Universiteit Leiden
The Netherlands

The advantages and disadvantages of bioorthogonal proteins

Groenewold, G.J.M.

Citation

Groenewold, G. J. M. (2021, February 17). *The advantages and disadvantages of bioorthogonal proteins*. Retrieved from https://hdl.handle.net/1887/3142384

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3142384

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

Leiden University Repository

The handle https://hdl.handle.net/1887/3142384 holds various files of this Leiden University dissertation.

Author: Groenewold, G.J.M.
Title: The advantages and disadvantages of bioorthogonal proteins
Issue Date: 2021-02-17

Chapter 1

Introduction

Abstract

It is possible nowadays to recombinantly express most proteins. However, these proteins will never be fully identical to those found in nature, due to differences in post-translational modifications. In order to mimic these modifications, various synthetic and semi-synthetic approaches have been developed, such as the *de novo* synthesis of proteins, their enzymatic modification or residue-specific chemical modification. An overview of these approaches will be given in this Chapter.

Introduction

Proteins play an important role in nearly all biological processes, such as catalysis [1], transport [2, 3], cell-cell interactions [4, 5] and our immune system [6, 7]. In early days, the study of a protein of interest required its isolation from its natural source. With the exception of highly abundant proteins, isolation of native proteins requires large – sometimes kilogram – quantities of tissue (plants, animals) and even then, often only small quantities of the desired protein are obtained [8]. Today it is possible – for most proteins – to either prepare a synthetic version [9] or to express and purify the protein of interest recombinantly [10]. Both approaches allow mutation of a protein of interest, which in turn allows the study of disease-associated mutations, but also the introduction of non-canonical amino acids including ones bearing chemical functionalities not encountered in nature [11]. This Chapter discusses current technologies for recombinant protein expression. Following a basic introduction into protein expression in general (section **1.1-1.3**), the main focus will be on the expression of proteins containing a non-canonical, non-natural functionality.

1.1 Basic principles of protein expression

Proteins are expressed from its corresponding coding sequence, a process which involves two major steps: transcription and translation [12]. During transcription the information encoded in deoxyribonucleic acid (DNA) is transcribed into pre-messenger ribonucleic acid (pre-mRNA) which is next processed by RNA polymerase II to form mature mRNA [13]. This mRNA serves as a template for protein synthesis with stretches of three specific bases encoding the twenty natural (also called canonical) amino acids as well as three signals to halt protein synthesis (stop codons) [14]. This process, called translation, occurs in the ribosomal compartment [15]. The ribosome consists of a small and a large subunit, of which the small subunit exhibits three different binding sites oriented from 5' to 3': the exit site (E), the peptidyl site (P) and the aminoacyl site (A) (Figure 1). When the mRNA is located inside the ribosome, the A-site binds the aminoacyl-transfer RNA (aa-tRNA, a complex which will be discussed in more detail later in this Chapter) bearing the complementary codon on the mRNA. The P-site holds the complex with the growing polypeptide chain and lastly, the E-site holds the tRNA without its amino acid, before it is released [16]. This process is terminated when a stop

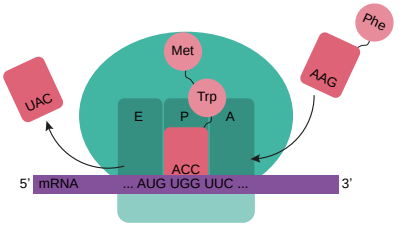


Figure 1. Stages of translation. The aa-tRNA binds to the A-site, hereafter the polypeptide chain will grow in the P-site. The empty aa-tRNA will leave via the E-site.

codon enters the A-site. Since there are no native aa-tRNA that recognize stop codons a termination factor will bind, releasing the polypeptide chain [17, 18]. Hereafter, the protein will either spontaneously or with the help of chaperones fold in its native state [19].

1.2 Protein expression hosts

Protein translation can be exploited to recombinantly produce proteins in a variety of expression systems, either prokaryotic or eukaryotic. The advantage of the former is the ease of use and scalability. The advantage of the latter is that proteins that have undergone mammalian post-translational modifications, in particular glycosylation, can be obtained. The first step in the production of recombinant proteins entails cloning of the underlying recombinant DNA, that codes for the protein of interest, into an expression host that supports gene expression [20]. Protein expression is most commonly conducted in bacteria, yeasts, insect cells or mammalian cells [21], but proteins can also be expressed in algae [22] or in cell free cultures [23]. Each of these expression systems have their advantages and disadvantages. The four most used expression systems, with their advantages and disadvantages, are discussed in some more detail below (Table 1).

Bacterial expression systems are commonly used because high expression levels can often be reached at low cost, thus allowing the production of large quantities of protein [24]. However, post-translational modifications (PTMs) such as may occur on mammalian proteins are not part of bacterial protein synthesis machineries (although some strains have been engineered to perform *N*-glycosylation [25]). Other features that complicate bacterial expression are the presence of endotoxins, such as lipopolysaccharides (LPS), as impurities can affect the immunological properties of a

Table 1: The advantages and disadvantages of different protein expression systems.

Expression systems	Advantages	Challenges	Most common applications
Bacterial	• Low cost • Rapid expression • Scalable • Most widely used expression system	• Inclusion bodies • No eukaryotic PTMs • Difficult to express higher MW proteins • Protein specific optimization	• Functional assays • Structural analysis • Bacterial proteins • Antigen proteins • Cytokines • Enzymes
Yeast	• Low cost • Rapid expression • High yield • Scalable • Diverse PTMs	• Improper glycosylation • High mannose modification • Growth optimization • Fermentation required for high yields	• Functional assays • Structural analysis • Antibody production • Lower MW proteins • Cytokines • Enzymes
Baculovirus - insect	• High capacity genes • Soluble proteins • Suitable for toxic proteins • PTMs similar to mammalian expression • Static or suspension cultures	• More demanding culture conditions • Lack of partial glycosylation • Time consuming	• Functional assays • Structural analysis • Cytoplasmic proteins • Toxic proteins • Transmembrane proteins (TP) • Protein complexes
Mammalian	• Soluble proteins • Lower endotoxin • Better bioactivity • PTMs • Transient and stable expression	• More demanding culture conditions • High-yield only in suspension conditions	• Functional assays • Structural analysis • Recombinant antibodies • Antibody production • Virus production • Expression of difficult proteins

protein sample. Their removal prior to use requires additional purification steps, whereas the production in an endotoxin free environment is also an option [26]. The reducing environment of the cytoplasm in for instance *Escherichia coli* (*E. coli*) [27] can also limit the yields of expression of proteins that require disulfide bridges for stable folding, and can result in the expression in inclusion bodies, vesicles containing partially or incorrectly folded proteins [28 - 30]. One of the possibilities to avoid the protein of interest ending up in inclusion bodies is by expressing the protein in the periplasm [31], where the oxidizing environment and the presence of proteins like disulfide oxidoreductase (DsbA) allow disulfide bonds to be formed more easily [27, 32].

Yeast expression systems support both intra- and extracellular expression of prokaryotic and eukaryotic proteins [33, 34]. The systems, in contrast to most contemporary bacterial ones, offer the possibility for expression levels in a minimal expression time in simple media [35]. Yeast expression allows for the introduction of post-translational modifications, including both *N*- and *O*-linked oligosaccharides, however often hyperglycosylated proteins are obtained [36].

Protein expression in insect cells makes use of baculovirus infection to introduce the gene of interest [37]. The system can be used for large scale production and allows for PTMs [38]. Post-translational modifications are chemical modifications of a protein,

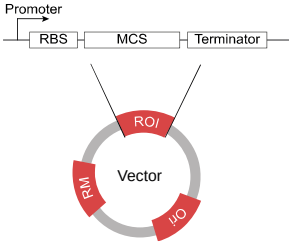


Figure 2. Minimal components of an expression vector. An expression vector, which at its region of interest (ROI), contains at least a promoter, a ribosomal binding site, a multiple cloning site and a corresponding terminator. Moreover, a vector contains an origin of replication (Ori) and a resistance marker (RM).

which play an important role in the human proteome, which will be described in more detail in section 1.4 [39]. The proteins will be, depending on the chosen system, either secreted intra- or extracellularly [40, 41]. On the downside, protein expression using this system is time-consuming at relatively high costs [37].

Despite the extensive developments in non-mammalian eukaryotic expression, most mammalian proteins, especially those that are endowed with post-translational modifications and/or those that rely on mammalian folding systems, are still best expressed in a mammalian expression system [42]. There are various ways to express a protein in mammalian cells: via plasmid transfection [43], via retrovirus or lentivirus transduction [44] or via modified mRNA transfection using both transient expression and stable cell cultures [45]. However, the use of mammalian cells comes with more challenging culture conditions, such as slow cell growth and high production costs by the use of expensive media and culture conditions [46]. High expression levels can only be obtained in suspension cultures, cultures in which cells are homogeneously suspended and grown in agitating liquid medium [47].

1.3 Recombinant protein expression

Once the expression host is selected, cloning the recombinant DNA encoding the protein of interest requires the selection and design of an appropriate expression vector, which should be compatible with the selected expression host.

For bacterial expression, two expression vector systems are commonly used, the pET vector system of Novagen/EMD Millipore [48, 49] and the pQE vector system of Qiagen [50]. The pET systems contain at least a T7 RNA polymerase promoter, a

ribosomal binding site (RBS) – the Shine-Dalgarno sequence –, a multiple cloning site (MCS) and a T7 terminator (Figure 2) [51]. When the gene is cloned into the MCS, the resulting plasmid is transformed into a lambda DE3 lysogen strain, containing the T7 RNA polymerase under a *lac* repressor, and subsequently protein expression is induced using isopropyl-β-D-1-thiogalactopyranoside (IPTG) [52]. The pQE vectors contain the same minimal components, with the exception that the gene will be cloned behind a T5 promoter [53]. This promoter will be recognized by the *E. coli* RNA polymerase and therefore does not require the T7 RNA polymerase [54]. Besides these two often-used vectors, other vectors are used as well which are inducible by other molecules [55-57].

The basic components, a promoter, a ribosomal binding site – called the Kozak sequence in mammalian cells – and a multiple cloning site, are the same for each of the other described expression systems in section **1.2**. Every expression host requires a different promoter. For instance, in mammalian protein expression a cytomegalovirus (CMV) promoter can be used [42]; in insect cells a polyhedron promoter [58]; and in yeast a galactose inducible GAL1 promoter [59].

1.4 Post-translational modifications of proteins

During and after protein expression, the nascent protein can be subjected to multiple modifications. These, so called post-translational modifications (PTMs) extend the repertoire of chemical functionalities beyond those available from the 20 canonical amino acids [60]. The changes in protein properties imbued by PTMs – which can be both permanent and reversible – modulate and regulate cellular processes [61]. The change in properties the introduction of *N*- and *O*-linked glycans has, plays important roles in protein folding and stability [62], mucosal barrier function [63], cell-cell communication [64], and immune activation [65], to name but a few examples. Phosphorylation of serine/threonine/tyrosine (and other) residues, GlcNAcylation, and ubiquitination are reversible PTMs that play key roles in transducing signals within the cell [66]. They affect processes ranging from cell division and growth to apoptosis [67] by regulating intracellular signalling, transcription, protein routing, and degradation (Figure 3) [68, 69].

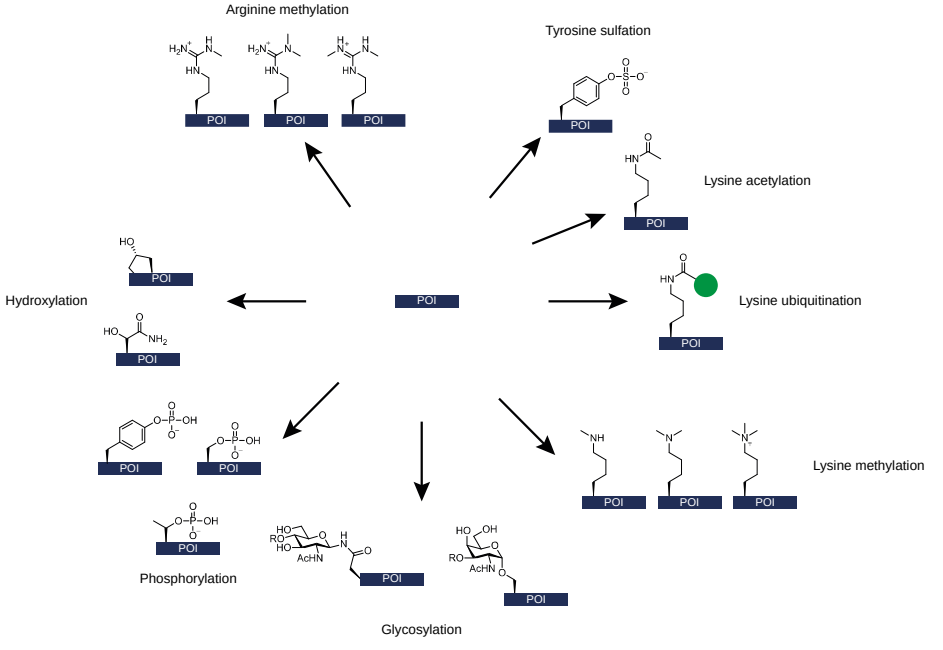


Figure 3. A few examples of post-translational modifications of proteins.

PTMs play a major role in a variety of human diseases [70], such as cancers [71], auto-immune diseases [72], heart diseases [73] and neurodegenerative diseases [74]. However, and unlike the amino acid backbone that is templated by the DNA of the original genetic code, PTMs are not encoded and therefore often heterogeneous, which makes the study of their precise function more difficult. This holds particularly true when multiple PTMs compete for the same site in the protein, such as for example serine and threonine residues that can be either phosphorylated or GlcNAcylated [75], – resulting in a heterogeneous protein mixture – or when, as is often the case, PTMs are reversible [76].

In recent years, extensive effort has gone into methods that allow the production of homogeneous PTM-modified proteins [77]. This has yielded three strategies for their production: chemical total synthesis of the PTM-modified protein [78]; site-selective enzymatic modification [79]; or the tag-and-modify approach, in which a small unique chemical functionality is introduced at a specific site in the protein and subsequently reacted with a PTM-reagent (or mimic) [80]. The latter approach has the advantages of its highly versatile and modular nature that gives full choice over the site of modification

[81]. One obvious caveat that the linkage of the PTM to backbone is non-native in most cases.

1.4.1 Synthesis of homogeneous N-linked glycosylated proteins through enzymatic reactions

The first two strategies to incorporate PTMs in proteins involve synthesizing the PTM-modified protein *de novo,* or via enzymatic reactions [82]. Both these methods have been largely used to produce proteins that carry glycans at specific sites. The synthesis of glycoproteins [83] can be performed combining solid phase peptide synthesis and native chemical ligation reactions [78], and has yielded some impressive examples such as synthetic single glycoforms of erythropoietin (EPO) [84] and the primary gp120 V1V2 glycopeptide of human immunodeficiency virus (HIV) [85].

The second approach to synthesize single glycoforms of proteins has been to engineer the existing mixtures of glycans on expressed proteins. In this approach, *N*-linked glycans (there is no universal protocol for site-specific *O*-linked glycan modification yet [86]) are hydrolyzed down to a single GlcNAc-asparagine. This minimal motif is then elaborated using engineered endoglycosidases that favor the reverse reaction, so called glycosynthases [87] (Scheme 1A). This method has been successfully used to synthesize single glycoforms of RNAse [88], but also of antibodies [89].

1.4.2. The “tag-and-modify” approach on native residues

The “tag-and-modify” method is the third method to produce glycoproteins and other PTMs, and compared to the above two methods is less cumbersome and more versatile. It is dependent on first introducing a reactive amino acid into the protein backbone at the prospective site of PTM, and the subsequent selective chemical modification of this tag [80]. This tag can be either a reactive natural amino acid (usually a cysteine [90]), or a noncanonical amino acid [81, 91].

1.4.3 Site-specific chemical ligations to canonical amino acids

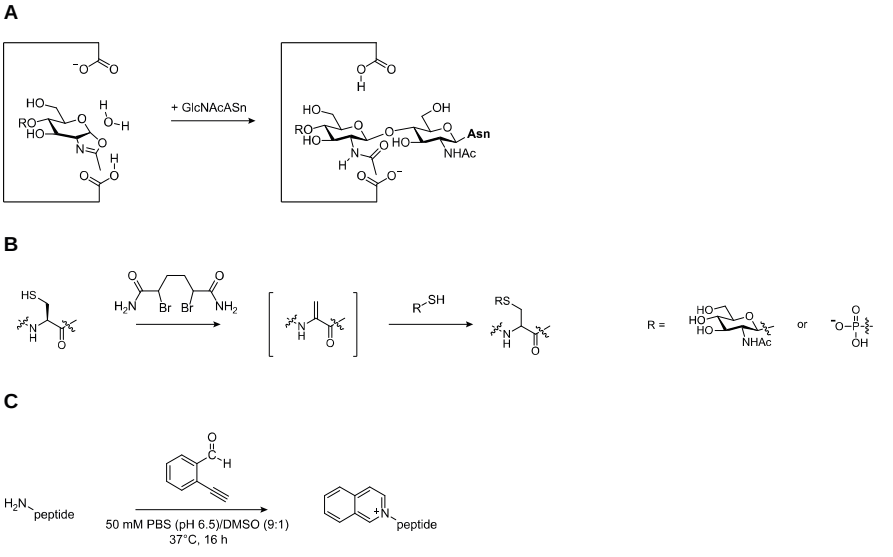
Bioconjugation has by and large relied on the reaction of electrophiles with the nucleophilic amino acids cysteine (Cys) and lysine (Lys) for (site-specific) modifications [92]. The scarcity of free cysteines in proteins has made this amino acid particularly

attractive for tag-and-modify approaches based on natural residues as this increases the chance of a predefined protein. However, other amino acids have also been studied, but – like lysines – their abundance has limited site-specific modification [93-99].

Cysteines are often used for modification as they contain a strong soft nucleophile under physiological conditions, that offers unique chemical reactivity. In combination with its low abundance in proteins [100] it is one of the more suitable amino acids that has been used for site-selective modification of a protein [90, 101]. One of the most commonly used cysteine modification techniques comprises alkylation with iodoacetamide-based reagents [102], and this technique is commonly used for capping free thiols for proteomics experiments [103]. This strategy has for instance been used by the Flitsch-group to create synthetic glycoproteins by reacting engineered thiols with glycosyl-iodoacetamides [104]. However, this reaction is not fully selective for thiols, with cross-reactivity to lysine residues observed at basic pH required to drive the reaction to completion. Maleimides react with a cysteine thiol in a Michael addition [105, 106], and has been employed for instance to produce Pneumococcal surface adhesin A (PsaA) mutants linked to a tetrasaccharide in a study on glycoconjugate vaccine candidates [107]. Moreover, a maleimide-thiol ligation reaction was used to site-specifically glycosylate the HIV-1 gp41 peptide, which is a potent anti-HIV agent [108].

Combining cysteine modification with the “tag-and-modify” approach, Davis and coworkers developed a toolbox to modify native and non-canonical amino acids, which allowed them to methylate, glycosylate, and phosphorylate proteins [81]. With this approach a library of PTM-modified histones – histones are known to undergo acetylation, phosphorylation, and ADP-ribosylation – were produced [109]. Except for histone H3, none of the human histones encompasses a natural cysteine, which makes this an obvious amino acid to introduce in recombinant histone proteins, and then modify. Cysteine-mutant histones were for instance reacted with a dibromide reagent to form a dehydroalanine (Dha), which served as precursor for the construction of close analogues (one sulfur substituting for one carbon) of six PTMs [110] (Scheme 1B).

Lysines can be efficiently modified by acylation using cyanates and isocyanates [111, 112], sulfonylation using sulfonyl fluorides [113] and, via reductive amination, with aldehydes [114]. Although aldehydes can be used for selective modification, reduction



Scheme 1. Schematic overview of different chemical amino acid modifications. A) Glycosylation of an asparagine via the oxazoline intermediate [116]. B) The modification of a cysteine via the Dha intermediate [110]. C) The modification of the N-terminal amine with the use of 2-EBA [121].

of the transiently formed imines can be detrimental to the structural, and therefore, functional integrity of the underlying protein [115]. Lysine is a highly abundant amino acid, and as such of much more limited use for site-selective modification than cysteines are.

The exception to this rule is perhaps the N-terminal amine residue, which under natural circumstances can be, for instance, acylated, methylated and ubiquitylated at the N-terminus [117-120]. The difference in pKa of the N-terminus compared to Lys-amines (pKa of about 8 for α- and 10 for ε-amino acids) allows for its selective modification at lower pH values [92, 120]. Under mildly acidic conditions, the use of 2-ethynylbenzaldehydes (2-EBA) has allowed selective modification of the N-termini [121], but modifications of this terminus with PTMs has not been reported (Scheme 1C). Moreover, the N-terminal amine residue can be selectively modified using 1.75 eq of a diazotransfer reagent – imidazole-1-sulfonyl azide – at pH 8.5 [122, 123].

1.5 Bioorthogonal chemistry for site-specific PTM introduction

All the above-described procedures are based on protein expression with the 20 proteinogenic amino acids. The next sections will describe methods to introduce additional chemical functionality during protein expression using non-canonical amino acids with functional groups the reactivity of which is orthogonal to those found in canonical amino acids. Three methods have been reported to achieve this: cell free translation [124, 125], the use of heterologous aaRS/tRNA pairs (for instance, amber codon suppression), and the use of non-canonical amino acids that are isosteric to canonical ones [126].

1.5.1 Incorporation of non-canonical amino acids via aaRS/tRNA pairs

Non-canonical amino acids can be incorporated in proteins using heterologous aaRS/tRNA pairs that bind certain stop codons [127]. This allows, amongst others, for site-specific introduction of PTMs [128], the introduction of amino acids containing fluorophores [129], photocrosslinking moieties [130, 131] as well as ligation handles [132].

The technique is centered around tRNAs – isolated from extremophilic bacteria – that recognise the codons normally assigned to bind the release factors RF1 or RF2 [133]. The three stop codons used for this method are known as amber (UAG), ochre (UAA) and opal/umber (UGA) [134]. For the incorporation of non-canonical amino acids via a heterologous aaRS/tRNA pair, the amber stop codon is most often used [135] (Figure 4), because of its limited use inside bacterial cells (~9%) [136, 137] and human cells (~23%) [138]. Several conditions have to be met to allow expression of a protein containing a bioorthogonal functionality with this system: the tRNA-synthase should not recognise any canonical amino acid, but should recognise the non-canonical one used and load it on the correct tRNA. This tRNA then, once charged with the desired amino acid should recognise the stop codon exclusively [127, 139]. It should be noted that the thus created UAC-tRNA bearing the non-canonical amino acid will be in competition with the release factor(s) that normally bind the stop codon, with a potentially reduced protein yield accompanied by the expression of truncated proteins as a result [140].

Early experiments on the termination codons were performed using an in vitro translation system [141]. Using this system, the *E. coli* chloramphenicol acetyltransferase

(cat) gene was modified [142] and mutating the anticodon sequence of the initiator tRNA from CAU to CUA could be used to initiate protein expression [143]. This research eventually led to the first active aaRS/tRNA pair for the incorporation of noncanonical amino acids (ncAAs) in *E. coli*: the tyrosyl pair of *Methanocaldococcus jannaschii,* that allowed incorporation of *O*-methyl-L-tyrosine in dihydrofolate reductase [144, 145]. Since then other aaRS/tRNA pairs have been developed, including the pyrrolysyl pair from *Methanosarcina barkeri* [146] and *Methanosarcina mazei* [147], the lysyl [148] and glutamyl [149] pair of *Pyrococcus horikoshii,* and the tryptophanyl pair of *Saccharomyces cerevisiae* [150]. The aminoacylation of these pairs was first shown using Northern blotting [151] or in stop codon read-through assays [152]. Quite recently, a tRNA extension method (tREX) was developed, utilizing a rapid screen and discovery of orthogonal aminoacyl-tRNA synthetase-tRNA pairs [153, 154]. This method uses an online database containing tRNA sequences, which can identify candidate orthogonal tRNAs [155].

Today, genetic code expansion technologies have reached maturity, at least for the design of non-canonical proteins in bacterial expression systems. A wide variety of proteins incorporating a non-canonical amino acid has seen the light in recent years and for a variety of purposes. Amongst these proteins it was green fluorescent protein (GFP), which was expressed with the incorporation of for instance p-azidophenylalanine (pAzpa) and p-propargyloxyphenylalanine (pPpa) in CHO and human 293T cells, using the amber codon suppressor - *Bst*RNA$^{Tyr}_{CUA}$ – from *B. stearothermophilus* [138].

Now that incorporation of one single non-canonical amino acid can be accomplished through genetic code expansion with considerable confidence, focus today is moving towards the incorporation of several non-canonical amino acids into a single protein [139]. Huang *et al.* were the first to introduce three acetylated lysine residues in model protein GFP. To do so, they introduced three amber stop codons at predetermined sites. To get useful quantities of the target protein, they had to suppress release factor 1, which was achieved by preparing a knockout of the N-terminal domain of ribosomal protein L11 [155]. Following these studies Xiao *et al.* showed that incorporation of two different non-canonical amino acids, namely O-methyltyrosine and ε-*tert*-Boc-lysine, could be achieved by using two aaRS/tRNA pairs in conjunction

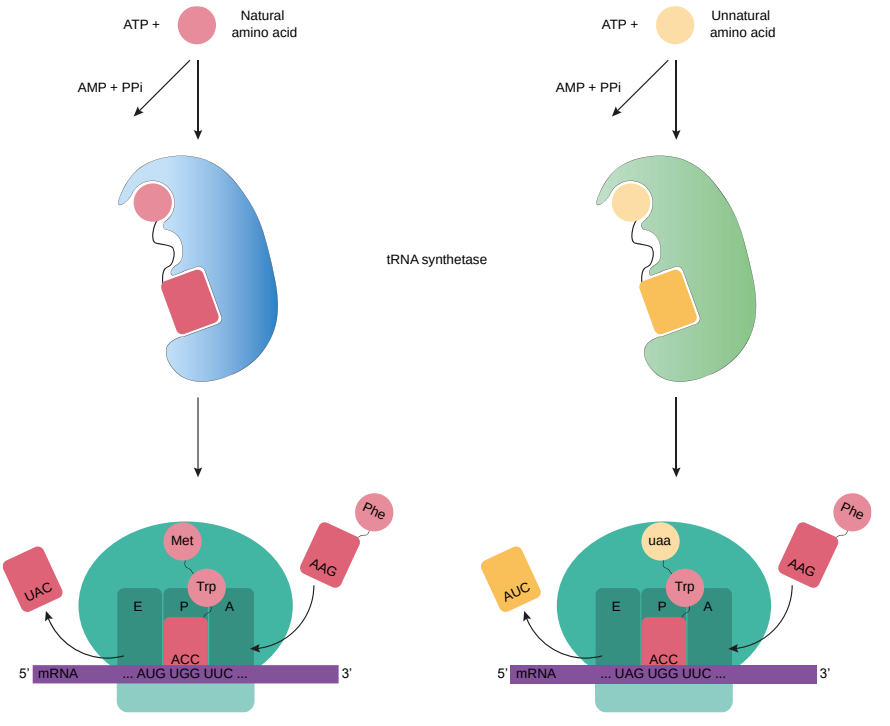


Figure 4. Schematic overview of the incorporation of unnatural amino acids into proteins using amber codon suppression. Left: The introduction of a natural amino acid. Right: The incorporation of an unnatural amino acid, using a modified tRNA synthetase.

with two different stop codons. These authors also showed that an antibody could be expressed with a p-acetylphenylalanine within the heavy chain and azidolysine within the light chain, using the amber (TAG) and ochre (TAA) stop codons respectively [156].

There are three natural stop codons that can be used to incorporate unnatural amino acids. This led the Chatterjee-group to expand genetic code expansion even further towards the incorporation of 5-hydroxytryptophan, p-azidophenylalanine and cyclopropene-lysine in GFP. With this, the limit of incorporation via manipulation of the three native stop codon usage was reached [157].

Genetic code expansion allows for the efficient incorporation of a variety of non-canonical amino acids in bacterial, yeast and mammalian cells [158, 159]. Protein yields proved a limiting factor in the early days [158, 160], but today proteins encompassing a single non-canonical amino acid can be obtained in good quantities: grams per liter

have been achieved in *E. coli* expression systems and 100s of milligrams in CHO cells [160].

1.5.2 Incorporation of non-canonical amino acids via metabolic labeling

The main alternative method to genetic code expansion for incorporating non-canonical amino acids into proteins is via metabolic labeling [127]. This method requires the activation of a non-canonical amino acid by a naturally occurring aminoacyl-tRNA synthetase that is in effect 'fooled' by the isosteric/isoelectronic nature of the new amino acid [161]. This method was first established with the incorporation of selenomethionine substituting for methionine to produce proteins suitable for structure determination by crystallography [162, 163]. Tirrell and coworkers then chose to use this methionine replacement method to attempt the incorporation of additional chemical functionalities into proteins to alter protein function. Even before the onset of bioorthogonal ligation chemistry they had reported the incorporation of amino acids, such as homoallylglycine [164], homopropargylglycine [165] and azidohomoalanine [166] into recombinant proteins for the use in conjugation reactions. The methodology is intrinsically more restricted in the diversity of amino acids that can be introduced in comparison with genetic code expansion: the non-canonical amino acid has to be structurally highly similar to the canonical one it is to substitute [167]. This canonical amino acid also should be absent, as otherwise it would compete with the unnatural analogue for incorporation [168], which dictates that auxotrophic strains are needed: ones that cannot produce the natural amino acid *de novo* [169]. Overcoming these limitations, Tirrell and coworkers revealed that proteins could be expressed with specific incorporation of non-canonical amino acids with a translation fidelity of >99% as shown by mass spectrometry [158].

Following the original report by Bertozzi and coworkers on the development of the Staudinger ligation as a bioorthogonal reaction [170], several groups employed Staudinger ligation chemistries for the modification of azide-containing proteins [171, 172]. Azide-containing proteins can also be modified using copper(I)-catalyzed azide alkyne [3+2] cycloaddition 'click' reactions [173, 174] or copper-free, strain-promoted varieties with for instance bicyclononyne (BCN) [175, 176]. Complementary chemistries

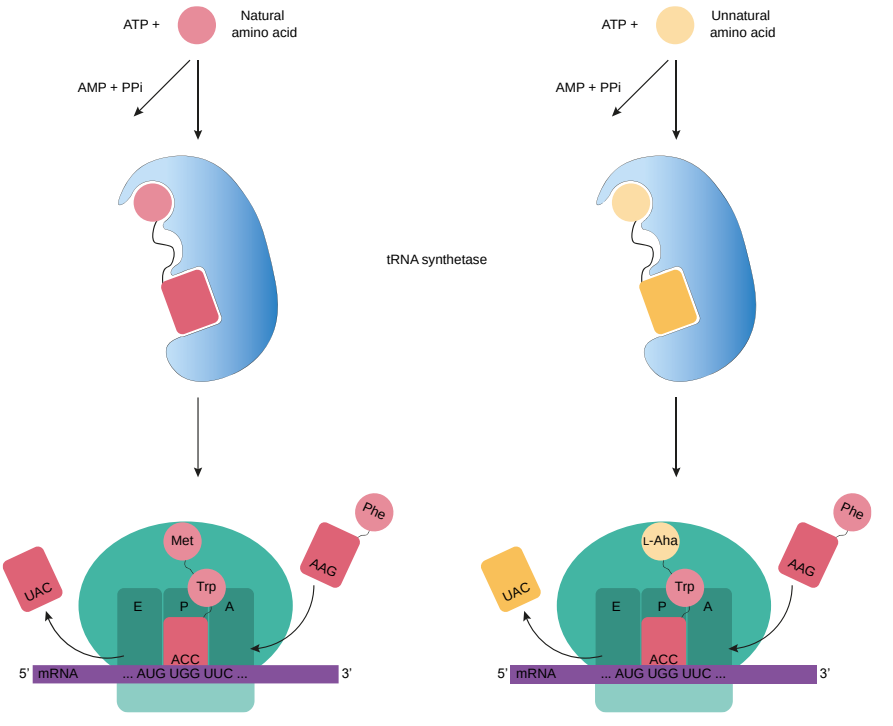


Figure 5. Schematic overview of the incorporation of unnatural amino acids in proteins using the native tRNA synthetase. Left: The introduction of a natural amino acid. Right: The incorporation of an unnatural amino acid, using the canonical tRNA synthetase.

have been reported as well and that make use of non-canonical amino acids introduced following the Tirrell protocol, for instance ones containing ketone moieties [177], which can react with functionalized hydrazides to form a hydrazine [178].

The above-described examples made use of methionine-auxotrophic *E. coli* expression strains. Since these reports, the approach has been expanded to other amino acid analogues, such as those of L-isoleucine [179], L-leucine [180], L-phenylalanine [181], L-proline [182] and L-tryptophan [183], using the corresponding auxotrophic expression strains.

1.5.3 Utilization of bioorthogonal amino acids incorporated in proteins.

Azidohomoalanine (L-Aha) and homopropargylglycine (L-Hpg) are extensively used as close L-methionine (L-Met) analogues. Their azide or alkyne functional groups can be used for bioorthogonal chemistry (Figure 5) [127, 184]. As the incorporation of

these isosteric amino acids in engineered proteins results in minimal differences in the protein construct compared to wild type counterparts, it offers a plethora of ligation possibilities.

Bioorthogonal labelling of proteins has been used by amongst others the groups of Bertozzi [92], Davis [185] and van Kasteren [186] in the visualization of specific proteins in biological systems. For instance, van Kasteren and Davies could site-selectively mutate all natural methionines in LacS into isoleucine residues and at the same time reintroduce a methionine at a preselected position. When the protein was expressed under Tirrell conditions (methionine-auxotroph strain, no methionine in the medium that did include azidohomoalanine) azide-modified LacS mutants were obtained that could be subsequently modified to carry glycan structures [186].

A variety of biological processes have been evaluated using the described metabolic labelling (section **1.5.2**). Among these is the labelling of immunogenic proteins for the visualization of antigen degradation inside antigen presenting cells [187]. Furthermore, metabolic labelling can be used for proteome labelling of bacteria, which can be subsequently used for detection purposes using correlated light-electron microscopy [188, 189] and for protein enrichment in secretome studies [190].

1.6 Aim of this Thesis

In this Thesis, results on the optimization of expression, and evaluation of bioorthogonal proteins in immunological systems is presented.

Chapter 2 of this Thesis describes the production and purification of ovalbumin (Ova), a common antigenic protein used for studies of both major histocompatibility complex (MHC) class I and II. In this Chapter it is attempted to optimize the production and purification of Ova containing the non-canonical amino acids L-Aha and L-Hpg instead of the natural methionine residue. This optimization was attempted using pET16b containing the following constructs: 6His-TEV-Ova, 10His-TEV-Ova and native Ova. It is shown that the first could be expressed in LB medium and with the incorporation of both unnatural amino acids. However, the yields proved to be low, most likely because of the formation of inclusion bodies during protein expression. For this reason, the second construct – 10His-TEV-Ova – was designed. Expressing this ovalbumin variant resulted in the formation of inclusion bodies even when the construct

was further extended with solubilization tags. For this reason, inclusion body purification was explored and although the yields for the methionine containing protein were substantial, it proved not applicable to Ova with the incorporation of L-Hpg. With this knowledge a final construct was designed, in which the purification tag was removed. This led to the expression of ovalbumin with the incorporation of L-Aha, but not L-Hpg and only in the in-house produced BL21::MetA auxotroph strain. **Chapter 2** ends with the comparison between the multiple expressed proteins.

As protein production as aimed for in **Chapter 2** proved troublesome, it was decided to switch attention to tetanus toxin C fragment (TTCF), results on which are described in **Chapter 3**. It was explored whether or not the MHC class II antigen presentation is altered when the antigenic epitope contains a ligation handle. For this, *TTCF* was modified at the antigenic region and the protein was produced with the incorporation of the unnatural amino acid L-Aha. After confirming the ligation possibilities, degradation of these modified proteins was visualized using bone marrow derived dendritic cells (BMDCs). Next, the antigen processing and presentation of the proteins is analyzed using the 2F2 T cell hybridoma. Presentation of each of the resulting 20 proteins proved possible, however, the IL-2 expression levels remained low. For this reason, the presentation was exemplified using ovalbumin MHC class II modified peptides and the use of A20s and DO11.10. Lastly, these azido-containing peptides were used for an initial immunoprecipitation assay.

In **Chapter 4**, horseradish peroxidase (HRP) was modified using a diazotransfer reaction on the lysine residues. This chemically modified protein was first assessed for its ability for chemical ligation reactions and afterwards used as a visualization tool. Using CuAAC ligation reactions, it was exemplified that alkylated proteins and bacteria can be ligated to the azidylated HRP. Moreover, it was attempted to ligate the protein to alkylated probes.

Finally, **Chapter 5** summarizes this Thesis and outlines some of the future aims using bioorthogonal proteins. It highlights the potential of this method, for instance by describing the potential protein-protein ligation, in which it is envisioned to homogenously ligate two proteins together by labelling one protein with an alkyne ligation handle and one with an azide handle. Furthermore, the site-specific ligation of

a glycan onto a protein is evaluated and initial antigen presentation studies are described in this Chapter. Moreover, alternative prospects of using protein ligation chemistry is described, of which the use of labelled fluorescent proteins is exemplified.

References

1. Chen, Z. and Zeng, A-P. (2016). Protein engineering approaches to chemical biotechnology. *Curr. Opin. Biotechnol. 42,* p.198-205
2. Reboul, E. and Borel, P. (2011). Proteins involved in uptake, intracellular tansport and basolateral secretion of fat-soluble vitamins carotenoids by mammalian enterocytes. *Prog. Lipid Res. 50,* p.388-402
3. Jaehme, M. and Slotboom, D.J. (2015). Diversity of membrane transport proteins for vitamins in bacteria and archaea. *Biochim. Biophys. Acta. 1850,* p.565-576
4. Movva, N.R., Nakamura, K. and Inouye M. (1980). Gene structure of the OmpA protein, a major surface protein of *Escherichia coli* required for cell-cell interaction. *J. Mol. Biol. 143,* p.317-328
5. Gloushankova, N.A., Rubtsova, S.N. and Zhitnyak, I.Y. (2017). Cadherin-mediated cell-cell interactions in normal and cancer cells. *Tissue Barriers. 5,* e1356900
6. Woo, S-R., Corrales, L. and Gajewski, T.F. (2015). Innate immune recognition of cancer. *Annu. Rev. Immunol. 33,* p.445-474
7. Lubbers, R., van Essen, M.F., van Kooten, C. and Trouw, L.A. (2017). Production of complement components by cells of the immune system. *Clin. Exp. Immunol. 188,* p.183-194
8. Rosano, G.L. and Ceccarelli, E.A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol. 5,* 172
9. Hartrampf, N., Saebi, A., Poskus, M., Gates, Z.P., Callahan, A.J., Cowfer, A.E., Hanna, S., Antilla, S., Schissel, C.K., Quartararo, A.J., Ye, X., Mijalis, A.J., Simon, M.D., Loas, A., Liu, S., Jessen, C., Nielsen, T.E. and Pentelute, B.L. (2020). Synthesis of proteins by automated flow cytometry. *Science. 368,* p.980-987
10. Young, C.L., Britton, Z.T. and Robinson, A.S. (2012). Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnol. J. 7,* p.620-634
11. Notarangelo, L.D., Kim, M-S., Walter, J.E. and Lee, Y.N. (2016). Human RAG mutations: Biochemistry and clinical implications. *Nat. Rev. Immunol. 16,* p.234-246
12. Crick, F. (1970). Central dogma of molecular biology. *Nature. 227,* p.561-563
13. Moore, M.J. and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell. 136,* p.688-700
14. Koonin, E.V. and Novozhilov, A.S. (2017). Origin and evolution of the universal genetic code. *Annu. Rev. Genet. 51,* p.45-62
15. Johnson, D.B.F and Wang, L. (2010). Imprints of the genetic code in the ribosome. *Proc. Natl. Acad. Sci. U.S.A. 107,* p.8298-8303
16. Rodina, M.V. and Wintermeyer, W. (2011). The ribosome as a molecular machine: The mechanism of tRNA-mRNA movement in translocation. *Biochem. Soc. Trans. 39,* p.658-662

17. Nakamura, Y. and Ito, K. (1998). How protein reads the stop codon and terminates translation. *Genes Cells. 3,* p.265-278
18. Dabrowski, M., Bukowy-Bieryllo, Z. and Zietkiewics, E. (2015). Translational readthrough potential of natural termination codons in eucaryotes – The impact of RNA sequence. *RNA Biol. 12,* p.950-958
19. Walter, S. and Buchner, J. (2002). Molecular chaperones – cellular machines for protein folding. *Angew. Chem. Int. Ed. Engl. 41,* p.1098-1113
20. Wang, R.F. and Kushner, S.R. (1991). Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli. Gene. 100,* p.195-199
21. Schmidt, F.R. (2004). Recombinant expression systems in the pharmaceutical industry. *Appl. Microbiol. Biotechnol. 65,* p.362-372
22. Mayfield, S.P., Franklin, S.E. and Lerner, R.A. (2003). Expression and assembly of a fully active antibody in algae. *Proc. Natl. Acad. Sci. U.S.A. 100,* 438-442
23. Sawasaki, T., Ogasawara, T., Morishita, R. and Endo, Y. (2002). A cell-free protein synthesis system for high-throughput proteomics. *Proc. Natl. Acad. Sci. U.S.A. 99,* p.14652-14657
24. Swartz, J.R. (2001). Advances in *Escherichia coli* production of therapeutic proteins. *Curr. Opin. Biotechnol. 12,* p.195-201
25. Valderrama-Rincon, J.D., Fisher, A.C., Merritt, J.H., Fan, Y-Y., Reading, C.A., Chhiba, K., Heiss, C., Azadi, P., Aebi, M. and DeLisa, M.P. (2012). An engineered eukaryotic protein glycosylation pathway in *Escherichia coli. Nat. Chem. Biol. 8,* p.434-436
26. Mamat, U., Wilke, K., Bramhill, D., Schromm, A.B., Lindner, B., Kohl, T.A., Corchero, J.L., Villaverde, A., Schaffer, L., Head, S.R., Souvignier, C., Meredith, T.C. and Woodard, R.W. (2015). Detoxifying *Escherichia coli* for endotoxin-free production of recombinant proteins. *Microb. Cell Fact. 14,* 57
27. Fabianek, R.A., Hennecke, H. and Thöny-Meyer, L. (1998). The active-site cysteines of the periplasmic thioredoxin-like protein CcmG of *Escherichia coli* are important but not essential for cytochrome *c* maturation in vivo. *J. Bacteriol. 180,* p.1947-1950
28. Fischer, B., Sumner, I. and Goodenough, P. (1993). Isolation, renaturation, and formation of disulfide bonds of eukaryotic proteins expressed in *Escherichia coli* as inclusion bodies. *Biotechnol. Bioeng. 41,* p.3-13
29. Carrío, M.M. and Villaverde, A. (2002). Construction and deconstruction of bacterial inclusion bodies. *J. Biotechnol. 96,* p.3-12
30. Singh, A., Upadhyay, V., Upadhyay, A.K., Singh, S.M. and Panda, A.K. (2015). Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microb. Cell. Fact. 14,* 41
31. Jong, W.S.P., Vikström, D., Houben, D., van den Berg van Saparoea, H.B., de Gier, J-W. and Luirink, J. (2017). Application of an *E. coli* signal sequence as a versatile inclusion body tag. *Microb. Cell Fact. 16,* 50

32. Missiakas, D. and Raina, S. (1997). Protein folding in the bacterial periplasm. *J. Bacteriol. 179,* p.2465-2471
33. Schwientek, T. and Ernst, J.F. (1994). Efficient intra- and extracellular production of human beta-1,4-galactosyltransferase in *Saccharomyces cerevisiae* is mediated by yeast secretion leaders. *Gene. 145,* p.299-303
34. Cregg, J.M., Cereghino, J.L., Shi, J. and Higgins, D.R. (2000). Recombinant protein expression in *Pichia pastoris. Mol. Biotechnol. 16,* p.23-52
35. Bonander, N. and Bill, R.M. (2012). Optimising yeast as a host for recombinant protein production (review). *Methods Mol Biol. 866,* p.1-9
36. Conde, R., Cueva, R., Pablo, G., Polaina, J. And Larriba, G. (2004). A search for hyperglycosylation signals in yeast glycoproteins. *J. Biol. Chem. 279,* p.43789-43798
37. Sampaio de Oliveira, K.B., Leite, M.L., Rodrigues, G.R., Morales Duque, H., Andrade da Costa, R., Cunha, V.A., de Loiola Costa, L., da Cunha, N.B., Franco, O.L. and Dias, S.C. (2020). Strategies for recombinant production of antimicrobial peptides with pharmacological potential. *Expert Rev. Clin. Pharmacol. 13,* p.367-390
38. Berger, I., Fitzgerald, D.J. and Richmond, T.J. (2004). Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol. 22,* p.1583-1587
39. Khoury, G.A., Baliban, R.C. and Floudas, C.A. (2011). Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep. 1,* 90
40. Cereghino, J.L. and Cregg, J.M. (2000). Heterologous protein expression in the methylotrophic yeasts *Pichia pastoris. FEMS Microbiology reviews. 24,* p.45-66
41. Vieira Gomes, A.M., Souza Carmo, T., Silva Carvalho, L., Mendonça Bahia, F. and Parachin, N.S. (2018). Comparison of yeasts as hosts for recombinant protein production. *Microorganisms. 6,* 38
42. Khan, K.H. (2013). Gene expression in mammalian cells and its application. *Adv. Pharm. Bull. 3,* p.257-263
43. Chen, C. and Okayama, H. (1987). High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell Biol. 7,* p.2745-2752
44. Cooray, S., Howe, S.J. and Thrasher, A.J. (2012). Retrovirus and lentivirus vector design and methods of cell conditioning. *Methods Enzymol. 507,* p.29-57
45. Kim, T.K. and Eberwine, J.H. (2010). Mammalian cell transfection: The present and the future. *Anal. Bioanal. Chem. 397,* p.3173-3178
46. Portolano, N., Watson, P.J., Fairall, L., Millard, C.J., Milano, C.P., Song, Y., Cowley, S.M. and Schwabe, J.W.R. (2014). Recombinant protein expression for structural biology in HEK 293F suspension cells: A novel and accessible approach. *J. Vis. Exp. 92,* 51897
47. Moore, G.E. and Ulrich, K. (1965). Suspension cultures of mammalian cells. A review. *J. Surg. Res. 5,* p.270-282
48. Studier, F.W. (1991). Use of the bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol. 219,* p.37-44

49. Shilling, P.J., Mirzadeh, K., Cumming, A.J., Widesheim, M., Köck, Z. and Daley, D.O. (2020). Improved designs for pET expression plasmids increase protein production yield in *Escherichia coli. Commun. Biol. 3,* 214
50. Yin, J., Li, G., Ren, X. And Herrler, G. (2007). Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *J. Biotechnol. 127,* p.335-347
51. Mierendorf, R.C., Morris, B.B., Hammer, B. and Novy, R.E. (1998). Expression and purification of recombinant proteins using the pET system. *Methods Mol. Med. 13,* p.257-292
52. Wurm, D.J., Veiter, L., Ulonska, S., Eggenreich, B., Herwig, C. and Spadiut, O. (2016). The *E. coli* pET expression system revisited – mechanistic correlation between glucose and lactose uptake. *Appl. Microbiol. Biotechnol. 100,* p.8721-8729
53. Bertani, I., Devescovi, G. and Venturi, V. (1999). Controlled specific expression and purification of 6x His-tagged proteins in *Pseudomonas. FEMS. Microbiol. Lett. 179,* p.101-106
54. Bujard, H., Gentz, R., Lanzer, M., Stueber, D., Mueller, M., Ibrahimi, I., Haeuptle, M.T. and Dobberstein, B. (1987). A T5 promoter-based transcription-translation system for the analysis of proteins *in vitro* and *in vivo. Methods Enzymol. 155,* p.416-433
55. Khlebnikov, A., Risa, Ø., Skaug, T., Carrier, T.A. and Keasling, J.D. (2000). Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J. Bacteriol. 182,* p.7029-7034
56. Bateman, B.T., Donegan, N.P., Jarry, T.M., Palma, M. and Cheung, A.L. (2001). Evaluation of a tetracycline-inducible promoter in *Staphylococcus aureus* in vitro and in vivo and its application in demonstrating the role of *sigB* in microcolony formation. *Infect. Immun. 69,* p.7851-7857
57. Wegerer, A., Sun, T. and Altenbuchner, J. (2008). Optimization of an *E. coli* L-rhamnose-inducible expression vector: Test of various genetic module combinations. *BMC Biotechnol. 8,* 2
58. Ooi, B.G., Rankin, C. and Miller, L.K. (1989). Downstream sequences augment transcription from the essential initiation site of a baculovirus polyhedrin gene. *J. Mol. Biol. 210,* p.721-736
59. Flick, J.S. and Johnston, M. (1990). Two systems of glucose repression of the GAL1 promoter in *Saccharomyces cerevisiae. Mol. Cell Biol. 10,* p.4757-4769
60. Uversky, V.N. (2013). Posttranslational modification. Brenner's Encyclopedia of Genetics. Academic Press. 2nd edition. p.425-430
61. Walsh, C.T., Garneau-Tsodikova, S. and Gatto, G.J. Jr. (2005). Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chem. Int. Ed. Engl. 44,* p.7342-7372

62. Shental-Bechor, D. and Levy, Y. (2008). Effect on glycosylation on protein folding: A close look at thermodynamic stabilization. *Proc. Natl. Acad. Sci. U.S.A. 105,* p.8256-8261
63. Corfield, A.P. (2014). Mucins: a biologically relevant glycan barrier in mucosal protection. *Biochim. Biophys. Acta. 1850,* p.236-252
64. Iwona, R. and Katarzyna, S. (2020). Glycosylation of proteins of human skin fibroblasts is changed by rosmarinic acid. *Naunyn Schmiedebergs Arch. Pharmacol. 393,* p.419-427
65. Marth, J.D. and Grewal, P.K. (2008). Mammalian glycosylation in immunity. *Nat. Rev. Immunol. 8,* p.874-887
66. Wang, Y-C., Peterson, S.E. and Loring, J.F. (2014). Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res. 24,* p.143-160
67. Pucci, B., Kasten, M. and Giordano, A. (2000). Cell cycle and apoptosis. *Neoplasia. 2,* p.291-299
68. Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev. 12,* p.599-606
69. Lecker, S.H., Goldberg, A.L. and Mitch, W.E. (2006). Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. *J. Am. Soc. Nephrol. 17,* p.1807-1819
70. Audgnotto, M. and Peraro, M.D. (2017). Protein post-translational modifications: *In silico* prediction tools and molecular modeling. *Comput. Struct. Biotechnol. J. 15,* p.307-319
71. Varier, R.A. and Timmers, H.T.M. (2011). Histone lysine methylation and demethylation pathways in cancer. *Biochim. Biophys. Acta. 1815,* p.75-89
72. Pugliese, A. (2017). Autoreactive T cells in type 1 diabetes. *J. Clin. Invest. 127,* p.2881-2891
73. Chung, H.S., Wang, S-B., Venkatraman, V., Murray, C.I. and van Eyk, J.E. (2013). Cysteine oxidative posttranslational modifications: Emerging regulation in the cardiovascular system. *Circ. Res. 112,* p.382-392
74. Longo, V.D. and Kennedy, B.K. (2006). Sirtuins in aging and age-related disease. *Cell. 126,* p.257-268
75. Hart, G.W., Slawson, C., Ramirez-Correa, G. and Lagerlof, O. (2011). Cross talk between O-GlcNAcylation and phosphorylation: Roles in signaling, transcription, and chronic disease
76. Craveur, P., Narwani, T.J., Rebehmed, J. and de Brevern, A.G. (2019). Investigation of the impact of PTMs on the protein backbone conformation. *Amino Acids. 51,* p.1065-1079
77. Barber, K.W. and Rinehart, J. (2018). The ABCs of PTMs. *Nat. Chem. Biol. 14,* p.188-192

78. Fernández-Tejada, A., Brailsford, J., Zhang, Q., Shieh, J-H., Moore, M.A.S. and Danishefsky, S.J. (2016). Total synthesis of glycosylated proteins. *Top. Curr. Chem. 362,* p.1-26
79. Zhang, Y., Park, K-Y., Suazo, K.F. and Distefano, M.D. (2018). Recent progress in enzymatic protein labelling techniques and their applications. *Chem. Soc. Rev. 47,* p.9106-9136
80. Chalker, J.M., Bernardes, G.J.L. and Davis, B.G. (2011). A "tag-and-modify" approach to site-selective protein modification. *Acc. Chem. Res. 44,* p.730-741
81. Wright, T.H., Bower, B.J., Chalker, J.M., Bernardes, G.J.L., Wiewiora, R., Ng, W-L., Raj, R., Faulkner, S., Vallée, M.R.J., Phanumartwiwath, A., Coleman, O.D., Thézénas, M-L., Khan, M., Galan, S.R.G., Lercher, L., Schombs, M.W., Gerstberger, S., Palm-Espling, M.E., Baldwin, A.J., Kessler, B.M., Claridge, T.D.W., Mohammed, S. and Davis, B.G. (2016). Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science. 354,* aag1465
82. Grotenberg, G. & Ploegh, H. (2007). Dressed-up proteins. *Nature. 446,* p.994-995
83. Davis, B.G. (2002). Synthesis of glycoproteins. *Chem. Rev. 102,* p.579-602
84. Wang, P., Dong, S., Shieh, J-H., Peguero, E., Hendrickson, R., Moore, M.A.S. and Danishefsky, S.J. (2013). Erythropoietin derived by chemical synthesis. *Science. 342,* p.1357-1360
85. Fernández-Tejada, A., Haynes, B.F. and Danishefsky, S.J. (2015). Designing synthetic vaccines for HIV. *Expert Rev. Vaccines. 14,* p.815-831
86. Gorelik, A. and van Aalten, D.M.F. (2020). Tools for functional dissection of site-specific O-GlcNAcylation. *RSC Chem. Biol. 1,* p.98-102
87. Chao, Q., Ding, Y., Chen, Z-H., Xiang, M-H., Wang, N. and Gao, X-D. (2020). Recent progress in chemo-enzymatic methods for the synthesis of N-glycans. *Front. Chem. 8,* 513
88. Huang, W., Li, C., Li, B., Umekawa, M., Yamamoto, K., Zhang, X. and Wang, L-X. (2010). Glycosynthases enable a highly efficient chemoenzymatic synthesis of N-glycoproteins carrying intact natural N-glycans. *J. Am. Chem. Soc. 131,* p.2214-2223
89. Shivatare, S.S., Huang, L-Y., Zeng, Y-F., Liao, J-Y., You, T-H., Wang, S-Y., Cheng, T., Chiu, C-W., Chao, P., Chen, L-T., Tsai, T-I., Huang, C-C., Wu, C-Y., Lin, N-H. and Wong, C-H. (2018). Development of glycosynthases with broad glycan specificity for the efficient glycol-remodeling of antibodies. *Chem. Commun. (Camb). 54,* p.6161-6164
90. Chalker, J.M., Bernardes, G.J.L., Lin, Y.A. and Davis, B.G. (2009). Chemical modification of proteins at cysteine: Opportunities in chemistry and biology. *Chem. Asian J. 4,* p.630-640
91. Spicer, C.D. and Davis, B.G. (2014). Selective chemical protein modification. *Nat. Commun. 5,* 4740
92. Sletten, E.M. and Bertozzi, C.R. (2009). Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed. Engl. 48,* p.6974-6998

93. Ward, C.C., Kleinman, J.I. and Nomura, D.K. (2017). NHS-esters as versatile reactivity-based probes for mapping proteome-wide ligandable hotspots. *ACS Chem. Biol. 12,* p.1478-1483
94. Bloom, S., Liu, C., Kölmel, D.K., Qiao, J.X., Zhang, Y., Poss, M.A., Ewing, W.R. and MacMillan, D.W.C. (2018). Decarboxylative alkylation: An approach to site-selective bioconjugation of native proteins via oxidation potentials. *Nat. Chem. 10,* p.205-211
95. deGruyter, J.N., Malins, L.R. and Baran, P.S. (2017). Residue-specific peptide modification: A chemist's guide. *Biochemistry. 56,* p.3863-3873
96. Ban, H., Nagano, M., Gavrilyuk, J., Hakamata, W., Inokuma, T. and Barbas, C.F. (2013). Facile and stabile linkages through tyrosine: Bioconjugation strategies with the tyrosine-click reaction. *Bioconjugate Chem. 24,* p.520-532
97. Cong, Y., Pawlisz, E., Bryant, P., Balan, S., Laurine, E., Tommasi, R., Singh, R., Dubey, S., Peciak, K., Bird, M., Sivasankar, A., Swierkosz, J., Muroni, M., Heidelberg, S., Farys, M., Khayrzad, F., Edwards, J., Badescu, G., Hodgson, I., Heise, C., Somavarapu, S., Liddell, J., Powel, K., Zloh, M., Choi, J-W., Godwin, A. and Brocchini, S. (2012). Site-specific PEGylation at histidine tags. *Bioconjugate Chem. 23,* p.248-263
98. Antos, J.M. and Francis, M.B. (2004). Selective tryptophan modification with rhodium carbenoids in aqueous solution. *J. Am. Chem. Soc. 126,* p.10256-10257
99. Taylor, M.T., Nelson, J.E., Suero, M.G. and Gaunt, M.J. (2018). A protein functionalization platform based on selective reactions at methionine residues. *Nature. 562,* p.563-568
100. Rudolph, T.K. and Freeman, B.A. (2009). Transduction of redox signaling by electrophile-protein reactions. *Sci. Signal. 2,* re7
101. Gunnoo, S.B. and Madder, A. (2016). Chemical protein modification through cysteine. *Chembiochem. 17,* p.529-553
102. Creighton, T.E. (1984). Disulfide bond formation in proteins. *Methods Enzymol. 107,* p.305-329
103. Sechi, S. and Chait, B.T. (1998). Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Anal. Chem. 70,* p.5150-5158
104. Macmillan, D., Bill, R.M., Sage, K.A., Fern, D. and Flitsch, S.L. (2001). Selective in vitro glycosylation of recombinant proteins: semi-synthesis of novel homogeneous glycoforms of human erythropoietin. *Chem. Biol. 8,* p.133-145
105. Nair, D.P., Podgórski, M., Chatani, S., Gong, T., Xi, W., Fenoli, C.R. and Bowman, C.N. (2014). The thiol-michael addition click reaction: A powerful and widely used tool in materials chemistry. *Chem. Mater. 26,* p.724-744
106. Ravesco, J.M.J.M., Faustino, H., Trindade, A. and Gois, P.M.P. (2019). Bioconjugation with maleimides: A useful tool for chemical biology. *Chemistry. 25,* p.43-59
107. Pillot, A., Defontaine, A., Fateh, A., Lambert, A., Prasanna, M., Fanuel, M., Pipelier, M, Csaba, N., Violo, T., Camberlain, E. and Grandjean, C. (2019). Site-specific

conjugation for fully controlled glycoconjugate vaccine preparation. *Front Chem. 7,* 726

108. Ni, J., Singh, S. and Wang, L-X. (2003). Synthesis of maleimide-activated carbohydrates as chemoselective tags for site-specific glycosylation of peptides and proteins. *Bioconjug. Chem. 14,* p.232-238
109. Nadal, S., Raj, R., Mohammed, S. and Davis, B.G. (2018). Synthetic post-translational modification of histones. *Curr. Opin. Chem. Biol. 45,* p.35-47
110. Chalker, J.M., Lercher, L., Rose, N.R., Schofield, C.J. and Davis, B.G. (2012). Conversion of cysteine into dehydroalanine enables access to synthetic histones bearing diverse post-translational modifications. *Angew. Chem. Int. Ed. Engl. 51,* p.1835-1839
111. Boutureira, O. and Bernardes, G.J.L. (2015). Advances in chemical protein modification. *Chem. Res. 115,* p.2174-2195
112. Stark, G.R., Stein, W.H. and Moore, S. (1960). Reactions of the cyanate present in aqueous urea with amino acids and proteins. *J. Biol. Chem. 235,* p.3177-3181
113. Narayanan, A. and Jones, L.H. (2015). Sulfonyl fluorides as privileged warheads in chemical biology. *Chem. Sci. 6,* p.2650-2659
114. Spears, R.J. and Fascione, M.A. (2016). Site-selective incorporation and ligation of protein aldehydes. *Org. Biomol. Chem. 14,* p.7622-7638
115. Apel, C., Kasper, M-A., Stieger, C.E., Hackenberger, C.P.R. and Christmann, M. (2019). Protein modification of lysine with 2-(2-styrylcyclopropyl)ethanal. *Org. Lett. 21,* p.10043-10047
116. Fairbanks, A.J. (2017). The ENGases: Versatile biocatalysts for the production of homogeneous *N*-linked glycopeptides and glycoproteins. *Chem. Soc. Rev. 46,* p.5128-5146
117. Meinnel, T., Serero, A. and Giglione, C. (2006). Impact of the N-terminal amino acid on targeted protein degradation. *Biol. Chem. 387,* p.839-851
118. Varland, S., Osberg, O. and Arnesen, T. (2015). N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. *Proteomics. 15,* p.2385-2401
119. Rosen, C.B. and Francis, M.B. (2017). Targeting the N terminus for site-selective protein modification. *Nat. Chem. Biol. 13,* p.697-705
120. Koniev, O. and Wagner, A. (2015). Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation. *Chem. Soc. Rev. 44,* p.5495-5551
121. Deng, J-R., Lai, N.C-H., Kung, K. K-Y., Yang, B., Chung, S-F., Leung, A.S-L., Choi, M-C., Leung, Y-C. and Wong, M-K. (2020). N-terminal selective modification of peptides and proteins using 2-ethynylbenzaldehydes. *Comm. Chem. 3,* 67
122. Lohse, J., Swier, L.Y.M., Oudshoorn, R.C., Médard, G., Kuster, B., Slotboom, D-J., Witte, M.D. (2017). Targeted diazotransfer reagents enable selective modification of proteins with azides. *Bioconjugate Chem. 28,* p.913-917

123. van Dongen, S.F., Teeuwen, R.L., Nallani, M., van Berkel, S.S., Cornelissen, J.J., Notte, R.J. and van Hest, J.C. (2009). Single-step azide introduction in proteins via an aqueous diazo transfer. *Bioconjug. Chem. 20,* p.20-23
124. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K. and Ueda, T. (2001). Cell-free translation reconstituted with purified components. *Nat. Biotechnol. 19,* p.751-755
125. Smolskaya, S., Logashina, Y.A. and Andreev, Y.A. (2020). *Escherichia coli* extract-based cell-free expression system as an alternative for difficult-to-obtain protein biosynthesis. *Int. J. Mol. Sci. 21,* 928
126. Melnikov, S.V. and Söll, D. (2019). Aminoacyl-tRNA synthetases and tRNAs for an expanded genetic code: What makes them orthogonal? *Int. J. Mol. Sci. 20,* 1929
127. Lang, K. and Chin, J.W. (2014). Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev. 114,* p.4764-4806
128. Mukai, T., Kobayashi, T., Hino, N., Ynagisawa, 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,* p.818-822
129. Katritzky, A.R. and Narindoshvili, T. (2009). Fluorescent amino acids: Advances in protein-extrinsic fluorophores. *Org. Biomol. Chem. 7,* p.627-634
130. Chin, J.W., Martin, A.B., King, D.S., Wang, L. and Schultz, P.G. (2002). Addition of photocrosslinking amino acid to the genetic code of *Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 99,* p.11020-11024
131. Guo, J., Melançon, C.E. III, Lee, H.S., Groff, D. and Schultz, P.G. (2009). Evolution of amber suppressor tRNAs for efficient bacterial production of unnatural amino acid-containing proteins. *Angew. Chem. Int. Ed. Engl. 48,* p.9148-9151
132. Kim, C.H., Axup, J.Y. and Schultz, P.G. (2015). Protein conjugation with genetically encoded unnatural amino acids. *Curr. Opin. Chem. Biol. 17,* p.412-419
133. Nakamura, Y. and Ito, K. (1998). How protein read the stop codon and terminates translation. *Genes Cells. 3,* p.265-278
134. Clark, D.P., Pazdernik, N.J. and McGehee, M.R. (2019). Chapter 26 – Mutations and Repair. *Molecular Biology* 3rd ed. p.832-879
135. Lin, X., Yu, A.C.S. and Chan, T.F. (2017). Efforts and challenges in engineering the genetic code. *Life. 7,* 12
136. Nakamura, Y., Gojobori, T. and Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: Status for the year 2000. *Nucleic Acids. Res. 28,* p.292
137. Korkmaz, G., Holm, M., Wiens, T. and Sanyal, S. (2014). Comprehensive analysis of stop codon usage in bacteria and its correlation with release factor abundance. *J. Biol. Chem. 289,* p.30334-30342

138. 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,* p.239-244
139. Xiao, H. and Schultz, P.G. (2016). At the interface of chemical and biological synthesis: An expanded genetic code. *Cold Spring Harb. Perspect Biol. 8,* a023945
140. Wang, K., Neumann, H., Peak-Chaw, S.Y. and Chin, J.W. (2007). Evolved orthogonal ribosomes enhance the effieciency of synthetic genetic code expansion. *Nat. Biotechnol. 25,* p.770-777
141. Anthony-Cahill, S.J., Griffith, M.C., Noren, C.J., Suich, D.J. and Schultz, P.G. (1989). Site-specific mutagenesis with unnatural amino acids. *Trends Biochem. Sci. 14,* p.400-403
142. Capone, J.P., Sedivy, J.M., Sharp, P.A. and RajBahndary, U.L. (1986). Introduction of UAG, UAA, and UGA nonsense mutations at a specific site in the *Escherichia coli* chloramphenicol acetyltransferase gene: Use in measurement of amber, ochre, and opal suppression in mammalian cells. *Mol. Cell. Biol. 6,* p.3059-3067
143. Varshney, U. and RajBhandary, U.L. (1990). Initiation of protein synthesis from a termination codon. *Proc. Natl. Acad. Sci. U.S.A. 87,* p.1586-1590
144. Wang, L., Magliery, T.J., Liu, D.R. and Schultz, P.G. (2000). A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins. *J. Am. Chem. Soc. 122,* p.5010-2011
145. Wang, L., Brock, A., Herberich, B. and Schultz, P.G. (2001). Expanding the genetic code of *Escherichia coli. Science. 292,* p.498-500
146. Ambrogelly, A., Gundllapalli, S., Herring, S., Polycarpo, C., Frauer, C. and Söll, D. (2007). Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc. Natl. Acad. Sci. U.S.A. 104,* p.3141-3146
147. Namy, O., Zhou, Y., Gundllapalli, S., Polycarpo, C.R., Denise, A., Rousset, J-P., Söll, D. and Ambrogelly, A. (2007). Adding pyrrolysine to the *Escherichia coli* genetic code. *FEBS Lett. 581,* p.5282-5288
148. 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,* p.7566-7571
149. Santoro, S.W., Anderson, J.C., Lakshman, V. and Schultz, P.G. (2003). An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in *Escherichia coli. Nucleic Acids Res. 31,* p.6700-6709
150. Chatterjee, A., Xiao, H., Yang, P-Y., Soundararajan, G and Schultz, P.G. (2013). A tryptophanyl-tRNA synthetase/tRNA pair for unnatural amino acid mutagenesis in *E. coli. Angew. Chem. Int. Ed. Engl. 52,* p.5106-5109
151. Varshney, U., Lee, C.P. and RajBhandary, U.L. (1991). Direct analysis of aminoacylation levels of tRNAs *in vivo. J. Biol. Chem. 266,* p.24712-24718

152. Wenthzel, A.M., Stancek, M. and Isaksson, L.A. (1998). Growth phase dependent stop codon readthrough and shift of translation reading frame in *Escherichia coli. FEBS Lett. 421,* p.237-242
153. Cervettini, D., Tang, S., Fried, S.D., Willis, J.C.W., Funke, L.F.H., Colwell, L.J. and Chin, J.W. (2020). Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase-tRNA pairs. *Nat. Biotechnol. Online ahead of print.*
154. Koehler, C., Girona, G.E., Reinkemeier, C.D. and Lemke, E.A. (2020). Inducible genetic code expansion in eukaryotes. *Chembiochem. Online ahead of print.*
155. Huang, Y., Russell, W.R., Wan, W., Pai, P-J., Hussell, D.H. and Liu, W. (2010). A convenient method for genetic incorporation of multiple noncanonical amino acids into one protein in *Escherichia coli. Mol. Biosyst. 6,* p.683-686
156. Xiao, H., Chatterjee, A., Choi, S-H., Bajjuri, K.M., Sinha, S.C. and Schultz, P.G. (2013). Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. *Angew. Chem. Int. Ed. Engl. 52,* p.14080-14083
157. Italia, J.S., Addy, P.S., Erickson, S.B., Peeler, J.C., Weerapana, E. and Chatterjee, A. (2019). Mutually orthogonal nonsense-suppression systems and conjugation chemistries for precise protein labeling at up to three distinct sites. *J. Am. Chem. Soc. 141,* p.6204-6212
158. Xie, J. and Schultz, P.G. (2005). An expanding genetic code. *Methods. 36,* p.227-238
159. Liu, C.C. and Schultz, P.G. (2010). Adding new chemistries to the genetic code. *Annu. Rev. Biochem. 79,* p.413-444
160. Chin, J.W. (2017). Expanding and reprogramming the genetic code. *Nature. 550,* p.53-60
161. Jafarnejad, S.M., Kim, S-H. and Sonenberg, N. (2018). Aminoacylation of proteins: New targets for the old ARSenal. *Cell Metab. 27,* p.1-3
162. Cohen, G.N. and Cowie, D.B. (1957). Total replacement of methionine by selenomethionine in proteins of *Escherichia coli. C. R. Hebd Seances Acad. Sci. 244,* p.680-683
163. Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science. 254,* p.51-58
164. van Hest, J.C.M., Kiick, K.L. and Tirrell, D.A. (2000). Efficient incorporation of unsaturated methionine analogues into proteins in vivo. *J. Am. Chem. Soc. 122,* p.1282-1288
165. Kiick, K.L., Saxon, E., Tirrell, D.A. and Bertozzi, C.R. (2002). Incorporation of azides into recombinant proteins for chemoselective modification by Staudinger ligation. *Proc. Natl. Acad. Sci. U.S.A. 99,* p.19-24
166. Kiick, K.L., van Hest, J.C.M. and Tirrell, D.A. (2000). Expanding the scope of protein biosynthesis by altering the methionyl-tRNA synthetase activity of a bacterial expression host. *Angew. Chem. Int. Ed. Engl. 39,* p.2148-2152
167. Budisa, N., Karnbrock, W., Steinbacher, S., Humm, A., Prade, L., Neuefeind, T., Moroder, L. and Huber, R. (1997). Bioincorporation of telluromethionine into

proteins: A promising new approach for X-ray structure analysis of proteins. *J. Mol. Biol. 270,* p.616-623

168. Tang, Y. and Tirrell, D.A. (2002). Attenuation of the editing activity of the *Escherichia coli* leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. *Biochemistry. 41,* p.10635-10645
169. Voloshchuk, N. and Montclare, J.K. (2010). Incorporation of unnatural amino acids for synthetic biology. *Mol. Biosyst. 6,* p.65-80
170. Saxon, E., Armstrong, J.I. and Bertozzi, C.R. (2000). A “traceless” Staudinger ligation for the chemoselective synthesis of amide bonds. *Org. Lett. 2,* p.22141-2143
171. Köhn, M. and Breinbauer, R. (2004). The staudinger ligation – a gift to chemical biology. *Angew. Chem. Int. Ed. Engl. 43,* p.3106-3116
172. Meyers, E.L. and Raines, R.T. (2011). A phosphine-mediated conversion of azides to diazo-compounds. *Angew. Chem. Int. Ed. Engl. 48,* p.2359-2363
173. 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. Eng. 41,* p.2596-2599
174. Wang, Q., Chan, T.R., Hilgraf, R., Fokin, V.V., Sharpless, K.B. and Finn, M.G. (2003). Bioconjugation by copper(I)-catalyzed azide-alkyne [3+2] cycloaddition. *J. Am. Chem. Soc. 125,* p.3192-3193
175. 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,* p.15046-15047
176. Lallana, E., Riguera, R. and Fernandez-Megia, E. (2011). Reliable and efficient procedures for the conjugation of biomolecules through Huisgen azide-alkyne cycloadditions. *Angew. Chem. Int. Ed. Engl. 50,* p.8794-8804
177. Datta, D., Wang, P., Carrico, I.S., Mayo, S.L. and Tirrell, D.A. (2002). A designed phenylalanyl-tRNA synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins. *J. Am. Chem. Soc. 124,* p.5652-5653
178. Ngo, J.T. and Tirrell, D.A. (2011). Non-canonical amino acids in the interrogation of cellular protein synthesis. *Acc. Chem. Res. 44,* p.677-685
179. Wang, P., Tang, Y. and Tirrell, D.A. (2003). Incorporation of trifluoroisoleucine into proteins in vivo. *J. Am. Chem. Soc. 125,* p.6900-6906.
180. Tang, Y., Ghirlanda, G., Petka, W.A., Nakajima, T., DeGrado, W.F. and Tirrell, D.A. (2001). *Angew. Chem. Int. Ed. Engl. 40,* p.1494-1496
181. Sharma, N., Furter, R., Kast, P. and Tirrell, D.A. (2000). Efficient introduction of aryl bromide functionality into proteins in vivo. *FEBS lett. 467,* p.37-40
182. Renner, C., Alefelder, S., Bae, J.H., Budisa, N, Huber, R. and Moroder, L. (2001). Fluoroprolines as tools for protein design and engineering. *Angew. Chem. Int. Ed. Engl. 40,* p.923-925

183. Budisa, N., Alefelder, S., Bae, J.H., Golbik, R., Minks, C., Huber, R. and Moroder, L. (2001). Proteins with ß-(thienopyrrolyl)alanines as alternative chromophores and pharmaceutically active amino acids. *Protein Sci. 10,* p.1281-1292
184. Fang, K.Y., Lieblich, S.A. and Tirrell, D.A. (2018). Incorporation of non-canonical amino acids into proteins by global reassignment of sense codons. *Methods Mol. Biol. 1798,* p.173-186
185. Dieterich, D.C., Link, A.J., Graumann, J., Tirrell, D.A. and Schuman, E.M. (2006). Selective identification of newly synthesized proteins in mammalian cells using biorthogonal noncanonical amino acid tagging (BONCAT). *PNAS, 103,* p.9482-9487
186. van Kasteren, S.I., Kramer, H.B., Jensen, H.H., Campbell, S.J., Kirkpatrick, J., Oldham, N.J., Anthony, D.C. and Davis, B.G. (2007). Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature. 446,* p.1105-1109
187. Araman, M.C., Pieper-Pournara, L., van Leeuwen, T., Kampstra, A.S.B., Bakkum, T., Marqvorsen, M.H.S., Nascimento, C.R., Groenewold, G.J.M., Wulp, W., Camps, M.G.M., Overkleeft, H.S., Ossendorp, F.A., Toes, R.E.M. and van Kasteren, S.I. (2019). Bioorthogonal antigens allow the unbiased study of antigen processing and presentation. *BioRxiv.* p.439323
188. van Elsland, D.M., Bos, E., de Boer, W., Overkleeft, H.S., Koster, A.J. and van Kasteren, S.I. (2016). Detection of bioorthogonal groups by correlative light and electron microscopy allows imaging of degraded bacteria in phagocytes. *Chem. Sci. 7,* p.752-758
189. van Elsland, D.M., Pujals, S., Bakkum, T., Bos, E., Oikonomeas-Koppasis, N., Berlin, I., Neefjes, J., Meijer, A.H., Koster, A.J., Albertazzi, L. and van Kasteren, S.I. (2018). Ultrastructural imaging of Salmonella-host interactions using super-resolution correlative light-electron microscopy of bioorthogonal pathogens. *Chembiochem. 19,* p.1766-1770
190. Eichelbaum, K., Winter, M., Berriel Diaz, M., Herzig, S. and Krijgsveld, J. (2012). Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nat. Biotechnol. 30,* p.984-990