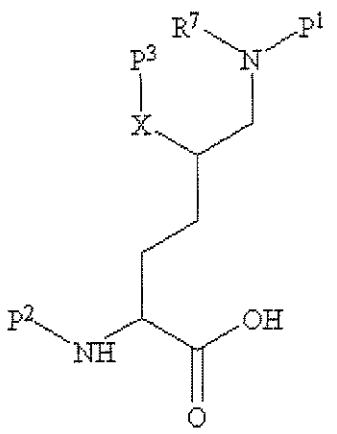
As will be clear from the above, despite the extraordinary interest in selectively modified proteins, each of the techniques currently available suffers from specific draw-backs. Over and above that, there is quite a lack of generally usable and flexible methods. Hence, additional chemical strategies for providing selective (bio-orthogonal) ligation handles are still needed to complement and expand upon methods from biology. It is the objective of the present invention, to provide such strategies.

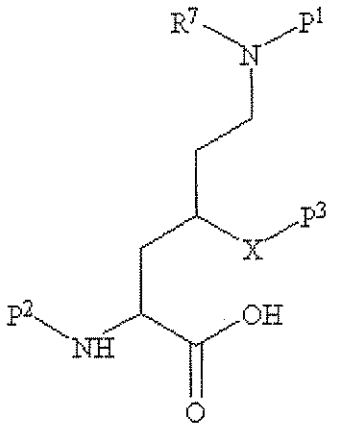
Summary of the ~~Invention~~Disclosure

The present ~~invention~~disclosure provides a new and generally applicable approach to site- and chemoselective modification of peptides and proteins, which is particularly fast, flexible and economical. In accordance with the ~~invention~~disclosure, certain novel lysine compounds are employed as building blocks in protein or peptide synthesis, providing the specific ligation handle(s) for site- and chemoselective modification.

In a first aspect, there is provided a lysine compound represented by formula (Ia) or (Ib):



(Ia)



(Ib)

wherein

-X- represents (i) sulphur or (ii) selenium;

- P^1 and -P^2 independently represent (i) hydrogen or (ii) an amine protecting group, preferably an amine protecting group independently selected from the group consisting of Carbobenzyloxy; *p*-Methoxybenzyl carbonyl; *tert*-Butyloxycarbonyl (Boc); 9-Fluorenylmethyloxycarbonyl (Fmoc); Benzyl; *p*-Methoxybenzyl; 3,4-Dimethoxybenzyl; *p*-methoxyphenyl; Tosyl; sulfonamides; allyloxycarbonyl trityl and methoxytrityl;

-P^3 represents (i) hydrogen; (ii) a group represented by the formula -X-R^6, wherein -R^6 represents an optionally substituted branched or straight chain, aliphatic or cyclic alkyl, heteroalkyl, alkenyl or heteroalkenyl moiety; (iii) a thiol or selenol protecting group, preferably a thiol or selenol protecting group selected from the group consisting of benzyl; 4-methoxybenzyl; trityl; methoxytrityl; t-butyl; t-butylthiol; acetyl; 3-nitro-2-pyridinesulphenyl; acetamidomethyl; methanethiol and 2-nitrobenzyl; (iv) a moiety –R^5 or (v) a moiety represented by the formula -C(=O)-$R^{5'}$;

-R^5 and –C(=O)-$R^{5'}$ represent a residue of a covalently bound peptide, lipid, carbohydrate, polymer, organic agent or inorganic agent; and

-R^7 represents (i) hydrogen or (ii) an optionally substituted branched or straight chain, aliphatic or cyclic alkyl, heteroalkyl, alkenyl or heteroalkenyl moiety,

or an ester, salt, solvate or hydrate of said lysine compound.

In a second aspect, there is provided a method of site- and chemoselective modification of a selected polypeptide or protein sequence comprising

i) synthesis or production of said selected polypeptide or protein sequence wherein at least one lysine compound according to formula (Ia) or (Ib) is incorporated by addition or substitution in said selected sequence, wherein –P^3 represents hydrogen or a protecting group as defined above;

ii) removing in the residue of said lysine compound the group represented by –P^3 and conjugating a group represented by the formula -R^5 or -C(=O)-$R^{5'}$, wherein –R^5 and -C(=O)-$R^{5'}$ have the same meaning as defined in relation to formulae (IIa-IIb), with the residue of said lysine compound at the sulphur or selenium atom.

In a third aspect, there is provided a method of synthesis of the lysine compounds as defined above, said method comprising the steps of:

(ia) treating a 4-hydroxylysine or 5-hydroxylysine compound with an organoborane compound to obtain an amino acid protected compound, which is subsequently reacted with an amine protecting group, yielding the protected 4-hydroxylsine or 5-hydroxylysine; or

(ib) treating a 4-chlorolysine, 5-chlorolysine, 4-bromolysine or 5-bromolysine compound with an organoborane compound to obtain an amino acid protected compound, which is subsequently reacted with an amine protecting group, yielding the protected 4-chlorolysine, 5-chlorolysine, 4-bromolysine or 5-bromolysine;

(iia) mesylating the compound obtained in step (ia) and subsequently reacting it with a suitable thiocarboxylic acid or selenocarboxylic acid or salt thereof to yield the corresponding thio- or seleno-ester; or

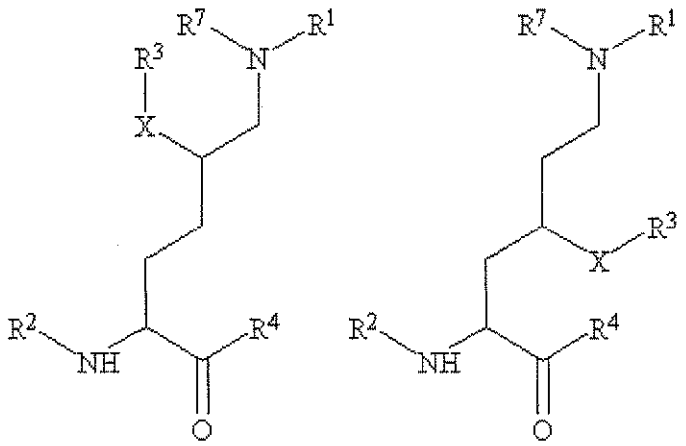
(iib) reacting the compound obtained in step (ib) with a suitable thiocarboxylic acid or selenocarboxylic acid or salt thereof to yield the corresponding thio- or seleno-ester;

(iii) hydrolysis of the thio- or selenoester obtained in step (iia) or (iib) with an aqueous solution of an alkali metal hydroxide, yielding the thiol or selenol compound, which is subesequently reacted with an agent that can transfer a group $-P^3$, as defined above, to the thiol or selenol group; and

(iv) removing the amino acid protecting organoborane group, and, optionally, subsequently reacting the compound with an amine protecting agent, to yield the compound according to formula (Ia) or (Ib) as defined here above.

In a fourth aspect, there is provided the use of a lysine compound as defined above as a building block in the synthesis of a selected peptide or protein to provide a ligation handle for the site- and chemoselective modification of said selected polypeptide or protein sequence.

In a fifth aspect, there is provided a non-naturally occurring peptide represented by formula (IIa) or (IIb):



wherein

$-R^1$ represents (i) hydrogen; (ii) $-P^1$ as defined above or (iii) a moiety represented by the formula $-C(=O)-R^{5'}$;

$-R^2$ represents (i) hydrogen; (ii) $-P^2$ as defined above, (iii) a C→N polypeptide chain or (iv) a moiety represented by the formula $-C(=O)-R^{5'}$;

$-R^3$ represents (i) hydrogen; (ii) $-P^3$ as defined above; (iii) a moiety $-R^5$; or (iv) a moiety represented by the formula $-C(=O)-R^{5'}$;

$-R^4$ represents (i) –OH or (ii) an N→C polypeptide chain;

-X- represents (i) sulphur or (ii) selenium;

$-R^5$ and $-C(=O)-R^{5'}$ represent a residue of a covalently bound peptide, lipid, carbohydrate, polymer, organic agent or inorganic agent; and

-R^7 represents (i) hydrogen or (ii) an optionally substituted branched or straight chain, aliphatic or cyclic alkyl, heteroalkyl, alkenyl or heteroalkenyl moiety;

with the proviso that in the respective formula at least one of $-R^2$ and $-R^4$ represents a polypeptide chain.

In a sixth aspect, there is provided a non-naturally occurring peptide as defined above, for use as a therapeutic or diagnostic agent.

In particular, the ~~invention~~disclosure entails the use of specific thiolysine and selenolysine compounds, notably the compounds 5-thiolysine (also referred to as δ-thiolysine) and 4-thiolysine (also referred to as γ-thiolysine) as well as 5-selenolysine (also referred to as δ-selenolysine) and 4-selenolysine (also referred to as γ-selenolysine), for the site- and chemoselective modification of a peptide or protein. The lysine compounds of the present ~~invention~~disclosure are easily synthesized in high yield from commercially available building blocks, involving, a relatively low number of synthetic steps, as will be illustrated herein.

The design of the present lysine compounds has two important features. First, the positioning of the thiol or selenol group at the respective carbon allows for a very efficient intramolecular transfer reaction to take place. Secondly, the thiol or selenol group can subsequently be removed using reported procedures, thereby restoring the native lysine structure, or be used as an additional conjugation handle. The methodology is fast and gives well-defined material.

The presence of the thiol or selenol residue in the vicinity of the ε-amine or, where applicable, the α-amine of lysine allows for very favourable S to N or Se to N transesterification reactions, whereby a selected functional agent, introduced in the peptide or protein through thio- or selenoesterification, is transferred to said ε-amine or α-amine.

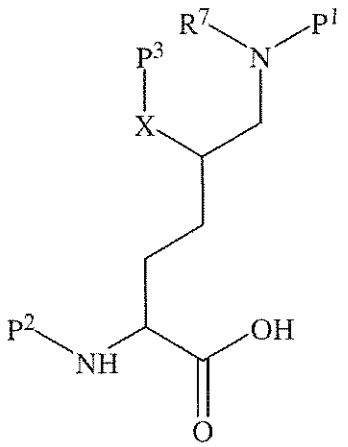
As will be explained in great detail in the examples, the present inventors have demonstrated the application of a lysine building block of the ~~invention~~disclosure in the

ubiquitylation of selected peptide targets. By employing an E1-mediated synthesis of full-length ubiquitin thioester and the newly developed amino acid 5-thiolysine, the inventors demonstrated site-specific ubiquitylation of lysine residues in selected target peptides, based on H2B, PCNA, PTEN and p53. This methodology is also readily adaptable to other functional agents or ligands. The use of mild desulfurization conditions after ligation do not affect the functional integrity of other building blocks present in the peptide or protein or the selected functional agent.

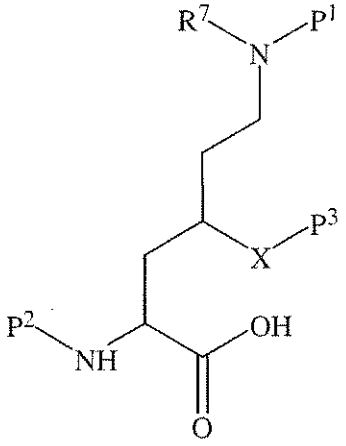
In summary, the present ~~invention~~disclosure concerns certain lysine compounds, methods for synthesizing them and their use as building block in protein or peptide synthesis, typically as bio-orthogonal target for site- and chemoselective modifications of polypeptide or protein sequence. The ~~invention~~disclosure also concerns methods of synthesizing and modifying selected protein or peptide sequences using the present lysine compounds. Furthermore, the ~~invention~~disclosure concerns the intermediate and end-products obtained in such methods. These and other aspects of the ~~invention~~disclosure, as defined in the appending claims, will be described and exemplified in more detail in the following description and examples.

Detailed description of the Invention

A first aspect of the invention concerns a lysine compound represented by formula (Ia) or (Ib):



(Ia)

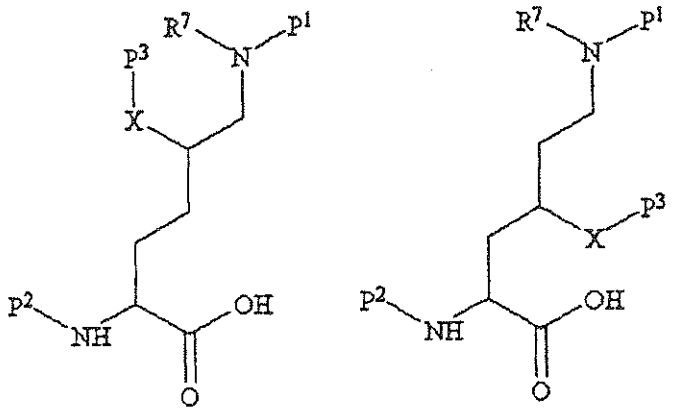


(Ib)

wherein -X- represents (i) sulphur or (ii) selenium; -P^1 and -P^2 independently represent (i) hydrogen or (ii) an amine protecting group; -P^3 represents (i) hydrogen;

~~Amended~~ Claims

1. Lysine compound represented by formula (Ia) or (Ib):



(Ia) (Ib)

wherein

-X- represents (i) sulphur or (ii) selenium;

- P^1 and -P^2 independently represent (i) hydrogen or (ii) an amine protecting group, preferably an amine protecting group independently selected from the group consisting of Carbobenzyloxy; *p*-Methoxybenzyl carbonyl; *tert*-Butyloxycarbonyl (Boc); 9-Fluorenylmethyloxycarbonyl (Fmoc); Benzyl; *p*-Methoxybenzyl; 3,4-Dimethoxybenzyl; *p*-methoxyphenyl; Tosyl; sulfonamides; allyloxycarbonyl trityl and methoxytrityl;

-P^3 represents (i) hydrogen; (ii) a group represented by the formula -X-R^6, wherein -R^6 represents an optionally substituted branched or straight chain, aliphatic or cyclic alkyl, heteroalkyl, alkenyl or heteroalkenyl moiety; (iii) a thiol or selenol protecting group, preferably a thiol or selenol protecting group selected from the group consisting of benzyl; 4-methoxybenzyl; trityl; methoxytrityl; t-butyl; t-butylthiol; acetyl; 3-nitro-2-pyridinesulphenyl; acetamidomethyl; methanethiol and 2-nitrobenzyl; (iv) a moiety –R^5 or (v) a moiety represented by the formula -C(=O)-$R^{5'}$;

-R^5 and –C(=O)-$R^{5'}$ represent a residue of a covalently bound peptide, lipid, carbohydrate, polymer, organic agent or inorganic agent; and

-R^7 represents (i) hydrogen or (ii) an optionally substituted branched or straight chain, aliphatic or cyclic alkyl, heteroalkyl, alkenyl or heteroalkenyl moiety,

or an ester, salt, solvate or hydrate of said lysine compound.