COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

Volume No: 9, Issue: 14, e201402001, http://dx.doi.org/10.5936/csbj.201402001

CSBJ

SITE-SPECIFIC FUNCTIONALIZATION OF PROTEINS AND THEIR APPLICATIONS TO THERAPEUTIC ANTIBODIES

Remko van Vught [a], Roland J Pieters [b], Eefjan Breukink [a,*]

Abstract: Protein modifications are often required to study structure and function relationships. Instead of the random labeling of lysine residues, methods have been developed to (sequence) specific label proteins. Next to chemical modifications, tools to integrate new chemical groups for bioorthogonal reactions have been applied. Alternatively, proteins can also be selectively modified by enzymes. Herein we review the methods available for site-specific modification of proteins and their applications for therapeutic antibodies.

MINI REVIEW ARTICLE

Introduction

Proteins are the working horses of a living cell. Within and around cells they perform a magnificently diverse set of functions. Besides providing structure and stability, proteins are involved in cell signaling, catalyzing reactions, storage and transport, and are therefore extensively studied. Over the years, tools have become available for researchers to reveal structure and function relationships, as well as localization and their interactions with other proteins.

A relatively new tool is based on novel and specific chemistry. By modifying existing amino acids or introducing unnatural amino acids, proteins can be manipulated at the single amino acid level. Several methods involving the site-specific modification of proteins have been reported in the last decade. This allows the spatial and temporal control of proteins *in vivo*, as well as single molecule tracking. Modifications are introduced during protein translation, as post translational modification or chemically, after protein isolation.

Besides their usefulness for *in vitro/vivo* research, site-specific modifications are also interesting for therapeutic applications. Pharmaceutical companies have been refocusing their pipeline towards biological medicines (mainly monoclonal antibodies) because of the high specificity and safety. The 'naked' monoclonal antibodies have shown to be very effective in blocking receptors. A next generation of biological medicines are the antibody drug conjugates (ADCs), which efficiently deliver the payload to the target limiting the off target effects. Interestingly, site-specific modifications have also been applied to improve the properties of these therapeutic proteins.

Here, we review the tools for site-specific modification of proteins, followed by their applications in the development of therapeutic antibodies.

[a]*Department of Membrane Biochemistry and Biophysics, Institute of Biomembranes, Utrecht University, Padualaan 8, 3584CH Utrecht, The Netherlands*

[b]*Department of Medicinal Chemistry and Chemical Biology. Utrecht Institute for Pharmaceutical Sciences, Utrecht University, P.O. Box 80082, 3508 TB Utrecht, The Netherlands*

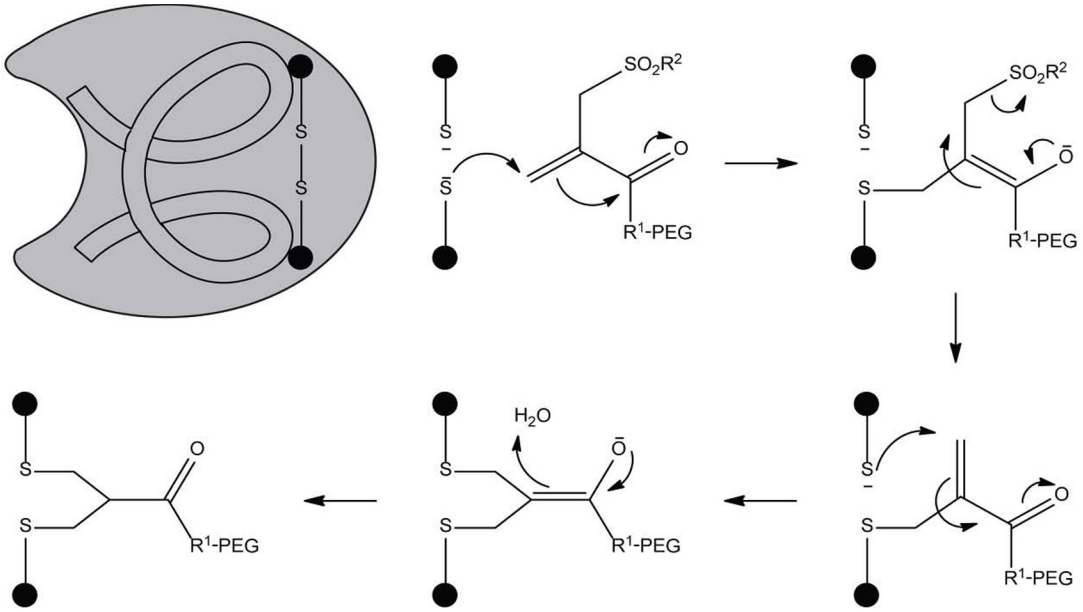
* Corresponding author.

E-mail address: Eefjan Breukink (E.J.Breukink@uu.nl)

Chemical modifications of proteins

The oldest and most straightforward method for labeling proteins is via the primary amino groups on lysine residues and at the N-terminus. In general, multiple accessible lysines and thus reactive amines are present on the protein surface, resulting in efficient labeling but inevitably leading to heterogeneous mixtures. Whether this method is applicable depends on the properties of the protein and the application. In the case of monoclonal antibodies, random labeling with fluorescent molecules hardly affects the antigen binding since many primary amines are present and only a small fraction may be important for this interaction. Smaller proteins such as antibody fragments are more likely to suffer from random conjugation due to a reduced number of lysines and the lack of an Fc region. There have been attempts to make this modification more specific by using preferential N-terminal labeling [1] or kinetically controlled lysine labeling [2]. Generally those methods suffer from low yields or require complex steps including the recycling of the original protein. Besides labeling the amino groups, similar obstacles exist for conjugation via carboxyl groups [3] and will therefore not be discussed in detail.

More selective is the labeling of proteins via sulfhydryl groups (also known as thiols). In proteins, most of the thiols are present in covalently linked pairs as disulfide bonds. The introduction of a cysteine by site-directed mutagenesis can be used for selective conjugation. Coupling reactions of maleimide groups with thiols have a high specificity over amines due to the lower pKa of the SH group (>1000 fold selectivity at pH 7.0) [4]. Therefore, cysteines are most commonly used for the site-selective modifications of proteins, though in some situations it is not feasible. One major drawback of introducing an extra cysteine is protein misfolding due to non-native disulfide bridge formation. In addition, thiol maleimide adducts have been reported to have limited stability *in vivo* [5]. Reactive thiols in albumin, free cysteine or glutathione can exchange with the existing thiol maleimide complex. Interestingly, hydrolysis of the succinimide ring prevented this exchange reaction [5]. Whether other alkylation reactions (with iodo/bromoacetamide analogs) also suffer from limited stability *in vivo* needs to be determined. Alternatively, an elegant double alkylation method by reducing disulfide bridges on the protein surface and subsequent conjugation with a PEG monosulfone-enone reagent was stable in human serum for over 30 hours and did not affect the protein stability (scheme 1) [6].



Scheme 1. Double alkylation of proteins by PEG monosulfone-enone.

Next to direct protein modification via alkylation, a reduced cysteine can be first converted to dehydroalanine. Subsequent nucleophilic addition by thiol modified biomolecules label the target protein via a thioether bond. This method is a straightforward route to natural occuring cysteine modifications including phosphor [7], farnesyl [8] and N-acetylhexosamine cysteine [9], and to structural mimics of post-translational modifications, but generates epimeric products due to loss of the stereocenter in the first step. Recently, several strategies for the conversion of cysteine to dehydroalanine have been evaluated [10].

Over the years, several site-specific chemical modifications methods have been reported for the N-terminal amino acids. N-terminal serine and threonine residues can selectively be oxidized by sodium periodate to form an aldehyde [11], followed by oxime ligation [12]. Besides oxime ligation, the oxidized serine was recently also used for the one step N-terminal dual protein functionalization using strain promoted alkyne–nitrone cycloaddition [13].

Proteins with N-terminal cysteines have been successfully used for reactions with thioesters [14] and applied for fusion proteins through native chemical ligation [15], which will be described in more detail later on.

More elegant methods are independent of the N-terminal amino acid. These approaches exploit the unique chemical properties of the N-terminus including the low pKa of the α-amino group of the N-terminus (8.9) compared to the pKa of the lysine ε-amino group (10.5). Kinetically controlled lysine labeling is performed in small steps, using multiple additions of the label and allowing the most reactive amino group to be preferentially labeled [2].

Other methods are based on the introduction of unique reactive groups. The diazotransfer reagent imidazole-1-sulfonyl azide was shown to specifically convert the N-terminal amino group into an azide group [16]. The N-terminus can also be converted into a ketone or aldehyde group by a transamination reaction [17]. Peptide library screening identified residues with high yields (A, G, D, E, N,), other amino acids were either not/less reactive or were prone to side reactions [18]. In more recent work the transamination reaction was demonstrated for labeling of a monoclonal antibody [19]. Alternatively, N-terminal modification based on ketenes was applied to introduce an alkyne in peptides and proteins [20]. This reaction is highly specific for most N-terminal amino acids but yields range from 9 to 94%.

Although these methods are generally straightforward for peptides, applications for proteins predominantly depend on the solvent accessibility of the N-terminus. Moreover, small modifications limit the usefulness of reactions with low yields due to difficulties in separating the modified from the unmodified proteins.

Metabolic modifications

Metabolic labeling of proteins involves the replacement of one or more canonical amino acids by non-canonical analogs. The first observations by Munier and Cohem showed the incorporation of phenylalanine and methionine analogs in bacterial proteins (red scheme in figure 1) [21]. Since then, many analogs have been synthesized and tested in auxotrophic bacterial hosts for incorporation at the expense of canonical amino acids [22]. The strict biological machinery accepts only minor modifications such as alkenes [23], alkynes [24] and azides [24] as amino acid side chains. The latter being of particular interest due to their compatibility with the Staudinger ligation and (copper-free) click chemistry [25].

The occurrence of multiple phenylalanine or methionine residues in proteins results in protein mixtures upon conjugation. Recently though, only one out of five azidohomoalanines of native CalB was shown to be surface accessible and reactive for functionalization [26].

Instead of designing amino acid analogs to be accepted by the biological machinery, advances have been made to manipulate the biosynthetic apparatus itself. Mutations in phenylalanyl-tRNA synthetase (PheRS) caused either an increase or decrease of the binding pocket size, and thus a change in the specificity towards phenylalanine analogs (blue scheme in figure 1) [27]. The unnatural amino acid *p*-chlorophenylalanine could be incorporated into *Photinus pyralis* luciferase by expression of the mutant PheRS (A294G) in *E. coli*, replacing all phenylalanines [28]. More recently, the same has been demonstrated for non-canonical analogs using mutations in LeuRS [29], PheRS [30] and ValRS [31].