# METHOD OF PRODUCING AN IMMUNOLIGAND/PAYLOAD CONJUGATE BY MEANS OF A SEQUENCE-SPECIFIC TRANSPEPTIDASE ENZYME

---

The present invention is related to methods of producing an immunoligand/payload conjugate

## Introduction

Currently, the predominant methods to label and/or to conjugate molecules to proteins, especially, when small-molecule payloads or labels are concerned, involve the chemical conjugation with specific linker molecules that covalently attach the payload to free lysine and/or cysteine amino acids of the proteins.

However, many proteins, like e.g. antibodies that are of particular interest for immunotargeting strategies, are fairly large proteins, that may contain several lysine and cysteine residues. Because linker-mediated, chemical conjugation is a stochastic process, linker-mediated chemical ligation of payloads leads to heterogeneous mixtures of conjugated proteins that may differ in their therapeutic efficacy and/or diagnostic utility. Obviously, mixtures of protein-payload conjugates also represent a significant challenge in the regulatory approval process for therapeutic conjuagtes, as batch-to-batch variation and/or variations in the active pharmaceutical ingredient (API) are negatively viewed by regulatory authorities due to potential safety concerns

In addition, if a defined ratio of payload to protein is desired, it is often necessary to purify the conjugate with the desired conjugation stoichiometry. This is not only tedious, but can significantly add to the cost-of-goods in the manufacturing process, as often only a fraction of the linker-mediated conjugated protein represents the desired ratio of payload

conjugation. This is particularly true for therapeutically relevant antibody/drug conjugates (ADCs), where depending on the toxin employed, 3 to 4 toxin molecules appear to be advantageous, but antibodies with no toxin coupled to up to 8 toxins per antibody coupled are found in typical linker-mediated chemical conjugation reactions (Panowski et al. (2014)).

Despite of the limitations described above, all antibody/drug conjugates currently in clinical trials, or approved by the health authorities for the therapy of disease, have been generated by linker-mediated chemical conjugation of toxic small-molecule drugs to antibodies (Lambert (2012) or Mullard (2013)).

It is widely acknowledged in the industry and by scientific experts in the field, that site-specific and stoichiometric conjugation of molecular payloads, including toxin or label molecules to immunoligands would have significant advantages in comparison to chemical, linker-mediated conjugation. This is evidenced by attempts to target the chemical conjugation to specific amino acids in the protein structure (Panowski et al. (2014)).

On one hand, this is attempted by mutating certain positions in the protein structure to delete unwanted and/or to provide desired conjugation sites (i.e. lysine and/or cysteine residues) to which the linker-ligation can be targeted (McDonagh et al. (2006) or Junutula et al. (2008)).

On the other hand, control of chemical conjugation to proteins is attempted by incorporation of unnatural amino acids at certain positions, like selenocysteine, p-azidophenylalanine, or acetylphenylalanine (Hofer et al. (2009), Axup et al. (2012), or Lemke (2011)).

However, all of these approaches change the primary amino acid sequence of the protein to be conjugated, and may result in undesired functional properties. Furthermore, the incorporation of unnatural amino acids, as described above, is often low efficient, and does not allow for a quantitative incorporation of specific labeling sites to proteins.

**Summary of the invention**

Therefore, there is an urgent need in the industry to overcome the known issues of stochastic conjugation methods in particular for the generation of therapeutically relevant immunoconjugates, including, but not limited to ADCs.

It is thus one object of the present invention to provide an efficient method for conjugating immunoligands and payloads, e.g., drugs, toxins, cytokines, markers, or the like, preferably full-length monoclonal antibodies to small-molecular weight toxins, for the generation of site-specifically conjugated antibody drug conjugates (ADCs).

It is another object of the present invention to create immunoligand/payload conjugates, which have better efficacy and/or can be produced with higher reproducibility.

It is another object of the present invention to allow the conjugation of payloads to immunoligands in a site-specific and/or sequence specific manner.

It is another object of the present invention to create immunoligand/payload conjugates which preserve the characteristic features of its components, e.g., target affinity, target specificity, target sensitivity, solubility, pharmacological function and the like
These objects are achieved by the subject matter of the independent claims, while the dependent claims as well as the specification disclose further preferred embodiments.

**Definitions**

As used herein, the term "immunoligand" is meant to define an entity, an agent or a molecule that has affinity to a given target, e.g., a receptor, a cell surface protein, a cytokine or the like. Such immunoligand may optionally block or dampen agonist-mediated responses, or inhibit receptor-agonist interaction. Most importantly, however, the immonoligand may serve as a shuttle to deliver a payload to a specific site, which is defined by the target recognized by said immunoligand. Thus, an immunoligand targeting, for instance, but not limited to a receptor, delivers its payload to a site which is characterized by abundance of said receptor. Immunoligands include, but are not limited to, antibodies, antibody fragments, antibody-based binding proteins, antibody mimetics, receptors, soluble decoy receptors, scaffold proteins with affinity for a given target and ligands of receptors.

"Antibodies", also synonymously called "immunoglobulins" (Ig), are generally comprising four polypeptide chains, two heavy (H) chains and two light (L) chains, and are therefore multimeric proteins, or an equivalent Ig homologue thereof (e.g., a camelid nanobody, which comprises only a heavy chain, single domain antibodies (dAbs) which can be either

be derived from a heavy or light chain); including full length functional mutants, variants, or derivatives thereof (including, but not limited to, murine, chimeric, humanized and fully human antibodies, which retain the essential epitope binding features of an Ig molecule, and including dual specific, bispecific, multispecific, and dual variable domain immunoglobulins; Immunoglobulin molecules can be of any class (e.g., IgG, IgE, IgM, IgD, IgA, and IgY), or subclass (e.g., IgG1, IgG2, IgG3, IgG4, IgA1, and IgA2) and allotype.

An "antibody-based binding protein", as used herein, may represent any protein that contains at least one antibody-derived $V_H$, $V_L$, or $C_H$ immunoglobulin domain in the context of other non-immunoglobulin, or non-antibody derived components. Such antibody-based proteins include, but are not limited to (i) $F_c$-fusion proteins of binding proteins, including receptors or receptor components with all or parts of the immunoglobulin $C_H$ domains, (ii) binding proteins, in which $V_H$ and or $V_L$ domains are coupled to alternative molecular scaffolds, or (iii) molecules, in which immunoglobulin $V_H$, and/or $V_L$, and/or $C_H$ domains are combined and/or assembled in a fashion not normally found in naturally occuring antibodies or antibody fragments.

An "antibody drug conjugate" (ADC), as used herein, relates to either an antibody, or an antibody fragment, or and antibody-based binding protein, coupled to a small molecular weight active pharmaceutical ingredient (API), including, but not limited to a toxin (including e.g., but not limited to, tubulin inhibitors, actin binders, RNA polymerase inhibitors, DNA-intercalating and modifying/damaging drugs), a kinase inhibitor, or any API that interferes with a particular cellular pathway that is essential for the survival of a cell and/or essential for a particular physiologic cellular pathway.

An "antibody derivative or fragment", as used herein, relates to a molecule comprising at least one polypeptide chain derived from an antibody that is not full length, including, but not limited to (i) a Fab fragment, which is a monovalent fragment consisting of the variable light ($V_L$), variable heavy ($V_H$), constant light ($C_L$) and constant heavy 1 ($C_H1$) domains; (ii) a F(ab')2 fragment, which is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; (iii) a heavy chain portion of a $F_{ab}$ ($F_d$) fragment, which consists of the $V_H$ and $C_H1$ domains; (iv) a variable fragment ($F_v$) fragment, which consists of the $V_L$ and $V_H$ domains of a single arm of an antibody, (v) a domain antibody (dAb) fragment, which comprises a single variable domain; (vi) an isolated complementarity

determining region (CDR); (vii) a single chain $F_v$ Fragment ($scF_v$); (viii) a diabody, which is a bivalent, bispecific antibody in which $V_H$ and $V_L$ domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with the complementarity domains of another chain and creating two antigen binding sites; and (ix) a linear antibody, which comprises a pair of tandem $F_v$ segments ($V_H$-$C_H1$-$V_H$-$C_H1$) which, together with complementarity light chain polypeptides, form a pair of antigen binding regions; and (x) other non-full length portions of immunoglobulin heavy and/or light chains, or mutants, variants, or derivatives thereof, alone or in any combination.

The term "modified antibody format", as used herein, encompasses antibody-drug-conjugates, Polyalkylene oxide-modified scFv, Monobodies, Diabodies, Camelid Antibodies, Domain Antibodies, bi- or trispecific antibodies, IgA, or two IgG structures joined by a J chain and a secretory component, shark antibodies, new world primate framework + non-new world primate CDR, IgG4 antibodies with hinge region removed, IgG with two additional binding sites engineered into the CH3 domains, antibodies with altered Fc region to enhance affinity for Fc gamma receptors, dimerised constructs comprising CH3+VL+VH, and the like.

The term "antibody mimetic", as used herein, refers to proteins not belonging to the immunoglobulin family, and even non–proteins such as aptamers, or synthetic polymers. Some types have an antibody-like beta-sheet structure. Potential advantages of "antibody mimetics" or "alternative scaffolds" over antibodies are better solubility, higher tissue penetration, higher stability towards heat and enzymes, and comparatively low production costs.

Some antibody mimetics can be provided in large libraries, which offer specific binding candidates against every conceivable target. Just like with antibodies, target specific antibody mimetics can be developed by use of High Throughput Screening (HTS) technologies as well as with established display technologies, just like phage display, bacterial display, yeast or mammalian display. Currently developed antibody mimetics encompass, for example, ankyrin repeat proteins (called DARPins), C-type lectins, A-domain proteins of S. aureus, transferrins, lipocalins, 10th type III domains of fibronectin, Kunitz domain protease inhibitors, ubiquitin derived binders (called affilins), gamma crystallin derived binders, cysteine knots or knottins, thioredoxin A scaffold based binders,

nucleic acid aptamers, artificial antibodies produced by molecular imprinting of polymers, peptide libraries from bacterial genomes, SH-3 domains, stradobodies, “A domains” of membrane receptors stabilised by disulfide bonds and Ca2+, CTLA4-based compounds, Fyn SH3, and aptamers (oligonucleic acid or peptide molecules that bind to a specific target molecules)

In case the immunoligand is not a protein or a peptide, e.g., if its an aptamer, it should preferably be provided with a peptide tag in order to provide a suitable substrate for the enzymatic conjugation disclosed further herein.

"Conjugation", as used herein, relates to the covalent association of a molecule to another molecule by formation of a covalent bond.

An "immunotoxin", as used herein, relates to an immunoligand conjugated to a protein or polypeptide representing a toxin, including, but not limited to bacterial toxins, e.g. diphteria-toxin A, Pseudomonas exotoxin, botulinum toxin, or e.g. proteinaceous venoms from invertebrates (e.g. but not limited spiders, scorpions, molluscs, jelly-fish), or vetrebrates (e.g., but not limited to snakes), or functional fragments thereof.

The term "low molecular-weight payload" as used herein, represents a payload with a molecular weight not exceeding 2'500 Dalton.

The term "payload", as used herein, represents any naturally occuring or synthetically generated molecule, including small-molecular weight molecules or chemical entities that can chemically be synthesized, and larger molecules or biological entities that need to be produced by fermentation of host cells and that confer a novel functionality to an immunoligand specific for binding to targets or antigens.

The term "small molecular weight toxin", as used herein, means a cytotoxic compound of small molecular weight not exceeding a molecular weight of 2'500 Dalton that is cytotoxic to mammalian cells.

A "transpeptidase", as used herein, is an enzyme or a catalytic domain of an enzyme or a protein that is able to catalyze the breakage of peptide bonds and subsequently either directly, or by way of several reaction intermediates, the formation of novel peptide bonds,

such that the energy of the first peptide bond is preserved during the reaction and transfered to a new peptide bond. Preferably, said transpeptidases preferably connect the C-terminus of one peptide or protein with the N-terminus of another peptide or protein. Due to the formation of a new peptide bond, these enzymes or functional domains are also refered to as "protein ligases", "peptide ligases", or nicknamed "protein or peptide staplers". Such protein ligases comprise, but are not limited to sortase enzymes, inteins and split-inteins.

As used herein, the term "sequence-specific transpepeptidase" is meant to define a transpeptidase which needs at least one substrate peptide or protein with a given peptide sequence as recognistion sequence (N-terminally and/or C-terminally) to connect said substrate peptide or protein to another peptide or protein, or a small-molecular weight compound containing a peptide or protein component

As used herein, the term "site-specific transpepeptidase" is meant to define a transpeptidase which has a specific site in at least one substrate peptide or protein which it uses to conjugate to another peptide or protein, or a small-molecular weight compound containing a peptide or protein component.

**Background and general description of the invention**

The invention discloses methods that utilize site-specific transpeptidases, e.g., sortase enzymes and split-inteins, to site-specifically and selectively conjugate payloads, preferably small molecular weight toxins to immunoligands, preferably antibodies, for the generation of immunoligand payloads, preferably antibody drug conjugates (ADCs). The preferred payloads are small molecular weight toxins modified with short, preferably less than 13 (thirteen) amino acid long synthetic amino acid sequence, which renders them as substrates for sortase enzymes or split intein mediated covalent conjugation either at the N- or C-terminus of the immunoligands (Figures 1 & 3). This conjugation is achieved in a site-specific manner and with defined stoichiometry, which is a distinguishing feature to conventional chemical conjugation of payloads to immunoligands, where the conjugation is a stochastic process, as disclosed further above.

The invention further discloses site specific transpeptidase, e.g. sortase or split-intein mediated conjugation of multimeric immunoligands, preferably antibodies specifically with two different toxin molecules or other labels using different modifications of the subunits of

the multimeric protein, e.g. antibody heavy and light chains, and different payloads modified with different, short amino acid stretches specific for different transpeptidases, in order to conjugate at least two different functional payloads to the multimeric immunoligand (Fig. 6).

The invention further discloses methods to add affinity purification and/or detection tags to the N- or C-termini of the immunoligands, which undergo enzyme-mediated transpeptidation. such that the removal of the affinity purification and/or detection tag can be utilized to select for immunologands with complete (100%) conjugation of the payload to the modified binding protein, by means of affinity resins that retain immunoligands that have not been completely conjugated, and therefore still retain the additional affinity purification and/or detection tag (Fig. 4).

The invention further discloses immunoligands in which a catalytic transpeptidase domain is directly fused to the N- or C-terminus of the protein to be conjugated, such that the transpeptidation activity is integral part of immunoligand to be conjugated, and no additional soluble sortase enzyme needs to be provided in the course of the transpeptidase-mediated conjugation reaction (Fig. 5).

All of these embodiments mentioned above allow the site-specific and stoichiometrically controlled conjugation of any payload, including small molecule toxins (chemical entities), toxic proteins, or fluorescent labels, preferably small molecular weight toxins to immunoligands, including preferably antibodies, which is superior to standard chemical conjugation of payloads to proteins by chemical linker chemistry methods, which cannot be controlled for conjugation ratio and site. Therefore, for the generation of antibody drug conjugates (ADCs) conjugation of toxic payloads by transpeptidases, preferably sortase enzymes and split inteins to antibodies will lead to more homogeneous products with expected improved therapeutic properties for cancer therapy (Fig. 12).

The enzymatic conjugation of payloads to immunoligands by sortase enzymes and split-intein allows site-specific and stoichiometric payload conjugation to proteins and immunoligands, lowering cost-of-goods and providing homogeneous immunoligand-payload conjugates, especially as the selectivity of the transpeptidases allows the conjugation of payloads to immunoligands in crude cell culture supernatant, and does not

require purified components as in traditional linker-mediated chemical conjugation. Therefore, the use of sequence-specific transpeptidases for conjugation of payloads to immunoligands could significantly lower the cost of goods in immunoligand-payload, and particularly ADC manufacturing.

The first type of transpeptidase disclosed herein, the sortase enzymes, has been identified in a variety of gram-positive bacteria, like *Staphylococcus, Streptococcus* and *Pneumococcus* species, and catalyse the coupling of virulence factors to cell wall proteoglycans, in order to change the surface signature of the bacteria for evading an efficient immune response by the infected host (Mazmanian et al. (1999)). Sortase A enzyme of the gram-positive bacterium *Staphylococcus aureus* has been characterized first (Ton-That et al. (1999)) and has subsequently been characterized further as a tool for many protein modifications (Tsukiji (2009)). The attraction of sortase enzymes is that the two molecules to be conjugated only require to be modified or expressed on one hand with a short 5 amino-acid long peptide tag (sortase tag, LPXTG in case of *Staphylococcus aureus* sortase A, X being any of the 20 naturally occuring aminoacids), and a short, preferably 3 to 5 amino acid long glycine stretch (Antos et al. (2009a)) (Fig. 1), which can easily be added to each of the molecules to achieve either N-terminal or C-terminal conjugation of proteins. This allows to utilize the system on one hand for the coupling or conjugation of two proteins, but also for the conjugation of smaller molecules to proteins.

The second type of transpeptidase resulting in peptide-bond cleavage and formation, is represented by the so-called inteins, which have originally been discovered as protein introns, that can remove (splice) themselves out of precursor proteins by cleavage of peptide bonds and formation of new peptide-bonds (Xu et al. (1993)) (Fig. 2a). Inteins can also occur separated into N-intein and C-intein domains (so-called split-inteins) and attached to independent proteins that can subsequently catalyze the *trans*-splicing of the extein domains (Fig. 2b). Split-inteins have been utilized for the covalent coupling of N-extein and C-extein moieties, and also the purification and/or circularization of proteins (Elleuche (2010)). However, in order to utilize split-inteins also for the conjugation of small molecule payloads, it is necessary to utilize split inteins that function if either the N-intein or the C-intein domain can be redused to few amino acids, that can easily be added to molecules of any size by chemical synthesis, similar to the short at preferably 3 glycine stretch required for sortase-mediated transpeptidation. With the development of the artificial Ssp GyrB S11

split-intein, in which the C-intein domain only comprises six amino acids (Sun et al. (2004)), this condition has been met and this split-intein has been utilized for the C-terminal labeling of proteins with biotin (Volkmann et al. (2009)) (Fig. 3a). Likewise, the development of a short 11 amino acid long N-intein from Ssp DnaX split-intein allows the N-terminal conjugation of proteins with any molecule, if such 11 amino acid long stretch is added by chemical synthesis to a payload of choice (Fig. 3b).

Therefore, one aspect of the invention is either to add a short, preferably 3 to 5 glycine amino-acid glycine stretch, or a short 12 amino-acid GVFVHNSXXXXX amino acid stretch (X any naturally occuring or artificial amino acids), containing a 6 amino acid C-int domain of Ssp GyrB or a 11 amino-acid N-int domain of Ssp DnaX to a payload-molecule, which is sufficient to allow the respective trasnpeptidase to conjugate the modified payload to proteins and immunoligands, preferably antibodies that, respectively, contain a sortase enzyme recognition motif, e.g. LPXTG in case of utilization of *Staphylococcus aureus* Sortase A, or a 150aa N-int domain in case of utilization of Ssp GyrB split intein, or a 139 aa C-int domain in case of utilization of Ssp DnaX split intein (see Figs 1 & 3).

The addition of short stretches of amino acids, like e.g. 3 or 5 glycine residues to a small molecular weight toxins as required for sortase mediated conjugation, or 12 amino acids as required for split-intein mediated conjugation, has been found to add to the water-solubility of certain hydrophobic toxin molecules (data not shown), such that the amino acid-toxin adduct can be dissolved in the physiologic buffer, ensuring optimal sortase or split-intein conjugation. This prevents stress on the strutural integrity of large protein molecules, particularly antibodies that can easily be denatured by exposure to organic solvents and non-physiologic pH often associated with traditional linker chemistry and conjugation. In addition, conjugation of hydrophobic toxin molecules to large proteins, particular antibodies can induce certain levels of protein aggregation. Also this may be improved by using transpeptidases, particularly sortase enzymes, because further hydrophilic amino acids remain in the enzymatically generated conjugate, reducing the propensity for aggregation of large protein, or antibody drug conjugates.

Sortase enzymes have been widely described in the prior art for protein-protein or protein-peptide ligations (Mao et al. (2004), Parthasarathy et al. (2007) or WO2011/133704A2), even including circularization of proteins (Antos et al. (2009b)). The applications of sortase

protein or peptide ligation also included protein or peptide ligation using antibody fragments, like Fab- and scFv-fragments with protein- or peptide labels (Möhlmann et al. (2011), Madej et al. (2012), or US2010/0055761A1 and WO2012/142659A1). Even two prior art documents were published, in which full-length antibodies have been sortase-ligated to proteins (Levary et al. (2011), e.g. EGFP, albumin, gelonin were conjugated to the light chain of an antibody), or in which full-length antibodies have been sortase-ligated to short peptides (Swee et al. (2013)). However, no prior art document could be identified demonstrating the sortase-mediated conjugation of small-molecular weight toxins, like e.g. auristatins or maytansins and the like, to full-length antibodies or antibody fragments. In particular no prior art documents could be identified, in which generation of ADCs with small molecular weight toxins has been been disclosed resulting in ADCs with small molecular weight toxins homogeneously conjugated to either IgH or IgL chains (drug-to-antibody ratio 2), or to IgH and IgL chains (drug to antibody ratio 4), as disclosed herein.

While the prior art also discloses the modification of non-protein substrates with glycine residues such that they could be used for sortase modification of simple, single-subunit proteins or peptides (Tsukiji (2009), or WO2007/108013A3, respectively), the more challenging homogeneous conjugation of non-protein substrates, preferably small molecular weight toxins, to multimeric proteins, preferrably antibodies, has not been described before, despite the fact that sortase enzyme mediated protein or peptide ligation has been in the prior art for many years.

Moreover, the conjugation of multimeric proteins, particularly full-length monoclonal antibodies with two different payloads, preferrably two different small molecular weight toxins as disclosed herein, has not been described in the prior art before, despite the fact that sortase enzyme mediated protein or peptide ligation has been in the prior art for many years (Panowski et al. (2014)).

It is known from the prior art that sortase enzymes may accept substrates that contain a minimum of 3 glycine amino acids (Parthasarathy et al. (2007), therefore the invention may include payloads that contain at least three (3) glycine amino acid residues added to the payload molecule of interest, although even one or two glycine residues may be sufficient, and should be comprised by the method disclosed herein. In case of small molecular weight payloads the addition of few glycine amino acid residues can be achieved by conventional

synthetic peptide chemistry, as described herein. In case of proteins glycine residues can be added either by adding codons for a number of glycine residues, preferably at least three glycine residues, in-frame to the open reading frame of the protein, or by conventional synthetic peptide chemistry such that the recombinant protein contains at least three N-terminal glycine amino acid residues.

It is known from the literature that different Sortase enzymes, e.g. Sortase B from *Staphylococcus aureus*, or Sortases from other gram-positive bacteria recognize different pentapeptide motifs, which differ from the LPXTG sortase A recognition motif (X = any amino acid) from *Staphylococcus aureus* (Spirig et al. (2011)). Therefore, the invention shall also include the concept of adding other sortase recognition motifs to proteins and immunoligands, including preferably antibodies, that differ from the *Staphylococcus aureus* sortase A recognition motif LPXTG, in order to prepare them for sortase conjugation with different cognate sortase enzyme of different gram-positive bacterial species. Therefore, proteins and immunoligands, preferably antibodies, can also be expressed with a different sortase recognition motif, e.g. a NPQTN pentapeptide motif specific for sortase B from *Staphylococcus aureus* which can then be conjugated to glycine modified payloads..

In an another aspect of the invention, multimeric immunoligands, preferably but not limited to antibodies, which are composed of immunoglobulin heavy and light chains, allow the utilization of said different sortase recognition sequences added to the different polypeptides of such multimeric proteins (in case of antibodies adding different sortase recognition sequences to the antibody heavy and light chains), in order to allow conjugation of different payloads to said different polypeptides by performing sequential conjugations with $Gly_n$-tagged payloads (n >2) in the presence of the respective sortase enzyme (Fig. 6b). For this, an antibody needs to be expressed with different C-terminal modifications at heavy and light chains comprising different sortase recognition motifs for different sortase enzymes. Such an antibody can then sequentially be conjugated to two different payloads containing a glycine modification as described further above.

This format may have the advantage that ADCs specifically be loaded with two different toxins, preferably interfering with a different cellular pathway will be more potent in cancer cell killing, because it is more difficult for a targeted cancer cell to evade the attack of two toxins comprised in the ADCs.

It is clear to a person skilled in the art, that a sortase pentapeptide recognition motif, like the *Staphylococcus aureus* sortase A LPXTG motif, can be added selectively to individual polypeptides of multimeric immunoligands, in order to provide desired conjugation sites. For instance, in the case of antibodies, this allows the generation of modified antibodies, either only containing sortase recognition motifs added to the heavy chains (resulting in two payloads per antibody conjugation), or only containing sortase recognition motifs added to the light chains (resulting in two payloads per antibody conjugation), or containing sortase recognition motifs added to the heavy and the light chains (resulting in four payloads per antibody conjugation). These designed variations will allow specific conjugation of payloads to antibodies by sortase enzymes either to the heavy chains alone (generating ADCs with drug to antibody ratio of 2, i.e. DAR2), or to the light chains alone (generating ADCs with drug to antibody ratio of 2, i.e. DAR2), or simultaneously to the heavy and the light chains (generating ADCs with drug to antibody ratio of 4, i.e. DAR4). This way, the conjugation sites and stoichiometries for antibodies can be varied in a controlled fashion, either generating two payload conjugations per antibody heavy or light chain, or generating four payload conjugations per antibody by addition of the payload to the heavy and the light chains.

Similar to the above-described variations in conjugation sites and stoichiometries using different sortase recognition motifs and sortase enzymes in multimeric proteins or immunoligands, it is a further aspect of the invention to conjugate different payloads to different polpeptide chains of multimeric proteins combining sortase-mediated and split-intein mediated conjugation. This concept allows the simultaneous conjugation of different payloads to different polypeptide chains of multimeric proteins and immunoconjugates in one step, because different transpeptidases and substrates are being employed (Fig. 6a).

It is to be understood that the above-mentioned conjugation of two different payloads to a multimeric protein, preferably an antibody, which is composed of each two disulphide linked heavy and lights chains, can either be accomplished by combining sortase enzyme mediated conjugation with split intein mediated conjugation, as depicted, in Fig. 6a, but that it is also possible to conjugate two different payloads to a multimeric protein, preferably an antibody, by utilizing two different sortase enzymes, recognizing different sortase peptide motifs, for instance sortase A and sortase B from *Staphylococcus aureus*, as mentioned

further above (Fig. 6b). However, this may also include sortase enzymes of other sortase classes (e.g. sortases C, D, E, F), or sortase enzymes from other bacterial species, differing in their sortase motif specificity.

Sortase-mediated conjugation of payloads to proteins and immunoligands can be achieved either by providing sortase recognition motif tagged proteins and at least tri-glycine tagged payloads and adding enzymatically active sortase enzyme or a functional fragment thereof as a soluble enzmye. In another aspect of the invention the enzymatically active domain of sortase enzyme can also be provided as a domain fused to either the N- or C-terminus of the protein. In this variation, is is advantageous, but not mandatory, to add the sortase enzymatic domain either N-terminal to an N-terminal sortase recognition motif, or C-terminal to a C-terminal sortase recognition motif. Both possibilities ensure that the after the reaction with a glycine-tagged payload, that the enzymatic sortase domain is removed from the protein in the course of the reaction (Fig. 5).

This variation of applying sortase-mediated conjugation of payloads to proteins is similar in concept to split-intein mediated conjugation of payloads, where the enzymatically active N-intein domains of split inteins are tethered to the protein to be conjugated, in order to define the conjugation site in the protein.

Similar to the large number of different sortase transpeptidases with different substrate specificity that have been identified in the literature (Spirig et al. (2011)), there is also a large and growing number of split-inteins known from different species and proteins with different N-intein and C-intein sequences required for transpeptidation that can be retrieved from the so-called InBase database (Perler (2002). Therefore, while the examples of split-intein mediated conjugation of immunoligands with payloads disclose the prefered Ssp GyrB S11 split intein (Volkmann et al. (2009)), because the C-intein domain can be reduced to a short, linear 6-mer amino acid stretch, split-intein mediated conjugation of payloads to proteins and immunoligands can also be achieved with other split inteins from the InBase database, as long as the N-intein or C-intein domains are short enough (preferably shorter than 13 amino acids) to easily allow peptide synthesis and addition to any payload molecule of choice. However, it is clear to a person skilled in the art that in the case of protein payloads, C-intein domains of any size may be fused to the protein payload by genetic fusion to the ORF of the protein payload of interest, and there is no mechanistic advantage of using split-inteins with small (< 13 amino acids) N-intein or C-intein domains.

However, if synthetic small-molecule payloads are to be conjugated to proteins and immunoligands, then a small N-int or C-int domain of less than 13 amino acids as disclosed herin are advantageous, as in the case of the prefered C-int of the Ssp GyrB S11 split intein, of the N-int of Ssp DnaX, because such a short peptide can synthetically be added to any synthetic small molecule weight payload by standard synthetic chemistry.

Sortase-mediated and split-intein mediated conjugation of payloads can be performed at either the N- or the C-termini of proteins and immunoligands. This is only dependent on how the sortase-motif/glycine stretch and N-intein/C-intein domains are positioned at protein and payload (Fig. 1).

In the case of antibodies, which are the preferred immunoligands, it is preferred to conjugate the payloads to the C-termini of the antibodies, because this positions the payloads most distally to the antigen-binding sites of the antibody. However, this preference shall not be interpreted by way of limitation, and it may be advantageous to conjugate payloads to the N-terminus of other immunoligand molecules, like e.g. antibody mimetics, in which the functional binding domains are not located at the N-terminus of the molecule.

Another aspect of the invention is to improve the efficiency of sortase and split-intein conjugation of payloads to proteins and immunoligands by adding affinity purification or detection tags, like e.g., but not limited to small peptide tags (e.g. histidine tags, strep-tag, MYC-tag or HA-tag) or larger protein affinity purification tags (e.g. maltose-binding protein (MBP) tag, Glutathione-S-transerase (GST) tag, or Chitin-binding tag) distal to the sortase recognition motif or the split-intein domain fused to the protein or immunoligand of interest. With this aspect of the invention the affinity purification tag will be removed from the immunoligand to be conjugated as part of the transpeptidation reaction. This can be exploited to enrich fully payload conjugated immunoligands, as unreacted proteins and immunoligands, that still contain the affinity purification tag, can be removed by binding to a suitable affinity resin, while completely payload conjugated proteins and immunoligands will no longer contain the affinity purification tag, and can thus be specifically separated from the unreacted immunoligand substrates. This aspect of the invention is particularly powerful in the context of multimeric proteins and immunoligands, like the preferred antibodies, in which several payloads need to be conjugated. The use of affinity purification

tags located distal to the sortase or intein transpeptidase conjugation site ensures that one can remove proteins and immunoligands in which the affinity purification tag is still present due to incomplete payload conjugation (Fig. 5).

In comparison to chemical conjugation, this provides a significant advantage in the process to obtain homogeneous immunoligand/payload conjugates, and preferably ADCs in which small molecular weight toxins are site specifically conjugated to the C-termini of antibody heavy and/or light chains.

Generally, the disclosed method provides a novel and efficient method to site-specifically and stoichiometrically conjugate payloads, preferably small molecular weight toxins to immunoligands, preferably antibodies, by which defined immunoligand/payload conjugates, preferably ADCs are generated, that are useful for the therapy of diseases, preferably of cancer. The method may also be utilized for the generation of immunoligand/payload conjugates useful for the diagnosis of diseases, preferably oncology diseases. The novel method allows generation covalent immunoligand/payload conjugates by utilization of peptide-bond breaking and forming enzymes (transpeptidases), including sortase enzymes and split-inteins, or catalytically active fragments thereof. Said enzymes can catalyze the covalent and site-specific conjugation of payloads containing short amino acid stretches (preferably shorter than 13 amino acids) either to the N- or C-termini of immunoligands which are suitably modified allowing sortase and split-inteins to break and to form peptide bonds in the course of the reaction. Immunoligands are preferably antibodies, for the site-specific conjugation of small molecular weight toxins, in order to generate antibody drug conjugates (ADCs) with defined antibody payload, or drug to antibody ratios.

**Embodiments of the invention**

According to the invention, a method of producing an immunoligand/payload conjugate is disclosed, which method encompasses conjugating a payload to an immunoligand by means of a sequence-specific transpeptidase, or a catalytic domain thereof.

According to a preferred embodiment of the invention, the payload and/or the immunoligand either

a) consists, entirely, of a protein or peptide
b) comprises at least one protein or peptide domain, or
c) comprises at least one peptide chain

and, further, the protein or peptide or domain comprises, preferably, an amino acid sequence that can be detected by the sequence-specific transpeptidase, or a catalytic domain thereof.

This means, for example, that, in case the payload and/or the immunoligand is a protein, it means that said protein comprises, at its N- or C-terminus, an amino acid sequence which can be detected by the sequence-specific transpeptidase. If such amino acid sequence is lacking to the naïve protein, it can be fused to the N- or C-terminus of said protein by recombinant methods known in the art.

In case the payload and/or the immunoligand is not a protein, such amino acid sequence which can be detected by the sequence-specific transpeptidase, is to be conjugated to the former by conventional chemical crosslinking methods known in the art.

Additional functionalities may be incorprated between the recognition sequence for a specific transpeptidase and the payload. This can be realized by chemical structures either being categorized by being cleavable (e.g. containing hydrazone, or disulfide chemistry, or specific peptide sequences for intracellular proteases) or being non-cleavable (e.g. containing thioether chemistry) following internalization into cells.

Chemical structures containing hydrazone chemistry can selectively be cleaved within the intracellular compartment of lysosomes (lower pH compared to the systemic blood circulation).

Peptide linkers have the potential to be selectively cleaved by lysosomal proteases (e.g. cathepsin-B) and have demonstrated increased serum stability and improved anti-tumor effects compared to hydrazone linkers. Valine-citruline (Val-Cit) pairs are the most commonly used peptide linkers and are ideally suited to work with the auristatin family of drugs such as monomethyl auristatin E (MMAE).

Non-cleavable Linkers have long been overlooked as researchers were convinced the cleaving of the linker was the most reasonable way to free the drug. However, conjugates

can, upon binding to a membrane receptor, get rapidly internalized and once internalized, the immunoligand can be degraded to the point where the payload, e.g., the drug is exposed. As one prominent example, thioether linkers, use the SMCC (*N*-succinimidyl-4-(*N*-maleimidomethyl)- cyclohexane-1-carboxylate) linker (Fig. .

All of theses appaoraches have in common that there is no true site-specificty of the coupling reaction. Because linker-mediated, chemical conjugation is a stochastic process, linker-mediated chemical ligation of payloads leads to heterogeneous mixtures of conjugated proteins that may differ in their therapeutic efficacy and/or diagnostic potential. Obviously, mixtures of protein-payload conjugates also represent a significant challenge in the regulatory approval process for therapeutic conjuagtes, as batch-to-batch variation and/or variations in the active pharmaceutical ingredient (API) are negatively viewed by regulatory authorities due to potential safety concerns.

Non-cleavable Linkers have long been overlooked as researchers were convinced the cleaving of the linker was the most reasonable way to free the drug. However, conjugates can, upon binding to a membarne receptor, get rapidly internalized and once internalized, the immunoligand can be degraded to the point where the payload, e.g., the drug is exposed. One prominent example, thioether linkers, use the SMCC (*N*-succinimidyl-4-(*N*-maleimidomethyl)- cyclohexane-1-carboxylate) linker (See Fig. 14a, structure 2).

All of theses approaches have in common that there is no true site-specificty of the coupling reaction. Because linker-mediated, chemical conjugation is a stochastic process, linker-mediated chemical ligation of payloads leads to heterogeneous mixtures of conjugated proteins that may differ in their therapeutic efficacy and/or diagnostic potential. Obviously, mixtures of protein-payload conjugates also represent a significant challenge in the regulatory approval process for therapeutic conjuagtes, as batch-to-batch variation and/or variations in the active pharmaceutical ingredient (API) are negatively viewed by regulatory authorities due to potential safety concerns.

According to another preferred embodiment of the invention, the immunoligand comprised in the immunoligand/payload conjugate is at least one selected from the group consisting of

- an antibody, modified antibody format, antibody derivative or fragment, and/or

- an antibody mimetic

Preferably, in this embodiment, a small molecular payload is rendered as substrate for the sequence-specific transpeptidase by coupling of a peptide of less than 13 amino acids to the small molecular payload, such that it can be conjugated by a transpeptidase to the C-termini of a monoclonal antibody containing C-terminal modifications recognized by said transpeptidases. Such C-terminal modifications may be contained on either both heavy chains, or both light chains, or of heavy and light chains of a full-length antibody, thereby allowing generation of a site-specifically conjugated ADC with either drug-to-antibody ratio of 2 or 4 (DAR2 or DAR4).

According to another preferred embodiment of the invention, the immunoligand binds at least one entity selected from the group consisting of

- a receptor
- an antigen,
- a growth factor
- a cytokine, and/or
- a hormone

As used herein, the term "receptor" means a cell surface molecule, preferably a cell surface molecule that (i) binds specific, or groups of specific, signalling molecules (i.e. a receptor, like, e.g., the VEGF receptor), and/or (ii) has no known ligand (i.e. an orphan receptor, like, e.g. HER2/neu). The natural receptors are expressed on the surface of a population of cells, or they merely represent the extracellular domain of such a molecule (whether such a form exists naturally or not), or a soluble molecule performing natural binding function in the plasma, or within a cell or organ. Preferably, such receptor is a member of a signalling cascade that is involved in a particular pathogenic process (e.g., a receptor that belongs to a signalling cascade of a growth factor), or is expressed on the surface of a cell or particle that is involved in a pathological process, e.g., a cancer cell.

As used herein, the term "antigen" means a substance that has the ability to induce a specific immune response, and may include surface proteins or protein complexes (e.g. ion channels). Often times, antigens are associated to pathogenic entities, e.g., a cancer cell.

As used herein, the term "cytokine" refers to small cell-signaling protein molecules that are secreted by numerous cells and are a category of signaling molecules used extensively in intercellular communication. Cytokines can be classified as proteins, peptides, or glycoproteins; the term "cytokine" encompasses a large and diverse family of regulators produced throughout the body by cells of diverse embryological origin.

As used herein, the term "growth factor" relates to naturally occurring substances capable of stimulating cellular growth, proliferation and cellular differentiation. Usually a growth factor is a protein or a steroid hormone. Growth factors are important for regulating a variety of cellular processes.

As used herein, the term "hormone" relates to a chemical released by a cell, a gland, or an organ in one part of the body that sends out messages that affect cells in other parts of the organism. The term encompasses peptide hormones, lipid and phospholipid-derived hormones including steroid hormones, and monoamines.

In case the immunoligand binds a receptor or an antigen, the immunoligand-payload conjugate can for example be directed to a specific site, e.g., to a pathogenic entity, e.g., a cancer cell, where the payload, e.g. a toxin or a chemotherapeutic agent, is delivered. Thus, the systemic toxicity of the toxin or the chemotherapeutic agent is reduced, while the local concentration of the latter at the site of action is increased, thus provding a better efficacy while side effects are reduced. Furthermore, a respective signalling cascade can be inhibited by the binding of the immunoligand. In case the payload is a marker the latter can thus be used to mark a specific site, e.g., a cancer cell charcterized by a given surface antigen detected by the immunoligand, for diagnosis.

In case the immunoligand binds a growth factor, a cytokine, and/or a hormone, the immunologand/payload conjugate can for example be directed to the site the growth factor cytokine or hormone usuall binds to, in oder to deliver the payload in a site-specific manner. Further, a respective signalling cascade can be inhibited by the binding of the immunoligand.

As used herein, the term "to bind" means the well-understood interaction or other nonrandom association between immunoligands, e.g., antibodies, or antibody fragments, and their targets. Preferably, such binding reaction is characterized by high specifity and/or sensitivity to the target. Preferably, the binding reaction is characterized by a dissociation constant (Kd) $\leq 10^{-3}$ M, preferably $\leq 10^{-4}$ M, $\leq 10^{-5}$ M, $\leq 10^{-6}$ M, $\leq 10^{-7}$ M, $\leq 10^{-8}$ M, $\leq 10^{-9}$ M, and most preferred $\leq 10^{-10}$.

Acccording to a preferred embodiment of the invention, it is provided that at least one catalytic domain of the sequence-specific transpeptidase is fused to the N-terminus or the C-terminus of either the immunoligand or the payload.

Such fusion may take place by recombinant engineering, or by chemical coupling. In this embodiment, the enzymatic activity leading to the site-specific conmjugation of the immunoligand to the payload does not need to be added to the reaction as a separate recombinant enzyme, but is rather part of protein substrate to be conjugated.

Preferably, the sequence-specific transpeptidase is at least one selected from the group consisting of

- a sortase, or one or more fragments or derivatives thereof
- a spilt-intein, or one or more fragments or derivatives thereof.

In a preferred embodiment, where the transpeptidase is a sortase, the payload, e.g., a toxin, is preferably rendered as substrate for sortase conjugation by addition of a small number of glycine amino acid residues, preferably 3 or 5 glycine residues.

In another preferred embodiment, where the transpeptidase is a split intein, e.g., a Ssp GyrB split intein, the payload, e.g., a toxin is rendered as substrate for split intein conjugation by addition of less than 13 amino acid residues of the sequence GVFVHN-$SX_n$, X being any amino acid and n being an integer between $\geq 0$ and $\leq 5$.

The use of transpeptidases, preferably sortase enzymes and split inteins for the generation of antibody drug conjugates, in which small molecular weight toxins are conjugated to full-length antibodies, has not yet been described in the prior art (Panowski et al. (2014)).

Sortase enzymes have been identified in a variety of gram-positive bacteria, like *Staphylococcus, Streptococcus* and *Pneumococcus* species, and catalyze, *in vivo*, the coupling of virulence factors to cell wall proteoglycans, in order to change the surface signature of the bacteria for evading an efficient immune response by the infected host (Mazmanian et al. (1999)).

The sortase A enzyme of the gram-positive bacterium *Staphylococcus aureus* has been characterized first (Ton-That et al. (1999)) and has subsequently been characterized further as a tool for many protein modifications (Tsukiji (2009)).

One beneficial feature of sortase enzymes is that the two molecules to be conjugated only require short peptide tags ("sortase tags"), which in case of *Staphylococcus aureus* sortase A is for example LPXTG at the C-terminus of one molecule (e.g., the payload), and a short 3 to 5 amino acid glycine stretch at the N-terminus of the other molecule (e.g., the immunoligand, see Fig. 1). These peptide tags can either be fused to the molecules, or conjugated thereto by means of conventional crosslinking chemistry. This allows to utilize the system on one hand for the ligation of two proteins, but also for the conjugation of small molecular weight compounds, preferably small molecular weight toxins to proteins. In case of *Staphylococcus aureus* sortase B, the respective sortase motif is NPQTN.

Inteins, which have originally been discovered as protein introns that can remove (splice) themselves out of precursor proteins by cleavage of peptide bonds and new peptide-bond formation (Xu et al. (1993)) (Fig. 2a).

Naturally occuring and artificial split-inteins involve that the intein coding region has been split into N-intein and C-intein domains, which can be attached to different proteins or peptides in such way that, subsequently the *trans*-splicing of the extein domains (Fig. 2b) leads to the conjugation of the two proteins

Split-inteins have thus been utilized for the covalent coupling of N-extein and C-extein moieties, and also for the purification and/or circularization of proteins (Elleuche (2010)). One embodiment disclosed herein is to utilize split-inteins for the conjugation of small molecular weight compounds, preferably small molecular weight toxins and other small molecule labels, in which a short C-extein peptide sequence of smaller than 13 amino acids

is coupled to molecules of any size, similar to the short glycine amino acid stretch required for sortase-mediated transpeptidation.

In case of sortase enzymes addition of a short glycine stretch (> 2 glycine residues) to a molecule of choice is sufficient to allow the molecule to be conjugated to immunoligands containing a penta-peptide sortase recognition motif, like e.g. LPXTG in case of sortase A of *S. aureus*. In case of split-inteins, minimally a short 12 amino-acid GVFVHNSAGSGK amino acid stretch containing a short, 6 amino acid C-intein (GVFVHN) from Ssp GyrB and a short C-extein (here: SAGSGK) are sufficient to modify any payload molecule, preferably a small molecular weight toxin, for split-intein mediated conjugation to immunoligands containing the N-intein domain of the Ssp GyrB split intein (Volkmann et al. (2009)). Other split inteins, in which functional intein domains can be reduced to small <13 amino acid long peptide stretches may be utilized as well.

Even if, in the literature, split-enzymes are not always referred to as enzymes, they qualify as such, because the reaction they catalyze results in the breakage of a peptide bond and the formation of a new peptide bond and this can be viewed as transpeptidases, because the energy of an existing peptide bond is transfered to a new peptide bond.

Other than chemical conjugation, the transpeptidase-mediated conjugation occurs under physiologic aqueous buffer conditions and physiologic temperatures, thereby minimally affecting the protein or antibody integrity in the conjugation reaction. This feature ensures optimal functionality of the resulting conjugate

Acccording to another preferred embodiment of the invention, it is provided that the payload comprised in the immunoligand/ payload conjugate is at least one selected from the group consisting of

- a marker
- a processing tag, and/or
- a drug.

The term "marker" (also called "detection tag"), as used herein, may refer to any molecule or moiety that comprises one or more appropriate chemical substances or enzymes, which

directly or indirectly generate a detectable compound or signal in a chemical, physical or enzymatic reaction.

The term "processing tag" as used herein, may encompass affinity tags, solubilization tags, chromatography tags and epitope tags. Affinity tags (also used as purification tags) are appended to proteins so that they allow purifification of the tagged molecule from their crude biological source using an affinity technique. These include chitin-binding protein (CBP), maltose binding protein (MBP), and glutathione-S-transferase (GST). The poly(His) tag, preferably a 6xHis tag, is a widely-used processing tag; it binds to metal matrices. Solubilization tags are used, especially for recombinant proteins expressed in chaperone-deficient species such as E. coli, to assist in the proper folding in proteins and keep them from precipitating. These include thioredoxin (TRX) and poly(NANP). Some affinity tags have a dual role as a solubilization agent, such as MBP, and GST.

Chromatography tags are used to alter chromatographic properties of the protein to afford different resolution across a particular separation technique. Often, these consist of polyanionic amino acids, such as FLAG-tag.

Epitope tags are short peptide sequences which are chosen because high-affinity antibodies can be reliably produced in many different species. Epitope tags are usually derived from viral genes, which explain their high immunoreactivity. Epitope tags include e.g. the V5-tag, MYC-tag, and HA-tag. These tags are particularly useful for western blotting, immunofluorescence and immunoprecipitation experiments, although they also find use in protein purification.

Processing tags find many other usages, such as specific enzymatic modification (such as biotin ligase tags) and chemical modification (FlAsH) tag. Often tags are combined to produce multifunctional modifications of the protein.
Preferably, said marker is at least one selected from the group consisting of

- a radiolabel, preferably a radioactively labelled peptide or protein
- a fluorescent label, preferably a fluorescent peptide or protein, and/or
- an enzyme label, preferably a peroxidase.

This enumeration of potential marker payloads is by no means restrictive. According to another preferred embodiment, said drug is at least one selected from the group consisting of

- a cytokine
- a radioactive agent
- an anti-inflammatory drug
- a toxin, and/or
- a chemotherapeutic agent

This enumeration of potential drug payloads is by no means restrictive. As used herein, the term "cytokine" refers to small cell-signaling protein molecules that are secreted by numerous cells and are a category of signaling molecules used extensively in intercellular communication. Cytokines can be classified as proteins, peptides, or glycoproteins; the term "cytokine" encompasses a large and diverse family of regulators produced throughout the body by cells of diverse embryological origin. In the present context, cytokines are for example meant to impair, or even kill, pathogenic entity, e.g., a cancer cell.

As used herein, the term "radioactive agent" relates to an entity which has at least one atom with an unstable nucleus, and which is thus prone to undergo radioactive decay, resulting in the emission of gamma rays and/or subatomic particles such as alpha or beta particles, which have a cell killing effect. In the present context, radioactive agents are meant to impair, or even kill, pathogenic entity, e.g., a cancer cell.

As used herein, the term "anti-inflammatory drug" relates to compounds that reduce inflammation. This can be, e.g., steroids, just like specific glucocorticoids (often referred to as corticosteroids), which reduce inflammation or swelling by binding to glucocorticoid receptors. The term further encompasses non-steroidal anti-inflammatory drugs (NSAIDs), which counteract the cyclooxygenase (COX) enzyme. On its own, COX enzyme synthesizes prostaglandins, creating inflammation. In whole, the NSAIDs prevent the prostaglandins from ever being synthesized, reducing or eliminating the pain. The term further encompasses Immune Selective Anti-Inflammatory Derivatives (ImSAIDs), which are a class of peptides that alter the activation and migration of inflammatory cells, which are immune cells responsible for amplifying the inflammatory response.

As used herein, the term "toxin" relates to a molecule which is toxic to a living cell or organism. Toxins may be peptides, or proteins or preferably small molecular weight compounds, that are meant to impair, or even kill, pathogenic entity, e.g., a cancer cell. Toxins, as meant herein, encompass, in particular, cellular toxins. Preferably, said toxin is a small molecular toxin, i.e., having a molecular weight of $\leq$ 2500 Da.

As used herein, the term "chemotherapeutic agent" relates to molecules that have the functional property of inhibiting a development or progression of a neoplasm, particularly a malignant (cancerous) lesion, such as a carcinoma, sarcoma, lymphoma, or leukemia. Inhibition of metastasis or angiogenesis is frequently a property of anti-cancer or chemotherapeutic agents. A chemotherapeutic agent may be a cytotoxic or chemotherapeutic agent. Preferably, said chemotherapeuic agent is a small molecular weight cytostatic agent, which inhibits or suppresses growth and/or multiplication of cancer cells.

Conjugating cytokines, radioactive agents, toxins or chemotherapeutic agents to an immunologand can help to reduce side effects and risks related to their administration, because

a) the immunoligand directs the conjugate to a specific site, e.g., to a pathogenic entity, e.g., a cancer cell where the payload effects its toxic function. Thus, the systemic toxicity of the payload is reduced, while the local concentration of the latter at the site of action is increased, thus provding a better efficacy while side effects are reduced.

b) it can be provided that the conjugate is internalized by the pathogenic entity, in such way that after internalization, the payload is released and only then develops its desired cytotoxic function, i.e., without affecting the surrounding cells or tissue.

The following table is a non restrictive list of potential targets/antigens (1$^{st}$ column) and examples for existing immunoligands targeting the former (2$^{nd}$ column). The 3$^{rd}$ column shows a non restrictive list of potential toxins, cytokines or chemotherapeutic agents. Note that the examples from the 1$^{st}$ and the 3$^{rd}$ column can be combined with one another *ad libitum*, while hundereds of further targets and payloads exist. Respective target/payload combinations not explicitly mentioned in the table are encompassed by the scope of the present invention.

| target/antigen | example of an existing immunoligand | payload |
|---|---|---|
| Endothelial Growth factor receptor (EGFR) | Cetuximab | Maytansinoides, e.g. Mertansine, Ansamitocin' Ravtansin, DM4, DM1 |
| CD20 | Rituximab, Ibritumomab, Tositumomab (mAb) | Calicheamicins, e.g. Ozogamicin |
| CD44 | | Doxorubicin |
| MUC1 | Cantuzumab (mAb) | bacterial Pseudomonas exotoxin PE38 |
| CD30 | Brentuximab (mAb) | Monomethyl Auristatin F (MMAF); Monomethyl Auristatin E (MMAE) |
| CD22 | inotuzumab (mAb) | Pyrrolobenzodiazepine (PBD) |
| transmembrane glycoprotein NMB (GPNMB) | Glembatumumab (mAb) | Interleukin-10 (IL10) (anti-inflammatory) |
| CD56 | Lorvotuzumab (mAb) | Diphtheria toxin |
| CanAg | huC242 (mAb) | Tumor necroris factor (TNF) |
| luteinizing hormone releasing hormone (LHRH) receptor | [D-Lys(6)] LHRH | RNase |
| Prostate-specific membrane antigen (PSMA) | | $Yttrium^{90}$ |
| CD74 | Milatuzumab (mAb) | $Iodine^{131}$ |
| CD70 | | $Lutetium^{177}$ |
| AGS-16 | | Cyclosporine |
| Integrin | | Methotrexate |
| CD19 | | Taxanes, e.g., Paclitaxel or Docetaxel |
| Nectin-4 | | |
| Interleukin 2 receptor | Interleukin-2 (Proleukin) | |
| CD3 | UCHT1 (mAb) | |
| extra domain B of fibronectin | L19-SIP (scFv fused in with the constant domain *CH4* | |
| SLAMF7 (CD319) | Elotuzumab (mAb) | |
| SDC1 | Indatuximab (mAb) | |
| Her-2/neu | Trastuzumab (mAb) | |
| CD33 | Gemtuzumab (mAb) | |

According to yet another embodiment of the present invention, the immunoligand comprises at least two subunits each being conjugated to a payload.

Preferably, at least two different payloads can be conjugated to the at least two subunits. This option provides a versatile toolbox with which a large variety of different

immunoligand-payload constructs can be created. For example, a bispecific dual-domain immunoligand can be conjugated with two different payloads, for example one marker and one toxin.

Preferably the at least two different payloads are toxic payloads interfering with one or more cellular pathways.

Such embodiment can be accomplished, e.g., by conjugating the two different payloads to each the 2 light chains of a full-length antibody, and to the 2 heavy chains of a full length antibody, respectively, by utilizing two different sortase enzymes, recognizing different sortase recognition motifs, plus an antibody that contains different C-terminal modifications at heavy and light chains comprising the respective recognition motifs for said different sortase enzymes.

In such way, an Antibody Drug Conjugate can be created which is composed of each two full-length Ig light chains and Ig heavy chains, containing different payloads covalently attached to said heavy and light chains.

Such embodiment results, preferably, in the synchronous conjugation of the at least two subunits for the generation of immunoligand payloads with equal payload conjugation to each of said subunits.

According to another preferred embodiment, said immunoligand with at least two subunits is being conjugated with at least 80% efficiency per conjugation site.

According to yet another preferred embodiment, said immunoligand with at least two subunits contains a peptide spacer sequence of at least two amino acids, preferably 2-5 amino acids, appended to the C-termini of at least one of the two subunits

This approach results, advantageously, in synchronous conjugation of the at least two subunits for the generation of immunoligand payloads with equal payload conjugation to each of said subunits.According to another embodiment of the present invention, the method allows a stoichiometrically defined relationship between immunoligand and payload. According to this embodiment, a strict quantitative relationship between immunoligand and

payload can be provided, thus improving the reproducibility and the overall performance of the respective immunoligand/payload conjugate particularly for clinical and/or therapeutic applications. This is accounted for by the sequence- and/or site specificity of the transpeptidase used.

According to a particularly prefered embodiment said stoichiometrically defined relationship between immunoligand and payload is achieved by removal of partially reacted C-terminally modified immunoligand substrate. Such removal can, for example, be carried out via affinity purification. Said approach results, preferably, in a homogeneous drug to immunoligand ratio.

Preferably, said removal is carried out by affinity purification using an affinity tag positioned C-terminal to the transpeptidase recognition motif or domain. Standard methods known to the skilled person can be used for this purpose, e.g., HIS tag, CBP tag, CYD (covalent yet dissociable NorpD peptide) tag, Strep II tag, FLAG tag, HPC (heavy chain of protein C) tag, and the GST and MBP protein fusion tags.

According to another embodiment of the present invention, the method allows a site-specific conjugation of a payload to the immunoligand. According to this embodiment, it is ensured that the conjugation process does not interfere with the activity of the immunoligand, or the payload, itself, thus improving the reproducibility and the overall performance of the respective immunoligand/payload conjugate particularly for clinical and/or therapeutic applications. This is accounted for by the sequence- and/or site specificity of the transpeptidase used. Other than with conventional binding chemistry, which is not site specific in most cases, or has limited site specificity (e.g, when the payload is conjugated to a free amino group, like in Arg, Lys, Asn or Gln), the binding site can thus be exactly determined, so that the characterizing features of the immunoligand (e.g., target specificity) or the payload (e.g., toxicity) are not affected.

The invention further provides an immunoligand/payload conjugate obtained with a method according to the above-mentioned embodiments.

Preferably, said immunoligand/payload conjugate is selected from the group consisting of an antibody/drug conjugate, and/or an antibody/marker conjugate.

The invention further provides the use of an immunoligand/payload conjugate according to the above mentioned embodiments for

- *in vitro* or *in vivo* diagnosis of a given pathologic condition
- *in vitro* or *in vivo* prediction or prognosis with respect to a given pathologic condition
- the treatment of a human or animal subject suffering from or being at risk of developing a given pathologic condition, and/or
- research and/or development purposes

Preferably, said pathologic condition is at least one selected from the group consisting of

- Neoplastic disease
- Autoimmune disease
- Neurodegenerative disease, and/or
- Infectious disease

In all these cases, the immunoligand/payload conjugate according to the invention can have beneficial effects, e.g, by directing the latter to a specific site, e.g., a cancer cell, a site of neuropathology, or a site of an autoimmune reaction.

The payload, e.g., a toxin, a chemotherapeutic agent, a cytokine or a drug is delivered at said site, e.g., to deplete a cancer cell, to act anti-proliferatively on a cancer cell, to dissolve a plaque, to inhibit autoantibodies, and the like.

In all these cases, the immunoligand/payload conjugate according to the invention can have beneficial effects, e.g, by directing the latter to a specific site, e.g., a cancer cell, where the payload, e.g. a toxin or a chemotherapeutic agent, is delivered, e.g., to deplete a cancer cell, to act anti-proliferatively on a cancer cell.

Thus, the systemic toxicity of the toxin or the chemotherapeutic agent is reduced, while the local of the latter at the site of action is increased, thus provding a better efficacy while side effects are reduced. Further, a respective signalling cascade can be inhibited by the binding

of the immunoligand. In case the payload is a marker the latter can thus be used to mark a specific site, e.g., a cancer cell charcterized by a given surface antigen detected by the immunoligand, for diagnosis.

The site-specifity of the conjugating process ensures a high reproducibility and overall performance of the respective immunoligand/payload conjugate particularly for clinical and/or therapeutic applications.

The term "neoplastic disease", as used herein, refers to an abnormal state or condition of cells or tissue characterized by rapidly proliferating cell growth or neoplasm. In a more specific meaning, the term relates to cancerous processes, e.g., tumors and/or leukemias.

The term "neuropathological diseases" encompasses, among others, neurodegenerative diseases, neuroinflammatory diseases or seizure disorders.

Neurodegenerative diseases are characterized by progressive loss of structure or function of neurons, including death of neurons. Many neurodegenerative diseases including Parkinson's, Alzheimer's, Huntington's, Amyotrophic lateral sclerosis and Multiple Sclerosis occur as a result of neurodegenerative processes. There are many parallels between different neurodegenerative disorders including atypical protein assemblies as well as induced cell death. Neurodegeneration can further be found in many different levels of neuronal circuitry ranging from molecular to systemic.

The terms "Neurodegenerative diseases" and "Neuroinflammatory diseases" have a partially overlapping scope. Inflammatory responses are a hallmark of neurodegenerative disease and participate, or contribute, through different mechanisms in the neuronal cell death. The tryptophan catabolism along the Kynurenine pathway (KP) represents one of these mechanisms.

Seizure disorders are brain disorders which are characterized by abnormal signaling between brain cells. Seizure disorders can affect part of the brain (Partial seizures) or the entire brain (Generalized seizures). The most prominent Seizure disorder is epilepsy.

The term "Autoimmune disease", as used herein, encompasses organ-specific autoimmune diseases, in which an autoimmune response is directed against a single tissue, such as Crohn's disease and ulcerative colitis, Type I diabetes mellitus, myasthenia gravis, vitiligo, Graves' disease, Hashimoto's disease, Addison' s disease and autoimmune gastritis and autoimmune hepatitis. The term also encompasses non-organ specific autoimmune diseases, in which an autoimmune response is directed against a component present in several or many organs throughout the body.

Such autoimmune diseases include, for example, rheumatoid arthritis, disease, systemic lupus erythematosus, progressive systemic sclerosis and variants, polymyositis and dermatomyositis.

Additional autoimmune diseases include pernicious anemia including some of autoimmune gastritis, primary biliary cirrhosis, autoimmune thrombocytopenia, Sjögren's syndrome, multiple sclerosis and psoriasis. One skilled in the art understands that the methods of the invention can be applied to these or other autoimmune diseases, as desired.

The term "infectious disease" as used herein, includes, but is not limited to any disease that is caused by an infectious organism. Infectious organisms may comprise viruses, (e.g., single stranded RNA viruses, single stranded DNA viruses, human immunodeficiency virus (HIV), hepatitis A, B, and C virus, herpes simplex virus (HSV), cytomegalovirus (CMV) Epstein-Barr virus (EBV), human papilloma virus (HPV)), parasites (e.g., protozoan and metazoan pathogens such as Plasmodia species, Leishmania species, Schistosoma species, Trypanosoma species), bacteria (e.g., Mycobacteria, in particular, M. tuberculosis, Salmonella, Streptococci, E. coli, Staphylococci), fungi (e.g., Candida species, Aspergillus species), *Pneumocystis carinii*, and prions.

The invention further provides a low molecular-weight payload modified with a $Gly_n$-modification, wherein, n>1, preferably n=3 or n=5.

As used herein, the term "$Gly_n$-modification" means that an oligo- or polypepide consisting of n Glycin residues has been added to said payload. As used herein, the term "low molecular-weight payload compound" shall encompass payloads that have a molecular weight of 2500 Da or less.

Said payload is, preferably, at least one selected from the group consisting of

- a marker,
- a processing tag, and/or
- a drug.

Said marker is at least one selected from the group consisting of

- a radiolabel, preferably a radioactively labelled peptide or protein
- a fluorescent label, preferably a fluorescent peptide or protein, and/or
- an enzyme label, preferably a peroxidase.

Said drug is at least one selected from the group consisting of

- a cytokine
- a radioactive agent
- a toxin, and/or
- a chemotherapeutic agent

As disussed above already, said toxin is preferably a small molecular toxin, i.e., having a molecular weight of $\leq$ 2500 Da. Preferably, said toxin is at least one selected from the group constisting of

- Maytansine
- Monomethyl auristatin, and/or
- Alpha-amanitin

or derivatives of the former. Examples for such $Gly_n$-modified toxions are shown in structures 1 to 9 of Fig. 14A – 14C

The invention further provides the use of a glycine-modified low molecular-weight payload for conjugation thereof to an immunoligand.

Preferably, and as mentioned above, the conjugation is a transpeptidease-mediated conjugation, preferably with a sortase and/or a split intein. Likewise preferably, the immunologand is an antibody.

Preferably, said immunoligand is an antibody. In such way, an antibody drug conjugate (ADC) can be provided.

Preferably, the immunologand-payload conjugation reaction is performed in crude cell culture supernatant. This means that, preferably, the conjugation reaction may take place with unpurified or only partially purified components.

**Experiments and Figures**

While the invention has been illustrated and described in detail in the drawings and foregoing description, such illustration and description are to be considered illustrative or exemplary and not restrictive; the invention is not limited to the disclosed embodiments. Other variations to the disclosed embodiments can be understood and effected by those skilled in the art in practicing the claimed invention, from a study of the drawings, the disclosure, and the appended claims. In the claims, the word "comprising" does not exclude other elements or steps, and the indefinite article "a" or "an" does not exclude a plurality. The mere fact that certain measures are recited in mutually different dependent claims does not indicate that a combination of these measures cannot be used to advantage. Any reference signs in the claims should not be construed as limiting the scope.

All amino acid sequences disclosed herein are shown from N-terminus to C-terminus; all nucleic acid sequences disclosed herein are shown 5'->3'.

**Example 1: Cloning of expression vectors and expression of a CD19 monoclonal antibody with C-terminal LPETG sortase tag and additional 6x-His and strepII affinity purification tags**

In order to perform the C-terminal conjugation of a payload to an antibody, first a recombinant antibody needs to be expressed that contains C-terminal modifications, including a recognition motif, e.g. for sortase A of *Staphylococcus aureus*.

For this, first ORFs for heavy and light chains of an anti-human CD19 specific antibody can be gene synthesized, e.g. at contract research organizations (CROs) offering such gene synthesis services, like e.g. Genscript (www.genscript.com, Piscataway, NJ, USA). As an example, the heavy and light chain sequences of a humanized anti-human CD19 antibody hBU12 can be found in patent US 8,242,252 B2 under Seq 53 (variant HF) and Seq 58 (variant LG). The $V_H$ and $V_L$ regions of this anti-human CD19 antibody are as follows:

SEQ ID NO 1 ($V_H$ coding region of humanized anti-human CD19 antibody hBU12):

```
ATGGGATGGAGCTGGATCTTTCTTTTCCTCCTGTCAGGAACTGCAGGTGTCCATTGTCAGGTTCAGCTGCAAGA
GTCTGGCCCTGGGTTGGTTAAGCCCTCCCAGACCCTCAGTCTGACTTGTACTGTGTCTGGGGGTTCAATCAGCA
CTTCTGGTATGGGTGTAGGCTGGATTAGGCAGCACCCAGGGAAGGGTCTGGAGTGGATTGGACACATTTGGTGG
GATGATGACAAGAGATATAACCCAGCCCTGAAGAGCAGAGTGACAATCTCTGTGGATACCTCCAAGAACCAGTT
TAGCCTCAAGCTGTCCAGTGTGACAGCTGCAGATACTGCTGTCTACTACTGTGCTAGAATGGAACTTTGGTCCT
ACTATTTTGACTACTGGGGCCAAGGCACCCTTGTCACAGTCTCCTCA
```

This translates to the following amino acid sequence (SEQ ID NO 2):

```
MGWSWIFLFLLSGTAGVHCQVQLQESGPGLVKPSQTLSLTCTVSGGSISTSGMGVGWIRQHPGKGLEWIGHIWW
DDDKRYNPALKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCARMELWSYYFDYWGQGTLVTVSS
```

SEQ ID NO 3 ($V_L$ coding region of humanized anti-human CD19 antibody hBU12)

```
ATGAAGTTGCCTGTTAGGCTGTTGGTGCTGATGTTCTGGATTCCTGCTTCCAGCAGTGAAATTGTTCTCACCCA
GTCTCCAGCAACCCTGTCTCTCTCTCCAGGGGAAAGGGCTACCCTGAGCTGCAGTGCCAGCTCAAGTGTAAGTT
ACATGCACTGGTACCAGCAGAAGCCAGGGCAGGCTCCCAGACTCCTGATTTATGACACATCCAAACTGGCTTCT
GGTATTCCAGCAAGGTTCAGTGGCAGTGGGTCTGGAACAGATTTTACACTCACAATCAGCAGCCTGGAGCCAGA
GGATGTTGCTGTCTATTACTGTTTTCAGGGGAGTGTATACCCATTCACTTTTGGCCAAGGGACAAAGTTGGAAA
TCAAA
```

This translates to the following amino acid sequence (SEQ ID NO 4):

```
MKLPVRLLVLMFWIPASSSEIVLTQSPATLSLSPGERATLSCSASSSVSYMHWYQQKPGQAPRLLIYDTSKLAS
GIPARFSGSGSGTDFTLTISSLEPEDVAVYYCFQGSVYPFTFGQGTKLEIK
```

These sequences can be fused to human $IgG_1$ constant heavy and constant light chain regions containing additional C-terminal tags, in order to realize the method disclosed herein.

In order to realize the invention, the human constant IgG1 heavy chain region can be synthesized with additional 3'-codons, encoding an LPETG *Staphylococcus aureus* sortase A recognition tag, followed by a 6xHis tag (HHHHHH), a MYC-tag (EQKLISEEDL) and a strep II tag (WSHPQFEK) resulting in a sequence, which is as follows:
SEQ ID NO 5 (human IgG1 heavy chain constant coding region with in-frame 3' extension encoding an LPETG sortase tag, an 6xHis tag and a strepII tag):

```
AGCACCAAGGGCCCATCTGTCTTCCCCCTGGCACCCTCCTCCAAGAGCACCTCTGGGGGCACAGCTGCCCTGGG
CTGCCTGGTCAAGGACTACTTCCCTGAACCTGTGACAGTGTCCTGGAACTCAGGCGCCCTGACCAGCGGCGTGC
ACACCTTCCCGGCTGTCCTACAGTCCTCAGGACTCTACTCCCTCAGCAGCGTGGTGACCGTGCCCTCCAGCAGC
TTGGGCACCCAGACCTACATCTGCAACGTGAATCACAAGCCCAGCAACACCAAGGTGGACAAGAAAGTTGAGCC
CAAATCTTGTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTCC
TCTTCCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTG
AGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCC
GCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGAATG
GCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAA
GGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCT
GACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACA
ACTACAAGACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAG
AGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGCACAACCACTACACACAGAA
GAGCCTCTCCCTGTCTCCGGGTAAACTGCCCGAGACCGGCCACCACCACCACCACCACGGCGAGCAGAAGCTGA
TCAGCGAGGAGGACCTGGGCTGGAGCCACCCCCAGTTCGAGAAGTAG
```

This translates to the following amino acid sequence (SEQ ID NO 6, amino acids of the tags are underlined):

```
STKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSS
LGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDV
SHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAK
GQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDK
SRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGKLPETGHHHHHHGEQKLISEEDLGWSHPQFEK•
```

Furthermore, the human constant IgG1 kappa light chain region can be synthesized with additional 3'-codons, encoding an LPETG *Staphylococcus aureus* sortase A recognition tag, followed by a 6xHis tag and a strep II tag (WSHPQFEK) resulting in a sequence, which is as follows:

SEQ ID NO 7 (human IgG1 kappa light chain constant coding region with in-frame 3' extension encoding an LPETG sortase tag, an 6xHis tag, a Myc tag, and a strepII tag):

```
ACGGTGGCTGCACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACTGCCTCTGTTGT
GTGCCTGCTGAATAACTTCTATCCCAGAGAGGCCAAAGTACAGTGGAAGGTGGATAACGCCCTCCAATCGGGTA
ACTCCCAGGAGAGTGTCACAGAGCAGGACAGCAAGGACAGCACCTACAGCCTCAGCAGCACCCTGACGCTGAGC
AAAGCAGACTACGAGAAACACAAAGTCTACGCCTGCGAAGTCACCCATCAGGGCCTGAGCTCGCCCGTCACAAA
GAGCTTCAACAGGGGAGAGTGTCTGCCCGAGACCGGCCACCACCACCACCACCACGGCGAGCAGAAGCTGATCA
GCGAGGAGGACCTGGGCTGGAGCCACCCCCAGTTCGAGAAGTAG
```

This translates to the following amino acid sequence (SEQ ID NO 8, amino acids of the tags are underlined):

TVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLS
KADYEKHKVYACEVTHQGLSSPVTKSFNRGECLPETGHHHHHHGEQKLISEEDLGWSHPQFEK•

The complete coding regions for LPETG sortase tag, 6xHis and strepII tagged heavy and light chains of the humanized anti-human CD19 antibody hBU12 are then as follows:

SEQ ID NO 9 (Complete human IgG1 $V_H$-$C_H$ heavy chain coding region for hBU12 with C-terminal LPETG sortase tag, 6xHis tag, Myc tag, and a strepII tag):

ATGGGATGGAGCTGGATCTTTCTTTTCCTCCTGTCAGGAACTGCAGGTGTCCATTGTCAGGTTCAGCTGCAAGA
GTCTGGCCCTGGGTTGGTTAAGCCCTCCCAGACCCTCAGTCTGACTTGTACTGTGTCTGGGGGTTCAATCAGCA
CTTCTGGTATGGGTGTAGGCTGGATTAGGCAGCACCCAGGGAAGGGTCTGGAGTGGATTGGACACATTTGGTGG
GATGATGACAAGAGATATAACCCAGCCCTGAAGAGCAGAGTGACAATCTCTGTGGATACCTCCAAGAACCAGTT
TAGCCTCAAGCTGTCCAGTGTGACAGCTGCAGATACTGCTGTCTACTACTGTGCTAGAATGGAACTTTGGTCCT
ACTATTTTGACTACTGGGGCCAAGGCACCCTTGTCACAGTCTCCTCAGCTAGCACCAAGGGCCCATCTGTCTTC
CCCCTGGCACCCTCCTCCAAGAGCACCTCTGGGGGCACAGCTGCCCTGGGCTGCCTGGTCAAGGACTACTTCCC
TGAACCTGTGACAGTGTCCTGGAACTCAGGCGCCCTGACCAGCGGCGTGCACACCTTCCCGGCTGTCCTACAGT
CCTCAGGACTCTACTCCCTCAGCAGCGTGGTGACCGTGCCCTCCAGCAGCTTGGGCACCCAGACCTACATCTGC
AACGTGAATCACAAGCCCAGCAACACCAAGGTGGACAAGAAAGTTGAGCCCAAATCTTGTGACAAAACTCACAC
ATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGGACCGTCAGTCTTCCTCTTCCCCCCAAAACCCAAGGACA
CCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAG
TTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCAC
GTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCT
CCAACAAAGCCCTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTG
TACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTA
TCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGC
TGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTC
TTCTCATGCTCCGTGATGCATGAGGCTCTGCACAACCACTACACACAGAAGAGCCTCTCCCTGTCTCCGGGTAA
ACTGCCCGAGACCGGCCACCACCACCACCACCACGGCGAGCAGAAGCTGATCAGCGAGGAGGACCTGGGCTGGA
GCCACCCCCAGTTCGAGAAGTAG

This translates to the following amino acid sequence (SEQ ID NO 10, amino acids of the tags are underlined):

MGWSWIFLFLLSGTAGVHCQVQLQESGPGLVKPSQTLSLTCTVSGGSISTSGMGVGWIRQHPGKGLEWIGHIWW
DDDKRYNPALKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCARMELWSYYFDYWGQGTLVTVSSASTKGPSVF
PLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYIC
NVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVK
FNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQV
YTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNV
FSCSVMHEALHNHYTQKSLSLSPGKLPETGHHHHHHGEQKLISEEDLGWSHPQFEK•

SEQ ID NO 11 (Complete human IgG1 $V_L$-$C_L$ kappa chain coding region for hBU12 with C-terminal LPETG sortase tag, 6xHis tag, Myc tag, and a strepII tag):

ATGAAGTTGCCTGTTAGGCTGTTGGTGCTGATGTTCTGGATTCCTGCTTCCAGCAGTGAAATTGTTCTCACCCA
GTCTCCAGCAACCCTGTCTCTCTCTCCAGGGGAAAGGGCTACCCTGAGCTGCAGTGCCAGCTCAAGTGTAAGTT
ACATGCACTGGTACCAGCAGAAGCCAGGGCAGGCTCCCAGACTCCTGATTTATGACACATCCAAACTGGCTTCT
GGTATTCCAGCAAGGTTCAGTGGCAGTGGGTCTGGAACAGATTTTACACTCACAATCAGCAGCCTGGAGCCAGA
GGATGTTGCTGTCTATTACTGTTTTCAGGGGAGTGTATACCCATTCACTTTTGGCCAAGGGACAAAGTTGGAAA
TCAAAAGAACTGTGGCTGCACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACTGCC
TCTGTTGTGTGCCTGCTGAATAACTTCTATCCCAGAGAGGCCAAAGTACAGTGGAAGGTGGATAACGCCCTCCA
ATCGGGTAACTCCCAGGAGAGTGTCACAGAGCAGGACAGCAAGGACAGCACCTACAGCCTCAGCAGCACCCTGA
CGCTGAGCAAAGCAGACTACGAGAAACACAAAGTCTACGCCTGCGAAGTCACCCATCAGGGCCTGAGCTCGCCC
GTCACAAAGAGCTTCAACAGGGGAGAGTGTCTGCCCGAGACCGGCCACCACCACCACCACCACGGCGAGCAGAA
GCTGATCAGCGAGGAGGACCTGGGCTGGAGCCACCCCCAGTTCGAGAAGTAG

This translates to the following amino acid sequence (SEQ ID NO 12, amino acids of the tags are underlined):

MKLPVRLLVLMFWIPASSSEIVLTQSPATLSLSPGERATLSCSASSSVSYMHWYQQKPGQAPRLLIYDTSKLAS
GIPARFSGSGSGTDFTLTISSLEPEDVAVYYCFQGSVYPFTFGQGTKLEIKRTVAAPSVFIFPPSDEQLKSGTA
SVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSP
VTKSFNRGECLPETGHHHHHHGEQKLISEEDLGWSHPQFEK•

The coding regions for the heavy and light chains of the anti-human CD19 specific antibody as disclosed in SEQ ID NOs 9 and 11, respectively, can then be synthesized with flanking restriction enzyme sites (e.g. HindIII and NotI) such that they can be cloned into a standard mammalian expression vector, such as pCDNA3.1-hygro (+) (Invitrogen), by standard molecular biology methods known in the art.

The complete DNA sequence of pCDNA3.1-hygro (+)-IgH chain expression vector for the tagged hBU12 anti-human CD19 antibody will be as follows:

SEQ ID NO 13 (coding region of human IgG1 $V_H$-$C_H$ heavy chain for hBU12 with C-terminal LPETG sortase tag, 6xHis tag and a strepII tag underlined, and HindIII and NotI cloning sites shaded):

GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGC
CAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCA
AGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCA
GATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCA
TATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT
GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATT
TACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGAC
GGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGT
ATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCAC
GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA
AAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG
AGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCC

AAGCTGGCTAGCGTTTAAACTTAAGCTTCCATGGGATGGAGCTGGATCTTTCTTTTCCTCCTGTCAGGAACTGC
AGGTGTCCATTGTCAGGTTCAGCTGCAAGAGTCTGGCCCTGGGTTGGTTAAGCCCTCCCAGACCCTCAGTCTGA
CTTGTACTGTGTCTGGGGGTTCAATCAGCACTTCTGGTATGGGTGTAGGCTGGATTAGGCAGCACCCAGGGAAG
GGTCTGGAGTGGATTGGACACATTTGGTGGGATGATGACAAGAGATATAACCCAGCCCTGAAGAGCAGAGTGAC
AATCTCTGTGGATACCTCCAAGAACCAGTTTAGCCTCAAGCTGTCCAGTGTGACAGCTGCAGATACTGCTGTCT
ACTACTGTGCTAGAATGGAACTTTGGTCCTACTATTTTGACTACTGGGGCCAAGGCACCCTTGTCACAGTCTCC
TCAGCTAGCACCAAGGGCCCATCTGTCTTCCCCCTGGCACCCTCCTCCAAGAGCACCTCTGGGGGCACAGCTGC
CCTGGGCTGCCTGGTCAAGGACTACTTCCCTGAACCTGTGACAGTGTCCTGGAACTCAGGCGCCCTGACCAGCG
GCGTGCACACCTTCCCGGCTGTCCTACAGTCCTCAGGACTCTACTCCCTCAGCAGCGTGGTGACCGTGCCCTCC
AGCAGCTTGGGCACCCAGACCTACATCTGCAACGTGAATCACAAGCCCAGCAACACCAAGGTGGACAAGAAAGT
TGAGCCCAAATCTTGTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACCGTCAG
TCTTCCTCTTCCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTG
GACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGAC
AAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGC
TGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAGCCCCCATCGAGAAAACCATCTCCAAA
GCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGT
CAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGG
AGAACAACTACAAGACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTG
GACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGCACAACCACTACAC
ACAGAAGAGCCTCTCCCTGTCTCCGGGTAAACTGCCCGAGACCGGCCACCACCACCACCACCACGGCGAGCAGA
AGCTGATCAGCGAGGAGGACCTGGGCTGGAGCCACCCCCAGTTCGAGAAGTAGGCGGCCGCTCGAGTCTAGAGG
GCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCG
TGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGT
CTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAG
CAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATC
CCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCC
AGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGC
TCTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGG
GTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTT
AATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGAT
TTTGGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAA
TGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAAT
TAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTA
GTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGC
CCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAG
TGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGA
TCAGCACGTGATGAAAAAGCCTGAACTCACCGCGACGTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCG
TCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATAT
GTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGC
GCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCAC
AGGGTGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTCGCGGAGGCCATGGAT
GCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAAGGAATCGGTCAATACAC
TACATGGCGTGATTTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAACTGTGATGGACGACACCG
TCAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTC
GTGCACGCGGATTTCGGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCATTGACTGGAGCGA
GGCGATGTTCGGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGC
AGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCGCCGCGGCTCCGGGCGTATATGCTCCGC
ATTGGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTTCGATGATGCAGCTTGGGCGCAGGGTCGATG
CGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACAAATCGCCCGCAGAAGCGCGGCCGTCTGGA
CCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGTCCGAGGGCAAAGGAATAG
CACGTGCTACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGC
CGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTT
ATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGT
GGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATC
ATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAA
AGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAG
TCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCG
CTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAA
AGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAG
GCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAA
TCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCC
TCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCG

CTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGA
ACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACT
TATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG
AAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT
CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGC
AGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG
AACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTA
AAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTG
AGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACG
ATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTT
ATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGT
CTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCT
ACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGT
TACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGG
CCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTT
TCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGC
GTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGC
GAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCA
GCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAG
GGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTC
TCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAA
GTGCCACCTGACGTC

The complete DNA sequence of pCDNA3.1-hygro (+)-IgL chain expression vector for the tagged hBU12 anti-human CD19 antibody will be as follows:

SEQ ID NO 14 (coding region of human IgG1 $V_L$-$C_L$ kappa light chain for hBU12 with C-terminal LPETG sortase tag, 6xHis tag, Myc tag, and a strepII tag underlined, and HindIII and NotI cloning sites shaded):

GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGC
CAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCA
AGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCA
GATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCA
TATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT
GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATT
TACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGAC
GGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGT
ATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCAC
GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA
AAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG
AGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCC
AAGCTGGCTAGCGTTTAAACTTAAGCTTCCATGAAGTTGCCTGTTAGGCTGTTGGTGCTGATGTTCTGGATTCC
TGCTTCCAGCAGTGAAATTGTTCTCACCCAGTCTCCAGCAACCCTGTCTCTCTCTCCAGGGGAAAGGGCTACCC
TGAGCTGCAGTGCCAGCTCAAGTGTAAGTTACATGCACTGGTACCAGCAGAAGCCAGGGCAGGCTCCCAGACTC
CTGATTTATGACACATCCAAACTGGCTTCTGGTATTCCAGCAAGGTTCAGTGGCAGTGGGTCTGGAACAGATTT
TACACTCACAATCAGCAGCCTGGAGCCAGAGGATGTTGCTGTCTATTACTGTTTTCAGGGGAGTGTATACCCAT
TCACTTTTGGCCAAGGGACAAAGTTGGAAATCAAAAGAACTGTGGCTGCACCATCTGTCTTCATCTTCCCGCCA
TCTGATGAGCAGTTGAAATCTGGAACTGCCTCTGTTGTGTGCCTGCTGAATAACTTCTATCCCAGAGAGGCCAA
AGTACAGTGGAAGGTGGATAACGCCCTCCAATCGGGTAACTCCCAGGAGAGTGTCACAGAGCAGGACAGCAAGG
ACAGCACCTACAGCCTCAGCAGCACCCTGACGCTGAGCAAAGCAGACTACGAGAAACACAAAGTCTACGCCTGC
GAAGTCACCCATCAGGGCCTGAGCTCGCCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTCTGCCCGAGACCGG
CCACCACCACCACCACCACGGCGAGCAGAAGCTGATCAGCGAGGAGGACCTGGGCTGGAGCCACCCCCAGTTCG
AGAAGTAGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTG
CCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGAC

AGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGA
AAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGG
TTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTC
GCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACG
GCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTC
GCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATC
TCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACA
AAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAG
GCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGC
AGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCT
AACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCG
CCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCG
GGAGCTTGTATATCCATTTTCGGATCTGATCAGCACGTGATGAAAAAGCCTGAACTCACCGCGACGTCTGTCGA
GAAGTTTCTGATCGAAAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTT
TCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGT
TATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAG
CCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTG
TTCTGCAGCCGGTCGCGGAGGCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCA
TTCGGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTTCATATGCGCGATTGCTGATCCCCATGTGTA
TCACTGGCAAACTGTGATGGACGACACCGTCAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGG
CCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTCGGCTCCAACAATGTCCTGACGGACAATGGC
CGCATAACAGCGGTCATTGACTGGAGCGAGGCGATGTTCGGGGATTCCCAATACGAGGTCGCCAACATCTTCTT
CTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGAT
CGCCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTTC
GATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACA
AATCGCCCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCC
CCAGCACTCGTCCGAGGGCAAAGGAATAGCACGTGCTACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGG
TTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTT
CGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATA
AAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCG
TCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAAT
TCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAA
TTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGC
GCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG
GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAA
AGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGG
CTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAG
ATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGT
CCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTC
GTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCG
TCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGA
GGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGT
ATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGC
TGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGA
TCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAA
AGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTG
GTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTG
CCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGG
TCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTA
ATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTC
AGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGG
TCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTC
TTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGT
ATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGT
AACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGA
AGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATA
TTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAA
TAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

These constructs allow upon transfection into mammalian cells, like e.g. - but not limited to -CHO cells, that are typically used for recombinant antibody expression, the expression of the anti-human CD19 specific humanized antibody hBU12 with C-terminal additions of a sortase A tag, a 6xHis tag, a Myc tag, and a strepII tag at both the IgH and IgL chains.

**Example 2: Cloning of expression vectors for monoclonal antibody with C-terminal N-intein domain of Ssp GyrB 11 split-intein with additional C-terninal 6xHis and strepII affinity purification tags**

Similar to the design of expression cassettes and vectors of *Staphylococcus aureus* sortase A tagged IgG1 heavy and light chains, the coding regions for a C-terminal fusion of N-intein domain of Ssp GyrB 11 split-intein to either the IgH and IgL chain can be designed as follows, in order to gene synthesize the genes by a qualified CRO (e.g. Genscript (www.genscript.com, Piscataway, NJ, USA), with the same elements for the anti-human CD19 antibody as disclosed further above.

The 150 amino acid sequence of the N-intein domain of Ssp GyrB 11 split-intein can be found in a publication by Appleby et al. (2009), and is as follows:

SEQ ID NO 15 (N-intein domain of Ssp GyrB 11 split-intein):

```
CFSGDTLVALTDGRSVSFEQLVEEEKQGKQNFCYTIRHDGSIGVEKIINARKTKTNAKVIKVTLDNGESIICTP
DHKFMLRDGSYKCAMDLTLDDSLMPLHRKISTTEDSGHMEAVLNYNHRIVNIEAVSETIDVYDIEVPHTHNFAL
AS
```

Reverse translation of that amino acid sequence with mammalian codon usage will result in the coding sequence for the N-intein domain of Ssp GyrB 11 split-intein as follows:

SEQ ID NO 16 (endocing sequence for N-intein domain of Ssp GyrB 11 split-intein):

```
TGCTTCAGCGGCGACACCCTGGTGGCCCTGACCGACGGCAGAAGCGTGAGCTTCGAGCAGCTGGTGGAGGAGGA
GAAGCAGGGCAAGCAGAACTTCTGCTACACCATCAGACACGACGGCAGCATCGGCGTGGAGAAGATCATCAACG
CCAGAAAGACCAAGACCAACGCCAAGGTGATCAAGGTGACCCTGGACAACGGCGAGAGCATCATCTGCACCCCC
GACCACAAGTTCATGCTGAGAGACGGCAGCTACAAGTGCGCCATGGACCTGACCCTGGACGACAGCCTGATGCC
CCTGCACAGAAAGATCAGCACCACCGAGGACAGCGGCCACATGGAGGCCGTGCTGAACTACAACCACAGAATCG
TGAACATCGAGGCCGTGAGCGAGACCATCGACGTGTACGACATCGAGGTGCCCCACACCCACAACTTCGCCCTG
GCCAGC
```

With this sequence information at hand, the complete IgG1 heavy chain coding region for anti-human CD19 antibody hBU12 with C-terminal extension, comprising the N-intein domain of Ssp GyrB 11 split-intein, followed by a 6xHis-tag and a strepII tag can be designed as disclosed in SEQ ID NO 17 below:

```
ATGAATTTTGGACTGAGGCTGATTTTCCTGGTGCTGACCCTGAAAGGCGTCCAGTGTCAGGTTCAGCTGCAAGA
GTCTGGCCCTGGGTTGGTTAAGCCCTCCCAGACCCTCAGTCTGACTTGTACTGTGTCTGGGGGTTCAATCAGCA
CTTCTGGTATGGGTGTAGGCTGGATTAGGCAGCACCCAGGGAAGGGTCTGGAGTGGATTGGACACATTTGGTGG
GATGATGACAAGAGATATAACCCAGCCCTGAAGAGCAGAGTGACAATCTCTGTGGATACCTCCAAGAACCAGTT
TAGCCTCAAGCTGTCCAGTGTGACAGCTGCAGATACTGCTGTCTACTACTGTGCTAGAATGGAACTTTGGTCCT
ACTATTTTGACTACTGGGGCCAAGGCACCCTTGTCACAGTCTCCTCAGCTAGCACCAAGGGCCCATCTGTCTTC
CCCCTGGCACCCTCCTCCAAGAGCACCTCTGGGGGCACAGCTGCCCTGGGCTGCCTGGTCAAGGACTACTTCCC
TGAACCTGTGACAGTGTCCTGGAACTCAGGCGCCCTGACCAGCGGCGTGCACACCTTCCCGGCTGTCCTACAGT
CCTCAGGACTCTACTCCCTCAGCAGCGTGGTGACCGTGCCCTCCAGCAGCTTGGGCACCCAGACCTACATCTGC
AACGTGAATCACAAGCCCAGCAACACCAAGGTGGACAAGAAAGTTGAGCCCAAATCTTGTGACAAAACTCACAC
ATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGGACCGTCAGTCTTCCTCTTCCCCCCAAAACCCAAGGACA
CCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAAG
TTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCAC
GTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCT
CCAACAAAGCCCTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTG
TACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTA
TCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGC
TGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTC
TTCTCATGCTCCGTGATGCATGAGGCTCTGCACAACCACTACACACAGAAGAGCCTCTCCCTGTCTCCGGGTAA
ATGCTTCAGCGGCGACACCCTGGTGGCCCTGACCGACGGCAGAAGCGTGAGCTTCGAGCAGCTGGTGGAGGAGG
AGAAGCAGGGCAAGCAGAACTTCTGCTACACCATCAGACACGACGGCAGCATCGGCGTGGAGAAGATCATCAAC
GCCAGAAAGACCAAGACCAACGCCAAGGTGATCAAGGTGACCCTGGACAACGGCGAGAGCATCATCTGCACCCC
CGACCACAAGTTCATGCTGAGAGACGGCAGCTACAAGTGCGCCATGGACCTGACCCTGGACGACAGCCTGATGC
CCCTGCACAGAAAGATCAGCACCACCGAGGACAGCGGCCACATGGAGGCCGTGCTGAACTACAACCACAGAATC
GTGAACATCGAGGCCGTGAGCGAGACCATCGACGTGTACGACATCGAGGTGCCCCACACCCACAACTTCGCCCT
GGCCAGCCACCATCACCATCACCATGGCTGGAGCCACCCCCAGTTCGAGAAGTAG
```

This translates to amino acid sequence SEQ ID NO 18 (amino acids of the N-intein domain are underlined, 6xHis tag and strepII tag are shaded):

```
MNFGLRLIFLVLTLKGVQCQVQLQESGPGLVKPSQTLSLTCTVSGGSISTSGMGVGWIRQHPGKGLEWIGHIWW
DDDKRYNPALKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCARMELWSYYFDYWGQGTLVTVSSASTKGPSVF
PLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYIC
NVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVK
FNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQV
YTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNV
FSCSVMHEALHNHYTQKSLSLSPGKCFSGDTLVALTDGRSVSFEQLVEEEKQGKQNFCYTIRHDGSIGVEKIIN
ARKTKTNAKVIKVTLDNGESIICTPDHKFMLRDGSYKCAMDLTLDDSLMPLHRKISTTEDSGHMEAVLNYNHRI
VNIEAVSETIDVYDIEVPHTHNFALASHHHHHHGWSHPQFEK•
```

Likewise, a complete IgG1 kappa light chain coding region for anti-human CD19 antibody hBU12 with C-terminal extension, comprising the N-intein domain of Ssp GyrB 11 split-intein, followed by a 6xHis-tag and a strepII tag can be designed as disclosed in SEQ ID NO 19 below:

ATGAATTTTGGACTGAGGCTGATTTTCCTGGTGCTGACCCTGAAAGGCGTCCAGTGTGACATTGTGCTGACCCA
ATCTCCAGCTTCTTTGGCTGTGTCTCTAGGGCAGAGGGCCACCATCTCCTGCAAGGCCAGCCAAAGTGTTGATT
TTGATGGTGATAGTTATATGAACTGGTACCAACAGAAACCAGGACAGCCACCCAAAGTCCTCATCTATGCTGCA
TCCAATCTAGAATCTGGGATCCCAGCCAGGTTTAGTGGCAGTGGGTCTGGGACAGACTTCACCCTCAACATCCA
TCCTGTGGAGGAGGAGGATGCTGCAACCTATTACTGTCAGCAAAGTAATGAGGATCCGTGGACGTTCGGTGGAG
GCACCAAGCTGGAAATCAAACGTACGGTGGCTGCACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTG
AAATCTGGAACTGCCTCTGTTGTGTGCCTGCTGAATAACTTCTATCCCAGAGAGGCCAAAGTACAGTGGAAGGT
GGATAACGCCCTCCAATCGGGTAACTCCCAGGAGAGTGTCACAGAGCAGGACAGCAAGGACAGCACCTACAGCC
TCAGCAGCACCCTGACGCTGAGCAAAGCAGACTACGAGAAACACAAAGTCTACGCCTGCGAAGTCACCCATCAG
GGCCTGAGCTCGCCCGTCACAAAGAGCTTCAACAGGGGAGAGTGCTTCAGCGGCGACACCCTGGTGGCCCTGAC
CGACGGCAGAAGCGTGAGCTTCGAGCAGCTGGTGGAGGAGGAGAAGCAGGGCAAGCAGAACTTCTGCTACACCA
TCAGACACGACGGCAGCATCGGCGTGGAGAAGATCATCAACGCCAGAAAGACCAAGACCAACGCCAAGGTGATC
AAGGTGACCCTGGACAACGGCGAGAGCATCATCTGCACCCCCGACCACAAGTTCATGCTGAGAGACGGCAGCTA
CAAGTGCGCCATGGACCTGACCCTGGACGACAGCCTGATGCCCCTGCACAGAAAGATCAGCACCACCGAGGACA
GCGGCCACATGGAGGCCGTGCTGAACTACAACCACAGAATCGTGAACATCGAGGCCGTGAGCGAGACCATCGAC
GTGTACGACATCGAGGTGCCCCACACCCACAACTTCGCCCTGGCCAGCCACCATCACCATCACCATGGCTGGAG
CCACCCCCAGTTCGAGAAGTAG

This translates to amino acid sequence SEQ ID NO 20 (amino acids of the N-intein domain are underlined, 6xHis tag and strepII tag are shaded):

MNFGLRLIFLVLTLKGVQCDIVLTQSPASLAVSLGQRATISCKASQSVDFDGDSYMNWYQQKPGQPPKVLIYAA
SNLESGIPARFSGSGSGTDFTLNIHPVEEEDAATYYCQQSNEDPWTFGGGTKLEIKRTVAAPSVFIFPPSDEQL
KSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQ
GLSSPVTKSFNRGECFSGDTLVALTDGRSVSFEQLVEEEKQGKQNFCYTIRHDGSIGVEKIINARKTKTNAKVI
KVTLDNGESIICTPDHKFMLRDGSYKCAMDLTLDDSLMPLHRKISTTEDSGHMEAVLNYNHRIVNIEAVSETID
VYDIEVPHTHNFALASHHHHHHGWSHPQFEK•

The coding regions for the N-intein modified heavy and light chains of the anti-human CD19 specific antibody as disclosed in SEQ ID NOs 17 and 19, respectively, can then be synthesized with flanking restriction enzyme sites (e.g. HindIII and NotI) such that they can be cloned into a standard mammalian expression vector, such as pCDNA3.1-hygro (+) (Invitrogen), by standard molecular biology methods known in the art.

The complete DNA sequence of pCDNA3.1-hygro (+)-IgH chain expression vector for the N-intein tagged hBU12 anti-human CD19 antibody is then as follows:

SEQ ID NO 21 (coding region of human IgG1 $V_H$-$C_H$ heavy chain for hBU12 with C-terminal N-intein domain of Ssp GyrB S11 split intein, followed by 6xHis tag strepII tag (underlined), and HindIII and NotI cloning sites (shaded)):

GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGC
CAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCA
AGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCA
GATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCA
TATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT
GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATT
TACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGAC

GGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGT
ATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCAC
GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA
AAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG
AGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCC
AAGCTGGCTAGCGTTTAAACTTAAGCTTCCATGAATTTTGGACTGAGGCTGATTTTCCTGGTGCTGACCCTGAA
AGGCGTCCAGTGTCAGGTTCAGCTGCAAGAGTCTGGCCCTGGGTTGGTTAAGCCCTCCCAGACCCTCAGTCTGA
CTTGTACTGTGTCTGGGGGGTTCAATCAGCACTTCTGGTATGGGTGTAGGCTGGATTAGGCAGCACCCAGGGAAG
GGTCTGGAGTGGATTGGACACATTTGGTGGGATGATGACAAGAGATATAACCCAGCCCTGAAGAGCAGAGTGAC
AATCTCTGTGGATACCTCCAAGAACCAGTTTAGCCTCAAGCTGTCCAGTGTGACAGCTGCAGATACTGCTGTCT
ACTACTGTGCTAGAATGGAACTTTGGTCCTACTATTTTGACTACTGGGGCCAAGGCACCCTTGTCACAGTCTCC
TCAGCTAGCACCAAGGGCCCATCTGTCTTCCCCCTGGCACCCTCCTCCAAGAGCACCTCTGGGGGCACAGCTGC
CCTGGGCTGCCTGGTCAAGGACTACTTCCCTGAACCTGTGACAGTGTCCTGGAACTCAGGCGCCCTGACCAGCG
GCGTGCACACCTTCCCGGCTGTCCTACAGTCCTCAGGACTCTACTCCCTCAGCAGCGTGGTGACCGTGCCCTCC
AGCAGCTTGGGCACCCAGACCTACATCTGCAACGTGAATCACAAGCCCAGCAACACCAAGGTGGACAAGAAAGT
TGAGCCCAAATCTTGTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGACCGTCAG
TCTTCCTCTTCCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTG
GACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGAC
AAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGC
TGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAGCCCCCATCGAGAAAACCATCTCCAAA
GCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGT
CAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGG
AGAACAACTACAAGACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTG
GACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGCACAACCACTACAC
ACAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGCTTCAGCGGCGACACCCTGGTGGCCCTGACCGACGGCAGAA
GCGTGAGCTTCGAGCAGCTGGTGGAGGAGGAGAAGCAGGGCAAGCAGAACTTCTGCTACACCATCAGACACGAC
GGCAGCATCGGCGTGGAGAAGATCATCAACGCCAGAAAGACCAAGACCAACGCCAAGGTGATCAAGGTGACCCT
GGACAACGGCGAGAGCATCATCTGCACCCCCGACCACAAGTTCATGCTGAGAGACGGCAGCTACAAGTGCGCCA
TGGACCTGACCCTGGACGACAGCCTGATGCCCCTGCACAGAAAGATCAGCACCACCGAGGACAGCGGCCACATG
GAGGCCGTGCTGAACTACAACCACAGAATCGTGAACATCGAGGCCGTGAGCGAGACCATCGACGTGTACGACAT
CGAGGTGCCCCACACCCACAACTTCGCCCTGGCCAGCCACCATCACCATCACCATGGCTGGAGCCACCCCCAGT
TCGAGAAGTAGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAG
TTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTT
CCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAG
GACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGC
GGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGG
TGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTT
CTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTT
ACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTT
TTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCT
ATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTA
ACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGG
CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCA
GGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCC
CCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGG
CCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTC
CCGGGAGCTTGTATATCCATTTTCGGATCTGATCAGCACGTGATGAAAAAGCCTGAACTCACCGCGACGTCTGT
CGAGAAGTTTCTGATCGAAAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTG
CTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGAT
CGTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGA
GAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCG
CTGTTCTGCAGCCGGTCGCGGAGGCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGC
CCATTCGGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTTCATATGCGCGATTGCTGATCCCCATGT
GTATCACTGGCAAACTGTGATGGACGACACCGTCAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCTGATGCTTT
GGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTCGGCTCCAACAATGTCCTGACGGACAAT
GGCCGCATAACAGCGGTCATTGACTGGAGCGAGGCGATGTTCGGGGATTCCCAATACGAGGTCGCCAACATCTT
CTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAG
GATCGCCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAAT
TTCGATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTAC
ACAAATCGCCCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGAC
GCCCCAGCACTCGTCCGAGGGCAAAGGAATAGCACGTGCTACGAGATTTCGATTCCACCGCCGCCTTCTATGAA
AGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTT

CTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAA
ATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATA
CCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCAC
AATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACAT
TAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAA
CGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT
TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCAT
AGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATA
AAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACC
TGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAG
GTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTA
TCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAG
CGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTT
GGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCAC
CGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTT
TGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCA
AAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAAC
TTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAG
TTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATA
CCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAG
TGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAG
TTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCA
TTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTT
CGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATT
CTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAG
TGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAA
AGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGA
TGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACA
GGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCA
ATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAAC
AAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

The complete DNA sequence of pCDNA3.1-hygro (+)-IgL chain expression vector for the Ssp GyrB S11 N-intein domain tagged hBU12 anti-human CD19 antibody will be as follows:

SEQ ID NO 22 (coding region of human IgG1 $V_L$-$C_L$ kappa light chain for hBU12 with C-terminal Ssp GyrB S11 N-intein domain, 6xHis tag and a strepII tag underlined, and HindIII and NotI cloning sites shaded):

GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGC
CAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCA
AGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCA
GATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCA
TATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT
GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATT
TACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGAC
GGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGT
ATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCAC
GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA
AAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG
AGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCC
AAGCTGGCTAGCGTTTAAACTTAAGCTTCCATGAATTTTGGACTGAGGCTGATTTTCCTGGTGCTGACCCTGAA
AGGCGTCCAGTGTGACATTGTGCTGACCCAATCTCCAGCTTCTTTGGCTGTGTCTCTAGGGCAGAGGGCCACCA

TCTCCTGCAAGGCCAGCCAAAGTGTTGATTTTGATGGTGATAGTTATATGAACTGGTACCAACAGAAACCAGGA
CAGCCACCCAAAGTCCTCATCTATGCTGCATCCAATCTAGAATCTGGGATCCCAGCCAGGTTTAGTGGCAGTGG
GTCTGGGACAGACTTCACCCTCAACATCCATCCTGTGGAGGAGGAGGATGCTGCAACCTATTACTGTCAGCAAA
GTAATGAGGATCCGTGGACGTTCGGTGGAGGCACCAAGCTGGAAATCAAACGTACGGTGGCTGCACCATCTGTC
TTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACTGCCTCTGTTGTGTGCCTGCTGAATAACTTCTA
TCCCAGAGAGGCCAAAGTACAGTGGAAGGTGGATAACGCCCTCCAATCGGGTAACTCCCAGGAGAGTGTCACAG
AGCAGGACAGCAAGGACAGCACCTACAGCCTCAGCAGCACCCTGACGCTGAGCAAAGCAGACTACGAGAAACAC
AAAGTCTACGCCTGCGAAGTCACCCATCAGGGCCTGAGCTCGCCCGTCACAAAGAGCTTCAACAGGGGAGAGTG
CTTCAGCGGCGACACCCTGGTGGCCCTGACCGACGGCAGAAGCGTGAGCTTCGAGCAGCTGGTGGAGGAGGAGA
AGCAGGGCAAGCAGAACTTCTGCTACACCATCAGACACGACGGCAGCATCGGCGTGGAGAAGATCATCAACGCC
AGAAAGACCAAGACCAACGCCAAGGTGATCAAGGTGACCCTGGACAACGGCGAGAGCATCATCTGCACCCCCGA
CCACAAGTTCATGCTGAGAGACGGCAGCTACAAGTGCGCCATGGACCTGACCCTGGACGACAGCCTGATGCCCC
TGCACAGAAAGATCAGCACCACCGAGGACAGCGGCCACATGGAGGCCGTGCTGAACTACAACCACAGAATCGTG
AACATCGAGGCCGTGAGCGAGACCATCGACGTGTACGACATCGAGGTGCCCCACACCCACAACTTCGCCCTGGC
CAGCCACCATCACCATCACCATGGCTGGAGCCACCCCCAGTTCGAGAAGTAGGCGGCCGCTCGAGTCTAGAGGG
CCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGT
GCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTC
TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC
AGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCC
CCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCA
GCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCT
CTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGG
TGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTA
ATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATT
TTGGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAAT
GTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATT
AGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAG
TCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCC
CCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGT
GAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGAT
CAGCACGTGATGAAAAAGCCTGAACTCACCGCGACGTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGT
CTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATG
TCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCG
CTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACA
GGGTGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTCGCGGAGGCCATGGATG
CGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAAGGAATCGGTCAATACACT
ACATGGCGTGATTTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAACTGTGATGGACGACACCGT
CAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCG
TGCACGCGGATTTCGGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCATTGACTGGAGCGAG
GCGATGTTCGGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCA
GCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCGCCGCGGCTCCGGGCGTATATGCTCCGCA
TTGGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTTCGATGATGCAGCTTGGGCGCAGGGTCGATGC
GACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACAAATCGCCCGCAGAAGCGCGGCCGTCTGGAC
CGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGTCCGAGGGCAAAGGAATAGC
ACGTGCTACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCC
GGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTA
TAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTG
GTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCA
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAA
GTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGT
CGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGC
TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAA
GGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGG
CCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAAT
CGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCT
CGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGC
TTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAA
CCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTT
ATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGA
AGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTC
GGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCA
GCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGA

ACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAA
AAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGA
GGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGA
TACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTA
TCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTC
TATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTT
ACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGC
CGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTT
CTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCG
TCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCG
AAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAG
CATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGG
GCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCT
CATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
TGCCACCTGACGTC

These pcDNA3.1-hygro(+) based expression vectors disclosed in SEQ ID NOs 21 and 22 allow upon transfection into mammalian cells, like e.g. but not limited to CHO cells, that are typically used for recombinant antibody expression, the expression of the anti-human CD19 specific humanized antibody hBU12 with C-terminal N-intein domain fused, followed by a 6xHis tag and a strepII tag at both the IgH and IgL chains.

**Example 3: Cloning and expression of recombinant Sortase A enzyme from *Staphylococcus aureus*.**

The ORF of Sortase A from *Staphylococcus aureus* is published in Genbank and can be found under entry: AF162687.1. The aa-sequence in that record reads is shown as SEQ ID NO 23 (amino acid sequence of sortase A from *Staphylococcus aureus*):

MKKWTNRLMTIAGVVLILVAAYLFAKPHIDNYLHDKDKDEKIEQYDKNVKEQASKDKKQQAKPQIPKDKSKVAG
YIEIPDADIKEPVYPGPATPEQLNRGVSFAEENESLDDQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGN
ETRKYKMTSIRDVKPTDVGVLDEQKGKDKQLTLITCDDYNEKTGVWEKRKIFVATEVK

The corresponding nucleotide sequence in this Genbank entry is provided as SEQ ID NO 24:

ATGAAAAAATGGACAAATCGATTAATGACAATCGCTGGTGTGGTACTTATCCTAGTGGCAGCATATTTGTTTGC
TAAACCACATATCGATAATTATCTTCACGATAAAGATAAAGATGAAAAGATTGAACAATATGATAAAAATGTAA
AAGAACAGGCGAGTAAAGATAAAAAGCAGCAAGCTAAACCTCAAATTCCGAAAGATAAATCGAAAGTGGCAGGC
TATATTGAAATTCCAGATGCTGATATTAAAGAACCAGTATATCCAGGACCAGCAACACCTGAACAATTAAATAG
AGGTGTAAGCTTTGCAGAAGAAAATGAATCACTAGATGATCAAAATATTTCAATTGCAGGACACACTTTCATTG
ACCGTCCGAACTATCAATTTACAAATCTTAAAGCAGCCAAAAAAGGTAGTATGGTGTACTTTAAAGTTGGTAAT
GAAACACGTAAGTATAAAATGACAAGTATAAGAGATGTTAAGCCTACAGATGTAGGAGTTCTAGATGAACAAAA
AGGTAAAGATAAACAATTAACATTAATTACTTGTGATGATTACAATGAAAAGACAGGCGTTTGGGAAAAACGTA
AAATCTTTGTAGCTACAGAAGTCAAATAA

Technical information with respect to the expression of an enzymatically active fragment of recombinant sortase A in E.coli, comprising amino acids 60-205 with 6xHis tag are disclosed in reference WO2007/108013A2. The coding region for a 6xHis tagged version of *Staphylococcus aureus* sortase A (aa60-205) is provided below as SEQ ID NO 25:

```
ATGCAAGCTAAACCTCAAATTCCGAAAGATAAATCGAAAGTGGCAGGCTATATTGAAATTCCAGATGCTGATAT
TAAAGAACCAGTATATCCAGGACCAGCAACACCTGAACAATTAAATAGAGGTGTAAGCTTTGCAGAAGAAAATG
AATCACTAGATGATCAAAATATTTCAATTGCAGGACACACTTTCATTGACCGTCCGAACTATCAATTTACAAAT
CTTAAAGCAGCCAAAAAAGGTAGTATGGTGTACTTTAAAGTTGGTAATGAAACACGTAAGTATAAAATGACAAG
TATAAGAGATGTTAAGCCTACAGATGTAGGAGTTCTAGATGAACAAAAAGGTAAAGATAAACAATTAACATTAA
TTACTTGTGATGATTACAATGAAAAGACAGGCGTTTGGGAAAAACGTAAAATCTTTGTAGCTACAGAAGTCAAA
CACCATCACCATCACCATTAA
```

This translates to amino acid sequence SEQ ID NO 26:

```
MQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNRGVSFAEENESLDDQNISIAGHTFIDRPNYQFTN
LKAAKKGSMVYFKVGNETRKYKMTSIRDVKPTDVGVLDEQKGKDKQLTLITCDDYNEKTGVWEKRKIFVATEVK
HHHHHH•
```

The coding region for the 6xHis tagged sortase A fragment of *Staphylococcus aureus*, as provided in SEQ ID NO 25, can be cloned into a standard bacterial expression vector, like e.g. pET29 (Novagen), in order to transform *E. coli* strain BL21(DE3) (Novagen) and to generate an *E. coli* clone that can be used for the bacterial production of recombinant sortase A according to standard methods known in the art. In short, *E. coli* BL21(DE3) transformed with pET29 expression plasmids for sortase A can be cultured at 37 °C in LB mediom with 50 μg/mL kanamycin until until an $OD_{600}$ = 0.5-0.8 is reached. IPTG can then be added to a final concentration of 0.4 mM and protein expression can be induced for three hours at 30 °C. The cells can then be harvested by centrifugation and resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl supplemented with 1 mM MgCl2, 2 units/mL DNAseI (NEB), 260 nM aprotinin, 1.2 μM leupeptin, and 1 mM PMSF). Cells can then be lysed by sonication and clarified supernatant can then be purified on Ni-NTA agarose following the manufacturer's instructions.

Fractions that are of >90% purity, as judged by SDS-PAGE, can then be consolidated and dialyzed against Tris-buffered saline (25 mM Tris pH 7.5, 150 mM NaCl), and the enzyme concentration can be calculated from the measured $A_{280}$ using the published extinction coefficient of 17,420 $M^{-1}$ $cm^{-1}$.The above-mentioned protocol has been followed and ca. 20 mg of > 90% pure recombinant enzymatically active fragment (of ca. 17 kD) sortase A of *Staphylococcus aureus* has been produced and the analysis of the recombinant protein by

SDS-PAGE and Western blotting is disclosed in Fig. 7.

**Example 4: Expression and purification of sortase tagged or N-intein tagged recombinant antibodies in CHO cells**

a.) CHO cell expression: Expression of recombinant IgG1 antibodies from the expression constructs disclosed under Examples 2 and 3 can be achieved by transient transfection using e.g. commercially available CHO expression systems, like the FreeStyle CHO system from Invitrogen following the instructions of the FreeStyle CHO manual.

In brief, about 1 day prior to transfection, CHO cells shall be seeded at 5-6 $x10^6$ cells/ml in FreeStyle CHO medium in shaker-flasks in order to expand them at 120 rpm on an orbital shaker at 37°C in a humidified incubator at 7.5% $CO_2$ atmosphere. The following day the cells can be transfected, when they reach a density of 1.2-1.5$x10^6$/ml. Cells then need to be diluted to $1x10^6$ cells/ml. 30 ml of such a cell suspension then needs to be added to a 125 ml shake flask and 40 µg of 1:1 mixed IgH and IgL expression plasmid DNA is added to 600µl OptiPro SF-medium (Invitrogen). At the same time, 40µl of FreeStyle MAX transfection reagent needs to be added to 600µl OptiPro SF-medium, and both samples need to be gently mixed, and incubated for 10 min at room temperature to allow DNA-transfection reagent complexes to form. Then the DNA-transfection reagent mix can be added slowly to the 125 ml CHO cell culture from above and the transfected cells are then grown for up to 6 days at 120 rpm on an orbital shaker at 37°C in a humidified incubator at 7.5% $CO_2$ atmosphere. Thereafter, cell culture supernatant can be collected and analyzed for antibody expression titer by appropriate methods known in the art (ELISA, Luminex, etc.).

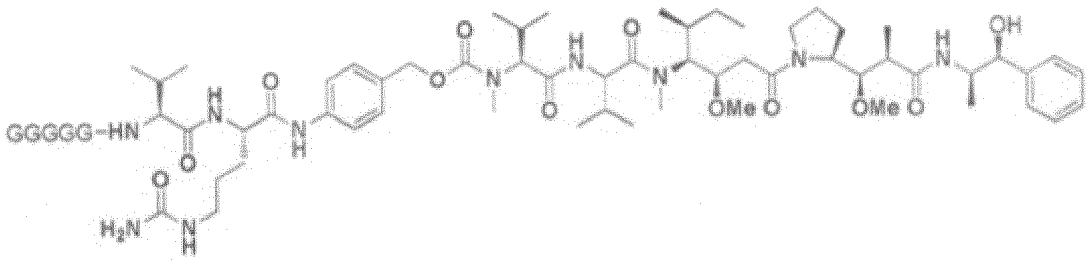
b.) Protein A purification: Protein A purification of recombinant antibodies from the CHO cell supernatant can be performed with commercially available protein A sepharose columns (Thermo Fisher, Pierce) according to instructions from the manufacturer.

In brief, cleared cell culture supernatant is run over a protein A column of appropriate size and capacity equilibrated with PBS. Residual medium is washed with PBS and eventually bound IgG can be eluted with low pH buffer, like 0.1 M citric acid-NaOH, pH 3.0. Eluted IgG should be neutralized immediately with 1/10th volume of 1M Tris/Cl, pH7.4. Combined fractions containing IgG can then be dialized against PBS over night at 4°C.

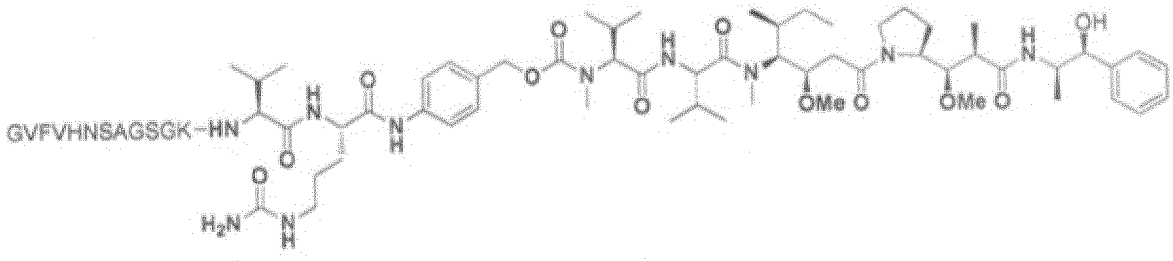
The protocols provided in Example 4 provide the skilled person in the art with the instruction to produce sufficient quantities of purified, recombinant antibodies from the constructs disclosed in Examples 1 and 2.

**Example 5: Generation of site-specifically C-terminally MMAE toxic payload conjugated monoclonal antibodies by sortase and split-intein mediated transpeptidation**

Monomethyl Auristatin A toxin coupled to a 5 amino acid glycine stretch and a 6 amino acid SSp GyrB S11 C-int split intein peptide according to the formulas provided below, can be custom ordered from qualified chemistry CROs.




**Formula 1: 5-glycine modified MMAE with vcPAB linker**
Me = methyl (-$CH_3$) group
G = glycine amino acid residue




**Formula 2: MMAE with vcPAB linker, modified with 6 amino acid C-intein domain GVFVHN and 6 amino acid SAGSGK C-extein peptide**
Me = methyl (-$CH_3$) group
GVFVHNSAGSGK = Gly-Val-Phe-Val-His-Asn-Ser-Ala-Gly-Ser-Gly-Lys stretch

a.) Toxic MMAE payload conjugation of LPETG sortaseA motif tagged recombinant IgG antibodies

Conjugation of 5 glycine amino acid modified MMAE toxic payload to LPETG sortase A tagged IgG1 antibody (that can be produced by following Examples 1 and 4) can be achieved by mixing appropriate ratios of LPETG tagged IgG1 antibody with the glycine-modified MMAE toxin disclosed in Formula 1 (e.g. at 1:1 ratio and 50 μM concentration) and with recombinant sortase A (production described in Example 3) (e.g. at 5 μM concentration), and using physiologic incubation buffer, like e.g.; 5 mM Tris/Cl, 15 mM NaCl, 6 mM $CaCl_2$, pH 8.0, and incubating at 37°C to 40°C for a minimum of 2 hours.

Efficiency of the conjugation can be monitored by analyzing the absence of the 6xHis tag and/or the strepII tag after stopping the reaction, e.g. by western-blot analysis or ELISA with anti-His-tag and/or anti strepII tag antibodies.

Completely conjugated product can be enriched by Nickel-NTA columns, or streptactin column binding, which bind to the 6xHis tag or strepII tag, respectively, which can only be present in incompletely reacted IgG1 substrate. Final IgG-payload conjugate can eventually be purified using protein A purification as described above.

b.) Toxic MMAE payload conjugation of SSp GyrB S11 N-intein tagged recombinant IgG antibodies

Conjugation of Ssp GyrB S11 C-intein amino acid modified MMAE toxic payload to N-intein tagged IgG1 antibody (that can be produced by following Examples 2 and 4) can be achieved by mixing appropriate ratios of N-intein tagged IgG1 antibody with the C-intein amino acid-modified MMAE toxin disclosed in Formula 2 (e.g. at 1:10 or 1:25 ratio at 5 μM concentration of the IgG antibody) using physiologic incubation buffer, like e.g.; 20 mM Tris/Cl, 250 mM NaCl, 1 mM EDTA, pH 8.5, and incubating at room temperature or at 37°C a minimum of 4 hours.

Efficiency of the conjugation can be monitored by analyzing the absence of the 6xHis tag and/or the strepII tag after stopping the reaction, e.g. by western-blot analysis or ELISA with anti-His-tag and/or anti strepII tag antibodies.

Completely conjugated product can be enriched by Nickel-NTA columns, or streptactin column binding, which bind to the 6xHis tag or strepII tag, respectively, which can only be present in incompletely reacted IgG1 substrate. Final IgG-payload conjugate can eventually be purified using protein A purification as described above.

In summary, the Examples 1-5 disclosed above allow a person skilled in the art to practice the invention of enzymatically conjugating a toxic payload site-specifically to the C-terminus either using sortase A mediated or split-intein mediated transpeptidation.

**Example 6: Production of Trastuzumab with C-terminal GS (glycine-serine) linker, LPETG Sortase motif and additional 6x-His and Strep II affinity purification tags on either heavy or light chain**

Antibody expression constructs encoding monoclonal antibody Trastuzumab (Tras) heavy and light chains, either untagged (SEQ ID NOs: 31 - 34) or C-terminally tagged with GS (glycine-serine) linker, LPETG Sortase tag, 6xHis tag, and Strep II tag (SEQ ID NOs: 35 - 38) were generated essentially as described in Example 1. Using these expression constructs, Tras-HC-GS-LHS and Tras-LC-GS-LHS (HC=heavy chain, LC=light chain, GS=glycine-serine, LHS=LPETG-tag+6xHis-tag+strepII-tag) were produced in CHO cells by co-transfection of the corresponding expression constructs. Tras-HC-GS-LHS is a Trastuzumab variant with an unmodified light chain (SEQ ID NOs: 35 - 36), and a heavy chain C-terminally tagged with GS (glycine-serine) linker, LPETG Sortase motif, 6xHis-tag, and strepII-tag (SEQ ID NOs: 33 - 34). Tras-LC-GS-LHS is a Trastuzumab variant with an unmodified heavy chain (SEQ ID NOs: 31 - 32), and a light chain C-terminally tagged with GS linker, LPETG Sortase motif, 6xHis-tag, and strepII-tag (SEQ ID NOs: 37 - 38). CHO cell transfection and affinity purification of antibodies by proteinA-sepharose chromatography was done essentially as described in Example 4.

**Example 7: Sortase A-mediated conjugation of heavy or light chain of Trastuzumab with Gly5-modified DM1 toxin**

Conjugation reactions containing $Gly_5$-modified DM1 toxin (ordered from Concortis, San Diego, CA, U.S., structure see Fig. 14 a.) and a 17kD recombinant sortase A fragment from

*Staphylococcus aureus* (see Example 3) were carried out with 10.5mg of each monoclonal antibody (mAb) (see Example 6) in 1x Sortase buffer (25mM Tris-HCl, pH8.2; 150mM NaCl; 7.5mM $CaCl_2$), as shown in Table II, below. The Tras-HC-GS-LHS conjugation reaction was incubated at 25°C for 2h; the Tras-LC-GS-LHS conjugation reaction was incubated at 25°C for 18h. Each reaction mixture was then passed over a Strep-Tactin® Sepharose columns (IBA Life-Sciences, Göttingen, Germany). For this, 1ml of Strep-Tactin Agarose was packed under gravity into a fritted column and equilibrated with 2 column volumes of equilibration buffer (100 mM Tris-HCl, pH 8.0; 150 mM NaCl; 1 mM EDTA). Each conjugation mixture was passed twice down the same column using gravity flow (to increase residence time on the resin). The resin was washed with an additional column volume of equilibration buffer to maximize conjugate yield and the pool then applied immediately to a protein A column. For this, a 1ml Protein A HiTrap column was equilibrated with 10 column volumes of buffer (25mM sodium phosphate pH 7.5). Each conjugation reaction was then applied to an equilibrated column and the column washed with a further 5 column volumes of buffer. Bound conjugate was eluted with 5 column volumes of elution buffer (0.1M succinic acid, pH 2.8) with 1 column volume fractions collected (into tubes containing 25% v/v 1M Tris Base to neutralise the acid) and analysed for protein content. Protein containing fractions were pooled and formulated by G25 column chromatography. For this, NAP 25 columns of an appropriate size for each scale of manufacture were used to formulate the conjugates for long term storage. The columns were equilibrated, loaded and eluted with 10mM Sodium Succinate pH 5.0, 100mg/mL Trehalose, 0.1% % w/v Polysorbate 20 (Formulation Buffer for Kadcyla® (T-DM1), marketed by Roche/Genentech) according to the manufacturer's instructions.

The Tras-HC-GS-LHS and Tras-LC-GS-LHS DM1-conjugate yields were, respectively, 8.0mg (76.2%) and 5.9mg (56.2%). The major process losses occurred during Protein A and G25 purification, most probably as a result of peak cutting to ensure maximal concentration of the product for each subsequent step or storage.

| **Reaction component** | **HC** | **LC** | **Final concentration** |
|---|---|---|---|
| Tras-HC-GS-LHS (5.3mg/ml) | 1981µl | - | 5µM |
| Tras-LC-GS-LHS (5.5mg/ml) | - | 1911µl | 5µM |

| | | | |
|---|---|---|---|
| $H_20$ | 7775.25µl | 7714µl | - |
| $Gly_5$-DM1 (1mM) | 1400µl | 1400µl | 100µM |
| Sortase A (0.85mg/ml = ca. 50µM) | 43.75µl | 175µl | 0.156/0.625µM |
| 5x Sortase buffer* | 2800µl | 2800µl | 1x |

**Table 2: Conjugation conditions for Tras-HC-GS-LHS and Tras-LC-GS-LHS:**

The drug loading was assessed by Hydrophobic Interaction Chromatography (HIC), and was performed on a TOSOH Butyl-NPR 4.6mm x 3.5cm, 2.5µm column run at 0.8mL/min with a 12 minute linear gradient between A - 1.5M $(NH_4)_2SO_4$, 25mM NaPi, pH=6.95±0.05 and B - 75% 25mM NaPi, pH=6.95±0.05, 25% IPA. The HIC profiles revealed that for both, Tras-HC-GS-LHS and Tras-LC-GS-LHS, there was no detectable unconjugated mAb left, and a major fraction of each mAb was loaded with 2 drugs (see Figure 8).

**Example 8: *In vitro* toxicity assay with sortase A - mediated Trastuzumab-DM1 conjugates**

Cytotoxicity of DM1-sortaseA-conjugated Tras-HC-GS-LHS and DM1-sortaseA-conjugated Tras-LC-GS-LHS was investigated and compared to Kadcyla® (Roche/Genentech) using SKBR3 cells, a human breast cancer cell line overexpressing the cognate antigen of trastuzumab (Tras) HER-2/neu, and T47D-5R cells, a breast cancer cell line naturally expressing low levels of HER-2/neu, engineered to be devoid of cell surface HER-2/neu (Graus-Porta et al. (1995)). Cells were plated on 96 well plates in 100µl complete DMEM (10'000 cells per well). After one day incubation, 50µl medium was carefully removed from each well and replaced by 50µl of 3.5-fold serial dilutions of each ADC in complete DMEM, resulting in ADC concentrations ranging from 20µg/ml to 0.25ng/ml. Each dilution was done in duplicates or triplicates. After 3 additional days incubation at 37°C in a humidified incubator at 5% $CO_2$ atmosphere, plates were removed from the incubator and equilibrated to room temperature. After approximately 30 minutes, 100µl CellTiter-Glo® Luminescent Solution (Promega, Cat.No G7570) was added to each well and, after shaking the plates at 450rpm for 5 min followed by a 10 min incubation without shaking, luminescence was measured on a Tecan Infinity F200 with an integration time of 1 second per well. All three ADCs were highly cytotoxic for the HER-2/neu overexpressing SKBR3 breast cancer cell line, but not for the HER-2/neu-negative T47D-5R breast cancer cell line

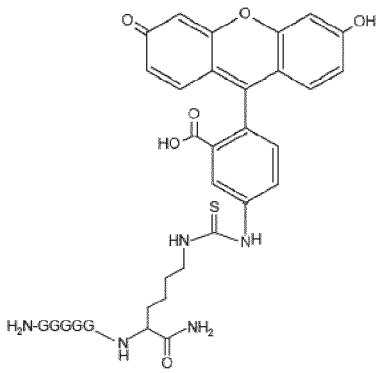
(see Fig. 9). The $EC_{50}$ values for Her-2/neu positive breast cancer cell line SKBR3 were: Kadcyla®, 32.4ng/ml; DM1-conjugated Tras-HC-GS-LHS, 45.6ng/ml; Tras-LC-GS-LHS, 51.4ng/ml, and thus are within similar range of potency in the *in vitro* tumor cell killing experiment. Conversely, no specific cellular toxicity was detectable with the Her-2/neu negative breast cancer cell line T47D-5R, demonstrating the functional equivalence of sortaseA, enzymatically conjugated ADC versus traditional, chemically conjugated ADC, when the comparison entails the same targeting antibody and the same toxin (DM1) (Fig. 9). However, it appears that the lower drug-to antibody ratio of ca. 1.80 (deducted from intergration of the DAR1 and DAR2 peaks in Fig. 8) for the Tras-HC-GS-LHS and Tras-LC-GS-LHS sortase A-conjugated ADCs, as compared to the DAR of ca. 3-4, reported for Kadcyla® does not translate into a proportionally different cellular cytotoxicity in the *in vitro* tumor cell killing assays (Fig. 9). This unexpected finding may be the result of a more defined and site-specific toxin-antibody conjugation mediated by sortase A in comparison to the less defined, stochastically, chemically conjugated Kadcyla®.

**Example 9: Optimization of synchronization of sortaseA mediated antibody heavy chain and light chain payload conjugation by variation of peptide-spacer length inserted between C-terminal end of antibody heavy chain and light chain and the sortaseA recognition motif**

The influence of peptide-spacer length positioned between the C-terminus of antibody heavy or light chain and LPETG sortase A recognition motif was investigated. For this, antibody heavy chain and light chain expression constructs encoding chimeric CD30-specific mAb Ac10 heavy and light chains (HC sequence derived from US 2008213289A1, Seq1, LC sequence derived from US 2008213289A1, Seq9), C-terminally modified with sequences comprising or not comprising a 2 amino acid GS (glycine-serine) spacer, and comprising a LPETG sortaseA recognition motif, and a strep-II purification tag (SEQ ID NOs: 39 - 46), have been cloned essentially according to instructions disclosed in Example 1. Using these expression constructs, mAbs Ac10-HC-GS-LHS/LC-GS-LHS and Ac10-HC-LS/LC-LS were produced in CHO cells by co-transfection of the corresponding plasmids. Ac10-HC-GS-LHS/LC-GS-LHS is an Ac10 variant with heavy and light chains modified at the C-termini of each HC and LC with a GS peptide spacer, a LPETG sortaseA motif, a 6xHis tag, and a strep-II tag (SEQ ID NOs:39 - 42; Table 3). Ac10-HC-LS/LC-LS is an Ac10 variant with heavy and light chains modified at the C-termini with LPETG Sortase motif and strep-

II tag without the 2-peptide GS linker (SEQ ID NOs: 43 - 46; Table 3). CHO cell transfection and affinity purification of antibodies by protein A-sepharose chromatography was done essentially as described in Example 4.

To investigate efficiency of conjugation, serial dilutions of Sortase A were used to conjugate penta-glycine-modified FITC ($Gly_5$-FITC, see Formula 3 below).




**Formula 3: penta-glycine modified FITC ($Gly_5$-FITC)**
**G = glycine residue**

For this, $Gly_5$-FITC was sortaseA conjugated to two Ac10 variants in 1x Sortase buffer (25mM Tris-HCl, pH8.2; 150mM NaCl; 7.5mM $CaCl_2$), as shown in Table 4. After 4h at 42°C, reaction products were analyzed by denaturing, reducing SDS-PAGE gel electrophoresis, and FITC was visualized by placing the gels on a UV box (Fig. 10). Conjugation to the heavy chain was found to be highly efficient irrespective of the presence absence of the GS-linker between heavy chain C-terminus and LPETG Sortase recognition motif. Unexpectedly, sortaseA mediated conjugation to the light chain was significantly less efficient in comparison to sortaseA mediated heavy chain conjugation. Furthermore, it was surprisingly found that coupling efficiency was dramatically affected by the presence or absence of the 2 peptide GS (glycine-serine) spacer positioned between the C-terminus of the antibody light chains and the LPETG sortaseA recognition motif. Whereas in the presence of the GS-linker, conjugation to the light chain took place with about 5-10x lower efficiency than to the heavy chain, it was about 50-100x less efficient in the absence of a linker. Therefore, it was concluded that increasing the peptide spacer length between the light chain and the LPETG Sortase recognition motif might further improve conjugation efficiency.

Therefore, the influence of increasing the length of the peptide spacer between light chain and LPETG Sortase A recognition motif on conjugation efficacy was investigated next. Expression constructs encoding mAb Ac10 light chains, C-terminally tagged with LPETG Sortase recognition motif and strep-II purification tag, with a 2 to 5 amino acid linker (SEQ ID NOs: 47 - 54), were generated essentially as described in Example 1. Using these expression constructs, mAbs Ac10-HC-LS/LC-GS-LS, Ac10-HC-LS/LC-GGS-LS, Ac10-HC-LS/LC-GGGS-LS and Ac10-HC-LS/LC-GGGGS-LS were produced in CHO cells by co-transfection of the corresponding expression constructs. In each of these antibodies, the heavy chain is C-terminally modified with an LPETG Sortase recognition motif and a strep-II purification tag (SEQ ID NOs: 43 - 44; Table 3). The light chain is C-terminally modified with an LPETG Sortase tag and strep-II tag containing either a GS, GGS, GGGS, or a GGGGS peptide spacer (SEQ ID NOs: 47 - 54; Table 3) in front of the LPETG motif. CHO cell transfection and affinity purification of antibodies by protein A-sepharose chromatography was done essentially as described in Example 4.

To investigate conjugation efficiency, serial dilutions of Sortase A were used to conjugate penta-glycine-modified FITC ($Gly_5$-FITC, see Formula 3, above) to the four different Ac10 mAb variants in 1x Sortase buffer (25mM Tris-HCl, pH8.2; 150mM NaCl; 7.5mM $CaCl_2$), as shown in Table 5. After 4h at 42°C, reaction products were analyzed by denaturing, reducing SDS-PAGE gel electrophoresis, and FITC was visualized by placing the gels on a UV box (Figure 11). As expected, conjugation to the heavy chain was equally efficient in all four antibody variants. In contrast, conjugation to the light chain was improved significantly by increasing peptide-spacer length. Significantly, with the longest peptide-spacer analyzed (GGGGS), light chain conjugation efficiency was equally efficient in comparison to conjugation of the heavy chain, thereby allowing synchronous conjugation of heavy and light chains of an antibody C-terminally modified at both heavy and light chain. It is concluded that this antibody format will facilitate Sortase A-mediated production of homogeneous ADCs loaded with 4 drugs per antibody (DAR4).

| Antibody | Heavy Chain modification | SEQ ID NOs | Light Chain modification | SEQ ID NOs |
|---|---|---|---|---|
| Ac10-HC-GS-LHS/LC-GS-LHS | GS-LPETG-G-HHHHHH-G-WSHPQFEK | 39,40 | GS-LPETG-G-HHHHHH-G-WSHPQFEK | 41, 42 |
| Ac10-HC-LS/LC-LS | LPETG-G-WSHPQFEK | 43, 44 | LPETG-G-WSHPQFEK | 45, 46 |
| Ac10-HC-LS/LC-GS-LS | LPETG-G-WSHPQFEK | 43, 44 | GS-LPETG-G-WSHPQFEK | 47, 48 |
| Ac10-HC-LS/LC-GGS-LS | LPETG-G-WSHPQFEK | 43, 44 | GGS-LPETG-G-WSHPQFEK | 49, 50 |
| Ac10-HC-LS/LC-GGGS-LS | LPETG-G-WSHPQFEK | 43, 44 | GGGS-LPETG-G-WSHPQFEK | 51, 52 |
| Ac10-HC-LS/LC-GGGGS-LS | LPETG-G-WSHPQFEK | 43, 44 | GGGGS-LPETG-G-WSHPQFEK | 53, 54 |

**Table 3: C-terminally modified monoclonal antibody Ac10 variants produced**

| Reaction component | 1-8 | 9-16 | Final concentration |
|---|---|---|---|
| Ac10-HC-GS-LHS/LC-GS-LHS (3.75mg/ml = 25µM) | 10 | - | 5µM |
| Ac10-HC-LS/LC-LS (3.75mg/ml = 25µM) | - | 10 | 5µM |
| $H_20$ | 20 | 20 | - |
| $Gly_5$-FITC (1mM) | 5 | 5 | 100µM |
| Sortase A (2x serial dil. of ca. 50µM) | 5 | 5 | 5 → 0.039µM |
| 5x Sortase buffer | 10 | 10 | 1x |

**Table 4: Conjugation conditions for mAbs Ac10-HC-GS-LHS/LC-GS-LHS and Ac10-HC-LS/LC-LS**

| Reaction component | 1-7 | 8-14 | 15-21 | 22-28 | Final conc. |
|---|---|---|---|---|---|
| Ac10-HC-LS/LC-GS-LS (3.75mg/ml=25µM) | 10 | - | - | - | 5µM |
| Ac10-HC-LS/LC-GGS-LS (3.75mg/ml=25µM) | - | 10 | - | - | 5µM |
| Ac10-HC-LS/LC-GGGS-LS (3.75mg/ml=25µM) | - | - | 10 | - | 5µM |
| Ac10-HC-LS/LC-GGGGS-LS (3.75mg/ml=25µM) | - | - | | 10 | 5µM |
| $H_20$ | 20 | 20 | 20 | 20 | - |
| $Gly_5$-FITC (1mM) | 5 | 5 | 5 | 5 | 100µM |
| Sortase A (2x serial dil. of ca. 25µM) | 5 | 5 | 5 | 5 | 2.5→0.039µM |
| 5x Sortase buffer | 10 | 10 | 10 | 10 | 1x |

**Table 5: Conjugation conditions for mAbs Ac10-HC-LS/LC-GS-LS, Ac10-HC-LS/LC-GGS-LS, Ac10-HC-LS/LC-GGGS-LS and Ac10-HC-LS/LC-GGGGS-LS.**

**Example 10: Generation of homogeneous ADC by strepII-tag affinity purification**

Sortase A mediated conjugation with $Gly_5$-labeled vc-PAB-MMAE (see Formula 1, Example 5) was performed with anti-CD30 antibody Ac10 modified at the C-termini of either the heavy chains, or the light chains with sequences comprising an LPETG sortase A motif and a strepII-affinity purification tag as provided in Table 6 below:

| Antibody | Heavy Chain modification | SEQ ID NOs | Light Chain modification | SEQ ID NOs |
|---|---|---|---|---|
| Ac10-HC-LS<br>Ac-10-LC | LPETG-G-WSHPQFEK | 43,44 | none | 29, 30 |
| Ac10-HC<br>Ac10-LC-GS-LHS | none | 27, 28 | GS-LPETG-G-HHHHHH-G-WSHPQFEK | 41,42 |

**Table 6: C-terminally modified antibody Ac10 with either HC or LC modification**

The expression vectors encoding the Ac10 heavy or light chain sequences of Table 4 have been constructed essentially as disclosed in Example 1. CHO cell transfection and affinity

purification of antibodies by protein A-sepharose chromatography was done essentially as described in Example 4.

Sortase A mediated conjugation of heavy or light chaing sortase motif tagged anti-CD30 antibodies with $Gly_5$-labeled vc-PAB-MMAE (see Formula 1, Example 5) was performed essentially according to the protocol provided in Example 7.

As described further above in the detailed description of the invention, unreacted antibody will retain the C-terminal strep-II affinity purification tag, which can be exploited to enrich fully reacted ADC with DAR2. Analysis of the heavy chain sortase A conjugation with vc-PAB-MMAE toxin via hydrophobicity interaction chromatography (HIC) (Fig. 12, panel A), shows that the majority of the sortase-motif modified heavy chains have been conjugated, but a certain percentage of unreacted substrate (DAR0 = drug to antibody ratio = zero), or partially reacted substrate (DAR1 = drug to antibody ratio = 1) was still detectable by HIC (Fig. 12, Panel A).

Therefore, the protein A purified vc-PAB-MMAE conjugate was passed 4 times times over a StrepTactin® affinity column (IBA Sciences, Göttingen, Germany), essentially as described in Example 7, in order to remove unreacted or partially reacted sortase A-modified antibody. Fig. 12, Panel B shows that upon several passages of the heterogeneous vc-PAB-MMAE antibody drug conjugate, completely reacted DAR2 ADCs (DAR2 = drug to antibody ratio = 2) could be highly enriched. This experiment demonstrates the feasibility to utilize additional affinity purification tags added C-terminally to the sortase A LPETG recognition motif to generate homogeneous ADC with a defined drugs per antibody ratio (here DAR2).

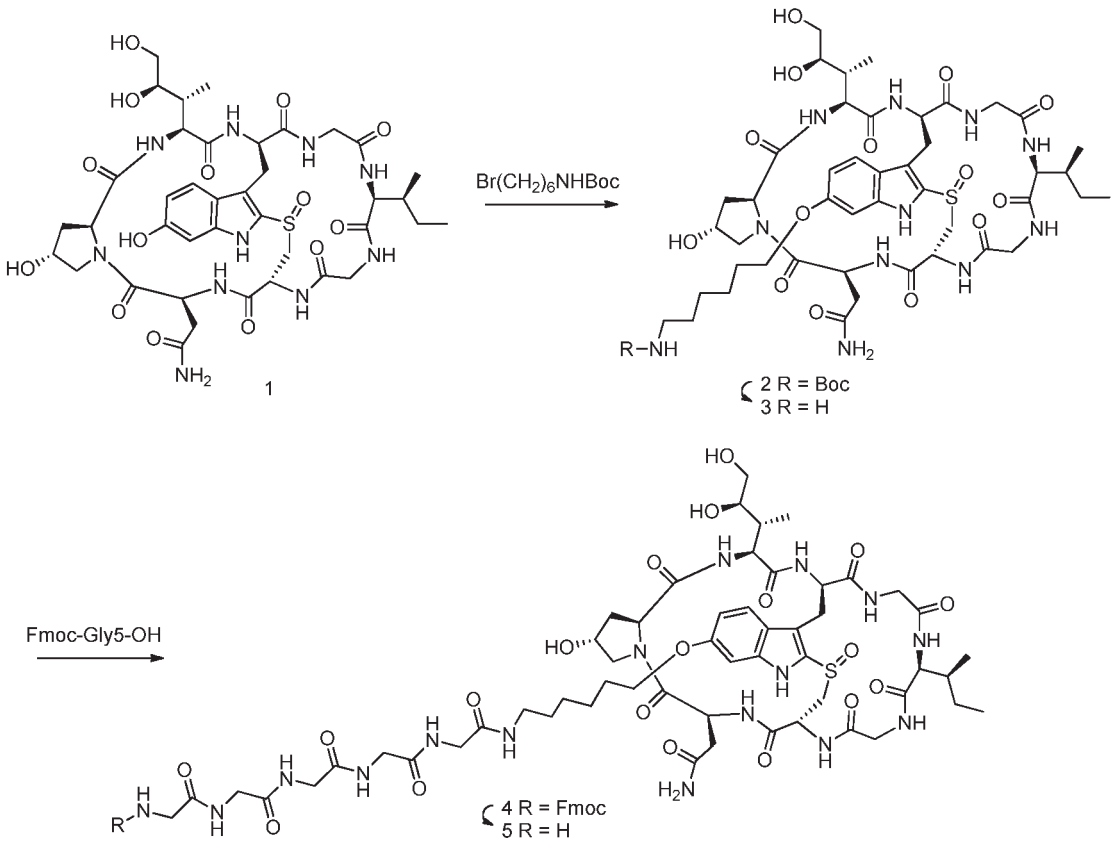
**Example 11: Synthesis of 5xglycine-modified maytansine and alpha-amanitin toxins**

In order to allow conjugation of two different payloads, preferably toxic payloads to a single antibody, modified with different sortase motifs at heavy and light chain C-termini, it is required to modify two different toxins with glycine residues, preferably toxins with different mode of actions, such that a cancer cell targeted with a dual payload conjugated ADC, is attacked with via two different, potentially synergistic routes. The synthesis of two different glycine-modified toxic payloads (here maytansine and alpha-amanitin) satisfying this requirement has been performed and is described herein.

**11.1 Synthesis of glycine-modified alpha-amanitin:**

30mg alpha-amanitin (Structure **1**) (Sigma-Aldrich, order # A2263) was dissolved in 1 ml anhydrous DMSO. To this solution 19mg NH-Boc-amino-hexylbromide were added, followed by potassium tert-butoxide (1M solution in THF, 35 µl). The reaction mixture was stirred at room temperature for 6 h and more potassium tert-butoxide (1M solution in THF, 20 µl) was added. The reaction was kept at room temperature for 16 h. Acetic acid (10 µl) was added and the crude mixture was purified by RP-HPLC directly (Sunfire C18 5µ 3 cm x 10 cm column, 50 mL/min, 5-50% acetonitrile/water 15 min gradient). The desired fraction was collected and lyophilized to give Structure **2** as a white powder (15 mg), which was treated with TFA/DCM solution (1/1, v/v, 1 ml) for 30 minutes at room temperature. The volatiles were removed under reduced pressure to give Structure **3** as a slightly yellowish gum, which was used in the next step without further purification.

Fmoc-Gly5-OH (8 mg) was dissolved in anhydrous DMF (0.5 ml). HATU (Sigma-Aldrich, order # 445460) (6 mg) was added, followed by DIEA (10 ml) (Sigma-Aldrich, order #496219). The mixture was agitated gently at room temperature for 30s and then transferred to a solution of compound **3** in DMF (0.5 ml). After 30 mins, LC/MS analysis showed that all of compound **3** was consumed. Piperidine (30 µl) was added and the progress of the reaction was monitored by LC/MS. Acetic acid was added to neutralize the reaction after 1 h and the mixture was purified by RP-HPLC (Sunfire C18 5µ 3 cm x 25 cm column, 50 mL/min, 2-40% acetonitrile/water 30 min gradient). The fractions were pooled and lyophilized to give structure **5** as a white powder (12 mg). Analytical data for compound **5** is provided in Fig. 13, panel A).



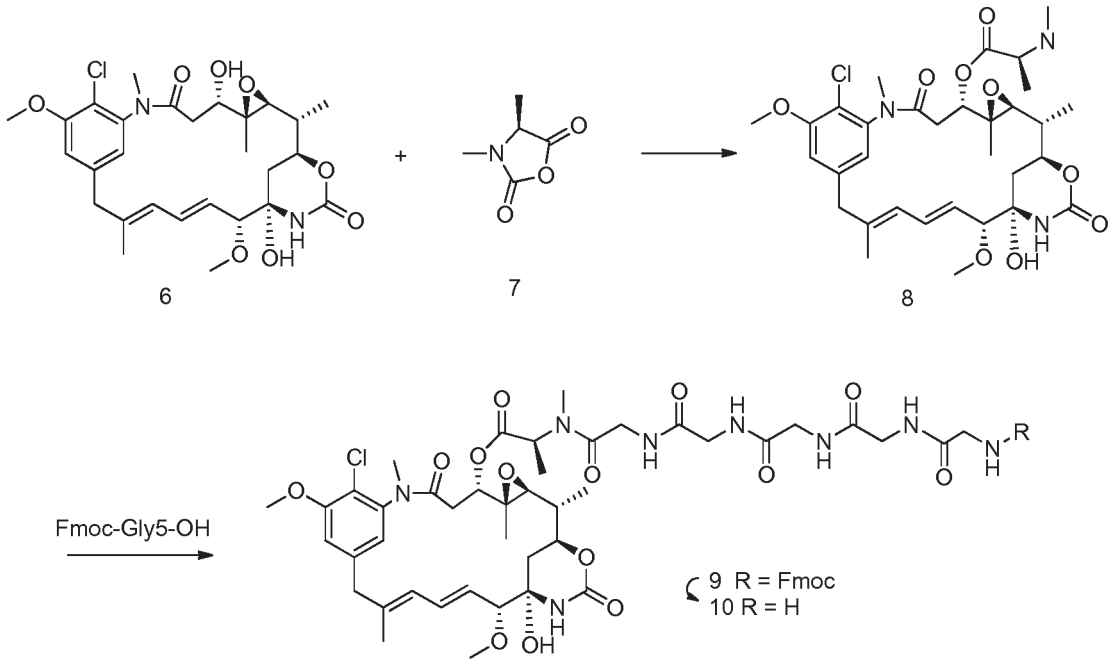

**Scheme 1: Synthesis of glycine-modified alpha-amanitin**

### 11.2. Synthesis of glycine-modified maytansine:

Maytansinol (0.6 g, 1.1 mmol) (Clearsynth Labs, Mumbai, India) was dissolved in anhydrous THF (6 ml) and anhydrous DMF (3 ml) after which 1,2 ml DIEA (Sigma-Aldrich, order #496219) was added. The solution was placed under argon atmosphere. Zinc triflate (1.2 g) and NMeAla NCA (0.7 g) were added in one portion. The mixture was sonicated until the solid was dissolved. The reaction mixture was stirred at room temperature for 2 days and then diluted with ethyl acetate (100 ml). It was washed with saturated $NaHCO_3$ (aq. solution, 2 x 50 ml) and brine (50 ml). The organic layer was dried (over $MgSO_4$) and concentrated to give the crude maytansinol 3-(S)-alpha-N-methylaminopropionate (**8**) which was used directly in the next step without further purification.

Fmoc-Gly5-OH (26 mg) was dissolved in anhydrous DMF (1 ml). HATU (Sigma-Aldrich, order # 445460) (19 mg) was added, followed by DIEA (18 µL). The mixture was agitated gently at room temperature for 30s and then transferred to a solution of compound **8** in THF

(1 ml). After 30 mins, LC/MS analysis showed that all compound **8** was consumed. Piperidine (40 µl) was added and the progress of the reaction was monitored by LC/MS. Ether (40 ml) was added to the reaction after 2 h and the precipitated solid was collected and washed with ether. The crude compound was purified by RP-HPLC (Sunfire C18 5µ 3 cm x 10 cm column, 50 ml/min, 10-60% acetonitrile/water 20 min gradient). The fractions were pooled and lyophilized to give compound **10** as a white powder (33 mg). Analytical data for compound 10 is provided in Fig. 13, Panel B.




**Scheme 2: Synthesis of glycine-modified maytansine**

Importantly, it is to be noted that in principle, any toxin can be functionalized for sortase mediated enzymatic conjugation, if either 5 glycines (as shown here), or any number of glycine residues greater or equal than one glycine, are attached to the toxins (see Fig. 14 ).

**Example 12: *In vivo* tumor inhibition of sortase A-conjugated Trastuzumab-DM1 in SKOV3 ovarial carcinoma xenograft models.**

$5x10^6$ SKOV3 tumor cells in 200µl PBS/Matrigel (1:1 ratio) were implanted subcutaneously into the left flanks of 5-6 weeks old female NMRI nude mice. Primary tumor volumes were

monitored by calipering. After a mean tumor volume of 100-200mm$^3$ was reached, tumor-bearing animals were randomized into 3 Groups according to tumor sizes (10 animals per group). On the day of randomization (day 0) and on day 21, animals of Groups 1, 2 and 3 were injected intravenously with, respectively, 5ml/kg PBS, 15 mg/kg Kadcyla®, or 15 mg/kg sortase A-conjugated Trastuzumab-DM1. Tumor volumes were measured bi-weekly by calipering (Fig. 15). The study was terminated after 39 days and animals were euthanized according to accepted animal experimentation guidelines.

In the course of the study, tumors in control animals mock-injected with PBS grew steadily to a volume of approximately 600mm$^3$. In contrast, tumors in Kadcyla®-treated animals shrank and were essentially undetectable on day 39. Anti-tumor activity of Sortase A-conjugated Trastuzumab-DM1 did not differ significantly from that of commercially available Kadcyla®, despite the fact that the sortase-conjugated T-DM1 exhibited a lower drug to antibody ratio of approximately 2, in comparison of a reported DAR of 3.5 of Kadcyla®. In combination with the data from Example 8, the results demonstrate that sortase conjugated ADCs, using identical antibody and toxin moiety, have comparable tumor killing activity in comparison to commercially available chemically conjugated Kadcyla® *in vitro* and *in vivo*, albeit at lower drug to antibody ratio.

**Example 13: Sortase A-mediated conjugation in crude CHO cell supernatant.**

The Trastuzumab variant Tras-HC-LS/LC-GGGGS-LS, consisting of heavy chains C-terminally tagged with LPETG Sortase motif and Strep II purification tag (SEQ ID NOs: 055-056), and light chains C-terminally tagged with a 5 amino acid $Gly_4$-Ser spacer (GGGGS), LPETG Sortase motif and Strep II tag (SEQ ID NOs: 057-058), was produced in CHO cells essentially as described in Example 4. The resulting serum-free crude cell supernatant contained approximately 157mg/L Tras-HC-LS/LC-GGGGS-LS and was directly used for conjugation essentially as described in Example 9, by adding Sortase buffer, $Gly_5$-FITC, and serial dilutions of Sortase A directly to the supernatant. In parallel, Tras-HC-LS/LC-GGGGS-LS purified by protein A affinity chromatography was also conjugated under otherwise identical conditions. After 4 hours at 42°C, the reactions were analyzed by denaturing, reducing SDS-PAGE gel electrophoresis. After visualizing FITC by placing the gel on a UV box, protein was stained using Coomassie Brilliant Blue (Fig. 16). The data shows the unexpected finding that Sortase A-mediated conjugation of antibodies in crude cell culture supernatant was as efficient as that of purified antibody. Further, the

conjugation reaction was highly specific and none of the protein contaminants present in crude CHO cell supernatant were non-specifically conjugated. Together, these data suggest that the robustness of the Sortase reaction may help facilitate ADC manufacturing by allowing to perform drug conjugation directly after production in CHO cells prior to purification and downstream processing.

## Figure legends

**Fig. 1:** This figure illustrates the principle of the sortase A mediated site-specific payload conjugation to an immunoligand (or binding protein), which can be performed at the N-terminus of a protein (a), or at the C-terminus of the protein (b). In order to achieve N-terminal conjugation, the payload needs to contain a sortase penta-peptide recognition motif (here LPXT**G**, the recognition motif of sortase A from *Staphylococcus aureus* (X representing any of the 20 natural amino acids), whereas the N-terminus of the immunologand/binding protein to be labeled needs to be expressed with an N-terminal extension of minimally 3 glycine residues, here indicated as $G_n$, (with n>2), that has a free N-terminal amino group (here indicated by the smaller $H_2N$- symbol). Typically 3-5 glycines are used in order to modify a substrate for sortase-mediated conjugation. Addition of recombinant sortas A enzyme from *Staphylococcus aureus*, as indicated here, then catalyzes the breakage of the peptide bond between the T and the C-terminal G residue in the LPXT**G** penta-peptide motif and forms a new peptide bond between the N-terminal glycine of the $G_n$ stretch (n>2) and the T residue. The C-terminal G residue of the LPXT**G** motif (here highlighted in boldface print) is removed in the transpeptidation reaction. (b) Conversely, in order to achieve C-terminal conjugation of a payload to a protein, which is the preferred method for conjugation of payloads, particularly toxins, to antibodies (see Fig. 6), the LPXT**G** sortase recognition penta-peptide motif needs to be added to the C-terminal end of the immunoligand/binding protein (e.g. by recombinant protein expression technology, as described in the Examples), and the payload needs to be modified with a short glycine stretch ($G_n$, with n>2, typically 3-5 glycines). As described under (a), addition of sortase A from *Staphylococcus aureus* will then catalyze the transpeptidation of the Gn-stretch to the LPXT**G** motif, whereby the terminal G residue of the LPXTG motif (in boldface) will be removed.

**Fig. 2:** This figure illustrates the principle of intein (a) and split-intein (b) mediated transpeptidation. (a) Inteins can occur as so-called "protein-introns" in precursor proteins, where they separate N-terminal and C-terminal parts of a mature protein, which are generally called N-extein and C-extein. The intein "protein-intron" can catalyze the breakage of the peptide bond between the intein and the C-extein and the formation of a new peptide bond between the N-extein and C-extein by transfering the N-terminal amino acid of the C-extein to the C-terminal amino acid of the N-extein in a transpeptidation reaction. The result of the reaction is the removal of the intein "protein-intron" from the precursor protein and the generation of a mature protein with a newly created peptide bond between the N-extein and C-intein domains. (b) The intein activity has also been described to be separable into distinct domains, that can be attached to different proteins, for which this intein variation has been termed split-intein. The N-int and C-int domains of the split intein form a non-covalent structural complex, that can perform the same transpeptidation reaction as a contiguous intein, on the attached N-extein and C-extein domains that are then in spatial proximity and part of the complex. The result of the transpeptidation of N-int and C-int split-intein reaction is then a "protein trans-splicing", or essentially a protein ligation between the N-extein and C-extein domains, by formation of a novel peptide bond.

**Fig. 3:** This figure illustrates how particular split inteins that are characterized by either an extremely short C-int domain or an extremely short N-int domain can be used to conjugate any payload to an immunoligand (or binding protein), including small molecular entities, because short amino acid stretches can be synthesisze chemically and can easily be attached to small molecular entities by conventional chemical coupling. (a) This part of the illustration shows the use of the Ssp GyrB S11 split intein (described in Appleby et al. (2009)) for the C-terminal conjugation of a payload to an immunoligand/binding protein. Here the C-int domain is only 6 amino acids long and comprises the amino acid sequence GVFVHN, as indicated. However, as there need to be some peptides that are the equivalent of an C-extein domain, additional amino acids need to be added, of which the first one needs to be a serine or cysteine amino acid residue, whereas the remaining amino acids can be chosen. This is indicated by the $SX_n$ symbol, which means that a short amino acid stretch lead at the N-terminal side by serine and followed by n amino acids (n>2, preferably 5), which can be any of the 20 naturally occuring amino acids (therefore indicated as X). Thus, as described in the Example, a short 12 amino acid stretch comprising a 6 amino acid mini C-int domain and 6 amino acid C-ext amino acid stretch are sufficient to allow the N-int/C-

int complex to catalyze the transpeptidation from the asparagine-serine peptide bond in the GVFVHN-$SX_n$ (X any amino acid, n>2, preferably 5) to the peptide bond between the N-extein and N-int transition. This will result in a C-terminally conjugated immunologand/binding protein with the payload attached via the short C-extein amino acid stretch. (b) This part of the illustration shoas the use of the Ssp DnaX split intein (described in Song et al. (2012)), which can be separate into a very short, 11 amino acid N-int domain and a 139 aa C-int domain for N-terminal conjugation of a payload to an immunologand/binding protein. As indicated here, this only requires the synthesis and coupling of a short 11 amino acid N-int domain to any payload (or the addition by recombinant protein technology), which then allows the specific conjugation of the payload to the N-terminus of any immunoligand or protein, that has a 139 amino acid long Ssp DnaX C-int domain fused to the N-terminus. The result of this reaction is then a N-terminally conjugated immunoligand/binding protein. Therefore, like in the case of sortase transpeptidation, where the N- or C-terminal conjugation only depends on the arrangement of the LPXTG and $G_n$ peptide motifs with regard to protein and payload, split inteins can also mediate site-specific N- and C-terminal conjugation of proteins with short peptide modified payloads, and by exploit short mini C-int, or mini N-int peptide domains, like those of Ssp GyrB and Ssp DnaX split inteins, respectively.

**Fig. 4:** (a) This Figure illustrates the utility of adding additional affinity purification and/or detection tags in addition to a sortase tag in the conjugation of payloads to immunoligands. (a) this part of the Figure shows how an additionally added amino acids representing a 6xHis purification tag (HHHHHH), a Myc-detection tag (EQKLISEEDL) and a strepII affinity purification tag (WSHPQFEK), as described in the Examples are removed in the course of the C-terminal payload conjugation via *Staphylococcus aureus* sortase A transpeptidase. This allows to select for the conjugated product, if Ni-NTA affinity resins (for the 6xHis-tag) or streptactin afifnity resins (for the strep II-tag) are employed to seperate non-conjugated substrate from conjugated product. This combination of tags is only provided by way of Example.

(b) This Figure illustrates that the use of affinity purification tags is particularly useful to select/purifiy completely conjugated product in the case of multimeric proteins, like antibodies as illustrated here. As also provided in the examples, antibodies can be modified with specific conjugations sites at heavy and light chains, and if the modification is targeted

to the C-termini of IgH and IgL chains, then up to four payloads may be conjugated to the antibody. The addition of (a) further affinity purification tag(s), e.g. as described in Fig 4(a) allows to bind incompletely conjugated product, that may only have one, two, or three (as illustrated here) payloads conjugated to the antibody, still bind to the respective affinity purification resin, and can thus easily be separated from the fully payload-conjugated product. This paradigm is of course also applicable to intein-modified immunologands, and not only to sortase-motif-modified immunologands, as depicted here.

**Fig. 5:** This figure illustrates a variation of the sortase-mediated conjugation that can also be applied, in which the sortase-enzyme is not added as a separate recombinant protein to the sortase tagged immunologand and glycine-stretch modified payload, but where the enzymatic sortase domain is expressed as a fusion protein C-terminal to the LPXTG sortase tag. The sortase enzyme domain will be inactive as long as it is not incubated with glycine-stretch modified payload (or substrate). As soon as glycine-stretch modified substrate (or here payload) is added to such a construct, the fused sortase domain will catalyze the transpeptidation of glycine-payload substrate to the LPXTG sortase tag, by cleaving the protein between the threonine-4 and glycine-5 position of the LPXTG tag, and thereby removing the sortase enzyme domain with additional affinity purification tags, that can be added optionally, as depicted here. This procedure has the advantage that, similar to the addition of catalytically active split-intein domains, the sortase enzyme domain can be expressed by recombinant protein technology as an integral component of the immunoligand to be conjugated.

**Fig. 6:** (a) This figure illustrates the use of different transpeptidases (here sortase and split-intein), in order to simultaneously conjugate different payloads to different subunits of a multimeric protein, like e.g., as depicted here, the heavy and the light chains of an antibody. In this selected example, the C-termini of the heavy chains are modified with the N-int domain of Ssp GyrB (as provided in Example 2), while the light chains are modified with the sortase A penta-peptide motif LPXTG (as provided in Example 1, the additional tags are omitted for simplicity). Incubation with a glycine-stretch modified payload A and with a C-int-domain modified payload B and sortase enzyme will allow the simultaneous and selective conjugation of payload B to the heavy chains and payload A to the light chains. If payloads A and B are toxins addressing different cellular pathways, this strategy could generate more potent anti-cancer drugs, as conventional ADCs, only containing a single

toxin moiety. (b) This figure illustrates the use of different sortase enzymes (here sortase A and sortase B from *Staphylococcus aureus*), in order to simultaneously conjugate different payloads to different subunits of a multimeric protein, like e.g., as depicted here, the heavy and the light chains of an antibody. In this selected example, the C-termini of the heavy chains are modified with the pentapeptide recognition motif for sortase B, NPQTN, while the light chains are modified with the sortase A penta-peptide motif LPXTG. Sequential conjugation of glycine-stretch modified payloads A and B with sortase A and sortase B will allow the simultaneous and selective conjugation of payload B to the heavy chains and payload A to the light chains (remaining peptide sequences from LPXTG and NPQTN are omitted in the conjugated structure for simplicity). If payloads A and B are toxins addressing different cellular pathways, this strategy could generate more potent anti-cancer drugs, as conventional ADCs, only containing a single toxin moiety.

**Fig. 7:** SDS-PAGE (a.) and Western-blot (b.) analysis of recombinant enzymatically active sortase A fragment of *Staphylococcus aureus*. (a.) Lane 1 in the SDS-PAGE contains BSA (ca. 66.4 kD), Lane $M_1$ contains protein molecular weight standard of Genscript (Cat.-Nr.: MO0505), Lane 2 contains His-tag purified recombinant sortase A fragment of *Staphylococcus aureus.* The proteins in the SDS-PAGE are stained with Commassie blue. (b.) The Western-blot was developed with an anti-His antibody (Genscript Cat.-Nr.: AO0186). Lane 3 contains His-tag purified recombinant sortase A fragment of *Staphylococcus aureus*. Lane $M_2$ contains molecular weight standard of Genscript (Cat.-Nr.: MM0908).

**Fig. 8:** Hydrophobic Interaction Chromatography (HIC) analysis of DM1-toxin conjugated Tras-HC-GS-LHS (A) and Tras-LC-GS-LHS (B). DAR1 indicates drug to antibody ratio of 1; DAR2 indicates a drug to antibody ratio of 2.

**Fig. 9:** Dose response of cytotoxic effects of the indicated ADCs on HER2-overexpressing SKBR3 (A) and HER2-negative T47D-5R cells (B). Cells were incubated with serial dilutions of ADCs for 3 days, after which cell viability was detected by CellTiter-Glo® Luminescent Solution (Promega). LC: DM1-sortaseA-conjugated Tras-LC-GS-LHS; HC: DM1-sortaseA-conjugated Tras-HC-GS-LHS.

**Fig. 10:** Sortase A-mediated conjugation of $Gly_5$-FITC to mAb Ac10 variants with or without GS peptide spacer. Serial dilutions of Sortase A were used to conjugate $Gly_5$-FITC to mAb Ac10-HC-GS-LHS/LC-GS-LHS (A) and mAb Ac10-HC-LS/LC-LS (B) under otherwise identical conditions. Reaction products were seperated by size on denaturing, reducing SDS-PAGE gels. FITC was visualized by placing the gels on a UV box. Sortase A concentrations used were: lanes 1, 9: 50µM; lanes 2, 10: 25µM; lanes 3, 11: 12.5µM; lanes 4, 12: 6.25µM; lanes 5, 13: 3.13µM; lanes 6, 14: 1.56µM; lanes 7, 15: 0.78µM; lanes 8, 16: 0.39µM.

**Fig. 11:** Influence of peptide spacer length on light chain conjugation efficiency. Serial dilutions of Sortase A were used to conjugate $Gly_5$-FITC to mAbs Ac10-HC-LS/LC-GS-LS (A, left), Ac10-HC-LS/LC-GGS-LS (A, right), Ac10-HC-LS/LC-GGGS-LS (B, left) and Ac10-HC-LS/LC-GGGGS-LS (B, right) under otherwise identical conditions. Reaction products were seperated by size on denaturing, reducing SDS-PAGE gels. FITC was visualized by placing the gels on a UV box. Sortase A concentrations used were: lanes 1, 8, 15, 22: 25µM; lanes 2, 9, 16, 23: 12.5µM; lanes 3, 10, 17, 24: 6.25µM; lanes 4, 11, 18, 25: 3.13µM; lanes 5, 12, 19, 26: 1.56µM; lanes 6, 13, 20, 27: 0.78µM; lanes 7, 14, 21, 28: 0.39µM

**Fig. 12:** Analysis of sortaseA vc-PAB-MMAE toxin heavy-chain-conjugated ADC of mAb Ac10 by hydrophobicity interaction chromatography (HIC), which is able to differentiate unreacted substrate (DAR0 = 0 drug to antibody ratio), substrate in which one of the two heavy chains has been conjugated (DAR1 = 1 drug to antibody ratio), and substrate in which both modified heavy chains have been conjugated (DAR2 = 2 drugs to antibody ratio), as indicated. Panel A shows the HIC profile after a standard sortase A mediated conjugation of HC modified Ac10 mAb, in which still DAR0 and DAR1 species are detectable, next to the desired DAR2 product. Panel B shows the HIC profile after 4 passes of the ADC preparation analyzed in Panel A over a StrepTactin® affinity purification column.

**Fig. 13:** Analysis of synthesized $Gly_5$-modified alpha-amanitin toxin (a.) and $Gly_5$-modified maytansin toxin (b.). In each of the panels a.) and b.) the snthesized structure is provided on top, with the five glycines highlighted by a box. The analysis of each compound by mass spectrometry and reverse-phase HPLC is provided below. a.) The expected mass of the $Gly_5$-modified alpha-amanitin toxin is 1302.07 D, the observed mass is 1325.38 D, corresponding to Ms + $Na^+$. The RP-HPLC profile indicates a purity of > 95%. b.) The

expected mass of the $Gly_5$-modified maytansine toxin is 991.41 D, the observed mass is 957.69 D, corresponding to Ms + $Na^+$. The RP-HPLC profile indicates a purity of > 95%.

**Fig. 14:** Structures of 5xGlycine ($Gly_5$) modified toxins that either have been synthesized by Concortis, San Diego, CA, U.S. (structures 1-6, and 9), or that can be synthesized (structures 7 & 8), demonstrating that any toxin can be functionalized for sortase mediated enzymatic conjugation, if either 5 glycines are attached to the toxins (as shown here), or any number of glycine residues greater or equal than one glycine. Glycine-modified toxins can either be synthesized containing additional validated linker/spacer structures as provided in structures 1-3 in Fig. 14 a.), potentially adding certain additional functionality (e.g. cleavability in certain subcellular compartments) or without additional linkers, as depicted in structures 4-6 in Fig. 14 b.). If several reactive groups are available at a given toxin, like e.g. in the case of alpha-amanitin toxin, glycine residues can be added to these different groups as exemplified in structures 7-9 in Fig. 14c.).

**Fig. 15:** Tumor volumes as determined in Example 12. The results demonstrate that sortase conjugated ADCs, using identical antibody and toxin moiety, have comparable tumor killing activity in comparison to commercially available chemically conjugated Kadcyla®.

**Fig. 16:** Gels stained with Coomassie blue as described in example 13 The data shows the unexpected finding that Sortase A-mediated conjugation of antibodies in crude cell culture supernatant was as efficient as that of purified antibody.

## References


Antos et al. (2009a) J. Am. Chem. Soc. 131, pp. 10800-10801

Antos et al. (2009b) J. Biol. Chem. 284, 16028-16036

Appleby et al. (2009) JBC 284, 6194-99

Axup et al. (2012) Proc. Natl. Acad. Sci USA 109, 16102-16106

Elleuche (2010) Appl. Microbiol. Biotechnol. 87, 479-489

Graus-Porta et al. (1995) Mol. Cell. Biol. 15, p1182ff

Hofer et al. (2009) Biochemistry 48, 12047-57

Junutula et al. (2008) Nat. Biotechol., 26, 925-932

Lambert (2012) British J Clin Pharmacol 76, 248-262,

Lemke (2011) Methods Mol. Biol. 751, 3-15

Levary et al. (2011) PLoS One 6, e18342

Madej et al. (2012) Biotechnol. Bioeng. 109, 1461-1470

Mao et al. (2004) J. Am. Chem. Soc. 126, 2670-2671,

Mazmanian et al. (1999) Science 285, 760-763

McDonagh et al. (2006) Prot. Engin. Design Selection 19, 299-307

Möhlmann et al. (2011) Chembiochem. 12, 1774-1780,

Mullard (2013) Nature Rev. Drug Discov. 12, 329-332).

Parthasarathy et al. (2007) Bioconjugate Chem. 18, 469-476

Perler (2002) Nucl. Acids Res. 30, 383-384

Song et al. (2012) PLoS One 7, e45355

Spirig et al. (2011) Molecular Microbiol. 82, 1044-1059

Sun et al. (2004) J. Biol. Chem. 279, 35281-35286

Swee et al. (2013) Proc. Natl. Acad. Sci USA 110, 1428-1433

Ton-That et al. (1999) Proc. Natl. Acad. Sci USA 96, 12424-12429

Tsukiji (2009) Chembiochem. 10, 787-798)

Volkmann et al. (2009) PLoS One 4, e8381

Xu et al. (1993) Cell 75, 1371-1377

**Claims**

1. A method of producing an immunoligand/payload conjugate, the method comprising
a) providing an immunoligand selected from the group consisting of:

(i) an antibody,
(ii) an antibody-based binding protein being a protein containing at least one antibody-derived VH, VL, or CH immunoglobulin domain,
(iii) an antibody fragment binding to a receptor, antigen, growth factor, cytokine and/or hormone, and/or
(iv) an antibody mimetic selected from the group consisting of DARPins, C-type lectins, A-domain proteins of S. aureus, transferrins, lipocalins, 10th type III domains of fibronectin, Kunitz domain protease inhibitors, affilins, gamma crystallin derived binders, cysteine knots or knottins, thioredoxin A scaffold based binders, nucleic acid aptamers, artificial antibodies produced by molecular imprinting of polymers, and stradobodies, and

b) enzymatically conjugating at least one payload to the immunoligand by a sequence-specific sortase or a catalytic domain thereof, wherein the payload is a toxin having a molecular weight not exceeding 2,500 Dalton,
wherein either

- the immunoligand comprises a sortase recognition motif and the toxin is modified with a Glyn-modification, wherein n>1, or
- the toxin comprises a sortase recognition motif and the immunoligand is modified with a Glyn-modification, wherein n>1,

thereby producing the immunoligand/payload conjugate.

2. The method of claim 1, wherein the immunoligand binds at least one entity selected from the group consisting of a receptor, an antigen, a growth factor, a cytokine, and a hormone.

3. The method of claim 1, wherein at least one catalytic domain of the sortase is fused to the C-terminus of either the immunoligand or the payload.

4. The method of claim 1, wherein comprising conjugating said immunoligand to at least two

different payloads, at least one of which is the toxin, and wherein said immunoligand has at least two subunits each being conjugated to a payload.

5. The method of claim 1, wherein said immunoligand comprises two subunits, at least one of which has a peptide spacer appended to the C-terminus thereof, said spacer comprises at least two amino acids.

6. The method of claim 5, wherein the spacer is 2-5 amino acids.

7. The method of claim 1, wherein the ratio of immunoligand and payload is stoichiometrically defined.

8. The method of claim 7, wherein said method comprises removal of partially reacted C-terminally modified immunoligand substrate.

9. The method of claim 1, wherein the method comprises an affinity purification step to separate incomplete conjugates from complete conjugates using an affinity purification tag which is attached to the immunoligand via a sortase recognition motif.

10. The method of claim 9, wherein the immunoligand contains at least two affinity purification tags.

11. The method of claim 1, wherein the conjugation of the payload to the immunoligand is site-specific.

12. The method of claim 1, wherein the conjugation comprises a transpeptidation of the sortase recognition motif.

13. The method of claim 1, wherein the sortase recognition motif is LPXTG (SEQ ID NO:27) or NPQTN (SEQ ID NO:28).

14. The method of claim 1, wherein the conjugation is performed in crude cell culture supernatant.