



(51) International Patent Classification:

C12Q 1/37 (2006.01) *C07K 14/81* (2006.01)
C12N 9/64 (2006.01) *C07C 15/04* (2006.01)

(21) International Application Number:

PCT/EP2018/066633

(22) International Filing Date:

21 June 2018 (21.06.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

17177384.9 22 June 2017 (22.06.2017) EP

(71) Applicant: AAA CHEMISTRY APS [DK/DK]; Margeritvej 2, 3650 Ølstykke (DK).

(72) Inventors: DINESS, Frederik; Margeritvej 2, 3650 Ølstykke (DK). MELDAL, Morten; Mosesvinget 78, 2400 Copenhagen NV (DK). EMBABY, Ahmed; Matthæusgade 50, 1666 Copenhagen NV (DK).

(74) Agent: AAGAARD, Louise; Høiberg P/S, Adelgade 12, 1304 Copenhagen K (DK).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

— with international search report (Art. 21(3))

(54) Title: TUNABLE PROBES FOR SELECTIVE PROTEIN LABELLING AND ENZYME INHIBITION

(57) Abstract: The present invention relates to methods of selectively reacting a compound of Formula (I) with a cysteine residue, which is contained in the sequence of a polypeptide. The methods may be employed in the fields of labelling polypeptides, detecting polypeptides, modulation of enzyme activity, isolation of polypeptides, methods of diagnostics, methods of prevention or treatment of clinical conditions or methods of transportation of compounds e.g. prodrugs.

Tunable probes for selective protein labelling and enzyme inhibition

Technical field

5

The present invention relates to the field of methods of covalently reacting cysteine residues with compound as well as to the compounds per se. The methods may be employed in the fields of labelling polypeptides, detecting polypeptides, modulation of enzyme activity, isolation of polypeptides, methods of diagnostics, methods of prevention or treatment of clinical conditions or methods of transportation of compounds e.g. prodrugs.

10

Background

Disease states are commonly linked to significant changes in protein expression patterns, either as the cause or a consequence of the disorder. The levels of enzymes are, as a consequence of their catalytic nature, among the most significant proteins to balance. In several diseases it appears to be of particularly importance to regulate the balance of the activity of different proteases involved in cellular processing and catalytic control. E.g. cysteine proteases, such as cathepsin B, K and L, and caspase 1, have crucial roles in the development of osteoporosis, cancer and arthritis. Thus, identification, quantification and regulation of these enzymes is essential as means of disease treatment, disease diagnostics and for investigation of mechanism of action.

15

20

Over the years, chemical biologists have designed a variety of electrophiles for modifying nucleophilic residues in a target protein. However, this approach is still facing major challenges with regards to the selectivity of such electrophiles among different types of amino acids (cysteine, lysine, etc.), and in particular among several residues of the same amino acid in different sites, e.g. differentiating surface exposed, buried and catalytic residues. Cysteine residues are, despite of their low occurrence in proteins (1.9%), often located in functionally important sites. In addition to their structural stability in the form of disulfide bonds, they play diverse roles in biological processing e.g. in catalysis, in redox reactions, in allosteric regulations, and in metal binding. The functional variety of cysteines may be attributed to the unique properties of the thiol group. The S-H bond has a low dissociation energy, which facilitates the ability of the thiol to act as a nucleophile in hydrolysis and in redox reactions. It is noteworthy that the acidity of cysteine in proteins greatly depend on the local protein

25

30

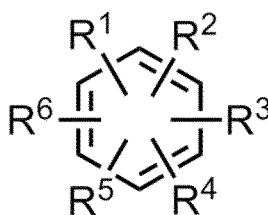
35

environment. At the surface of proteins, cysteine thiols have a pKa of ~ 8.5; whereas, it may be as low as 2.5 for a catalytic thiol in an active site. Reactive electrophiles designed to probe cysteine residues have been described, including chloromethylketone (CMK), acylomethylketone (ACMK), epoxides, sulfonate esters, halocetamides as well as a range of Michael acceptors, including maleimides, acrylamides, vinyl sulfonamides, amino methyl acrylate, and methyl vinyl sulfones. Regardless of the significant efforts in this area, cross reactivity with other nucleophilic amino acids side chains such as the amine of lysines, insufficient reactivity and/or lack of selectivity among cysteine residues are still major issues. Recently, aryl halides such as chloronitrobenzenes and dichlorotriazines, have been investigated for cysteine labeling. However, these have not been explored with regards to promiscuity towards other amino acid nucleophiles, or selective reactivity among different cysteine residues.

Summary

The present invention provides selective reactive probes comprising a benzene core. The benzene core provides six positions for spatial orientation of an electrophilic carbon, tuning of reactivity, and attachment of linkers with other functionalities of interests. This is in contrast to the other reactive electrophiles listed above, which have no or little room for attenuation of reactivity by structural variation in close proximity to the carbon attacked by the thiol of cysteine (Figure 1).

More specifically, the present invention provides methods for reacting a compound of Formula I with a cysteine residue, thereby forming a covalent bond, wherein the cysteine is contained in the sequence of a polypeptide, wherein said compound has the following structure:



Formula I

wherein:

R^3 is a leaving group

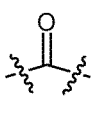
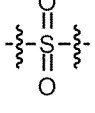
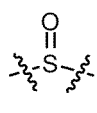
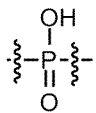
R^1 is R^7 ;

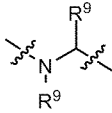
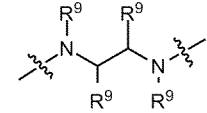
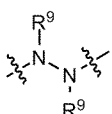
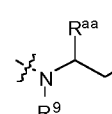
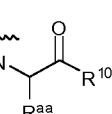
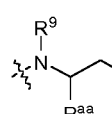
R^7 is selected from the group consisting of $-X-Z-R^8$, $-Z-T-R^8$, and $-Z-T-(R^8)_2$;

5

wherein

X is selected from the group consisting of a bond, $-\text{CH}_2-\text{N}(\text{R}^9)-$, and $-\text{CH}_2-\text{O}-$;

Z is selected from the group consisting of a bond, , , , and ;

T is selected from the group consisting of a bond, , , , , , and ;

10

R^8 is individually selected from the group consisting of amino acids, amino acid derivatives, peptides, peptide derivatives, peptidomimetic, alkylamines, protecting groups, $-\text{NH}-\text{linker}-\text{Y}$, $-\text{linker}-\text{Y}$ and $-\text{linker}-\text{R}^{14}$, wherein said peptide optionally may be N- and/or C-terminally modified;

15

Y is a labelling molecule, a drug molecule or a prodrug;

R^9 is individually selected from the group consisting of $-\text{H}$, alkyl-COOH and an amino acid side chain;

R^{10} is selected from the group consisting of $-\text{H}$, $-\text{OH}$ and R^8 ;

R^{14} is a reactive group;

20

R^2 , R^4 , R^5 , R^6 , are individually selected from the group consisting of $-\text{H}$, R^7 , an amino acid side chain and an electron withdrawing group; under the proviso that at least one of R^2 , R^4 , R^5 , and R^6 is an electron withdrawing group.

25

In one aspect, the present invention concerns a method of modulating the activity of a protein, said method comprising performing the method described herein.

In one aspect, the present invention concerns a method for labelling a polypeptide with a labelling molecule, said method comprising performing the method described herein.

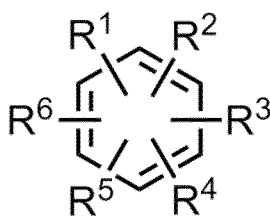
In one aspect, the present invention concerns a method of transportation of a drug molecule or a prodrug, said method comprising performing the method as described
5 herein, using a compound of formula I, wherein Y is a drug molecule or a prodrug.

In one aspect, the present invention concerns a method for detecting a polypeptide, said method comprises the steps of a) performing the method described herein using a
10 compound of formula I, wherein R^8 is –linker-Y or –NH-linker-Y; and b) detecting Y.

In one aspect, the present invention concerns a method for diagnosis of a clinical condition associated with a polypeptide in an individual at risk of acquiring said clinical condition, said method comprising the steps of a) performing the method as described herein on a sample from said individual using a compound of formula I, wherein R^8 is –
15 linker-Y or –NH-linker-Y, and wherein the polypeptide is associated with said clinical condition; and b) detecting Y; thereby determining the presence, absence and/or level of said polypeptide in said sample.

In one aspect, the current invention concerns a method for treatment or prevention of a clinical condition associated with a polypeptide in an individual in need thereof, said
20 method comprising performing the method described herein.

In one aspect, the current invention concerns a compound of Formula I:



Formula I

25

wherein:

R^3 is a leaving group

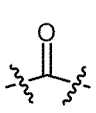
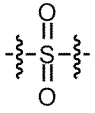
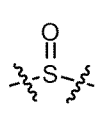
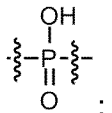
R^1 is R^7 ;

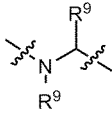
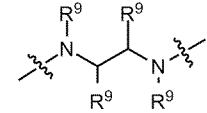
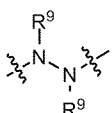
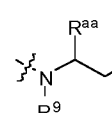
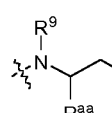
R^7 is selected from the group consisting of –X-Z- R^8 , –Z-T- R^8 , and –Z-T-(R^8)₂;

30

wherein

X is selected from the group consisting of a bond, –CH₂-N(R^9)–, and –CH₂-O–;

Z is selected from the group consisting of a bond, , , , and ;

T is selected from the group consisting of a bond, , , , , and ;

5 R^8 is individually selected from the group consisting of amino acids, amino acid derivatives, peptides, peptide derivatives, peptidomimetic, alkylamines, protecting groups, -NH-linker-Y and -linker-Y, wherein said peptide optionally may be N- and/or C-terminally modified;

Y is a labelling molecule, a drug molecule or a prodrug;

10 R^9 is individually selected from the group consisting of -H, alkyl-COOH and an amino acid side chain;

R^{10} is selected from the group consisting of -H, -OH and R^8 ;

R^2 , R^4 , R^5 , R^6 , are individually selected from the group consisting of -H, R^7 , an amino acid side chain and an electron withdrawing group;

15 under the proviso that at least one of R^2 , R^4 , R^5 , and R^6 is an electron withdrawing group, and that no more than four of R^2 , R^3 , R^4 , R^5 , and R^6 are -F.

20 Description of Drawings

Figure 1. A) General approach for selective cysteine modification by tuning the reactivity of fluoroaryls in comparison to non-tunable unselective traditional reactive groups as iodacetamide. **B)** General scheme for protein modification through a
25 nucleophilic aromatic substitution reaction between cysteine and fluoroaryls.

Figure 2. The reactivity screening of compound **2m** with Boc-L-cysteine in different conditions; Method A (a) and Method B (b).

Figure 3. Selective cysteine modification of peptide **4**, which contains all possible nucleophilic sidechains of the natural amino acids, by compound **2g**. **A)** TOF-MS shows a full conversion of peptide **4** to **5**. MS (MALDI-TOF) m/z calcd. for $C_{63}H_{83}F_4N_{18}O_{16}S_2^+ [M+H]^+$ 1487.5606, found 1487.5907. **B)** MS/MS fragmentation of peak 1487 shows that the presence of all residues except cysteine that is modified with compound **2g** to give the corresponding fragment with mass $C_{14}H_{14}F_4N_2O_3S_2$ 398.0382, found 398.03659.

Figure 4. A) LC of the pure peptide **4**. **B)** The reaction of peptide **4** with compound **2g** according to method B (AcN/water (1:1), pH=11). **C)** No reaction of peptide **4** with Probes 2b according to method B.

Figure 5. Labeling of protein sulfhydryl groups. **A)** Synthesis of azide probe **8a** and labeling of cysteine-containing eGFP and BSA by Cu-catalyzed click chemistry is visualized by coomassie blue stain and in-gel fluorescence scanning. **B)** In-gel fluorescence scanning of eGFP and albumin fluorescence labeled with probe **10**. **C)** Western blot of eGFP and albumin labeled with probe **11** and HRP–streptavidin.

Figure 6. A) The ability of compounds **2a** and **2g** to inhibit papain (cysteine protease) while in **B)** no such effect was observed with subtilisin (serine protease) after incubation for 60 min at 37 °C. **C)** Fluorescence intensity measured on beads incubated with buffer or different compounds (**2a** and **2g**) after treatment with papain (blue) or subtilisin (red).

Figure 7. A) Schematic presentation of the attenuation of reactivity of the compounds **2g**, **2h** and **2c** to cysteine residues in TEV protease; compound **2h** was able to selectively arylate the catalytic cysteine, in contrast to **2g** that reacted with both the active-site and surface-exposed cysteine and **2c** that did not react at all. **B)** The activity of TEV protease was measured by FRET-substrate cleavage. Shown in red is the average from triplet determinations of FRET-substrate incubated with TEV protease. In blue is shown the average of triplet determinations of sample without enzyme (negative control). **C)** The overlay of HRMS spectra shows peaks corresponding to cysteine-containing tryptic fragments of TEV protease treated with compound **2g** and **2h** before and after the reaction.

Figure 8. A) Azide-functionalized derivatives tested for ability of labelling TEV protease.

B) Activity-based protein profiling (ABPP) of TEV protease: Only compound **8b** and **8c** could label TEV protease in an activity-dependent manner, since preheated TEV protease did not show any fluorescence. Compound **8a** could label both active and denatured TEV protease while compound **8e** did not label any of them.

C) The tunability and selectivity of compounds **8a**, **8b** and **8d** toward cysteine-containing proteins as the pretreatment with iodacetamide inhibited their activity

D) Compound **8b** was able to label TEV protease (spiked) and a few other proteins from which chloramphenicol acetyl transferase enzyme was identified by HRMS.

E) MS spectrum of the protein isolated from the intense band observed just below the 25-kDa protein marker after tryptic digestion. The very intense peptide peaks that were used for identification of the unknown protein band were encircled.

Figure 9. A) Shows inhibition assay of caspase-1 by compounds **23-25** in comparison to **AcYVAD-CMK**. **B)** TEV inhibition assay shows that compounds **23-25** has no effect on TEV activity at the same concentration.

Figure 10. A) Labelling of cysteine-containing Human Serum Albumin with **27** by Cu-catalyzed click chemistry as described in Example 9.

B) MS spectrum of Human Serum Albumin (upper) and Human Serum Albumin labelled with **27** (lower).

Detailed description

Definitions

A waved line ($\text{---}\overset{\text{~}}{\underset{\text{~}}{\text{S}}}\text{---}$) indicates the point of attachment of the substituent.

A substituent or moiety indicated by a general formula comprising two waved lines may be linked to the compounds of the invention in any direction. Accordingly, a substituent

or moiety of the general formula $\text{---}\text{A---B---}$ may equally well be denoted with the

general formula $\text{---}\text{B---A---}$. Similarly, a substituent or moiety indicated by a general

formula comprising two free bonds may be linked to the compounds of the invention in any direction. Accordingly, a substituent or moiety of the general formula –A-B- may equally well be denoted with the general formula –B-A-.

- 5 The term “active site” as used herein refers to the region of an enzyme where substrate molecules bind and undergo a chemical reaction.

The term “acetylated” means linked to an acetyl group: .

- 10 The term “alkane” refers to saturated linear or branched carbonhydrides of the general formula C_nH_{2n+2} .

The term “alkenyl” as used herein refers to a substituent derived from an alkene by removal of one –H. An alkene may be any acyclic carbonhydride comprising at least one double bond. Frequently, alkenyl will have the general formula $-C_nH_{2n-1}$.

15

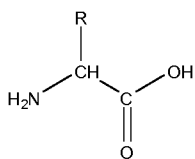
The term “alkyl” refers to a substituent derived from an alkane by removal of one –H.

The term amine refers to a compound comprising an –N-. Primary amines comprises – NH_2 .

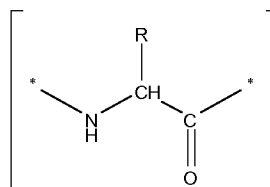
20

The term “alkynyl” as used herein refers to a substituent derived from an alkyne by removal of one –H. An alkyne may be any acyclic carbonhydride comprising at least one triple bond. Frequently, alkynyl will have the general formula $-C_nH_{2n-3}$.

- 25 The term “amino acid” as used herein refers to a compound of the following general

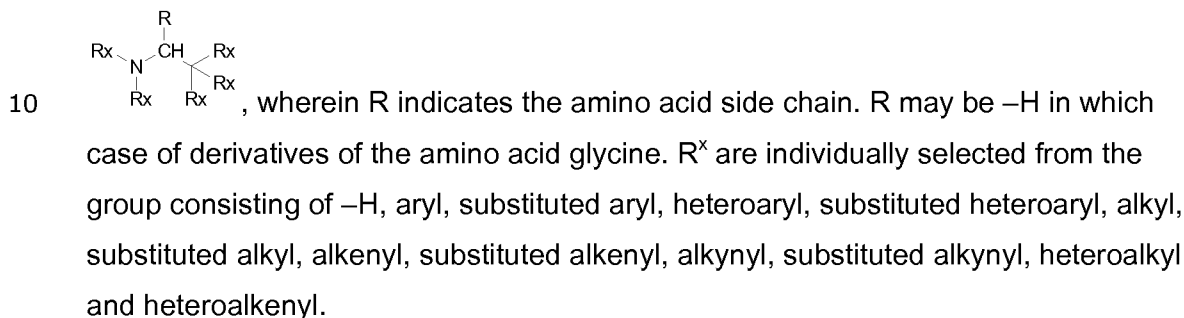


structure: , wherein R indicates the amino acid side chain. R may be –H in which case the amino acid is glycine. Furthermore, the term “amino acid” as used herein also covers amino acids linked to other amino acids in a peptide or polypeptide. Thus, amino acids may be bound to each other by peptide bonds to form peptides or



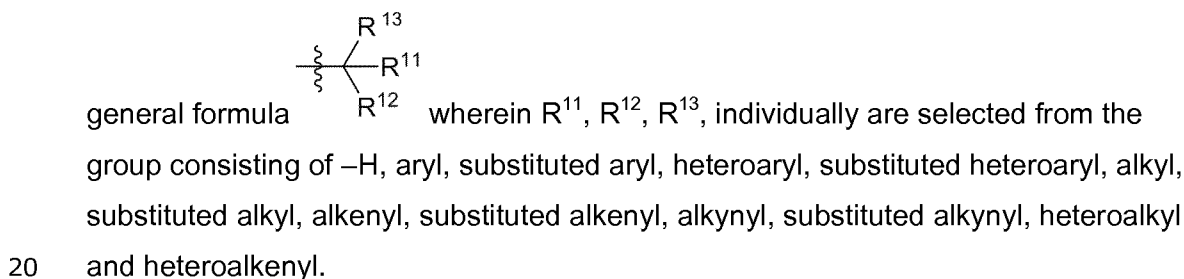
polypeptides of the following general structure: $\left[\begin{array}{c} \text{R} \\ | \\ \text{*} - \text{N} - \text{CH} - \text{C} - \text{*} \\ | \quad \quad || \\ \text{H} \quad \quad \text{O} \end{array} \right]_n$, wherein n is an integer and * indicates the point of attachment to the next amino acid residue. Amino acids may be standard amino acids, but also includes other amino acids of aforementioned general structure. Amino acids may be D-stereo-isomers (referred to as D-amino acids herein) or may be L-stereo-isomers (referred to as L-amino acids herein).

The term “amino acid derivative” as used herein refers to a compound which may be synthesised from an amino acid. The compound has the following general structure:



15

The term “amino acid side chain” as used herein refers to -H or a substituent of the



Preferred “amino acid side chains” are the side chains of the standard amino acids, for example the side chains of the 21 proteinogenic α-amino acids found in eukaryotes.

25 The term “arene” as used herein refers to aromatic mono- or polycyclic carbonhydrides.

The term “aromatic” refers to a chemical substituent characterised by the following:

- contains a delocalized conjugated π system, most commonly an arrangement of alternating single and double bonds
- has a coplanar structure, with all the contributing atoms in the same plane
- the contributing atoms are arranged in one or more rings
- 5 • it contains a number of π delocalized electrons that is even, but not a multiple of 4.

The term "aryl" as used herein refers to a substituent derived from an arene by removal of one $-H$ from a C in the ring. Examples of useful aryls to be used with the present invention comprise phenyl, naphthyl, anthracenyl, phenanthrenyl, and pyrenyl.

The term "C-terminally modified" refers to a peptide being covalently linked to a moiety at the C-terminus. A non-limiting example of a C-terminal modification may be a C-terminal amidation or esterification, e.g. the C-terminal OH group of the peptide may be exchanged with an amine group or an alkoxy group.

The term "recognition sequence for peptide cleaving enzyme" as used herein refers to an amino acid sequence recognised and hydrolysed by a peptide cleaving enzyme, such as a protease. A "recognition sequence" may comprise or consists of a consensus sequence rather than of a specific amino acid sequence. Several databases are available to aid in the identification of recognition sequences for peptide cleaving enzymes, including the PROSPER database available at <https://prosper.erc.monash.edu.au> and described in Song J et al., *PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites*, *PLoS ONE*, 2012, 7(11): e50300 or the "PeptideCutter" of the ExPASy database available at http://web.expasy.org/peptide_cutter/. In one embodiment the recognition sequence for the peptide cleaving enzyme comprises the amino acid sequence ENLYFQGY. In another embodiment the recognition sequence for the peptide cleaving enzyme comprises the amino acid sequence E N L Y F Q G K. This may in particular be the case when the peptide cleaving enzyme is TEV protease.

The term "detectable label" as used herein refers to any label, which can be detected, i.e. capable of generating detectable signals. Detectable labels are also commonly known in the art as "detectable moieties". Examples of detectable labels include tags and probes. The detectable label may for example be selected from the group

consisting of radiolabels, biotin, fluorescent labels, luminescent labels and coloured labels. Examples of detectable labels include chromogenic moieties, fluorescent moieties, radioactive moieties and electrochemically active moieties.

5 A chromogenic moiety is a moiety which is coloured, which becomes coloured when it is incorporated into a conjugate, or which becomes coloured when it is incorporated into a conjugate and the conjugate subsequently interacts with a secondary target species (for example, where the conjugate comprises a protein which then interacts with another target molecule). Typically, the term "chromogenic moiety" refers to a
10 group of associated atoms which can exist in at least two states of energy, a ground state of relatively low energy and an excited state to which it may be raised by the absorption of light energy from a specified region of the radiation spectrum. Often, the group of associated atoms contains delocalised electrons. Examples include porphyrins, polyenes, polyynes and polyaryls.

15 A fluorescent moiety is a moiety which comprises a fluorophore, which is a fluorescent chemical moiety. Examples of fluorescent compounds include: the Alexa Fluor® dye family available from Invitrogen; cyanine and merocyanine; the BODIPY (boron-dipyrromethene) dye family, available from Invitrogen; the ATTO dye family
20 manufactured by ATTO-TEC GmbH; fluorescein and its derivatives; rhodamine and its derivatives; naphthalene derivatives such as its dansyl and prodan derivatives; pyridyloxazole, nitrobenzoxadiazole and benzoxadiazole derivatives; coumarin and its derivatives; pyrene derivatives; and Oregon green, eosin, Texas red, Cascade blue and Nile red, available from Invitrogen.

25 A radioactive moiety is a moiety that comprises a radionuclide. Examples of radionuclides include iodine-131, iodine-125, bismuth-212, yttrium-90, yttrium-88, technetium-99m, copper-67, rhenium-188, rhenium-186, gallium-66, gallium-67, indium-111, indium-114m, indium-114, boron-10, tritium (hydrogen-3), carbon-14,
30 sulfur-35, fluorine-18 and carbon-11. Fluorine-18 and carbon-11, for example, are frequently used in positron emission tomography. In one embodiment, the radioactive moiety may consist of the radionuclide alone. In another embodiment, the radionuclide may be incorporated into a larger radioactive moiety, for example by direct covalent bonding to a linker group (such as a linker containing a thiol group) or by forming a co-
35 ordination complex with a chelating agent. Suitable chelating agents known in the art

include DTPA (diethylenetriamine- pentaacetic anhydride), NOTA (1,4,7-triazacyclononane-N,N',N"-triacetic acid), DOTA (1, 4,7,10-tetraazacyclododecane-N,N',N",N'''-tetraacetic acid), TETA (1,4,8,11-tetraazacyclotetra-decane-N,N',N",N'''-tetraacetic acid), DTTA (N'- ϕ - isothiocyanatobenzyl)-diethylene-triamine-N¹,N²,N³-tetraacetic acid) and DFA (N'- [5 - [[5- [5-acetylhydroxyamino)pentyl] amino] - 1,4-dioxobutyl]hydroxyamino]pentyl] - N-(5-aminopentyl)-N-hydroxybutanediamide).

An electrochemically active moiety is a moiety that comprises a group that is capable of generating an electrochemical signal in an electrochemical method such as an amperometric or voltammetric method. Typically, an electrochemically active moiety is capable of existing in at least two distinct redox states.

The term "electron withdrawing group" as used herein refers to a substituent which has a positive Hammett Meta Substituent Constant (σ_m) and a positive Hammett Para Substituent Constant (σ_p). (As listed in table I in Chemical Reviews 1991, 91, 165-195)

Preferred "electron withdrawing groups": X, SX₅, SO₂NR₂, CX₃, OCX₃, SCX₃, SOCX₃, SO₂CX₃, SOR, SO₂R, CN, CHX₂, COR, CONR₂ and COOR, wherein X is a halogen and R is alkyl.

The term halogen as used herein refers to a substituent selected from the group consisting of -F, -Cl, -Br and -I.

The term "heteroalkenyl" refers to a straight- or branched-chain alkenyl group, of which one or more carbon has been replaced by a heteroatom selected from S, O and N. Exemplary heteroalkenyls include alkyl esters, ketones, aldehydes, amides, carbamates, ureas, guanidines, and sulfoxides.

The term "heteroalkyl" refers to a straight- or branched-chain alkyl group, of which one or more carbon has been replaced by a heteroatom selected from S, O and N. Exemplary heteroalkyls include alkyl ethers, secondary and tertiary alkyl amines, and alkyl sulfides.

The term "heteroaryl" as used herein refers to a substituent derived from an heteroarene by removal of one -H from an atom in the ring structure of said

heteroarene. Heteroarenes are mono- or polycyclic aromatic compounds comprising one or more heteroatoms in the ring structure. Said heteroatoms are preferably selected from the group consisting of S, N and O. Non limiting examples of useful heteroaryls to be used with the present invention comprise azolyl, pyridinyl, pyrimidinyl, furanyl, and thiophenyl.

The term "individual" as used herein refers to any individual, preferably a mammal, and more preferably a human being.

The term "Labelling molecule" refers to any moiety, which can be used to label a compound. Preferably, a labelling molecule is a detectable label. The labelling molecule may be selected from the group consisting of a fluorescent labelling molecule, a radioisotope labelling molecule, an affinity molecule, an azide, a terminal alkyne, a spin label, a prodrug, a mass tag, and a photoreactive group.

The term "linker" as used herein refers to a chemical moiety linking two other chemical moieties. Preferred linkers to be used with the invention are described herein elsewhere.

The term "N-terminally modified" refers to a peptide being covalently linked to a moiety at the N-terminus. A non-limiting example of an N-terminal modification may be that said peptide may contain an N-terminal acetylation, carboxylation, or alkylation, e.g. the N-terminal amine may be acetylated, benzoylated, carbobenzoxyated, methylated or benzylated.

The term "peptide" as used herein refers to a shorter sequence of amino acid residues linked by peptide bonds. For example, a peptide may consist of in the range of 2 to 40 amino acids. Peptides may be N- and/or C-terminally modified.

The term "polypeptide" as used herein refers to a sequence of amino acids linked by peptide bonds. In general a polypeptide comprises at least 4 amino acid residues.

The term pKa as used herein refers to the negative logarithmic of the dissociation constant K_a for an acid in a given solvent: $pK_a = -\text{Log}_{10} K_a$.

K_a , also called the *acidity constant*, is defined as:

$$K_a = \frac{[A^-][SH^+]}{[HA][S]}$$

for the reaction:



wherein S is the solvent and HA is an acid that dissociates into A^- , known as the
 5 conjugate base of the acid, and a hydrogen ion which combines with a solvent
 molecule. When the concentration of solvent molecules can be taken to be constant,
 K_a is:

$$K_a = \frac{[A^-][H^+]}{[HA]}$$

- 10 The term “small molecule” as used herein refers to organic molecules with molecular
 weights of no more than 900 g/mol.

The term “standard amino acid” refers to the 20 amino acids encoded by the standard
 genetic code. The amino acids are referred to herein using standard IUPAC

- 15 nomenclature. Standard amino acids are all L-amino acids.

The term “substituted” as used herein in relation to chemical compounds refers to
 hydrogen group(s) being substituted with another moiety. Thus, “substituted with X” as
 used herein in relation to chemical compounds refers to hydrogen group(s) being
 20 substituted with X. Similarly, “substituted X” refers to X, wherein one hydrogen group
 has been substituted with another moiety. By way of example “substituted alkyl” refers
 to alkyl-R, wherein R is any moiety but –H.

The term “substituent” as used herein in relation to chemical compounds refers to an
 25 atom or group of atoms substituted in place of a hydrogen atom.

Method of reacting a compound of formula I with a cysteine residue

In one aspect, the present invention concerns a method for reacting a compound of
 Formula I with a cysteine residue, thereby forming a covalent bond, wherein the
 30 cysteine is contained in the sequence of a polypeptide. The compound of formula I may
 be any of the compounds of formula I described herein below in the section
 “Compound of formula I”. In particular, the compound of formula I may be any

compound of one of the more specific formulas II, III, IV, VI, VII, VIII, IX or X described herein below in the section compound of formulas II to X.

5 The methods of the invention may be performed in any manner allowing for reaction between the compound of formula I and the cysteine residue. The skilled person will be able to select appropriate conditions.

10 In some embodiments the methods of the invention are performed in an aqueous environment, for example in an aqueous solvent.

In some embodiments the methods of the invention are performed in vivo. This may in particular be the case in embodiments, where the cysteine residue is contained in the sequence of a polypeptide present in vivo. In such embodiments, the methods are performed under in vivo conditions.

15 Thus, in some embodiments, the reaction of compound of Formula I with a cysteine residue is conducted at the surface of or inside a living organism.

20 In some embodiments the methods may be performed in or on a sample obtained from an individual.

In some embodiments the methods are performed in vitro. In such embodiments, the methods may comprise the steps of:

- Providing a mixture of the polypeptide in a solvent;
- 25 • Optionally adjusting pH of the mixture;
- Providing the compound of Formula I;
- Reacting said polypeptide in the mixture with said compound of formula I;
- Optionally purifying or detecting the product.

30 The steps of said method may be performed in any order. Preferably, the steps of said method are performed in the above mentioned order.

The pH may optionally be adjusted to any pH useful for performing the method, for example the pH may be adjusted to in the range of 6 to 9.

35

In a preferred embodiment, the reacting step is performed in the absence of a catalyst.

The method may be performed in a solvent, which is a protic solvent mixture. In one embodiment, the solvent is an aqueous solvent mixture. The solvent may be a solvent
 5 containing material from living organisms e.g. body fluids and/or tissues. The body fluid may for example be plasma. Accordingly, said polypeptide may be provided in the form of a sample from an individual.

In other embodiments, the solvent may be a solvent compatible with hosting living
 10 organisms, e.g. a culture medium or a buffer with physiological salinity and pH.

Preferred examples of general reaction schemes for reacting compounds of formula I with a polypeptide are shown in Example 1 in Method A and Method B. The scheme shows reaction of a protected cysteine with fluoro-benzene, however, the scheme can
 15 easily be adapted to a reaction between any polypeptide comprising cysteine and any compound of formula I.

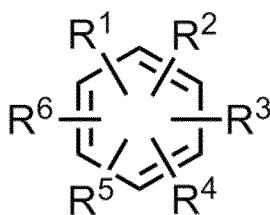
Another example of a general reaction schemes for reacting compounds of formula I with a polypeptide are shown in Example 1 in the general procedure (1) for S-arylation. This scheme can also be adapted as described above.

20

Compound of formula I

The invention relates to methods of reacting a compound of formula I with a cysteine residue. Furthermore, the invention relates to compounds of formula I per se.

25 The compound of Formula I, may be any compound having the following structure:



Formula I

30 wherein:


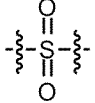

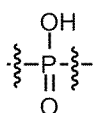
R^3 is a leaving group

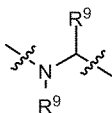
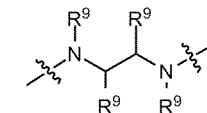
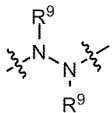
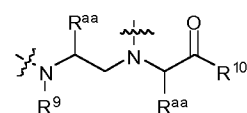
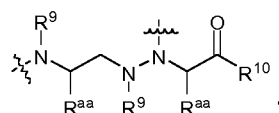
R^1 is R^7 ;

R^7 is selected from the group consisting of $-X-Z-R^8$, $-Z-T-R^8$, and $-Z-T-(R^8)_2$;

wherein

5 X is selected from the group consisting of a bond, $-\text{CH}_2-\text{N}(\text{R}^9)-$, and $-\text{CH}_2-\text{O}-$;

Z is selected from the group consisting of a bond, , , , and ;

T is selected from the group consisting of a bond, , , , , and ;

10 R^8 is individually selected from the group consisting of amino acids, amino acid derivatives, peptides, peptide derivatives, peptidomimetic, alkylamines, protecting groups, $-\text{NH}-\text{linker}-\text{Y}$, $-\text{linker}-\text{Y}$, and $-\text{linker}-\text{R}^{14}$, wherein said peptide optionally may be N- and/or C-terminally modified;

Y is a labelling molecule, a drug molecule or a prodrug;

15 R^9 is individually selected from the group consisting of $-\text{H}$, alkyl-COOH and an amino acid side chain;

R^{10} is selected from the group consisting of $-\text{OH}$ and R^8 ;

R^{14} is a reactive group;

20 R^2 , R^4 , R^5 , R^6 , are individually selected from the group consisting of $-\text{H}$, R^7 , an amino acid side chain and an electron withdrawing group;

under the proviso that at least one of R^2 , R^4 , R^5 , and R^6 is an electron withdrawing group.

25 In one embodiment, R^8 is individually selected from the group consisting of amino acids, amino acid derivatives, peptides, peptide derivatives, peptidomimetic, alkylamines, protecting groups, $-\text{NH}-\text{linker}-\text{Y}$, and $-\text{linker}-\text{Y}$, wherein said peptide optionally may be N- and/or C-terminally modified.

In one embodiment, no more than four of R^2 , R^3 , R^4 , R^5 , R^6 are $-F$.

In preferred embodiments of the invention R^2 , R^4 , R^5 , R^6 , are individually selected from the group consisting of $-H$, R^7 , an amino acid side chain and an electron withdrawing group with the proviso that at least one of R^2 , R^4 , R^5 , and R^6 is an electron withdrawing group, and that no more than four of R^2 , R^3 , R^4 , R^5 , and R^6 are $-F$.

As mentioned above, moieties comprising two waved lines or two free bonds may be linked to the compounds of the invention in any direction. Accordingly, individual X, Z and T described above may be placed in the compounds of the invention in any direction.

In one embodiment, R^1 is para to R^3 . In another embodiment, R^1 is ortho to R^3 .

In some embodiments, R^3 is selected from the group consisting of $-F$, $-Cl$, $-Br$, $-I$, $-NO_2$ and SO_2 -alkyl. Preferably, R^3 is selected from the group consisting of $-F$ and $-Cl$. In a preferred embodiment, R^3 is $-F$.

In one embodiment, the electron withdrawing groups individually are selected from the group consisting of $-F$, $-Cl$, $-CF_3$, $-CCl_3$, SO_2 -alkyl and $-C\equiv N$. For example, the electron withdrawing group may be selected from the group consisting of $-F$ and $-Cl$.

In one embodiment, at least one, such as at least two of R^2 , R^4 , R^5 , and R^6 are $-F$ or $-Cl$. In one embodiment, at least one, such as at least two, but at the most three of R^2 , R^4 , R^5 , and R^6 are $-F$ or $-Cl$. In another embodiment, all of R^2 , R^4 , R^5 and R^6 are $-H$, $-F$ or $-Cl$, with the proviso that at least one, such as at least two of R^2 , R^4 , R^5 , and R^6 are $-F$ or $-Cl$. In yet another embodiment, all of R^2 , R^4 , R^5 and R^6 are $-H$ or $-F$, with the proviso that at least two of R^2 , R^3 , R^4 , R^5 , and R^6 are $-F$. In one embodiment, at the most three of R^2 , R^4 , R^5 , and R^6 are $-F$ or $-Cl$. In another embodiment, all of R^2 , R^4 , R^5 and R^6 are $-H$, $-F$ or $-Cl$, with the proviso that at the most three of R^2 , R^4 , R^5 , and R^6 are $-F$ or $-Cl$. In yet another embodiment, all of R^2 , R^4 , R^5 and R^6 are $-H$ or $-F$, with the proviso that at the most three of R^2 , R^3 , R^4 , R^5 , and R^6 are $-F$.

In a preferred embodiment, all of R^3 , R^4 , R^5 and R^6 are $-F$, and R^2 is $-F$ or R^9 . In another preferred embodiment 3 of R^3 , R^4 , R^5 and R^6 are $-F$, one of R^3 , R^4 , R^5 and R^6 is $-H$ and R^2 is $-F$ or R^9 .

In one embodiment, at least one R^9 is $-H$. In another embodiment, all R^9 is $-H$.

R^8 may individually selected from the group consisting of amino acids, amino acid
5 derivatives, peptides, peptide derivatives, peptidomimetic, alkylamines, $-NH$ -linker- Y
and $-linker-Y$, wherein said peptide optionally may be N- and/or C-terminally modified.
In some embodiments R^8 may individually selected from the group consisting of amino
acids, peptides, alkylamines, $-NH$ -linker- Y and $-linker-Y$, wherein said peptide
optionally may be N- and/or C-terminally modified.

10

In one embodiment R^8 may individually be selected from the group consisting of $-NH$ -
linker- Y and $-linker-Y$.

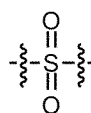
The linker may be an alkyl wherein one or more $-C(H_2)-$ have been replaced with $-$
15 $C(O)-$, $-N(H)-$, $-O-$, or $-S-$. In one embodiment, the linker is selected from the group
consisting of peptides, oligosaccharides and steroids. In a preferred embodiment, the
linker is or comprises $-N(H)-(PEG)_n$, wherein n is an integer from 0 to 10. In one
embodiment, the linker is or comprises $-N(H)-(CH_2-CH_2-O)_n-(CH_2)_m-$, wherein n is an
integer from 0 to 20 and m is an integer from 0 to 5. In one embodiment, the linker is or
20 comprises $-(CH_2-CH_2-O)_n-(CH_2)_m-$. Preferably, n is an integer from 4 to 20. In one
embodiment, n is an integer from 3 to 10, such as from 3 to 5, such as 3. Preferably, m
is 2. In one embodiment, the linker comprises a triazole moiety.

In one embodiment, R^8 is $-linker-R^{14}$, wherein R^{14} is a first reactive group. The term
25 "first reactive group" as used herein refers to groups that are capable, under suitable
conditions, of reacting with a second reactive group. Said first reactive group may be
reacted with any compound comprising the second reactive group. In one embodiment,
a compound comprising a second reactive group further comprises Y . In such
embodiments Y may be linked to the compound of formula I following a reaction
30 between the first reactive group and the second reactive group. The reaction between
a compound of Formula I wherein R^8 is $-linker-R^{14}$, and a compound comprising Y and
a second reactive group, may occur before or after the compound of Formula I has
reacted with said cysteine residue.

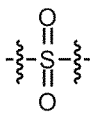
Preferably, the reactive group is an azide group or an alkyne. In one embodiment, said first reactive group is an azide group, and said second reactive group is an alkyne.

Said azide group may be reacted with the alkyne group to form a triazole moiety through a 1,3-dipolar cycloaddition, optional in the presence of a transition metal

- 5 catalyst. The reaction between a compound of Formula I, wherein R^8 is a linker with a terminal azide group, and a compound comprising Y and an alkyne group, may occur before or after the compound of Formula I has reacted with said cysteine residue, see e.g. Examples 3 and 9.



- 10 In some embodiments it may be preferred that Z is $\begin{array}{c} \text{O} \\ \parallel \\ \text{---S---} \\ \parallel \\ \text{O} \end{array}$. This may in particular be the case in embodiments, wherein at the most 4, such as at the most 3 of R^2 , R^3 , R^4 , R^5 , and R^6 are -F.



- 15 In a preferred embodiment, Z is $\begin{array}{c} \text{O} \\ \parallel \\ \text{---S---} \\ \parallel \\ \text{O} \end{array}$, and at the most 4, such as at the most 3, of R^2 , R^3 , R^4 , R^5 , and R^6 are -F.

The labelling molecule Y may be any labelling molecule. In particular, the labelling

- 20 molecule may be a detectable label. For example, the labelling molecule Y may be selected from the group consisting of a fluorescent labelling molecule, a radioisotope labelling molecule, an affinity molecule, an azide, a terminal alkyne, a spin label, a prodrug, a mass tag, and a photoreactive group.

- 25 In one embodiment Y is a protective group, for example a protective group selected from the group consisting of Boc and Fmoc.

In one embodiment, Y is a reactive group, which can be detected by reacting the reactive group with a directly detectable compound. For example, the labelling

- 30 molecule may be azide, which can be detected by reacting the azide moiety with a

directly detectable compound by click chemistry. The directly detectable compound may for example be a fluorescent molecule or a dye, such as cyanine dye.

In a preferred embodiment, the labelling molecule Y is biotin.

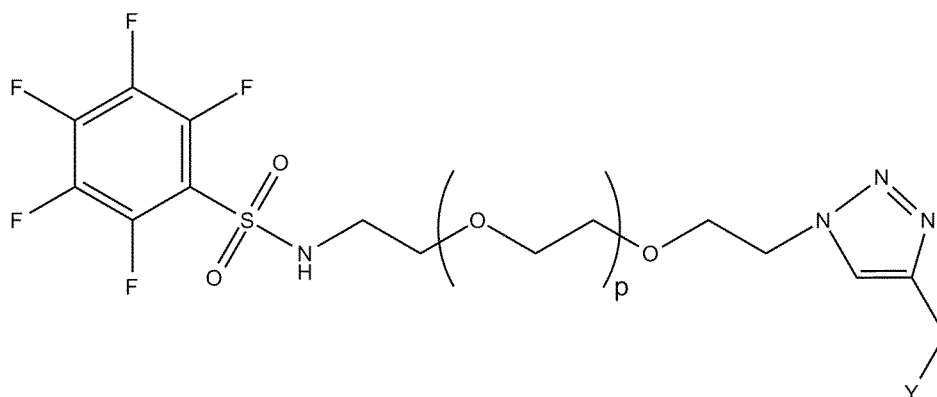
5

The fluorescent labelling molecule may be any fluorescent molecule, e.g. rhodamine or sulforhodamine.

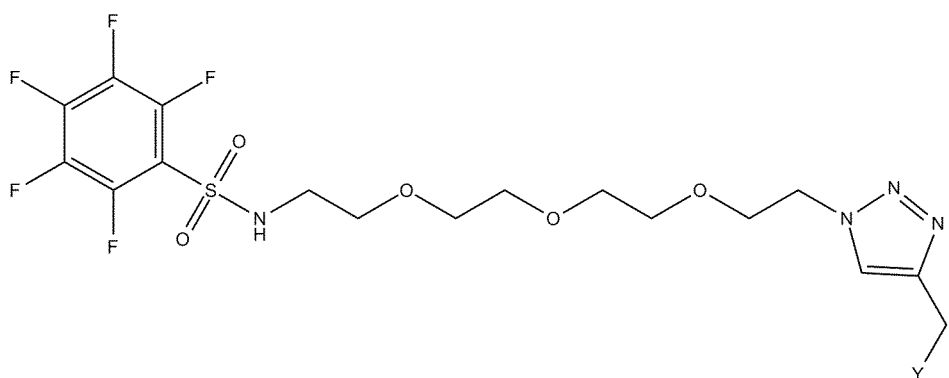
In some embodiments, the Y is a prodrug or a drug molecule. Preferably, said prodrug
10 or drug molecule is a small molecule. The person of skill in the art is well aware of
prodrugs and drug molecules that are suitable as Y of the present invention. For
example, such a small molecule may be ibuprofen, as in Example 9.

15 In some embodiments, R⁸ may individually be selected from the group consisting of amino acids and peptides. In a preferred embodiment, R⁸ is a peptide. In some embodiments, said peptide is linked to the benzene, Z or T via the N- terminus or the C-terminus.

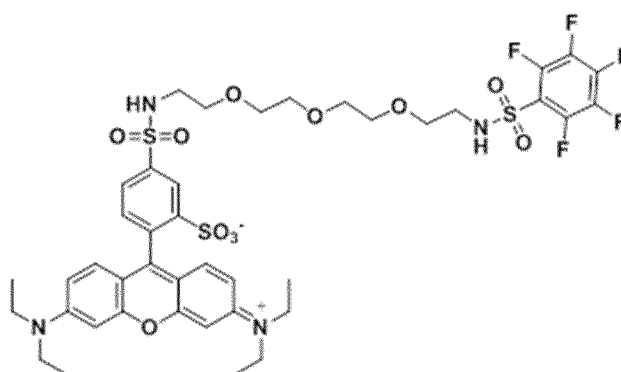
In one embodiment, the compound of Formula I is:



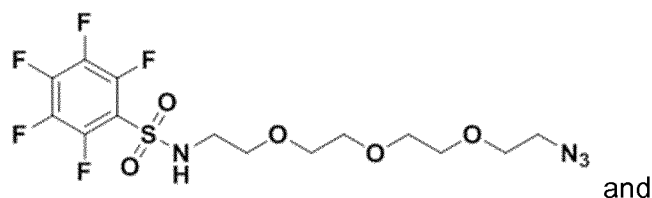
wherein p is an integer from 0 to 10, such as 1, such as 2, such as 3, such as 4, such as 5, such as 6, such as 7, such as 8, such as 9, such as 10. Thus, in one embodiment, the compound of Formula I is:



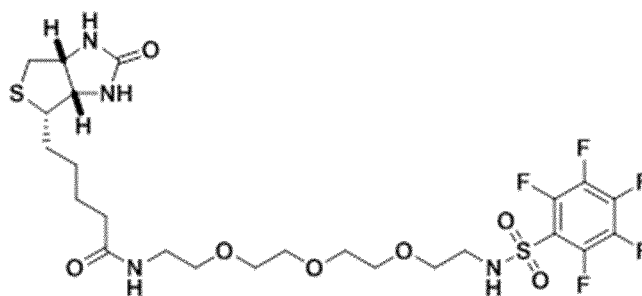
In one embodiment, the compound of Formula I is selected from the group consisting of:



5



and



In one embodiment, the compound of Formula I is selected from the group consisting of compounds **2g**, **8a**, **8b**, **8c**, **8d**, **10**, and **11** described in the Examples below. In one embodiment of the invention, the compound of formula I is selected from the group consisting of compounds **22c**, **22d**, **22e**, **22f**, **23** and **24** described in the Examples below.

10

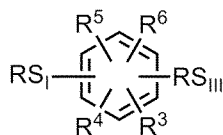
Compounds of formulas II to X

The compound of formula I may in preferred embodiments be any of the compounds of one of the more specific formulas II, III, IV, VI, VII, VIII, IX or X described in this section.

In one embodiment, the benzene of the compound of Formula I is integrated into the backbone of a peptide, which may optionally be C- or N-terminally modified. In particular, the benzene may be integrated into the backbone of a peptide, wherein the peptide may comprise a recognition sequence for a peptide cleaving enzyme in a manner, wherein said benzene replaces at least part of said recognition sequence. Said recognition sequence may for example be represented by the general formula $RS_I-RS_{II}-RS_{III}$. RS_{II} may for example comprise the cleavage site. The benzene of the compound of Formula I may in some embodiments be incorporated into a peptide comprising or consisting of the sequence of the sequence $RS_I-RS_{II}-RS_{III}$, wherein the benzene replaces RS_{II} , or RS_I-RS_{II} or $RS_{II}-RS_{III}$. Preferably, the peptide $RS_I-RS_{II}-RS_{III}$ is a substrate for a peptide cleaving enzyme.

In one embodiment, R^1 of the compound of Formula I is RS_I and R^2 of the compound of Formula I is RS_{III} , wherein RS_I and RS_{III} individually are selected from the group consisting of amino acids and peptides, wherein said amino acid or peptide optionally may be N- or C-terminally modified. As described above, when said RS_I is linked to said RS_{III} via RS_{II} , which is selected from the group consisting of a bond, amino acids and peptides, they together form the peptide $RS_I-RS_{II}-RS_{III}$, which is a substrate for an enzyme. Preferably, said enzyme is a peptide cleaving enzyme.

Thus, the compound of Formula I may be a compound of the general formula II



Formula II

30

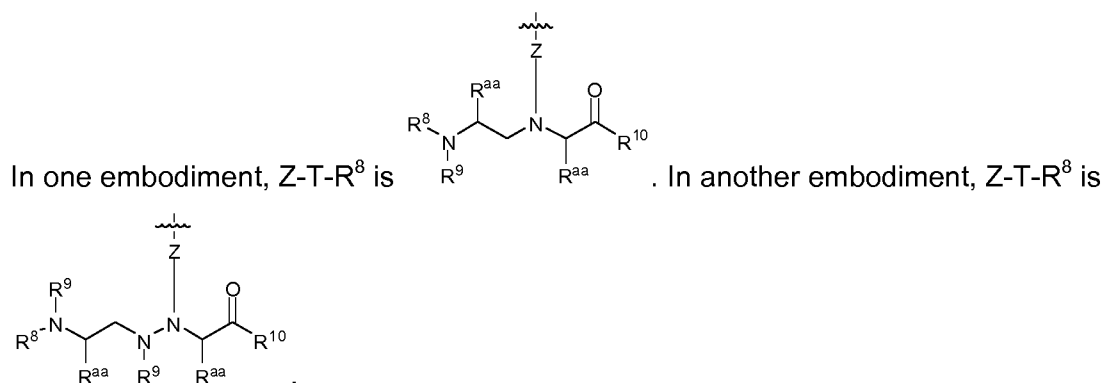
wherein

RS_I when linked to RS_{III} via RS_{II} together forms the peptide $RS_I-RS_{II}-RS_{III}$, which is a substrate for a peptide cleaving enzyme; and

RS_I and RS_{III} individually are selected from the group consisting of amino acids and peptides, wherein said amino acid or peptide may be N- or C-terminally modified; and

RS_{II} is selected from the group consisting of a bond, amino acids and peptides;

5 and R³, R⁴, R⁵ and R⁶ are as defined in item 1.

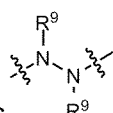


10 In a preferred embodiment, R¹⁰ is -OH.

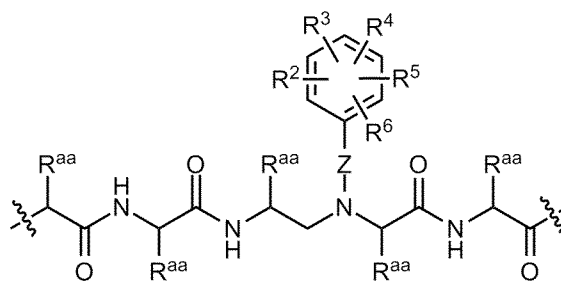
In one embodiment, R⁸ is individually selected from the group consisting of amino acids, amino acid derivatives, peptides, peptide derivatives, peptidomimetic, alkylamines, protecting groups, -NH-linker-Y and -linker-Y.

In one embodiment, R⁸ is individually selected from the group consisting of amino acids, peptides and alkylamines, wherein said amino acid or peptide optionally may be N- and/or C-terminally modified. Preferably, R⁸ is an amino acid or a peptide. Said peptide may consist of 2 to 5 amino acids, such as 2 to 3 amino acids.

In one embodiment, R⁸ is a protecting group, which may be selected from the group consisting of Boc and Fmoc.

In a preferred embodiment, T is  and one R⁹ is -H and the other R⁹ is alkyl-COOH, preferably, -CH₂-COOH.

In a preferred embodiment, the compound of Formula I is bonded to a nitrogen of the peptide backbone as in Formula X:



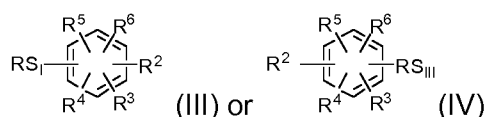
Formula X

5

In a further embodiment, R¹ of the compound of Formula I is RS_I or RS_{III}, wherein RS_I and RS_{III} individually are selected from the group consisting of amino acids and peptides, wherein said amino acid or peptide optionally may be N- or C-terminally modified. As described above, when said RS_I is linked to said RS_{III} via RS_{II}, which is selected from the group consisting of a bond, amino acids and peptides, they together form the peptide RS_I-RS_{II}-RS_{III}, which is a substrate for an enzyme. Preferably, said enzyme is a peptide cleaving enzyme.

15

Thus, in one embodiment, the benzene of the compound of Formula I is linked to the N- or C-terminal of a peptide, such as depicted in the general formula III or IV:



20

wherein

RS_I when linked to RS_{II}-RS_{III} or RS_{III} when linked to RS_I-RS_{II} together forms the peptide RS_I-RS_{II}-RS_{III}, which is a substrate for a peptide cleaving enzyme; and RS_I and RS_{III} individually are selected from the group consisting of amino acids and peptides, wherein said amino acid or peptide may be N- or C-terminally modified; and

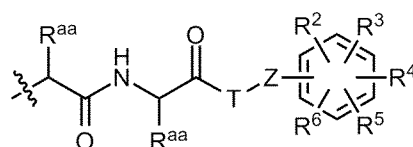
25

RS_{II} is selected from the group consisting of a bond, amino acids and peptides; and R², R³, R⁴, R⁵ and R⁶ are as defined in item 1.

RS_I-RS_{II}-RS_{III} denotes a peptide sequence, which is a substrate for a peptide cleaving enzyme. RS_I-RS_{II}-RS_{III} may be a substrate for any peptide cleaving enzyme, for example it may be a substrate for any of the hydrolases containing a cysteine in its active site described herein below in the section "Polypeptides".

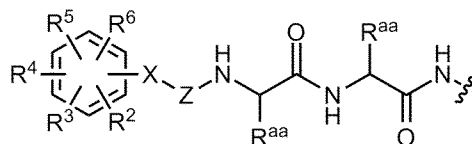
5

In one embodiment, the compound of Formula I is a compound of any one of Formulas VI to VIII:

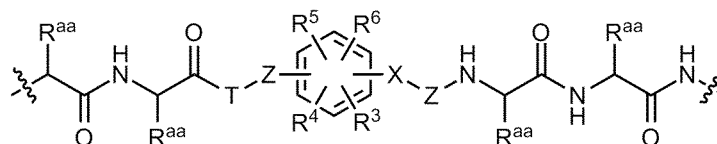


Formula VI

10



Formula VII



Formula VIII

15

wherein R^{aa} is individually selected from the group consisting of amino acid side chains. In a preferred embodiment, R^{aa} is individually selected from the group consisting of the side chains of the 20 standard amino acid.

20 The waved lines indicate points of attachment. For example, the waved line may indicate a point of attachment to an amino acid or a peptide, wherein said amino acids of peptides optionally may be C- or N-terminally modified.

Polypeptides

25 The present invention provides methods for reacting a compound of Formula I with a cysteine residue, thereby forming a covalent bond, wherein the cysteine is contained in the sequence of a polypeptide.

Said polypeptide may be any polypeptide comprising at least one cysteine in its sequence. Thus, the polypeptide may also be referred to as a "polypeptide comprising the cysteine residue in its sequence".

- 5 In some embodiments, the polypeptide comprising the cysteine residue in its sequence is an enzyme. In particular, the polypeptide comprising the cysteine residue in its sequence may be an enzyme. In a preferred embodiment, said cysteine residue is positioned within the active site of the enzyme, e.g. a cysteine protease. Accordingly, the polypeptide comprising the cysteine residue in its sequence may be a hydrolase
10 containing a cysteine in its active site. In preferred embodiments the polypeptide comprising the cysteine residue in its sequence may be a cysteine protease.

- In such embodiments the methods may preferably comprise reacting the cysteine residue positioned in the active site of said hydrolase (e.g. said protease) with a
15 compound of Formula I.

- In some embodiments, the hydrolase containing a cysteine in its active site may be selected from the group consisting of cathepsin B (EC 3.4.22.1), papain (EC 3.4.22.2), ficain (EC 3.4.22.3), chymopapain (EC 3.4.22.6), asclepain (EC 3.4.22.7), clostripain
20 (EC 3.4.22.8), cerevisin (EC 3.4.21.48), streptopain (EC 3.4.22.10), insulysin (EC 3.4.24.56), γ -glutamyl hydrolase (EC 3.4.19.9), actinidain (EC 3.4.22.14), cathepsin L (EC 3.4.22.15), cathepsin H (EC 3.4.22.16), prolyl oligopeptidase (EC 3.4.21.26), thimet oligopeptidase (EC 3.4.24.15), proteasome endopeptidase complex (EC 3.4.25.1), saccharolysin (EC 3.4.24.37), kexin (EC 3.4.21.61), Cathepsin T (EC
25 3.4.22.24), Glycyl endopeptidase (EC 3.4.22.25), Cancer procoagulant (EC 3.4.22.26), cathepsin S (EC 3.4.22.27), picornain 3C (EC 3.4.22.28), picornain 2A (EC 3.4.22.29), Caricain (EC 3.4.22.30), Ananain (EC 3.4.22.31), Stem bromelain (EC 3.4.22.32), Fruit bromelain (EC 3.4.22.33), Legumain (EC 3.4.22.34), Histolysain (EC 3.4.22.35), caspase-1 (EC 3.4.22.36), Gingipain R (EC 3.4.22.37), Cathepsin K (EC 3.4.22.38),
30 adenain (EC 3.4.22.39), bleomycin hydrolase (EC 3.4.22.40), cathepsin F (EC 3.4.22.41), cathepsin O (EC 3.4.22.42), cathepsin V (EC 3.4.22.43), TEV nuclear-inclusion-a endopeptidase (EC 3.4.22.44), helper-component proteinase (EC 3.4.22.45), L-peptidase (EC 3.4.22.46), gingipain K (EC 3.4.22.47), staphopain (EC 3.4.22.48), separase (EC 3.4.22.49), V-cath endopeptidase (EC 3.4.22.50), cruzipain
35 (EC 3.4.22.51), calpain-1 (EC 3.4.22.52), calpain-2 (EC 3.4.22.53), calpain-3 (EC

3.4.22.54), caspase-2 (EC 3.4.22.55), caspase-3 (EC 3.4.22.56), caspase-4 (EC 3.4.22.57), caspase-5 (EC 3.4.22.58), caspase-6 (EC 3.4.22.59), caspase-7 (EC 3.4.22.60), caspase-8 (EC 3.4.22.61), caspase-9 (EC 3.4.22.62), caspase-10 (EC 3.4.22.63), caspase-11 (EC 3.4.22.64), peptidase 1 (mite) (EC 3.4.22.65), calicivirin
5 (EC 3.4.22.66), zingipain (EC 3.4.22.67), Ulp1 peptidase (EC 3.4.22.68), SARS coronavirus main proteinase (EC 3.4.22.69), sortase A (EC 3.4.22.70), sortase B (EC 3.4.22.71), and cathepsin X (EC 3.4.18.1).

10 In one embodiment, the polypeptide comprising the cysteine residue in its sequence is a human protein.

In one embodiment, the polypeptide comprising the cysteine residue in its sequence is an albumin, for example human serum albumin.

15 In one embodiment, the polypeptide comprising the cysteine residue in its sequence is an antibody or an antigen-binding fragment, which are capable of binding to a specific antigen via an epitope on the antigen. By “an antibody or an antigen-binding fragment” according to the invention is a polypeptide or protein capable of recognising and binding an antigen, said polypeptide comprising at least one antigen binding site. Said
20 antigen binding site preferably comprises at least one CDR. The antibody may be a naturally occurring antibody, a fragment of a naturally occurring antibody or a synthetic antibody.

25 The term “naturally occurring antibody” refers to heterotetrameric glycoproteins capable of recognising and binding an antigen and comprising two identical heavy (H) chains and two identical light (L) chains inter-connected by disulfide bonds. Each heavy chain comprises a heavy chain variable region (abbreviated herein as V_H) and a heavy chain constant region (abbreviated herein as C_H). Each light chain comprises a light chain variable region (abbreviated herein as V_L) and a light chain constant region
30 (abbreviated herein as C_L). The V_H and V_L regions can be further subdivided into regions of hypervariability, termed complementarity determining regions (CDRs), interspersed with regions that are more conserved, termed framework regions (FRs). Antibodies may comprise several identical heterotetramers.

The antibody may in particular be a monoclonal antibody. The antibody may also be a humanised antibody or a human antibody. Thus, the antibody may be a monoclonal, humanised or human antibody.

5 Antibodies according to the invention may for example be monoclonal antibodies, chimeric antibodies, humanised antibodies, isolated human antibodies, single chain antibodies, bi-epitopic antibodies, bispecific antibodies, antibody heavy chains, antibody light chains, homodimers and heterodimers of antibody heavy and/or light chains, and antigen-binding fragments and derivatives of the same. Suitable antigen-
10 binding fragments and derivatives include, but are not necessarily limited to, Fv fragments (e.g. single chain Fv and disulphide-bonded Fv), Fab-like fragments (e.g. Fab fragments, Fab' fragments and F(ab)₂ fragments), single variable domains (e.g. VH and VL domains) and domain antibodies (dAbs, including single and dual formats [i.e. dAb-linker-dAb]).

15

The techniques and procedures described or referenced herein are generally well understood and commonly employed using conventional methodology by those skilled in the art, such as, for example, the widely utilized methodologies described in *The Antibodies* (M. Zanetti and J. D. Capra, eds., Harwood Academic Publishers, 1995).

20

In embodiments of the invention where the polypeptide is a peptide cleaving enzyme, e.g. a cysteine protease, the compound of Formula I may comprise at least part of the recognition sequence for said peptide cleaving enzyme. For example, the compound of Formula I may be a compound of any one of Formulas II, III or IV, wherein RS_I-RS_{II}-
25 RS_{III} together forms a recognition sequence for said peptide cleaving enzyme.

In one embodiment of the invention, the polypeptide is a peptide cleaving enzyme. For example the polypeptide may be papain. When the polypeptide is papain, the compound of formula I may for example be compound **2g** described in the Examples below. For example the polypeptide may be TEV protease. When the polypeptide is
30 TEV protease the compound of formula I may for example be selected from the group consisting of compounds **2g**, **2h**, **8a**, **8b** and **8c** described in the Examples below. For example the polypeptide may be caspase-1. When the polypeptide is caspase-1 the compound of formula I may for example be selected from the group consisting of compounds **22c**, **22e**, **22f**, **23** and **24** described in the Examples below. In particular

the compound of formula I may be selected from the group consisting of compounds **23** and **24** described in the Examples below.

Method of modulating the activity of a protein

5 In one aspect, the current invention concerns a method of modulating the activity of a protein. The methods of modulating the activity of a protein in general involve reacting a compound of Formula I (e.g. any of the compounds described herein above in the sections "Compound of Formula I" and "Compounds of Formula II to X) with a cysteine
10 residue contained in the sequence of a polypeptide constituting at least part of said protein. Thus, said protein may comprise one or more polypeptides, and the methods may involve reacting at least one cysteine residue contained in at least one of said polypeptides, with a compound of Formula I.

In one embodiment, the protein is an enzyme. Preferably, said enzyme is an enzyme
15 with a cysteine residue in the active site. In a preferred embodiment, the enzyme is a hydrolase comprising a cysteine in the active site, for example the enzyme may be a cysteine hydrolase. In one preferred embodiment, the enzyme is a cysteine protease.

In such embodiments the methods may preferably comprise reacting the cysteine
20 residue positioned in the active site of said hydrolase (e.g. said protease) with a compound of Formula I.

Accordingly, it is preferred that the compound of Formula I is capable of binding to the active site of said enzyme, for example said cysteine hydrolase, such as said cysteine
25 protease. For example, if the polypeptide is a cysteine protease, the compound of Formula I may comprise at least part of the recognition sequence for said cysteine protease. For example, the compound of Formula I may be a compound of any one of Formulas II, III or IV, wherein RS_I - RS_{II} - RS_{III} together forms a recognition sequence for said peptide cleaving enzyme.

30 The method of modulating the activity of a protein, may modulate the activity in any manner, however, in particular, the method may be a method of reducing or even inhibiting the activity of said protein, e.g. a method reducing or even inhibiting the activity of said enzyme, such as a method of reducing or even inhibiting the activity of
35 said cysteine hydrolase.

Method of treatment

In one aspect, the present invention concerns a method for treatment or prevention of a clinical condition associated with a polypeptide comprising a cysteine residue in an individual in need thereof. Such methods in general involve reacting a compound of
5 Formula I (e.g. any of the compounds described herein above in the sections “Compound of Formula I” and “Compounds of Formula II to X) with a cysteine residue contained in the sequence of the polypeptide associated with the clinical condition. Thus, the method may comprise administering a compound of Formula I capable of reacting with said polypeptide to said individual in a therapeutically effective amount.

10

The clinical condition may be associated with a polypeptide in various manners. For example the clinical condition may be associated with the presence of said polypeptide. The clinical condition may also be associated with elevated levels of said polypeptide.

15 The clinical condition may also be associated with aberrant activity of said polypeptide. For example, the polypeptide may be a protein or may constitute part of a protein, in which case, the clinical condition may be associated with aberrant activity of said protein, e.g. of an enzyme. In some embodiments, the clinical condition may be associated with elevated activity of said protein, e.g. of an enzyme.

20

In such embodiments, the methods of treatment or prevention of a clinical condition may comprise modulating the activity of said protein, for example by any of the methods of modulating activity of a protein described herein above in the section “Method of modulating the activity of a protein”.

25

The clinical condition may be any clinical condition associated with a polypeptide. The polypeptide may for example be any of the polypeptides described herein above in the section “Polypeptides”. In one embodiment the clinical condition is selected from the group consisting of osteoporosis, cancer and arthritis. In such embodiments, the
30 polypeptide may for example be a cysteine protease, such as cathepsin B, K or L or caspase 1.

In one aspect, the method of treatment or prevention of a clinical condition involves transporting a prodrug to the site of the clinical condition. In such embodiments, the
35 methods may involve use of a compound of Formula I, wherein Y is a prodrug. Said

prodrug preferably is a prodrug of a drug useful in the treatment of prevention of said clinical condition.

Method of labelling

5 In one aspect, the present invention concerns a method for labelling a polypeptide with a label, e.g. with a labelling molecule. The methods of labelling a polypeptide in general involve reacting a compound of Formula I with a cysteine residue contained in the sequence of said polypeptide, wherein the compound of Formula I comprises a labelling molecule.

10

Thus, in embodiments of the invention concerning labelling a polypeptide, then the compound of Formula I may for example be any of the compounds of Formula I described herein above in the sections "Compound of Formula I" and "Compounds of Formula II to X", wherein R^8 is selected from the group consisting of -NH-linker-Y and -linker-Y, wherein Y is a labelling molecule.

15

The molecule Y may for example be selected from the group consisting of a fluorescent labelling molecule, a radioisotope labelling molecule, an affinity molecule, an azide, a terminal alkyne, a spin label, a prodrug, a mass tag, and a photoreactive group. For example, the labelling molecule may be a detectable label.

20

The methods may be used for labelling any polypeptide comprising a cysteine residue. For example, the methods may be used for labelling any of the polypeptides described herein above in the section "Polypeptides".

25

Once said polypeptide has been labelled, said label may be used for various purposes. For example, the label may be used for detecting the polypeptide. Accordingly, in one aspect, the current invention concerns a method for detecting a polypeptide, said method comprises the steps of

30

- a. Reacting said polypeptide with a compound of Formula I, wherein R^8 is -linker-Y or -NH-linker Y; and
- b. detecting Y.

35

In other embodiments, the labelling molecule may be used in a method of isolation of said polypeptide. Thus, said polypeptide may for example be isolated using an isolation

method based on affinity isolation using a moiety with affinity for said labelling molecule. E.g. if the labelling molecule is tag, the method of isolation may be based on use of antibodies specifically binding said tag. If the labelling molecule is biotin said method of isolation may be based on use of streptavidin and/or avidin.

5

Compounds of formula I useful for methods of labelling include compounds of Formula I, wherein R^8 is -linker-Y or -NH-linker Y. Non-limiting examples of such compounds includes compounds **8a**, **8b**, **8c**, **8d**, **10**, **11**, **17** and **18** described in the Examples below, and in particular compounds **8a**, **8b**, **8c**, **8d**, **10** or **11**.

10

Diagnostic method

In one aspect, the present invention concerns a method for diagnosis of a clinical condition associated with a polypeptide in an individual at risk of acquiring said clinical condition.

15

In general, said method comprises the steps of:

- a. incubating a compound of Formula I with a sample from said individual, wherein said compound of Formula I is capable of reacting with a cysteine residue in said polypeptide associated with said clinical condition, and wherein said compound of Formula I otherwise may be any of the compounds of Formula I described herein above in the sections "Compound of Formula I" and "Compounds of Formula II to X", wherein R^8 is selected from the group consisting of -NH-linker-Y and -linker-Y, and wherein Y is a labelling molecule; and

20

25

- b. detecting Y;

thereby determining the presence, absence and/or level of said polypeptide in said sample.

- Step a. of said method may for example be performed under conditions allowing for reaction between the compound of Formula I and a cysteine residue in said polypeptide.

30

The clinical condition may be associated with a polypeptide in various manners. For example the clinical condition may be associated with the presence of said polypeptide,

35

in which case the method may determine the presence of said polypeptide in said sample. The presence of said polypeptide in the sample may be indicative of said individual suffering from said clinical condition.

- 5 The clinical condition may also be associated with elevated levels of said polypeptide in which case the method may determine the level of said polypeptide in said sample. An elevated level of said polypeptide in the sample may be indicative of said individual suffering from said clinical condition.
- 10 The clinical condition may also be associated with the absence of said polypeptide, in which case the method may determine the presence of said polypeptide in said sample. The absence of said polypeptide in the sample may be indicative of said individual suffering from said clinical condition.
- 15 The clinical condition may also be associated with the reduced levels of said polypeptide, in which case the method may determine the levels of said polypeptide in said sample. Reduced levels of said polypeptide in the sample may be indicative of said individual suffering from said clinical condition.
- 20 The methods for determining the presence, absence or level of said polypeptide may for example comprise performing on a sample from said individual any of the methods of labelling said polypeptide described herein above in the section "Method of labelling" followed by detecting the labelling molecule.
- 25 The skilled person will be aware of useful methods for detection of the labelling molecule depending on the labelling molecule employed.

Method of preparing compounds

30

The compounds of formula I may be prepared by any useful method available to the skilled person.

- The methods may comprise preparing a benzene substituted with one or more electron withdrawing groups, e.g. a fluoro-benzene. This may also be referred to as the aryl part
- 35 of the compound.

The aryl part of the compound, i.e. the benzene part of a compound of Formula I including T, Z, X and/or non-peptide R¹-R⁶, can be synthesized according to conventional methods.

5

If the compound comprises a fluoro-benzene moiety, the fluoro-benzene may be prepared by conventional methods or they may be purchased.

10 In some embodiments, a fluorophenyl acid or a fluorophenyl acid chloride may be used as starting compound. Such compounds are commercially available. The starting compound may be reacted with additional substituents R¹ to R⁶ according to methods known to the skilled person.

15 If the compound of formula I comprises an amide bond, this may be prepared according to the general procedure (2) for amide bond formation described in Example 1 below.

Preparation of compounds of Formula I not comprising a peptide part can be conducted by conventional methods, such as those presented in example 6.

20

To prepare compounds of Formula I comprising peptide part(s), the peptide part(s) of the compound and the aryl part of the compound can be synthesized separately, and then coupled together. The peptide part of the compounds can be prepared by conventional solid-phase peptide synthesis. Such methods are well known to the skilled person and may for example be standard SPPS synthesis or it may be conducted essentially as described in Examples 2 or 8 herein below.

25

The aryl part may be prepared using substituted a phenyl acid or a phenyl acid chloride substituted with the appropriate electron withdrawing groups, e.g. fluorophenyl acid or fluorophenyl acid chloride as starting compound. The starting compounds may be reacted with e.g. substituent T, wherein T may be protected. For example the starting compounds may be reacted with a protected amino-amino compound yielding an intermediate compound. The protected compound may comprise one or more protecting groups, e.g. a benzyl and/or Boc This may be done as described in the General Procedure A described in Example 8.

35

The intermediate compound be deprotected by removal of one or more of the protecting groups, e.g. by removal of the benzyl. This may be done as described in the general procedure B in Example 8. The aryl part thus prepared may be useful as a
5 compound of formula I or it may further be coupled to a peptide part.

The coupling between the peptide part and the aryl part may be done by esterification of the aryl part to a HMBA linker followed by deprotection, e.g. deprotection of the Boc group. Then reductive amination of the terminal amine with an boc protected amino
10 acid derivative may be performed followed by standard SPPS protocol. Alternatively, instead of preparing the peptide, the individual amino acids of the peptide may be added one after the other to the aryl part by standard SPPS protocol. The amino acid derivative may be an amine, e.g. an amino acid lacking the carboxyl group.

15

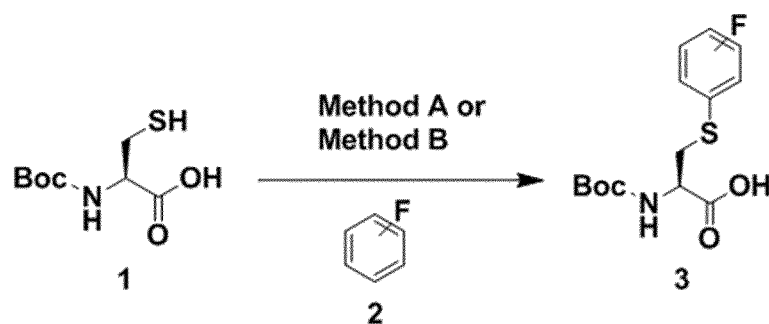
Examples

The invention is further illustrated by the following examples, which however should not be construed as being limiting for the invention.

20

Example 1 – Tuning fluorobenzene derivatives towards cysteine arylation

In order to investigate the feasibility of using fluorobenzenes as tunable reactive probes, screening of a wide range of different derivatives was performed. It was
25 hypothesized that the reactivity of the ideal compound for selectivity among different cysteines in a protein would react fully with cysteine under basic conditions in promoting organic solvents such as DMF, but selectively in aqueous buffer at only slightly basic conditions. Hence, the reactivity towards the model substrate tertbutyloxycarbonyl protected cysteine (Boc-L-cysteine, **1**) was assessed under two
30 sets of reaction conditions (see Scheme 1).

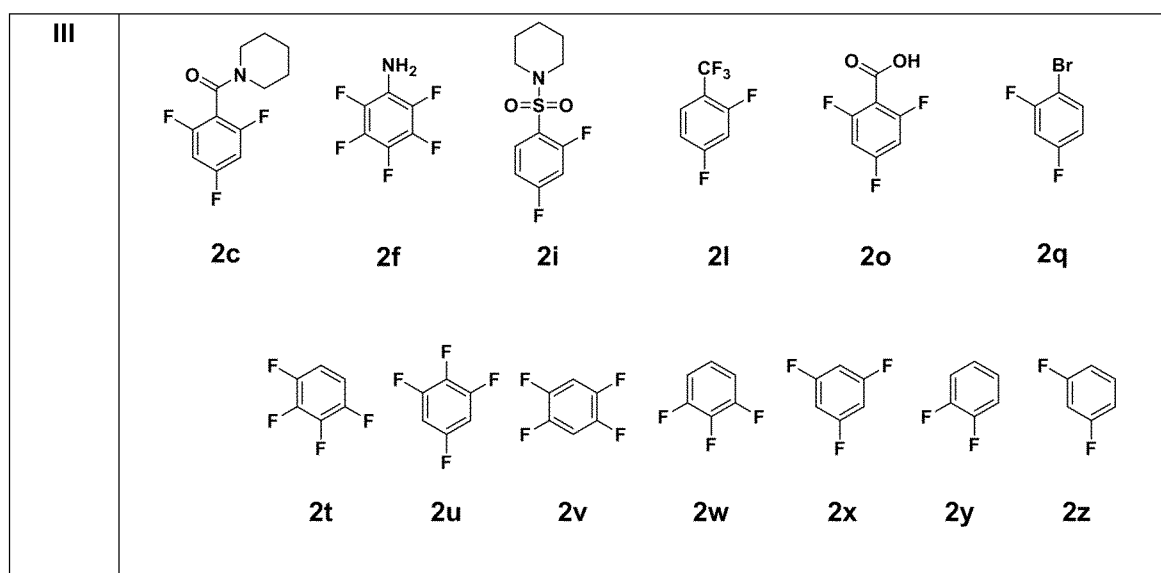


Scheme 1. Screening of Boc-L-Cys-OH (**1**) toward fluorobenzenes in different conditions. **Method A:** The fluoroaryl compound (0.06 mmol) was added to a solution of **1** (11.5 mg, 0.05 mmol) in DMF (0.5 mL), DIPEA (350 μ L, 0.24 mmol). After 16 h at 22 °C, the reaction was evaluated by LC-MS. **Method B:** The fluoroaryl compound (0.06 mmol) was added to a solution of **1** (11.5 mg, 0.05 mmol) in AcN/BPS buffer, pH=7-11, (0.5 mL). After 16 h at 22 °C, the reaction was evaluated by LC-MS.

First, each of 26 selected fluorobenzene derivatives (compounds **2a-z**, see Table 1) were tested for reaction with Boc-L-cysteine in DMF using diisopropylethyl amine (DIPEA) as base (method A). The reactions were followed by LC-MS. Only 13 compounds displayed reactivity (groups I and II of Table 1). The remaining derivatives did not react under the same conditions.

Table 1. Screened fluoroaryls (**2**). Compounds in row I (fluoroaryls with highest reactivity) react both in conditions A and B. Compounds in row II (fluoroaryls with moderate reactivity) reacts only in condition B. Compounds in row III (fluoroaryls with lowest reactivity) don not react in either condition A or B.

Group	Compounds
I	<p>2a 2d 2g 2j 2m</p>
II	<p>2b 2e 2h 2k 2n 2p 2r</p>



The 13 reactive fluorobenzenes (groups I and II of Table 1) were then tested for reactivity in an aqueous mixture of AcN/water (1:1) at pH 8.5 (method B). Only five candidates (**2a**, **2d**, **2g**, **2j** and **2w**) showed partial reactivity that increased with

5 increased pH (see Figure 2). The reactivity of the pentafluorobenzene derivatives could be directly linked to the electron withdrawing capacity of the last substituent. The nitro-, cyano-, sulfonamide, and trifluoromethyl-substituents all promoted reactivity both in water and DMF, and reactivity correlated to a Hammett σ_p -constant above 0.5. On the other hand amidoyl-, carboxy-, chloro- or bromo-substituted pentafluorobenzene

10 derivative did not react in water, but only in DMF. These substituents have a Hammett σ_p -constant between 0.5 and 0.2. Pentafluorobenzene with a Hammett σ_p -constant for the hydrogen substituent of 0.0 reacted only slowly in DMF (**2e**) and as expected, pentafluoroaniline (σ_p -constant of $-\text{NH}_2 = -0.63$) was completely unreactive under any conditions. Reducing the number of fluoro-substituents from five to three reduced the

15 reactivity significantly for all derivatives. The sulfonamide- and trifluoromethyl-substituted trifluorobenzenes were not reactive with unactivated sulfhydryl groups in water opposite to their corresponding pentafluorobenzene derivatives (**2h** and **2k**). The amidoyl-substituted trifluorobenzene did not react with unactivated sulfhydryl groups under any of the tested conditions while the corresponding pentafluorobenzene reacted

20 in DMF (**2b** and **2c**). Finally, the reactivity may be reduced even further by omitting additional fluoro-substituents. Thus, the sulfonamide or trifluoromethyl-substituted difluorobenzene were not reactive even in DMF (**2i** and **2l**). This initial screening experiment indicated that the reactivity of fluoroaryls towards cysteine could accurately

controlled by changing the number of fluorine atoms and the electron-withdrawing properties of other substituents.

General procedure (1) for S-arylation

- 5 To a solution of Boc-L-cysteine (110.5 mg, 0.5 mmol) in DMF (5.0 mL) DIPEA (350 μ L 2.4 mmol), and the appropriate fluoroaryl derivative (0.6 mmol) was added. The reaction was stirred at room temperature for 16 hours or until complete conversion as monitored by TLC. The solvent was evaporated and the crude residue was purified by column chromatography (MeOH: DCM / 0-10:100-90) to afford the product.

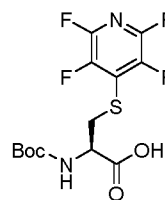
10

General procedure (2) for amide bond formation

- To a solution of the amine (0.5 mmol, 1 equiv.) in dichloromethane (5.0 ml), DIPEA (350 μ L, 2 mmol, 4 equiv.), and acid chloride (0.75 mmol, 1.5 equiv.) were added in that order. The reaction was stirred at room temperature for 3 hours. The solvent was
15 evaporated and the crude residue was purified by column chromatography to afford the product.

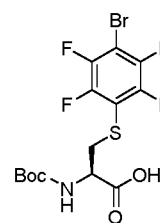
***N*-(tert-butoxycarbonyl)-*S*-(tetrafluoropyridin-4-yl)-*L*-cysteine (3m)**

- 20 Synthesized by general procedure (1) using Boc-L-cysteine (110.5 mg, 0.5 mmol) in DMF (5.0 mL), DIPEA (350 μ L, 2.4 mmol) and pentafluoropyridine (66.0 μ L, 0.6 mmol,) to afford **3m** as white solid (166 mg, 90 %), ^1H NMR (500 MHz, DMSO- d_6) δ 6.56 (s, 1H), 3.69 (dd, J = 13.4, 4.2 Hz, 1H), 3.41 (d, J = 7.9 Hz, 1H), 3.38 (d, J = 7.9 Hz, 1H), 1.32 (s, 9H); ^{13}C NMR (126 MHz, DMSO) δ 171.62, 154.81, 143.65, 143.51, 143.39, 142.53, 142.27, 141.72, 141.59, 141.47, 141.45, 140.52, 140.25, 131.48, 131.34, 131.20, 77.99, 54.93, 35.60, 27.93; ^{19}F NMR (470 MHz, DMSO- d_6) δ -93.42 – -93.59 (m), -136.86 – -137.03 (m); MS (ESI+) m/z calcd. for $\text{C}_{13}\text{H}_{15}\text{F}_4\text{N}_2\text{O}_4\text{S}^+$ $[\text{M}+\text{H}]^+$ 371.0683, found 371.068773.



- 30 ***S*-(4-bromo-2,3,5,6-tetrafluorophenyl)-*N*-(tert-butoxycarbonyl)-*L*-cysteine (3s)**

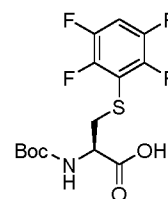
- Synthesized by general procedure (1) using Boc-L-cysteine (110.5 mg, 0.5 mmol) in DMF (5.0 mL), DIPEA (350 μ L, 2.4 mmol) and 1-bromo-pentafluorobenzene (75.0 μ L, 0.6 mmol) to afford **3p** as white solid (194 mg, 87%), ^1H NMR (500 MHz, DMSO- d_6) δ 6.58 (d, J = 7.3 Hz, 1H), 3.97 (td, J = 7.9, 4.0 Hz, 1H), 3.43 (dd, J = 13.8, 4.0 Hz, 1H), 3.26 – 3.19 (m, 1H), 1.31 (s, 9H); ^{13}C NMR (126 MHz, DMSO) δ 171.63, 154.83, 147.89,



145.91, 145.32, 143.41, 78.03, 54.65, 35.31, 27.92; ^{19}F NMR (470 MHz, DMSO- d_6) δ -131.70 (dd, J = 27.2, 9.7 Hz), -134.18 (dd, J = 27.0, 8.6 Hz); MS (ESI+) m/z calcd. for $\text{C}_{14}\text{H}_{15}\text{BrF}_4\text{NO}_4\text{S}^+ [\text{M}+\text{H}]^+$ 447.9836, found 447.985609.

5 ***N*-(*tert*-butoxycarbonyl)-*S*-(2,3,5,6-tetrafluorophenyl)-*L*-cysteine (3e)**

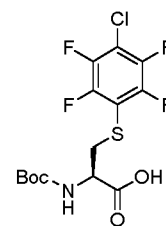
Synthesized by general procedure (1) using Boc-*L*-cysteine (110.5 mg, 0.5 mmol) in DMF (5.0 mL), DIPEA (350 μL , 2.4 mmol) and pentafluorobenzene (66.5 μL , 0.6 mmol) to afford **3e** as white solid (156 mg, 85 %), ^1H NMR (500 MHz, DMSO- d_6) δ 8.17 – 7.53 (m, 1H),



10 6.81 (s, 1H), 3.98 (td, J = 8.5, 4.1 Hz, 1H), 3.41 (dd, J = 13.6, 4.2 Hz, 1H), 3.18 (dd, J = 13.6, 8.9 Hz, 1H), 1.34 (s, 9H); ^{13}C NMR (126 MHz, DMSO) δ 172.36, 154.64, 147.32, 147.21, 146.31, 145.39, 145.28, 144.47, 144.44, 144.38, 144.35, 144.26, 115.13, 114.97, 114.81, 106.96, 106.77, 106.58, 77.80, 54.92, 48.54, 36.43, 27.95; ^{19}F NMR (470 MHz, DMSO- d_6) δ -133.75 (dd, J = 23.6, 11.2 Hz),
15 -138.96 (dd, J = 25.4, 12.3 Hz); MS (ESI+) m/z calcd. for $\text{C}_{14}\text{H}_{16}\text{F}_4\text{NO}_4\text{S}^+ [\text{M}+\text{H}]^+$ 370.0731, found 370.074362.

***N*-(*tert*-butoxycarbonyl)-*S*-(4-chloro-2,3,5,6-tetrafluorophenyl)-*L*-cysteine (3r)**

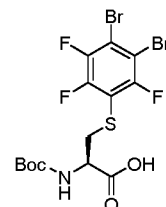
20 Synthesized by general procedure (1) using Boc-*L*-cysteine (110.5 mg, 0.5 mmol) in DMF (5.0 mL), DIPEA (350 μL , 2.4 mmol) and 1-chloro-pentafluorobenzene (78.0 μL , 0.6 mmol) to afford **3r** as white solid (183 mg, 91 %), ^1H NMR (500 MHz, DMSO- d_6) δ 6.89 (d, J = 8.3 Hz, 1H), 4.08 – 3.94 (m, 1H), 3.40 (dd, J = 13.9, 4.1 Hz, 1H), 3.18 (dd, J = 13.9, 9.4 Hz, 1H),
25 1.34 (s, 9H); ^{13}C NMR (126 MHz, DMSO) δ 171.59, 154.90, 148.00, 147.96, 147.86, 146.05, 146.02, 145.91, 144.50, 144.37, 142.52, 142.38, 113.18, 113.02, 112.85, 111.35, 111.20, 111.05, 78.08, 54.49, 35.20, 27.92; ^{19}F NMR (470 MHz, DMSO- d_6) δ -131.94 (dd, J = 24.3, 7.0 Hz), -141.65 – -141.79 (m); MS (ESI+) m/z calcd. for $\text{C}_{14}\text{H}_{15}\text{ClF}_4\text{NO}_4\text{S}^+ [\text{M}+\text{H}]^+$ 404.0341, found 405.037691.



30

***N*-(*tert*-butoxycarbonyl)-*S*-(3,4-dibromo-2,5,6-trifluorophenyl)-*L*-cysteine (3p)**

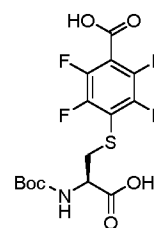
35 Synthesized by general procedure (1) using Boc-*L*-cysteine (110.5 mg, 0.5 mmol) in DMF (5.0 mL), DIPEA (350 μL , 2.4 mmol) and 1,2-dibromotetrafluorobenzene (83.0 μL , 0.6 mmol) to afford **3p** as white



solid (220 mg, 87 %), ^1H NMR (500 MHz, DMSO- d_6) δ 6.76 (s, 1H), 3.98 (dd, J = 8.2, 4.0 Hz, 1H), 3.39 (dd, J = 13.8, 4.1 Hz, 2H), 3.19 (dd, J = 13.8, 8.6 Hz, 1H), 1.32 (s, 9H); ^{13}C NMR (126 MHz, DMSO) δ 171.54, 156.32, 154.82, 154.39, 150.80, 150.77, 150.68, 150.64, 148.83, 148.79, 148.71, 148.67, 145.94, 144.14, 144.11, 144.01, 143.98, 114.04, 113.87, 113.50, 113.34, 113.30, 113.14, 108.21, 108.17, 107.99, 107.96, 78.05, 78.05, 54.56, 35.33, 27.95; ^{19}F NMR (470 MHz, DMSO- d_6) δ -95.86 (d, J = 12.2 Hz), -125.61 (d, J = 25.6 Hz), -127.23 – -127.98 (m); MS (ESI $^+$) m/z calcd. for $\text{C}_{14}\text{H}_{15}\text{Br}_2\text{F}_3\text{NO}_4\text{S}^+$ $[\text{M}+\text{H}]^+$ 507.9035, found 507.910844.

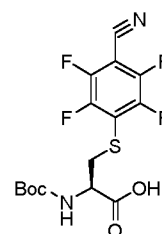
10 N-(tert-butoxycarbonyl)-S-(4-carboxy-2,3,5,6-tetrafluorophenyl)-L-cysteine (3n)

Synthesized by general procedure (1) using Boc-L-cysteine (110.5 mg, 0.5 mmol) in DMF (5.0 mL), DIPEA (350 μL , 2.4 mmol) and pentafluorobenzoic acid (130, 0.6 mmol) to afford **3n** as white solid (178 mg, 89 %); ^1H NMR (500 MHz, DMSO- d_6) δ 7.90 (tt, J = 10.5, 7.5 Hz, 1H), 6.58 (d, J = 7.4 Hz, 1H), 3.95 (dd, J = 7.8, 4.1 Hz, 1H), 3.42 (td, J = 12.1, 10.6, 4.0 Hz, 1H), 3.21 (dd, J = 13.5, 8.1 Hz, 1H), 1.32 (s, 9H); ^{13}C NMR (126 MHz, DMSO) δ 172.05, 154.72, 147.36, 147.24, 146.41, 146.31, 146.22, 145.42, 145.31, 144.35, 131.67, 131.54, 128.61, 114.89, 114.73, 114.56, 107.05, 106.87, 106.68, 77.88, 54.70, 36.10, 27.96; ^{19}F NMR (470 MHz, DMSO- d_6) δ -133.73 (dd, J = 23.1, 9.7 Hz), -138.76 – -139.16 (m); MS (ESI $^+$) m/z calcd. for $\text{C}_{15}\text{H}_{16}\text{F}_4\text{NO}_5\text{S}^+$ $[\text{M}+\text{H}]^+$ 398.0680, found 395.07254.



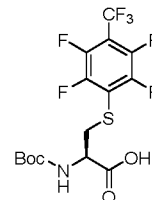
25 N-(tert-butoxycarbonyl)-S-(4-cyano-2,3,5,6-tetrafluorophenyl)-L-cysteine (3d)

Synthesized by general procedure (1) using Boc-L-cysteine (110.5 mg, 0.5 mmol) in DMF (5.0 mL), DIPEA (350 μL , 2.4 mmol) and 1-cyanofluorobenzene (76.0 μL , 0.6 mmol) to afford **3d** as white solid (128 g, 65 %); ^1H NMR (500 MHz, DMSO- d_6) δ 6.97 (d, J = 8.6 Hz, 1H), 4.09 (td, J = 9.0, 4.1 Hz, 1H), 3.62 – 3.47 (m, 1H), 3.35 – 3.21 (m, 1H), 1.35 (s, 9H); ^{13}C NMR (126 MHz, DMSO) δ 171.48, 162.24, 154.99, 147.55, 147.41, 147.17, 147.08, 145.49, 145.35, 145.22, 145.13, 122.98, 122.82, 122.66, 108.11, 108.08, 108.06, 92.40, 92.26, 92.12, 78.21, 54.40, 35.19, 27.94; ^{19}F NMR (470 MHz, DMSO- d_6) δ -131.02 – -131.28 (m), -134.44 (dd, J = 24.2, 10.4 Hz); MS (ESI $^+$) m/z calcd. for $\text{C}_{15}\text{H}_{15}\text{F}_4\text{N}_2\text{O}_4\text{S}^+$ $[\text{M}+\text{H}]^+$ 395.0683, found 395.395.073087.

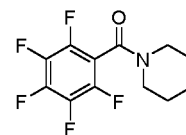


***N*-(*tert*-butoxycarbonyl)-*S*-(2,3,5,6-tetrafluoro-4-(trifluoromethyl)phenyl)-*L*-cysteine (**3j**)**

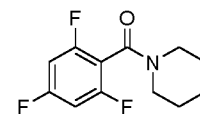
Compound **3j** was synthesized by general procedure (1) using Boc-L-cysteine (110.5 mg, 0.5 mmol) in DMF (5.0 mL), DIPEA (350 μ L, 2.4 mmol) and octafluorotoluene (85.0 μ L, 0.6 mmol) to afford **3j** as white solid (185 mg, 85 %); ^1H NMR (500 MHz, DMSO- d_6) δ 6.91 (d, J = 8.5 Hz, 1H), 4.11 (td, J = 9.1, 3.9 Hz, 1H), 3.52 (dd, J = 13.9, 3.9 Hz, 1H), 3.26 (dd, J = 13.9, 9.1 Hz, 1H), 1.34 (s, 9H). ^{13}C NMR (126 MHz, DMSO) δ 171.49, 154.94, 147.81, 147.71, 145.86, 145.76, 144.43, 144.29, 142.38, 142.24, 124.01, 121.83, 120.63, 120.48, 120.32, 119.66, 117.48, 107.26, 107.16, 106.99, 106.89, 106.61, 78.08, 54.70, 35.00, 27.87; ^{19}F NMR (470 MHz, CDCl_3) δ -56.30, -56.35, -56.40, -131.07, -131.33, -140.02. MS (ESI $^+$) m/z calcd. for $\text{C}_{15}\text{H}_{15}\text{F}_7\text{NO}_4\text{S}^+$ $[\text{M}+\text{H}]^+$ 438.0605, found 438.06235.



piperidin-1-yl(pentafluorophenyl)methanone (2b**)**
 Piperidine (148 μ L, 1.5 mmol) was added to a solution of pentafluorobenzoylchloride (72 μ L, 0.5 mmol) in DCM (5 mL). The reaction was stirred at 22 $^\circ\text{C}$ for 30 min. The solvent was evaporated and the crude residue was purified by column chromatography (EtOAc: pentane / 10:90) to afford the product **2b** as white solid (132 mg, 95 %); ^1H NMR (500 MHz, DMSO- d_6) δ 3.68 – 3.59 (m, 2H), 3.31 – 3.25 (m, 2H), 1.67 – 1.60 (m, 2H), 1.56 (td, J = 6.8, 6.3, 4.4 Hz, 2H), 1.50 – 1.42 (m, 2H); ^{13}C NMR (126 MHz, DMSO) δ 155.56, 143.02, 142.06, 141.17, 141.07, 140.99, 140.17, 140.06, 138.38, 138.26, 138.17, 138.13, 136.38, 136.27, 111.37, 111.19, 111.02, 47.19, 42.42, 26.21, 25.21, 23.64; ^{19}F NMR (470 MHz, DMSO- d_6) δ -143.15 (d, J = 18.0 Hz), -154.05 (t, J = 21.3 Hz), -160.81 (t, J = 21.3 Hz); MS (ESI $^+$) m/z calcd. For $\text{C}_{12}\text{H}_{11}\text{F}_5\text{NO}^+$ $[\text{M}+\text{H}]^+$ 280.0755, found 280.0762.



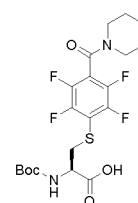
piperidin-1-yl(2,4,6-trifluorophenyl)methanone (2c**)**
 Piperidine (148 μ L, 1.5 mmol) was added to a solution of 2,4,6-trifluorobenzoylchloride (62 μ L, 0.5 mmol) in DCM (5 mL). The reaction was stirred 22 $^\circ\text{C}$ for 30 min. The solvent was evaporated and the crude residue was purified by column chromatography (EtOAc: pentane / 10:90) to afford the product **2c** as white solid (110 mg, 90 %); ^1H NMR (500 MHz, Methanol- d_4) δ 6.94 – 6.86 (m, 2H), 3.67 – 3.56 (m, 2H), 3.22 (s, 2H), 1.64 – 1.59 (m, 2H), 1.56 (tq, J = 7.6, 3.8 Hz, 2H), 1.49 – 1.44 (m, 2H); ^{13}C NMR (126 MHz, MeOD) δ



166.02, 165.91, 165.90, 165.78, 164.03, 163.91, 163.90, 163.79, 161.62, 161.53, 161.50, 161.41, 161.07, 159.63, 159.55, 159.51, 159.43, 111.93, 111.89, 111.74, 111.70, 111.69, 111.55, 111.51, 102.19, 102.16, 102.13, 102.00, 101.98, 101.95, 101.92, 101.79, 101.76, 101.73, 44.11, 27.61, 26.70, 25.30¹⁹F NMR (470 MHz, CDCl₃)
 5 δ -104.35, -108.92; MS (ESI⁺) m/z calcd. For C₁₂H₁₁F₅NO⁺ [M+H]⁺ 243.0871, found 280.0897

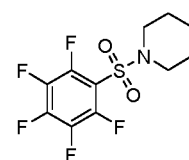
***N*-(*tert*-butoxycarbonyl)-*S*-(2,3,5,6-tetrafluoro-4-(piperidine-1-carbonyl)phenyl)-*L*-cysteine (**3b**)**

10 Compound **3b** was synthesized by general procedure (1) using Boc-L-cysteine (110.5 mg, 0.5 mmol) in DMF (5.0 mL), DIPEA (350 μL, 2.4 mmol) and (pentafluorophenyl)(piperidin-1-yl)methanone (167.0 mg, 0.6 mmol) to afford **3b** as white solid (179 mg, 86 %); ¹H NMR (500 MHz,
 DMSO-*d*₆) δ 6.83 (d, *J* = 8.2 Hz, 1H), 4.02 (tt, *J* = 12.8, 6.3 Hz, 1H), 3.65 (dtd, *J* = 36.3,
 15 12.8, 5.2 Hz, 4H), 3.44 (d, *J* = 4.2 Hz, 1H), 3.20 – 3.14 (m, 1H), 1.63 (qd, *J* = 5.9, 3.2 Hz, 2H), 1.57 – 1.53 (m, 2H), 1.48 – 1.44 (m, 2H), 1.35 (s, 9H); ¹³C NMR (126 MHz, DMSO) δ 171.79, 162.25, 156.18, 154.96, 147.55, 147.44, 145.61, 145.49, 142.56, 140.60, 116.01, 115.83, 115.65, 115.01, 114.85, 78.01, 54.51, 47.14, 35.48, 27.96, 25.22, 23.67.; MS (ESI⁺) m/z calcd. for C₂₀H₂₅F₄N₂O₅S⁺ [M+H]⁺ 418.1415, found
 20 418.14268.



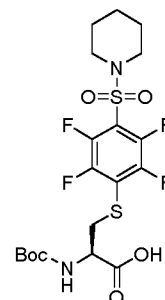
1-((pentafluorophenyl)sulfonyl)piperidine (2g**)**

Piperidine (148 μL, 1.5 mmol) was added to a solution of 1-((pentafluorophenyl)sulfonyl)piperidine (72.0 μL, 0.5 mmol) in DCM (5
 25 mL). The reaction was stirred at 22 °C for 30 min. The solvent was evaporated and the crude residue was purified by column chromatography (EtOAc: pentane / 10:90) to afford the product **2g** as white solid (146 mg, 93 %); ¹H NMR (500 MHz, DMSO-*d*₆) δ 3.16 (t, *J* = 5.6 Hz, 4H), 1.59 (h, *J* = 11.2, 8.4 Hz, 4H), 1.51 – 1.40 (m, 2H); ¹³C NMR (126 MHz, DMSO) δ 145.16, 145.07,
 30 143.09, 143.04, 143.01, 142.38, 138.71, 138.55, 136.69, 136.56, 112.34, 51.27, 45.84, 24.61, 22.60; ¹⁹F NMR (470 MHz, DMSO-*d*₆) δ -135.82 (d, *J* = 22.5 Hz), -147.15 (t, *J* = 22.1 Hz), -159.62 (d, *J* = 21.0 Hz); MS (ESI⁺) m/z calcd. for C₁₁H₁₁F₅NO₂S⁺ [M+H]⁺ 316.0425, found 316.044078.



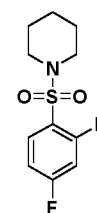
***N*-(*tert*-butoxycarbonyl)-*S*-(2,3,5,6-tetrafluoro-4-(piperidin-1-ylsulfonyl)phenyl)-*L*-cysteine (2*R*)-2-[[(*tert*-butoxy)carbonyl]amino]-3-[[2,3,5,6-tetrafluoro-4-(piperidine-1-sulfonyl)phenyl]sulfanyl]propanoic acid (3g)**

Compound **3g** was synthesized by general procedure 1 method A using
 5 Boc-*L*-cysteine (110.5 mg, 0.5 mmol) in DMF (5.0 mL), DIPEA (350 μ L, 2.4 mmol) and 1-(pentafluorobenzenesulfonyl)piperidine (189.0 μ L, 0.6 mmol) at rt for 16 h to afford **3g** as white solid (232 mg, 90%); ^1H NMR (500 MHz, DMSO- d_6) δ 6.97 (d, J = 8.7 Hz, 1H), 4.11 (td, J = 9.2, 3.9 Hz, 1H), 3.53 (dd, J = 13.9, 4.1 Hz, 1H), 3.25 (dd, J = 13.8, 9.9 Hz, 1H),
 10 3.16 (t, J = 5.6 Hz, 4H), 1.59 (p, J = 5.5 Hz, 4H), 1.46 (dp, J = 11.8, 4.8, 4.2 Hz, 2H), 1.35 (s, 9H); ^{13}C NMR (126 MHz, DMSO) δ 171.59, 147.80, 147.69, 145.84, 145.72, 144.25, 144.12, 142.21, 142.08, 120.17, 120.01, 119.85, 115.68, 115.55, 115.42, 78.17, 54.42, 45.86, 39.14, 27.92, 24.64, 22.59; ^{19}F NMR (470 MHz, DMSO- d_6) δ -131.61 (d, J = 19.7 Hz), -136.74 (d, J = 18.3 Hz); MS (ESI $^+$) m/z calcd. for
 15 $\text{C}_{19}\text{H}_{25}\text{F}_4\text{N}_2\text{O}_6\text{S}_2^+$ [M+H] $^+$ 517.1085, found 517.115716.



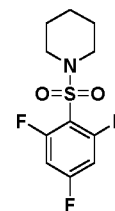
1-(2,4-difluorophenyl)sulfonyl)piperidine (2i)

Piperidine (444.0 μ L, 4.5 mmol) was added to a solution of 2,4-difluorophenylsulfonyl chloride (200 μ L, 1.5 mmol) in DCM (10 mL). The
 20 reaction was stirred at 22 $^{\circ}\text{C}$ for 30 min. The solvent was evaporated and the crude residue was purified by column chromatography (EtOAc: pentane / 10:90) to afford the product **2i** as white solid (238 mg, 61%); ^1H NMR (500 MHz, DMSO- d_6) δ 7.85 (td, J = 8.6, 6.3 Hz, 1H), 7.62 – 7.55 (m, 1H), 7.36 – 7.28 (m, 1H), 3.08 – 2.99 (m, 4H), 1.54 (h, J = 5.5 Hz, 4H), 1.46 – 1.38 (m, 2H); ^{13}C NMR (126
 25 MHz, DMSO) δ 166.13, 166.03, 164.10, 164.01, 160.13, 160.02, 158.09, 157.98, 132.93, 132.92, 132.85, 132.83, 121.32, 121.29, 121.20, 121.17, 112.59, 112.56, 112.42, 112.39, 106.41, 106.20, 105.99, 46.08, 24.79, 22.76; ^{19}F NMR (470 MHz, CDCl_3) δ -101.26, -102.32, -102.34, -102.36; MS (ESI $^+$) m/z calcd. for $\text{C}_{11}\text{H}_{14}\text{F}_2\text{NO}_2\text{S}^+$ [M+H] $^+$ 262.0708, found 262.071990.



1-(2,4,6-trifluorobenzenesulfonyl)piperidine (2h)

To a solution of 2,4,6-trifluorophenylsulfonylchloride (200 μ L, 1.5 mmol) in DCM (10 mL), piperidine (444.0 μ L, 4.5 mmol) was added. The reaction
 35 residue was purified by column chromatography (EtOAc: pentane / 10:90) to afford the product **2h** as white solid (234 mg, 56%); ^1H NMR (500 MHz,



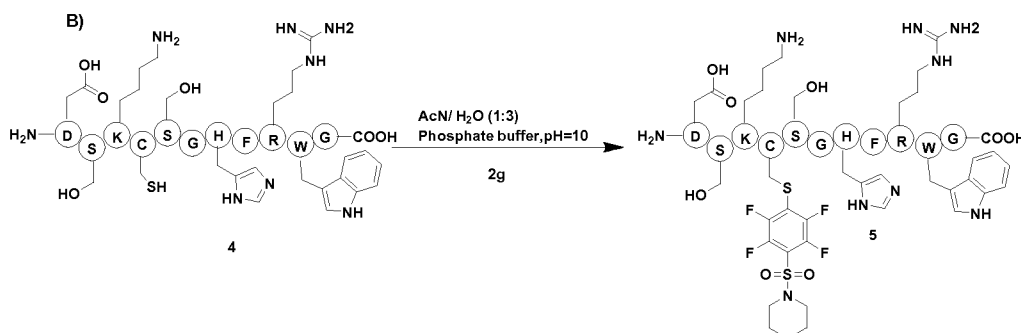
Chloroform-*d*) δ 6.81 – 6.59 (m, 2H), 3.15 (t, J = 5.5 Hz, 4H), 1.59 (q, J = 5.8 Hz, 4H), 1.49 – 1.41 (m, 2H); ^{13}C NMR (126 MHz, CDCl_3) δ 165.96, 165.84, 165.72, 163.92, 163.79, 163.67, 161.87, 161.81, 161.75, 161.69, 159.80, 159.75, 159.68, 159.63, 112.99, 112.95, 112.85, 112.81, 112.71, 112.67, 102.21, 102.18, 102.01, 101.98, 101.96, 101.79, 101.76, 46.50, 25.28, 23.57; ^{19}F NMR (470 MHz, CDCl_3) δ -99.38, -99.39, -99.42, -99.43, -100.95, -100.98, -100.98, -100.98; MS (ESI $^+$) m/z calcd. for $\text{C}_{11}\text{H}_{12}\text{F}_3\text{NO}_2\text{S}^+ [\text{M}+\text{H}]^+$ 280.0614, found 280.062602

10

Example 2 – Chemoselective cysteine arylation of peptides

Among the members of the most reactive compounds, the **2g** was found most interesting as potential general compound for chemo-selective labeling of cysteines in proteins under aqueous conditions. This fluorobenzene can be linked to other molecules through the sulfonamide bond and is hydrophilic compared to e.g. a trifluoromethyl substitution. To test for selectivity towards cysteine over other amino acids, peptide **4** (scheme 2) containing all nucleophilic protein functionalities, was synthesized. The peptide was reacted with an excess (3 eq.) of **2g** under aqueous condition (see Scheme 2). Only the monosubstituted adduct **5** could be observed and was identified by LC-MS. The structure of **5** was confirmed by MS/MS sequencing and only the target cysteine was modified by the compound (see Figure 3). Under the same conditions, compound **2b** was found not to react (see Figure 4).

25



Scheme 2. Selective modification of cysteine in peptide **4** (D S K C S G H F R W G-OH), which contains all nucleophilic amino acid residues, under aqueous conditions by compound **2g**.

30 General procedures for peptide couplings

1H-Benzotriazolyl tetramethyluronium tetrafluoroborate (TBTU) couplings were performed by dissolving the amino acid (3 equiv.) in DMF containing 4-ethylmorpholine (NEM) (3 equiv.), followed by addition of TBTU (3 equiv.). The resulting solution was left to pre-activate for 3 min before being added to the resin with gentle agitation.

- 5 Coupling reactions were allowed a reaction time of 2–3 h. Peptide couplings were generally performed in an amount of solvent sufficient to just cover the resin and washings were conducted slowly. After the reaction, the resin was washed with DMF (×6) and completion of coupling was assessed by the Kaiser test, described in Kaiser et al., *Analytical Biochemistry* **1970**, 34, 595-598.

10

General procedure for Fmoc deprotection

Fmoc deprotection was achieved by two treatments of 5 min with 20% piperidine in DMF (v/v), followed by washing of the resin with DMF (×10). The free amine content was analyzed using the Kaiser test.

15

General procedure for Boc protection

An excess of 2% di-tert-butyl dicarbonate (Boc₂O) in DMF and NEM was added to the peptide containing a free amine and the mixture was left to react for 1 h. After washing with DMF (×10) the completion of reaction was verified using the Kaiser test.

20

General procedure for cleavage peptides off the resin

Cleavage of peptides off the resin was achieved with 0.1 M aqueous NaOH for 1 h followed by neutralisation with 0.1 M aqueous HCl. The resin was washed with 70% MeCN in water to collect all the peptide and the combined extracts were concentrated by lyophilization.

25

Example 3 – Developing probes for labelling of cysteine containing proteins

Based on these results labeling probes for cysteine-containing proteins were designed.

- 30 Pentafluorophenylsulfonylchloride, **7** was reacted with amino-functionalized PEG linkers containing a tag, such as an azide group for click chemistry (CuAAC or SPACC) (**8a**), the sulforhodamine B fluorophore (**10**) or biotin affinity probe (**11**) (see Figure 5). The generated tag-containing compounds (**8a**, **10**, or **11**) were reacted with enhanced green fluorescent protein (eGFP) comprising two cysteines, not involved in disulfide bonds; and with bovine serum albumin, which has 17 conserved disulfide bonds and one free thiol (Cys 34). All three types of reporter-linked compounds were capable of
- 35

labeling eGFP and albumin in a concentration-dependent manner (see Figure 5). The reactive fluoroaryl compounds did not interfere with the functional click, fluorescence or affinity tags under the protein labeling conditions.

5

Labeling of eGFP and bovine serum albumin (BSA) with azide probe 8a followed by click reaction (in-gel fluorescence scanning)

After incubation of eGFP (5 μ L, 12 μ M), expressed and purified as described previously 2 or BSA (3 μ M, from Thermo Scientific: Pierce™ BSA standard ampules, 2 mg/ml) 10 with fluoroaryl azide probe 8a (5 μ L, 20 and 200 μ M) in tris buffer (pH 8.3) at 37 °C for 16 h, click mixture consisting of CuSO₄ (1 μ L, 45 mM), THPTA (1 μ L, 90 mM) and sodium ascorbate (1 μ L, 60 mM) as well as alkyne cyanine dye 718 (Sigma Aldrich 30154) (3 μ L, 1 mM) were added and the mixture was left to react at 22 oC for 2 hours. The reaction mixture (15 μ L) was analyzed by SDS-PAGE electrophoresis followed by 15 in-gel fluorescence scanning at λ_{ex} 650 nm and λ_{em} 680 nm on a Typhoon FLA 7000 laser scanner, and by staining with Coomassie Brilliant Blue.

Labeling of eGFP and BSA with fluorophore probe 10 (in-gel fluorescence scanning)

20 After incubation of eGFP (5 μ L, 12 μ M) or bovine serum albumin (5 μ L, 3 μ M) with fluoroaryl probes 10 (5 μ L, 20 and 200 μ M) in tris buffer (pH=8.3) at 37 °C for 16 h, the reaction mixture was analyzed by SDS-PAGE electrophoresis followed by in-gel fluorescence scanning at λ_{ex} 530 nm and λ_{em} 580 nm on a Typhoon FLA 7000 laser scanner, and by staining with Coomassie Brilliant Blue.

25

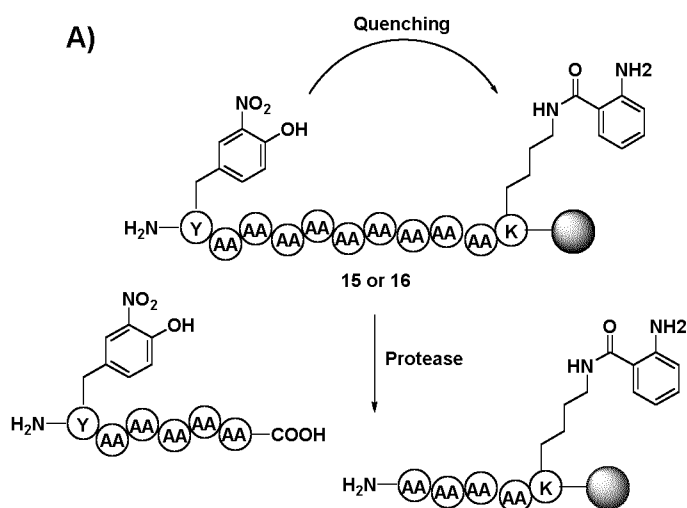
Labeling of eGFP and BSA with biotin probe 11 (Western blot analysis)

After incubation of eGFP (5 μ L, 12 μ M) or bovine serum albumin (5 μ L, 3 μ M) with fluoroaryl probes 11 (5 μ L, 20 and 200 μ M) in tris buffer (pH=8.3) at 37 °C for 6 h, the reaction mixture was separated by SDS-PAGE electrophoresis. Following gel 30 electrophoresis the proteins were transferred to a nitrocellulose membrane using a Trans-Blot Turbo Transfer system (Bio-Rad). The blot was blocked in 3% BSA and 0.5% Tween-20 in TBS (50 mM Tris-base, 150 mM NaCl, pH 7.6), 1–2 h at 22 oC. Labeling was performed with SAV-HRP (1 μ g/mL, Sigma Aldrich S5512) in 0.2% BSA and 0.05% Tween-20 in TBS, 1 h at 22 oC. Bands were visualized by the HRP- 35 catalyzed conversion of luminol (Santa Cruz Biotechnology luminol reagent sc-2048).

An image of the chemiluminescent signal was acquired with a normal photocamera in a dark room. The image was processed in ImageJ (conversion from RGB color to 32-b, and inverted). Brightness and contrast were adjusted in ImageJ and Microsoft PowerPoint. Aliquots of the labeled proteins and negative controls were loaded on a second gel, which was stained with Coomassie Brilliant Blue R250 to verify that protein was present in all samples.

Example 4 – Selective cysteine protease inhibition

In addition to the labelling applications it was proposed that the compounds may also be applied for selective inhibition of cysteine proteases over other classes of proteases including the closely related serine proteases. To explore this area, an on-bead inhibition screening assay towards the cysteine protease papain and the serine protease subtilisin was conducted using compounds **2a**, **2b**, **2g**, **2h** and **2f**. FRET substrates for the two proteases were synthesized attached to polymer beads of PEGA₁₉₀₀ resin (see Scheme 4). The enzymes (100 nM) were pre-incubated with the compounds (500 μ M) for 1 h at 37°C to allow for potential covalent modification. These mixtures were then added to the substrate beads. Beads started to light up in wells with papain containing blank, **2b**, **2h** or **2f** after 30 minutes while the beads in the wells with compounds **2a** or **2g** stayed dark even when they were left overnight at room temperature. None of the compounds were able to inhibit subtilisin activity at the same conditions as those used for inhibition of papain (see Figure 6).



Scheme 4. Schematic representation of on-bead papain and subtilisin activity assay. The black circle represents the solid phase matrix of the PEGA₁₉₀₀ resin.

These studies demonstrate that the developed compounds may be applied as selective inhibitors for knocking out cysteine proteases without affecting a serine protease like subtilisin.

5 Example 5 – Residue selective arylation of cysteine in the catalytic site

Three fluorobenzene derivatives with different reactivities (**2g**, **2h** and **2c**) were investigated for their ability to react selectively with one among several cysteine residues in the same protein, as well as their ability to modify cysteine proteases in an activity-based manner (see Figure 7A). Tobacco Etch Virus (TEV) protease was chosen as a model protease as it contains unpaired surface exposed cysteines in addition to the cysteine (Cys170) in the catalytic site. TEV protease has a sequence-specific activity and is an important tool for the removal of fusion tags or the cleavage of fusion proteins *in vitro* and *in vivo*. The activity of the protease was confirmed using a FRET peptide with the sequence Y(3-NO₂) E N L Y F Q G K(Abz) G-OH (**12**), as substrate (see Figure 7B). To test reactivity towards the compounds, the enzyme was incubated with fluoroaryl compounds **2g**, **2h** and **2c** at a final concentration of 100 µM in the assay buffer. As negative control only assay buffer was added. After 2 h of incubation with the most reactive compound **2g**, the mono- (**13a**) and the disubstituted (**13b**) adducts of TEV protease were identified; the latter (**13b**) being the major product (see Table 2). As used herein substituted adduct mean TEV protease covalently linked to fluoroaryl.

Table 2. LCMS with ESI⁺-TOF analysis of reaction mixture (TEV + **2g**). After incubation for 2 h at 37 °C: TEV peak with a Deconvoluted mass 28,843 KDa almost disappeared and three major peaks that correspond to the mono- and di- substituted adduct of TEV and di- substituted adduct of α-N-gluconoylTEV appeared.

	Calculated isotopically averaged mass (kDa)	Found Deconvoluted mass (kDa)
TEV peak	28843.5488	28,843.3028
di-substituted adduct 13b of TEV	29,433.6224	29,433.9419
mono-substituted adduct of TEV 13a	29,138.5886	29,138.8873
Di-substituted adduct of (+178-Da) α -N-gluconoyl TEV 13e	29,611	29,611.3510

After 10 h of incubation, the mono- (**13a**), the di- (**13b**), the tri- (**13c**), and the tetrasubstituted (**13d**) adducts were detected; at this time point the tetrasubstituted (**13d**) was the major product (see Table 3).

- 5 **Table 3.** LCMS with ESI⁺-TOF analysis of reaction mixture (TEV + **2g**). After incubation for 10 h at 37 °C: TEV peak at 28,843 completely disappeared and five major peaks that correspond to the di, tri and tetra- adduct of TEV or α -*N*-gluconoyl TEV appeared.

	Calculated isotopically averaged mass (kDa)	Found Deconvoluted mass (kDa)
tetra-substituted adduct 13d of TEV	30,023.726	30,022.7451
Di-substituted adduct 13b of TEV	29,433.6224	29,433.9419
Tri-substituted adduct 13c of TEV	29,728.6592	29,728.2317
Tri-substituted adduct 13f of (+178-Da) α - <i>N</i> -gluconoyl TEV	29,907	29,907.1397
Tetra-substituted adduct 13g of (+178-Da) α - <i>N</i> -gluconoyl TEV	30,201	30,201.9647

- 10 These results demonstrate that the most reactive compound has little selectivity among cysteine residues, but support the claim of chemo-selectivity of this compound toward cysteine over all other amino acid residues. Gratifyingly, when the less reactive compound **2h** was incubated with TEV protease for 10 h, only the monosubstituted adduct (**14**) was identified (Calculated isotopically averaged mass 29,102.6045 KDa, found Deconvoluted mass 29,102.8660 kDa). Lastly, TEV protease was not modified by the least reactive compound **2c** under the same conditions. The selective arylation of the cysteine in the catalytic site by compound **2h** was confirmed by in-solution and in-gel tryptic digestion followed by high-resolution mass spectrometry. The peptide fragment containing the unmodified cysteine in the active site (Cys170) was found at 1219.59 (red) in the native TEV protease (negative control). This signal did not appear after the treatment with both **2g** (green) and **2h** (blue), but new mass peaks appeared at 1514.62 and 1478.65, respectively, corresponding to modified peptides fragments (see Figure 7C). In the native TEV protease, the peptide fragment with the unmodified surface-exposed cysteine (Cys129) was found at 1267.68. This fragment was not
- 15
- 20

found in the sample reacted with compound **2g**. Instead, the modified fragment was found at 1562.73. In contrast, the digest of TEV protease exposed to compound **2h** still contained the fragment at 1267.68, being the unmodified surface-exposed cysteine, while a fragment with the expected mass equivalent to the compound-functionalized cysteine was not detectable.

These results demonstrate the versatile concept of being able to tune the reactivity of fluorobenzenes accurately as compounds for residue-selective reaction with only one of several nucleophilic cysteine residues. The only differences between the selective compound **2h** and the non-selective compound **2g** is the number of fluoro substituents on the benzene core. It should be noted that due to the upper limit of the mass range of the MS-detection method no fragments above 2250 appeared in the MS spectrum of the digested TEV.

15 **TEV protease expression and purification**

BL21 (DE3) competent cells (50 μ L) were transformed with 1 μ L pET15 plasmid carrying the gene for TEV protease with a N-terminal His6-tag. Cells were plated on LB-agar plates supplemented with ampicillin (100 μ g/mL). LB medium supplemented with ampicillin (100 μ g/mL) was inoculated with a single colony and grown overnight at 37 °C. A 200 mL expression culture was prepared in LB medium supplemented with ampicillin (100 μ g/mL) and the addition of overnight culture to an absorbance at 600 nm of 0.1. When the culture reached an absorbance at 600 nm of 0.6, it was induced with IPTG to a final concentration of 1 mM and expression was allowed to proceed at 37 °C for 4 h. The culture was thereafter harvested and stored at 20 °C.

The cell pellet from a 200 mL cell culture was suspended and incubated on ice in 10 mL 50 mM NaH₂PO₄, 300 mM NaCl, 1 mg/mL lysozyme, pH 7.5. The suspension was then sonicated twice on ice with a sequence of 6 x 10 s at 50% amplitude with 20 s pauses. The lysate was centrifuged for 10 min at 20,000 x g, 4 °C and the cleared lysate isolated. DTT and imidazole were added to a final concentration of 5 mM and 20 mM, respectively. The solution was loaded onto a 1-mL HisTrap HP column (GE Healthcare), and after washing with wash buffer (50 mM NaH₂PO₄, 500 mM NaCl, 20 mM imidazole, pH 7.5) until the absorbance signal did not further decrease, the protein was eluted in 50 mM NaH₂PO₄, 500 mM NaCl, 250 mM imidazole, pH 7.5. Fractions were stored at 20 °C. Relevant HisTrap fractions were pooled and DTT added to a final concentration of 5 mM. The solution was then centrifuged for 15 min at 11,000 x g,

4 °C and the resulting supernatant loaded onto a HiLoad 16/600 Superdex 75 pg (GE Healthcare) with 20 mM NaH₂PO₄, 150 mM NaCl, pH 7.5 as buffer. Relevant fractions were concentrated using an Amicon ultracentrifugation unit with a molecular weight cut-off of 10 kDa (Millipore) and the concentrated samples were diluted to 0.3 mg/mL in storage buffer containing 20 mM NaH₂PO₄, 300 mM NaCl, 5 mM DTT, 50% glycerol, pH 7.5. Fractions derived from the affinity and size-exclusion chromatography steps, respectively, were checked for purity by SDS-PAGE analysis. For this purpose, samples were mixed with SDS sample buffer and incubated for 3 min at 96 °C and then loaded on pre-casted TGX™ any-kD gels (Bio-Rad), which were stained with Coomassie Brilliant Blue R250 afterwards (see SI-Figure 20). Protein concentrations were determined by measuring the absorbance at 280 nm on a NanoDrop™ 2000 spectrophotometer (Thermo Scientific).

LC-MS analysis of wild-type TEV and modified TEV

Active TEV protease (20 µL, ~ 40 µM) was incubated with probes 2g and 2h (20 µL, 200 µM) in TEV buffer (50 Mm Tris, 0.5 mM EDTA and 1.0 mM TCEP, pH=8.5) at 37 °C. The reaction was followed by LC-MS after 2 and 10 h. LCMS with ESI+-TOF detection of TEV (20 µM solution, inject 2 µL). LCMS with ESI+-TOF detection of TEV; Calculated isotopically averaged mass 28,843.5488, found deconvoluted mass 28,843.3028 Da. Also a post-translational modified (+178-Da) α-N-gluconoyl TEV was occasionally observed[3]. LCMS with ESI+-TOF detection of (+178-Da) α-N-gluconoyl TEV; Calculated isotopically averaged mass 29,021. kDa, found deconvoluted mass 29,021.2480 kDa.

25 Tryptic digestion and MALDI-TOF analysis

Portions of 10 µL purified TEV protease were mixed with 2 µL 0.5 µg/µL trypsin (sequencing-grade, Roche) in 40 µL 0.1 M (NH₄)₂CO₃ and incubated overnight at 37 °C. To the solution was then added formic acid to 1% final concentration. The tryptic peptides were desalted using C18 ZipTips according to the manufacturer's protocol (Millipore) and analyzed by MALDI-TOF using αCHCA as matrix.

Example 6 – Synthesis of compounds

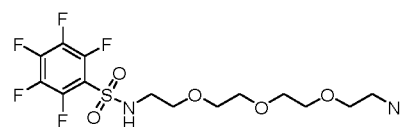
General methods

All reagents and solvents used were purchased from commercial sources and used without further purification. Dry solvents were obtained directly from a Pure Solv system

(Innovative Technology inc.) for dry solvents. NMR spectra were recorded on a Bruker 500 MHz instrument. ^1H -NMR chemical shifts (δ) are reported in parts per million (ppm) relative to a residual proton peak of the solvent, $\delta = 2.50$ for $(\text{CD}_3)_2\text{SO}$, $\delta = 3.31$ for CD_3OD and $\delta = 7.26$ for CDCl_3 . Acetone was used as reference in water. ^{13}C -NMR chemical shifts (δ) are reported in parts per million (ppm) relative to $(\text{CD}_3)_2\text{SO}$ ($\delta = 39.52$), CD_3OD ($\delta = 49.00$) or CDCl_3 ($\delta = 77.16$). Coupling constants are reported in Hertz (Hz). Multiplicities of peaks are given as: s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet). LC spectra were recorded on Agilent 1100 series. MS (ESI) spectra were recorded with a Bruker Esquire 3000 plus mass spectrometer, or on a Bruker Microflex QIII instrument fitted with a UPLC Dionex Ultimate 3000 from Thermo Fisher. MALDI MS were recorded on a Bruker Solaris XR ICR-instrument. TLC plates used were Fluka silica gel F254 on aluminum. Column chromatography was performed with Merck silica 60 (0.015 – 0.040 mm). Chromatographic purifications (flash) were performed with Silica Gel 60 from Fluka (0.015 – 0.040 mm). Preparative RP-HPLC was performed on a Gilson 215 semi-prep HPLC-system using an XTerra® Prep column (10 μM 19 x 150 mm, flow rate 15 mL min $^{-1}$). Analytical HPLC was performed on an Agilent HP1100 instrument using an XBridge® column (10 μM 4.6 x 100 mm, flow rate 1.0 mL min $^{-1}$). Compounds were detected by UV absorption at 215 and 254 nm. Eluents for both the semi-prep and analytical RP-HPLC were: A: water with 0.05 % TFA, B: Acetonitrile/water (90/10) with 0.05 % TFA.

***N*-(2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)-2,3,4,5,6-pentafluorophenylsulfonamide (8a)**

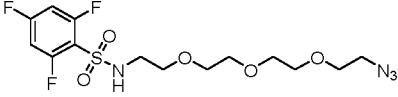
Synthesized by general procedure 2 of Example 1 using 1-[2-(2-aminoethoxy)ethoxy]-2-(2-azidoethoxy)ethane (99.0 μL , 0.5 mmol) in DMF (5.0



mL), DIPEA (350 μL , 2 mmol) and of pentafluorobenzenesulfonylchloride (112 μL , 0.75 mmol) and purified by column chromatography with a gradient of EtOH: heptane (20-60 %) to afford **8a** as oily liquid (150 mg, 67%); ^1H NMR (500 MHz, Chloroform-*d*) δ 6.20 (d, $J = 5.7$ Hz, 1H), 3.61 (m, 6H), 3.54 (m, 6H), 3.34 (t, $J = 5.0$ Hz, 2H), 3.27 (q, $J = 5.2$ Hz, 2H); ^{13}C NMR (126 MHz, CDCl_3) δ 145.58, 145.54, 145.50, 145.47, 145.44, 145.41, 145.37, 144.87, 144.83, 144.79, 144.76, 144.72, 144.68, 144.65, 144.62, 144.58, 143.53, 143.49, 143.45, 143.42, 143.39, 143.36, 143.32, 142.79, 142.75, 142.71, 142.68, 142.64, 142.60, 142.58, 142.54, 142.50, 138.98, 138.94, 138.92, 138.88, 138.84, 138.80, 138.74, 138.69, 136.94, 136.90, 136.84, 136.80, 136.76,

136.72, 136.70, 136.66, 116.86, 116.84, 116.82, 116.76, 116.74, 116.72, 116.70, 116.68, 116.61, 116.60, 116.58, 116.56, 70.60, 70.57, 70.40, 70.02, 69.08, 50.68, 43.45; ^{19}F NMR (470 MHz, CDCl_3) δ -136.85, -136.89, -146.62, -146.66, -146.70, -158.90, -158.95, -158.99; MS (ESI^+) m/z calcd. for $\text{C}_{14}\text{H}_{18}\text{F}_5\text{N}_4\text{O}_5\text{S}^+$ $[\text{M}+\text{H}]^+$ 449.0913, found 449.0953, MS (ESI^+) m/z calcd. for $\text{C}_{14}\text{H}_{18}\text{F}_5\text{N}_4\text{O}_5\text{S Na}^+$ $[\text{M}+\text{Na}]^+$ 471.0732, found 471.07439.

***N*-(2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)-2,4,6-trifluorophenylsulfonamide (8b)**

10 Synthesized by general procedure 2 using 1-[2-(2-aminoethoxy)ethoxy]-2-(2-azidoethoxy)ethane (99 μL , 0.5 mmol) in DCM (5.0 ml), DIPEA (350 μL , 2 mmol) and of 2,4,6-trifluorosulfonylchloride (109 μL , 0.75 mmol) and purified by column chromatography with a gradient of EtOH: heptane (20-60 %) to afford **8b** as oily liquid (137 mg, 67%); ^1H NMR (500 MHz, Chloroform-*d*) δ 6.80 – 6.67 (m, 2H), 5.81 (t, J = 5.8 Hz, 1H), 3.64 – 3.58 (m, 6H), 3.58 – 3.54 (m, 2H), 3.55 – 3.50 (m, 4H), 3.33 (t, J = 5.0 Hz, 2H), 3.23 (q, J = 5.3 Hz, 2H); ^{13}C NMR (126 MHz, CDCl_3) δ 165.93, 165.80, 165.68, 163.88, 163.76, 163.63, 161.56, 161.50, 161.46, 161.44, 161.38, 159.50, 159.45, 159.38, 159.32, 115.55, 115.51, 115.41, 115.37, 115.28, 115.24, 102.18, 102.15, 101.98, 101.96, 101.95, 101.93, 101.76, 101.73, 70.64, 70.57, 70.46, 70.06, 69.18, 50.68, 43.31; ^{19}F NMR (470 MHz, CDCl_3) δ -99.47, -103.85; MS (ESI^+) m/z calcd. for $\text{C}_{14}\text{H}_{20}\text{F}_3\text{N}_4\text{O}_5\text{S}^+$ $[\text{M}+\text{H}]^+$ 413.1101, found 413.126005.

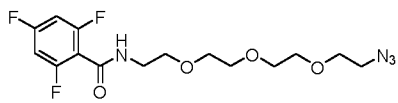
***N*-(2-[2-[2-(2-azidoethoxy)ethoxy]ethoxy]ethyl)-2,3,4,5,6-pentafluorobenzamide (8c)**

25 Compound **8c** was synthesized by general procedure (2) using 1-[2-(2-aminoethoxy)ethoxy]-2-(2-azidoethoxy)ethane (99.0 μL , 0.5 mmol) in DCM (5.0 mL), DIPEA (350 μL , 2.0 mmol) and pentafluorobenzoylchloride (108.0 μL , 0.75 mmol). The product was purified by column chromatography with a gradient of EtOH: heptane (20–60 %) to afford **8c** as oily liquid (124 mg, 60%); ^1H NMR (500 MHz, Chloroform-*d*) δ 3.60 (d, J = 4.3 Hz, 8H), 3.58 – 3.53 (m, 7H), 3.30 – 3.27 (m, 2H); ^{13}C NMR (126 MHz, CDCl_3) δ 157.45, 145.13, 143.19, 143.15, 141.14, 138.60, 136.62, 128.57, 128.13, 111.96, 111.79, 70.61, 70.46, 70.37, 69.97, 69.35, 50.65, 45.83, 40.11; ^{19}F NMR (470 MHz, CDCl_3) δ -13.83, -102.68, -103.89, -104.58, -108.06, -108.77, -183.38, -192.87; ^{19}F NMR (470 MHz, DMSO-*d*₆) δ -143.75 (d, J = 18.0 Hz), -154.60 (t,

$J = 21.3$ Hz), -160.28 (t, $J = 21.3$ Hz); MS (ESI⁺) m/z calcd. for $C_{15}H_{17}F_5N_4O_4^+$ [M+H]⁺ 413.1243, found 413.126310.

5

***N*-(2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)-2,4,6-trifluorobenzamide (8d)**

Synthesized by general procedure 2 using 1-[2-(2-aminoethoxy)ethoxy]-2-(2-azidoethoxy)ethane (99.0 μ L, 0.5 mmol) in DCM (5.0 mL), DIPEA (350 μ L, 2 mmol) and of 2,4,6-trifluorobenzoylchloride (112.0 μ L, 0.75 mmol) and purified by column chromatography with a gradient of EtOH: heptane (20-60 %) to afford **8d** as oily liquid (132 mg, 70%); ¹H NMR (500 MHz, Chloroform-*d*) δ 6.67 – 6.61 (m, 2H), 3.60 (d, $J = 2.0$ Hz, 8H), 3.59 – 3.53 (m, 7H), 3.28 (t, $J = 5.1$ Hz, 2H); ¹³C NMR (126 MHz, CDCl₃) δ 164.44, 164.32, 164.20, 162.43, 162.31, 162.19, 161.64, 161.56, 161.52, 161.44, 159.62, 159.55, 159.51, 159.43, 111.42, 111.39, 111.26, 111.22, 111.10, 111.06, 101.10, 101.07, 101.04, 100.89, 100.88, 100.87, 100.85, 100.68, 100.65, 70.65, 70.63, 70.52, 70.38, 70.02, 69.56, 50.64, 39.86; ¹⁹F NMR (470 MHz, CDCl₃) δ -104.29, -108.88; MS (ESI⁺) m/z calcd. for $C_{15}H_{20}F_3N_4O_4^+$ [M+H]⁺ 377.1431, found 377.144867.

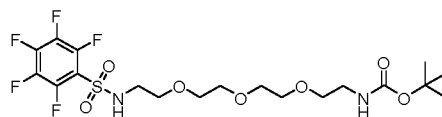
20

***tert*-butyl *N*-{2-[2-(2-(2-**

[(pentafluorophenyl)formamido]ethoxy}ethoxy)ethoxy}ethyl}carbamate (17)

Synthesized by general procedure (2) using *N*-

Boc-2-{2-[2-(2-amino-ethoxy)-ethoxy]-ethoxy} ethylamine (398 μ L, 1.5 mmol) in DCM (15.0 mL),



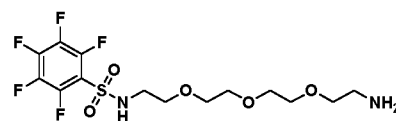
DIPEA (350 μ L, 2 mmol) and pentafluorobenzene sulfonyl chloride (432.0 μ L, 3.0 mmol) and purified by column chromatography with a gradient of EtOH: heptane (20-40 %) to afford **17** as oily liquid (678.0 mg, 87%); ¹H NMR (500 MHz, Chloroform-*d*) δ 3.59 – 3.51 (m, 10H), 3.48 (dd, $J = 5.6, 4.8$ Hz, 2H), 3.28 (q, $J = 4.9$ Hz, 2H), 3.23 (t, $J = 5.2$ Hz, 2H), 1.37 (s, 9H); ¹³C NMR (126 MHz, CDCl₃) δ 167.74, 156.10, 145.55, 145.52, 145.48, 145.45, 145.42, 145.38, 145.35, 144.79, 144.75, 144.69, 144.65, 144.61, 144.58, 144.54, 144.50, 143.50, 143.47, 143.43, 143.39, 143.38, 143.33, 143.30, 142.68, 142.61, 142.57, 142.53, 142.46, 138.94, 138.91, 138.84, 138.81, 138.77, 138.71, 138.66, 136.91, 136.86, 136.81, 136.76, 136.73, 136.68, 136.63, 132.42,

35

130.87, 128.78, 117.01, 116.89, 116.76, 79.39, 70.40, 70.31, 70.13, 70.00, 50.58, 43.34, 40.51, 38.72, 30.34, 28.91, 28.36, 23.73, 22.96, 14.03, 10.94; ^{19}F NMR (470 MHz, CDCl_3) δ -136.82, -146.82, -159.08; MS (ESI^+) m/z calcd. for $\text{C}_{19}\text{H}_{28}\text{F}_5\text{N}_2\text{O}_7\text{S}^+$ $[\text{M}+\text{H}]^+$ 523.1532, found 523.15723. MS (ESI^+) m/z calcd. for $\text{C}_{19}\text{H}_{28}\text{F}_5\text{N}_2\text{O}_7\text{S Na}^+$ $[\text{M}+\text{Na}]^+$ 545.1351, found 545.13994.

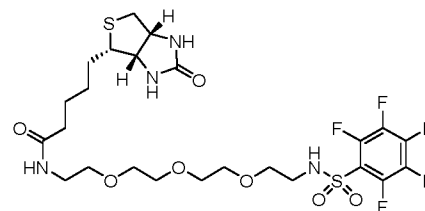
***N*-(2-{2-[2-(2-aminoethoxy)ethoxy]ethoxy}ethyl)-2,3,4,5,6-pentafluorobenzamide (18)**

Trifluoroacetic acid (1.0 mL) was added to a solution of **17** (522.0 mg, 1 mmol) in DCM (10.0 mL), and the reaction mixture was stirred at room temperature for 1 hour. After the solvent was evaporated, the crude residue was purified by column chromatography (MeOH: DCM / 0-10:100-90) to afford **18** as oily liquid (422 mg, quant.); ^1H NMR (500 MHz, Chloroform- d) δ 7.70 (s, 3H), 7.39 (d, J = 6.5 Hz, 1H), 3.93 (s, 3H), 3.68 (t, J = 4.9 Hz, 2H), 3.64 – 3.46 (m, 10H), 3.24 (q, J = 4.5 Hz, 2H), 3.14 (s, 2H); ^{13}C NMR (126 MHz, CDCl_3) δ 162.61, 162.33, 162.04, 161.77, 145.48, 145.45, 145.42, 145.38, 145.34, 144.80, 144.69, 144.59, 143.43, 143.40, 143.37, 143.33, 143.30, 142.73, 142.66, 142.62, 142.58, 142.52, 138.96, 138.92, 138.87, 138.83, 138.78, 138.73, 138.68, 136.93, 136.88, 136.83, 136.79, 136.75, 136.69, 136.65, 119.68, 117.36, 116.76, 116.66, 116.64, 116.52, 115.05, 112.74, 70.05, 69.98, 69.74, 69.65, 69.22, 66.61, 50.24, 42.83, 39.64; ^{19}F NMR (470 MHz, DMSO) δ -73.43, -148.62, -160.44; MS (ESI^+) m/z calcd. for $\text{C}_{14}\text{H}_{19}\text{F}_5\text{N}_2\text{O}_5\text{S}^+$ $[\text{M}+\text{H}]^+$ 423.1008, found 423.10221.



5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl)-N-(2-(2-(2-(2-(pentafluorophenyl)sulfonamido)ethoxy)ethoxy)ethoxy)ethyl)pentanamide (11)

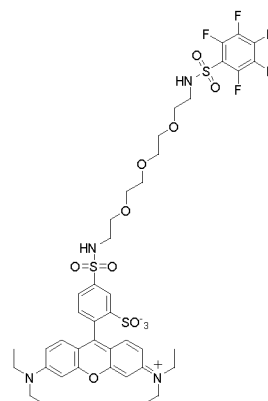
Compound **11** was synthesized by general procedure 2 using (+)-Biotin-(PEO)₄-amine (50.0 mg, 0.12 mmol) in DCM (2.0 mL), DIPEA (87.0 μL , 0.5 mmol) and pentafluorobenzenesulfonylchloride (36.0 μL , 0.24 mmol) and purified by column chromatography with gradient chromatography (MeOH: DCM / 0-10:100-90) to afford **11** as white solid (550 mg, 85%); ^1H NMR (500 MHz, Methanol- d_4) δ 4.39 (dd, J = 7.9, 4.8 Hz, 1H), 4.21 (dd, J = 7.9, 4.4 Hz, 1H), 3.54 – 3.43 (m, 12H), 3.26 (t, J = 5.5 Hz,



2H), 3.23 – 3.16 (m, 2H), 3.14 – 3.06 (m, 2H), 2.97 (t, $J = 2.6$ Hz, 2H), 2.83 (dd, $J = 12.7, 5.0$ Hz, 1H), 2.60 (d, $J = 12.7$ Hz, 1H), 2.12 (t, $J = 7.4$ Hz, 2H), 1.74 – 1.41 (m, 5H), 1.34 (qt, $J = 8.5, 4.2$ Hz, 2H); ^{13}C NMR (126 MHz, MeOD) δ 176.14, 166.09, 147.31, 147.22, 147.17, 147.15, 146.88, 146.85, 146.82, 145.93, 145.20, 144.85, 143.88, 142.91, 142.79, 141.01, 140.96, 140.83, 140.42, 140.28, 140.18, 140.14, 138.37, 138.27, 138.13, 138.13, 136.33, 136.25, 136.17, 118.71, 118.56, 118.42, 111.36, 111.23, 111.11, 71.61, 71.58, 71.52, 71.25, 71.23, 71.19, 70.76, 70.74, 70.63, 63.37, 61.64, 57.01, 54.85, 44.08, 43.93, 43.29, 43.25, 43.21, 41.08, 40.35, 40.33, 36.76, 29.77, 29.52, 26.86; ^{19}F NMR (470 MHz, DMSO) δ -73.43, -148.62, -160.44; MS (ESI⁺) m/z calcd. for $\text{C}_{24}\text{H}_{34}\text{F}_5\text{N}_4\text{O}_7\text{S}_2^+$ [M+H]⁺ 649.1784, found 649.18032.

2-(6-(diethylamino)-3-(diethyliminio)-3H-xanthen-9-yl)-5-(N-(2-(2-(2-(2-(pentafluorophenyl)sulfonamido)ethoxy)ethoxy)ethoxy)ethyl)sulfamoyl)benzene sulfonate (10)

Compound **10** was synthesized by general procedure 2, using **18** (211.0 mg, 0.5 mmol) in DCM (10.0 mL), DIPEA (350 μL , 2 mmol) and sulforhodamine B chloride (577.0 mg, 0.75 mmol) and purified by column chromatography with gradient chromatography (MeOH: DCM / 0-10:100-90) to afford **10** as dark red powder (290 mg, 60 %); ^1H NMR (500 MHz, DMSO- d_6) δ 8.80 (t, $J = 5.6$ Hz, 1H), 8.05 (t, $J = 5.9$ Hz, 1H), 7.95 (dd, $J = 7.9, 1.9$ Hz, 1H), 7.05 (dd, $J = 9.6, 2.5$ Hz, 2H), 6.98 (d, $J = 9.5$ Hz, 2H), 6.95 (d, $J = 2.4$ Hz, 2H), 3.65 (dt, $J = 11.6, 7.2$ Hz, 8H), 3.48 (d, $J = 5.4$ Hz, 6H), 3.45 – 3.40 (m, 6H), 3.17 (q, $J = 5.6$ Hz, 2H), 3.03 (q, $J = 5.9$ Hz, 2H), 1.22 (t, $J = 7.0$ Hz, 12H); ^{13}C NMR (126 MHz, DMSO) δ 157.47, 157.07, 154.98, 147.97, 144.77, 142.73, 141.54, 132.95, 132.62, 130.54, 126.49, 125.64, 113.59, 113.44, 95.35, 69.60, 69.58, 69.49, 69.09, 68.84, 45.22, 42.39, 39.00, 12.43; ^{19}F NMR (470 MHz, DMSO) δ -73.43, -148.62, -160.44; MS (ESI⁺) m/z calcd. for $\text{C}_{41}\text{H}_{48}\text{F}_5\text{N}_4\text{O}_{11}\text{S}_3^+$ [M+H]⁺ 962.2318, found 962.24230.



Example 7 – Activity-based protein profiling of TEV protease

The fluoroaryl ligation concept also was investigated for activity-based protein profiling (ABPP). A series of azide-functionalized derivatives (**8a**, **8b**, **8c** and **8d**, Figure 8A) was

tested. Native or heat-denatured TEV protease was treated with compound **8a**, **8b**, **8c** or **8d** at 10 μ M and subsequently, after gel electrophoresis, exposed to a cyanine alkyne “click” dye for visualization. Gratifyingly, compound **8b** and **8c** demonstrated ability of labeling TEV protease in an activity-dependent manner, as only the native
5 TEV protease and not the preheated sample displayed fluorescence. Compound **8a** was able to label both active and denatured TEV protease while compound **8d** did not label any of the samples. This clearly demonstrates accurate tuning of the fluoro-substituents is an important feature of the method. It is readily achieved by simple change of the nature of the strongly electron-withdrawing group in the para
10 position and/or the overall number of fluoro-substituents. That said, the excessively reactive pentafluorobenzene sulphonamide compound (**8a**) may be “tuned down” by either replacing the sulphonamide with the less electron-withdrawing amide (**8c**) or by removing the two ortho fluoro-substituents (**8b**). However, combining both effects as found in **8d** reduces the reactivity of the compound too much to be applicable for the
15 labeling of even the catalytic cysteine of the active protease (see Figure 8B).

To demonstrate the application of the fluoroaryls to ABPP in a native environment, bacterial cell lysates were treated with compound **8a**, **8b** and **8d** at 10 μ M with or without pretreatment with iodacetamide (1 mM). Interestingly, these compounds
20 displayed markedly different labelling abilities, as the more reactive compounds labelled many more proteins in the cellular lysate. In-gel fluorescence scanning showed that the most active compound (**8a**), labelled many bands with great intensity, whereas compound **8b** appeared much more selective. Azide compound **8d**, the least active, was not able to label any protein in the cell lysate. The pretreatment with iodacetamide
25 inhibited the labeling with the active compounds (**8a** and **8b**) completely, which is another strong indication that these compounds selectively target cysteine residues (see Figure 8C). Moreover, compound **8b** was clearly able to label TEV protease within a cell lysate when spiked with this protease (see Figure 8D). In the gels of both the spiked and non-spiked reactions another major protein was also labeled by compound
30 **8b**. This prominent band was cut from the cell lysate and analyzed by tryptic digestion and MS analysis. Gratifyingly, the protein proved to be an enzyme containing cysteine in the active site. Protein blast at NCBI identified it as chloramphenicol acetyl transferase (CAT1) (See Figure 8D and 8E). Retrospectively, we could expect to find this protein since the applied B834[DE3]pLysS bacteria contained the pLysS plasmid
35 with the gene encoding for CAT1 in order to maintain the plasmid.

TEV protease activity assay

The activity of TEV protease was measured by proteolytic cleavage of the Abz-E N L Y F Q G Y(NO₂) G-OH FRET substrate. TEV protease (170 nM) was incubated with the substrate (160 μM) in aqueous buffer containing 50 mM Tris-HCl, 0.5 mM EDTA, 1 mM DTT at pH 8.3. Hydrolysis was monitored in triplo at 25 °C using a microtiter plate reader (Synergy H4 hybrid reader). A sample without enzyme and one with buffer only were included as negative controls.

10 **Labeling of TEV protease with fluoroaryl probes 2g and 2h followed by trypsin digestion and MALDI-TOF analysis**

Active TEV protease (10 μL, ~ 4 μM) was incubated with iodoacetamide, 2g or 2h (10 μL, 200 μM) in TEV assay buffer (50 mM Tris, 0.5 mM EDTA and 1 mM TCEP, pH 8.5) at 37 °C for 2-16 h. Excess DTT (5 μL, 100 mM) was added for 30 minutes followed by in-solution trypsin digestion. That is, 2 μL 0.5 μg/μL trypsin in 40 μL 0.1 M (NH₄)₂CO₃ was added to the solution and incubated overnight at 37 °C. The tryptic digested peptides were desalted using C18 ZipTips according to the manufacturer's protocol (Millipore). Desalted samples were analyzed by MALDI-TOF using αCHCA as matrix.

20 **Labeling of active and heat-inactivated TEV protease with azide probes 8a, 8b, 8c and 8d, followed by click chemistry-mediated fluorescent labeling (in-gel fluorescence scanning)**

Active and heat-inactivated (left at 95 °C for 15 min) TEV protease (10 μL, ~ 4 μM) were incubated with fluoroaryl azide probes 8a, 8b, 8c and 8d (10 μL, 20 μM) in TEV assay buffer (50 mM Tris, 0.5 mM EDTA and 1 mM TCEP) at 37 °C for 16 h. Subsequently, click mixture consisting of CuSO₄ (1 μL, 45 mM), THPTA (1 μL, 90 mM) and sodium ascorbate (1 μL, 60 mM) as well as alkyne cyanine dye 718 (Sigma Aldrich 30154) (3 μL, 1 mM) were added at 22 °C and left for 2 h. The reaction mixture was analyzed by SDS-PAGE electrophoresis followed by in-gel fluorescence scanning at 30 λex 650 nm and λem 680 nm on a Typhoon FLA 7000 laser scanner, and staining with Coomassie Brilliant Blue.

Preparation and labeling of a bacterial cell lysate spiked with TEV protease with probe 8b (in-gel fluorescence scanning)

A culture of B834 (DE3)pLysS bacteria (Merck Millipore) was grown overnight at 37 °C in LB medium supplemented with chloramphenicol. Cells were harvested by centrifugation, lysed by resuspension in lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, pH 8.0) and sonicated on ice (6 x 10 s). A cleared lysate was obtained by centrifugation at 13,000 rpm. After addition of active TEV protease (5 µL, ~ 4 µM) to an aliquot of bacterial cell lysate (10 µL), fluoroaryl azide probe 8b (10 µL, 20 µM) in TEV buffer (50 mM Tris, 0.5 mM EDTA and 1 mM TCEP) was added and allowed to react at 37 °C for 16 h. Subsequently, click mixture consisting of CuSO₄ (1 µL, 45 mM), THPTA (1 µL, 90 mM) and sodium ascorbate (1 µL, 60 mM) as well as alkyne cyanine dye 718 (Sigma Aldrich 30154) (3 µL, 1 mM) were added followed by incubation at 22 °C for 2 h. The reaction mixture was analyzed by SDS-PAGE electrophoresis followed by in-gel fluorescence scanning at λ_{ex} 650 nm and λ_{em} 680 nm on a Typhoon FLA 7000 laser scanner, and staining with Coomassie Brilliant Blue.

The intense band observed just below the 25-kDa protein marker band was excised from the destained gel and cut into small pieces. The pieces were further destained by soaking them in 0.1 M aq. sodium bicarbonate/AcN (1:1). Subsequently, the pieces were incubated in 100 µL trypsin digestion buffer (10 ng/µL trypsin in 0.1 M aq. sodium bicarbonate) at 37 °C overnight. The tryptic digested peptides were desalted using C18 ZipTips according to the manufacturer's protocol (Millipore). Desalted samples were analyzed by MALDI-TOF using α CHCA as matrix. An in-gel digestion after reduction and treatment with iodoacetamide was performed and analyzed by MALDI-TOF MS. The mass list was generated from the MALDI-TOF data.

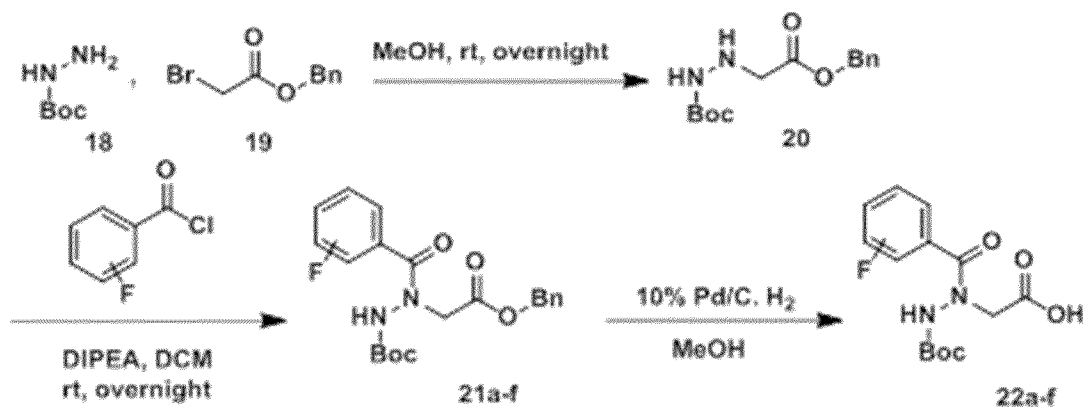
25

Example 8 – Inhibition of Caspase-1

Selective covalent inhibitor of Caspase-1 protease that relies on combining a tuned reactive group with an anchor and/or recognition peptide was also explored. It was hypothesized that those probes with low or none reactivity toward cysteine in aqueous conditions could be very useful for achieving optimum selectivity and potency toward caspase-1 if positioned well at the active site by an anchor and/or recognition peptide. Caspase-1, interleukin-1 β converting enzyme, which is a potential target in drug discovery, is cysteine protease that has a critical role in inflammatory responses. Diseases like Osteoarthritis (OA) and Rheumatoid arthritis (RA) are a direct result of the dysregulation of inflammatory processes and fluctuations in cytokines levels which

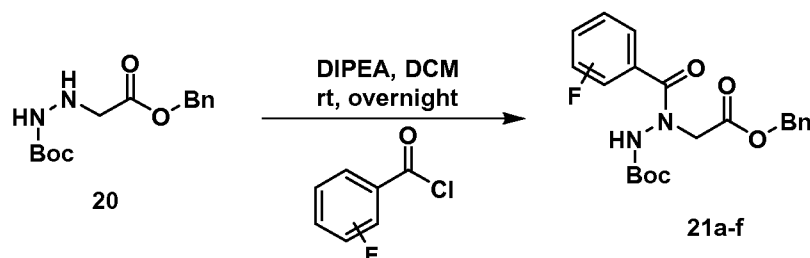
35

are regulated by caspase-1. Hence, six small molecules that combined selected reactive groups with an aspartate analog that should be recognizable by the active site of caspase 1 as this has an absolute preference for aspartate in P1 position. Based on known substrate specificity potential inhibitors **22a-f** and **23-25** were designed (see scheme 5). They are readily accessible by reacting benzylbromoacetate **18** with boc-hydrazine **19** that yielded **20** that was then coupled to fluoroaryls acid or acid chloride that afford **21a-f**. Deprotection of **21a-f** by catalytic hydrogenation afforded **22a-f**. The caspase-1 inhibition assay using Ac-YVAD-AMC as substrate shows the ability of these compounds **22a-f** (data no shown) to inhibit Caspase 1 activity at low micromolar range which was in comparable level to Ac-YVAD-CMK (**41**), a standard caspase 1 inhibitor. To aim for enhanced selectivity and potency a peptide motif based on caspase-1 specificity profile was attached to **22b**, **22c** or **22f**. This was done by esterifying **22b**, **22c** or **22f** to HMBA linker functionalized PEGA-800 resin followed by deprotection of the Boc group. Then reductive amination the terminal amine with boc-alaninal followed by standard SPPS protocol and cleavage by 0.1 N NaOH to afforded **23-25**. Applying **23**, **24** or **25** to the Caspase-1 inhibition assay showed a strong inhibition of the enzyme activity at 10 or 100 nM, all have better or comparable potency to Ac-YVAD-CMK. Moreover **23**, **24** or **25** were not able to inhibit TEV protease at the same concentration, see Figure 9 A and B.



Scheme 5. Synthesis and chemical structures of fluoroaryls based inhibitors **22a-f** and **23-25**.

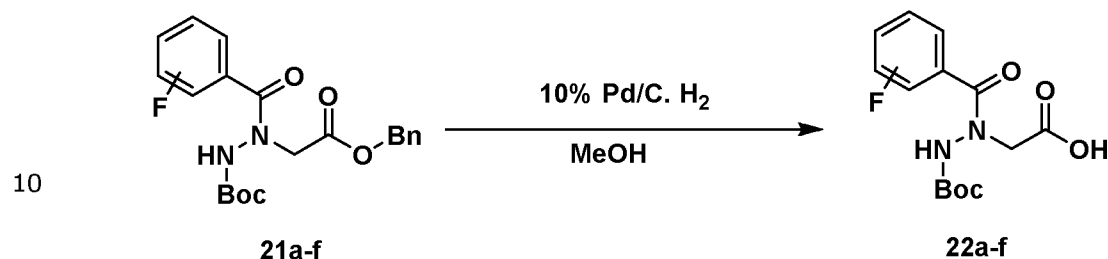
General procedure A



To a solution of **20** (840 mg, 3 mmol) in DCM (20 ml), Acid chloride (3.0 mmol) was added. After 10 minutes, DIPEA (521 μ l, 3.0 mmol) was added. The reaction was stirred at rt for 3 h. The solvent was evaporated and the crude residue was purified by column chromatography (EtOAc: heptane / 15:85) to afford the product.

5

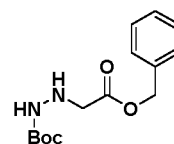
General procedure B



Protected acid (0.1 mmol) was dissolved in degassed MeOH (5.0 ml), 10 % Pd/C (0.05 mmol), the system was degassed in vacuo in presence of hydrogen gas. The reaction mixture was stirred under hydrogen atmosphere at rt overnight. The reaction mixture was filtered over a celite pad and washed with MeOH. The solvent was evaporated and the residue was purified by flash chromatography (MeOH: dichloromethane /20:80) to afford the product.

Benzyl 2-({[(tert-butoxy)carbonyl]amino}amino)acetate **20**

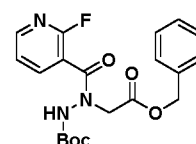
To a solution of (tert-butoxy)carbohydrazide (1.98 g, 15.0 mmol in EtOH (30.0 mL), benzyl 2-bromoacetate (1.15 g, 5.0 mmol) was added gradually. The reaction was stirred at rt for 5 h. The solvent was evaporated *in vacuo* and the crude residue was purified by column chromatography (EtOAc: heptane / 40:100) to afford the product **20** (0.67 g, 49%). ¹H NMR (500 MHz, Methanol-*d*₄) δ 7.47 – 7.25 (m, 5H), 5.20 (s, 2H), 3.62 (s, 2H), 1.46 (s, 9H); ¹³C NMR (126 MHz, MeOD) δ 172.40, 158.51, 137.32, 129.57, 129.34, 129.31, 81.21, 67.58, 53.54, 28.66.; MS (ESI⁺) *m/z* calcd. for C₁₄H₂₁N₂O₄⁺ [M+H]⁺ 281.1496, found 281.1510.



30

Benzyl 2-(N-({[(tert-butoxy)carbonyl]amino}-1-(2-fluoropyridin-3-yl)formamido)acetate **21a**

To a solution of 2-fluoropyridine-3-carboxylic acid (169.0 mg, 1.2 mmol) in DMF (5.0 mL), Pyridine (0.8 mL, 10 mmol), N-(3-

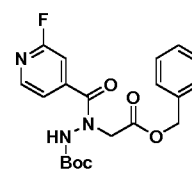


35

Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC) (960.0 mg, 5.0 mmol) were added. After 10 min **20** (280.0 mg, 1.0 mmol) was added to the reaction mixture that was subsequently stirred at rt overnight. The solvent was evaporated in vacuo and the crude residue was purified by column chromatography (EtOAc: heptane / 40:100) to afford the product **21a** (270.0 mg, 67%).¹H NMR (500 MHz, Methanol-*d*₄) δ 8.31 (dd, *J* = 5.0, 1.9 Hz, 1H), 7.98 (ddd, *J* = 9.2, 7.3, 2.0 Hz, 1H), 7.44 – 7.32 (m, 6H), 5.26 (s, 2H), 1.31 (s, 9H).¹³C NMR (126 MHz, MeOD) δ 169.25, 161.27, 159.36, 150.26, 150.15, 141.63, 137.06, 129.61, 129.41, 129.36, 122.62, 122.58, 119.40, 119.14, 82.81, 68.24, 51.33, 28.30; ¹⁹F NMR (470 MHz, CDCl₃) δ -66.49; MS (ESI⁺) *m/z* calcd. for C₂₀H₂₃FN₃O₅⁺ [M+H]⁺ 404.1616, found 404.1642.

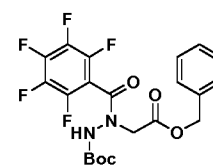
Benzyl 2-(N-[[tert-butoxy)carbonyl]amino]-1-(2-fluoropyridin-4-yl)formamido)acetate 21b

To a solution of 2-fluoropyridine-4-carboxylic acid (317.0 mg, 2.25 mmol) in DMF (5.0 mL), Pyridine (1.2 mL, 15.0 mmol), N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC) (1.4 g, 7.5 mmol) were added. After 10 min **20** (420.0 mg, 1.0 mmol) was added to the reaction mixture that was subsequently stirred at rt overnight. The solvent was evaporated in vacuo and the crude residue was purified by column chromatography (EtOAc: heptane / 40:100) to afford the product **21b** (417.0 mg, 69%), ¹H NMR (500 MHz, Methanol-*d*₄) δ 8.29 (d, *J* = 5.1 Hz, 1H), 7.43 – 7.33 (m, 6H), 7.15 (s, 1H), 5.25 (s, 2H), 1.34 (s, 9H); ¹³C NMR (126 MHz, MeOD) δ 171.57, 169.38, 165.57, 163.66, 149.78, 149.72, 149.05, 148.94, 137.03, 129.65, 129.48, 129.40, 120.67, 120.64, 108.82, 108.51, 82.94, 68.31, 51.53, 28.37; ¹⁹F NMR (470 MHz, CDCl₃) δ -66.32; MS (ESI⁺) *m/z* calcd. For C₂₀H₂₃FN₃O₅⁺ [M+H]⁺ 404.1616, found 404.1642.



Benzyl 2-(N-[[tert-butoxy)carbonyl]amino]-1-(pentafluorophenyl)formamido)acetate 21c

Synthesized by general procedure A using **20** (840.0 mg, 3.0 mmol) in DCM (20.0 mL), pentafluorobenzoyl chloride (800.0 μL, 3.0 mmol) and DIPEA (521.0 μL, 3.0 mmol) at rt for 3 h to afford **21c** (1.3 g, 93%), ¹H NMR (500 MHz, Chloroform-*d*) δ 7.33 – 7.28 (m, 5H), 7.06 (s, 1H), 5.16 (s, 2H), 1.29 (s, 9H); ¹³C NMR (126 MHz, CDCl₃) δ 168.42, 168.00,

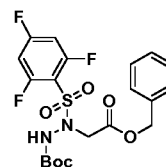


161.71, 153.23, 144.42, 143.16, 142.28, 141.13, 138.47, 136.38, 134.64, 128.80, 128.76, 128.53, 110.17, 110.04, 109.86, 83.36, 67.80, 48.59, 27.81; MS (ESI⁺) m/z calcd. for C₂₁H₂₀F₅N₂O₅⁺ [M+H]⁺ 475.1287, found 475.1313.

5

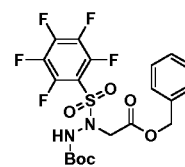
Benzyl 2-(N-[[tert-butoxy]carbonyl]amino}2,4,6-trifluorobenzenesulfonamido)-acetate **21d**

Synthesized by general procedure A using **20** (840.0 mg, 3.0 mmol) in DCM (20.0 mL), 2,4,6-trifluorobenzene-1-sulfonyl chloride (436.0 μL, 3.0 mmol) and DIPEA (521.0 μL, 3.0 mmol) at rt for 3 h to afford **21d** (1.35 g, 95%), ¹H NMR (500 MHz, Chloroform-*d*) δ 7.41 – 7.22 (m, 5H), 6.81 – 6.59 (m, 2H), 5.13 (s, 2H), 4.46 (s, 1H), 1.27 (s, 9H); ¹³C NMR (126 MHz, CDCl₃) δ 168.48, 166.64, 164.59, 162.50, 162.38, 160.42, 160.38, 153.03, 134.83, 128.72, 128.45, 102.04, 101.82, 101.61, 82.49, 67.61, 51.73, 27.86; ¹⁹F NMR (470 MHz, CDCl₃) δ -97.60, -101.14; MS (ESI⁺) m/z calcd. for C₂₀H₂₂F₃N₂O₆S⁺ [M+H]⁺ 475.1145, found 475.1196.



Benzyl 2-(N-[[tert-butoxy]carbonyl]amino}pentafluorobenzenesulfonamido)-acetate **21e**

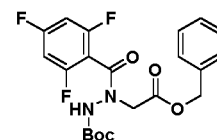
Synthesized by general procedure A using **20** (840.0 mg, 3.0 mmol) in DCM (20.0 mL), pentafluorobenzene-1-sulfonyl chloride (800.0 μL, 3.0 mmol) and DIPEA (521.0 μL, 3.0 mmol) at rt for 3 h to afford **21e** (1.3 g, 93%), ¹H NMR (500 MHz, Chloroform-*d*) δ 7.37 – 7.23 (m, 5H), 5.14 (s, 2H), 4.45 (s, 1H), 1.29 (s, 9H); ¹³C NMR (126 MHz, CDCl₃) δ 168.15, 153.02, 146.33, 144.31, 143.38, 138.75, 135.29, 135.29, 134.63, 130.88, 128.79, 128.77, 128.75, 128.57, 128.54, 128.51, 128.46, 83.04, 67.82, 51.89, 27.80; ¹⁹F NMR (470 MHz, CDCl₃) δ -134.36, -144.68, -159.34. MS (ESI⁺) m/z calcd. for C₂₀H₂₀F₅N₂O₆S⁺ [M+H]⁺ 511.0957, found 511.0992.



30

Benzyl 2-(N-[[tert-butoxy]carbonyl]amino}-1-(2,4,6-trifluorophenyl)formamido)-acetate **21f**

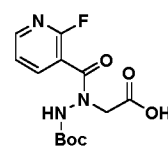
Synthesized by general procedure A using **20** (420.0 mg, 1.50 mmol) in DCM (20.0 mL), 2,4,6-trifluorobenzoyl chloride (294.0



35

μL , 2.25 mmol) and DIPEA (261.0 μL , 1.5 mmol) at rt for 3 h to afford **21f** (1.3 g, 93 ^1H NMR (500 MHz, Methanol- d_4) δ 7.44 – 7.32 (m, 4H), 6.98 (t, J = 8.8 Hz, 2H), 5.26 (s, 2H), 1.34 (s, 9H); ^{13}C NMR (126 MHz, MeOD) δ 169.07, 166.28, 166.16, 166.04, 165.95, 164.29, 164.17, 164.05, 162.21, 160.20, 156.00, 137.06, 129.57, 129.37, 129.33, 110.96, 110.92, 110.78, 110.74, 110.61, 110.57, 101.54, 82.82, 68.21, 51.13, 28.22.; ^{19}F NMR (470 MHz, CDCl_3) δ -104.29, -108.88; MS (ESI $^+$) m/z calcd. for $\text{C}_{21}\text{H}_{22}\text{F}_3\text{N}_2\text{O}_5$ $[\text{M}+\text{H}]^+$ 439.1475, found 439.1514.

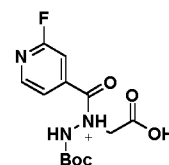
2-(N-([(tert-butoxy)carbonyl]amino)-1-(2-fluoropyridin-3-yl)formamido)acetate 22a
 Synthesized by general procedure B using **21a** (100.0 mg, 0.25 mmol) in MeOH (10.0 ml), 10 % Pd/C (132.0 mg, 0.125 mmol) at rt overnight to afford **23a** as a white solid (75.0 mg, quant.), ^{19}F NMR (470 MHz, MeOD) δ -69.47; MS (ESI $^+$) m/z calcd. for $\text{C}_{13}\text{H}_{17}\text{FN}_3\text{O}_5$ $[\text{M}+\text{H}]^+$ 314.1147, found 314.1172.



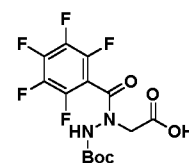
15

Benzyl 2-(N-([(tert-butoxy)carbonyl]amino)-1-(2-fluoropyridin-4-yl)formamido)-acetate 22b

Synthesized by general procedure B, using **21b** (241.0 mg, 0.6 mmol) in MeOH (10.0 ml), 10% Pd/C (132.0 mg, 0.125 mmol) at rt overnight to afford **22b** (75.0 mg, quant.), ^1H NMR (500 MHz, Methanol- d_4) δ 8.16 (d, J = 5.1 Hz, 1H), 7.30 (d, J = 5.2 Hz, 1H), 7.07 (s, 1H), 3.65 – 3.43 (m, 1H), 1.28 – 1.09 (m, 9H); ^{13}C NMR (126 MHz, MeOD) δ 176.01, 171.56, 165.55, 163.65, 156.20, 150.53, 150.47, 148.89, 148.78, 120.68, 108.78, 108.46, 82.94, 52.77, 28.36.; ^{19}F NMR (470 MHz, CDCl_3) δ -66.40; MS (ESI $^+$) m/z calcd. for $\text{C}_{13}\text{H}_{17}\text{FN}_3\text{O}_5$ $[\text{M}+\text{H}]^+$ 314.1147, found 314.1172.

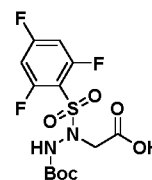


2-(N-([(tert-butoxy)carbonyl]amino)-1-(pentafluorophenyl)formamido)acetate 22c
 Synthesized by general procedure B using **21c** (100.0 mg, 0.2 mmol) in MeOH (10.0 ml), 10 % Pd/C (126.0 mg, 0.10 mmol) at rt overnight to afford **22c** (75.0 mg, quant.), MS (ESI $^+$) m/z calcd. for $\text{C}_{14}\text{H}_{14}\text{F}_5\text{N}_2\text{O}_5$ $[\text{M}+\text{H}]^+$ 385.0817, found 385.08403



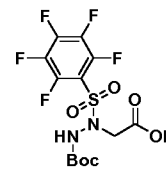
**2-(N-[[[(tert-butoxy)carbonyl]amino]2,4,6-trifluorobenzenesulfonamido)acetate
22d**

Synthesized by general procedure B using **21d** (50.0 mg, 0.1 mmol) in MeOH (3.0 ml), 10 % Pd/C (126.0 mg, 0.10 mmol) at rt overnight to afford **22d** as a white solid (38.0 mg, quant¹H NMR (500 MHz, Methanol-*d*₄) δ 6.96 (t, *J* = 9.1 Hz, 2H), 1.25 (s, 9H); ¹³C NMR (126 MHz, MeOD) δ 176.18, 168.34, 168.21, 168.09, 166.30, 166.18, 166.05, 166.05, 163.99, 163.94, 163.86, 163.82, 161.93, 161.88, 161.80, 161.74, 155.85, 114.90, 103.20, 102.98, 102.76, 82.85, 54.25, 28.31; ¹⁹F NMR (470 MHz, CDCl₃) δ -96.91, -101.06 MS (ESI⁺) *m/z* calcd. For C₁₃H₁₆F₃N₂O₆S⁺ [M+H]⁺ 385.0676, found 385.0696



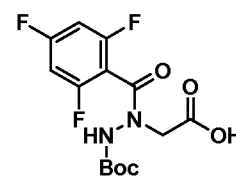
2-(N-[[[(tert-butoxy)carbonyl]amino]pentafluorobenzenesulfonamido)acetate 22e

Synthesized by general procedure B using **21e** (51.0 mg, 0.1 mmol) in MeOH (3.0 ml), 10 % Pd/C (53.0 mg, 0.05 mmol) at rt overnight to afford **22e** (31.0 mg, 74%), MS (ESI⁺) *m/z* calcd. For C₁₃H₁₄F₅N₂O₆S⁺ [M+H]⁺ 421.0487, found 421.0629.



2-(N-[[[(tert-butoxy)carbonyl]amino]-1-(2,4,6-trifluorophenyl)formamido)acetic acid 22f

Synthesized by general procedure B using **21f** (248.0 mg, 0.65 mmol) in MeOH (30.0 ml), 10 % Pd/C (410.0 mg, 0.325 mmol) at rt overnight to afford **22f** (181.0 mg, 80%), ¹H NMR (500 MHz, Methanol-*d*₄) δ 7.05 – 6.89 (m, 2H), 3.97 – 3.65 (m, 1H), 1.35 (s, 9H); ¹³C NMR (126 MHz, MeOD) δ 169.07, 166.28, 166.16, 166.04, 165.95, 164.29, 164.17, 164.05, 162.10, 160.12, 156.00, 137.06, 129.57, 129.37, 129.33, 110.96, 110.92, 110.78, 110.74, 110.61, 110.57, 101.54, 82.82, 68.21, 51.13, 28.22; MS (ESI⁺) *m/z* calcd. For C₁₄H₁₆F₃N₂O₅S⁺ [M+H]⁺ 349.1006, found 349.1031.



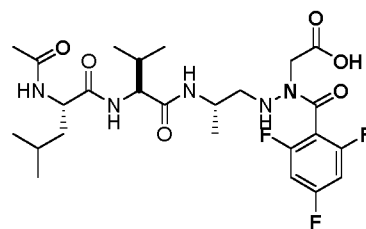
Synthesis of peptide inhibitors compounds 23-25

The peptide synthesis is performed according to standard SPPS synthesis. PEGA₈₀₀ was coupled with 4-hydroxymethyl benzoic acid (HMBA) using the TBTU activation. Upon the esterification (1 h) in dry DCM of the PEGA-linker, using modified building blocks **22b**, **22c** and **22f** (1 equiv) activation with mesitylene-sulfonyl-5-nitro-1,2,4-triazole (MSNT) and *N*-methyl imidazole (Melm), after Boc Deprotection using a cocktail of 95% TFA, 3%water and 2%TIPS, the resulting resin was reductively alkylated with Boc-alaninal (2 equiv.) in 2% acetic acid/DMF and then addition of NaCNBH₃ (10 equiv.) for 2 hours. After Boc Deprotection, L-Boc- amino acids were coupled using the general procedure for peptide couplings described below and finally the free amine was acetylated using acetic acid. The peptide was cleaved from the resin using 5% TEA/water for 2 hours.

General procedure for peptide couplings: TBTU couplings were performed by dissolving the amino acid (3 equiv.) in DMF with 4-ethylmorpholine (NEM) (3 equiv.), followed by addition of 1*H*-benzotriazolyl tetramethyluronium tetrafluoroborate (TBTU) (3 equiv.). The resulting solution was left to pre-activate for 3 min before being added to the resin. Coupling reactions were allowed a reaction time of 2–3 h). Peptide couplings were generally performed in an amount of solvent just enough to cover the resin and washings were conducted slowly and without back mixing. After reaction, the resin was washed with DMF (×6) and finally checked using the Kaiser test.

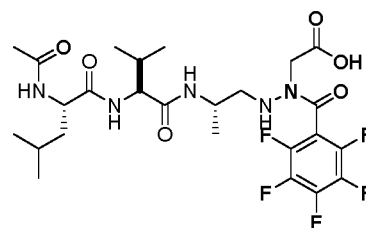
H-leu-Val-Ala*22f (23)

Compound **23** was synthesized from **22f** as described above and purified by HPLC with the following linear gradient: 0–20 min, 0 → 60% B; 20–22 min 60 → 100% B, *t*_R = 10.0 min. FT-ICR MS (ESI+) *m/z* calcd. for C₂₅H₃₆F₃N₅O₆⁺ [M+H]⁺ 560.2690, found 560.2693



H-leu-Val-Ala*22c (24)

Compound **24** was synthesized as described above and purified by HPLC with the following linear gradient: 0–20 min, 0 → 50% B; 20–22 min 50 → 100% B, *t*_R = 6.0 min.

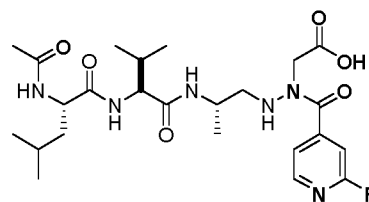


FT-ICR MS (ESI+) m/z calcd. for $C_{25}H_{34}F_5N_5O_6^+$ $[M+H]^+$ 596.2502, 596.2636.

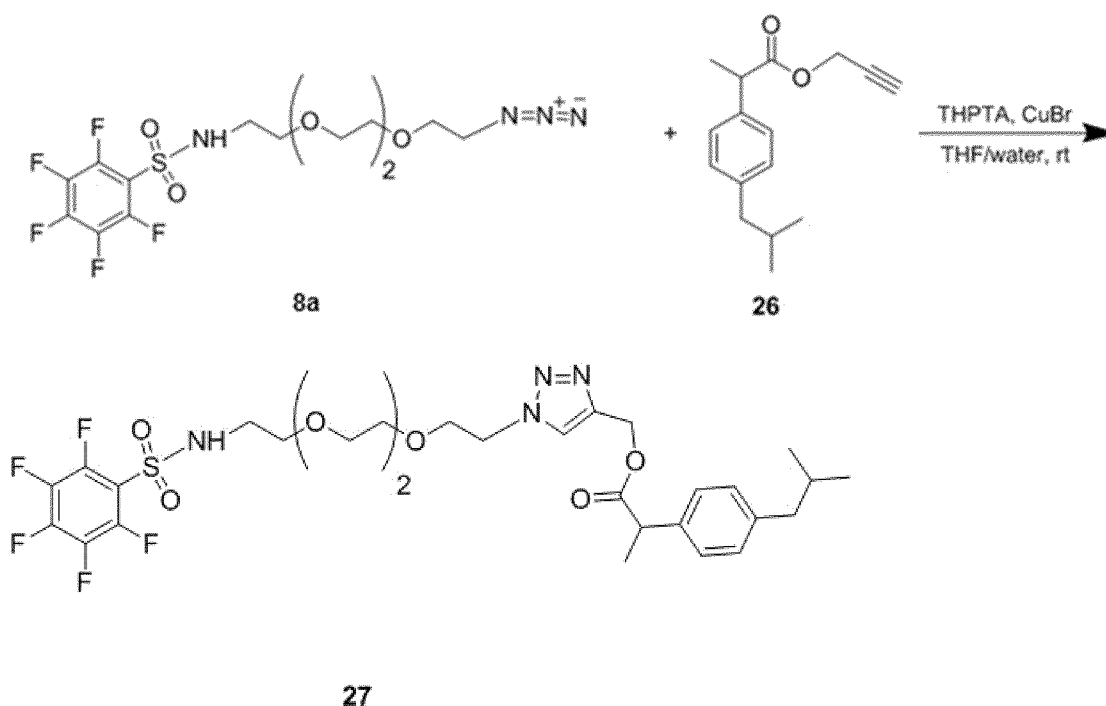
H-leu-Val-Ala*22b (25)

Compound **25** was synthesized as described above and purified by HPLC with the following linear gradient: 0–20 min, 0 → 50% B; 20–22 min 50 → 100% B, t_R = 6.0 min.

FT-ICR MS (ESI+) m/z calcd. for $C_{24}H_{38}FN_5O_6^+$ $[M+H]^+$ 525.2831, 525.2831, found 525.3027



10 Example 9 – Labelling of Human Serum Albumin



Synthesis of compound 27

- 15 Copper(I) bromide (5.28 μ mol, 0.2 equiv.) was added to a solution of Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) (2.64 μ mol, 0.1 equiv.) in water (300 μ l) and shaken very well. THF (300 μ l) was added and the mixture was purged with nitrogen for ~10 minutes. The mixture was added to a solution of alkyne **26** (26.45 μ mol, 2 equiv.) and compound **8a** (26.45 μ mol, 1 equiv.) in 500 μ l THF. The reaction
- 20 was stirred at room temperature overnight. The product was isolated by column chromatography. Yield: 95 %. 1H NMR (500 MHz, Chloroform- d) δ 7.64 (s, 1H), 7.17 (dd, J = 8.0, 1.9 Hz, 2H), 7.09 – 7.04 (m, 2H), 6.43 (s, 1H), 5.30 – 5.10 (m, 2H), 3.95 –

3.26 (m, 17H), 2.43 (d, $J = 7.2$, 2H), 1.89 -1.76 (m, , 1H), 1.47 (d, $J = 7.2$, 3H), 0.88 (d, $J = 6.1$ Hz, 6H). ^{13}C NMR (126 MHz, CDCl_3) δ 174.79, 140.77, 137.52, 129.46, 127.33, 70.69, 70.59, 70.57, 70.45, 69.51, 69.22, 58.10, 50.39, 45.16, 45.09, 43.50, 29.51, 22.84, 14.26.

5

Conjugation human serum albumin with prodrug

A solution of Human Serum Albumin in PBS buffer (pH = 8.5, 100 μM , 70 μL , 7.0 nmol) was added to a solution of compound **27** in acetonitrile (93.3 μL , 28.0 nmol, 4 equiv.).

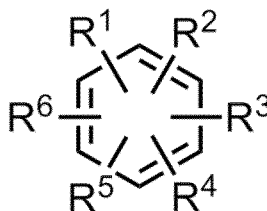
10 PBS buffer in 30% ACN (pH = 8.5, 450 μL) was added in order to obtain a concentration of 11 μM of compound **27**. The mixture was shaken at 31 $^{\circ}\text{C}$ for 4 days. The mixture was transferred to an Amicon^R Ultra-0.5 filter device and centrifuged at 14000 G for 12 minutes at 4 $^{\circ}\text{C}$. The filtration was repeated a total of 3 times, where a 17% acetonitrile in PBS (pH 8.5) was added to the filter device after each round of centrifugation. After
15 filtration the product was once again confirmed by LCMS, see Figure 10.

Items

20 The invention may also be defined by the following items.

1. A method for reacting a compound of Formula I with a cysteine residue, thereby forming a covalent bond, wherein the cysteine is contained in the sequence of a polypeptide, wherein said compound has the following structure:

25



Formula I

wherein:

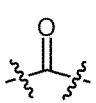
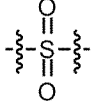
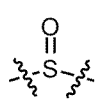
30 R^3 is a leaving group

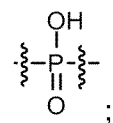
R^1 is R^7 ;

R^7 is selected from the group consisting of $-X-Z-R^8$, $-Z-T-R^8$, and $-Z-T-(R^8)_2$;

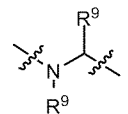
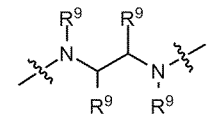
wherein

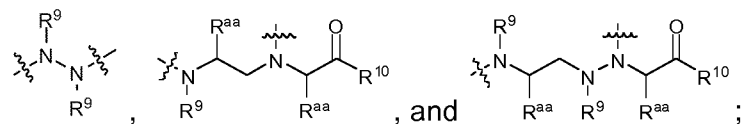
X is selected from the group consisting of a bond, $-\text{CH}_2-\text{N}(\text{R}^9)-$, and $-\text{CH}_2-\text{O}-$;

Z is selected from the group consisting of a bond, , , , and



5

T is selected from the group consisting of a bond, , ,



R^8 is individually selected from the group consisting of amino acids, amino acid derivatives, peptides, peptide derivatives, peptidomimetic, alkylamines, protecting groups, $-\text{NH}-\text{linker}-\text{Y}$, $-\text{linker}-\text{Y}$ and $-\text{linker}-\text{R}^{14}$, wherein said peptide optionally may be N- and/or C-terminally modified;

10

Y is a labelling molecule, a drug molecule or a prodrug;

R^9 is individually selected from the group consisting of $-\text{H}$, alkyl-COOH and an amino acid side chain;

15

R^{10} is selected from the group consisting of $-\text{OH}$ and R^8 ;

R^{14} is a reactive group;

R^2 , R^4 , R^5 , R^6 , are individually selected from the group consisting of $-\text{H}$, R^7 , an amino acid side chain and an electron withdrawing group;

under the proviso that at least one of R^2 , R^4 , R^5 , and R^6 is an electron withdrawing group.

20

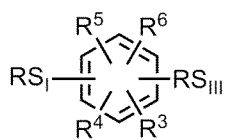
2. The method according to item 1, wherein R^1 is para to R^3 .

3. The method according to item 1, wherein R^1 is ortho to R^3 .

25

4. The method according to any one of the preceding items, wherein R^8 is a peptide, wherein said peptide optionally may be N- and/or C-terminally modified.

5. The method according to item 4, wherein the peptide is linked to the benzene core of the compound of Formula I, Z or T via the N- terminus or the C-terminus.
- 5 6. The method according to any one of the preceding items, wherein the benzene of the compound of Formula I is integrated into the backbone of a peptide, wherein said peptide optionally may be N- and/or C-terminally modified.
7. The method according to any one of the preceding items, wherein the
10 compound of Formula I is a compound of the general formula II

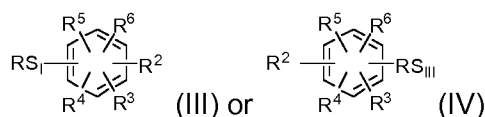


Formula II

wherein

- 15 RS_I when linked to RS_{III} via RS_{II} together forms the peptide RS_I-RS_{II}-RS_{III}, which is a substrate for a peptide cleaving enzyme; and
RS_I and RS_{III} individually are selected from the group consisting of amino acids and peptides, wherein said amino acid or peptide optionally may be N- or C-terminally modified; and
- 20 RS_{II} is selected from the group consisting of a bond, amino acids and peptides; and R³, R⁴, R⁵ and R⁶ are as defined in item 1.
8. The method according to any one of items 1 to 6, wherein
R¹ is RS_I;
25 R² is RS_{III};
RS_I and RS_{III} individually are selected from the group consisting of amino acids and peptides, wherein said amino acid or peptide optionally may be N- or C-terminally modified;
RS_I when linked to RS_{III} via RS_{II} together forms the peptide RS_I-RS_{II}-RS_{III}, which
30 is a substrate for a peptide cleaving enzyme; and
RS_{II} is selected from the group consisting of a bond, amino acids and peptides.

9. The method according to any one of items 1 to 6, wherein the compound of Formula I is a compound of the general formula III or IV:



5 wherein

RS_I when linked to RS_{II}-RS_{III} or RS_{III} when linked to RS_I-RS_{II} together forms the peptide RS_I-RS_{II}-RS_{III}, which is a substrate for a peptide cleaving enzyme, ; and RS_I and RS_{III} individually are selected from the group consisting of amino acids and peptides, wherein said amino acid or peptide optionally may be N- or C-

10 terminally modified; and

RS_{II} is selected from the group consisting of a bond, amino acids and peptides; and R², R³, R⁴, R⁵ and R⁶ are as defined in item 1.

10. The method according to any one of items 1 to 6, wherein

15 R¹ is RS_I or RS_{III};

RS_I and RS_{III} individually are selected from the group consisting of amino acids and peptides, wherein said amino acid or peptide optionally may be N- or C-terminally modified;

RS_I when linked to RS_{III} via RS_{II} together forms the peptide RS_I-RS_{II}-RS_{III}, which is a substrate for a peptide cleaving enzyme; and

20

RS_{II} is selected from the group consisting of a bond, amino acids and peptides.

11. The method according to any one of items 7 to 10, wherein the substrate for a peptide cleaving enzyme is the recognition sequence for a cysteine protease

25 selected from the group consisting of cathepsin B (EC 3.4.22.1), papain (EC 3.4.22.2), ficain (EC 3.4.22.3), chymopapain (EC 3.4.22.6), asclepain (EC 3.4.22.7), clostripain (EC 3.4.22.8), cerevisin (EC 3.4.21.48), streptopain (EC 3.4.22.10), insulysin (EC 3.4.24.56), γ -glutamyl hydrolase (EC 3.4.19.9),

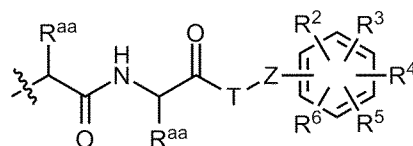
actinidain (EC 3.4.22.14), cathepsin L (EC 3.4.22.15), cathepsin H (EC

30 3.4.22.16), prolyl oligopeptidase (EC 3.4.21.26), thimet oligopeptidase (EC 3.4.24.15), proteasome endopeptidase complex (EC 3.4.25.1), saccharolysin

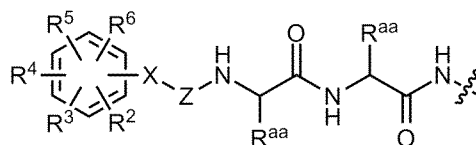
(EC 3.4.24.37), kexin (EC 3.4.21.61), Cathepsin T (EC 3.4.22.24), Glycyl endopeptidase (EC 3.4.22.25), Cancer procoagulant (EC 3.4.22.26), cathepsin S (EC 3.4.22.27), picornain 3C (EC 3.4.22.28), picornain 2A (EC 3.4.22.29),

- Caricain (EC 3.4.22.30), Ananain (EC 3.4.22.31), Stem bromelain (EC 3.4.22.32), Fruit bromelain (EC 3.4.22.33), Legumain (EC 3.4.22.34), Histolysain (EC 3.4.22.35), caspase-1 (EC 3.4.22.36), Gingipain R (EC 3.4.22.37), Cathepsin K (EC 3.4.22.38), adenain (EC 3.4.22.39), bleomycin hydrolase (EC 3.4.22.40), cathepsin F (EC 3.4.22.41), cathepsin O (EC 3.4.22.42), cathepsin V (EC 3.4.22.43), nuclear-inclusion-a endopeptidase (EC 3.4.22.44), helper-component proteinase (EC 3.4.22.45), L-peptidase (EC 3.4.22.46), gingipain K (EC 3.4.22.47), staphopain (EC 3.4.22.48), separase (EC 3.4.22.49), V-cath endopeptidase (EC 3.4.22.50), cruzipain (EC 3.4.22.51), calpain-1 (EC 3.4.22.52), calpain-2 (EC 3.4.22.53), calpain-3 (EC 3.4.22.54), caspase-2 (EC 3.4.22.55), caspase-3 (EC 3.4.22.56), caspase-4 (EC 3.4.22.57), caspase-5 (EC 3.4.22.58), caspase-6 (EC 3.4.22.59), caspase-7 (EC 3.4.22.60), caspase-8 (EC 3.4.22.61), caspase-9 (EC 3.4.22.62), caspase-10 (EC 3.4.22.63), caspase-11 (EC 3.4.22.64), peptidase 1 (mite) (EC 3.4.22.65), calicivirin (EC 3.4.22.66), zingipain (EC 3.4.22.67), Ulp1 peptidase (EC 3.4.22.68), SARS coronavirus main proteinase (EC 3.4.22.69), sortase A (EC 3.4.22.70), sortase B (EC 3.4.22.71), and cathepsin X (EC 3.4.18.1).

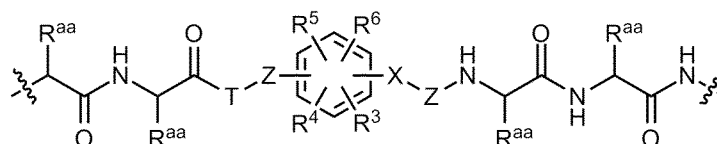
12. The method according to any one of items 1 to 6, wherein the compound of Formula I comprising any one of Formulas VI to VIII:



Formula VI



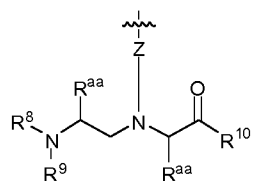
Formula VII



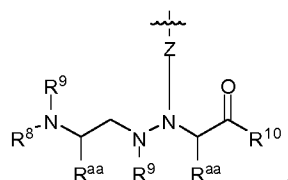
Formula VIII

wherein R^{aa} is individually selected from the group consisting of amino acid side chains.

- 5 13. The method according to any one of the preceding items, wherein $Z-T-R^8$ is



- 10 14. The method according to any one of the preceding items, wherein $Z-T-R^8$ is



- 15 15. The method according to any one of the preceding items, wherein R^{10} is $-\text{OH}$.

16. The method according to any one of the preceding items, wherein R^8 is individually selected from the group consisting of amino acids, peptides and alkylamines, wherein said peptide optionally may be N- and/or C-terminally modified.

- 20 17. The method according to any one of the preceding items, wherein R^8 is a peptide, wherein said peptide optionally is N- and/or C-terminally modified.

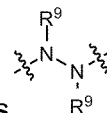
- 25 18. The method according to any one of items 1 to 16, wherein R^8 is an amino acid, wherein said amino acid optionally is N-terminally modified.

19. The method according to any one of items 1 to 16, wherein R^8 is a peptide consisting of in the range of 2 to 5 amino acids, such as of in the range 2 to 3 amino acids, wherein said peptide optionally is N- and/or C-terminally modified.

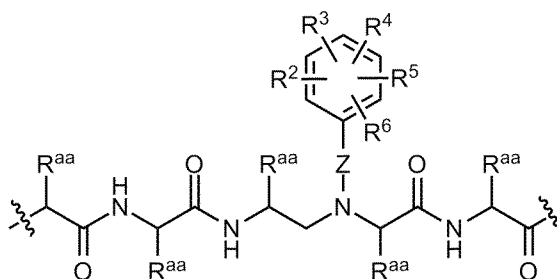
- 30 20. The method according to any one of items 1 to 15, wherein R^8 is individually selected from the group consisting of amino acids, amino acid derivatives,

peptides, peptide derivatives, peptidomimetic, alkylamines, protecting groups, -NH-linker-Y and -linker-Y, wherein said peptide optionally may be N- and/or C-terminally modified.

21. The method according to any one of items 1 to 15, wherein R^8 is a protecting group and the protecting group is selected from the group consisting of Boc, Cbz and Fmoc.



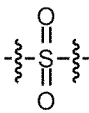
22. The method according to any one of the preceding items, wherein T is and one R^9 is -H and the other R^9 is alkyl-COOH, preferably, -CH₂-COOH.
23. The method according to any one of items 1 to 6, wherein the compound of Formula I is bonded to a nitrogen of the peptide backbone as in Formula X:



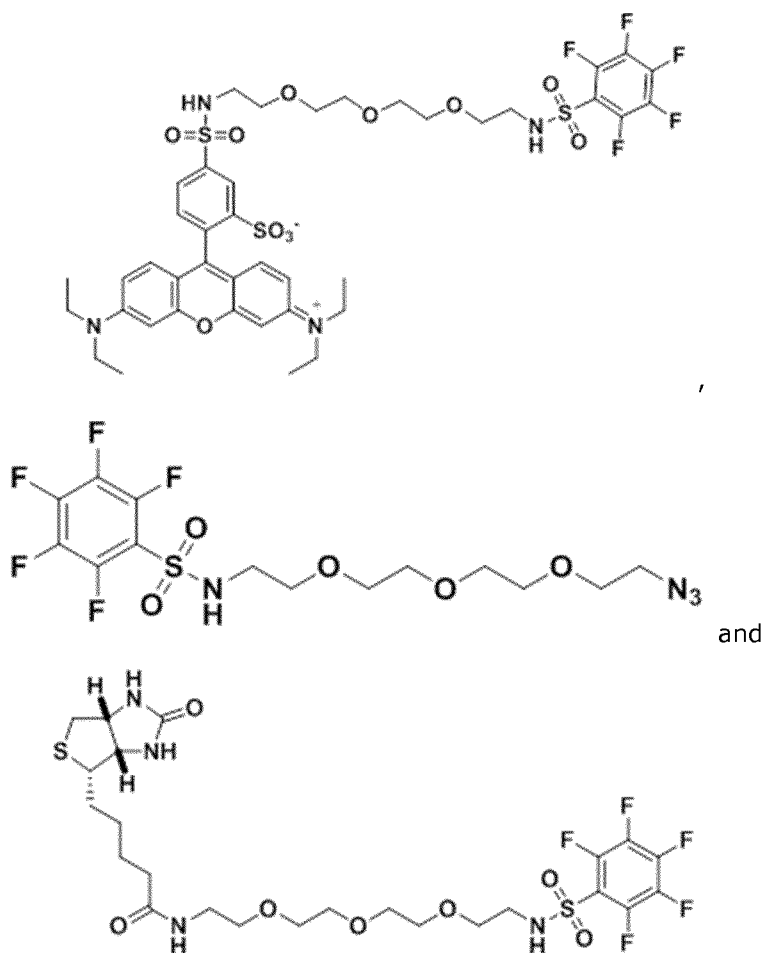
Formula X

- wherein R^{aa} is individually selected from the group consisting of amino acid side chains.
24. The method according to any one of items 12 to 14 or 23, wherein R^{aa} is individually selected from the group consisting of the side chains of the 20 standard proteinogenic amino acid.
25. The method according to any one of the preceding items, wherein R^3 is selected from the group consisting of -F, -Cl, -Br, -I, -NO₂ and SO₂-alkyl.
26. The method according to any one of the preceding items, wherein R^3 is selected from the group consisting of -F and -Cl.
27. The method according to any one of the preceding items, wherein R^3 is -F.

28. The method according to any one of the preceding items, wherein the electron withdrawing group individually are selected from the group consisting of $-F$, $-Cl$, $-CF_3$, $-CCl_3$, SO_2 -alkyl and $-C\equiv N$.
- 5 29. The method according to any one of the preceding items, wherein at least one, such as at least two of R^2 , R^4 , R^5 , and R^6 are $-F$ or $-Cl$.
30. The method according to any one of the preceding items, wherein at least one, such as at least two, but at the most three of R^2 , R^4 , R^5 , and R^6 are $-F$ or $-Cl$.
- 10 31. The method according to any one of the preceding items, wherein all of R^2 , R^4 , R^5 and R^6 are $-H$, $-F$ or $-Cl$, with the proviso that at least one, such as at least two of R^2 , R^4 , R^5 , and R^6 are $-F$ or $-Cl$.
- 15 32. The method according to any one of the preceding items, wherein all of R^2 , R^4 , R^5 and R^6 are $-H$ or $-F$, with the proviso that at least two of R^2 , R^3 , R^4 , R^5 , and R^6 are $-F$.
- 20 33. The method according to any one of the preceding items, wherein at the most three of R^2 , R^3 , R^4 , R^5 , and R^6 are $-Cl$ or $-F$.
34. The method according to any one of the preceding items, wherein all of R^3 , R^4 , R^5 and R^6 are $-F$, and R^2 is $-F$ or R^9 , wherein R^9 is as defined in item 1.
- 25 35. The method according to any one of the preceding items, wherein at least one R^9 is $-H$.
36. The method according to any one of the preceding items, wherein all R^9 is $-H$.
- 30 37. The method according to any one of the preceding items, wherein linker is selected from the group consisting of alkyl wherein one or more $-C(H_2)-$ have been replaced with $-C(O)-$, $-N(H)-$, $-O-$, or $-S-$.
- 35 38. The method according to any one items 1 to 36, wherein linker is selected from the group consisting of peptides, oligosaccharides and steroids.

39. The method according to any one of items 1 to 37, wherein linker is $-N(H)-(CH_2-CH_2-O)_n$, wherein n is an integer from 0 to 10.
40. The method according to any one of items 1 to 37, wherein linker is $-N(H)-(CH_2-CH_2-O)_n-(CH_2)_m-$, wherein n is an integer from 0 to 10 and m is an integer from 0 to 5.
41. The method according to any one of items 1 to 37, wherein linker is $-N(H)-(CH_2-CH_2-O)_n-(CH_2)_m-NH-$, wherein n is an integer from 0 to 10 and m is 2.
42. The method according to any one of the preceding items, wherein Z is , and wherein at the most 4, such as at the most 3 of R^2 , R^3 , R^4 , R^5 , and R^6 are -F.
43. The method according to any one of the preceding items, wherein Y is selected from the group consisting of a fluorescent labelling molecule, a radioisotope labelling molecule, an affinity molecule, an azide, a terminal alkyne, a spin label, a drug molecule, a prodrug, a mass tag, and a photoreactive group.
44. The method according to any one of the preceding items, wherein the labelling molecule Y is biotin, fluorescent labelling molecule or azide.
45. The method according to any one of the preceding claims, wherein the reactive group is $-N_3$.
46. The method according to any one of the preceding items, wherein the polypeptide comprising the cysteine residue in its sequence is an enzyme, e.g. a cysteine protease.
47. The method according to any one of the preceding items, wherein the polypeptide comprising the cysteine residue in its sequence is an enzyme and said cysteine is positioned within the active site of the enzyme, e.g. a cysteine protease.

48. The method according to any one of items 46 to 47, wherein the enzyme is a cysteine protease.
- 5 49. The method according to any one of items 1 to 48, wherein the polypeptide comprising the cysteine residue in its sequence is a human protein.
50. The method according to any one of items 1 to 45, wherein the polypeptide comprising the cysteine residue in its sequence is an albumin.
- 10 51. The method according to item 50, wherein the albumin is human serum albumin.
52. The method according to any one of items 1 to 45, wherein the polypeptide comprising the cysteine residue in its sequence is an antibody or an antigen-binding fragment thereof.
- 15 53. The method according to item 52, wherein the antibody is a naturally occurring antibody.
- 20 54. The method according to item 52, wherein the antibody is a monoclonal antibody.
55. The method according to item 52, wherein the antibody is a humanised antibody or a human antibody.
- 25 56. The method according to item 1, wherein the compound of Formula I is selected from the group consisting of:



and

57. The method according to any one of the preceding items, wherein said method comprises the steps of:

- Providing a mixture of the polypeptide in a solvent;
- Optionally adjusting pH of the mixture;
- Providing the compound of Formula I;
- Reacting said polypeptide in the mixture with said compound of formula I;
- Optionally purifying the product.

58. The method according to item 57, wherein the reacting step is performed in the absence of a catalyst.

15

59. The method according to any one of the preceding items, wherein the method is performed in a solvent, which is a protic solvent mixture.

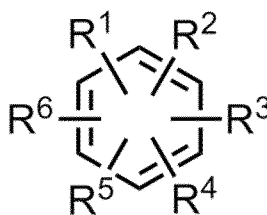
60. The method according to any one of the preceding items, wherein the method is performed in a solvent, which is an aqueous solvent mixture.
- 5 61. The method according to any one of the preceding items, wherein the method is performed in a solvent containing material from living organisms e.g. plasma or tissue.
62. The method according to any one of the preceding items, wherein the method is performed in a solvent compatible with hosting living organisms.
- 10 63. The method according to any one of the preceding items, wherein the reaction of compound of Formula I with a cysteine residue is conducted at the surface of or inside a living organism.
- 15 64. The method according to any one of the preceding items, wherein the method is performed in vitro.
- 20 65. The method according to item 57, wherein the pH is adjusted to in the range of 6 to 9.
- 25 66. A method of modulating the activity of a protein, said method comprising performing the method according to any one of the preceding items, wherein the polypeptide is said protein.
- 30 67. A method of modulating the activity of a protein, said method comprising reacting a compound of Formula I with a cysteine residue of said protein using the method according to any one of the preceding items.
68. The method according to any one of items 66 to 67, wherein the protein is an enzyme.
- 35 69. The method according to any one of items 66 to 68, wherein the polypeptide is an enzyme with a cysteine residue in the active site.

70. The method according to any one of items 68 to 69, wherein the enzyme is a cysteine hydrolase, for example a cysteine protease.
71. The method according to any one of items 66 to 70, wherein the method
5 comprises performing the method according to any one of items 1 to 65.
72. The method according to any one of item 66 to 71, wherein the compound of formula I is capable of binding to the active site of said enzyme, for example said cysteine hydrolase, such as said cysteine protease.
- 10 73. The method according to any one of items 66 to 72, wherein the method is a method of inhibiting the activity of said enzyme.
74. The method according to any one of items 66 to 73, wherein the method
15 comprises performing the method according to any one of items 7 to 10, wherein RS_I - RS_{II} - RS_{III} is a substrate for said cysteine protease.
75. A method for labelling a polypeptide with a labelling molecule, said method comprising performing the method according to any one of items 1 to 65.
- 20 76. The method according to item 75, wherein the labelling is used in a method of isolation of polypeptides.
77. The method according to any one of items 75 to 76, wherein R^1 is as defined in
25 any one of the preceding items.
78. The method according to any one of items 75 to 77, wherein R^8 is -NH-linker-Y, and -linker-Y or -linker- R^{14} .
- 30 79. The method according to any one of items 75 to 78, wherein the label is a labelling molecule Y as defined in any one of items 1 to 45.
80. The method according to any one of items 75 to 78, wherein Y is a prodrug or a drug molecule.
- 35 81. The method according to any one of items 75 to 78, wherein Y is selected from the group consisting of a fluorescent labelling molecule, a radioisotope labelling molecule, an affinity molecule and a photoreactive group.

82. The method according to any one of items 75 to 81, wherein the polypeptide is a protein.
83. The method according to any one of items 75 to 81, wherein the polypeptide is an antibody or an antigen-binding fragment thereof.
84. A method of conjugating a prodrug or a drug molecule to a polypeptide, said method comprising reacting a compound of Formula I wherein Y is a prodrug or a drug molecule with a cysteine residue of said polypeptide using the method according to any one of the preceding items.
85. A method of transportation of a prodrug or a drug molecule, said method comprising performing the method according to any one of the preceding items, wherein Y is a prodrug or a drug molecule.
86. A method for detecting a polypeptide, said method comprising the steps of
- performing the method according to any one of the preceding items, wherein R^8 is $-\text{linker-Y}$ or $-\text{NH-linker-Y}$; and
 - detecting Y.
87. A method for detecting a polypeptide, said method comprising the steps of
- performing the method according to any one of the preceding items, wherein R^8 is $-\text{linker-R}^{14}$;
 - reacting R^{14} with a compound comprising Y; and
 - detecting Y.
88. A method for diagnosis of a clinical condition associated with a polypeptide in an individual at risk of acquiring said clinical condition, said method comprising the steps of:
- performing the method according to any one of the preceding items on a sample from said individual, wherein R^8 is $-\text{linker-Y}$ or $-\text{NH-linker-Y}$, and wherein the polypeptide is associated with said clinical condition; and
 - detecting Y;

thereby determining the presence, absence and/or level of said polypeptide in said sample.

89. A method for diagnosis of a clinical condition associated with a polypeptide in an individual at risk of acquiring said clinical condition, said method comprising the steps of:
- performing the method according to any one of the preceding items on a sample from said individual, wherein R^8 is $-\text{linker}-R^{14}$; and wherein the polypeptide is associated with said clinical condition;
 - reacting R^{14} with a compound comprising Y; and
 - detecting Y;
- thereby determining the presence, absence and/or level of said polypeptide in said sample
90. The method according to any one of items 86 to 88, wherein Y is a labelling molecule.
91. A method for treatment or prevention of a clinical condition associated with a polypeptide in an individual in need thereof, said method comprising performing the method according to any one of the preceding items.
92. The method according to item 91, wherein said clinical condition is associated with the presence of said polypeptide.
93. A compound of Formula I:



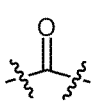
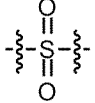
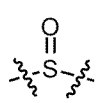
Formula I

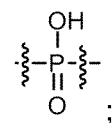
- wherein:
- R^3 is a leaving group
- R^1 is R^7 ;

R^7 is selected from the group consisting of $-X-Z-R^8$, $-Z-T-R^8$, and $-Z-T-(R^8)_2$;

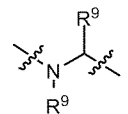
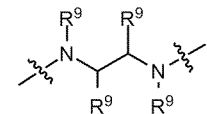
wherein

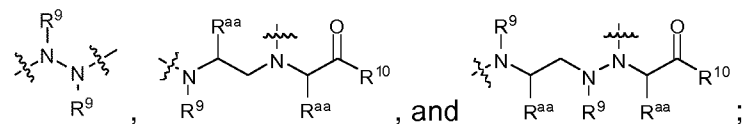
X is selected from the group consisting of a bond, $-\text{CH}_2-\text{N}(\text{R}^9)-$, and $-\text{CH}_2-\text{O}-$;

Z is selected from the group consisting of a bond, , , , and



5

T is selected from the group consisting of a bond, , ,



R^8 is individually selected from the group consisting of amino acids, amino acid derivatives, peptides, peptide derivatives, peptidomimetic, alkylamines, protecting groups, $-\text{NH}-\text{linker}-\text{Y}$ and $-\text{linker}-\text{Y}$, wherein said peptide optionally may be N- and/or C-terminally modified;

10

Y is a labelling molecule, a drug molecule or a prodrug;

R^9 is individually selected from the group consisting of $-\text{H}$, alkyl-COOH and an amino acid side chain;

15

R^{10} is selected from the group consisting of $-\text{H}$, $-\text{OH}$ and R^8 ;

R^2 , R^4 , R^5 , R^6 , are individually selected from the group consisting of $-\text{H}$, R^7 , an amino acid side chain and an electron withdrawing group;

under the proviso that at least one of R^3 , R^4 , R^5 , and R^6 is an electron withdrawing group, and that no more than four of R^2 , R^3 , R^4 , R^5 , R^6 are $-\text{F}$.

20

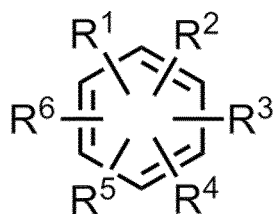
94. The compound according to item 93, wherein the compound of Formula I is as defined in any one of items 1 to 45.

95. The compound according to any one of items 93 to 94, wherein the compound is for use in the treatment or prevention of a clinical condition in an individual in need thereof, wherein said clinical condition is associated with a polypeptide, wherein the polypeptide is a polypeptide comprising a cysteine residue and the compound is capable of forming a covalent bond with said cysteine residue.

25

Claims

1. A method for reacting a compound of Formula I with a cysteine residue, thereby forming a covalent bond, wherein the cysteine is contained in the sequence of a polypeptide, wherein said compound has the following structure:



Formula I

wherein:


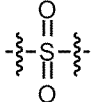

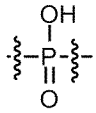
R^3 is a leaving group

R^1 is R^7 ;

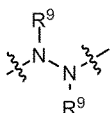
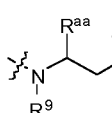
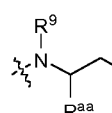
R^7 is selected from the group consisting of $-X-Z-R^8$, $-Z-T-R^8$, and $-Z-T-(R^8)_2$;

wherein

X is selected from the group consisting of a bond, $-\text{CH}_2-\text{N}(\text{R}^9)-$, and $-\text{CH}_2-\text{O}-$;

Z is selected from the group consisting of a bond, , , , and ;

T is selected from the group consisting of a bond, , , ,

, , and ;

R^8 is individually selected from the group consisting of amino acids, amino acid derivatives, peptides, peptide derivatives, peptidomimetic, alkylamines, protecting groups, $-\text{NH-linker-Y}$, $-\text{linker-Y}$ and $-\text{linker-R}^{14}$, wherein said peptide optionally may be N- and/or C-terminally modified;

Y is a labelling molecule, a drug molecule or a prodrug;

R⁹ is individually selected from the group consisting of –H, alkyl-COOH and an amino acid side chain;

R¹⁰ is selected from the group consisting of –OH and R⁸;

5 R¹⁴ is a reactive group, such as an azide or an alkyne;

R², R⁴, R⁵, R⁶, are individually selected from the group consisting of –H, R⁷, an amino acid side chain and an electron withdrawing group;

under the proviso that at least one of R², R⁴, R⁵, and R⁶ is an electron withdrawing group.

10

2. The method according to claim 1, wherein

R¹ is RS_I;

R² is RS_{III};

15

RS_I and RS_{III} individually are selected from the group consisting of amino acids and peptides, wherein said amino acid or peptide optionally may be N- or C-terminally modified;

RS_I when linked to RS_{III} via RS_{II} together forms the peptide RS_I-RS_{II}-RS_{III}, which is a substrate for a peptide cleaving enzyme; and

RS_{II} is selected from the group consisting of a bond, amino acids and peptides.

20

3. The method according to claim 1, wherein

R¹ is RS_I or RS_{III};

25

RS_I and RS_{III} individually are selected from the group consisting of amino acids and peptides, wherein said amino acid or peptide optionally may be N- or C-terminally modified;

RS_I when linked to RS_{III} via RS_{II} together forms the peptide RS_I-RS_{II}-RS_{III}, which is a substrate for a peptide cleaving enzyme; and

RS_{II} is selected from the group consisting of a bond, amino acids and peptides.

30

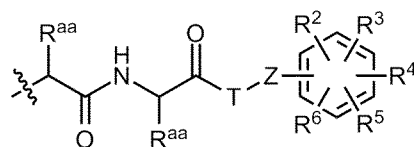
4. The method according to any one of claims 2 to 3, wherein the substrate for a peptide cleaving enzyme is the recognition sequence for a cysteine protease selected from the group consisting of cathepsin B (EC 3.4.22.1), papain (EC 3.4.22.2), ficain (EC 3.4.22.3), chymopapain (EC 3.4.22.6), asclepain (EC 3.4.22.7), clostripain (EC 3.4.22.8), cerevisin (EC 3.4.21.48), streptopain (EC

35

3.4.22.10), insulysin (EC 3.4.24.56), γ -glutamyl hydrolase (EC 3.4.19.9), actinidain (EC 3.4.22.14), cathepsin L (EC 3.4.22.15), cathepsin H (EC 3.4.22.16), prolyl oligopeptidase (EC 3.4.21.26), thimet oligopeptidase (EC 3.4.24.15), proteasome endopeptidase complex (EC 3.4.25.1), saccharolysin (EC 3.4.24.37), kexin (EC 3.4.21.61), Cathepsin T (EC 3.4.22.24), Glycyl endopeptidase (EC 3.4.22.25), Cancer procoagulant (EC 3.4.22.26), cathepsin S (EC 3.4.22.27), picornain 3C (EC 3.4.22.28), picornain 2A (EC 3.4.22.29), Caricain (EC 3.4.22.30), Ananain (EC 3.4.22.31), Stem bromelain (EC 3.4.22.32), Fruit bromelain (EC 3.4.22.33), Legumain (EC 3.4.22.34), Histolysain (EC 3.4.22.35), caspase-1 (EC 3.4.22.36), Gingipain R (EC 3.4.22.37), Cathepsin K (EC 3.4.22.38), adenain (EC 3.4.22.39), bleomycin hydrolase (EC 3.4.22.40), cathepsin F (EC 3.4.22.41), cathepsin O (EC 3.4.22.42), cathepsin V (EC 3.4.22.43), nuclear-inclusion-a endopeptidase (EC 3.4.22.44), helper-component proteinase (EC 3.4.22.45), L-peptidase (EC 3.4.22.46), gingipain K (EC 3.4.22.47), staphopain (EC 3.4.22.48), separase (EC 3.4.22.49), V-cath endopeptidase (EC 3.4.22.50), cruzipain (EC 3.4.22.51), calpain-1 (EC 3.4.22.52), calpain-2 (EC 3.4.22.53), calpain-3 (EC 3.4.22.54), caspase-2 (EC 3.4.22.55), caspase-3 (EC 3.4.22.56), caspase-4 (EC 3.4.22.57), caspase-5 (EC 3.4.22.58), caspase-6 (EC 3.4.22.59), caspase-7 (EC 3.4.22.60), caspase-8 (EC 3.4.22.61), caspase-9 (EC 3.4.22.62), caspase-10 (EC 3.4.22.63), caspase-11 (EC 3.4.22.64), peptidase 1 (mite) (EC 3.4.22.65), calicivirin (EC 3.4.22.66), zingipain (EC 3.4.22.67), Ulp1 peptidase (EC 3.4.22.68), SARS coronavirus main proteinase (EC 3.4.22.69), sortase A (EC 3.4.22.70), sortase B (EC 3.4.22.71), and cathepsin X (EC 3.4.18.1).

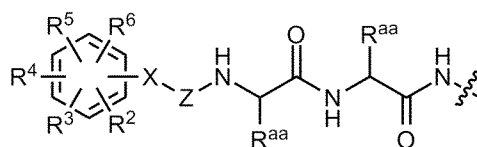
25

5. The method according to claim 1, wherein the compound of Formula I comprising any one of Formulas VI to VIII:

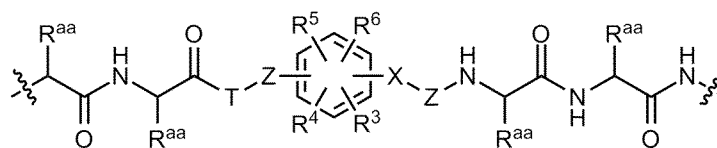


Formula VI

30



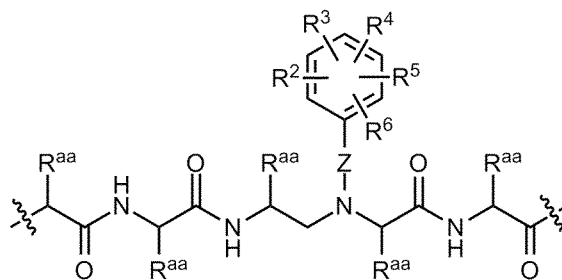
Formula VII



Formula VIII

wherein R^{aa} is individually selected from the group consisting of amino acid side chains.

6. The method according to claim 1, wherein the compound of Formula I is bonded to a nitrogen of the peptide backbone as in Formula X:



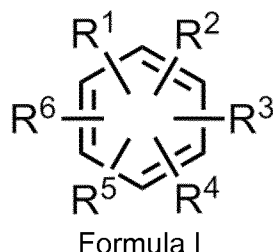
Formula X

7. The method according to any one of the preceding claims, wherein the polypeptide comprising the cysteine residue in its sequence is an enzyme and said cysteine is positioned within the active site of the enzyme, e.g. a cysteine protease.
8. The method according to any one of the preceding claims, wherein said method comprises the steps of:
- Providing a mixture of the polypeptide in a solvent;
 - Optionally adjusting pH of the mixture;
 - Providing the compound of Formula I;

- Reacting said polypeptide in the mixture with said compound of formula I;
- Optionally purifying the product.

- 5 9. A method of modulating the activity of a protein, said method comprising reacting a compound of Formula I with a cysteine residue of said protein using the method according to any one of the preceding claims.
- 10 10. A method of conjugating a prodrug or a drug molecule to a polypeptide, said method comprising reacting a compound of Formula I wherein Y is a prodrug or a drug molecule with a cysteine residue of said polypeptide using the method according to any one of claims 1 to 8.
- 15 11. A method for labelling a polypeptide with a labelling molecule, said method comprising performing the method according to any one of claims 1 to 8.
- 20 12. A method for detecting a polypeptide, said method comprises the steps of
- a. performing the method according to any one of claim 1 to 8, wherein R⁸ is –linker-Y or –NH-linker-Y; and
 - b. detecting Y.
- 25 13. A method for diagnosis of a clinical condition associated with a polypeptide in an individual at risk of acquiring said clinical condition, said method comprising the steps of:
- a. performing the method according to any one of claims 1 to 8 on a sample from said individual, wherein R⁸ is –linker-Y or –NH-linker-Y, and
 - 30 wherein the polypeptide is associated with said clinical condition; and
 - b. detecting Y;
- thereby determining the presence, absence and/or level of said polypeptide in said sample.
- 35 14. A method for treatment or prevention of a clinical condition associated with a polypeptide in an individual in need thereof, said method comprising performing the method according to any one of claims 1 to 8.

15. A compound of Formula I:



5 wherein:

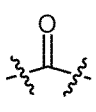
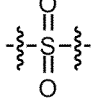
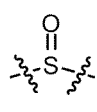
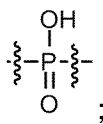
R^3 is a leaving group

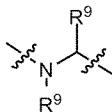
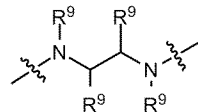
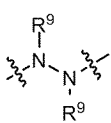
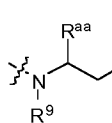
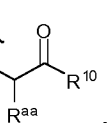
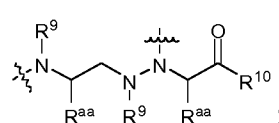
R^1 is R^7 ;

R^7 is selected from the group consisting of $-X-Z-R^8$, $-Z-T-R^8$, and $-Z-T-(R^8)_2$;

wherein

10 X is selected from the group consisting of a bond, $-\text{CH}_2-\text{N}(\text{R}^9)-$, and $-\text{CH}_2-\text{O}-$;

Z is selected from the group consisting of a bond, , , , and ;

T is selected from the group consisting of a bond, , , , , , and ;

15 R^8 is individually selected from the group consisting of amino acids, amino acid derivatives, peptides, peptide derivatives, peptidomimetic, alkylamines, protecting groups, $-\text{NH}-\text{linker}-\text{Y}$ and $-\text{linker}-\text{Y}$, wherein said peptide optionally may be N- and/or C-terminally modified;

Y is a labelling molecule, a drug molecule or a prodrug;

20 R^9 is individually selected from the group consisting of $-\text{H}$, alkyl-COOH and an amino acid side chain;

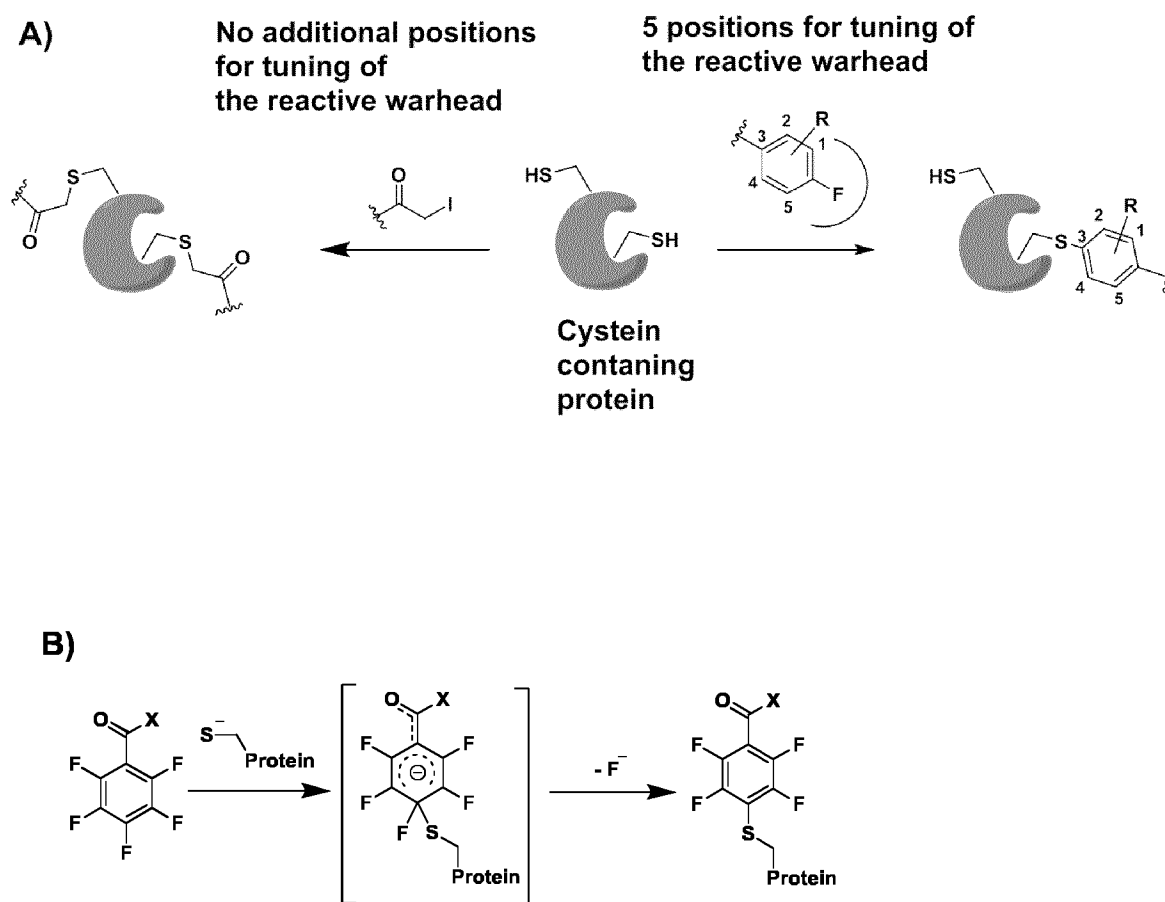
R^{10} is selected from the group consisting of $-\text{OH}$ and R^8 ;

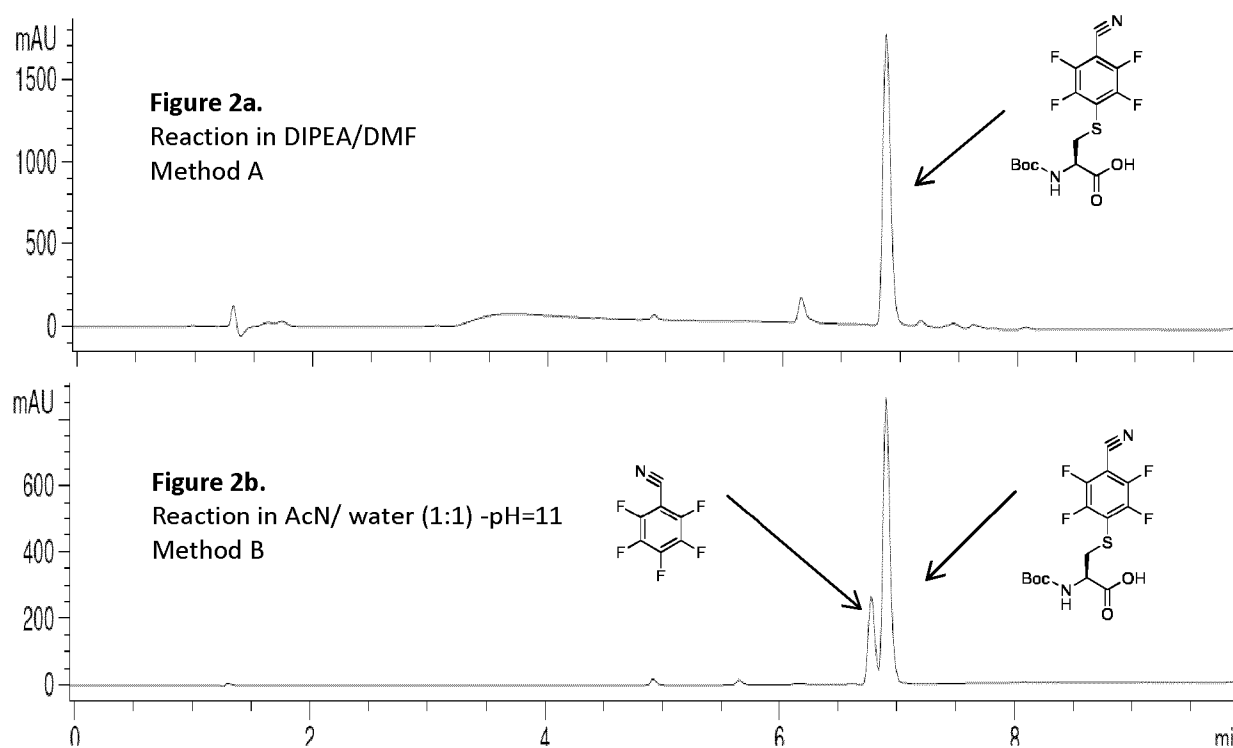
R^2 , R^4 , R^5 , R^6 , are individually selected from the group consisting of $-\text{H}$, R^7 , an amino acid side chain and an electron withdrawing group;

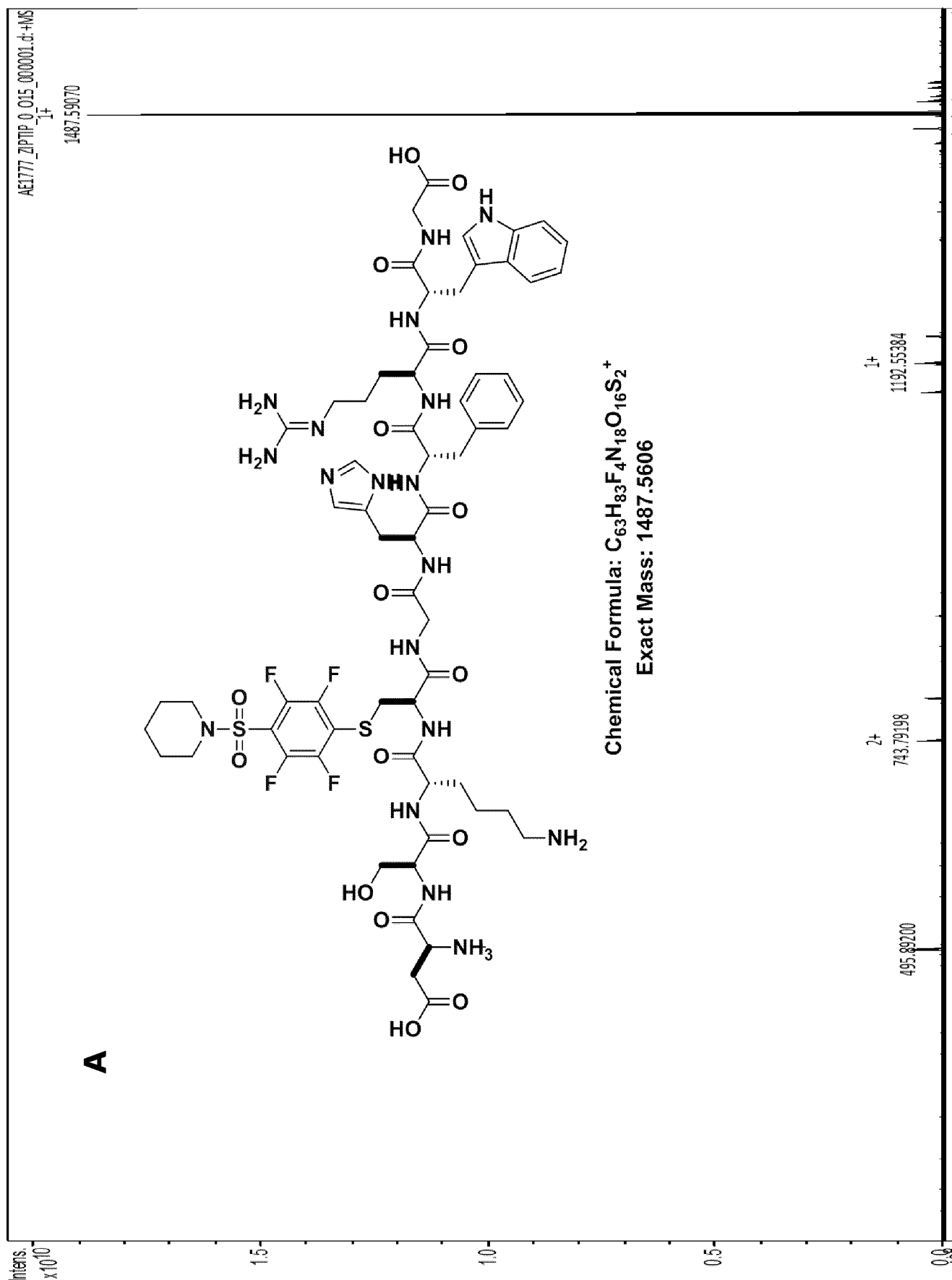
under the proviso that at least one of R^3 , R^4 , R^5 , and R^6 is an electron withdrawing group, and that no more than four of R^2 , R^3 , R^4 , R^5 , R^6 are -F.

5

10

**Figure 1**

**Figure 2**

**Figure 3**

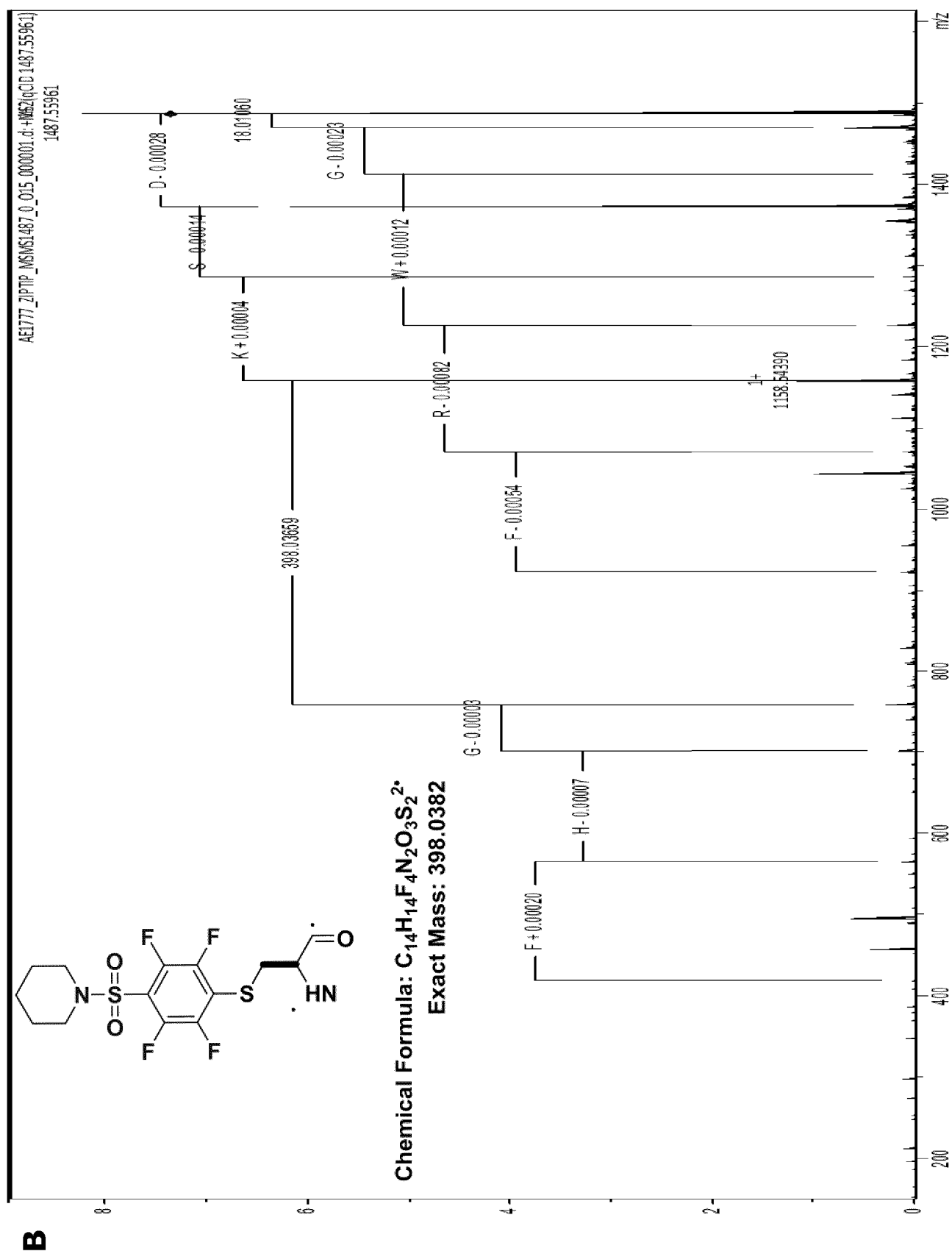
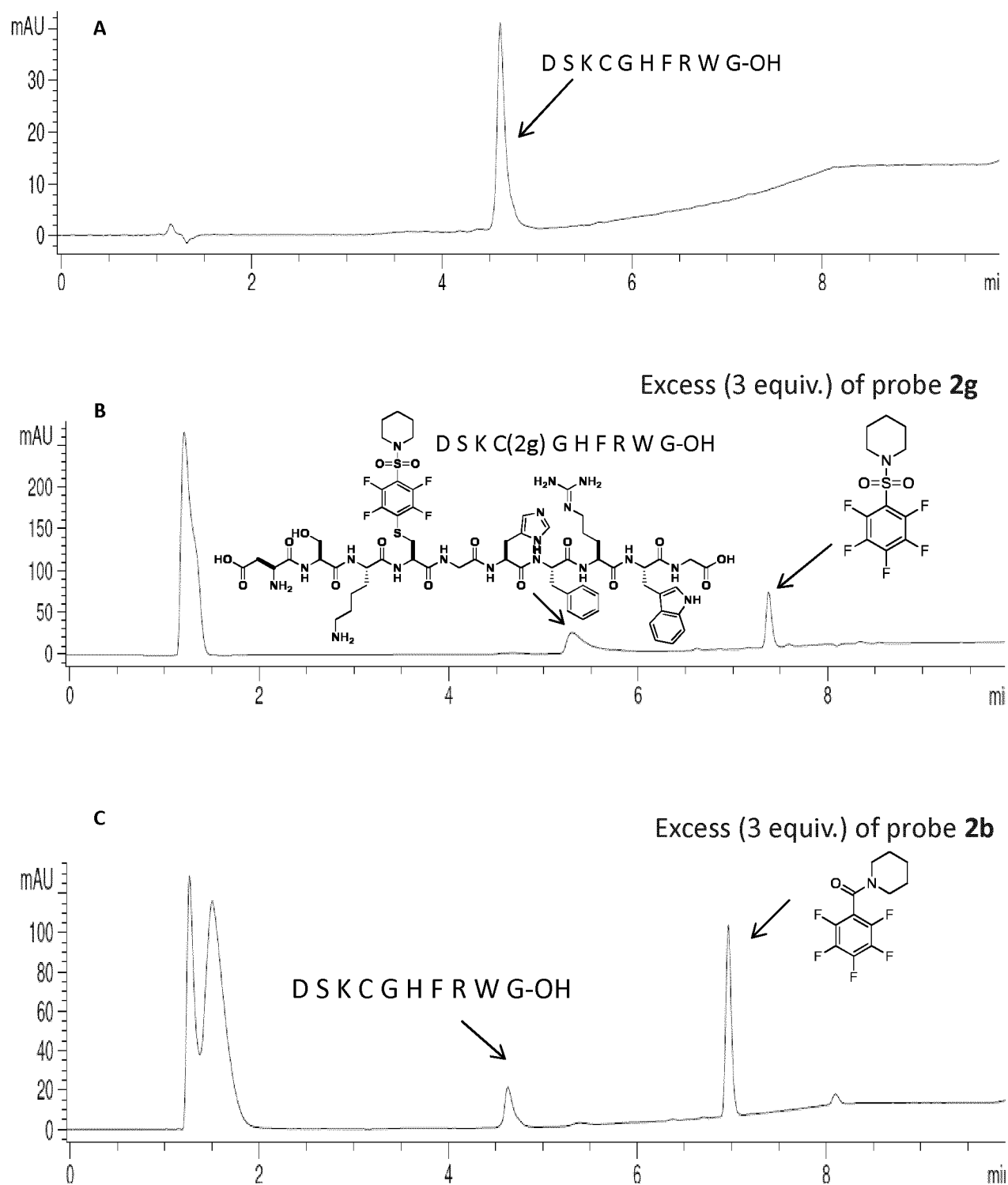


Figure 3 cont.

**Figure 4**

A

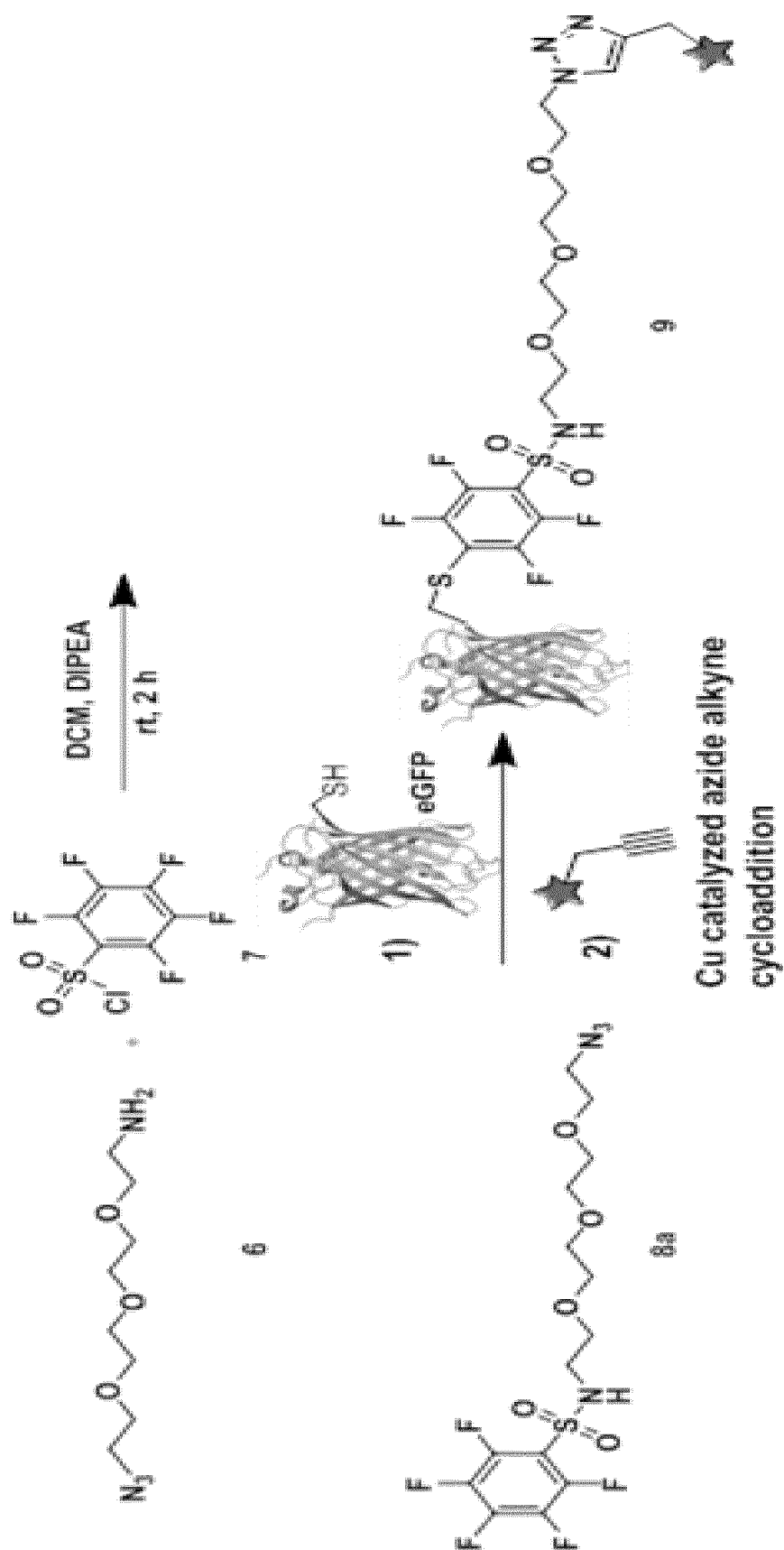
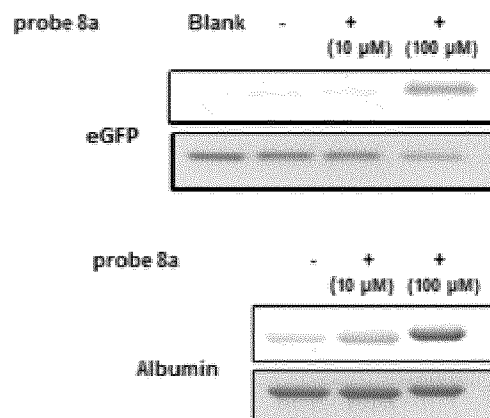
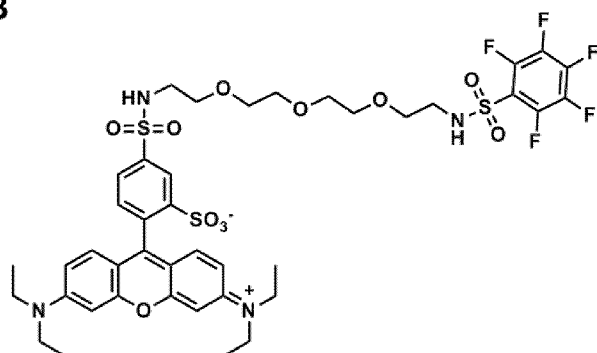
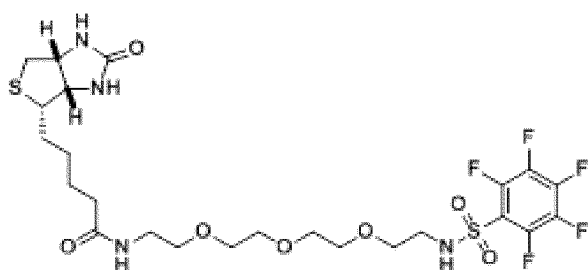
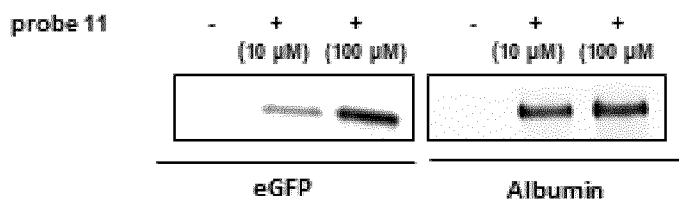
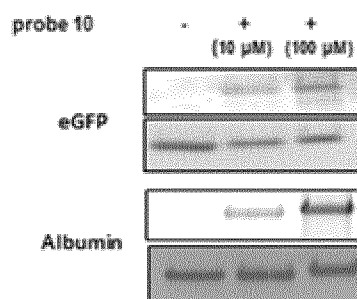
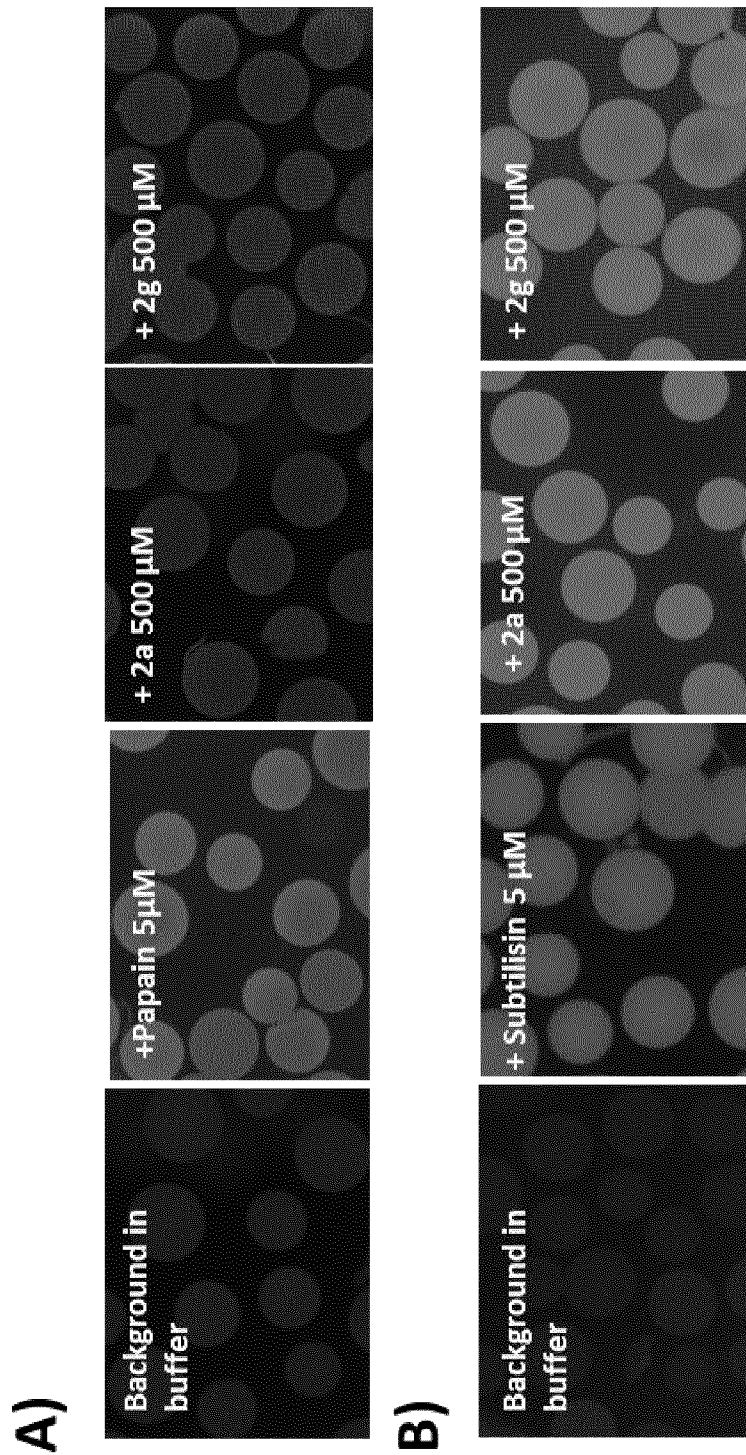


Figure 5

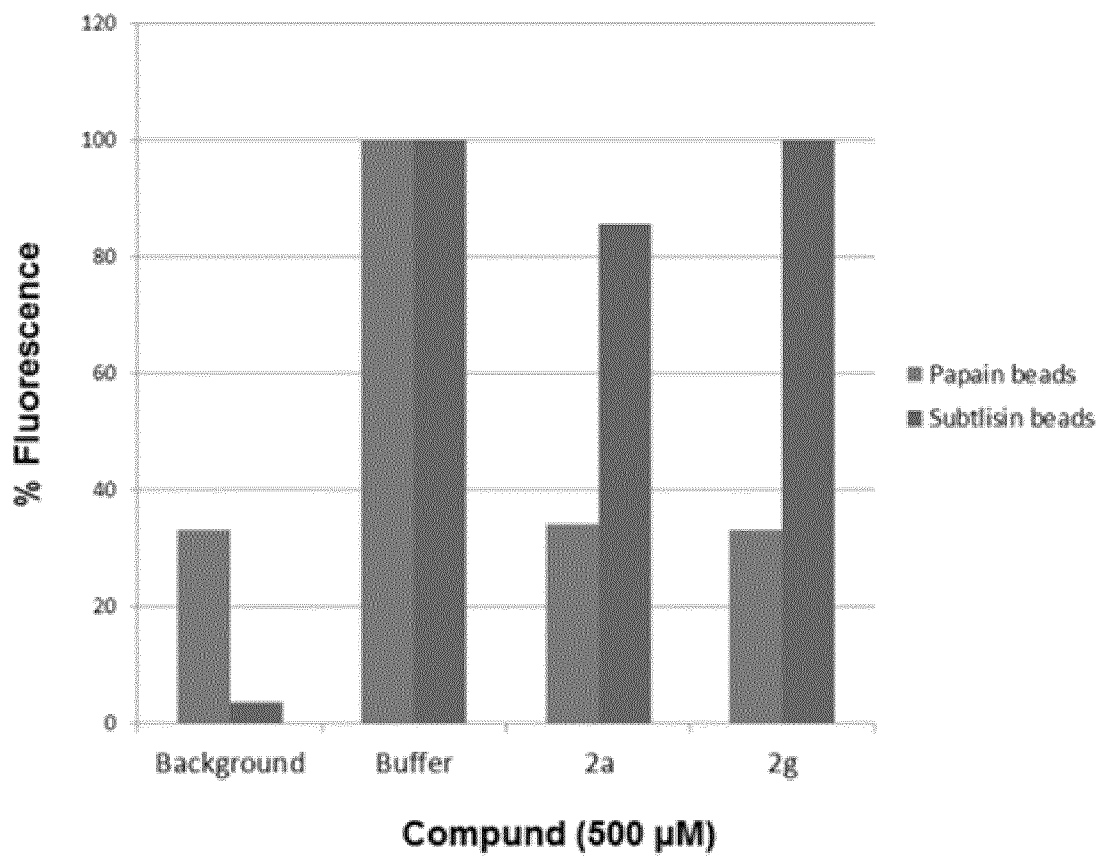
B**C**

11

**Figure 5
cont.**

**Figure 6**

C)

**Figure 6 cont.**

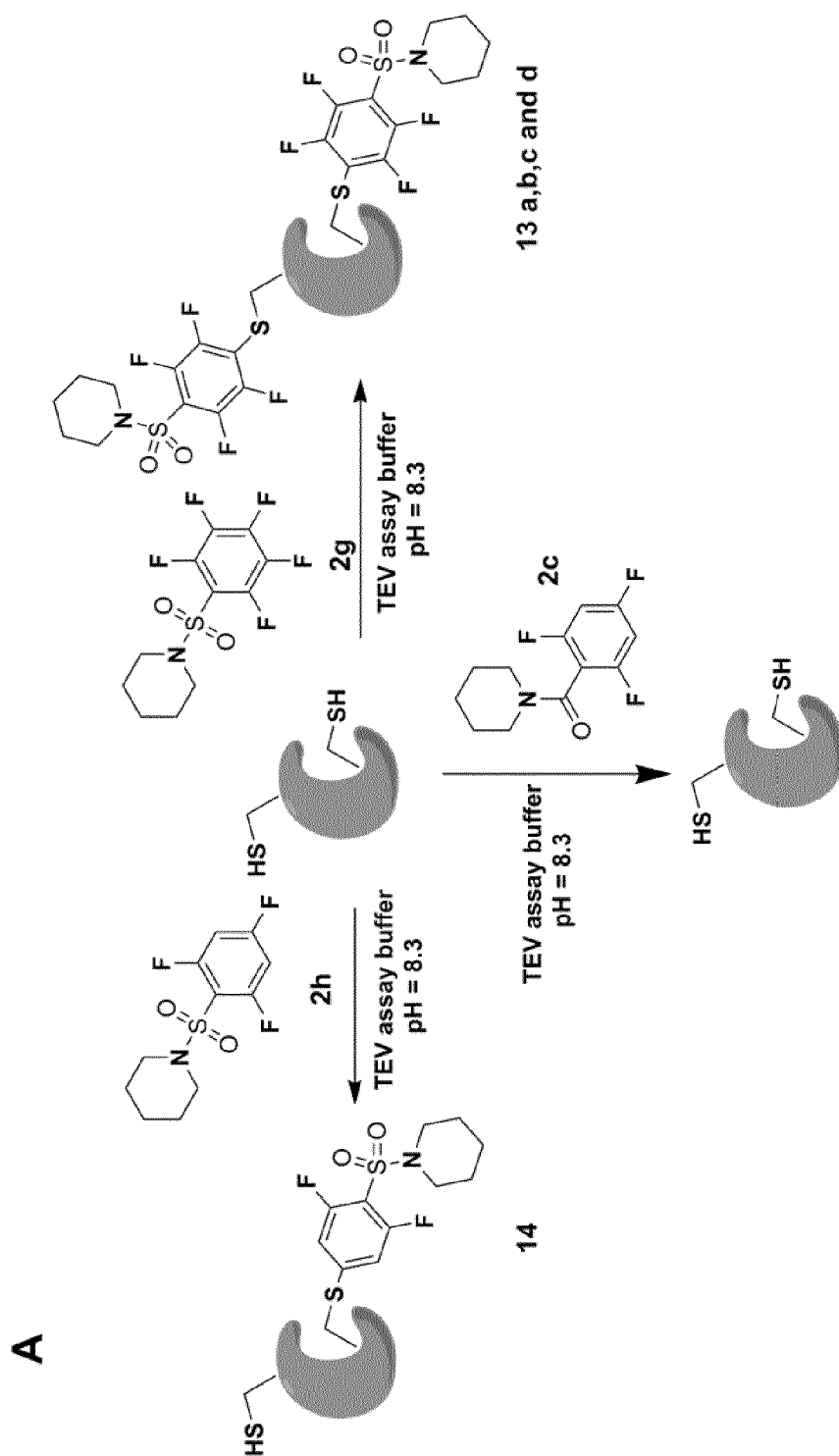
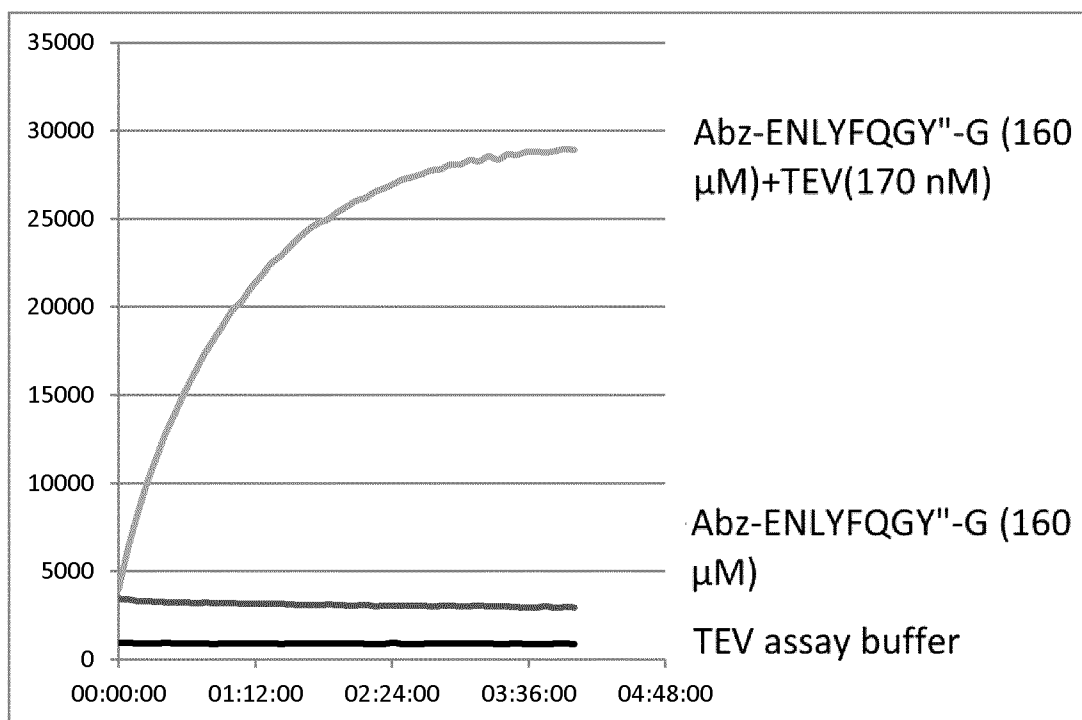
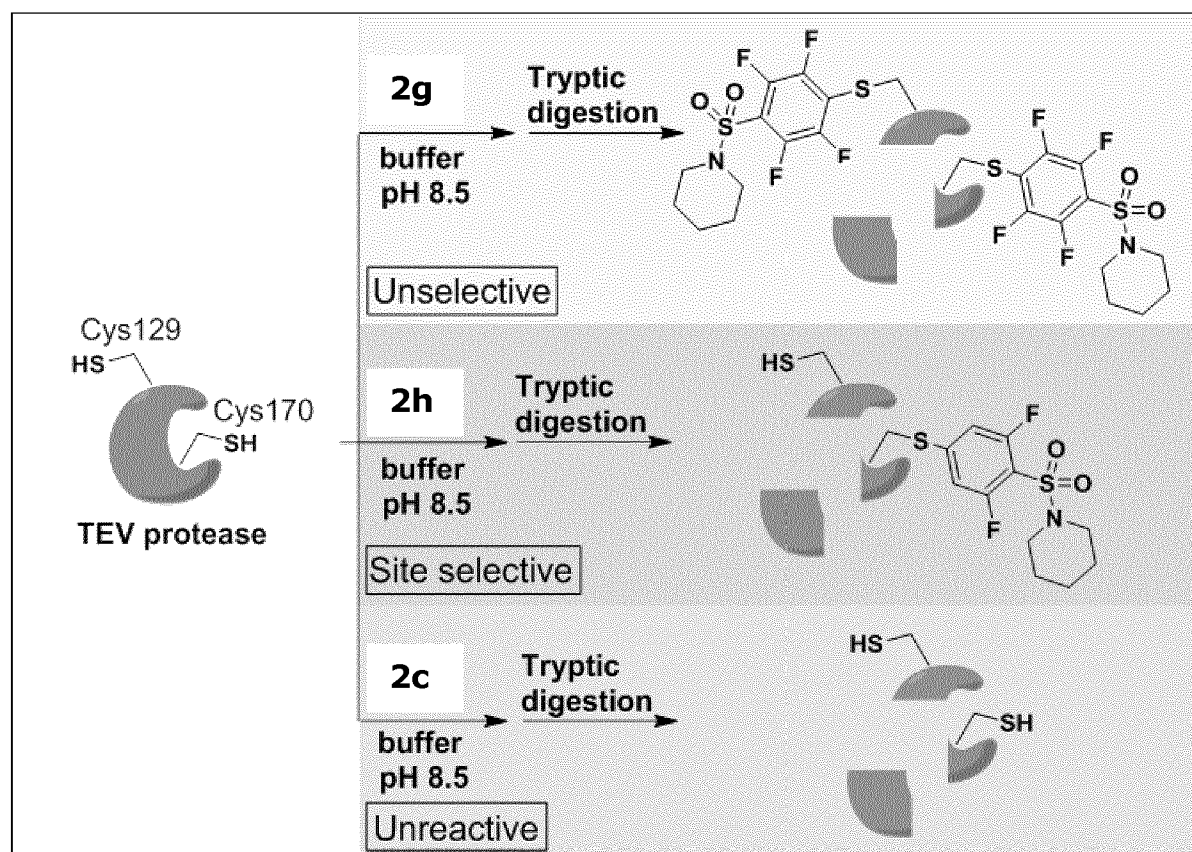


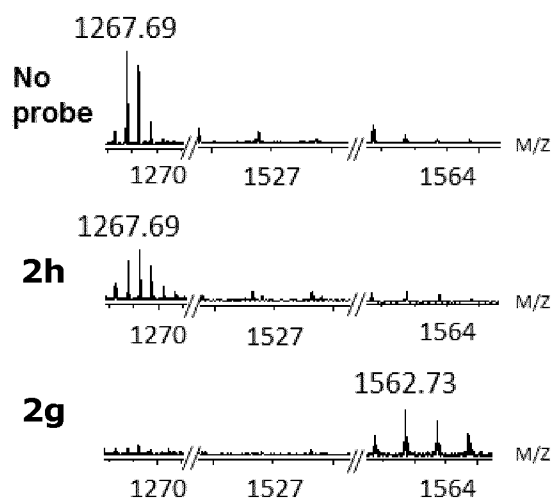
Figure 7

B**Figure 7 cont.**

C



Fragment with surface cysteine (Cys129)



Fragment with catalytic cysteine (Cys170)

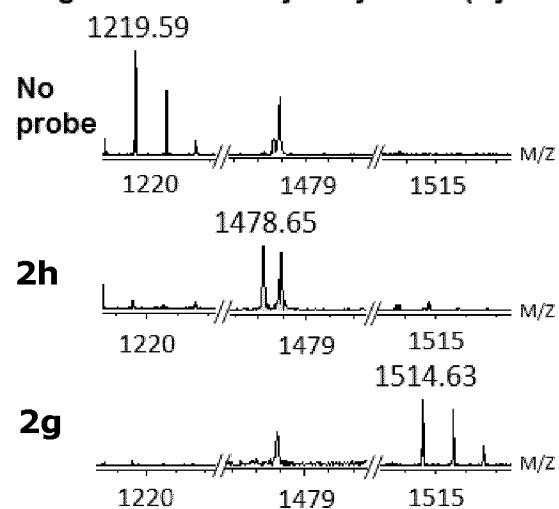
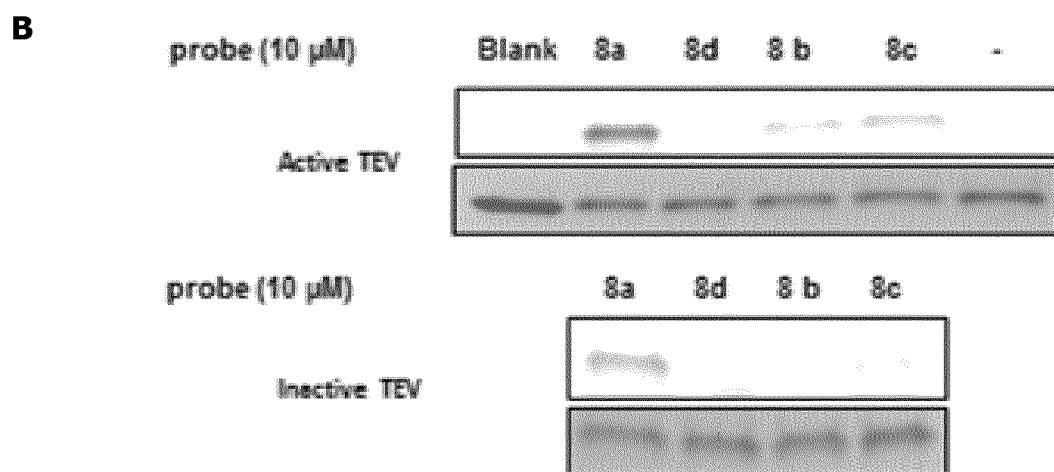
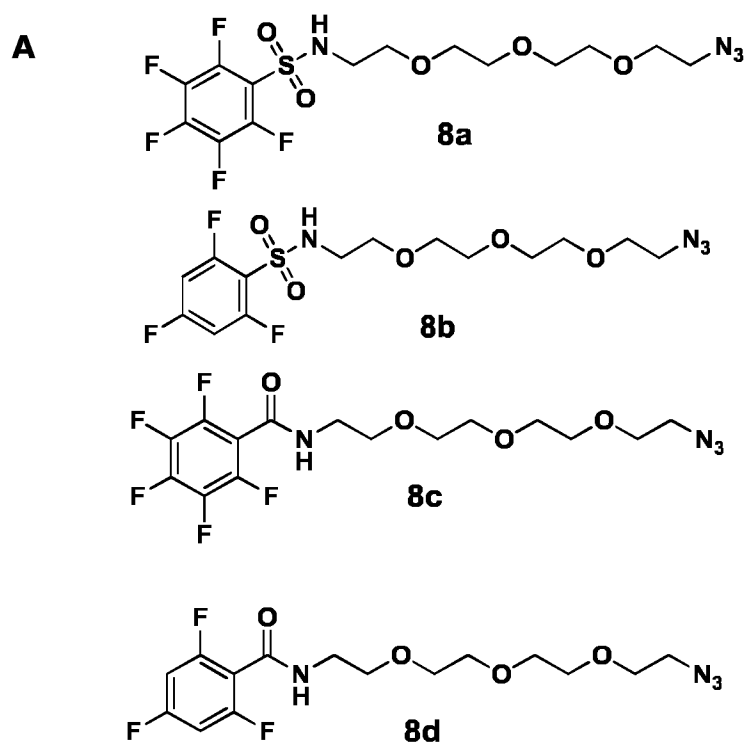
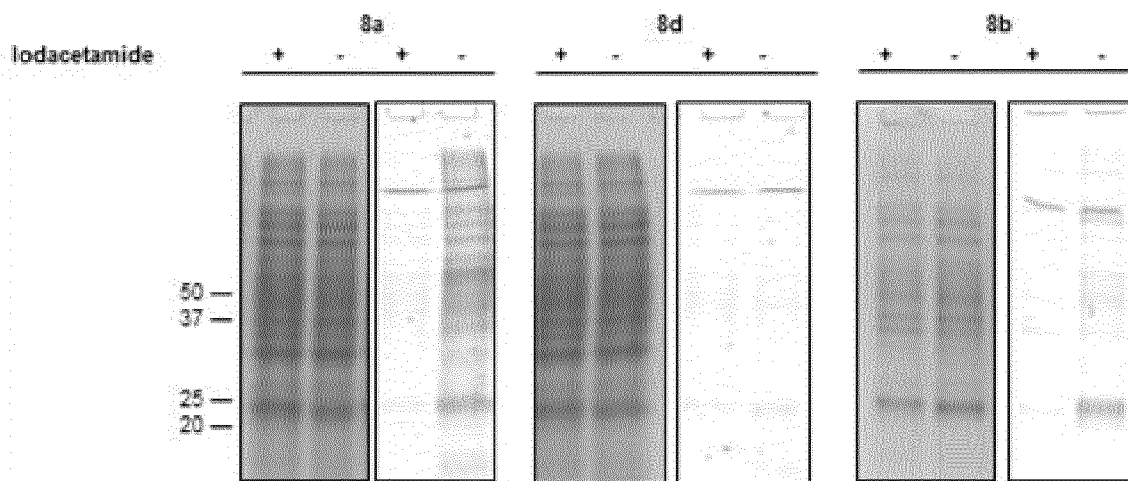
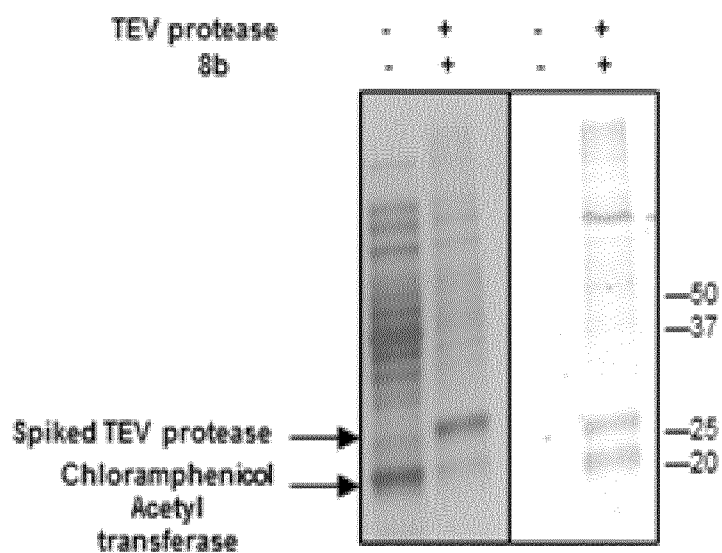
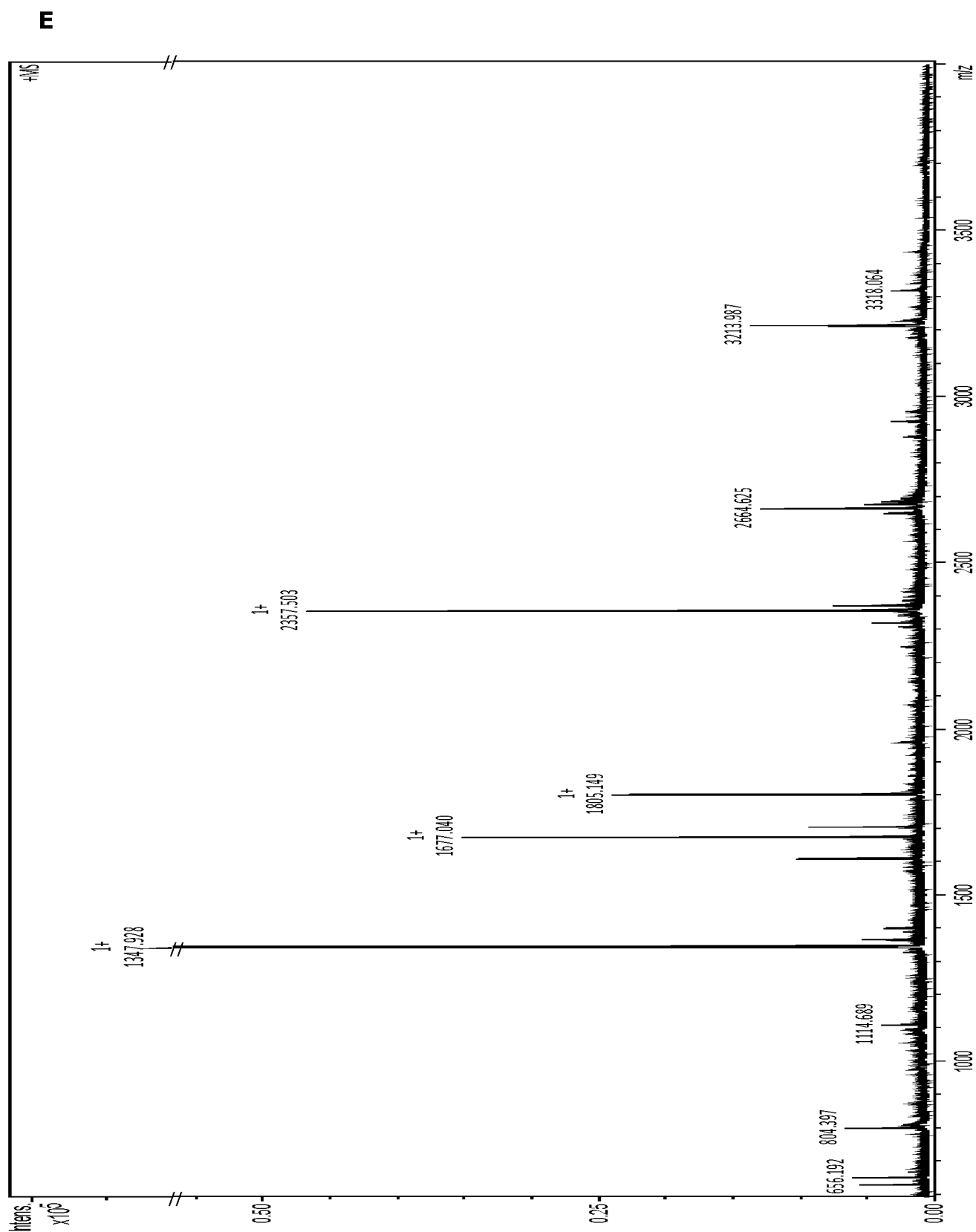
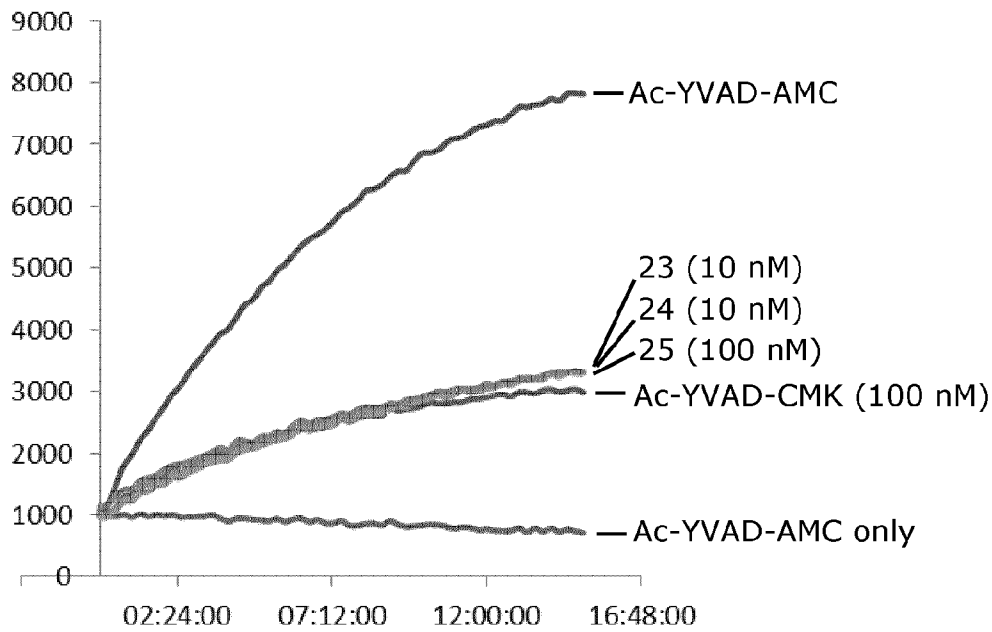
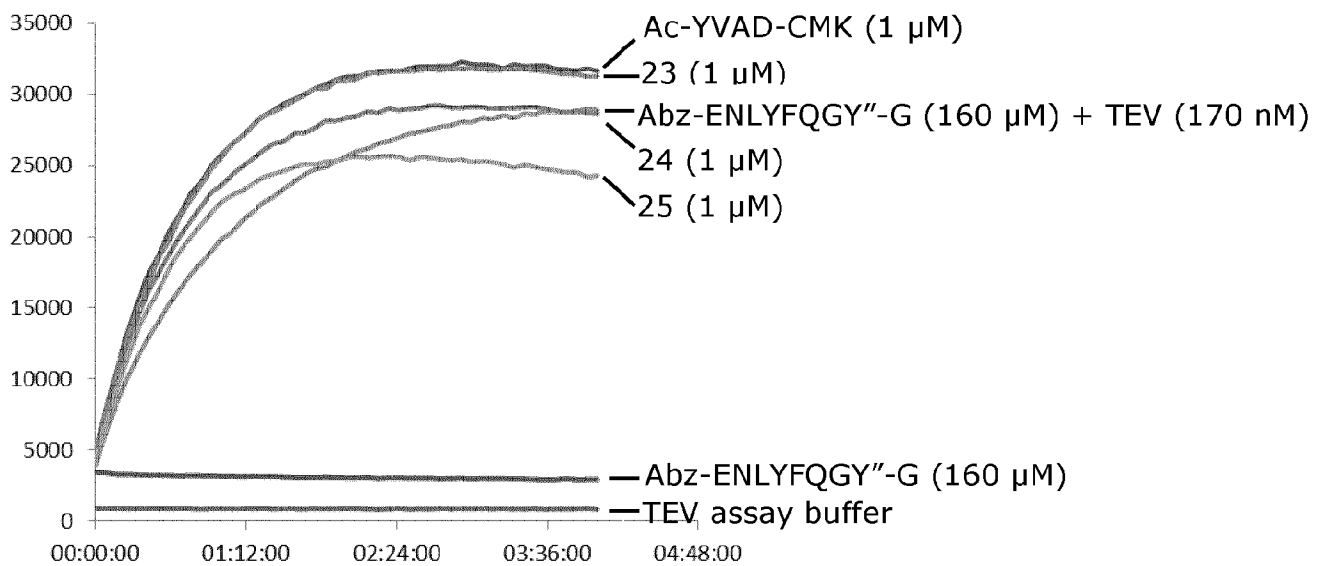


Figure 7 cont.

**Figure 8**

C**D****Figure 8 cont.**

**Figure 8 cont.**

A**B****Figure 9**

A

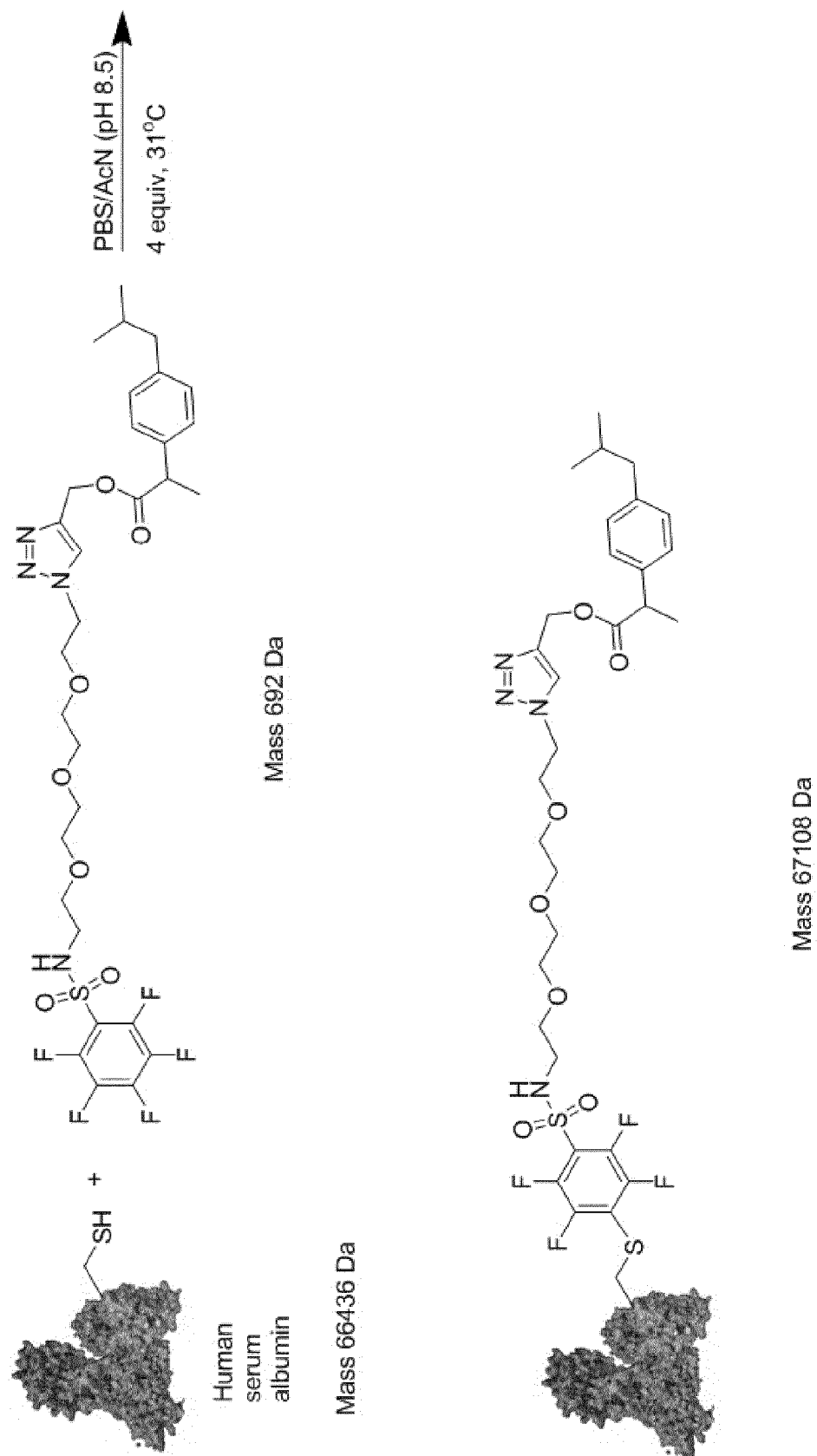
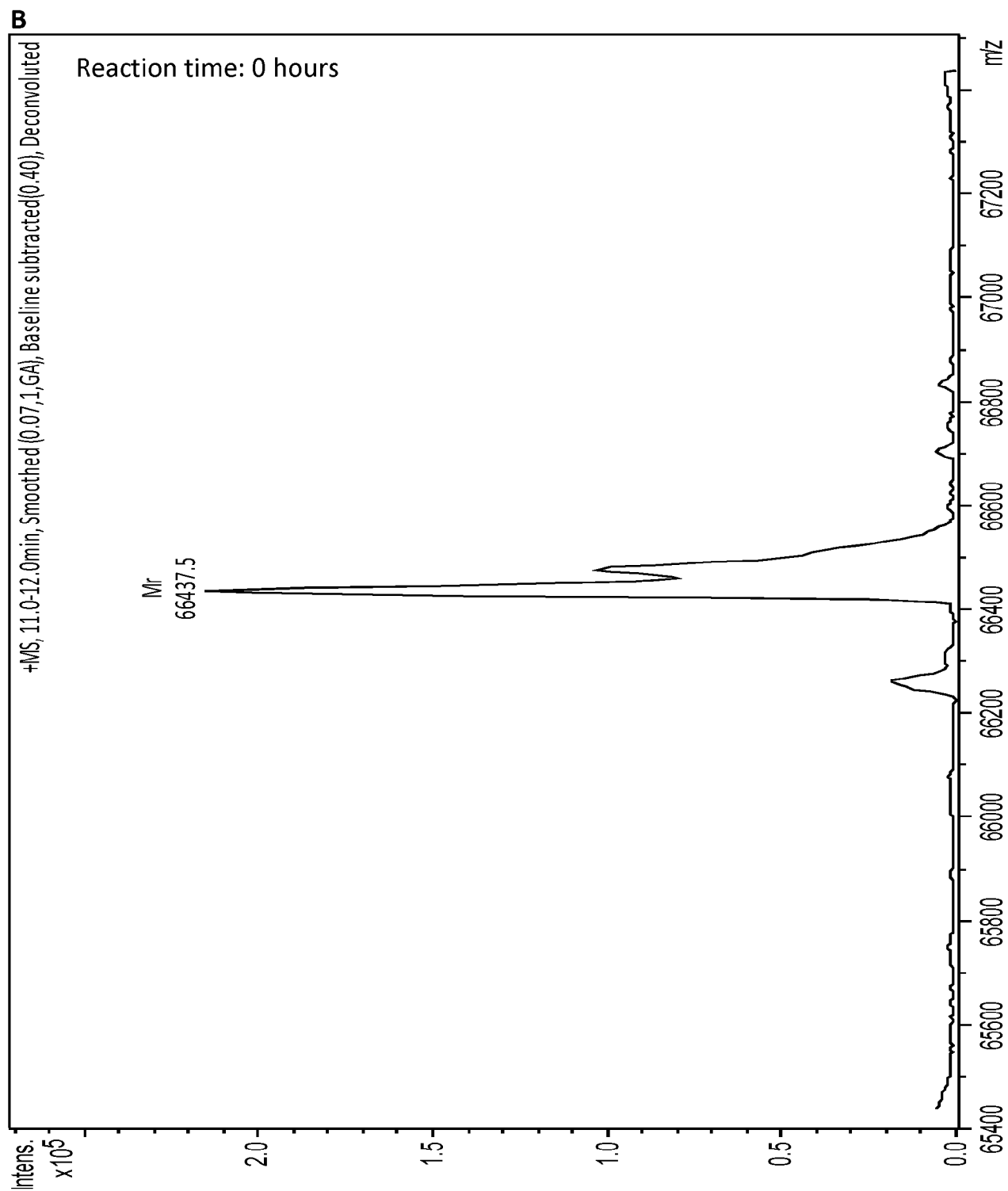
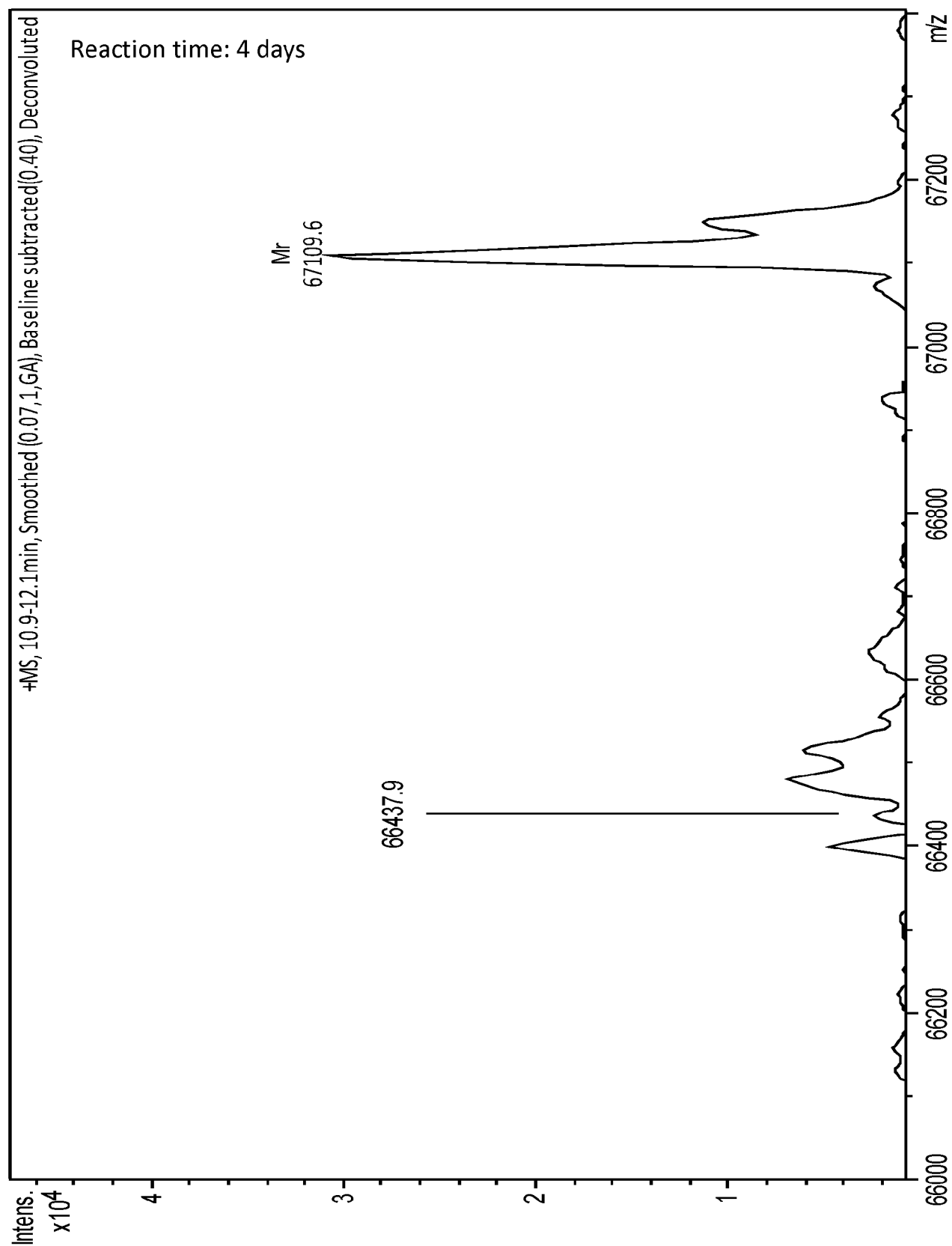


Figure 10

**Figure 10 cont.**

B cont.**Figure 10 cont.**

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2018/066633

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/37 C12N9/64 C07K14/81
ADD. C07C15/04

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07C C12Q C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, EMBASE, FSTA, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	D. ALEXANDER SHANNON ET AL: "Investigating the Proteome Reactivity and Selectivity of Aryl Halides", JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 136, no. 9, 24 February 2014 (2014-02-24), pages 3330-3333, XP055494151, US ISSN: 0002-7863, DOI: 10.1021/ja4116204 p.3331 col.1 par.1, 2, col.2 par.1, Fig.2 p.3331 col.1 par.1, p.3333 col.1 par.1 -----	1-13,15
X	WO 2006/114190 A1 (MERCK PATENT GMBH [DE]; LARDY CLAUDE [FR]; LECOMTE MARC [FR]; CONVARD) 2 November 2006 (2006-11-02) p.6 l.16- p.7 l.22 p.23 l.24-32 ----- -/-	1-9,14, 15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

23 July 2018

Date of mailing of the international search report

10/08/2018

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Bonello, Steve

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2018/066633

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>STEFAN G. KATHMAN ET AL: "A Fragment-Based Method to Discover Irreversible Covalent Inhibitors of Cysteine Proteases", JOURNAL OF MEDICINAL CHEMISTRY, vol. 57, no. 11, 28 May 2014 (2014-05-28), pages 4969-4974, XP055494263, ISSN: 0022-2623, DOI: 10.1021/jm500345q p.4969 col.2 par.1 . p.4970 col.1 par.2, Fig.1, p.4970 col.2 par.2</p> <p>-----</p>	1-15
A	<p>ALEXANDER M. SPOKOYNY ET AL: "A Perfluoroaryl-Cysteine S N Ar Chemistry Approach to Unprotected Peptide Stapling", JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 135, no. 16, 24 April 2013 (2013-04-24), pages 5946-5949, XP055237282, US ISSN: 0002-7863, DOI: 10.1021/ja400119t p.5946 col.1 par.2, col.2 par.1, Fig.1 p.5948 col.2 par.3-p.5949 col.1</p> <p>-----</p>	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2018/066633

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2006114190	A1	02-11-2006	
		AR 053366 A1	02-05-2007
		AT 502916 T	15-04-2011
		AU 2006239557 A1	02-11-2006
		CA 2605748 A1	02-11-2006
		EP 1881960 A1	30-01-2008
		ES 2362111 T3	28-06-2011
		FR 2884820 A1	27-10-2006
		JP 5210153 B2	12-06-2013
		JP 2008538774 A	06-11-2008
		US 2008182906 A1	31-07-2008
		US 2013090385 A1	11-04-2013
		WO 2006114190 A1	02-11-2006
