Supplementary data

Engineering a Catalytically Efficient Recombinant Protein Ligase.

Renliang Yang[1,2], Yee Hwa Wong[1,2], Giang K T Nguyen[1], James P Tam[1], Julien Lescar[1,2]*, Bin Wu[1,2]*

EXPERIMENTAL PROCEDURES

Protein sequence analysis,

A structure-based alignment of complete amino-acid sequences for human legumain, OaAEP1 and butelase 1 is shown in **fig. S1A.** These three enzymes share significant amino-acid sequence homology especially in their core catalytic domains and also an overall conserved molecular architecture that can be broken into a core enzymatic domain, a linker and prodomain (or cap) region (**fig. S1B and Fig. 2A**). Amino-acid sequence identities between these three proteins are as follows: OaAEP1/butelase 1: 311/471 (66%); OaAEP1/hlegum: 175/435 (40%); butelase 1/hlegum: 163/434 (38%). However, as shown in the present work, substitutions at a few key residues positions result in very different enzymatic activities (Asparaginyl endoprotease, moderately active peptide ligase and fast acting ligase respectively). Note that Cys247 is not conserved in butelase 1.

Protein expression, purification and activation

A gene with codons optimized for expression in *E. coli*, encoding a protein composed of a N-terminal hexa-His tag, the 76 amino-acid residues human ubiquitin and residues 24-474 of OaAEP1 (devoid of the OaAEP1 signal peptide) was synthesized by Genescript, USA. The nucleotide sequence is shown below:

CCATGGGCATGGCGCACCACCACCACCACCACATGCAGATCTTCGTTAAAACCCTGACCGGCAAGACCAT
TACCCTGGAAGTGGAACCGAGCGACACCATCGAGAACGTGAAAGCGAAGATCCAAGACAAAGAAGGTATT
CCGCCGGATCAGCAACGTCTGATTTTTGCGGGCAAGCAGCTGGAGGACGGTCGTACCCTGAGCGATTACA
ACATCCAAAAAGAAAGCACCCTGCATCTGGTGCTGCGTCTGCGTGGTGGCGCGCGTGACGGTGATTATCT
GCACCTGCCGAGCGAGGTGTCTCGTTTCTTTCGTCCGCAGGAGACCAACGACGATCACGGCGAAGACAGC
GTGGGTACCCGTTGGGCGGTTCTGATTGCGGGTAGCAAGGGCTACGCGAACTATCGTCATCAGGCGGGCG
TGTGCCATGCGTACCAAATCCTGAAACGTGGTGGCCTGAAGGATGAGAACATCGTGGTTTTCATGTACGA
CGATATTGCGTATAACGAAAGCAACCCGCGTCCGGGTGTTATCATTAACAGCCCGCACGGCAGCGATGTG
TATGCGGGTGTTCCGAAAGATTATACCGGCGAGGAAGTGAACGCGAAGAACTTCCTGGCGGCGATCCTGG
GTAACAAAAGCGCGATTACCGGTGGCAGCGGCAAGGTGGTTGACAGCGGTCCGAACGATCACATCTTTAT
TTACTATACCGACCACGGTGCGGCGGGCGTTATCGGTATGCCGAGCAAGCCGTACCTGTATGCGGACGAG
CTGAACGATGCGCTGAAGAAAAAGCACGCGAGCGGTACCTACAAAAGCCTGGTGTTCTATCTGGAGGCGT

GCGAAAGCGGCAGCATGTTTGAGGGTATCCTGCCGGAAGATCTGAACATTTACGCGCTGACCAGCACCAA
CACCACCGAAAGCAGCTGGTGCTACTATTGCCCGGCGCAGGAGAACCCGCCGCCGCCGGAATATAACGTG
TGCCTGGGCGACCTGTTCAGCGTTGCGTGGCTGGAGGACAGCGATGTGCAGAACAGCTGGTACGAAACCC
TGAACCAGCAATATCACCACGTTGATAAGCGTATTAGCCACGCGAGCCACGCGACCCAATACGGTAACCT
GAAACTGGGCGAGGAAGGTCTGTTTGTGTACATGGGCAGCAACCCGGCGAACGACAACTATACCAGCCTG
GATGGTAACGCGCTGACCCCGAGCAGCATCGTGGTTAACCAGCGTGACGCGGATCTGCTGCACCTGTGGG
AGAAATTCCGTAAGGCGCCGGAAGGCAGCGCGCGTAAAGAGGAAGCGCAGACCCAAATTTTTAAGGCGAT
GAGCCACCGTGTTCACATCGACAGCAGCATCAAACTGATTGGCAAGCTGCTGTTCGGTATCGAGAAATGC
ACCGAAATTCTGAACGCGGTTCGTCCGGCGGGTCAACCGCTGGTTGACGATTGGGCGTGCCTGCGTAGCC
TGGTGGGTACCTTTGAGACCCACTGCGGCAGCCTGAGCGAATACGGTATGCGTCACACCCGTACCATCGC
GAACATTTGCAACGCGGGTATTAGCGAGGAACAGATGGCGGAAGCGGCGAGCCAAGCGTGCGCGAGCATC
CCGTAACATATG

The clone is OaAEP1 one amino acid different from OaAEP1b (E371V). OaAEP1 was cloned from mRNA and OaAEP1b from DNA sequence. Nevertheless, our wild type constructs behaves essentially the same as previously reported.

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^{o} 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 and Ub-AEP-FL 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^{o}. 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. The primers used for mutagenesis are:

AEP mutagenesis primers:

5-AEP-G182L-Mut

GGCGGGCGTTATCCTTATGCCGAGCAAGCCG

5-AEP-Y190A-Mut:

GCCGAGCAAGCCGTACCTGGCTGCGGACGAGCTGAACGATGC

5-AEP-C217S-Mut:

GGTGTTCTATCTGGAGGCGAGCGAAAGCGGCAGCATGTTTGAG

5-AEP-E218R:

GGTGTTCTATCTGGAGGCGTGCAGAAGCGGCAGCATGTTTGAG

5-AEP-W246A_Mut:

GCACCAACACCACCGAAAGCAGCGCGTGCTACTATTGCCCGGCGCAGGAG

5-AEP-C247A-Mut :

CACCACCGAAAGCAGCTGGGCCTACTATTGCCCGGCGC

5-AEP-C247G-Mut

CACCACCGAAAGCAGCTGGGGCTACTATTGCCCGGCGC

5-AEP-C247S_Mut:

CACCACCGAAAGCAGCTGGTCCTACTATTGCCCGGCGC

5-AEP-C247T_Mut:

CACCACCGAAAGCAGCTGGACCTACTATTGCCCGGCGC

5-AEP-C247L_Mut

CACCACCGAAAGCAGCTGGCTCTACTATTGCCCGGCGC

5-AEP-C247I_Mut

CACCACCGAAAGCAGCTGGATCTACTATTGCCCGGCGC

5-AEP-C247V_Mut:

CACCACCGAAAGCAGCTGGGTCTACTATTGCCCGGCGC

5-AEP-C247M_Mut:

CACCACCGAAAGCAGCTGGATGTACTATTGCCCGGCGC

5-AEP-C264Y-Mut:

CCGCCGCCGGAATATAACGTGTACCTGGGCGACCTGTTCAGCG

5-AEP-N346A-Mut:

CCCCGAGCAGCATCGTGGTTGCCCAGCGTGACGCGGATCTG

5-AEP-Q347A-Mut:

CGAGCAGCATCGTGGTTAACGCGCGTGACGCGGATCTG

5-AEP-NQ346_347AA_Mut:

CCCCGAGCAGCATCGTGGTTGCCGCGCGTGACGCGGATCTG

5-AEP-D349E_Mut:

GCATCGTGGTTAACCAGCGTGAGGCGGATCTGCTGCACCTGTGGGAG

5-AEP-D349_351AA_Mut:

GCATCGTGGTTAACCAGCGTGCGGCGGCTCTGCTGCACCTGTGGGAG

5-AEP-N346D349D351-A3-Mut:

CGAGCAGCATCGTGGTTGCCCAGCGTGCCGCGGCTCTGCTGCACCTGTGGGAGAAA

TTC5-AEP-N346Q347D349D351-A4-Mut:

CGAGCAGCATCGTGGTTGCCGCGCGTGCCGCGGCTCTGCTGCACCTGTGGGAGAAATTC

5-AEP-N327D328N329D334N336-A5-Mut:

CAGCAACCCGGCGGCCGCCGCCTATACCAGCCTGGCTGGTGCCGCGCTGACCCCGAGCAG

5-AEP- N346D349D351-QEE(QE3)-Mut:

CGAGCAGCATCGTGGTTCAGCAGCGTGAAGCGGAGCTGCTGCACCTGTGGGAGAAATTC

5-AEP- N327D328N329D334N336-QEQEQ(QE5)-Mut:

CAGCAACCCGGCGCAGGAACAATATACCAGCCTGGAGGGTCAGGCGCTGACCCCGAGCAG

5-AEP-R348A-Mut:

CCGAGCAGCATCGTGGTTAACCAGGCTGACGCGGATCTGCTGCACC

5-AEP-R348K-Mut:

CCGAGCAGCATCGTGGTTAACCAGAAAGACGCGGATCTGCTGCACC

Loop Swap primer F:

T CCT CCA GTT ACT CAC CTT GAC CTC ACC CCA TCC CCA GAT GTC CCT CTGCTGCACCTGTG

Loop Swap primer R:

GTGAGTAACTGGAGGAAGACTGATCGGGCTGCTAGCCCGGTGTTTGCTGCCCATGTACAC

Ubiquitin primers:

5-Ub-NGL insertion:

CCTGCGTCTGAAAGGTGGTAATGGTCTT TGAGGATCCTAACTCGAGGC

5-Ub-AA-NGL-Mut:

GCACTTGGTCCTGCGTCTGCGCGCTGCTAATGGTCTTTGAGGATCCTAACTCG
AG

5-Ub-GLPV-F:

CTTTAAGAAGGAGATATACATATGGGTCTCCCCGTGCAGATCTTCGTGAAAA
CCC

SNAP primers:

5-SNAP-ASTM/GLPV-Mut:

CTTTAAGAAGGAGATATACATATGGGTCTCCCCGTGGATATCAAGCTTACCG
GTG

SNAP-R:

GTTATTGCTCAGCGGTGGCAGCAGCCAAC

3-SNAP-His6-R:

CGGATCAATGCGGCCGCTTAGTGATGATGGTGATGGTGGGATCCGCCTGCAG
GTCCC

5_AEP_loop_Flag_rpl_F: CAACTATACCAGCCTGGATTATAAAGAT
GACGACGATAAAAGCAGCATCGTGGTTAACC

5-Ub-AA-NHV-Mut:
GCACTTGGTCCTGCGTCTGCGCGCTGCTAATCATGTTTGAGGATCCTAACTCG
AG

All mutants were confirmed using sequencing services from 1st Base (Singapore).

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-µl reaction mixtures

containing 20 mM phosphate buffer, pH 6.0, ligases (5 to 700 nM) and peptide substrates (10 to 300 μM). Each reaction was performed in triplicate at 37 °C and quenched by adding 5 μl 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.

Crystallization, data collection and structural determination.

The zymogenic OaAEP1 protein at a concentration of 12 mg/ml was subjected to robotic crystallization trials using the sitting drop/vapor diffusion method and a Mosquitoe instrument (TTP Biotech UK). This initial screen yielded small crystals at an acidic pH close to the activation pH. Crystal growth optimization was performed manually using the hanging drop method in 15 wells trays, using commercially available additive screens. Further improvement could also be obtained by macro-seeding. The best diffracting crystals grow from a precipitation solution containing 200 mM NH_4NO_3, pH ~ 4.5, 13-15% (w/v) PEG 3,350 with 10% (v/v) glycerol or ethylene glycol. These crystals were suitable for X-ray crystallography analysis at the Swiss Light Source. The space group was C2 with unit-cell dimensions a=145.73 Å, b=70.21 Å, c=118.28 Å, β =117.14°. Although crystals tend to suffer from radiation damage, a complete data set could be obtained from a single crystal. A total of 720 images of 0.25° oscillation each, was processed with the CCP4 package. Diffraction intensities were integrated with program Imosflm and scaled with program SCALA from the CCP4 suite (**table S1**) (*24*). The crystal structure of OaAEP1 in its zymogenic form was determined by molecular

replacement using the human zymogenic legumain structure as a search probe (PDB codes 4FGU and 4NOK). Model rebuilding sessions at the computer graphics, using program Coot (*25*) were interspersed with refinement to a resolution of 2.56 Å using program Buster and Phenix (*26*). Tight ncs restraints were used between the two oaAEP1 molecules present in the asymmetric unit. Buried surface areas were calculated with the PISA server (www.ebi.ac.uk/pdbe/pisa) and the Molprobity structure validation server (molprobity.biochem.duke.edu) was used to check the model stereochemistry. Figure illustrations were produced using Pymol.

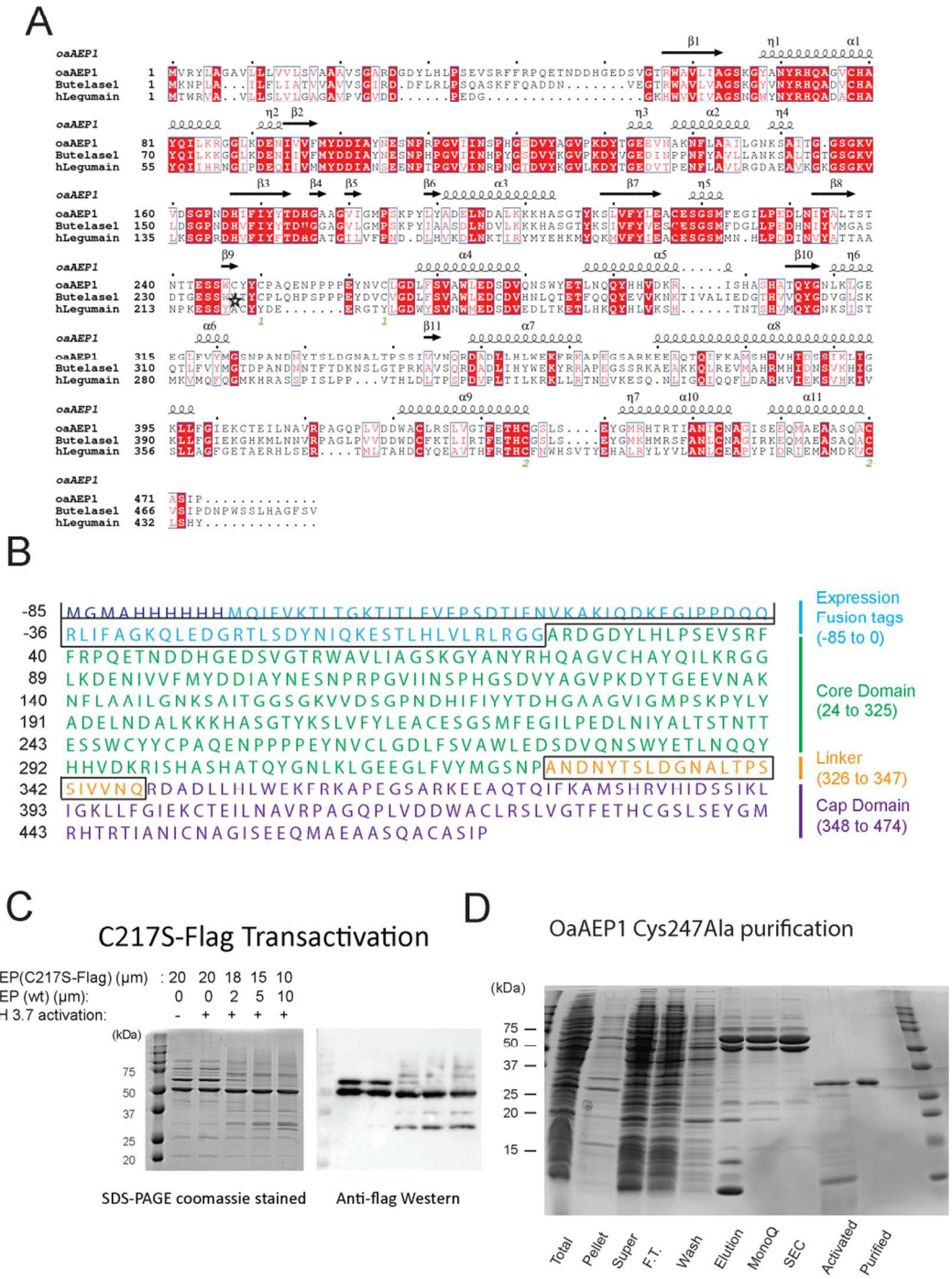

A
B
-85 MGMAHHHHHHMQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQ
-36 RLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGARDGDYLHLPSEVSRF
40 FRPQETNDDHGEDSVGTRWAVLIAGSKGYANYRHQAGVCHAYQILKRGG
89 LKDENIVVFMYDDIAYNESNPRPGVIINSPHGSDVYAGVPKDYTGEEVNAK
140 NFLAAILGNKSAITGGSGKVVDSGPNDHIFIYYTDHGAAGVIGMPSKPYLY
191 ADELNDALKKKHASGTYKSLVFYLEACESGSMFEGILPEDLNIYALTSTNTT
243 ESSWCYYCPAQENPPPPEYNVCLGDLFSVAWLEDSDVQNSWYETLNQQY
292 HHVDKRISHASHATQYGNLKLGEEGLFVYMGSNPANDNYTSLDGNALTPS
342 SIVVNQRDADLLHLWEKFRKAPEGSARKEEAQTQIFKAMSHRVHIDSSIKL
393 IGKLLFGIEKCTEILNAVRPAGQPLVDDWACLRSLVGTFETHCGSLSEYGM
443 RHTRTIANICNAGISEEQMAEAASQACASIP
Expression Fusion tags (-85 to 0)
Core Domain (24 to 325)
Linker (326 to 347)
Cap Domain (348 to 474)
C
C217S-Flag Transactivation
AEP(C217S-Flag) (μm) : 20 20 18 15 10
AEP (wt) (μm): 0 0 2 5 10
pH 3.7 activation: - + + + +
(kDa)
75
50
37
25
20
SDS-PAGE coomassie stained
Anti-flag Western
D
OaAEP1 Cys247Ala purification
(kDa)
75
50
37
25
20
15
Total
Pellet
Super
F.T.
Wash
Elution
MonoQ
SEC
Activated
Purified

Figure S1. Sequence analysis. A. Structure-based alignment of amino-acid sequences for OaAEP1 (Genbank accession code: KR259377), butelase 1 (Genbank accession code: KF918345.1) and human legumain (UniProtKB/Swiss-Prot: Q99538.1). Both signal peptide and prodomain (cap) region sequences are included in the sequence alignment. Secondary structure elements (derived from the OaAEP1 3D structure, this work) are indicated above the sequence with β-strands depicted as arrows and helices as springs. Strictly conserved residues are boxed and shaded in red while residues displaying chemically conservative substitutions between the three enzymes are colored in pink. Active site residues His175, Cys217 are indicated with red stars. Residue Cys247, a crucial determinant for the kinetics of ligase activity (this work), is indicated by a black star. **B.** Complete amino acid sequence of the composite protein expressed in this study. Residues numbered-85 to 0 encompasses the N-terminally fused hexa-histidine and ubiquitin tags (colored in cyan). This segment is immediately followed by residue Ala24 of AEP, which is the predicted N-terminus of the OaAEP1 protein following signal peptide cleavage. Residues 24-325) forming the catalytic core domain of OaAEP1 are colored in green. The linker residues (326-347) are highlighted in orange, and the cap domain (348-474) is displayed in purple. The same color code was used to produce Fig. 2A. **C**. Additional data to demonstrate the trans-activation mechanism. An internally flag-tagged OaAEP1 Cys217Ser inactive mutant was co-incubated with wild type enzyme at different concentrations. Western-blot analysis convincingly showed that this inactive mutant could be cleaved by other OaAEP1, thus proving trans-activation cleavage is possible. **D**. The purification of OaAEP1 Cys247Ala mutant. This mutant protein was purified with at least 95% purify before being used for ligation kinetic test.

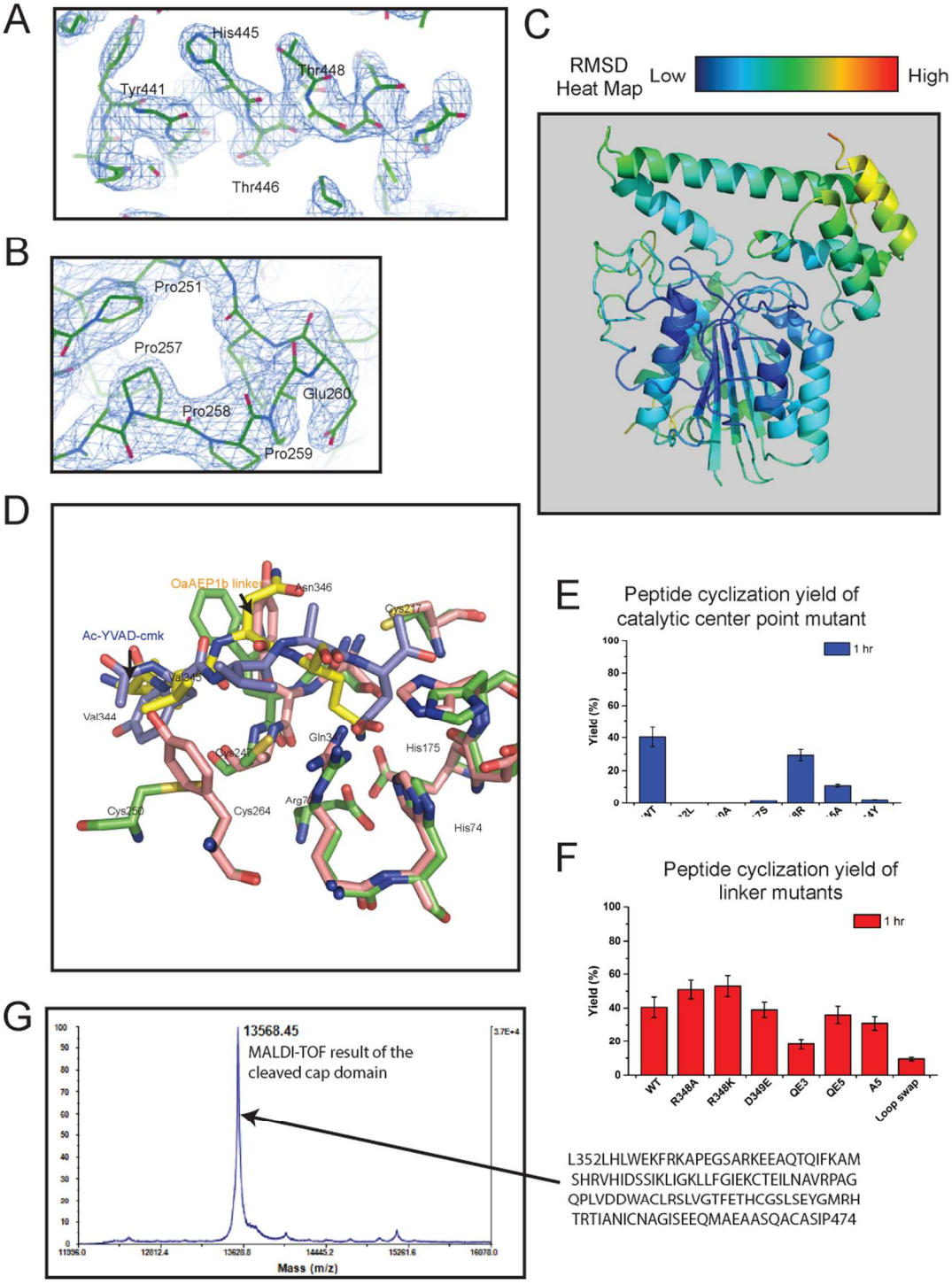
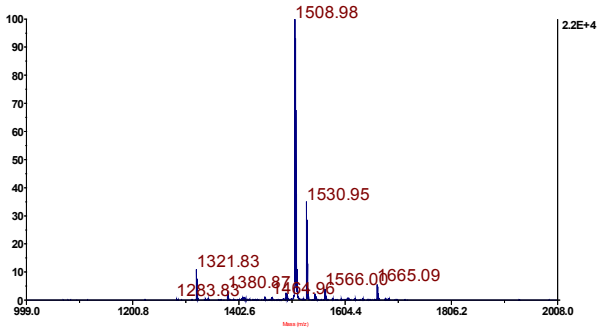
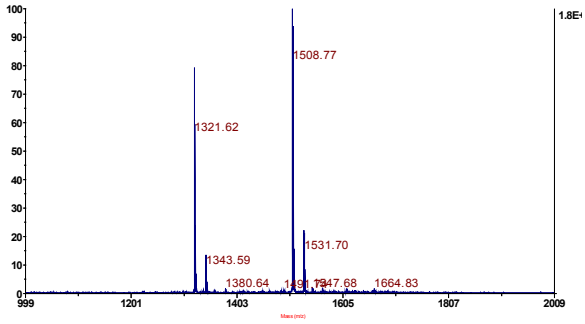

A
His445
Thr448
Tyr441
Thr446
B
Pro251
Pro257
Glu260
Pro258
Pro259
C
RMSD
Heat Map
Low
High
D
OaAEP1b linker
Asn346
Cys247
Ac-YVAD-cmk
Val345
Val344
Gln3
His175
Cys247
Cys250
Cys264
Arg7
His74
E
Peptide cyclization yield of
catalytic center point mutant
1 hr
Yield (%)
F
Peptide cyclization yield of
linker mutants
1 hr
Yield (%)
WT
R348A
R348K
D349E
QE3
QE5
A5
Loop swap
G
13568.45
MALDI-TOF result of the
cleaved cap domain
Mass (m/z)
L352LHLWEKFRKAPEGSARKEEAQTQIFKAM
SHRVHIDSSIKLIGKLLFGIEKCTEILNAVRPAG
QPLVDDWACLRSLVGTFETHCGSLSEYGMRH
TRTIANICNAGISEEQMAEAASQACASIP474

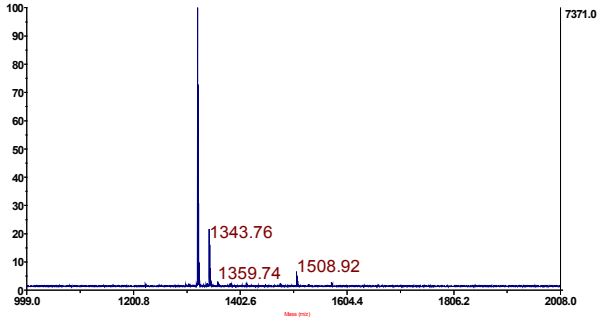
Figure S2. Views of the 2Fo-Fc electron density map with phases from the refined model, highlighting the quality of the fit of the model to the map. A. Close-up view of the OaAEP1 α_{10} helix structure: this region of the cap domain departs from the human legumain structure. The good definition of the map suggests well-assigned electron density. **B.** Close-up view of the Pro-rich loop Pro256-Pro257-Pro258-Pro259 within the core domain of OaAEP1. This Pro-rich loop, which is inserted in the sequence of both OaAEP1 and butelase 1 but is absent in the human legumain structure, appears to be a unique feature of AEP ligases. The density is also unambiguously assigned in this region. **C.** The RMSD heat map of OaAEP1 crystal structure. The catalytic pocket region is more stable in the crystal, while the cap domain and the residues close to the linker are more flexible. **D**. The linker region 344-347 of the OaAEP1 protein (depicted as yellow sticks) adopts an overall orientation closely similar (but shifted by about half the size of a peptide bond) to the AcTyr-Val-Ala-Asp-CMK inhibitor covalently bound to human legumain (blue sticks, PDB code 4AWA) via Cys189 (corresponding to Cys217 in OaAEP1). Note the good spatial overlap between the amide side-chain of Gln347 (yellow sticks OaAEP1) that inserts in the active site pocket of OaAEP1 and the carboxyl group of the corresponding Asp of the inhibitor. **E.** Peptide cyclization efficiency of multiple point mutations of OaAEP1. These results highlight the importance of an intact catalytic center of OaAEP1 for ligation activity. **F.** Peptide cyclization efficiency of several linker mutants. Only mutating multiple Asn/Asp/Gln residues within the 346/351 segment of the linker can significantly impact on the OaAEP1 activation and ligation activities, suggesting a complex activation profile. **G**. The cleaved cap domain was subjected to MALDI-TOF analysis. The results demonstrated that the primary cleavage site during activation is likely to be Asp351. There are a few minor peaks with slightly larger molecular weight, suggesting possible alternative cleavage sites within 347-351 stretch.

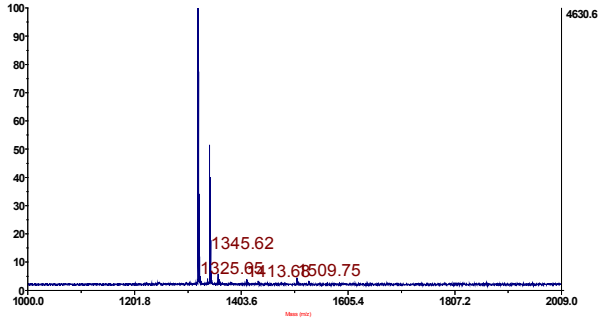
Cys247, 10 min and 1 hour ligation



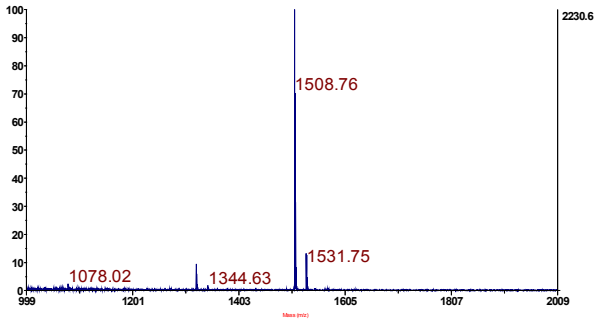


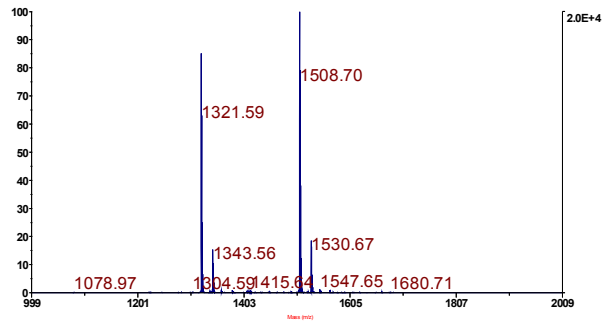
Cys247Gly, 10 min and 1 hour ligation



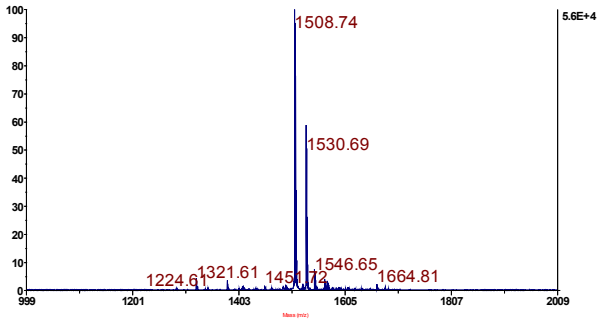


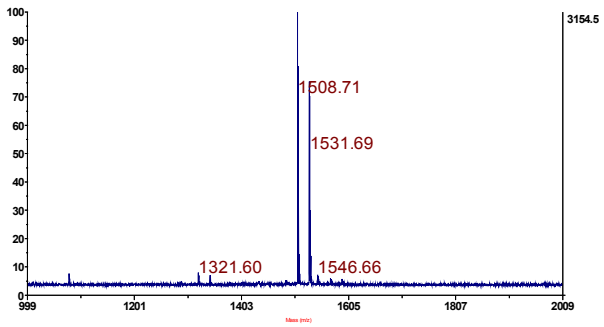
Cys247Ser, 10 min and 1 hour ligation



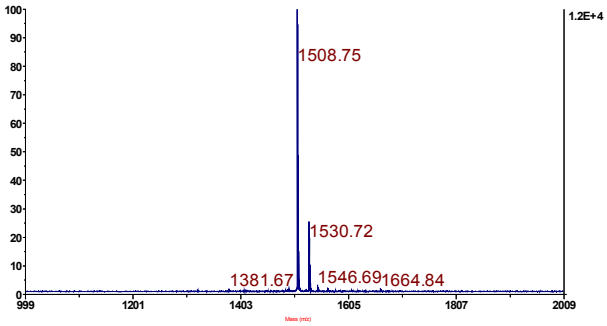


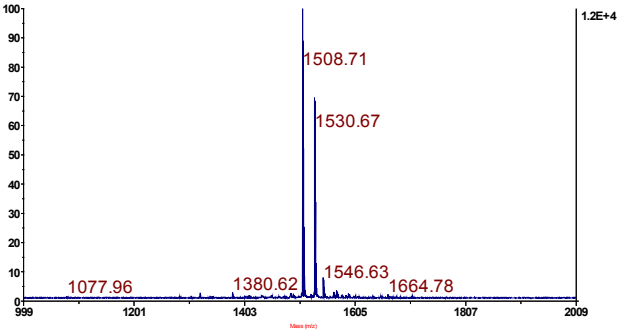
Cys247Thr, 10 min and 1 hour ligation



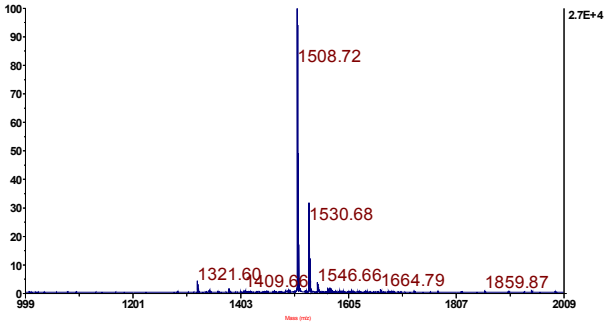


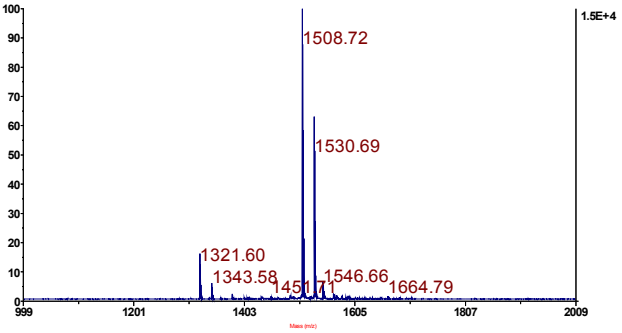
Cys247Met, 10 min and 1 hour ligation



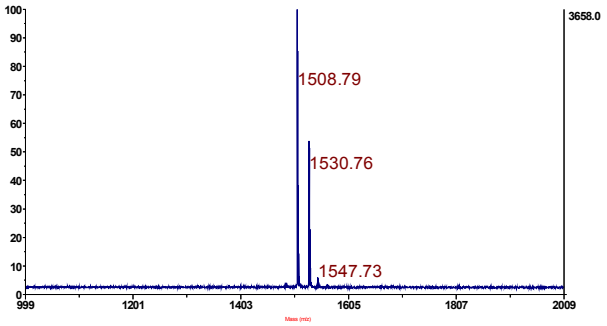


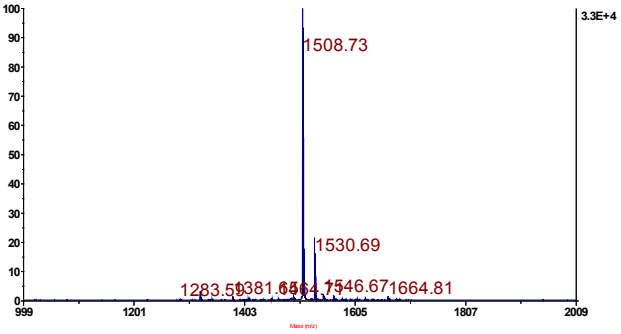
Cys247Val, 10 min and 1 hour ligation



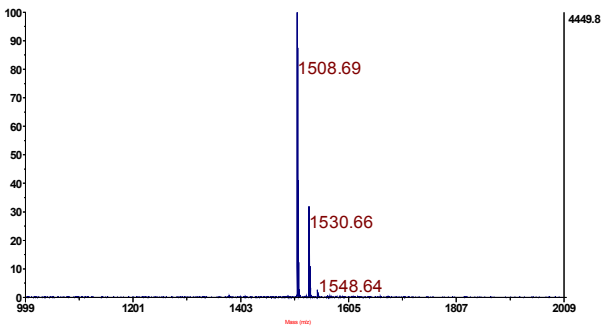


Cys247Leu, 10 min and 1 hour ligation





Cys247Ile, 10 min and 1 hour ligation



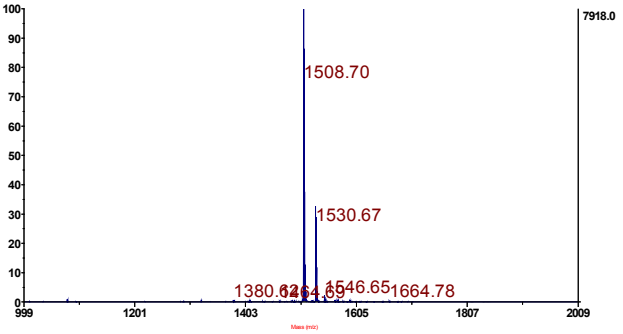


Figure S3. Raw mass-spectrometric results for the WT OaAEP1 and the series of "gate-keeper" mutants studied in this work. Experiments were done using MALDI-TOF MS as described in the text and the surface areas were evaluated at two time points: 10 min and 1hr for Cys247, Cys247Gly, Cys247Ser, Cys247Thr, Cys247Met, Cys247Val, Cys247Leu, Cys247Ile, respectively.

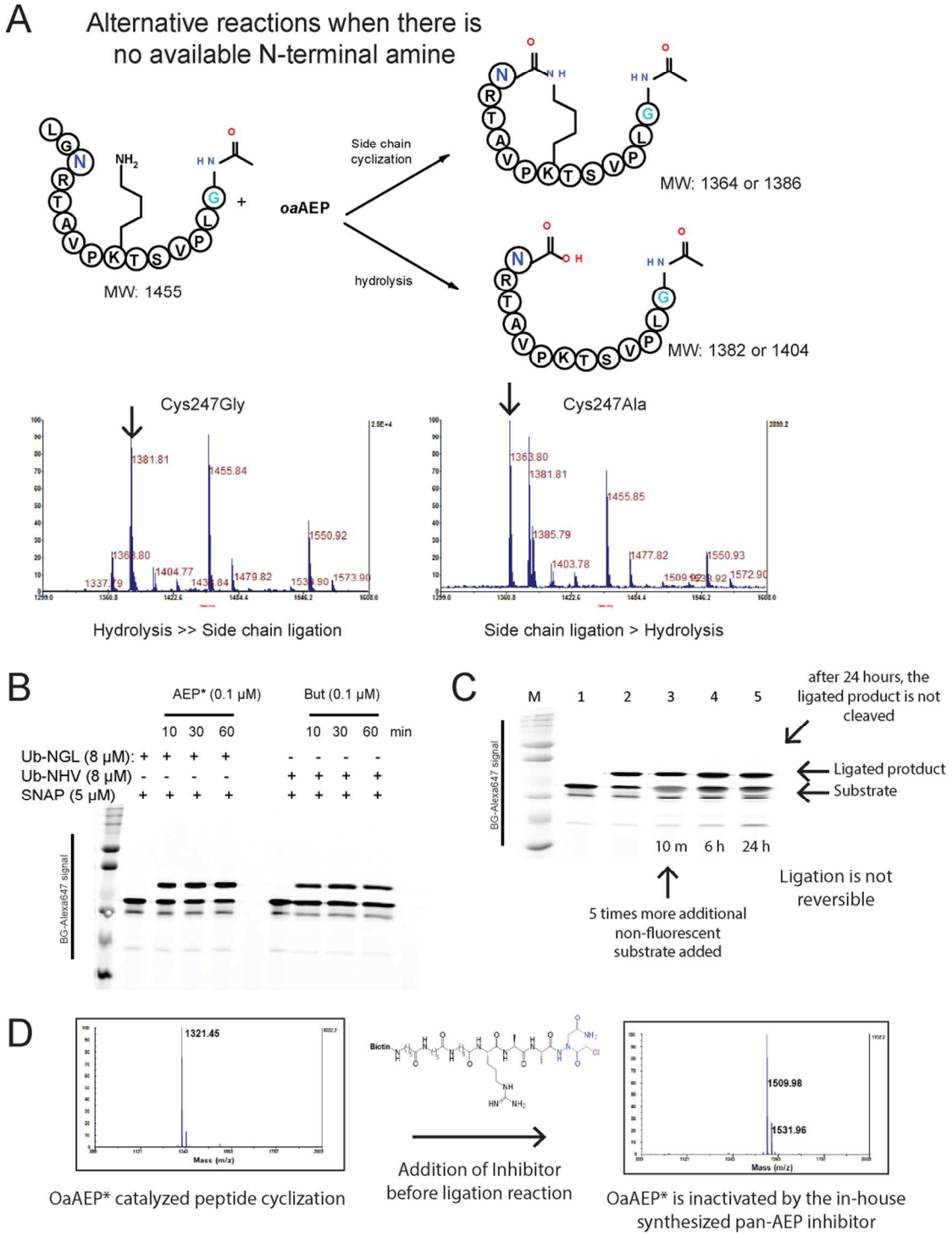

A
Alternative reactions when there is no available N-terminal amine
MW: 1455
oaAEP
Side chain cyclization
hydrolysis
MW: 1364 or 1386
MW: 1382 or 1404
Cys247Gly
Cys247Ala
Hydrolysis >> Side chain ligation
Side chain ligation > Hydrolysis
B
AEP* (0.1 μM)
But (0.1 μM)
10 30 60
10 30 60 min
Ub-NGL (8 μM): + + + + - - - -
Ub-NHV (8 μM) - - - - + + + +
SNAP (5 μM) + + + + + + + +
BG-Alexa647 signal
C
M 1 2 3 4 5
after 24 hours, the ligated product is not cleaved
Ligated protduct
Substrate
BG-Alexa647 signal
10 m 6 h 24 h
Ligation is not reversible
5 times more additional non-fluorescent substrate added
D
1321.45
Mass (m/z)
Biotin
Addition of Inhibitor before ligation reaction
1509.98
1531.96
Mass (m/z)
OaAEP* catalyzed peptide cyclization
OaAEP* is inactivated by the in-house synthesized pan-AEP inhibitor

Figure S4. Potential side products of the OaAEP1 catalyzed ligation reaction. A. When the N-terminal amine of the substrate is protected by an acetyl-group, side-chain ligation and proteolysis reactions are observed in the presence of either the Cys247Gly or Cys247Ala mutant. The smaller size of the side chain of Cys247Gly mutant presumably permits water molecules to participate in the reaction and is likely to favor the proteolysis reaction. **B**. Side by side comparison of OaAEP1 Cys247Ala and Butelase 1 in protein-protein ligation activity. The optimum substrates were used for the comparison, Ub-NGL for OaAEP1* and Ub-NHV for Butelase 1, respectively. **C**. Additional enzyme and non-labeled substrate were used to test the reversibility of the ligation reaction. Up to 24 hours, there is no visible hydrolysis or reversed ligation seen. **D**. An in-house synthesized pan-AEP inhibitor was very effective in quenching the peptide cyclization reaction catalyzed by OaAEP1*, demonstrating the catalytic pocket residues mentioned in our discussion is indeed responsible for the ligation reaction.

Table S1: Data collection and refinement statistics

	OaAEP1 (5H0I)
Resolution range (Å)	71.72 - 2.56 (2.70 - 2.56)*
Space group	C 2
Unit cell dimensions a, b, c (Å) Angles α, β, γ, (°)	145.73 70.21 118.28 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	95.7
allowed	3.4
outliers	0.6
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.