C) By introducing a series of bulky UAAs (4-methoxy-L-phenylalanine (**16**), 4-cyano-L-phenylalanine (**17**), *O-tert*-butyl-L-tyrosine (**18**) and 4-phenyl-L-phenylalanine (**19**)) in place of a native tryptophan at position 222 in a diketoreductase (4E12),[89] the enantiomeric excess of the enzyme could be altered.

By using structure-guided rational design, they pinpointed Trp149 and Trp222 as important residues for substrate binding.[89] Trp222 was replaced by several canonical amino acids (valine, leucine, methionine, phenylalanine, tyrosine) or UAAs (4-cyano-L-phenylalanine, 4-methoxy-L-phenylalanine, 4-phenyl-L-phenylalanine, *O-tert*-butyl-L-tyrosine). The kinetic behaviour of the variants revealed that the enantiomeric excess of the wild-type enzyme for the *R*-isomer of 2-chloro-1-phenylethanone could be increased from 9.1% for the wild-type enzyme to 10.5%, 29.7% and 33.7% when tryptophan was replaced, by 4-methoxy-L-phenylalanine, 4-phenyl-L-phenylalanine and *O-tert*-butyl-L-tyrosine, respectively. This is an excellent example of an UAA providing properties not otherwise accessible by the 20 canonical amino acids. In this case, the greater enantiomeric excess was a result of greater

steric bulk of the residue at this position as it controls access of the respective isomers into the active site. Thus, tryptophan, which is the largest canonical amino acid by volume (147.1 Å^3), could not be improved upon by any canonical amino acid, but could be more favourably replaced by the larger UAAs 4-methoxy-L-phenylalanine (166.7 Å^3), 4-phenyl-L-phenylalanine (181.9 Å^3) and *O-tert*-butyl-L-tyrosine (185.3 Å^3).

6.4.4 UAAs to Probe Enzyme Function and Mechanism

One of the most valuable applications of UAA mutagenesis and incorporation, given the practical challenges that still need to be overcome for broad utilization in industrial biocatalysis,[90] is for the UAAs to be used as mechanistic probes to allow us to better understand enzyme structure, function and catalytic mechanisms. One of the earliest examples of the use of UAAs to probe enzyme structure and function was from Dupureur *et al.*, who used residue-specific incorporation of *o*-, *m*- or *p*-fluorophenylalanine into the *Pvu*II restriction endonuclease.[91] Notably, the incorporation of *m*-fluorophenylalanine did not affect conformational stability but caused a doubling of specific activity. Analysis of the effects that substitutions that were remote from the active site had on the catalytic activity of the enzyme allowed the authors to conclude that relatively subtle changes in enzyme conformation due to remote mutations can have significant impact on endonuclease activity and suggests novel ways to influence catalytic behaviour (Figure 6.5).

In a recent and particularly elegant study, Boxer and colleagues have used UAA mutagenesis to dissect the proton delocalization in the hydrogen bond network of the model enzyme ketosteroid isomerase.[92] This work was carried out to examine the role of extended hydrogen bond networks in enzymes due to their common occurrence. Extending work that suggested quantum delocalization of protons across a triad of Tyr16, Tyr32, and Tyr57 residues with very close bonding distances and an unusually low pK_a value in the ketosteroid isomerase active site was an important factor in the understanding of the catalytic mechanism of the enzyme.[93,94] Using site-specific UAA mutagenesis, the tyrosine residues were replaced, separately with 3-chlorotyrosine analogues, to investigate the delocalization of protons in the triad (since the reduced pK_a value of the 3-chlorotyrosine disrupts the proton affinity balance, but not the H-bonding). Biophysical analysis with X-ray crystallography, UV-Vis spectroscopy and ^{13}C-NMR revealed that the proton delocalization within the triad was affected by the replacements, highlighting the complexity of enzymes and the role of quantum effects such as proton delocalization in enzyme catalysis.

Recently, Otting and colleagues have used site-specific UAA incorporation to introduce *O-tert*-butyltyrosine into the DnaB helicase from *Stearothermophilus.*[95] DnaB unwinds double stranded DNA during DNA replication and functions as part of the larger replisome complex.[96] *O-tert*-Butyltyrosine was particularly useful as a NMR probe, because the nine

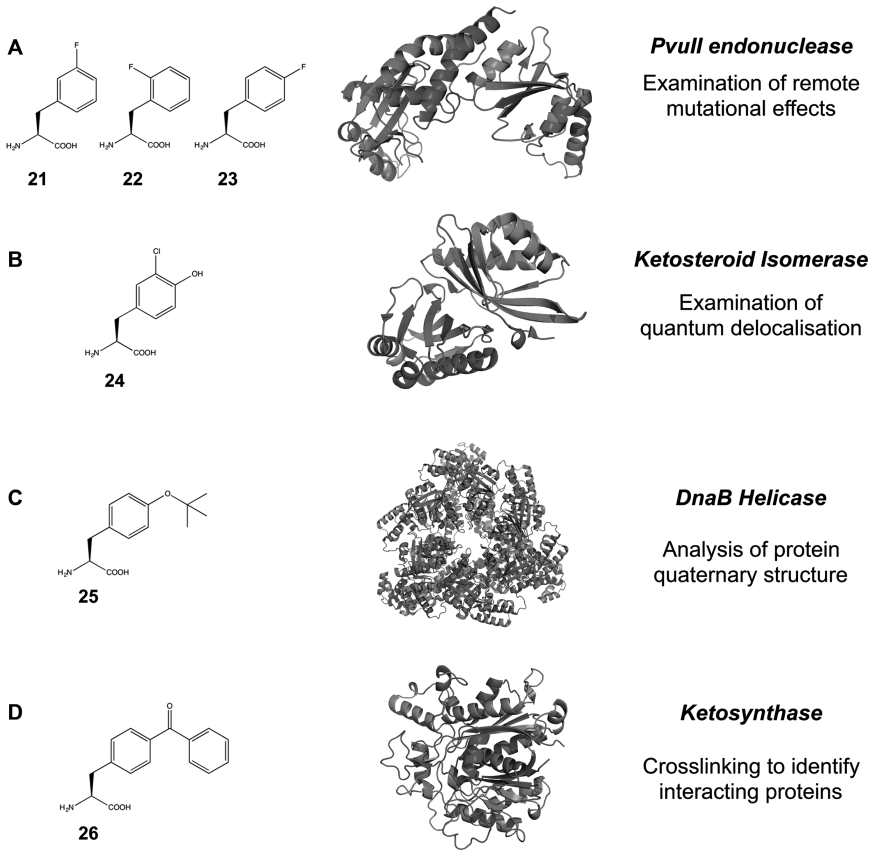

Figure 6.5 The use of UAAs in engineered enzymes to study structure and function. (A) The residue-specific incorporation of *o*-, *m*- or *p*-fluorophenylalanine (**21**, **22**, **23**) in place of phenylalanine in the *Pvu*II restriction endonuclease (1PVU)[121] allowed for analysis of subtle, remote, structural effects on enzyme activity. (B) In studying the contribution of quantum delocalization to the catalytic activity of a ketosteroid isomerase (5D82),[92] the replacement of certain tyrosine residues with 3-chlorotyrosine (**24**) was useful because the UAA has a significantly lower pK_a than the native amino acid, allowing selective disruption of the proton delocalization. (C) The use of the UAA *O-tert*-butyltyrosine (**25**) in the study of the DnaB helicase from *E. coli* (represented here by DnaB from *A. aeolicus*)[122] allowed the quaternary structure of the enzyme to studied by NMR spectroscopy. (D) Using *p*-benzoyl-L-phenylalanine (**26**), an UAA capable of undergoing photo-induced crosslinking, the interaction between acyl carrier proteins and ketosynthase (represented by the ketosynthase-chain length factor heterodimer; 1KAS)[123] could be studied in unprecedented detail.

protons of a solvent-exposed *tert*-butyl group produce a narrow and intense single signal which stands out from the background of other methyl resonances in the protein. This unique chemical property means that the

O-tert-butyl group can be easily observed in one-dimensional proton NMR spectra, eliminating the need for expensive isotope labeling and time consuming multi-dimensional NMR experiments. Remarkably, even though NMR is most often used for analysis of relatively small proteins (<50 kDa), the 320 kDa hexamer of DnaB could be shown to adopt 3-fold, rather than 6-fold, symmetry in solution.

6.4.5 UAAS to Control Enzyme Activity and Interactions

Just as UAAs can act as chemical tags and handles that allow us to gain a deeper understanding of a protein's structure and mechanism, they can also be used to control enzyme activity and interactions. This is of tremendous value in allowing us to better understand enzyme function in complex biological environments.

Photo-labile UAAs have found application in the study of enzyme function due to their ability to act as figurative ''cages'' for chemical reactivity: in their caged form they have no activity and it is only after the UAA is irradiated with UV light that it undergoes chemical decomposition to yield the active amino acid. Consequently, enzyme function can be tightly controlled through the use of photo-chemistry.[97,98] This approach has recently been used in enzymology, allowing the precise control of T7 RNA polymerase with a photocaged lysine derivative that was incorporated through the use of engineered, orthogonal pyrrolysine synthetase/tRNA pair in mammalian cells.[99] The ability to control T7 RNA polymerase activity allowed Hemphill *et al.* to control the transcription of short RNA hairpins that could subsequently be used for RNA interference of other genes, creating a powerful optogenetic system for light-controlled manipulation of biology. This work has been extended further with the development of small molecule caging approaches. In work by Deiters *et al.*, phosphine-mediated Staudinger reduction was used to uncage an unnatural azidobenzyloxycarbonyl amino acid to activate protein function.[100] This UAA was again genetically encoded using an orthogonal pyrrolysyl tRNA synthetase/tRNA pair in mammalian cells. In this application, the authors successfully demonstrated control over the activity of luciferase to induce fluorescence, as well CRISPR/Cas9 gene editing.

UAAs have also found a niche in the study of enzyme:protein interactions, primarily through their applicability to photo-cross-linking. For example, Williams *et al.* have used the photo-cross-linking UAA *p*-benzoyl-L-phenylalanine to probe the interactions between acyl carrier proteins and ketosynthase subunit in fatty acid synthase.[101] By using photo-cross-linking, a variety of acyl carrier proteins were tested, highlighting the role of residues within the conserved helix II region of the acyl carrier proteins. This work is notable because it demonstrated that UAAs are of particular value in moderate throughput screening experiments owing to their specificity and bioorthogonality.

6.4.6 UAAs for Chemical Modification of Enzymes

Protein chemical modification is of increasing interest to the enzymology community. For example, chemical modification of non-catalytic scaffolds can allow the attachment of reactive groups to generate a catalyst,[102] specific labelling of an enzyme with a fluorescent tag to monitor function,[103] and PEGylation to protect enzymes and increase their circulatory lifetime and stability.[104] One of the biggest challenges in the area of enzyme chemical modification is the need for specificity—it is often necessary to chemically modify a single protein in a complex mixture, or a single position in a protein. The development of new labelling methodologies through UAA mutagenesis appears to have become one of the largest sub-fields in the applications of UAAs to enzyme engineering. The value of UAAs in this approach is their bio-orthogonality: whereas almost all proteins include amine groups, and most include thiol groups, the addition of an UAA results in the incorporation of a unique amino acid, with reactivity not otherwise present among the proteome, thereby allowing specific labelling (Figure 6.6).[105]

One of the most commonly used chemical modifications of enzymes is PEGylation, which has been shown to increase the stability and circulatory lifetime of medically useful enzymes.[104] In one of the earliest examples of UAA incorporation, Schultz and colleagues have shown that *p*-azidophenylalanine, which can be incorporated in a site-specific manner in proteins, can be used in a $[3+2]$ cycloaddition reaction with an alkyne derivatised PEG reagent for PEGylation of superoxide dismutase,[106] building on previous work demonstrating that an azide–alkyne pair can be used in copper-mediated Huisgen $[3+2]$ cycloaddition.[107]

Similar work has also exploited copper-catalysed alkyne–azide cycloaddition click reactions to allow labelling of T4 lysozyme with fluorescent probes. By labelling one cysteine amino acid with a fluorescent dye using maleimide chemistry, and incorporating UAAs such as propargyllysine, which contains an alkyne functionality that reacts with azide-containing dyes,[105] two dyes could be attached that can then undergo Förster Resonance Energy Transfer, or FRET (previous work utilized *p*-acetylphenylalanine, which contains a ketone functionality that reacts with hydroxylamine-containing dyes)[108] to allow the enzyme to be studied through single molecule spectroscopy. This provided new insight into the folding landscape of T4 lysozyme.

Enzyme function in many biological situations is dependent on post-translational modification, such as phosphorylation or acetylation of important amino acids. Studying the effects of post-translational modification is extremely challenging owing to the heterogeneity of biological systems, where there is often a mixture of differently modified enzymes. To define the effects of the modification on enzyme activity, or to study the biological effects of the mutation, it is necessary to generate a homogeneous population of the modified protein. Thus, UAA incorporation is a powerful technique that allows specific amino acids within a protein to be replaced with an UAA that mimics or contains the post-translational modification, generating a

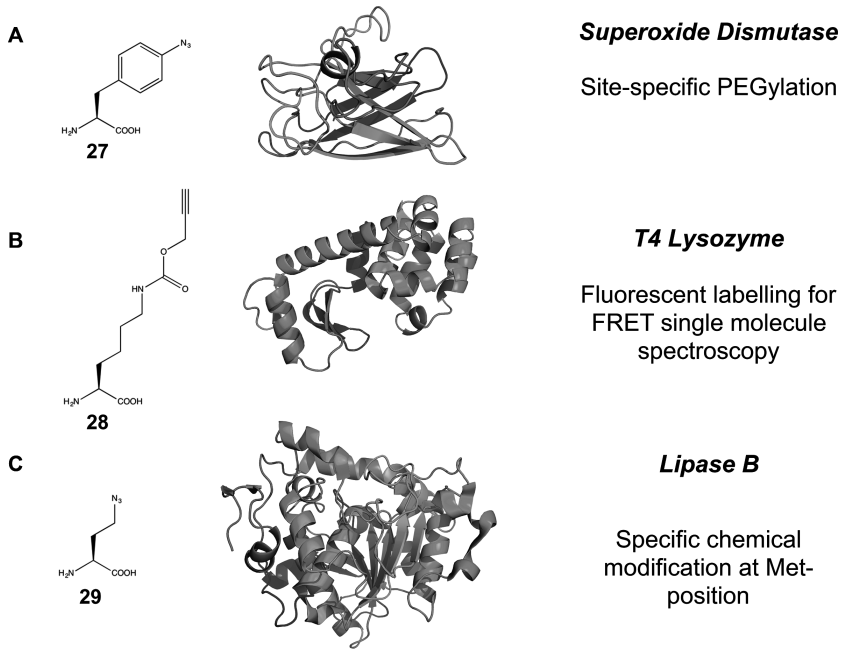

Figure 6.6 The use of UAA in engineering enzymes to facilitate chemical modification. (A) The site-specific introduction of *p*-azidophenylalanine (**27**) into superoxide dismutase (1SOS)[124] allowed site-specific PEGylation of the enzyme. (B) Single-molecule FRET spectroscopy of T4 lysozyme (2LZM)[125] was made possible by the site-specific attachment of fluorescent dyes at positions where the UAA propargyllysine (**28**) was incorporated. (C) Chemical modification at specific positions of lipase B from *Candida antarctica* (4ZV7)[126] was made possible through the residue-specific replacement of methionine by azidohomoalanine (**29**), exploiting the fact that only one methionine residue was in a solvent exposed position.

homogeneous sample. Successes in this area include the incorporation of *N*(ε)-acetyllysine in the enzyme manganese superoxide dismutase by Chin *et al.* using an orthogonal pyrrolysine synthetase/tRNA pair from *Methanosarcina barkeri* in *E. coli*.[109] This allowed them to measure and define the effect of lysine acetylation on manganese superoxide dismutase, revealing that the enzyme kinetics were essentially the same as the unacetylated form and suggesting acetylation must affect another aspect of its function. More recently, Chin *et al.* have engineered an orthogonal pathway for phosphoserine and demonstrated it was successfully incorporated into ubiquitin and the kinase Nek7.[110] The technological advance that allowed the challenging incorporation of phosphoserine involved exploiting the mechanism used by some methanogenic bacteria for cysteine incorporation, which involves aminoacylation of phosphoserine onto $tRNA^{Cys}_{GCA}$ by the tRNA synthetase SepRS (conversion of phosphoserine to cysteine happens subsequently).[111]

This orthogonal archael SepRS/tRNA pair was introduced into *E. coli* after converting the GCA anticodon to CUA, which significantly reduced the efficiency of amino acylation. Thus, extensive evolution and manipulation of the regions around the anticodon in the tRNA was performed to facilitate recognition and decoding of the tRNA by the bacterial ribosome, and SepRS was also subsequently evolved for improved recognition of the new anticodon.

Finally, van Hest and colleagues have used UAA mutagenesis to site-specifically modify the commonly used industrial lipase B from *Candida antarctica*.[112] By using residue-specific replacement of methionine by azidohomoalanine, they were able to carry out specific functionalisation of these UAAs. Notably, four methionines in this enzyme are buried in the hydrophobic core of the enzyme and are therefore not solvent accessible. The one remaining and solvent accessible methionine was therefore targeted for replacement by UAA mutagenesis. Again using a copper-catalysed cycloaddition reaction, a monofunctionalised, and active, enzyme could be obtained.

6.4.7 Engineering Enzymes for UAA Biosynthesis

One of the greatest hurdles to the large-scale use of UAAs is the need to chemically synthesise the majority of the UAAs that are incorporated, which vastly increases the cost of the experiment and requires that these UAAs be effectively fed to the organism or cell-free system into which they will be incorporated.[90] In contrast, most naturally occurring amino acids are inexpensively biosynthesised. To address this, Turner and colleagues have performed engineering of the bacterial ammonia lyase EncP, which allowed them to biosynthetically produce in high enantiomeric excess, a range of phenylalanine derivatives.[113] Similar work has been carried out to allow the whole-cell biosynthesis of L-homoalanine.[114] Although the primary focus of that work was to reduce the cost of L-homoalanine production and use as a chiral precursor for the pharmaceuticals levetiracetam, brivaracetam, and ethambutol, it demonstrated that engineering bacteria to produce additional UAAs is feasible. In that work, they altered the substrate specificity of glutamate dehydrogenase to work more efficiently with 2-ketobutyrate, rather than the natural substrate, 2-ketoglutarate. Through coexpression with a threonine dehydratase in a modified threonine-hyperproducing *Escherichia coli* strain, substantial amounts of L-homoalanine could be produced.

6.5 Future Perspectives

In writing this chapter, we were surprised at the relatively small number of examples of UAAs being used to improve our understanding of enzymes, as well as their catalytic activity and physical properties. In contrast to the phenomenal growth of enzyme engineering in general, the application of UAAs in this field appears to be lagging, despite the enormous potential that

they have. The reasons for this are probably two-fold: (i) the technique of UAA mutagenesis, especially site-specific UAA mutagenesis, is still significantly more demanding than traditional mutagenesis, which has led to its use being generally confined to research groups that have developed expertise in UAA mutagenesis; (ii) the approach is unfortunately limited by economic factors, especially the need to chemically synthesise UAAs and the intrinsically low yields of the enzymes that are produced through site-specific mutagenesis. This has mostly limited their application to academic research or proof-of-principle style studies that have, importantly, demonstrated the potential of UAA incorporation, or yielded new insight into enzyme function, but cannot be translated into industrial scale without difficulty, owing to the yield and cost.

There are therefore two areas that need particular attention in the coming years to further advance the use of UAAs in enzymology. First, the continuing efforts to simplify and make the technique widely available should see it become adopted by a wider range of research groups and industry, as a standard laboratory approach, rather than a specialist technique. Second, the economic limitations of the technique must continue to be addressed, perhaps through the development of biosynthetic pathways for UAAs to allow them to be produced as cheaply and simply, *in vivo*, as canonical amino acids. However, this requires the development of separate biosynthetic pathways for every amino acid, although this might be less demanding than it initially appears, since the majority of UAAs are derivatives of tyrosine and lysine.

6.6 Conclusion

In conclusion, UAA incorporation and site-specific mutagenesis are ideally suited to use in modern enzymology, from providing unique biochemical and biophysical handles and probes to advance our understanding of enzymes, through to more applied outcomes, such as exploiting the novel chemistry that can be incorporated into UAAs to push back the limits of catalysis and improving the physical properties of enzymes. The use of UAAs could potentially lead to the development of true designer enzymes with vastly superior catalytic power and biophysical robustness or versatility than we see in the natural repertoire of enzymes. However, to get to this point, more work is required to increase the usability of this technique and grow the size of the community of UAA–enzyme engineers, as well as to address some of the economic barriers that currently limit the industrial applications of this technique.

Acknowledgements

C. J. J. and T. H. acknowledge support from the Australian Research Council (DP170100181, FT140101059, DP150100383).

References

1. B. J. Lee, P. J. Worland, J. N. Davis, T. C. Stadtman and D. L. Hatfield, *J. Biol. Chem.*, 1989, **264**, 9724.
2. B. Hao, W. Gong, T. K. Ferguson, C. M. James, J. A. Krzycki and M. K. Chan, *Science*, 2002, **296**, 1462.
3. W. Eckhart, M. A. Hutchinson and T. Hunter, *Cell*, 1979, **18**, 925.
4. F. R. Bettelheim, *J. Am. Chem. Soc.*, 1954, **76**, 2838.
5. R. Harwood, M. E. Grant and D. S. Jackson, *Biochem. J.*, 1975, **152**, 291.
6. B. Imperiali and S. E. O'Connor, *Curr. Opin. Chem. Biol.*, 1999, **3**, 643.
7. A. Warshel, *Annu. Rev. Biophys. Biomol. Struct.*, 2003, **32**, 425.
8. E. Campbell, M. Kaltenbach, G. J. Correy, P. D. Carr, B. T. Porebski, E. K. Livingstone, L. Afriat-Jurnou, A. M. Buckle, M. Weik, F. Hollfelder, N. Tokuriki and C. J. Jackson, *Nat. Chem. Biol.*, 2016, **12**, 944.
9. C. von der Osten, S. Branner, S. Hastrup, L. Hedegaard, M. D. Rasmussen, H. Bisgard-Frantzen, S. Carlsen and J. M. Mikkelsen, *J. Biotechnol.*, 1993, **28**, 55.
10. F. Van, de Werf, S. R. Bergmann, K. A. Fox, H. de Geest, C. F. Hoyng, B. E. Sobel and D. Collen, *Circulation*, 1984, **69**, 605.
11. G. Dalbadie-McFarland, L. W. Cohen, A. D. Riggs, C. Morin, K. Itakura and J. H. Richards, *Proc. Natl. Acad. Sci. U. S. A.*, 1982, **79**, 6409.
12. W. P. Stemmer, *Nature*, 1994, **370**, 389.
13. D. Rothlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik and D. Baker, *Nature*, 2008, **453**, 190.
14. O. Khersonsky, D. Röthlisberger, O. Dym, S. Albeck, C. J. Jackson, D. Baker and D. S. Tawfik, *J. Mol. Biol.*, 2010, **396**, 1025.
15. N. J. Turner and M. D. Truppo, *Curr. Opin. Chem. Biol.*, 2013, **17**, 212.
16. A. Karlin, *J. Gen. Physiol.*, 1969, **54**, 245.
17. H. Bozler, K. D. Jany and G. Pfleiderer, *Biochim. Biophys. Acta*, 1983, **749**, 238.
18. X.-C. Su, T. Huber, N. E. Dixon and G. Otting, *ChemBiochem*, 2006, **7**, 1599.
19. C. Mayer, D. G. Gillingham, T. R. Ward and D. Hilvert, *Chem. Commun.*, 2011, **47**, 12068.
20. N. V. Katre, *J. Immunol.*, 1990, **144**, 209.
21. A. Koyama, H. Inage, M. Kobayashi, H. Nakamura, M. Narita and S. Tojo, *Immunology*, 1986, **58**, 535.
22. F.-X. Gallat, A. P. S. Brogan, Y. Fichou, N. McGrath, M. Moulin, M. Härtlein, J. Combet, J. Wuttke, S. Mann, G. Zaccai, C. J. Jackson, A. W. Perriman and M. Weik, *J. Am. Chem. Soc.*, 2012, **134**, 13168.
23. A. Ochoa-Solano and C. Gitler, *J. Nutr.*, 1968, **94**, 243.
24. S. M. Hecht, B. L. Alford, Y. Kuroda and S. Kitano, *J. Biol. Chem.*, 1978, **253**, 4517.
25. W. A. Hendrickson, J. R. Horton and D. M. LeMaster, *EMBO J.*, 1990, **9**, 1665.

26. S. K. Blight, R. C. Larue, A. Mahapatra, D. G. Longstaff, E. Chang, G. Zhao, P. T. Kang, K. B. Green-Church, M. K. Chan and J. A. Krzycki, *Nature*, 2004, **431**, 333.
27. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith and P. G. Schultz, *Science*, 1989, **244**, 182.
28. J. D. Bain, E. S. Diala, C. G. Glabe, T. A. Dix and A. R. Chamberlin, *J. Am. Chem. Soc.*, 1989, **111**, 8013.
29. L. Pauling, *Arbeiten aus dem Gebiet der Naturstoffchemie, Festschrift Arthur Stroll.*, Birkhauser Verlag, Basel, Switzerland, 1957, pp. 597–602.
30. K. E. Splan, K. Musier-Forsyth, M. T. Boniecki and S. A. Martinis, *Methods*, 2008, **44**, 119.
31. N. Cvetesic and I. Gruic-sovulj, *Methods*, 2016, 1.
32. A. P. Mascarenhas and S. A. Martinis, *FEBS Lett.*, 2009, **583**, 3443.
33. M. H. Schwartz and T. Pan, *Nucleic Acids Res.*, 2016, **44**, 294.
34. A. J. Link, M. K. S. Vink, N. J. Agard, J. A. Prescher, C. R. Bertozzi and D. A. Tirrell, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 10180.
35. S. Zheng and I. Kwon, *Biotechnol. J.*, 2012, **7**, 47.
36. C. Minks, S. Alefelder, L. Moroder, R. Huber and N. Budisa, *Tetrahedron*, 2000, **56**, 9431.
37. G. N. Cohen and D. B. Cowie, *C. R. Hebd. Seances Acad. Sci.*, 1957, **244**, 680.
38. A. J. Link and D. A. Tirrell, *Methods*, 2005, **36**, 291.
39. Y. Ravikumar, S. P. Nadarajan, T. Hyeon Yoo, C. S. Lee and H. Yun, *Biotechnol. J.*, 2015, **10**, 1862.
40. K. L. Kiick, J. C. M. van Hest and D. A. Tirrell, *Angew Chem. Int. Ed.*, 2000, **39**, 2148.
41. Y. Ravikumar, S. P. Nadarajan, T. Hyeon Yoo, C. S. Lee and H. Yun, *Trends Biotechnol.*, 2015, **33**, 462.
42. Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado and D. A. Tirrell, *Angew Chem. Int. Ed.*, 2001, **40**, 1494.
43. L. Wang and P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **44**, 34.
44. J. D. Bain, E. S. Diala, C. G. Glabe, D. A. Wacker, M. H. Lyttle, T. A. Dix and A. R. Chamberlin, *Biochemistry*, 1991, **30**, 5411.
45. R. Furter, *Protein Sci.*, 1998, **7**, 419.
46. L. Wang, A. rock, B. Herberich and P. G. Schultz, Science, 2001, **292**, 498.
47. T. Kobayashi, O. Nureki, R. Ishitani, A. Yaremchuk, M. Tukalo, S. Cusack, K. Sakamoto and S. Yokoyama, *Nat. Struct. Biol.*, 2003, **10**, 425.
48. R. J. Ernst, T. P. Krogager, E. S. Maywood, R. Zanchi, V. Beranek, T. S. Elliott, N. P. Barry, M. H. Hastings and J. W. Chin, *Nat. Chem. Biol.*, 2016, **12**, 776.
49. W. Liu, A. Brock, S. Chen, S. Chen and P. G. Schultz, *Nat. Methods*, 2007, **4**, 239.
50. S. Chen, P. G. Schultz and A. Brock, *J. Mol. Biol.*, 2007, **371**, 112.
51. I. N. Arthur, J. E. Hennessy, D. Padmakshan, D. J. Stigers, S. Lesturgez, S. A. Fraser, M. Liutkus, G. Otting, J. G. Oakeshott and C. J. Easton, *Chem. – Eur. J.*, 2013, **19**, 6824.

52. F. Katzen, G. Chang and W. Kudlicki, *Trends Biotechnol.*, 2005, **23**, 150.
53. K. Ozawa, M. J. Headlam, P. M. Schaeffer, B. R. Henderson, N. E. Dixon and G. Otting, *Eur. J. Biochem.*, 2004, **271**, 4084.
54. R. W. Alexander, in *Encyclopedia of Biological Chemistry*, 2013, pp. 441–444.
55. M. Abou-Nader and M. J. Benedik, *Bioeng. Bugs*, 2010, **1**, 337.
56. R. C. Cadwell and G. F. Joyce, *PCR Methods Appl.*, 1992, **2**, 28.
57. J. Meiler and D. Baker, *Proteins: Struct., Funct., Genet.*, 2006, **65**, 538.
58. W. Ren, T. M. Truong and H. Ai, *Sci. Rep.*, 2015, **5**, 12632.
59. R. Sun, H. Zheng, Z. Fang and W. Yao, *Biochem. Biophys. Res. Commun.*, 2010, **391**, 709.
60. A. Currin, N. Swainston, P. J. Day and D. B. Kell, *Chem. Soc. Rev.*, 2015, **44**, 1172.
61. J. Xie and P. G. Schultz, *Nat. Rev. Mol. Cell Biol.*, 2006, **7**, 775.
62. R. D. Socha and N. Tokuriki, *FEBS J.*, 2013, **280**, 5582.
63. E. Neil and G. Marsh, *Chem. Biol.*, 2000, **7**, R153.
64. C. Odar, M. Winkler and B. Wiltschi, *Biotechnol. J.*, 2015, **10**, 427.
65. L. Merkel, M. Schauer, G. Antranikian and N. Budisa, *ChemBioChem*, 2010, **11**, 1505.
66. K. Deepankumar, M. Shon, S. P. Nadarajan, G. Shin, S. Mathew, N. Ayyadurai, B. G. Kim, S. H. Choi, S. H. Lee and H. Yun, *Adv. Synth. Catal.*, 2014, **356**, 993.
67. B. Holzberger and A. Marx, *J. Am. Chem. Soc.*, 2010, **132**, 15708.
68. K. Ohtake, A. Yamaguchi, T. Mukai, H. Kashimura, N. Hirano, M. Haruki, S. Kohashi, K. Yamagishi, K. Murayama, Y. Tomabechi, T. Itagaki, R. Akasaka, M. Kawazoe, C. Takemoto, M. Shirouzu, S. Yokoyama and K. Sakamoto, *Sci. Rep.*, 2015, **5**, 9762.
69. B. Borgo and J. J. Havranek, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 1494.
70. E. Sugrue, P. D. Carr, C. Scott and C. J. Jackson, *Biochemistry*, 2016, **55**, 6304.
71. C. M. Miton and N. Tokuriki, *Protein Sci.*, 2016, **25**, 1260.
72. C. Mitchinson and J. A. Wells, *Biochemistry*, 1989, **28**, 4807.
73. C. J. Jackson, C. W. Coppin, P. D. Carr, A. Aleksandrov, M. Wilding, E. Sugrue, J. Ubels, M. Paks, J. Newman, T. S. Peat, R. J. Russell, M. Field, M. Weik, J. G. Oakeshott and C. Scott, *Appl. Environ. Microbiol.*, 2014, **80**, 4003.
74. M. G. Hoesl, C. G. Acevedo-Rocha, S. Nehring, M. Royter, C. Wolschner, B. Wiltschi, N. Budisa and G. Antranikian, *ChemCatChem*, 2011, **3**, 213.
75. J. C. Jackson, S. P. Duffy, K. R. Hess and R. A. Mehl, *J. Am. Chem. Soc.*, 2006, **128**, 11124.
76. W. A. Denny, *Curr. Pharm. Des.*, 2002, **8**, 1349.
77. J. N. Kolev, J. M. Zaengle, R. Ravikumar and R. Fasan, *ChemBioChem*, 2014, **15**, 1001.
78. C. J. Jackson, J. L. Foo, N. Tokuriki, L. Afriat, P. D. Carr, H. K. Kim, G. Schenk, D. S. Tawfik and D. L. Ollis, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 21631.

79. C. J. Jackson, K. Weir, A. Herlt, J. Khurana, T. D. Sutherland, I. Horne, C. Easton, R. J. Russell, C. Scott and J. G. Oakeshott, *Appl. Environ. Microbiol.*, 2009, **75**, 5153.
80. C. J. Jackson, C. Scott, A. Carville, K. Mansfield, D. L. Ollis and S. B. Bird, *Biochem. Pharmacol.*, 2010, **80**, 1075.
81. C. J. Jackson, A. Carville, J. Ward, K. Mansfield, D. L. Ollis, T. Khurana and S. B. Bird, *Toxicology*, 2014, **317**, 1.
82. C. J. Jackson, J.-L. Foo, N. Tokuriki, L. Afriat, P. D. Carr, H.-K. Kim, G. Schenk, D. S. Tawfik and D. L. Ollis, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 21631.
83. I. N. Ugwumba, K. Ozawa, Z.-Q. Xu, F. Ely, J.-L. Foo, A. J. Herlt, C. Coppin, S. Brown, M. C. Taylor, D. L. Ollis, L. N. Mander, G. Schenk, N. E. Dixon, G. Otting, J. G. Oakeshott and C. J. Jackson, *J. Am. Chem. Soc.*, 2011, **133**, 326.
84. P.-Y. Kim, D. J. Pollard and J. M. Woodley, *Biotechnol. Prog.*, 2007, **23**, 74.
85. S. Zheng and I. Kwon, *Biotechnol. Bioeng.*, 2013, **110**, 2361.
86. R. Wohlgemuth, *Curr. Opin. Microbiol.*, 2010, **13**, 283.
87. H. Ma, X. Yang, Z. Lu, N. Liu and Y. Chen, *PLoS One*, 2014, **9**, e103792.
88. Y. Chen, C. Chen and X. Wu, *Chem. Soc. Rev.*, 2012, **41**, 1742.
89. M. Lu, Y. Huang, M. A. White, X. Wu, N. Liu, X. Cheng and Y. Chen, *Chem. Commun.*, 2012, **48**, 11352.
90. W. H. Zhang, G. Otting and C. J. Jackson, *Curr. Opin. Struct. Biol.*, 2013, **23**, 581.
91. M. A. J. Dominguez, K. C. Thornton, M. G. Melendez and C. M. Dupureur, *Proteins*, 2001, **45**, 55.
92. Y. Wu, S. D. Fried and S. G. Boxer, *Biochemistry*, 2015, **54**, 7110.
93. A. T. Fafarman, P. A. Sigala, J. P. Schwans, T. D. Fenn, D. Herschlag and S. G. Boxer, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, E299.
94. L. Wang, S. D. Fried, S. G. Boxer and T. E. Markland, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 18454.
95. W.-N. Chen, K. V. Kuppan, M. D. Lee, K. Jaudzems, T. Huber and G. Otting, *J. Am. Chem. Soc.*, 2015, **137**, 4581.
96. J. S. Lewis, S. Jergic and N. E. Dixon, *Enzymes*, 2016, **39**, 31.
97. H.-M. Lee, D. R. Larson and D. S. Lawrence, *ACS Chem. Biol.*, 2009, **4**, 409.
98. A. Deiters, D. Groff, Y. Ryu, J. Xie and P. G. Schultz, *Angew. Chem., Int. Ed. Engl.*, 2006, **45**, 2728.
99. J. Hemphill, C. Chou, J. W. Chin and A. Deiters, *J. Am. Chem. Soc.*, 2013, **135**, 13433.
100. J. Luo, Q. Liu, K. Morihiro and A. Deiters, *Nat. Chem.*, 2016, **8**, 1027.
101. Z. Ye and G. J. Williams, *Biochemistry*, 2014, **53**, 7494.
102. T. Matsuo and S. Hirota, *Bioorg. Med. Chem.*, 2014, **22**, 5638.
103. M. J. Jezewska, S. Rajendran, D. Bujalowska and W. Bujalowski, *J. Biol. Chem.*, 1998, **273**, 10515.
104. G. DeSantis and J. B. Jones, *Curr. Opin. Biotechnol.*, 1999, **10**, 324.

105. S. Tyagi and E. A. Lemke, *Curr. Opin. Struct. Biol.*, 2015, **32**, 66.
106. A. Deiters, T. A. Cropp, D. Summerer, M. Mukherji and P. G. Schultz, *Site-specific PEGylation of Proteins Containing Unnatural Amino Acids*, 2004, vol. 14.
107. A. Deiters, T. A. Cropp, M. Mukherji, J. W. Chin, J. C. Anderson and P. G. Schultz, *J. Am. Chem. Soc.*, 2003, **125**, 11782.
108. E. M. Brustad, E. A. Lemke, P. G. Schultz and A. A. Deniz, *J. Am. Chem. Soc.*, 2008, **130**, 17664.
109. H. Neumann, S. Y. Peak-Chew and J. W. Chin, *Nat. Chem. Biol.*, 2008, **4**, 232.
110. D. T. Rogerson, A. Sachdeva, K. Wang, T. Haq, A. Kazlauskaite, S. M. Hancock, N. Huguenin-Dezot, M. M. K. Muqit, A. M. Fry, R. Bayliss and J. W. Chin, *Nat. Chem. Biol.*, 2015, **11**, 496.
111. A. Sauerwald, W. Zhu, T. A. Major, H. Roy, S. Palioura, D. Jahn, W. B. Whitman, J. R. Yates, M. Ibba and D. Söll, *Science*, 2005, **307**, 1969.
112. S. Schoffelen, M. H. L. Lambermon, M. B. van Eldijk and J. C. M. van Hest, *Bioconjugate Chem.*, 2008, **19**, 1127.
113. N. J. Weise, F. Parmeggiani, S. T. Ahmed and N. J. Turner, *J. Am. Chem. Soc.*, 2015, **137**, 12977.
114. K. Zhang, H. Li, K. M. Cho and J. C. Liao, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 6234.
115. S. Angkawidjaja, C. Tayyab, M. Rashid, N. Kanaya, Protein Data Bank, 2012.
116. K. S. Midelfort, R. Kumar, S. Han, M. J. Karmilowicz, K. McConnell, D. K. Gehlhaar, A. Mistry, J. S. Chang, M. Anderson, A. Villalobos, J. Minshull, S. Govindarajan and J. W. Wong, *Protein Eng., Des. Sel.*, 2013, **26**, 25.
117. E. Johansson, G. N. Parkinson, W. A. Denny and S. Neidle, *J. Med. Chem.*, 2003, **46**, 4009.
118. C. J. Jackson, J. L. Foo, H. K. Kim, P. D. Carr, J. W. Liu, G. Salem and D. L. Ollis, *J. Mol. Biol.*, 2008, **375**, 1189.
119. V. Cody, J. R. Luft and W. Pangborn, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2005, **61**, 147.
120. H. M. Girvan, H. E. Seward, H. S. Toogood, M. R. Cheesman, D. Leys and A. W. Munro, *J. Biol. Chem.*, 2007, **282**, 564.
121. A. Athanasiadis, M. Vlassi, D. Kotsifaki, P. A. Tucker, K. S. Wilson and M. Kokkinidis, *Nat. Struct. Biol.*, 1994, **1**, 469.
122. M. S. Strycharska, E. Arias-Palomo, A. Y. Lyubimov, J. P. Erzberger, V. L. O'Shea, C. J. Bustamante and J. M. Berger, *Mol. Cell*, 2013, **52**, 844.
123. W. Huang, J. Jia, P. Edwards, K. Dehesh, G. Schneider and Y. Lindqvist, *EMBO J.*, 1998, **17**, 1183.
124. H. E. Parge, R. A. Hallewell and J. A. Tainer, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 6109.
125. L. H. Weaver and B. W. Matthews, *J. Mol. Biol.*, 1987, **193**, 189.
126. P. Strzelczyk, G. D. Bujacz, P. Kielbasinski and J. Blaszczyk, *Acta Biochim. Pol.*, 2016, **63**, 1065.