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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 March 2018 (29.03.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/056899 A1

(51) International Patent Classification:
*C07K 14/415* (2006.01) *C12P 21/04* (2006.01)
*C12N 9/00* (2006.01)

(21) International Application Number:
PCT/SG2017/050458

(22) International Filing Date:
12 September 2017 (12.09.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10201607951V 23 September 2016 (23.09.2016) SG

(71) Applicant: **NANYANG TECHNOLOGICAL UNIVERSITY** [SG/SG]; 50 Nanyang Avenue, Singapore 639798 (SG).

(72) Inventors: **WU, Bin**; c/o Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore 639798 (SG). **YANG, Renliang**; c/o Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore 639798 (SG). **WONG, Yee Hwa**; c/o Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore 639798 (SG). **LESCAR, Julien**; c/o Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore 639798 (SG). **LIU, Chuan Fa**; c/o Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore 639798 (SG). **TAM, James P**; c/o Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore 639798 (SG). **NGUYEN, Kien Truc Giang**; c/o Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore 639798 (SG). **LONG, Ziqi**; c/o Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore 639798 (SG).

(74) Agent: **VIERING, JENTSCHURA & PARTNER LLP**; P.O Box 1088, Rochor Post Office, Rochor Road, Singapore 911833 (SG).

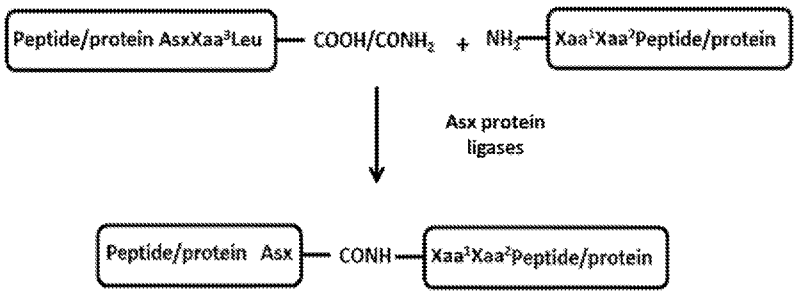
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,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

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

(54) Title: METHODS FOR ENZYMATIC PEPTIDE LIGATION

Figure 1

a




(57) Abstract: The present invention relates to a method of ligating a first peptide via its C-terminus to the N-terminus of a second peptide, wherein the reaction is catalyzed by an asparagine/aspartate (Asx) peptide ligase OaAEPI Cys247Ala having the amino acid sequence of SEQ ID NO: 1. Further encompassed are a method of preparing a dimer, oligomer, or multimer of one or more peptides of interest and a method of modifying or tagging the surface of a target cell by one or more peptides of interest. Also encompassed in the invention are the ligated peptides and/or tagged target cells obtainable according to any of the methods, the peptide ligase OaAEPI Cys247Ala having the amino acid sequence of SEQ ID NO: 1, as well as kits comprising said peptide ligase.

WO 2018/056899 A1

*[Continued on next page]*

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

**Published:**

— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

# METHODS FOR ENZYMATIC PEPTIDE LIGATION

## CROSS-REFERENCE TO RELATED APPLICATION

This application makes reference to and claims the benefit of priority of the Singapore Patent Application No. 10201607951V filed on 23 September 2016, the content of which is incorporated herein by reference for all purposes, including an incorporation of any element or part of the description, claims or drawings not contained herein and referred to in Rule 20.5(a) of the PCT, pursuant to Rule 4.18 of the PCT.

## FIELD OF THE INVENTION

The present invention relates generally to Asx-specific peptide ligases, methods of enzymatically ligating peptides using said ligases and the thus generated peptides and peptide-containing complexes.

## BACKGROUND OF THE INVENTION

Unlike proteases that are abundantly found in nature, peptide ligases are relatively rare. The first asparagine/aspartate (Asx) peptide ligase reported was butelase 1, purified from the cyclotide-producing plant *Clitoria ternatea* (Nguyen, G.K., et al., Nat Chem Biol, 2014. 10(9): p. 732-8). Compared to bacterial sortase which is commonly used to catalyze transpeptidation reactions *in vitro*, butelase 1 shows exceptionally high efficiency in catalyzing both peptide and protein ligation reactions, opening a wide range of applications in biotechnology, protein engineering, chemoenzymatic synthesis and protein labeling. Recently an asparaginyl endopeptidase (AEP) evolutionary related to butelase 1 named OaAEP1b, also having the ability to link the N- and C-termini of peptidyl substrates, was isolated from the plant *Oldenlandia affinis* and expressed using *E. coli* in an active form following activation at acidic pH. Both butelase 1 and OaAEP1b use the amino acid(s) Asx at the ligation site and have 65% amino-acid sequence identity. However, the reported catalytic efficiency of OaAEP1b is markedly lower than butelase 1 (Harris, K.S., et al., Nat Commun, 2015. 6: p. 10199). In addition, the structural and catalytic mechanisms underlying this family of protein ligases were poorly understood. Therefore, there is still need in the art for new developments that overcome the drawbacks of existing techniques, in particular new methods that allow efficient peptide ligation.

## SUMMARY OF THE INVENTION

The present invention satisfies the aforementioned need in the art by providing the methods described herein.

In a first aspect, the present invention relates to a method of generating a peptide of Formula (I)

$P^1$-Asx-$Xaa^1$-$Xaa^2$-$P^2$ (I),

by ligating a first peptide of Formula (II)

$P^1$-Asx-$Xaa^3$-Leu-$COOH/CONH_2$ (II)

to a second peptide of Formula (III)

$H_2N$-$Xaa^1$-$Xaa^2$-$P^2$ (III),

wherein $P^1$ and $P^2$ are each independently any peptide, modified or unmodified, and optionally can combine such that the peptides of formula (II) and (III) are the termini of the same peptide; Asx is Asp or Asn; $Xaa^1$ is any naturally occurring amino acid; $Xaa^2$ is any naturally occurring amino acid with the exception of Pro, preferably Leu or Ile; and $Xaa^3$ is any naturally occurring amino acid, preferably selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu,
by enzymatically cleaving the bond between "Asx" and "$Xaa^3$" in the first peptide of Formula (II) and ligating the fragment $P^1$-Asx of the first peptide via its C-terminus to the N-terminus of the second peptide of Formula (III) to form a ligated peptide of Formula (I), wherein the enzymatic cleavage and ligation reaction is catalyzed by a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1, also termed "Quicklase" hereinafter) under conditions suitable for said cleavage and ligation reaction.

In various embodiments, the first and second peptides are the termini of the same peptide (i.e. $P^1$ and $P^2$ combine to form a single core peptide sequence) such that the method cyclizes said peptide.

In various embodiments, the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) comprises or consists of:

(a) the amino acid sequence set forth in SEQ ID NO:1;

(b) an amino acid sequence that shares at least 65%, preferably at least 75%, even more preferably at least 85%, most preferably at least 95% sequence identity with, or at least 80%, preferably at least 90%, more preferably at least 95% sequence homology with the amino acid

sequence as set forth in SEQ ID NO:1, provided that said peptide ligase comprises the amino acid sequence set forth in SEQ ID NO:2 at the positions corresponding to residues 247-264 of SEQ ID NO:1;

(c) a functional fragment of (a) or (b); or

(d) an amino acid sequence containing either (a) or (b) or (c) as its essential component,

with the proviso that said peptide ligase is not the wild-type OaAEP1 having the amino acid sequence set forth in SEQ ID NO:3 or butelase 1 having the amino acid sequence set forth in SEQ ID NO:4.

In various embodiments, the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) comprises or consists of:

(a) the amino acid sequence set forth in SEQ ID NO:1;

(b) the amino acid sequence set forth in SEQ ID NO:5;

(c) the amino acid sequence set forth in SEQ ID NO:6;

(d) the amino acid sequence set forth in SEQ ID NO:7;

(e) the amino acid sequence set forth in SEQ ID NO:8;

(f) the amino acid sequence set forth in SEQ ID NO:9;

(g) the amino acid sequence set forth in SEQ ID NO:10; or

(h) the amino acid sequence set forth in SEQ ID NO:11.

In various embodiments, the first and/or second peptide further comprises a labeling moiety.

In various embodiments, the labeling moiety is an affinity tag, therapeutic agent, detectable label, or scaffold molecule.

In various embodiments, the first and/or second peptide may be coupled to a solid support material.

In various embodiments, the first and/or second peptide is a cellular surface protein such that the method results in the modification or tagging of the cellular surface protein and in result the cellular surface as such.

In a second aspect, the present invention relates to a method of preparing a dimer, oligomer, or multimer of one or more peptides of interest, the method comprising the steps of:

(a) providing one or more peptides of interest having C-terminal Asx-Xaa$^3$-Leu-COOH/CONH$_2$ residues and a scaffold molecule having two or more copies of N-terminal H$_2$N-Xaa$^1$-Xaa$^2$ residues or, alternatively, providing one or more peptides of interest having N-terminal H$_2$N-Xaa$^1$-Xaa$^2$ residues and a scaffold molecule having two or more copies of C-terminal Asx-Xaa$^3$-

Leu-COOH/$CONH_2$ residues, wherein Asx is Asp or Asn; $Xaa^1$ is any naturally occurring amino acid; $Xaa^2$ is any naturally occurring amino acid with the exception of Pro, preferably Leu or Ile; and $Xaa^3$ is any naturally occurring amino acid, preferably selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu;

(b) providing a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1);

(c) preparing a mixture of the one or more peptides of interest, the scaffold molecule, and the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1);

(d) subjecting the mixture to conditions that allow the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) to catalyze the ligation of the one or more peptides of interest to the scaffold molecule.

In a third aspect, the present invention relates to a method of modifying or tagging the surface of a target cell by one or more peptides of interest, the method comprising the steps of:

(a) providing the one or more peptides of interest having C-terminal Asx-$Xaa^3$-Leu-COOH/$CONH_2$ residues and/or having N-terminal $H_2N$-$Xaa^1$-$Xaa^2$ residues, wherein Asx is Asp or Asn; $Xaa^1$ is any naturally occurring amino acid; $Xaa^2$ is any naturally occurring amino acid with the exception of Pro, preferably Leu or Ile; and $Xaa^3$ is any naturally occurring amino acid, preferably selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu;

(b) providing a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1);

(c) contacting the target cell with the one or more peptides of interest and the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1);

(d) subjecting the target cell to conditions that allow the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) to catalyze the ligation of the one or more peptides of interest to a cellular surface protein of the target cell.

In various embodiments, the method further comprises removing the unligated one or more peptides from the taget cell after step (d).

In various embodiments, the one or more peptides of interest comprise a labeling moiety, as already described above.

In various embodiments, the target cell recombinantly expresses a surface polypeptide having N-terminal $H_2N$-$Xaa^1$-$Xaa^2$ residues or C-terminal Asx-$Xaa^3$-Leu-COOH residues for ease of tagging by the peptide of interest.

In various embodiments of all afore-described aspects of the invention,

(a) Asx is Asn or Asp; and/or

(b) $Xaa^1$ is all naturally occouring amino acids; and/or

(c) $Xaa^2$ is Leu or Ile; and/or

(d) $Xaa^3$ is selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu.

In preferred embodiments, two or more of features (a)-(d) are met, i.e. (a) and (b), (a) and (c), (a) and (d), (b) and (c), (b) and (d), or three or more features are met, i.e. (a), (b) and (c), (a), (b) and (d), (a), (c) and (d), (b), (c) and (d) or, most preferably all four are met.

In preferred embodiments, Asx is Asn, $Xaa^1$ is any naturally occurring amino acid, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ser, Cys, Gly or Ala.

In more preferred embodiments, Asx is Asn, $Xaa^1$ is Arg or Gly, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ala or Gly.

In a fourth aspect, the invention relates to the ligated peptides and/or tagged target cells obtainable according to any of the methods of the invention.

In a fifth aspect, the invention relates to the peptide ligase OaAEP1 Cys247Ala having the amino acid sequence of SEQ ID NO:1 and other peptide ligases having the activity thereof, as described above, as well as kits comprising any of said peptide ligases, in particular OaAEP1 Cys247Ala (SEQ ID NO:1).

In still another aspect, the present invention relates to the use of a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) for the ligation of two peptides, as described herein.

It is understood that all embodiments disclosed herein in relation to one aspect of the invention, are similarly applicable to all other aspects of the invention.

## BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings.

Figure 1 shows a schematic illustration of (a) OaAEP1 Cys247Ala (SEQ ID NO:1)-mediated peptide ligation and (b) a preferred embodiment thereof.

Figure 2 shows purification and activation of OaAEP1b. (a) Schematic illustration of the His-Ub-OaAEP1 construct used with the boundaries of the three regions of the protein indicated and the deduced activation segment. 6-His tag and Ubiquitin tag followed by OaAEP1 (24-474) was synthesized and inserted into pET28b vector. After Ni-NTA affinity purification, about half of Ub-AEP-FL lost the N-terminal fusion tags and was purified in AEP-FL form (53-474), which was eventually crystallized. The enzymatic active form of OaAEP1b, AEP-holo (53-347/351) is the acidic induced self-cleavage product of AEP-FL, which lost the C-terminal auto inhibitory cap domain. (b) SDS PAGE analysis of fractions obtained over the course of chromatographic purification and of dissolved crystals. (c) Size-exclusion chromatography profiles of OaAEP1 before and after activation showing Ub-AEP-FL and AEP-FL heterodimer and the AEP-holo monomer. The elution volumes of standard proteins are indicated for comparison. (d) SDS PAGE analysis of OaAEP1 pH-dependent activation. The activation was performed for 3 hours, and the peak of activation is in between pH 3.4 to 4. (e) *Trans* activation/maturation of the inactive Cys217Ser active site mutant by the active recombinant mature OaAEP1 enzyme. The active site Cys217Ala mutant is devoid of self-activation activity. Addition of 5% AEP-FL wt enable trans-cleavage and activation of most of the AEP-FL Cys217Ala after 18 hours incubation at pH 3.7.

Figure 3 shows a crystal structure of OaAEP1 in its zymogenic form. (a) Overall "front" view of the OaAEP1 structure represented as "cartoon" (with $\alpha$-helices as ribbons and $\beta$-strands as arrows) with the catalytic core domain colored in green, the pro-domain in purple. The flexible linker spanning residues 325-343 (not visible) is depicted as dashed line and residues 344-347 that are trapped in the OaAEP1 active site are depicted as sticks. The molecular surface views in panels (b) and (c) were obtained by separating the core domain from the cap domain and rotating by an angle of 90° in opposite directions respectively (indicated by arrows). (d) The OaAEP1 inactive dimer is stabilized via intermolecular interactions provided by the cap region. Residues 344-347 that are bound in the active site are depicted as sticks and labeled. Compared to human legumain, these linker residues are bound at the catalytic center in an orientation similar to peptide based legumain inhibitors. From this comparative view, the putative channel that accommodates the incoming amine could be identified on the core domain surface. The orientation of an incoming substrate poised to undergo the ligation reaction is indicated by an arrow along the channel at the protein surface (see text). (e) The hlegum molecular surface is shown in the same orientation as OaAEP1 with the peptide Ac-Tyr-Val-Ala-Asp-CMK (sticks) covalently bound to the active site Cys189 (from PDB code: 4AWA). Here, the incoming amine channel is completely blocked, explaining the lack of significant ligase activity. (f) A cartoon highlight the two critical structural features related to the substrates binding and catalytic activities of OaAEP1 Cys247Ala. P1 binding site is responsible for efficient cleavage, and P2 binding site is related to efficient protein ligation.

Figure 4 shows the OaAEP1 active-site and the activation segment. (a) Close-up view of the OaAEP1 active site with active site residues discussed in the text displayed as sticks and labeled: The canonical AEP catalytic pair His175 and Cys217 and the linker residues 344-347 bound at the catalytic center are depicted as sticks. Also shown the disulfide bond formed by Cys250 and Cys264 (conserved in butelase 1), which forms an open amine incoming channel and the "gate-keeper" Cys247 a major determinant for the ligase activity. Hydrogen bonds formed by P1 residue Gln347 which is bound in the S1 pocket are depicted by dashed lines and interatomic distances are displayed in Å. Panel (b) displays an electron density map with the segment 344-347 of the polypeptide omitted from the phase calculation and contoured at $3\sigma$ level. (c) Residual activation of OaAEP1 mutants following site-directed mutagenesis targeting the activation segment 346-351. Enzymatic behavior of linker mutants: Single or double mutations of Asn346, Gln347, Asp349 and Asp351 demonstrated moderate loss of self-cleavage induced activation. Triple mutant A3 (Asn346Ala, Asp349Ala and Asp351Ala) and quadruple mutant A4 (Asn346, Gln347, Asp349 and Asp351) have more severe loss of function phenotype, indicating the site of activation could be any residue with this stretch of linker. (d) An histogram of activation efficiency after 6 hours is shown for the linker mutants. Asp349 and Asp351 could be the more preferred cleavage site, while Asn346 and Gln347 are potential cleavage sites as well.

Figure 5 shows comparison of peptide cyclization activity of WT OaEP1b, Cys247Ala and butelase 1 and the influence of the gate-keeper on ligase activity. (a) Kinetic parameters for the cyclization of "GLPVSTKPVATRNGL" by WT OaEP1b, Cys247Ala and butelase 1. MS analysis of the reaction is given in panel (b). (c) Role of the gate-keeper in OaAEP1 activation. (d) Comparison of cyclization efficiency for various residues at the gate-keeper Cys247 position. The Gate-keeper residue Cys247 governs the ligation reaction kinetics. (a) The kinetic measurement of wild type OaAEP1 and Cys247Ala mutant, in comparison with Butelase 1 freshly extracted from plant. Cys247Ala has the Kcat/Km 160 times faster than wt OaAEP1b. (b) The schematic view of the peptide used to determine the ligation efficiency, together with the mass-spec results demonstrating the backbone ligation achieved by wild type OaAEP1 and Cys247Ala mutant. (c) The self-activation of the gate-keeper mutants. The self-activation is not much affected by mutations of the gate-keeper residue. Bulky side chain mutants have difficulties been fully activated. (d) The ligation catalytic activity is reversely correlated with the size of the side chain of Cys247 mutants.

Figure 6 shows protein ligase activity of Cys247Ala. (a) Fluorescence detection using O2 benzylcytosine-Alexa647 of the SNAP protein and SDS PAGE analysis with Comassie blue staining detection of the protein substrate and ligated products. The reaction is schematically

depicted as inset. Cys247Ala ligates well folded protein well. (a) 100 nM of Cys247Ala, but not 4000 nM of wild type OaAEP1b, efficiently ligated 5 μM C-terminal modified SNAP tag protein with N-terminal modified Ubiquitin. (b) As little as 30 nM Cys247Ala catalyze ligation of 100 μM Ubiquitin and 500 μM biotin labelled peptide efficiently in 30 min, at room temperature, in pH 7.4 PBS buffer.

Figure 7 shows (a) sequence alignment of OaAEP1b, Butelase 1 and human legumain. (b) The exact amino acid sequence encoded in our OaAEP1 construct. Residues -85 to 0 is the fusion tags (colored in cyan), while the first 23 amino acids in AEP coding was omitted. The four residues inserted at the active site pocket observed in our crystal structure are colored in orange, and the cap domain is displayed in violet.

Figure 8 shows the 2FO-FC electron density map, highlighting the good fitting of the model. (a) side chains in the cap domain are well fit with the density. (b) the novel PPP loop in the core domain, which is unique to AEP ligases are nicely fit with the density.

Figure 9 shows the raw mass-spectrometry results of Gate-Keeper mutants.

Figure 10 shows the potential side product of the ligation reaction. If there is no available N-terminal amine, side chain ligation and proteolysis could happen in Cys247Gly and Cys247Ala mutants. The smaller size of the side chain of Cys247Gly mutant permitted water molecules to participate in the reaction and results proteolysis.

Figure 11 shows a screen for substrate sequence specificity of the C- or N- terminal end peptide/protein: (a) Asn-Xaa$^3$-Leu-COOH; (b) Asn-Gly-Xaa$^2$-COOH; and (c) $H_2N$-Xaa$^1$-Leu. 100 μM substrate, with 50 nM Quicklase, reacted for 5 mins at room temperature, under neutral pH (pH 7.4) standard reaction buffer. The % of self-ligated (cyclized) peptides could be used as qualitative indications of Kcat. The raw Mass-spec reading of the assay readout were also included in Figure 11(d).

Figure 12 shows (a) a graphic summary of a Basic Local Alignment Search Tool (BLAST) search in the National Center for Biotechnology Information (NCBI) database performed with the amino acid sequence set forth in SEQ ID NO:1 (OaAEP1 Cys247Ala) and (b) the sequences producing significant alignments.

Figure 13 shows that AEP_C13CO4 having the amino acid sequence set forth in SEQ ID NO:8 is a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1).

Figure 14 shows schemes for (a) scaffold-based one-step homodimerization, (b) scaffold-based two-step heterodimerization, (c) scaffold-based one-step tetramerization, and (d) modular ultra-oligomerization. Additionally shown are schemes for preparing (e, f) NGL head-to-head dimers, (g) GL linker dimers and tetramers, (h) NGL multimers, and (i) GL multimers.

Figure 15 shows (a) a schematic illustration of the direct cellular modification procedure mediated by OaAEP1 Cys247Ala (SEQ ID NO:1). (b) Fluorecently tagged SNAP (New England Biolabs) protein was conjugated to the intrinsic cellular surface proteins of two different cell lines, RAW cell (human macrophage cell line) and HEK293T (human epithelial cell line). 10 μM modified SNAP protein (with additional C-terminal NGL residues, lane 1 sample) with or without (lane 2 and lane 5 samples) coincubation of 0.1 μM Quicklase were added to cell culture, pre-washed with PBS buffer. After 30 minutes ~ 1 hour incubation, cells were rinsed with additional PBS buffer (lane 3 and lane 6 were first wash samples), two times. The remaining cell was harvested and lysed with SDS loading buffer (lane 4 and 7 samples). All samples were examined using SDS-PAGE, and only the covalently fluorescently tagged SNAP protein carried the signal on the gel, which were shown as dark bands in (b). Additional bands in lane 4 and 7 indicated that significant amount of intrinsic proteins in RAW and HEK283T cells were conjugated with SNAP protein, in a Quicklase dependant way. (c) Surface modified RAW cells were examined using a fluorescent microscope, bright spots on the cellular surface indicated successful surface labelling of intrinsic cellular membrane. Certain intrinsic proteins with amino acids residue sequences closely related to formula III could be recognized as Quicklase, due to its superior enzymatic activity and broader sequence regonition ability.

## DETAILED DESCRIPTION OF THE INVENTION

The following detailed description refers to, by way of illustration, specific details and embodiments in which the invention may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention. Other embodiments may be utilized and structural, and logical changes may be made without departing from the scope of the invention. The various embodiments are not necessarily mutually exclusive, as some embodiments can be combined with one or more other embodiments to form new embodiments.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art. The singular terms "a," "an," and "the" include plural referents unless context clearly indicates otherwise. Similarly, the word "or" is intended to include "and" unless the context clearly indicates otherwise. The term "comprises"

means "includes." In case of conflict, the present specification, including explanations of terms, will control.

Breaking and forming peptidyl bonds are fundamental biochemical reactions in protein chemistry. Unlike proteases that are abundantly available, fast-acting ligases are rare. OaAEP1 is an enzyme isolated from the cyclotide-producing plant oldenlandia affinis that displayed weak peptide cyclase activity, despite having a similar structural fold with other asparaginyl endopeptidases (AEP). The inventors of the present application report herein the first atomic structure of OaAEP1, at a resolution of 2.56 Å, in its pre-activation form. The structure and biochemical analysis of this enzyme reveals its activation mechanism as well as structural features important for its ligation activity. Importantly, through structure-based mutagenesis of OaAEP1, the inventors obtained an ultra-fast variant having hundreds of times faster catalytic kinetics, capable of ligating well-folded protein substrates using only sub-micro molar concentration of enzyme. In contrast, the protein-protein ligation activity in the original wild-type OaAEP1 enzyme described previously is extremely weak. Thus, the structure-based identification of a biochemical motif described herein leads to characterization of a unique and novel recombinant tool (OaAEP1 Cys247Ala having the amino acid sequence set forth in SEQ ID NO:1) and its related structural homologs that can now be used to conduct various protein labeling and modifications that were extremely challenging before. For example, more efficient peptide cyclization and peptides/proteins can be enzymatically catalyzed. Furthermore, this new type of recombinant enzyme, and its structural homologs enable novel biochemical applications, like modifications of intrinsic cellular surface proteins.

The object of the present invention is to provide a technique for ligating peptides, taking advantage of the superior site specificity and catalytic efficiency of OaAEP1 Cys247Ala-like Asx-specific protein ligases.

To this end, provided in a first aspect of the present invention is a method of generating a peptide of Formula (I)

$$P^1\text{-Asx-Xaa}^1\text{-Xaa}^2\text{-}P^2 \quad (I),$$

by ligating a first peptide of Formula (II)

$$P^1\text{-Asx-Xaa}^3\text{-Leu-COOH/CONH}_2 \quad (II)$$

to a second peptide of Formula (III)

$H_2N$-$Xaa^1$-$Xaa^2$-$P^2$ (III),

wherein $P^1$ and $P^2$ are each independently any peptide, modified or unmodified, and optionally can combine such that the peptides of formula (II) and (III) are the termini of the same peptide; Asx is Asp or Asn; $Xaa^1$ is any naturally occurring amino acid; $Xaa^2$ is any naturally occurring amino acid with the exception of Pro, preferably Leu or Ile; and Xaa3 is any naturally occurring amino acid, preferably selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu,

by enzymatically cleaving the bond between "Asx" and "$Xaa^3$" in the first peptide of Formula (II) and ligating the fragment $P^1$-Asx of the first peptide via its C-terminus to the N-terminus of the second peptide of Formula (III) to form a ligated peptide of Formula (I), wherein the enzymatic cleavage and ligation reaction is catalyzed by a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) under conditions suitable for said cleavage and ligation reaction.

In the context of the whole application, the terms "peptide", "polypeptide", and "protein" are used interchangeably to refer to polymers of amino acids of any length connected by peptide bonds. The polymer may comprise modified amino acids, it may be linear or branched, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified naturally or artificially; for example, by disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation to a labeling moiety. However, in preferred embodiments, these terms relate to polymers of naturally occurring amino acids, as defined below, which may optionally be modified as defined above, but does not comprise non-amino acid moieties in the polymer backbone.

The term "amino acid" refers to natural and/or unnatural or synthetic amino acids, including both the D and L optical isomers, amino acid analogs (for example norleucine is an analog of leucine) and derivatives known in the art. The term "naturally occurring amino acid", as used herein, relates to the 20 naturally occurring L-amino acids, namely Gly, Ala, Val, Leu, Ile, Phe, Cys, Met, Pro, Thr, Ser, Glu, Gln, Asp, Asn, His, Lys, Arg, Tyr, and Trp. The term "peptide bond" refers to a covalent amide linkage formed by loss of a molecule of water between the carboxyl group of one amino acid and the amino group of a second amino acid. The "-" between $P^1$ and Asx in Formula (I), as well as all "-" in Formula (I) represent peptide bonds. In addition, "Leu-$COOH/CONH_2$" represents Leucine or Leucinamide, so the "-" between "Leu" and "$COOH/CONH_2$" in Formula (II) represents the covalent bond between the $\alpha$–carbon and the carboxyl or carboxamide group of Leucine/Leucinamide. Likewise, "$H_2N$-$Xaa^1$" in Formula (III) represents amino acid residue $Xaa^1$, so the "-" between "$NH_2$" and "$Xaa^1$" represents the covalent bond between the $\alpha$–carbon and the amino group of $Xaa^1$. It is understood that the carboxyl/carboxamide group in the peptide of formula (II) and the amino group in the peptide of

formula (III) form part of the terminal amino acid and are not additional functional groups. Generally, in all formulae depicted herein, the peptides are shown in the N- to C-terminal orientation.

Without wishing to be bound to any particular theory, it is believed that the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) as described herein can also catalyze the ligation between any one substance or object comprising a terminal "Asx-$Xaa^3$-Leu-$COOH/CONH_2$" motif and any one substance or object comprising a terminal "$H_2N$-$Xaa^1$-$Xaa^2$" motif. In this connection, it should be noted that the method described herein also applies *mutatis mutandis* to embodiments wherein one or both of $P^1$ and $P^2$ are any substance or object other than peptide, which is also within the scope of the present application.

In various embodiments,

(a) Asx is Asn or Asp; and/or

(b) $Xaa^1$ is all naturally occouring amino acids; and/or

(c) $Xaa^2$ is Leu or Ile; and/or

(d) $Xaa^3$ is selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu.

In preferred embodiments, two or more of features (a)-(d) are met, i.e. (a) and (b), (a) and (c), (a) and (d), (b) and (c), (b) and (d), or three or more features are met, i.e. (a), (b) and (c), (a), (b) and (d), (a), (c) and (d), (b), (c) and (d) or, most preferably all four are met.

In preferred embodiments, Asx is Asn, $Xaa^1$ is any naturally occurring amino acid, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ser, Cys, Gly or Ala.

In more preferred embodiments, Asx is Asn, $Xaa^1$ is Arg or Gly, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ala or Gly, meaning that Asx is Asn, $Xaa^1$ is Arg, $Xaa^2$ is Leu, $Xaa^3$ is His; Asx is Asn, $Xaa^1$ is Arg, $Xaa^2$ is Leu, $Xaa^3$ is Ala; Asx is Asn, $Xaa^1$ is Arg, $Xaa^2$ is Leu, $Xaa^3$ is Gly; Asx is Asn, $Xaa^1$ is Gly, $Xaa^2$ is Leu, $Xaa^3$ is His; Asx is Asn, $Xaa^1$ is Gly, $Xaa^2$ is Leu, $Xaa^3$ is Ala; or Asx is Asn, $Xaa^1$ is Gly, $Xaa^2$ is Leu, $Xaa^3$ is Gly.

It should be noted that, according to Figure 11(b), the "Leu" residue of the first peptide of Formula (II), i.e. $P^1$-Asx-$Xaa^3$-Leu-$COOH/CONH_2$, may also be replaced by another naturally occurring amino acid, e.g. Ile, Pro, Phe, Cys, Gln, or Lys, provided that said peptide in this form is still sutiable for the method described herein. Such embodiments in combination with any of the embodiments described herein are also fully encompassed by the present application.

It is also envisaged that, when the first peptide and the second peptide are termini of the same peptide, i.e. $P^1$ and $P^2$ combine to form the core sequence of a peptide terminated by the sequences of formula (II) and (III), the presently disclosed method cyclizes said peptide.

The peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) for use in the present invention may be any oaAEP family enzymes or homologous thereof having the desired activity. The peptide ligase has an ability to site-specifically break a peptide bond and then reform a new bond with an incoming nucleophile. It is Asx-specific in that the C-terminal amino acid to which ligation occurs, i.e. the C-terminal end of the peptide that is ligated, is either Asn or Asp, preferably Asn. As set forth above, it recognizes the motif Asx-$Xaa^3$-Leu-$COOH/CONH_2$, at the C-terminus of the first peptide of Formula (II), and mediates peptide ligation by cleaving off the sorting signal $Xaa^3$-Leu-$COOH/CONH_2$ and ligating $P^1$-Asx to the N-terminal residue of the second peptide $H_2N$-$Xaa^1$-$Xaa^2$-$P^2$ to form a ligated peptide $P^1$-Asx-$Xaa^1$-$Xaa^2$-$P^2$.

In various embodiments, the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) in accordance with the present application comprises, consists of, or essentially consists of:

(a) the amino acid sequence set forth in SEQ ID NO:1 (OaAEP1 Cys247Ala);

(b) an amino acid sequence that shares at least 65%, preferably at least 75%, even more preferably at least 85%, most preferably at least 95% sequence identity with, or at least 80%, preferably at least 90%, more preferably at least 95% sequence homology with the amino acid sequence as set forth in SEQ ID NO:1, provided that said peptide ligase comprises the amino acid sequence set forth in SEQ ID NO:2 at the positions corresponding to residues 247-264 of SEQ ID NO:1, which defines a unique surface structural feature of the enzyme serving as the binding site for approaching N-terminal amino group;

(c) a functional fragment of (a) or (b); or

(d) an amino acid sequence containing either (a) or (b) or (c) as its essential component,

with the proviso that said peptide ligase is not the wild-type OaAEP1 having the amino acid sequence set forth in SEQ ID NO:3 or butelase 1 having the amino acid sequence set forth in SEQ ID NO:4.

In various embodiments, the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) comprises, consists of, or essentially consists of:

(a) the amino acid sequence set forth in SEQ ID NO:1;

(b) the amino acid sequence set forth in SEQ ID NO:5;

(c) the amino acid sequence set forth in SEQ ID NO:6;

(d) the amino acid sequence set forth in SEQ ID NO:7;

(e) the amino acid sequence set forth in SEQ ID NO:8;

(f) the amino acid sequence set forth in SEQ ID NO:9;

(g) the amino acid sequence set forth in SEQ ID NO:10; or

(h) the amino acid sequence set forth in SEQ ID NO:11.

In preferred embodiments, the peptide ligase comprises or consists of the amino acid sequence as set forth in SEQ ID NO:1 (OaAEP1 Cys247Ala).

The identity of nucleic acid or amino acid sequences is generally determined by means of a sequence comparison. This sequence comparison is based on the BLAST algorithm that is established in the existing art and commonly used (cf. for example Altschul et al. (1990) "Basic local alignment search tool", J. Mol. Biol. 215:403-410, and Altschul et al. (1997): "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs"; Nucleic Acids Res., 25, p. 3389-3402) and is effected in principle by mutually associating similar successions of nucleotides or amino acids in the nucleic acid sequences and amino acid sequences, respectively. A tabular association of the relevant positions is referred to as an "alignment." Sequence comparisons (alignments), in particular multiple sequence comparisons, are commonly prepared using computer programs which are available and known to those skilled in the art.

A comparison of this kind also allows a statement as to the similarity to one another of the sequences that are being compared. This is usually indicated as a percentage identity, which is calculated in relation to a reference sequence and its entire length. The term "sequence identity" refers to the extent that sequences are identical on a nucleotide-by-nucleotide or amino acid-by-amino acid basis over a window of comparison. Thus, a "percentage of sequence identity"' is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The more broadly construed term "homology", in the context of amino acid sequences, also incorporates consideration of the conserved amino acid exchanges, i.e. amino acids having a similar chemical activity, since these usually perform similar chemical activities within the protein. The similarity of the compared sequences can therefore also be indicated as a "percentage homology" or "percentage similarity." Indications of identity and/or homology can be encountered over entire polypeptides or genes, or only over individual regions. Homologous and identical regions of various nucleic acid sequences or amino acid sequences are therefore defined by way of matches in the sequences. Such regions often exhibit identical functions. They can be small, and can encompass only a few nucleotides or amino acids. Small regions of

this kind often perform functions that are essential to the overall activity of the protein. It may therefore be useful to refer sequence matches only to individual, and optionally small, regions. Unless otherwise indicated, however, indications of identity and homology herein refer to the full length of the respectively indicated nucleic acid sequence or amino acid sequence.

While it is recognized that various peptide ligases as described above may be suitable for the practice of the present invention, it is preferable to use one with potent protein ligase actively. In various embodiments, this means that it can ligate a given peptide with an efficiency of at least 50%, preferably at least 70%, more preferably at least 90%, most preferably at least 95%. Methods to determine such efficiency by, for example, ligating substrate (100 μM) in the presence of said peptide ligase (50 nM) for 30 mins in a standard reaction buffer at neutral pH and room temperature, are well known in the art and can be routinely applied by those skilled in the art, for example. It is preferred that the peptide ligases of the invention have at least 50%, more preferably at least 70%, most preferably at least 90% of the protein ligase activity of the enzyme having the amino acid sequence of SEQ ID NO:1.

Peptide ligases having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) according to the present application can comprise amino acid modifications, in particular amino acid substitutions, insertions, or deletions. Such peptide ligases are, for example, further developed by targeted genetic modification, i.e. by way of mutagenesis methods, and optimized for specific purposes or with regard to special properties (for example, with regard to their catalytic activity, stability, etc.). The objective may be to introduce targeted mutations, such as substitutions, insertions, or deletions, into the known molecules in order, for example, to alter substrate specificity and/or improve the catalytic activity. For this purpose, in particular, the surface charges and/or isoelectric point of the molecules, and thereby their interactions with the substrate, can be modified. Alternatively or additionally, the stability of the peptide ligase can be enhanced by way of one or more corresponding mutations, and its catalytic performance thereby improved. Advantageous properties of individual mutations, e.g. individual substitutions, can supplement one another.

In various embodiments, the peptide ligase may be characterized in that it is obtainable from a peptide ligase as described above as an initial molecule by single or multiple conservative amino acid substitution. The term "conservative amino acid substitution" means the exchange (substitution) of one amino acid residue for another amino acid residue, where such exchange does not lead to a change in the polarity or charge at the position of the exchanged amino acid, e.g. the exchange of a nonpolar amino acid residue for another nonpolar amino acid residue. Conservative amino acid substitutions in the context of the invention encompass, for example, G=A=S, I=V=L=M, D=E, N=Q, K=R, Y=F, and S=T.

Alternatively or additionally, the peptide ligase may be characterized in that it is obtainable from a peptide ligase contemplated herein as an initial molecule by fragmentation or by deletion, insertion, or substitution mutagenesis, and encompasses an amino acid sequence that matches the initial molecule over a length of at least 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, or 300 continuously connected amino acids, provided that said peptide ligase comprises the amino acid sequence set forth in SEQ ID NO:2 at the positions corresponding to residues 247-264 of SEQ ID NO:1.

In various embodiments, the present invention thus also relates to functional fragments of the peptide ligases described herein, with said fragments retaining enzymatic activity. It is preferred that they have at least 50%, more preferably at least 70%, most preferably at least 90% of the protein ligase activity of the initial molecule, preferably of the peptide ligase having the amino acid sequence of SEQ ID NO:1. The functional fragments are preferably at least 150 amino acids in length, more preferably at least 180 or 200, most preferably at least 250.

In various embodiments, the peptides to be ligated in accordance with the present application may be modified by, for example, conjugation to a labeling moiety, either covalently or non-covalently. A labeling moiety may be any molecules such as, without limitation, an affinity tag, therapeutic agent, detectable label, or scaffold molecule.

In various embodiments, the first and/or second peptide may be coupled to a solid support material.

The term "affinity tag" as used herein refers to a moiety such as biotin that can be used to separate a molecule to which the affinity tag is attached from other molecules that do not contain the affinity tag.

The term "detectable label" is intended to mean at least one label capable of directly or indirectly generating a detectable signal. In non-limiting examples, a detectable label can be an enzyme producing a detectable signal, for example by colorimetry, fluorescence, or luminescence; a chromophore, such as a fluorescent, luminescent or dye compound, e.g. GFP; a group with an electron density detectable by electron microscopy or by virtue of their electrical property, such as conductivity, amperometry, voltammetry or impedance; detectable group, for example the molecules of which are sufficiently large to induce detectable modifications of their physical and/or chemical characteristics (this detection can be carried out by optical methods such as diffraction, surface plasmon resonance, surface variation or contact angle variation, or physical

methods such as atomic force spectroscopy or the tunnel effect; or a radioactive molecule such as $^{32}P$, $^{35}S$ or $^{125}I$.

The term "scaffold molecule" as used herein refers to a compound to which other moieties are attached, covalently or non-covalently. Various scaffold molecules such as dendrimers are well known in the art.

The term "solid support material" as used herein refers to a solid or semi-solid (e.g., a hydrogel) material onto which the peptide can be immobilized. Non-limiting examples include solid supports for peptide synthesis, magnetic beads, glass fibers, and resins.

Peptides modified by the labeling moiety may be prepared using standard techniques known to those skilled in the art of synthetic organic chemistry, or may be deduced by reference to the pertinent literature.

In various embodiments, the first and/or second peptide is a cellular surface protein such that the method results in the modification or tagging of the cellular surface protein and in result the cellular surface as such.

In a second aspect, the invention relates to a method of preparing a dimer, oligomer, or multimer of one or more peptides of interest, the method comprising the steps of:

(a) providing one or more peptides of interest having C-terminal Asx-$Xaa^3$-Leu-$COOH/CONH_2$ residues and a scaffold molecule having two or more copies of N-terminal $H_2N$-$Xaa^1$-$Xaa^2$ residues or, alternatively, providing one or more peptides of interest having N-terminal $H_2N$-$Xaa^1$-$Xaa^2$ residues and a scaffold molecule having two or more copies of C-terminal Asx-$Xaa^3$-Leu-$COOH/CONH_2$ residues, wherein Asx is Asp or Asn; $Xaa^1$ is any naturally occurring amino acid; $Xaa^2$ is any naturally occurring amino acid with the exception of Pro, preferably Leu or Ile; and $Xaa^3$ is any naturally occurring amino acid, preferably selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu;

(b) providing a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1), as described above;

(c) preparing a mixture of the one or more peptides of interest, the scaffold molecule, and the peptide ligase having the activity of OaAEP1 Cys247Ala(SEQ ID NO:1);

(d) subjecting the mixture to conditions that allow the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) to catalyze the ligation of the one ore more peptides of interest to the scaffold molecule.

In various embodiments,

(a) Asx is Asn or Asp, and/or

(b) $Xaa^1$ is all naturally occouring amino acids, and/or

(c) $Xaa^2$ is Leu or Ile, and/or

(d) $Xaa^3$ is selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu.

In preferred embodiments, two or more of features (a)-(d) are met, i.e. (a) and (b), (a) and (c), (a) and (d), (b) and (c), (b) and (d), or three or more features are met, i.e. (a), (b) and (c), (a), (b) and (d), (a), (c) and (d), (b), (c) and (d) or, most preferably all four are met.

In preferred embodiments, Asx is Asn, $Xaa^1$ is any naturally occurring amino acid, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ser, Cys, Gly or Ala.

In more preferred embodiments, Asx is Asn, $Xaa^1$ is Arg or Gly, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ala or Gly.

Figures 14 (a)-(d) show how hetero- or homo- dimers, oligomers, or multimers of one or more peptides of interest can be prepared according to some non-limiting embodiments. Figures 14 (e)-(i) show some non-limiting examples of how scaffold molecules for use in the present invention can be prepared.

In a third aspect, the invention relates to a method of modifying or tagging the surface of a target cell by one or more peptides of interest, the method comprising the steps of:

(a) providing the one or more peptides of interest having C-terminal Asx-$Xaa^3$-Leu-$COOH/CONH_2$ residues and/or having N-terminal $H_2N$-$Xaa^1$-$Xaa^2$ residues, wherein Asx is Asp or Asn; $Xaa^1$ is any naturally occurring amino acid; $Xaa^2$ is any naturally occurring amino acid with the exception of Pro, preferably Leu or Ile; and $Xaa^3$ is any naturally occurring amino acid, preferably selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu;

(b) providing a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1), as described above;

(c) contacting the target cell with the one or more peptides of interest and the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1);

(d) subjecting the target cell to conditions that allow the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) to catalyze the ligation of the one or more peptides of interest to a cellular surface protein of the target cell.

In certain embodiments, the wild-type OaAEP1 having the amino acid sequence of SEQ ID NO:3 may also be used in this method, probably with a compromised efficiency.

In various embodiments,

(a) Asx is Asn or Asp; and/or

(b) $Xaa^1$ is all naturally occouring amino acids; and/or

(c) $Xaa^2$ is Leu or Ile; and/or

(d) $Xaa^3$ is selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu.

In preferred embodiments, two or more of features (a)-(d) are met, i.e. (a) and (b), (a) and (c), (a) and (d), (b) and (c), (b) and (d), or three or more features are met, i.e. (a), (b) and (c), (a), (b) and (d), (a), (c) and (d), (b), (c) and (d) or, most preferably all four are met.

In preferred embodiments, Asx is Asn, $Xaa^1$ is any naturally occurring amino acid, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ser, Cys, Gly or Ala.

In more preferred embodiments, Asx is Asn, $Xaa^1$ is Arg or Gly, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ala or Gly.

In various embodiments, the method further comprises removing the unligated one or more peptides from the taget cell after step (d).

The one or more peptides of interest may be functionalized to bind a variety of cargo molecules. In various embodimetns, the one or more peptides of interest comprise a labeling moiety such as an affinity tag, therapeutic agent, detectable label, or scaffold molecule, as already described above. The one or more peptides of interest may also be coupled to a solid support material.

Figure 15(a) shows how a target cell can be tagged according to a non-limiting embodiment of the invention.

The method described herein may be applicable to all types of cells *in vitro*, *ex vivo*, or *in vivo*. The target cell may be any prokaryotic or eukaryotic cell, e.g. a bacterial, yeast, plant, or human cell; it may be a cancer cell, oocyte, embryonic stem cell, hematopoietic stem cell, or any other differentiated or undifferentiated cell, provided that the target cell expresses a surface polypeptide having C-terminal Asx-$Xaa^3$-Leu-COOH/$CONH_2$ or N-terminal $H_2N$-$Xaa^1$-$Xaa^2$ residues that are accessible to the one or more peptides of interest.

In preferred embodiments, the target cell expresses endogenous surface proteins suited for ligation to the one or more peptides of interest. However, the target cell may also recombinantly express a surface polypeptide having N-terminal $H_2N$-Xaa$^1$-Xaa$^2$ or C-terminal Asx-Xaa$^3$-Leu-COOH residues for ease of tagging by the peptide of interest.

The term "recombinantly express" as used herein refers to the expression of said polypeptide by recombinant DNA technology.

It is also within the scope of the present invention that the one or more peptides of interest may be polypeptides endogenously or recombinantly expressed on the surface of the target cell, in which case no additional said peptides need to be provided and the method described herein results in hetero- or homo- dimerization, oligomerization, or multimerization of surface proteins of the target cell.

It is therefore believed that the present invention provides a versatile and fast-acting technology for modifying or tagging cell surface by attaching modified or unmodified peptides of interest. Compared to conventional chemical labeling strategies, this method enables specific and fast conjugation to the N- and/or C-terminus of surface proteins.

In a fourth aspect, the invention relates to the ligated peptides and/or tagged target cells obtainable according to any of the methods of the invention.

In a fifth aspect, the invention relates to the peptide ligase OaAEP1 Cys247Ala having the amino acid sequence of SEQ ID NO:1 and other peptide ligases having the activity thereof, as described above, as well as kits comprising any of said peptide ligases, in particular OaAEP1 Cys247Ala (SEQ ID NO:1).

In still another aspect, the present invention relates to the use of a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) for the ligation of two peptides, as described herein.

The present invention is further illustrated by the following examples. However, it should be understood, that the invention is not limited to the exemplified embodiments.

## EXAMPLES

### *Materials and methods*

**Quicklase preparation:**

OaAEP1 was cloned and expressed in *E. coli* as an ubiquitin fusion protein. A gene with codons optimized for expression in E. coli, encoding a protein (SEQ ID NO:12) composed of a N-terminal hexa-His tag (SEQ ID NO:13), the 76 amino-acid residues human ubiquitin (SEQ ID NO:14), and residues 24-474 of OaAEP1 (SEQ ID NO:15, devoid of the OaAEP1 signal peptide) was synthesized by Genescript, USA.

The clone is OaAEP1 one amino acid different from OaAEP1b (SEQ ID NO:3) (with E371V mutation). OaAEP1 was cloned from mRNA and OaAEP1b from genomic DNA sequence. Nevertheless, the wild type constructs (OaAEP1, SEQ ID NO:3) of the inventors behaves essentially the same as previously reported (OaEP1b, SEQ ID NO:3 with E371V mutation).

The amino acid sequence of this composite protein construct is shown in Fig. S1B. The gene was inserted into the coding region (NcoI-NdeI) of the pET-28b (+) vector and amplified. Recombinant expression of OaAEP1 in E. coli (using either Bl21DE3 or T7 Shuffle from New England Biolabs, USA) was performed following the protocol reported previously, using a concentration of 0.4 mM IPTG to induce protein expression at 16 °C for 20 hours. Cells were then harvested by centrifugation at 4,000 g for 15 min at 4 °C and resuspended in a lysis buffer containing 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.05% (v/v) CHAPS, 10 % (v/v) glycerol. Lysis was done by passing three times the homogenized cells through an Emulsiflex-C3 (Avestin, USA) high-pressure apparatus, at 1500 psi. The supernatant fraction, which contains His-Ub-AEP-FL (SEQ ID NO:12) and Ub-AEP-FL (SEQ ID NO:15) was further purified by metal affinity using a Ni-NTA column (Bio-Rad). His tag-containing proteins bound to the column were eluted using a linear imidazole gradient from 0 to 500 mM in a buffer containing 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.05% (v/v) CHAPS, 10 % (v/v) glycerol. Ni-NTA elution fractions containing the protein were diluted and loaded onto two 5 ml HiTrap Q Sepharose high performance columns connected in series (GE Healthcare; 2 ml sample per ml of resin). Bound proteins were eluted using a continuous salt gradient of 0–30% of buffer B: 20 mM Bis-Tris propane, 2 M NaCl, pH 7. Finally, the protein was purified through a SEC column that had been pre-equilibrated in PBS buffer.

To self-activate OaAEP1, 1 mM EDTA and 0.5 mM Tris (2-carboxyethyl) phosphine hydrochloride were added to the immature protein and the pH of the solution was adjusted to 4 with glacial acetic acid. Fractions containing the protein (as analyzed by SDS-PAGE) were

pooled and then incubated for 3 to 16 hours at room temperature or 37 °C. Protein precipitation at this pH allowed removal of the bulk of the contaminating proteins by centrifugation. Activated proteins were concentrated by ultracentrifugation using a 50 kDa cutoff concentrator (Sartorius) and stored at -80 °C.

Wild type and mutant ubiquitin (containing the additional "NGL" sequence at its C-terminal end) and SNAP tag (New England Biolabs) were cloned into pET47b vector and expressed in Bl21(DE3) E. coli cells. Purification was done by metal affinity using Ni-NTA as described above. This was followed by prescission protease (GE healthcare) digestion to remove the N-terminal hexa-histidine tag and SEC purification. Mutagenesis was performed using Kapa-Hifi polymerase (Kapa Biosystems) and two primers far away from each other with opposite directions.

Cys247Ala mutant was obtained through mutagenesis PCR of using primer AEP-C247A-Mut: CACCACCGAAAGCAGCTGGGCCTACTATTGCCCGGCGC (SEQ ID NO:16).

**Peptides and peptide cyclization assay**

Native and modified amino acids were purchased from Sigma-Aldrich (USA). All peptides used in this manuscript were synthesized in-house using the solid-phase method and HPLC purified. Cyclization assays were performed in 50-$\mu$l reaction mixtures containing 20 mM phosphate buffer, ligases (5 to 700 nM) and peptide substrates (10 to 300 $\mu$M). Each reaction was performed in triplicate at 37 °C and quenched by adding 5 $\mu$l of 1 M HCl solution. The peptides were separated by using a reversed-phase C18 analytical column (150 x 2.1 mm, Vydac) with a linear gradient from 10% to 50% acetonitrile over 15 min on a Nexera UHPLC system (Shimadzu). For kinetic analysis, the concentrations of WT OaAEP1, mutant OaAEP1, and butelase 1 were fixed at 700 nM, 50 nM, and 20 nM, respectively. The cyclization velocities were calculated by converting the HPLC-peak areas of remained linear precursors or the cyclized products into concentrations. The identity of each HPLC peak was analyzed by MALDI-TOF MS (ABI 4800 MALDI TOF/TOF). The velocities were input into GraphPad Prism (GraphPad Software, San Diego) to obtain the Michaelis-Menten curve and the kinetic parameters (kcat and Km). Comparison for the cyclization efficiency of various OaAEP1 mutants were performed in a volume of 20 µl of the reaction mixture and analyzed by MALDI-TOF MS after 10 mins or 1 hour ligation reaction.

For quick accessment of the substrate specificity of quicklase, 100 µM peptide substrate ($Xaa^1Xaa^2$YRRGRLYRRN$Xaa^3Xaa^4$, (SEQ ID NO:17) were used, with 50 nM Quicklase, reacted for 5 mins at room temperature, under neutral pH (pH 7.4) standard reaction buffer. The % of self-ligated (cyclized) peptides could be used as qualitative indications of Kcat. In each screen,

only one of the residue is modified, and the default amino acid at $Xaa^1$ position is Gly, at $Xaa^2$ position is Leu, and at $Xaa^3$ position is Gly, and at the $Xaa^4$ position is Leu, in this screen (Figure 11).

**List of Sequence ID NOs and detailed amino acid sequences described herein**

1. SEQ ID NO:1
Common Name: OaAEP1 Cys247Ala (Quicklase)
Source: *Ovaloparmena affinis*
Sequence:
MVRYLAGAVLLLVVLSVAAAVSGARDGDYLHLPSEVSRFFRPQETNDDHGEDSVGTRWAVLIA
GSKGYANYRHQAGVCHAYQILKRGGLKDENIVVFMYDDIAYNESNPRPGVIINSPHGSDVYAG
VPKDYTGEEVNAKNFLAAILGNKSAITGGSGKVVDSGPNDHIFIYYTDHGAAGVIGMPSKPYLYA
DELNDALKKKHASGTYKSLVFYLEACESGSMFEGILPEDLNIYALTSTNTTESSWAYYCPAQEN
PPPPEYNVCLGDLFSVAWLEDSDVQNSWYETLNQQYHHVDKRISHASHATQYGNLKLGEEGL
FVYMGSNPANDNYTSLDGNALTPSSIVVNQRDADLLHLWEKFRKAPEGSARKEEAQTQIFKAM
SHRVHIDSSIKLIGKLLFGIEKCTEILNAVRPAGQPLVDDWACLRSLVGTFETHCGSLSEYGMRH
TRTIANICNAGISEEQMAEAASQACASIP

2. SEQ ID NO:2
Common Name: critical structural motif responsible for efficient protein ligation at the positions corresponding to residues 247-264 of SEQ ID NO:1;
Source: Based on structural analysis of this application
Sequence:
$(G/A/S/C)^{247}$-X-Y-$C^{250}$-P-X-X-X-X-X-P-P-P-E-Y-X-X-$C^{264}$
*X means any amino acids that do not disrupt the structural fold of the remaining conserved amino acids.
**Small amino acids (Gly, Ala, Ser or Cys) at position 247 of this structural motif allows efficient protein ligation. An Cys disulfide bond between Cys250 and Cys264 stablize this structural motif.

3. SEQ ID NO:3
Common Name: wild type OaAEP1
Source: *Ovaloparmena affinis*
Sequence:
MVRYLAGAVLLLVVLSVAAAVSGARDGDYLHLPSEVSRFFRPQETNDDHGEDSVGTRWAVLIA
GSKGYANYRHQAGVCHAYQILKRGGLKDENIVVFMYDDIAYNESNPRPGVIINSPHGSDVYAG
VPKDYTGEEVNAKNFLAAILGNKSAITGGSGKVVDSGPNDHIFIYYTDHGAAGVIGMPSKPYLYA

DELNDALKKKHASGTYKSLVFYLEACESGSMFEGILPEDLNIYALTSTNTTESSWCYYCPAQEN
PPPPEYNVCLGDLFSVAWLEDSDVQNSWYETLNQQYHHVDKRISHASHATQYGNLKLGEEGL
FVYMGSNPANDNYTSLDGNALTPSSIVVNQRDADLLHLWEKFRKAPEGSARKEEAQTQIFKAM
SHRVHIDSSIKLIGKLLFGIEKCTEILNAVRPAGQPLVDDWACLRSLVGTFETHCGSLSEYGMRH
TRTIANICNAGISEEQMAEAASQACASIP

4. SEQ ID NO:4
Common Name: AEP1 from *Clitoria ternatea,* Butelase1
Source: *Clitoria ternatea*
Sequence:
IRDDFLRLPSQASKFFQADDNVEGTRWAVLVAGSKGYVNYRHQADVCHAYQILKKGGLKDENII
VFMYDDIAYNESNPHPGVIINHPYGSDVYKGVPKDYVGEDINPPNFYAVLLANKSALTGTGSGK
VLDSGPNDHVFIYYTDHGGAGVLGMPSKPYIAASDLNDVLKKKHASGTYKSIVFYVESCESGS
MFDGLLPEDHNIYVMGASDTGESSWVTYCPLQHPSPPPEYDVCVGDLFSVAWLEDCDVHNLQ
TETFQQQYEVVKNKTIVALIEDGTHVVQYGDVGLSKQTLFVYMGTDPANDNNTFTDKNSLGTP
RKAVSQRDADLIHYWEKYRRAPEGSSRKAEAKKQLREVMAHRMHIDNSVKHIGKLLFGIEKGH
KMLNNVRPAGLPVVDDWDCFKTLIRTFETHCGSLSEYGMKHMRSFANLCNAGIRKEQMAEAS
AQACVSIPDNPWSSLHAGFSV

5. SEQ ID NO:5
Common Name: uncharacterized AEP from coffee [CDP08231.1]
Source: *Coffea canephora*
Sequence:
MMRYATAALLLLALSIIAVAEARDNFLKLPSEIADFFHPKERSDAGGDSVGTRWAVLIAGSNGY
WNYRHQADVCHAYQILKRGGLKDENIVVFMYDDIAYNEENPRPGVIINSPHGADVYQGVPKDY
TGDDVNAKNFLAAILGDKTAITGGSGKVVDSGPNDHIFIYYTDHGGPGVLGTPSGPYLYADDLN
EVLKKKHASGTYKSLVFYLEACESGSIFEGLLPEDLNIYATTASNAEESSWGTYCPGEYPSPPP
EYETCLGDLYSVAWMEDSEIHNLHTETLKQQYHLVKKRTSSSNSAYGSHVMQYGDLKLSLEDL
FLYMGTNPANDNYTFVDENSLRPSSKAVNQRDADLLHFWDKFRKAPEGSARKVEAQKQVVEA
MSHRMHIDNSVKLIGKLLFGIEKGSEILNSVRPAGHPLADDWDCLKSLVRTFETHCGSLSQYGM
KHMRSIANICNAGIKKDQMAEAAAQACVSLPSNSWSSLHRGFSA

6. SEQ ID NO:6
Common Name: uncharacterized enzyme [XP_017229093.1]
Source: *Daucus carota subsp. sativus*
Sequence:
MVRYLAGAVALLVVVSISAVVESRRDIVGDVLKLPSEVSSFFRPVAEEEDSVGTRWAVLIAGSN
GYWNYRHQADICHAYQLLRRGGVKEENIVVFMYDDIAYDEENPRPGVIINSPHGSDVYKGVPK

DYTGEDVTVNNVFAAILGDKSATTGGSGKVVDSGPNDHIFIYYSDHGGPGVLGMPTNPYMYAG
DLVDVLKKKHASGTYKSMVFYLEACESGSIFEGLLPEGLNIYATTASNAYESSWGTYCPGEYPS
PPPEYETCLGDLYSVAWMEDSDIHNLRTETLRQQYQQVKKRTSNDNSGWGSHVMQYGDLKL
STEELFMYMGTNPANDNFTFVDDNSLRLSSSKAVNQRDADLLHFWDKYRKAPEGSDRKIAAQ
KQFSEAMSHRMHLDNSIQLIGKLLFGIDTASEVLTTVRPSGQPLVDDWLCLKKLVRTFETYCGS
LSQYGMKHMRSIANICNAGISEEQMSEASAQACVTFPSNPWSSVNKGFTA

7. SEQ ID NO:7
Common Name: Peptidase C13, legumain [OMO66906.1]
Source: *Corchorus capsularis*
Sequence:
MTRLVAGVILLLLSVTGIVSAGRDATGDVLRLPSEASRFFRPSDDDEVGTRWAVLIAGSNGYWN
YRHQADVCHAYQLLKKGGLKDENIIVFMYDDIAYNYENPRQGIIINSPHGDDVYQGVPKDYTGE
EVTVHNFLAAILGNKTAITGGSGKVVDSGPNDHIFIYYTDHGGPGVLGMPTYPYLYADELIDTLK
KKHASGTYKSLVFYLEACESGSIFEGLLPEGLNIYATTASNADESSWGTYCPGEYPSPPPEYET
CLGDLYSVAWMEDSDLHNLRTETLHQQYELVKRRTLNGNSAYGSHVMQYGDVGLAKEHLFLY
LGTNPANDNFTFIDENSLQPPAKAVNQRDADLVHFWDKYRKAPDGSARKVEAQKQVVEAMSH
RMHVDNSIQLIGKLLFGIERGADVLKTVRPAGQPLVDDWKCLKSMVRTFETHCGSLAQYGMKH
MRSIANICNAGIQTEQIAEASAQACVSIPSGQWSSIQKGFSA

8. SEQ ID NO:8
Common Name: Peptidase C13, legumain [OMO86616.1]
Source: *Corchorus olitorius*
Sequence:
MTRLVAGVILLLLSVTGIASAGRDATGDVLRLPSEASRFFRPSDDDEVGTRWAVLIAGSNGYWN
YRHQADVCHAYQLLKKGGLKDENIIVFMYDDIAFNYENPRQGIIINSPHGDDVYQDVPKDYTGE
EVTVHNFLAAILGNKTAIKGGSGKVVDSGPNDHIFIYYTDHGGPGVLGMPTYPYLYADELIDTLK
KKHASGTYKSLVFYLEACESGSIFEGLLPEGLNIYATTASNADESSWGTYCPGEYPSPPPEYET
CLGDLYSVAWMEDSDLHNLRTETLHQQYELVKRRTLNGNSAYGSHVMQYGDVGLAKEHLFLY
LGTNPANDNFTFIDENSLQPPAKAVNQRDADLVHFWDKYRKAPDGSVRKVEAQKQVVEAMSH
RMHVDNSIQLIGKLLFGIERGADVLKTVRPAGQPLVDDWKCLKSMVRTFETHCGSLAQYGMKH
MRSIANICNAGIQTEQMAEASSQACVSIPSGQWSSIQKGFSA

9. SEQ ID NO:9
Common Name: vacuolar-processing enzyme [XP_012077326.1]
Source: *Jatropha curcas*
Sequence:

MTRLATGVILLLLALCVVSSAGSRDIVGDVLRLPSEASRFFRPGGAHVAKEDDSTGTRWAILIAG
SNGYWNYRHQADVCHAYQLLKKGGLKDENIIVFMYDDIAFNKENPRPGVIINNPYGEDVYKGV
PKDYTGEDVNVNNFFAAILGNKTAITGGSGKVVDSGPNDHIFIYYTDHGGPGVLGMPTNPYLYA
NDLIDVLIKKHASGTYKSLVFYLEACESGSIFEGLLPEGLNIYATTAANAEESSWGTYCPGEYPS
PPPEYETCLGDLYSVAWMEDSDVHNLQTETLRQQYQLVKRRTANGNSAYGSHVMQYGDVGL
SKDNLFLYMGTNPANENYTFVDENSLRPPSKAVNQRDADLVHFWDKYRKAPDGSTRKIQAQK
QFVEAMSHRMHLDHSMKLIGKLLFGIGKGSEVLNAIRPAGQPLVDDWVCLKTLVRTFETHCGS
LSQYGMKHMRSLANLCNAGIREDQMAEASAQACVSIPSGPWSSLHKGFSA

10. SEQ ID NO:10

Common Name: uncharacterized enzyme [AGC94758.1]

Source: *Malus hupehensis*

Sequence:

MTRLASAVVLLFLASVLASAAGSRDLIGDVLRLPSEASRFFGRGDDAPDQQDDGTVGTRWAVL
IAGSNGYWNYRHQADICHAYQLLKKGGLKDENIVVFMYDDIAYNEENPRQGVIINSPHGSDVYE
GVPKDYTGEDVTVNNFFAAILGNKTALTGGSGKVVDSGPNDHIFIYYTDHGGPGILGMPTSPYI
YANDLIEVLKKKHAAGTYRSLVFYLEACESGSIFEGLLPEGLNIFATTASNAEESSWGTYCPGEY
PSPPPEYDTCLGDLYSVAWMEDSDVHNLRSETLHQQYELVKTRAANDNSGSGSHVMQYGDV
GLSKNNLFVYMGTNPANDNYTFLGENSLRPSSKAVNQRDADLLHFWHKYRKAPEGSARKIQA
QKDFVEAMSHRMHIDQTMKLIGKLLFGIEKGPQVLNAVRPAGQPLVDDWDCLKTMVRSFETHC
GSLSQYGMKHMRSLANICNAGMTQEQMAEASAQACVSAPSGRWSSLHRGFSA

11. SEQ ID NO:11

Common Name: vacuolar-processing enzyme-like [XP_009361606.1]

Source: *Pyrus x bretschneideri*

Sequence:

MTRFAGAVVLLFLASVLASAAGSRDLIGDVLRLPSEASRFFGRGDDAPDQQDDGTVGTRWAVL
IAGSNGYWNYRHQADICHAYQLLKKGGLKDENIVVFMYDDIAYNEENPRQGVIINNPQGSDVYE
GVPKDYTGEDVTVNNFFAAILGNKTALTGGSGKVVDSGPNDHIFIYYTDHGGPGVLGMPTSPYI
YANDLIEVLKKKHAAGTYKSLVFYLEACESGSIFEGLLPEGLNIFATTASNAEESSWGTYCPGEY
PSPPPEYETCLGDLYSVAWMEDSDIHNLRSETLHQQYELVKTRTANDNYGFGSHVMQYGDVG
LSKNNLFVYMGTNPANDNYTFLGENSLRPSSKAVNQRDADLLHFWHKYRKAPEGSARKVQAQ
KDFVEAMSHRMHIDQTMKLIGKLLFGIEKGPQVLNAVRPAGQPLVDDWDCLKTMVRSFETHCG
SLSQYGMKHMRSLANICNAGMTQEQMAEASAQACVSAPSSRWSSLHRGFSA

12. SEQ ID NO:12

Common Name: wild type OaAEP1 (SEQ ID NO:3) Construct for recombinant expression

Sequence:

MGMAHHHHHHMQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRT
LSDYNIQKESTLHLVLRLRGGARDGDYLHLPSEVSRFFRPQETNDDHGEDSVGTRWAVLIAGS
KGYANYRHQAGVCHAYQILKRGGLKDENIVVFMYDDIAYNESNPRPGVIINSPHGSDVYAGVPK
DYTGEEVNAKNFLAAILGNKSAITGGSGKVVDSGPNDHIFIYYTDHGAAGVIGMPSKPYLYADEL
NDALKKKHASGTYKSLVFYLEACESGSMFEGILPEDLNIYALTSTNTTESSWCYYCPAQENPPP
PEYNVCLGDLFSVAWLEDSDVQNSWYETLNQQYHHVDKRISHASHATQYGNLKLGEEGLFVY
MGSNPANDNYTSLDGNALTPSSIVVNQRDADLLHLWEKFRKAPEGSARKEEAQTQIFKAMSHR
VHIDSSIKLIGKLLFGIEKCTEILNAVRPAGQPLVDDWACLRSLVGTFETHCGSLSEYGMRHTRTI
ANICNAGISEEQMAEAASQACASIP

13. SEQ ID NO:13
Common Name: Hexa-His tag (recombinant)
Sequence:
MGMAHHHHHH

14. SEQ ID NO:14
Common Name: Ubiquitin tag for assisting recombinant protein expression (recombinant)
Sequence:
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKEST
LHLVLRLRGG

15. SEQ ID NO:15
Common Name: pre-activation form of wild type OaAEP1 (construct being crystalized in this application, PDB code: 5H0I)
Sequence:
ARGDYLHLPSEVSRFFRPQETNDDHGEDSVGTRWAVLIAGSKGYANYRHQAGVCHAYQILKR
GGLKDENIVVFMYDDIAYNESNPRPGVIINSPHGSDVYAGVPKDYTGEEVNAKNFLAAILGNKS
AITGGSGKVVDSGPNDHIFIYYTDHGAAGVIGMPSKPYLYADELNDALKKKHASGTYKSLVFYL
EACESGSMFEGILPEDLNIYALTSTNTTESSWCYYCPAQENPPPPEYNVCLGDLFSVAWLEDS
DVQNSWYETLNQQYHHVDKRISHASHATQYGNLKLGEEGLFVYMGSNPANDNYTSLDGNALT
PSSIVVNQRDADLLHLWEKFRKAPEGSARKEEAQTQIFKAMSHRVHIDSSIKLIGKLLFGIEKCTE
ILNAVRPAGQPLVDDWACLRSLVGTFETHCGSLSEYGMRHTRTIANICNAGISEEQMAEAASQ
ACASIP

16. SEQ ID NO:16
Common Name: AEP-C247A-Mut, primer used for mutagenesis, converting Seq ID NO:3 to Seq ID NO:1.
Sequence: CACCACCGAAAGCAGCTGGGCCTACTATTGCCCGGCGC

17. SEQ ID NO:17

Common Name: Peptide template used for characterizing recognition sequence specificity of Quicklase

Sequence: $Xaa^1Xaa^2YRRGRLYRRNXaa^3Xaa^4$

* In each screen, only one of the residue is modified, and the default amino acid at $Xaa^1$ position is Gly, at $Xaa^2$ position is Leu, and at $Xaa^3$ position is Gly, and at the $Xaa^4$ position is Leu, in this screen.

***Example 1: Quicklase OaAEP1 mediates peptide ligation***

AEP-like plant peptide ligases share a conserved protein architecture also found in human legumain endopeptidase (hlegum) (Ishii, S., Methods Enzymol, 1994. 244: p. 604-15; Chen, J.M., et al., J Biol Chem, 1997. 272(12): p. 8090-8). Plant AEPs have an N-terminal signal sequence of ~20-30 amino acids that direct them to the plant vacuolar compartment, followed by an enzymatic core domain. At the C-terminus of the core domain, a ~130-residues "cap" or pro-domain entirely covers the active site, keeping the immature protein in an inactive zymogenic state. The core and pro-domain are connected by a "linker" defined as spanning residues 325-346 of the protein (Figure 2A). Activation of plant AEP enzymatic activities requires proteolytic maturation at an acidic pH of about 3.7, at which the pro-domain is cleaved either *in cis* by the proteolytic activity of the core domain itself or *in trans* by activated Asx proteases present in the milieu, and released from the core domain. The precise cleavage site(s) within the linker and pro-domain, leading to the formation of a mature active plant AEP ligase remain poorly characterized. In the case of butelase 1, Asn383 was hypothesized to constitute the site of cap domain cleavage (Nguyen, G.K., et al., Nat Chem Biol, 2014. 10(9): p. 732-8), while several potential cleavage sites spanning region 328-351 of the protein were proposed for OaAEP1 (Figure 2A and Figure 7) (Harris, K.S., et al., Nat Commun, 2015. 6: p. 10199). Amino-acid sequence alignments of plant AEP-like ligases with AEP peptidases such as hlegum also reveal the presence of several evolutionary conserved residues in the catalytic pocket (Figure 7) (Nguyen, G.K., et al., Nat Chem Biol, 2014. 10(9): p. 732-8; Harris, K.S., et al., Nat Commun, 2015. 6: p. 10199; Dall, E., et al., Angew Chem Int Ed Engl, 2015. 54(10): p. 2917-21; Dall, E. and H. Brandstetter, Biochimie, 2016. 122: p. 126-50; Saska, I., et al., J Biol Chem, 2007. 282(40): p. 29721-8). Therefore, what determines how these closely similar enzymes preferentially function either as ligase or protease remains elusive. Herein reported is a crystal structure of OaAEP1 in its zymogenic form revealing how the amide group of Gln347 a residue at the N-terminus of the pro-domain can be accommodated in the catalytic pocket. Based on the OaAEP1 structure, a series of OaAEP1 mutants was generated to better understand the activation mechanism and catalytic activities of plant AEP ligases. It was found that that the segment 346-351 plays a critical role for the maturation of OaAEP1b, because

mutating Asx residues within this segment severely disrupts AEP ligase activation. A major determinant governing the efficiency of peptide ligation activities of OaAEP1 is the side chain at residue 247, which is located near the catalytic Cys217-His175 dyad: larger side-chains at this position completely abolish ligase enzymatic activities, while smaller side-chains allow ligation reactions to take place at a much faster rate. Accordingly, it was demonstrated that the single Cys247Ala mutant of OaAEP1 has peptide ligase catalytic activity ~160 times faster than the wild type OaAEP1 enzyme, comparable to the butelase 1 enzyme extracted from plant. It was also shown that Cys247Ala can efficiently catalyze the ligation of two well-folded proteins: ubiquitin and SNAP tag protein, using only sub-micro molar concentrations of the recombinant enzyme.

In more detail:

OaAEP1 was cloned and expressed in *E. coli* as an ubiquitin fusion protein as reported earlier (Figure 2) (Harris, K.S., et al., Nat Commun, 2015. 6: p. 10199). Using ion exchange and size exclusion chromatography (SEC), sufficient amounts of pure protein suitable for enzymatic and structural analysis were obtained (Figure 2B). At the neutral pH of purification (pH=7.4), OaAEP1 was mainly present as a heterodimer, where both monomers are in their zymogenic form, but only one monomer had retained in addition the ubiquitin and N-terminal His-tag, while the fusion moiety was cleaved in the other monomer (labeled as "His-Ub-AEP-FL" and "AEP-FL" in Figure 2B and Figure 2C). This heterodimer was subjected to crystallization trials for structural studies (see below). In agreement with reference (Harris, K.S., et al., Nat Commun, 2015. 6: p. 10199), activation of OaAEP1 was optimum for pH values in the range of 3.4 – 4.0 (Figure 2D). Following activation, OaAEP1 converts to a monomer in solution according to SEC analysis ("AEP-holo" in Figure 2C). To test the ability of the protein to be activated *in trans*, the catalytic residue Cys217 was mutated into Ser217. As anticipated, the Cys217Ser mutant only displayed background level activation and ligase activity (Figure 2E). However, activation of the inactive mutant could be rescued *in trans* by addition of only 10% of wild type OaAEP1 to the inactive enzyme (Figure 2E). This suggested that similar *trans*-cleavage processing events could also be at play during OaAEP1 activation and that Cys217 is the active site residue for both *cis* and *trans*-activation.

Crystals of OaAEP1 diffracting to a resolution of 2.56 Å at a 3$^{rd}$ generation synchrotron source were obtained (Table 1 and Figure 3). Although a OaAEP1 heterodimer was used as starting material, the crystal structure revealed a homo-dimer where both N-terminal $(His)_6$ tag and Ubiquitin fusion protein as well as the signal peptide sequence had been released from the two monomers present in the asymmetric unit, suggesting N-terminal proteolysis during crystallization (Figure 3A and D). The OaAEP1 dimer is stabilized by intermolecular interactions

between the two cap domains (Figure 3D). This is in agreement with SEC analysis (Figure 2C) showing that the pre-activated form of the enzyme containing the cap region is dimeric at neutral pH, while the activated form, which is devoid of cap, is monomeric. Accordingly, SDS PAGE of dissolved OaAEP1 crystals (Figure 2B) revealed the presence of both the enzymatic core and the C-terminal pro-domain with a total Mr of approximately 49.7 kDa, supporting the hypothesis that the crystallized enzyme is in an immature form. The OaAEP1 core is formed by a 6-stranded $\beta$-sheet surrounded by six $\alpha$-helices at its periphery (Dall, E., et al., Angew Chem Int Ed Engl, 2015. 54(10): p. 2917-21; Dall, E. and H. Brandstetter, Proc Natl Acad Sci U S A, 2013. 110(27): p. 10940-5; Zhao, L., et al., Cell Res, 2014. 24(3): p. 344-58). An overall view of the structure shown in Figure 3A reveals that the $\alpha$-helical cap domain (spanning residues 347-474) entirely covers the catalytic cleft and suggests that release of the cap is required to grant access to peptide or protein substrates. The OaAEP1 catalytic site contains several residues present in other AEPs including the Cys217-His175 catalytic dyad (Cys189-His148 in hlegum) and His73-Asp174 that superimpose with His45-Asp147 of hlegum bound to cystatin E (PDB code 4N6O) (Figure 4). A clear difference captured here between the zymogenic forms of hlegum (PDB code: 4NOK) and OaAEP1 resides within the linker region: in the case of the prolegumain structure, the linker is short and well ordered, whilst it is long and flexible in the present OaAEP1 structure (Figure 3A): no electron density was present for the 19 residues between 325 to 343, although this segment is present in the crystallized protein (Figure 2B).

Table 1: Data collection and refinement statistics

| | **oAEP1b** |
|---|---|
| Resolution range (Å) | 71.72 - 2.56 (2.70 - 2.56)* |
| Space group | C 2 |
| Unit cell dimensions a, b, c (Å)<br>Angles α, β, γ (°) | 145.73 70.21 118.28<br>90 117.14 90 |
| Measured reflections | 104,721 (14,189) |
| Unique reflections | 34,420 (5,029) |
| Multiplicity | 3.0 (2.8) |
| Completeness (%) | 99.7 (99.6) |
| Mean I/sigma(I) | 7.0 (1.7) |
| $R_{merge}$[a] | 0.095 (0.700) |
| $R_{pim}$[b] | 0.065 (0.501) |
| $R_{work}$[c] | 0.186 (0.212) |
| $R_{free}$[d] | 0.221 (0.212) |
| Number of non-hydrogen atoms | 6,498 |

| | |
|---|---|
| macromolecules | 6,200 |
| water | 298 |
| Protein residues | 851 |
| RMS (bonds, Å) | 0.010 |
| RMS (angles, °) | 1.19 |
| Ramachandran plot (%) | |
| favoured | 93.8 |
| allowed | 5.1 |
| outliers | 1.1 |
| Clash score | 6.37 |
| Average B-factor | |
| Monomer A,B | 71.8, 66.1 |
| solvent | 62.0 |

*The numbers in parentheses refer to the last (highest) resolution shell.

[a] $R_{merge}$ = $\sum|I_j - < I > |/\sum I_j$, where $I_j$ is the intensity of an individual reflection, and $< I >$ is the average intensity of that reflection.

[b] $R_{pim} = \frac{\sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_{j=1}^{n} |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_{j} I_{hkl,j}}$

[c] $R_{work}$ = $\sum||F_o| - |F_c||/\sum|F_c|$, where $F_o$ denotes the observed structure factor amplitude, and $F_c$ the structure factor amplitude calculated from the model.
[d] $R_{free}$ is as for $R_{work}$ but calculated with 5% (3044) of randomly chosen reflections omitted from the refinement.

While residues 325-343 of the linker region are not visible in the electron density map due to their flexibility, residues Val344-Val345-Asn346-Gln347 are trapped at the interface between the cap and the core domain (Figure 3C and 4A). Remarkably, their orientation closely overlaps with the Ac-Tyr-Val-Ala-Asp-chloromethylketone inhibitor that was covalently linked to the catalytic Cys189 of hlegum (Figure 3E, PDB code: 4AWA) (Dall, E. and H. Brandstetter, Proc Natl Acad Sci U S A, 2013. 110(27): p. 10940-5.). Here however, residue Gln347 penetrates deeply into the S1 pocket such that its amide group occupies the position of the carboxylic group of the Asp residue in the Ac-Tyr-Val-Ala-Asp-CMK protease inhibitor that was covalently bound to hlegum and forming several interactions with the surrounding active site residues

(Figure 4A and 4B). Thus the present crystal structure can serve as a useful model to picture a pre-ligation conformation, with linker residues 344-347 mimicking a N-terminal substrate acceptor bound to the active site.

Although in other AEPs, Asx is the most favored residue for processing at the P1 position, the carbonyl group of Gln347 lies at a distance of 5.2 Å from the attacking Cys217 sulfur center, suggesting that the segment 346-351 of the pro-protein might contain auto-cleavage sites used during maturation. Since the crystallization conditions are compatible with OaAEP1 activation, the present crystal structure is likely to have captured one of the conformational states occurring along the activation pathway. To determine whether Gln347 belongs to the segment recognized and cleaved during self-activation, mutagenesis experiments targeting region 346-351 of the OaAEP1 protein were performed (Figure 4C). The role of Gln347 in OaAEP1 auto-activation was particularly assessed by mutating this residue to Ala, and whether enzyme activation was abolished was examined. Only little impact was found on the mutant enzyme activation (Figure 4C). Residues Asn346, Asp349 and Asp351 that are next to Gln347 were then targeted. While the two OaAEP1 double mutants Asn346Ala-Gln347Ala and Asp349Ala -Asp351Ala and the Asn346Ala-Asp349Ala-Asp351Ala triple mutant can still be activated, the Asn346Ala-Gln347Ala-Asp349Ala-Asp351Ala had the most significantly impaired auto-activation activity (lane "A4" in Figure 4C). Importantly, activation of the four-Ala mutant could not be rescued *in trans* when wild type enzyme was added to the reaction mixture. Thus, *trans*-activation requires the presence of at least one of the four Asx/Gln residues (with a slight preference for Asp349 or Asp351) found in this segment of the linker. Taken together, these data suggest that the present crystal structure represents an intermediate conformational state during which the linker region is trapped at the catalytic site, mimicking a substrate-bound pre-ligation complex.

The peptide cyclization activity of mature OaAEP1 was then analyzed using MS and the peptide substrate "GLPVSTKPVATRNGL" (Figure 5A). The turn-over rate ($k_{cat}$) for OaAEP1 is 0.052 $s^{-1}$, a value in good agreement with the previous report (Harris, K.S., et al., Nat Commun, 2015. 6: p. 10199). Likewise, the cyclization activity of butelase 1 extracted from plant was measured using the 15 residues peptide "GLPVSTKPVATRNHV" (Figure 5A) leading to a $k_{cat}$ about 90 times faster than OaAEP1b.

The pre-ligation conformation of OaAEP1 captured in the present structure was then analyzed to identify several key features that could differentiate an AEP protease from an AEP-like ligase. A side-by-side comparison with a previously reported proteolytic AEP-inhibitor complex (PDB code: 4AWA) revealed interesting structural features unique to AEP-like ligases. In contrast to AEP proteases, the surface of OaAEP1 facing the catalytic Cys217 appears to be more widely open. Three cysteine residues (Cys247, Cys250, Cys264) are aligned along a shallow cleft at

the surface of the enzyme (Figure 3C and 3E). Unlike the surface of AEP proteases that contains several bulges, both butelase 1 (using a homology model based on OaAEP1b) and OaAEP1 have residues with smaller side-chains along this patch. A disulfide bond is formed between Cys250 and Cys264, which are conserved in AEP-like ligases butelase 1 and OaAEP1b. This covalent bond is likely to stabilize the local structure and provide an adequate surface allowing the approach of an N-terminal amine incoming group. Conversely, the corresponding surface patch of AEP proteases (PDB: 4AWA) is lined with residues having bulky side-chains like Tyr (Figure 3E) blocking this channel. It was therefore hypothesized that residue Cys247 which is located at the extremity of the substrate channel is likely to function as a final nucleophile filter (or "gate-keeper"), allowing attack from the N-terminal amine to complete the ligation reaction. To test the impact of various side chains on ligase behavior at the "gate keeper" position, a series of point mutagenesis was performed whereby Cys247 was mutated into Gly, Ala, Val, Ser, Thr, Met, Leu and Ile (Figure 5). The activity of the corresponding mutant proteins revealed that the size of the side chain indeed plays a critical role on ligase activity. Remarkably, the Cys247Ala mutant displayed significantly improved enzymatic properties compared to WT OaAEP1 (Figure 5D). This phenotype was attribute to the presence of a smaller side chain that is more appropriate to accommodate an incoming amine group, while the Cys247Gly mutation destabilizes the local conformation and negatively affects the ligase activity. Further investigations showed that Cys247Gly catalyzes peptide hydrolysis as alternative reaction, if the available N-terminal amine group was acetylated (Figure 10). While in the case of Cys247Ala mutant, side chain ligation is more preferred if no available N-terminal amine is present, Ala is less hydrophilic which disfavors water entering the catalytic patch causing hydrolysis of the intermediate thio-ester bond. When the N-terminal amine is not protected, backbone ligation is much more favored than any of the alternative reactions.

Remarkably Cys247Ala demonstrated improved enzymatic properties over the parent wild-type OaAEP1 protein: its $k_{cat}$ approximately 160 times higher than WT OaAEP1b, makes it an attractive tool for a variety of challenging protein engineering and labeling applications. The ability of Cys247Ala to ligate ubiquitin with a peptide was first tested. The three residues"Asn-Gly-Leu" were added to the C-terminus of ubiquitin to make it recognizable by Cys247Ala. Another peptide having N-terminal residues GlyLeu was used as the add-on component. At a protein: peptide molar ratio of 1:5, Cys247Ala catalyzed the ligation of more than 90% ubiquitin with the peptide within 10 min (Figure 6). An amount of 30 nM of Cys247Ala was enough for the reaction compared to 10-30 μM of sortase A typically required to perform productive ligations. To further demonstrate the potential of Cys247Ala, ligations between two well-folded proteins were also tested. Ubiquitin having the same "Asn-Gly-Leu" modification at its C-terminal end and a N-terminal modified SNAP (NEB) tag protein were used as substrates. Using 5 μM of SNAP tag protein and 8 μM ubiquitin, 100 nM Cys247Ala was able to ligate more than 60% of

the individual substrate proteins into a ligated single polypeptide product, indicating that the catalyzed protein ligation is highly efficient and irreversible (Figure 6). Protein ligation can be performed within minutes at neutral pH and at room temperature. Remarkably, the efficiency of Cys247Ala to catalyze the ligation of two folded proteins is comparable to its ability to ligate one protein with a peptide, suggesting that its open binding site can accommodate two bulky protein substrates without hindering the ligation process. Control ligation experiments showed that protein ligation activity could only be observed in the presence of the Cys247Ala protein, but not with the OaAEP1 wild type construct (Figure 6). This further demonstrated the important role played by the gate-keeper residue in regulating access to protein substrate.

Both sortase A and butelase 1 have been used to perform peptide ligations. However, both enzymes suffered from several disadvantages. Sortase A requires $Ca^{2+}$, is slow and recognizes a longer -more flexible- C-terminal sequence (LPXTG). Although through extensive mutagenesis studies, several mutants were identified improving its kinetics, the catalysis was still quite inefficient in ligating two well-folded proteins (Chen, I., B.M. Dorr, and D.R. Liu, Proc Natl Acad Sci U S A, 2011. 108(28): p. 11399-404). In contrast, butelase 1 extracted from plant, is extremely efficient and has a shorter recognition amino acids sequence (NHV) (Nguyen, G.K., et al., Nat Chem Biol, 2014. 10(9): p. 732-8). However, optimization and engineering of the recombinant construct of Butelase 1 still remains challenging because of the difficulty to produce an active recombinant enzyme. Here, to the knowledge of the inventors the first X-ray crystallographic structure of OaAEP1 was reported that can serve as a template to understand the family of plant AEP ligases and will facilitate the design of faster protein ligases with alternative substrate specificities. Our structure reveals the mode through which self-cleavage activation is achieved in AEP-like ligases and also points to key residues and structural features accounting for the functional divergence among AEP-like proteins: AEPs that act primarily as proteases lack the flat surface near the catalytic pocket conducive to ligation, while AEP-like proteins that are efficient ligases have a wide and open surface able to accommodate the incoming amine group. These structural observations led us to engineer a modified ligase Cys247Ala with improved biochemical properties, which can efficiently ligate proteins. Given its fast ligation catalytic kinetics, it is believed that the present Cys247Ala recombinant protein ligase can be employed to perform attractive biotechnological applications both for protein/peptide synthesis and specific protein labeling (Guimaraes, C.P., et al., Nat Protoc, 2013. 8(9): p. 1787-99; Swee, L.K., et al., Proc Natl Acad Sci U S A, 2013. 110(4): p. 1428-33; Theile, C.S., et al., Nat Protoc, 2013. 8(9): p. 1800-7; Witte, M.D., et al., Nat Protoc, 2013. 8(9): p. 1808-19; Wagner, K., et al., Proc Natl Acad Sci U S A, 2014. 111(47): p. 16820-5; Nguyen, G.K., et al., Angew Chem Int Ed Engl, 2015. 54(52): p. 15694-8; Nguyen, G.K., et al., J Am Chem Soc, 2015. 137(49): p. 15398-401).

***Example 2: Quicklase OaAEP1 homologues Identified by NCBI-BLAST***

A Blast search of the amino acid sequence of SEQ ID NO:1 was carried out by submitting the sequence to the NCBI blast server using standard blast criteria. Top hits as displayed in Figure 12 were selected for sequence analysis. All of them were found to contain the conserved structural motif of residues 247-264 of OaAEP1b.

***Example 3: AEP_C13CO4 (SEQ ID NO:8) is a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1).***

Based on the prediction of the sequence alignment, AEP_C13CO4 contained the conserved structural motif identical to Quicklase (247-264). This indicated that AEP_C13CO4 might be functional protein ligase. The cDNA of AEP_C13CO4 was then subsequently synthesized and subcloned into pet28b in a similar way as Quicklase. This new potential ligase was then expressed, purified and activated in the same way as Quicklase, and exhibited both proteolytic and protein ligation activities. Although the wild type of AEP_C13CO4 is not an ideal protein ligase, with much slower enzymatic catalytic efficiency as Quicklase, and having significant byproducts of proteolytic reactions (Figure 13), it appears to a good starting point for further engineering of an alternative protein ligase.

***Example 4: QuicklaseOaAEP1 -Catalyzed Peptide Dimerization, Oligomerization, and Multimerization***

In relation to the schemes illustrated in Figure 14, the following exemplary experiments were carried out:

Figure 14 (a):

The branched 2-mer linker peptide was synthesized based on standard Fmoc-Solid Phase Peptide Synthesis (SPPS) chemistry on Rink Amide MBHA resin (GL Biochem, Shanghai, China). The branch was introduced by incorporation of Fmoc-L-Lys(Fmoc)-OH. After removing of Fmoc groups at both $\alpha$- and $\epsilon$-amino groups, the remaining amino acids (GLGG) were introduced according to the standard Fmoc-SPPS procedure.

For one-step homodimerization reaction, 50 μM of C-terminal NGL-tagged Protein of Interest (POI-NGL), Ub-AA-NGL in this case, was reacted with various concentrations of 2-mer linker ranging from 0-200 μM catalyzed by 0.5 μM of AEP at room temperature. The reaction was monitored by SDS-PAGE after 1 h.

Figure 14(b):

For two-step heterodimerization strategy, the first step was the enzyme (0.5 μM) catalyzed reaction between POIA-NGL (50 μM) with excessive amount of 2-mer linker (200 μM) to enable

only one branch of the dimer linker was labelled. After reaction, the POIA-linker was isolated by removing 2-mer linker peptide through buffer exchange with 3500 MW cut off concentrator.

For the second ligation step, POIB-NGL (100, 200 μM ) was reacted with POIA-linker (50 μM) catalyzed by 0.5 μM of AEP to get the final heterodimer product.

Figure 14(c):
To generate the 4-mer linker, the 2-mer linker was dimerized through crosslinking of the C-terminal cysteine residue. See figure g for the reaction detail.

One-step protein tetramerization was achieved by AEP catalyzed reaction between POI-NGL and the 4-mer linker. With ubiquitin as an example, to optimize the ligation reaction, 20 μM of Ub-AA-NGL (POI-NGL) was ligated with various concentrations (0-80 μM) of 4-mer linker in the presence of 0.5 μM of AEP at pH 7.4, room temperature. After 1 h, the reaction was analyzed by SDS-PAGE, which indicated that maximal amount of Ub-tetramer was yielded when 2.5-5 μM of 4-mer linker was used.

Figure 14(d):
To achieve protein ultra-oligomerization, the oligomer peptide platform was first synthesized via two steps: i) a peptide containing multiple cysteine residues which were separated by flexible spacer of 7-8 amino acids was synthesized directly or through native chemical ligation; ii) the cysteine residues in the peptide was then alkylated by a bromoacetyl peptide with a free N-terminal GL residues for AEP catalyzed ligation later. After producing the oligomer linker, AEP catalyzed reaction between POI-NGL and the oligomer linker was performed to generated the protein ultra-oligomer.

Figures 14(e, f):
Schemes for the generation of 2-mer-NGL linker used for protein head-to-head dimerization:
The synthesis of 2-mer-NGL liner was achieved via S-alkylation of the N-terminal acetylated cysteine residue of the monomer-NGL peptide by either 1,3-dichloroacetone or m-Xylylene dibromide (Sigma-aldrich). Briefly, monomer-NGL peptide was dissolved in a buffer containing 20 mM Tris·HCl, 150 mM NaCl, 5 mM TCEP, pH 8.0 at a final concentration of 1 mg/mL. 1 eq. of 1,3-dichloroacetone or m-Xylylene dibromide was added into the solution. After 3 h reaction under dark condition, the desired 2-mer-NGL linker was isolated by HPLC purification with C18-reverse phase column.

Figure 14(g):
Schemes for the generation of GL-tetramer linker used for protein tail-to-tail tetramerization:

The synthesis of GL-tetramer linker was achieved via S-alkylation of the C-terminal cysteine residue of the GL-dimer linker peptide by either 1,3-dichloroacetone or m-Xylylene dibromide. Briefly, GL-dimer peptide was dissolved in a buffer containing 20 mM Tris·HCl, 150 mM NaCl, 5 mM TCEP, pH 8.0 at a final concentration of 1 mg/mL. 1 eq. of 1,3-dichloroacetone or m-Xylylene dibromide was added into the solution. After 3 h reaction under dark condition, the desired GL-tetramer linker was isolated by HPLC purification with C18-reverse phase column.

Figure 14(h):

Scheme for the generation of NGL-oligomer linker used for protein head-to-head oligomerization: NGL-oligomer linker platform was synthesized via the S-alkylation reaction between a peptide containing multiple cysteine residues which were separated by flexible spacer of 7-8 amino acids and another N-terminal bromoacetyl peptide with C-terminal NGL residues for AEP catalyzed ligation later. For the alkylation reaction, briefly, 1 mg/mL of cysteinyl peptide was reacted with 1.5 mg/mL of bromoacetyl peptide in buffer containing 20 mM Tris·HCl, 150 mM NaCl, 5 mM TCEP, pH 8.0, at room temperature avoid of light. After 2 hours, the desired NGL-oligomer linker was isolated by HPLC purification with C18-reverse phase column.

Figure 14(i):

Scheme for the generation of GL-oligomer linker used for protein tail-to-tail oligomerization: GL-oligomer linker platform was synthesized via the S-alkylation reaction between a peptide containing multiple cysteine residues which were separated by flexible spacer of 7-8 amino acids and another bromoacetyl peptide (bromoacetyl attached to lysine side chain) with N-terminal GL residues for AEP catalyzed ligation later. For the alkylation reaction, briefly, 1 mg/mL of cysteinyl peptide was reacted with 1.5 mg/mL of bromoacetyl peptide in buffer containing 20 mM Tris·HCl, 150 mM NaCl, 5 mM TCEP, pH 8.0, at room temperature avoid of light. After 2 hours, the desired GL-oligomer linker was isolated by HPLC purification with C18-reverse phase column.

***Example 5: QuicklaseOaAEP1 -Catalyzed Cell Surface Labeling***

In relation to Figure 15, the following exemplary experiments were carried out:

0.1 to 1 μM of OaAEP1 Cys247Ala was added to cellular culture in standard cell culture conditions. 1-20 μM labeling protein tags were added to the reaction afterwards, and incubated for 30-45 mins at 37 °C and standard culturing condition, before both excess enzymes and labelling tags were washed away. Triple washed cellular fractions were harvested to perform the fluorescent SDS-PAGE analysis to determine the labelling efficiency.

In the case of imaging analysis, the images were acquired using Nikon Ti microscope or equivalant (Nikon) with the laser source X-Cite 120 (Excelitas), or equivalent, and digital camera C11440 (Hamamatsu), or equivalant. The programme with which cells were viewed and photographed was MetaMorph (Molecular Devices) or equivalent. Images were compiled using Image J, or equivalent.

The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein. Other embodiments are within the following claims. In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. Further, it will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention. The compositions, methods, procedures, treatments, molecules and specific compounds described herein are presently representative of preferred embodiments are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention are defined by the scope of the claims. The listing or discussion of a previously published document in this specification should not necessarily be taken as an acknowledgement that the document is part of the state of the art or is common general knowledge.

The invention illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms “comprising”, “including,” containing”, etc. shall be read expansively and without limitation. The word "comprise" or variations such as "comprises" or "comprising" will accordingly be understood to imply the inclusion of a stated integer or groups of integers but not the exclusion of any other integer or group of integers. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be

understood that although the present invention has been specifically disclosed by exemplary embodiments and optional features, modification and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

The content of all documents and patent documents cited herein is incorporated by reference in their entirety.

## CLAIMS

What is claimed is:

1. Method of forming a peptide of Formula (I)

$P^1$-Asx-$Xaa^1$-$Xaa^2$-$P^2$ (I),

by ligating a first peptide of Formula (II)

$P^1$-Asx-$Xaa^3$-Leu-$COOH/CONH_2$ (II)

to a second peptide of Formula (III)

$H_2N$-$Xaa^1$-$Xaa^2$-$P^2$ (III),

wherein $P^1$ and $P^2$ are each independently any peptide, modified or unmodified, and optionally can combine such that the peptides of formula (II) and (III) are the termini of the same peptide; Asx is Asp or Asn; $Xaa^1$ is any naturally occurring amino acid; $Xaa^2$ is any naturally occurring amino acid with the exception of Pro, preferably Leu or Ile; and $Xaa^3$ is any naturally occurring amino acid, preferably selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu,
by enzymatically cleaving the bond between "Asx" and "$Xaa^3$" in the first peptide of Formula (II) and ligating the fragment $P^1$-Asx of the first peptide via its C-terminus to the N-terminus of the second peptide of Formula (III) to form a ligated peptide of Formula (I), wherein the enzymatic cleavage and ligation reaction is catalyzed by a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) under conditions suitable for said cleavage and ligation reaction.

2. The method of claim 1, wherein:
   (a) Asx is Asn or Asp; and/or
   (b) $Xaa^1$ is any naturally occurring amino acids; and/or
   (c) $Xaa^2$ is Leu or Ile; and/or
   (d) $Xaa^3$ is selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu.

3. The method of claim 2, wherein two or more of features (a)-(d) are met, i.e. (a) and (b), (a) and (c), (a) and (d), (b) and (c), (b) and (d), or three or more features are met, i.e.

(a), (b) and (c), (a), (b) and (d), (a), (c) and (d), (b), (c) and (d) or, most preferably all four are met.

4. The method of any one of claims 1-3, wherein Asx is Asn, $Xaa^1$ is any naturally occurring amino acid, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ser, Cys, Gly, or Ala.

5. The method of any one of claims 1-4, wherein Asx is Asn, $Xaa^1$ is Arg or Gly, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ala, or Gly.

6. The method of any one of claims 1-5, wherein the first and second peptides are termini of the same peptide (i.e. $P^1$ and $P^2$ combine to form a single core peptide sequence) such that the method cyclizes said peptide.

7. The method of any one of claims 1-6, wherein the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) comprises or consists of:
(a) the amino acid sequence set forth in SEQ ID NO:1;
(b) an amino acid sequence that shares at least 65%, preferably at least 75%, even more preferably at least 85%, most preferably at least 95% sequence identity with, or at least 80%, preferably at least 90%, more preferably at least 95% sequence homology with the amino acid sequence as set forth in SEQ ID NO:1, provided that said peptide ligase comprises the amino acid sequence set forth in SEQ ID NO:2 at the positions corresponding to residues 247-264 of SEQ ID NO:1;
(c) a functional fragment of (a) or (b); or
(d) a polypeptide sequence containing functional fragments of (a) or (b),
with the proviso that said peptide ligase is not the wild-type OaAEP1 having the amino acid sequence set forth in SEQ ID NO:3 or butelase 1 having the amino acid sequence set forth in SEQ ID NO:4.

8. The method of any one of claims 1-7, wherein the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) comprises or consists of:
(a) the amino acid sequence set forth in SEQ ID NO:1;
(b) the amino acid sequence set forth in SEQ ID NO:5;
(c) the amino acid sequence set forth in SEQ ID NO:6;
(d) the amino acid sequence set forth in SEQ ID NO:7;
(e) the amino acid sequence set forth in SEQ ID NO:8;
(f) the amino acid sequence set forth in SEQ ID NO:9;
(g) the amino acid sequence set forth in SEQ ID NO:10; or
(h) the amino acid sequence set forth in SEQ ID NO:11.

9. The method of any one of claims 1-8, wherein the first and/or second peptide further comprises a labeling moiety.

10. The method of claim 9, wherein the labeling moiety is an affinity tag, therapeutic agent, detectable label, or scaffold molecule.

11. The method of any one of claims 1-10, wherein the first and/or second peptide is coupled to a solid support material.

12. The method of any one of claims 1-11, wherein the first and/or second peptides is a cellular surface protein such that the method results in the modification or tagging of the cellular surface protein and in result the cellular surface as such.

13. Method of preparing a dimer, oligomer, or multimer of one or more peptides of interest, the method comprising the steps of:

(a) providing one or more peptides of interest having C-terminal Asx-$Xaa^3$-Leu-$COOH/CONH_2$ residues and a scaffold molecule having two or more copies of N-terminal $H_2N$-$Xaa^1$-$Xaa^2$ residues or, alternatively, providing one or more peptides of interest having N-terminal $H_2N$-$Xaa^1$-$Xaa^2$ residues and a scaffold molecule having two or more copies of C-terminal Asx-$Xaa^3$-Leu-$COOH/CONH_2$ residues, wherein Asx is Asp or Asn; $Xaa^1$ is any naturally occurring amino acid; $Xaa^2$ is any naturally occurring amino acid with the exception of Pro, preferably Leu or Ile; and $Xaa^3$ is any naturally occurring amino acid, preferably selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu;

(b) providing a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1);

(c) preparing a mixture of the one or more peptides of interest, the scaffold molecule, and the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1);

(d) subjecting the mixture to conditions that allow the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) to catalyze the ligation of the one ore more peptides of interest to the scaffold molecule.

14. The method of claim 13, wherein:

(a) Asx is Asn or Asp; and/or

(b) $Xaa^1$ is any naturally occurring amino acids; and/or

(c) $Xaa^2$ is Leu or Ile; and/or

(d) $Xaa^3$ is selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu.

15. The method of claim 14, wherein two or more of features (a)-(d) are met, i.e. (a) and (b), (a) and (c), (a) and (d), (b) and (c), (b) and (d), or three or more features are met, i.e. (a), (b) and (c), (a), (b) and (d), (a), (c) and (d), (b), (c) and (d) or, most preferably all four are met.

16. The method of any one of claims 13-15, wherein Asx is Asn, $Xaa^1$ is any naturally occurring amino acid, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ser, Cys, Gly, or Ala.

17. The method of any one of claims 13-16, wherein Asx is Asn, $Xaa^1$ is Arg or Gly, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ala, or Gly.

18. The method of any one of claims 13-17, wherein the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) comprises or consists of:
(a) the amino acid sequence set forth in SEQ ID NO:1;
(b) an amino acid sequence that shares at least 65%, preferably at least 75%, even more preferably at least 85%, most preferably at least 95% sequence identity with, or at least 80%, preferably at least 90%, more preferably at least 95% sequence homology with the amino acid sequence as set forth in SEQ ID NO:1, provided that said peptide ligase comprises the amino acid sequence set forth in SEQ ID NO:2 at the positions corresponding to residues 247-264 of SEQ ID NO:1;
(c) a functional fragment of (a) or (b); or
(d) a polypeptide sequence containing functional fragments of (a) or (b),
with the proviso that said peptide ligase is not the wild-type OaAEP1 having the amino acid sequence set forth in SEQ ID NO:3 or butelase 1 having the amino acid sequence set forth in SEQ ID NO:4.

19. The method of any one of claims 13-18, wherein the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) comprises or consists of:
(a) the amino acid sequence set forth in SEQ ID NO:1;
(b) the amino acid sequence set forth in SEQ ID NO:5;
(c) the amino acid sequence set forth in SEQ ID NO:6;
(d) the amino acid sequence set forth in SEQ ID NO:7;
(e) the amino acid sequence set forth in SEQ ID NO:8;
(f) the amino acid sequence set forth in SEQ ID NO:9;
(g) the amino acid sequence set forth in SEQ ID NO:10; or

(h) the amino acid sequence set forth in SEQ ID NO:11.

20. Method of modifying or tagging the surface of a target cell by one or more peptides of interest, the method comprising the steps of:

(a) providing the one or more peptides of interest having C-terminal Asx-$Xaa^3$-Leu-$COOH/CONH_2$ residues and/or having N-terminal $H_2N$-$Xaa^1$-$Xaa^2$ residues, wherein Asx is Asp or Asn; $Xaa^1$ is any naturally occurring amino acid; $Xaa^2$ is any naturally occurring amino acid with the exception of Pro, preferably Leu or Ile; and $Xaa^3$ is any naturally occurring amino acid, preferably selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu;

(b) providing a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1);

(c) contacting the target cell with the one or more peptides of interest and the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1);

(d) subjecting the target cell to conditions that allow the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) to catalyze the ligation of the one or more peptides of interest to a cellular surface protein of the target cell.

21. The method of claim 20, wherein:

(a) Asx is Asn or Asp; and/or

(b) $Xaa^1$ is any naturally occurring amino acids; and/or

(c) $Xaa^2$ is Leu or Ile; and/or

(d) $Xaa^3$ is selected from the group consisting of His, Ala, Ser, Cys, Asn, Gly, Arg, Met, Lys, Gln, Leu, and Glu.

22. The method of claim 21, wherein two or more of features (a)-(d) are met, i.e. (a) and (b), (a) and (c), (a) and (d), (b) and (c), (b) and (d), or three or more features are met, i.e. (a), (b) and (c), (a), (b) and (d), (a), (c) and (d), (b), (c) and (d) or, most preferably all four are met.

23. The method of any one of claims 20-22, wherein Asx is Asn, $Xaa^1$ is any naturally occurring amino acid, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ser, Cys, Gly or Ala.

24. The method of any one of claims 20-23, wherein Asx is Asn, $Xaa^1$ is Arg or Gly, $Xaa^2$ is Leu, and $Xaa^3$ is His, Ala, or Gly.

25. The method of any one of claims 20-24, wherein the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) comprises or consists of:

(a) the amino acid sequence set forth in SEQ ID NO:1;

(b) an amino acid sequence that shares at least 65%, preferably at least 75%, even more preferably at least 85%, most preferably at least 95% sequence identity with, or at least 80%, preferably at least 90%, more preferably at least 95% sequence homology with the amino acid sequence as set forth in SEQ ID NO:1, provided that said peptide ligase comprises the amino acid sequence set forth in SEQ ID NO:2 at the positions corresponding to residues 247-264 of SEQ ID NO:1;

(c) a functional fragment of (a) or (b); or

(d) a polypeptide sequence containing functional fragments of (a) or (b),

with the proviso that said peptide ligase is not the wild-type OaAEP1 having the amino acid sequence set forth in SEQ ID NO:3 or butelase 1 having the amino acid sequence set forth in SEQ ID NO:4.

26. The method of any one of claims 20-25, wherein the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) comprises or consists of:
    (a) the amino acid sequence set forth in SEQ ID NO:1;
    (b) the amino acid sequence set forth in SEQ ID NO:5;
    (c) the amino acid sequence set forth in SEQ ID NO:6;
    (d) the amino acid sequence set forth in SEQ ID NO:7;
    (e) the amino acid sequence set forth in SEQ ID NO:8;
    (f) the amino acid sequence set forth in SEQ ID NO:9;
    (g) the amino acid sequence set forth in SEQ ID NO:10; or
    (h) the amino acid sequence set forth in SEQ ID NO:11.

27. The method of any one of claims 20-26, wherein the method further comprises removing the unligated one or more peptides of interest from the taget cell after step (d).

28. The method of any one of claims 20-27, wherein the one or more peptides of interest comprise a labeling moiety.

29. The method of any one of claim 28, wherein the labeling moiety is an affinity tag, therapeutic agent, detectable label, or scaffold molecule.

30. The method of any one of claims 20-29, wherein the one or more peptides of interest are coupled to a solid support material.

31. The method of any one of claims 20-30, wherein the target cell recombinantly expresses a surface polypeptide having N-terminal $H_2N$-$Xaa^1$-$Xaa^2$ residues or C-terminal Asx-$Xaa^3$-Leu-COOH residues for ease of tagging by the one or more peptides of interest.

32. Ligated peptide and/or tagged target cell obtainable according to any of the methods of claims 1-31.

33. Peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) comprising or consisting of:
(a) the amino acid sequence set forth in SEQ ID NO:1;
(b) an amino acid sequence that shares at least 65%, preferably at least 75%, even more preferably at least 85%, most preferably at least 95% sequence identity with, or at least 80%, preferably at least 90%, more preferably at least 95% sequence homology with the amino acid sequence as set forth in SEQ ID NO:1, provided that said peptide ligase comprises the amino acid sequence set forth in SEQ ID NO:2 at the positions corresponding to residues 247-264 of SEQ ID NO:1;
(c) a functional fragment of (a) or (b); or
(d) a polypeptide sequence containing functional fragments of (a) or (b),
with the proviso that said peptide ligase is not the wild-type OaAEP1 having the amino acid sequence set forth in SEQ ID NO:3 or butelase 1 having the amino acid sequence set forth in SEQ ID NO:4.

34. The method of claim 33, wherein the peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) comprises or consists of:
(a) the amino acid sequence set forth in SEQ ID NO:1;
(b) the amino acid sequence set forth in SEQ ID NO:5;
(c) the amino acid sequence set forth in SEQ ID NO:6;
(d) the amino acid sequence set forth in SEQ ID NO:7;
(e) the amino acid sequence set forth in SEQ ID NO:8;
(f) the amino acid sequence set forth in SEQ ID NO:9;
(g) the amino acid sequence set forth in SEQ ID NO:10; or
(h) the amino acid sequence set forth in SEQ ID NO:11.

35. Kit for use in any of the methods of claims 1-31, wherein the kit comprises a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) according to claim 33 or 34, in particular OaAEP1 Cys247Ala (SEQ ID NO:1).

36. Use of a peptide ligase having the activity of OaAEP1 Cys247Ala (SEQ ID NO:1) according to claim 33 or 34 for the ligation of two peptides.

Figure 1

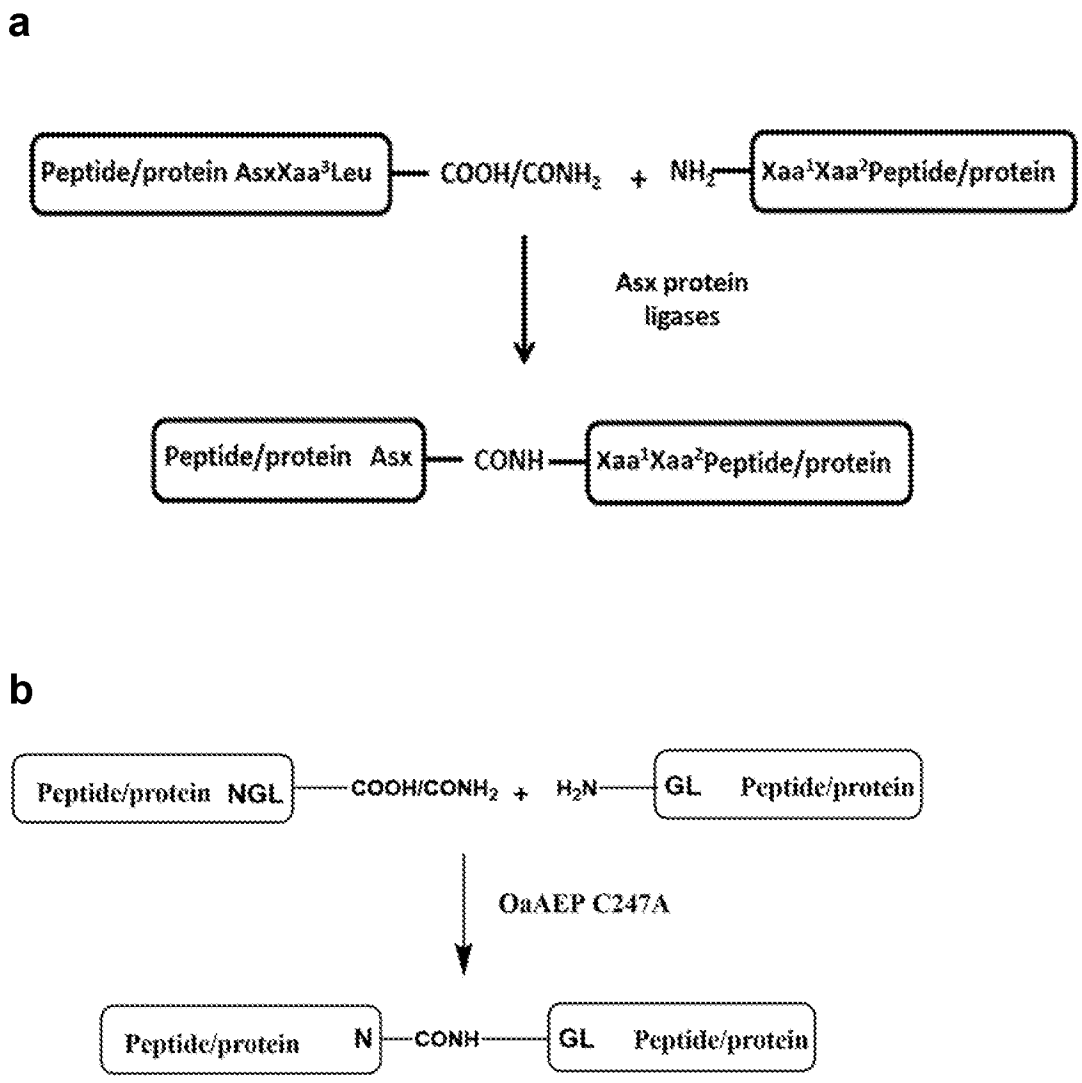

a
Peptide/protein AsxXaa³Leu
COOH/CONH₂
+
NH₂
Xaa¹Xaa²Peptide/protein
Asx protein ligases
Peptide/protein Asx
CONH
Xaa¹Xaa²Peptide/protein
b
Peptide/protein NGL
COOH/CONH₂
+
H₂N
GL Peptide/protein
OaAEP C247A
Peptide/protein N
CONH
GL Peptide/protein

Figure 2

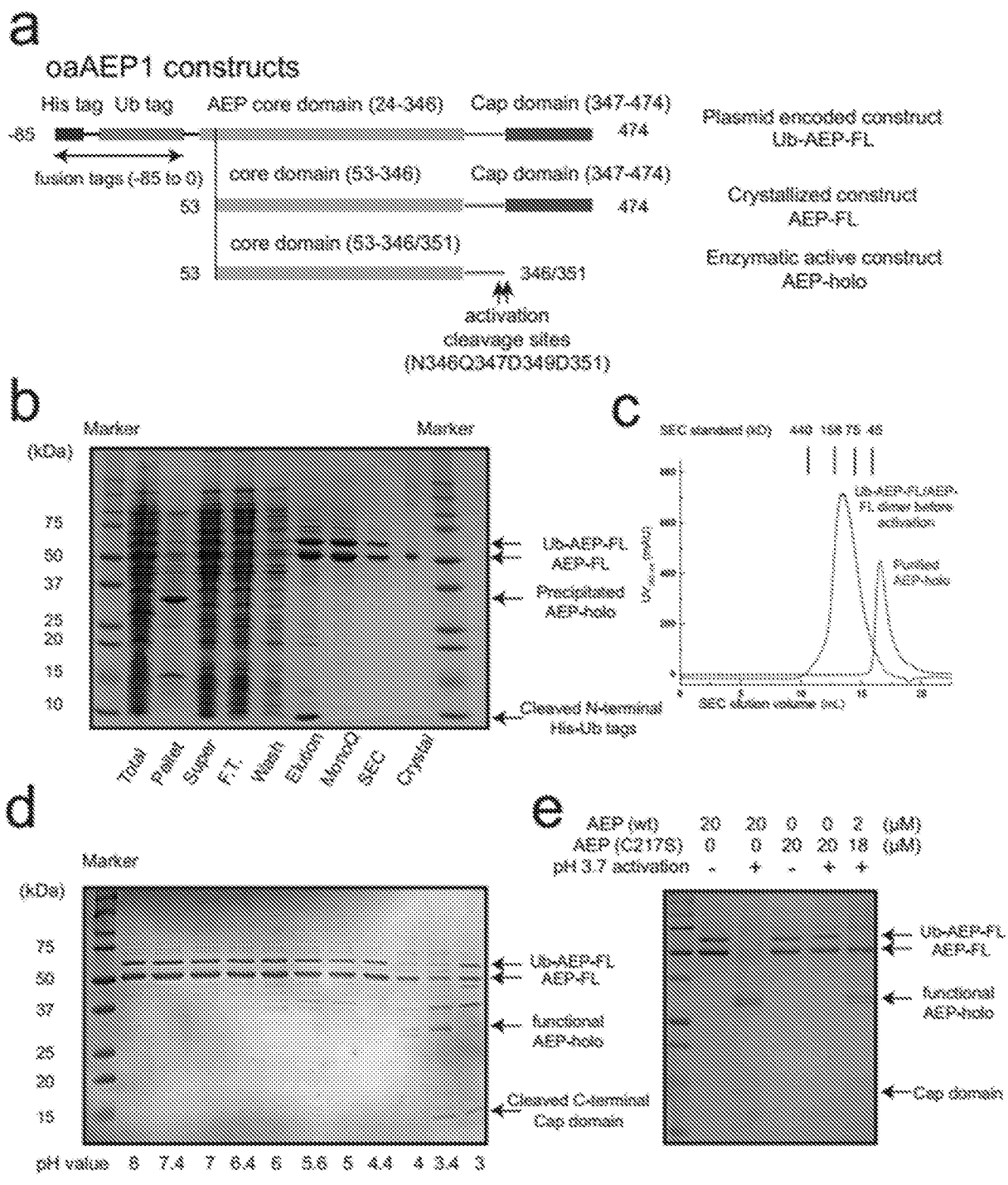

a
oaAEP1 constructs
His tag
Ub tag
AEP core domain (24-346)
Cap domain (347-474)
-85
474
fusion tags (-85 to 0)
core domain (53-346)
53
core domain (53-346/351)
346/351
activation cleavage sites (N346Q347D349D351)
Plasmid encoded construct Ub-AEP-FL
Crystallized construct AEP-FL
Enzymatic active construct AEP-holo
b
Marker
(kDa)
75
50
37
25
20
15
10
Ub-AEP-FL
AEP-FL
Precipitated AEP-holo
Cleaved N-terminal His-Ub tags
Total
Pellet
Super
F.T.
Wash
Elution
MonoQ
SEC
Crystal
c
SEC standard (kD)
440
158
75
45
Ub-AEP-FL/AEP-FL dimer before activation
Purified AEP-holo
SEC elution volume (mL)
d
Marker
(kDa)
75
50
37
25
20
15
Ub-AEP-FL
AEP-FL
functional AEP-holo
Cleaved C-terminal Cap domain
pH value 8 7.4 7 6.4 6 5.6 5 4.4 4 3.4 3
e
AEP (wt) 20 20 0 0 2 (μM)
AEP (C217S) 0 0 20 20 18 (μM)
pH 3.7 activation - + - + +
Ub-AEP-FL
AEP-FL
functional AEP-holo
Cap domain

Figure 3

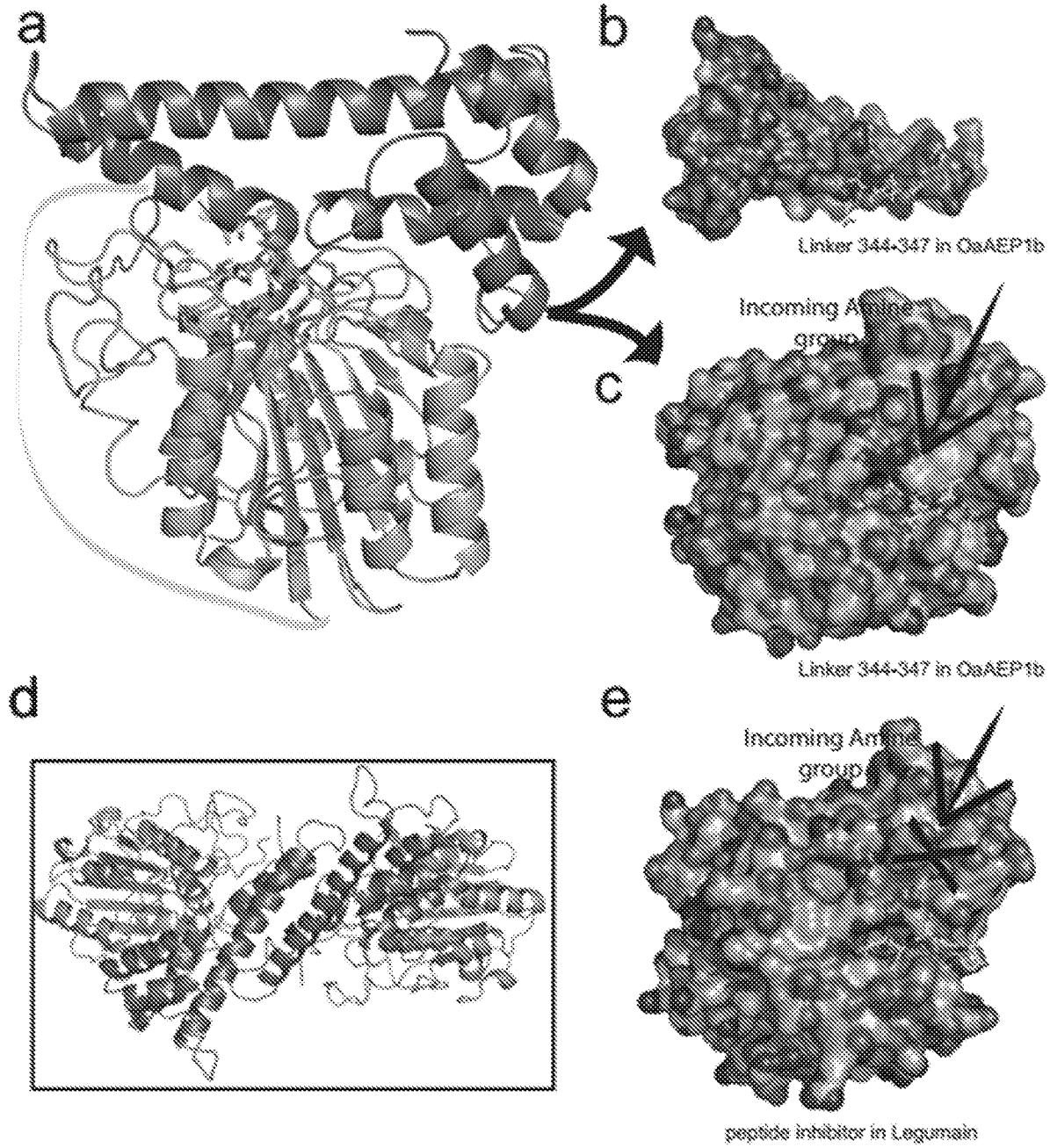

a
b
c
Linker 344-347 in OaAEP1b
Linker 344-347 in OaAEP1b
d
e
peptide inhibitor in Legumain

Figure 3f

**f**

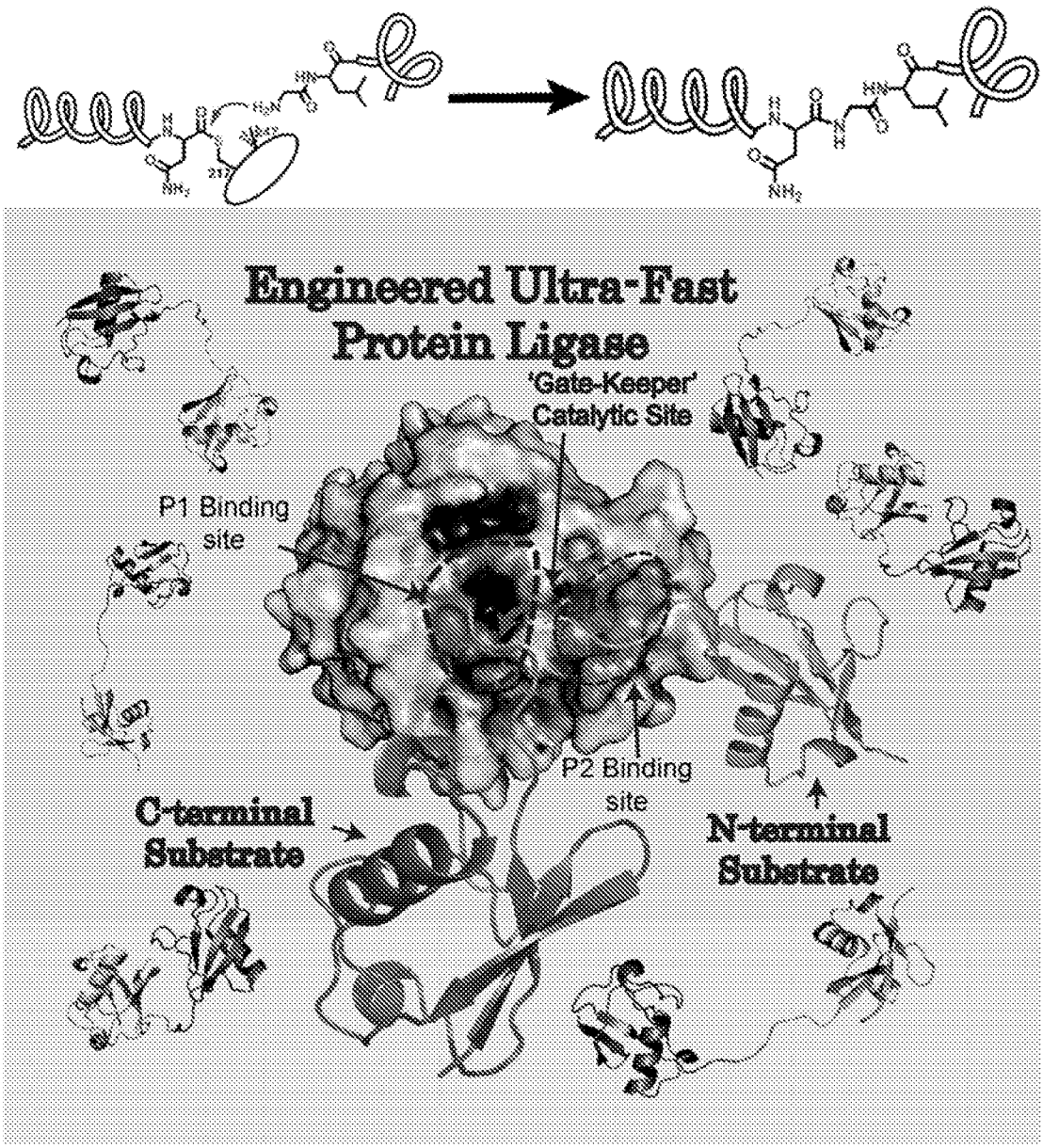
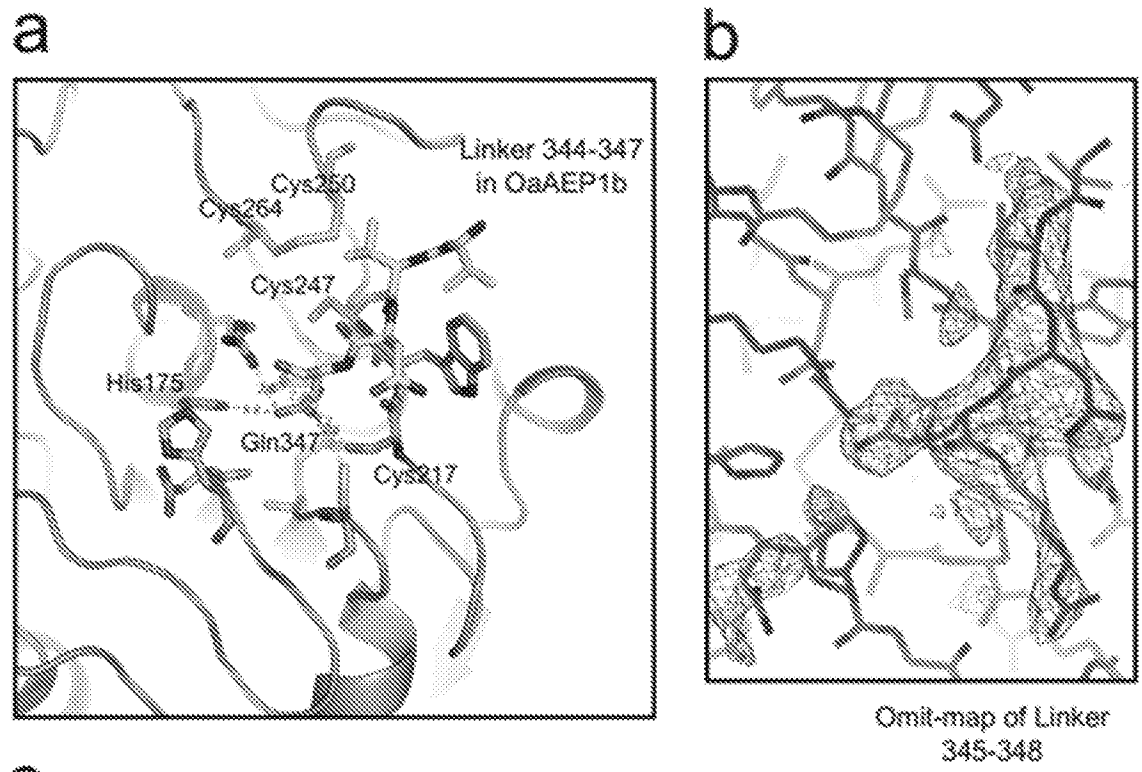

Engineered Ultra-Fast Protein Ligase
'Gate-Keeper' Catalytic Site
P1 Binding site
P2 Binding site
C-terminal Substrate
N-terminal Substrate

Figure 4


a
b
Linker 344-347
in OaAEP1b
Cys250
Cys264
Cys247
His175
Gln347
Cys217
Omit-map of Linker
345-348

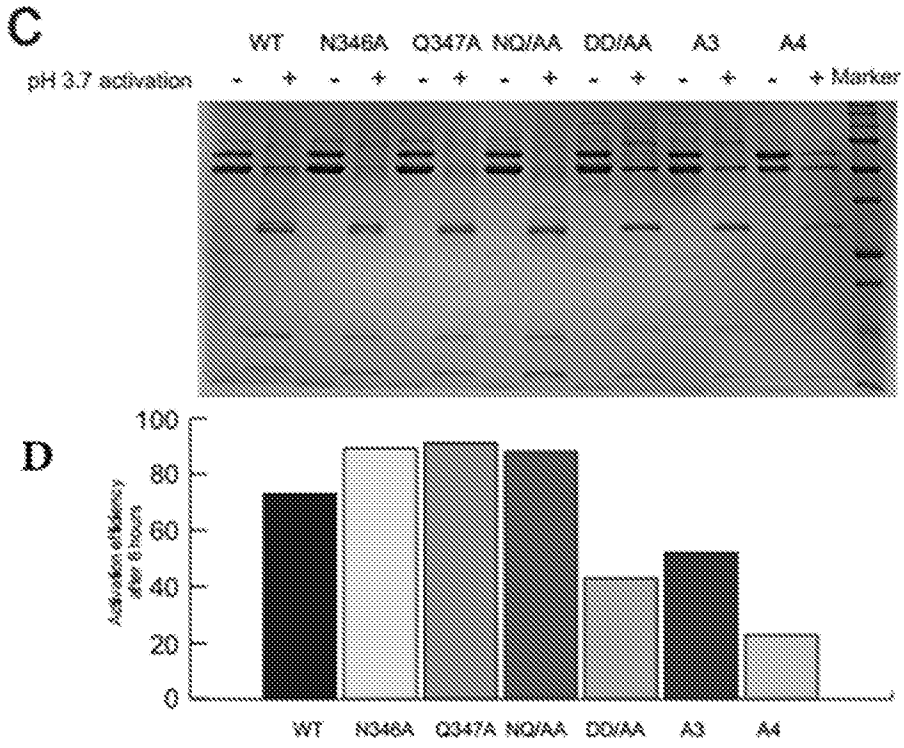
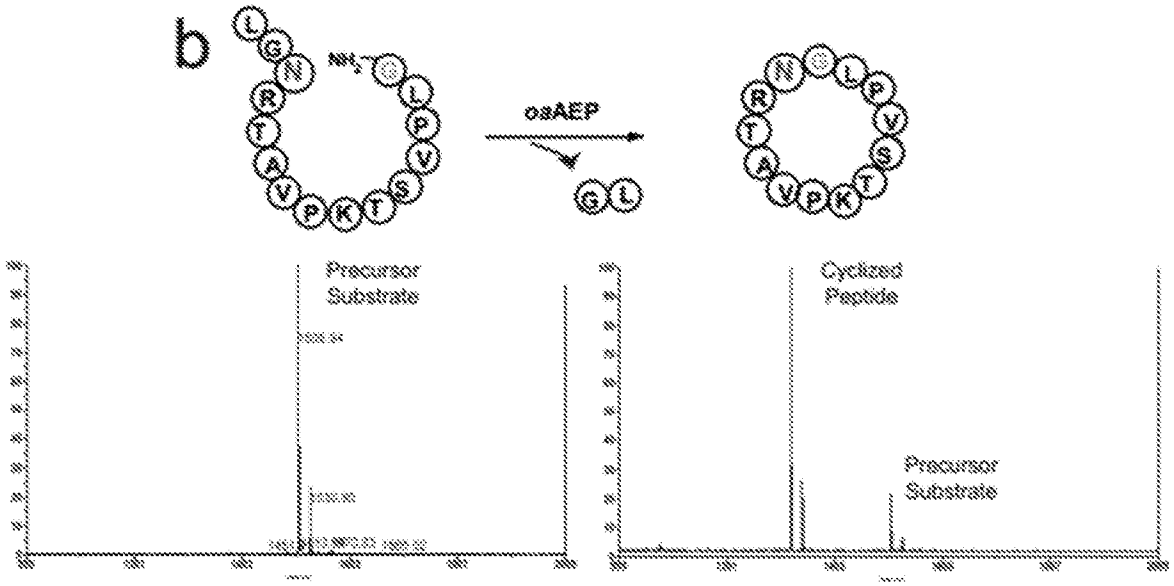
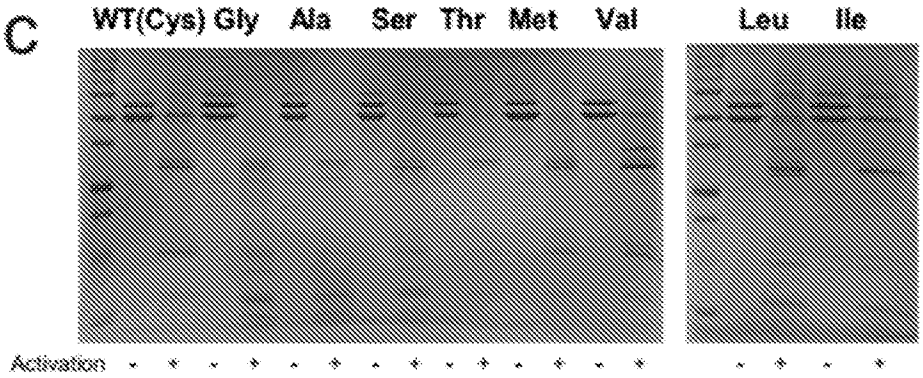

C
WT
N346A
Q347A
NQ/AA
DD/AA
A3
A4
pH 3.7 activation
- + - + - + - + - + - + - + Marker
D
100
80
60
40
20
0
WT
N346A
Q347A
NQ/AA
DD/AA
A3
A4

Figure 5

a

| Ligase | $k_{cat}$ ($s^{-1}$) | $K_m$ (μM) | $k_{cat}/K_m$ ($M^{-1} s^{-1}$) |
|---|---|---|---|
| WT AEP | 0.052 ± 0.008 | 245 ± 60 | 215 |
| AEP (Cys247Ala) | 13.9 ± 1.1 | 407 ± 49 | 34,209 |
| Butelase 1 | 4.83 ± 0.62 | 48.4 ± 14 | 99,793 |



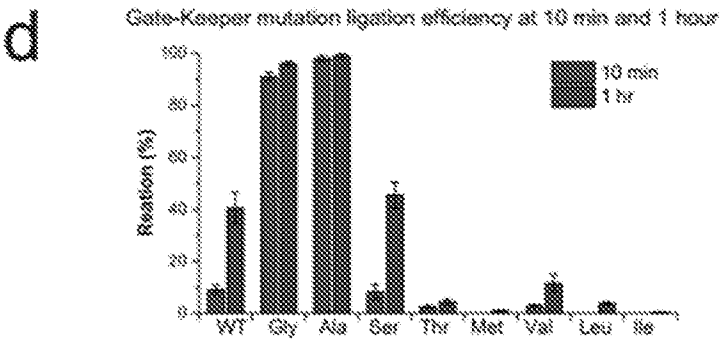



Figure 6

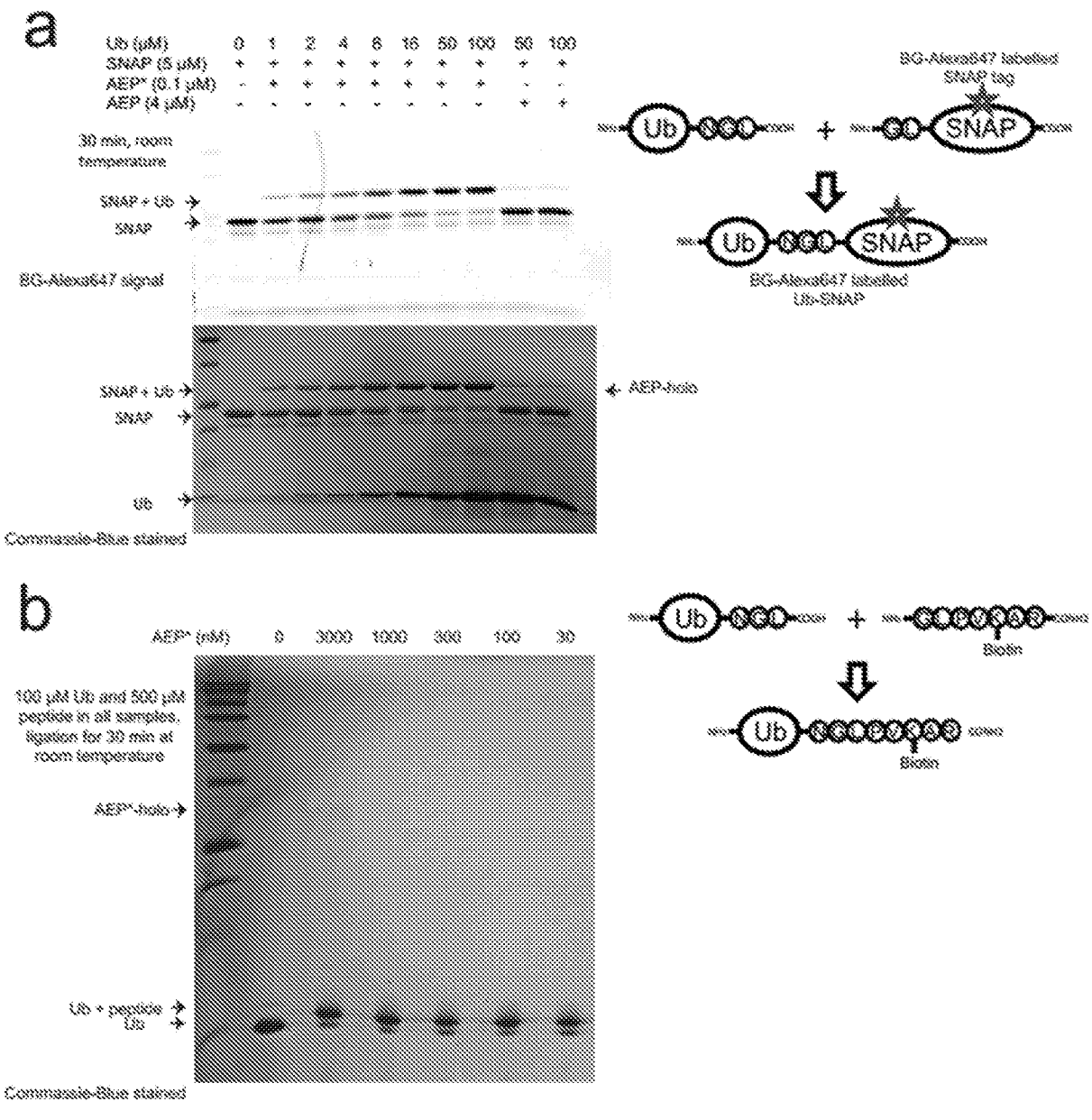
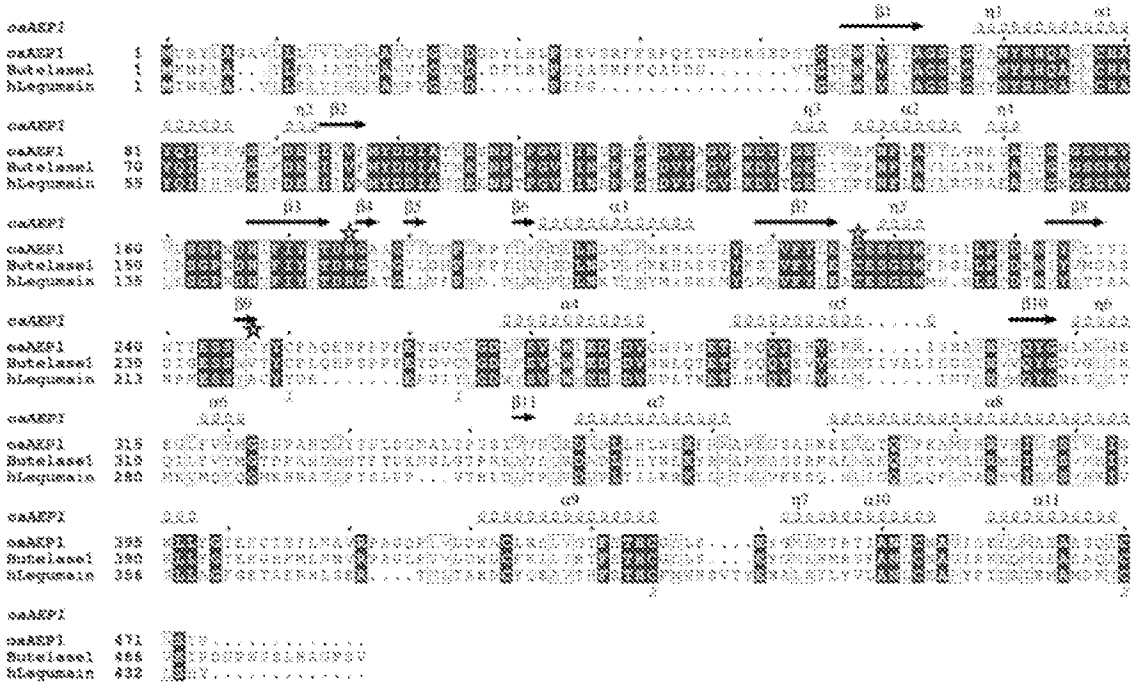

a
Ub (μM) 0 1 2 4 8 16 50 100 50 100
SNAP (5 μM) + + + + + + + + + +
AEP* (0.1 μM) - + + + + + + + - -
AEP (4 μM) - - - - - - - - + +
30 min, room temperature
SNAP + Ub
SNAP
BG-Alexa647 signal
SNAP + Ub
SNAP
Ub
Coomassie-Blue stained
AEP-biotin
BG-Alexa647 labelled SNAP tag
Ub
SNAP
BG-Alexa647 labelled Ub-SNAP
b
AEP* (nM) 0 3000 1000 300 100 30
100 μM Ub and 500 μM peptide in all samples, ligation for 30 min at room temperature
AEP*-biotin
Ub + peptide
Ub
Coomassie-Blue stained
Ub
Biotin
Ub
Biotin

Figure 7

A

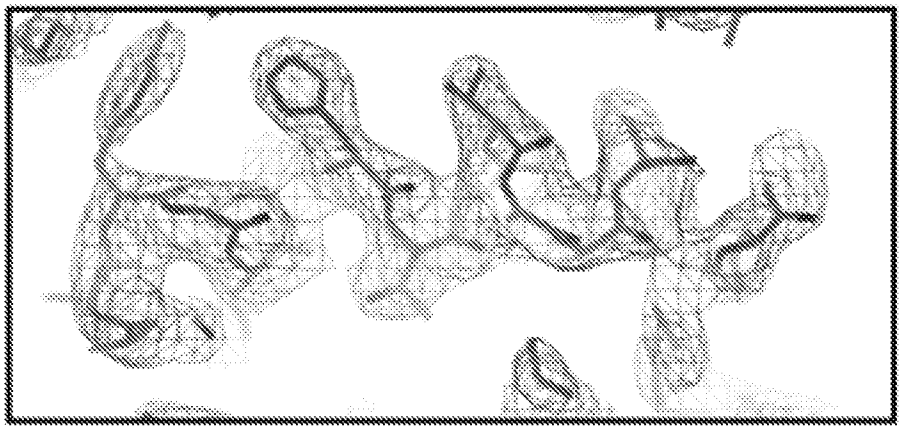


B


-85 MGMAHNHNHNHMQIFVETLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQ
-36 RLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGARDGDYLMLPSEVSRF
40 FRFQETNDDHGEDSVGTRWAVLIAGSKGYANYRHQAGVCHAYQILKRGG
89 LKDENIVVFMYDDIAYNESNPRPGVIINSPHGSDVYAGVPKDYTGEEVNAK
140 NFLAAILGNKSAITGGSGKVVDSGPNDHIFIYYTDHGAAGVIGMPSKPYLY
191 ADELNDALKKKHASGTYKSLVFYLEACESGSMFEGILPEDLNIYALTSTNTT
243 ESSWCYYCPAQENPPPPEYNVCLGDLFSVAWLEDSDVQNSWYETLNQQY
292 HHVQKRISHASHATQYGNLKLGEEGLFVYMGSNPANDNYTSLDGNALTPS
342 SIVVNGNDADLLHLWEKFRKAPEGSARKEEAQTQIFKAMSHRVHIDSSIKL
393 IGKLLFGIEKCTEILNAVRPAGQPLVDDWACLRSLVGTFETHCGSLSEYGM
443 RHTRTIANICNAGISEEQMAEAASQACASIP
Expression Fusion tags (-85 to 0)
Core Domain (24 to 325)
Linker (326 to 347)
Cap Domain (348 to 474)

Figure 8

a



b

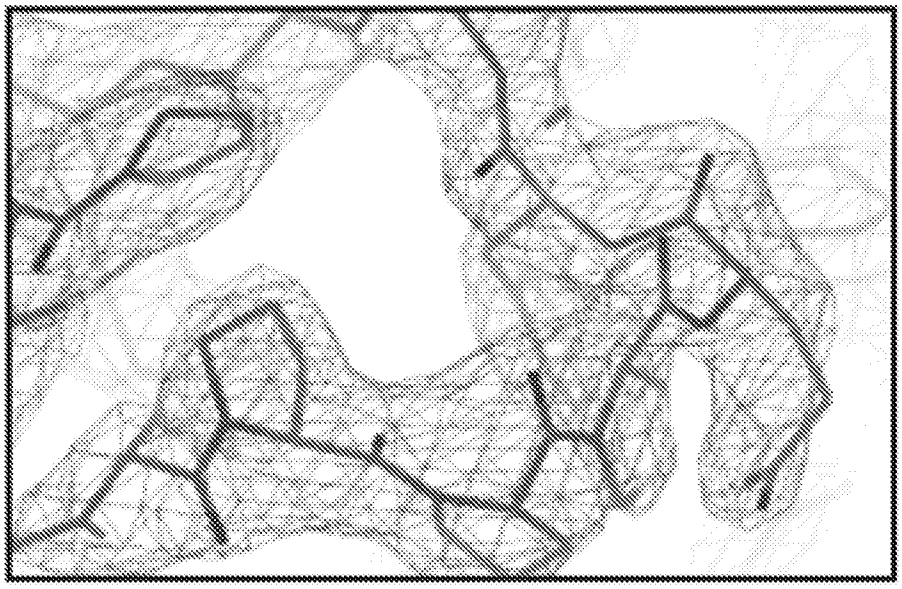
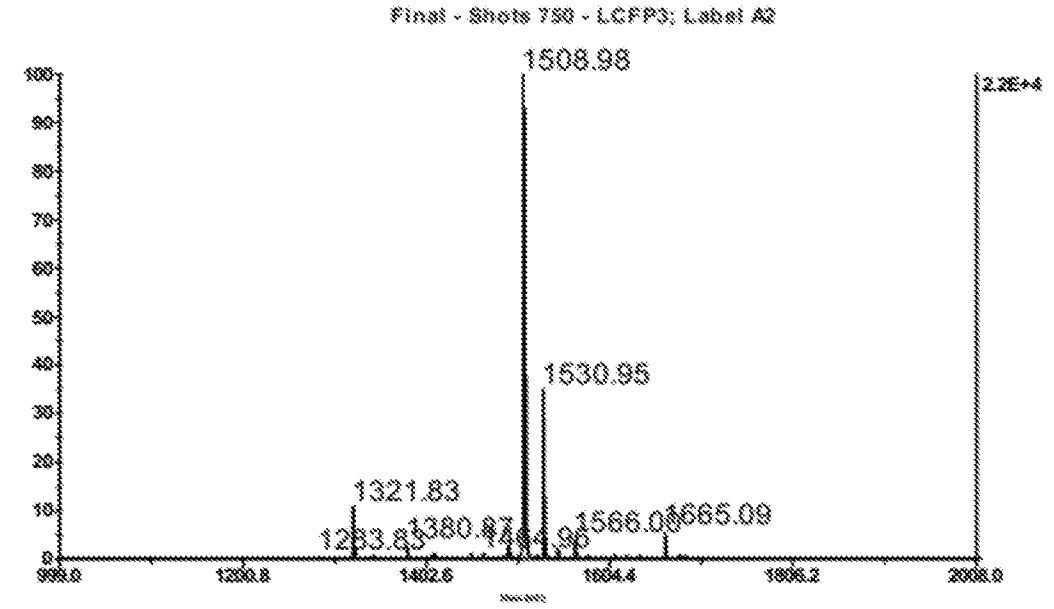
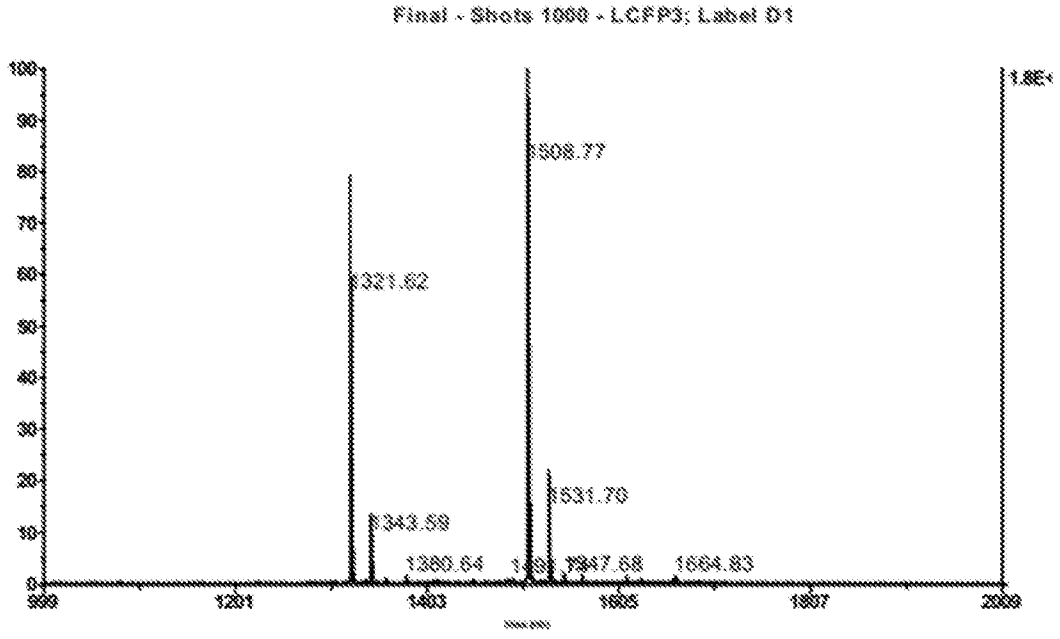


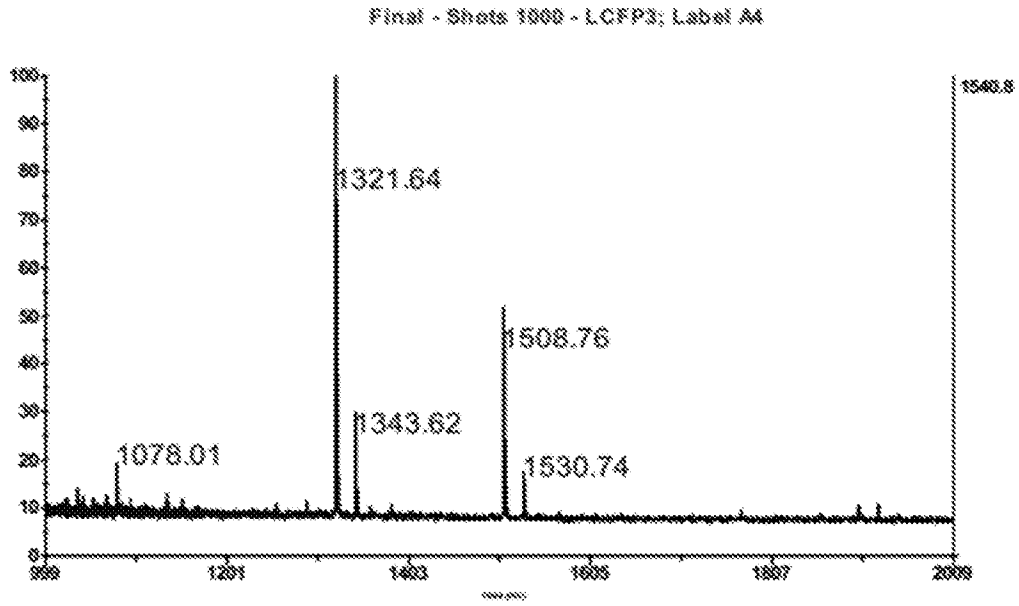
Figure 9



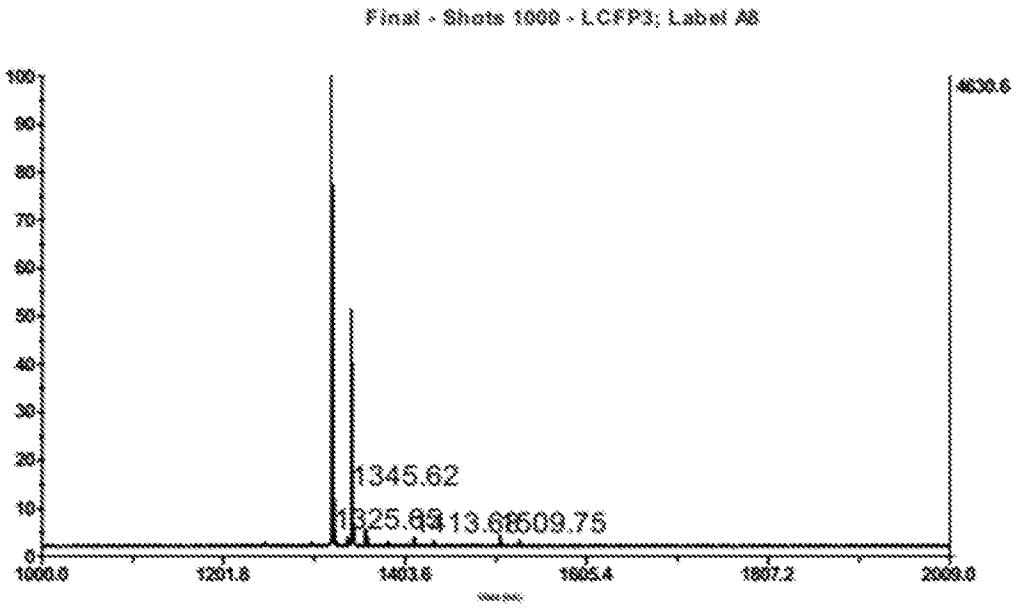

WT, 10MIN



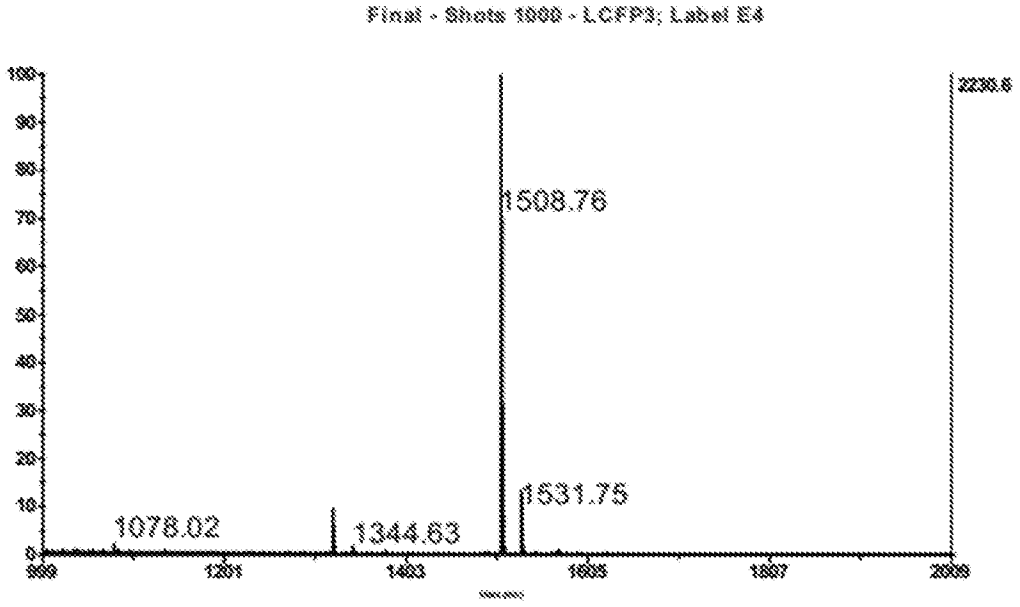

AEP WT, 1H



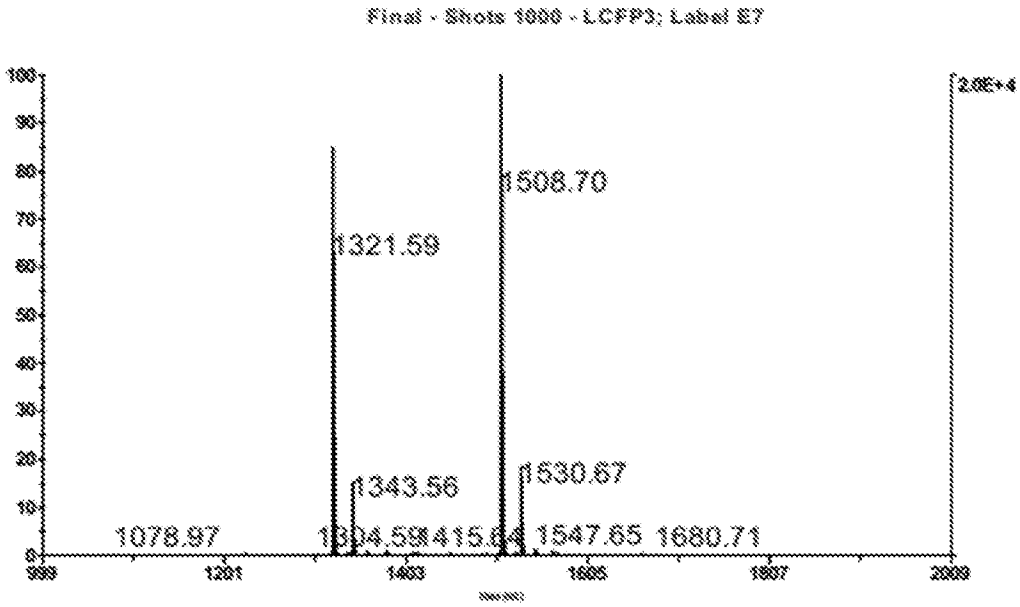

Gly, 10 min



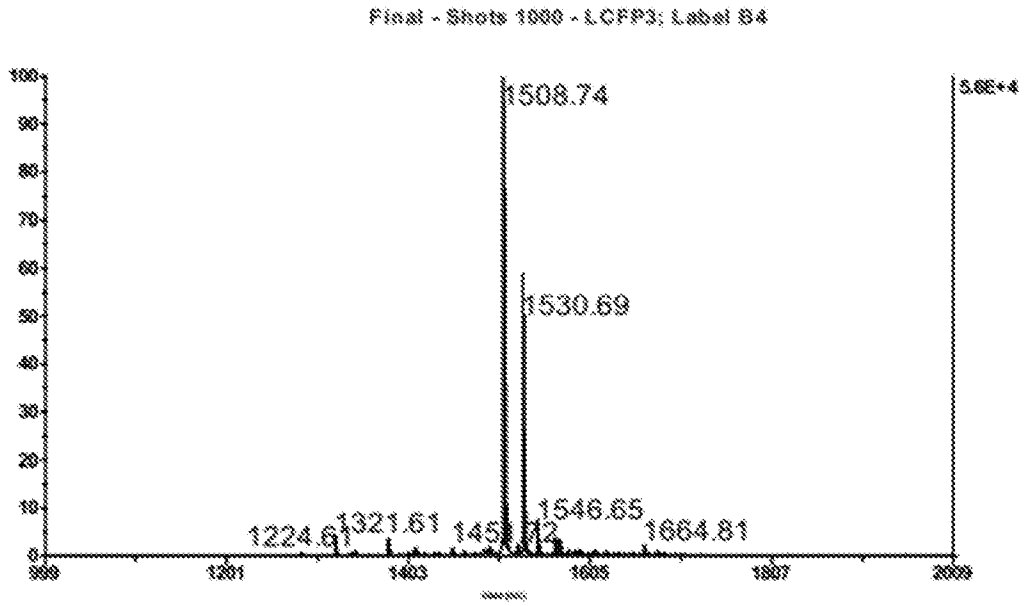

Gly, 1 h



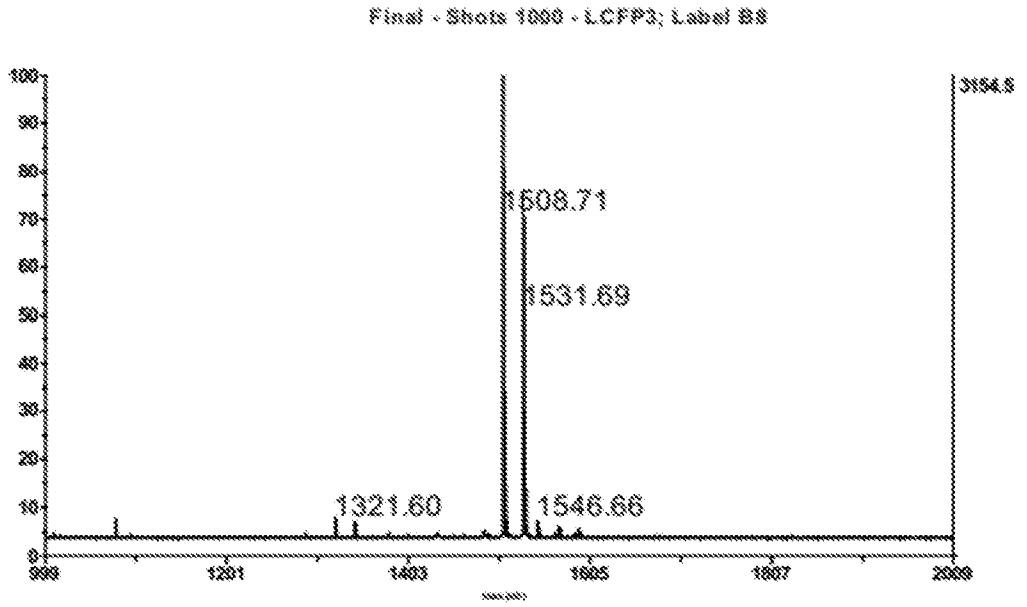

Ser, 10 min



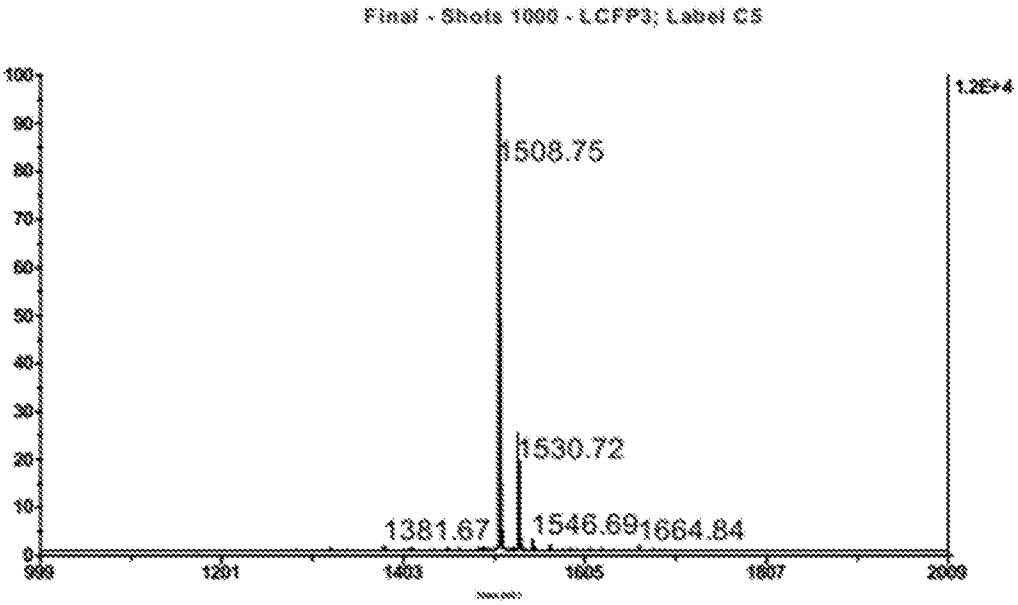

Ser, 1 h



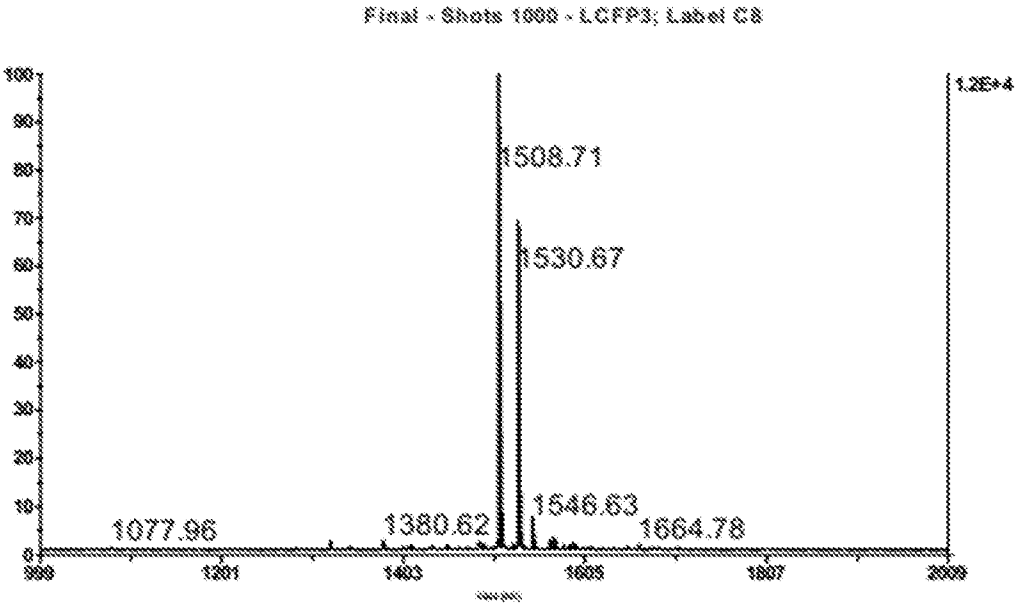

Thr, 10 min



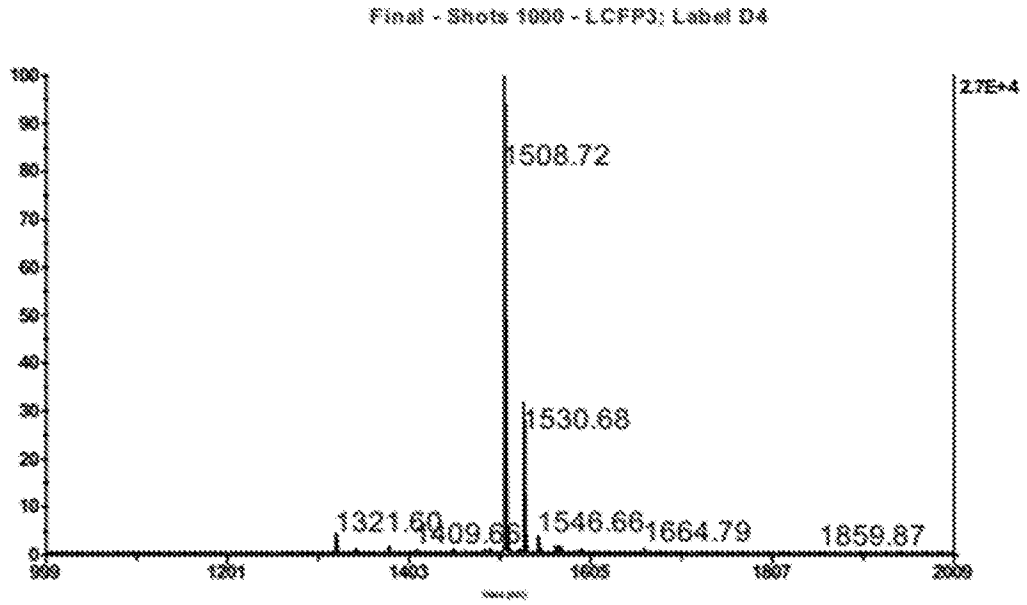

Thr, 1 h



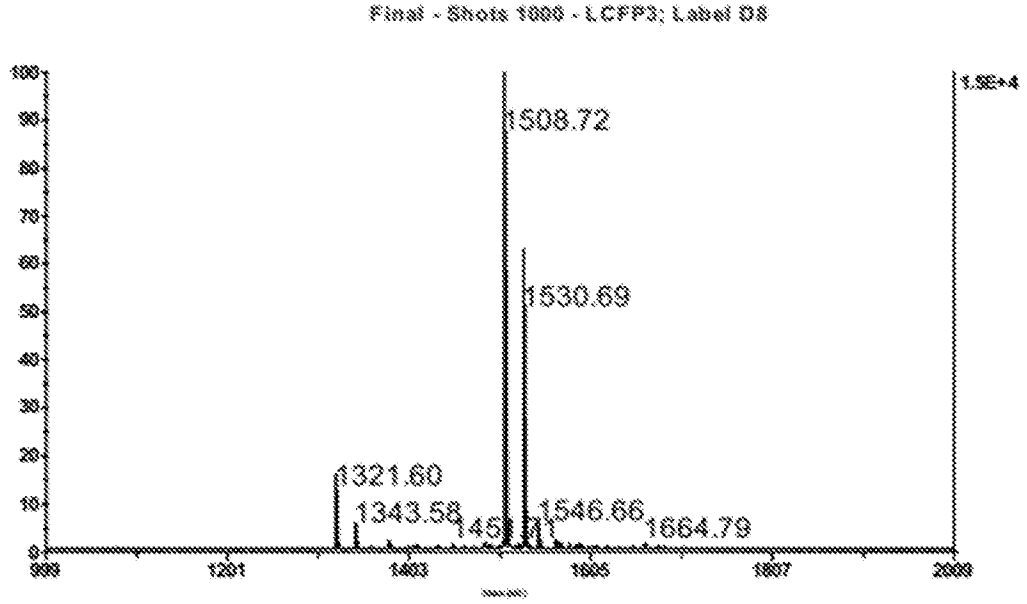

Met, 10 min



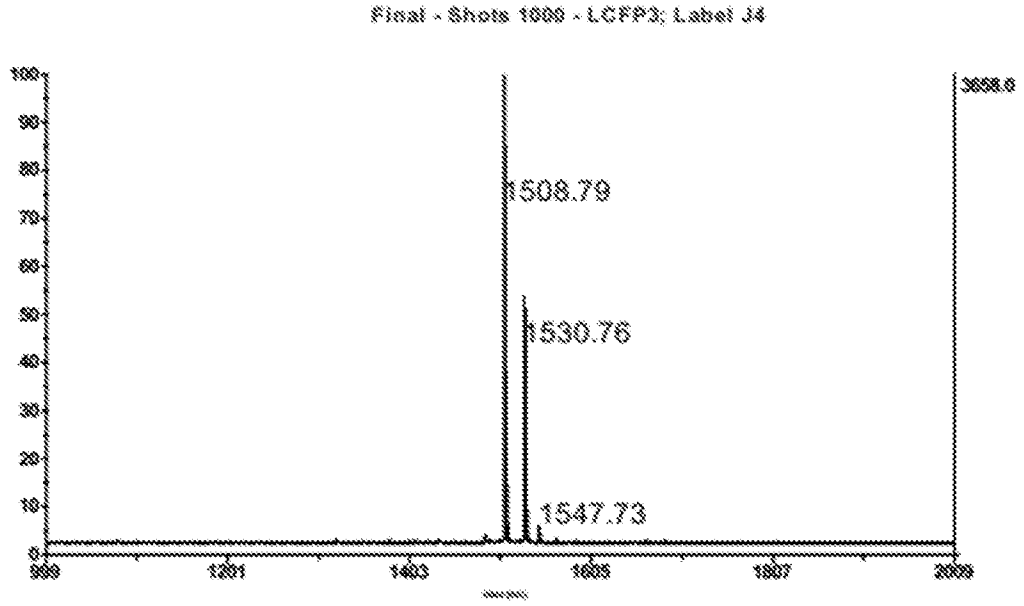

Met, 1 h



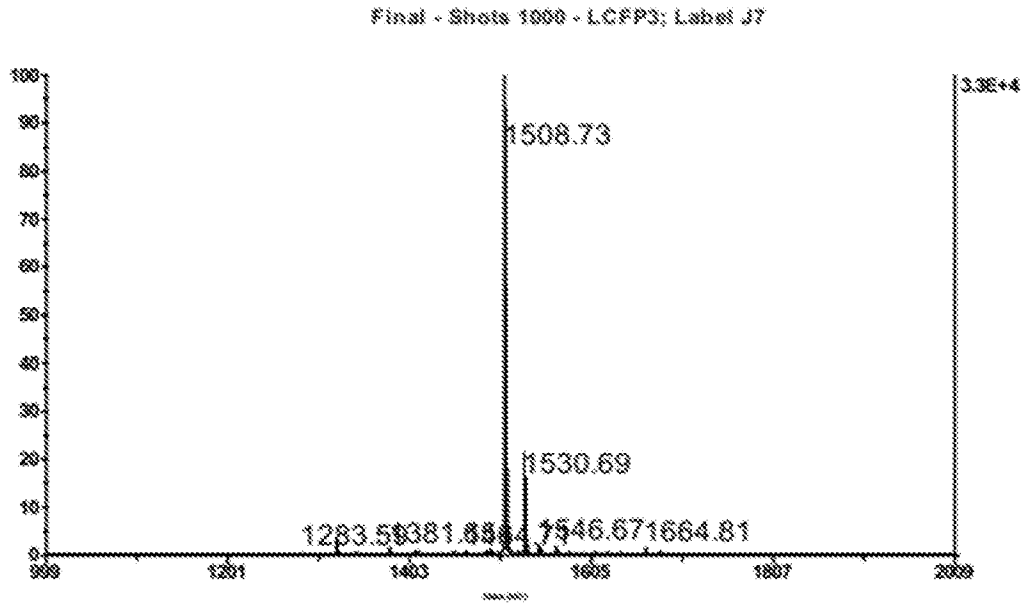

Val, 10 min



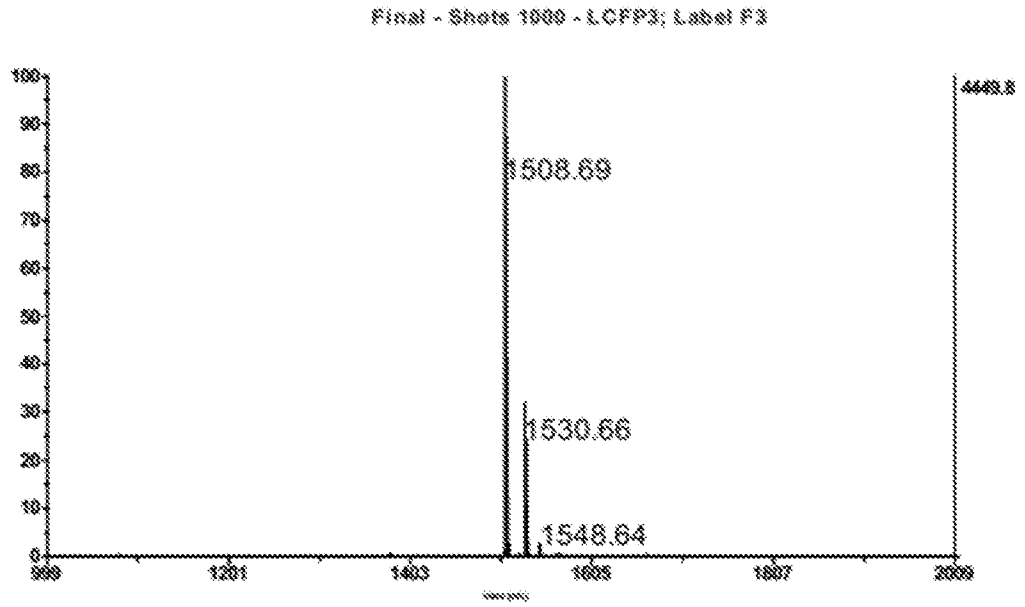

Val, 1 h



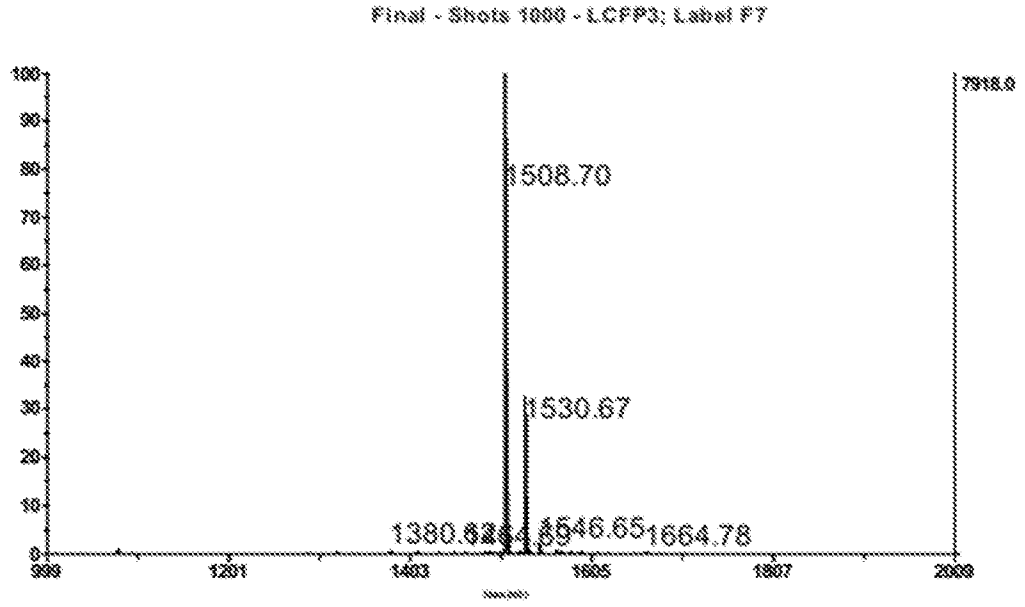

Leu, 10 min




Leu, 1 h




Ile, 10 min




Ile, 1 h

Figure 10

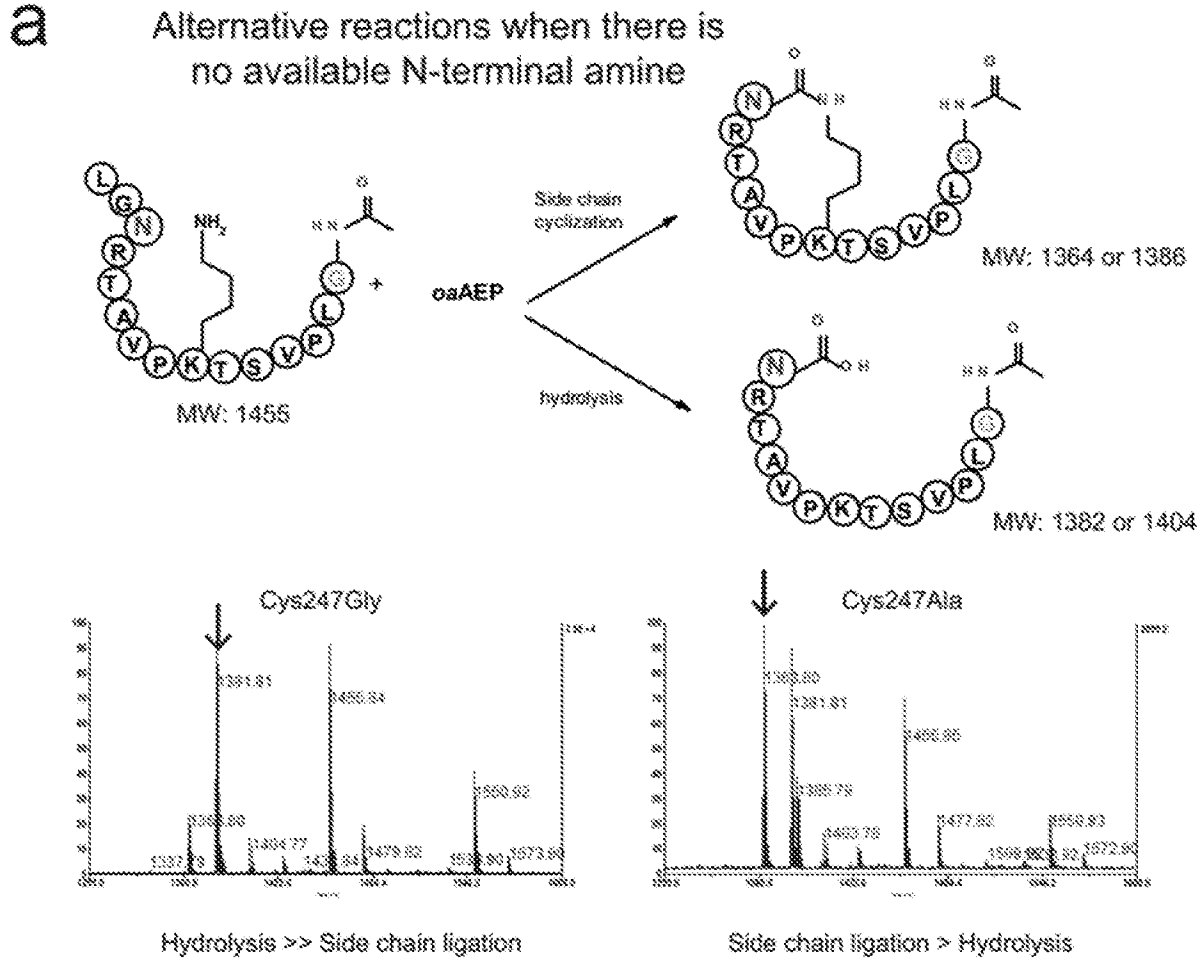
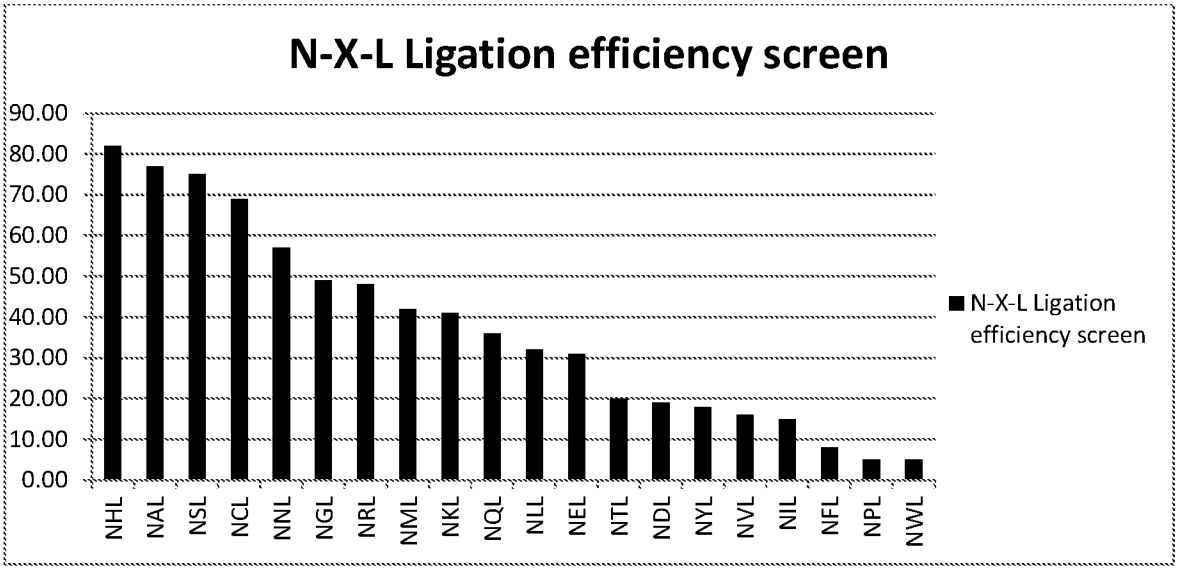
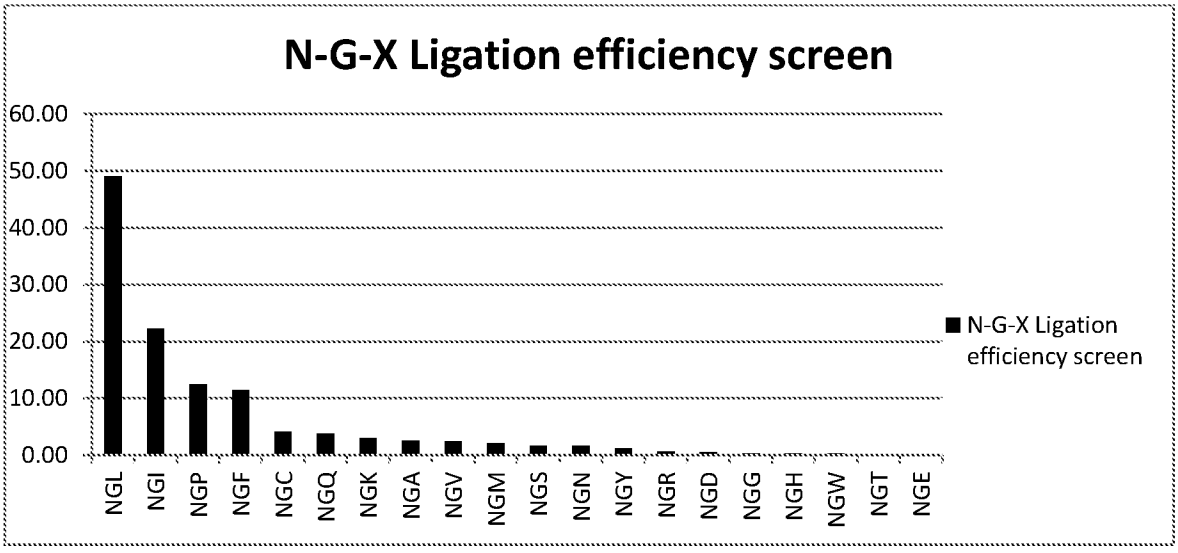

a
Alternative reactions when there is no available N-terminal amine
MW: 1455
oaAEP
Side chain cyclization
hydrolysis
MW: 1384 or 1386
MW: 1382 or 1404
Cys247Gly
Cys247Ala
Hydrolysis >> Side chain ligation
Side chain ligation > Hydrolysis

Figure 11

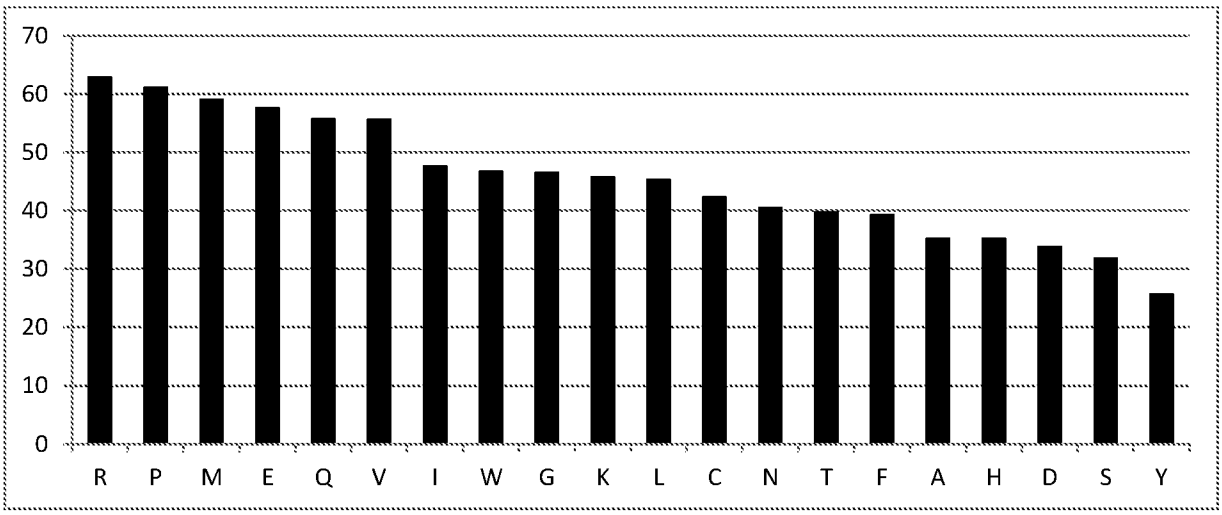
**a**


N-X-L Ligation efficiency screen
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
NHL
NAL
NSL
NCL
NNL
NGL
NRL
NML
NKL
NQL
NLL
NEL
NTL
NDL
NYL
NVL
NIL
NFL
NPL
NWL
N-X-L Ligation efficiency screen

**b**

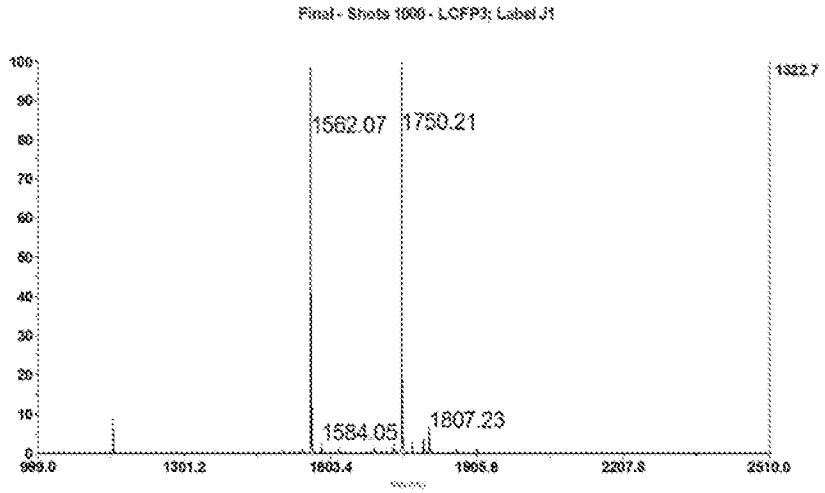

N-G-X Ligation efficiency screen
60.00
50.00
40.00
30.00
20.00
10.00
0.00
NGL
NGI
NGP
NGF
NGC
NGQ
NGK
NGA
NGV
NGM
NGS
NGN
NGY
NGR
NGD
NGG
NGH
NGW
NGT
NGE
N-G-X Ligation efficiency screen

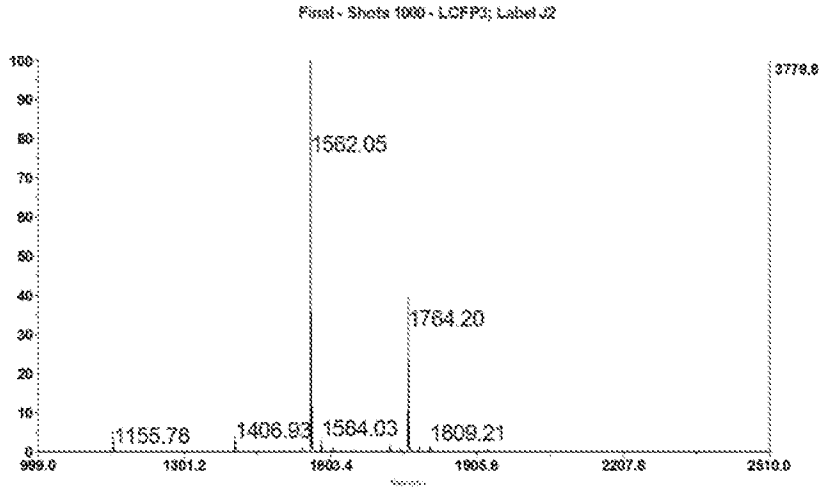
**c**

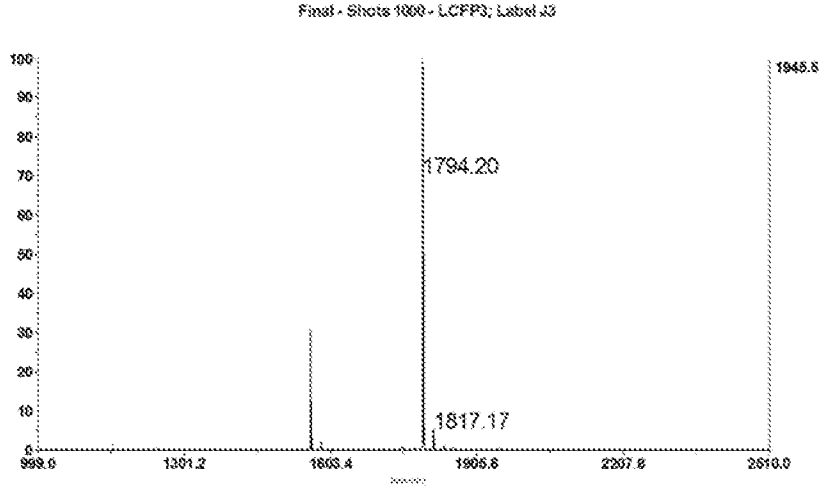


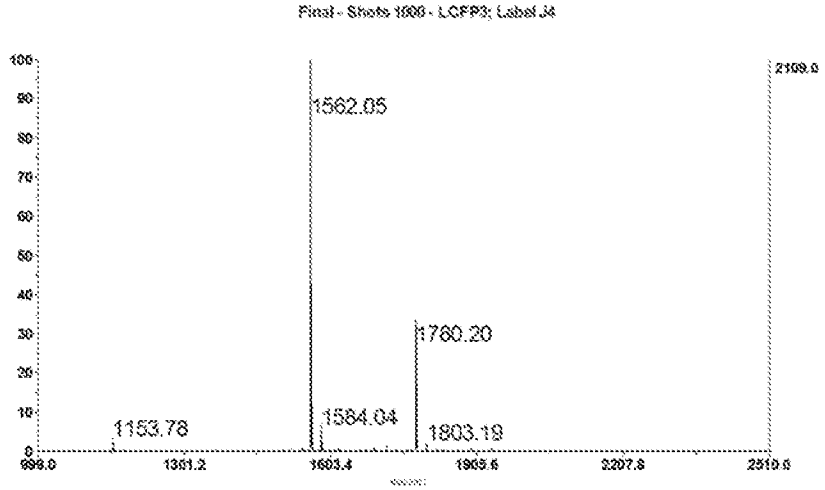

**d**

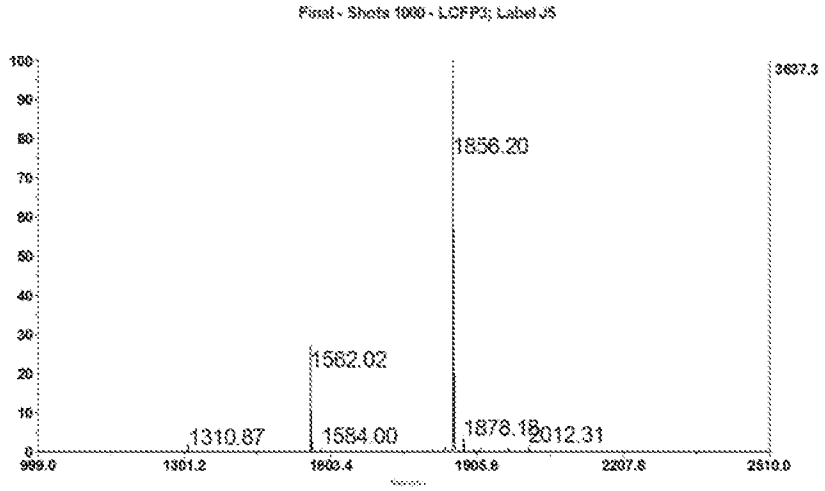
GLYRRGRLYRRNXL (J1-J20: NGL, NAL, NTL, NSL, NYL, NCL, NEL, NPL, NLL, NIL, NNL, NWL, NRL, NHL, NDL, NVL, NML, NKL, NQL, NFL, total 20 peptides)

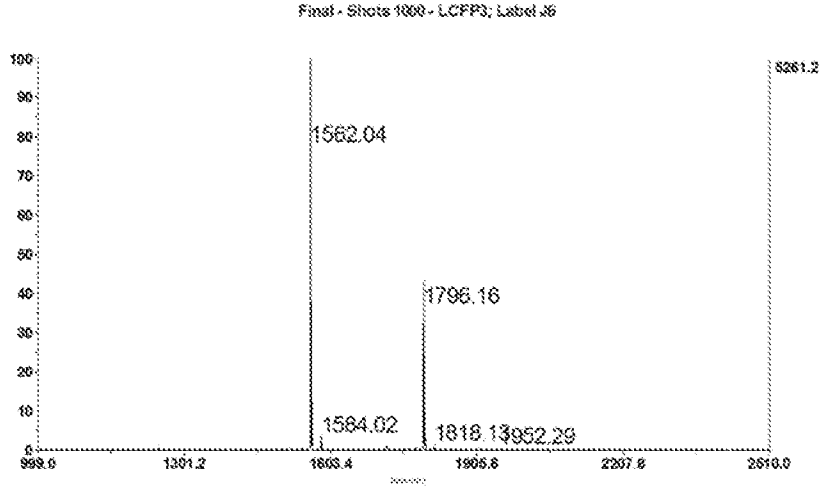


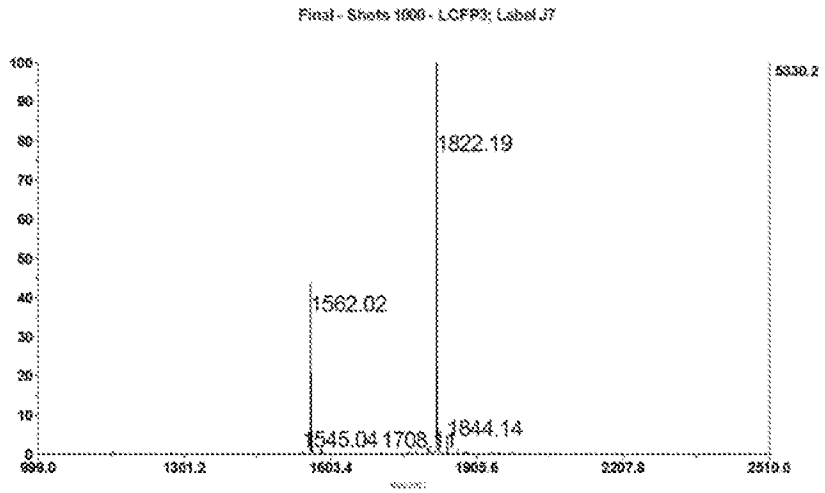

Final - Shots 1000 - LCFP3; Label J2
3779.8
1562.05
1764.20
1155.78
1406.93
1584.03
1809.21
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

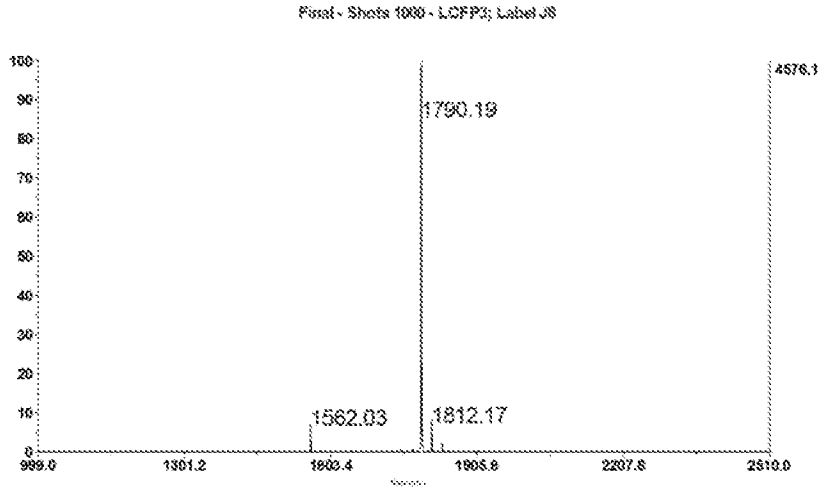

Final - Shots 1000 - LCFP3; Label J3
1945.8
1794.20
1817.17
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

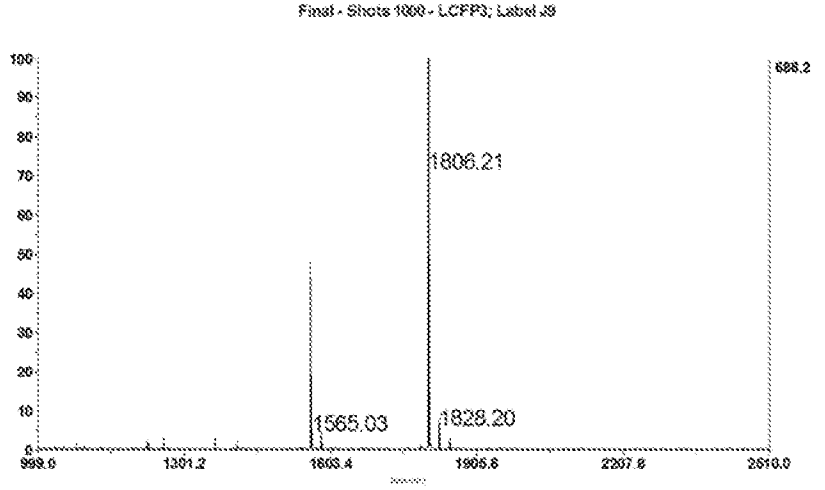

Final - Shots 1000 - LCFP3; Label J4
2109.0
1562.05
1760.20
1153.78
1584.04
1803.19
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

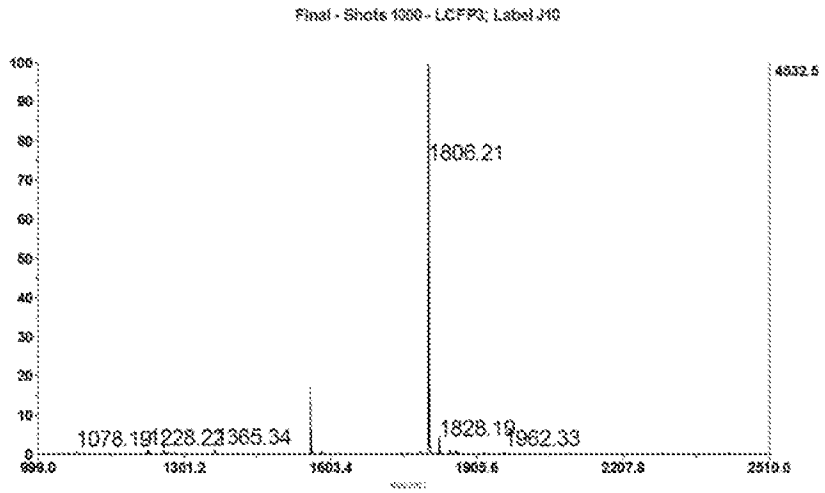

Final - Shots 1000 - LCFP3; Label J5
3687.3
1856.20
1562.02
1310.87
1584.00
1878.18
2012.31

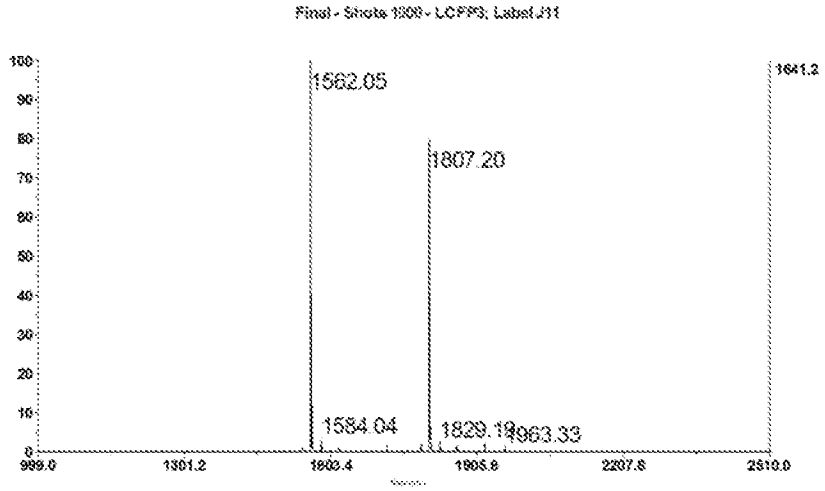

Final - Shots 1000 - LCFP3; Label J6
5261.2
1562.04
1796.16
1584.02
1818.13
952.29

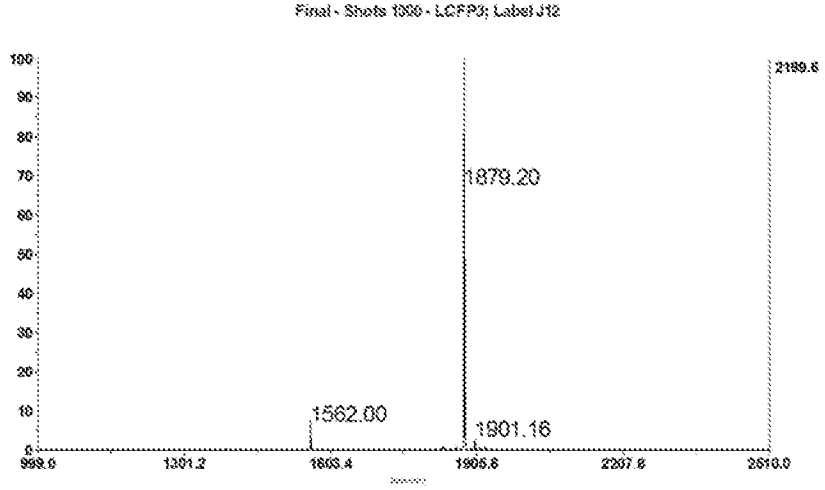

Final - Shots 1000 - LCFP3; Label J7
5330.2
1822.19
1562.02
1545.04
1708.11
1844.14

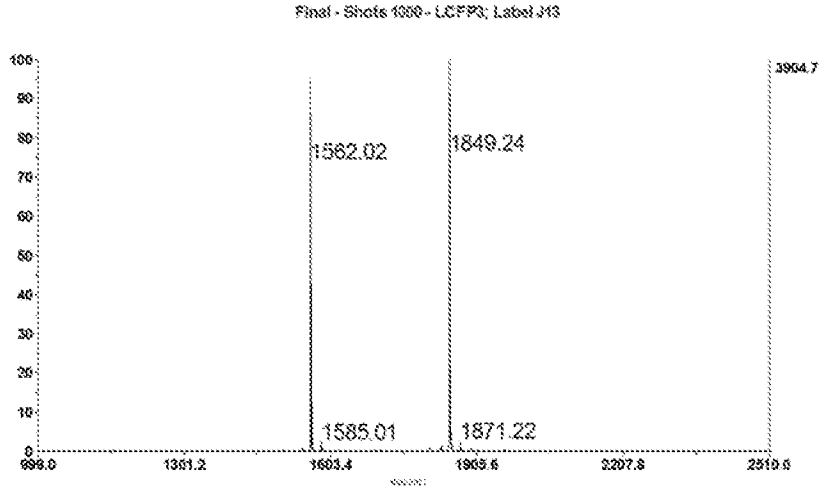

Final - Shots 1000 - LCFP3; Label J8
4576.1
1790.19
1562.03
1812.17
999.0
1301.2
1603.4
1905.8
2207.8
2510.0

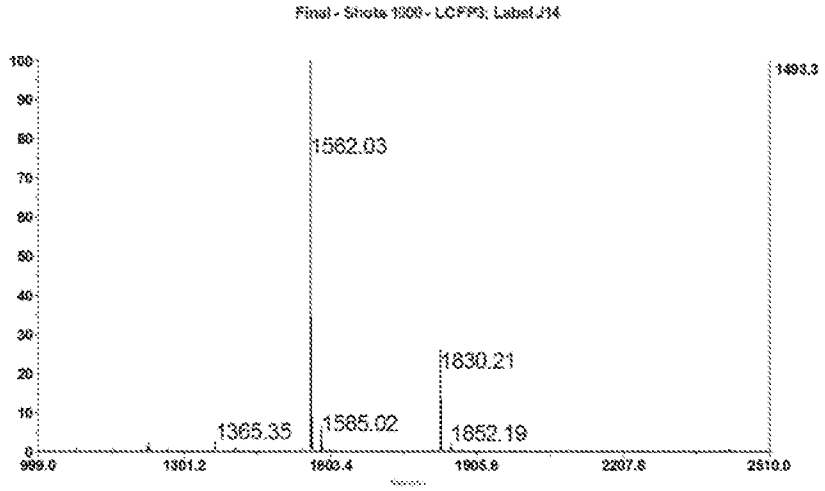

Final - Shots 1000 - LCFP3; Label J9
686.2
1806.21
1565.03
1828.20
999.0
1301.2
1603.4
1905.8
2207.8
2510.0

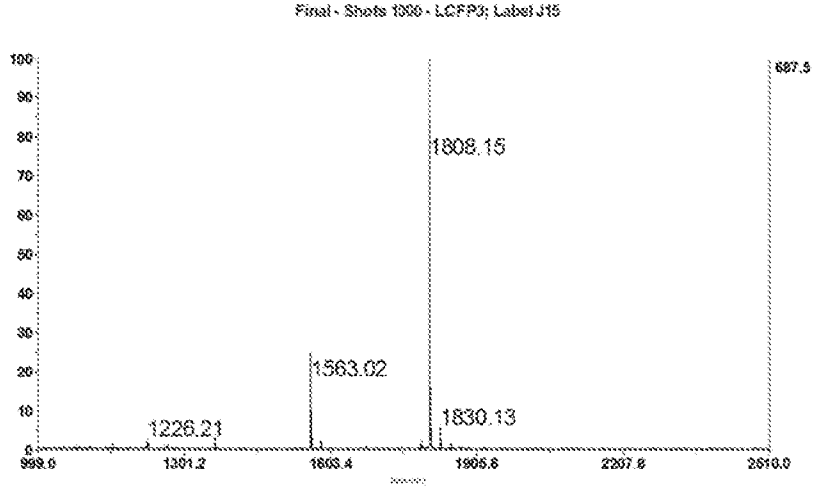

Final - Shots 1000 - LCFP3; Label J10
4632.5
1806.21
1078.19
1228.23
1365.34
1828.19
1962.33
999.0
1301.2
1603.4
1905.8
2207.8
2510.0

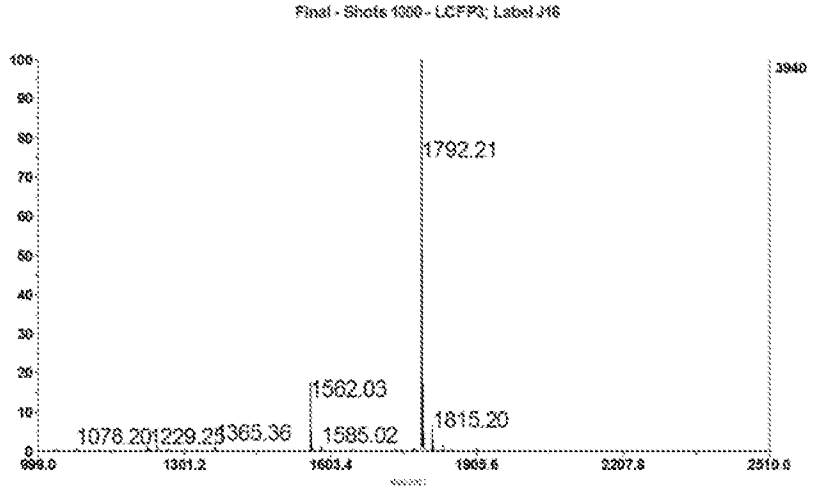

Final - Shots 1000 - LCFP3; Label J11
1562.05
1807.20
1584.04
1829.19
963.33
1641.2
999.0
1301.2
1603.4
1905.8
2207.6
2510.0

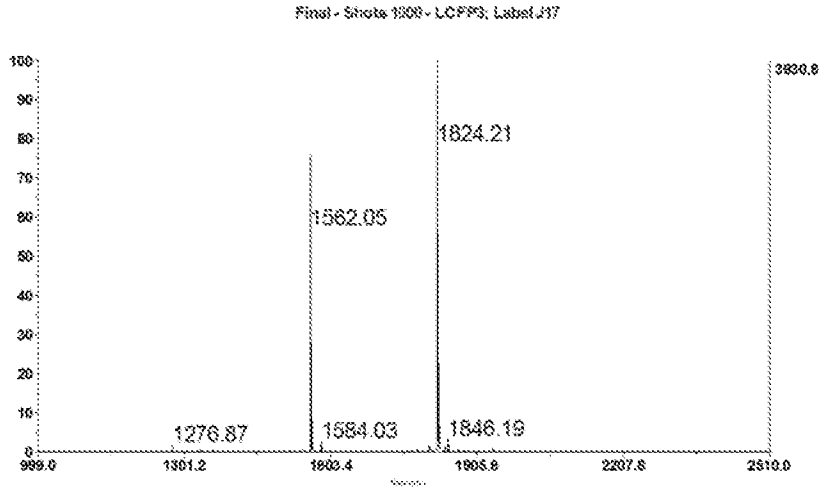

Final - Shots 1000 - LCFP3; Label J12
1879.20
1562.00
1901.16
2189.6
999.0
1301.2
1603.4
1905.8
2207.6
2510.0

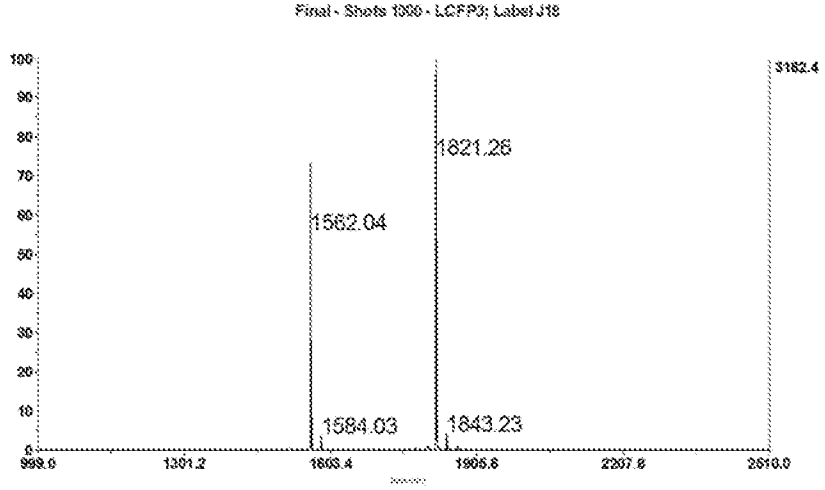

Final - Shots 1000 - LCFP3; Label J13
1562.02
1849.24
1585.01
1871.22
3904.7
999.0
1301.2
1603.4
1905.8
2207.6
2510.0

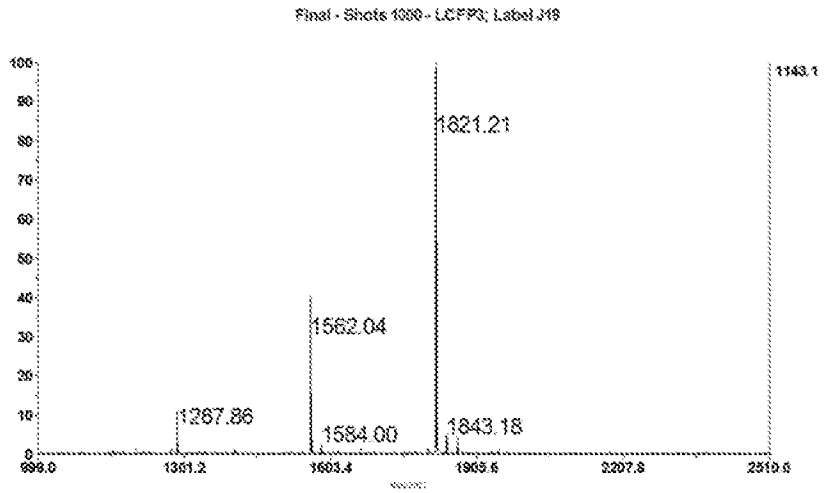

Final - Shots 1000 - LCFP3; Label J14
100
90
80
70
60
50
40
30
20
10
0
1562.03
1493.3
1830.21
1365.35
1585.02
1852.19
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

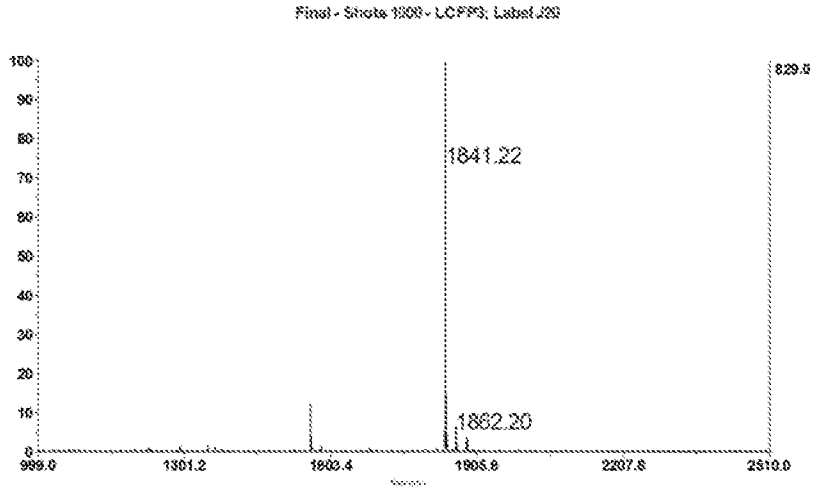

Final - Shots 1000 - LCFP3; Label J15
100
90
80
70
60
50
40
30
20
10
0
1808.15
687.5
1563.02
1830.13
1226.21
999.0
1301.2
1603.4
1905.6
2207.8
2510.0


Final - Shots 1000 - LCFP3; Label J16
100
90
80
70
60
50
40
30
20
10
0
1792.21
3940
1562.03
1815.20
1078.20
1229.25
1365.36
1585.02
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

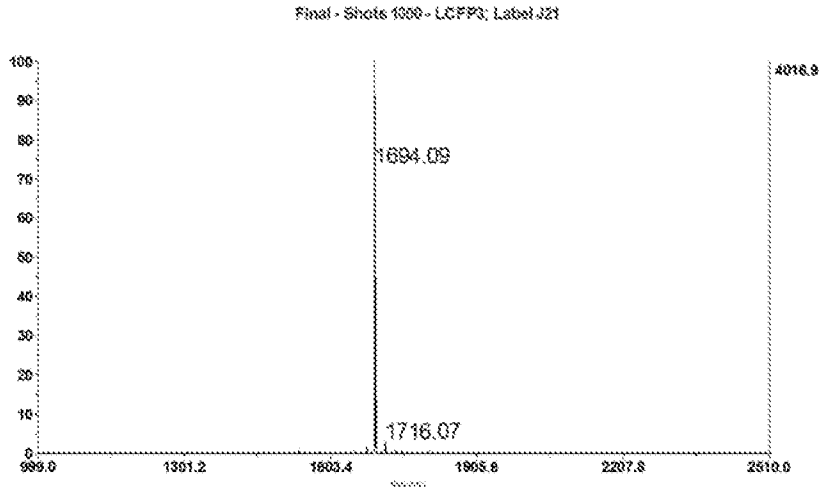

Final - Shots 1000 - LCFP3; Label J17
1824.21
1562.05
1276.87
1584.03
1846.19

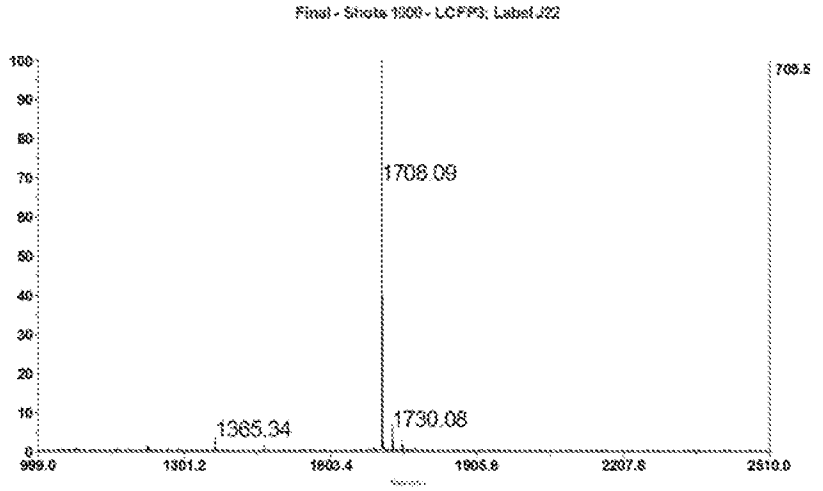

Final - Shots 1000 - LCFP3; Label J18
1821.26
1562.04
1584.03
1843.23

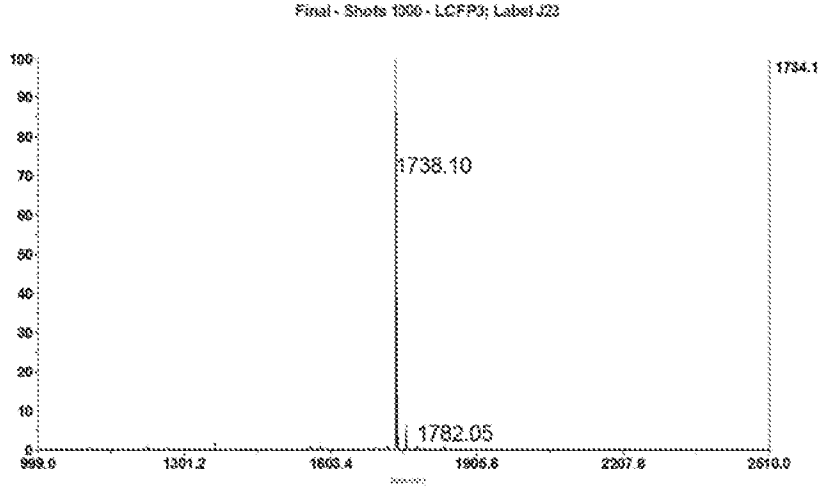

Final - Shots 1000 - LCFP3; Label J19
1821.21
1562.04
1287.86
1584.00
1843.18

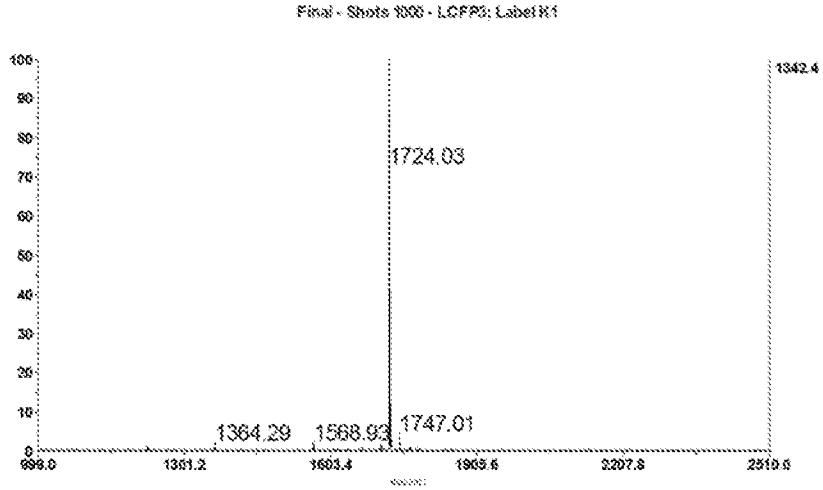


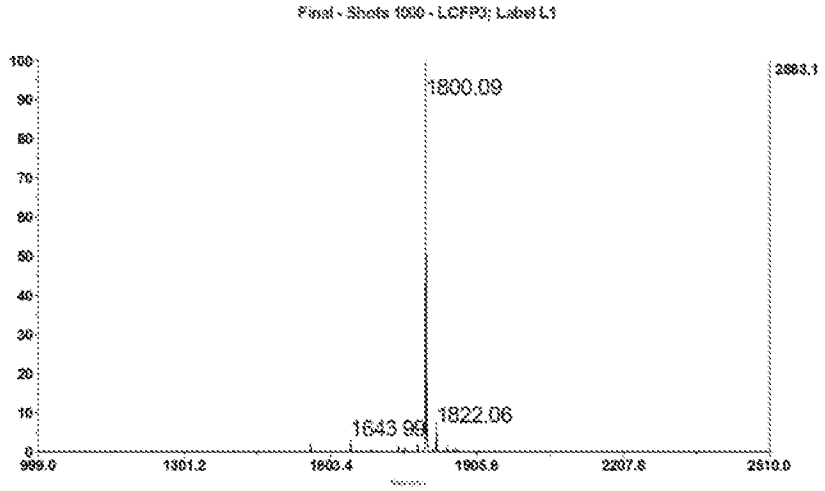

GLYRRGRLYRRNGX (J21-L16: NGG, NGA, NGT, NGS, NGY, NGC, NGE, NGP, NGI, NGN, NGW, NGR, NGH, NGD, NGV, NGM, NGK, NGQ, NGF total 19 peptides, NGL is the same peptide as the first plot-J1)

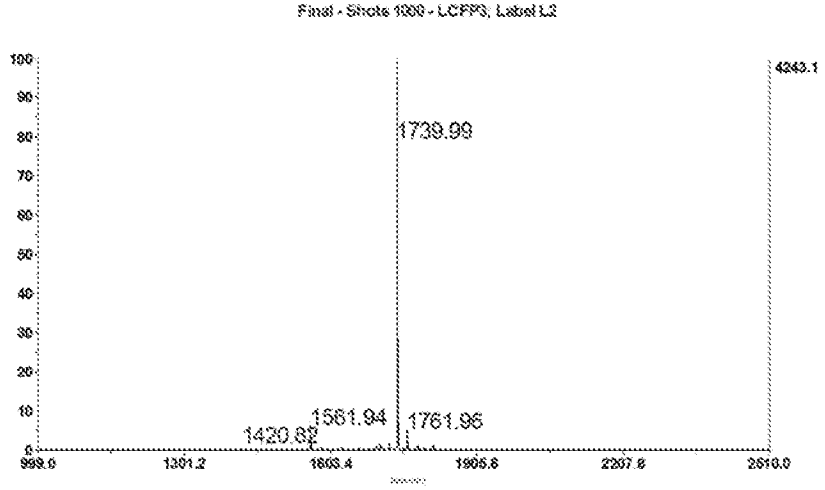


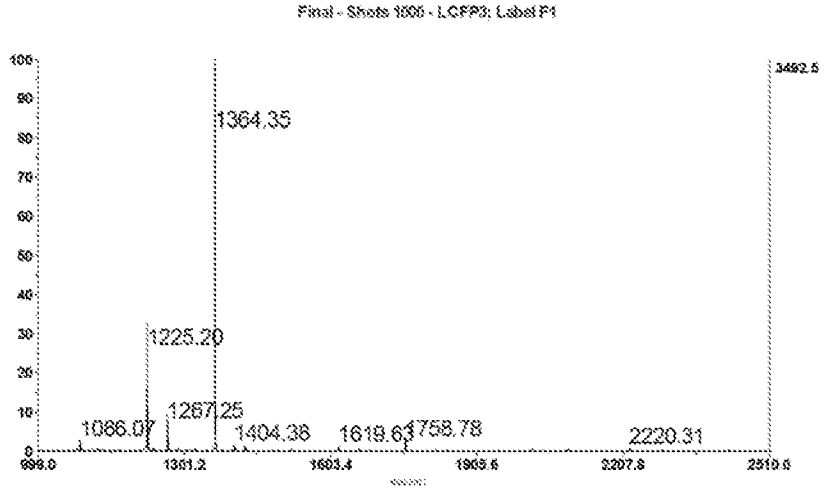

Final - Shots 1000 - LCFP3; Label J22
1706.09
1365.34
1730.08
709.5
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

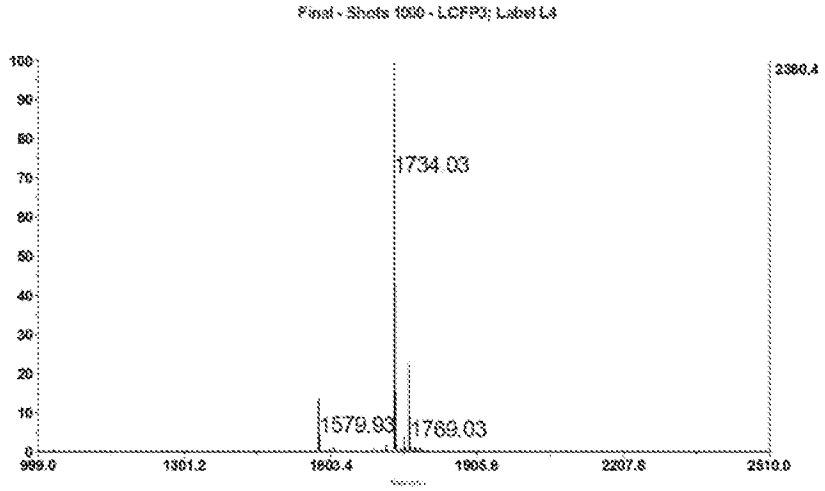

Final - Shots 1000 - LCFP3; Label J23
1738.10
1782.05
1734.1
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

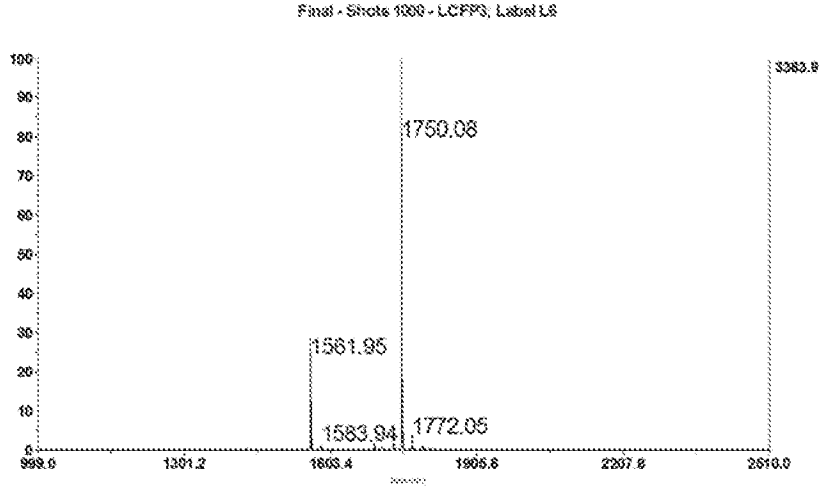

Final - Shots 1000 - LCFP3; Label K1
1724.03
1364.29
1568.93
1747.01
1342.4
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

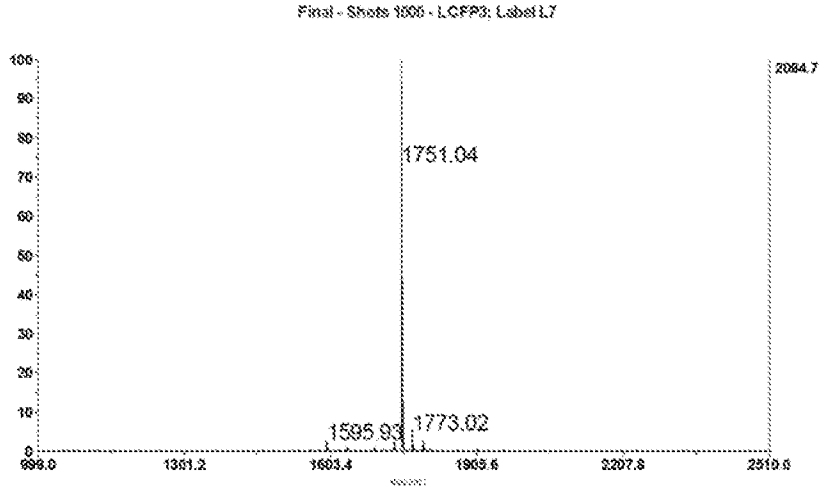

Final - Shots 1000 - LCFP3; Label L1
1800.09
1822.06
1643.96
2883.1
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

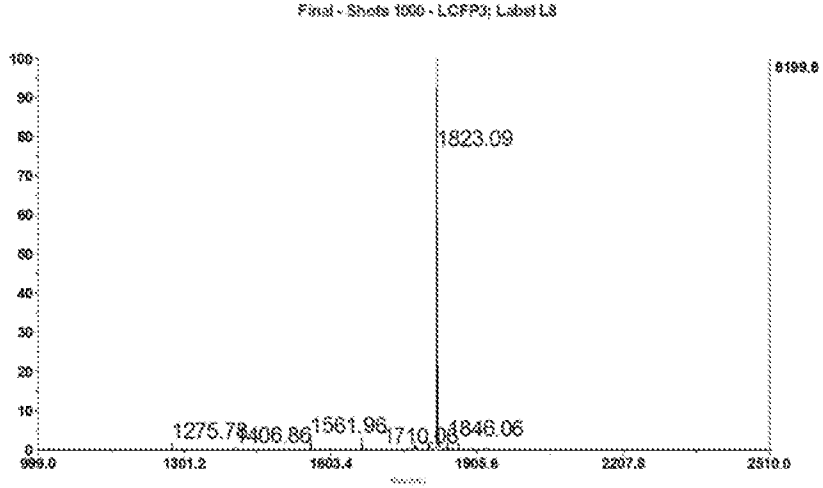

Final - Shots 1000 - LCFP3; Label L2
1739.99
1761.96
1561.94
1420.89
4243.1
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

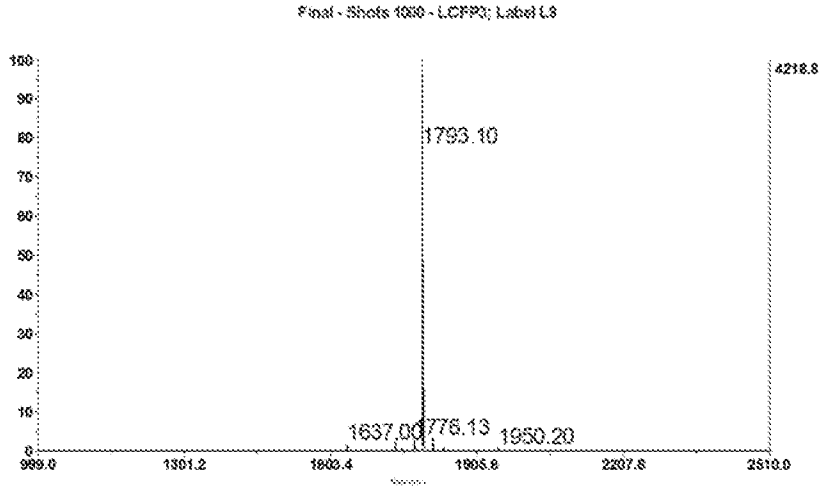

Final - Shots 1000 - LCFP3; Label F1
1364.35
1225.20
1267.25
1086.07
1404.38
1619.63
1758.78
2220.31
3492.5
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

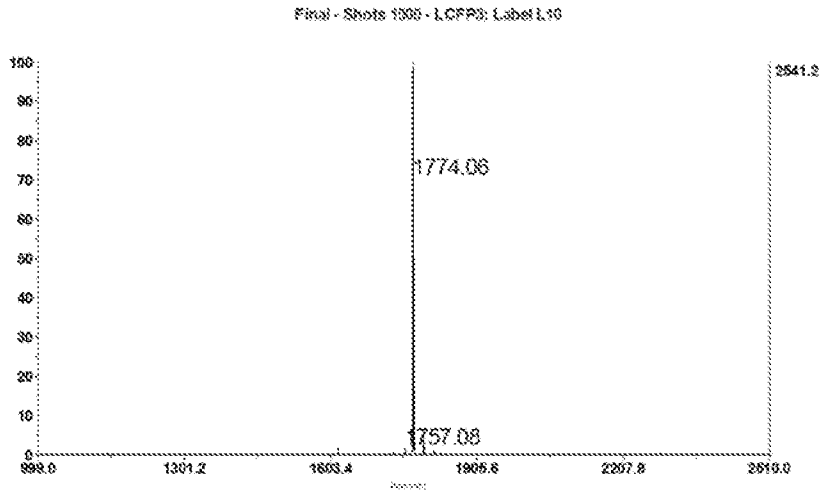

Final - Shots 1000 - LCFP3; Label L8
100
90
80
70
60
50
40
30
20
10
0
1734.03
1579.93
1769.03
2380.4
999.0
1301.2
1603.4
1905.8
2207.6
2510.0

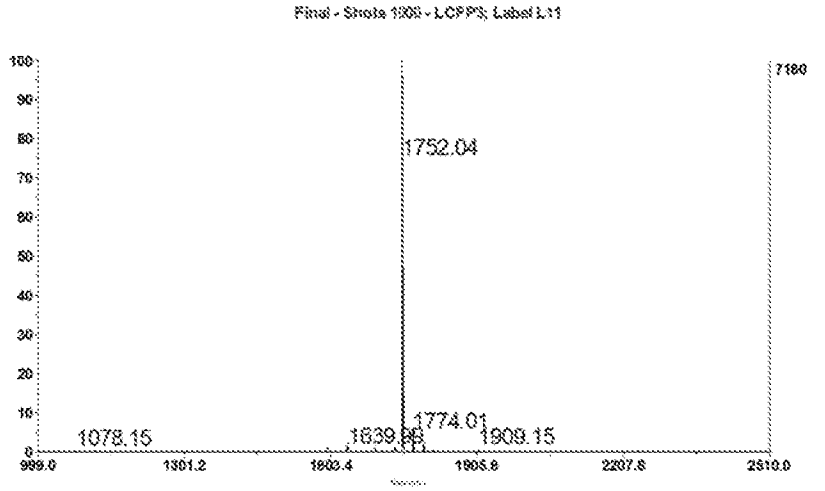

Final - Shots 1000 - LCFP3; Label L8
100
90
80
70
60
50
40
30
20
10
0
1750.08
1561.95
1583.94
1772.05
3383.8
999.0
1301.2
1603.4
1905.8
2207.6
2510.0

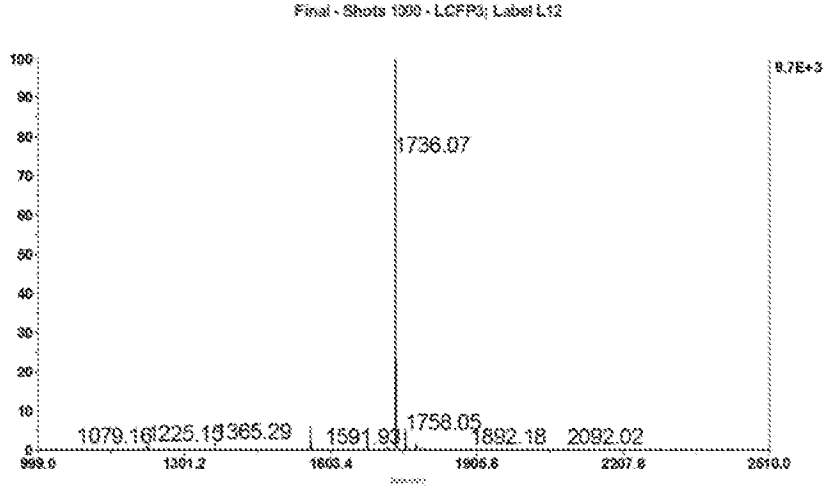

Final - Shots 1000 - LCFP3; Label L7
100
90
80
70
60
50
40
30
20
10
0
1751.04
1595.93
1773.02
2084.7
999.0
1301.2
1603.4
1905.8
2207.6
2510.0

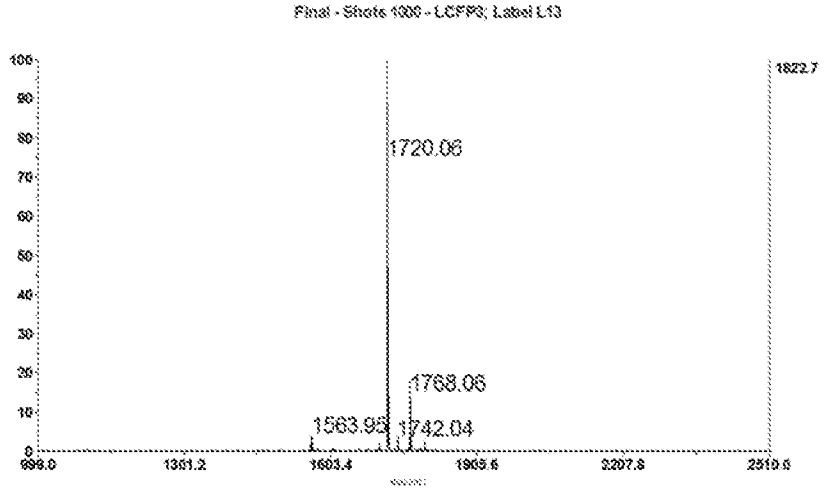

Final - Shots 1000 - LCFP3; Label L8
8199.8
1823.09
1275.78
1406.86
1561.96
1710.08
1846.06
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

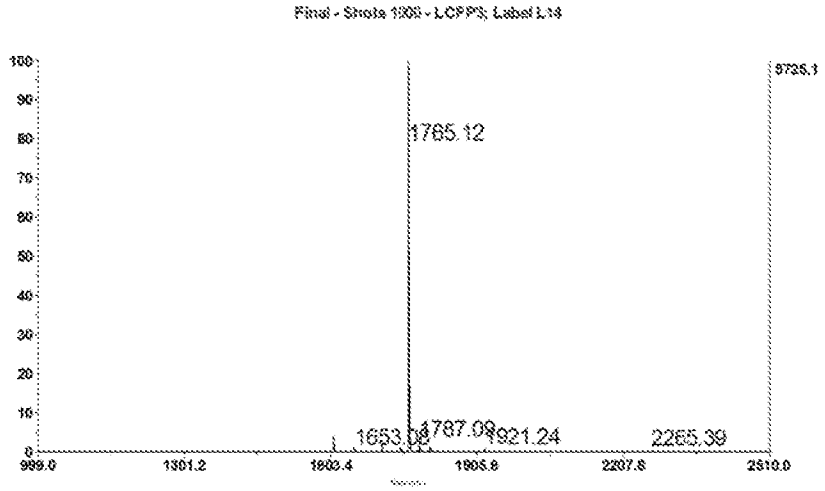

Final - Shots 1000 - LCFP3; Label L8
4218.8
1793.10
1637.00
1776.13
1950.20
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

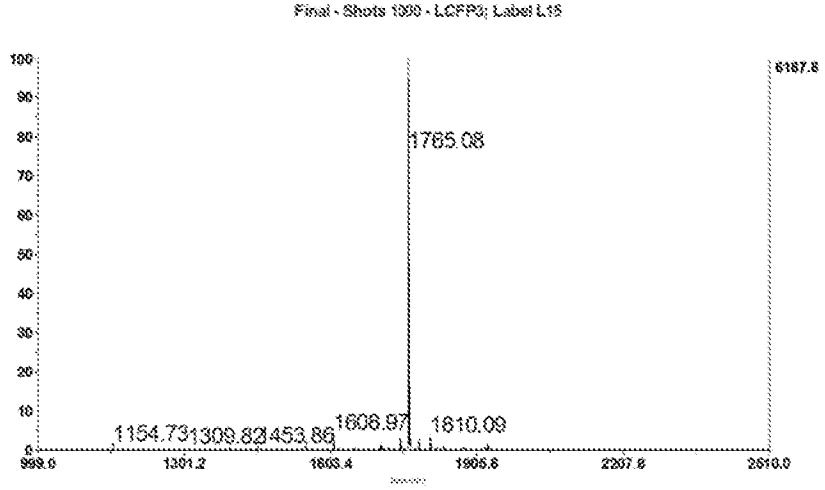

Final - Shots 1000 - LCFP3; Label L10
2541.2
1774.06
1757.08
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

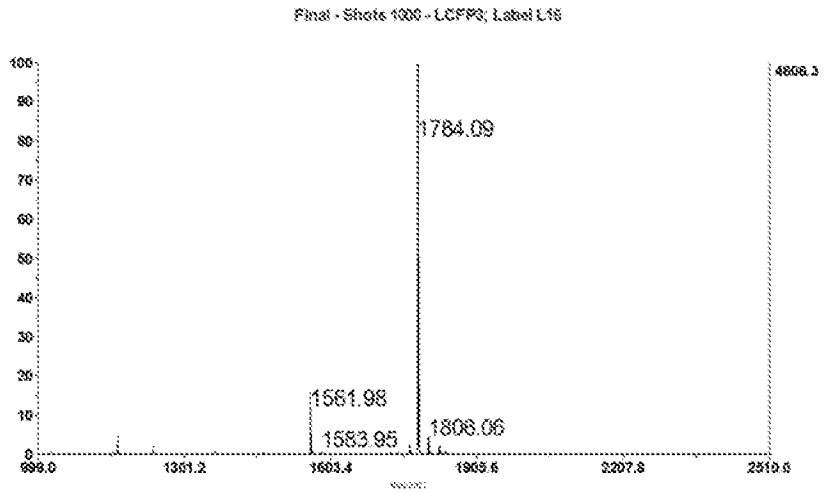

Final - Shots 1000 - LCFP3; Label L11
7180
1752.04
1774.01
1078.15
1639.88
1909.15
999.0
1301.2
1603.4
1905.6
2207.8
2510.0


Final - Shots 1000 - LCFP3; Label L12
9.7E+3
1736.07
1758.05
1079.16
1225.15
1365.29
1591.93
1892.18
2092.02
999.0
1301.2
1603.4
1905.6
2207.8
2510.0

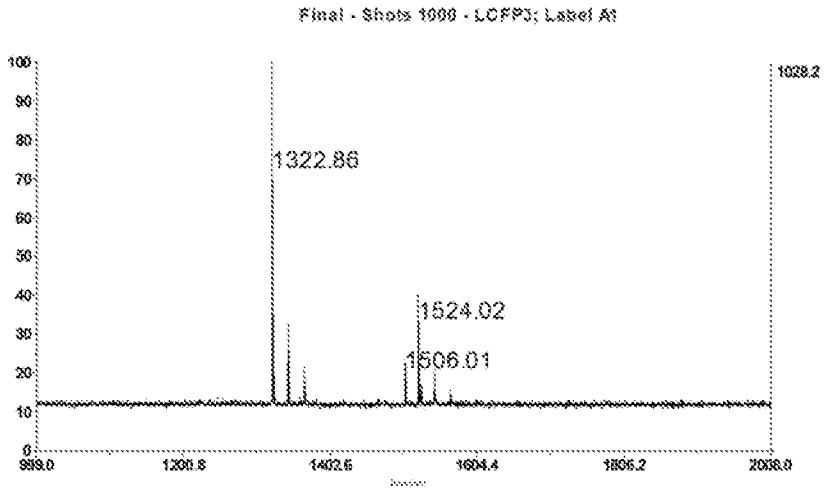

Final - Shots 1000 - LCFP3; Label L13
1822.7
1720.06
1768.06
1563.95
1742.04
999.0
1301.2
1603.4
1905.6
2207.8
2510.0


Final - Shots 1000 - LCFP3; Label L14
1765.12
1653.08
1787.09
1921.24
2265.39

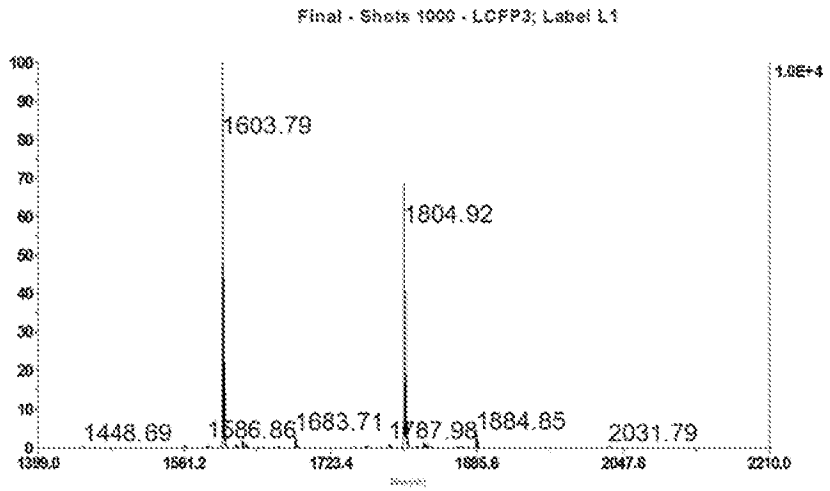

Final - Shots 1000 - LCFP3; Label L15
1765.08
1154.73
1309.82
1453.86
1608.97
1810.09

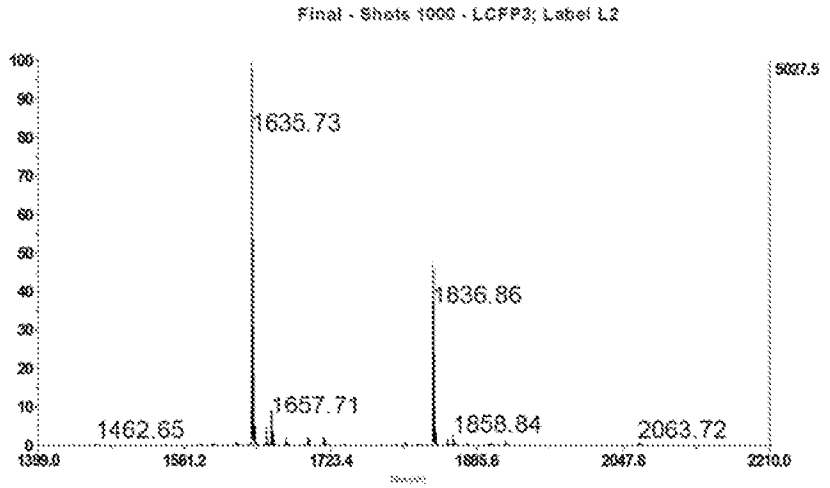

Final - Shots 1000 - LCFP3; Label L16
1784.09
1561.98
1583.95
1808.06

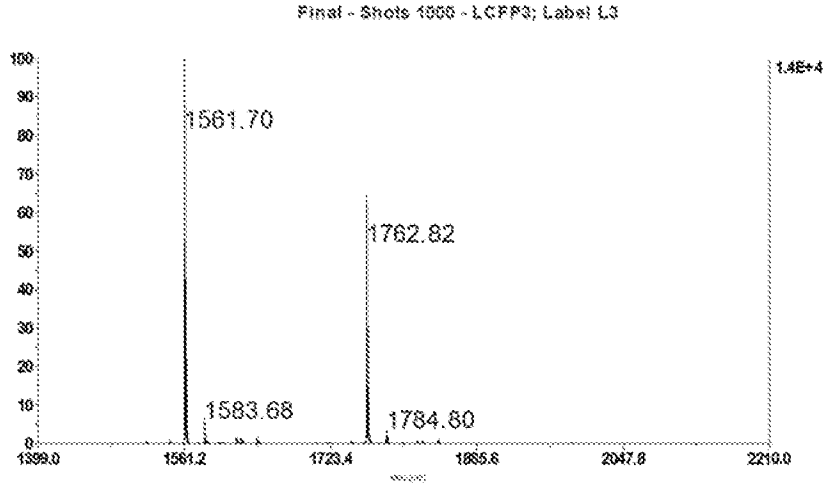
GL PVSTKPVATRDAL cyclization for 5 min, when replacing the Asn in NAL sequence with Asp in DAL sequence. The ligation is less efficient but still robust enough.

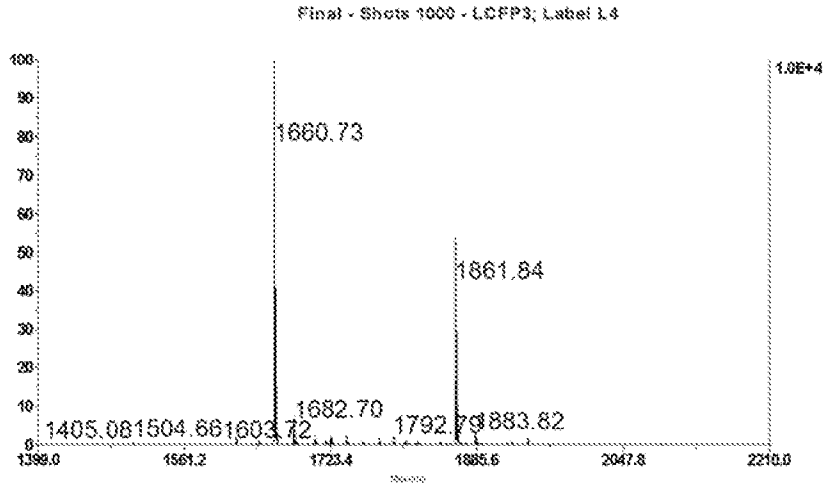


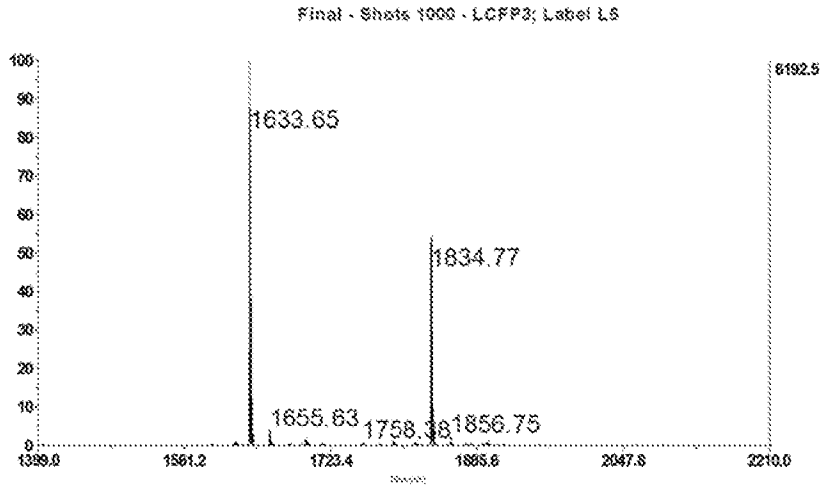

XLYRRGRLYRRNAL (L1-L20: VL, ML, GL, RL, QL, WL, EL, PL, DL, TL, SL, AL, IL, NL, KL, CL, YL, HL, FL, LL)

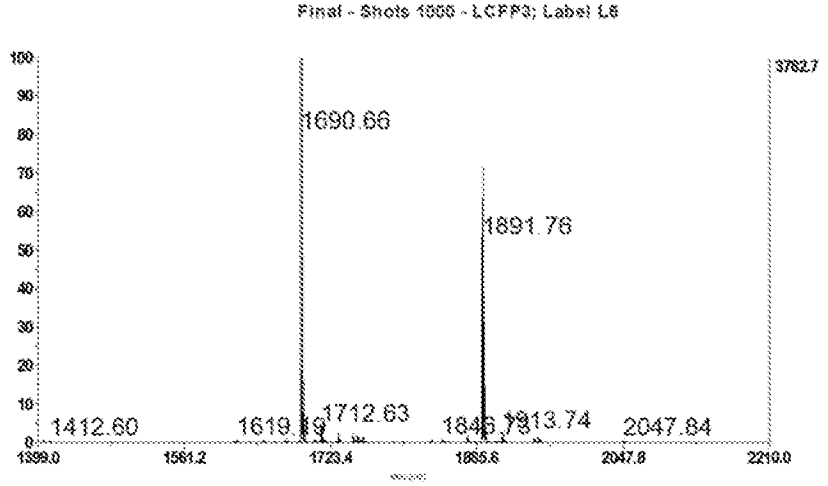


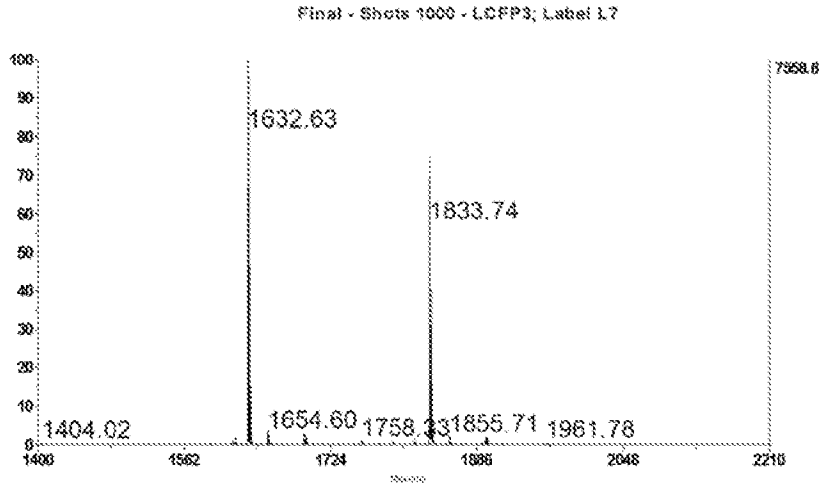

Final - Shots 1000 - LCFP3; Label L2
5027.5
1635.73
1636.86
1657.71
1462.65
1858.84
2063.72
1399.0
1561.2
1723.4
1885.6
2047.8
2210.0

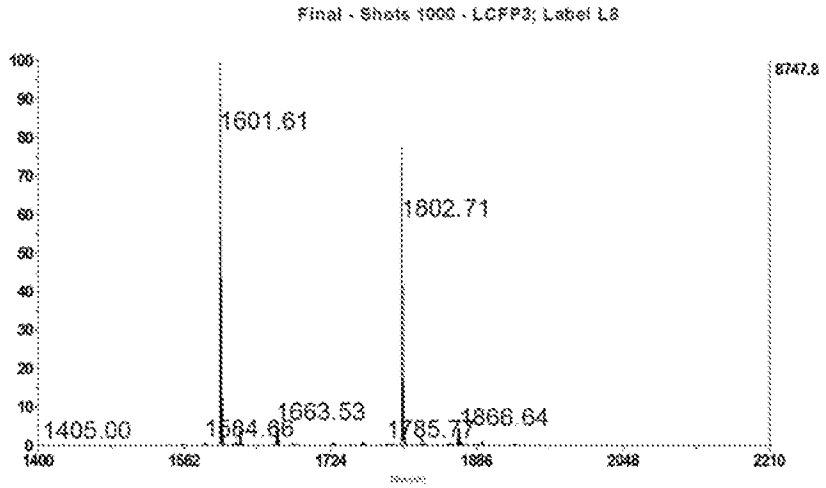

Final - Shots 1000 - LCFP3; Label L3
1.4E+4
1561.70
1762.82
1583.68
1784.80
1399.0
1561.2
1723.4
1885.6
2047.8
2210.0

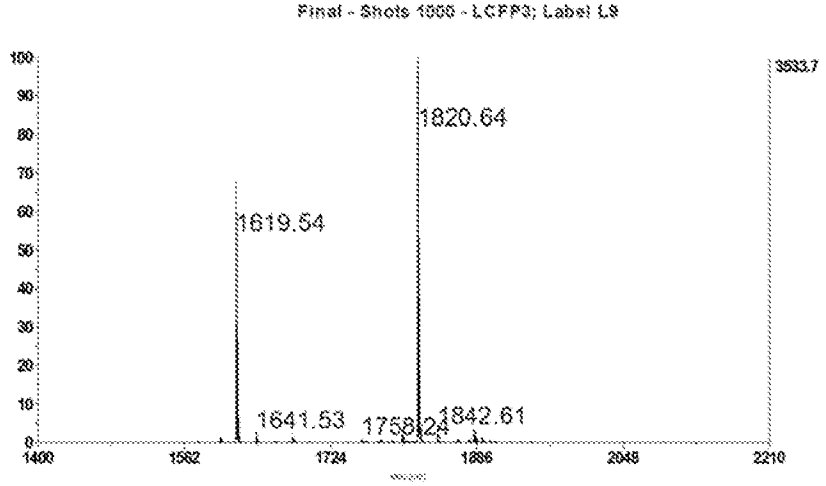

Final - Shots 1000 - LCFP3; Label L4
1.0E+4
1660.73
1861.84
1682.70
1405.08
1504.66
1603.72
1792.79
1883.82
1399.0
1561.2
1723.4
1885.6
2047.8
2210.0

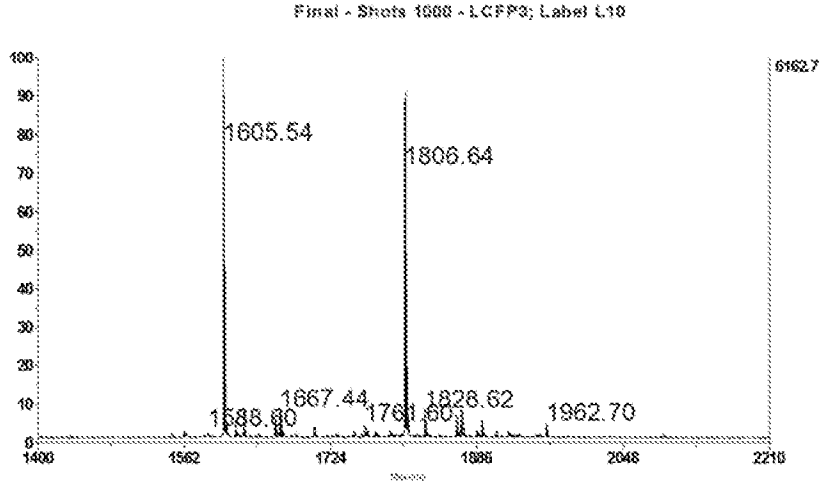

Final - Shots 1000 - LCFP3; Label L5
6192.5
1633.65
1834.77
1655.63
1758.38
1856.75
1399.0
1561.2
1723.4
1885.6
2047.8
2210.0

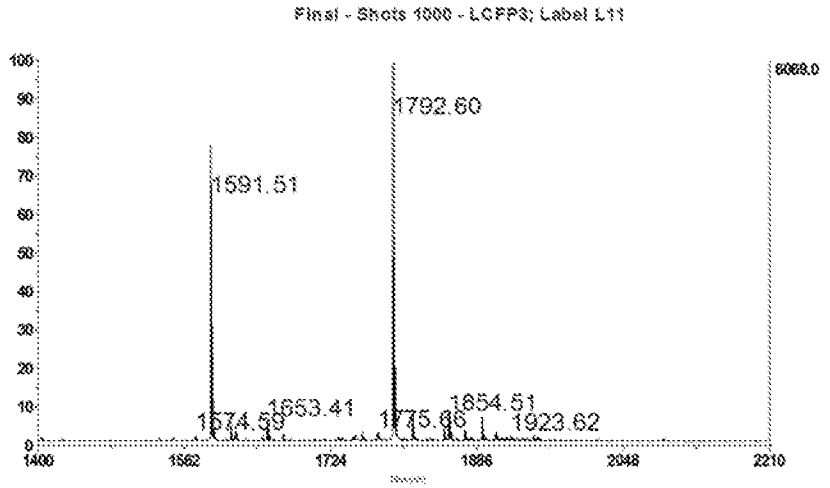

Final - Shots 1000 - LCFP3; Label L6
3762.7
1690.66
1891.76
1412.60
1619.69
1712.63
1846.78
1913.74
2047.84
1399.0
1561.2
1723.4
1885.6
2047.8
2210.0

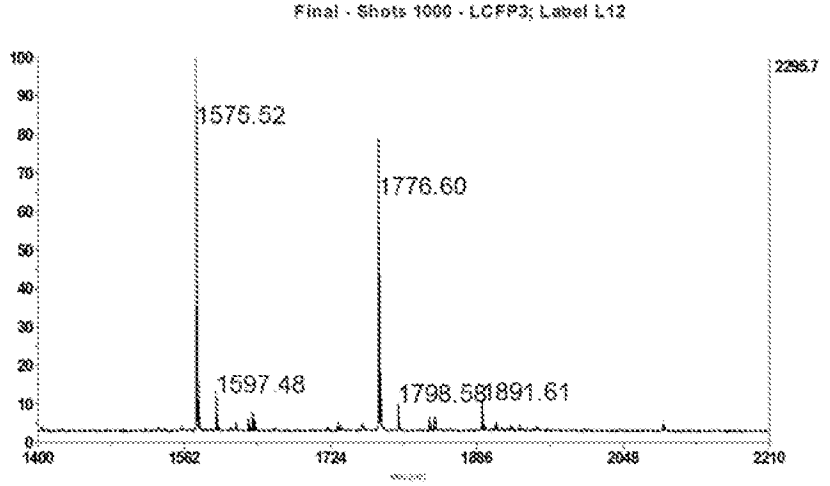

Final - Shots 1000 - LCFP3; Label L7
7358.6
1632.63
1833.74
1404.02
1654.60
1758.33
1856.71
1961.78
1400
1562
1724
1886
2048
2210

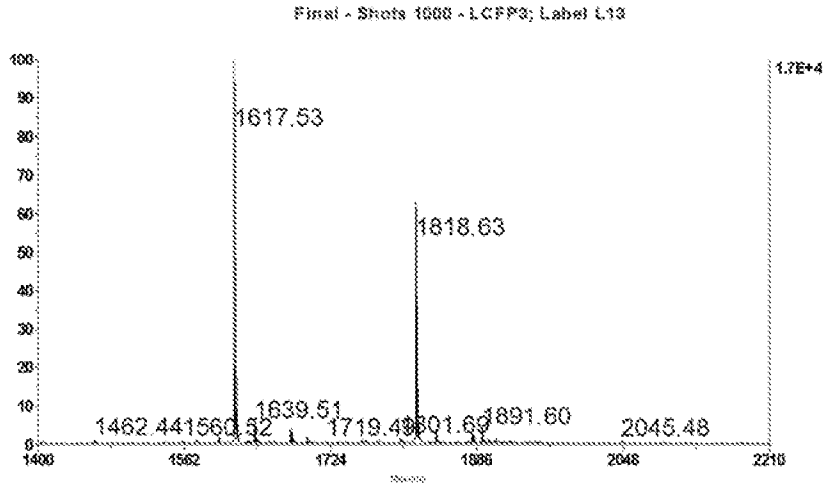

Final - Shots 1000 - LCFP3; Label L8
8747.8
1601.61
1802.71
1663.53
1866.64
1405.00
1400
1562
1724
1886
2048
2210

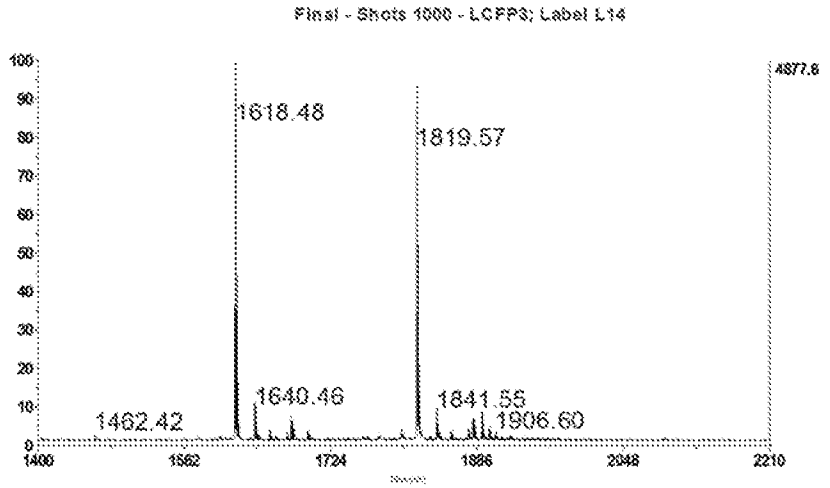

Final - Shots 1000 - LCFP3; Label L9
3533.7
1820.64
1619.54
1641.53
1842.61
1400
1562
1724
1886
2048
2210

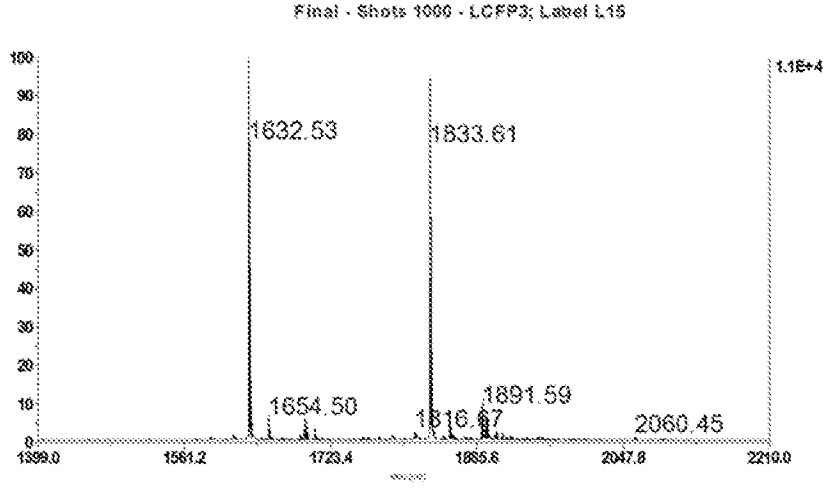

Final - Shots 1000 - LCFP3; Label L10
6162.7
1605.54
1806.64
1667.44
1828.62
1962.70
1400
1562
1724
1886
2048
2210

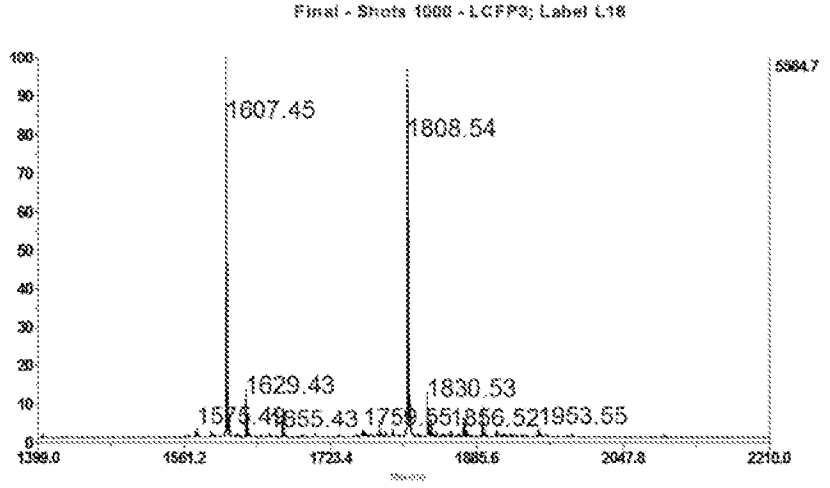

Final - Shots 1000 - LCFP3; Label L11
6069.0
1792.60
1591.51
1653.41
1854.51
1923.62
1400
1562
1724
1886
2048
2210

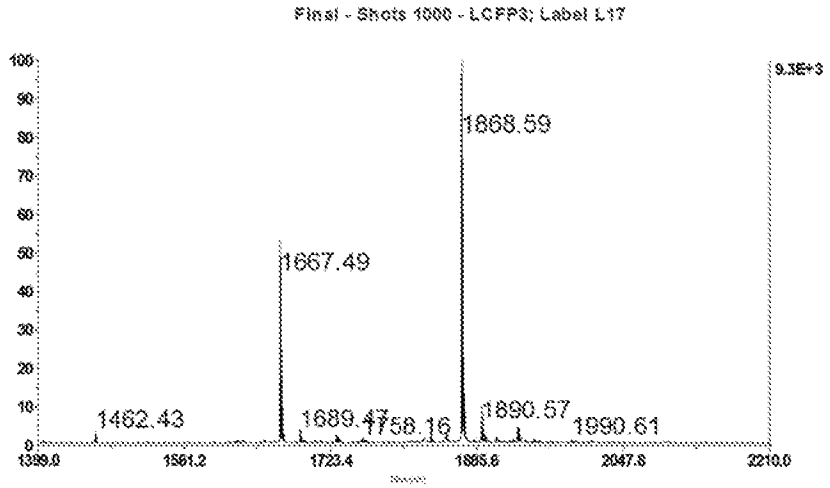

Final - Shots 1000 - LCFP3; Label L12
2295.7
1575.52
1776.60
1597.48
1798.58
1891.61
1400
1562
1724
1886
2048
2210

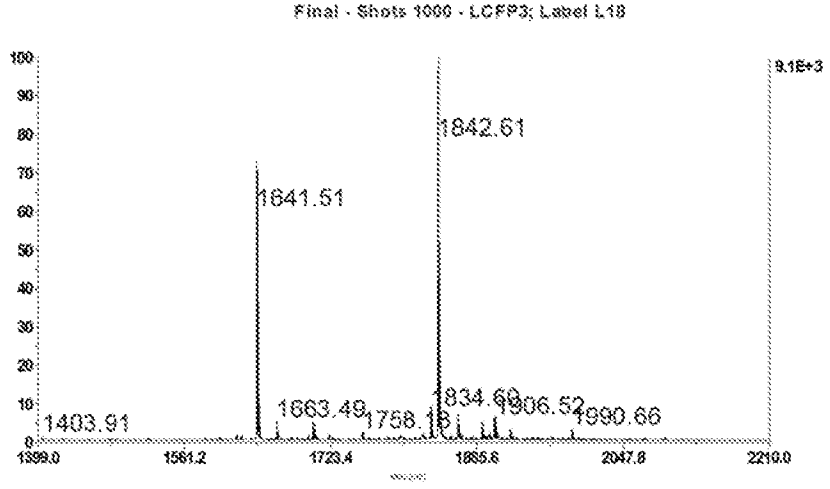

Final - Shots 1000 - LCFP3; Label L13
1.7E+4
1617.53
1818.63
1639.51
1462.44
1560.52
1719.43
1801.60
1891.60
2045.48
1400
1562
1724
1886
2048
2210

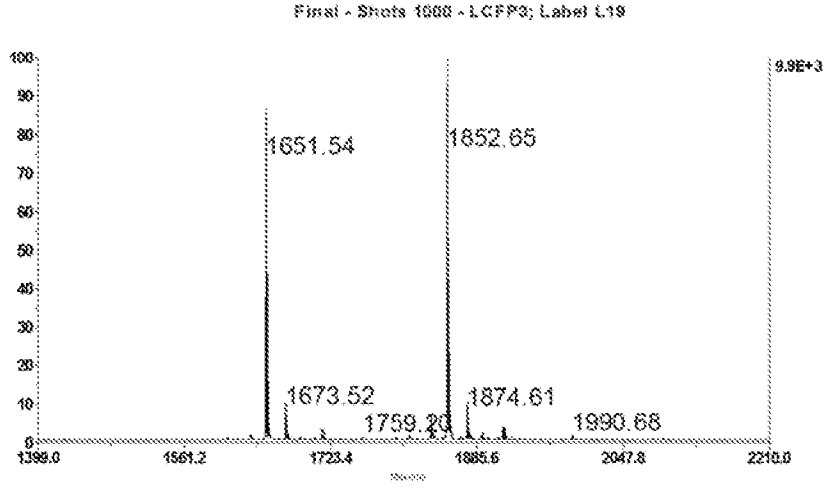

Final - Shots 1000 - LCFP3; Label L14
4877.8
1618.48
1819.57
1640.46
1841.55
1462.42
1906.60
1400
1562
1724
1886
2048
2210


Final - Shots 1000 - LCFP3; Label L15
1.1E+4
1632.53
1833.61
1654.50
1891.59
2060.45
1399.0
1561.2
1723.4
1885.6
2047.8
2210.0


Final - Shots 1000 - LCFP3; Label L16
5564.7
1607.45
1808.54
1629.43
1830.53
1953.55
1399.0
1561.2
1723.4
1885.6
2047.8
2210.0


Final - Shots 1000 - LCFP3; Label L17
9.3E+3
1868.59
1667.49
1462.43
1689.47
1758.16
1890.57
1990.61
1399.0
1561.2
1723.4
1885.6
2047.8
2210.0


Final - Shots 1000 - LCFP3; Label L18
9.1E+3
1842.61
1641.51
1403.91
1663.49
1758.18
1834.60
1906.52
1990.66
1399.0
1561.2
1723.4
1885.6
2047.8
2210.0


Final - Shots 1000 - LCFP3; Label L19
9.9E+3
1651.54
1852.65
1673.52
1759.20
1874.61
1990.68
1399.0
1561.2
1723.4
1885.6
2047.8
2210.0

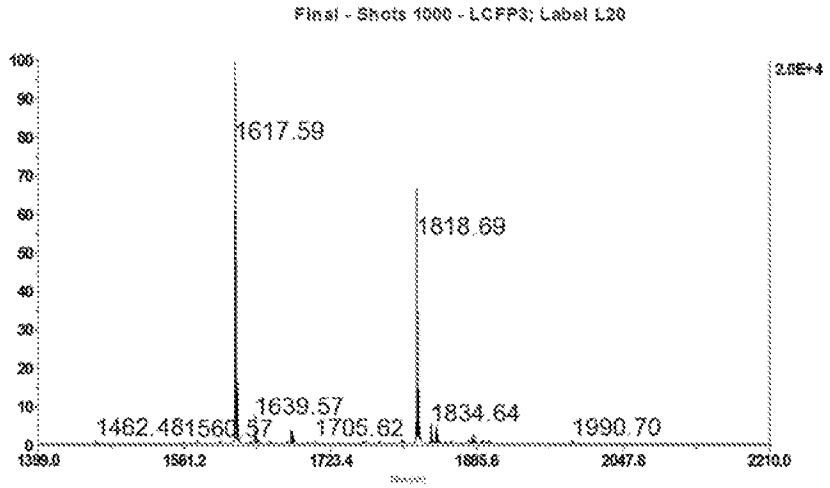

Final - Shots 1000 - LCFP3; Label L20
100
90
80
70
60
50
40
30
20
10
0
1617.59
1818.69
1639.57
1462.481560
57
1705.62
1834.64
1990.70
1399.0
1561.2
1723.4
1885.6
2047.8
2210.0

Figure 12a

**a**

**BLAST Results**

**Blast 2 sequences**

**Job title: Protein Sequence (474 letters)**

**RID** H8CRXSS011N (Expires on 05-19 13:30 pm)

| | | | |
|---|---|---|---|
| **Query ID** | lcl\|Query_162141 | **Subject ID** | 7 subjects |
| **Description** | None | **Description** | See details |
| **Molecule type** | amino acid | **Molecule type** | amino acid |
| **Query Length** | 474 | **Subject Length** | 3432 |
| | | **Program** | BLASTP 2.6.1+ |

New Analyze your query with SmartBLAST

Graphic Summary

**Distribution of the top 7 Blast Hits on 7 subject sequences**

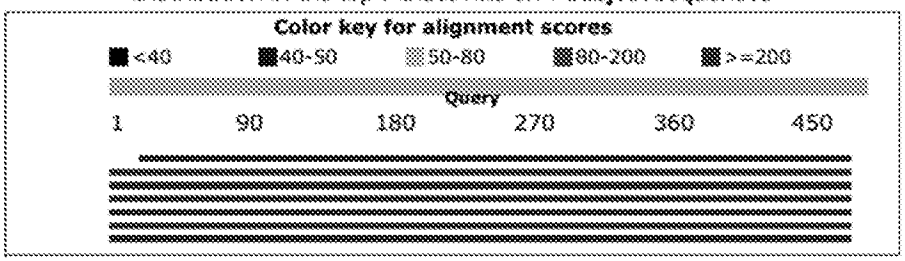



Figure 12b

**b**

Sequences producing significant alignments:

| Description | Max score | Total score | Query cover | E value | Ident | Accession |
|---|---|---|---|---|---|---|
| unnamed protein product [Coffea canephora] | 730 | 730 | 85% | 0.0 | 76% | CDP06321.1 |
| PREDICTED: vacuolar-processing enzyme-like [Daucus carota subsp. sativus] | 708 | 708 | 100% | 0.0 | 76% | XP_017228852.1 |
| Peptidase C13, legumain [Corchorus capsularis] | 707 | 707 | 100% | 0.0 | 76% | OMO69896.1 |
| Peptidase C13, legumain [Corchorus olitorius] | 702 | 702 | 100% | 0.0 | 69% | OMO96618.1 |
| vacuolar-processing enzyme [Jatropha curcas] | 701 | 701 | 100% | 0.0 | 68% | XP_012087228.1 |
| vacuolar processing enzyme [Malus hupehensis] | 699 | 699 | 100% | 0.0 | 69% | AGC94758.1 |
| PREDICTED: vacuolar processing enzyme-like [Pyrus x bretschneideri] | 698 | 698 | 100% | 0.0 | 69% | XP_009361866.1 |

Figure 13

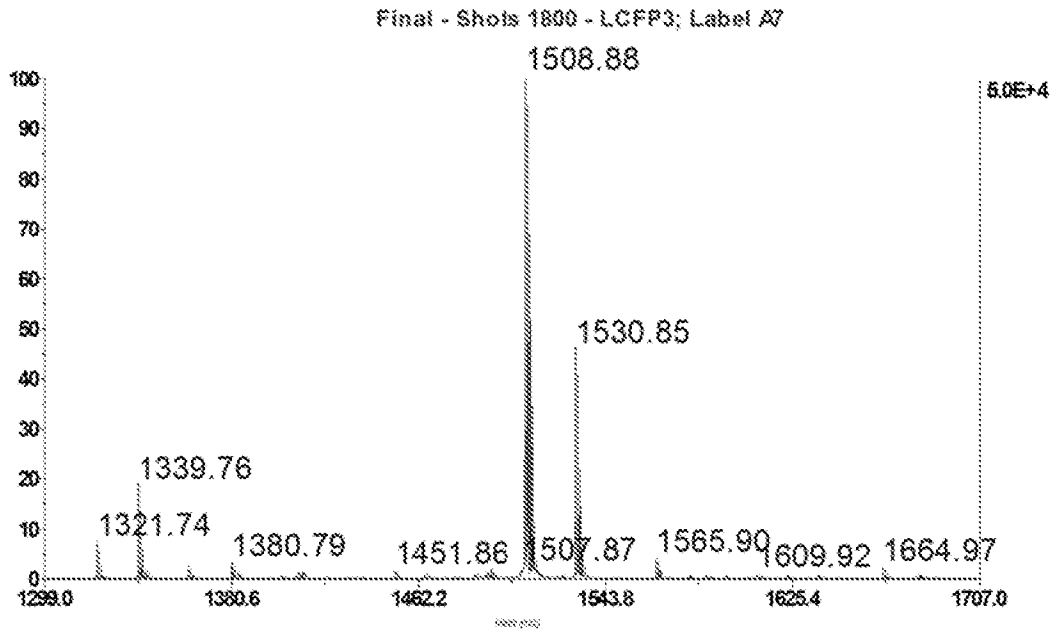
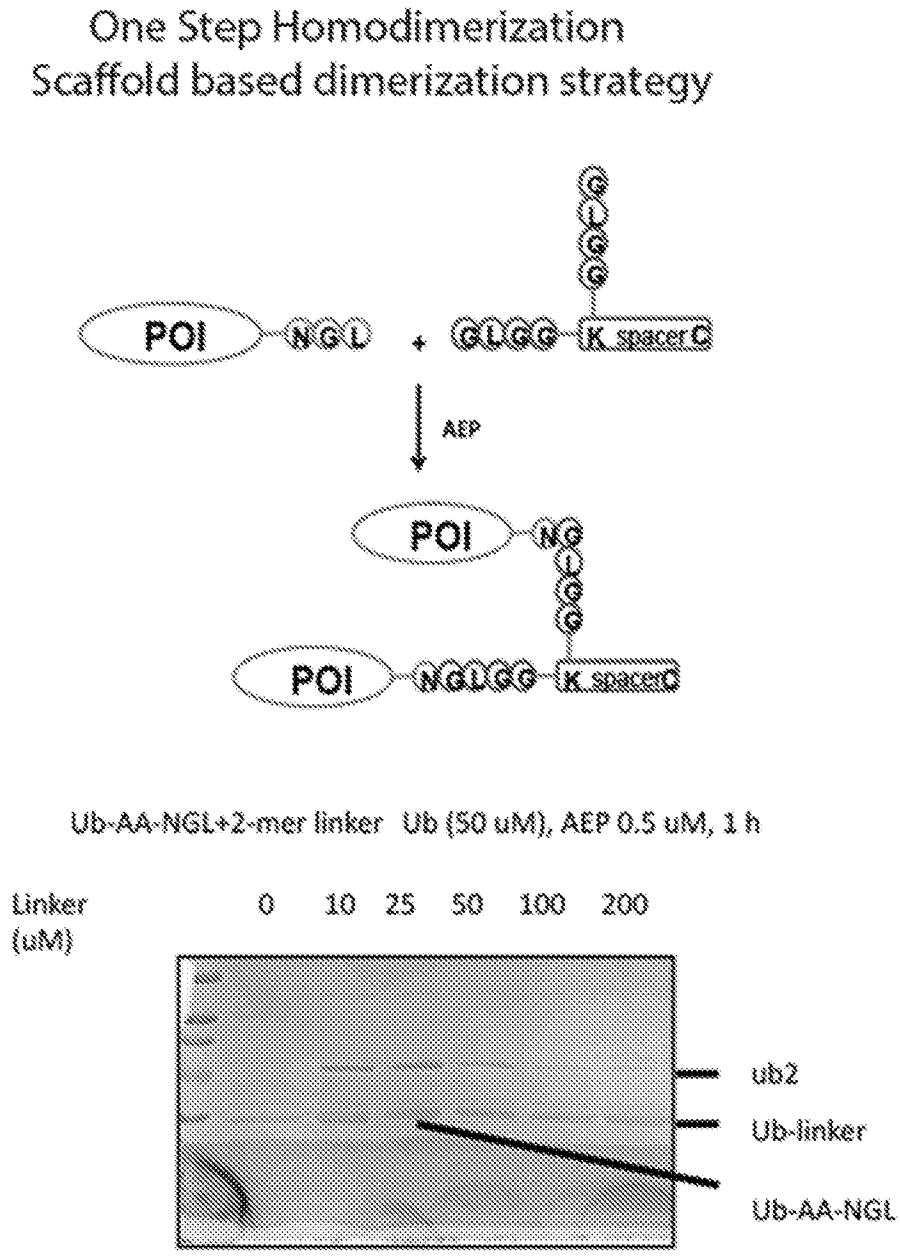
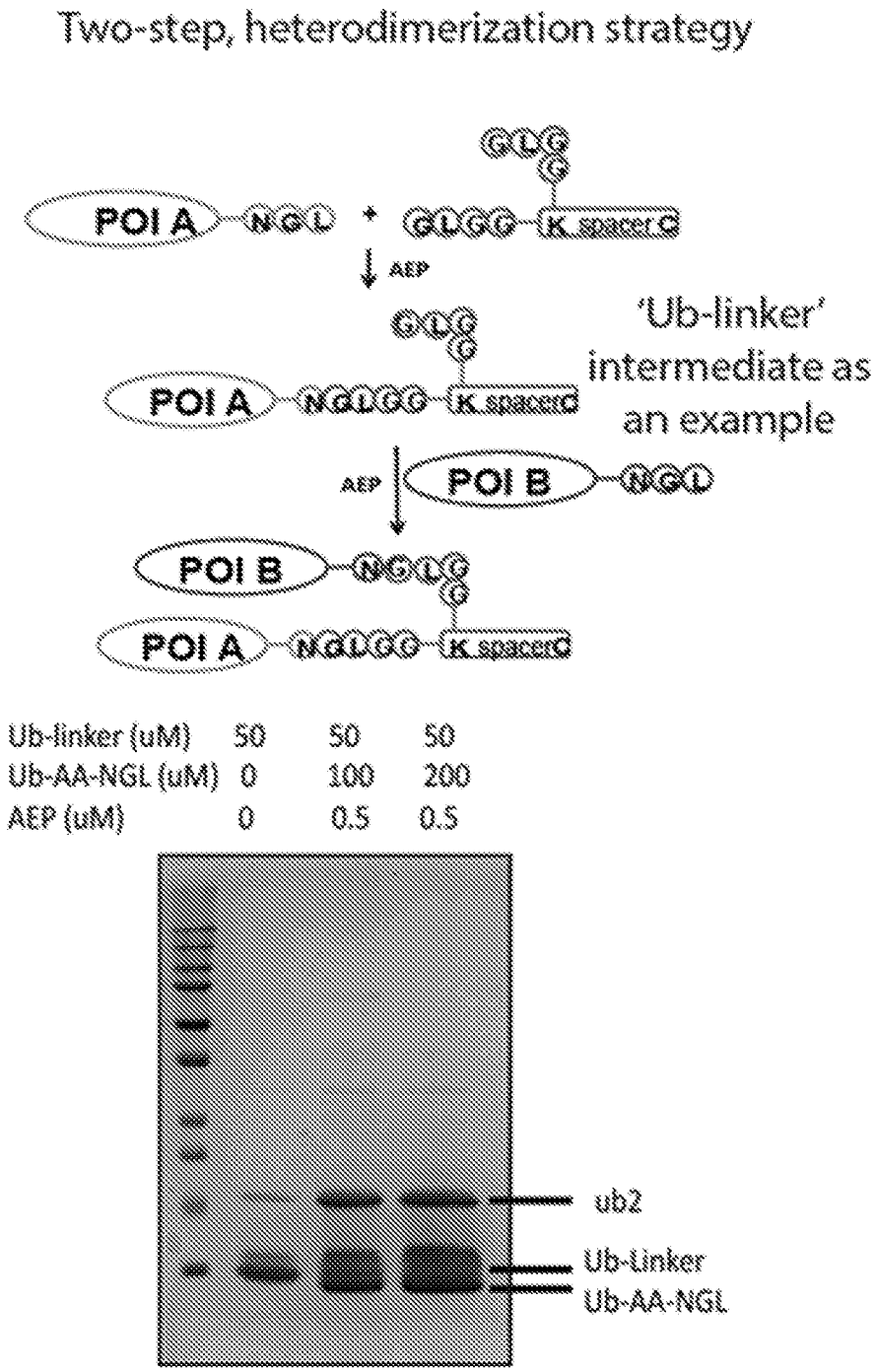

Final - Shots 1800 - LCFP3; Label A7
1508.88
5.0E+4
100
90
80
70
60
50
40
30
20
10
0
1530.85
1339.76
1321.74
1380.79
1451.86
1507.87
1565.90
1609.92
1664.97
1299.0
1380.6
1462.2
1543.8
1625.4
1707.0

Figure 14

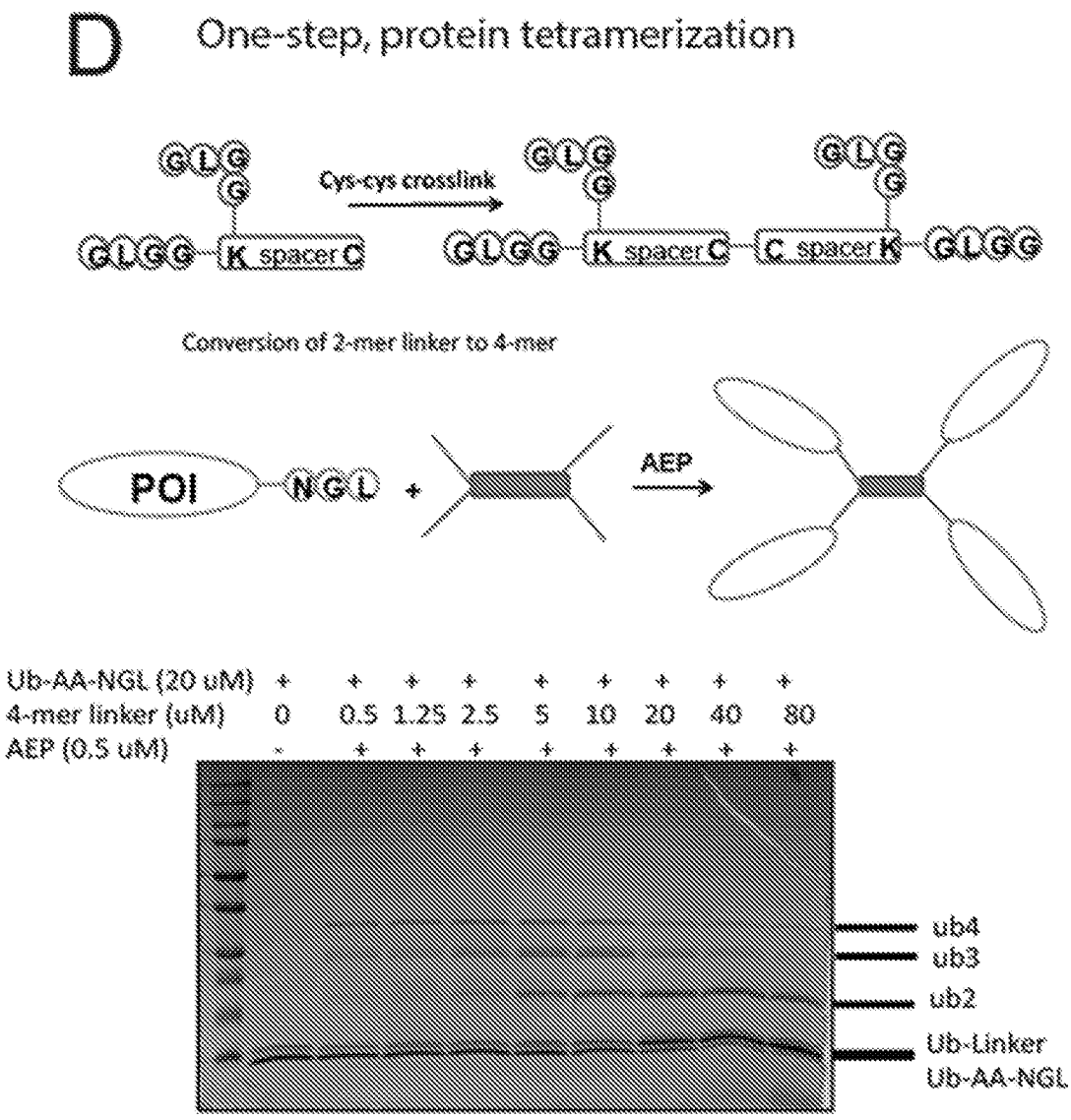
**a**


One Step Homodimerization
Scaffold based dimerization strategy
POI
NGL
GLGG
K spacer C
AEP
POI
NG
GLGG
POI
NGLGG
K spacer C
Ub-AA-NGL+2-mer linker Ub (50 uM), AEP 0.5 uM, 1 h
Linker (uM)
0 10 25 50 100 200
ub2
Ub-linker
Ub-AA-NGL

**b**

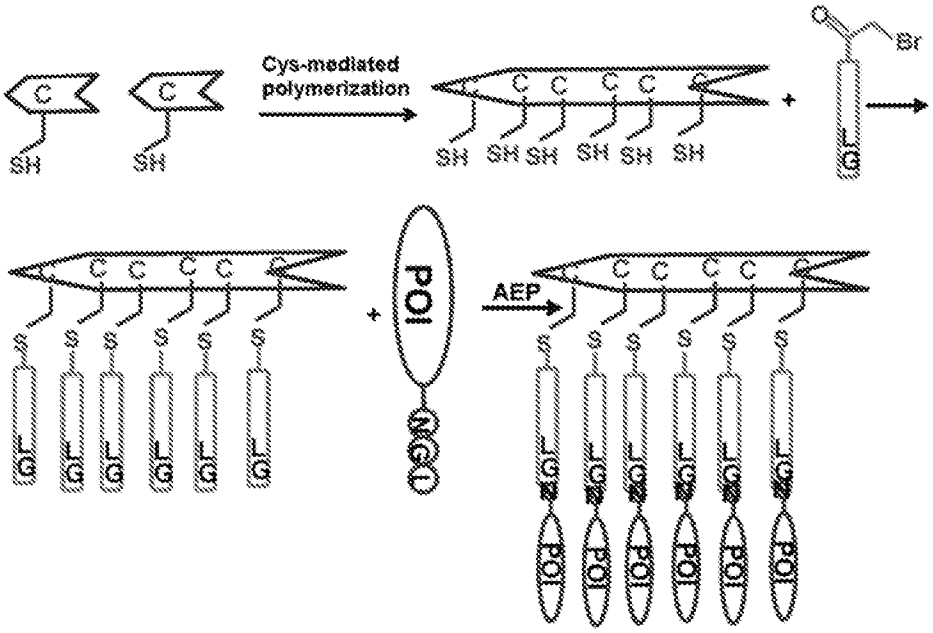

Two-step, heterodimerization strategy
POI A
NGL
GLGG
K spacer C
AEP
'Ub-linker' intermediate as an example
POI B
Ub-linker (uM) 50 50 50
Ub-AA-NGL (uM) 0 100 200
AEP (uM) 0 0.5 0.5
ub2
Ub-Linker
Ub-AA-NGL

C

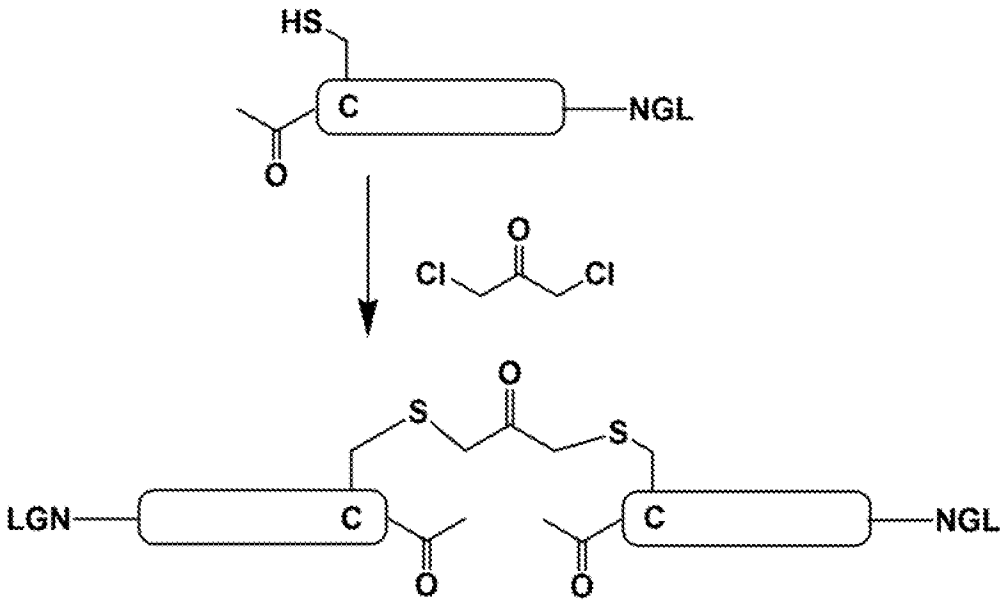

D
One-step, protein tetramerization
GLG
G
Cys-cys crosslink
GLGG
K spacer C
C spacer K
Conversion of 2-mer linker to 4-mer
POI
NGL
AEP
Ub-AA-NGL (20 uM)
4-mer linker (uM)
AEP (0.5 uM)
0
0.5
1.25
2.5
5
10
20
40
80
ub4
ub3
ub2
Ub-Linker
Ub-AA-NGL

**d**

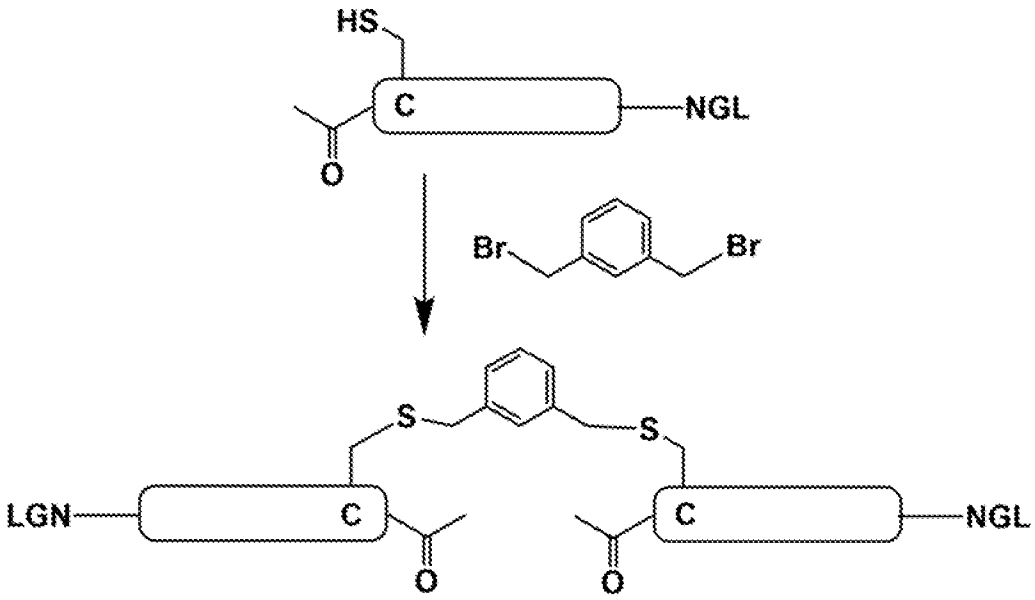
Schematic Illustration of protein ultra-oligomer strategy



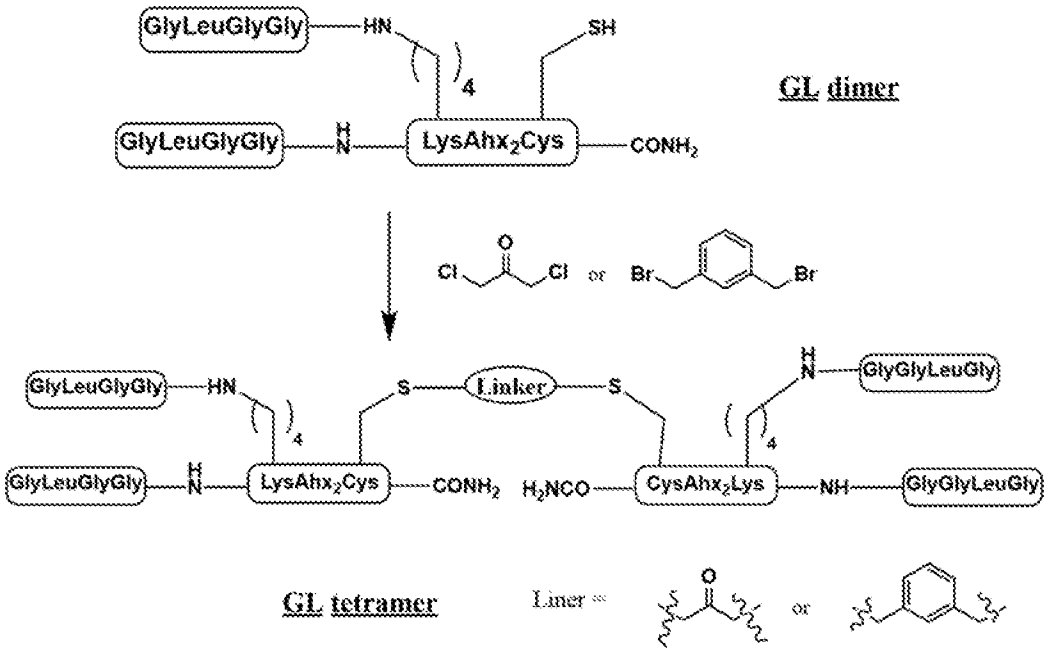

**e**



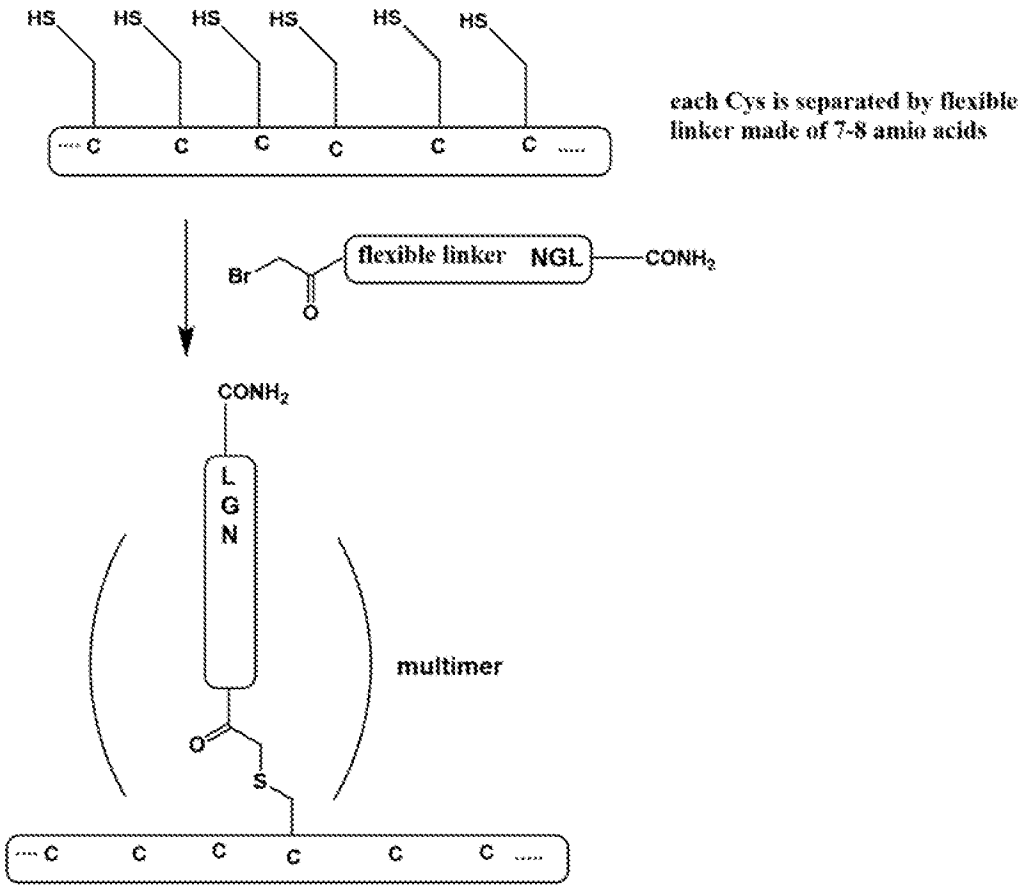
f


HS
C
NGL
O
Br
Br
S
S
LGN
C
C
NGL
O
O

g


GlyLeuGlyGly
HN
4
SH
GlyLeuGlyGly
H N
LysAhx2Cys
CONH2
GL dimer
Cl
O
Cl
or
Br
Br
GlyLeuGlyGly
HN
4
S
Linker
S
H N
GlyGlyLeuGly
4
GlyLeuGlyGly
H N
LysAhx2Cys
CONH2
H2NCO
CysAhx2Lys
NH
GlyGlyLeuGly
GL tetramer
Liner =
O
or

h


HS
HS
HS
HS
HS
HS
C
C
C
C
C
C
each Cys is separated by flexible linker made of 7-8 amio acids
Br
O
flexible linker
NGL
CONH2
CONH2
L
G
N
multimer
O
S
C
C
C
C
C
C

i

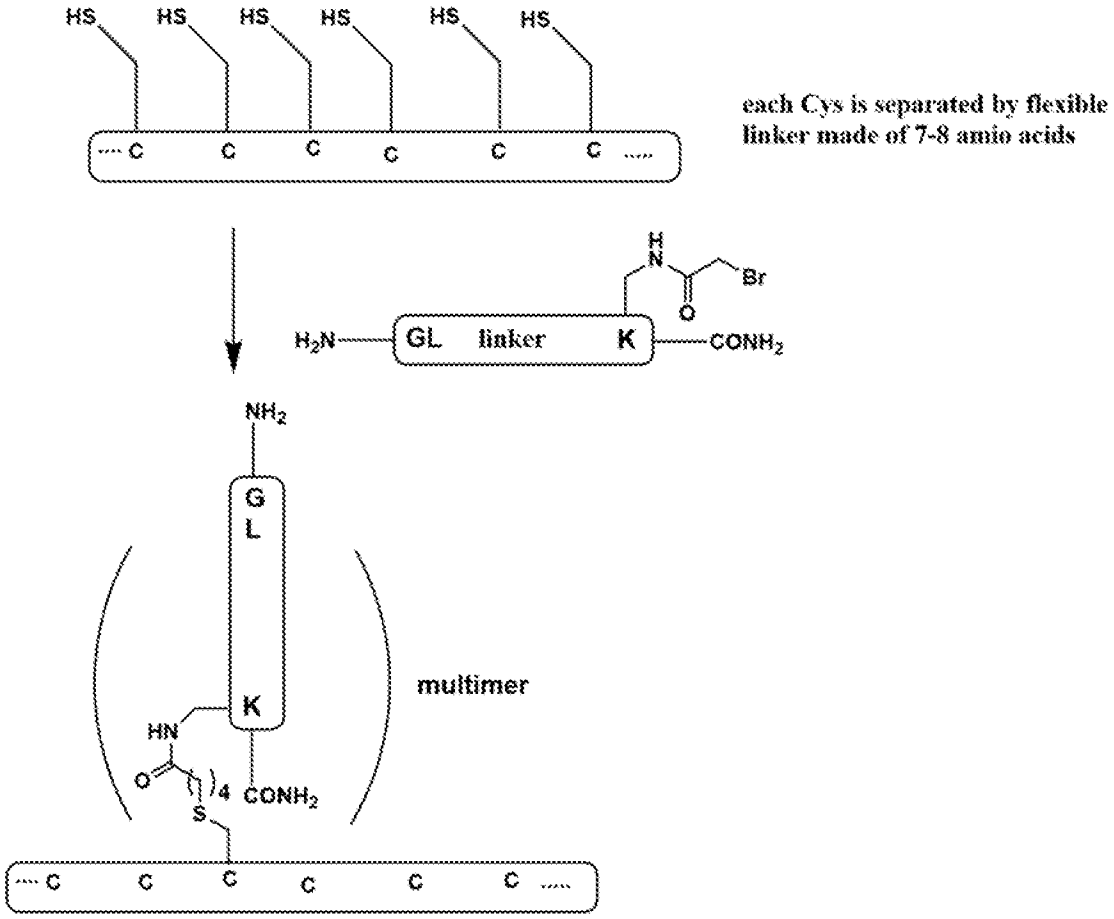
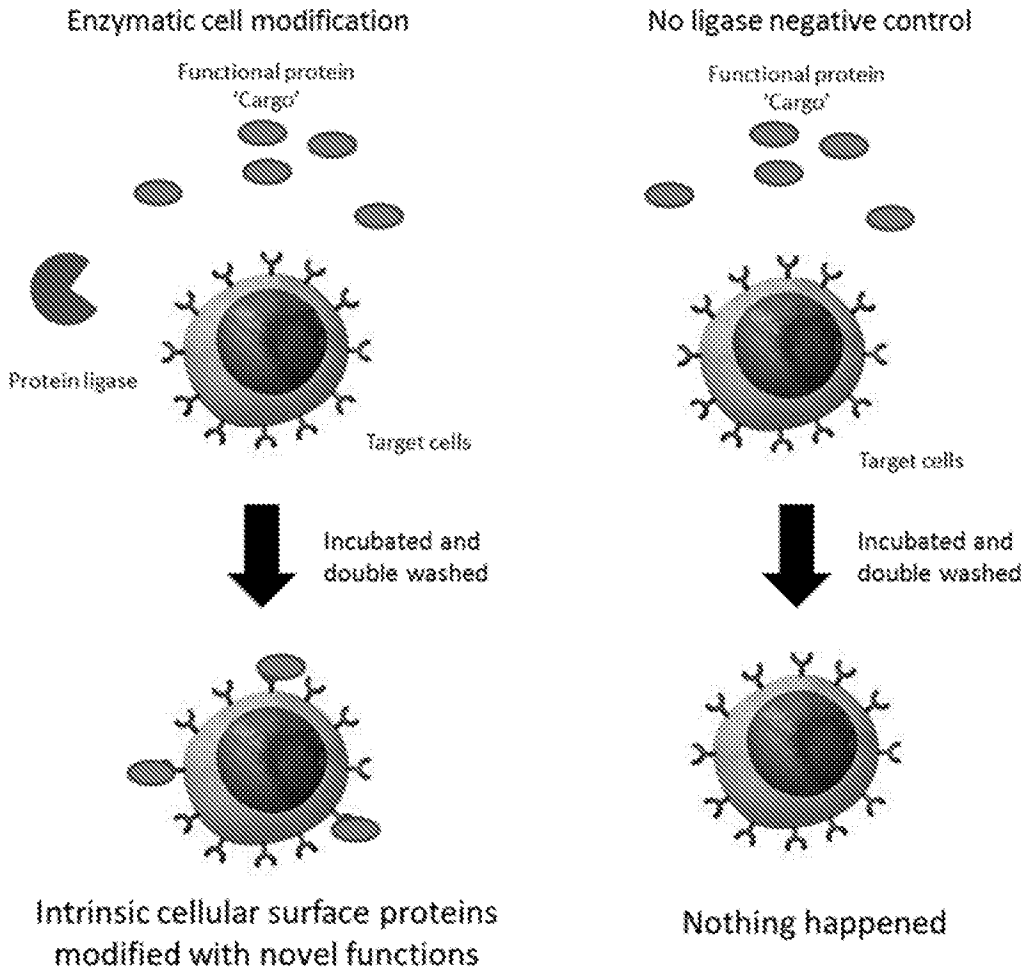
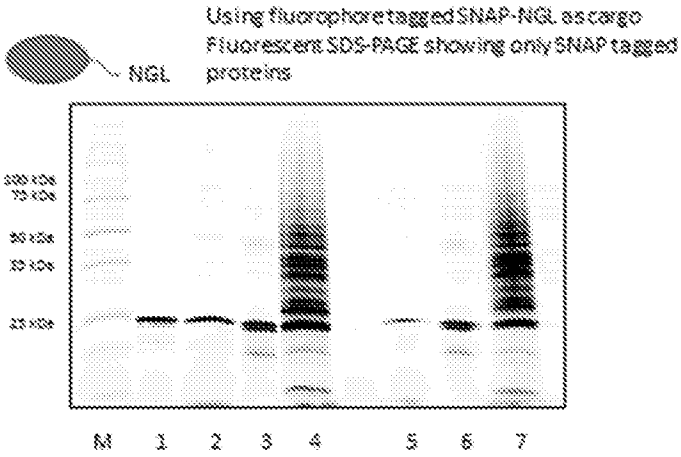

HS
HS
HS
HS
HS
HS
C
C
C
C
C
C
each Cys is separated by flexible linker made of 7-8 amio acids
H
N
Br
O
$H_2N$
GL
linker
K
$CONH_2$
$NH_2$
G
L
K
HN
O
4
$CONH_2$
S
multimer
C
C
C
C
C
C

Figure 15

**a**

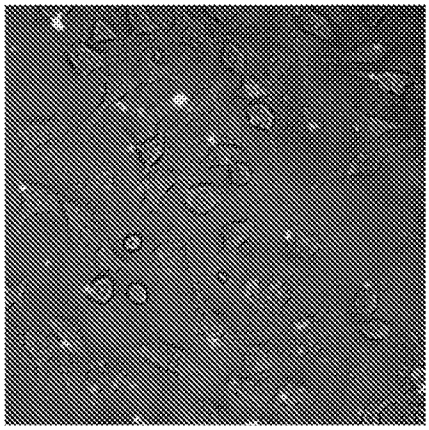

Enzymatic cell modification
Functional protein 'Cargo'
Protein ligase
Target cells
Incubated and double washed
Intrinsic cellular surface proteins modified with novel functions
No ligase negative control
Functional protein 'Cargo'
Target cells
Incubated and double washed
Nothing happened

**b**




M = Molecular weight
1 = SNAP-NGL protein control
2 = SNAP-NGL added to RAW cells, no ligase, cells after wash
3 = SNAP-NGL added to RAW cells, with ligase, wash solution
4 = SNAP-NGL added to RAW cells, with ligase, cells after wash
5 = same as 2, using HEK cells
6 = same as 3, using HEK cells
7 = same as 4, using HEK cells

**c**



# INTERNATIONAL SEARCH REPORT

International application No.
**PCT/SG2017/050458**

## A. CLASSIFICATION OF SUBJECT MATTER

***C07K 14/415 (2006.01)*** ***C12N 9/00 (2006.01)*** ***C12P 21/04 (2006.01)***

According to International Patent Classification (IPC)

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

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)

FAMPAT/CAPLUS/BIOSIS/EMBASE/MEDLINE: Asparagine/aspartate (Asx) peptide ligase, asparaginyl endopeptidase, ligation and similar terms

REGISTRY: SEQ ID NO: 1

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X<br>A | HARRIS K.S. ET AL., Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat Commun*, 18 December 2015, Vol. 6, pages 10199: 1-10<br>[Retrieved on 2017-11-15] <DOI: 10.1038/NCOMMS10199><br>Results, in particular Fig. 3 | 1-6, 9-17, 20-24 AND 27-32<br>7, 8, 18, 19, 25, 26 AND 33-36 |
| X<br>A | BERNATH-LEVIN K. ET AL., Peptide Macrocyclization by a Bifunctional Endoprotease. *Chem Biol*, 7 May 2015, Vol. 22, No. 5, pages 571-582<br>[Retrieved on 2017-11-15] <DOI: 10.1016/J.CHEMBIOL.2015.04.010><br>Results, in particular Fig. 1A, Fig. 6A & 6B | 1-6, 9-17, 20-24 AND 27-32<br>7, 8, 18, 19, 25, 26 AND 33-36 |
| X | SASKA I. ET AL., An Asparaginyl Endopeptidase Mediates *in Vivo* Protein Backbone Cyclization. *J Biol Chem*, 13 August 2007, Vol. 282, No. 40, pages 29721-29728<br>[Retrieved on 2017-11-15] <DOI: 10.1074/JBC.M705185200><br>Fig. 3 | 32 |

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

*Special categories of cited documents:

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

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

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

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

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

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

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

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

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 15/11/2017 *(day/month/year)* | 23/11/2017 *(day/month/year)* |


Name and mailing address of the ISA/SG

IPOS INTELLECTUAL PROPERTY OFFICE OF SINGAPORE

**Intellectual Property Office of Singapore**
51 Bras Basah Road
#01-01 Manulife Centre
Singapore 189554

Email: pct@ipos.gov.sg

Authorized officer

Sung Ying Ying (Dr)

IPOS Customer Service Tel. No.: (+65) 6339 8616

# INTERNATIONAL SEARCH REPORT

International application No.
**PCT/SG2017/050458**

**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 2015/163818 A1 (NANYANG TECHNOLOGICAL UNIVERSITY) 29 October 2015<br>Fig. 2 | 32 |
| X | NGUYEN G.K.T. ET AL., Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat Chem Biol*, 20 July 2014, Vol. 10, No. 9, pages 732-738<br>[Retrieved on 2017-11-15] <DOI: 10.1038/NCHEMBIO.1586><br>Fig. 1 | 32 |
| P,X | YANG R. ET AL., Engineering a Catalytically Efficient Recombinant Protein Ligase. *J Am Chem Soc*, 15 February 2017, Vol. 139, No. 15, pages 5351-5358<br>[Retrieved on 2017-11-15] <DOI: 10.1021/JACS.6B12637><br>Whole document | 1-36 |
| P,X | WO 2017/054044 A1 (HEXIMA LIMITED AND THE UNIVERSITY OF QUEENSLAND) 6 April 2017 | 1-6, 9-17, 20-24 AND 27-32 |
| P,A | Whole document | 7, 8, 18, 19, 25, 26 AND 33-36 |

# INTERNATIONAL SEARCH REPORT

International application No.
**PCT/SG2017/050458**

**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:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

   Since only one version or copy of a sequence listing has been filed or furnished, the statements under item 2 are not required.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

**PCT/SG2017/050458**

*Note: This Annex lists known patent family members relating to the patent documents cited in this International Search Report. This Authority is in no way liable for these particulars which are merely given for the purpose of information.*

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 2015/163818 A1 | 29/10/2015 | CN 106459160 A | 22/02/2017 |
| | | EP 3134427 A1 | 01/03/2017 |
| | | JP 2017515468 A | 15/06/2017 |
| | | US 2017/0044515 A1 | 16/02/2017 |
| WO 2017/054044 A1 | 06/04/2017 | NONE | |