

Herein is reported an enzymatic ligation method wherein a first polypeptide comprising a sortase motif and a second polypeptide that has a cysteine residue at its N-terminus are incubated with a sortase A which results in the formation of a thioester between the first and second polypeptide.

5 **Background of the Invention**

Sortase A (SrtA) is a membrane bound enzyme which attaches proteins covalently to the bacterial cell wall. The specific recognition motif on the SrtA substrate is LPXTG, whereby the enzyme cleaves between the residues threonine and glycine. The recognition motif on the peptidoglycan is a pentaglycine motif. It has been
10 shown that a triglycine and even a diglycine motif on the N-terminus is sufficient to support the SrtA reaction (Clancy, K.W., et al., Peptide science 94 (2010) 385-396). The reaction proceeds through a thioester acyl-enzyme intermediate, which is resolved by the attack of an amine nucleophile from the oligoglycine, covalently linking peptidoglycan to a protein substrate and regenerating SrtA. SrtA can be
15 used to covalently conjugate chemically synthesized peptides to recombinantly expressed proteins.

In WO 2010/087994 methods for ligation and uses thereof are reported. Recombinant approaches to IgG-like bispecific antibodies are reported by Marvin, J.S., et al. (Acta Pharmacol. Sinica 26 (2005) 649-658). In WO 2013/003555 the
20 use of sortases to install click chemistry handles for protein ligation is reported.

Strijbis, K. et al (Traffic 13 (2012) 780-789) report protein ligation in living cells using sortase. It has been stated by them that the Ca^{2+} -dependent *S. aureus* sortase A is not functional intracellularly, but that the Ca^{2+} -independent *S. pyogenes* sortase A is functional in the cytosol and endoplasmic reticulum (ER) lumen of
25 both *Saccharomyces cerevisiae* and mammalian HEK293T cells.

Levary, D.A., et al., report protein-protein fusion catalyzed by Sortase A (PLOS ONE 6 (2011)). Engineering of an anti-epidermal growth factor receptor antibody to single chain format and labeling by sortase A-mediated protein ligation is reported by Madej, M.P., et al. (Biotechnol. Bioeng. 109 (2012) 1461-1470). Ta,
30 H.T., et al., report enzymatic single-chain antibody tagging as a universal approach to targeted molecular imaging and cell homing in cardiovascular diseases (Cir. Res.

109 (2011) 365-373). Popp, M., et al., report making and breaking peptide bonds - protein engineering using sortase (Angew. Chem. Int. Ed. Eng. 50 (2011) 5024-5032). Engineered proteins with high affinity for DOTA chelates are reported in WO 2010/099536.

5 Different efforts to block the revers reactions of Sortase have been reported. Yamamura, Y., et al. (Chem. Commun. 47 (2011) 4742-4744) reported enhancement of sortase A-mediated protein ligation by inducing a beta-hairpin structure around the ligation site by introducing a β -hairpin around the recognition site of the substrate.

10 Sorting of LPXTG peptides by archetypal sortase A, role of invariant substrate residues in modulating the enzyme dynamics and conformational signature of a productive substrate was reported by Biswas, T., et al. (Biochem. 53 (2014) 2515-2524).

15 Li, Y.M., et al. report irreversible site-specific hydrazinolysis of proteins by use of Sortase (Angew. Chem. Int. Ed. Engl. 53 (2014) 2198-2202).

A remarkable development in protein conjugation was Native Chemical Ligation (NCL) (Dawson, P. E., et al., Science 266 (1994) 776-779), which allowed the ligation of two unprotected peptides. Prerequisite for NCL is a thioester on the N-terminal peptide. The synthesis of thioester is an elaborate work and additionally
20 thioesters are not stable for long time especially under aquatic conditions. The side specific in situ production of thioesters at C-termini of peptides or even proteins would enhance the power of NCL (Dawson, P. E. and Kent, S. B. Annual review of biochemistry 69 (2000) 923-960; Aimoto, S., Tanpakushitsu kakusan koso. Protein, nucleic acid, enzyme 52 (2007) 1804-1805).

25 WO 2013/016653 reports methods for detecting the concurrent presence of at least two targets within a biological sample. The method includes contacting said biological sample with a first binding agent, said first binding agent operably linked to a first sortase molecule, wherein said first binding agent specifically binds to a first target; contacting said biological sample with a second binding agent, said
30 second binding agent operably linked to a first sortase recognition sequence peptide, wherein said second binding agent specifically binds to a second target; adding a sortase substrate under conditions where a first sortase-mediated ligation of the sortase substrate to the first sortase recognition sequence will produce a ligation product, and detecting the ligation product, wherein detection of said ligation

product indicates the concurrent presence of the first target and the second target in the biological sample.

Sortase-mediated ligations for the site-specific modification of proteins is reported by Schmohl, L. and Schwarzer, D. (Curr. Opin. Chem. Biol. 22 (2014) 122-128).
5 Race, P.R., et al. (J. Biol. Chem. 284 (2009) 6924-6933) reported the crystal structure of streptococcus pyogenes sortase and implications for sortase mechanism.

Ling and co-workers showed the introduction of a thioester comprising group via a sortase (Ling, J. J. J., et al., J. Am. Chem. Soc. 134 (2012) 10749-10752).

Summary

10 It has been found that sortase A accepts as nucleophile a polypeptide comprising at its N-terminus a cysteine amino acid residue. The resulting enzymatic conjugation product that is released from the enzyme is a thioester, which can undergo a rearrangement thereafter. It has further been found that site specific, in situ
15 generated C-terminal thioester between the sortase motif and the Sortase itself can be used for native chemical ligation (NCL).

As reported herein is a method for the enzymatic formation/production of a thioester (formation of a new thioester bond between the alpha carbonic acid group of an amino acid and the thiol group of a cysteine) comprising the following step

- incubating

- 20 i) a first polypeptide comprising (optionally within the 100 C-terminal amino acid residues) the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue),
- ii) a second polypeptide that has at its N-terminus a cysteine amino acid residue or is a cysteinyl compound (i.e. a compound that comprises a
25 cysteine amino acid residue with free alpha amino group, e.g. as NH₂ or NH₃⁺, and a carboxy group at position 1, which is part of a peptide bond), and
- iii) a third polypeptide that is a sortase A or a catalytically active fragment thereof (i.e. that has sortase A activity),

and thereby producing a thioester.

One aspect as reported herein is a method for the enzymatic formation of a thioester bond comprising the following step

- incubating

- 5 i) a first polypeptide comprising the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue),
- ii) a second polypeptide that has an amino acid sequence CGGG (SEQ ID NO: 02) or CAAA (SEQ ID NO: 03) at its N-terminus, and
- 10 iii) a third polypeptide that is a sortase A or a catalytically active fragment thereof,

and thereby forming a thioester bond.

In one embodiment the third polypeptide is derived from *Staphylococcus aureus* sortase A or *Listeria monocytogenes* Sortase A.

15 In one embodiment the method is for the enzymatic conjugation of two polypeptides.

In one embodiment the second polypeptide has at its N-terminus a cysteine amino acid residue followed by one to three glycine or alanine amino acid residues. In one preferred embodiment the second polypeptide has the amino acid sequence CGG, CGGG (SEQ ID NO: 02), CAA or CAAA (SEQ ID NO: 03) at its N-terminus.

20 In one embodiment the incubating is further in the presence of a thiol additive. In one embodiment the thiol additive is selected from the group consisting of thiophenol, 4-mercaptophenylacetic acid (MPAA), 2-mercaptoethanesulfonate (MESNA) and combinations thereof. In one preferred embodiment the incubating is in the presence of 2-mercaptoethanesulfonate.

25

In one embodiment the first polypeptide comprises at its C-terminus the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue). In one embodiment the first polypeptide comprises at its C-terminus the amino acid sequence LPETG (SEQ ID NO: 04).

- 5 In one embodiment the first polypeptide and the second polypeptide are independently of each other selected from an antibody variable domain, an antibody heavy chain Fab-fragment, an antibody Fc-region, a tag, and a peptide comprising the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue), a linker and a non-sortase motif moiety.
- 10 In one embodiment the third polypeptide has the amino acid sequence of SEQ ID NO: 05 or SEQ ID NO: 06.

Detailed Description of the Invention

The current invention is based at least in part on the finding that a sortase A accepts as nucleophile a polypeptide that has a cysteine residue at its N-terminus.

I. DEFINITIONS

5 In the present specification and claims the numbering of the residues in an immunoglobulin heavy chain Fc-region is that of the EU index of Kabat (Kabat, E.A., et al., Sequences of Proteins of Immunological Interest, 5th ed., Public Health Service, National Institutes of Health, Bethesda, MD (1991), NIH Publication 91-3242, expressly incorporated herein by reference).

10 The term "a cysteinyl compound" denotes a compound that comprises a cysteine amino acid residue with free alpha amino group, e.g. as NH_2 or NH_3^+ , and a carboxy group at position 1 that is in/part of a peptide bond with another moiety, whereby the moiety can be any amino group containing moiety, such as an isolated amino acid residue, a peptide, a polypeptide, a protein, a small molecule, a dye or a (chemical or peptidic) linker.

15 The term "alteration" denotes the mutation, addition, or deletion of one or more amino acid residues in a parent amino acid sequence, e.g. of an antibody or fusion polypeptide comprising at least an FcRn binding portion of an Fc-region, to obtain a variant antibody or fusion polypeptide.

20 The term "amino acid mutation" denotes a modification in the amino acid sequence of a parent amino acid sequence. Exemplary modifications include amino acid substitutions, insertions, and/or deletions. In one embodiment the amino acid mutation is a substitution. The term "amino acid mutations at the position" denotes the substitution or deletion of the specified residue, or the insertion of at least one amino acid residue adjacent the specified residue. The term "insertion adjacent to a specified residue" denotes the insertion within one to two residues thereof. The insertion may be N-terminal or C-terminal to the specified residue.

25 The term "amino acid substitution" denotes the replacement of at least one amino acid residue in a predetermined parent amino acid sequence with a different "replacement" amino acid residue. The replacement residue or residues may be a "naturally occurring amino acid residue" (i.e. encoded by the genetic code) and selected from the group consisting of: alanine (Ala); arginine (Arg); asparagine (Asn); aspartic acid (Asp); cysteine (Cys); glutamine (Gln); glutamic acid (Glu); glycine (Gly); histidine (His); isoleucine (Ile); leucine (Leu); lysine (Lys); methionine (Met); phenylalanine (Phe); proline (Pro); serine (Ser); threonine (Thr); tryptophan (Trp); tyrosine (Tyr); and valine (Val). In one embodiment the replacement residue is not cysteine. Substitution with one or more non-naturally

30

occurring amino acid residues is also encompassed by the definition of an amino acid substitution herein. A “non-naturally occurring amino acid residue” denotes a residue, other than those naturally occurring amino acid residues listed above, which is able to covalently bind adjacent amino acid residues(s) in a polypeptide chain. Examples of non-naturally occurring amino acid residues include norleucine, ornithine, norvaline, homoserine, alpha-amino isobutyric acid and other amino acid residue analogues such as those described in Ellman, et al., Meth. Enzym. 202 (1991) 301-336. To generate such non-naturally occurring amino acid residues, the procedures of Noren, et al. (Science 244 (1989) 182) and/or Ellman, et al. (supra) can be used. Briefly, these procedures involve chemically activating a suppressor tRNA with a non-naturally occurring amino acid residue followed by in vitro transcription and translation of the RNA. Non-naturally occurring amino acids can also be incorporated into peptides via chemical peptide synthesis and subsequent fusion of these peptides with recombinantly produced polypeptides, such as antibodies or antibody fragments.

The term “amino acid insertion” denotes the incorporation of at least one additional amino acid residue into a predetermined parent amino acid sequence. While the insertion will usually consist of the insertion of one or two amino acid residues, the present application contemplates larger “peptide insertions”, e.g. insertion of about three to about five or even up to about ten amino acid residues. The inserted residue(s) may be naturally occurring or non-naturally occurring as defined above.

The term “amino acid deletion” denotes the removal of at least one amino acid residue at a predetermined position in an amino acid sequence.

Within this application whenever an amino acid alteration is mentioned it is a deliberated amino acid alteration and not a random amino acid modification.

The term “tag” denotes a sequence of amino acid residues connected to each other via peptide bonds that has specific binding properties. In one embodiment the tag is an affinity or purification tag. In one embodiment the tag is selected from Arg-tag, His-tag, Flag-tag, 3xFlag-tag, Strep-tag, HA-tag, Nano-tag, SBP-tag, c-myc-tag, S-tag, SNUT-Tag, NusA, T7, thioredoxin, calmodulin-binding-peptide, cellulose-binding-domain, chitin-binding-domain, GST-tag, or MBP-tag (see, e.g., Amau, J., et al., Prot. Expr. Purif. 48 (2006) 1-13).

In one embodiment the tag is selected from SEQ ID NO: 07 (RRRRR), or SEQ ID NO: 08 (RRRRRR), or SEQ ID NO: 09 (HHHHHH), or SEQ ID NO: 10

(KDHLIHNVHKEFHAAHANK), or SEQ ID NO: 11 (DYKDDDDDK), or SEQ ID NO: 12 (DYKDHDGDYKDHDIDYKDDDDDK), or SEQ ID NO: 13 (AWRHPQFGG), or SEQ ID NO: 14 (WSHPQFEK), or SEQ ID NO: 15 (MDVEAWLGAR), or SEQ ID NO: 16 (MDVEAWLGARVPLVET), or SEQ ID NO: 17 (MDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP), or SEQ ID NO: 18 (EQKLISEEDL), or SEQ ID NO: 19 (KETAAAKFERQHMDSD), or SEQ ID NO: 20 (KRRWKKNFIAVSAANRFKKISSSGAL), or SEQ ID NO: 21 (cellulose binding domain), or SEQ ID NO: 22 (cellulose binding domain), or SEQ ID NO: 23 (TNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEP SNVPALWQLQ), or SEQ ID NO: 24 (GST-tag), or SEQ ID NO: 25 (MBP-tag).

An "effective amount" of an agent, e.g., a pharmaceutical formulation, refers to an amount effective, at dosages and for periods of time necessary, to achieve the desired therapeutic or prophylactic result.

The term "individual" or "subject" denotes a mammal. Mammals include, but are not limited to, domesticated animals (e.g. cows, sheep, cats, dogs, and horses), primates (e.g., humans and non-human primates such as monkeys), rabbits, and rodents (e.g., mice, rats, and hamsters). In certain embodiments, the individual or subject is a human.

The term "pharmaceutical formulation" refers to a preparation which is in such a form as to permit the biological activity of an active ingredient contained therein to be effective, and which contains no additional components which are unacceptably toxic to a subject to which the formulation would be administered.

A "pharmaceutically acceptable carrier" refers to an ingredient in a pharmaceutical formulation, other than an active ingredient, which is nontoxic to a subject. A pharmaceutically acceptable carrier includes, but is not limited to, a buffer, excipient, stabilizer, or preservative.

The term "position" denotes the location of an amino acid residue in the amino acid sequence of a polypeptide. Positions may be numbered sequentially, or according to an established format, for example the EU index of Kabat for antibody numbering.

As used herein, "treatment" (and grammatical variations thereof such as "treat" or "treating") refers to clinical intervention in an attempt to alter the natural course of the individual being treated, and can be performed either for prophylaxis or during

the course of clinical pathology. Desirable effects of treatment include, but are not limited to, preventing occurrence or recurrence of disease, alleviation of symptoms, diminishment of any direct or indirect pathological consequences of the disease, preventing metastasis, decreasing the rate of disease progression, amelioration or palliation of the disease state, and remission or improved prognosis. In some embodiments, antibodies of the invention are used to delay development of a disease or to slow the progression of a disease.

The term "sortase A enzymatic activity" as used herein denotes a polypeptide that shows transpeptidation activity in a Reporter immobilization assay (i.e. an assay according to Example 4; and as reported in (WO 2016/096740 incorporated herein by reference).

II. NATIVE CHEMICAL LIGATION

The term native chemical ligation (NCL) denotes a synthetic protocol in which a large polypeptide is constructed by the assembly of two or more not-protected small polypeptide fragments, i.e. the reactive groups of the small polypeptides not intended to react in this reaction are not masked by a protection group (see e.g. Kemp, D.S., Biopolymers 20 (1981) 1793-1804; Schnolzer, M. and Kent, S. B., Science 256 (1992) 221-225; Dawson, P.E., et al., Science. 266 (1994) 776-779). NCL is especially useful for the synthesis of polypeptides of 300 amino acid residues or less.

In the hitherto known chemical form of native chemical ligation a first polypeptide comprising an N-terminal free cysteine residue and a second polypeptide comprising a C-terminal thioester are reacted with each other. In the first step of the reaction the thiol group of an N-terminal cysteine residue as a nucleophile attacks the activated carbonyl group carbon atom of the C-terminal thioester whereby an intermolecular thioester is formed. This reaction is generally performed in a buffer aqueous solution at about neutral pH of 7 in the temperature range of from 20 °C to 37 °C. This first step is reversible but chemo- as well as regioselective. Due to the steric orientation of the intermolecular thioester an intramolecular S,N-acyl shift is possible. The S,N-acyl shift results in the rearrangement of the thioester bond to a natural amide bond (peptide bond) at the conjugation site of the first and second polypeptide.

The first step of the NCL reaction is reversible, especially in the presence of a free thiol as catalyst. But after the S,N-acyl shift the reaction is no longer reversible.

5 The NCL reaction can be performed with high yields even if the first and/or second polypeptide comprise further cysteine residues in their respective amino acid sequence (internal cysteine residues). This is due to the S,N-acyl shift based rearrangement of the labile thioester into a stable amide bond (labile and stable refer to neutral pH conditions in the absence of bond breaking reagents). Other functional groups might be present in the first and second polypeptide, such as e.g. acid groups, basic amino groups, or phenolic hydroxyl groups.

10 The NCL reaction can be catalyzed by thiol additives such as a combination of thiophenol, 4-mercaptophenylacetic acid (MPAA), or 2-mercaptoethanesulfonate (MESNA).

Due to the S,N-acyl shift the primary reaction product is removed from the reaction equilibrium and the reaction generally gives very high yields, frequently even a quantitative conversion of the isolated polypeptides in a conjugate.

15 The resulting polypeptide comprises a cysteine residue at the site of the ligation. This cysteine residue can be modified to give other amino acid residues, such as e.g. alanine (by desulfurization). Beside cysteine also homo-cysteine, seleno-cysteine and beta-thiol amino acids have been used in NCL reactions.

20 Compared to other amide bond formation reaction the NCL reaction has the advantages that it does not require that other reactive groups present in the two polypeptides to be ligated have to be masked/protected, that almost exclusively the intended reaction product is formed despite the low molar concentrations which normally result in by-product formation, and that no racemization prone activation of an acyl group has to be performed.

25 The main drawback of the NCL reaction is the need to provide a not-protected polypeptide with a C-terminal thioester. Such thioesters can be synthesized using a BOC-based approach as the FMOC-based approach requires the use of a nucleophilic base which is not possible if a thioester is to be synthesized. Additionally protective groups that result in the formation of aldehydes or ketones upon their release are also not suitable as the released aldehyde or ketone might
30 react with the N-terminal cysteine residue.

Variants of the native chemical ligation approach (conjugation of modified synthetic polypeptides with recombinantly produced polypeptides) are expressed protein ligation (see e.g. Pellois, J.-P. and Muir, T. W., Curr. Opin. Chem. Biol. 10

(2006) 487; Muir, T. W., et al., Proc. Nat. Acad. Sci. USA 95 (1998) 6705-6710) and intein mediated protein ligation (see e.g. Saleh, L. and Perler, F. B., Chem. Rec. 6 (2006) 183-193; Noren, C. J., et al., Angew. Chem. Int. Ed. 39 (2000) 451-466).

5 Ling and co-workers showed the introduction of a thioester via a sortase (Ling, J. J. J., et al., J. Am. Chem. Soc. 134 (2012) 10749-10752). They used the normal Sortase reaction with a pentaglycine containing a C-terminal thioester. The resulting product has to be purified before using it for NCL.

III. ENZYMATIC CONJUGATION USING SORTASE A

10 A covalent conjugate comprising two in vivo not covalently associated entities can be obtained in vitro by using the enzyme sortase, especially sortase A.

Transamidases in general catalyze the formation of a peptide bond (amide bond) between an acyl donor and a nucleophilic acyl acceptor. In order to form a peptide bond the acyl acceptor has to contain a NH₂-CH₂-moiety. Gram-positive bacteria include the following genera: Actinomyces, Bacillus, Bifidobacterium, 15 Cellulomonas, Clostridium, Corynebacterium, Micrococcus, Mycobacterium, Nocardia, Staphylococcus, Streptococcus and Streptomyces.

Sortases have been classified into 4 classes, designated A, B, C, and D, based on sequence alignment and phylogenetic analysis of 61 sortases from gram-positive bacterial genomes (Dramsi, S., et al., Res. Microbiol. 156 (2005) 289-297). These 20 classes correspond to the following subfamilies, into which sortases have also been classified by Comfort and Clubb (Comfort, D. and Clubb, R.T., Infect. Immun. 72 (2004) 2710-2722): Class A (Subfamily 1), Class B (Subfamily 2), Class C (Subfamily 3), Class D (Subfamilies 4 and 5). The aforementioned references disclose numerous sortases and recognition motifs (see also Pallen, M. J., et al., 25 Trends Microbiol. 9 (2001) 97-101). With this information a person skilled in the art can assign a sortase to the correct class based on its sequence and/or other characteristics such as those described in Dramsi (supra).

Sortase A (SrtA) is a membrane bound enzyme has transamidase activity. It has been identified and isolated from gram-positive bacteria. In vivo Sortase A attaches 30 proteins covalently to the bacterial cell wall. The specific recognition motif on the SrtA substrate is LPXTG, whereby the enzyme cleaves between the residues threonine and glycine. The recognition motif on the peptidoglycan is a pentaglycine motif. It has been shown that a triglycine and even a diglycine motif on the N-

terminus is sufficient to support the SrtA reaction (Clancy, K.W., et al., *Peptide Science* 94 (2010) 385-396). The reaction proceeds through a thioester acyl-enzyme intermediate, which is resolved by the attack of an amine nucleophile from the oligoglycine, covalently linking peptidoglycan to a protein substrate and regenerating SrtA. SrtA can be used to covalently conjugate chemically synthesized peptides to recombinantly expressed proteins.

Many gram-positive bacteria use sortase to covalently anchor a variety of surface proteins including virulence factors to their cell wall (peptidoglycan). Sortases are membrane associated enzymes. The wild-type *Staphylococcus aureus* sortase A (SrtA) is a polypeptide of 206 amino acids with an N-terminal membrane-spanning region. In a first step, sortase A recognizes substrate proteins that contain a LPXTG (SEQ ID NO: 01) amino acid sequence motif and cleaves the amide bond between the Thr and Gly by means of an active-site Cys. This peptide cleaving reaction results in a sortase A-substrate thioester intermediate. In a second step the thioester acyl-enzyme intermediate is resolved by nucleophilic attack of an amino group of an oligoglycine containing second substrate polypeptide (corresponding to the pentaglycine unit of peptidoglycan in *S. aureus*) leading to a covalently linked cell wall protein and the regeneration of sortase A. In the absence of oligoglycine nucleophiles, the acyl-enzyme intermediate can be hydrolyzed by a water molecule.

Sortase-mediated ligation/conjugation has begun to be applied for a variety of protein engineering and bioconjugation purposes. This technique enables the introduction of natural and synthetic functionalities into LPXTG-tagged recombinant or chemically synthesized polypeptides. Examples include the covalent attachment of oligoglycine derivatized polymers (e.g. PEG), fluorophores, vitamins (e.g. biotin and folate), lipids, carbohydrates, nucleic acids, synthetic peptides and proteins (e.g. GFP) (see e.g. Tsukiji, S. and Nagamune, T., *ChemBioChem* 10 (2009) 787-798; Popp, M.W.L. and Ploegh, H.L., *Angew. Chem. Int. Ed. Engl.* 50 (2011) 5024-5032).

year / citation	content
1990 Fischetti et al. Mol. Microbiol. 4 (1990) 1603- 1605	- LPETG sortase motif
1999 Mazmanian et al. Science 285 (1999) 760- 763	- Staphylococcus aureus strain OS2 - conserved Leu-Pro-X-Thr-Gly (LPXTG) motif - srtA gene specifies a protein of 206 amino acids with a potential NH2-terminal signal peptide/membrane anchor sequence and a presumed active-site cysteine at position 184 - srtA homologs are present in Actinomyces naeslundii, Bacillus subtilis, Enterococcus faecalis, Staphylococcus aureus, Streptococcus mutans, Streptococcus pneumoniae, and Streptococcus pyogenes - AF162687 discloses coding sequence (frame 1 reading) MKKWTNRLMT IAGVVLILVA AYLFAKPHID NYLHDKDKDE KIEQYDKNVK EQASKDKKQQ AKPQIPKDKS KVAGYIEIPD ADIKEPVYPG PATPEQLNRG VSFAEENESL DDQNISIAGH TFIDRPNYQF TNLKAAKKGS MVYFKVGNET RKYKMTSIRD VKPTDVGVLVD EQKGKDKQLT LITCDDYNEK TGVWEKRKIF VATEVK
1999 Ton-That et al. Proc. Natl. Acad. Sci. USA 96 (1999) 12424-12429	- Staphylococcus aureus strain OS2 - Staphylococcus aureus Sortase A - residues 2-25 deleted (N-terminal signal sequence) - MW: 22139 Da - mutation C184S abolishes catalytic activity - sortase homologs from Streptococcus pyogenes (Spyo), Enterococcus faecalis (Efea), Actinomyces naeslundii (Anei), Streptococcus mutans (Smut), Bacillus subtilis (Bsub), and Streptococcus pneumoniae (SpnA, SpnB, and SpnC) - water can resolve acyl-enzyme intermediate
2000 Dhar et al. Biochem. 39 (2000) 3725- 3733	- Listeria monocytogenes has a peptidoglycan cross-bridge (m-Dpm) which is chemically distinct and much shorter than the pentaglycine cross-bridge of staphylococci - Listeria monocytogenes Sortase A has the same sortase motif as Staphylococcus aureus Sortase A: LPXTG (LPTTG)

year / citation	content
2000 Ton-That et al. J. Biol. Chem. 275 (2000) 9876- 9881	- purified recombinant <i>Staphylococcus aureus</i> Sortase A - hydrolyzes peptides with LPXTG motif - triple G-motif - in presence of H₂N-GGG exclusive transpeptidation - in the presence of amino-donors sortase mediated LPXTG motif cleavage rate was increased - Sortase is a 206-amino acid polypeptide with an N-terminal signal sequence/stop transfer domain, is anchored in the cytoplasmic membrane of <i>staphylococci</i> - residues 1-25 correspond to N-terminal signal sequence - reaction conditions: - 5 mM amino group nucleophile - 4.71 μM SrtADN - 150 mM NaCl, 5 mM CaCl₂, 50 mM Tris-HCl, pH 7.5) - volume of 520 μl - reaction conditions: - 10 μM fluorescent peptide - 5 mM amino group nucleophile H₂NGGG - 15 μM SrtADN - 150 mM NaCl, 5 mM CaCl₂, 50 mM Tris-HCl, pH 7.5) - volume of 520 μl - 37 °C, 16 h

TABLE III Kinetic analysis of SrtA _{ΔN}				
Kinetic constants K_m , V_{max} , and k_{cat} were calculated from the curve fit for the Michaelis-Menten equation using the Lineweaver-Burk plot. Reaction conditions are described in the legend to Fig. 4.				
Nucleophile	K_m	V_{max}	K_{cat}	K_m/K_{cat}
	μ M	μ M/s	l/s	l/ μ M · s
H ₂ O	10.88	5.08×10^{-5}	1.06×10^{-5}	9.77×10^{-7}
NH ₂ -Gly ₃	16.48	1.08×10^{-4}	2.27×10^{-5}	1.38×10^{-6}

TABLE IV The effect of different nucleophiles on the rate of LPXTG peptide cleavage by sortase (SrtA _{ΔN})	
Nucleophile	M (s ⁻¹) ^a
H ₂ O	1.84 (±0.11)
NH ₂ OH	1.91 (±0.07)
NH ₂ -Gly	1.95 (±0.05)
NH ₂ -Gly ₂	2.03 (±0.13)
NH ₂ -Gly ₃	2.91 (±0.03)

^a Slope of the kinetic curves as shown in Fig. 4. The substrate peptide d-QALPETGEE-e was incubated with SrtA_{ΔN} and various nucleophiles. Substrate cleavage between the threonine and the glycine was measured as an increase in fluorescence. With the exception of water, all nucleophiles were added at a concentration of 5 mM. Averages were calculated from three independent experiments and standard deviations are reported (parentheses).

year / citation	content
2001 Bolken et al. Infect. Immun. 69 (2001) 75-80	- sortase A from <i>Streptococcus gordonii</i> - 252 amino acid residues with N-terminal signal sequence - cysteine at position 210 - 12-amino-acid extension at the carboxy-terminus of the <i>S. gordonii</i> protein compared to <i>S. aureus</i>
2001 Glaser et al. Science 294 (2001) 849-852	- <i>Listeria monocytogenes</i> genome contains 41 proteins containing an LPXTG - <i>Lm</i> genome contains more LPXTG proteins than any other gram-positive bacterium (<i>Strep. pyogenes</i>: 13; <i>Staph. aureus</i>: 18)
2001 Ilangoan et al. Proc. Natl. Acad. Sci. USA 98 (2001) 6056-6061	- NMR structure of <i>Staph. aureus</i> Sortase A - unique b-barrel structure including two short helices and several loops - the active-site sulfhydryl of cysteine-184 is poised for ionization by histidine-120, presumably enabling the resultant thiolate to attack the LPXTG peptide - conservation of H120 and C184 - calcium binding near the active site stimulates catalysis - stimulation by about 2 mM calcium ions - magnesium and manganese ions can substitute for calcium ions - sortase with residues 1-29 deleted - sortase with residues 1-59 deleted - average mass of 16,595.12 Da observed - reaction conditions: - 2 mM H₂N-GGG - 5 μM sortase N-terminal deletion variant - 150 mM NaCl, 5 mM CaCl₂, 50 mM Tris*HCl pH 7.5 - reaction volume 520 μL - <i>S. aureus</i> SrtA residues 26–59 display no amino acid conservation; core SrtA residues 60–206 present in all sortase homologs examined
2001 Mazmanian et al. Mol. Microbiol. 40 (2001) 1049-1057	- <i>Staph. aureus</i> sorting motifs: LPETG, LPDTG, LPKTG, LPNTG, PLAAG, LPAAG, LPKAG, LPQTG - <i>Actinomyces naeslundii</i>, <i>Bacillus anthracis</i>, <i>Bacillus subtilis</i>, <i>Clostridium acetabutylicum</i>, <i>Corynebacterium diphtheriae</i>, <i>Enterococcus faecalis</i>, <i>Listeria monocytogenes</i>, <i>Streptococcus mutans</i>, <i>Streptococcus pneumoniae</i>, <i>Streptococcus pyogenes</i>

year / citation	content
2001 Pallen et al. TRENDS Microbiol. 9 (2001) 97- 101	- LPXTG like motif, followed shortly by a membrane-spanning hydrophobic domain and a charged carboxy-terminal tail SC7A8.19 VAGHV DNA-----EGPAVFYRLGALENGSAIEIDRRDGGV- SCM10.23 VVGHVDNQ-----QGPAVFYGLGALKKGNKVEVHRQOGKT- SC5C11.07 IAGHVDTK-----TSAAVFARLDQLDKGDQFQVRRADGRS- SCE20.15c MVGHVDTE-----TRPAVFYQLSTLEPGQTIRVARDDDEV- BH3596_bachd LSGHRDT-----VFRDMGKLEIGDDLTVMMPYGS-- yhcs_bacsu LSGHRDT-----VFRRTGELEKGDQLRLLLSYGE-- BH4010_bachd LSGHRDT-----VFRLEGEVGVGDLLIVETATGT-- slp2_bacan LSGHRDT-----VFTDLGQLKEKDTLVLEYDNKT-- BH2127_bachd IAAHRSR-----TYGRQFNRLDEVEVGQVITVTNNHM-- slp_cloab LAGHSY-----TFGEYFNRLGEIGSGDEIDVETVNGT-- SCH69.20c LAGHRN-----THGEFTRYINKLEPGDFIVVETQDKY-- SCH69.19c LAAHRD-----GHGARFHNIDKIERGDFIVFETKDTW-- fap2_actna ITGHRGL-----AEATMFNLDKVKITGDSLIVEVFGEV-- slp4_cordi ITAHRGL-----AEATMFNLNKVGVDRTFIEVVGGEV-- slp2_strpn ITAHTGL-----PTAKMFTDLTKLVGDKFYVHNIEK-- slp4_strpn ITAHRGL-----PTAELFSQLDKMKKGDFYLVLDQV-- slp2_enfae ISGHRGL-----PQARLFTDLPELKKGDEFYIEVNGKT-- slp_streq ISGHRGL-----PSAKLFTNIDKLRINDTFTITSLNRT-- slp2_strpy ISAHRGL-----PSAEMFNLNKLVKRGDTFYFRVLNKKV-- slp1_cordi ITGHSGL-----ANATLFDNLLEDVKHDPFYITVQGET-- slp2_cordi ITGHTGL-----ANSTMFDHLNKAERKDTFYVQVSGEK-- slp3_cordi LSAHTGL-----QNATLWDNLIQIKKGDPFVYVAAAGEK-- slp5_cordi LTAHSGI-----QKSTFFDNLEKVKRGDAIYVRNIGET-- BH2015_bachd IAGHRGYRGNR-----HFSRLPDVTIGDEVFLHTKEET-- BH0362_bachd IAGHRGYRGNR-----HFSRLPDVTIGDEVFLHTKEET-- slp_stimu LASHHVFGMTG-----SSQMLFSPLERAKEGMEIYLTDKNKV-- slp1_strpy LASHHIFGITG-----SSQMLFSPLERAQNGMSIYLTDKKEKI-- slp1_strpn LASHHIFGVDN-----ANKMLFSPLDNAKNGMKIYLTDKNKV-- slp1_enfae LASHRTEDGVS-----LFSPLERTKDELIIYITDLSTV-- BH3294_bachd VDHHEGFYYDT-LYNRYDVEVFSAYVTTT--DFYYIETEFPS-I slp3_bacan FMSHRKLYYDT-LFEGYDLEVSUYTTTT--DFYYIETDFSS-I slp_staau YEKHKIIIEFDN-KYGKYQLQVFSAYKTTT--KDNYIRTDFEN-I slp3_strpy FNKHKKEFSIETKTRQKLKINIFACIQTDAFDLSLLFNPIDVDI-- slp4_strpy FNKHNKAIIEKTRKKLTVTIIFACLKTDADFQDLVFNFNNAITN-- slp3_strpn IAGHRAE-----PSHVFFRHLQDLKVGDAIYDNGQEI-- slp1_bacan LAGHNMS-----KKGVLFSDIASLKKGDKIYLYDNENE-- ywpe_bacsu LAGHHLK-----QKNLLFGPLENIKTGAQIVITDFKKD-- slp3_enfae LASHNAG-----YEGLLFTSLNKVSVGDLVXLNDRECHS-- sortase_staau IAGHTFID-----RPNYQFTNLKAAKKGSMVYFKVGNET-- slp_shepu IAGHRDT-----HFAILKGMTVGRRLALQTAAGKE-- SCH22A.15c VLGHVTVG-----RYDGVFRHLAGRR-GERIEARENGT--- slp_clodi IYGHNMKN-----KTMFNNLNKFEDADFFKNNKIKITLNGKE-- MTH1829_meth ILGHRTT-----YSGPFRKIGALRRGDRVIEDASSI

year / citation	content
	-AVFTVDAVEVYAADAFFDEKVGGAAD-----RPELRVITCGGPFYSR-----STGYQGNVV -AVFEIYGVIEVFENNFPGDRVYGSRG-----SPELRVITCGGFTK-----QNGYDGNVV -ATFVVDGLETTAKDEFFSDRVYGDAD-----RFEVRLITCAGDYDE-----KVDYDTDNLV -AEFTVDDVQVLTRDGFDAQAYGPRDTG-----RSELRLITCGGTFDQ-----TTDSYTANVV -YTYRIVDTEIVDAN-----DTSVIRSTAF-----DEVLTSLTCYFFNF-----IGSAPERYIYY -FTYEIVKTKIVDKD-----DTSIITLQHE-----KEELILITCYPFSY-----VGNAPKRYIYY -HTYRVRKVRIVDED-----DRTVIVPKF-----RATLTVSTCTPFD-----IGSAPERYIIE -YTYEIQKIWIHAD-----DRTVIIKKE-----EPILTLITCYPFDY-----IGDAPDRYIIE -YRYTVYSITVVEPT-----NIDILQHDG-----TAPVLTSLITCDFVKDP-----THERLIVQAEI -FKYKYSTKVVLPF-----EVHVLDTQ-----KDFMTLTVCTPIRIA-----THERLIKAKR -FVYKMASILPVTSPS-----NVSULDVPFKQSGFKGPGRYITLITCTPFTS-----KYRMIVWGEM -YVYKTYAVLPETSKY-----NVEVLOGIPKESGKKKAGHYITLITCTPVYTS-----RYRYVWVWEL -LTYRVSTKVVPE-----ETEALRVEE-----GKDLTLVTCTPLGIN-----THERILLTGER -LTTEVRETRVVSFE-----DTRFLQTD-----DRDLVTLVTCTPLGIN-----THERILVTAER -MAYQVDQVKVIEPT-----NFDOLLIVP-----GHDYVTLITCTPYMIN-----THERLLVRGHR -LAYQVDQIVTVEPN-----DFEPVLIQH-----GEDYATLLITCTPYMIN-----SHERLLVRGHR -LAYQVDQIKVIEPT-----DTRDLHIES-----GQDLVTLITCTPYMIN-----SHERLLVRGHR -MTYQVDKIAIVLPD-----DVSLLRIEF-----GKDLVTLVTCTPYGVN-----THERLLVRGHR -LAYQVDQILTVEPD-----QVTSLSGVM-----GKDYATLVTCTPYGVN-----THERLLVRGHR -LKYVDAINVVLPE-----DTRLLAPDF-----NKDQITLITCTPYAVN-----SHERLLVRAHR -LKYVDQIKVVLPT-----ETEDLRPEQ-----GKDYITLITCTPYGIN-----THERLMVRGHR -LKYEVNIEVVTDP-----KTSLLRRTS-----NKDQVTLITCTPYGIN-----THERLIITAER -LKYQVRDIEIIRPA-----EIDRIQPIF-----DRDLTLVTCTPYGIN-----THERLLVTAER -FVYKVTDISIIEPT-----DVDVLDORD-----GKHEITMITCTRSK-----QRVAVRGEL -FVYKVTDISIIEPT-----DVDVLDORD-----GKHEITMITCTRSK-----QRVAVRGEL -YTYVISEVKVTPE-----HVEVIDNRP-----GQNEVTLVTCTDAGAT-----ARTIVBGTYK -YEYIINDVFTVAPE-----RVDVIDDTA-----GLKEVTLVTCTDIEAT-----ERIIVRGELK -YTYEIREVKRVTPD-----RVDEVDDORD-----GVNEITLVTCTEDLAAT-----ERIIVRGDLK -YTYKITSVEKIEPT-----RVELIDDVP-----GQNMITLITCGDLQAT-----TRIAVQGTIA KDDYKAFNLNKKRSV-----VQTNVEVGE-----EDQIITLSTCDYRLDRD-----RGRLLVVEGKL DTEYTSFLERIQEKS-----YKTDITVTA-----GDQIVTLSTCDYALDFE-----AGRLVHAKL DQDYQQFLDETNRKS-----INSDVNVTV-----KDKIMTLSTCEDAYSET-----TKRIVVVAKI -SSKNEFLNHIKQSV-----QYREILTIN-----ESRFVALSTCEDMTT-----DGRIIVIGQI -QDQQQLVDYISKRS-----QKFPVKLKH-----HTKFVAFSTCEMFST-----DNRVIVVGTI -VEYQMMOTEIILPS-----EWEKLESVS-----SKNIMTLITCDFIPTFN-----KRLLVNFER -YEYAVTGVSEVTPD-----KWEVVEDHG-----KDEITLITCVSVKDN-----SKRYVAVGDL -YIYSVTSKDIISEM-----DADVVEETN-----KKEITLITCDKAVKT-----EGRLVVRGEL -FIYKVKQKHVDMT-----DTTMLNLTR-----KPTLTLITCDQATKT-----TGRIVVIAEL -RKYRMTSIRDVKPT-----DVGVLDEQKG-----KDNQTLITCDDYNEK-----TGVWEKRIK -IVYQVVATKVHES-----QTELMAPSD-----DNRLTLITCYPFDALQGVAELEFVVQAVF -TAEFTTAVRTVAKDF-----PTDDVYGVA-----PELRLITCGPRDQGE-----YRDNVIVAEI -FLYDVFSAYIVESDYDLKTNFNHESD-----YQNYINDITSKSLYKSP-----IKVNSNDKI RYIYTVTSNGDDIRWDY-----RTNFVRFSQS-----GDARLMLITCYPFGQK-----KAAWITCKEL Staph. aureus: LPKTG Staph. pyogenes: LPITG Bac. anthracis: LPKTG Bac. subtilis: LPDTA Clos. difficile: SPKTG Clos. acetabutylicum: LPKTG S. coelicolor: LAETG, LAATG, LAHTG, LASTG Arthrobacter sp.: LASTG A. naeslundii: LPLTG A. viscosus: LPLTG, LSRTG S. pneumoniae: LPETG, VPDTG, IPQTG, YPRTG C. diphtheriae: LPMTG, LALTG, LPKTG, LGNTG, LPLTG, LAFTG S. putrefasciens: LPQTS

year / citation	content
2002 Garandeau et al. Infect. Immun. 70 (2002) 1382-1390	- in silico identification of sortase from <i>L. monocytogenes</i> - 222 amino acid residues - TLXTC consensus motif for sortase active site <pre> MLKKTIAAAA LAAGLLLI FS PFIKNGIVKY MSGHETIEQY KASDIKKNNE KDATFDFESV QLPSMTSVIK GAANYDKDAV VGSIAVPSVD VNLLVFKGTN TANLLAGATT MRSDQVMGKG NYPLAGHHMR DESMLFGPIM KVKKGDKIYL TDLENLYEYT VTETKTIDET EVSVIDNTKD ARITLITCDK PTETTKRFVA VGELEKTEKL TKELENKYFP SK </pre>
2002 Bierne et al. Mol. Microbiol. 43 (2002) 869-881	- <i>Listeria monocytogenes</i> - BLAST analysis: one sequence that is 28% identical to <i>S. aureus</i> SrtA and encodes a protein of 222 amino acids with an expected molecular weight of 24.7 kDa - contains a putative signal peptide/transmembrane region, the expected TLXTC sequence and two stretches of 13 and 31 amino acids that are not present in SrtA
2004 Mao et al. J. Am. Chem. Soc. 126 (2004) 2670-2671	- peptide with single aminoglycoside as nucleophile - attack 50 times faster than water - yield not affected by number of glycine residues - 30 % yield after 30 min, 50 % after 6 h, 90 % after 24 h - conjugates D-amino acid containing peptide; rate half that for L-amino acid peptides - conjugation of small molecules coupled to triglycine (folate) - branched peptide as efficiently coupled as linear peptide

year / citation	content
2004 Kruger et al. Biochem. 43 (2004) 1541- 1551	- SrtA in <i>Staphylococcus aureus</i>, <i>Streptococcus gordonii</i>, <i>Listeria monocytogenes</i>, <i>Streptomyces coelicolor</i>, <i>Streptococcus pyogenes</i>, <i>Streptococcus suis</i> - <i>S. aureus</i> two sortase isoforms: SrtA, SrtB (MPQTN motif) - LPXTG motif highly conserved among all gram-positive bacteria - NPQTN motif appears to be conserved only among at least three bacteria containing the heme iron acquisition <i>isd</i> gene locus (<i>Bacillus anthracis</i>, <i>Bacillus halodurans</i>, and <i>S. aureus</i>) - <i>Staphylococcus aureus</i> Sortase A - residues 2-24 deleted - transpeptidase reaction conditions: - 100 μL reaction volume - 150 mM NaCl, 300 mM Tris, 5 mM CaCl₂ (pH 7.5), pentaglycine (2 mM), SrtAΔN24 (840 nM), and 0 to 10 mM Abz-LPETG-Dap(Dnp)-NH₂ - 37 °C, 30 min - transpeptidase reaction conditions: - 100 μL reaction volume - 150 mM NaCl, 300 mM Tris, 5 mM CaCl₂ (pH 7.5) - SrtAΔN24 (15 μM), Gly5 (2 mM), peptide (300 μM) - 37 °C for 30 min - quenched by 1 N HCl (50 μL) - transpeptidase reaction conditions: - 100 μL reaction volume - 150 mM NaCl, 300 mM Tris, 5 mM CaCl₂ (pH 7.5) - SrtAΔN24 (60 μM), Gly5 (2 mM), peptide (300 μM) - 37 °C for 360 min - no reaction with inverted sortase motif or SrtB motif - initial velocity motif: LPXTG (X=any except P,C; X=M fastest) - end point motif: L/M-P-X-A/L/S/T/V-G (X=any except P,C; L better than M; T and A comparable, then S and V and L comparable) - Table 3 sorting signals: LM: IPKTG, IPALG, LAASS, LPATG, LPKAG, LPISS, IPALG, LPKTS

- 19 -

year / citation	content
2007 Parthasarathy et al. Bioconjug. Chem. 18 (2007) 469- 476	- conjugation to surface of non-protein species (polystyrene beads, PEG) - alkylamine as nucleophile - making of cyclic peptides - use of 159 amino acid sortase - transpeptidase reaction conditions: - 50 mM Tris, 150 mM NaCl, pH 8, 0.1% Tween-20, 6 mM CaCl₂, 3 mM beta-mercaptoethanol - 37°C, 3 h - 15 µM eGFP, 10 µM sortase - add 10 mM EDTA to stop the reaction - transpeptidase reaction conditions: - 50 mM Tris, 150 mM NaCl, pH 8 - 37°C, 3 h - 12 µM eGFP-LPETG, 40 µM sortase, 36 µM GGG-eGFP - with beads higher concentration of sortase has to be used for comparable yield - use of a cleavable upstream tag to protect nucleophile (GGG) - enzyme displays a strong preference for glycine in the first position; alanine and valine can apparently substitute for glycine in the second position, although the reaction is not as efficient
2007 Chan et al. PlosOne 11 (2007) e1164	- conjugation to solid supports - tetraglycine beads react faster than diglycine beads which react faster than monoglycine beads - reaction conditions: - 85 µM eGFP-LPETGG-His6 - 40 nM His6-sortase - 50 mM Tris-HCl, 150 mM NaCl, 5 mM CaCl₂, pH 7.5 - sortase amplified for <i>S. aureus</i> genome - N-terminal membrane targeting sequence removed - 30 kDa on SDS gel
2008 Samantaray et al. J. Am. Chem. Soc. 130 (2008) 2132- 2133	- peptide-sugar conjugation - 6-aminohexoses - peptide antibiotic ligation (aminoglycosides) - conjugates between antibiotics and peptides with yields of 35 to 70% for kanamycin class, about 18-30% for ribostamycin class - YALPET-sugar adduct - YALPMTGK- sugar adduct - LPNTG motif with <i>S. aureus</i> sortase and peptide

For the enzymatic conjugation a soluble truncated sortase A lacking the membrane-spanning region (SrtA; amino acid residues 60-206 of *Staphylococcus aureus* SrtA) can be used (SEQ ID NO: 05; see also Ton-That, H., et al., Proc. Natl. Acad. Sci. USA 96 (1999) 12424-12429; Ilangovan, H., et al., Proc. Natl. Acad. Sci. USA 98 (2001) 6056-6061).

The sortase A-mediated reaction results in the ligation of species containing a sortase motif (sequence) with those bearing one or more N-terminal glycine residues. The sortase motif can be the amino acid sequence LPXTG, but can also differ therefrom (see below). However, a drawback of using such sequences as acyl donors is that the transfer of the LPXT unit to a nucleophilic acyl acceptor liberates a stoichiometric amount of a corresponding fragment containing at least one N-terminal glycine residue. The liberated glycine-containing fragment competes with the intended acyl acceptor for the enzymatic intermediate and works against the progress of the enzymatic ligation reaction. Additionally the hydrolytic cleavage of the enzymatic intermediate as well as the LPXTG containing substrate, although a relatively slow process, compete with the reaction. In the beginning of the use of the sortase-mediated reaction useful levels of ligation could only be obtained using concentrations of at least 5 mM of the acyl donor comprising the sortase-motif.

The general sortase-motif has the amino acid sequence LPXT, wherein X can be any amino acid residue, i.e. a naturally occurring amino acid residue or a non-naturally occurring amino acid residue. In some embodiments, X is selected from the group of amino acid residues comprising or consisting of (in one letter code) D, E, A, N, Q, K, and R. In some embodiments, the sortase-motif is selected from the group comprising or consisting of the amino acid sequences LPXT, LPXA, SPXT, LAXT, LSXT, NPXT, VPXT, IPXT, LGXT, and YPXR. In some embodiments, the sortase motif is selected from the group of amino acid sequences consisting of LPST, LPKT, LPIT, LPDT, SPKT, LAET, LAAT, LAET, LAST, LAET, LPLT, LSRT, LPET, VPDT, IPQT, YPRR, LPMT, LPLT, LAFT, and LPQT. In certain embodiments in which sortase A is used, the sortase-motif comprises the amino acid sequence X1PX2X3, wherein i) X1 is selected from the group consisting of the amino acid residues leucine, isoleucine, valine and methionine, ii) X2 is any amino acid, and iii) X3 is selected from the group consisting of threonine, serine and alanine. In specific embodiments, as noted above X1, is leucine and X3 is threonine. In certain embodiments X2 is selected from the group consisting of aspartate, glutamate, alanine, glutamine, lysine and methionine.

In some embodiments the sortase-motif is selected from the group of amino acid sequences comprising or consisting of LPKTG, LPITG, LPDTA, SPKTG, LAETG, LAATG, LAHTG, LASTG, LAETG, LPLTG, LSRTG, LPETG, VPDTG, IPQTG, YPRRG, LPMTG, LPLTG, LAFTG, and LPQTS. In some embodiments of the invention the sortase is a sortase A (SrtA). SrtA recognizes a sortase-motif with the

amino acid sequence LPXTG. Common sortase-motif amino acid sequences are, e.g., LPKTG, LPATG, LPETG and LPNTG. In some embodiments LPETG is used. However, sortase-motifs not in line with this consensus sortase-motif amino acid sequence may also be recognized. For example, in some embodiments the sortase-motif comprises the amino acid residue A rather than the amino acid residue T at position 4, e.g. LPXAG or LPNAG. In some embodiments the sortase-motif comprises the amino acid residue A rather than the amino acid residue G at position 5, e.g. LPXTA or LPNTA. In some embodiments the sortase-motif comprises the amino acid residue G rather than the amino acid residue P at position 2, e.g. LGXTG or LGATG. In some embodiments the sortase-motif comprises the amino acid residue I rather than the amino acid residue L at position 1, e.g., IPXTG or IPNTG or IPETG.

In some embodiments, where the sortase-motif is LPXTG or LPXT, X is selected from the group consisting of D, E, A, N, Q, K, and R. In some embodiments X is selected from the group of amino acid residues consisting of K, E, N, Q, and A in an LPXTG or LPXT motif where the sortase is a sortase A. In one embodiment the sortase-motif is LPET or LPETG or LPETA.

In certain embodiments where sortase A from *Staphylococcus aureus* (Staph. SrtA) is used the sortase-motif has the amino acid sequence LPX1TX2, wherein i) X1 is selected from the group of amino acid residues consisting of D, E, A, N, Q, K, and R, and ii) X2 is selected from the group of amino acid residues consisting of alanine and glycine. In certain embodiments the sortase-motif of Staph. SrtA is LPX1TA. In other embodiments the sortase-motif of Staph. SrtA is LPX1TG. X1 has the meaning as outlined before.

Streptococcus pyogenes sortase A (Strep. SrtA) will accept di-alanine based nucleophiles. This sortase will efficiently cleave the sortase-motif amino acid sequence LPXTA between the threonine and the alanine residue and install modified alanine-based nucleophiles. Strep. SrtA will also recognize and cleave LPXTG motifs, albeit with reduced efficiency.

Staphylococcus aureus sortase A (Staph. SrtA) will not significantly cleave LPXTA motifs or accept alanine based nucleophiles.

In one embodiment, a polypeptide is contacted with Strep. SrtA and an alanine-containing nucleophile. The polypeptide comprises a sortase-motif amino acid sequence that can be recognized by Strep. SrtA at or near its C-terminus and the

nucleophile comprises one or more amino acids capable of serving as nucleophile for a Staph. SrtA-mediated reaction at or near its N-terminus (e.g., (G)_n, where n is between 1 and 10, e.g., between 1 and 5). This leads to the formation of an LPXTA sequence at the reactive site, a motif refractory to cleavage by Staph. SrtA. This allows for example Staph. SrtA to act on the N-terminus without affecting the C-terminal modification installed with Strep. SrtA.

Sortase fragments having sortase transamidation activity can be used in the methods as reported herein. Sortase fragments can be identified by producing fragments of sortase, for example, by recombinant techniques or proteolytic digestion of full length sortase, and determining the rate of peptide bond formation, i.e. ligation. The fragment can comprise about 80 % of amino acid sequence of full-length sortase, about 70 %, about 60 %, about 50 %, about 40 % or about 30 % of the amino acid sequence of full-length sortase such as that of *S. aureus* Sortase A (GenBank Accession number AAD48437). In some embodiments the fragment lacks an N-terminal portion of the full-length sortase amino acid sequence that is not essential to the catalytic activity of sortase, for example the fragment lacks the N-terminal portion extending to the end of the membrane anchor sequence. In some embodiments the fragment comprises the C-terminus of a full-length sortase amino acid sequence. In some embodiments, the fragment comprises the catalytic core region of a sortase. In one embodiment the core region is from about position 60 to about position 206 of SrtA, e.g., *S. aureus* SrtA, or about from position 82 to about position 249 of Strep. SrtA.

Sortases from other organisms also can be utilized in the processes as reported herein. Such sortases often are encoded by nucleotide sequences substantially identical or similar to the nucleotide sequences that encode SrtA. A similar or substantially identical nucleotide sequence may include modifications to the native sequence, such as substitutions, deletions, or insertions of one or more nucleotides. Included are nucleotide sequences that are at least 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, or 85 % or more identical to a native nucleotide sequence, and often are 90 % or 95 % or more identical to the native nucleotide sequence (each identity percentage can include a 1 %, 2 %, 3 % or 4 % variance). One test for determining whether two nucleic acids are substantially identical is to determine the percentage of identical nucleotide positions shared between two nucleic acids.

SrtA nucleotide sequences may be used as "query sequences" to perform a search against public databases to identify related sequences. Such searches can be

performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (J. Mol. Biol. 215 (1990) 403-410). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, word-length = 12 to obtain homologous nucleotide sequences. To obtain gapped alignments for comparison purposes, gapped BLAST can be utilized as described in Altschul, et al. (Nuc. Acids Res. 25 (1997) 3389-3402). When utilizing BLAST and Gapped BLAST programs, default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used (see e.g. www.ncbi.nlm.nih.gov).

A variant amino acid sequence departs from a native amino acid sequence. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, helix-forming properties and/or amphipathic properties and the resulting variants are screened for enzymatic activity with a suitable assay, such as that reported in European patent application EP14198535. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar or non-polar head groups having similar hydrophilicity values include leucine, isoleucine, valine, glycine, alanine, asparagine, glutamine, serine, threonine, phenylalanine, and tyrosine. In certain embodiments, conservative substitutions may be made, according to the following Table. Amino acids in the same block in the second column and in the same line in the third column may be substituted for one another other in a conservative substitution. Certain conservative substitutions are substituting an amino acid in one row of the third column corresponding to a block in the second column with an amino acid from another row of the third column within the same block in the second column.

aliphatic amino acid residues	non-polar	G, A, P
		I, L, V
	polar, non-charged	C, S, T, M
		N, Q
	polar, charged	D, E
		K, R
aromatic		H, F, W, Y

In certain embodiments homologous substitution may occur, which is a substitution or replacement of like amino acids, such as basic for basic, acidic for acidic, polar for polar amino acids, and hydrophobic for hydrophobic, for example. Non-homologous substitutions can be introduced to a native sequence, such as from one

class of residue to another (e. g. a non-hydrophobic to a hydrophobic amino acid), or substituting a naturally occurring amino acid with an unnatural amino acids or non-classical amino acid replacements.

5 In the methods as reported herein the sortase, the sortase-motif comprising polypeptide (i.e. the acyl donor), and the nucleophile (i.e. the acyl acceptor) are incubated together under conditions suitable to effect the formation of a peptide bond between the N-terminal part of the sortase-motif comprising polypeptide and the nucleophile. As used herein, the term "incubating" or grammatical equivalents thereof denotes that the components of the process are brought in close proximity
10 to one another to allow contact between the molecules. Incubating can be done by adding them to one reaction vessel, for example. The components in the system may be mixed in a variety of manners, such as by oscillating a vessel, subjecting a vessel to a vortex generating apparatus, or repeated mixing with a pipette or pipettes, for example. The components may be added in any order to the system.

15 The sortase reaction may be performed in any convenient vessel (e.g., tubes such as microfuge tubes, flask, dish), microtiter plates (e.g., 96-well or 384-well plates), glass slides, silicon chips, filters, or any solid or semisolid support having surface (optionally coated) having molecules immobilized thereon and optionally oriented in an array (see, e.g., US 6,261,776 and Fodor, Nature 364 (1993) 555-556), and
20 microfluidic devices (see, e.g., US 6,440,722; US 6,429,025; US 6,379,974; and US 6,316,781).

The reaction mixture is generally cell free and further does not include bacterial cell wall components or intact bacterial cell walls. In some embodiments, the sortase-motif comprising polypeptide and/or the nucleophile are expressed by one
25 or more recombinant nucleotide sequences in a cell, which nucleotide sequences are integrated into the cell genome or non-integrated (e.g., in a plasmid).

The reaction mixture is maintained at any convenient temperature at which the sortase reaction can be performed. In some embodiments, the sortase reaction is performed at a temperature between and including about 15 °C and about 50 °C. In
30 some embodiments, the sortase reaction is performed at a temperature between and including about 23 °C and about 37 °C. In certain embodiments, the temperature is room temperature (i.e. about 20 °C to 25 °C). The temperature can be optimized by repetitively performing the same sortase reaction at different temperatures and determining ligation rates.

Any convenient volume and component ratio can be used.

5 In certain embodiments, a (molar) ratio of 1:1000 or greater of sortase enzyme to sortase-motif comprising polypeptide is utilized, or a (molar) ratio of 1:1000 or greater of sortase enzyme to nucleophile is utilized. In specific embodiments, ratios of sortase enzyme to sortase-motif comprising polypeptide or enzyme to nucleophile is about 1:1, including 1:2 or greater, 1:3 or greater, 1:4 or greater, 1:5 or greater, 1:6 or greater, 1:7 or greater, 1:8 or greater, and 1:9 or greater.

10 In some embodiments, the sortase-motif comprising polypeptide is present at a concentration ranging from about 10 μ M to about 10 mM. In some embodiments, the sortase-motif comprising polypeptide is present at a concentration ranging from about 100 μ M to about 1 mM. In some embodiments, the sortase-motif comprising polypeptide is present at a concentration ranging from about 100 μ M to about 50 mM. In some embodiments, the sortase-motif comprising polypeptide is present at a concentration ranging from about 200 μ M to about 10 mM. In some
15 embodiments, the sortase-motif comprising polypeptide is present at a concentration ranging from about 200 μ M to about 800 μ M. In some embodiments, the sortase-motif comprising polypeptide is present at a concentration ranging from about 400 μ M to about 600 μ M.

20 In certain embodiments the nucleophile is present in excess with respect to the sortase-motif comprising polypeptide. In certain embodiments, the nucleophile is present in 10-fold excess with respect to the sortase-motif polypeptide. In certain embodiments, the nucleophile is present in 25-fold excess with respect to the sortase-motif polypeptide. In certain embodiments, the nucleophile is present in 50-fold excess with respect to the sortase-motif polypeptide. In certain
25 embodiments, the nucleophile is present in 100-fold excess with respect to the sortase-motif polypeptide. In certain embodiments, the nucleophile is present in 250-fold excess with respect to the sortase-motif polypeptide.

30 In certain embodiments, the nucleophile is present at a concentration ranging from about 1 μ M to about 50 mM. In certain embodiments, the nucleophile is present at a concentration ranging from about 15 μ M to about 1500 μ M. In certain embodiments, the nucleophile is present at a concentration ranging from about 25 μ M to about 1000 μ M. In certain embodiments, the nucleophile is present at a concentration ranging from about 40 μ M to about 250 μ M.

In certain embodiments, the sortase is present at a concentration ranging from about 1 μ M to about 500 μ M. In certain embodiments, the sortase is present at a concentration ranging from about 15 μ M to about 150 μ M. In certain embodiments, the sortase is present at a concentration ranging from about 25 μ M to about 100 μ M. In certain embodiments, the sortase is present at a concentration ranging from about 40 μ M to about 60 μ M.

In certain embodiments, the method is performed in a reaction mixture comprising an aqueous environment. Water with an appropriate buffer and/or salt content often may be utilized. An alcohol or organic solvent may be included in certain embodiments. The amount of an organic solvent often does not appreciably esterify a protein or peptide in the ligation process (e.g., esterified protein or peptide often increase only by 5 % or less upon addition of an alcohol or organic solvent). Alcohol and/or organic solvent contents sometimes are 20 % or less, 15 % or less, 10 % or less or 5 % or less, and in embodiments where a greater amount of an alcohol or organic solvent is utilized, 30 % or less, 40 % or less, 50 % or less, 60 % or less, 70 % or less, or 80 % or less alcohol or organic solvent is present. In certain embodiments, the reaction mixture includes only an alcohol or an organic solvent, with only limited amounts of water if it is present.

In some embodiments, the reaction mixture comprises a buffer. A person skilled in the art will be familiar with a variety of buffers that could be used in accordance with the methods as reported herein. In some embodiments, the buffer solution comprises calcium ions. In certain embodiments, the buffer solution does not contain substances that precipitate calcium ions. In some embodiments, the buffer solution does not include phosphate ions. In some embodiments, the buffer solution does not contain chelating agents.

In some embodiments, the method is performed at a pH value in the range of from 6 to 8.5. In some embodiments, the method is performed at a pH value in the range of from 6 to 8. In some embodiments, the method is performed at a pH value in the range of from 6 to 7.5. In some embodiments, the method is performed at a pH value in the range of from 6.5 to 8.5. In some embodiments, the method is performed at a pH value in the range of from 7 to 8.5. In some embodiments, the method is performed at a pH value in the range of from 7.5 to 8.5. In some embodiments, the method is performed at a pH value in the range of from 7.0 to 8.5. In some embodiments, the method is performed at a pH value in the range of from 7.3 to 7.8.

One or more components of the reaction mixture or the product may be immobilized to a solid support. The attachment between the reaction mixture component and the solid support may be covalent or non-covalent (see, e.g., US 6,022,688 for non-covalent attachments). The solid support may be one or more
5 surfaces of the system, such as one or more surfaces in each well of a microtiter plate, a surface of a glass slide or silicon wafer, BIAcore chip, a surface of a particle, e.g., a bead (see e.g., Lam, Nature 354 (1991) 82-84) that is optionally linked to another solid support, or a channel in a microfluidic device, for example. Types of solid supports, linker molecules for covalent and non-covalent
10 attachments to solid supports, and methods for immobilizing molecules to solid supports are known (see, e.g., US 6,261,776; US 5,900,481; US 6,133,436; US 6,022, 688; WO 2001/18234). Any material may be used, e.g., plastic (e.g., polystyrene), metal, glass, cellulose, gels (e.g., formed at least in part from organic polymers such as PDMS), etc. In some embodiments the solid support is semi-solid
15 and/or gel-like, deformable, flexible, or the like.

Any polypeptide, eventually after introduction of a sortase-motif or an oligoglycine or -alanine, may be used as sortase-motif comprising polypeptide or nucleophile in the methods as reported herein.

Summarizing the above, the first substrate, also denoted as donor, comprises the
20 sortase recognition motif. It is cleaved by the sortase after the threonine residue in the recognition motif. Thereby a C-terminal activated carboxyl group (acyl intermediate) is generated. The second substrate, also denoted as acceptor or nucleophile, provides a free N-terminal amino group. Between the free amino group and the activated carboxyl group a peptide bond is formed in the sortase
25 catalyzed transpeptidation reaction.

Thus, for the enzymatic sortase mediated transpeptidation reaction it is only required that a donor comprising a sortase recognition motif and an acceptor comprising an N-terminal free glycine, alanine, cysteine or an equivalent functional
30 group is incubated with a polypeptide having sortase A catalytic activity. The remainder of the donor as well as of the acceptor does not interfere with the reaction.

Thus, a sortase mediated transpeptidation reaction can be performed with virtually any protein or small molecule independently of each other as donor or acceptor as long as these comprise a pair of sortase recognition sequence and nucleophile.

This is confirmed by the art.

For example, Marraffini et al. (Microbiol. Mol. Biol. Rev. 70 (2006) 192-221) reported that sortase A can be used to incorporate chemicals containing glycine residues with a free amino group to the LPXTG motif of recombinant proteins, i.e. without limitation of the protein. Presented examples are the conjugation of triglycyl-lysine-folate with (GFP or Cre or p27)-LPETG-His6 with high efficiency, the incorporation of the branched peptide AT-P-022 into polypeptides, and the self-cleavage of chimeras of His6-sortase-LPETG-target protein (the fusion cleaves itself once the enzyme has been activated by the addition of calcium and triglycine).

Further, Antos et al. (J. Am. Chem. Soc. 131 (2009) 10800-10801) reported that the transpeptidation reaction catalyzed by sortase A allows site-specific derivatization of proteins with virtually any type of functional material. Target proteins are engineered to contain the recognition site (LPXTG) near their C terminus, thus allowing a trans-acylation reaction in which the residues C-terminal to threonine are exchanged for a synthetic oligoglycine peptide. It is reported that the terminal G residue of the sortase recognition motif can be replaced by a methyl ester without imparting the reaction. In this document nucleophiles comprising either a fluorescent label or a protein were used for the conjugation to cholera toxin B subunit.

Further, Popp et al. (Proc. Natl. Acad. Sci. USA 108 (2011) 3169-3174) reported the use of Sortase for polypeptide cyclization and PEGylation. The method is general and applicable to a wide variety of proteins. The sortase transpeptidase reaction allows facile site-specific PEGylation of multiple distinct proteins, as exemplified using interferon $\alpha 2$, GCSF, and erythropoietin. In all cases tested, the site-specific C-terminal PEGylation proceeded efficiently.

In EP 2 990 423 a self-cleaving sortase construct is reported. In this construct the sortase recognition sequence LPETG and the catalytic sortase domain have been combined in the same molecule. As protein comprising the sortase recognition sequence any protein, such as e.g. a protein selected from the group comprising polymer proteins, glycoproteins, cytokines, growth factor, blood preparations, vaccines, hormones, enzymes, antibodies and parts or fragments thereof (isolated light or heavy chains).

IV. The method as reported herein

It has been found that sortase A accepts as nucleophile a polypeptide comprising at its N-terminus a cysteine amino acid residue. The resulting enzymatic conjugation product that is released from the enzyme is a thioester. It has further been found that site specific, in situ generated C-terminal thioester between the sortase motif and the Sortase itself can be used for native chemical ligation (NCL).

One aspect as reported herein is a method for the enzymatic production (=formation) of a thioester by forming a thioester bond / a method for the enzymatic formation of a thioester comprising the following step

- incubating

- i) a first polypeptide comprising (optionally within the 100 C-terminal amino acid residues) the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue),
- ii) a second polypeptide that has at its N-terminus a cysteine amino acid residue or is a cysteinyl compound, and
- iii) a third polypeptide with sortase A activity,

and thereby producing a thioester.

In one embodiment the cysteinyl compound is a compound that comprises a cysteine amino acid residue with free alpha amino group (in one embodiment a NH_2 or NH_3^+), and a carboxy group at position 1, which is part of a peptide bond.

In one embodiment the polypeptide with sortase A activity is derived from *Staphylococcus aureus* sortase A or *Listeria monocytogenes* Sortase A.

In one embodiment the third polypeptide is a sortase A or a sortase A fragment that has sortase A catalytical activity. In one embodiment sortase A catalytical activity is determined using a bond forming assay. In one embodiment the bond forming assay is the assay according to example 4.

In one embodiment the method is for the enzymatic conjugation of two polypeptides.

In one embodiment the second polypeptide has at its N-terminus a cysteine amino acid residue followed by one to three glycine or alanine amino acid residues. In one preferred embodiment the second polypeptide has the amino acid sequence CGG, CGGG (SEQ ID NO: 02), CAA or CAAA (SEQ ID NO: 03) at its N-terminus.

5 In one embodiment the incubating is further in the presence of a thiol additive. In one embodiment the thiol additive is selected from the group consisting of thiophenol, 4-mercaptophenylacetic acid (MPAA), 2-mercaptoethanesulfonate (MESNA) and combinations thereof. In one preferred embodiment the incubating is in the presence of 2-mercaptoethanesulfonate.

10 In one embodiment the first polypeptide comprises within the 250 C-terminal amino acid residues the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue). In one embodiment the first polypeptide comprises within the 100 C-terminal amino acid residues the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue). In one embodiment
15 the first polypeptide comprises within the 25 C-terminal amino acid residues the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue). In one embodiment the first polypeptide comprises within the 10 C-terminal amino acid residues the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue).

20 In one embodiment the first polypeptide comprises at its C-terminus the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue). In one embodiment the first polypeptide comprises at its C-terminus the amino acid sequence LPETG (SEQ ID NO: 04).

25 In one embodiment the first polypeptide and the second polypeptide are independently of each other selected from an antibody variable domain, an antibody heavy chain Fab-fragment, an antibody Fc-region, a tag, and a peptide comprising the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue), a linker and a non-sortase motif moiety.

30 In one embodiment the third polypeptide has the amino acid sequence of SEQ ID NO: 05 or SEQ ID NO: 06.

The first or second polypeptide

The sortase-motif (amino acid sequence) may be conjugated to or incorporated in, if it is not directly comprised in one of these molecules, a therapeutic agent (drug), a cytotoxic agent (e.g. a toxin such as doxorubicin or pertussis toxin), a fluorophore such as a fluorescent dye like fluorescein or rhodamine, a chelating agent for an imaging or radiotherapeutic metal, a peptidyl or non-peptidyl label, a tag, or a clearance-modifying agent such as various isomers of polyethylene glycol, a peptide that binds to a third component, another carbohydrate or lipophilic agent, or a small molecule, such as e.g. a synthetic small molecule (e.g. acetyl salicylic acid). If the motif is incorporated via conjugation the conjugation can be either directly or via an intervening linker. Furthermore the first and/or second polypeptide can either be recombinantly produced or can be synthetic or semi-synthetic, i.e. recombinantly produced and thereafter chemically modified.

a) Therapeutic agents

The therapeutic agent can be any compound, moiety or group which has a therapeutic effect, such as e.g. an antibody, a cytotoxic or cytostatic compound. The antibody can be a full length or complete antibody or an antigen-binding fragment thereof.

A number of therapeutic antibodies directed against cell surface molecules and their ligands are known, such as Rituxan/MabThera/Rituximab, 2H7/Ocrelizumab, Zevalin/Ibrizumomab, Arzerra/Ofatumumab (CD20), HLL2/Epratuzumab, Inotuzomab (CD22), Zenapax/Daclizumab, Simulect/Basiliximab (CD25), Herceptin/Trastuzumab, Pertuzumab (Her2/ERBB2), Mylotarg/Gemtuzumab (CD33), Raptiva/Efalizumab (Cd11a), Erbitux/Cetuximab (EGFR, epidermal growth factor receptor), IMC-1121B (VEGF receptor 2), Tysabri/Natalizumab (α 4-subunit of α 4 β 1 and α 4 β 7 integrins), ReoPro/Abciximab (gpIIb-gpIIa and α v β 3-integrin), Orthoclone OKT3/Muromonab-CD3 (CD3), Benlysta/Belimumab (BAFF), Tolerx/Oteliximab (CD3), Soliris/Eculizumab (C5 complement protein), Actemra/Tocilizumab (IL-6R), Panorex/Edrecolomab (EpCAM, epithelial cell adhesion molecule), CEA-CAM5/Labetuzumab (CD66/CEA, carcinoembryonic antigen), CT-11 (PD-1, programmed death-1 T-cell inhibitory receptor, CD-d279), H224G11 (c-Met receptor), SAR3419 (CD19), IMC-A12/Cixutumumab (IGF-1R, insulin-like growth factor 1 receptor), MEDI-575 (PDGF-R, platelet-derived growth factor receptor), CP-675, 206/Tremelimumab (cytotoxic T lymphocyte

antigen 4), RO5323441 (placenta growth factor or PGF), HGS1012/Mapatumumab (TRAIL-R1), SGN-70 (CD70), Vedotin(SGN-35)/Brentuximab (CD30), and ARH460-16-2 (CD44).

5 The conjugates obtained with the method as reported herein can be used in the preparation of medicaments for the treatment of e.g. an oncologic disease, a cardiovascular disease, an infectious disease, an inflammatory disease, an autoimmune disease, a metabolic (e.g., endocrine) disease, or a neurological (e.g. neurodegenerative) disease. Exemplary non-limiting examples of these diseases are Alzheimer's disease, non-Hodgkin's lymphomas, B-cell acute and chronic
10 lymphoid leukemias, Burkitt lymphoma, Hodgkin's lymphoma, hairy cell leukemia, acute and chronic myeloid leukemias, T-cell lymphomas and leukemias, multiple myeloma, glioma, Waldenstrom's macroglobulinemia, carcinomas (such as carcinomas of the oral cavity, gastrointestinal tract, colon, stomach, pulmonary tract, lung, breast, ovary, prostate, uterus, endometrium, cervix, urinary bladder, pancreas, bone, liver, gall bladder, kidney, skin, and testes), melanomas, sarcomas, gliomas, and skin cancers, acute idiopathic thrombocytopenic purpura, chronic
15 idiopathic thrombocytopenic purpura, dermatomyositis, Sydenham's chorea, myasthenia gravis, systemic lupus erythematosus, lupus nephritis, rheumatic fever, polyglandular syndromes, bullous pemphigoid, diabetes mellitus, Henoch-Schonlein purpura, post-streptococcal nephritis, erythema nodosum, Takayasu's arteritis, Addison's disease, rheumatoid arthritis, multiple sclerosis, sarcoidosis, ulcerative colitis, erythema multiforme, IgA nephropathy, polyarteritis nodosa, ankylosing spondylitis, Goodpasture's syndrome, thromboangitis obliterans, Sjogren's syndrome, primary biliary cirrhosis, Hashimoto's thyroiditis,
20 thyrotoxicosis, scleroderma, chronic active hepatitis, polymyositis/dermatomyositis, polychondritis, pemphigus vulgaris, Wegener's granulomatosis, membranous nephropathy, amyotrophic lateral sclerosis, tabes dorsalis, giant cell arteritis/polymyalgia, pernicious anemia, rapidly progressive glomerulonephritis, psoriasis, or fibrosing alveolitis.

30 A number of cell surface markers and their ligands are known. For example cancer cells have been reported to express at least one of the following cell surface markers and or ligands, including but not limited to, carbonic anhydrase IX, alpha fetoprotein, alpha-actinin-4, A3 (antigen specific for A33 antibody), ART-4, B7, Ba-733, BAGE, BrE3-antigen, CA125, CAMEL, CAP-1, CASP-8/m, CCCL19,
35 CCCL21, CD1, CD1a, CD2, CD3, CD4, CDS, CD8, CD1-1A, CD14, CD15, CD16, CD18, CD19, CD20, CD21, CD22, CD23, CD25, CD29, CD30, CD32b, CD33,

CD37, CD38, CD40, CD40L, CD45, CD46, CD54, CD55, CD59, CD64, CD66a-e, CD67, CD70, CD74, CD79a, CD80, CD83, CD95, CD126, CD133, CD138, CD147, CD154, CDC27, CDK-4/m, CDKN2A, CXCR4, CXCR7, CXCL12, HIF-l-alpha, colon-specific antigen-p (CSAp), CEA (CEACAM5), CEACAM6, c-met, DAM, EGFR, EGFRvIII, EGP-1, EGP-2, ELF2-M, Ep-CAM, Flt-1, Flt-3, folate receptor, G250 antigen, GAGE, GROB, HLA-DR, HM1.24, human chorionic gonadotropin (HCG) and its subunits, HER2/neu, HMGB-1, hypoxia inducible factor (HIF-1), HSP70-2M, HST-2or 1a, IGF-1R, IFN-gamma, IFN-alpha, IFN-beta, IL-2, IL-4R, IL-6R, IL-13R, IL-15R, IL-17R, IL-18R, IL-6, IL-8, IL-12, IL-15, IL-17, IL-18, IL-25, insulin-like growth factor-1 (IGF-1), KC4-antigen, KS-1-antigen, KS1-4, Le-Y, LDR/FUT, macrophage migration inhibitory factor (MIF), MAGE, MAGE-3, MART-1, MART-2, NY-ESO-1, TRAG-3, mCRP, MCP-1, MIP-1A, MIP-1B, MIF, MUC1, MUC2, MUC3, MUC4, MUC5, MUM-1/2, MUM-3, NCA66, NCA95, NCA90, pancreatic cancer mucin, placental growth factor, p53, PLAGL2, prostatic acid phosphatase, PSA, PRAME, PSMA, P1GF, ILGF, ILGF-1R, IL-6, IL-25, RS5, RANTES, T101, SAGE, S100, survivin, survivin-2B, TAC, TAG-72, tenascin, TRAIL receptors, TNF-alpha, Tn-antigen, Thomson-Friedenreich antigens, tumor necrosis antigens, VEGFR, ED-B fibronectin, WT-1, 17-1A-antigen, complement factors C3, C3a, C3b, C5a, C5, an angiogenesis marker, bcl-2, bcl-6, Kras, cMET, an oncogene marker and an oncogene product (see, e.g., Sensi, et al., Clin. Cancer Res. 12 (2006) 5023-5032; Parmiani, et al, J. Immunol. 178 (2007) 1975-1979; Novellino, et al., Cancer Immunol. Immunother. 54 (2005) 187-207).

Thus, antibodies recognizing specific cell surface receptors including their ligands can be used for specific and selective targeting and binding to a number/multitude of cell surface markers that are associated with a disease. A cell surface marker is a polypeptide located on the surface of a cell (e.g. a disease-related cell) that is e.g. associated with signaling event or ligand binding.

In one embodiment, for the treatment of cancer/tumors multispecific binding molecules/bispecific antibodies are produced that target tumor-associated antigens, such as those reported in Herberman, "Immunodiagnosis of Cancer", in Fleisher (ed.), "The Clinical Biochemistry of Cancer", page 347 (American Association of Clinical Chemists (1979)) and in US 4,150,149; US 4,361,544; and US 4,444,744.

Reports on tumor associated antigens (TAAs) include Mizukami, et al., (Nature Med. 11 (2005) 992-997); Hatfield, et al., (Curr. Cancer Drug Targets 5 (2005)

229-248); Vallbohmer, et al., (J Clin. Oncol. 23 (2005) 3536-3544); and Ren, et al., (Ann. Surg. 242 (2005) 55-63), each incorporated herein by reference with respect to the TAAs identified.

Where the disease involves a lymphoma, leukemia or autoimmune disorder, targeted antigens may be selected from the group consisting of CD4, CD5, CD8, CD14, CD15, CD19, CD20, CD21, CD22, CD23, CD25, CD33, CD37, CD38, CD40, CD40L, CD46, CD54, CD67, CD74, CD79a, CD80, CD126, CD138, CD154, CXCR4, B7, MUC1 or Ia, HM1.24, HLA-DR, tenascin, VEGF, P1GF, ED-B fibronectin, an oncogene, an oncogene product (e.g., c-met or PLAGL2), CD66a-d, necrosis antigens, IL-2, T101, TAG, IL-6, MIF, TRAIL-R1 (DR4) and TRAIL-R2 (DR5).

A number of bispecific antibodies are known directed against two different targets, such as BCMA/CD3, different antigens of the HER family in combination (EGFR, HER2, HER3), CD19/CD3, IL17RA/IL7R, IL-6/IL-23, IL-1-beta/IL-8, IL-6 or IL 6R/ IL-21 or IL-21R, first specificity directed to a glycoepitope of an antigen selected from the group consisting of Lewis x-, Lewis b- and Lewis y-structures, Globo H-structures, KH1, Tn-antigen, TF-antigen and carbohydrate structures of Mucins, CD44, glycolipids and glycosphingolipids, such as Gg3, Gb3, GD3, GD2, Gb5, Gm1, Gm2, sialyltetraosylceramide and a second specificity directed to an ErbB receptor tyrosine kinase selected from the group consisting of EGFR, HER2, HER3 and HER4, GD2 in combination with a second antigen binding site is associated with an immunological cell chosen from the group consisting of T lymphocytes NK cell, B-lymphocytes, dendritic cells, monocytes, macrophages, neutrophils, mesenchymal stem cells, neural stem cells, ANG2/VEGF, VEGF/PDGFR-beta, Vascular Endothelial Growth Factor (VEGF) acceptor 2/CD3, PSMA/CD3, EPCAM/CD3, combinations of an antigen is selected from a group consisting of VEGFR-1, VEGFR-2, VEGFR-3, FLT3, c FMS/CSF1R, RET, c-Met, EGFR, Her2/neu, HER3, HER4, IGFR, PDGFR, c-KIT, BCR, integrin and MMPs with a water-soluble ligand is selected from the group consisting of VEGF, EGF, PIGF, PDGF, HGF, and angiopoietin, ERBB-3/C-MET, ERBB-2/C-MET, EGF receptor 1/CD3, EGFR/HER3, PSCA/CD3, C-MET/CD3, ENDOSIALIN/CD3, EPCAM/CD3, IGF-1R/CD3, FAPALPHA/CD3, EGFR/IGF-1R, IL 17A/F, EGF receptor 1/CD3, and CD19/CD16.

Toxic drug moieties include: (i) chemotherapeutic agents, which may function as microtubule inhibitors, mitosis inhibitors, topoisomerase inhibitors, or DNA

intercalators; (ii) protein toxins, which may function enzymatically; and (iii) radioisotopes.

Exemplary toxic drug moieties include, but are not limited to, a maytansinoid, an auristatin, a dolastatin, a trichothecene, CC1065, a calicheamicin and other enediyne antibiotics, a taxane, an anthracycline, and stereoisomers, isosters, analogs or derivatives thereof.

Protein toxins include diphtheria-A chain, non-binding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain (Vitetta et al (1987) *Science*, 238:1098), abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP -5), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes (WO 93/21232).

Therapeutic radioisotopes include ³²P, ³³P, ⁹⁰Y, ¹²⁵I, ¹³¹I, ¹³¹In, ¹⁵³Sm, ¹⁸⁶Re, ¹⁸⁸Re, ²¹¹At, ²¹²B, ²¹²Pb, and radioactive isotopes of Lu.

The radioisotope or other labels may be incorporated in known ways (Fraker et al (1978) *Biochem. Biophys. Res. Commun.* 80: 49-57; "Monoclonal Antibodies in Immunoscintigraphy" Chatal, CRC Press 1989). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triamine pentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of a radionuclide to the complex (WO 94/11026).

b) Labels

The non-sortase motif moiety can be a label. Any label moiety which can be covalently attached to the sortase amino acid sequence can be used (see e.g. Singh et al (2002) *Anal. Biochem.* 304:147-15; Harlow E. and Lane, D. (1999) *Using Antibodies: A Laboratory Manual*, Cold Springs Harbor Laboratory Press, Cold Spring Harbor, N.Y.; Lundblad R. L. (1991) *Chemical Reagents for Protein Modification*, 2nd ed. CRC Press, Boca Raton, Fla.). The label may function to: (i) provide a detectable signal; (ii) interact with a second label to modify the detectable signal provided by the first or second label, e.g. to give FRET (fluorescence resonance energy transfer); (iii) affect mobility, e.g. electrophoretic mobility or cell-permeability, by charge, hydrophobicity, shape, or other physical parameters, or (iv) provide a capture moiety, e.g. to modulate ionic complexation.

Conjugates comprising a haptenylated label as reported herein may be useful in diagnostic assays, e.g., for detecting expression of an antigen of interest in specific cells, tissues, or serum. For diagnostic applications, a bispecific antibody will be used wherein the first binding specificity binds to a target and the second binding specificity binds to a haptenylated label. The hapten will typically be labeled with a detectable moiety. Numerous labels are available which can be generally grouped into the following categories:

(a) Radioisotopes (radionuclides), such as ^3H , ^{11}C , ^{14}C , ^{18}F , ^{32}P , ^{35}S , ^{64}Cu , ^{68}Ga , ^{86}Y , ^{89}Zr , $^{99\text{Tc}}$, ^{111}In , ^{123}I , ^{124}I , ^{125}I , ^{131}I , ^{133}Xe , ^{177}Lu , ^{211}At , or ^{131}Bi . Radioisotope labeled conjugates are useful in receptor targeted imaging experiments. The antigen (hapten) can be labeled with ligand reagents that bind, chelate or otherwise complex a radioisotope metal using the techniques described in Current Protocols in Immunology, (1991) Volumes 1 and 2, Coligen et al, Ed. Wiley-Interscience, New York, N.Y., Pubs. Chelating ligands which may complex a metal ion include DOTA, DOTP, DOTMA, DTPA and TETA (Macrocyclics, Dallas, Tex.). Radionuclides can be targeted via complexation with the complex as reported herein (Wu et al, Nature Biotechnology 23(9) (2005) 1137-1146). Receptor target imaging with radionuclide labeled complexes can provide a marker of pathway activation by detection and quantification of progressive accumulation of complexes or corresponding therapeutic antibodies in tumor tissue (Albert et al (1998) Bioorg. Med. Chem. Lett. 8:1207-1210).

Metal-chelate complexes suitable as labels for imaging experiments (US 2010/0111856; US 5,342,606; US 5,428,155; US 5,316,757; US 5,480,990; US 5,462,725; US 5,428,139; US 5,385,893; US 5,739,294; US 5,750,660; US 5,834,456; Hnatowich et al, J. Immunol. Methods 65 (1983) 147-157; Meares et al, Anal. Biochem. 142 (1984) 68-78; Mirzadeh et al, Bioconjugate Chem. 1 (1990) 59-65; Meares et al, J. Cancer (1990), Suppl. 10:21-26; Izard et al, Bioconjugate Chem. 3 (1992) 346-350; Nikula et al., Nucl. Med. Biol. 22 (1995) 387-90; Camera et al., Nucl. Med. Biol. 20 (1993) 955-62; Kukis et al., J. Nucl. Med. 39 (1998) 2105-2110; Verel et al., J. Nucl. Med. 44 (2003) 1663-1670; Camera et al., J. Nucl. Med. 21 (1994) 640-646; Ruegg et al, Cancer Res. 50 (1990) 4221-4226; Verel et al., J. Nucl. Med. 44 (2003) 1663-1670; Lee et al, Cancer Res. 61 (2001) 4474-4482; Mitchell, et al., J. Nucl. Med. 44 (2003) 1105-1112; Kobayashi et al Bioconjugate Chem. 10 (1999) 103-111; Miederer et al., J. Nucl. Med. 45 (2004) 129-137; DeNardo et al, Clinical Cancer Research 4 (1998) 2483-90; Blend et al, Cancer Biotherapy & Radiopharmaceuticals 18 (2003) 355-363; Nikula et al J.

Nucl. Med. 40 (1999) 166-76; Kobayashi et al., J. Nucl. Med. 39 (1998) 829-36; Mardirossian et al., Nucl. Med. Biol. 20 (1993) 65-74; Roselli et al., Cancer Biotherapy & Radiopharmaceuticals, 14 (1999) 209-20).

5 (b) Fluorescent labels such as rare earth chelates (europium chelates), fluorescein types including FITC, 5-carboxyfluorescein, 6-carboxy fluorescein; rhodamine types including TAMRA; dansyl; Lissamine; cyanines; phycoerythrins; Texas Red; and analogs thereof. The fluorescent labels can be conjugated to the antigen (hapten) using the techniques disclosed in Current Protocols in Immunology, supra, for example. Fluorescent dyes and fluorescent label reagents include those which
10 are commercially available from Invitrogen/Molecular Probes (Eugene, Oregon, USA) and Pierce Biotechnology, Inc. (Rockford, Ill.).

Detection labels such as fluorescent dyes and chemiluminescent dyes (Briggs et al "Synthesis of Functionalised Fluorescent Dyes and Their Coupling to Amines and Amino Acids," J. Chem. Soc., Perkin-Trans. 1 (1997) 1051-1058) provide a
15 detectable signal and are generally applicable for labeling, especially with the following properties: (i) the labeled conjugate should produce a very high signal with low background so that small quantities of conjugate can be sensitively detected in both cell-free and cell-based assays; and (ii) the labeled conjugate should be photostable so that the fluorescent signal may be observed, monitored
20 and recorded without significant photo bleaching. For applications involving cell surface binding of labeled conjugates to membranes or cell surfaces, especially live cells, the labels should (iii) have good water-solubility to achieve effective conjugate concentration and detection sensitivity and (iv) are non-toxic to living cells so as not to disrupt the normal metabolic processes of the cells or cause
25 premature cell death.

(c) Various enzyme-substrate labels are available or disclosed (see e.g. US 4,275,149). The enzyme generally catalyzes a chemical alteration of a chromogenic substrate that can be measured using various techniques. For example, the enzyme may catalyze a color change in a substrate, which can be measured
30 spectrophotometrically. Alternatively, the enzyme may alter the fluorescence or chemiluminescence of the substrate. The chemiluminescent substrate becomes electronically excited by a chemical reaction and may then emit light which can be measured (using a chemiluminometer, for example) or donates energy to a fluorescent acceptor. Examples of enzymatic labels include luciferases (e.g., firefly luciferase and bacterial luciferase; US 4,737,456), luciferin, 2,3-
35

- dihydrophthalazinediones, malate dehydrogenase, urease, peroxidase such as horseradish peroxidase (HRP), alkaline phosphatase (AP), (3-galactosidase, glucoamylase, lysozyme, saccharide oxidases (e.g., glucose oxidase, galactose oxidase, and glucose-6-phosphate dehydrogenase), heterocyclic oxidases (such as uricase and xanthine oxidase), lactoperoxidase, microperoxidase, and the like. Techniques for conjugating enzymes to polypeptides are described in O'Sullivan et al "Methods for the Preparation of Enzyme-Antibody Conjugates for use in Enzyme Immunoassay", in Methods in Enzym. (ed. by J. Langone & IT Van Vunakis), Academic Press, New York, 73 (1981) 147-166.
- 10 Examples of enzyme-substrate combinations (US 4,275,149; US 4,318,980) include, for example:
- (i) Horseradish peroxidase (HRP) with hydrogen peroxide as a substrate, wherein the hydrogen peroxide oxidizes a dye precursor (e.g., orthophenylene diamine (OPD) or 3,3',5,5'-tetramethylbenzidine hydrochloride (TMB));
- 15 (ii) alkaline phosphatase (AP) with para-nitrophenyl phosphate as chromogenic substrate; and
- (iii) (3-D-galactosidase ((3-D-Gal) with a chromogenic substrate (e.g., p-nitro phenyl-(3-D-galactosidase) or fluorogenic substrate 4-methylumbelliferyl-(3-D-galactosidase.
- 20 The labeled conjugate as reported herein may be employed in any known assay method, such as ELISA, competitive binding assays, direct and indirect sandwich assays, and immunoprecipitation assays (Zola, Monoclonal Antibodies: A Manual of Techniques (1987) pp. 147-158, CRC Press, Inc.).
- 25 Labeled conjugates as reported herein are useful as imaging biomarkers and probes by the various methods and techniques of biomedical and molecular imaging such as: (i) MRI (magnetic resonance imaging); (ii) MicroCT (computerized tomography); (iii) SPECT (single photon emission computed tomography); (iv) PET (positron emission tomography) Tinianow, J. et al, Nuclear Medicine and Biology, 37(3) (2010) 289-297; Chen et al, Bioconjugate Chem. 15 (2004) 41-49;
- 30 US 2010/0111856 (v) bioluminescence; (vi) fluorescence; and (vii) ultrasound. Immunoscintigraphy is an imaging procedure in which conjugates labeled with radioactive substances are administered to an animal or human patient and a picture is taken of sites in the body where the conjugate localizes (US 6,528,624). Imaging

biomarkers may be objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention. Biomarkers may be of several types: Type 0 markers are natural history markers of a disease and correlate longitudinally with known clinical indices, e.g. MRI assessment of synovial inflammation in rheumatoid arthritis; Type I markers capture the effect of an intervention in accordance with a mechanism-of-action, even though the mechanism may not be associated with clinical outcome; Type II markers function as surrogate endpoints where the change in, or signal from, the biomarker predicts a clinical benefit to "validate" the targeted response, such as measured bone erosion in rheumatoid arthritis by CT. Imaging biomarkers thus can provide pharmacodynamic (PD) therapeutic information about: (i) expression of a target protein, (ii) binding of a therapeutic to the target protein, i.e. selectivity, and (iii) clearance and half-life pharmacokinetic data. Advantages of in vivo imaging biomarkers relative to lab-based biomarkers include: non-invasive treatment, quantifiable, whole body assessment, repetitive dosing and assessment, i.e. multiple time points, and potentially transferable effects from preclinical (small animal) to clinical (human) results. For some applications, bioimaging supplants or minimizes the number of animal experiments in preclinical studies.

Peptide labeling methods are well known. See Haugland (2003) *Molecular Probes Handbook of Fluorescent Probes and Research Chemicals*, Molecular Probes, Inc.; Brinkley (1992) *Bioconjugate Chem.* 3:2; Garman, (1997) *Non-Radioactive Labeling: A Practical Approach*, Academic Press, London; Means (1990) *Bioconjugate Chem.* 1:2; Glazer et al *Chemical Modification of Proteins. Laboratory Techniques in Biochemistry and Molecular Biology* (T. S. Work and E. Work, Eds.) American Elsevier Publishing Co., New York; Lundblad, R. L. and Noyes, C. M. (1984) *Chemical Reagents for Protein Modification*, Vols. I and II, CRC Press, New York; Pfeleiderer, G. (1985) "Chemical Modification of Proteins", *Modern Methods in Protein Chemistry*, H. Tschesche, Ed., Walter DeGruyter, Berlin and New York; and Wong (1991) *Chemistry of Protein Conjugation and Cross-linking*, CRC Press, Boca Raton, Fla.); DeLeon-Rodriguez et al, *Chem. Eur. J.* 10 (2004) 1149-1155; Lewis et al, *Bioconjugate Chem.* 12 (2001) 320-324; Li et al, *Bioconjugate Chem.* 13 (2002) 110-115; Mier et al *Bioconjugate Chem.* 16 (2005) 240-237.

c) Linker

The term "linker" denotes a bifunctional or multifunctional moiety which can be used to conjugate (link) a first moiety with a second moiety. Linked conjugates can be conveniently prepared using a linker having two reactive functionalities.

5 In one embodiment, a linker has a reactive site which has an electrophilic group that is reactive to a nucleophilic group present in the sortase amino acid sequence. Useful electrophilic groups include, but are not limited to, another thiol, maleimide and haloacetamide groups (see e.g. conjugation method at page 766 of Klussman et al, Bioconjugate Chemistry 15(4) (2004) 765-773).

10 Examples of thiol-reaction functional groups include, but are not limited to, thiol, maleimide, and alpha-haloacetyl.

The linker may comprise amino acid residues which link the sortase amino acid sequence to the non-sortase motif moiety. The amino acid residues may form a dipeptide, tripeptide, tetrapeptide, pentapeptide, hexapeptide, heptapeptide,
15 octapeptide, nonapeptide, decapeptide, undecapeptide or dodecapeptide unit. Amino acid residues include those occurring naturally, as well as non-naturally occurring amino acid analogs, such as e.g. citrulline or β -amino acids, such as e.g. β -alanine, or ω -amino acids such as 4-amino-butyric acid.

In another embodiment, the linker has a reactive functional group which has a
20 nucleophilic group that is reactive to an electrophilic group present in the sortase amino acid sequence. Useful electrophilic groups include, but are not limited to, aldehyde and ketone carbonyl groups. The heteroatom of a nucleophilic group of a linker can react with an electrophilic group in the sortase amino acid sequence and form a covalent bond to the sortase amino acid sequence. Useful nucleophilic
25 groups on a linker include, but are not limited to, hydrazide, oxime, amino, hydrazine, hydrazine carboxylate, and arylhydrazide. The electrophilic group on an antigen (hapten) provides a convenient site for attachment to a linker.

Typically, peptide-type linkers can be prepared by forming a peptide bond between two or more amino acids and/or peptide fragments. Such peptide bonds can be
30 prepared, for example, according to the liquid phase synthesis method (E. Schroder and K. Lubke "The Peptides", volume 1 (1965) 76-136, Academic Press) which is well known in the field of peptide chemistry.

In another embodiment, the linker may be substituted with groups which modulated solubility or reactivity. For example, a charged substituent such as sulfonate (SO₃⁻) or ammonium or a polymer such as PEG, may increase water solubility of the reagent and facilitate the coupling reaction of the linker reagent with the antigen (hapten) or the drug moiety, or facilitate the coupling reaction depending on the synthetic route employed.

The conjugates comprising a non-sortase motif moiety as reported herein expressly contemplate, but are not limited to, complexes prepared with linker reagents: BMPEO, BMPS, EMCS, GMBS, HBVS, LC-SMCC, MBS, MPBH, SBAP, SIA, SIAB, SMCC, SMPB, SMPH, sulfo-EMCS, sulfo-GMBS, sulfo-KMUS, sulfo-MBS, sulfo-SIAB, sulfo-SMCC, and sulfo-SMPB, and SVSB (succinimidyl-(4-vinylsulfone) benzoate), and including bis-maleimide reagents: DTME, BMB, BMDB, BMH, BMOE, BM(PEO)₃, and BM(PEO)₄, which are commercially available from Pierce Biotechnology, Inc. Bis-maleimide reagents allow the attachment of e.g. a thiol group to a thiol-containing drug moiety, label, or linker intermediate, in a sequential or concurrent fashion. Other functional groups besides maleimide, which are reactive with e.g. a thiol group, include iodoacetamide, bromoacetamide, vinyl pyridine, disulfide, pyridyl disulfide, isocyanate, and isothiocyanate.

Exemplary linker include a valine-citrulline (val-cit or vc) dipeptide linker reagent having a maleimide stretcher and a para-aminobenzylcarbamoyl (PAB) self-immolative spacer, and a phe-lys(Mtr) dipeptide linker reagent having a maleimide Stretcher unit and a p-amino benzyl self-immolative spacer.

Cysteine thiol groups are nucleophilic and capable of reacting to form covalent bonds with electrophilic groups on linker reagents and the non-sortase motif moiety or the sortase amino acid sequence including: (i) active esters such as NHS esters, HOBt esters, haloformates, and acid halides; (ii) alkyl and benzyl halides, such as haloacetamides; (iii) aldehydes, ketones, carboxyl, and maleimide groups; and (iv) disulfides, including pyridyl disulfides, via sulfide exchange. Nucleophilic groups on a haptenylated compound include, but are not limited to: amine, thiol, hydroxyl, hydrazide, oxime, hydrazine, hydrazine carboxylate, and arylhydrazide groups capable of reacting to form covalent bonds with electrophilic groups on linker moieties and linker reagents.

V. RECOMBINANT METHODS

Any polypeptide domain (e.g. a single chain antigen binding polypeptide such as a scFv, a scFab, or a darpin, or a multi chain antigen binding polypeptide such as a dsFv or a Fab) comprising an nucleophilic amino acid sequence at its N-terminus, such as e.g. an oligoglycine motif (GG (SEQ ID NO: 28), GGG (SEQ ID NO: 29), GGGG (SEQ ID NO: 30), GGGGG (SEQ ID NO: 31)), can be expressed und purified from the supernatant of eukaryotic cells (e.g. HEK293 cells, CHO cells). It does not matter if the polypeptide is an isolated polypeptide or comprised in a multimeric or heteromeric entity.

Suitable host cells for cloning or expression/secretion of polypeptide-encoding vectors include prokaryotic or eukaryotic cells described herein. For example, polypeptides may be produced in bacteria, in particular when glycosylation is not needed (see, e.g., US 5,648,237, US 5,789,199 and US 5,840,523, Charlton, Methods in Molecular Biology 248 (2003) 245-254 (B.K.C. Lo, (ed.), Humana Press, Totowa, NJ), describing expression of antibody fragments in *E. coli*). After expression, the polypeptide may be isolated from the bacterial cell paste in a soluble fraction or may be isolated from the insoluble fraction so called inclusion bodies which can be solubilized and refolded to bioactive forms. Thereafter the polypeptide can be further purified.

In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeasts are suitable cloning or expression hosts for polypeptide-encoding vectors, including fungi and yeast strains whose glycosylation pathways have been “humanized”, resulting in the production of a polypeptide with a partially or fully human glycosylation pattern (see e.g. Gerngross, Nat. Biotech. 22 (2004) 1409-1414, and Li, et al., Nat. Biotech. 24 (2006) 210-215).

Suitable host cells for the expression of glycosylated polypeptides are also derived from multicellular organisms (invertebrates and vertebrates). Examples of invertebrate cells include plant and insect cells. Numerous baculoviral strains have been identified which may be used in conjunction with insect cells, particularly for transfection of *Spodoptera frugiperda* cells.

Plant cell cultures can also be utilized as hosts (see, e.g., US 5,959,177, US 6,040,498, US 6,420,548, US 7,125,978 and US 6,417,429 (describing PLANTIBODIESTM technology for producing antibodies in transgenic plants)).

Vertebrate cells may also be used as hosts. For example, mammalian cell lines that are adapted to grow in suspension may be useful. Other examples of useful mammalian host cell lines are the COS-7 cell line (monkey kidney CV1 cell transformed by SV40); the HEK293 cell line (human embryonic kidney); the BHK cell line (baby hamster kidney); the TM4 mouse sertoli cell line (TM4 cells as described, e.g., in Mather, Biol. Reprod. 23 (1980) 243-251); the CV1 cell line (monkey kidney cell); the VERO-76 cell line (African green monkey kidney cell); the HELA cell line (human cervical carcinoma cell); the MDCK cell line (canine kidney cell); the BRL-3A cell line (buffalo rat liver cell); the W138 cell line (human lung cell); the HepG2 cell line (human liver cell); the MMT 060562 cell line (mouse mammary tumor cell); the TRI cell line (e.g. described in Mather, et al., Anal. N.Y. Acad. Sci. 383 (1982) 44-68); the MRC5 cell line; and the FS4 cells-line. Other useful mammalian host cell lines include the CHO cell line (Chinese hamster ovary cell), including DHFR negative CHO cell lines (see e.g. Urlaub, et al., Proc. Natl. Acad. Sci. USA 77 (1980) 4216), and myeloma cell lines such as Y0, NS0 and Sp2/0 cell line. For a review of certain mammalian host cell lines suitable for polypeptide production, see, e.g., Yazaki, and Wu, Methods in Molecular Biology, Antibody Engineering 248 (2004) 255-268 (B.K.C. Lo, (ed.), Humana Press, Totowa, NJ).

Description of the Figures

- Figure 1** Scheme showing the native chemical ligation reaction.
- Figure 2** Scheme showing the use of Sortase in NCL.
- Figure 3** Chromatogram of the reaction product of Example 3.
- Figure 4** Mass-spectrum of the peak of the chromatogram of Figure 3; A:1, B:2, C:3, D:4, E:5, F:6, G:7, H:8.
- Figure 5** Enzymatic activity of Sa-SrtA using a REIA with different Nucleophiles.
- Figure 6** Enzymatic activity of Lm-SrtA using a REIA with different Nucleophiles.

The following examples, figures and sequences are provided to aid the understanding of the present invention, the true scope of which is set forth in the appended claims. It is understood that modifications can be made in the procedures set forth without departing from the spirit of the invention.

ExamplesRecombinant DNA techniques

Standard methods were used to manipulate DNA as described in Sambrook, J. et al., Molecular cloning: A laboratory manual; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989. The molecular biological reagents were used according to the manufacturer's instructions.

Gene and oligonucleotide synthesis

Desired gene segments were prepared by chemical synthesis at Geneart GmbH (Regensburg, Germany). The synthesized gene fragments were cloned into an E. coli plasmid for propagation/amplification. The DNA sequences of subcloned gene fragments were verified by DNA sequencing. Alternatively, short synthetic DNA fragments were assembled by annealing chemically synthesized oligonucleotides or via PCR. The respective oligonucleotides were prepared by metabion GmbH (Planegg-Martinsried, Germany).

Description of the basic/standard mammalian expression plasmid

For the expression of a desired gene/protein (e.g. full length antibody heavy chain, full length antibody light chain, or an Fc-chain containing an oligoglycine at its N-terminus) a transcription unit comprising the following functional elements is used:

- the immediate early enhancer and promoter from the human cytomegalovirus (P-CMV) including intron A,
- a human heavy chain immunoglobulin 5'-untranslated region (5'UTR),
- a murine immunoglobulin heavy chain signal sequence,
- a gene/protein to be expressed (e.g. full length antibody heavy chain), and
- the bovine growth hormone polyadenylation sequence (BGH pA).

Beside the expression unit/cassette including the desired gene to be expressed the basic/standard mammalian expression plasmid contains

- an origin of replication from the vector pUC18 which allows replication of this plasmid in E. coli, and
- a beta-lactamase gene which confers ampicillin resistance in E. coli.

Protein determination

The protein concentration of purified polypeptides was determined by determining the optical density (OD) at 280 nm, using the molar extinction coefficient calculated on the basis of the amino acid sequence of the polypeptide.

5 **Example 1**

Generation of an expression plasmid for soluble *S. aureus* sortase A

The sortase gene encodes an N-terminally truncated *Staphylococcus aureus* sortase A (60-206) molecule (amino acid sequence of SEQ ID NO: 05).

10 The expression plasmid for the expression of soluble sortase in *E. coli* cells comprised besides the soluble sortase expression cassette an origin of replication from the vector pUC18, which allows replication of this plasmid in *E. coli*, and the URA3 gene as selectable marker, and the LacI gene to allow induction of transcription using IPTG.

15 The transcription unit of the soluble sortase comprised the following functional elements:

- a T5 promoter,
- a purification tag,
- an N-terminally truncated *S. aureus* sortase A encoding nucleic acid, and
- the To and fd termination sequences.

20 The expression plasmid for the transient expression of soluble sortase in HEK293 cells comprised besides the soluble sortase expression cassette an origin of replication from the vector pUC18, which allows replication of this plasmid in *E. coli*, and a beta-lactamase gene which confers ampicillin resistance in *E. coli*. ^

25 The transcription unit of the soluble sortase comprised the following functional elements:

- the immediate early enhancer and promoter from the human cytomegalovirus (P-CMV) including intron A,
- a human heavy chain immunoglobulin 5'-untranslated region (5'UTR),
- a murine immunoglobulin heavy chain signal sequence,
- 30 - a purification tag encoding nucleic acid,
- an N-terminally truncated *S. aureus* sortase A encoding nucleic acid, and
- the bovine growth hormone polyadenylation sequence (BGH pA).

The amino acid sequence of the mature soluble sortase is

QAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNRGVSF AEENESLD
DQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRDV
KPTDVGVLDEQKGKDKQLTLITCDDYNEKTGVWEKRKIFVATEVK
(SEQ ID NO: 05).

The purification tag has the amino acid sequence MRGSHHHHHHGS (SEQ ID NO: 32).

Example 2

Transient expression and analytical characterization

E.coli:

The recombinant production of Sortase was performed by growing E.coli cells transformed with the respective Sortase expression plasmids to an OD₅₇₈ of approx. 0.9 at 37 °C (pre-culture). At this OD₅₇₈ of approx. 0.9 protein expression was induced by adding 2 mM IPTG and growing the cells for an additional 24 hours at 28 °C. Thereafter, cells were harvested by centrifugation and lysed via high pressure using a homogenizer. Cell lysates were centrifuged to remove cell debris and subsequently the cell lysates were stored at reduced temperature (e.g. -80°C) until purification. Soluble Sortase was purified using Ni-NTA chromatography followed by size exclusion chromatography. For depletion of endotoxins an anion exchange chromatography was performed in flow through mode. The protein concentration of sortase preparations was determined by measuring the optical density (OD) at 280 nm, using the molar extinction coefficient calculated on the basis of the amino acid sequence. Purity and integrity of sortase was determined by SDS-PAGE in the presence and absence of a reducing agent (5 mM 1,4-dithiotreitol) and staining with Coomassie brilliant blue.

The protein concentration was determined by measuring the optical density (OD) at 280 nm, using the molar extinction coefficient calculated on the basis of the amino acid sequence. Purity was analyzed by SDS-PAGE in the presence and absence of a reducing agent (5 mM 1,4-dithiotreitol) and staining with Coomassie brilliant blue.

HEK:

The recombinant production was performed by transient transfection of HEK293 cells (human embryonic kidney cell line 293-derived) cultivated in F17 Medium

(Invitrogen Corp.). For transfection "293-Fectin" Transfection Reagent (Invitrogen) was used. Transfection was performed as specified in the manufacturer's instructions. Cell culture supernatants were harvested three to seven (3-7) days after transfection. Supernatants were stored at reduced temperature (e.g. -80°C).

- 5 General information regarding the recombinant expression of human immunoglobulins in e.g. HEK293 cells is given in: Meissner, P. et al., Biotechnol. Bioeng. 75 (2001) 197-203.

10 The protein concentration was determined by measuring the optical density (OD) at 280 nm, using the molar extinction coefficient calculated on the basis of the amino acid sequence. Purity was analyzed by SDS-PAGE in the presence and absence of a reducing agent (5 mM 1,4-dithiotreitol) and staining with Coomassie brilliant blue.

Example 3

Sortase mediated conjugation

15 A reaction mixture comprising 0.5 mM of the polypeptide LCR640-ULPETGGGRRRC (U: LCR640 fluorophore conjugated beta-alanine; SEQ ID NO: 33) Fc-region fragment comprising a LPETG sortase motif (SEQ ID NO: 04), 1.5 mM of an N-terminal biotinylated N-terminal cysteine comprising peptide with the C-terminally biotinylated amino acid sequence CAAA (SEQ ID NO: 03) and 50 μ M Staphylococcus aureus Sortase A in 50 mM Tris pH 7.5, 150 mM NaCl, 10 mM CaCl₂ was incubated at 37 °C for 18 hours.

In the samples were analyzed without stopping the reaction.

25 The samples (10 μ l) were injected on a Vydac C18 column of an LC-MS system and separated with a 30 min. linear gradient to 100 % buffer B (buffer A (v/v): 95% water, 5% acetonitrile, 0.1% trifluoro acetic acid (TFA); buffer B (v/v): 5% water, 95% Acetonitrile, 0.1 % TFA). The respective chromatogram is shown in Figure 3.

The Analysis of the reaction mixture with LC-ESI-TOF-MS in positive ion mode shows in peak 4 the product of the native chemical ligation reaction with the mass of 2155 Da.

30 The respective fragment pattern and masses are shown in the following Table.

LCR-----U-----L-----P-----E-----T-----G-----G-----G-----R-----R---C-NH											
1840											
726	1114										
1237 (1254)						603					
726	512					603					
1	Educt:LPETG				1840						
2	LCR-GGRR				1329						
3	Educt: LPETG dimer				3676						
4	Educt: G-Bio				830						
5	Educt: C-Bio				919						
6	LCR-LPETG-Bio				2067						
7	LCR-LPETC-Bio				2156						

Example 4**Reporter immobilization assay**

More detailed analysis of Sortases generating a thioester for native chemical ligation was done using a reporter immobilization assay (REIA) as reported in European Patent application EP14198535 and as outlined below.

Reaction mixture:

20 μ M Staphylococcus aureus sortase A (Sa-SrtA)

100 μ M nucleophile (GGGG/AAAA/CAAA)

20 μ M glucose dehydrogenase with C-terminal sortase motif (LPXTG)

250 mM MESNA

0.5 mM TCEP

or

100 μ M Listeria monocytogenes sortase A (Lm-SrtA)

100 μ M nucleophile (GGGG/AAAA/CAAA)

20 μ M glucose dehydrogenase with C-terminal sortase motif (LPXTG)

250 mM MESNA.

The glucose dehydrogenase was expressed and purified as described in WO 2007/118647.

Both reaction mixtures were prepared in 50 mM Tris-HCl buffer pH 7.5, 150 mM NaCl, 10 mM CaCl_2 .

- 49 -

The reaction mixture was incubated at 37 °C for up to 60 hours. The reaction was stopped by addition of a 60-fold excess of inhibition buffer (50 mM Tris, pH 7.5, 200 mM NaCl, 10 mM CaCl₂, 5 mM iodoacetamide). The stopped reaction mixture was centrifuged for 10 min at 5000 x g. The supernatant (50 µL) was added to 100 µL of 50 mM Tris buffer (pH 7.5) comprising 200 mM NaCl, 10 mM CaCl₂ and streptavidin coated magnetic beads. The mixture was incubated for 30 min at 30 °C with shaking at 200 rpm. Thereafter the magnetic beads were washed five times with 300 µL washing buffer each (50 mM Tris, pH 7.5, 200 mM NaCl, 10 mM CaCl₂, 5 mg/mL BSA, 0.1% Triton X-100) in V-bottom micro-titer-plates using a magnet and a vacuum pump. Afterwards the beads were resuspended in 100 µL citrate buffer and 80 µL thereof was transferred to a new well. Thereto 150 µL test buffer (0.2 M sodium citrate, pH 5.8, 0.3 g/L 4-nitrosoaniline, 1 mM CaCl₂, 30 mM glucose) were added. The kinetic of the reporter enzyme was measured over a time period of 5 min at 620 nm.

The results are shown in Figures 5 and 6.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, the descriptions and examples should not be construed as limiting the scope of the invention.

Patent Claims

1. A method for the enzymatic formation of a thioester bond comprising the following step
 - incubating
 - 5 i) a first polypeptide comprising the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue),
 - ii) a second polypeptide that has an amino acid sequence CGGG (SEQ ID NO: 02) or CAAA (SEQ ID NO: 03) at its N-terminus , and
 - 10 iii) a third polypeptide that is a sortase A or a catalytically active fragment thereof,
 and thereby forming a thioester bond.
2. The method according to claim 1, wherein the third polypeptide is Staphylococcus aureus sortase A or Listeria monocytogenes Sortase A or a catalytically active fragment thereof.
- 15 3. The method according to any one of claims 1 to 2, wherein the method is for the enzymatic conjugation of two polypeptides.
4. The method according to any one of claims 1 to 3, wherein the second polypeptide has at its N-terminus a cysteine amino acid residue followed by one to three glycine or alanine amino acid residues.
- 20 5. The method according to any one of claims 1 to 4, wherein the incubating is further in the presence of a thiol additive.
6. The method according to claim 5, wherein the thiol additive is selected from the group consisting of thiophenol, 4-mercaptophenylacetic acid (MPAA), 2-mercaptoethanesulfonate (MESNA) and combinations thereof.
- 25 7. The method according to any one of claims 1 to 6, wherein the first polypeptide comprises at its C-terminus the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue).
8. The method according to any one of claims 1 to 7, wherein the first polypeptide comprises at its C-terminus the amino acid sequence LPETG (SEQ ID NO: 04).
- 30

- 5
9. The method according to any one of claims 1 to 8, wherein the first polypeptide and the second polypeptide are independently of each other selected from an antibody variable domain, an antibody heavy chain Fab-fragment, an antibody Fc-region, a tag, and a peptide comprising the amino acid sequence LPXTG (SEQ ID NO: 01, wherein X can be any amino acid residue), a linker and a non-sortase motif moiety.
10. The method according to any one of claims 1 to 9, wherein the third polypeptide has the amino acid sequence of SEQ ID NO: 05 or SEQ ID NO: 06.