

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization

International Bureau

(43) International Publication Date  
10 April 2025 (10.04.2025)



(10) International Publication Number  
**WO 2025/074091 A1**

(51) International Patent Classification:

C07K 14/315 (2006.01) C07K 7/00 (2006.01)

(21) International Application Number:

PCT/GB2024/052526

(22) International Filing Date:

02 October 2024 (02.10.2024)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

2315107.9 02 October 2023 (02.10.2023) GB

(71) Applicants: **OXFORD UNIVERSITY INNOVATION LIMITED** [GB/GB]; Buxton Court, 3 West Way, Oxford Oxfordshire OX2 0JB (GB). **KING'S COLLEGE LONDON** [GB/GB]; Strand, London Greater London WC2R 2LS (GB).

(72) Inventors: **READING, Eamonn**; c/o King's College London, Strand, London Greater London WC2R 2LS (GB). **HAMMERSCHMID, Dietmar**; c/o King's College London, Strand, London Greater London WC2R 2LS (GB). **HOWARTH, Mark**; Department of Biochemistry, University of Oxford, South Parks Rd, Oxford Oxfordshire OX1 3QU (GB). **KEEBLE, Anthony**; Department of Biochemistry, University of Oxford, South Parks Rd, Oxford Oxfordshire OX1 3QU (GB).

(74) Agent: **WILKINS, Christopher**; Dehns, 10 Old Bailey, London Greater London EC4M 7NG (GB).

(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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV,

GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: POLYPEPTIDES AND PEPTIDE TAGS CAPABLE OF SPONTANEOUS ISOPEPTIDE BOND FORMATION AT LOW TEMPERATURE AND LOW PH AND USES THEREOF

(57) Abstract: The present invention relates to a two-part linker comprising a peptide (peptide tag) and a polypeptide (protein) that is capable of spontaneously forming an isopeptide bond, particularly in low temperature and low pH conditions. Recombinant or synthetic polypeptides comprising the polypeptide or peptide tag, nucleic acid molecules encoding the polypeptide, peptide tag or recombinant polypeptides, vectors comprising said nucleic acid molecules, and host cells comprising said vectors and nucleic acid molecules are provided herein. A kit and apparatus comprising said polypeptide and peptide tag and various methods and uses of the polypeptide and peptide tag are also provided.



WO 2025/074091 A1

Polypeptides and peptide tags capable of spontaneous isopeptide bond formation  
at low temperature and low pH and uses thereof

FIELD OF THE INVENTION

5           The present invention relates to a two-part linker comprising a peptide  
(peptide tag) and a polypeptide (protein) that is capable of spontaneously forming  
an isopeptide bond. In particular, the two-part linker is able to spontaneously form  
an isopeptide bond efficiently in low temperature and low pH conditions, such as  
quench conditions used in protein analysis methods, e.g. Hydrogen-Deuterium  
10       eXchange mass spectrometry (HDX-MS). In particular, the polypeptide and peptide  
may be viewed as a peptide tag and polypeptide binding partner cognate pair that  
can be conjugated via a covalent bond when contacted under conditions that allow  
the spontaneous formation of an isopeptide bond between the peptide tag and its  
polypeptide binding partner.

15           Also provided herein are recombinant or synthetic polypeptides comprising  
the polypeptide or peptide tag, nucleic acid molecules encoding the polypeptide,  
peptide tag or recombinant polypeptides, vectors comprising said nucleic acid  
molecules, and host cells comprising said vectors and nucleic acid molecules. A kit  
and apparatus comprising said polypeptide and peptide tag and various methods  
20       and uses of the polypeptide and peptide tag are also provided, e.g. use in protein  
analysis methods, such as HDX-MS.

BACKGROUND TO THE INVENTION

25           Hydrogen-Deuterium eXchange mass spectrometry (HDX-MS) is a valuable  
technique used to investigate the dynamics of proteins, typically in isolated systems  
although investigations within cells have been conducted. HDX is a chemical  
reaction which relies on the replacement of covalently bound hydrogen atoms (such  
as those found in amide groups or polar side chains) with deuterium atoms derived  
from the reaction solvent, or *vice versa*, without the requirement of a catalyst. The  
30       exchange of hydrogen to deuterium occurs at different rates depending on the  
structure of the protein of interest (POI), as tightly hydrogen-bonded sections of the  
protein will undergo a very slow hydrogen to deuterium exchange, whilst exchange  
will happen rapidly in more disordered sections. By monitoring this solvent  
exchange over time, HDX-MS provides valuable insight into the structure and  
35       dynamics of peptides and proteins.

Although the field of HDX-MS has made tremendous progress since its inception, especially in automation and instrumentation, there are very few examples of HDX-MS successfully being used to study protein dynamics in cells and related complex environments (Lin *et al.*, 2022 *Protein Science* 31, e4402; 5 Kaldmäe *et al.*, 2020 *FEBS J* 287, 2823–2833; Fang *et al.*, 2023 *Anal Chem* 95, 1805–1810). These pioneering studies revealed that cellular HDX was possible and could provide valuable information on protein conformation and interactions, e.g. factors that are known to be influenced by environmental factors but that are difficult to recapitulate *in vitro*, such as local macromolecular crowding, ionic strength, and 10 pH. However, these studies lacked the ability to examine proteins selectively.

Selective protein enrichment post-HDX would allow for the selective examination of proteins, but would need to take place under the essential acidic and cold quench conditions required to preserve the Deuterium-label (D-Label; pH 2.3-3.0 and 0-4 °C). Such conditions provide a highly challenging environment for 15 protein enrichment as it is well-known that such extreme conditions disrupt common interaction chemistries and most antibodies undergo conformational changes at acidic pH, leading to ligand dissociation. For example, the enrichment technique immobilised metal-affinity chromatography (IMAC) is only active over a pH range between 6 and 9. Similarly, barnase/barstar is inactive below pH 4. Indeed, 20 conventional phage display experiments utilise the disruption of protein-protein interactions at pH 2 to elute the phage from the bait proteins.

Streptavidin-biotin is a well-known and exploited method of protein enrichment/purification and has been shown to be effective for biotinylated protein depletion post-HDX within a purified *in vitro* system. However, for *in situ* HDX-MS 25 utilisation of this method would require the introduction of an additional biotinylating step of an AviTagged construct by a biotin ligase. This additional step would further require characterisation for each construct as this reaction may often not reach completion, even in strains in which biotin ligase (BirA) is overexpressed. Additionally, biotin-binding partners, streptavidin or avidin (and their derivatives), 30 rely on maintenance of non-covalent interactions which can be easily disrupted and are therefore not good fusion partners.

Systems for capturing proteins via covalent bonds have been developed. For instance, proteins that are capable of spontaneous isopeptide bond formation (so-called “isopeptide proteins”) have been advantageously used to develop 35 peptide tag/polypeptide binding partner pairs (i.e. two-part linkers) which covalently

bind to each other and provide irreversible interactions (see e.g. WO2011/098772, WO 2016/193746, WO 2018/197854, WO2020/183198 and WO2022/214795 all herein incorporated by reference). In this respect, proteins which are capable of spontaneous isopeptide bond formation may be expressed as separate fragments, to give a peptide tag and a polypeptide binding partner for the peptide tag, where the two fragments are capable of covalently reconstituting by isopeptide bond formation, thereby linking molecules or components fused to the peptide tag and its polypeptide binding partner.

Isopeptide bonds are amide bonds formed between carboxyl/carboxamide and amino groups, where at least one of the carboxyl or amino groups is outside of the protein main-chain (the backbone of the protein). Such bonds are chemically irreversible under typical biological conditions and they are resistant to most proteases. Since isopeptide bonds are covalent in nature, they result in the some of the strongest measured protein interactions. The isopeptide bond formed by a peptide tag and its polypeptide binding partner is stable under conditions where non-covalent interactions would rapidly dissociate, e.g. over long periods of time (e.g. weeks), at high temperature (to at least 95 °C), at high force, or with harsh chemical treatment (e.g. pH 2-11, organic solvent, detergents or denaturants).

A peptide tag/binding partner pair, termed SpyTag/SpyCatcher, was derived from the CnaB2 domain of the *Streptococcus pyogenes* FbaB protein (Zakeri *et al.*, 2012, Proc Natl Acad Sci U S A 109, E690-697) and has been used in diverse applications, including biomaterials (Botyanszki *et al.*, 2015, Biotechnology and bioengineering 112, 2016-2024; Chen *et al.*, 2014, Proc Natl Acad Sci U S A 108, 11399-11404), next generation sequencing (Stranges *et al.*, 2016, Proc Natl Acad Sci U S A 113, E6749-E6756), enzyme stabilisation (Schoene *et al.*, 2016, Scientific Reports 6, 21151) and vaccine development (Brune *et al.*, 2016, Scientific Reports 6, 19234; Thrane *et al.*, 2016, Journal of Nanobiotechnology 14, 30). Furthermore, peptide tag/binding partner pairs with improved reaction rates, termed SpyTag002/SpyCatcher002 and SpyTag003/SpyCatcher003, have also been described (see e.g. WO2018/197854 and WO2020/183198). However, as the parent *Streptococcus pyogenes* FbaB protein has evolved to react at 37 °C and neutral pH, the existing peptide tag/binding partner pairs based on this protein show peak reactivity at physiological temperatures and around neutral pH, e.g. about pH 5-6. Thus, these existing peptide tag/binding partner pairs are not suitable for

- 4 -

workflows which are required to operate at the extreme pH and low temperatures required for many proteomic analysis methods, such as HDX-MS.

Accordingly, the need for an effective protein enrichment system that functions effectively in conditions typical for proteomic analysis methods, such as HDX-MS, remains unmet.

#### SUMMARY OF THE INVENTION

In work leading to the present invention, the inventors have developed a protein enrichment system that is able to function efficiently under cold and acidic conditions, i.e. such as those using HDX-MS. On the basis that isopeptide bonds are robust and stable under a wide range of pH and temperatures, the inventors identified the SpyTag/SpyCatcher system as a potential starting point for the development of a protein enrichment system, i.e. a covalent capture system. As noted above, several generations of the system have been developed. While each new generation of tag and catcher is capable of reacting with the previous version (e.g. SpyCatcher003 reacts with SpyTag and SpyTag002, in addition to SpyTag003), the latest generation, i.e. the SpyTag003/SpyCatcher003 pair, has been shown to have the highest reaction rate; more than two orders of magnitude greater than the original SpyTag/SpyCatcher pair. Moreover, the SpyTag003/SpyCatcher003 pair functions well at low concentrations. As such, it was unexpected that when combinations of SpyTag and SpyCatcher polypeptides were tested for activity in the acidic and cold conditions required to limit D-to-H back-exchange in HDX-MS methods (e.g. pH 2.5 and 0 °C), the combination of SpyTag002 (SEQ ID NO: 26) and SpyCatcher003 (SEQ ID NO: 27) was found to be the pairing with the best functionality.

Moreover, as discussed in the Examples, using this pairing as a starting point, the inventors surprisingly determined that a single modification (i.e. mutation) of the amino acid sequence of SpyCatcher003 (R32 to a non-basic amino acid) significantly improved its reaction rate with SpyTag002 in the extreme conditions of an HDX quench (acidic pH and low temperature). Furthermore, the inventors determined that the reaction rate of the pair in these conditions could be improved further by introducing additional modifications to the SpyCatcher003 mutant (SEQ ID NO: 5) and by modifying the SpyTag002 sequence. Notably, and surprisingly, it was found that although the modifications improve activity in acidic and cold conditions, they did not impair the activity in other conditions. However, due to the

reactivity of the modified catcher/tag pair in lemon juice, which has a pH of ~2-3, this new system was termed LemonTag/LemonCatcher.

In addition to demonstrating that the LemonTag/LemonCatcher system can capture a POI under HDX quench conditions, the inventors advantageously further developed the covalent capture system such that it can be integrated into an HDX-MS workflow. HDX-MS analysis typically relies on short digestion, peptide trapping and chromatography to limit D-to-H back exchange (<15-30 minutes) which, in the case of a complex sample background, increases peptide co-elution and consequently hampers the identification and analysis of the target protein/peptides. The introduction of a unique cysteine on the opposite side of the polypeptide to the reactive centre facilitates surface coupling of the polypeptide via disulfide bridge formation. As shown in the Examples, it was established that a POI, fused with LemonTag (SEQ ID NO: 4), could be pulled-down by LemonCatcher (SEQ ID NO: 1) coupled to CarboxyLink resin (diamino-dipropylamine activated, crosslinked beaded agarose) by a Sulfo-LC-SPDP linker. Whilst the bead-captured POI can be digested in solution using acid proteases, e.g. pepsin, it was determined that elution of the entire complex (LemonCatcher:LemonTag-POI) can be achieved with a reducing agent, e.g. tris(2-carboxyethyl)phosphine (TCEP), a common quench additive tolerated at high concentrations by pepsin (up to 100-200 mM). Release of the POI permits its digestion by immobilised protease systems commonly used to improve digestion efficiency, reproducibility, throughput, and reduce background from autolytic products that are prevalent with typical solution protease digests.

While the system has been exemplified using conditions typically used in HDX-MS methods, and will facilitate the study of proteins in increasingly complex environments via covalent capture HDX-MS, it will be understood that the polypeptide and peptide pair provided herein may find utility in wide range of applications. The advantageous properties of the peptide and polypeptide (i.e. LemonTag/LemonCatcher and variants thereof) provided herein facilitate their use in many settings. As the polypeptide and peptide are both relatively small, they do not materially increase sample complexity and the peptide can be cloned easily into the protein of interest (POI). Moreover, the polypeptide provided herein is readily expressed in protein expression systems and purified in high yields (10-20 mg per L of culture). As noted above, the polypeptide provided herein can be cross-linked to amine-derivatised beads via a unique cysteine residue and subsequently eluted. Thus, the polypeptide and peptide provided herein will find utility in the study of

protein dynamics in a complex cellular environment and in large protein assembly machineries, which would be too complex to tackle with existing technology (e.g. ribosome, proteasome). Furthermore, the polypeptide and peptide may find use in other settings that require a highly acidic pH. For instance, highly acidic conditions  
5 are commonly used to inhibit the activity of many proteases or other modifying enzymes and are relevant to analysis of gastric function.

Furthermore, in view of the fact that the SpyCatcher polypeptide can be truncated at its N-terminus and C-terminus without substantially affecting its activity (Li et al., 2014, J Mol Biol.; 426(2): 309–317), it is contemplated that the polypeptide  
10 exemplified herein (e.g. SEQ ID NO: 1-3 or 5-13) may be truncated at the N-terminus and/or at the C-terminus without substantially reducing the activity of the polypeptide. In particular, SEQ ID NOs: 1-3 or 5-13 may be truncated by up to 21 amino acids at the N-terminus (e.g. 5, 10, 15 or 20 amino acids) and by up to 9 amino acids at the C-terminus (e.g. 1, 2, 3, 4 or 5, 6, 7, 8 or 9 amino acids),  
15 preferably by up to 5 amino acids. In preferred embodiments, the polypeptide exemplified herein (e.g. SEQ ID NOs: 1-3 or 5-13) is truncated only at the N-terminus (e.g. SEQ ID NOs: 14-25). As shown in the Examples, none of the mutations that provide the advantageous properties of the polypeptide provided herein are located in the regions that may be removed via truncation. Alternatively  
20 viewed, the truncated polypeptide (i.e. portion of the exemplified polypeptide) contains all of the residues that may be modified to provide the advantageous properties of the polypeptide provided herein.

Thus, in one aspect, the present invention provides a polypeptide comprising:

- 25 (a) an amino acid sequence with at least 80% sequence identity to a sequence as set forth in SEQ ID NO: 2, wherein the amino acid sequence comprises lysine at position 31, glutamic acid at position 77 and wherein:
- (i) X at position 32 is not a basic amino acid;
  - (ii) X at position 49 is any amino acid, preferably a polar amino acid or  
30 cysteine;
  - (iii) X at position 56 is a hydrophobic amino acid or a polar amino acid;
  - (iv) X at position 59 is a polar amino acid;
  - (v) X at position 94 is a hydrophobic amino acid; and
  - (vi) X at position 95 is a polar amino acid,

- 7 -

and wherein the polypeptide is capable of spontaneously forming an isopeptide bond with a peptide comprising an amino acid sequence as set forth in SEQ ID NO: 4, wherein the isopeptide bond forms between the aspartic acid residue at position 8 of SEQ ID NO: 4 and the lysine residue at position 31 of the polypeptide; or

(b) an amino acid sequence with at least 80% sequence identity to a sequence as set forth in SEQ ID NO: 15, wherein the amino acid sequence comprises lysine at position 10, glutamic acid at position 56 and wherein:

(i) X at position 11 is not a basic amino acid;

(ii) X at position 28 is any amino acid, preferably a polar amino acid or cysteine;

(iii) X at position 35 is a hydrophobic amino acid or a polar amino acid;

(iv) X at position 38 is a polar amino acid;

(v) X at position 73 is a hydrophobic amino acid; and

(vi) X at position 74 is a polar amino acid,

and wherein the polypeptide is capable of spontaneously forming an isopeptide bond with a peptide comprising an amino acid sequence as set forth in SEQ ID NO: 4, wherein the isopeptide bond forms between the aspartic acid residue at position 8 of SEQ ID NO: 4 and the lysine residue at position 10 of the polypeptide.

Alternatively viewed, provided herein is a polypeptide comprising:

(a) an amino acid sequence with at least 80% sequence identity to a sequence as set forth in SEQ ID NO: 3, wherein X at position 32 is not a basic amino acid, and the amino acid sequence comprises lysine at position 31, glutamic acid at position 77 and one or more of the following:

(i) alanine at position 56;

(ii) threonine at position 59;

(iii) isoleucine at position 94; and

(iv) threonine at position 95,

wherein the specified amino acid residues are at positions equivalent to the positions in SEQ ID NO: 3

and wherein the polypeptide is capable of spontaneously forming an isopeptide bond with a peptide comprising an amino acid sequence as set forth in SEQ ID NO: 4, wherein the isopeptide bond forms between the aspartic acid

- 8 -

residue at position 8 of SEQ ID NO: 4 and the lysine residue at position 31 of the polypeptide; or

(b) an amino acid sequence with at least 80% sequence identity to a sequence as set forth in SEQ ID NO: 16, wherein X at position 11 is not a basic amino acid, and the amino acid sequence comprises lysine at position 10, glutamic acid at position 56 and one or more of the following:

- (i) alanine at position 35;
- (ii) threonine at position 38;
- (iii) isoleucine at position 73; and
- (iv) threonine at position 74,

wherein the specified amino acid residues are at positions equivalent to the positions in SEQ ID NO: 16

and wherein the polypeptide is capable of spontaneously forming an isopeptide bond with a peptide comprising an amino acid sequence as set forth in SEQ ID NO: 4, wherein the isopeptide bond forms between the aspartic acid residue at position 8 of SEQ ID NO: 4 and the lysine residue at position 10 of the polypeptide.

It will be understood that the polypeptide defined in part (b) of the statements above may be viewed as a portion of the polypeptide defined in part (a), e.g. a portion of the polypeptide defined in (a) comprising an amino acid sequence with at least 80% sequence identity to a sequence as set forth in SEQ ID NO: 15, wherein the amino acid sequence comprises lysine at position 10, glutamic acid at position 56 and wherein:

- (i) X at position 11 is not a basic amino acid;
- (ii) X at position 28 is any amino acid, preferably a polar amino acid or cysteine;
- (iii) X at position 35 is a hydrophobic amino acid or a polar amino acid;
- (iv) X at position 38 is a polar amino acid;
- (v) X at position 73 is a hydrophobic amino acid; and
- (vi) X at position 74 is a polar amino acid,

and wherein the polypeptide is capable of spontaneously forming an isopeptide bond with a peptide comprising an amino acid sequence as set forth in SEQ ID NO: 4, wherein the isopeptide bond forms between the aspartic acid residue at position 8 of SEQ ID NO: 4 and the lysine residue at position 10 of the polypeptide; or

a portion of the polypeptide defined in (a) comprising an amino acid sequence with at least 80% sequence identity to a sequence as set forth in SEQ ID NO: 16, wherein X at position 11 is not a basic amino acid, and the amino acid sequence comprises lysine at position 10, glutamic acid at position 56 and one or more of the following:

- (i) alanine at position 35;
- (ii) threonine at position 38;
- (iii) isoleucine at position 73; and
- (iv) threonine at position 74,

wherein the specified amino acid residues are at positions equivalent to the positions in SEQ ID NO: 16

and wherein the polypeptide is capable of spontaneously forming an isopeptide bond with a peptide comprising an amino acid sequence as set forth in SEQ ID NO: 4, wherein the isopeptide bond forms between the aspartic acid residue at position 8 of SEQ ID NO: 4 and the lysine residue at position 10 of the polypeptide.

As discussed above, it has been determined that variants of SpyCatcher polypeptides typically are capable of spontaneously forming an isopeptide bond with all SpyTag peptide variants, e.g. SpyTag (SEQ ID NO: 28), SpyTag002 (SEQ ID NO: 26) and SpyTag003 (SEQ ID NO: 29). Accordingly, it will be understood that the polypeptide provided herein also may be capable of spontaneously forming an isopeptide bond with such peptides. Thus, the polypeptide may be capable of spontaneously forming an isopeptide bond with a peptide comprising an amino acid sequence as set forth in SEQ ID NO: 4, 26, 28 and/or 29, wherein the isopeptide bond forms between the aspartic acid residue at: (i) position 8 of SEQ ID NO: 4 or 26; (ii) position 7 of SEQ ID NO: 28; or (iii) position 10 of SEQ ID NO: 29; and the lysine residue at position 31 of the polypeptide (or position 10 of the portion of the polypeptide). The polypeptide provided herein may be capable of spontaneously forming an isopeptide bond with variants of said peptides, e.g. variants comprising 1-5 mutations, e.g. substitutions, deletions or additions as defined elsewhere herein.

As previously described, the polypeptide and peptide provided herein function effectively in cold and acidic conditions, e.g. a temperature of about 0-4 °C and a pH of about 2.0-3.0. While the polypeptide and peptide also function effectively in a broader range of temperature and pH conditions as described in

- 10 -

detail below, it is preferred that a polypeptide variant as defined herein is functionally equivalent to an exemplified polypeptide (e.g. SEQ ID NO: 1) under cold and acidic conditions. In a representative example, the polypeptide is capable of spontaneously forming an isopeptide bond with a peptide comprising an amino acid sequence as set forth in SEQ ID NO: 4 as described above, at a temperature of about 0-4 °C (e.g. 0 °C) and/or a pH of about 2.0-3.0 (e.g. 2.5). Alternatively viewed, a polypeptide variant as defined herein is functionally equivalent to an exemplified polypeptide (e.g. SEQ ID NO: 1) under HDX quench conditions.

A further aspect provided herein is a peptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 4.

Thus, another aspect provided herein is a two-part linker comprising:

- (a) the polypeptide as defined herein; and
- (b) the peptide as defined herein,

wherein the peptide and polypeptide are capable of spontaneously forming an isopeptide bond between an aspartic acid residue at a position corresponding to position 8 of SEQ ID NO: 4 and a lysine residue at a position corresponding to position 31 of SEQ ID NO: 1.

A further aspect provided herein is a recombinant or synthetic polypeptide comprising a peptide or polypeptide linked to either the polypeptide or the peptide as described herein.

Also provided herein is a nucleic acid molecule comprising a nucleotide sequence which encodes the polypeptide, peptide or recombinant or synthetic polypeptide as described herein. Also provided is a vector comprising the nucleic acid molecule as described herein.

Another aspect provided herein is a kit comprising:

- (1)
  - (i) the peptide or the recombinant polypeptide as defined herein; and
  - (ii) the polypeptide or the recombinant polypeptide as defined herein; and/or
- (2)

(i) a nucleic acid molecule, particularly a vector, encoding the peptide or recombinant polypeptide as defined in (1)(i); and

(ii) a nucleic acid molecule, particularly a vector, encoding a polypeptide or recombinant polypeptide as defined in (1)(ii).

Still another aspect provided herein is the use of the polypeptide as defined herein to conjugate two molecules or components via an isopeptide bond,

wherein the molecules or components conjugated via an isopeptide bond comprise:

a) a first molecule or component comprising a polypeptide as defined herein; and

5 b) a second molecule or component comprising a peptide as defined herein, particularly a recombinant polypeptide comprising a peptide as defined herein,

and wherein the isopeptide bond forms between an aspartic acid residue at a position corresponding to position 8 of SEQ ID NO: 4 and a lysine residue at a position corresponding to position 31 of SEQ ID NO: 1.

10 Still further provided herein is a process for conjugating two molecules or components via an isopeptide bond comprising:

a) providing a first molecule or component comprising a polypeptide as defined herein;

15 b) providing a second molecule or component comprising a peptide as defined herein, particularly a recombinant polypeptide comprising a peptide as defined herein;

c) contacting the first and second molecules or components under conditions that enable the spontaneous formation of an isopeptide bond between the polypeptide and peptide, thereby conjugating the first molecule or component to

20 the second molecule or component via an isopeptide bond to form a complex,

wherein the isopeptide bond forms between an aspartic acid residue at a position corresponding to position 8 of SEQ ID NO: 4 and a lysine residue at a position corresponding to position 31 of SEQ ID NO: 1.

Further provided herein is the use of the polypeptide, peptide and/or recombinant polypeptide as defined herein in a protein analysis method, optionally where in the protein analysis method uses mass spectrometry, particularly Hydrogen-Deuterium eXchange (HDX) mass spectrometry.

25

Yet further provided herein is an apparatus suitable for use in the process or use as defined herein comprising a solid substrate on which a polypeptide as defined herein is immobilised.

30

Also provided is a kit suitable for use in preparing a solid substrate on which a polypeptide as defined herein is immobilised, comprising:

a) a polypeptide as defined herein; and

b) means for immobilising the polypeptide of a) on a solid substrate.

- 12 -

Yet further provided herein is a resin and/or beads for use the process or use as defined herein on which a polypeptide as defined herein is immobilised.

Still further provided herein is a process or method for analysing a polypeptide of interest (POI) comprising:

- 5 (i) providing a sample comprising the POI linked to the peptide as defined herein;
- (ii) subjecting the sample to hydrogen-deuterium exchange to provide a sample comprising a deuterium-labelled POI linked to the peptide as defined herein;
- 10 (iii) contacting the sample from (ii) with a polypeptide as defined herein under conditions that enable the spontaneous formation of an isopeptide bond between the polypeptide and peptide linked to the POI, thereby conjugating the POI to the polypeptide via an isopeptide bond to form a complex, wherein the conditions are also HDX quench conditions (i.e. conditions that slow or minimise the H/D
- 15 exchange in the sample);
- (iv) isolating the complex under HDX quench conditions; and
- (v) analysing the complex, e.g. using mass spectrometry.

#### DETAILED DESCRIPTION

20 The polypeptide provided may be viewed as a polypeptide binding partner of a two-part linker as mentioned above. In brief, a two-part linker, i.e. a peptide tag and its polypeptide binding partner (a so-called peptide tag/binding partner pair) are derived from a protein capable of spontaneously forming an isopeptide bond (an isopeptide protein), wherein the domains of the protein are expressed separately to

25 produce a peptide tag that comprises one of the residues involved in the isopeptide bond (e.g. an aspartate or asparagine) and a polypeptide binding partner (or “catcher”) that comprises the other residue involved in the isopeptide bond (e.g. a lysine) and at least one other residue required to form the isopeptide bond (e.g. a glutamate). Mixing the peptide tag and polypeptide binding partner results in the

30 spontaneous formation of an isopeptide bond between the tag and binding partner. Thus, by separately conjugating the peptide tag and polypeptide binding partner to different molecules or entities, e.g. a protein of interest (POI) and a solid support (e.g. bead or resin), it is possible to covalently link said molecules/entities together

35 via an isopeptide bond formed between the peptide tag and polypeptide binding partner, i.e. to form a polypeptide-based linker between the molecules/entities

conjugated to the peptide tag and polypeptide binding partner. The peptide tag and polypeptide binding partner that react with each other (e.g. specifically and efficiently) to form an isopeptide bond may be defined as a pair of peptide linkers, particularly a cognate pair of peptide linkers. Alternatively, the polypeptide binding partner may be viewed as a polypeptide comprising a peptide-binding domain that reacts its cognate peptide tag to form an isopeptide bond. The terms “polypeptide”, “polypeptide binding partner” and “catcher” are used interchangeably herein.

As described in further detail in the Examples and above, it was surprisingly determined that a substitution of the arginine residue at position 32 of SpyCatcher003 to a non-basic amino acid residue resulted in significantly increased reactivity of the polypeptide with SpyTag002 under cold and acidic conditions. In this respect, at neutral pH most proteins will have a mixture of positive and negative charges on their surface, whereas in acidic conditions, such as those used to quench H-D exchange (e.g. pH 2-3) there may be almost no negative charges remaining, which may lead to positive-positive electrostatic repulsions that reduce protein stability. Accordingly, it was hypothesised that the stability of SpyCatcher003 in acidic conditions may be improved by mutating residues that are likely to be positively charged in such conditions. As shown in the Examples, several residues that are likely to be positively charged at pH 2-3 in SpyCatcher003 and SpyTag002 were identified (K28, R32, R37, R47, K52, H62, K72, and H112 in SpyCatcher003 and K11, R12 and K14 in SpyTag002). However, it was found that mutation of these residues to non-basic amino acids did not improve reactivity, except for R32. In this respect, it was demonstrated that the characteristics of the substituted amino acid at this position were not critical to the improved activity, as mutants containing amino acids with structurally unrelated R-groups at this position, i.e. threonine and valine, both showed significantly improved reactivity (e.g. SEQ ID NO: 5, SpyCatcher003 R32V). Thus, it is contemplated that the substitution of arginine at position 32 with any non-basic amino acid, i.e. an amino acid other than a basic amino acid, will improve the reactivity of SpyCatcher003 in acidic conditions. Alternatively viewed, the substitution of arginine at position 32 with any non-basic amino acid will increase the pH range in which the polypeptide is functional.

The term “basic amino acid” as used herein refers to amino acids which contain side chains (R-groups) which are cationic (positively charged) at neutral pH. Basic amino acids have a high pKa which allows them to bind protons, thus gaining

a positive charge. Conventional basic amino acids, i.e. amino acids coded for by the standard genetic code, are lysine, arginine and histidine. Accordingly, in the polypeptide provided herein, X at position 32 of SEQ ID NO: 2 or position 11 of SEQ ID NO: 15 is not lysine, arginine or histidine.

5           Thus, alternatively viewed, X at position 32 of SEQ ID NO: 2 or position 11 of SEQ ID NO: 15 may be selected from alanine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, isoleucine, leucine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine or valine. In some embodiments, X at position 32 of SEQ ID NO: 2 or position 11 of SEQ ID NO: 15 is  
10 not cysteine.

          The X at position 32 of SEQ ID NO: 2 or position 11 of SEQ ID NO: 15 may be a hydrophobic and/or a polar amino acid, particularly an aliphatic amino acid or polar amino acid. The term "hydrophobic amino acid" as used herein refers to an amino acid which has a hydrophobic side chain predominantly comprised of carbon  
15 and hydrogen. A hydrophobic amino acid may be selected from alanine, cysteine, glycine, isoleucine, leucine, methionine, tryptophan, phenylalanine, proline and valine. An aliphatic amino acid may be selected from alanine, isoleucine, leucine, proline and valine.

          The term "polar amino acid" as used herein refers to an amino acid which  
20 contains side chains that are polar, and thus have an uneven distribution of electrons on their surface but are not charged. These side chains are usually hydrophilic in nature, thus allowing polar amino acids to participate in hydrogen bonding. A polar amino acid may be selected from serine, threonine, cysteine, asparagine, glutamine and tyrosine.

25           Thus, the hydrophobic amino acid at position 32 of SEQ ID NO: 2 or position 11 of SEQ ID NO: 15 may be valine, isoleucine, leucine, or alanine, preferably valine. The polar amino acid at position 32 of SEQ ID NO: 2 or position 11 of SEQ ID NO: 15 may be threonine or serine, preferably threonine.

          As discussed in the Examples, a directed evolution approach was utilised to  
30 identify additional mutations in the SpyCatcher003 R32V polypeptide (SEQ ID NO: 5) that further improved the reactivity of the polypeptide with SpyTag002 under cold and acidic conditions. While it is contemplated that each substitution in the polypeptide relative to the amino acid sequence of the original SpyCatcher003 polypeptide may separately improve the reactivity of the polypeptide under cold and  
35 acidic conditions, it will be understood that the mutation at position 32 alone is

sufficient to provide a polypeptide with improved reactivity in acidic and cold conditions relative to SpyCatcher003. Thus, while it may be preferred that the polypeptide contains one or more of the other residues identified herein as improving reactivity, it is contemplated that if the polypeptide does not contain all of the residues specified above, it is preferred that in the specified positions the variants contain the amino acid residues at the equivalent position in the SpyCatcher003 polypeptide (SEQ ID NO: 27). The equivalent positions readily can be determined by comparing the amino acid sequence of the polypeptide with SEQ ID NO: 27, e.g. using the BLASTP algorithm.

Thus, by way of example, if the polypeptide comprises an amino acid sequence with at least 80% sequence identity to a sequence as set forth in SEQ ID NO: 2 and the residue at position 49 (or the equivalent position) is not cysteine, it is preferred that the residue is serine. Similarly, if the residue at position 56 (or the equivalent position) is not alanine, it is preferred that the residue is threonine. If the residue at position 59 (or the equivalent position) is not threonine, it is preferred that the residue is serine. If the residue at position 94 (or the equivalent position) is not isoleucine, it is preferred that the residue is valine. If the residue at position 95 (or the equivalent position) is not threonine, it is preferred that the residue is asparagine. This applies to other residues specified below. However, it will be understood that substitutions, particularly conservative substitutions, at the specified positions may be tolerated without significantly affecting the reactivity of the polypeptide.

Thus, for example, the polypeptide may comprise an amino acid sequence with at least 80% sequence identity to an amino acid sequence as set forth in SEQ ID NO: 5, wherein the amino acid at position 31 is lysine, the amino acid at position 32 is arginine and the amino acid at position 77 is glutamic acid and, optionally, the amino acids at positions 49, 56, 59, 94 and 95 are as defined herein, wherein the specified amino acid residues are at positions equivalent to the positions in SEQ ID NO: 5.

Accordingly, X at position 49 of SEQ ID NO: 2 or position 28 of SEQ ID NO: 15 may be any amino acid, particularly any conventional amino acid. In preferred embodiments, X at position 49 of SEQ ID NO: 2 or position 28 of SEQ ID NO: 15 may be a polar amino acid or a cysteine. Thus, alternatively viewed, in some embodiments, the X at position 49 of SEQ ID NO: 2 or position 28 of SEQ ID NO: 15 may be serine, threonine, cysteine, asparagine, glutamine or tyrosine. In a

particularly preferred embodiment, X at position 49 of SEQ ID NO: 2 or position 28 of SEQ ID NO: 15 may be cysteine, serine, threonine or tyrosine, particularly cysteine, serine or threonine. As discussed further below, the presence of a cysteine residue at position 49 of SEQ ID NO: 2 or position 28 of SEQ ID NO: 15 provides the polypeptide with an additional functionality, i.e. it facilitates the coupling of the polypeptide to a solid support (e.g. bead or resin) via a covalent bond, e.g. a disulfide bond. While this additional functionality is useful, particularly in the protein analysis methods described herein, it is not an essential feature of the polypeptide which may find utility in various in-solution methods. Thus, in some aspects, X at position 49 is a polar amino acid other than cysteine, e.g. an amino acid selected from serine, threonine, asparagine, glutamine and tyrosine, preferably serine or threonine. In a particularly preferred embodiment, X at position 49 of SEQ ID NO: 2 or position 28 of SEQ ID NO: 15 may be cysteine or serine, preferably cysteine.

In a further aspect, X at position 56 of SEQ ID NO: 2 or position 35 of SEQ ID NO: 15 may be a hydrophobic and/or polar amino acid, particularly an aliphatic amino acid or polar amino acid. Thus, X at position 56 of SEQ ID NO: 2 or position 35 of SEQ ID NO: 15 may be alanine, cysteine, glycine, isoleucine, leucine, methionine, tryptophan, phenylalanine, proline, valine, serine, threonine, asparagine, glutamine, or tyrosine. In particular, X at position 56 of SEQ ID NO: 2 or position 35 of SEQ ID NO: 15 may be selected from: (i) alanine, isoleucine, leucine, proline and valine, particularly alanine; or (ii) serine, threonine, cysteine, asparagine, glutamine and tyrosine, particularly threonine and serine. Thus, X at position 56 of SEQ ID NO: 2 or position 35 of SEQ ID NO: 15 may be alanine or threonine, preferably alanine.

The X at position 59 of SEQ ID NO: 2 or position 38 of SEQ ID NO: 15 is a polar amino acid. Thus, X at position 59 of SEQ ID NO: 2 or position 38 of SEQ ID NO: 15 may be selected from serine, threonine, cysteine, asparagine, glutamine and tyrosine, particularly serine, threonine, asparagine, glutamine and tyrosine. In particular, X at position 59 of SEQ ID NO: 2 or position 38 of SEQ ID NO: 15 may be threonine or serine, preferably threonine.

The X at position 94 of SEQ ID NO: 2 or position 73 of SEQ ID NO: 15 is a hydrophobic amino acid, preferably an aliphatic amino acid. Thus, X at position 94 of SEQ ID NO: 2 or position 73 of SEQ ID NO: 15 may be alanine, cysteine, glycine, isoleucine, leucine, methionine, tryptophan, phenylalanine, proline or

- 17 -

valine. In particular, X at position 94 of SEQ ID NO: 2 or position 73 of SEQ ID NO: 15 may be selected from alanine, isoleucine, leucine, proline and valine. Thus, in a preferred aspect it may be valine or isoleucine, preferably isoleucine.

5 The X at position 95 of SEQ ID NO: 2 or position 74 of SEQ ID NO: 15 is a polar amino acid. Thus, X at position 95 of SEQ ID NO: 2 or position 74 of SEQ ID NO: 15 may be serine, threonine, cysteine, asparagine, glutamine or tyrosine, particularly serine, threonine, asparagine, glutamine and tyrosine. In particular, X at position 95 of SEQ ID NO: 2 or position 74 of SEQ ID NO: 15 may be threonine, serine, asparagine or glutamine, preferably threonine or serine, e.g. threonine.

10 It will be understood that the amino acids at the positions specified herein may be independently selected from the lists of amino acid specified above and all possible combinations are contemplated herein. However, for the purpose of brevity, not all possible combinations are explicitly listed.

In a representative example:

15 (i) X at position 49 of SEQ ID NO: 2 or position 28 of SEQ ID NO: 15 is cysteine or serine, preferably cysteine;

(ii) X at position 56 of SEQ ID NO: 2 or position 35 of SEQ ID NO: 15 is alanine or threonine, preferably alanine;

20 (iii) X at position 59 of SEQ ID NO: 2 or position 38 of SEQ ID NO: 15 is threonine or serine, preferably threonine;

(iv) X at position 94 of SEQ ID NO: 2 or position 73 of SEQ ID NO: 15 is valine or isoleucine, preferably isoleucine; and/or

(v) X at position 95 of SEQ ID NO: 2 or position 74 of SEQ ID NO: 15 is threonine, serine, asparagine or glutamine, preferably threonine or serine.

25 As detailed above, any modification or combination of modifications may be made to SEQ ID NO: 2 to produce a variant polypeptide provided herein (i.e. a sequence identity related variant as defined herein), provided that the variant polypeptide comprises a lysine residue at a position equivalent to position 31 of SEQ ID NO: 2, a glutamic acid residue at a position equivalent to position 77 of  
30 SEQ ID NO: 2 and all of the amino acid residues at positions equivalent to positions 32, 49, 56, 59, 94 and 95 of SEQ ID NO: 2 as detailed above. Furthermore, the variant polypeptide must retain the functional characteristics defined above, i.e. the resultant polypeptide is capable of spontaneously forming an isopeptide bond with a peptide tag comprising of an amino acid sequence as set forth in SEQ ID NO: 4 and  
35 optionally has an equivalent or higher yield, reaction rate, e.g. rate constant,

temperature and/or buffer range relative to a polypeptide having an amino acid sequence as set forth in SEQ ID NO: 1.

An equivalent position in the polypeptide provided herein is preferably determined by reference to the amino acid sequence of SEQ ID NO: 1 or 2. The homologous or corresponding position can be readily deduced by lining up the sequence of the homologue (mutant, variant or derivative) polypeptide and the sequence of SEQ ID NO: 1 or 2 based on the homology or identity between the sequences, for example using a BLAST algorithm.

Thus, the polypeptide as described herein may comprise:

(a) an amino acid sequence with at least 80% sequence identity (e.g. at least 85, 90, 95, 96, 97, 98 or 99% sequence identity) to a sequence as set forth in SEQ ID NO: 3, wherein X at position 32 is not a basic amino acid, and the amino acid sequence comprises lysine at position 31, glutamic acid at position 77 and one or more of the following:

- (i) alanine at position 56;
- (ii) threonine at position 59;
- (iii) isoleucine at position 94; and
- (iv) threonine at position 95,

wherein the specified amino acid residues are at positions equivalent to the positions in SEQ ID NO: 3; or

(b) an amino acid sequence with at least 80% sequence identity (e.g. at least 85, 90, 95, 96, 97, 98 or 99% sequence identity) to a sequence as set forth in SEQ ID NO: 16, wherein X at position 11 is not a basic amino acid, and the amino acid sequence comprises lysine at position 10, glutamic acid at position 56 and one or more of the following:

- (i) alanine at position 35;
- (ii) threonine at position 38;
- (iii) isoleucine at position 73; and
- (iv) threonine at position 74,

wherein the specified amino acid residues are at positions equivalent to the positions in SEQ ID NO: 16.

It is contemplated that the polypeptide of the invention may comprise any one or any combination of the specified amino acid residues defined above (e.g. any combination of two, three or four of the amino acid residues specified above),

e.g. i) and ii, i) and ii), i) and iv), ii) and iii), ii) and iv) etc., i), ii) and iii), i), iii) and iv), etc.

Thus, the polypeptide may comprise:

5 (a) an amino acid sequence with at least 80% sequence identity (e.g. at least 85, 90, 95, 96, 97, 98 or 99% sequence identity) to a sequence as set forth in SEQ ID NO: 3, wherein X at position 32 is not a basic amino acid, and the amino acid sequence comprises lysine at position 31, glutamic acid at position 77 and all of the following:

- 10 (i) alanine at position 56;  
(ii) threonine at position 59;  
(iii) isoleucine at position 94; and  
(iv) threonine at position 95,

wherein the specified amino acid residues are at positions equivalent to the positions in SEQ ID NO: 3; or

15 (b) an amino acid sequence with at least 80% sequence identity (e.g. at least 85, 90, 95, 96, 97, 98 or 99% sequence identity) to a sequence as set forth in SEQ ID NO: 16, wherein X at position 11 is not a basic amino acid, and the amino acid sequence comprises lysine at position 10, glutamic acid at position 56 and all of the following:

- 20 (i) alanine at position 35;  
(ii) threonine at position 38;  
(iii) isoleucine at position 73; and  
(iv) threonine at position 74,

25 wherein the specified amino acid residues are at positions equivalent to the positions in SEQ ID NO: 16.

The X at position 32 of SEQ ID NO: 3 or at position 11 of SEQ ID NO: 16 may be as defined above. In particular, it may be a hydrophobic and/or a polar amino acid, particularly an aliphatic amino acid or polar amino acid. The hydrophobic amino acid residue may be selected from valine, isoleucine, leucine and alanine, preferably valine. The polar amino acid residue may be selected from  
30 threonine or serine, preferably threonine.

As detailed above, sequence identity-related variants of the exemplified polypeptides are provided herein. Thus, said sequence identity-related variants contain the above described amino acid residues at the equivalent positions in the  
35 exemplified polypeptides, particularly SEQ ID NO: 1 or 13. The homologous or

corresponding position can be readily deduced by lining up the sequence of the homologue (mutant, variant or derivative) polypeptide and the sequence of SEQ ID NO: 1 based on the homology or identity between the sequences, for example using a BLAST algorithm.

5            Thus, any mutations that are present in the polypeptide provided herein (i.e. the sequence identity-related variants) relative to the exemplified polypeptides (e.g. SEQ ID NO: 1 or 2) may be conservative amino acid substitutions, unless specified otherwise. A conservative amino acid substitution refers to the replacement of an amino acid by another which preserves the physicochemical character of the  
10            polypeptide (e.g. D may be replaced by E or vice versa, N by Q, or L or I by V or vice versa). Thus, generally the substituting amino acid has similar properties, e.g. hydrophobicity, hydrophilicity, electronegativity, bulky side chains etc. to the amino acid being replaced. Isomers of the native L-amino acid e.g. D-amino acids may be incorporated. In particular, it is preferred that mutations at positions other than  
15            those specified herein are conservative substitutions.

              Standard amino acid nomenclature is used herein. Thus, the full name of an amino acid residue may be used interchangeably with one letter code or three letter abbreviations. For instance, lysine may be substituted with K or Lys, isoleucine may be substituted with I or Ile, and so on. Moreover, the terms aspartate and aspartic  
20            acid, and glutamate and glutamic acid are used interchangeably herein and may be replaced with Asp or D, or Glu or E, respectively.

              It is contemplated that a polypeptide variant provided herein may differ from any one of SEQ ID NOs: 1, 5-14 or 17-25 by, for example, 1 to 22, 1 to 20, 1 to 18, 1 to 16, 1 to 14, 1 to 12, 1 to 10, 1 to 15, 1 to 10, 1 to 8, 1 to 6, 1 to 5, 1 to 4, e.g. 1,  
25            2 or 3 amino acid substitutions, insertions and/or deletions, preferably 1 to 22, 1 to 20, 1 to 15, 1 to 10, 1 to 8, 1 to 6, 1 to 5, 1 to 4, e.g. 1, 2 to 3 amino acid substitutions and/or 1 to 22, 1 to 20, 1 to 15, 1 to 10, 1 to 8, 1 to 6, 1 to 5, 1 to 4, e.g. 1, 2 or 3 amino acid deletions. It is preferred that deletions are at the N-terminus, i.e. truncations, preferably N-terminal truncations, thereby generating  
30            polypeptide portions of SEQ ID NO: 1 and 5-13 as defined above, e.g. SEQ ID NOs: 14 or 17-25. Thus, it is preferred that variants of the polypeptide portions provided herein do not contain deletions relative to the exemplified sequences, i.e. SEQ ID NOs: 14 or 17-25. Thus, it is preferred that variants of the polypeptide portions contain only substitutions relative to the exemplified sequences,

particularly conservative substitutions, e.g. 1 to 18, 1 to 16, 1 to 14, 1 to 12, 1 to 10, 1 to 15, 1 to 10, 1 to 8, 1 to 6, 1 to 5, 1 to 4, e.g. 1, 2 or 3 amino acid substitutions.

As detailed above, mutant forms of the exemplified polypeptide are provided herein (i.e. referred to herein as homologues, variants or derivatives) and are  
5 structurally similar to the exemplified polypeptides set forth in SEQ ID NO: 1 and 13. The amino acid sequences of specific variants provided herein are set forth in SEQ ID NOs: 5-12, 14 and 17-25. The polypeptide variants are able to function as a binding partner to a peptide tag, i.e. capable of spontaneously forming an isopeptide bond between the aspartic acid at position 8 of the peptide as set forth  
10 SEQ ID NO: 4 and the lysine at position 31 (or an equivalent position, based on the numbering of SEQ ID NO: 1) of the polypeptide variant under suitable conditions as defined herein.

In cases where a polypeptide variant comprises mutations, e.g. deletions or insertions, relative to SEQ ID NOs: 1 or 13, the residues specified above are  
15 present at equivalent amino acid positions in the variant polypeptide sequences. Thus, in some embodiments, deletions in the polypeptide variants provided herein are not N-terminal truncations.

As noted above, polypeptides provided by SEQ ID NOs: 1-3 and 5-13 may be truncated by up to 21 amino acids at the N-terminus (e.g. by 5, 10, 15 or 20  
20 amino acids). Thus, the term variant as used herein includes truncation variants of the exemplified polypeptide, such as those set forth in SEQ ID NOs: 14-25.

As discussed above, the polypeptide provided herein is derived from the SpyCatcher003 polypeptide, which contains a number of residues that may contribute to the reaction rate of the polypeptide. Thus, it may be preferred that  
25 polypeptide variants do not contain substitutions at positions 2, 9, 13, 19, 34, 37, 50, 62, 69, 86, 89, 91, 97, 103, 105, 108 or 113 of SEQ ID NO: 1 or that such substitutions are conservative substitutions. Thus, a polypeptide provided herein may contain any one or more of the following:

- 1) threonine or alanine at position 2;
- 30 2) glycine at position 9;
- 3) proline at position 13;
- 4) threonine at position 19;
- 5) glutamic acid at position 34;
- 6) arginine at position 37;
- 35 7) serine at position 50;

- 22 -

- 8) histidine at position 62;  
9) tyrosine or serine at position 69;  
10) valine at position 86;  
11) proline at position 89;  
5 12) glutamic acid at position 91;  
13) aspartic acid at position 97;  
14) aspartic acid at position 103;  
15) glutamic acid at position 105;  
16) glutamic acid at position 108; and  
10 17) threonine at position 113.

Thus, it is contemplated that the polypeptide may comprise a polypeptide having amino acid sequence with at least 80% sequence identity to an amino acid sequence as set forth in SEQ ID NO: 1, wherein the amino acid sequence comprises:

- 15 1) lysine at position 31;  
2) valine, isoleucine, leucine, alanine, serine or threonine at position 32, preferably valine or threonine at position 32;  
3) cysteine or serine at position 49;  
4) alanine or threonine at position 56;  
20 5) threonine or serine at position 59;  
6) glutamic acid at position 77;  
7) isoleucine or valine at position 94; and  
8) threonine, serine, asparagine or glutamine at position 95, preferably threonine or serine at position 95; and  
25 9) any combination of the specified amino acid residues 1)-17) defined above. However, some particularly preferred combinations include:  
a) 1) threonine or alanine at position 2;  
2) proline at position 13;  
3) lysine at position 31;  
30 4) valine or threonine at position 32;  
5) arginine at position 37;  
6) cysteine or serine at position 49;  
7) alanine or threonine at position 56;  
8) threonine or serine at position 59;  
35 9) histidine at position 62;

- 23 -

- 10) glutamic acid at position 77;  
11) glutamic acid at position 91;  
12) isoleucine or valine at position 94;  
13) threonine or serine at position 95;  
5 14) aspartic acid at position 103;  
15) glutamic acid at position 105;  
16) glutamic acid at position 108; and  
17) threonine at position 113;
- b) 1) threonine or alanine at position 2;  
10 2) proline at position 13;  
3) lysine at position 31;  
4) valine or threonine at position 32;  
5) arginine at position 37;  
6) cysteine or serine at position 49;  
15 7) alanine or threonine at position 56;  
8) threonine or serine at position 59;  
9) histidine at position 62;  
10) glutamic acid at position 77;  
11) proline at position 89;  
20 12) glutamic acid at position 91;  
13) isoleucine or valine at position 94;  
14) threonine or serine at position 95;  
15) aspartic acid at position 103;  
16) glutamic acid at position 105;  
25 17) glutamic acid at position 108; and  
18) threonine at position 113;
- c) 1) threonine or alanine at position 2;  
2) proline at position 13;  
3) lysine at position 31;  
30 4) valine or threonine at position 32;  
5) arginine at position 37;  
6) cysteine or serine at position 49;  
7) alanine or threonine at position 56;  
8) threonine or serine at position 59;  
35 9) histidine at position 62;

- 24 -

- 10) glutamic acid at position 77;  
11) proline at position 89;  
12) glutamic acid at position 91;  
13) isoleucine or valine at position 94;  
5 14) threonine or serine at position 95;  
15) aspartic acid at position 97;  
16) aspartic acid at position 103;  
17) glutamic acid at position 105;  
18) glutamic acid at position 108; and  
10 19) threonine at position 113; and
- d) 1) threonine or alanine at position 2;  
2) proline at position 13;  
3) lysine at position 31;  
4) valine or threonine at position 32;  
15 5) arginine at position 37;  
6) cysteine or serine at position 49;  
7) alanine or threonine at position 56;  
8) threonine or serine at position 59;  
9) histidine at position 62;  
20 10) glutamic acid at position 77;  
11) proline at position 89;  
12) glutamic acid at position 91;  
13) isoleucine or valine at position 94;  
14) threonine or serine at position 95;  
25 15) aspartic acid at position 97;  
16) aspartic acid at position 103;  
17) glutamic acid at position 105;  
18) glutamic acid at position 108; and  
19) threonine at position 113;

30 wherein the specified amino acid residues are at positions equivalent to the positions in SEQ ID NO: 1.

Thus, a polypeptide may comprise an amino acid sequence with at least 80% (e.g. at least 85, 90, 95, 96, 97, 98 or 99%) sequence identity to an amino acid sequence as set forth in SEQ ID NO: 1, wherein the polypeptide comprises a lysine  
35 at position 31, a glutamic acid at position 77 and all of the following:

- 25 -

- 1) threonine or alanine at position 2, preferably threonine;
- 2) glycine at position 9;
- 3) proline at position 13;
- 4) threonine at position 19;
- 5) valine or threonine at position 32, preferably valine;
- 6) glutamic acid at position 34;
- 7) arginine at position 37;
- 8) cysteine or serine at position 49;
- 9) serine at position 50;
- 10) alanine or threonine at position 56, preferably alanine;
- 11) threonine or serine at position 59, preferably threonine;
- 12) threonine at position 59;
- 13) histidine at position 62;
- 14) tyrosine at position 69;
- 15) glycine at position 83;
- 16) valine at position 86;
- 17) proline at position 89;
- 18) glutamic acid at position 91;
- 19) valine or isoleucine at position 94, preferably isoleucine;
- 20) threonine or serine at position 95, preferably threonine;
- 21) aspartic acid at position 97;
- 22) aspartic acid at position 103;
- 23) glutamic acid at position 105;
- 24) glutamic acid at position 108; and
- 25) threonine at position 113;

wherein the specified amino acid residues are at positions equivalent to the positions in SEQ ID NO: 1.

Further amino acid residues that it may be preferable to conserve (i.e. not mutate) in the polypeptide variants provided herein are described in WO2020/183198 (incorporated herein by reference).

Thus, in some embodiments, the polypeptide as described herein may comprise:

(a) an amino acid sequence selected from a sequence as set forth in any one of SEQ ID NOs: 1 and 5-14 or 17-25, preferably SEQ ID NOs: 1, 13, 14 or 25;

- 26 -

(b) an amino acid sequence selected from a sequence as set forth in any one of SEQ ID NOs: 5-13, wherein the serine residue at position 49 is substituted with cysteine; or

5 (c) an amino acid sequence selected from a sequence as set forth in any one of SEQ ID NOs: 17-25, wherein the serine residue at position 28 is substituted with cysteine.

As described above, the polypeptide may comprise:

10 (a) an amino acid sequence with at least 80% sequence identity (e.g. at least 85, 90, 95, 96, 97, 98 or 99% sequence identity) to an amino acid sequence selected from a sequence as set forth in any one of SEQ ID NOs: 1 and 5-14 or 17-25, preferably SEQ ID NOs: 1, 13, 14 or 25;

15 (b) an amino acid sequence with at least 80% sequence identity (e.g. at least 85, 90, 95, 96, 97, 98 or 99% sequence identity) to an amino acid sequence selected from a sequence as set forth in any one of SEQ ID NOs: 5-13, wherein the serine residue at position 49 is substituted with cysteine; or

(c) an amino acid sequence with at least 80% sequence identity (e.g. at least 85, 90, 95, 96, 97, 98 or 99% sequence identity) to an amino acid sequence selected from a sequence as set forth in any one of SEQ ID NOs: 17-25, wherein the serine residue at position 28 is substituted with cysteine,

20 wherein the polypeptides contain the residues specified in the lists above.

Alternatively, the polypeptide may comprise:

(a) an amino acid sequence selected from a sequence as set forth in any one of SEQ ID NOs: 1 and 5-14 or 17-25, preferably SEQ ID NOs: 1, 13, 14 or 25;

25 (b) an amino acid sequence selected from a sequence as set forth in any one of SEQ ID NOs: 5-13, wherein the serine residue at position 49 is substituted with cysteine; or

(c) an amino acid sequence selected from a sequence as set forth in any one of SEQ ID NOs: 17-25, wherein the serine residue at position 28 is substituted with cysteine,

30 wherein the polypeptide contains 1-22, such as 1-15, 1-12 or 1-10, amino acid substitutions relative to the recited sequences, wherein the amino acid substitutions are not in the positions specified in the lists above.

35 Sequence identity may be determined by any suitable means known in the art, e.g. using the SWISS-PROT protein sequence databank using FASTA pep-cmp with a variable pamfactor, and gap creation penalty set at 12.0 and gap extension

- 27 -

penalty set at 4.0, and a window of 2 amino acids. Other programs for determining amino acid sequence identity include the BestFit program of the Genetics Computer Group (GCG) Version 10 Software package from the University of Wisconsin. The program uses the local homology algorithm of Smith and Waterman with the default  
5 values: Gap creation penalty = 8, Gap extension penalty = 2, Average match = 2.912, Average mismatch = -2.003.

Preferably said comparison is made over the full length of the sequence (which may be a corresponding fragment or portion), but may be made over a smaller window of comparison, e.g. less than 100 or 50 contiguous amino acids.

10 Preferably, the polypeptide variants described herein (e.g. sequence identity-related variants) are functionally equivalent to the polypeptide having a sequence as set forth in SEQ ID NO: 1 or 13. As referred to herein, "functional equivalence" refers to variants of the polypeptide provided herein that may show some reduced efficacy in the spontaneous formation of an isopeptide bond with its  
15 respective partner (e.g. lower expression yield, lower reaction rate, or activity in a limited range of reaction conditions (e.g. narrower temperature range, such as 0-10 °C etc.)) relative to the parent molecule (i.e. the molecule with which it shows sequence homology, e.g. SEQ ID NO: 1 or 13), but preferably are as efficient or are more efficient.

20 The term "functional variant" means that the variant polypeptide must retain the ability to spontaneously form an isopeptide bond with its cognate peptide tag under suitable conditions.

Thus, alternatively viewed, a functional variant of a polypeptide provided herein may have at least 80, 85, 90, 95, 96, 97, 98 or 99% sequence identity to the  
25 sequence to which it is compared (e.g. SEQ ID NOs: 1 or 13).

A mutant or variant polypeptide with activity that is "equivalent" to the activity of a polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 1 or 13 may have activity that is similar (i.e. comparable) to the activity of a polypeptide comprising or consisting of an amino acid sequence as set forth in  
30 SEQ ID NO: 1 or 13, i.e. such that the practical applications of the polypeptide are not significantly affected, e.g. within a margin of experimental error. Thus, an equivalent polypeptide activity means that the mutant or variant polypeptide is capable of spontaneously forming an isopeptide bond with a peptide (e.g. comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 4)  
35 with a similar reaction rate (i.e. rate constant as discussed below) and/or yield to a

polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 1 or 13 under the same conditions.

The activity of the polypeptides (e.g. SEQ ID NO: 1 or 13 versus their respective mutants) measured under the same reaction conditions, e.g.

5 temperature, pH, substrates (i.e. peptide tag sequence) and their concentration, buffer, salt etc. as exemplified above, can be readily compared to determine whether the activity for each polypeptide is higher, lower or equivalent.

In particular, the polypeptide variants may have an equivalent rate constant with the peptide provided herein (i.e. SEQ ID NO: 4) to the polypeptide having a  
10 sequence as set forth in SEQ ID NOs: 1 or 13. The rate constant refers to the coefficient of proportionality relating the rate of the reaction (the formation of an isopeptide bond) at a given temperature to the product of the concentrations of reactants (i.e. the product of the concentration of the peptide tag and polypeptide of the invention).

15 Thus, the activity, e.g. rate constant, of the variant (e.g. mutant) polypeptide (polypeptide binding partner) may be at least 50%, e.g. at least 60, 70, 75, 80, 85 or 90% of the activity, e.g. rate constant, of a polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 1 or 13, such as at least 91, 92, 93, 94, 95, 96, 97, 98 or 99% of the activity, e.g. rate constant, of a polypeptide  
20 comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 1 or 13. Alternatively viewed, the activity of the mutant polypeptide may be no more than 50% lower than the activity, e.g. rate constant, of a polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 1 or 13, e.g. no more than 45, 40, 35, 30, 25 or 20% lower than the activity, e.g. rate constant, of a  
25 polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 1 or 13, such as no more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1% lower than the activity, e.g. rate constant, of a polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 1 or 13.

Notably, the rate constant of the reaction of the polypeptide and peptide tag  
30 provided herein may be lower than the values described in the Examples when the polypeptide or peptide tag are fused to large molecules or components (e.g. proteins, beads), which diffuse slower than the isolated polypeptide and peptide tag. Moreover, the rate constant may be reduced if the molecules or components to which the polypeptide and/or peptide tag are fused cause steric hindrance to the  
35 reaction. Accordingly, when measuring the rate constant of the reaction of the

polypeptide variants, it is preferred that measurement is performed using isolated polypeptides and peptide tags, i.e. polypeptides and peptide tags that are not fused or conjugated to other molecules or components. However, it will be understood that it may be convenient to use polypeptides and peptide tags that are fused or  
5 conjugated to other molecules or components, e.g. polypeptides, when measuring the rate constant. In particular, the peptide tag may be in the form of a fusion protein with a polypeptide, e.g. maltose binding protein (MBP).

The reaction rate and rate constant can be assessed by any suitable means known in the art and as described in the Examples. For instance, the reaction rate  
10 may be monitored by assessing the mobility of the reaction products on SDS-PAGE after boiling in SDS or other strong denaturing treatment that would disrupt all non-covalent interactions or by mass spectrometry.

As previously described, the polypeptide and peptide provided herein function effectively in cold and acidic conditions, e.g. a temperature of about 0-4 °C  
15 and a pH of about 2.0-3.0. While the polypeptide and peptide also function effectively in a broader range of temperature and pH conditions as described in detail below, it is preferred that a polypeptide variant as defined herein is functionally equivalent to an exemplified polypeptide (e.g. SEQ ID NO: 1) under cold and acidic conditions. Thus, when assessing the reaction rate and rate  
20 constant of a polypeptide variant, it is preferable to use conditions in which the exemplified polypeptide (e.g. SEQ ID NO: 1) shows improved characteristics relative to its parent molecule, i.e. SpyCatcher003 (SEQ ID NO: 27). Thus, in a representative example, a polypeptide variant is functionally equivalent to an exemplified polypeptide (e.g. SEQ ID NO: 1) as defined above at a temperature of  
25 about 0-4 °C (e.g. 0 °C) and a pH of about 2.0-3.0 (e.g. 2.5). Alternatively viewed, a polypeptide variant as defined herein is functionally equivalent to an exemplified polypeptide (e.g. SEQ ID NO: 1) under HDX quench conditions. Further suitable conditions for determining the reaction rate of a polypeptide provided herein are described below.

30 As described in the Examples, it may be useful to immobilise the polypeptide on a solid substrate, e.g. to facilitate the isolation or purification of a molecule or component (e.g. POI) for further analysis, and it will be evident that this may be achieved in any convenient way. Thus, the manner or means of immobilisation and the solid support may be selected from any number of

immobilisation means and solid supports as are widely known in the art and described in the literature.

Thus, the polypeptide may be directly bound to the support, for example via a domain or moiety of the polypeptide (e.g. chemically cross-linked). In some  
5 embodiments, the polypeptide may be bound indirectly by means of a linker group, or by an intermediary binding group(s) (e.g. by means of a biotin-streptavidin interaction). Thus, the polypeptide may be covalently or non-covalently linked to the solid support. In certain embodiments the polypeptide is immobilised on a solid substrate via a covalent bond.

10 The linkage may be a reversible (e.g. cleavable) or irreversible linkage. Thus, in some embodiments, the linkage may be cleaved enzymatically, chemically, or with light, e.g. the linkage may be a light-sensitive linkage.

Linking groups of interest may vary widely depending on the nature of the solid support. The linking group, when present, is in many embodiments  
15 biologically inert.

Many linking groups are known to those of skill in the art and may find use in immobilising the polypeptide to a solid support or other molecule or entity. A linking group generally may be at least about 50 daltons, usually at least about 100 daltons and may be as large as 1000 daltons or larger, for example up to 1000000 daltons  
20 if the linking group contains a spacer, but generally will not exceed about 1000 daltons and usually will not exceed about 600 daltons. Generally, such linkers will comprise a spacer group terminated at either end with a reactive functionality capable of covalently bonding to the polypeptide other molecule or entity, e.g. solid support.

25 Spacer groups of interest may include aliphatic and unsaturated hydrocarbon chains, spacers containing heteroatoms such as oxygen (ethers such as polyethylene glycol) or nitrogen (polyamines), peptides, carbohydrates, cyclic or acyclic systems that may possibly contain heteroatoms. Spacer groups may also be comprised of ligands that bind to metals such that the presence of a metal ion  
30 coordinates two or more ligands to form a complex. Specific spacer elements include: sulfosuccinimidyl 6-(3'-(2-pyridyldithio)propionamido)hexanoate, 1,4-diaminohexane, xylylenediamine, terephthalic acid, 3,6-dioxaoctanedioic acid, ethylenediamine-N,N-diacetic acid, 1,1'-ethylenebis(5-oxo-3-pyrrolidinecarboxylic acid), 4,4'-ethylenedipiperidine, oligoethylene glycol and polyethylene glycol.

35 Potential reactive functionalities include nucleophilic functional groups (amines,

alcohols, thiols, hydrazides), electrophilic functional groups (aldehydes, esters, vinyl ketones, epoxides, isocyanates, maleimides), functional groups capable of cycloaddition reactions, forming disulfide bonds, or binding to metals. Specific examples include primary and secondary amines, hydroxamic acids, N-  
5 hydroxysuccinimidyl esters, N-hydroxysuccinimidyl carbonates, oxycarbonylimidazoles, nitrophenylesters, trifluoroethyl esters, glycidyl ethers, vinylsulfones, and maleimides. Specific linker groups that may find use linking the polypeptide to a solid support include heterofunctional compounds, such as  
10 azidobenzoyl hydrazide, N-[4-(p-azidosalicylamino)butyl]-3'-[2'-pyridyldithio]propionamid), bis-sulfosuccinimidyl suberate, dimethyladipimidate, disuccinimidyltartrate, N-maleimidobutyryloxysuccinimide ester, N-hydroxy sulfosuccinimidyl-4-azidobenzoate, N-succinimidyl [4-azidophenyl]-1,3'-dithiopropionate, N-succinimidyl [4-iodoacetyl]aminobenzoate, glutaraldehyde, and  
15 succinimidyl-4-[N-maleimidomethyl]cyclohexane-1-carboxylate, 3-(2-pyridyldithio)propionic acid N-hydroxysuccinimide ester (SPDP), 4-(N-maleimidomethyl)-cyclohexane-1-carboxylic acid N-hydroxysuccinimide ester (SMCC), and the like. For instance, a spacer may be formed with an azide reacting with an alkyne or formed with a tetrazine reacting with a trans-cyclooctene or a norbornene.

20 A polypeptide as described herein may be provided with means for immobilisation (e.g. an affinity binding partner, e.g. biotin or a hapten, capable of binding to its binding partner, i.e. a cognate binding partner, e.g. streptavidin or an antibody) provided on the support. The interaction between the polypeptide and the solid support must be robust enough to allow for washing steps, i.e. the interaction  
25 between the polypeptide and solid support is not disrupted (significantly disrupted) by the washing steps. For instance, it is preferred that with each washing step, less than 5%, preferably less than 4, 3, 2, 1, 0.5 or 0.1% of the polypeptide is removed or eluted from the solid phase.

30 As detailed above and discussed in detail in the Examples, the inventors have determined that the cysteine at position 49 of the polypeptide facilitates the coupling of the polypeptide to a solid support with minimal disruption to SpyTag peptide binding and capture. Without wishing to be bound by theory, it is hypothesised that coupling the polypeptide to a solid support via the cysteine  
35 residue at position 49 of SEQ ID NO: 1 maximised accessibility of the polypeptide to its cognate peptide tag (e.g. LemonTag and variants thereof).

Thus, the polypeptide may be immobilised on a solid support by means of covalent bond, e.g. a covalent bond that may be provided by a linkage group as described above.

Thus, the polypeptide as described herein may be immobilised on a solid substrate via a covalent bond between a cysteine residue and the solid substrate, wherein the cysteine residue is at a position equivalent to position 49 in SEQ ID NO: 1. The covalent bond may be formed by reacting the thiol group on the cysteine residue with an appropriate reactive group on the solid substrate. For instance, the solid substrate may comprise an iodoacetyl group, e.g. the solid substrate may be an iodoacetyl-activated substrate. Alternatively, the solid substrate may comprise a carboxy group, e.g. the solid substrate may be a diamino-dipropylamine activated substrate. In an advantageous embodiment, the polypeptide may be immobilised on a solid substrate via a reversible linkage, particularly a chemically cleavable linkage, e.g. a disulfide bond. In a representative embodiment, the polypeptide may be immobilised on a solid substrate using a crosslinker for amine-to-sulfhydryl conjugation, e.g. via NHS-ester and pyridyldithiol reactive groups, that forms a cleavable (reducible) disulfide bond with the thiol group of a cysteine residue. In particular, the crosslinker may be Sulfo-succinimidyl 6-(3'-[2-pyridyldithio]-propionamido)hexanoate (Sulfo-LC-SPDP).

A chemically cleavable linkage may be cleaved using any suitable means. For instance, a linkage comprising a disulfide bond may be cleaved by the addition of a reducing agent, such as tris(2-carboxyethyl)phosphine (TCEP) or dithiothreitol (DTT).

Whilst it may be advantageous to immobilise the polypeptide on a solid support via the cysteine residue at position 49 of SEQ ID NO: 1, this is not essential.

Thus, the manner or means of immobilisation and the solid support may be selected, according to the choice of the skilled person, from any number of immobilisation means and solid supports as are widely known in the art and described in the literature.

The solid support (phase or substrate) may be any of the well-known supports or matrices which are currently widely used or proposed for immobilisation, separation etc. These may take the form of particles (e.g. beads which may be magnetic, para-magnetic or non-magnetic), sheets, gels, filters,

membranes, fibres, capillaries, slides, arrays or microtitre strips, tubes, plates or wells etc.

The support may be made of glass, silica, latex or a polymeric material, e.g. polyacrylamide or a polysaccharide polymer material, such as agarose (e.g. sepharose) or dextran. Suitable are materials presenting a high surface area for binding of the polypeptide. Such supports may have an irregular surface and may be for example porous or particulate, e.g. particles, fibres, webs, sinters or sieves. Particulate materials, e.g. beads, are useful due to their greater binding capacity, particularly polymeric (e.g. agarose) beads.

Conveniently, a particulate solid support used according to the invention will comprise spherical beads. The size of the beads is not critical, but they may for example be of the order of diameter of at least 1  $\mu\text{m}$  and preferably at least about 2  $\mu\text{m}$ , 5  $\mu\text{m}$ , 10  $\mu\text{m}$  or 20  $\mu\text{m}$  and have a maximum diameter of preferably not more than about 500  $\mu\text{m}$ , and e.g. not more than about 100  $\mu\text{m}$ .

Monodisperse particles, that is those which are substantially uniform in size (e.g. size having a diameter standard deviation of less than 5%) have the advantage that they provide very uniform reproducibility of reaction. Representative monodisperse polymer particles may be produced by the technique described in US-A-4336173.

However, to aid manipulation and separation, magnetic beads are advantageous. The term "magnetic" as used herein means that the support is capable of having a magnetic moment imparted to it when placed in a magnetic field, and thus is displaceable under the action of that field. In other words, a support comprising magnetic particles may readily be removed by magnetic aggregation, which provides a quick, simple and efficient way of separating the particles following the isopeptide bond formation steps.

In some embodiments, the solid support is a resin, e.g. a thiol-reactive resin. In some embodiments, the solid support is beads. In further embodiments, the solid support may be both a resin and beads.

Thus, the polypeptide of the present invention may be immobilised on a solid substrate.

Thus, in a further embodiment, the present invention provides a kit suitable for use in preparing a solid substrate on which a polypeptide as defined herein is immobilised, comprising:

a) a polypeptide as defined herein; and

- 34 -

b) means for immobilising the polypeptide of a) on a solid substrate (e.g. a crosslinker, such as Sulfosuccinimidyl 6-(3'-[2-pyridyldithio]-propionamido)hexanoate (Sulfo-LC-SPDP)).

5 In a further embodiment, the kit as described above may additionally comprise a solid substrate on which the polypeptide as described herein is immobilised.

As described in detail within the Examples, a SpyTag002 variant (i.e. mutant) peptide tag (SEQ ID NO: 4) was identified that exhibited improved reactivity with the polypeptide defined herein in cold and acidic conditions. Therefore, also provided herein is a peptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 4.

As detailed above, the present invention may alternatively be seen as providing a two-part linker comprising the polypeptide and peptide provided herein. The term "linker" as used herein refers to molecules that function to link, i.e. conjugate or join, two molecules or components together, preferably by a covalent bond, e.g. an isopeptide bond. Thus, the polypeptide and peptide provided herein may be viewed as a two-part linker, wherein formation of the isopeptide bond between the first part, i.e. polypeptide, and second part, i.e. peptide tag, reconstitutes the linker, thereby joining molecules or components fused or conjugated to said first and second parts of the linker. Alternatively stated, the polypeptide and peptide provided herein may be viewed as a cognate pair that functions as a linker, i.e. a polypeptide and peptide tag cognate pair or a peptide and binding partner cognate pair. These terms are used interchangeably throughout the description.

25 Alternatively viewed, the polypeptide and peptide provided herein may be viewed as a covalent capture system, particularly when the polypeptide is immobilised on a solid support. As shown in the Examples, the immobilised polypeptide finds particular utility in reacting with a polypeptide of interest (POI), e.g. target polypeptide, comprising a peptide provided herein, e.g. a recombinant polypeptide comprising a POI and the peptide provided herein, in a complex sample. The formation of the isopeptide bond therefore captures the POI on the solid support via the polypeptide. The solid support facilitates the isolation of the POI from the sample. Thus, as noted above, the reaction of polypeptide and peptide functions to form a linker, i.e. to link the POI to the solid support.

The term “cognate” refers to components that function together. Thus, in the context of the present invention, a cognate pair refers to a polypeptide and peptide provided herein that react together spontaneously to form an isopeptide bond.

Thus, a two-part linker comprising a polypeptide and peptide tag that react together efficiently to form an isopeptide bond under conditions that enable the spontaneous formation of said isopeptide bond can also be referred to as being a  
5 “complementary pair”, i.e. a polypeptide and peptide complementary pair.

Thus, the present invention may be seen as providing a two-part linker comprising:

10 (a) a polypeptide as defined herein; and

(b) a peptide as defined herein,

wherein the peptide and polypeptide are capable of spontaneously forming an isopeptide bond between the aspartic acid residue at a position 8 of SEQ ID NO: 4 and a lysine residue at a position corresponding to position 31 of SEQ ID NO: 1.

15 As noted above, the polypeptide provided herein polypeptide may be capable of spontaneously forming an isopeptide bond with all SpyTag peptide variants. Thus, the peptide provided in the two-part linker or kit disclosed herein may comprise an amino acid sequence as set forth in SEQ ID NO: 4, 26, 28, 29 or a variant thereof (e.g. a peptide variant comprising 1-5 substitutions, deletions or  
20 additions relative to the recited sequences, such 1, 2 or 3 substitutions), wherein the isopeptide bond forms between the aspartic acid residue at: (i) position 8 of SEQ ID NO: 4 or 26; (ii) position 7 of SEQ ID NO: 28; or (iii) position 10 of SEQ ID NO: 29; and the lysine residue at position 31 of the polypeptide (or position 10 of the portion of the polypeptide). It will be evident that the position of the aspartic acid  
25 may differ in the peptide variants mentioned above, and thus the formation of the isopeptide bond in the variants will result from an aspartic acid in an equivalent position to those mentioned above.

As previously detailed, the polypeptide and peptide provided herein spontaneously form an isopeptide bond between the aspartic acid residue at  
30 position 8 of SEQ ID NO: 4 and the lysine residue at position 31 of SEQ ID NO: 1 under various conditions including those explained below that are suitable for the formation of an isopeptide bond between said peptide tag and polypeptide. As is evident from the Examples below, the polypeptide and peptide provided herein are active and able to form a spontaneous isopeptide bond under a wide range of  
35 conditions.

For instance, the polypeptide and peptide are active in a variety of buffers including phosphate buffers (e.g. phosphate buffered saline (PBS)), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), HEPES buffered saline (HBS), Tris, Tris buffered saline (TBS), both with and without EDTA. As detailed above, the present invention finds particularly utility in HDX-MS workflows. Thus, the peptide tag and polypeptide of the present invention are active in buffers utilised in these work flows including quench buffers. The term "quench buffer" as used herein refers to a buffer which is able to quench the hydrogen-deuterium exchange reaction. Commonly used quench buffers include phosphate buffers, e.g. buffers comprising  $\text{NaH}_2\text{PO}_4$ , having a pH of about 4.0 or less, such as about 3.5, 3.0 or less (e.g. about 2.0 or 2.5). As it is very important to control the pH precisely in HDX-MS workflows, it may be particularly advantageous to perform the HD exchange in a "weak" buffer (i.e. low concentration) and subsequently quench the exchange using a "strong" buffer (i.e. high concentration). The differential in concentrations ensures that the pH will drop to a suitable acidic pH and remain at that pH. Thus, a quench buffer may be a more concentrated buffer than an HD exchange buffer, e.g. 2, 3, 4 or more times more concentrated. In a representative example, an HD exchange buffer may be a phosphate buffer with a concentration of about 10-30 mM (e.g. 25 mM) at pH 6.0-8.0 (e.g. about pH 7.0) and a corresponding quench buffer may be a phosphate buffer with a concentration of about 80-120 mM (e.g. 100 mM) at pH 2.0-4.0 (e.g. about pH 2.5 or 3.0).

HDX quench conditions refer to conditions used in HDX workflows that are required to limit (i.e. minimise) D-to-H back-exchange. The precise conditions and formulation of the buffers used in the conditions may vary depending on the POI, but typically HDX quench conditions involve a low pH and low temperature. In a representative example, HDX quench conditions involve subjecting a sample containing a POI to a temperature of about 10 °C or less, such as about 0-10 °C, about 0-6 °C or about 0-4 °C and a pH of about 4.5 or less, e.g. about 1.0-4.5, 1.5-4.0, 2.0-3.5, 3.0-4.0 or 2.0-3.0, such as about 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8 or 2.9. Any combination of temperature and pH may be selected from the ranges above. For instance, HDX quench conditions may be a pH of about 2.5 and temperate of 0-4 °C.

The skilled person will understand that the composition of the buffer in which the polypeptide and peptide are used will be dependent on the workflow or protocol in which they are being used.

In certain protocols in which the polypeptide and peptide may find utility, the reaction conditions (e.g. buffers) may be supplemented with various additives. For example, a reducing agent, such as tris(2-carboxyethyl)phosphine (TCEP), may be included in the reaction conditions, e.g. the buffer may contain a reducing agent, such as TCEP or DTT. TCEP is a common HDX-MS quench additive that reduces (i.e. cleaves) or prevents the formation of disulfide bonds. As described in the Examples, while a POI captured on a solid support via the polypeptide can be digested in solution using acid proteases (e.g. pepsin) for analysis (e.g. using mass spectrometry), it was found that the addition of TCEP facilitated the elution of the entire polypeptide:peptide:POI complex (i.e. LemonCatcher:LemonTag-POI complex) from the solid support. It will be understood that the inclusion of a reducing agent may depend on the workflow being performed. For instance, where the polypeptide is conjugated to a solid support via a reducible linker, a reducing agent may not be added until after the complex has been formed, i.e. until the isopeptide bond has formed.

Other additives which may be included in the reaction conditions (i.e. reaction mixture) and may be beneficial to the reactivity between the polypeptide and peptide include glycerol, sucrose, methanol and glycol, e.g. ethylene glycol. For instance, the reaction mixture may contain glycerol at a concentration of about 5-25% (v/v), such as about 10-20% or 12-18% (v/v), e.g. about 15% (v/v). Additionally or alternatively, the reaction mixture may contain sucrose at a concentration of about 5-30% (w/v), such as about 5-25% or 5-20% (w/v), e.g. about 5-15% (w/v). As described below, the reaction conditions may contain a cryoprotectant, e.g. ethylene glycol or an alcohol (e.g. methanol), to facilitate the use of very low temperatures. Thus, the reaction mixture may contain a cryoprotectant, e.g. ethylene glycol or methanol, at a concentration of about 10-50% (v/v), such as about 15-45% or 20-45% (v/v), e.g. about 25-40% (v/v). The skilled person would readily be able to determine other suitable concentrations and/or additives that would be beneficial to the specific workflow in question.

The polypeptide and peptide provided here are also active in the presence of the commonly used detergents, such as Tween 20 and Triton X-100, e.g. up to a concentration of about 1% (v/v), and in the presence of urea, e.g. up to a concentration of about 3 M. The skilled person would readily be able to determine other suitable detergents and their concentrations.

As described in the Examples, it was determined that the mutations to SpyCatcher003 and SpyTag002 as described herein resulted in increased reactivity compared to the original SpyCatcher003:SpyTag002 pairing in acidic and cold conditions (e.g. HDX quench conditions). However, it was also found that the mutated polypeptide and peptide (i.e. LemonCatcher (SEQ ID NO: 1) and LemonTag (SEQ ID NO: 4)) exhibited improved reactivity at neutral pH (pH 7.5) compared to the original SpyCatcher003:SpyTag002 pairing. Thus, LemonTag and LemonCatcher may be viewed as expanding the range of pH and temperature functionality compared to previously developed Catcher/Tag technologies.

Moreover, it will be understood that the utility of the polypeptide and peptide linker system provided herein is not limited to use in HDX quench conditions. As the system is active in a wide range of conditions, it has a diverse number of utilities.

Thus, conditions that are suitable for the formation of an isopeptide bond between a polypeptide as disclosed herein and its cognate peptide tag (e.g. a peptide comprising of an amino acid sequence as set forth in SEQ ID NOs: 4) includes any conditions in which contacting the cognate peptide tag and polypeptide as disclosed herein results in the spontaneous formation of an isopeptide bond between said cognate peptide tag and polypeptide as disclosed herein, particularly between the aspartic acid residue at position 8 of SEQ ID NO: 4 and the lysine residue at position 31 of SEQ ID NO: 1 or a functional variant thereof (e.g. SEQ ID NOs: 5-14 or 17-25). For instance, contacting the polypeptide and peptide tag may occur in buffered conditions, e.g. in a buffered solution or on a solid phase (e.g. column) that has been equilibrated with a buffer, such as a phosphate buffer.

The step of contacting may be at any suitable pH, such as about pH 1.0-8.0. In some embodiments, the step of contacting may be at an acidic pH, e.g. about 1.0-5.5, 2.5-4.5, 3.0-4.0 or 2.0-3.0, such as about pH 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0, 3.2, 3.4, 3.6, 3.8, 4.0, 4.5, 5.0 or 5.5. In some embodiments, the step of contacting may be at a neutral, mildly alkaline or mildly acidic pH, e.g. about 6.0-8.0, e.g. about 6.0, 6.5, 7.0, 7.5 or 8.0. Additionally or alternatively, the step of contacting may be at any suitable temperature, such as about -25 °C to 40 °C. In some embodiments, the step of contacting may be at a low temperature, e.g. about -25 °C to 10 °C, -20 °C to 8 °C, -15 °C to 6 °C, -10 °C to 4 °C, such as about -5 °C to 5 °C, 0-10 °C, 0-6 °C or 0-4 °C, e.g. about -10, -5, -2, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 °C. For instance, such conditions are common in protein analysis workflows,

such as HDX-MS. However, in some embodiments, it may be advantageous to use higher temperatures. Thus, the step of contacting may be at a temperature of about 10-39, 12-38, 13-37, 14-36, 15-35, 16-34, 17-33, 18-32, 19-31 or 20-30 °C, e.g. about 20, 22, 25, 28, 30, 33, 35 or 37 °C.

5           Contacting the polypeptide and peptide "under conditions that enable the spontaneous formation of an isopeptide bond" may include contacting the polypeptide and peptide in the presence of a chemical chaperone, e.g. a molecule that enhances or improves the reactivity of the peptide and polypeptide. The chemical chaperone may be TMAO (trimethylamine N-oxide), which may be  
10 present in the reaction at a concentration of at least about 0.2 M, e.g. at least 0.3, 0.4, 0.5, 1.0, 1.5, 2.0 or 2.5 M, e.g. about 0.2-3.0 M, 0.5-2.0 M, 1.0-1.5 M.

          As noted above, the formation of the isopeptide bond between the polypeptide and peptide provided herein is spontaneous. In this respect, the polypeptide comprises a glutamic acid at position 77 (or an equivalent position,  
15 based on the numbering of SEQ ID NO: 1) that facilitates, e.g. induces, promotes or catalyses, the formation of the isopeptide bond between the lysine residue at position 31 (or an equivalent position, based on the numbering of SEQ ID NO: 1) of the polypeptide and aspartic acid residue at position 8 of the peptide (SEQ ID NO:  
4).

20           The term "spontaneous" as used herein refers to an isopeptide bond, which can form in a protein or between peptides or proteins (e.g. between two peptides or a peptide and a protein, i.e. the polypeptide and peptide provided herein) without any other agent (e.g. an enzyme catalyst) being present and/or without chemical modification of the protein or peptide, e.g. without native chemical ligation or  
25 chemical coupling using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC). Thus, native chemical ligation to modify a peptide or protein having a C-terminal thioester is not carried out.

          Thus, a spontaneous isopeptide bond can form between a polypeptide and peptide of the invention when in isolation and without chemical modification of the  
30 polypeptide and/or peptide. A spontaneous isopeptide bond may therefore form of its own accord in the absence of enzymes or other exogenous substances and without chemical modification of the polypeptide and/or peptide tag of the invention.

          A spontaneous isopeptide bond may form almost immediately after contact of the polypeptide and peptide, e.g. within 1, 2, 3, 4, 5, 10, 15, 20, 25 or 30 minutes,  
35 or within 1, 2, 4, 8, 12, 16, 20 or 24 hours.

The speed of isopeptide formation will be dependent on the concentration of the polypeptide and peptide reactants and the conditions of the reaction, e.g. temperature and/or pH. Spontaneous isopeptide bond formation may complete for about 80% or more of the reactants in about 120 minutes, e.g. where the reactants are each present at a concentration of about 10  $\mu\text{M}$  at a reaction temperature of about 0  $^{\circ}\text{C}$  and pH of about 7.0. Spontaneous isopeptide bond formation may complete for about 50% or more of the reactants in about 120 minutes, e.g. where the reactants are each present at a concentration of about 1  $\mu\text{M}$  at a reaction temperature of about 0  $^{\circ}\text{C}$  and pH of about 7.0.

The other reaction conditions, e.g. buffer, etc. used to determine the speed of reaction defined above may be any conditions defined herein. In some embodiments, the reaction conditions are those used in the Examples. For instance, in some embodiments, the spontaneous isopeptide bond formation is complete in the amounts specified above in PBS buffer at a pH of about 7.0, particularly in buffer comprising 5 mM TCEP.

The term "peptide" is used herein interchangeably with the term "peptide tag", "peptide linker", or "tag".

There is no standard definition regarding the size boundaries between what is meant by peptide. Typically, a peptide may be viewed as comprising between 2-39 amino acids. Accordingly, a polypeptide may be viewed as comprising at least 40 amino acids, preferably at least 50, 60, 70, 80, 90, 100, 110 or 120 amino acids.

A peptide tag provided herein may be viewed as comprising at least 14 amino acids, e.g. 14-39 amino acids, such as e.g. 14-35, 15-34, 16-33, 17-31, 18-30 amino acids in length, e.g. it may comprise or consist of 14, 15, 16, 17, 18, 19 or 20 amino acids.

A polypeptide provided herein may be viewed as comprising at least 83 amino acids, e.g. 83-150, such as e.g. 80-140, 80-130, 80-120 amino acids in length, e.g. it may comprise or consist of 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119 or 120 amino acids.

As discussed above, two-part linkers as defined herein have a large number of utilities and the polypeptide and peptide provided herein find particular utility in conjugating (i.e. joining or linking) two molecules or components via an isopeptide bond. For instance, the polypeptide and peptide tag may be separately conjugated or fused to molecules or components of interest and subsequently contacted

together under conditions suitable to allow the spontaneous formation of an isopeptide bond between the polypeptide and peptide tag, thereby joining (i.e. linking or conjugating) the molecules or components via an isopeptide bond. In some embodiments, the polypeptide may be linked to a solid support (e.g. bead),  
5 i.e. immobilised on solid support, and used to capture a POI containing the peptide tag on the solid support, i.e. via formation of an isopeptide bond between the peptide and immobilised polypeptide.

Thus, in some embodiments, the invention may be seen to provide the use of a polypeptide and peptide as defined herein to conjugate two molecules or  
10 components (e.g. entities) via an isopeptide bond,

wherein the molecules or components conjugated via an isopeptide bond comprise:

- a) a first molecule or component comprising a polypeptide as defined herein; and
- 15 b) a second molecule or component comprising a peptide as defined herein, e.g. a recombinant polypeptide comprising a peptide comprising or consisting of the amino acid sequence as set forth in SEQ ID NO: 4,

wherein the isopeptide bond forms between an aspartic acid residue at a position corresponding to position 8 of SEQ ID NO: 4 and a lysine residue at a  
20 position corresponding to position 31 of SEQ ID NO: 1.

It will be evident that the use the polypeptide and peptide tag (i.e. two-part linker) described above comprises contacting said first and second molecules under conditions suitable to enable (e.g. promote or facilitate) the spontaneous formation of an isopeptide bond between said polypeptide and peptide tag as described  
25 above.

Thus, alternatively viewed, the invention provides a process for conjugating two molecules or components via an isopeptide bond comprising:

- a) providing a first molecule or component comprising a polypeptide as defined herein;
- 30 b) providing a second molecule or component comprising a peptide as defined herein, e.g. a recombinant polypeptide comprising or consisting of the amino acid sequence as set forth in SEQ ID NO: 4;
- c) contacting the first and second molecules or components under conditions that enable the spontaneous formation of an isopeptide bond between

the polypeptide and peptide, thereby conjugating the first molecule or component to the second molecule or component via an isopeptide bond to form a complex,

wherein the isopeptide bond forms between an aspartic acid residue at a position corresponding to position 8 of SEQ ID NO: 4 and a lysine residue at a position corresponding to position 31 of SEQ ID NO: 1.

In some embodiments, the first molecule or component as described in the use and process above may comprise an immobilised polypeptide as defined herein. In a further optional embodiment, the covalent bond that immobilises the polypeptide on the solid substrate is a reducible covalent bond, e.g. a disulfide bond.

Alternatively viewed, the process for conjugating two molecules or components via an isopeptide bond as defined above may be seen as a process for purifying or isolating the second molecule or component, the process comprising:

(a) contacting the immobilised polypeptide as defined herein with a sample comprising the second molecule or component (e.g. POI containing the peptide as defined herein) under conditions that enable the spontaneous formation of an isopeptide bond between the immobilised polypeptide and the peptide in the second molecule or component (e.g. POI) thereby forming a complex;

(b) separating the complex from other molecules or components in the sample, thereby isolating or purifying the second molecule or component; and optionally

(c) releasing the complex from the solid substrate.

In some embodiments, the step of releasing the complex from the solid substrate comprises contacting the complex with a reducing agent. In further optional embodiments, said reducing agent is TCEP or DTT.

In some embodiments, such as in the context of protein analysis workflows, e.g. MS, HDX-MS, releasing the complex from the solid substrate may comprise proteolytic digestion of the complex, e.g. contacting the complex with a protease (e.g. acidic protease, such as pepsin). It will be understood that contacting the complex with a protease may result in the release of a plurality of portions of the complex, i.e. the complex may not be released from the solid support intact. Furthermore, a portion of the complex may remain on the solid substrate. Thus, alternatively viewed, the step of releasing the complex may be incomplete or may be a partial release of the complex.

As described above, the polypeptide and peptide are capable of spontaneously forming an isopeptide in a wide range of conditions, e.g. temperature and pH conditions, and any such conditions may be used in the process defined above. Furthermore, each step of the process may be performed  
5 under the same conditions or different conditions. For instance, the contacting step may be performed at the higher end of the functional temperature range for the polypeptide and peptide, e.g. 25-30°C, to maximise the activity of the polypeptide, and subsequent separating and release steps may be performed at lower temperatures, e.g. to minimise degradation of the captured molecule or component,  
10 i.e. as is typical in protein purification workflows. However, as discussed herein, the polypeptide and peptide provided herein find particular utility in protein analysis methods, such as HDX-MS, e.g. in the capture of a deuterated POI. In this utility, all of the steps may be performed under HDX quench conditions as defined herein.

Thus, in the use or the process as detailed above, the steps may be  
15 performed at:

- i) a temperature of about -20°C to about 10 °C, such as about -15°C to about 10 °C, about -10°C to about 5 °C, about 0-6°C or about 0-4°C; and/or
- ii) an acidic pH, preferably a pH of about 1.0-5.5, such as about 2.5-4.5, about 3.0-4.0, about 2.0-3.0 or about 2.5-3.5.

20 Alternatively, in the use or the process as detailed above, the steps may be performed at:

- i) a temperature of about 10-40°C, optionally about 20-40°C or about 25-35°C; and/or
- ii) a pH of about 1.0-8.0, optionally about 3.0-7.5 or about 4.0-7.0.

25 It will be understood that the step of separating the complex from other molecules or components in the sample may be achieved using any suitable means and will depend on the solid support on which the polypeptide is immobilised. Typically, the step will involve separating the solid support from the rest of the sample and optionally washing the solid support.

30 The step of washing the solid support may comprise contacting the solid support with a wash solution prior to releasing said molecule or component from the solid support. The washing step may utilise any conditions suitable to maintain the covalent interaction between the polypeptide and peptide, e.g. conditions used in the contacting step.

The wash solution may be selected based on the molecules or components conjugated or linked to the peptide tag. Furthermore, the step of washing the solid support may be repeated multiple times, e.g. 2, 3, 4, 5 or more times. Alternatively viewed, in some embodiments the process comprises multiple wash steps, wherein  
5 the same or different washing conditions may be used in each step.

Where the solid support comprises beads (e.g. agarose-based beads) the volume of buffer used in the wash steps may be at least about 2 times the volume of the beads, e.g. at least about 3, 4, 5, 6, 7, 8, 9 or 10 times the volume of the  
10 beads.

The solid substrate may be subjected to stringent washing conditions. The nature of the stringent washing conditions will depend on the molecules or components conjugated or linked to the peptide tag and/or the composition of the solid substrate. The skilled person could select such conditions as a matter of routine.  
15

The washing steps should be conducted in conditions that minimise release of the molecule or component comprising the peptide tag from the solid substrate. Suitable "non-eluting" conditions may be determined readily by a person of skill in the art based on routine experimentation and may depend on the nature of the molecule or component being isolated or purified. In some embodiments, the  
20 washing steps are performed at about 10 °C or less, e.g. 9, 8, 7, 6, 5 or 4 °C or less.

In a further aspect, the invention provides an apparatus for use in the process or use hereinbefore defined comprising a solid substrate (solid support) on which a polypeptide defined herein is immobilised.  
25

The apparatus may comprise a chromatography column comprising the solid substrate on which a polypeptide as defined herein is immobilised. The apparatus may further comprise means for contacting the solid substrate with the sample, washing and release buffers/solutions and/or means for removing (e.g. aspirating) or collecting liquids (e.g. wash-through, eluted fractions) from the solid  
30 substrate.

In a further aspect, the invention provides a kit, particularly a kit for use in preparing a solid substrate on which a polypeptide as defined herein is immobilised, comprising:

- a) a polypeptide as defined herein; and
- 35 b) means for immobilising the polypeptide of a) on a solid substrate.

In a further embodiment, the kit further comprises a solid substrate as defined above, e.g. resin and/or beads (e.g. agarose beads).

Means for immobilising the polypeptide on a solid substrate may comprise reagents for activating the solid substrate (e.g. resin) and/or polypeptide, reagents  
5 for coupling the polypeptide to the solid substrate (e.g. Sulfo-LC-SPDP) and/or reagents for blocking the solid substrate.

Thus, a further aspect provided herein is a resin and/or beads for use the process or use defined above on which a polypeptide as defined herein is immobilised

10 The terms “conjugating” or “linking” in the context of the present invention with respect to connecting two or more molecules or components to form a complex refers to joining or conjugating said molecules or components, e.g. proteins, via a covalent bond, particularly an isopeptide bond which forms between the polypeptide and peptide tag pair that are incorporated in, or fused to, said molecules or  
15 components, e.g. the polypeptide and peptide tag pair may form domains of proteins to be conjugated or linked together. In particular aspects, the peptide forms a domain of a POI to be captured by an immobilised polypeptide as defined herein, particularly wherein the POI has been subjected to HD exchange, i.e. the POI is a deuterated polypeptide comprising the peptide as defined herein.

20 As mentioned above, the polypeptide and peptide tag may be fused or conjugated to other molecules or to other components or entities. Such molecules or components (i.e. entities) may be a protein, peptide, nucleic acid molecule, small-molecule organic compound, fluorophore, metal-ligand complex, polysaccharide, nanoparticle, 2D monolayer (e.g. graphene), nanotube, polymer,  
25 cell, virus, virus-like particle or a combination thereof. In preferred embodiments, the peptide tag is fused or conjugated to a protein or peptide, e.g. a POI, and/or the polypeptide is conjugated to a solid support, e.g. a bead or resin.

The terms “polypeptide of interest”, “protein of interest” and “POI” are used interchangeable herein to refer to a polypeptide or protein that it is desirable to  
30 conjugate to the polypeptide provided herein. For instance, the POI may be part of a complex sample, e.g. a sample containing a plurality of different polypeptides such as a cell lysate, and it is desirable to isolate the POI from the other polypeptides. Thus, the POI may alternatively be viewed as a target polypeptide or target protein.

The cell may be a prokaryotic or eukaryotic cell. In some embodiments, the cell is a prokaryotic cell, e.g. a bacterial cell. In some embodiments, the cell is a eukaryotic cell, such as an animal cell, e.g. a human cell.

5 The polypeptide and/or peptide tag may be conjugated or fused to a compound or molecule which has a therapeutic or prophylactic effect, e.g. an antibiotic, antiviral, vaccine, antitumour agent, e.g. a radioactive compound or isotope, cytokines, toxins, oligonucleotides and nucleic acids encoding genes or nucleic acid vaccines.

10 The polypeptide and/or peptide tag may be conjugated or fused to a label, e.g. a radiolabel, a fluorescent label, luminescent label, a chromophore label as well as to substances and enzymes which generate a detectable substrate, e.g. horseradish peroxidase, luciferase or alkaline phosphatase. This detection may be applied in numerous assays where antibodies are conventionally used, including Western blotting/immunoblotting, histochemistry, enzyme-linked immunosorbent  
15 assay (ELISA), or flow cytometry (FACS) formats. Labels for magnetic resonance imaging, positron emission tomography probes and boron 10 for neutron capture therapy may also be conjugated to the peptide tag and/or polypeptide. Particularly, the polypeptide and/or peptide may be fused or produced with another peptide, for example a purification tag, e.g. His tag, and/or may be fused or produced with  
20 another protein, for example with the purpose of enhancing recombinant protein expression by fusing to Maltose Binding Protein.

In a particularly useful embodiment, the peptide and/or polypeptide may be produced as part of another peptide or polypeptide using recombinant techniques as discussed below, i.e. as a recombinant or synthetic protein or polypeptide. In  
25 embodiments in which the peptide or polypeptide is fused or conjugated to another peptide or polypeptide, the peptide or polypeptide is not a peptide or polypeptide from which the peptide tag or polypeptide of the invention is derived (e.g. the peptide or polypeptide is not an isopeptide protein, i.e. the isopeptide protein from which the polypeptide of the invention is derived (the CnaB2 domain of the  
30 *Streptococcus pyogenes* FbaB protein)).

It will be evident that the polypeptide and/or peptide tag of the invention may be fused to any peptide, proteins or polypeptide. The peptide, protein or polypeptide may be derived or obtained from any suitable source. For instance, the protein may be *in vitro* translated or purified from biological and clinical samples, e.g. any cell or  
35 tissue sample of an organism (eukaryotic, prokaryotic), or any body-fluid or

preparation derived therefrom, as well as samples such as cell cultures, cell preparations, cell lysates etc. Proteins may be derived or obtained, e.g. purified from environmental samples, e.g. soil and water samples or food samples are also included. The samples may be freshly prepared or they may be prior-treated in any  
5 convenient way e.g. for storage.

Thus, in a preferred embodiment, the peptide, protein or polypeptide fused to the peptide or polypeptide of the invention may be produced recombinantly and thus the nucleic acid molecules encoding said recombinant proteins may be derived or obtained from any suitable source, e.g. any viral or cellular material, including all  
10 prokaryotic or eukaryotic cells, viruses, bacteriophages, mycoplasmas, protoplasts and organelles. Such biological material may thus comprise all types of mammalian and non-mammalian animal cells, plant cells, algae including blue-green algae, fungi, bacteria, protozoa, viruses etc. In some embodiments, the proteins may be synthetic proteins. For example, the peptide and polypeptide  
15 (proteins) disclosed herein may be produced by chemical synthesis, such as solid-phase peptide synthesis.

The position of the polypeptide or peptide tag within a recombinant or synthetic polypeptide is not particularly important. Thus, in some embodiments the polypeptide or peptide may be located at the N-terminus or C-terminus of the  
20 recombinant or synthetic polypeptide. In some embodiments, the polypeptide or peptide may be located internally within the recombinant or synthetic polypeptide. Thus, in some embodiments the polypeptide or peptide may be viewed as an N-terminal, C-terminal or internal domain of the recombinant or synthetic polypeptide.

In some preferred embodiments, the polypeptide is preferably located at the  
25 N-terminus or C-terminus of the recombinant or synthetic polypeptide. Thus, in some embodiments the polypeptide may be viewed as an N-terminal or C-terminal domain of the recombinant or synthetic polypeptide.

In some embodiments, it may be useful to include one or more spacers, e.g. a peptide spacer, between the peptide, protein or polypeptide to be joined or  
30 conjugated with the polypeptide or peptide of the invention. Thus, the peptide, protein or polypeptide and the polypeptide or peptide of the invention may be linked directly to each other or they may be linked indirectly by means of one or more spacer sequences. Thus, a spacer sequence may interspace or separate two or  
35 more individual parts of the recombinant or synthetic polypeptide. In some embodiments, a spacer may be N-terminal or C-terminal to the peptide tag and/or

polypeptide. In some embodiments, spacers may be at both sides of the peptide tag and/or polypeptide.

The precise nature of the spacer sequence is not critical and it may be of variable length and/or sequence, for example it may have 1-40, more particularly 2-20, 1-15, 1-12, 1-10, 1-8, or 1-6 residues, e.g. 6, 7, 8, 9, 10 or more residues. By way of representative example, the spacer sequence, if present, may have 1-15, 1-12, 1-10, 1-8 or 1-6 residues etc. The nature of the residues is not critical and they may for example be any amino acid, e.g. a neutral amino acid, or an aliphatic amino acid, or alternatively they may be hydrophobic, or polar or charged or structure-forming e.g. proline. In some preferred embodiments, the linker is a serine and/or glycine-rich sequence.

Exemplary spacer sequences thus include any single amino acid residue, e.g. S, G, L, V, P, R, H, M, A or E or a di-, tri- tetra- penta- or hexa-peptide composed of one or more of such residues.

Thus, in some embodiments, the invention provides a recombinant or synthetic polypeptide comprising the polypeptide or peptide of the invention as defined above, i.e. a recombinant or synthetic polypeptide comprising a peptide or polypeptide (e.g. a heterologous peptide or polypeptide, i.e. a peptide or polypeptide that is not normally associated with the peptide or polypeptide of the invention, e.g. from a different organism) fused to a polypeptide or peptide of the invention. Thus, the recombinant or synthetic polypeptide may optionally comprise a spacer as defined above.

The recombinant or synthetic polypeptide of the invention may also comprise purification moieties or tags to facilitate their purification (e.g. prior to use in the methods and uses provided herein). Any suitable purification moiety or tag may be incorporated into the polypeptide and such moieties are well known in the art. For instance, in some embodiments, the recombinant or synthetic polypeptide may comprise a peptide purification tag or moiety, e.g. a His-tag or C-tag sequence. Such purification moieties or tags may be incorporated at any position within the polypeptide. In some preferred embodiments, the purification moiety is located at or towards (i.e. within 5, 10, 15, 20 amino acids of) the N- or C-terminus of the polypeptide.

As noted above, an advantage of the present invention arises from the fact that the polypeptide or peptide tag of the invention incorporated in a peptide, protein or polypeptide (e.g. the recombinant or synthetic polypeptides of the invention) may

be completely genetically encoded. Thus, in a further aspect, the invention provides a nucleic acid molecule encoding a peptide tag, polypeptide or recombinant or synthetic polypeptide as defined above.

The nucleic acid molecules of the invention may be made up of  
5 ribonucleotides and/or deoxyribonucleotides as well as synthetic residues, e.g. synthetic nucleotides, that are capable of participating in Watson-Crick type or analogous base pair interactions. Preferably, the nucleic acid molecule is DNA or RNA.

The nucleic acid molecules described above may be operatively linked to an  
10 expression control sequence, or a recombinant DNA cloning vehicle or vector containing such a recombinant DNA molecule. This allows cellular expression of the polypeptides and peptides of the invention as a gene product, the expression of which is directed by the gene(s) introduced into cells of interest. Gene expression is directed from a promoter active in the cells of interest and may be inserted in any  
15 form of linear or circular nucleic acid (e.g. DNA) vector for incorporation in the genome or for independent replication or transient transfection/expression. Suitable transformation or transfection techniques are well described in the literature. Alternatively, the naked nucleic acid (e.g. DNA or RNA, which may include one or more synthetic residues, e.g. base analogues) molecule may be  
20 introduced directly into the cell for the production of peptides and polypeptides of the invention. Alternatively, the nucleic acid may be converted to mRNA by *in vitro* transcription and the relevant proteins may be generated by *in vitro* translation.

Appropriate expression vectors include appropriate control sequences such as for example translational (e.g. start and stop codons, ribosomal binding sites)  
25 and transcriptional control elements (e.g. promoter-operator regions, termination stop sequences) linked in matching reading frame with the nucleic acid molecules of the invention. Appropriate vectors may include plasmids and viruses (including both bacteriophage and eukaryotic viruses). Suitable viral vectors include baculovirus and also adenovirus, adeno-associated virus, herpes and vaccinia/pox  
30 viruses. Many other viral vectors are described in the art. Examples of suitable vectors include bacterial and mammalian expression vectors pGEX-KG, pEF-neo and pEF-HA.

As noted above, the recombinant or synthetic polypeptide of the invention may comprise additional sequences (e.g. peptide/polypeptides tags to facilitate  
35 purification of the polypeptide) and thus the nucleic acid molecule may conveniently

be fused with DNA encoding an additional peptide or polypeptide, e.g. His-tag, maltose-binding protein, to produce a fusion protein on expression.

Thus, viewed from a further aspect, the present invention provides a vector comprising a nucleic acid molecule as defined above.

5 Nucleic acid molecules of the invention, preferably contained in a vector, may be introduced into a cell by any appropriate means. Suitable transformation or transfection techniques are well described in the literature. Numerous techniques are known and may be used to introduce such vectors into prokaryotic or eukaryotic cells for expression. Preferred host cells for this purpose include insect cell lines,  
10 yeast cells (e.g. *Pichia pastoris* (also known as *Komagataella phaffii*)), mammalian cell lines or *E. coli*, such as strain BL21/DE3.

By "recombinant" is meant that the nucleic acid molecule and/or vector has been introduced into the host cell. The host cell may or may not naturally contain an endogenous copy of the nucleic acid molecule, but it is recombinant in that an  
15 exogenous or further endogenous copy of the nucleic acid molecule and/or vector has been introduced. Thus, a recombinant polypeptide refers to a polypeptide encoded by a recombinant nucleic acid molecule or vector.

In some embodiments, the polypeptide or peptide of the invention, or for use in the method and uses of the invention, may be generated synthetically, e.g. by  
20 ligation of amino acids or smaller synthetically generated peptides, or more conveniently by recombinant expression of a nucleic acid molecule encoding said polypeptide as described hereinbefore.

Nucleic acid molecules of the invention may be generated synthetically by any suitable means known in the art.

25 Thus, the polypeptide or peptide tag of the invention may be an isolated, purified, recombinant or synthesised peptide or polypeptide.

Similarly, the nucleic acid molecules of the invention may be an isolated, purified, recombinant or synthesised nucleic acid molecule.

Thus, alternatively viewed, the peptide tag, polypeptides and nucleic acid  
30 molecules of the invention preferably are non-native, i.e. non-naturally occurring, molecules.

The term "polypeptide" is used herein interchangeably with the term "protein". As noted above, the term polypeptide or protein typically includes any amino acid sequence comprising at least 40 consecutive amino acid residues, e.g.

at least 50, 60, 70, 80, 90, 100, 150 amino acids, such as 40-1000, 50-900, 60-800, 70-700, 80-600, 90-500, 100-400 amino acids.

Whilst it is envisaged that the polypeptide and peptide provided herein may be produced recombinantly, and this is a preferred embodiment of the invention, it will be evident that the polypeptide or peptide may be conjugated to proteins or other entities, e.g. molecules or components, as defined above by other means. In other words, the polypeptide or peptide and other molecule, component or entity, e.g. protein, may be produced separately by any suitable means, e.g. recombinantly, and subsequently conjugated (joined) to form a peptide tag-other component conjugate or polypeptide-other component conjugate that can be used in the methods and uses of the invention. For instance, the polypeptide or peptide of the invention may be produced synthetically or recombinantly, as described above, and conjugated to another component, e.g. a protein via a non-peptide linker or spacer, e.g. a chemical linker or spacer.

Thus, the polypeptide or peptide and other component, e.g. protein, may be joined together either directly through a bond or indirectly through a linking group. Where linking groups are employed, such groups may be chosen to provide for covalent attachment of the polypeptide or peptide and other entity, e.g. protein, through the linking group. Suitable linking groups are defined above. It will be evident that the polypeptide and peptide tag have a wide range of utilities. Alternatively viewed, the polypeptide and peptide tag of the invention may be employed in a variety of industries.

For instance, as detailed above, the polypeptide and peptide tag find utility in a protein enrichment method for an HDX-MS workflow (i.e. a covalent capture system). In this respect, the protein of interest (POI) for analysis by HDX-MS may be provided as a recombinant polypeptide comprising the peptide tag as a domain of the protein, e.g. an N- or C-terminal domain. This enables cellular expression of the POI meaning that the HD exchange can be performed on the POI in its native setting, where it may interact with other molecules, e.g. proteins, that will impact on the hydrogen atoms available for exchange. Following HD exchange on the cell, the cell can be lysed under quench conditions and the lysate (i.e. sample) containing the recombination POI (i.e. LemonTag-POI fusion protein) may be contacted with the polypeptide of the invention (e.g. SEQ ID NO: 1) immobilised on a solid support under HDX quench conditions, which are suitable to allow the spontaneous formation of an isopeptide bond between the polypeptide and peptide tag. This

results in the covalent capture of the POI on the solid support under quench conditions, which can be isolated and processed for analysis, e.g. mass spectrometry analysis.

Thus, it will be appreciated that the invention provides the use of:

5

- (i) a polypeptide as defined herein;
- (ii) a peptide as defined herein; and/or
- (iii) a recombinant polypeptide as defined herein,

in a protein analysis method, optionally where in the protein analysis method uses mass spectrometry, particularly Hydrogen-Deuterium eXchange (HDX) mass spectrometry.

10

More particularly, the invention provides a process or method for analysing a polypeptide of interest (POI) comprising:

15

(i) providing a sample comprising the POI linked to the peptide as defined herein;

(ii) subjecting the sample to hydrogen-deuterium exchange to provide a sample comprising a deuterium-labelled POI linked to the peptide as defined herein;

20

(iii) contacting the sample from (ii) with a polypeptide as defined herein under conditions that enable the spontaneous formation of an isopeptide bond between the polypeptide and peptide linked to the POI, thereby conjugating the POI to the polypeptide via an isopeptide bond to form a complex, wherein the conditions are also HDX quench conditions (i.e. conditions that slow or minimise the H/D exchange in the sample);

25

(iv) isolating the complex under HDX quench conditions; and

(v) analysing the complex, e.g. using mass spectrometry.

Thus, the sample may be a cell comprising or expressing the POI linked to the peptide, wherein the cell may be any cell as defined herein.

The step of subjecting the sample to hydrogen-deuterium exchange may be performed using any suitable means known in the art.

30

The step of contacting the sample with a polypeptide may involve a step of processing the sample to enable the polypeptide to interact with the peptide in the POI. For instance, the step may involve lysing the cell containing the POI linked to the peptide. The lysis step may be separate from the step of contacting the sample with the polypeptide or the lysis and contacting steps may be performed

35

simultaneously, e.g. the lysis buffer may contain the polypeptide. The lysis step

may involve contacting the cell with a lysis buffer comprising Polymyxin B (e.g. 20-80 µg/mL, such as about 50 µg/mL Polymyxin B sulfate salt), melittin from bee venom (e.g. 5-20 µM, such as about 10 µM Melittin (≥85% purity)) and Octyl β-D-glucopyranoside (OG) (e.g. 1-5% w/v, such as about 2% w/v). The pH of the lysis buffer may be about 2.0-4.0, such as about 2.5.

The step of isolating the complex may involve isolating the solid support on which the polypeptide is immobilised as described above. However, it will be understood that the polypeptide may not be immobilised when it is contacted with the sample. For example, the polypeptide may comprise a moiety, e.g. purification tag, that facilitates the isolation of the complex from the sample. As discussed above, the step of isolating the complex may involve a step of releasing the complex from a solid support, e.g. cleaving a covalent bond linking the polypeptide to the solid support, such as contacting the solid support with a reducing agent, e.g. TCEP or DTT.

The step of analysing the complex may use any suitable means and may involve additional steps, such as processing steps. For instance, the complex may be processed using enzymatic digestion, e.g. using pepsin, for analysis by mass spectrometry. It will be understood that the step of releasing the complex from a solid support and processing the complex may be performed simultaneously. In particular, the complex may be released by proteolytic digestion (e.g. using an acidic protease, such as pepsin), thereby obviating the requirement for separate release and processing steps. Alternatively viewed, the processing step may be performed on the solid support. Thus, the step of isolating the complex may not involve a releasing step. Alternatively viewed, the releasing step may form part of analysing the complex, e.g. where the releasing and processing steps are performed simultaneously.

While the present invention is particularly useful in workflows such as HDX-MS, it will be appreciated by the skilled person that other workflows in which the present invention may find particular utility may include processes that involve an acidic pH and/or low temperature. Such processes may use an acidic pH to inhibit the activity of proteases or other modifying enzymes. Alternatively, the invention may be particularly applicable to protocols utilised to analyse gastric function. Some further representative examples of processes in which the invention may find utility are presented below.

For instance, the peptide and polypeptide may find utility in targeting fluorescent or other biophysical probes or labels to specific proteins. In this respect, a protein of interest (POI) may be modified to incorporate the peptide tag and the fluorescent or other biophysical probe or label may be fused or conjugated to the polypeptide (e.g. SEQ ID NO: 1 or 13). The modified protein and probe or label may be contacted together under conditions suitable to allow the spontaneous formation of an isopeptide bond between the peptide tag and polypeptide, thereby labelling the protein with the label or probe via an isopeptide bond. For instance, the labelled polypeptide of the invention may find utility in an antibody-free Western blot, i.e. where the labelled polypeptide is used to detect a polypeptide containing the peptide provided herein without the need for a separate labelled antibody.

The peptide tag and polypeptide may find utility in protein immobilisation for proteomics. In this respect, a POI may be modified to incorporate a peptide tag and the polypeptide may be immobilised on a solid substrate. The modified protein and solid substrate may be contacted together under conditions suitable to allow the spontaneous formation of an isopeptide bond between the peptide tag and polypeptide, thereby immobilising the proteins on the solid substrate via an isopeptide bond. It will be evident that the peptide tag may be used to simultaneously immobilise multiple proteins on a solid phase/substrate, i.e. in a multiplex reaction.

In a further representative example, the peptide tag and polypeptide may find utility in conjugation of antigens to virus-like particles, viruses, bacteria or multimerisation scaffolds for vaccination. For instance, the production of virus-like particles, viruses or bacteria that display the polypeptide on the surface would facilitate the conjugation of antigens comprising the peptide tag to their surface via an isopeptide bond. In this respect, antigen multimerisation gives rise to greatly enhanced immune responses. Thus, the molecule or component fused to the polypeptide or peptide may be a viral protein, e.g. a viral capsid protein, and/or the molecule or component fused to the peptide or polypeptide may be an antigen, e.g. an antigen associated with a particular disease, e.g. infection, an autoimmune disease, allergy or cancer.

The peptide tag and polypeptide may be used to cyclise an enzyme, e.g. by fusing the peptide tag and polypeptide to each end of the enzyme and subsequently allowing the spontaneous formation of the isopeptide bond between the peptide tag

and polypeptide. In this respect, cyclisation of enzymes has been shown to increase enzyme resilience.

In particular, cyclisation of enzymes or enzyme polymers (fusion proteins) may improve the thermostability of the protein or protein units in the enzyme polymer. In this respect, enzymes are valuable tools in many processes but are  
5 unstable and hard to recover. Enzyme polymers have greater stability to temperature, pH and organic solvents and there is an increased desire to use enzyme polymers in industrial processes. However, enzyme polymer generation commonly uses a glutaraldehyde non-specific reaction and this will damage or  
10 denature (i.e. reduce the activity of) many potentially useful enzymes. Site-specific linkage of proteins into chains (polymers) through isopeptide bonds using the peptide tag and polypeptide of the present invention is expected to enhance enzyme resilience, such as in diagnostics or enzymes added to animal feed.

The peptide tag and polypeptide provided herein also could be used to link  
15 multiple enzymes into pathways to promote metabolic efficiency, as described in WO 2016/193746. In this respect, enzymes often come together to function in pathways inside cells and traditionally it has been difficult to connect multiple enzymes together outside cells (*in vitro*). Thus, the peptide tag and polypeptide could be used to couple or conjugate enzymes to produce fusion proteins and  
20 therefore enhance activity of multi-step enzyme pathways, which could be useful in a range of industrial conversions and for diagnostics.

The peptide tag and polypeptide also may find utility in the production of antibody polymers. In this respect, antibodies are one of the most important class of pharmaceuticals and are often used attached to surfaces. However, antigen mixing  
25 in a sample, and therefore capture of said antigen in said sample, is inefficient near surfaces. By extending chains of antibodies, it is anticipated that capture efficiency will be improved. This will be especially valuable in circulating tumour cell isolation, which at present is one of the most promising ways to enable early cancer diagnosis.

In another example, the peptide tag and polypeptide may find utility in the  
30 production of drugs for activating cell signalling. In this respect, many of the most effective ways to activate cellular function are through protein ligands. However, in nature a protein ligand will usually not operate alone but with a specific combination of other signalling molecules. Thus, the peptide tag and polypeptide allows the  
35 generation of tailored fusion proteins (i.e. protein teams), which could give optimal

activation of cellular signalling. These fusion proteins (protein teams) might be applied for controlling cell survival, division, or differentiation.

In yet a further example, the peptide tag and polypeptide may find utility in the generation of hydrogels for growth of eukaryotic cells, e.g. neurons, stem cells,  
5 preparation of biomaterials, antibody functionalisation with dyes or enzymes and stabilising enzymes by cyclisation.

The invention will now be described in more detail in the following non-limiting Examples with reference to the following drawings in which:

Figure 1 shows SDS-PAGE gels visualised with Coomassie staining  
10 showing the reactivity of SpyTag variants (SpyTag, SpyTag002 or SpyTag003 linked to maltose-binding protein, MBP) with (A) SpyCatcher002 and (B) SpyCatcher003. 10  $\mu$ M of each SpyTag variant was incubated with 10  $\mu$ M of SpyCatcher002 or SpyCatcher003 for 30 minutes in quenched PBS at pH 2.5 and 0 °C; and (C) shows a graph depicting the reaction rate of the best reacting pair  
15 (SpyTag002:SpyCatcher003) at 0 °C within quenched PBS + 5 mM TCEP at pH 2.5. The data shown in (C) is the mean reconstitution product of reactions carried out in triplicate  $\pm$  1 s.d.

Figure 2 shows (A) a bar chart depicting the reconstitution product efficiency of SpyTag002:SpyCatcher003 tested within commonly used HDX quench solutions  
20 (PBS,  $\text{K}_2\text{HPO}_4$  and Glycine) that buffer at pH 2.5 performed at 0 °C for 15 minutes using 5  $\mu$ M each of SpyTag002-MBP and SpyCatcher003; and (B) a bar chart depicting the influence of additives (glycerol, NaCl and sucrose) on reconstitution product efficiency relative to averaged reconstitution product achieved in standard PBS conditions. The addition % (v/v) of glycerol, mM of NaCl, or % (w/v) of sucrose  
25 with PBS pH 2.5 is indicated on the x-axis. The data shown are the mean of reactions carried out in duplicate, with each data point shown as a square and the bar representing the mean.

Figure 3 shows SDS-PAGE gels visualised with Coomassie staining under reductive conditions showing (A) the reaction products of SpyTag002-MBP and  
30 rationally designed SpyCatcher003 variants (R32T, R37N, R47T, K52E, H62Q, K72E and H11Q) compared to previous tests with SpyCatcher003 WT; (B) the improved reactivity of SpyCatcher003 following the introduction of a hydrophobic residue (R32V) compared to R32T; (C) the reaction product of LemonCatcher (SEQ ID NO: 1) and rational design SpyTag002-MBP mutants (K120E, R121T, K120E-  
35 R121T – this numbering refers to the amino acid positions in the protein from which

the peptide is derived and corresponds to K12E, R13T and K12E-R13T using the numbering of SEQ ID NO: 26); and (D) the reduced amidation efficacy of SpyTag002-MBP following deletion of the KRYK C-terminus (KRYK del). In all experiments, 5  $\mu$ M of SpyTag002-MBP, SpyCatcher003 mutants and  
5 LemonCatcher were incubated for 15 minutes with their respective partner at 0 °C within quenched PBS + 5 mM TCEP at pH 2.5. The SDS-PAGE gels shown are representative of reactions carried out in triplicate.

Figure 4 shows an SDS-PAGE gel visualised with Coomassie staining under reductive conditions showing (A) the reactivity of the individual single-point mutants combined in the R32V/S59T/V94I SpyCatcher003 construct (SEQ ID NO: 10) used  
10 as a template for directed evolution. 1  $\mu$ M of SpyTag002-MBP and the SpyCatcher003 mutants were incubated for 15 minutes at 0 °C within quenched PBS + 5 mM TCEP at pH 2.5. The SDS-PAGE gel is the representative of reactions carried out in triplicate; and (B) shows the reaction efficiency of the SpyCatcher003  
15 triple mutant (SpyCatcher003 TM, SEQ ID NO: 10) to SpyTag002 compared to the original best reacting pair (SpyTag002/SpyCatcher003). 1  $\mu$ M of SpyTag002-MBP and Catchers were incubated for 15 minutes at 0 °C within quenched PBS + 5 mM TCEP at pH 2.5. The data shown is the mean reconstitution product of reactions carried out in triplicate  $\pm$  1 s.d.

Figure 5 shows SDS-PAGE gels showing the reactivity of various  
20 Tag/Catcher pairs in cold and acidic conditions: (A) SpyTag002:SpyCatcher003; (B) SpyTag002:SpyCatcher003 TM; (C) SpyTag002:LemonCatcher; (D-E) LemonTag:LemonCatcher; and (F) shows the reactivity of the individual triple mutant alone (TM), or with additional point mutations identified in the phage display  
25 screen. R121P refers to LemonTag, which contains a R121P mutation (using the numbering from the protein from which the peptide is derived). 1  $\mu$ M of SpyTag002-MBP, LemonTag and Catchers were incubated at 0 °C within quenched PBS + 5 mM TCEP at pH 2.5 (A-D and F) or 3.0 (E). The SDS-PAGE gel shown is the representative of reactions carried out in triplicate.

Figure 6 shows (A) a bar chart demonstrating that the M13K07 helper phage  
30 incubated at pH 2.5 or 7.5 at 4 °C, before titering by *E. coli* infection, remains infective for up to 8 hours; and (B) a bar chart depicting the results of model selection for phage panning, where phage displaying wildtype, E77A or SpyCatcher003 triple mutant (SEQ ID NO: 10) were incubated with biotinylated  
35 SpyTag002-MBP and pulled down on streptavidin-beads, before quantifying the

number of plaque-forming units (pfu) per mL presented (mean represented by the bar, with each data-point shown with a square,  $n = 2$ ).

Figure 7 shows (A) the reconstitution of LemonCatcher (SEQ ID NO: 1) and LemonTag (SEQ ID NO: 4) linked to MBP at 1  $\mu$ M each analysed by SDS-PAGE with Coomassie staining; and the percentage of product formed at 0 °C in quenched PBS sample buffer + 5 mM TCEP of (B) LemonTag/LemonCatcher and SpyTag002/SpyCatcher003 at 10  $\mu$ M; and (C) LemonTag/LemonCatcher, SpyTag002/LemonCatcher and SpyTag002/SpyCatcher003 at 1  $\mu$ M. The data shown are the mean reconstitution product of reactions carried out in triplicate  $\pm$  1 s.d.

Figure 8 shows (A) the reactivity of 1  $\mu$ M LemonTag and LemonCatcher in lemon juice supplemented with 2.5 mM TCEP at  $\sim$ pH 2-3 and at 0 °C or 25 °C for 1, 2 or 4 hours compared to reactivity quenched PBS + 5 mM TCEP at pH 2.5 or 3.0 and at 25 °C for 1 hour by SDS-PAGE with Coomassie staining; and (B) the pH-dependent reconstitution of LemonCatcher/LemonTag compared to SpyTag002/SpyCatcher003. Tags and Catchers were incubated at 0 °C in quenched or unquenched PBS sample buffer + 5 mM TCEP (pH 2.0, 2.5, 3.0, 5.0 and 7.5). Data are the mean reconstitution product  $\pm$ 1 s.d.,  $n = 3$ .

Figure 9 shows the differential scanning calorimetry (DSC) analysis of: (A) LemonCatcher (SEQ ID NO: 1); (B-C) SpyCatcher003 and SpyCatcher003 triple mutant (SEQ ID NO: 10) at pH 2.5 and 7.5; (D) LemonTag (SEQ ID NO: 4) and LemonCatcher; and (E) SpyTag002 and LemonCatcher, performed in quenched or unquenched PBS sample buffer + 5 mM TCEP at a range of pH values and temperatures. The melting temperature ( $T_m$ ) and Full Width Half Maximum (FWHM) are tabulated below each graph.

Figure 10 shows (A) the results of chemical cell lysis optimisation experiments under HDX-MS quench conditions (acidic pH and low temperature) to determine the lysis efficacy of select additives (Melittin, polymyxin B or ZrO<sub>2</sub> beads) and detergents (Octyl  $\beta$ -D-glucopyranoside, OG; n-Dodecyl-beta-Maltoside, DDM; and n-Dodecyl-phosphocholine, FC-12); (B) the efficacy of *in vivo* chemical lysis utilising an optimised chemical cocktail of 2% OG, 10  $\mu$ M Melittin, 50  $\mu$ g/mL Polymyxin B in 0.5 $\times$  PBS, 50 mM NaH<sub>2</sub>PO<sub>4</sub> at pH 2.5. Lysis was assessed by Western blot using an anti-His antibody against AcrB-His<sub>6</sub> (left, end-point) or His<sub>6</sub>-LemonTag-MBP (right, time-course). Calculations to determine the relative percentage of protein found in filtered lysate compared to whole cell was performed using ImageJ densitometry analysis of

the dot spots. (C) shows a bar chart depicting the results of a bead-based assay utilised to capture LemonTag-MBP (LT-MBP) captured from *E. coli* cell lysate via LemonCatcher beads. The graph depicts the number of identified peptides for LemonCatcher (SEQ ID NO: 1) (light bars) and LemonTag-MBP (including sequence coverage) (dark bars) relative to standard measurements of purified LemonTag-MBP, negative control (cell lysate without LemonTag-MBP), and cell lysate (cell lysate of *E. coli* C41(DE3) with overexpressed LemonTag-MBP).

Figure 11 shows SDS-PAGE gels visualised with Coomassie staining under reductive conditions showing the reactivity between LemonTag-MBP and LemonCatcher present at 1  $\mu$ M and reacted on ice or at -20 °C, for the time points reported, in quenched PBS sample buffer containing 40% ethylene glycol (v/v) + 5 mM TCEP at pH 2.5 (A) or pH 3.0 (B).

Figure 12 shows SDS-PAGE gels visualised with Coomassie staining under reductive conditions showing the reactivity LemonTag-MBP and LemonCatcher present at 1  $\mu$ M and reacted on ice or at -20 °C, for the time points reported, in quenched PBS sample buffer containing 30% methanol (v/v) + 5 mM TCEP at pH 2.5.

## EXAMPLES

20 Example 1 – Development of a two-part linker system that functions in low pH and low temperature conditions

HDX-MS is a valuable technique used to study protein dynamics and provides valuable information regarding protein conformation and interaction. HDX-MS also enables the study of protein dynamics within cells and related complex environments. However, thus far there have been a lack of developments to existing workflows which allow for selective examination of proteins post-HDX. Selective protein enrichment post-HDX would satisfy this need, but the extreme conditions required for the HDX quench to preserve the D-label (pH 2.3-3.0 and 0-4 °C) greatly limits the enrichment methods that can be utilised because many existing purification methods do not work under these conditions. It was hypothesised that the existing SpyTag/SpyCatcher technology could be adapted for use under these extreme conditions.

First, experiments were performed to determine if any of the existing SpyTag/SpyCatcher pairs (original, 002 or 003; see WO2011/098772,

WO2018/197854 and WO2020/183198 herein incorporated by reference) were able to react in the acidic and cold conditions required to limit D-to-H back-exchange during HDX-MS.

5 The reconstitution levels of the original, 002 and 003 SpyTag/SpyCatcher constructs were measured by SDS-PAGE analysis of the covalent complex within quenched PBS sample buffer at pH 2.5 and 0 °C (**Figure 1A-B**). The reactivity between different permutations of the SpyTag/SpyCatcher constructs (e.g., the reactivity of SpyTag with SpyCatcher002) were investigated. It was found that the highest reconstitution level under HDX quench conditions was observed for  
10 SpyCatcher003 in combination with SpyTag002 (**Figure 1C**). Given the excellent reactivity previously observed between SpyTag003 and SpyCatcher003 in neutral pH and physiological temperature conditions, the results observed were unexpected. SpyTag002 (SEQ ID NO: 26) and SpyCatcher003 (SEQ ID NO: 27) were chosen as the basis for further experiments.

15 As the presence of additives has previously been demonstrated to impact the reactivity between peptides and proteins, several commonly used HDX quench reaction conditions were screened to determine if the presence of certain buffers or additives would beneficially impact upon the rate of reactivity of SpyTag002 and SpyCatcher003 (**Figure 2**).

20 It was found that low levels of glycerol (5-15% v/v) or sucrose (5-15% w/v) enhanced SpyTag002/SpyCatcher003 product formation, although their benefit was ablated at higher concentrations (25-30%). Moreover, it was found that the addition of sodium chloride beyond its standard concentration in PBS buffer (137 mM) decreased reactivity. A combination of these optimised buffer conditions was only  
25 able to provide a system with a reconstitution efficiency of  $6 \pm 1\%$ , even when the reaction progressed longer than the 15-minute reaction time desired (120 min at 1  $\mu\text{M}$ ) (**Figure 1C**). Therefore, further engineering was required to establish a robust solution for protein enrichment under HDX-MS conditions.

### 30 Example 2 – Rational mutation of SpyTag002 and SpyCatcher003

It was attempted to enhance the reactivity of SpyTag002 and SpyCatcher003 within the extreme conditions of an HDX quench using rational mutations of the sequences.

35 It is well-known that the net charge of a protein is dependent on its local environment. The isoelectric point (pI) of a protein is the point at which the net

charge of a protein is zero, as the positive and negative charges of the amino acid residues are balanced. When in a solution with a pH lower than its pI, the surface of a protein is predominantly positively charged resulting in electrostatic repulsion which can result in reduced protein stability. As HDX-MS quench conditions require a pH of 2-3, it was hypothesised that a high degree of protonation of the carboxylates from Asp/Glu residues may be causing destabilisation of both binding partners and reducing reactivity. Accordingly, substitution of positively charged amino acids with uncharged amino acids was sought as a means to reduce the number of positively charged groups. Eight residues in SpyCatcher003 (K28, R32, R37, R47, K52, H62, K72 and H112) and two residues in SpyTag002 (K11 and R12) were selected for mutation.

It was found that mutation of R32, the residue next to the reactive K31 in SpyCatcher003 led to enhanced reactivity with SpyTag002 (~20-25% within 120 minutes). In initial experiments, a mutation from arginine to threonine at position 32 (R32T) was investigated as it is known that threonine is well tolerated in  $\beta$ -strands. While this mutation provided improved reactivity with SpyCatcher003 (**Figure 3A**), the introduction of the hydrophobic residue valine at position 32 (R32V, SEQ ID NO: 5) further improved reactivity ( $52 \pm 4\%$  compared to R32T; **Figure 3B**).

Although the mutation of arginine at position 32 to threonine or valine (R32T or R32V) successfully improved reactivity, the mutation of other amino acids likely to be positively charged at pH 2-3 within SpyCatcher003 (K28I, R37N, R47T, K52E, H62Q, K72E, and H112Q) and SpyTag002 (K120E, R121T, K120E-R121T) did not result in improved reactivity (**Figure 3A and C**).

In addition to these findings, it was determined that the deletion of the positively charged KRYK (SEQ ID NO: 30) at the C-terminus of SpyTag002 led to striking reduction in amidation efficiency, thereby preventing isopeptide bond formation (**Figure 3D**).

These data support the conclusion that a strategy beyond rational positive charge deletion or substitution was required for development of a Tag/Catcher system which works in at pH 2.3-3.0 and 0-4 °C.

### Example 3 – Directed evolution of SpyTag002 (SEQ ID NO: 26) and SpyCatcher003 (SEQ ID NO: 27)

A phage library based on SpyCatcher003 was generated and panned for covalent bond formation under acidic and cold conditions. From this library, two

additional mutations which improved reactivity were discovered: S59T and V94I ( $39 \pm 3\%$  and  $21 \pm 6\%$  respectively; **Figure 4A**). These two mutations were combined with the R32V mutation described in Example 2 to provide a triple mutant SpyCatcher003 construct (R32V/S59T/V94I; SEQ ID NO: 10) which was found to greatly improve reactivity with SpyTag002 (**Figure 4B** and **5F**).

Using SpyCatcher003 R32V/S59T/V94I (SEQ ID NO: 10) as a template (also known as SpyCatcher003 TM (triple mutant)), a further round of phage display with multiple rounds of selection in HDX quench conditions was conducted (**Figure 6A-B**). These experiments identified two further mutations (T56A and N95T) which improved the amidation rate of SpyCatcher003 R32V/S59T/V94I (SEQ ID NO: 10) (**Figure 5F** and **Table 1**). The combination of the five identified mutations (R32V/T56A/S59T/V94I/N95T; SEQ ID NO: 13) resulted in a greatly improved rate of reactivity with SpyTag002 (**Figure 5F**).

**Table 1:** The amino acid sequences of selected clones from the SpyCatcher003 triple mutant library selections.

Clone	Mutations relative to R32V/S59T/V94I SpyCatcher003				
L1R2C5	D16H	D22E			E108V
L1R2C6					E108G
L1R3C4					E108D
L2R3C10				D97V	
L1R4C4			V76D	D97E	E108G
L1R4C7		Y69H		D97E	G104D
L1R4C11	T56A			N95T	
L1R4C13				N95S	
	Additions to R32V/S59T/V94I SpyCatcher003 tested				
Mutation	T56A	V76I	N95T	D97V	E108G
Rxn effect	↑	---	↑	---	---

The reactivity of SpyTag002 with the R32V/T56A/S59T/V94I/N95T SpyCatcher003 mutant (SEQ ID NO: 13) was tested and it was found that the mutation R12P provided marked improvement in reactivity ( $20 \pm 8\%$ ) (**Figure 5F** and **Figure 7**).

Therefore, a combination of rational design and phage selection was required to develop a protein linker system adapted to the extreme conditions of HDX quench. The peptide and polypeptide were termed LemonTag (SEQ ID NO: 4) and LemonCatcher (SEQ ID NO: 1 and 13), respectively, as this pairing reacts

within lemon juice which has a matching pH of 2-3 (**Figure 8A**). As discussed below, the polypeptide can be further mutated to introduce a cysteine residue to facilitate the directed coupling of the polypeptide to a solid support without affecting its activity. Accordingly, both SEQ ID NO: 1 and SEQ ID NO: 13 may be termed  
5 LemonCatcher polypeptides.

Example 4 – Validation of LemonTag and LemonCatcher across range of conditions

In view of LemonTag/LemonCatcher's excellent reactivity under acidic and cold conditions, the system was assessed further to determine if the pair exhibited improved reactivity in less extreme conditions, i.e. across a wider range of pH values and temperatures. Whilst the activity of LemonTag/LemonCatcher in HDX quench conditions is highly desirable, it should be noted that applicability of this system to a range of processes or systems, which may not necessarily require  
10 acidic conditions and/or low temperatures, is also desirable.

It was found that LemonTag/LemonCatcher retained reactivity across a range of acidic and neutral pH values (pH 2-7.5). This is particularly desirable within the field of HDX-MS because it is beneficial for the purification system utilised to have functionality from pH 2.0 to 3.0 as it has been found that quenching up to pH  
15 3.0 is beneficial within some workflows. Surprisingly, it was determined that LemonTag/LemonCatcher reacted at an accelerated rate at pH 7.5 compared to the SpyTag002/SpyCatcher003 pair (**Figure 8B**).

Differential scanning calorimetry (DSC) revealed that the melting temperatures for LemonTag/LemonCatcher compared to SpyTag002/SpyCatcher003 are similar at both pH 2.5 and 7.5 ( $T_m$ 's of 51.2-56.1 °C), which suggests that effects of acidic and neutral contexts on potential Catcher shelf-life could be modest (**Figure 9A-C**). Nonetheless, the full width at half maximum (FWHM) values are wider within acid conditions, which suggests less cooperative protein unfolding at low pH. As seen for SpyCatcher003, the reaction of LemonCatcher with either LemonTag (**Figure 9D**) or SpyTag002 (**Figure 9E**) drastically improved thermal stability: up to 98.4 °C  
25 (LemonTag) and 98.9 °C (SpyTag002) at pH 2.5, and 105.3 °C (LemonTag) at pH 3.0. Such stabilisation upon Catcher and Tag reaction is thought to be important thermodynamically in driving isopeptide bond formation towards completion.

Example 5 - Assessing the reactivity of LemonTag/LemonCatcher at extremely low temperatures

The ability of LemonTag/LemonCatcher to react at extremely low temperatures was assessed by adding cryoprotectants to the buffer.

5 LemonCatcher and LemonTag-MBP were buffer exchanged to sample buffer (1xPBS + 10% glycerol + 40% ethylene glycol or 30% methanol (v/v), pH 7.5) and diluted to 4  $\mu$ M. Proteins and quench buffer (100 mM  $\text{NaH}_2\text{PO}_4$  + 40% ethylene glycol or 30% methanol (v/v) + 10 mM TCEP pH 2.35 or pH 2.81 for final pH of 2.5 or 3.0, respectively) were equilibrated for 2 hrs at -20 °C. LemonCatcher and  
10 LemonTag-MBP were quenched 1:1 in quench buffer and kept for further 15 min at -20 °C. Reaction was started by 1:1 mix of LemonCatcher and LemonTag-MBP in pre-cooled (-20 °C) tubes and incubated at -20 °C for various time points. Reaction was stopped by adding 20  $\mu$ L reaction volume to 5  $\mu$ L 5x Laemmli buffer (pre-heated to -95 °C) and incubation for 5 min at 95 °C. SDS-PAGE was performed on  
15 NuPAGE 12% Bis-Tris protein gels.

The addition of ethylene glycol or methanol to the reaction buffer prevented the buffer from freezing and, as shown in **Figures 11 and 12**, LemonTag and LemonCatcher were able to react even at temperatures as low as -20°C and in acidic conditions (pH 2.5, **Figures 11A and 12**, and pH3.0, **Figure 11B**).

20

Example 6 – Protein purification following HDX-MS utilising LemonTag/LemonCatcher

While the efficacy of LemonTag/LemonCatcher was demonstrated under the extreme conditions of an HDX quench, full integration into an HDX-MS workflow  
25 requires a suitable enrichment strategy. HDX-MS analysis typically relies on short digestion, peptide trapping and chromatography to limit D-to-H back exchange (<15-30 minutes) which, in case of a complex sample background, increases peptide co-elution and consequently hampers the identification and analysis of the target protein/peptides.

30 The LemonCatcher construct contains a unique cystine residue (S49C), which allows the Catcher to be bound to a suitable structure through formation of a disulfide bridge. Utilising this cysteine residue, it was established that a protein of interest (POI), fused with LemonTag, could be pulled down by LemonCatcher coupled to CarboxyLink resin (diamino-dipropylamine activated, crosslinked beaded agarose) by a Sulfo-LC-SPDP linker. Whilst the bead-captured POI can be digested  
35

in solution using acid proteases (e.g. pepsin), a strategy was devised that enables elution of the entire complex (LemonCatcher:LemonTag-POI) with tris(2-carboxyethyl)phosphine (TCEP), a common quench additive that is tolerated at high concentrations by pepsin (up to 100-200 mM). Release of the POI permits its digestion by immobilised protease systems commonly used to improve digestion efficiency, reproducibility, throughput, and reduce background from autolytic products that are prevalent with typical solution protease digests.

The utility of this adapted workflow was demonstrated utilising maltose-binding protein (MBP) equipped with LemonTag by capturing the LemonTag-MBP complex from a pre-purified solution at different concentrations to determine the limit of detection. This workflow required a pH of 3.0 to reliably capture LemonTag-MBP utilising the LemonCatcher beads, which is speculated to be due to excessive charging of the protein backbone at pH <2.5 which may hamper interaction with LemonCatcher when affixed to the bead surface.

15

#### Example 7 – *In vivo* protein labelling utilising LemonTag/LemonCatcher

To further improve the utility and flexibility of the HDX-MS workflow utilising LemonTag/LemonCatcher, a chemical lysis protocol under HDX quench conditions suitable for *in vivo* HDX-MS was developed.

20

A range of membrane-active substances including detergents (OG, DDM and FC-12) and additives (Melittin, Polymyxin B and ZrO<sub>2</sub> beads) were tested (**Figure 10A**), and it was identified that a mixture of 2% (w/v) OG, 10 µM Melittin, and 50 µg/mL Polymyxin B was efficient to perform cell lysis within 1 min at pH 2.5 and 0 °C (**Figure 10B**). This discovery is particularly beneficial as this lysis step can be integrated directly into the bead-based workflow by spin-filtering the obtained cell lysate onto a solution containing LemonCatcher beads.

25

Further tests were conducted to confirm the viability of this adapted *in vivo* HDX-MS workflow in complex mixtures. Sample complexity was increased by exploring enrichment from cell lysate of *E. coli* C41 (DE3) with overexpressed LemonTag-MBP or *E. coli* C41(DE3) cell lysate supplemented with purified LemonTag-MBP. For the latter, a dilution series was performed to vary the concentration of LemonTag-MBP from 1/8 to 7/8 of a total of 1 mg/mL protein. Independent of the dilution test, approximately 100 LemonTag-MBP peptides (1/3 compared to measurements of purified LemonTag-MBP) were successfully obtained with a sequence coverage of 85 to 95% (**Figure 10C**). This clearly

35

demonstrates the workflows viability in complex mixtures which further expands the applicability of this invention.

### Methods

#### 5 Plasmid constructs

Site-directed mutagenesis was carried out by standard methods using Q5 High-Fidelity Polymerase (NEB). All mutations were validated by Sanger sequencing (Eurofins Genomics).

10 pET28a-SpyTag-MBP (Addgene plasmid ID 35050), pET28a-SpyTag002-MBP (GenBank MF974389, Addgene plasmid ID 102831), pET28a-SpyTag003-MBP (GenBank MN433888, Addgene plasmid ID 133450), pET28a AviTag-SpyTag002-MBP, pDEST14-SpyCatcher (GenBank JQ478411, Addgene plasmid ID 35044), pDEST14-SpyCatcher002 (GenBank MF974388, Addgene plasmid ID 102827), and pDEST14-SpyCatcher003 (GenBank MN433887, Addgene plasmid ID 133447) have  
15 been described previously. pET28a-LemonTag-MBP was generated from pET28a-SpyTag002-MBP by site-directed mutagenesis incorporating the following mutation: R12P. pDEST14-LemonCatcher was generated from pDEST14-SpyCatcher003 by site-directed mutagenesis incorporating the following mutations: R32V, S49C, T56A, S59T, V94I and N95T. pET28a-MBP-sTEV (GenBank MZ365307, Addgene plasmid  
20 ID 171782) encodes an MBP fusion to superTEV protease - a variant with high stability that is active without requiring added reducing agent.

#### Bacterial protein expression and purification by Ni-NTA of SpyCatcher variants and LemonCatcher

25 pDEST14 vectors containing Catcher variants were transformed into chemically competent *E. coli* C41 (DE3) cells. Single colonies were picked into 100 mL LB medium containing 100 µg/mL ampicillin and grown for 16-18 h at 37 °C and 220 rpm. 1 L LB medium with appropriate antibiotic was inoculated with 7 mL of the saturated overnight culture and grown at 37 °C and 220 rpm. At OD<sub>600</sub> = 0.5-0.6,  
30 protein overexpression was induced by addition of 0.42 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated for 4-5 h at 30 °C and 220 rpm. Cells were harvested and washed with TBS (50 mM Tris-Cl, pH 7.5, 150 mM NaCl). Cells were resuspended and lysed by sonication in Ni-NTA buffer A (50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, pH 7.5) containing protease inhibitors (Pierce protease

- 67 -

inhibitor tablet, EDTA-free, Thermo Scientific) and 1 mM phenylmethylsulfonyl fluoride (PMSF). Cell lysate was centrifuged for 35 min at 30,000 g. Supernatant was purified with super Ni-NTA affinity resin (ProteinArk) by equilibration/wash with buffer A and elution with buffer B (buffer A + 250 mM imidazole). Purified protein was buffer-exchanged to PBS sample buffer (PBS + 10% (v/v) glycerol, pH 7.5) using a PD-10 desalting column (Cytiva) before being further purified by size exclusion chromatography using a Superdex 200 Increase 10/300 GL column (Cytiva). Protein concentration was determined by absorbance measurements at 280 nm using the extinction coefficient from ExPASy ProtParam.

10

#### Expression and purification of SpyTag-MBP variants and LemonTag-MBP

pET28a vector containing LemonTag-MBP (or other sequences along the evolution) was used. Protein expression and purification was performed as described for LemonCatcher using the appropriate antibiotic (30 µg/mL kanamycin) for the overnight and main culture.

15

#### LemonTag/LemonCatcher reconstitution assay

Both binding partners, i.e. LemonCatcher and LemonTag-MBP were diluted into sample buffer (PBS + 10% (w/v) glycerol, pH 7.5) at their desired concentrations before quenching with 1:1 (v/v) quench buffer 3.0 (100 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM TCEP, pH 2.72) to reach pH 3.0, and 1:1 (v/v) quench buffer 2.5 (100 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM TCEP, pH 2.3) to reach pH 2.5. TCEP addition prevented S49C mediated disulphide bond formation of Catcher dimers during Tag/Catcher reconstitution assays. pH measurements were conducted on ice using a four-point calibration (pH solutions of 1.68, 4.01, 7.01 and 10.01) with a HI-11310 pH Edge Electrode (Hanna instruments). Quenched protein binding partners were then reacted for defined time-points before being stopped by adding 1× Laemmli sample buffer (62.5 mM Tris-HCl, pH 6.8, 10% (v/v) glycerol, 2% (w/v) SDS, 0.01% (w/v) bromophenol blue) and samples were heated for 5 min at 95 °C. Samples were loaded on 12% (w/v) Bis-Tris protein gel (NuPAGE, Thermo Scientific) and proteins were separated by electrophoresis. Proteins were stained with Quick Coomassie Stain (Generon). The rate of amidation, i.e. isopeptide bond formation, was determined based on protein band intensities analysed by an in-house written software.

20

25

30

35

#### Generation of LemonCatcher variants by error-prone PCR

The phage display vector was based on pBAD-DsbA(ss)-HA tag-SpyDock2.0 C49S-pIII (GenBank ON131078). Gibson assembly was used to generate pBAD-DsbA(ss)-HA tag-SpyCatcher003-pIII and pBAD-DsbA(ss)-HA tag-SpyCatcher003 EA-pIII (where the E77A mutation prevents isopeptide bond formation). For library creation, the vector backbone was amplified using KOD polymerase (MilliporeSigma) with oligonucleotide primers flanking SpyCatcher-TM. The vector was amplified using KOD polymerase (MilliporeSigma) with oligonucleotide primers flanking SpyCatcher-TM.

Error-prone PCR was performed on SpyCatcher003 R32V S59T V94I using GeneMorph II Random Mutagenesis Kit (Agilent). Both PCR-amplified fragments were assembled by Gibson assembly using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). Assembly reactions were run on an agarose gel. The DNA was extracted by using Monarch DNA Gel Extraction Kit (New England Biolabs) and eluted in nuclease-free water. TG1 phage display electrocompetent *E. coli* cells (Lucigen) were transformed with 250 ng of library DNA (8 aliquots of 25  $\mu$ L in total). Electroporations were performed in 0.2 mm cuvettes with a Micro-Pulser (Bio-Rad) using the EC2 program. Each electroporation was immediately recovered in 1 mL recovery medium (Lucigen) and incubated for 1 h at 37 °C and 200 rpm. Recovered cells were plated onto two bioassay dishes (245 mm  $\times$  245 mm, Nunc) with LB agar containing 0.8% (w/v) glucose and 100  $\mu$ g/mL carbenicillin and incubated overnight at 30 °C. Cells were carefully collected in 2 $\times$ YT medium containing 0.8% (w/v) glucose and 100  $\mu$ g/mL carbenicillin. Collected cells were centrifuged and stored in 2 $\times$ YT medium containing 20% (v/v) glycerol at -80 °C.

#### 25 Phage production and purification

100 mL 2 $\times$ YT medium containing 2% (w/v) glucose, 0.2% (v/v) glycerol and 100  $\mu$ g/mL carbenicillin were inoculated from an overnight culture of the library (*E. coli* TG1). *E. coli* TG1 cells were grown at 37 °C and 200 rpm until OD<sub>600</sub> reached 0.5. Cultures were infected with M13KO7 helper phage (New England Biolabs) at 20 $\times$  multiplicity of infection and incubated for 45 min at 37 °C and 70 rpm. Infected cells were harvested at 3,000 g and 4 °C for 10 min, subsequently resuspended in the same volume of 2 $\times$ YT medium containing 0.2% (w/v) L-arabinose, 0.2% (v/v) glycerol, and 100  $\mu$ g/mL carbenicillin, and then incubated for 30 min at 18 °C and 200 rpm. 100  $\mu$ g/mL kanamycin was added, and the culture was incubated for 16 h at 18

- 69 -

°C and 200 rpm. Cells were pelleted at 4,000 g and 4 °C for 15 min, and phage was precipitated from supernatant by incubation with 4% (w/v) PEG8000 (Thermo Fisher) and 0.5 M NaCl for 2 h on ice. Phage were pelleted at 15,000 g and 4 °C for 45 min, resuspended in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>), pH 7.5, and centrifuged again at 15,000 g and 4 °C to remove insoluble material. Phage precipitation was repeated twice, and purified phage were stored in PBS pH 7.5 + 20% (v/v) glycerol at -80 °C. Purified phage were titered relative to a serial dilution of M13K07 by qPCR with 2× SensiMix SYBR Hi-ROX (Bioline) master mix. qPCR was performed on an ABI7500 FAST real-time PCR machine (Thermo Fisher).

10

#### Phage survival assay

Six samples of 25 µL phage buffer (0.66× PBS + 20% (v/v) glycerol) were mixed 1:1 with quench buffer (25 µL of 100 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 2.4) to bring the solutions to pH 2.5. Six control samples at pH 7.5 were made using 50 µL of phage buffer. 10<sup>8</sup> plaque forming units (pfu) of M13K07 helper phage were added to each sample by adding 1 µL of stock M13K07 solution at a concentration of 10<sup>11</sup> pfu/mL, and pipetting and vortexing to mix. The samples and corresponding controls were incubated on ice for 15 min, 30 min, 1 hr, 2 hrs, 4 hrs or 8 hrs. The reactions were stopped by addition of 50 µL of neutralisation buffer (1 M Tris-HCl, pH 7.6) and incubated for 5 mins to bring the reaction pH to 7.5. Each 100 µL reaction (all samples and controls) was added to 1 mL of *E. coli* TG1 cell culture at an OD600 of 0.53, and are incubated at 37°C, 70 rpm for 45 mins. Serial dilutions of 1:10-1:10<sup>6</sup> were prepared using YT medium, and 100 µL of each dilution was plated out on kanamycin plates and incubated at 37°C overnight. Colonies were counted and concentrations of phage in plaque forming units (pfu) per mL were calculated.

25

#### Phage display

Biotinylated AviTag-SpyTag002-MBP equipped with a TEV protease cleavage site was used as the bait to react with the LemonCatcher phage library.

To simulate HDX-MS quench conditions, i.e. pH 2.5 and 0 °C, the reaction was performed in 1:1 (v/v) mix of reaction buffer [PBS pH 7.5, 3% (w/v) BSA, 0.05 (v/v) Tween-20] and quench buffer (100 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 2.3). Both bait and phage were quenched (brought to pH 2.5 and 0 °C) before mixing. In the first panning round, 3.0 µM bait was mixed with 1 × 10<sup>12</sup> phage and reacted for 6 h at 1,000 rpm and 4 °C. The reaction was stopped by adding a large excess of unbiotinylated SpyTag002-

35

- 70 -

MBP (60  $\mu$ M final). Three further panning rounds were performed by reacting (i) 0.5  $\mu$ M bait and  $2 \times 10^{11}$  phage for 1 h; (ii) 0.25  $\mu$ M bait and  $1 \times 10^{11}$  phage for 30 min; and (iii) 0.1  $\mu$ M bait and  $1 \times 10^{11}$  phage for 10 min in the presence of *E. coli* C41 (DE3) cell lysate (4 mg/mL total protein concentration). Phage were purified from unreacted bait by precipitation with 4% (w/v) PEG8000 (Thermo Fisher) and 0.5 M NaCl for 1 h on ice. Phage were pelleted at 15,000 g and 4 °C for 15 min and resuspended in 100  $\mu$ L blocking buffer (PBS pH 7.5 containing 3% (w/v) BSA and 0.1 (v/v) Tween-20). The phage bound to biotinylated AviTag-SpyTag002-MBP were captured by addition of 25  $\mu$ L Dynabeads Biotin Binder (Thermo Scientific) in blocking buffer and incubating at 500 rpm and 4 °C for 1 h. The beads were captured using a MagRack 6 (Cytiva) and washed with 150  $\mu$ L blocking buffer. Weakly bound phage was removed by one wash with 150  $\mu$ L 200 mM glycine-HCl pH 2.2, one wash with 150  $\mu$ L 100 mM triethylamine pH 11.0, 3 washes with 150  $\mu$ L TBS + 0.05% (v/v) Tween-20, and a last wash with 150  $\mu$ L PBS pH 7.5 + 0.1% (w/v) BSA. The supernatant was removed, and phage were eluted from beads by addition of 50  $\mu$ L 50  $\mu$ M MBP-sTEV in PBS pH 7.5, 10% (v/v) glycerol, and 0.5 mM ethylenediamine tetraacetic acid (EDTA). TEV protease digestion was performed at 34 °C and 1,000 rpm for 2 h. Eluted phage were rescued by infection of 1 mL *E. coli* TG1 cell culture at 37 °C and 70 rpm for 45 min. 10  $\mu$ L of the culture were used for plating a serial dilution on LB agar plates containing 100  $\mu$ g/mL carbenicillin to quantify the eluted phage. The remaining culture was transferred into 25 mL 2 $\times$ YT supplemented with 100  $\mu$ g/mL carbenicillin and grown at 37 °C and 200 rpm for 16 h. Cells were harvested, and after addition of 20% (v/v) glycerol, cells were flash-frozen in liquid nitrogen and stored at -80 °C.

25

#### Differential scanning calorimetry

SpyTag002 (SEQ ID NO: 26) and LemonTag (SEQ ID NO: 4), each containing an additional N-terminal glycine, were solid-phase synthesised by Insight Biotechnology at >95% purity. Experiments were performed with 20  $\mu$ M SpyCatcher003, SpyCatcher003 triple mutant, and LemonCatcher at pH 7.5, 3.0 or 2.5. Buffers and solution conditions for the different pH were as follows: sample buffer (PBS + 10% (w/v) glycerol) for pH 7.5; 1:1 (v/v) mixture of sample buffer and quench buffer 3.0 (100 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 2.72) for pH 3.0; and 1:1 (v/v) mixture of sample buffer and quench buffer 2.5 (100 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 2.3) for pH 2.5. At acidic pH, LemonCatcher was pre-reacted with 40  $\mu$ M SpyTag002 (pH 2.5) or 40  $\mu$ M LemonTag

35

(pH 2.5 and 3.0) under gentle agitation using a rotation mixer (Cole-Parmer) at 4 °C for 16-18 hr, followed by dialysis with three changes of either sample buffer at pH 7.5, 3.0, or 2.5. Thermal transitions were monitored from 20 to 110 °C at a scan rate of 3 °C/min at 3 atm on a MicroCal PEAQ-DSC (Malvern). Data were processed with  
5 MicroCal PEAQ-DSC analysis software version 1.22. The appropriate blank buffer was subtracted from the experimental sample and corrected for concentration and volume, followed by baseline subtraction. The observed transition was fitted to a two-state model, to obtain the melting temperature  $T_M$ , area under the peak (the enthalpy of unfolding,  $\Delta H_m$ ), and Full Width Half Maximum using MicroCal PEAQ-DSC analysis  
10 software (version 1.22) and Origin 2019b (OriginLab).

#### Preparation of LemonCatcher beads

150  $\mu$ L CarboxyLink coupling gel (Thermo Scientific) were transferred into a Pierce spin column (Thermo Scientific) and washed 3 $\times$  with 500  $\mu$ L ddH<sub>2</sub>O  
15 (centrifugation for 1 min at 500 g). 500  $\mu$ L of 5 mg/mL sulfosuccinimidyl 6-[3'-(2-pyridyl)dithio]propionamido]hexanoate (Sulfo-LC-SPDP; prepared fresh in ddH<sub>2</sub>O; Sigma-Aldrich) were added and incubated for 1 h with shaking at 1,500 rpm and 25 °C. Activated beads were washed 3 times with 500  $\mu$ L ddH<sub>2</sub>O, before being  
20 equilibrated with 3  $\times$  500  $\mu$ L equilibration buffer [PBS, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100, pH 7.5]. 250  $\mu$ L LemonCatcher (362.8  $\mu$ M; PBS, 10% (v/v) glycerol, pH 7.5) were supplemented with 0.1% (v/v) Triton X-100 and 1 mM EDTA, before being added to activated beads and incubated for 24 h at 1,500 rpm and 25 °C. Beads were  
25 washed 3 times with 500  $\mu$ L equilibration buffer. 500  $\mu$ L of L-cysteine (1 mg/mL in equilibration buffer) were added and incubated for 16-18 hrs at 1,500 rpm and 25 °C, to block free sulfhydryl-reactive sites. Excess of L-cysteine was removed by washing  
30 3 times with 500  $\mu$ L equilibration buffer. Beads were washed 5 times with 500  $\mu$ L 1:1 (v/v) sample buffer [PBS, 10% (v/v) glycerol, pH 7.5] and quench buffer (100 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 2.3) to remove non-specifically bound proteins. Beads were equilibrated by washing 3 times with 500  $\mu$ L sample buffer, before adding 250  $\mu$ L sample buffer for storage at 4 °C.

#### Bead-based capture assay using cell lysate

Both LemonCatcher beads and sample (purified LemonTag-MBP, cell lysate of *E. coli* C41(DE3) with overexpressed LemonTag-MBP, or *E. coli* C41(DE3) cell

- 72 -

lysate provided with LemonTag-MBP) were quenched by 1:1 (v/v) addition of quench buffer (100 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 2.72; final pH 3.0). Quenched LemonCatcher beads and quenched sample were reacted for 15 min at 1,500 rpm and 4 °C (5 μL of bead suspension pre-quench in 100 μL final reaction volume). Reacted beads were spin-  
5 filtered through pre-wet Spin-X centrifuge tube filters (Costar) and washed with 100 μL 1:1 (v/v) mix of sample and quench buffer (centrifugations performed for 15 s at 15,000 g). The entire complex (LemonCatcher:LemonTag-MBP) was eluted by reduction with TCEP. 100 μL elution buffer (100 mM TCEP in 1:1 (v/v) PBS sample and quench buffer) were added to the beads on the filter, incubated for 1 min on ice,  
10 before being centrifuged for 15 s at 15,000 g. Flow-through was collected for liquid chromatography-mass spectrometry (LC-MS) analysis.

#### Chemical lysis optimisation under HDX quench conditions

A previously constructed pET15b-AcrB-His<sub>6</sub> plasmid was transformed into  
15 C43(DE3)Δ*acrAB* *E. coli* cells and overexpressed. Polymyxin B sulfate salt, melittin from bee venom (≥85% purity) and Octyl β-D-glucopyranoside (OG) were all purchased from Sigma-Aldrich. n-Dodecyl-beta-Maltoside (DDM, Anatrace) and n-Dodecyl-phosphocholine (Fos-choline-12 (FC-12)) were purchased from Generon. Zirconium oxide (ZrO<sub>2</sub>) beads were purchased from Supelco.

20 *E. coli* C43(DE3)Δ*acrAB* cells were prepared in a 20 % w/v slurry in PBS, pH 7.4. The cell slurry was then mixed 1:1 with a Lysis Cocktail buffer [0.1 M Na<sub>2</sub>HPO<sub>4</sub>, pH 2.2] containing two-times the concentration of lysis additives either alone or in combination [final additive concentrations being 1% (w/v) DDM, 0.5 % (w/v) OG, 0.5% (w/v) FC-12, 10 μM melittin, 50 μg/mL polymyxin B, 100 mg/mL  
25 ZrO<sub>2</sub> beads, 1.5 mg/mL lysozyme). After lysis for 1-10 minutes, the sample was then spin filtered through a 0.22 μm Corning™ Costar™ Spin-X™ Centrifuge Tube Filter at 3,200 x g at 4 °C in a cooling microcentrifuge (Thermo Scientific). Relative abundance of AcrB-His<sub>6</sub> or MBP-LemonTag-His<sub>6</sub> within the lysed cell and filtered fractions was assessed by dot blot or Western blot analysis. Blotting analysis was  
30 performed using a nitrocellulose membrane (Amersham Protran) and visualised using an anti-His antibody conjugated to horseradish peroxidase (HRP) (Sigma-Aldrich, product code: A7058). The membrane was blocked for 1 hour at 25 °C using blocking solution of 5 % (w/v) skim milk powder (Millipore) in PBS-T [PBS with 1 % (v/v) Tween-20]. The membrane was then incubated with the primary anti-

His antibody at a 1:10,000 dilution in blocking solution for 2 hours at 25 °C. The membrane was then washed with PBS-T at least four times for 30 min each before chemiluminescent detection by Amersham™ ECL Select™ Western Blotting Detection Reagent.

5

#### Liquid Chromatography-Mass Spectrometry (LC-MS) analysis

Samples (eluted LemonCatcher:LemonTag-MBP complex) were injected on an ultraperformance liquid chromatography (UPLC) system (nanoACQUITY, Waters, Wilmslow, UK) coupled to an electrospray ionisation quadrupole time-of-flight (ESI-Q-ToF) mass spectrometer (Xevo G2-XS, Waters, Wilmslow, UK). The HDX manager of the nanoACQUITY system was equipped with a Vanguard column (BEH C18, 130 Å, 1.7 µm, 2.1 mm × 5 mm; Waters) and an Acquity UPLC column (BEH C18, 130 Å, 1.7 µm, 1.0 mm × 100 mm; Waters) for peptide trapping and separation, respectively. Protein digestion was performed online with the UPLC chromatographic system using an in-house packed protease column (immobilised pepsin agarose resin, Thermo Scientific) at 15 °C. The generated peptides were trapped and washed with solvent A [0.23% (v/v) formic acid in H<sub>2</sub>O, pH 2.5] at 200 µL/min for 3 min. Subsequently, peptides were separated by applying a 7.5 min linear gradient from 8 to 35% solvent B [0.23% (v/v) formic acid in acetonitrile] at 40 µL/min. Peptides were measured in positive ion mode between 50 and 2,000 m/z on the Xevo G2-XS mass spectrometer. For comparison, purified LemonCatcher and LemonTag-MBP were separately measured in triplicates under the same buffer conditions and instrument settings.

25

#### Assessing the reactivity of LemonTag/LemonCatcher at extremely low temperatures

LemonCatcher and LemonTag-MBP were buffer exchanged to sample buffer (PBS + 10% (v/v) glycerol + 40% (v/v) ethylene glycol, pH 7.5) and diluted to 4 µM. Proteins and quench buffer (100 mM NaH<sub>2</sub>PO<sub>4</sub> + 40% (v/v) ethylene glycol + 10 mM TCEP at pH 2.35 or pH 2.81 for a final pH of 2.5 or 3.0, respectively) were equilibrated for 2 hrs at -20 °C. LemonCatcher and LemonTag-MBP were quenched 1:1 in quench buffer and kept for a further 15 min at -20 °C. Reaction was started by 1:1 mix of LemonCatcher and LemonTag-MBP in pre-cooled (-20 °C) tubes and incubated at -20 °C for various time-points. Reaction was stopped by adding 20 µL reaction volume

30

- 74 -

to 5  $\mu$ L 5 $\times$  Laemmli buffer (pre-heated to 95 °C) and incubation for 5 min at 95 °C. SDS-PAGE was performed on NuPAGE 12% Bis-Tris protein gels.

#### Data processing

5 Protein identification and peptide filtering were performed with ProteinLynx Global Server 3.0 (PLGS) and DynamX 3.0, respectively (Waters, Wilmslow, UK). PLGS workflow parameters for peptide identification were as follows: peptide tolerance: automatic; fragment tolerance: automatic; min fragment ion matches per peptide: 2; minimum fragment ion matches per protein: 7; minimum peptide matches  
10 per protein: 3; maximum protein mass 250,000 Da; primary digest reagent: nonspecific; and false discovery rate: 100. PLGS output files were then consulted for further validation by DynamX. DynamX parameters for peptide filtering were as follows: minimum intensity: 1481; minimum sequence length: 5; maximum sequence length: 25; minimum products per amino acid: 0.11; minimum consecutive products:  
15 1; minimum PLGS score: 6.62; maximum MH<sup>+</sup> error (ppm): 5.

### Claims

1. A polypeptide comprising:

5 (a) an amino acid sequence with at least 80% sequence identity to a sequence as set forth in SEQ ID NO: 2, wherein the amino acid sequence comprises lysine at position 31, glutamic acid at position 77 and wherein:

(i) X at position 32 is not a basic amino acid;

(ii) X at position 49 is any amino acid, preferably a polar amino acid or cysteine;

10 (iii) X at position 56 is a hydrophobic amino acid or a polar amino acid;

(iv) X at position 59 is a polar amino acid;

(v) X at position 94 is a hydrophobic amino acid; and

(vi) X at position 95 is a polar amino acid,

15 and wherein the polypeptide is capable of spontaneously forming an isopeptide bond with a peptide comprising an amino acid sequence as set forth in SEQ ID NO: 4, wherein the isopeptide bond forms between the aspartic acid residue at position 8 of SEQ ID NO: 4 and the lysine residue at position 31 of the polypeptide; or

20 (b) an amino acid sequence with at least 80% sequence identity to a sequence as set forth in SEQ ID NO: 15, wherein the amino acid sequence comprises lysine at position 10, glutamic acid at position 56 and wherein:

(i) X at position 11 is not a basic amino acid;

(ii) X at position 28 is any amino acid, preferably a polar amino acid or cysteine;

25 (iii) X at position 35 is a hydrophobic amino acid or a polar amino acid;

(iv) X at position 38 is a polar amino acid;

(v) X at position 73 is a hydrophobic amino acid; and

(vi) X at position 74 is a polar amino acid,

30 and wherein the polypeptide is capable of spontaneously forming an isopeptide bond with a peptide comprising an amino acid sequence as set forth in SEQ ID NO: 4, wherein the isopeptide bond forms between the aspartic acid residue at position 8 of SEQ ID NO: 4 and the lysine residue at position 10 of the polypeptide.

2. The polypeptide of claim 1, wherein X at position 32 of the polypeptide of part (a) or position 11 of the polypeptide of (b) is a hydrophobic amino acid or a polar amino acid.
- 5 3. The polypeptide of claim 2, wherein:  
(a) the hydrophobic amino acid is valine, isoleucine, leucine or alanine, preferably valine; or  
(b) the polar amino acid is threonine or serine, preferably threonine.
- 10 4. The polypeptide of any one of claims 1 to 3, wherein:  
(i) X at position 49 of the polypeptide of part (a) or position 28 of the polypeptide of (b) is cysteine or serine, preferably cysteine;  
(ii) X at position 56 of the polypeptide of part (a) or position 35 of the polypeptide of (b) is alanine or threonine, preferably alanine;  
15 (iii) X at position 59 of the polypeptide of part (a) or position 38 of the polypeptide of (b) is threonine or serine, preferably threonine;  
(iv) X at position 94 of the polypeptide of part (a) or position 73 of the polypeptide of (b) valine or isoleucine, preferably isoleucine; and/or  
(v) X at position 95 of the polypeptide of part (a) or position 74 of the  
20 polypeptide of (b) is threonine, serine, asparagine or glutamine, preferably threonine or serine.
5. The polypeptide of claim 1, wherein the polypeptide comprises:  
(a) an amino acid sequence with at least 80% sequence identity to a  
25 sequence as set forth in SEQ ID NO: 3, wherein X at position 32 is not a basic amino acid, and the amino acid sequence comprises lysine at position 31, glutamic acid at position 77 and one or more of the following:  
(i) alanine at position 56;  
(ii) threonine at position 59;  
30 (iii) isoleucine at position 94; and  
(iv) threonine at position 95,  
wherein the specified amino acid residues are at positions equivalent to the positions in SEQ ID NO: 3; or  
(b) an amino acid sequence with at least 80% sequence identity to a  
35 sequence as set forth in SEQ ID NO: 16, wherein X at position 11 is not a basic amino

- 77 -

acid, and the amino acid sequence comprises lysine at position 10, glutamic acid at position 56 and one or more of the following:

- (i) alanine at position 35;
- (ii) threonine at position 38;
- 5 (iii) isoleucine at position 73; and
- (iv) threonine at position 74,

wherein the specified amino acid residues are at positions equivalent to the positions in SEQ ID NO: 16.

10 6. The polypeptide of claim 5, wherein the polypeptide contains two or more of the amino acids specified in (i)-(iv).

7. The polypeptide of claim 5 or 6, wherein the polypeptide contains three or more of the amino acids specified in (i)-(iv).

15

8. The polypeptide of any one of claims 5 to 7, wherein the polypeptide contains all of the amino acids specified in (i)-(iv).

9. The polypeptide of any one of claims 5 to 8, wherein X at position 32  
20 or position 11 is as defined in claim 2 or 3.

10. The polypeptide of any one of claims 1 to 9, wherein the polypeptide comprises:

25 (a) an amino acid sequence selected from a sequence as set forth in any one of SEQ ID NOs: 1 and 5-14 or 17-25, preferably SEQ ID NOs: 1, 13, 14 or 25;

(b) an amino acid sequence selected from a sequence as set forth in any one of SEQ ID NOs: 5-13, wherein the serine residue at position 49 is substituted with cysteine; or

30 (c) an amino acid sequence selected from a sequence as set forth in any one of SEQ ID NOs: 17-25, wherein the serine residue at position 28 is substituted with cysteine.

11. The polypeptide of any one of claims 1 to 10, wherein the polypeptide comprises cysteine at a position equivalent to position 49 in SEQ ID NO: 3.

35

12. The polypeptide of any one of claims 1 to 11, wherein the polypeptide is immobilised on a solid substrate, optionally wherein the polypeptide is immobilised on the solid substrate via a covalent bond.

5 13. The polypeptide of claim 12, wherein the polypeptide is immobilised on a solid substrate via a covalent bond between a cysteine residue and the solid substrate, wherein the cysteine residue is at a position equivalent to position 49 in SEQ ID NO: 3.

10 14. A peptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 4.

15 15. The peptide of claim 14, conjugated to a protein, peptide, nucleic acid molecule, small-molecule organic compound, fluorophore, metal-ligand complex, polysaccharide, nanoparticle, 2D monolayer (e.g. graphene), nanotube, polymer, cell, virus, virus-like particle or a combination thereof, preferably a protein or peptide.

16. A recombinant or synthetic polypeptide comprising a peptide or polypeptide linked to:

20 (a) a polypeptide as defined in any one of claims 1 to 13; or  
(b) a peptide as defined in claim 14.

17. A two-part linker comprising:

25 (a) a polypeptide as defined in any one of claims 1 to 13; and  
(b) a peptide as defined in claim 14 or 15,

wherein the peptide and polypeptide are capable of spontaneously forming an isopeptide bond between an aspartic acid residue at a position corresponding to position 8 of SEQ ID NO: 4 and a lysine residue at a position corresponding to position 31 of SEQ ID NO: 1.

30

18. A kit comprising:

(1)

(i) a peptide of claims 14 or 15 or a recombinant polypeptide of claim 16 comprising a peptide as defined in claim 14; and

(ii) a polypeptide of any one of claims 1 to 13 or a recombinant polypeptide of claim 16 comprising a polypeptide as defined in any one of claims 1 to 13; and/or

(2)

5 (i) a nucleic acid molecule, particularly a vector, encoding a peptide or recombinant polypeptide as defined in (1)(i); and

(ii) a nucleic acid molecule, particularly a vector, encoding a polypeptide or recombinant polypeptide as defined in (1)(ii).

10 19. A nucleic acid molecule comprising a nucleotide sequence which encodes:

(i) a polypeptide as defined in any one of claims 1 to 13;

(ii) a peptide as defined in claim 14 or 15; or

(iii) a recombinant polypeptide as defined in claim 16.

15 20. A vector comprising the nucleic acid molecule of claim 19.

21. Use of a polypeptide as defined in any one of claims 1 to 13 to conjugate two molecules or components via an isopeptide bond,

20 wherein the molecules or components conjugated via an isopeptide bond comprise:

a) a first molecule or component comprising a polypeptide of any one of claims 1 to 13; and

25 b) a second molecule or component comprising a peptide of claim 14 or 15, particularly a recombinant polypeptide of claim 16 comprising a peptide as defined in claim 14,

and wherein the isopeptide bond forms between an aspartic acid residue at a position corresponding to position 8 of SEQ ID NO: 4 and a lysine residue at a position corresponding to position 31 of SEQ ID NO: 1.

30 22. A process for conjugating two molecules or components via an isopeptide bond comprising:

a) providing a first molecule or component comprising a polypeptide of any one of claims 1 to 13;

- 80 -

b) providing a second molecule or component comprising a peptide of claim 14 or 15, particularly a recombinant polypeptide of claim 16 comprising a peptide as defined in claim 14;

5 c) contacting the first and second molecules or components under conditions that enable the spontaneous formation of an isopeptide bond between the polypeptide and peptide, thereby conjugating the first molecule or component to the second molecule or component via an isopeptide bond to form a complex,

10 wherein the isopeptide bond forms between an aspartic acid residue at a position corresponding to position 8 of SEQ ID NO: 4 and a lysine residue at a position corresponding to position 31 of SEQ ID NO: 1.

23. The use of claim 21 or process of claim 22, wherein the first molecule or component comprises an immobilised polypeptide as defined in claim 12 or 13, optionally wherein the covalent bond that immobilises the polypeptide on the solid  
15 substrate is a reducible covalent bond.

24. The process of claim 23, being a process for purifying or isolating the second molecule or component, the process comprising:

20 (a) contacting the immobilised polypeptide with a sample comprising the second molecule or component under conditions that enable the spontaneous formation of an isopeptide bond between the immobilised polypeptide and the peptide in the second molecule or component thereby forming a complex;

(b) separating the complex from other molecules or components in the sample, thereby isolating or purifying the second molecule or component; and  
25 optionally

(c) releasing the complex from the solid substrate.

25. The process of claim 24, wherein the step of releasing the complex from the solid substrate comprises contacting the complex with a reducing agent, optionally wherein the reducing agent is tris(2-carboxyethyl)phosphine (TCEP).  
30

26. The use of claim 21 or 23 or the process of any one of claims 22 to 25, wherein the steps are performed at:

35 i) a temperature of about -20 °C to about 10 °C, optionally about -15 °C to about 10 °C, about -10 °C to about 5 °C, about 0-6 °C or about 0-4 °C; and/or

ii) an acidic pH, preferably a pH of about 1.0-5.5, optionally about 2.5-4.5 or about 3.0-4.0.

27. The use of claim 21 or 23 or the process of any one of claims 22 to 5 25, wherein the steps are performed at:

i) a temperature of about 10-40 °C, optionally about 20-40 °C or about 25-35 °C; and/or

ii) a pH of about 1.0-8.0, optionally about 3.0-7.5 or about 4.0-7.0.

10 28. Use of:

(i) a polypeptide as defined in any one of claims 1 to 13;

(ii) a peptide as defined in claim 14 or 15; and/or

(iii) a recombinant polypeptide as defined in claim 16,

15 in a protein analysis method, optionally where in the protein analysis method uses mass spectrometry, particularly Hydrogen-Deuterium eXchange (HDX) mass spectrometry.

29. An apparatus suitable for use in the process of any one of claims 22 to 26 or use of any one of claims 21, 23, 26 to 28 comprising a solid substrate on 20 which a polypeptide as defined in any one of claims 1 to 13 is immobilised.

30. A kit suitable for use in preparing a solid substrate on which a polypeptide as defined in any one of claims 1 to 13 is immobilised, comprising:

25 a) a polypeptide as defined in any one of claims 1 to 13; and

b) means for immobilising the polypeptide of a) on a solid substrate.

31. The kit of claim 30 further comprising a solid substrate.

32. A resin and/or beads for use in the process of any one of claims 22 to 30 27 or the use of any one of claims 21, 23, 26 to 28 on which a polypeptide as defined in any one of claims 1 to 13 is immobilised.

33. A process for analysing a polypeptide of interest (POI) comprising:

35 (i) providing a sample comprising the POI linked to a peptide as defined in claim 14;

- 82 -

(ii) subjecting the sample to hydrogen-deuterium exchange to provide a sample comprising a deuterium-labelled POI linked to the peptide;

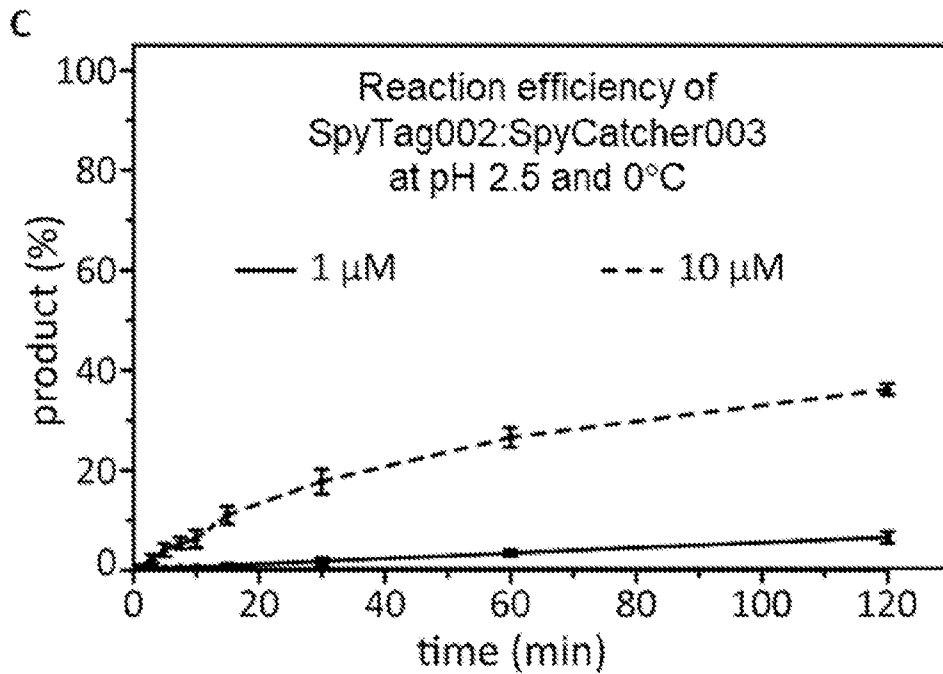
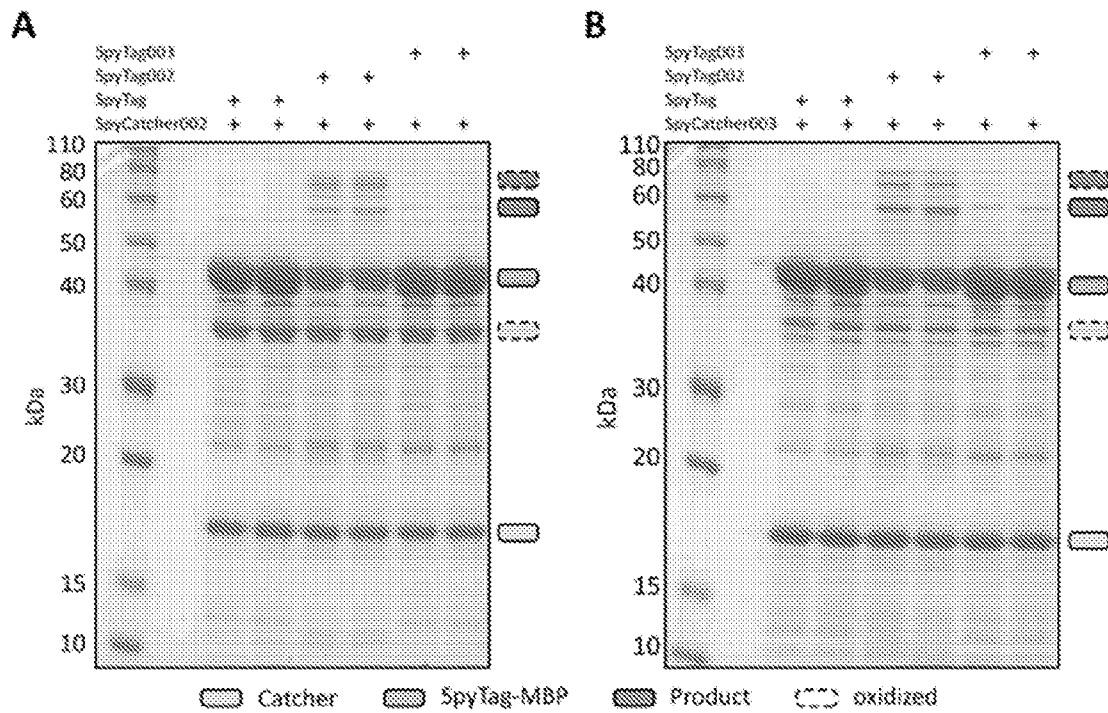
(iii) contacting the sample from (ii) with a polypeptide as defined in any one of claims 1 to 13 under conditions that enable the spontaneous formation of an isopeptide bond between the polypeptide and peptide linked to the POI, thereby  
5 conjugating the POI to the polypeptide via an isopeptide bond to form a complex, wherein the conditions are also HDX quench conditions;

(iv) isolating the complex under HDX quench conditions; and

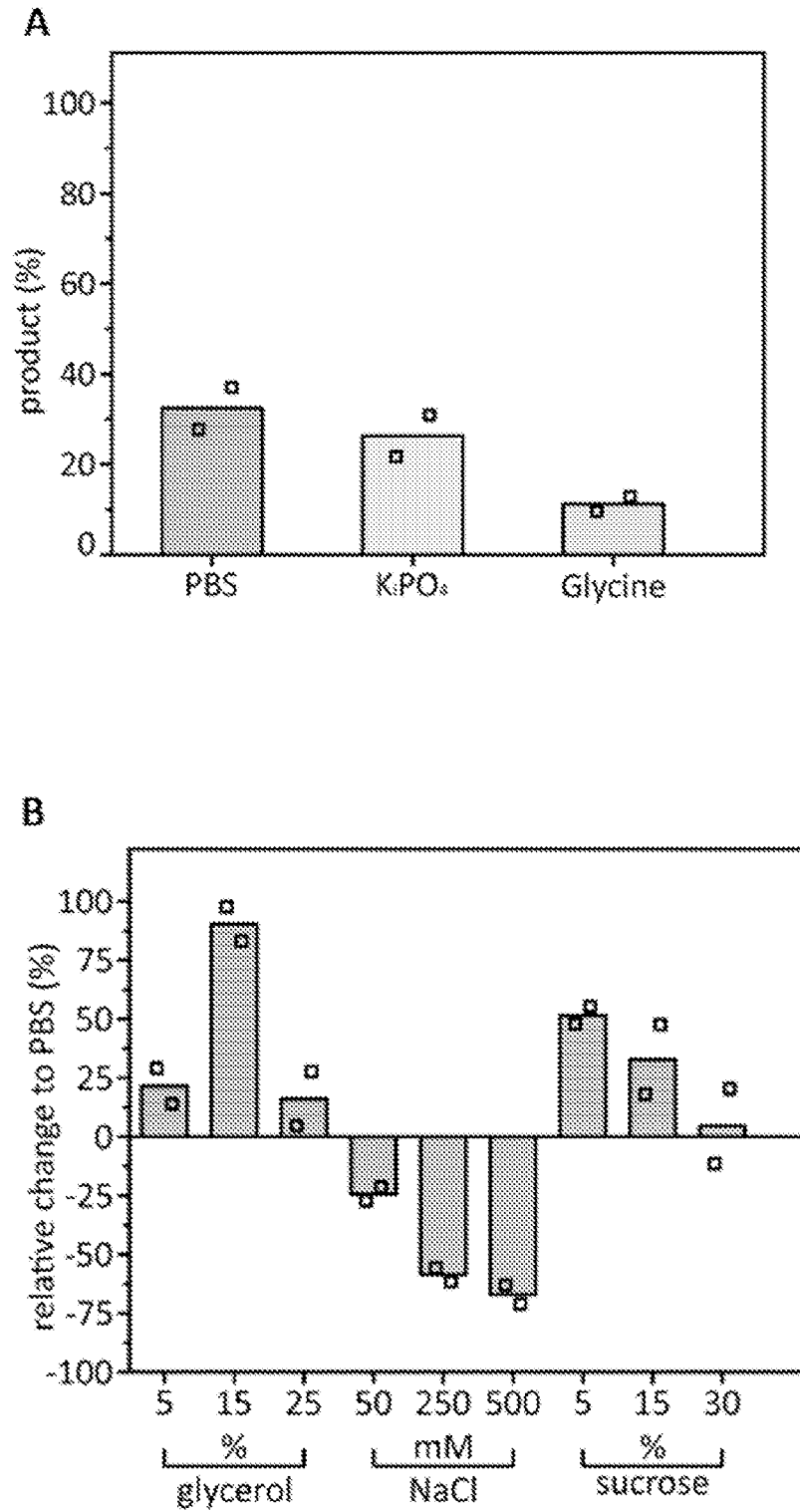
(v) analysing the complex.

10

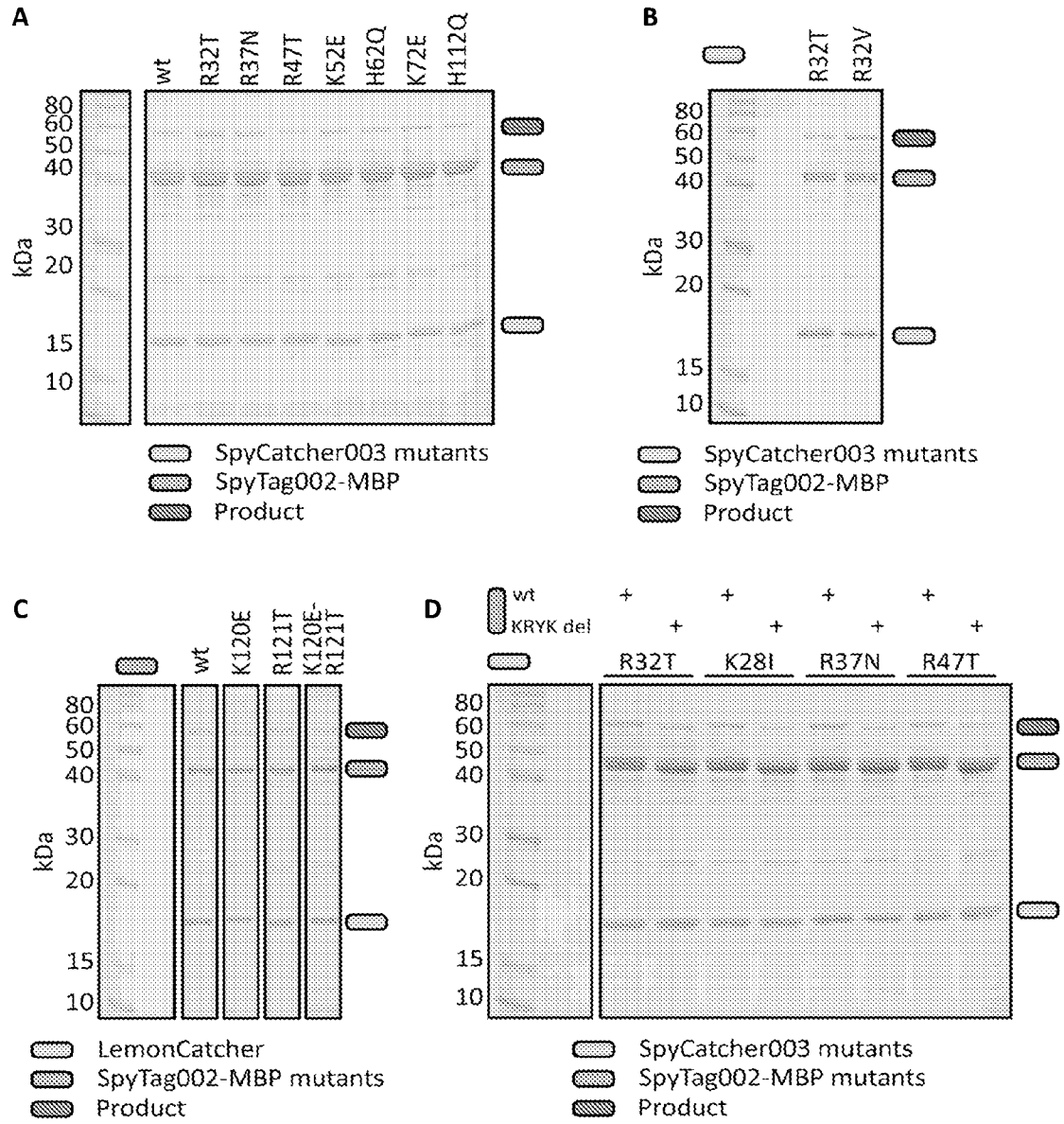
**Figure 1**



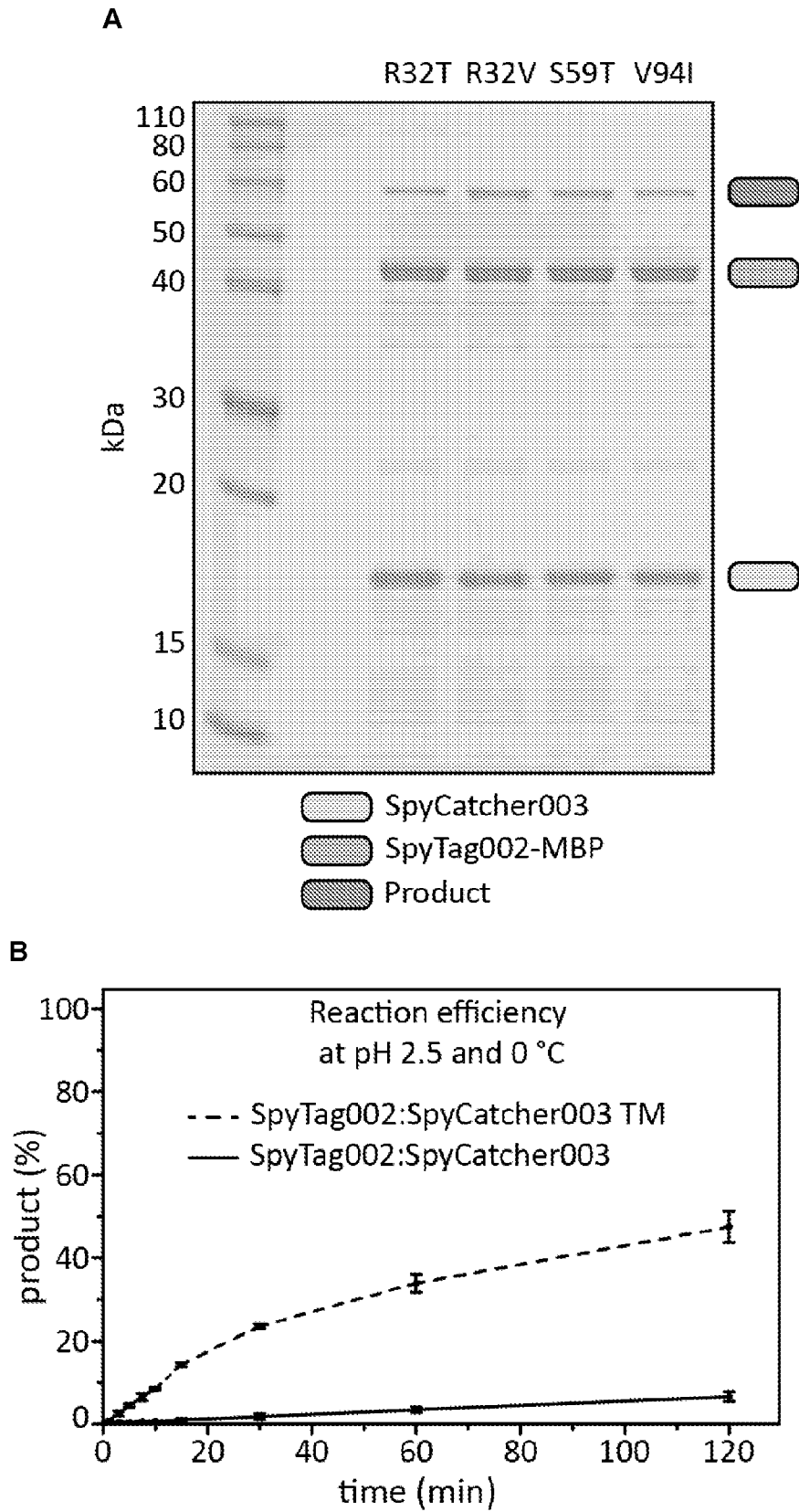
**Figure 2**



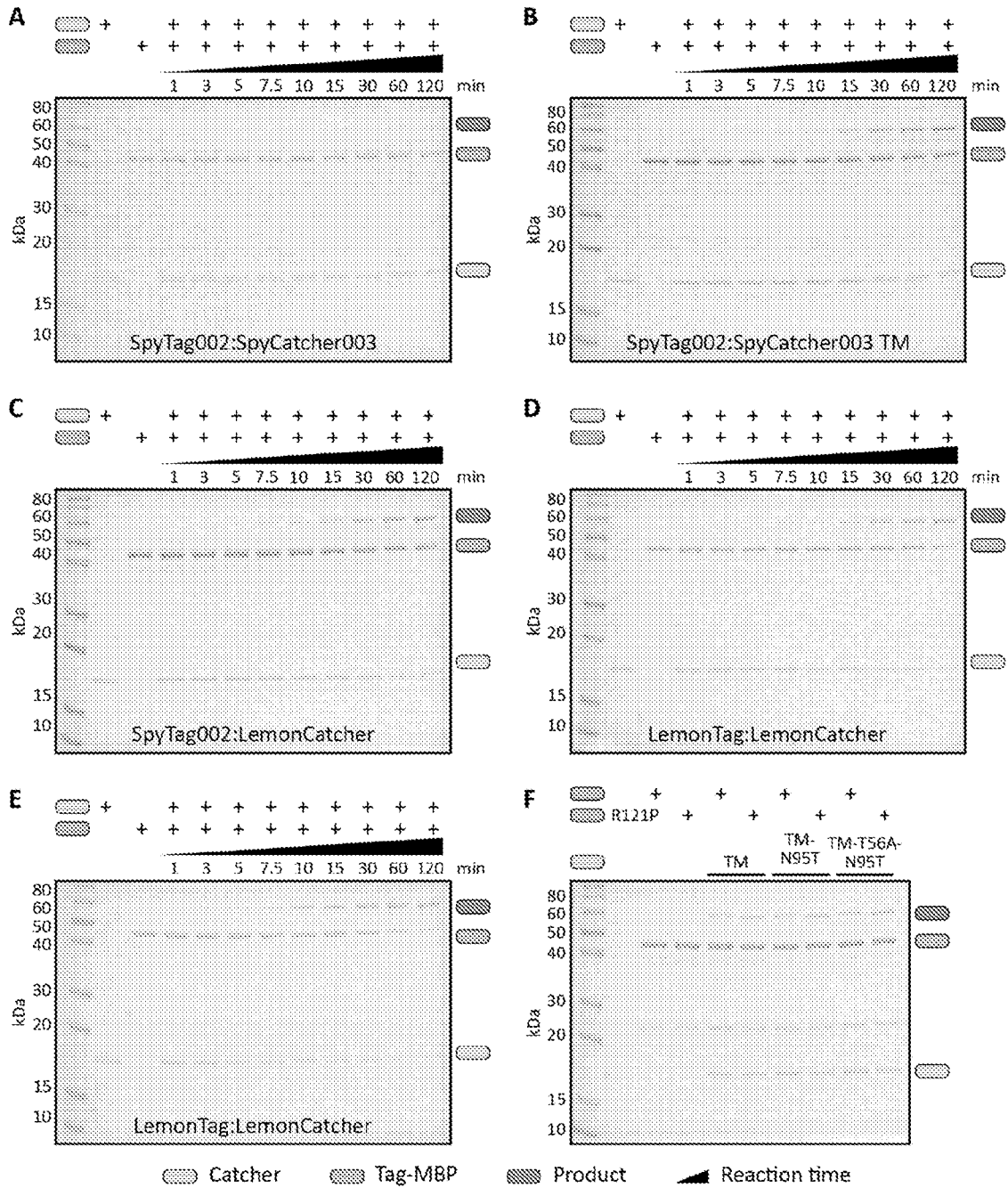
**Figure 3**



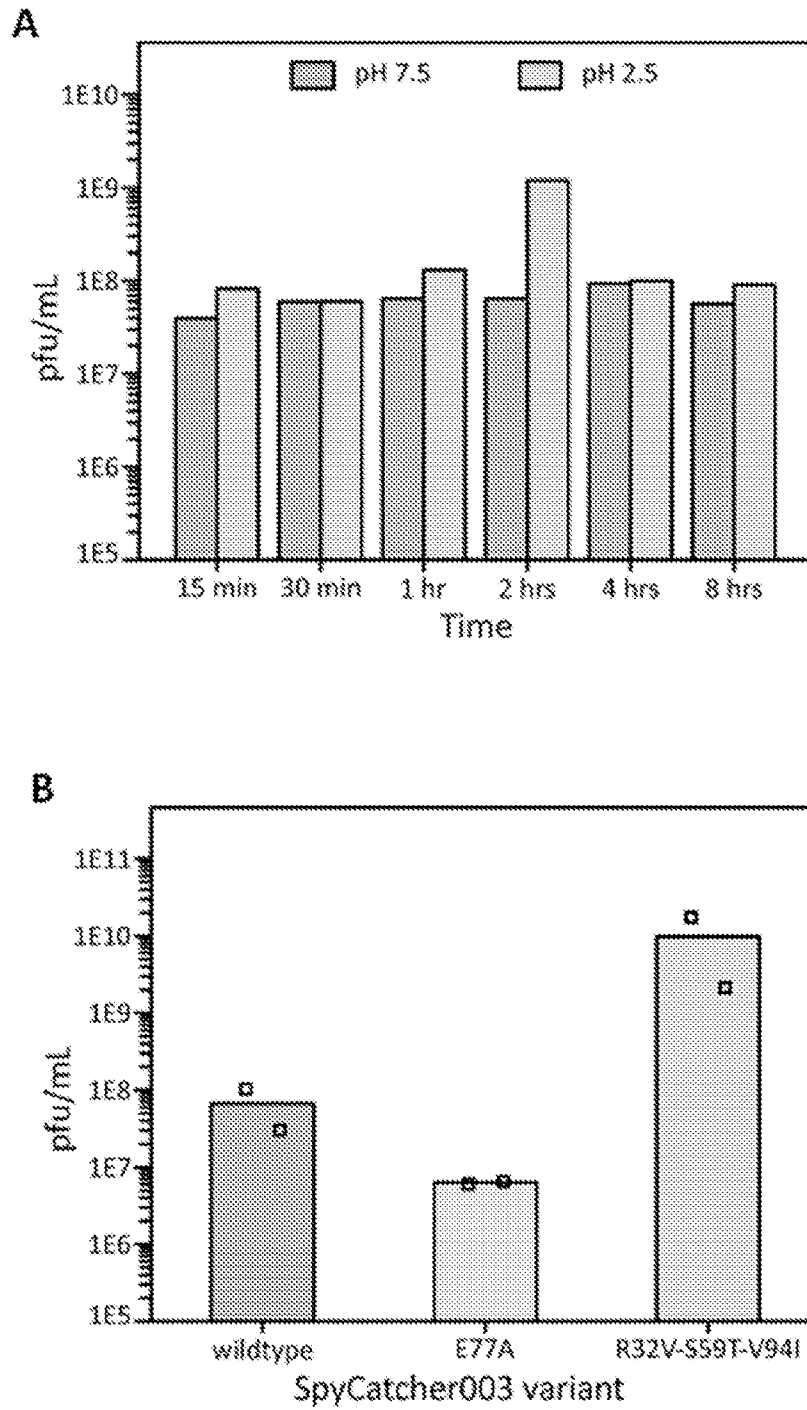
**Figure 4**



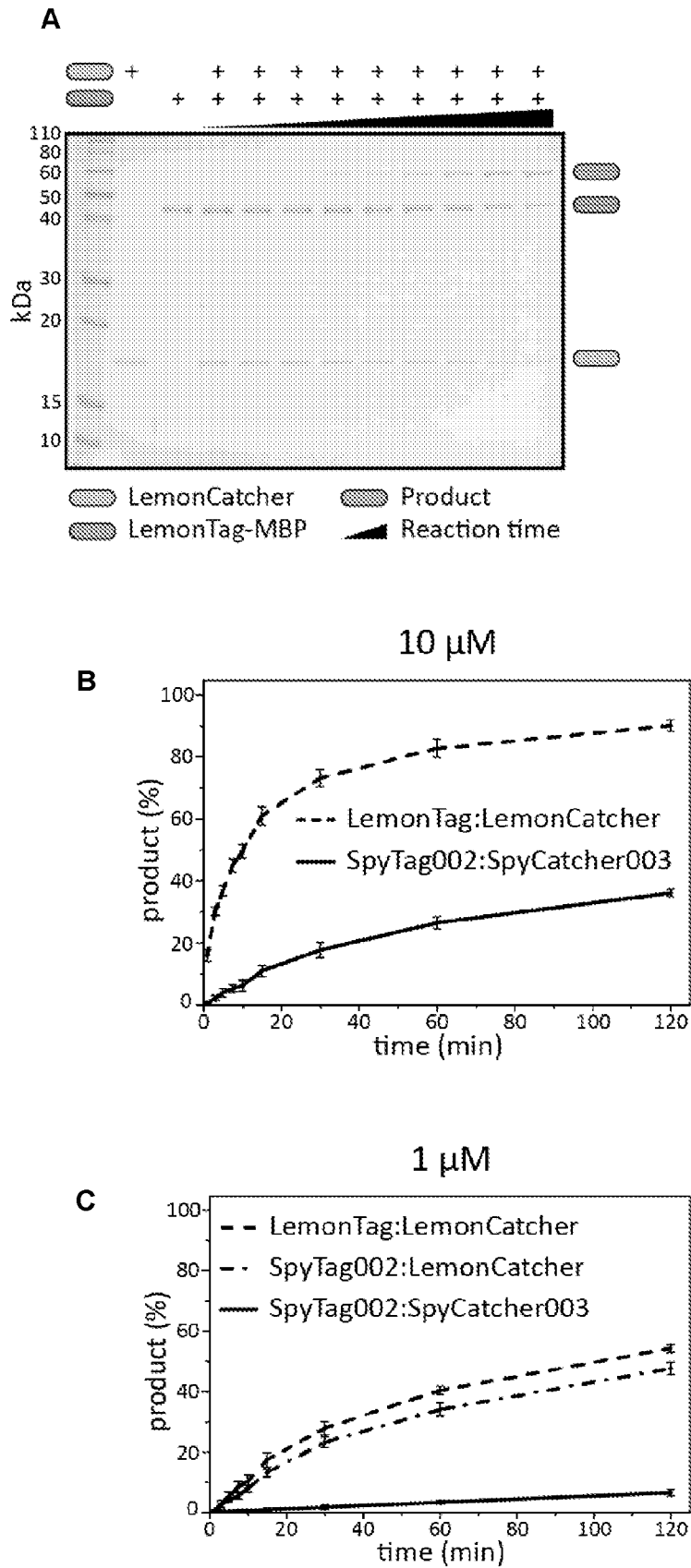
**Figure 5**



**Figure 6**

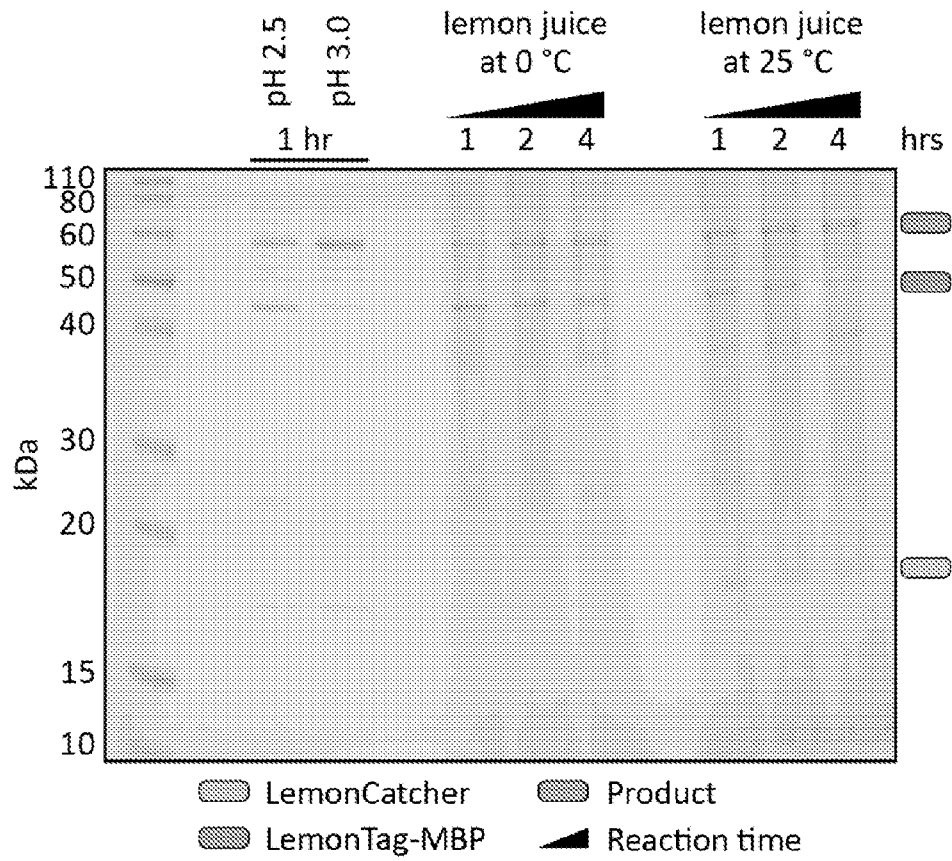


**Figure 7**

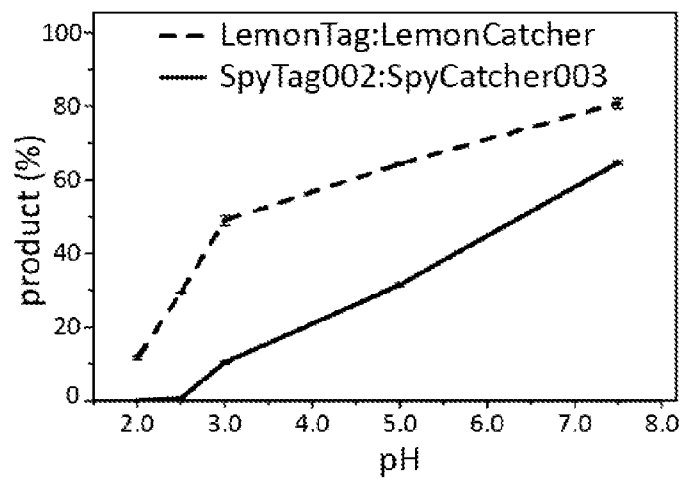


**Figure 8**

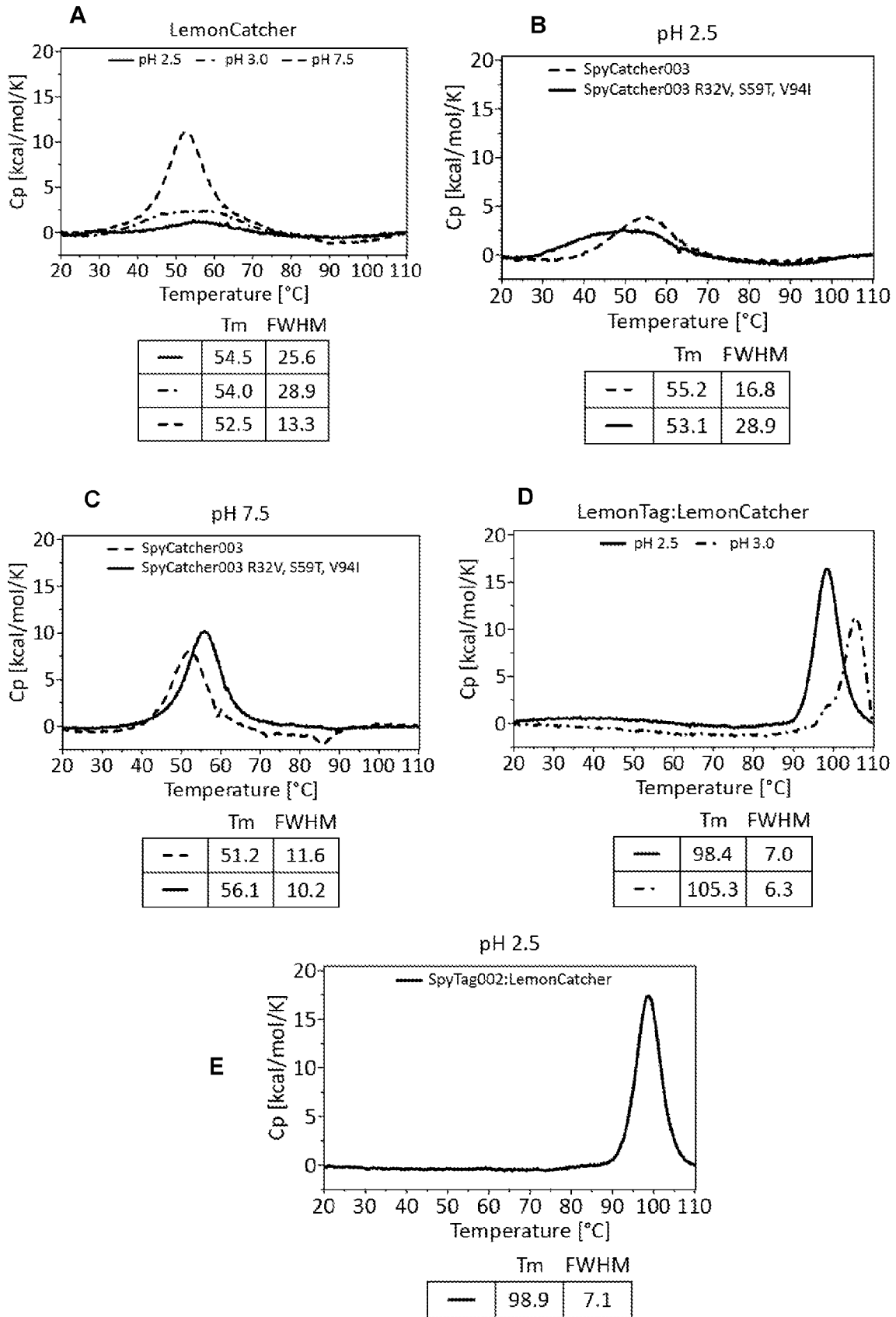
**A**



**B**

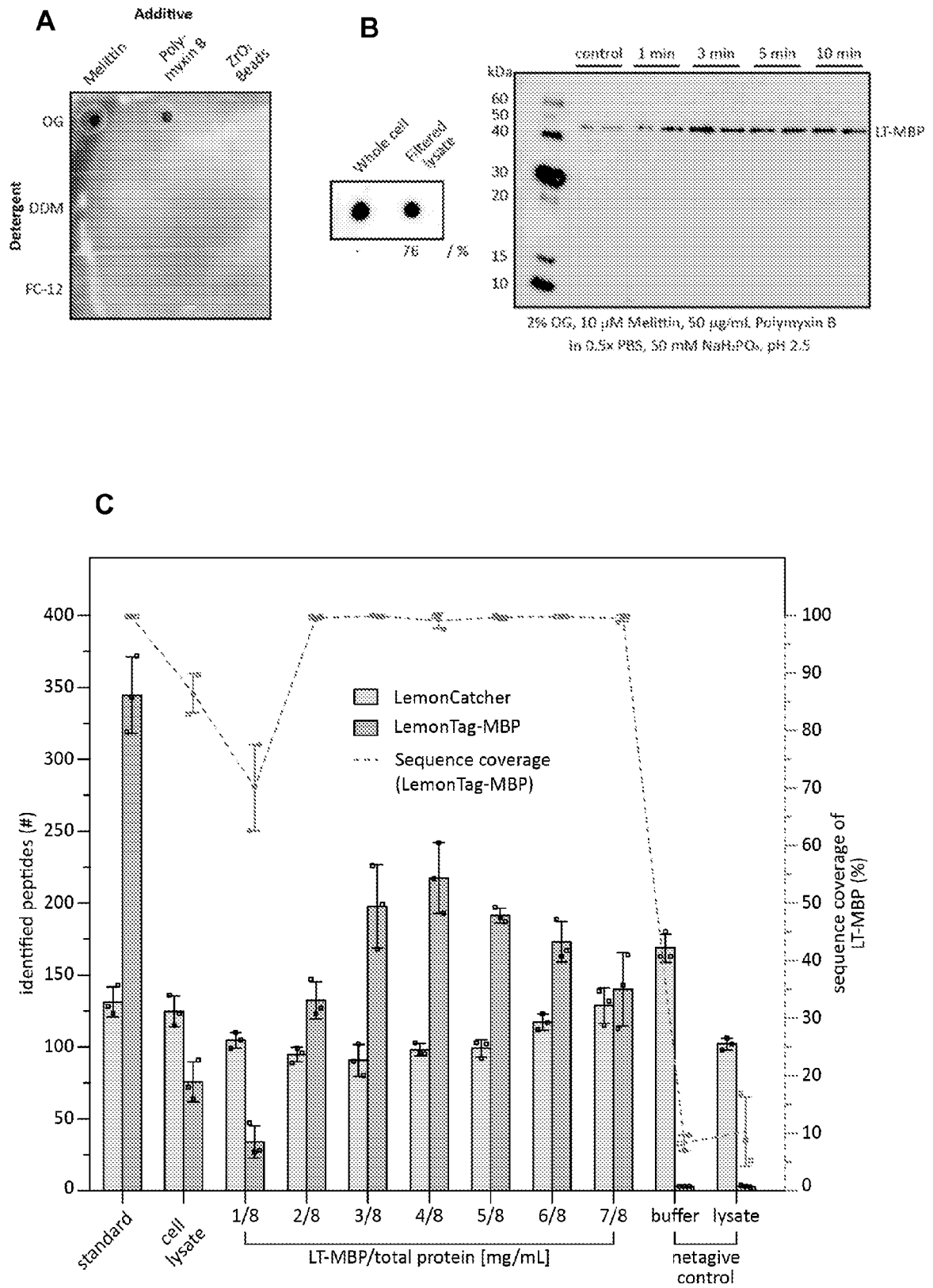


**Figure 9**

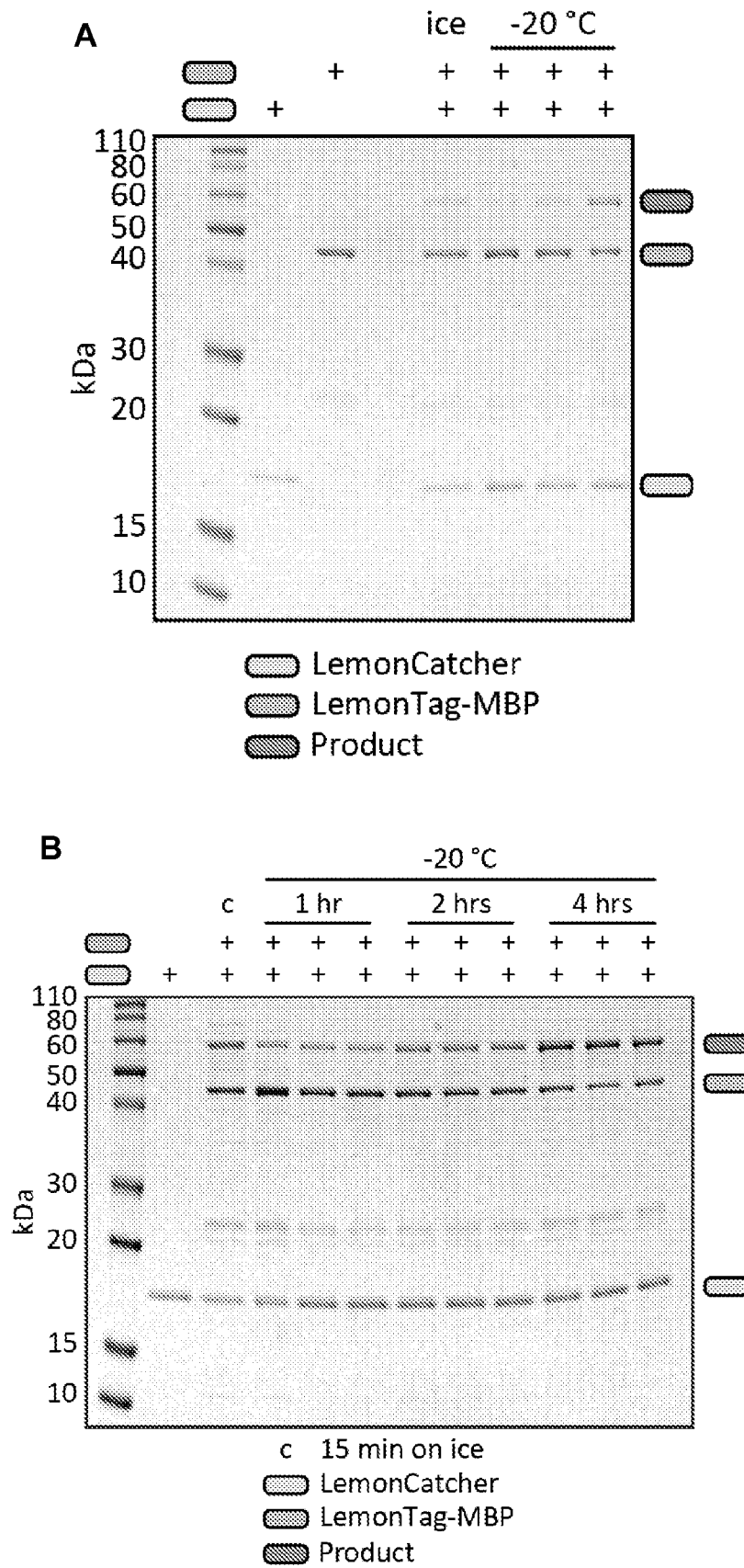


10/12

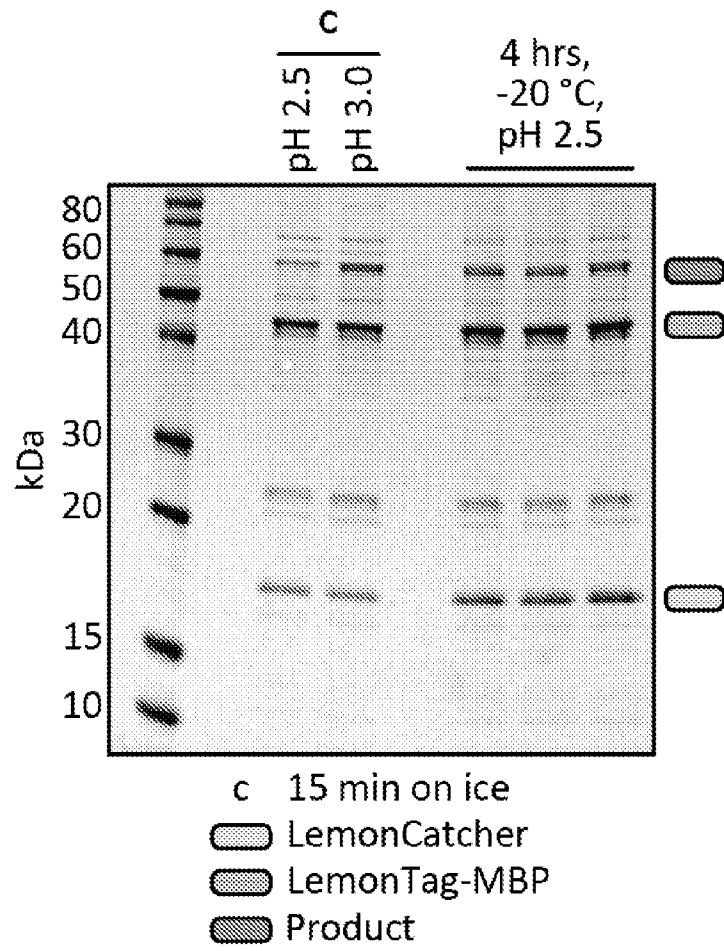
**Figure 10**



**Figure 11**



**Figure 12**



# INTERNATIONAL SEARCH REPORT

International application No  
PCT/GB2024/052526

**A. CLASSIFICATION OF SUBJECT MATTER**  
 INV. C07K14/315 C07K7/00  
 ADD.

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)  
 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, WPI Data, BIOSIS, EMBASE

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	VESTER SUSAN K. ET AL: "SpySwitch enables pH- or heat-responsive capture and release for plug-and-display nanoassembly", NATURE COMMUNICATIONS, vol. 13, no. 1, 28 June 2022 (2022-06-28), XP093251103, UK ISSN: 2041-1723, DOI: 10.1038/s41467-022-31193-8 Retrieved from the Internet: URL:https://www.nature.com/articles/s41467-022-31193-8> abstract page 2, right-hand column figures 2c, d  <div style="text-align: center;">-----</div> <div style="text-align: center;">-/-</div>	1-13, 16-33

Further documents are listed in the continuation of Box C.       See patent family annex.

\* Special categories of cited documents :

<p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"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)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"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</p> <p>"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</p> <p>"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</p> <p>"&amp;" document member of the same patent family</p>
---	---

Date of the actual completion of the international search	Date of mailing of the international search report
11 March 2025	21/03/2025

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  <p style="text-align: center;"><b>Sonnerat, Isabelle</b></p>
--	--

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/GB2024/052526

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2020/183198 A1 (UNIV OXFORD INNOVATION LTD [GB]) 17 September 2020 (2020-09-17) cited in the application	14-33
A	the whole document -----	1-13
X	WO 2020/115252 A1 (UNIV OXFORD INNOVATION LTD [GB]) 11 June 2020 (2020-06-11)	14-33
A	the whole document -----	1-13
A	IRSYAD N. A. KHAIRIL ANUAR ET AL: "Spy&Go purification of SpyTag-proteins using pseudo-SpyCatcher to access an oligomerization toolbox", NATURE COMMUNICATIONS, vol. 10, no. 1, 15 April 2019 (2019-04-15), page 1734, XP055657572, DOI: 10.1038/s41467-019-09678-w the whole document -----	1-33

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/GB2024/052526

## Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
  - a.  forming part of the international application as filed.
  - b.  furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*:1(a)).
    - accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2.  With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2024/052526

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2020183198 A1	17-09-2020	AU 2020234063 A1	04-11-2021
		CN 113767111 A	07-12-2021
		EP 3938380 A1	19-01-2022
		JP 2022525160 A	11-05-2022
		US 2022135628 A1	05-05-2022
		WO 2020183198 A1	17-09-2020
-----			
WO 2020115252 A1	11-06-2020	EP 3891172 A1	13-10-2021
		US 2022041663 A1	10-02-2022
		WO 2020115252 A1	11-06-2020
-----			