The advantages and disadvantages of bioorthogonal proteins
Groenewold, G.J.M.

Citation

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).
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
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
Introduction

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

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

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,
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

### 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).
Introduction

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].
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.
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 \textit{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
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 \(\alpha\)- and 10 for \(\varepsilon\)-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 \[122, 123\].

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 \[121\].
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
Introduction

(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 - BsfRNA\textsubscript{Tyr, CUA} - from *B. stea\textit{thermophilus* [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
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 noncanonical 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.
Introduction

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
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.
Chapter 1

References

Introduction


Chapter 1


~ 30 ~
Introduction

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


Introduction


Introduction


Chapter 1


