We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Protein Engineering with Non-Natural Amino Acids

Aijun Wang, Natalie Winblade Nairn,
Marcello Marelli and Kenneth Grabstein
Allozyne
USA

1. Introduction

Among all biomolecules in living cells, proteins represent a group with highly divergent biological functions. These include maintaining structural integrity, catalyzing most chemical reactions *in vivo*, regulating cellular responses via signal transduction, and targeting foreign molecules via the immune system. Because of these diverse functions and associated applications, protein engineering has been a field for intense academic research and biopharmaceutical drug development. However, proteins made by nature are generally composed of only 20 canonical amino acids. The recently developed technology to incorporate non-natural amino acids (NNAAs) into proteins at specifically chosen sites represents an extremely powerful tool for protein engineering. Unlike traditional protein engineering with the 20 canonical amino acids as building blocks, protein engineering with NNAAs has nearly unlimited novel side-chain structures to choose from. Coupled with rapid advances in protein structure computation and bioorthogonal chemical reactions, proteins with tailored properties or novel functions beyond natural or directed evolution are becoming a reality.

In this book chapter, we will first review technologies being developed to incorporate NNAAs into proteins. While NNAAs can be incorporated via chemical synthesis, or *in vitro* cell-free protein translation systems, we will focus our review on cell-based protein expression systems because of their capacity for large scale manufacturing. Examples of different ways to assign genetic codons for NNAAs will be reviewed and compared with each other in terms of incorporation efficiency at desired site(s), sequence fidelity of the engineered protein, protein yield, and scalability. Then we will provide examples of chemical conjugation reactions that are orthogonal to the naturally occurring functional groups present in proteins. Topics include reaction kinetics, effect of the reaction conditions on protein, and stability of the final conjugate. At the end, we will highlight a few potential applications of proteins engineered with NNAAs. A special focus will be on therapeutic proteins for clinical use. It is our hope that with this review, more scientists in the protein engineering field will harness this powerful technology and apply it to meet the challenges in their work.

2. Methodologies for incorporating non-natural amino acids into proteins

Protein translation is the process where the genetic information, stored as triplet codon sequence in the mRNA, is used to dictate the polypeptide synthesis. The components required in protein translation were first postulated in the original Adaptor hypothesis by Crick in 1958, and were subsequently identified experimentally (Ibba *et al.*, 2000). During the course of protein synthesis, a ribosome moves along an mRNA molecule and pairs each triplet codon with the adaptor aminoacyl-tRNA containing the complementary anticodon, and catalyzes the formation of a peptide bond with the growing polypeptide chain (Figure 1).
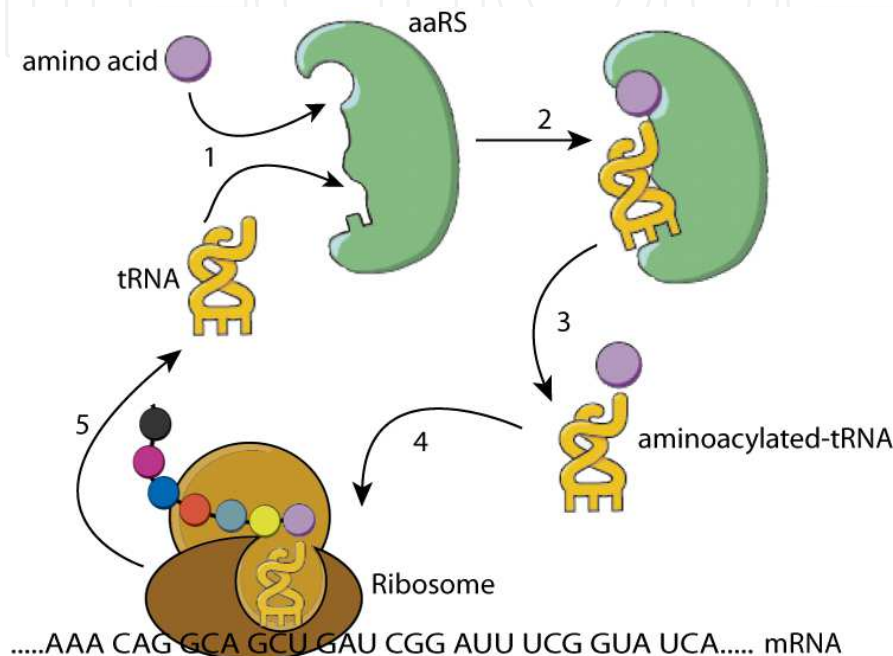

Fig. 1. Natural protein translation system. Aminoacyl-tRNA synthetases (aaRSs) play a key role in maintaining the fidelity of protein translation by demonstrating strict specificity for their cognate amino acid and their tRNA. (1) Each aaRS recognizes unique structural elements of its cognate tRNA(s) and amino acid. (2) The aaRS joins the amino acid to the tRNA, forming the aminoacyl-tRNA (3) that is released into the cytosol. (4) As the ribosome moves along an mRNA molecule, codons are paired with aminoacylated tRNAs containing the complementary anticodons. The ribosome catalyses the sequential peptide-bond formation of amino acids delivered by tRNAs, generating the nascent protein. At the end of translation, release factors occupy stop codons and trigger the termination of protein synthesis. (5) Deaminoacylated tRNA is recycled for further use.

Translation is a highly regulated process, and protein sequence fidelity is largely controlled at two steps: 1) the correct coupling of the amino acid with the tRNA to generate the aminoacyl-tRNA; 2) the correct pairing of the anticodon in the aminoacyl-tRNA with the codon in mRNA. The former is governed by the function of the aminoacyl-tRNA synthetases (aaRSs). The substrate selectivity for tRNA is accomplished through structural recognition elements that are unique to each tRNA and often include the anticodon site. The substrate specificity for the amino acid is achieved by a highly selective amino acid binding pocket. Some aaRSs cannot distinguish between structurally similar amino acids, but they have evolved an editing function to eliminate mis-acylated tRNAs. In most cells, there are 20

aaRS enzymes, one for each amino acid. The 20 canonical amino acids are encoded by 61 codons, with most amino acids encoded by more than one codon. Protein translation is terminated at one of the three stop codons by the protein release factors (RF). It should be mentioned that there are exceptions to nearly every rule in the general description of the translation process. These exceptions actually provide some of the tools that protein engineers have used to introduce NNAAs into proteins, and will be described in subsequent sections.

Several methods have been developed for genetically encoding NNAAs into recombinant proteins. These methodologies range from simple exploitation of the promiscuity of the endogenous protein translation system in the host cells, to the introduction of an exogenous aaRS/tRNA pair specifically designed for a NNAA. All these methods manipulate the natural protein translation process and rely on the reassignment of a genetic codon (such as a sense codon or a stop codon) to the NNAA of interest. As a result, incorporation of the NNAA is intrinsically in competition with certain natural processes: for example, competition from the natural amino acid for a sense codon, or from the release factor for a stop codon. This competition impacts on the efficiency of NNAA incorporation, sequence fidelity as well as the expression yield of target proteins. The ability to reduce or totally eliminate such competition will significantly expedite protein engineering with NNAAs. Recent progress in synthetic biology, which enables the generation of artificial cells with synthetic genomes, shows great promise for such a goal. In this section, we will first describe methods to evolve aaRSs and tRNAs for NNAAs, and then review NNAA incorporation methods based on codon reassignment.

2.1 Methods for evolution of aaRSs and tRNAs for NNAAs

Site-specific introduction of NNAAs requires the expression of an aaRS/tRNA pair specific to a NNAA in the expression host. While this pair must integrate seamlessly with the host translational apparatus, controlled insertion of NNAAs at desired sites requires that it be orthogonal to the host cell. That is, the aaRS/tRNA pair should have the following properties: the introduced aaRS should aminoacylate only NNAA to its cognate tRNA, but not any of the host tRNAs. This avoids mis-incorporation of the NNAA at other codons; in turn, the introduced tRNA should be activated only by its cognate aaRS, but not any of the host aaRSs. This ensures that only the NNAA is incorporated at its assigned codon. The orthogonality of the aaRS/tRNA pair within the expression host cell is essential to maintain sequence fidelity of the target proteins. By exploiting the evolutionary divergence between species, orthogonal aaRS/tRNA pairs have been obtained by transferring an aaRS/tRNA pair from one species into another. An alternate approach requires the re-engineering of recognition elements of the aaRS and its tRNA *de novo* to achieve orthogonality (Neumann *et al.*, 2010a).

While most cells contain 20 aaRSs, some species have evolved additional aaRSs to allow for the utilization of other naturally occurring, but non-canonical amino acids. Pyrrolysine (Pyl) is used by several species of the methanogenic archea, *Methanosarcina*. In its native context, the PylRS/tRNA evolved alongside the 20 canonical aaRSs and developed natural specificity for its cognate amino acid, making it an ideal pair for transfer into other expression hosts. Indeed, the PylRS/tRNA has been shown to be fully orthogonal in *E. coli*,

yeast and mammalian cells (reviewed by Voloshchuk & Montclare, 2010). Furthermore, Pyl is naturally encoded by the amber stop codon, the most commonly reassigned codon to specify a NNAA, and thus requires no engineering to adapt it to this purpose. However, the Pyl-tRNA is not hardwired for amber codon suppression, and can be adapted to decode a number of other codons (Ambrogelly *et al.*, 2007). In addition, the PylRS has been shown to readily accept a variety of side chain structures (Polycarpo *et al.*, 2006;Yanagisawa *et al.*, 2008;Neumann *et al.*, 2008;Chen *et al.*, 2009;Nguyen *et al.*, 2009;Li *et al.*, 2009;Hancock *et al.*, 2010;Ou *et al.*, 2011;Plass *et al.*, 2011;Takimoto *et al.*, 2011) as well as a set of non-alpha amino derivatives (Kobayashi *et al.*, 2009). This feature makes the PylRS/tRNA pair an ideal candidate for site specific integration of NNAAs.

2.1.1 Positive/negative selection

Alternating positive and negative screens were pioneered by Schultz's group to evolve orthogonal aaRS/tRNA pairs from large mutational libraries for NNAA incorporation at a stop codon (reviewed by Liu & Schultz, 2010). Figure 2 illustrates the application for evolving aaRSs: First, an aaRS mutant library, along with its cognate tRNA and a conditionally required gene, whose open reading frame is interrupted by a stop codon, is introduced to the host cells for positive selection. Cells are grown the presence of NNAA and only cells with a functional aaRS/tRNA pair (for either NNAA or canonical amino acids) will enable the expression of the essential gene for survival. The selected aaRS mutants (along with the suppressor tRNA) are then shuffled into the host containing a toxic gene (containing an in frame stop codon) for negative selection. When cells are grown in the absence of NNAA, cells containing the aaRS variants that activate canonical amino acids will express the full-length lethal protein and be eliminated; cells containing aaRSs that activate NNAA but not canonical amino acids will survive. The resulting aaRS mutants will be shuffled back for multiple iterations of positive and negative selections until the library converges and/or the desired functionality of the aaRS is obtained. While the alternating positive and negative screen has proven to be extremely successful, the strategy requires plasmid shuttling that is time consuming and affects library diversity. In addition, the common practice of using antibiotic resistance genes for positive selection, while effective, can lead to unexpected variability of identified mutants (Pastrnak & Schultz, 2001;Young *et al.*, 2010). Furthermore, this scheme does not eliminate the mis-acylation of host tRNA(s) with the NNAA by some aaRS variants, leading to potential non-orthogonality of the aaRS/tRNA pair and the introduction of NNAA at undesired sites.

The positive/negative selection (or its various modified formats) is the most widely practiced method for evolving orthogonal aaRS/tRNA pairs. While it was initially developed for an *E. coli* expression system, its principle has been applied successfully in yeast (Chin *et al.*, 2003). Development of orthogonal aaRS/tRNA pairs for mammalian expression faces some additional challenges. Due to the low efficiency of mammalian cell transfection and their relatively slow growth, it is impractical to screen large libraries and evolve aaRSs directly in mammalian cells. Instead, aaRS/tRNA pairs derived from prokaryotes and developed for yeast hosts have been adapted successfully for mammalian expression. For example, both the *E. coli* TyrRS/tRNA pair and *E. coli* TyrRS paired with *Bacillus stearothermophilus* (*B.St*) tRNA were successfully transferred to mammalian cells (Liu *et al.*, 2007;Wang *et al.*, 2007b;Takimoto *et al.*, 2009)
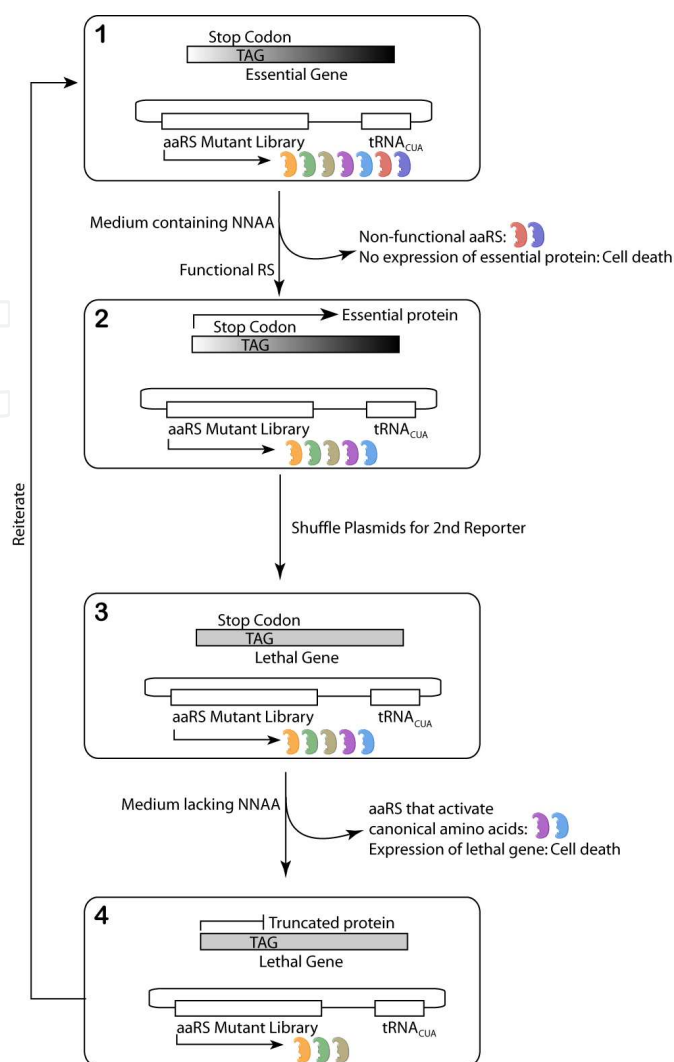
Fig. 2. An alternating positive and negative screen for evolving orthogonal aaRSs and tRNAs. (1) A library of aaRS mutants and its cognate tRNA$_{CUA}$ are introduced into an expression host containing a conditionally essential gene (e.g. antibiotic resistance marker) with an amber stop codon interrupting its open reading frame and grown in medium containing NNAA. (2) Cells containing functional aaRSs express the full-length essential gene and are able to survive under selective conditions (e.g., presence of antibiotic). (3) Surviving cells are harvested and the aaRS/tRNA$_{CUA}$ variants are introduced into cells containing a gene (with an amber codon) that when expressed is lethal (e.g. Barnase), and grown in the absence of NNAA. (4) Cells containing orthogonal NNAA aaRS/tRNA$_{CUA}$ pairs express truncated non-functional Barnase and get selected for the next round of screening.

2.1.2 Green fluorescent protein (GFP)-based selection

A more streamline scheme utilizing GFP for both positive and negative selection was developed to evolve aaRS/tRNA pair in *E. coli* (Santoro *et al.*, 2002). The method depends on antibiotic resistance and the expression of GFP, and enables high throughput screening of library variants by FACS. A single plasmid carrying two reporter genes, the chloramphenicol resistance gene containing an amber codon, and a GFP gene that depends

on amber suppression for its expression, was used to select *Methanococcus jannaschii* (*Mj*.) TyrRS mutants in the presence of its cognate $tRNA_{CUA}$. Positive selection was performed in medium containing chloramphenicol and a NNAA, and the isolation of live green bacteria led to bacteria cells carrying functional *Mj.* TyrRS. Negative selection was performed in medium lacking both the antibiotic and the NNAA. By isolating non-fluorescent cells, *Mj.* TyrRS mutants that incorporated natural amino acids were eliminated. The resulting clones were then retested for growth on chloramphenicol in the presence of the NNAA. This screen led to the identification of mutations that preferred *p*-isopropyl-Phe, *p*-amino-Phe, and *O*-allyl-Tyr (Santoro *et al.*, 2002).

2.1.3 Surface-display-based selection

Phage (or yeast, or any cell surface) display utilizes the special properties of NNAA-containing peptides displayed on the cell surface for screening and selection of functional aaRSs (or tRNAs) from a vast number of variants in a library. NNAAs with special binding properties to a particular type of immobilized resin or ligand may be used for selection. For example, proteins containing a boronate amino acid, which forms strong and reversible complexes with polyols and *cis*-diols, could be selected on *N*-methylglucamine resin (Brustad *et al.*, 2008). Alternatively, a monoclonal antibody that binds specifically to a NNAA-containing peptide may be used for selection (Pastrnak & Schultz, 2001); conjugation of a florescent tag to the NNAA (containing a special chemical reactive group) displayed on the cell surface, followed by cell sorting via flow cytometry, may also lead to selection of the aaRS or tRNA. In fact, MetRS mutants with improved activities toward azidonorleucine and better discrimination against Met were generated via surface labelling and cell sorting (Tanrikulu *et al.*, 2009). The surface-display based selection can be applied to evolve aaRS/tRNA for either a sense codon or a stop codon. The limitation of the method is that the method may not be readily applicable to any NNAA, as it relies on the special properties of the NNAA or on antibody reagents.

2.1.4 Rational design for selection

Rational design predicts structural changes needed for NNAA activation based on a solved (or simulated) aaRS structure for its cognate amino acid. By computationally simulating the binding (or editing) sites and binding energy for a NNAA, putative mutations that favour the introduction of a NNAA over the natural amino acid can be identified and tested experimentally. This approach avoids the laborious generation and screening of a large aaRS mutation library, and is expected to be utilized more frequently as structural prediction algorithms improve. An example of this approach was the successful development of an orthogonal yeast PheRS/tRNA pair for the incorporation of a NNAA at an amber stop codon in *E coli*. By sequence and structural analysis of PheRS from various species, a key residue in yeast PheRS (T415) was identified that could relax the specificity of the amino acid binding pocket and allow for the incorporation of NNAAs (Kwon *et al.*, 2006). In combination with a yeast amber suppressor Phe-$tRNA_{CUA-UG}$ (which was designed to limit the cross-reactivity of the tRNA with the *E. coli* LysRS), the mutant PheRS (T415G or A) enabled the incorporation of several Trp or Phe analogs in a site specific manner, with greater than 98% NNAA occupancy at the amber site (Kwon *et al.*, 2006; Kwon & Tirrell, 2007).

Similarly, rational design on tRNA can lead to a smaller number of candidates for the generation of orthogonal tRNAs. For example, the amber suppressor Trp-tRNA$_{CUA}$ from yeast was found to be mis-acylated by endogenous LysRS in *E. coli* (Hughes & Ellington, 2010). Analysis of the tRNA structure revealed that the yeast amber suppressor Trp-tRNA$_{CUA}$ shares 73% sequence homology with *E. coli* Lys-tRNA$_{UUU,}$ and contains several key identity elements for the recognition of *E. coli* Lys-tRNA$_{UUU}$ by its cognate aaRS. Furthermore, structural analysis revealed that the modification of the G:C content of the anticodon stem may reduce the structural flexibility of this stem and eliminate mis-acylation. Combined with experimental testing, an orthogonal and functional yeast Trp-tRNA$_{CUA}$ was obtained for the incorporation of NNAAs (Hughes & Ellington, 2010).

2.2 Methods for incorporation of NNAAs into recombinant proteins

2.2.1 Reassigned sense codon

The simplest method to introduce NNAAs into proteins exploits the promiscuity of aaRSs of the host cells. That is, when a particular aaRS can charge a NNAA non-specifically to its cognate tRNA, the activated NNAA-tRNA may be utilized by the ribosomal translation machinery, and the NNAA added to the nascent polypeptide chain according to the sense codon specified for the canonical amino acid. The sense codon is reassigned to the NNAA, resulting in a recombinant protein composed of 19 canonical amino acids and one NNAA. Thus, this is a residue-specific global substitution method.

The successful substitution of a canonical amino acid with a NNAA relies on the use of auxotrophic expression hosts deficient in the biosynthesis of that amino acid. Employment of such hosts limits competition from the canonical amino acid for the reassigned sense codon, and improves the incorporation efficiency and yield of target proteins. For example, the MetRS has a relatively relaxed substrate binding pocket, and a number of Met analogs have been successfully incorporated into target proteins (reviewed by Hendrickson *et al.*, 2004;Voloshchuk & Montclare, 2010). A special consideration worth mentioning when using Met analogs for protein engineering is the *in vivo* processing of NNAAs at the N-terminus of proteins. It is well know that most protein translation starts with a Met codon, and that Met at the N-terminus of proteins is subjected to enzyme-mediated processing, resulting either in the retention or cleavage of N-terminal Met. The rule for such N-terminal processing depends on the identity of the 2nd amino acid in the protein sequence. A set of similar but not identical rules have been established for the N-terminal processing of NNAAs in recombinant interferon (Wang *et al.*, 2008) as well as other proteins (unpublished data). As a result, the N-terminal NNAA can be either retained or cleaved by choosing an appropriate 2nd residue (Table 1).

2nd residual	Aha cleavage%	Hpg cleavage%
Ala	96	91
Ser	80	80
Gly	52	33
Gln	8	0
Glu	0	0
His	0	0

Table 1. N-terminal processing of Met analog Aha or Hpg (structures provided in Fig 4) in *E. coli*.

Besides the Met codon, many other sense codons (such as codons for Pro, Tyr, Phe, Leu, Val etc.) have also been reassigned, and more than 60 NNAAs have been incorporated into proteins via global substitution (reviewed by Hendrickson *et al.*, 2004;Voloshchuk & Montclare, 2010). In many cases, the intracellular level of wild-type aaRS may need to be augmented in order to compensate for the low activity of native aaRS towards the NNAA. In other cases, engineered aaRSs with higher catalytic activity towards the NNAA could be supplied to the expression host cells. An impressive example is with a MetRS mutant, which was identified via cell-surface display. The mutant MetRS activated azidonorleucine two-fold more efficiently than Met. As a result, the target protein was expressed with azidonorleucine at higher levels than with Met (Tanrikulu *et al.*, 2009). By utilizing auxotrophic host cells for two or more canonical amino acids, it may be possible to produce recombinant proteins with more than one type of NNAA by this strategy. Furthermore, the global substitution method may be applied in other expression hosts, such as mammalian cells.

It should be mentioned that while native or mutant aaRSs can utilize NNAAs to replace a canonical amino acid *in vivo*, most NNAAs don't support the normal growth and expansion of host cells, presumably due to the impairment of some protein functions when a NNAA is incorporated. As a result, cell density (and thus protein expression yield) is affected during the induction phase when a NNAA is added into the media. Adaption of host strains to the NNAA may provide a solution. By gradually increasing the amount of NNAA in the growth media until complete replacement was accomplished, *E. coli* hosts capable of surviving in media containing 4-fluoro-Trp were evolved (Bacher & Ellington, 2001). These evolved *E. coli* hosts contain 3 mutant enzymes involved in the uptake and utilization of Trp, and are capable of fully substituting 4-fluoro-Trp in their proteomes. Utilization of such adapted expression hosts represents another approach to improve the yield and NNAA incorporation efficiency.

The main limitation of this technology is that the NNAA will replace the canonical amino acid throughout the protein sequence, which may restrict its application if such global substitution is undesired. One solution to this problem is to mutate undesired sites with other canonical amino acids so that the desired site(s) is reserved for NNAA. To this end, functional target proteins lacking a particular amino acid have been generated, such as Met-free human interferon beta or scFv (unpublished data) and Phe-free GFP (Goltermann *et al.*, 2010). With these proteins, the NNAA can be site-specifically incorporated at any desired site. Alternatively, site-specific incorporation can be accomplished via the introduction of an aaRS/tRNA pair specifically designed for the NNAA, as described in the following sections.

The global substitution technology offers many advantages, which include: 1) the efficiency of NNAA incorporation, which can be near 100% at lab scale using a media exchange procedure. At production scale with a media depletion procedure, the incorporation efficiency can still reach 85-90% (unpublished data); 2) the target protein sequence is tightly controlled. In general, the NNAA incorporates only at the reassigned sense codon; the protein sequence at all other codons is strictly controlled and not affected by the reassigned sense codon; 3) target protein expression level is generally very high, even for proteins containing multiple copies of the NNAA. Protein yields at 10-100mg/L have been reported (Link & Tirrell, 2005) in shaker flasks, and yields of up to several gram/L can be achieved via high-density fermentation in bioreactors (unpublished data); 4) the expression system is simple and well-controlled, and the fermentation process is highly scalable. Thus, this technology is amendable for large-scale manufacture.

2.2.2 Wobble codon

Wobble codons refer to codons that are decoded by tRNA via non-classical Watson-Crick base-pairing. As there are 61 sense codons to encode 20 canonical amino acids, but fewer than 50 tRNAs in many organisms (Sharp *et al.*, 2005) (for a tRNA database on different organisms: http://gtrnadb.ucsc.edu/), some tRNA must pair with more than one codon. The non-classical (or wobble) pairing is enabled through modification at the tRNA's 1st anticodon base (which pairs with the 3rd base to the codon triplet), as proposed in the "Wobble Hypothesis"(Crick, 1966). For example, many organisms have only one Phe-$tRNA_{GAA}$ to decode two codons for Phe: UUU and UUC. As a result, the GAA anticodon binds to the UUC codon via Watson-Crick base-pairing, and to the UUU codon via "wobble" base-pairing. While many more modifications on tRNA have since been identified over the last 40 years, and the "Wobble Hypothesis" has been modified and expanded, the essential premise remains the same: through various base modifications, the anticodon structure is shaped in a way that enables it to bind to cognate and wobble bases at the ribosome (Agris *et al.*, 2007).

Because of the wobble pairing between codon and anticodon, one tRNA may pair with several codons, and a given codon may pair with more than one tRNAs. However, pairing at a wobble codon may be less effective to support protein translation (Curran & Yarus, 1989;Sharp *et al.*, 2005). Taking advantage of this property, a wobble codon may be assigned to a NNAA to generate recombinant protein that contains 20 canonical amino acids and one NNAA, as described below.
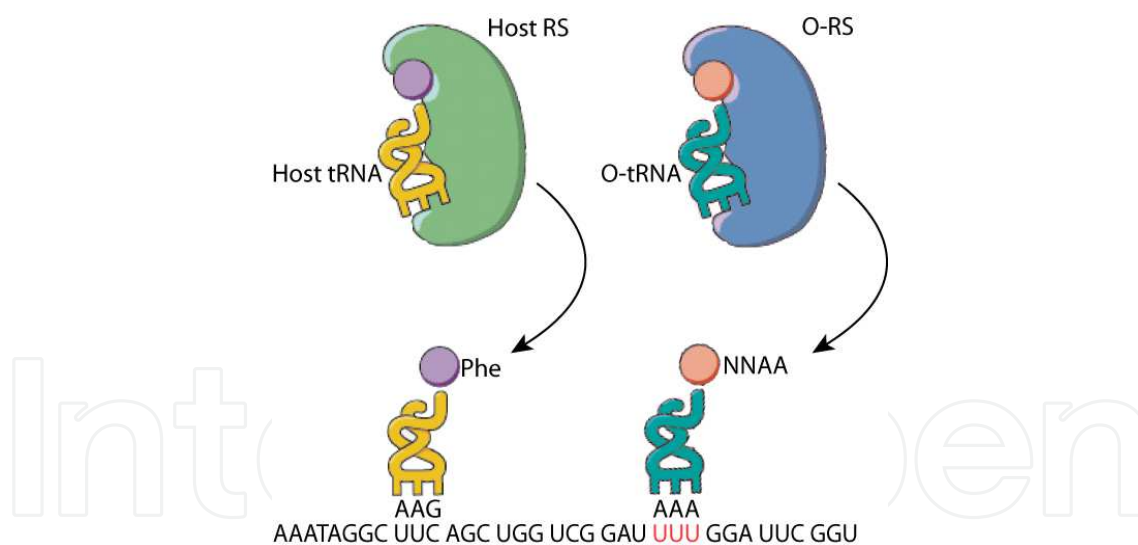

Fig. 3. Site-specific integration of NNAAs at wobble codons. Phe is normally encoded by two codons UUC and UUU, with both codons recognized by the $tRNA_{GAA}$. By expressing an orthogonal aaRS/$tRNA_{AAA}$ (O-RS and O-tRNA) pair, with specificity for a NNAA and containing the "AAA" anticodon, efficient introduction of the NNAA at UUU codons can be achieved.

To assign a wobble codon to a NNAA, a heterologous PheRS/$tRNA_{AAA}$ pair from the yeast *S. cerevisiae* was introduced into an *E. coli* Phe auxotrophic expression host containing the target gene under a strong inducible promoter (Kwon *et al.*, 2003). In this system, the yeast

PheRS was engineered to activate 2-naphthylalanine to the yeast Phe-tRNA$_{AAA}$, and the target gene was designed with a UUC codon for all desired Phe sites, and a UUU wobble codon was reserved at specific sites for 2-naphthylalanine (Figure 3).

With this method, Phe can be essentially quantitatively assigned to the UUC codon, and NNAA to the UUU wobble codon at approximately 80% efficiency in *E. coli* expression host (Link & Tirrell, 2005). Furthermore, multiple copies of a NNAA can be introduced site-specifically into a target protein. The expression level of target protein, however, is typically an order of magnitude lower than the global sense codon reassignment method, on the order of 2-5 mg DHFR per liter with shaker flasks. The applicability of this method in other expression hosts such as yeast and mammalian cells remains to be investigated.

2.2.3 Bias codon

It is well known that degenerate codons are not used with equal frequency in many organisms. The preferred codons are the ones that match the tRNAs that are most abundant in the cells, and this codon preference is more evident in rapidly multiplying organisms or in highly expressed genes. Presumably, codon usage and tRNA gene content coevolved to match each other, and in doing so, organisms effectively minimize the energy and time costs for protein translation and cell growth (Higgs & Ran, 2008).

The preferred codons differ between organisms, and even between different tissues or cell types of the same organism. The cellular content of tRNA species is a determining factor on the rates and amounts of protein synthesized (Gu *et al.*, 2004). As a consequence, recombinant protein production in heterologous host cells is often codon-optimized to match the preferred host cell codon bias (The codon usage database for different organisms and codon analysis of a given gene can be found at: http://www.kazusa.or.jp/codon/).

The biased codon usage provides another method to introduce NNAAs into recombinant proteins. For example, out of the six degenerate codons for Arg, AGG and AGA are rarely used in *E. coli*. Introduction of an orthogonal ArgRS/tRNA$_{CCU}$ pair into an *E. coli* expression host may enable the NNAA-activated orthogonal tRNA$_{CCU}$ to compete effectively with the low endogenous *E. coli* Arg-tRNA$_{CCU}$. This method has been proven feasible with an *in vitro* cell-free based system, where chemically synthesized NNAA-tRNA$_{CCU}$ was incorporated at AGG codons (Hohsaka *et al.*, 1994). The method could be adapted to *E. coli* cell-based expression system if orthogonal ArgRS/tRNA pair can be engineered.

Similarly, a bias codon may be assigned to a NNAA in mammalian cells that exhibit codon bias. For example, through study of human papillomavirus gene expression in different mammalian cells, Frazer and his colleagues have found that papillomavirus protein expression is determined by the codon usage and tRNA availability. Substantial differences in the tRNA pools were discovered between differentiated and undifferentiated keratinocytes (Zhao et al., 2005), and the observed bias in their tRNA may be the reason that papillomavirus replicates exclusively in epithelial cells. In CHO and Cos1 cells, wild-type papillomavirus L1 gene is effectively transcribed, but not translated. Introduction of an extra copy of Ser-tRNA$_{CGA}$ to the CHO or Cos1 (but not the empty vector) enabled the expression of L1 protein (Gu et al., 2004). Thus, it seems that TCG is a bias codon in these mammalian hosts and thus might be assigned to a NNAA.

As the codon bias phenomenon is wide-spread in different organisms, utilization of such codons for site-specific incorporation of NNAAs could be applied in many expression host cells. The limitation would be the engineering of the orthogonal aaRS/tRNA pair that matches the bias codon. The competition from a canonical amino acid for the bias codon could be minimized by using auxotrophic host cells, and/or by limiting the canonical amino acid concentration in the induction medium. Deletion (or conditional inactivation) of the tRNA genes corresponding to the bias codon in the host genome may abolish the competition. Recent advances in synthetic biology to generate artificial organisms may improve this method dramatically, as it holds promise to express target proteins at high level and with different NNAA molecules (see section 2.2.6).

2.2.4 Stop codon

Generally, protein translation terminates at one of the three stop codons (encoded by UAG (amber), UAA (ochre) and UGA (opal)) by the action of protein release factors (RF). However, occasional read-through of a stop codon with an amino acid has been observed to happen naturally in a variety of species. The suppression is caused by either mutations in the tRNA anticodon or mismatches of the codon-anticodon (Beier & Grimm, 2001). Utilization of stop codon suppression represents another way to engineer proteins containing NNAAs, and generally involves the introduction of an orthogonal aaRS/tRNA pair.

A growing number of orthogonal aaRS/tRNA pairs have been developed to introduce NNAAs site-specifically at amber codons, as it is the least frequently used stop codon in both eukaryotic (23% in humans) and prokaryotic genomes (7% in *E. coli*) (Liu *et al.*, 2007). However, ochre and opal suppressors have been evolved for the introduction of NNAAs as well (Köhrer et al., 2004). So far, over 70 NNAAs have been site-specifically incorporated into target proteins (Liu & Schultz, 2010). Typically, over 95% NNAA incorporation efficiency (defined as occupancy rate of NNAA in the full-length product) at the desired site can be obtained, making it one of the most frequently used methods for NNAA incorporation.

The main limitation of this method is protein yield. In general, yields in the range of 1-10 mg/L have been obtained, but 0.5-0.8 g/L has been reported using optimized fed-batch *E. coli* systems (Cho *et al.*, 2011). One reason for the limited yields is the low overall suppression efficiency (defined as the percentage of stop codon read-though), generally in the range of 10-40%. The low suppression efficiency also limits the number of integration sites for NNAAs in a given gene. As the number of inserted stop codons increases, the overall suppression efficiency decreases multiplicatively. Several approaches have been tested to increase the overall suppression efficiency.

One way to increase suppression efficiency is to eliminate the competition from RF. In prokaryote organisms, RF1 terminates gene translation at UAA and UAG stop codons, RF2 at UAA and UGA stop codons, and RF3 facilitates the function of RF1 and RF2 (Nakamura *et al.*, 2000). The principle of deleting RF1 in *E. coli* for amber stop codon reassignment was demonstrated by Sakamoto and his colleagues (Mukai *et al.*, 2010). As expected, deletion of the *prfA* gene, which encodes RF1, is lethal to the organism. The *prfA*- bacteria were rescued by 1) mutation of seven essential genes containing amber UAG stop codon to UAA stop codon and supplementation of these mutant genes with a bacterial artificial chromosome

(BAC) plasmid; 2) introduction of an amber suppressor tRNA such as Gln-tRNA or Tyr-tRNA. The requirement for an amber suppressor tRNA with a canonical amino acid can be extended to amber suppressor tRNA designated for a NNAA. In fact, when an archaebacterial TyrRS/tRNA$_{CUA}$ pair (engineered for 3-iodo-L-Tyr) was introduced to replace the *E. coli* amber suppressors, 3-iodo-L-Tyr was incorporated to the target protein containing six copies of the amber stop codon (Mukai *et al.*, 2010). The target protein expression level was not reported, but the full-length protein expression was demonstrated by western blot analysis. Such a result is very impressive given that six amber codons were read in the target protein, and it was only made possible by eliminating competition from the RF-1. However, the introduction of multiple mutated components in an *E. coli* host strain may be problematic for commercial manufacturing. The advent of designer organisms, containing artificial genomes, may provide the right solution to address these concerns and offer unprecedented potential for NNAA applications.

Deleting RF is not feasible in yeast and mammalian cells since eukaryotes have a single RF that functions for all three stop codons. One can minimize competition from RF by increasing the overall abundance of functional tRNA levels. However, the expression of most heterologus (usually prokaryotic) tRNA in eukaryotes is problematic due to their divergent transcriptional regulation. Eukaryotic tRNAs are transcribed by RNA Polymerase III (Pol-III) that is regulated by intragenic promoter regions known as A- and B- box elements. Most prokaryotic tRNAs do not contain these consensus sequences and are poorly transcribed in eukaryotic cells. Careful selection of a prokaryotic tRNA that contains consensus A- and B- box elements, or are permissive to mutation in order to include them, has been used successfully (Sakamoto *et al.*, 2002;Zhang *et al.*, 2004). However, the strict recognition elements of aaRSs for tRNAs often prevents modifications of the tRNA sequence to introduce these elements. Several extragenic promoters for Pol-III have been shown to allow for high levels of heterologous tRNA expression in both yeast and mammalian cells. In mammalian cells, the promoters for the small nuclear RNA U6 and H1 have been used successfully (Mukai *et al.*, 2008). But, RNA expressed by the U6 promoter has been shown to mislocalize to nucleoli (reviewed by Haeusler & Engelke, 2006) and may affect the processing and targeting of the transcript, both important processes required for the generation of functional tRNA. In yeast, RNA Pol-III -based promoters for the tRNA (e.g. SUP4), and other RNAs (RPR1 and SNP52), were shown to efficiently express tRNA genes. Levels of expression could be increased by coupling these promoters with RNA Pol-II-based promoters (e.g. PGK1). Interestingly, correct processing of the tRNA, and not absolute expression levels, is essential to generate functional tRNA and increase NNAA incorporation (reviewed by Wang *et al.*, 2009).

Another limiting factor to incorporate NNAAs at amber codon is a mechanism that degrades mRNA containing premature stop codons, called nonsense mediated mRNA decay (NMD). Since the most prevalent site-specific integration method includes amber codon read-through, the degradation of the target mRNA can have a significant impact on the yield. The importance of this process was highlighted when Wang and co-workers developed a yeast strain lacking an essential component of the NMD mechanism, UPF1. Strains lacking this gene exhibited a two-fold increase in target gene expression when compared with wild type cells (Wang & Wang, 2008). A similar approach has not been accomplished in higher eukaryotic cells.

2.2.5 Four-base-pair codon or non-natural base pair

The discovery of frame shift suppressor tRNA has opened another door to expand codons for NNAAs. In principle, a four-base codon can be used at a desired site in a target gene and read by a complementary suppressor tRNA (containing a four-base anticodon), leading to a reading frame shift and synthesis of a full length protein. Four-base codons have been used to incorporate L-homoglutamine in *E. coli*, using the orthogonal mutant *Pyrococcus horikoshii* LysRS/tRNA$_{AGGA}$ pair (Anderson *et al.*, 2004). The observed incorporation efficiency is similar to those seen for amber codon suppression (>95%) but the overall yields, in the order of 1-2 mg/L, are significantly lower. The yield of target protein with NNAA is affected by the competition of the host tRNA (containing a triplet codon) for the first 3 bases in the 4-base codon. As such, competition gives rise to a different reading frame and a protein of different sequence. To minimize this competition by the host tRNA, the 4-base codon often targets a rare codon (Anderson *et al.*, 2004). When the *P. horikoshii* LysRS/tRNA$_{AGGA}$ pair was co-expressed with an *Mj*.TyrRS/tRNA$_{CUA}$ pair for *O*-methyl-L-Tyr, target protein containing two NNAAs was generated with no appreciable loss in yield or efficiency (Anderson *et al.*, 2004). Additional variants of this approach using five-base codons have been successfully demonstrated *in vitro* (reviewed by Hohsaka & Sisido, 2002). Together, the use of four- and five-base codons will greatly increase the number determinants for NNAAs and provides one method that skirts the limitations of stop codon suppression and allows for the introduction of multiple NNAAs into a single protein. However, this approach suffers from the relative inefficiency of ribosomes to translate an extended codon. But, one can envision the development of modified, orthogonal ribosomes with increased efficiency at these codons, as described in section 2.2.7.

An alternate strategy to encode NNAAs involves the generation of codons made of new base pairs. Pioneering work by Benner and colleagues showed that new codon-anticodon pairs could be functionally introduced to support translation. By introducing a new letter to the genetic lexicon, the number of coding triplets can effectively be increased from 64 to 216. Using two complementary base-pairs, termed isoC and isoG to generate a new codon, Benner and colleagues introduced *p*-iodo-Tyr into a short peptide sequence in a cell free expression system utilizing chemically synthesized mRNA and tRNA (Bain *et al.*, 1992). But to expand the capacity for protein expression this technology requires an expansion into *in vivo* systems. This is a significant hurdle as the new base-pairs need to be orthogonal to the existing nucleotides and be accepted by DNA and RNA polymerases for faithful replication and transcription. Some of these challenges are beginning to be addressed: Benner and colleagues showed that the ribosome could translate a 65th codon (Bain *et al.*, 1992). In addition, recent progress in the design of new nucleotide pairs has been made that are specific and stable, like 2-amino-6-(2-thienyl)purine and pyridine-2-one, and can be faithfully integrated by polymerases into tRNA and mRNA, and be translated *in vitro (Hirao et al., 2002;Hirao et al., 2004)* . By combining these features with an orthogonal aaRS for the aminoacylation of tRNA with a synthetic base at its anticodon all the elements for *in vivo* transcription-translation would be in place.

2.2.6 Reserved codon in artificial cells

Incorporation of NNAAs into recombinant protein *in vivo* requires the reassignment of a codon (such as a sense codon or a stop codon) to the NNAA, and this inevitably introduces competition from the existing endogenous components: the canonical amino acid for the

sense codon and protein release factor for the stop codon. Such competition will not only reduce the NNAA incorporation efficiency, but may also reduce the target protein expression yield. Recent breakthroughs in *de novo* synthesis and implantation of an entire prokaryotic genome to create a living artificial organism (Gibson *et al.*, 2010) provide the ultimate solution for the codon competition problem.

Generation of an artificial synthetic organism via synthetic biology has many applications, as the entire genome can be tailor-designed. One such application is for NNAA technology by reserving specific codons for NNAAs and by removing the endogenous competition components. For example, the host genome may be designed with fewer than the 61 sense codons. The tRNA genes might also be minimized, potentially avoiding tRNAs such as those with wobble base-pair properties as much as possible. Such design may enable the complete reassignment of particular sense codons to a NNAA. To make better use of stop codons, host genes might be designed with as few as one type of stop codon so that the unused ones may be reserved exclusively to encode the NNAA. This will avoid the toxic effect often observed in the existing expression system with stop codon suppression, presumably caused by the functional inactivation of essential genes due to NNAA-mediated translation extension at stop codons. Furthermore, a release factor (RF) may be eliminated from the host genome to avoid premature translation termination of target proteins at the stop codon intended for the NNAA. In prokaryotes, the UAA stop codon is recognized by both RF1 and RF2, while the other two stop codons are each recognized by a single RF (UAG by RF1 and UGA by RF2). RF1 may be deleted from the artificial genome if all host genes are terminated with UGA/UAA, reserving UAG for the NNAA; alternatively, RF2 may be deleted if all artificial host genes are terminated with UAG/UAA, so that UGA is reserved for the NNAA. Deletion of RF for eukaryote organisms may need some manipulation of the wild-type eRF1, as eukaryote cells have only one eRF1 for all three stop codons.

It should be mentioned that for the majority of current expression systems, the NNAA is supplied in the culture media, from where it is transported into host cells and taken up by the aaRS. However, some NNAAs may be synthesized in the host cells by manipulating metabolic pathway(s). For example, L-homoalanine can be made biosynthetically in large quantity when a mutant glutamate dehydrogenase and an exogenous threonine dehydratase were introduced to threonine-hyperproducing *E. coli* (Zhang *et al.*, 2010). By introducing a biosynthetic pathway for the *p*-amino-Phe (pAF) from *S. venezuelae* into *E. coli* host (which also contains an orthogonal pair of pAF-RS/Tyr-tRNA$_{CUA}$), target protein with *p*-AF incorporated at the amber codon was produced (Mehl *et al.*, 2003). With deep understanding of various metabolic pathways, synthetic cells may be designed to integrate NNAA biosynthesis with NNAA incorporation, thus substantially reduce the cost of the target proteins.

2.2.7 Orthogonal ribosome

Incorporation of NNAAs via orthogonal ribosomes was lead by Chin and his colleagues, to overcome cell toxicity and low suppression efficiency associated with the amber suppression method, and to expand the numbers of NNAAs that can be introduced to target proteins. Through screening of a ribosome mutational library, they have identified ribosome

variants that recognize unique ribosome binding sequences (o-RBS). The orthogonal ribosome (o-Ribo) functions in parallel with, but independently from endogenous ribosome in host cells: o-Ribo does not translate cellular mRNAs, and the target gene cloned with o-RBS is not translated by the endogenous ribosome (Rackham & Chin, 2005). Subsequent evolution identified a mutant Ribo-x (containing U531G and U534A mutation in 16s rRNA) that improved the amber tRNA suppression of the target gene (Wang *et al.*, 2007a). By combining the Ribo-x, o-RBS-target gene that contains an amber codon, and an orthogonal aaRS/tRNA pair for NNAA into *E. coli*, these investigators were able to significantly increase the amber suppression efficiency (from 20% to >60% for a single amber codon, or from <1% to >20% for double amber codons) (Wang et al., 2007a). In addition, a ribosome variant (Ribo-Q1) that efficiently decodes both quadruplet codons and amber codons was evolved (Neumann *et al.*, 2010b). By introducing two mutually orthogonal aaRS/tRNA pairs in *E. coli* host containing Ribo-Q1, they could incorporate two different NNAAs molecules into one target protein (Neumann *et al.*, 2010b).

The protein yield obtained with orthogonal ribosome system is very impressive when compared with results obtained with wild-type ribosome. For example, full-length protein with two NNAA molecules (assigned to amber codon and a quadruplet codon, respectively) was produced at 0.5 mg/L (Neumann *et al.*, 2010b). However, the expression system contains multiple components, and its application in large scale manufacturing remains to be explored.

3. Orthogonal chemical reactions at the side chain of non-natural amino acids

One of the major advantages of incorporating NNAAs into proteins is the ability to introduce orthogonal, site-specific chemistries for modification of the recombinant protein. These modifications are often desired for applications such as PEGylation, drug conjugation, or labelling (see section 4).

Techniques for chemical modification of proteins have steadily increased in their sophistication. Early methods used electrophilic addition to the ε-amine groups of lysines. However, the large number of reactive lysines in a protein gives rise to a heterogeneous reaction product. For instance, PEG-Intron™ contains 15 positional isomers of PEGylation sites, in addition to multi-PEGylated species (Wang *et al.*, 2002). The various isomers possess 6-37% bioactivity compared to the unPEGylated protein. Conjugation technology then shifted to reductive amination at the N-terminus or thiol-specific additions at engineered cysteine residues (Harris & Chess, 2003). However, even these approaches do not result in complete limitation to a single reaction site, and they also introduce significant limitations in conjugation site and/or requirements for more protein engineering.

With the introduction of NNAAs at defined sites in a protein, a whole new world is opened up to the conjugation chemist (Sletten & Bertozzi, 2009). A new array of functional groups is available and reactions that are orthogonal to (do not react with) the functional groups in canonical amino acids present the opportunity for truly site specific conjugations. The requirements are that the necessary functional group can be incorporated into the

recombinant protein, the complementary functional group can be installed on the other molecule to be conjugated, the reaction can occur in conditions acceptable for the protein (typically aqueous, 5-37°C, mild pH, preferably in air), the reaction does not cause unacceptable side reactions (e.g., metal-catalyzed degradation of amino acids, disulfide shuffling of a folded protein), and that the chemistry is suitably bioorthogonal. Useful functional groups available in commonly incorporated NNAAs include azides, alkynes, alkenes, ketones, and aryl halides (Figure 4). The following sections highlight the most frequently harnessed NNAA conjugation chemistries in this rapidly evolving field.
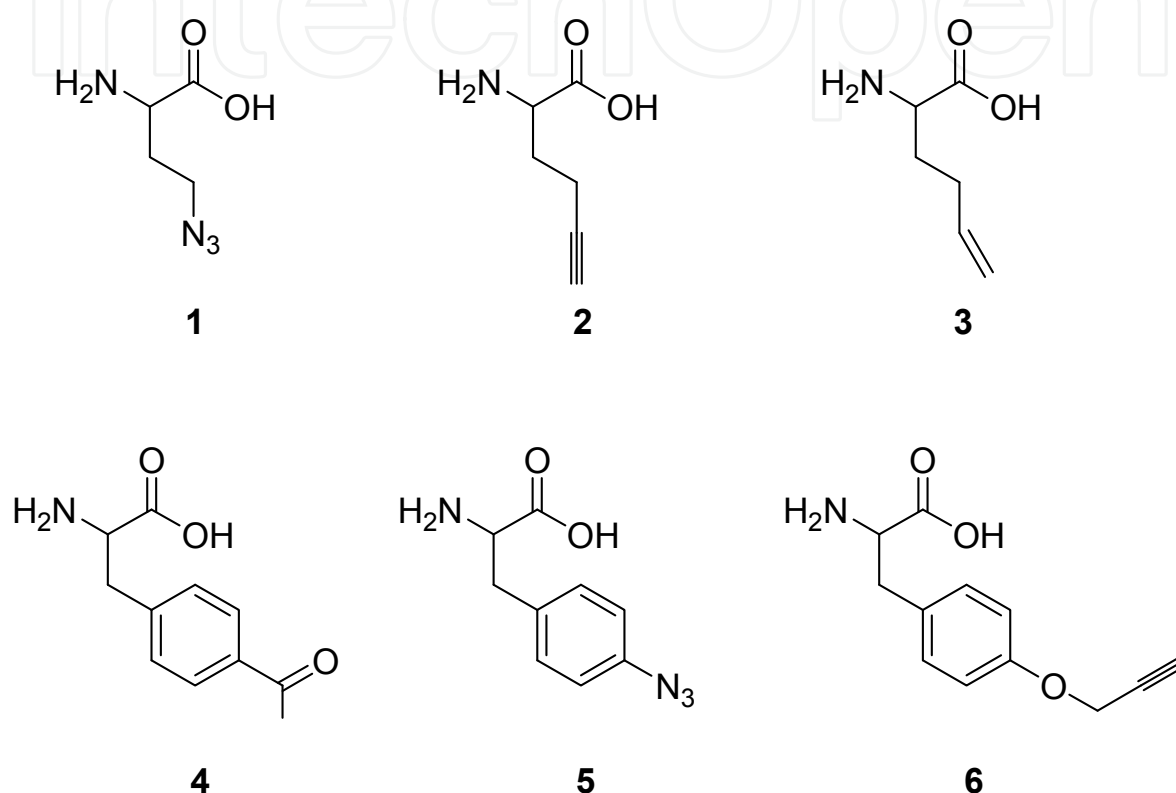
Fig. 4. NNAAs frequently used for bioconjugation. Met analogs: 1 azidohomoalanine (Aha, azide group), 2 homopropargylglycine (Hpg, alkyne group), 3 homoallylglycine (Hag, alkene group). Phe analogs: 4 p-acetyl-Phe (ketone group), 5 p-azido-Phe (azide group), 6 p-propargyloxy-Phe (alkyne group).

3.1 Copper-catalyzed azide-alkyne cycloaddition ("click" chemistry)

Copper-catalyzed azide-alkyne cycloaddition (often referred to as "click" chemistry or CuAAC) was first reported simultaneously by both Sharpless and Meldal and co-workers (Rostovtsev et al., 2002; Tornoe et al., 2002). Click chemistry is a term coined by the Sharpless group referring to a philosophy of efficient chemical synthesis, but has since been co-opted to also refer specifically to CuAAC. CuAAC joins a terminal alkyne and an azide, via a Cu(I) catalyst, to form an aromatic triazole ring (Figure 5). Huisgen had previously demonstrated catalyst-free cycloaddition of azides and alkynes at high temperature (e.g., 100°C), but with the addition of the copper catalyst, the reaction occurs readily at room temperature and forms only the 1,4 regioisomer. It is essential that the alkyne is terminal, so the copper acetylide can be formed as a reaction intermediate.
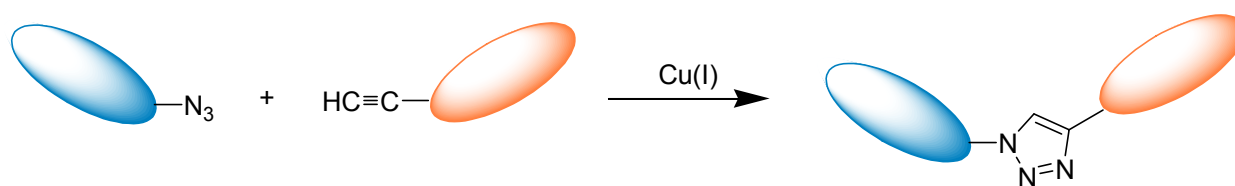
Fig. 5. Copper-catalyzed azide-alkyne [3+2] cycloaddition ("click" chemistry or CuAAC). An azide and a terminal alkyne combine to form an aromatic triazole linkage.

There are several advantages to CuAAC for bioconjugation. Azides and alkynes do not react with canonical amino acids, are kinetically stable, and yet are highly primed to react when presented with the appropriate catalyst. Either functional group is readily installed by the simple codon reassignment method with commercially available NNAAs. At physiological temperatures, the reaction does not occur in the absence of the catalyst. Importantly, the resultant linkage between the protein and the conjugated molecule is a triazole group, which is highly stable due to its aromaticity. The reaction can be performed in many aqueous buffers at a wide range of pHs, in many organic solvents, in the presence of SDS (Grabstein et al., 2010), and, in most cases, in the presence of oxygen. Importantly, the reaction displays especially rapid kinetics (Presolski *et al.*, 2010).

As stated, CuAAC is catalyzed by Cu(I). Cu(I) is not stable in water, but rapidly disproportionates into Cu(II) and Cu(0). Cu(I) salts such as CuBr and $[Cu(NCCH_3)_4][PF_6]$ have been used successfully, though the literature suggests ultra high purity may be required (Link *et al.*, 2004). More frequently, the cupric salt $CuSO_4$ is used in conjunction with a reducing agent such as TCEP or elemental copper. Thiol-containing reducing agents such as DTT can also be used (Grabstein et al., 2010). Finn and colleagues have recently developed the use of aminoguanidine as an additive to prevent deleterious side reactions when using sodium ascorbate as a reducing agent (Hong *et al.*, 2009).

A ligand is typically added to greatly increase reaction yields. The ligand is thought to complex and stabilize the Cu(I) and intermediates and also to protect the protein from deleterious effects of copper. The most commonly used ligand is TBTA, shown in Figure 6. Significant protein breakdown can be seen after reaction with no ligand, whereas breakdown can be prevented by TBTA (unpublished data). TBTA is very hydrophobic, but solubility is not required for the reaction. Many other ligands have been investigated in the literature, including water soluble analogs such as THPTA and benzimidazoles. Finn and colleagues have had good success with a water soluble bathophenanthroline ligand (Gupta *et al.*, 2005), but this requires careful removal of oxygen from the system. A thorough structure-function investigation of ligand acceleration generated recommendations for different systems (Presolski *et al.*, 2010). Dilute reactions in aqueous media have very different requirements than concentrated reactions in organic solvents due to the varying degree of copper binding competition in the reaction. The literature and the authors' own experience suggests that the best ligand must be determined empirically for any given reaction, which will depend on the protein, solvent, and other conditions.

While CuAAC can be readily employed in many systems, there are some side reactions that can be problematic. First, copper is well known to create oxidative species in aqueous solution, including hydrogen peroxide and free radicals. These effects can be mediated by

choice of ligand, reducing agent, and other reaction conditions. Copper's toxicity can also be problematic for reactions on live cells or in living organisms. Second, the presence of reducing agents (or even the copper itself) can disrupt disulfide bonds present in a folded protein. This can lead to misfolding or disulfide-linked aggregates, though it has been shown that some proteins such as antibodies can be exposed to reducing agents and then oxidized to re-form the correct disulfide bonds (Sun *et al.*, 2005). Third, Back and co-workers have shown that reducing agents can result in azide reduction to an amine group or can cleave the peptide bond C-terminal to Aha, wherein the Aha is modified to a homoserine lactone which can then react with amine groups (Back *et al.*, 2005). High pH can also result in direct replacement of the azide group with a hydroxyl group. We have developed reaction conditions that avert all of the above issues, apart from toxicity in living systems, for various protein conjugation systems (unpublished data). Recent publications have also demonstrated conditions that allow short-term CuAAC in the presence of live cells or organisms (Hong *et al.*, 2010; del Amo *et al.*, 2010).
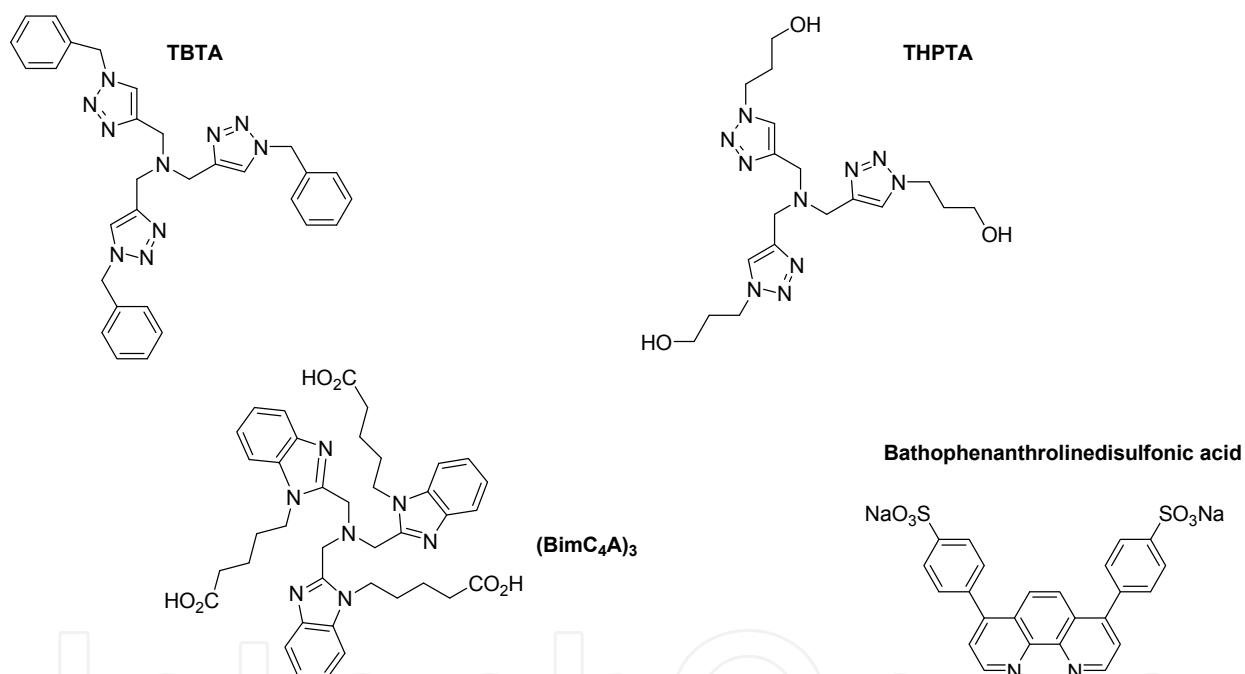

Fig. 6. Popular ligands for CuAAC

3.2 Strain-promoted azide-alkyne cycloaddition

Strain-promoted azide-alkyne cycloaddition (SPAAC) is popular as a metal-free alternative to CuAAC (Jewett & Bertozzi, 2010). While CuAAC uses a copper catalyst to access the reactivity of the alkyne, the [3+2] cycloaddition activation barrier can also be lowered by placing strain on the alkyne group. Bertozzi and colleagues first pioneered this approach by using a highly strained (18 kcal/mol of ring strain) cyclooctyne ring (Agard *et al.*, 2004). As shown in Figure 7, the alkyne is internal rather than terminal, the resultant linkage is comprised of two fused rings, and no metals, reducing agents, or ligands are necessary. SPAAC is of particular interest for labelling applications in living systems in order to avoid a toxic copper catalyst.
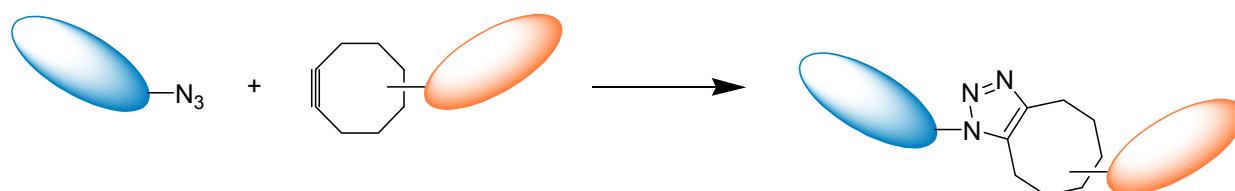
Fig. 7. Strain-promoted azide-alkyne [3+2] cycloaddition (SPAAC). An azide and a strained cyclooctyne ring combine to form a fused ring linkage.

SPAAC has undergone significant improvements beyond the slow reactivity and difficult synthesis of the original cyclooctyne ring. Installation of electron withdrawing groups on the cyclooctyne or additional ring strain provided by fusing ring(s) to the cyclooctyne has improved reaction kinetics, and a practical focus has improved synthetic accessibility of the starting materials. A sampling of structures is given in Figure 8. An initial advance occurred with DIFO, where the introduction of two propargylic fluorine atoms resulted in more than an order of magnitude increase in the second-order rate constant. Second generation DIFO analogs displayed similar reactivity but synthetic yields were improved from 1% in 12 steps to 28-36% in 6 steps. Increasing ring strain by fusing two aromatic rings with the cyclooctyne has also increased reactivity to similar levels, such as with DIBO. Addition of a nitrogen atom (DIBAC) or amide group (BARAC) within the central cyclooctyne ring increases the reactivity even further (Debets *et al.*, 2010;Jewett *et al.*, 2010). Alternatively, ring strain introduced by fusion of a 3-membered ring with the cyclooctyne generates a bicyclononyne (BCN) with similarly high reactivity (Dommerholt *et al.*, 2010). This molecule was made in 4 steps with an overall synthetic yield of 46% and, due to its symmetry, its cycloaddition reaction product is a single regioisomer. These improved reagents have now brought SPAAC reactivity to a similar level as CuAAC. A much more water soluble reagent was generated by the addition of 3 heteroatoms to the single cyclooctyne ring (DIMAC), but the reactivity was significantly lower than other recent analogs.
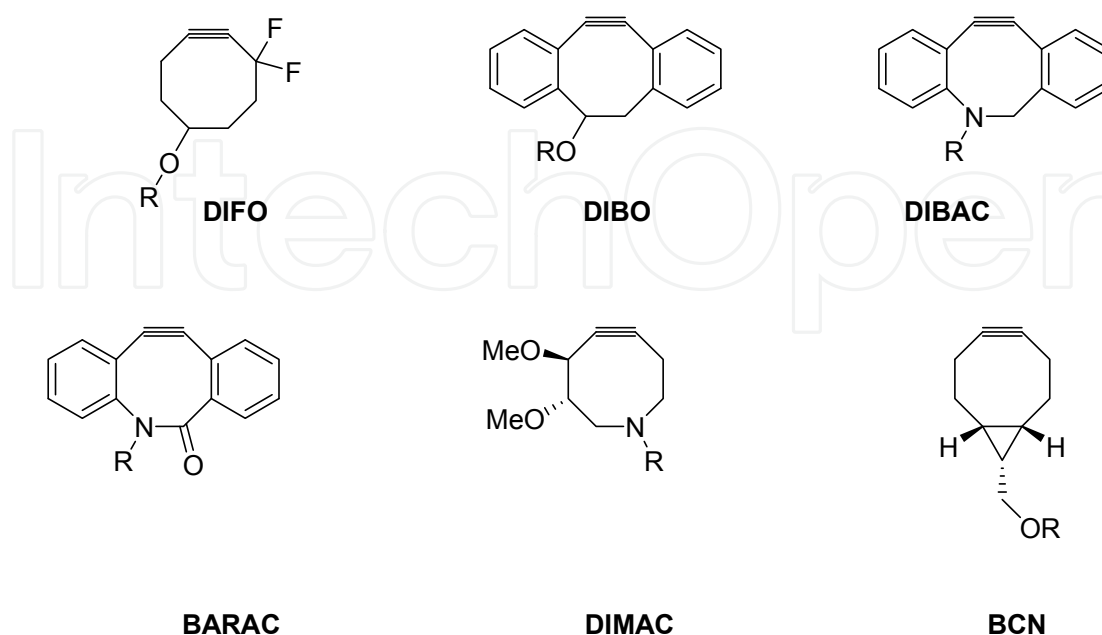

Fig. 8. Improved strained alkynes for use in SPAAC.

SPAAC can expand the application of azide-alkyne cycloaddition to metal-sensitive systems, but it does present some challenges. Unfortunately, many of the modifications that reduce the activation barrier enough for fast reaction rates also reduce the barrier for side reactions with canonical functional groups such as cysteinyl thiols. This loss of orthogonality can be addressed by tailoring of the activated alkyne and reaction conditions, but ultimately the stringency of the orthogonality requirement is dependent upon the application. In addition, the cyclooctyne ring is a large hydrophobic structure with limited aqueous solubility. It has been shown to nonspecifically bind to proteins even in the presence of SDS (Sletten and Bertozzi, 2008). The fusion of additional rings only serves to magnify this problem (Agard *et al.*, 2006;Codelli *et al.*, 2008). Commercial availability of cyclooctyne reagents is also limited. For applications such as therapeutic protein conjugates, other considerations include the stability of the activated alkyne, its synthetic availability and purity, and the production of regioisomers in the reaction product. Importantly, the final fused ring linker that is generated is bulky, rigid, hydrophobic, and still contains some strain. Whether this presents any challenges for immunogenicity or stability of a therapeutic conjugate remains to be investigated.

3.3 Carbonyl chemistries

Carbonyl chemistries for creating oxime or hydrazone linkages are well established methods for protein conjugation. With this approach, a ketone or aldehyde handle is reacted with an aminooxy or hydrazide group to form an oxime or hydrazone linkage (Figure 9). The reaction is extremely selective with the exception that N-terminal cysteines will undergo thiazolidine formation. The reaction is typically performed around pH 4-5 to facilitate protonation of the carbonyl but maintain the nucleophilicity of the aminooxy group. No metal catalysts are required and the final linker is a small inconspicuous group. The recent discovery of aniline as a nucleophilic catalyst of this reaction was key to extending the utility of this reaction, even making conjugation at neutral pH viable (Dirksen & Dawson, 2008).
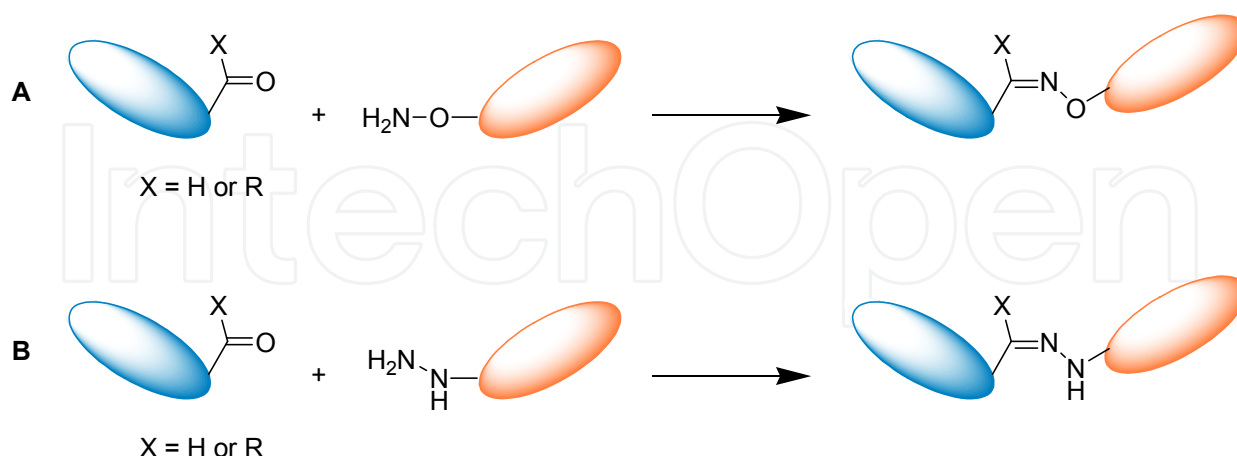

Fig. 9. Carbonyl chemistries, X = H for aldehydes, X = R for ketones. (A) A carbonyl and an aminooxy group combine to form an oxime linkage. (B) A carbonyl and a hydrazide group combine to form a hydrazone linkage.

The final oxime and hydrazone linkages can degrade, undergoing hydrolysis to regenerate the starting materials (Kalia & Raines, 2008). Hydrazone half lives can be on the order of

minutes to hours, so oximes are the linkage of choice for stable bioconjugates. The half life of a simple oxime was reported as ~25 day at pD 7.0 but dropped to 15.7h at pD 5.0 due to the acid catalysis of this degradation. The chemistry of neighboring groups (electron withdrawing or donating, aromaticity) can be tailored for desired stability. Importantly, researchers at Ambrx have recently reported stability for more than 1 year at pH 7.3 of a PEGylated growth hormone containing an oxime linkage through *p*-acetyl-Phe (Cho *et al.*, 2011). On the other hand, the instability of the hydrazone can be used as a tool, such as for temporary labelling or purification purposes. The ability of the final linkage to hydrolyze can also be a potential advantage if a therapeutic conjugate is desired to eventually degrade *in vivo*. Another problem when working with proteins containing carbonyl groups is reversible imine formation between the carbonyl group and lysinyl amino groups.

While the optimal reaction pH of 4-5 is not extreme, it is not desirable for some proteins, especially if this is near the pI. The reaction is often performed at a suboptimal pH in order to keep the protein in solution. When hGH (pI ~5) was PEGylated by oxime chemistry, 15% DMF was added and the reaction was performed at pH 6 to prevent precipitation, necessitating an amazing 10 days at 30°C for the reaction (Peschke *et al.*, 2007). When the same investigators utilized CuAAC, the reaction took place over 22h with a much higher final yield of protein. The final conjugate also had a higher *in vitro* potency after CuAAC, perhaps due to the suboptimal conditions the protein was exposed to for the oxime conjugation.

3.4 Staudinger ligation

Bertozzi and colleagues have pioneered the Staudinger ligation as another orthogonal conjugation chemistry reactive to azide-containing NNAAs (Saxon & Bertozzi, 2000). In this reaction, a triarylphosphine moiety forms an aza-ylide intermediate (iminophosphorane) with the NNAA (Figure 10). A methyl ester group *ortho* to the phosphorous atom provides an electrophilic trap for the nitrogen, causing rearrangement to the final amide linkage.

Raines and colleagues have since developed a traceless Staudinger ligation (Figure 10), wherein a phosphinothioester reacts with the azide to form an aza-ylide, which undergoes a rearrangement to form an amide bond with release of the phosphine oxide (Tam & Raines, 2009). Thus, no "trace" of the phosphine group remains in the conjugated product. Various water soluble phosphinothiols are now available, incorporating amine groups that also favor the ligation over hydrolysis.

The Staudinger ligation has benefits including the lack of a metal catalyst. No ligation occurs with canonical amino acids and the reaction has been shown to work even on crude cell lysates. Particularly appealing in the case of the traceless ligation is that a simple amide bond is the reaction product.

Challenges with the Staudinger ligation include competing side reactions and the nature of the phosphine reagent. The design of the phosphine group is important to drive ligation rather than the competing classical Staudinger reaction. In the classical reaction, hydrolysis of the aza-ylide results in reduction of the azide to an amine and release of the phosphine oxide. The ligation works best around pH 8, presumably due to higher hydrolysis competition at acidic pH. Phosphines are also known to reduce disulfide bonds. Fortunately, triarylphosphines are mild reducing agents, and were shown not to induce significant

disulfide reduction on a cell surface (Saxon & Bertozzi, 2000). Aryl phosphines are, however, very bulky (affecting reaction rates on protein surfaces), hydrophobic (affecting solubility of the reagent), and still prone to oxidation by air (requiring special handling). In the case of the nontraceless ligation, the aryl phosphine leaves a bulky hydrophobic signature in the linkage.
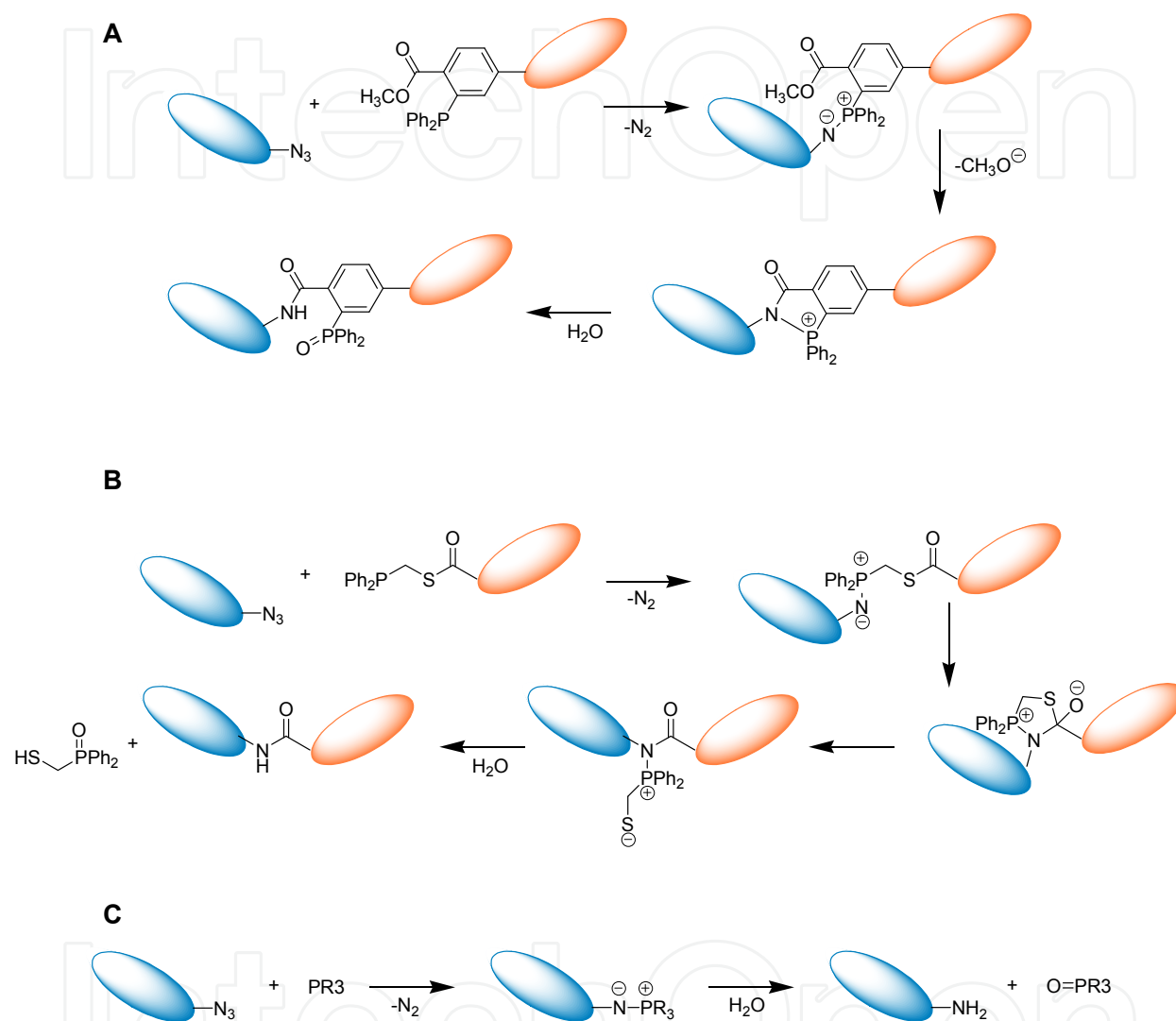

Fig. 10. (A) Staudinger ligation – an azide and a triarylphosphine moiety bearing an electrophilic trap combine to form an amide linkage bearing the triarylphosphine oxide. (B) Traceless Staudinger ligation – an azide and a phosphinothioester combine to form an amide linkage. (C) Classical Staudinger reaction – an azide reacts with a phosphine to form an amine and a phosphine oxide.

Reaction kinetics are perhaps the most significant shortcoming of the Staudinger ligation. The rate is orders of magnitude slower than other available chemistries. Attempts to accelerate the reaction by increasing the electron density of the phosphine substituent also increased the air instability of the reagent. The slow reaction rates, combined with air instability and competing hydrolysis, may be why the Staudinger ligation is used less frequently than other conjugation chemistries.

3.5 Other reactions

Due to the abundant applications, new chemistries are constantly in development for site specification conjugation to proteins containing NNAAs. Summarized below are a number of interesting approaches, some of which await further refinement before implementation as a readily available bioorthogonal strategy.

3.5.1 Metal-free

A tandem 1,3-dipolar cycloaddition-retro-Diels-Alder (tandem crDA) has recently emerged as an intriguing metal-free alternative to CuAAC (van Berkel *et al.*, 2008). In this reaction, an azide group is reacted with a strained and electron deficient oxanorbornadiene derivative. The azide undergoes [3+2] cycloaddition with one of the double bonds, which is followed by elimination of a furan in a retro-Diels-Alder reaction, leading to formation of a stable 1,2,3-triazole. While an elegant approach, the reaction can take days to complete and side reactions with basic canonical amino acids can occur. Most importantly, the addition can also take place over the unsubstituted double bond in the strained ring; this results in elimination of the desired substituent and the conjugate is not formed. Hopefully further refinements will enable this chemistry for more general use.

Cycloadditions between alkenes and tetrazines or tetrazoles have also been an active area of research for bioorthogonal conjugation. In the inverse electron demand Diels-Alder cycloaddition, a strained alkene dienophile undergoes cycloaddition with a tetrazine, leading to a fused ring conjugate and elimination of N_2 (Blackman *et al.*, 2008;Devaraj *et al.*, 2008). The reaction is fast, selective, and metal-free. However, the reaction requires an alkene (such as *trans*-cyclooctene or norbornene) with much more strain than NNAAs that have been incorporated *in vivo* (such as homoallylglycine). One solution has been to use a tetrazole that can be photoactivated to produce a nitrile imine intermediate that then reacts with unactivated alkenes (Song *et al.*, 2008). This has been used successfully with a Z-domain protein containing *O*-allyl-Tyr expressed from *E. coli*.

The cycloadditions of strained alkynes or alkenes with nitrile oxide have also been investigated recently (Gutsmiedl *et al.*, 2009). The reaction also occurred with an unstrained alkyne and *in situ* formation of the nitrile oxide by hypervalent iodine, but was sluggish (Jawalekar *et al.*, 2011). This approach suffers from a lack of incorporatable NNAAs containing the ideal functional groups and also the nitrile oxide can undergo side reactions with other functional groups or homodimerize with itself.

As an alternative to SPAAC, appendage of electron-withdrawing groups neighboring an alkyne can also enable azide-alkyne cycloaddition without a metal catalyst (Li *et al.*, 2004). Similarly to SPAAC, refinement of the local environment of the alkyne is necessary to promote a desirable reaction rate while maintaining orthogonality, but this approach avoids the use of bulky rings in the reactants.

Various photoactivated chemistries have been used, either with photoreactive NNAAs or by generation of reactive functional groups through irradiation. However, photoirradiation is well known to generate other reactive species such as free radicals that could cause side reactions with proteins and can also be difficult to scale up for a manufacturing process, and thus is not discussed in depth in this practical review.

Thiol-ene coupling to form a thioether linkage between thiols and alkenes has also gained recent popularity (Dondoni, 2008). Photoirradiation or free radical initiators are used to generate a reactive thiyl radical. This approach has been recently used to generate various bovine serum albumin conjugates. A similar thiol-yne coupling with alkynes can also be utilized (Minozzi *et al.*, 2011). For the application of thiol-ene/yne chemistry to NNAA-containing proteins, it would be possible to incorporate an alkene- or alkyne- containing NNAA and conjugate it to a thiol-containing substituent. However, given the prevalence and importance of cysteine residues in most proteins, and the side reactions to be expected by deliberate free radical generation, the generalizability of this approach for bioorthogonal reactions with NNAA-containing proteins is limited.

3.5.2 Metal-catalyzed

The family of palladium-catalyzed cross coupling reactions has great potential for bioorthogonal conjugations. These include couplings between an aryl or vinyl halide and an alkyne (Sonogashira), alkene (Heck), or boronic acid (Suzuki) group (Genet & Savignac, 1999). The Sonogashira coupling typically also includes a Cu(I) cocatalyst. The simple carbon-carbon linkage that is formed is very appealing with respect to stability and bulk. Classically these reactions are performed in organic solvent, under inert atmosphere, and with significant heat. However, recent advances in this rapidly evolving area have enabled these reactions for mild aqueous conditions.

These palladium-catalyzed reactions have now been employed for protein modification with varying success. Ras protein containing 4-iodo-Phe has been conjugated, via Heck or Sonogashira couplings, with vinylated or propargylated biotin (Kodama *et al.*, 2007). The reactions yielded only 2 - 25% product and the addition of DMSO, $MgCl_2$, and tyramine was required to reduce protein degradation and to suppress cysteine interference. More success has been seen with Suzuki couplings, performed with up to quantitative yields with 4-iodo-Phe-containing proteins (Ojida *et al.*, 2005;Chalker *et al.*, 2009). Issues with protein precipitation, dehalogenation, and cysteine interference could still be seen. Less success was seen when coupling a boronic acid-containing protein by Suzuki coupling (Brustad *et al.*, 2008). An important limitation to all Pd-catalyzed reactions is that the catalyst is easily poisoned by thiols and the presence of free cysteines (and potentially even cystines or methionines) can often block the reaction completely.

Ruthenium has emerged as a different metal that can catalyze the cycloaddition of azides and alkynes (Zhang *et al.*, 2005). Unlike CuAAC, the reaction is catalyzed by the metal in the +2 oxidation state and internal or terminal alkynes can be used. Also, depending on the catalyst complex used, the reaction product can be the 1,4- or the 1,5- disubstituted regioisomer of the triazole. One potential side reaction is oligomerization of the alkyne. RuAAC has been used to ligate pancreatic ribonuclease with a *cis*-peptide bond surrogate (Tam *et al.*, 2007).

Olefin cross-metathesis, the exchange of fragments between double-bond-containing molecules, has recently been applied to protein conjugations (Lin *et al.*, 2009). This has been enabled by the development of ruthenium-based catalysts that are air stable and active in water. Davis and co-workers recently demonstrated the first cross-metathesis coupling reactions with a protein (Lin *et al.*, 2010). A protein containing *S*-allylcysteine reacted quantitatively with alkene-containing glycosides after short times at 25-37°C. The reactions,

however, took place in 30% *tert*-butyl alcohol and required addition of $MgCl_2$ as a mild Lewis acid to prevent non-productive chelation. The proximity of the sulfur seems to be critical for rate enhancement. Unfortunately, homoallylglycine, which can be readily incorporated into proteins, is much less reactive. Olefin metathesis is attractive due to the simple alkene bond that is formed between the substituents. Potential issues with olefin cross metathesis include self-metathesis homodimerization side reactions and interference from sulfur atoms (cysteine, cystine, methionine).

4. Potential applications of proteins engineered with non-natural amino acids

Protein engineering with NNAAs harnesses the vast divergent array of new protein building blocks and brings forth many new applications, such as monitoring protein functions *in vivo* via probe-labelling, trapping transient protein-protein interactions via cross-linking, facilitating structural determination and proteomic studies for newly synthesized proteins. In addition, it may also find practical applications in protein purification, materials science, diagnostics, drug screening, and green energy. In this section, we will highlight its applications on protein therapeutics and protein functional evolution.

4.1 Protein therapeutics

Protein therapeutics represent a $72 billion market with more than 130 products approved (Leader *et al.*, 2008). The ability of complex proteins to exquisitely target disease pathways continues to drive rapid development despite their high cost of goods. While the proteins themselves can offer a strong therapeutic effect, additional modifications such as polymer or drug conjugation can improve or add new pharmacodynamic effects. A number of conjugated biotherapeutics are already approved for clinical use, including PEGylated proteins, antibody-drug conjugates, and saccharide-protein conjugate vaccines.

First generation conjugate therapeutics suffer from a lack of control over the conjugation site and current research is focused on enabling precise direction to a chosen site of modification. The expression of proteins with NNAAs at predetermined locations lends itself readily to these applications. While many relevant NNAA bioconjugations have been reported in the literature, a number of issues must be addressed when considering translation to a clinical product. The conjugation reaction must use conditions that are amenable to scale-up, use reagents that can be obtained with sufficient purity and scale, have cost effective reactant ratios with high yields, and avoid side reactions that can be problematic at even very low levels. In addition, the nature of the linkage that is formed between the protein and the conjugated species is important with respect to shelf life and *in vivo* stability as well as to immunogenicity of the product.

In the following sections we provide an overview of NNAA conjugate therapeutic approaches that are readily contemplated or have already been reduced to practice. We also provide an example of therapeutic effect derived from incorporation of an unmodified NNAA.

4.1.1 Modifications for improved pharmacokinetics

Increasing the residence time of protein therapeutics in the body has been a focus of biotherapeutic development for some time. Because these pharmaceuticals are typically administered by injection, decreasing the dosing frequency from daily to weekly or monthly

is highly desirable. This increases patient comfort and compliance, and the maintenance of a more steady plasma concentration of the drug, rather than frequent peaks and troughs, can increase the therapeutic effect (Harris & Chess, 2003).

Conjugation of polyethylene glycol (PEG) is the current gold standard for protein half-life extension (Fishburn, 2008). There are now at least eight approved PEGylated protein therapeutics in the US. PEGylation of proteins is known to decrease renal clearance, decrease enzymatic degradation, reduce immunogenicity of the protein, and increase serum half-life due to the steric hindrance of a PEG "cloud" around the protein. Upon PEGylation, the half-life of a protein may increase by a few-fold to even a thousand-fold. Consequently, even though PEGylation typically results in decreased *in vitro* activity, *in vivo* efficacy is usually greatly enhanced. Alternative polymers for conjugation are under investigation, including hydroxyethyl starch.

The site of conjugation to a protein has important implications. PEGylation near a binding or catalytic site of the protein can strongly reduce binding affinity and *in vitro* potency. As discussed in sections 3 and 4.1.3, protein PEGylation can occur through reactive groups such as lysinyl amines or cysteinyl thiols, but this provides poor control over the site of and number of modification(s). Even reductive amination, which is designed to be selective for the N-terminal amine due to its lower pKa, still results in PEGylation at additional sites on the protein (Gronke et al., 2010). While multi-PEGylated species can often be purified from mono-PEGylated species, separation of positional PEG isomers (monoPEGylation at different sites on the protein) is unfeasible in commercial production. The resultant heterogeneous product contains isomeric molecules with a range of potencies and provides a challenge for batch-to-batch reproducibility. Thus, the field is rapidly moving toward true site specific modification to allow production of a single designed homogenous product.

PEGylation has now been performed through NNAA conjugation, and a few companies are advancing these products in the clinic. Allozyne has reported a PEGylated interferon beta-1b (IFNβ) produced by incorporating a single Aha as a reassigned sense codon for Met in *E. coli* and then attaching a PEG-alkyne via CuAAC (Nairn et al., 2007). We have since produced a PEG-IFNβ candidate under cGMP at >800L fermentation scale (unpublished data). Up to 98% conjugation of Aha-containing IFNβ was achieved using a low molar excess of reactive 40 kDa branched PEG. No PEGylation was seen at any site besides the Aha and residual copper in the final product was <0.1 ppm with no special efforts for removal. After 18 months storage at 5°C, there was no detected unPEGylated IFNβ in the Drug Product, indicating a very stable linkage. Phase Ia trials showed the resultant half-life extension may allow monthly dosing, whereas the unPEGylated analog Betaseron® is dosed every other day. Ambrx recently reported clinical data of a PEGylated human growth hormone that supports weekly dosing. This molecule was made by amber suppression in *E. coli* with incorporation of *p*-acetyl-Phe at position 35 followed by reaction with linear 30 kDa aminooxy-PEG to create an oxime linker (Cho *et al.*, 2011). As has been seen for other PEGylated proteins, variations in the PEGylation site had a significant effect on binding affinity and *in vitro* bioactivity; a systematic comparison *in vitro* and *in vivo* was performed in order to select the site taken into the clinic. Fermentation was performed at >1000L scale with 0.5-0.8g/L titers and the linkage was shown to be stable for more than a year at 5°C.

4.1.2 Bispecific protein conjugates

Bispecific and multivalent proteins are receiving growing interest as therapeutic candidates (Marvin & Zhu, 2005;Filpula, 2007). Antibody-like molecules with specificity for two different epitopes show promise for targeted cancer therapy, such as by bringing cancer cells and effector cells into close contact. Combination of two antibody binding domains also has potential for immunotherapy, such as by simultaneously binding two cytokines or to regulate multiple pathways by simultaneously binding cell surface receptors. Other applications include toxin conjugates, where a proteinaceous toxin is linked to a targeting protein such as an antibody binding domain or a targeting ligand. As another approach, multivalency of the same protein can improve avidity, targeting, and function.

Bispecific or multivalent proteins are typically constructed by expression of genetic fusions or by conjugation through an engineered cysteine residue. Limitations of genetic fusions can include mis-folding, mis-pairing of antibody chains, low fermentation yields, and a lack of choices for relative orientation between the proteins, as the linkage can only occur between the N-terminus of one protein and the C-terminus of the other (Griffiths *et al.*, 2004;Marvin & Zhu, 2005). Engineered cysteine residues can result in mis-disulfide bond formation (Kim *et al.*, 2008) and a complex reaction product (Natarajan *et al.*, 2005).

The use of NNAAs for protein-protein coupling enables production of a single conjugate with any desired orientation. It requires that the two target proteins be produced in separate fermentations with 2 complementary NNAAs. This can allow independent folding of the two proteins prior to conjugation if desired to avoid mis-folding events.

There are a few examples of protein-protein conjugation through incorporated NNAAs. DHFR and GFP (or 2 GFPs) have been directly conjugated through CuAAC between incorporated *p*-propargyloxy-Phe and *p*-azido-Phe residues (Bundy & Swartz, 2010). Up to 62% conjugation yield was achieved with a 1:1 molar reactant ratio. Not surprisingly, use of a ligand to protect the protein from the copper and removal of the copper prior to assay were important to demonstrate that bioactivity was retained. An scFv and a luciferase reporter protein have also been conjugated by CuAAC between incorporated Hpg and *p*-azido-Phe residues (Patel *et al.*, 2009). Another interesting approach was to react *p*-acetyl-Phe genetically incorporated in an anti-HER2 Fab domain with hydroxylamine biotin to form an oxime linkage (Hutchins *et al.*, 2011). This conjugate was then assembled into a tetravalent species through incubation with deglycosylated avidin. By incorporating the NNAA at various sites, the authors showed that the orientation of the protein assemblage is important, with 5-50 fold increased in phosphorylation inhibition versus the monovalent parent Fab depending on the conjugation site.

4.1.3 Antibody-drug conjugates and radioimmunoconjugates

The chemical conjugation of highly potent cytotoxic drugs to monoclonal antibodies has proven to be effective means of targeted delivery of chemotherapeutics that would otherwise be too toxic for systemic administration (Alley *et al.*, 2010). The key attributes of antibody-drug conjugates (ADCs) include the specificity of the antibody for the tumor, the potency of the toxin, and the linker, which must release the toxin conditionally upon internalization into the tumor cell. ADCs are typically linked via acylation of lysines, alkylation of reduced interchain disulfides, or alkylation of introduced unpaired Cys

residues. Each of these methods leads to some degree of heterogeneity of the product. Modification at native Cys residues produces a mixture of molecules including 0, 2, 4, 6, and 8 drugs, where the final conjugate also contains less stabilizing disulfide bonds than the parent (Sun *et al.*, 2005). The therapeutic index will be highest for a molecule with the ideal number of drugs attached, for example, the maximum tolerated dose and clearance rate improved by 4-5 fold when comparing ADCs containing exactly 2 auristatin analogs versus 8 (Hamblett *et al.*, 2004). One approach to limit product heterogeneity is to replace some of the natural cysteines with serines (McDonagh *et al.*, 2006). Engineering in new Cys residues has also allowed more control over the degree of conjugation, but the reaction product still carries a distribution of species (Junutula *et al.*, 2008).

The ability to introduce a NNAA at any location in an antibody (or Fab, scFv, etc.) provides several advantages for ADC construction. The NNAA creates specific sites where the toxin is conjugated, leaving all other residues and the disulfide bonds unmodified. These sites can be positioned at favorable locations that are permissive for full function of the antibody. The number of toxins per antibody can also be efficiently adjusted without sacrificing the yield or homogeneity of the product. There are, to our knowledge, no literature reports of ADC construction with genetically incorporated NNAAs, but Allozyne and Ambrx are both pursuing undisclosed ADC targets for oncology.

Radioimmunoconjugates for radioimmunotherapy (RIT) are similar to ADCs, but a chelating agent for a radionuclide is conjugated instead of a toxin (Martin *et al.*, 2010). The design principles are similar to ADCs, where the use of NNAAs to control the location and number of attached chelating agents can be expected to substantially improve the product. One significant advantage of RIT is that internalization is not required. Thus, there are a greater number of available targets. In addition, smaller targeting proteins, such as scFvs may be used, which should provide for better tumor penetration.

4.1.4 Cancer vaccines

Cancer vaccines hold great promise to treat metastatic tumors effectively and to elicit long-term protection against recurrence (Farkas & Finn, 2010). However, because most tumors are derived from normal cells, the identification of true tumor specific antigens (absent in all normal tissues) is difficult. Instead, many tumor associated antigens, especially those self-proteins over-expressed in tumor, have been identified and used for immunotherapy. While it has been shown that cancer vaccines can be effective, it is still a challenge to elicit effective immunity against the over-expressed tumor associated antigens because of self-tolerance.

Introduction of NNAAs into tumor antigens offers potential strategies to break such self-tolerance, since the NNAA may be recognized as a foreign entity by the immune system. Such a theory is supported by animal data (Grunewald *et al.*, 2009): when the wild-type murine tumor necrosis factor-α was used as an immunogen, mice failed to generate IgG antibodies; however, when the mutant protein containing only one NNAA (*p*-nitro-Phe) was used, sustained polyclonal IgG antibodies against both the wild-type as well as NNAA-containing protein antigens were generated. The resulting immune response protected mice from severe endotoxemia induced by lipopolysaccharide (LPS) challenge. Therefore, vaccination with NNAA-modified self-protein can break self-tolerance due to epitope spreading, resulting in polyclonal antibodies against the native protein.

Incorporation of NNAAs into antigens may also improve vaccine efficacy by increasing the antigen stability *in vivo*. This may be of special importance to peptide-based vaccines due to the short half-life of peptides. Peptides designed with NNAAs may adopt bioactive conformation with enhanced stability, and thus elicit immune response to wild-type tumor antigens (Vichier-Guerre *et al.*, 2004;Corzana *et al.*, 2011).

4.2 Functional evolution of proteins with NNAAs

Engineering proteins for desired biophysical properties, enhanced or novel functions is of central importance, not only for basic scientific research, but also for biopharmaceutical drug development. The availability of the vast number of NNAAs as building blocks greatly expands our ability to customize proteins for different purposes. Several strategies for functional evolution of proteins with NAA are described below.

Rational-design (based on the known structural information of target proteins) is an efficient way to evolve proteins with NNAAs. For example, based on the hyper-hydrophobic and near isosteric properties of fluorinated amino acids to their natural counterparts, trifluoroleucine (Tfl) introduced to leucine zipper peptides (replacing all Leu residues) has been shown to increase the stability while maintaining their DNA-binding function (Tang *et al.*, 2001). Molecular dynamics simulation of the thermodynamic properties of the fluorinated leucine zipper peptide indicated that Tfl-substituted structure is indeed more stable than the wild-type, correlating well with experimental data (Tang *et al.*, 2001). Another impressive example involves replacing a single amino acid (Tyr 309) in the bacterial enzyme phosphotriesterase with NNAAs (L-(7-hydroxycoumarin-4-yl)ethylglycine and L-(7-methycoumarin-4-yl)ethylglycine amino acid) (Ugwumba *et al.*, 2011). This enzyme catalyzes the hydrolysis of the pesticide paraoxon with very fast turnover rate, which was thought to be near the evolutionary limit. However, through a single, rationally designed mutation with NNAA, which increased the electrostatic repulsion of the negatively charged hydrolysis product, and thus improved the rate-limiting product release step, about a 10-fold improvement of this already highly efficient enzyme was achieved. Such result is in great contrast to the difficulty in improving this enzyme's activity through screening large mutation library with canonical amino acids.

In the absence of structural information required for rational design, library screening methods can be utilized to evolve NNAA-containing proteins for the desired properties or functions. Briefly, a target gene containing various mutations is expressed in the presence of the NNAA, and the resultant NNAA-containing protein variants are screened and selected for the desired properties. While the designs of library (ranging from total random mutation library, to the limited mutation library at a few codons), the methods to introduce NNAAs into proteins (as reviewed in previous sections), as well as the library screening strategies can vary from case to case, the basic principle for functional evolution of proteins with NNAAs can be applied broadly for protein engineering. For example, sense codon global substitution with a NNAA can often result in the loss of a protein's structure and function. By screening random libraries with a sense codon global substitution method to replace Leu with Tfl, a mutant Tfl-containing enzyme with enhanced thermal stability (Montclare & Tirrell, 2006) and a mutant Tfl-containing GFP with increased fluorescence intensity (Yoo *et al.*, 2007) were obtained. By screening a phage-display scFv antibody library, where six residues in the V_H CDR3 were randomized, one scFv containing sulfotyrosine was obtained

that binds gp120 more effectively than any similarly evolved scFv molecules containing only canonical amino acids (Liu *et al.*, 2008). By screening a peptide library based on cellular viability, cyclic peptide (containing NNAA) protease inhibitors were derived (Young *et al.*, 2011). Other NNAA-containing proteins, such as boron-containing proteins for carbohydrate binding (Liu *et al.*, 2009), have been evolved as well. Therefore, functional evolution of proteins with NNAAs can offer a selective advantage beyond the traditional directed evolution with only the canonical amino acids.

5. Conclusion

NNAAs greatly expand the amino acid repertoire for protein engineering. The ability to introduce NNAAs into proteins is a very important technical advance in the fields of both discovery research as well as commercial production of medicines and industrial enzymes. There are several established methods to accomplish the incorporation of NNAAs. In most cases, the NNAA serves as a point of protein modification. For biopharmaceutical applications, the ability to introduce the NNAA at a specific position becomes a very important advantage, as protein function is very sensitive to the position of conjugation. These new technologies are currently being used in the development of new and better protein therapeutics that may soon be on the market. The future will likely include a host of new medicines based on this technology, and its many diverse applications.

6. References

Agard,N.J., Baskin,J.M., Prescher,J.A., Lo,A., and Bertozzi,C.R. (2006). A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol. 1*, 644-648.

Agard,N.J., Prescher,J.A., and Bertozzi,C.R. (2004). A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc. 126*, 15046-15047.

Agris,P.F., Vendeix,F.A., and Graham,W.D. (2007). tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol. 366*, 1-13.

Alley,S.C., Okeley,N.M., and Senter,P.D. (2010). Antibody-drug conjugates: targeted drug delivery for cancer. *Curr. Opin. Chem. Biol. 14*, 529-537.

Ambrogelly,A., Gundllapalli,S., Herring,S., Polycarpo,C., Frauer,C., and Soll,D. (2007). Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc. Natl. Acad. Sci. U. S. A 104*, 3141-3146.

Anderson,J.C., Wu,N., Santoro,S.W., Lakshman,V., King,D.S., and Schultz,P.G. (2004). An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U. S. A 101*, 7566-7571.

Bacher,J.M. and Ellington,A.D. (2001). Selection and characterization of Escherichia coli variants capable of growth on an otherwise toxic tryptophan analogue. *J. Bacteriol. 183*, 5414-5425.

Back,J.W., David,O., Kramer,G., Masson,G., Kasper,P.T., de Koning,L.J., de Jong,L., van Maarseveen,J.H., and de Koster,C.G. (2005). Mild and chemoselective peptide-bond cleavage of peptides and proteins at azido homoalanine. *Angew. Chem. Int. Ed Engl. 44*, 7946-7950.

Bain,J.D., Switzer,C., Chamberlin,A.R., and Benner,S.A. (1992). Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. *Nature 356*, 537-539.

Beier,H. and Grimm,M. (2001). Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Res. 29*, 4767-4782.

Blackman,M.L., Royzen,M., and Fox,J.M. (2008). Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *J. Am. Chem. Soc. 130*, 13518-13519.

Brustad,E., Bushey,M.L., Lee,J.W., Groff,D., Liu,W., and Schultz,P.G. (2008). A genetically encoded boronate amino acid. *Angew. Chem. Int. Ed Engl. 47*, 8220-8223.

Bundy,B.C. and Swartz,J.R. (2010). Site-specific incorporation of

Chalker,J.M., Wood,C.S., and Davis,B.G. (2009). A convenient catalyst for aqueous and protein Suzuki-Miyaura cross-coupling. *J. Am. Chem. Soc. 131*, 16346-16347.

Chen,P.R., Groff,D., Guo,J., Ou,W., Cellitti,S., Geierstanger,B.H., and Schultz,P.G. (2009). A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chem. Int. Ed Engl. 48*, 4052-4055.

Chin,J.W., Cropp,T.A., Anderson,J.C., Mukherji,M., Zhang,Z., and Schultz,P.G. (2003). An expanded eukaryotic genetic code. *Science 301*, 964-967.

Cho,H., Daniel,T., Buechler,Y.J., Litzinger,D.C., Maio,Z., Putnam,A.H., Kraynov,V.S., Sim,B., Bussell,S., Javahishvili,T., Kaphle,S., Viramontes,G., Ong,M., Chu,S., GC,B., Lieu,R., Knudsen,N., Castiglioni,P., Norman,T.C., Axelrod,D.W., Hoffman,A.R., Schultz,P.G., DiMarchi,R.D., and Kimmel,B.E. (2011). Optimized clinical performance of growth hormone with an expanded genetic code. *Proc. Natl. Acad. Sci. U. S. A 108*, 9060-9065.

Codelli,J.A., Baskin,J.M., Agard,N.J., and Bertozzi,C.R. (2008). Second-generation difluorinated cyclooctynes for copper-free click chemistry. *J. Am. Chem. Soc. 130*, 11486-11493.

Corzana,F. *et al.* (2011). Rational design of a Tn antigen mimic. *Chem. Commun. (Camb.) 47*, 5319-5321.

Crick,F.H. (1966). Codon--anticodon pairing: the wobble hypothesis. *J. Mol. Biol. 19*, 548-555.

Curran,J.F. and Yarus,M. (1989). Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. *J. Mol. Biol. 209*, 65-77.

Debets,M.F., van Berkel,S.S., Schoffelen,S., Rutjes,F.P., van Hest,J.C., and van Delft,F.L. (2010). Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition. *Chem. Commun. (Camb.) 46*, 97-99.

del Amo,D.S., Wang,W., Jiang,H., Besanceney,C., Yan,A.C., Levy,M., Liu,Y., Marlow,F.L., and Wu,P. (2010). Biocompatible copper(I) catalysts for in vivo imaging of glycans. *J. Am. Chem. Soc. 132*, 16893-16899.

Devaraj,N.K., Weissleder,R., and Hilderbrand,S.A. (2008). Tetrazine-based cycloadditions: application to pretargeted live cell imaging. *Bioconjug. Chem. 19*, 2297-2299.

Dirksen,A. and Dawson,P.E. (2008). Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. *Bioconjug. Chem. 19*, 2543-2548.

Dommerholt,J., Schmidt,S., Temming,R., Hendriks,L.J., Rutjes,F.P., van Hest,J.C., Lefeber,D.J., Friedl,P., and van Delft,F.L. (2010). Readily accessible bicyclononynes for bioorthogonal labeling and three-dimensional imaging of living cells. *Angew. Chem. Int. Ed Engl. 49*, 9422-9425.

Dondoni,A. (2008). The emergence of thiol-ene coupling as a click process for materials and bioorganic chemistry. Angew. *Chem. Int. Ed Engl. 47*, 8995-8997.

Farkas,A.M. and Finn,O.J. (2010). Vaccines based on abnormal self-antigens as tumor-associated antigens: immune regulation. *Semin. Immunol. 22*, 125-131.

Filpula,D. (2007). Antibody engineering and modification technologies. Biomol. Eng *24*, 201-215.

Fishburn,C.S. (2008). The pharmacology of PEGylation: balancing PD with PK to generate novel therapeutics. *J. Pharm. Sci. 97*, 4167-4183.

Genet,J.P. and Savignac,M. (1999). Recent developments of palladium(0) catalyzed reaactions in aqueous medium. *Journal of Organometallic Chemstry 576*, 305-317.

Gibson,D.G. *et al.* (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science 329*, 52-56.

Goltermann,L., Larsen,M.S., Banerjee,R., Joerger,A.C., Ibba,M., and Bentin,T. (2010). Protein evolution via amino acid and codon elimination. *PLoS. One. 5*, e10104.

Grabstein, K., Wang, A., Nairn,N.W., Graddis, T.J. (2010). Methods of modifying polypeptides comprisig non-natural amino acids, *US patent 7,829,659*

Griffiths,G.L., Chang,C., McBride,W.J., Rossi,E.A., Sheerin,A., Tejada,G.R., Karacay,H., Sharkey,R.M., Horak,I.D., Hansen,H.J., and Goldenberg,D.M. (2004). Reagents and methods for PET using bispecific antibody pretargeting and 68Ga-radiolabeled bivalent hapten-peptide-chelate conjugates. *J. Nucl. Med. 45*, 30-39.

Gronke, R.S., Jaquez, O. and Kiistala, M. (2010). Challenges of developing a PEGylaed Interferon beta-1a. 239 ACS National Meeting, San Francisco, CA

Grunewald,J. *et al.* (2009). Mechanistic studies of the immunochemical termination of self-tolerance with unnatural amino acids. *Proc. Natl. Acad. Sci. U. S. A 106*, 4337-4342.

Gu,W., Li,M., Zhao,W.M., Fang,N.X., Bu,S., Frazer,I.H., and Zhao,K.N. (2004). tRNASer(CGA) differentially regulates expression of wild-type and codon-modified papillomavirus L1 genes. *Nucleic Acids Res. 32*, 4448-4461.

Gupta,S.S., Kuzelka,J., Singh,P., Lewis,W.G., Manchester,M., and Finn,M.G. (2005). Accelerated bioorthogonal conjugation: a practical method for the ligation of diverse functional molecules to a polyvalent virus scaffold. *Bioconjug. Chem. 16*, 1572-1579.

Gutsmiedl,K., Wirges,C.T., Ehmke,V., and Carell,T. (2009). Copper-free "click" modification of DNA via nitrile oxide-norbornene 1,3-dipolar cycloaddition. *Org. Lett. 11*, 2405-2408.

Haeusler,R.A. and Engelke, D.R. (2006) Spatial organization of transcription by RNA polymerase III. Nucleic Acids Res. 34, 4826-4836

Hamblett,K.J., Senter,P.D., Chace,D.F., Sun,M.M., Lenox,J., Cerveny,C.G., Kissler,K.M., Bernhardt,S.X., Kopcha,A.K., Zabinski,R.F., Meyer,D.M., and Francisco,J.A. (2004). Effects of drug loading on the antitumor activity of a monoclonal antibody drug conjugate. *Clin. Cancer Res. 10*, 7063-7070.

Hancock,S.M., Uprety,R., Deiters,A., and Chin,J.W. (2010). 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.

Harris,J.M. and Chess,R.B. (2003). Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug Discov. 2*, 214-221.

Hendrickson,T.L., de Crecy-Lagard,V., and Schimmel,P. (2004). Incorporation of nonnatural amino acids into proteins. *Annu. Rev. Biochem. 73*, 147-176.

Higgs,P.G. and Ran,W. (2008). Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. *Mol. Biol. Evol. 25*, 2279-2291.

Hirao,I. *et al.* (2002). An unnatural base pair for incorporating amino acid analogs into proteins. *Nat. Biotechnol. 20*, 177-182.

Hirao,I., Harada,Y., Kimoto,M., Mitsui,T., Fujiwara,T., and Yokoyama,S. (2004). A two-unnatural-base-pair system toward the expansion of the genetic code. *J. Am. Chem. Soc. 126*, 13298-13305.

Hohsaka,T. and Sisido,M. (2002). Incorporation of non-natural amino acids into proteins. *Curr. Opin. Chem. Biol. 6*, 809-815.

Hohsaka,T., Sato,K., Sisido,M., Takai,K., and Yokoyama,S. (1994). Site-specific incorporation of photofunctional nonnatural amino acids into a polypeptide through in vitro protein biosynthesis. *FEBS Lett. 344*, 171-174.

Hong,V., Presolski,S.I., Ma,C., and Finn,M.G. (2009). Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew. Chem. Int. Ed Engl. 48*, 9879-9883.

Hong,V., Steinmetz,N.F., Manchester,M., and Finn,M.G. (2010). Labeling live cells by copper-catalyzed alkyne--azide click chemistry *Bioconjug. Chem. 21*, 1912-1916.

Hughes,R.A. and Ellington,A.D. (2010). Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA. *Nucleic Acids Res. 38*, 6813-6830.

Hutchins,B.M., Kazane,S.A., Staflin,K., Forsyth,J.S., Felding-Habermann,B., Schultz,P.G., and Smider,V.V. (2011). Site-specific coupling and sterically controlled formation of multimeric antibody fab fragments with unnatural amino acids. *J. Mol. Biol. 406*, 595-603.

Ibba,M., Becker,H.D., Stathopoulos,C., Tumbula,D.L., and Soll,D. (2000). The adaptor hypothesis revisited. *Trends Biochem. Sci. 25*, 311-316.

Jawalekar,A.M., Reubsaet,E., Rutjes,F.P., and van Delft,F.L. (2011). Synthesis of isoxazoles by hypervalent iodine-induced cycloaddition of nitrile oxides to alkynes. *Chem. Commun. (Camb.) 47*, 3198-3200.

Jewett,J.C. and Bertozzi,C.R. (2010). Cu-free click cycloaddition reactions in chemical biology. *Chem. Soc. Rev. 39*, 1272-1279.

Jewett,J.C., Sletten,E.M., and Bertozzi,C.R. (2010). Rapid Cu-free click chemistry with readily synthesized biarylazacyclooctynones. *J. Am. Chem. Soc. 132*, 3688-3690.

Junutula,J.R. *et al.* (2008). Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat. Biotechnol. 26*, 925-932.

Kalia,J. and Raines, R.T. (2008). Hydrolytic stability of hydrazones and oximes. *Angew.Chem.Int.Ed. 47*, 7523-7526.

Kim,K.M., McDonagh,C.F., Westendorf.L., Brown,L.L., Sussman,D., Feist,T., Lyon,R., Alley,S.C., Okeley,N.M., Zhang,X., Thompson,M.C., Stone,I., Gerber,H., and Carter,P.J. (2008). Anti-CD30 diabody-drug conjugates with potent antitumor activity. *Mol. Cancer Ther. 7*, 2486-2497.

Kobayashi,T., Yanagisawa,T., Sakamoto,K., and Yokoyama,S. (2009). Recognition of non-alpha-amino substrates by pyrrolysyl-tRNA synthetase. *J. Mol. Biol. 385*, 1352-1360.

Kodama,K., Fukuzawa,S., Nakayama,H., Sakamoto,K., Kigawa,T., Yabuki,T., Matsuda,N., Shirouzu,M., Takio,K., Yokoyama,S., and Tachibana,K. (2007). Site-specific

functionalization of proteins by organopalladium reactions. *Chembiochem. 8*, 232-238.

Köhrer,C., Sullivan,E.L., and RajBhandary, U.L. (2004) Complete set of orthogonal 21st aminoacyl-tRNA synthetase-amber, ochre and opal suppressor tRNA pairs: concomitant suppression of three different termination codons in an mRNA in mammalian cells. *Nucleic Acids Res*. 32, 6200-6211

Kwon,I. and Tirrell,D.A. (2007). Site-specific incorporation of tryptophan analogues into recombinant proteins in bacterial cells. *J. Am. Chem. Soc. 129*, 10431-10437.

Kwon,I., Kirshenbaum,K., and Tirrell,D.A. (2003). Breaking the degeneracy of the genetic code. *J. Am. Chem. Soc. 125*, 7512-7513.

Kwon,I., Wang,P., and Tirrell,D.A. (2006). Design of a bacterial host for site-specific incorporation of p-bromophenylalanine into recombinant proteins. *J. Am. Chem. Soc. 128*, 11778-11783.

Leader,B., Baca,Q.J., and Golan,D.E. (2008). Protein therapeutics: a summary and pharmacological classification. *Nat. Rev. Drug Discov. 7*, 21-39.

Li,W.T., Mahapatra,A., Longstaff,D.G., Bechtel,J., Zhao,G., Kang,P.T., Chan,M.K., and Krzycki,J.A. (2009). Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs. *J. Mol. Biol. 385*, 1156-1164.

Li,Z., Seo,T.S., and Ju,J. (2004). 1,3-Dipolar cycloaddition of azides with electron-deficient alkynes under mild condition in water. *Tetrahedron Letters 45*, 3143-3146.

Lin,Y.A., Chalker,J.M., and Davis,B.G. (2009). Olefin metathesis for site-selective protein modification. *Chembiochem. 10*, 959-969.

Lin,Y.A., Chalker,J.M., and Davis,B.G. (2010). Olefin cross-metathesis on proteins: investigation of allylic chalcogen effects and guiding principles in metathesis partner selection. *J. Am. Chem. Soc. 132*, 16805-16811.

Link,A.J. and Tirrell,D.A. (2005). Reassignment of sense codons in vivo. *Methods 36*, 291-298.

Link,A.J., Vink,M.K., and Tirrell,D.A. (2004). Presentation and detection of azide functionality in bacterial cell surface proteins. *J. Am. Chem. Soc. 126*, 10598-10602.

Liu,C.C. and Schultz,P.G. (2010). Adding new chemistries to the genetic code. *Annu. Rev. Biochem. 79*, 413-444.

Liu,C.C., Mack,A.V., Brustad,E.M., Mills,J.H., Groff,D., Smider,V.V., and Schultz,P.G. (2009). Evolution of proteins with genetically encoded "chemical warheads". *J. Am. Chem. Soc. 131*, 9616-9617.

Liu,C.C., Mack,A.V., Tsao,M.L., Mills,J.H., Lee,H.S., Choe,H., Farzan,M., Schultz,P.G., and Smider,V.V. (2008). Protein evolution with an expanded genetic code. *Proc. Natl. Acad. Sci. U. S. A 105*, 17688-17693.

Liu,W., Brock,A., Chen,S., Chen,S., and Schultz,P.G. (2007). Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods 4*, 239-244.

Martin,M.E., Parameswarappa,S.G., O'Dorisio,M.S., Pigge,F.C., and Schultz,M.K. (2010). A DOTA-peptide conjugate by copper-free click chemistry. *Bioorg. Med. Chem. Lett. 20*, 4805-4807.

Marvin,J.S. and Zhu,Z. (2005). Recombinant approaches to IgG-like bispecific antibodies. *Acta Pharmacol. Sin. 26*, 649-658.

McDonagh,C.F. Turcott,E., Westendorf,L., Webster,J.B., Alley,S.C., Kim,K., Andreyka,J., Stone,I., Hamblett,K.J., Francisco,J.A., and Carter,P. (2006). Engineered antibody-

drug conjugates with defined sites and stoichiometries of drug attachment. *Protein Eng Des Sel 19*, 299-307.

Mehl,R.A., Anderson,J.C., Santoro,S.W., Wang,L., Martin,A.B., King,D.S., Horn,D.M., and Schultz,P.G. (2003). Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc. 125*, 935-939.

Minozzi,M., Monesi,A., Nanni,D., Spagnolo,P., Marchetti,N., and Massi,A. (2011). An insight into the radical thiol/yne coupling: the emergence of arylalkyne-tagged sugars for the direct photoinduced glycosylation of cysteine-containing peptides. *J. Org. Chem. 76*, 450-459.

Montclare,J.K. and Tirrell,D.A. (2006). Evolving proteins of novel composition. *Angew. Chem. Int. Ed Engl. 45*, 4518-4521.

Mukai,T., Hayashi,A., Iraha,F., Sato,A., Ohtake,K., Yokoyama,S., and Sakamoto,K. (2010). Codon reassignment in the Escherichia coli genetic code. *Nucleic Acids Res. 38*, 8188-8195.

Mukai,T., Kobayashi,T., Hino,N., Yanagisawa,T., Sakamoto,K., and Yokoyama,S. (2008). Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun. 371*, 818-822.

Nairn, N.W., Graddis, T.J., Wang, A., Shanebeck, K., and Grabstein, K. (2007). Site-specific PEGylation of interferon-beta by Cu(I)-catalyzed cycloaddition. *ACS National Meeting 2007, Boston, MA*

Nakamura,Y., Ito,K., and Ehrenberg,M. (2000). Mimicry grasps reality in translation termination. *Cell 101*, 349-352.

Natarajan,A., Xiong,C.Y., Albrecht,H., DeNardo,G.L., and DeNardo,S.J., (2005). Characterization of site-specific ScFv PEGylation for tumor-targeting pharmaceuticals Bioconjug. Chem. *16*, 113-121.

Neumann,H., Peak-Chew,S.Y., and Chin,J.W. (2008). Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat. Chem. Biol. 4*, 232-234.

Neumann,H., Slusarczyk,A.L., and Chin,J.W. (2010a). De novo generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs. *J. Am. Chem. Soc. 132*, 2142-2144.

Neumann,H., Wang,K., Davis,L., Garcia-Alai,M., and Chin,J.W. (2010b). Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature 464*, 441-444.

Nguyen,D.P., Lusic,H., Neumann,H., Kapadnis,P.B., Deiters,A., and Chin,J.W. (2009). 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.

Ojida,A., Tsutsumi,H., Kasagi,N., and Hamachi,I. (2005). Suzuki coupling for protein modification. *Tetrahedron Letters 46*, 3301-3305.

Ou,W. *et al.* (2011). Site-specific protein modifications through pyrroline-carboxy-lysine residues. *Proc. Natl. Acad. Sci. U. S. A.* 108, 3301-3305.

Pastrnak,M. and Schultz,P.G. (2001). Phage selection for site-specific incorporation of unnatural amino acids into proteins in vivo. *Bioorg. Med. Chem. 9*, 2373-2379.

Patel,K.G., Ng,P.P., Kuo,C.C., Levy,S., Levy,R., and Swartz,J.R. (2009). Cell-free production of Gaussia princeps luciferase--antibody fragment bioconjugates for ex vivo detection of tumor cells. *Biochem. Biophys. Res. Commun. 390*, 971-976.

Peschke,B., Zundel,M., Bak,S., Clausen,T.R., Blume,N., Pedersen,A., Zaragoza,F., and Madsen,K. (2007). C-Terminally PEGylated hGH-derivatives. *Bioorg. Med. Chem.* 15, 4382-4395.

Plass,T., Milles,S., Koehler,C., Schultz,C., and Lemke,E.A. (2011). Genetically encoded copper-free click chemistry. *Angew. Chem. Int. Ed Engl.* 50, 3878-3881.

Polycarpo,C.R., Herring,S., Berube,A., Wood,J.L., Soll,D., and Ambrogelly,A. (2006). Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* 580, 6695-6700.

p-propargyloxyphenylalanine in a cell-free environment for direct protein-protein click conjugation. *Bioconjug. Chem.* 21, 255-263.

Presolski,S.I., Hong,V., Cho,S.H., and Finn,M.G. (2010). Tailored ligand acceleration of the Cu-catalyzed azide-alkyne cycloaddition reaction: practical and mechanistic implications. *J. Am. Chem. Soc.* 132, 14570-14576.

Rackham,O. and Chin,J.W. (2005). A network of orthogonal ribosome x mRNA pairs. *Nat. Chem. Biol.* 1, 159-166.

Rostovtsev,V.V., Green,L.G., Fokin,V.V., and Sharpless,K.B. (2002). A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed Engl.* 41, 2596-2599.

Sakamoto,K. *et al.* (2002). Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.* 30, 4692-4699.

Santoro,S.W., Wang,L., Herberich,B., King,D.S., and Schultz,P.G. (2002). An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* 20, 1044-1048.

Saxon,E. and Bertozzi,C.R. (2000). Cell surface engineering by a modified Staudinger reaction. *Science 287*, 2007-2010.

Sharp,P.M., Bailes,E., Grocock,R.J., Peden,J.F., and Sockett,R.E. (2005). Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res.* 33, 1141-1153.

Sletten,E.M. and Bertozzi,C.R. (2008). A hydrophilic azacyclooctyne for Cu-free click chemistry. *Org. Lett. 10*, 3097-3099.

Sletten,E.M. and Bertozzi,C.R. (2009). Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed Engl.* 48, 6974-6998.

Song,W., Wang,Y., Qu,J., and Lin,Q. (2008). Selective functionalization of a genetically encoded alkene-containing protein via "photoclick chemistry" in bacterial cells. *J. Am. Chem. Soc. 130*, 9654-9655.

Sun,M.M., Beam,K.S., Cerveny,C.G., Hamblett,K.J., Blackmore,R.S., Torgov,M.Y., Handley,F.G., Ihle,N.C., Senter,P.D., and Alley,S.C. (2005). Reduction-alkylation strategies for the modification of specific monoclonal antibody disulfides. *Bioconjug. Chem. 16*, 1282-1290.

Takimoto,J.K., Adams,K.L., Xiang,Z., and Wang,L. (2009). Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells. *Mol. Biosyst. 5*, 931-934.

Takimoto,J.K., Dellas,N., Noel,J.P., and Wang,L. (2011). Stereochemical Basis for Engineered Pyrrolysyl-tRNA Synthetase and the Efficient in Vivo Incorporation of Structurally Divergent Non-native Amino Acids. *ACS Chem. Biol.* Ahead of publication

Tam,A. and Raines,R.T. (2009). Protein engineering with the traceless Staudinger ligation. *Methods Enzymol.* 462, 25-44.

Tam,A., Arnold,U., Soellner,M.B., and Raines,R.T. (2007). Protein prosthesis: 1,5-disubstituted[1,2,3]triazoles as cis-peptide bond surrogates. *J. Am. Chem. Soc. 129*, 12670-12671.

Tang,Y., Ghirlanda,G., Vaidehi,N., Kua,J., Mainz,D.T., Goddard III,W.A., DeGrado,W.F., and Tirrell,D.A. (2001). Stabilization of coiled-coil peptide domains by introduction of trifluoroleucine. *Biochemistry 40*, 2790-2796.

Tanrikulu,I.C., Schmitt,E., Mechulam,Y., Goddard,W.A., III, and Tirrell,D.A. (2009). Discovery of Escherichia coli methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc. Natl. Acad. Sci. U. S. A 106*, 15285-15290.

Tornoe,C.W., Christensen,C., and Meldal,M. (2002). Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem. 67*, 3057-3064.

Ugwumba,I.N. *et al.* (2011). Improving a Natural Enzyme Activity through Incorporation of Unnatural Amino Acids. *J. Am. Chem. Soc.* 133, 326-333

van Berkel,S.S., Dirks,A.T., Meeuwissen,S.A., Pingen,D.L., Boerman,O.C., Laverman,P., van Delft,F.L., Cornelissen,J.J., and Rutjes,F.P. (2008). Application of metal-free triazole formation in the synthesis of cyclic RGD-DTPA conjugates. *Chembiochem. 9*, 1805-1815.

Vichier-Guerre,S., Lo-Man,R., Huteau,V., Deriaud,E., Leclerc,C., and Bay,S. (2004). Synthesis and immunological evaluation of an antitumor neoglycopeptide vaccine bearing a novel homoserine Tn antigen. *Bioorg. Med. Chem. Lett. 14*, 3567-3570.

Voloshchuk,N. and Montclare,J.K. (2010). Incorporation of unnatural amino acids for synthetic biology. *Mol. Biosyst. 6*, 65-80.

Wang,A., Winblade,N.N., Johnson,R.S., Tirrell,D.A., and Grabstein,K. (2008). Processing of N-terminal unnatural amino acids in recombinant human interferon-beta in Escherichia coli. *Chembiochem. 9*, 324-330.

Wang,K., Neumann,H., Peak-Chew,S.Y., and Chin,J.W. (2007a). Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol. 25*, 770-777.

Wang,Q. and Wang,L. (2008). New methods enabling efficient incorporation of unnatural amino acids in yeast. *J. Am. Chem. Soc. 130*, 6066-6067.

Wang,Q., Parrish,A.R., and Wang,L. (2009). Expanding the genetic code for biological studies. *Chem. Biol. 16*, 323-336.

Wang,W., Takimoto,J.K., Louie,G.V., Baiga,T.J., Noel,J.P., Lee,K.F., Slesinger,P.A., and Wang,L. (2007b). Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat. Neurosci. 10*, 1063-1072.

Wang,Y.S., Youngster,S., Grace,M., Bausch,J., Bordens,R., and Wyss,D.F. (2002). Structural and biological characterization of pegylated recombinant interferon alpha-2b and its therapeutic implications. *Adv. Drug Deliv. Rev. 54*, 547-570.

Yanagisawa,T., Ishii,R., Fukunaga,R., Kobayashi,T., Sakamoto,K., and Yokoyama,S. (2008). Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol. 15*, 1187-1197.

Yoo,T.H., Link,A.J., and Tirrell,D.A. (2007). Evolution of a fluorinated green fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A 104*, 13887-13890.

Young,T.S., Ahmad,I., Yin,J.A., and Schultz,P.G. (2010). An enhanced system for unnatural amino acid mutagenesis in E. coli. *J. Mol. Biol. 395*, 361-374.

Young,T.S., Young,D.D., Ahmad,I., Louis,J.M., Benkovic,S.J., and Schultz,P.G. (2011). Evolution of cyclic peptide protease inhibitors. *Proc. Natl. Acad. Sci. U. S. A.*

Zhang,K., Li,H., Cho,K.M., and Liao,J.C. (2010). Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proc. Natl. Acad. Sci. U. S. A 107*, 6234-6239.

Zhang,L., Chen,X., Xue,P., Sun,H.H., Williams,I.D., Sharpless,K.B., Fokin,V.V., and Jia,G. (2005). Ruthenium-catalyzed cycloaddition of alkynes and organic azides. *J. Am. Chem. Soc. 127*, 15998-15999.

Zhang,Z., Alfonta,L., Tian,F., Bursulaya,B., Uryu,S., King,D.S., and Schultz,P.G. (2004). Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A 101*, 8882-8887.

Zhao,K.N., Gu,W., Fang,N.X., Saunders,N.A., and Frazer,I.H. (2005). Gene codon composition determines differentiation-dependent expression of a viral capsid gene in keratinocytes in vitro and in vivo. *Mol. Cell Biol. 25*, 8643-8655.
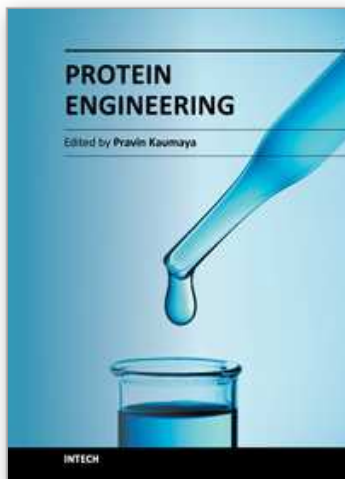
Protein Engineering

Edited by Prof. Pravin Kaumaya

A broad range of topics are covered by providing a solid foundation in protein engineering and supplies readers with knowledge essential to the design and production of proteins. This volume presents in-depth discussions of various methods for protein engineering featuring contributions from leading experts from different counties. A broad series of articles covering significant aspects of methods and applications in the design of novel proteins with different functions are presented. These include the use of non-natural amino acids, bioinformatics, molecular evolution, protein folding and structure-functional insight to develop useful proteins with enhanced properties.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Aijun Wang, Natalie Winblade Nairn, Marcello Marelli and Kenneth Grabstein (2012). Protein Engineering with Non-Natural Amino Acids, Protein Engineering, Prof. Pravin Kaumaya (Ed.), ISBN: 978-953-51-0037-9, InTech, Available from: http://www.intechopen.com/books/protein-engineering/protein-engineering-with-non-natural-amino-acids

INTECH
open science | open minds