Improving the use of asparaginyl endopeptidase for biocatalytic applications

Submitted to Cardiff University for the degree of Doctor of Philosophy by

Tsz Man Simon Tang

Supervisor: Dr. Louis Y. P. Luk

2021

ABSTRACT

Asparaginyl endopeptidases (AEP) are cysteine proteases found in mammalian and plant cells. Several AEP isoforms from plant species were found to exhibit transpeptidase activity, which is integral for the key head-to-tail cyclisation reaction during the biosynthesis of cyclotides. AEPs isolated from plants exhibit excellent enzyme kinetics for peptide ligation *via* a relatively short substrate recognition sequence. Hence, AEPs could offer powerful applications as a biocatalyst in the production of modified proteins and peptides. However, the uptake of AEPs for biocatalytic application is limited by key challenges. In this thesis, an engineered AEP variant from *Oldenlandia affinis*, OaAEP1-C247A, is employed as a model enzyme to explore strategies which can address some vital limitations to the application of AEP catalysis.

Fundamental enzymology of AEPs are discussed and a new protocol which simplifies the recombinant preparation of OaAEP1-C247A is described. To address the challenges posed by the reversible reactivity of the enzyme, a chemo-enzymatic approach to AEP-mediated peptide and protein bioconjugation is reported. Employing an affordable chemical reagent as a scavenger, this method takes advantage of the non-specific substrate recognition of OaAEP1-C247A. Accumulation of active proteases in the cytoplasm can be detrimental to the viability of the host cell. Therefore, strategies to encapsulate OaAEP1-C247A within a self-assembled protein capsid (AaLS-13) were explored. The caged protein complex serves as an artificial organelle for protein and peptide processing which mimics organelles and protein microcompartments found in nature. This synthetic biology approach offers the potential to enable AEP-catalysis in *E. coli*.

ACKNOWLEDGEMENTS

This thesis marks the completion of an incredible journey, through which I have learnt a great deal both in and outside of scientific research. My achievements would not be possible without the help and support of all the fantastic people around me, and for this, I would like to express my most sincere gratitude.

First, my thanks to Dr. Louis Luk for giving me this opportunity to study and develop as a scientist under his guidance and supervision. I appreciate the way he continually challenges me to improve and giving me the freedom to pursue my research ideas and interests. Thank you to my co-supervisor, Dr. Yuhsuan Tsai, and also Dr. Yi Jin for their advice and inputs towards my research.

Thank you to Prof. Donald Hilvert for hosting me at ETH for my short visit, for giving me the opportunity to learn about the lumazine synthase projects. His suggestions regarding the role of the AEP pro-domain inspired me to perform the split-AEP experiments. Mikail Lavasseur for his support in the lab at ETH and all the members in the Hilvert/Kast group who are incredibly friendly and welcomed me into their research team.

A multitude of thank you's are in order to all the past and present members from the Luk, Tsai, Jin and Allemann groups with whom I have learnt, shared and experienced the good times and the bad. In particular, Tom Williams for his extensive knowledge and expertise in analytical chemistry and chemical biology (not to mention the countless mass spectrometry experiments!). To Davide Cardella, Alex Lander and Dr. XueFei Li, my Italian, Welsh and Chinese peptide synthesisers. I could not ask for a better support team in the delivery of our publication. To Patrick Baumann, Luke Spear and Dr. Alex Nödling, for all the discussions, sage advices and questionable music choices in the lab. To Emily Mills and Heather Hayes for the Friday 3pm shenanigans.

Thank you to my family. My mum, dad and sister for their unconditional and unwavering support in every choice I've made. First, when I left home at a young age, and then when I decided to remain in the UK and pursue a PhD.

Finally, thank you to my incredible girlfriend, Ellie, who has been there with me for this journey. I am grateful to have her to celebrate the wins with me and to give me a rational outsider's view when the going gets tough.

TABLE OF CONTENTS

LIST OF FIGURES AND SCHEMES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abz   2-aminobenzoyl

AcOH   Acetic acid

AEP   Asparaginyl endopeptidase

AtLEGγ   *Arabadopsis thaliana* legumain γ

CA   Carbonic anhydrase

CMK   Chloromethyl ketone

cOaAEP1   *Oldenlandia affinis* asparaginyl endopeptidase 1b [C247A] core domain

cpAaLS   Circularly permutated *Aquifex aeolicus* lumazine synthase

DMF   N, N-dimethyl formamide

Dnp   2,4-dinitrophenyl

EDTA   Ethylenediaminetetraacetic acid

ER   Endoplasmic reticulum

ESI-MS   Electrospray ionisation mass spectrometry

fOaAEP1   Full length *Oldenlandia affinis* asparaginyl endopeptidase 1b [C247A]

FPBA   2-formylphenylboronic acid

FRET   Förster resonance energy transfer

GFP   Green fluorescent protein

GK   Gatekeeper

hLEG   human legumain

IMAC   Immobilised metal affinity chromatography

IPTG   Isopropyl-β-D-1-thioglactopyranoside

LAD   Ligase activity determinant

LB   Lysogeny broth

LC-MS   Liquid chromatography mass spectrometry

LS   Lumazine synthase

MCoTI-II   *Momordica cochinchinensis* trypsin inhibitor 2

MLA   Marker of ligase activity

MSP2   Merozoite surface protein 2

NMR   Nuclear magnetic resonance

OaAEP1   *Oldenlandia affinis* asparaginyl endopeptidase 1b [C247A]

PAL   Peptide asparaginyl ligase

PCR   Polymerase chain reaction

POC   Proof of concept

PPL   Poly-proline loop

RS   Riboflavin synthase

RuBisCo   Ribulose-1,5-bisphosphate carboxylase/oxygenase

SAX   Strong anion exchange

SDS-PAGE   Sodium dodecylsulfate poly acrylamide gel electrophoresis

SEC   Size exclusion chromatography

sOaAEP1   simplified *Oldenlandia affinis* asparaginyl endopeptidase 1b [C247A]

TCEP   Tris(2-carboxyethyl) phosphine

TEM   Transmission electron microscopy

TEVp   Tobacco etch virus protease

tOaAEP1   truncated *Oldenlandia affinis* asparaginyl endopeptidase 1b [C247A]

TOP   Topaz yellow fluorescent protein

UV   Ultra-violet

VPE   Vacuolar processing enzyme

1 Introduction

1.1   Asparaginyl endopeptidase

Proteases have a central role in modification and homeostasis of proteins. Asparaginyl endopeptidase (AEP) are proteases identified from the vacuole of plant cells.[1] These enzymes are also known as legumain,[2,3] vacuolar processing enzyme[1] or peptide asparaginyl ligase.[4] AEPs react selectively at an internal asparagine (Asn) or aspartate (Asp) residue of a peptide backbone *via* a catalytic cysteine residue. The structure and activity of AEPs are similar to caspases.[5–7] Moreover, AEPs were also identified in other organisms, including humans (hLEG) and mice (mLEG).[2,8] The enzymes are folded in the endoplasmic reticulum (ER) as a zymogen with a pro-domain which prevents substrate access to the active site.[9–11] Pro-AEPs have been shown to accumulate in ER bodies within the epidermal cells of *Arabidopsis thaliana* seedlings.[9] In response to stress, ER bodies fuse and release zymogenic AEPs into the vacuole.[9] AEPs have also been reported to process seed storage proteins, 2S albumins, in prevacuolar multivesicular bodies (MVB) before reaching the protein storage vacuole in A. thaliana embryos.[12] The acidic environment within the vacuole triggers a proteolytic maturation to afford the active AEP (figure 1.1).[2,9,11–14]
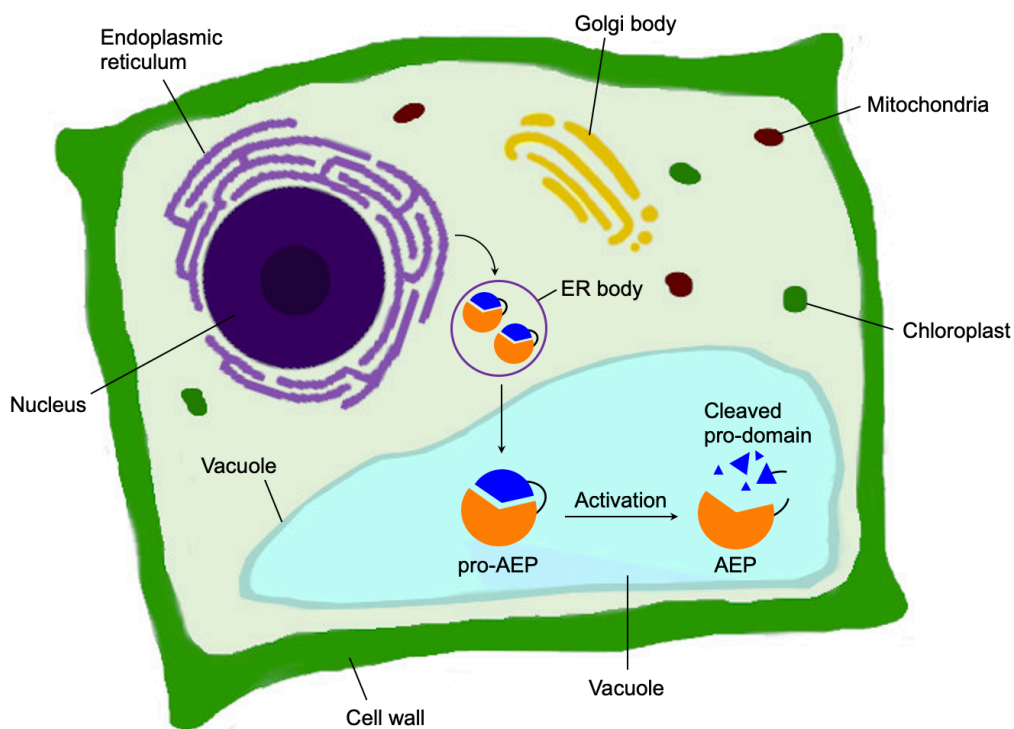
Figure 1.1 Biosynthesis of AEP in plant cells. Zymogenic AEPs are produced in the ER and stored in ER bodies. The pro-enzyme is trafficked to the vacuole, where the acidic environment enables cleavage of the pro-domain resulting in activation.

Matured AEPs primarily function as peptidases which are associated with several critical cellular processes (figure 1.2).[2,15] Purified AEP from the seed of castor beans (*Ricinus communis*) have been shown to mediate the proteolytic processing pro-proteins to afford mature seed storage proteins including 2S albumin and 11S globulin.[1,16] In vegetative organs, AEPs have been shown to facilitate programmed cell death.[9,17] When an environmental stress (concentrated salt solution) was applied to healthy seedlings of *A. thaliana*, ER bodies were found to fuse and deliver AEPs to the vacuole, leading to cell death.[9] Mammalian orthologues of AEPs are associated with antigen processing which moderate antigen presentation and T-cell activity.[18–20] Myelin basic protein (MBP), an autoantigen associated with inflammatory diseases, was shown to be cleaved by AEP. Overexpression of a gene encoding for AEP in mammalian cells was shown to inhibit presentation of MBP and T-cell activation.[19] On the other hand, AEP-mediated proteolytic processing of the microbial tetanus toxin antigen was found to enable antigen presentation by class II major histocompatibility complex molecules.[18]
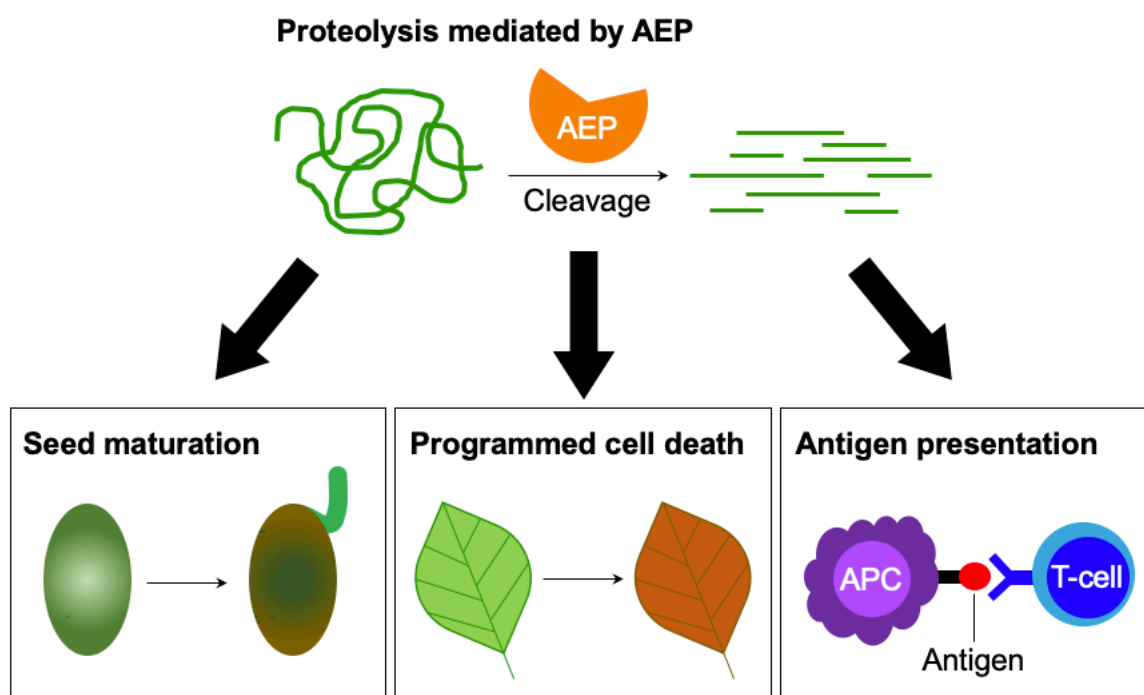
Figure 1.2 The peptidase activities of AEPs are involved in several critical cellular processes including seed maturation, programmed cell death and the regulation of antigen presentation.

Although most AEPs were shown to facilitate peptide cleavage, some AEPs isolated from plant species were found to exhibit peptide ligase activity.[4,21–27] Various plant species have been found to utilise this peptide ligase activity to facilitate post translational modification to produce backbone cyclised peptides (figure 1.3).[4,22–24,26] The cyclisation step during the biosynthesis of kalata B1, a cystine knotted macrocyclic peptide, was shown to be catalysed by an AEP (OaAEP1b) in *Oldenlandia affinis*.[22] Cyclic peptides found in plants exhibit a range of bioactivities,[28,29] and demonstrate superior stability compared to their linear counterparts.[30] Kalata B1 extracted from *O. affinis*, a native African plant species, demonstrate insecticidal activities which strongly inhibit the growth of *Helicoverpa punctigera* larvae.[31,32] Furthermore, the cyclic kalata B1 was shown to display high tolerance towards thermal (370 K, 97 °C) and chemical (8 M urea) denaturation.[30] Another example of a cystine knotted cyclic peptide is Cter-M, which is cyclised by an AEP, dubbed butelase 1, in butterfly pea plants (*Clitoria ternatea*).[33] Other AEPs such as HaAEP1 from common sunflowers (*Helianthus annuus*) and MCoAEP2 from gac fruit plants (*Momordica cochinchinensis*) have been reported to facilitate backbone cyclisation during the biosynthesis of cyclic trypsin inhibitor peptides SFTI-1 and MCoTI-II, respectively (figure 1.3).[27,34]
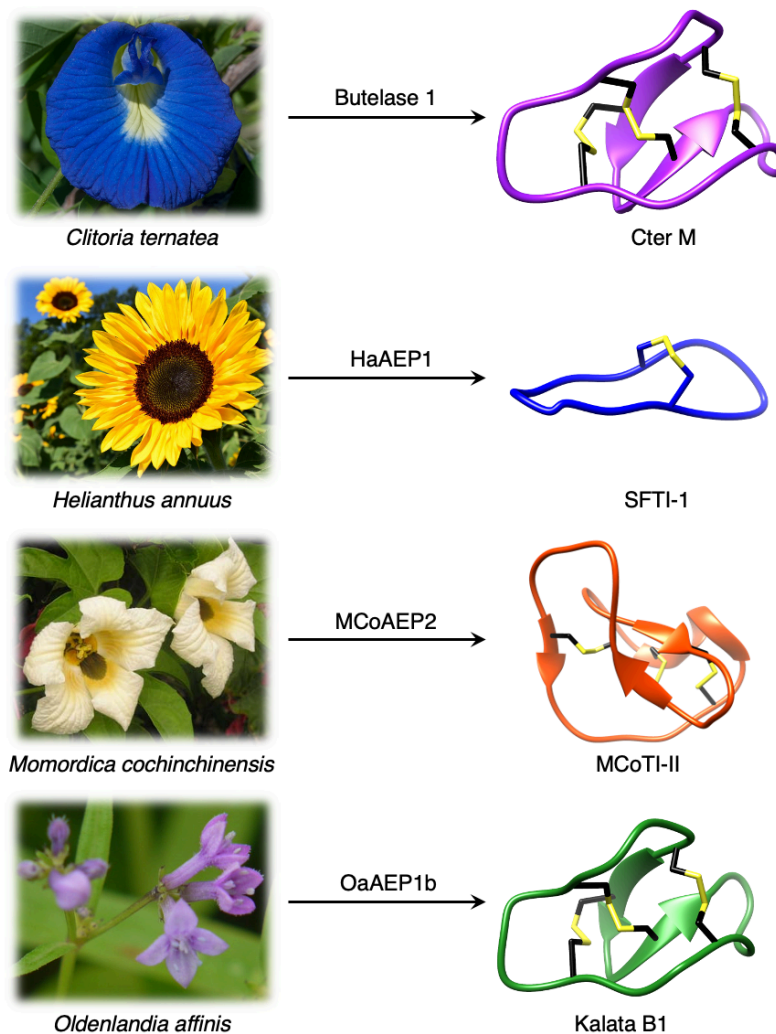
Figure 1.3 Plant species utilise AEPs with ligase activities to generate backbone cyclised peptides with enhanced stability and bioactivity. Cter M (PDB: 2LAM) from *C. ternatea*, SFTI-1 (PDB: 1JBL) from *H. annuus*, MCoTI-II (PDB: 1IB9) from *M. cochinchinensis* and kalata B1 (PDB: 1NB1) from *O. affinis*.

1.2 Reaction mechanisms of AEPs

Since the initial discovery of a AEP in the vacuole of maturing seeds from castor beans,[1] there has been an expansion of this research area to offer greater understanding of AEPs. The availability of X-ray crystallographic data from a number of AEPs originating from plant and mammalian species have provided some fundamental knowledge and rationales for the localisation and activity of these enzymes.[4,11,14,35,36] Prominent structural features include the catalytic core domain which adopts a protein fold with a six-stranded β-sheet surrounded by five α-helices. A further two anti-parallel β-sheets serve as the substrate binding domains near the active site (figure 1.4). In the proenzyme, the catalytic site is covered by the pro-

domain which is C-terminal to the core domain (figure 1.4).[4,11,14,35,36] Studies performed on an AEP from *A. thaliana* (AtLEGγ) revealed that pro-AEP adopts a dimeric state at neutral pH which has been shown to prevent enzymatic activity.[11] The enzyme adopts a monomeric state at pH 4.0 and the protonation of salt bridges surrounding the active site of the AEP were reported to increase flexibility, enabling the auto-proteolytic cleavage of the pro-domain to generate the active enzyme.[9,11,22,35] Taken away from the acidic and reducing environments of the plant vacuole, it has been shown the cleaved core and pro-domains of AtLEGγ remain bound in a two-chain state.[11] The non-covalent association of the cleaved AEP pro-domain was suggested to confer stability to the catalytic core domain at neutral pH.[11]
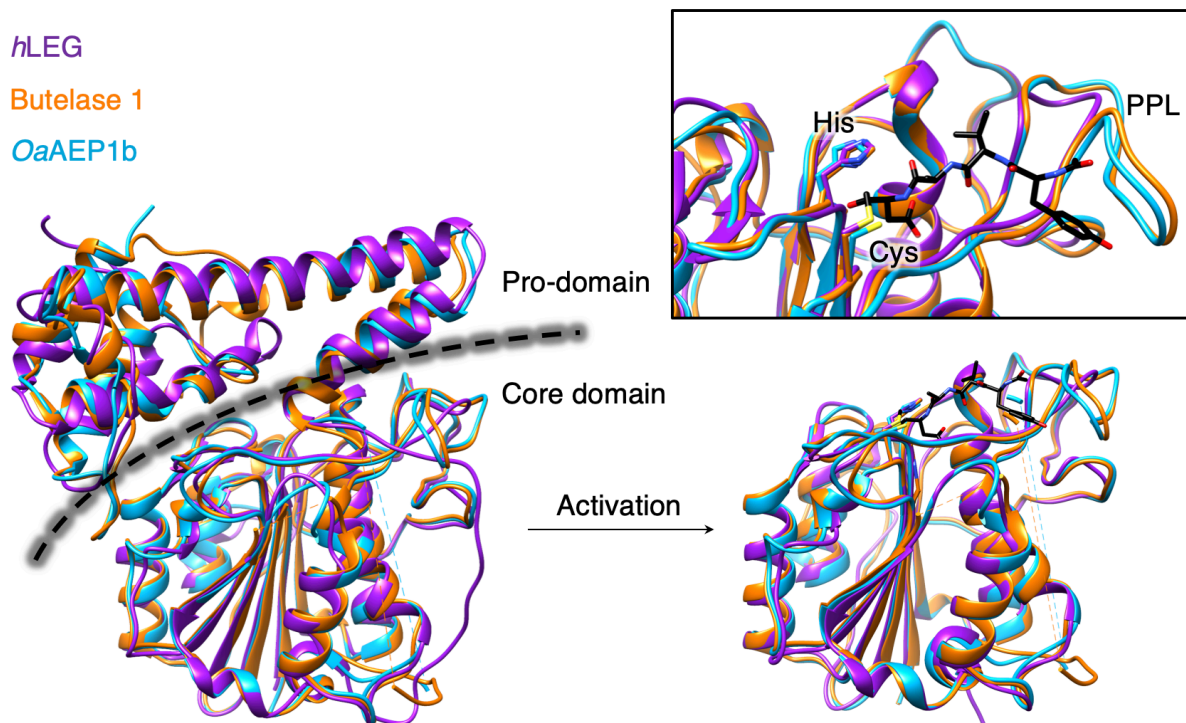
Figure 1.4 Overlaid ribbon representations of human pro-legumain (PDB: 4FGU, purple), butelase 1 (PDB: 6DHI, orange), and OaAEP1b (PDB: 5H0I, blue) structures deduced from X-ray crystallography.[14,35,36] The black dotted line indicates the separation between the pro-domain and the core domain. After activation, modified butelase 1 (PDB: 6DHI) and OaAEP1b (PDB: 5H0I) structures showing only the catalytic core domains were overlaid with the crystal structure of activated human legumain (PDB: 4AWA, purple). In the box, an enlarged image of the active sites, which reveals the respective catalytic diads (Cys and His), the conformation adopted by the peptide-chloromethyl ketone inhibitor (YVAD-CMK) in complex with activated human legumain, and a plant specific loop region referred to as the poly-proline loop (PPL).[26]

The catalytically active core domain of AEPs have been proposed to operate via a mechanism that recruits the catalytic cysteine residue as a nucleophile.The catalytically active core domain of AEPs has been proposed to operate *via* the canonical cysteine protease mechanism.[5,21,35] The catalytic diad (Cys and His) mediates a nucleophilic attack on the amide carbonyl of an internal Asn residue to form a thioester acyl-enzyme intermediate (scheme 1.5).[7,22,23,25,35] The proton transfer between the catalytic diad has been reported to proceed indirectly via either a water molecule or the substrate peptide.[7] For most AEPs, nucleophilic attack from a water molecule completes the peptide bond hydrolysis.[35] For ligase-type AEPs, the N-terminal amine of an incoming peptide adopts the role of the nucleophile to resolve the thioester intermediate (scheme 1.5).[22,23] The nucleophilic amine can be either from the same peptide, resulting in an intramolecular cyclisation,[22,23] or from a different peptide, resulting in an intermolecular ligation (Scheme 1.5).[23,37] Peptide ligation proceeds with a direct nucleophilic attack from the N-terminal amine to the thioester intermediate with no participation of solvent water molecules. A mechanistic study of AEP-mediated peptide ligation employing ^{18}O-labelled water found no isotopic shift in the ligated peptide product.[22,24] This finding indicated that water is likely to be excluded from the active site during peptide ligation by AEPs.

A succinimide motif not commonly seen in other enzymes has been observed by X-ray crystallography in active sites of AEP isoforms hLEG, HaAEP1 and AtLEGγ from human, sunflower and *A. thaliana*, respectively.[25,35,38] The metastable structure resulted from the condensation between the sidechain carboxylate of Asp and a backbone amide NH. Nevertheless, Haywood *et al.* have shown that a mutation of the succinimide forming aspartate residue to an alanine bears no influence on the ligation activity of HaAEP1.[25]
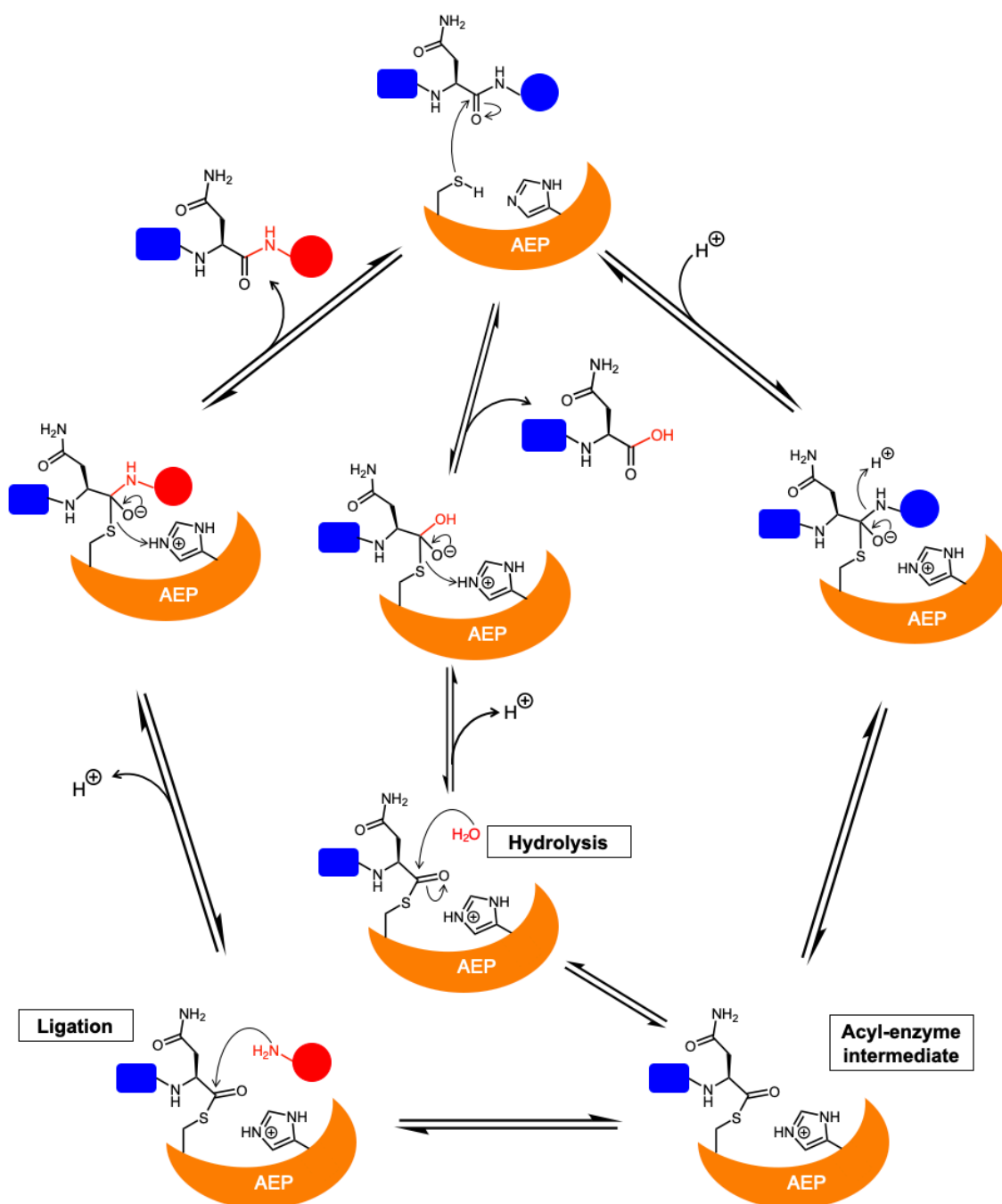
Scheme 1.5 Proposed mechanism of AEPs catalysis with a Cys/His catalytic diad. AEP recognises and cleaves a peptide substrate containing Asn to form an acyl-enzyme intermediate *via* a tetrahedral intermediate. The proton transfer between the catalytic diad has been reported to proceed indirectly via either a water molecule or the substrate peptide.[7] A nucleophilic attack resolves the thioester intermediate to afford the product peptide and the AEP is regenerated. The nucleophile can be a water molecule or the N-terminus of a peptide, resulting in peptide hydrolysis or ligation, respectively.

1.3 Reactivity and substrate recognition of AEP

A comparison of AEPs from the same, or across multiple organisms, reveals that enzymes with ligase activities are uncommon compared to the putative proteases.[4,21–23] Various AEP isoforms exhibit a remarkable difference in activity (protease *vs.* ligase) despite only subtle differences in sequence and structure.[4,26,35,36] Furthermore, the activity of AEPs were reported to be influenced by the amino acid sequence and length of substrate peptides as well as reaction conditions such as pH.[22,23,39]

1.3.1 Protease and ligase-type AEPs

The growing library of AEPs identified from numerous organisms, accompanied by the increasing availability of sequence and structural information, has enabled comparative studies between different AEP isoforms to identify features that influence the preference towards ligase or protease activity.[11,22,23,26,27,35,36,39] Substrate accessibility to the active site of AEPs are governed by a gatekeeper residue, also referred to as ligase-activity determinant 1 (LAD1),[4] has been highlighted as a factor that determines activity (figure 1.6).[14,26] The identity of this residue, with particular emphasis on the size of its sidechain, was suggested to influence substrate access to the active site. OaAEP1b, isolated from *O. affinis*, is an AEP variant which exhibits peptide ligase activity.[22] Mutagenesis studies of OaAEP1b demonstrated that substitutions of the gatekeeper residue (Cys247) to larger residues such as Ile, Leu and Met abolished ligase activity. In contrast, replacements with smaller Ala and Gly residues at the same position resulted in an increase in peptide cyclisation activity. In addition, the Cys247Gly mutation was shown to enhance peptide hydrolysis when compared to the wild-type and Cys247Ala variants.[14] The smaller and less hydrophobic Gly gatekeeper was proposed to allow water molecules to access the active site, resulting in protease activity.

Figure 1.6 Key features associated with the determination AEP ligase and protease activity. Multiple sequence alignment of AEPs, from *O. affinis, P. x hybrida, C. ternatea and H. sapien,* centred around amino acid residues reported to influence activity. Catalytic diads (Cys/His) are shown in white letters with black highlight, N-glycosylation site is highlighted in green, the overlaid red boxes show LAD1 (gate keeper/GK) and LAD2, blue box shows the plant-specific poly-proline loop and the green box highlights the MLA.[4,26] In the box, an enlarged overlaid image of OaAEP1 (PDB: 5H0I),[14] butelase 1 (PDB: 6DHI)[36] and human legumain (PDB: 4AWA)[35] active sites. Key features (catalytic diad, LAD1, LAD2, PPL and MLA) are rendered blue, orange and purple for OaAEP1, butelase 1, and human legumain respectively. In black, the conformation adopted by the inhibitor (YVAD-CMK) in complex with human legumain.

Further investigations into the sequence and structures of different AEP isoforms revealed additional polymorphic positions named ligase-activity determinant 2 (LAD2)[4] and marker of ligase activity (MLA)[26] which are located within, and adjacent to, the

substrate binding pocket, respectively. The MLA was identified by a study comparing the sequences of protease and ligase AEPs from garden petunia (*Petunia x hybrida*) and describes a five amino acid deletion found in AEPs with ligase activity (figure 1.6).[26] The impact of the MLA on the peptide cyclisation activity *in planta* was investigated using tobacco plants (*Nicotiana benthamiana*) which co-expressed transgenes encoding for AEPs and corresponding peptide substrates. Relative cyclisation efficiencies *in planta* were measured by matrix-assisted laser desorption ionisation mass spectrometry (MALDI-MS).[26] The insertion of five amino acid to the MLA region of OaAEP1b, a putative AEP ligase, was shown to significantly impede its ability to mediate peptide cyclisation.[26] However, when the experiments were repeated *in vitro*, the MLA modification was found to only reduce the rate of peptide cyclisation by OaAEP1b rather than a distinct shift to protease activity.[26] The truncation described as the MLA is also not found in the amino acid sequence of butelase 1, an AEP ligase from *C. ternatea* (figure 1.7). Thus, while the MLA offer an indication to the protease *vs.* ligase preference of AEPs, it is not a critical determinant of enzyme activity.

LAD2 describes two amino acid residues lining the substrate binding pocket of AEPs. Mutation experiments performed in an AEP isolated from *Viola yedoensis* (VyPAL3) demonstrated the effect of amino acid substitutions at LAD2. Replacing Tyr with the smaller Gly residue was reported to improve the *in vitro* ligase activities of VyPAL3.[4] Furthermore, a Tyr to Ala substitution at LAD2 of a protease AEP from *Viola canadensis* (VcAEP) was shown to shift enzyme activity from protease to ligase.[4] However, the introduction of protease favouring residues at LAD positions did not switch the activity preference of *Vy*PAL2, an AEP ligase from *V. yedoensis*, to favour protease activity.[4]

None of the identified structural features were found to provide absolute control over the protease vs. ligase preference of AEPs. Nevertheless, mutations at gatekeeper, MLA and LAD2 have all been shown to significantly impact AEP activity.[4,26] These findings suggest that key structural features may operate synergistically to dictate AEP activity, and further investigations are required to fully establish the key determinants of AEP protease and ligase activity.

1.3.2 Substrate recognition

Amino acid residues surrounding the peptide cleavage site for substrates of AEPs are commonly labelled following the numbering conventions adopted by Schechter and Berger for substrates of proteases (scheme 1.8).[40] From the cleavage site to the peptide N-terminus, amino acid residues are numbered in ascending order starting from P1. On the other side of the cleavage site, amino acid residues are denoted with a single prime and numbered in ascending order towards the C-terminus starting with P1'. For peptide ligation, the amino acid positions on the incoming peptide are labelled with a double prime and numbered in ascending order from the N-terminus starting with P1''.
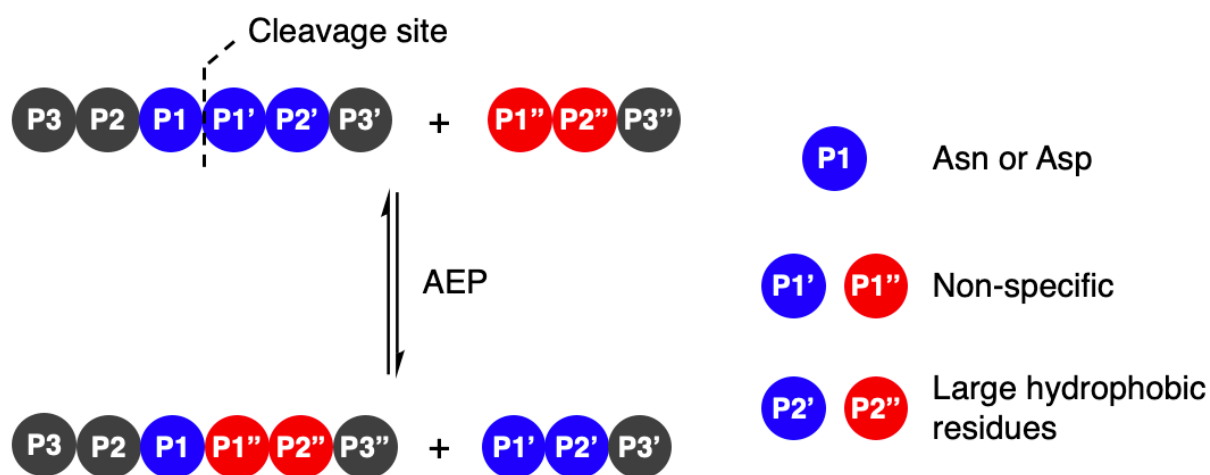
Scheme 1.7 AEP-mediated transpeptidation, with amino acid residues numbered according to Schechter and Berger.[40] The putative recognition sequences P1-P1'-P2' and P1''-P2'' are coloured in blue and red, respectively.

Initial studies performed with an AEP isolated from castor beans have identified strict substrate specificity towards Asn at the P1 position.[1,2,8,18,22,23] The understanding of substrate recognition by AEPs have subsequently expanded with experiments in AEPs isolated from *C. ternatea* (butelase 1),[23,41,42] *O. affinis* (OaAEP1b and OaAEP3–5)[22,39] and other plant species.[24,27] Substrate recognition by AEPs were shown to be determined mostly by amino acids at positions P1, P1' and P2' (scheme 1.8). While many AEPs have a strict preference towards Asn at P1, some isoforms such as

HaAEP1 and OaAEP1b were found to also accept Asp at P1.[22,24] The preferences toward P1' and the ensuing P1'' from the nucleophilic peptide for ligation are highly non-specific. A peptide ligation assay which employed a library of peptides, with varying amino acids at the N-terminus, demonstrated that butelase 1 accepts all canonical amino acid residues at P1'' with the exception of Pro.[23] Furthermore, butelase 1 was reported to cyclise peptides with D-amino acids at P1'' and recognise peptides with a thioester linkage between the P1 and P1' residues.[41,42] AEPs from *O. affinis* (OaAEP1b, OaAEP3–5) have been shown to process peptides with several different amino acid residues at P1'.[22,39,43] In contrast, recognition at P2' and P2'' demonstrate greater specificity. While peptides with His or Ala at P2' were not processed by OaAEP1b, peptide cyclisation was detected when substrates with Leu or Phe at P2' were used.[22,39] Furthermore, butelase 1 was reported to ligate peptides with large, hydrophobic side chains such as Leu, Ile and Val at P2''.[23]

Although most investigations towards the substrate specificity of AEPs have focused on the putative recognition sequences at P1-P1'-P2' and P1''-P2'' (scheme 1.8), some studies suggested enzyme activity can be affected by other positions in the substrate peptide.[22,27,44] A substitution of Arg with Ala at the P2 position was found to significantly reduce the rate of OaAEP1b-mediated peptide cleavage.[22] Moreover, the use of Pro at P2 in a short peptide (5–9 amino acids) was shown to favour intermolecular ligation over intramolecular cyclisation by butelase 1. Intermolecular ligation of short peptides followed by head-to-tail cyclisation results in the formation of cyclo-oligomers (scheme 1.8 A).[44] Notably, it was highlighted that butelase 1 was not able to recognise the corresponding cyclo-oligomeric peptide products despite the presence of a suitable recognition sequence.[44] Since AEP catalysis is not known to involve co-factors or other external sources of energy, the irreversible process may be rationalised by conformational stabilisation of the cyclised peptide and an entropic gain from the release of the cleaved recognition peptide motif.
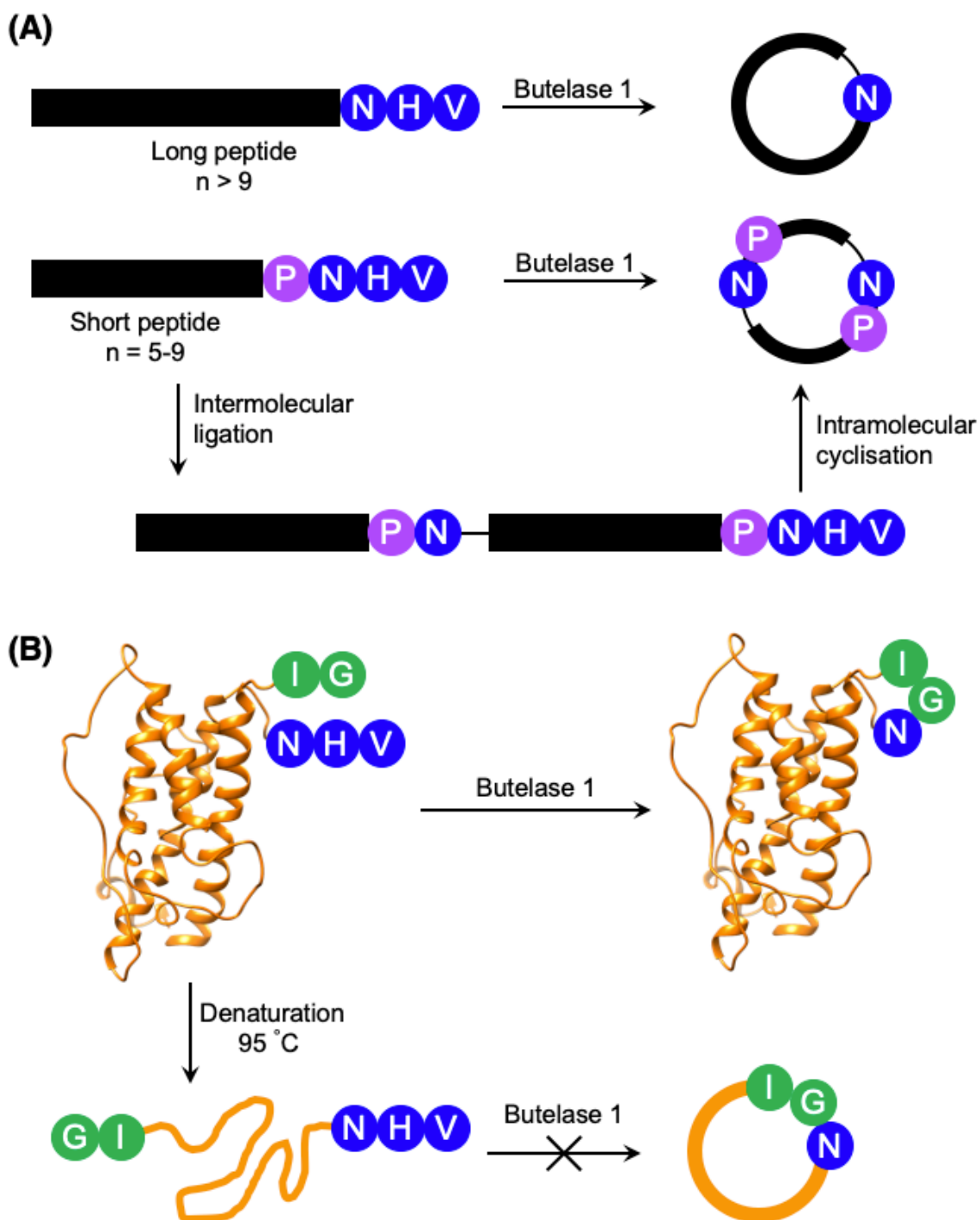
Scheme 1.8 Peptide cyclisation catalysed by butelase 1. **(A)** Peptides greater than 9 amino acid residues in length and bearing the butelase 1 substrate recognition sequence (NHV) are cyclised by the AEP. Whereas peptides of 5-9 amino acid residues in length, with Pro at P2 and bearing the butelase 1 substrate recognition sequence (NHV) are ligated, then cyclised by the AEP to afford cyclic oligomers.[44] (only cyclo-dimer shown here) **(B)** Butelase 1 was reported to cyclise the folded Somatropin (PDB: 1HGU), but not the denatured protein.[45]

The proximity of the N and C-termini was shown to influence the efficiency of AEP-mediated cyclisation of larger peptide and protein substrates. While butelase 1 was reported to effectively cyclise recombinantly prepared interleukin-1 receptor antagonist (IL-1Ra) and human growth hormone (somatropin), no cyclic proteins were detected when IL1-Ra and somatropin were denatured prior to incubation with the enzyme (scheme 1.8 B).[37,45] A similar effect can also be observed in OaAEP1-mediated protein cyclisation of an intrinsically disordered protein, merozoite surface protein 2 (MSP2), which proceeds in low yield.[39]

1.4 Rate of peptide cyclisation by AEPs

The reported kinetic parameters for peptide cyclisation catalysed by various AEPs have been summarised in Table 1. As investigations have conducted peptide cyclisation assays with different reaction conditions and substrates, it is difficult to draw direct comparisons. Nevertheless, the catalytic efficiencies of AEPs are relatively high compared to sortase A, a widely used transpeptidase with catalytic efficiency (k_{cat}/K_M) values ranging from 0.02 to 2.80×10^4 $M^{-1}{\cdot}s^{-1}$ for various engineered variants.[46–48]

With k_{cat}/K_M up to 131×10^4 $M^{-1}{\cdot}s^{-1}$, butelase 1 extracted from *C. ternatea* is one of the fastest peptide ligases reported in the literature.[23,37] The enzyme was reported to cyclise >95% of linear substrate peptide within 1 h at 37 ºC (0.125 µM butelase 1 and 50 µM peptide).[23] On the other hand, OaAEP1b exhibits a lower k_{cat}/K_M than butelase 1.[22] However, by replacing the gatekeeper (Cys247) near the active site with a smaller Ala residue, the engineered protein was shown to cyclise a model substrate peptide with k_{cat}/K_M value of 3.42×10^4 $M^{-1}{\cdot}s^{-1}$ at pH 6.0 and 37 ºC.[14]

Subsequently, a suite of AEP ligases from *O. affinis* (OaAEP3, OaAEP4 and OaAEP5) have been identified.[39] OaAEP4 was reported to cyclise >95% of linear substrate peptide within 30 min at room temperature (0.185 µM AEP and 280 µM peptide) and good tolerance to organic co-solvents (acetone, dimethylformamide and methanol). In directly comparable experiments, OaAEP3, 4 and 5 all demonstrated greater k_{cat}/K_M than the prototypic OaAEP1b at pH 5.0 and room temperature. Further differences in the activity of AEPs from *O. affinis* were shown when the peptide cyclisation assay

was performed at varying pH values. OaAEP4, the most active among all of the tested AEPs, offered optimal conversion of the linear peptide at pH 4.2 and remained active over a relatively narrow pH range (3.7-5.0). In contrast, OaAEP1b demonstrated optimal activity at pH 5.0 with a broader tolerance to pH compared to OaAEP4 (4.2-7.5). A 50% conversion to the cyclic peptide was detected when linear peptide substrates were incubated with OaAEP1b for 30 min at pH 7.5 (0.528 µM AEP and 280 µM peptide).[39]

Table 1.1 Kinetic parameters of AEP for peptide backbone cyclisation

Enzyme	Origin	Recognition motif	k_{cat} / s^{-1}	k_{cat}/K_M / 10^4 M^{-1} s^{-1}	Assay conditions		Ref.
					Temperature / °C	pH	
Butelase 1	*C. ternatea*	GI-X$_n$-NHV	26.55	131	42	6.0	23,45
OaAEP1b	*O. affinis*	GL-X$_n$-NGL	0.99	0.68	r.t.	5.0	39
OaAEP1-C247A	*O. affinis*	GL-X$_n$-NGL	13.9	3.42	37	6.0	14
OaAEP3	*O. affinis*	GL-X$_n$-NGL	0.61	33.0	r.t.	5.0	39
OaAEP4	*O. affinis*	GL-X$_n$-NGL	0.76	98.4	r.t.	5.0	39
OaAEP5	*O. affinis*	GL-X$_n$-NGL	0.59	24.7	r.t.	5.0	39
CeAEP1	*C. ensiformis*	GR-X$_n$-NGL	N.R.	0.58	37	5.0	24
MCoAEP2	*M. cochinchinensis*	GG-X$_n$-NAL	19.86	62.1	22	6.0	27

N.R. – Not reported. r.t. – room temperature

1.5 Applications of AEP ligase activity

The growing importance and diverse applications of protein science in contemporary research encompasses a range of research topics including protein labelling for the study of post-translational modification;[49–52] protein-drug conjugation for the production of biopharmaceuticals;[53–55] peptide ligation for the semi-synthesis of small

proteins;[56] bioconjugation to solid supports for the immobilization of proteins.[57,58] The ability to facilitate peptide ligation and backbone cyclisation *via* a relatively short substrate recognition sequence, with good catalytic efficiency and tolerance to a broad pH range highlight the significant potential of AEPs in applications.

1.5.1 Intramolecular ligation by AEP results in backbone cyclisation

Cyclisation often reduces conformational flexibility and imparts greater thermal and chemical stability.[30,39,45,59] The joining of a peptide's termini can provide resistance towards degradation by exopeptidases and improve the oral bioavailability of peptide products.[29,30,45,60] Chemical approaches towards backbone cyclisation, such as native chemical ligation, are limited by strict requirements for N-terminal cysteines (or Cys-surrogates), C-terminal thioesters and denaturing conditions (scheme 1.9 A).[59,61,62] Similarly, the introduction of large split intein domains often impair solubility which limits the versatility of *trans*-protein splicing strategies.[63,64] Cyclisation *via* small synthetic molecules have been employed to generate diverse libraries of cyclic peptides (scheme 1.9 B).[65] Chemical linkers with two thiol reactive functional groups were reported to bridge across two cysteine residues to afford cyclic peptide species and allow the insertion of non-native motifs.[65] However, this "chemical bridging" approach exploits the thiol reactivity of cysteine residues, which have limited selectivity when applied to cysteine-rich protein constructs.[65] Moreover, the termini of bridged cyclic peptides remain unmodified, and thus susceptible to degradation by exopeptidases. The use of peptide ligases offers an alternative approach to peptide cyclisation. With excellent reaction kinetics and a short recognition sequence that is orthogonal to the existing enzymes such as the prolyl oligopeptidase, PCY1,[66] and transpeptidase, sortase A,[59,67] the addition of AEPs complements existing strategies.

Peptide backbone cyclisation has been identified as the native function of AEP ligases isolated from plant species which produce cyclic peptides such as *O. affinis* (OaAEP1b),[21,22] and *C. ternatea* (butelase 1).[23] In nature, AEP ligases process the linear peptide precursors to afford the cyclic product which are typically around 30 amino acids in length.[60] Native and engineered AEPs derived from various plants have also been successfully employed to facilitate backbone cyclisation of peptides derived

from other plant and non-plant species such as kalata B1;[21–23] sunflower trypsin inhibitor 1;[23,41,68] Histatin-3;[23] anti-malarial peptide R1;[39] MCoTI-II;[27] AS-48,[68] and proteins such as GFP;[14,45,69] somatropin;[45] MSP2;[39] p53 binding domain.[70] Butelase 1 was reported to generate the cyclic antimicrobial peptide θ-defensin and the cyclic conotoxin MrIA with >95% yield in 1 minute at 42 °C and pH 6.0.[41]
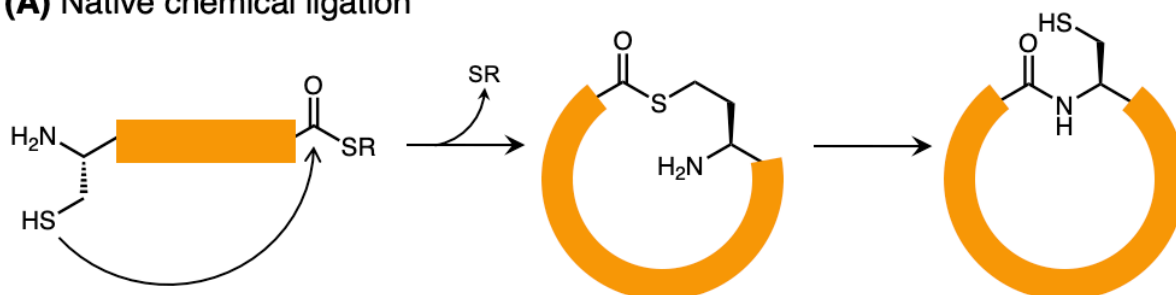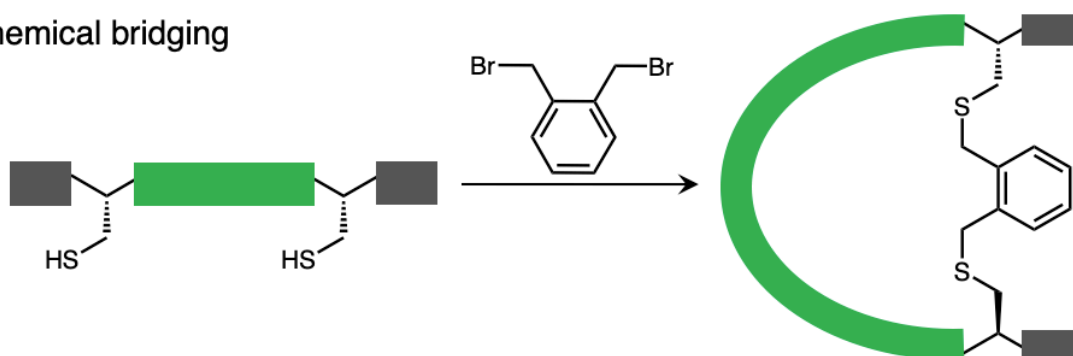
Scheme 1.9 Chemical approaches to peptide backbone cyclisation by **(A)** native chemical ligation and **(B)** chemical bridging

Cyclotides, backbone cyclised and cysteine knotted peptides, have been touted as a potential scaffold for the development of novel therapeutics.[27,29,45] The emergence of convenient and effective peptide cyclisation methods by AEPs have enabled the synthesis of grafted peptides, whereby bioactive peptides are inserted into cyclotides (figure 1.10).[71] Exemplified by the cyclisation of MCoSST-01 using MCoAEP2,[27,72] this approach employs the highly stable cyclic cysteine knot architecture from cyclotides as a scaffold to deliver peptide pharmacophores.
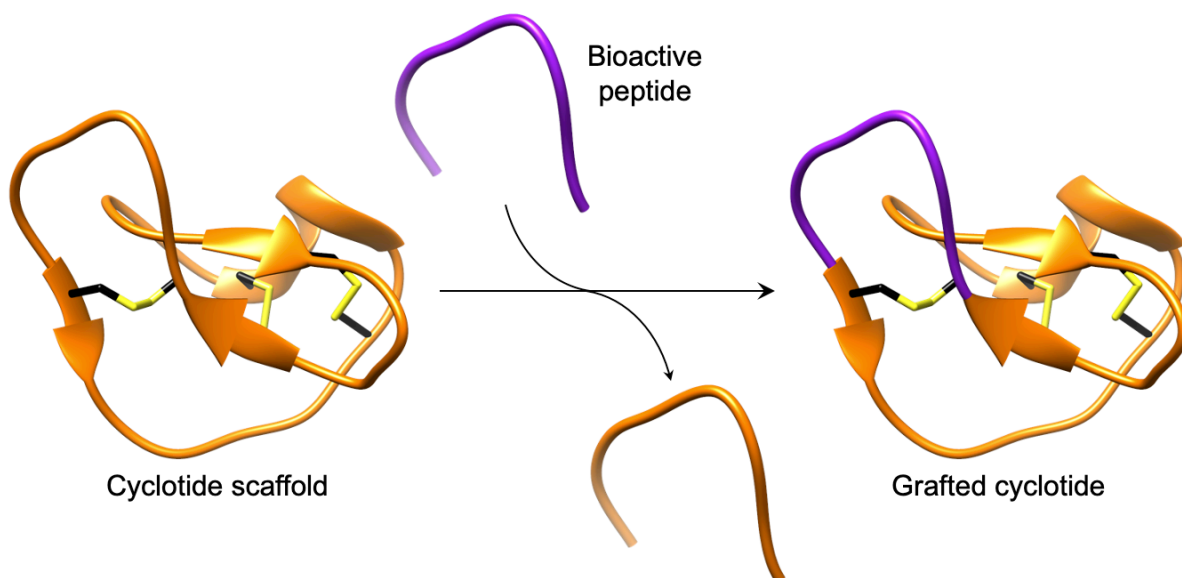
Figure 1.10 Cartoon representation of a grafted cyclotide. A bioactive peptide is inserted into a cyclotide backbone, shown here using the structure of MCoTI-II (PDB: 1IB9).

Applications of AEPs in protein research was demonstrated by the butelase 1-mediated cyclisation of the p53 binding domain, a potential target for anti-cancer therapeutics. The cyclic protein was reported to aid structural investigations by NMR and X-ray crystallography, methods that may be hindered by conformational flexibility induced line broadening and poor crystal formation, respectively.[70] Furthermore, greater conformational stability in the cyclic protein was suggested to offer improved ligand binding due to a reduction of the associated entropic penalty.[70] AEP-mediated protein cyclisation was also applied to MSP2 from *Plasmodium falciparum*, a 25 kDa protein for malarial vaccine development.[39] When produced recombinantly, the protein is disordered and its antigenic behaviour no longer represents the native protein.[73–75] OaAEP1b-mediated cyclisation of MSP2 was suggested to reduce disorder and potentially force the protein into adopting a conformation similar to the native antigen.[39]

Further developments in AEP-based technology can also be found in a recently published article that explored the effects of enzyme immobilisation (figure 1.11).[68] The study illustrated that AEPs from *C. ternatea* (butelase 1) and *V. yedoensis* (VyPAL2) which were conjugated to a solid support offered greater stability in storage and repeated use. Significantly, immobilisation enabled the use of these enzymes at

higher concentrations, and to perform peptide cyclisation reactions in continuous flow (figure 1.11 C).[68] The findings reported by Hemu *et al*. represent an advance towards adapting AEP catalysis for large scale and industrial applications.[58]

(A) Immobilisation by NeutrAvidin-biotin binding
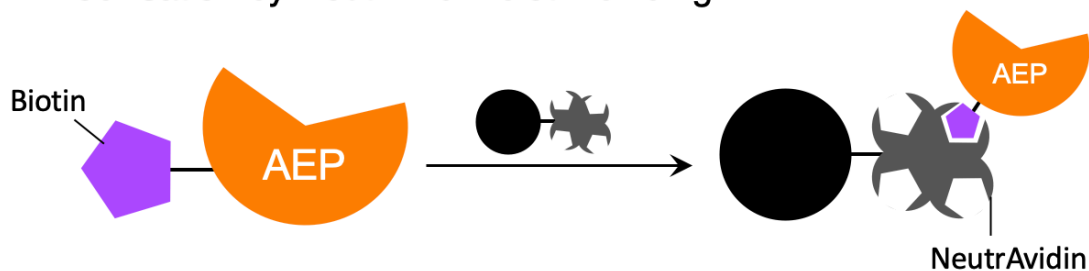
(B) Immobilisation by coupling to NHS-ester
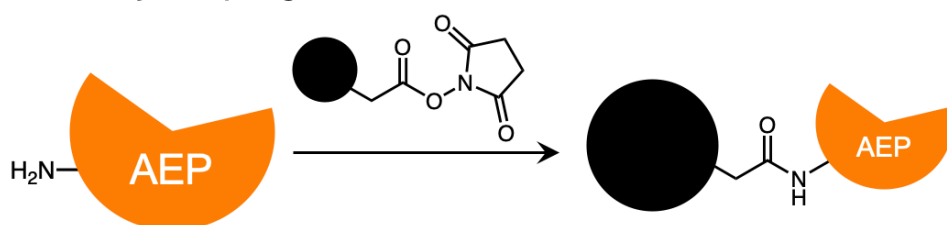
(C) AEP catalysis in continuous flow
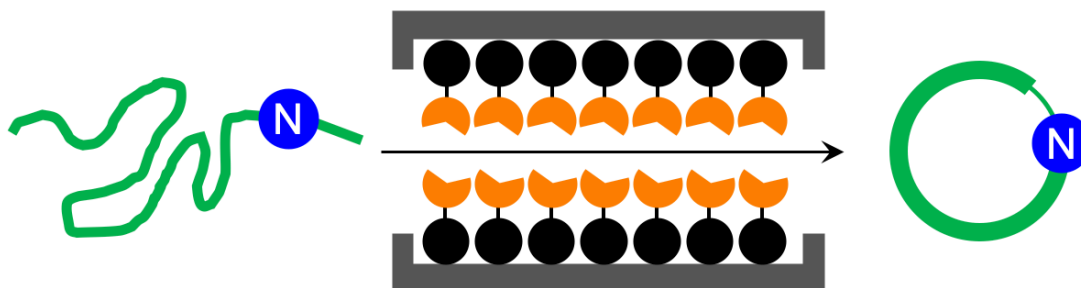
Figure 1.11 Immobilisation of AEPs on solid support. AEPs, butelase1 and VyPAL2, have been immobilised on agarose beads *via* **(A)** NeutrAvidin-biotin affinity binding and **(B)** direct coupling to N-hydroxysuccinimide (NHS) ester with primary amines presented by lysine residues. **(C)** The immobilised AEPs were reported to facilitate ligation of proteins and peptides in continuous flow.

1.5.2   Intermolecular ligation by AEP for peptide and protein bioconjugation

Naturally occurring proteins commonly feature non-ribosomal modifications; glycosylation, methylation, and lipidation can influence protein stability, localisation and interactions with other biological components.[50,52,76–78] Protein modifications are routinely employed to study and engineer proteins beyond their native functions for

academic research in addition to medicinal and industrial applications.[51,79,80] Consequently, the development of novel polypeptide ligation technologies have been a growing theme in biological and organic chemistry research.[47,48,54,81–96]

Currently there is a range of chemical and biological approaches to bioconjugation.[47,48,54,81–96] Chemical approaches tend to be more affordable, simpler to perform, but can be hindered by poor selectivity or the availability of specific reactive handles.[90,91] Strategies employing enzymes offer an attractive alternative, whereby site-specific reactions can be catalysed under relatively mild conditions.[92,93] Transpeptidase from Gram-postitive bacteria (Sortase A),[47,48,94] and other enzymes such as subtiligase,[81,95,96] biotin ligase,[82,83] transferases[84–86] and formyl glycine generating enzymes[54,87–89] have been employed for protein bioconjugation. Excellent kinetic parameters and relatively short recognition sequences indicate that AEP methodologies would be a valuable addition to supplement the existing approaches. AEPs have been employed to modify protein substrates including GFP,[42,43,97] ubiquitin,[37,42,98] ompA,[99] DARPin,[98] maltose binding proteins,[43,100] and nanobodies[69] with synthetic labels such as unnatural amino acids, click handles, modified residues, polyethylene glycol, fluorophores, biotin and drug molecules.

Introduction of an isotopically labelled segment into protein backbone by OaAEP1b-mediated ligation was shown to enable NMR studies of maltose binding protein.[43] The reported method uses a nicked protein, or self-assembling protein domains, to ensure close proximity between the ligating motifs, one of the factors that affects AEP-mediated ligation (figure 1.12).[43,97] While preparation of labelled proteins by recombinant gene expression is routinely performed,[101–103] this enzymatic approach offers the ability to selectively label specific regions within the protein. In addition, segmental isotopic labelling of proteins may find other applications such as elucidating protein dynamics.[104]
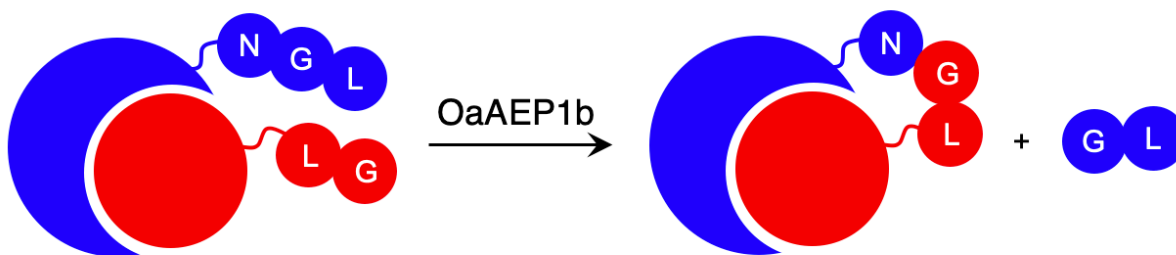
Figure 1.12 Ligation of self-assembling protein domains by OaAEP1b. Pre-organisation of the ligating substrates bring the reactive sites into close proximity, which was shown to affect AEP catalysis

Enzymes commonly target specific substrate recognition motifs defined by the active site. While AEPs have relatively non-specific recognition at certain positions (P1'-P2' and P1''-P2''), changes in the recognition sequence have been shown to influence ligation efficiency.[22,23,27,39,69] It was reported that although Gly-Val was accepted by OaAEP1 at the P1''-P2'' positions for ligation, the corresponding product (Asn-Gly-Val at P1-P1'-P2') was poorly recognised for further enzymatic reactions.[69] This differential recognition of Val at P2' *vs.* P2'' was utilised to construct a protein bearing two different modifications at the N- and C-terminus mediated by a single AEP.[69] Butelase 1 has also been employed in conjunction with sortase A to modify immunoglobulin molecules (IgG1) with two fluorescent tags (5,6-carboxyfluorescein and AlexaFluor 647) a one-pot dual labelling reaction.[105] These reports highlight the potential use of AEPs in the semi-synthesis of more complex, multiply labelled protein such as antibody-drug conjugates.

Finally, the reported application of butelase 1 for protein labelling on the surface of live bacteria exemplifies the biocompatibility of enzyme catalysis.[99] The outer membrane protein A (OmpA) of *Escherichia coli* strain BL21(DE3), modified with the recognition sequence of butelase 1 at the C-terminus, was successfully labelled with tags such as 5,6-carboxyfluorescein, biotin and the fluorescent protein mCherry.[99] The ability to selectively modify proteins on the cell surface can offer an insight towards signalling mechanisms between cells and the extracellular matrix as well as intercellular communications.

1.6 Limitations of AEP catalysis

1.6.1 Preparation of recombinant AEP

While the utility of AEPs have been demonstrated through a number of examples, access to active enzyme has hindered the use of AEPs in biocatalytic applications. Butelase 1 was obtained *via* an arduous extraction and purification process from the native plant host, butterfly pea, for initial experiments.[23,37] Subsequently, procedures for the preparation of recombinant butelase 1 from bacteria (*E. coli*)[36] and yeast (*Pichia pastoris*)[70] were reported. Similarly, the preparation of recombinant OaAEP1b from *E. coli* culture have been reported in a number of studies.[14,22,39,69]

Mimicking the production and maturation of AEPs in nature, the recombinant AEPs are produced as a zymogen which can be activated at low pH (pH 3.6–4.0). The acidic environment simulates the plant vacuole, where active AEPs are known to accumulate in cyclotide-producing plant species.[11,22,24,25] The gene construct employed to prepare recombinant OaAEP1b encodes for four distinct features, His_6 tag, ubiquitin, AEP core domain and pro-domain, producing a His_6-ubiquitin-AEP fusion protein. Located at the N-terminus, the His_6 tag facilitates purification by immobilised metal affinity chromatography (IMAC). Between the His_6 tag and AEP, the ubiquitin sequence has been employed to enhance solubility and expression levels of OaAEP1b and other AEPs.[22,27] The pro-domain, a C-terminal section of the zymogenic AEP blocks substrate access to the active site, and thus silences enzyme activity. The pro-enzyme was isolated by IMAC, strong anion exchange (SAX) and size exclusion chromatography (SEC). Incubation at low pH (pH 4.0) was required to trigger the auto-proteolytic activation process (figure 1.13). In some studies, the activated AEP was subjected to a further purification step by ion-exchange chromatography.[22,39,100] The reported yield of the activated OaAEP1b ranged between 1.8 and 2.0 mg/L from lysogeny-broth (LB) media.[22]

The current procedures for the preparation of AEPs involve multiple purification steps and an incubation step at low pH. The time consuming and low yielding aspects of the

procedure present a considerable limitation to the uptake of AEPs for bioconjugation applications.
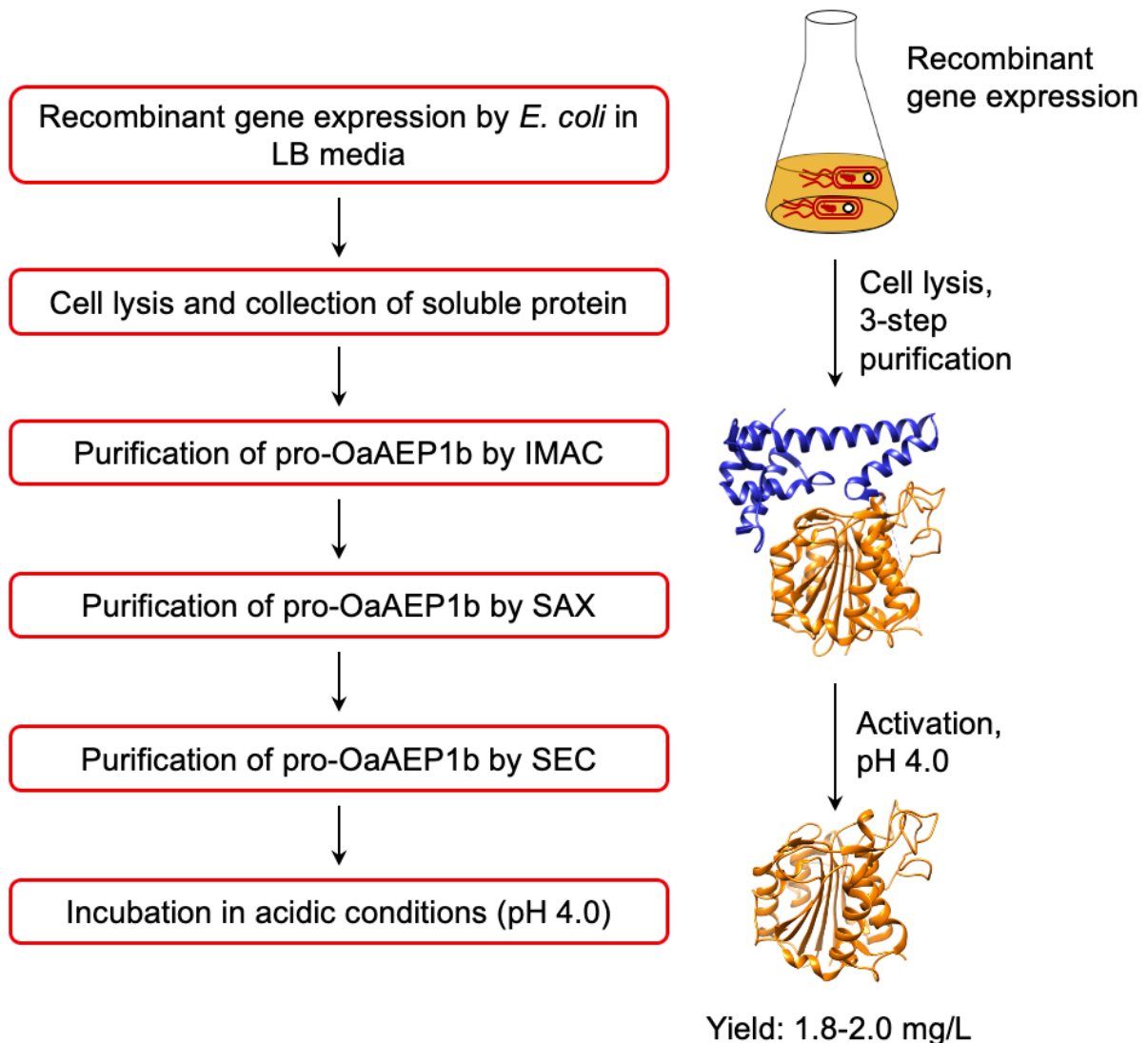
Figure 1.13 Procedure for the preparation of recombinant OaAEP1b (PDB: 5H0I) involves multiple purification steps and an incubation period in acidic conditions (pH 4.0). (IMAC) Immobilised metal affinity chromatography, (SAX) strong anion exchange and (SEC) size exclusion chromatography.

1.6.2 Equilibrium in AEP-mediated peptide ligation

The ligation activity of AEPs have also been employed to facilitate intermolecular ligation at either the N or C-terminus of proteins and peptides (see Chapter 1.4.2). While AEP mediated intermolecular ligation has been successfully employed in some cases, it operates with limited efficiency. Similar to other transpeptidases such as sortase A and trypsiligase,[94,106–108] the challenge in preventing efficient AEP-mediated

intermolecular ligation is that the by-product dipeptide (P1'-P2') acts as a competing nucleophile to reverse the ligation reaction.[42,69] The efficiency of AEP-mediated ligations is affected by the proximity of the ligation partners and the rate of undesired reverse ligation.[37] Unlike intramolecular cyclisation, the ligating termini are unlikely to be in proximity. Moreover, the entropic penalty from intermolecular ligation negates any entropy gained from the release of the cleaved recognition peptide motif. Consequently, a large excess of the labelling peptide nucleophile is required to drive the ligation reaction to completion.[37,42,69,98,105] The labelling of a nanobody protein using OaAEP1-C247A employed 25 equivalents of a fluorescent peptide label (250 µM), with respect to the protein (10 µM).[69] This rendered transpeptidase-mediated ligations uneconomical, especially when expensive or non-commercially available reagents are used (e.g. radioactive, isotopic or fluorescent labels).[43,97,109,110]

Attempts to address this limitation were made by exploiting the non-specific substrate recognition nature of AEPs.[23,42,43,69,97,98] Butelase 1 has been reported to exhibit considerable reactivity towards synthetic peptide analogues bearing ester and thioester linkages (depsipeptide and thiodepsipeptide, respectively).[42,98] This was exploited to develop an irreversible butelase 1-mediated ligation strategy (scheme 1.14 A).[42] While the depsipeptide approach has been employed to ligate a synthetic peptide to the N-terminus of ubiquitin and green fluorescent protein (GFP) with good yields in 2.5 h, the asparaginyl thiodepsipeptide motif have a short half-life of around 45–75 min at pH 6.5 (scheme 1.14 B).[42,61] Owing to the hydrolysis of thiodepsipeptides, a five-fold excess of this unstable reagent was required to effectively label the target protein substrates.[42] The use of an unnatural thioester motif further limits the application of this method to the labelling of protein N-termini.[42]
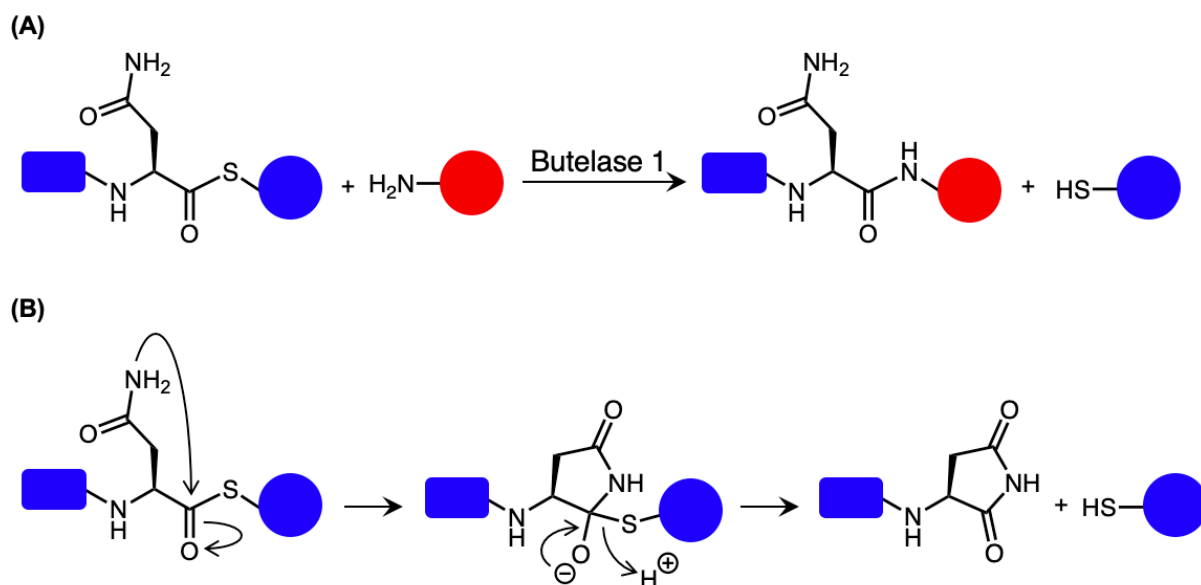
Scheme 1.14 Use of thiodepsipeptide for butelase 1-mediated ligation **(A)** Thioester linked Asn is accepted by butelase 1. Peptide ligation generates a peptide by-product with a α-thiol which is not recognised as a nucleophilic peptide substrate. As a result, the reaction was rendered irreversible. **(B)** Proposed mechanism for the hydrolysis of the unstable thioester.

Without using unnatural (thio)depsipeptide motifs, the reverse reaction mediated by OaAEP1-C247A can be prevented by taking advantage of the subtle difference in reactivity between two recognition sequences. Valine was reported to be an amino acid which is recognised at the P2'' position, but poorly at P2'. Thus, Gly-Val has been used as a recognition sequence on the nucleophilic peptide to generate a product peptide that is not readily hydrolysed by the enzyme (figure 1.15).[69] Nevertheless, a large excess of the nucleophilic peptide (20-fold excess) was required to obtain the ligated product in good yields.[69]
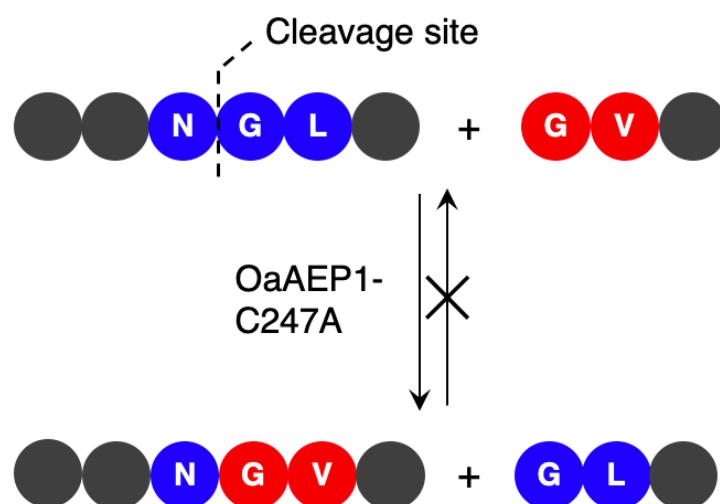
Scheme 1.15 OaAEP1-C247A recognises Val at P2'', but not at P2'. Consequently, this feature in the substrate recognition of OaAEP1-C247A has been exploited to prevent the undesired reverse reaction.

Recently, a nucleophile quenching strategy was reported to enable effective protein bioconjugation by OaAEP1-C247A at near-equimolar concentrations of label and protein substrates.[100] By employing a N-terminal tripeptide motif with a His residue at the third position, the peptide by-product from the OaAEP1-C247A reaction can be sequestered by divalent cations such as Ni^{2+} (figure 1.16). Quenching of the nucleophilic by-product prevents the reverse peptide ligation by OaAEP1-C247A, thus driving the reaction equilibrium towards product formation.[100] However, Ni^{2+} and more commonly Fe^{2+} are employed by enzymes as metal cofactors which are critical to activity.[111] For example, urease from *Helicobacter pylori* utilises a Ni^{2+} catalytic centre to generate ammonia.[112] Consequently, the use of transition metal ions in this approach may not be suitable for labelling enzymes which naturally bind metals such as Ni^{2+} or Fe^{2+}.

Finally, the incorporation of self-assembling protein domains such as nicked-maltose binding protein or designed armadillo repeat protein was the only published approach which enabled effective conjugation without using an excess of labelling reagent.[43,97] This method employed protein self-assemblies to bring the ligation partners into close proximity, and thus improve the ligation efficiency of OaAEP1b (figure 1.12). However, the insertion of a large protein domain into the substrates limits the versatility of this

approach. Therefore, a versatile method to enable effective AEP-mediated intermolecular ligation would unlock the tremendous potential of AEPs to a range of applications in protein sciences.
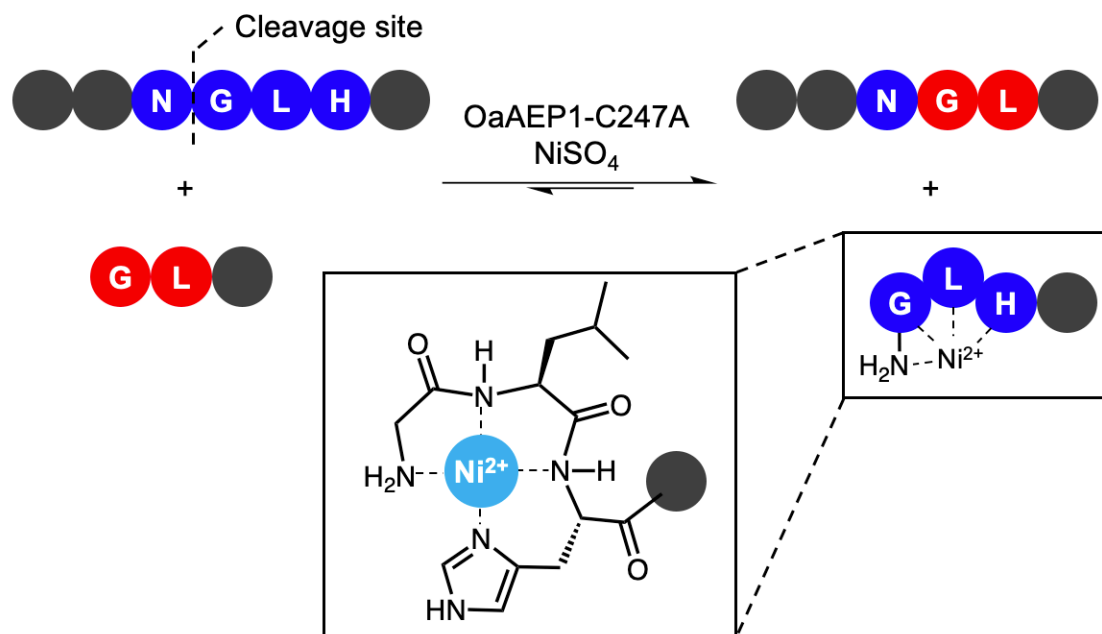
Figure 1.16 Divalent Ni^{2+} quenches the nucleophilic peptide by-product generated by OaAEP1-C247A, thus driving the reaction equilibrium towards product formation. **(Box)** The tripeptide motif Gly-Leu-His at the N-terminus coordinates to Ni^{2+}

1.6.3 Toxicity limits *in vivo* applications of AEP

The applications of AEP catalysis were discussed in Chapter 1.4. While some experiments have been performed *in planta*, AEPs are mostly employed *in vitro*. The relatively short and non-specific substrate recognition of AEPs increases the probability of a compatible substrate sequence being present in proteins. This can lead to undesired side reactions and limit the development of *in vivo* applications. In nature, the broad substrate scope of AEPs are controlled by spatial segregation in organelles. Activated AEPs naturally accumulate in the vacuole of plant cells and they are associated with vacuole collapse and protein degradation during programmed cell death.[2,9,11,13,14] In the absence of organelles in prokaryotic recombinant host cells such as *E. coli*, the accumulation of active AEPs in the cytoplasm can be toxic.

Encapsulation is a potential strategy to enable cellular applications of AEPs. Compartmentalisation is a ubiquitous feature among biological systems and has become increasingly popular in the field of metabolic pathway engineering.[113–115] The encapsulation of biochemical processes within lipid membrane or protein bound compartments offer considerable advantages, these include control over unwanted side reactions, containment of toxic intermediates and reduction in diffusion distances.

While the presence of membrane bound organelles is a hallmark of eukaryotes, compartments derived from multimeric protein assemblies can be found in bacteria, archaea and viruses.[116–118] Various protein compartments including shell proteins which form viral particles,[119,120] ferritin,[121] carboxysome[122,123] and lumazine synthase[124–126] have been studied. Self-assembling protein compartments have numerous roles in nature, ranging from the regulation of enzyme activity in carboxysomes[122,127] to storage and transport of iron by ferretin.[121] The carboxysome is one of the earliest bacterial microcompartments to be identified (figure 1.16 A), where the polyhedral shell protein was found to encapsulate two successive enzymes within the carbon fixation pathway of cyanobacteria.[117,122,123] The co-localization of carbonic anhydrase (CA) and ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) in an enclosed environment limits diffusion of CO_2, the pathway intermediate (figure 1.17 A). Consequently, high local concentrations of CO_2 drives the carboxylation of ribulose-1,5-bisphosphate (RBP) by RuBisCo to afford 3-phosphoglycerate (3-PGA).[123,128,129] The diverse applications and programmability of protein-based capsids have attracted significant interest from biotechnology and synthetic biology reserach.[130–133] The capacity to modify proteins by genetic as well as chemical methods present a major advantage over lipid-based microcompartments.[130–133]

Lumazine synthase (LS) is another protein which forms microcompartments consisting of identical protein subunits and encapsulate riboflavin synthase (RS).[126,134] Together, the protein complex catalyses the final two reactions in the biosynthetic pathway of riboflavin (figure 1.17 B).[135] Intimate association of these enzymes enables substrate channelling and recycling, which enhances the catalytic process, particularly at low

substrate concentrations.[135,136] LS has been shown to adopt several quaternary states with complexes ranging from pentamers to large icosahedral assemblies of pentameric subunits.[124,126] LS from *Aquifex aeolicus* (AaLS) forms triangulation number T = 1 or T = 3 icosahedral capsids composed of 12 or 36 pentamers, respectively.[125] A 12-amino acid peptide sequence located at C-terminus of RS from *A. aeolicus* (AaRS) was shown to mediated its selective association to AaLS. Foreign guest molecules such as GFP appended with the peptide tag was reported to be sequestered by recombinant AaLS capsids in *E. coli*.[137]
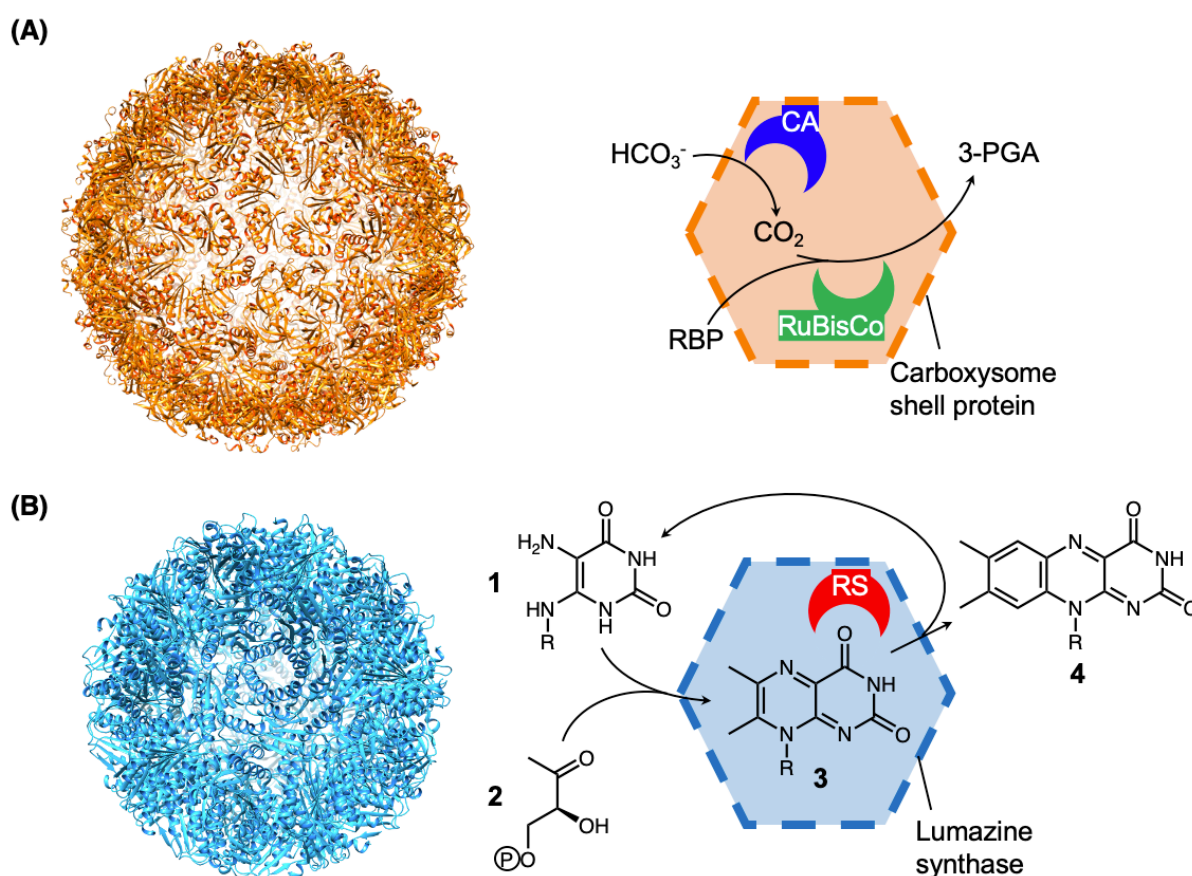
Figure 1.17 The carboxysome is a bacterial microcompartment which facilitates carbon fixation. **(A)** β-carboxysome shell (PDB: 6OWG),[138] and a schematic of 3-phosphoglycerate production from carbonic acid and ribulose-1,5-bisphosphate in a carboxysome. **(B)** Lumazine synthase (PDB: 5MPP), and a schematic showing the biosynthesis of riboflavin (4) from 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione (1) and 4-dihydroxy-2-butanone 4-phosphate (2) *via* the dismutation of ,7-dimethyl-8-ribityl-lumazine (3). Ribityl-side chain is abbreviated to R.

Engineering efforts have been reported to refine and tailor the properties of AaLS to various applications.[130–132,139] In particular, an engineered variant of AaLS (AaLS-13) was reported to form cage structures capable of hosting active enzymes within its negatively supercharged cavity (figure 1.18 A).[130,140–142] Capsid formation and the uptake of positively charged molecules by AaLS-13 have been observed both in bacterial cells and *in vitro*.[130,139–141] The rapid loading of supercharged fluorescent proteins, GFP(+36) and TOP(+36), has been shown to proceed at a diffusion limited rate at room temperature (figure 1.18 B).[140]
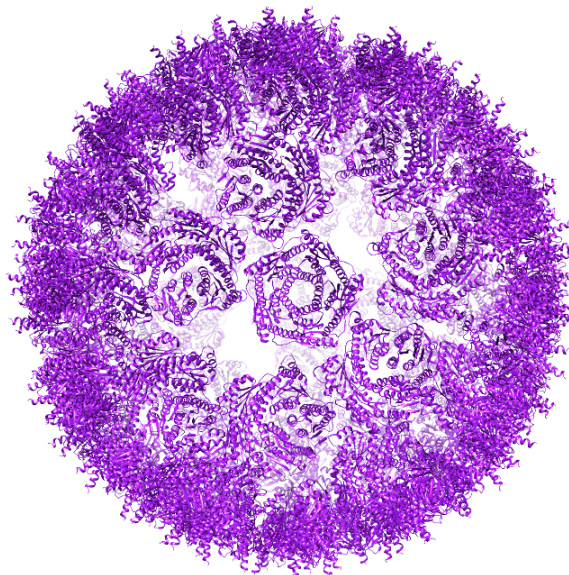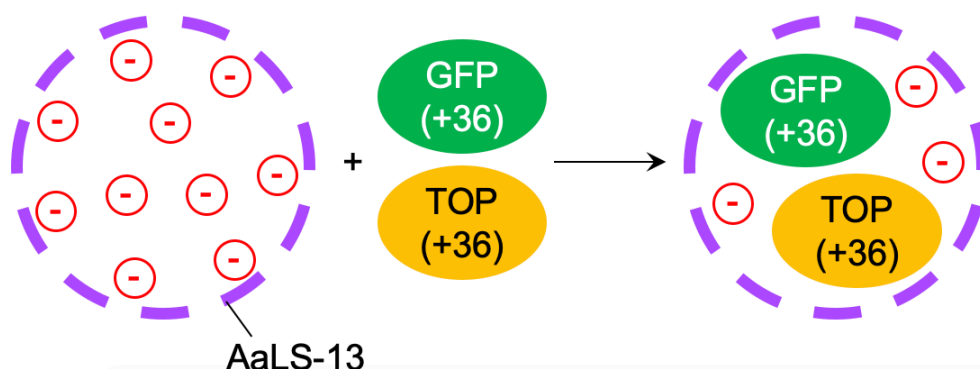
Figure 1.18 AaLS-13 forms protein capsids with a negatively charged cavity which has been shown to host positively charged proteins. **(A)** Ribbon representation of the AaLS-13 protein cage assembly (PDB: 5MQ7) **(B)** Cartoon illustration of the rapid uptake of charged fluorescent proteins, GFP(+36) and TOP(+36), by AaLS-13.

The application of AaLS-13 protein capsids may be adapted for the encapsulation of AEPs and presents an opportunity to construct an artificial organelle which imitate the AEP containing vacuoles observed in plants. This synthetic biology approach provides a potential strategy to incur additional specificity to the substrate recognition of AEPs, which may enable *in vivo* applications of AEP catalysis such as whole cell catalysis.

1.7 Thesis Aims

AEPs derived from plant species such as *C. ternatea*, *O. affinis* and others exhibit excellent enzyme kinetics for peptide ligation *via* a relatively short substrate recognition sequence. These enzymes have been applied to mediate the cyclisation and conjugation of proteins and peptides. Nevertheless, questions and challenges regarding recombinant preparation of AEPs, reversibility during intermolecular ligation and cytotoxicity remain. An engineered AEP variant from *O. affinis*, OaAEP1-C247A (herein referred to as OaAEP1), was employed as the model AEP for all of the experiments described in this thesis. This enzyme was selected due to the availability of a recombinant preparation procedure. Furthermore, this OaAEP1 variant was reported to exhibit superior catalytic parameters compared to the wild-type enzyme. This thesis aims to address three key challenges which limit the application of AEPs.

1. **Validate the current methods and establish a simplified protocol for the preparation of recombinant AEPs.** The reported procedure for the preparation of OaAEP1 was reproduced. The peptide cyclisation activity of the enzyme was validated by steady state enzyme kinetic analysis. The pH tolerance of AEPs is known to vary between isoforms. Therefore, the peptide cyclisation kinetics of OaAEP1 was examined at different pH values. New gene constructs for the production of recombinant OaAEP1 was created to explore more streamlined preparation procedures. The activity of OaAEP1 prepared from each method was confirmed with the cyclisation of a model peptide substrate.

2. **Develop a strategy to address the reversibility of AEP-driven intermolecular peptide ligation.** The method should enable effective AEP-mediated protein labelling with low equivalents of the nucleophilic labelling substrate. Substrate recognition by OaAEP1 at the P1' and P2' positions were examined using a peptide ligation assay. A chemo-enzymatic approach to AEP-catalysed protein labelling, which exploits the low specificity towards the P1' position, is described. The method was validated through peptide ligation as well as protein labelling experiments. Factors affecting the protease activity of AEPs, resulting in hydrolytic side reactions, were also examined and discussed.

3. **Explore the applications of AaLS-13 protein capsids as an artificial organelle to sequester and regulate the activity of activated AEPs.** Active enzymes have been successfully encapsulated in AaLS-13 compartments. Protein cargo have also been incorporated into cpAaLS/AaLS-13 patchwork protein capsids *via* genetic fusion to cpAaLS. The application of both methods to the encapsulation of OaAEP1 were explored. Protein encapsulation was analysed by negative stain transmission electron microscopy (TEM). The activity of the caged OaAEP1 was examined and compared against the free enzyme.

2 Preparation of recombinant *Oldenlandia affinis* asparaginyl endopeptidase 1 – C247A (OaAEP1)[I]

2.1 Preface

Good kinetic parameters for peptide ligation and short substrate recognition sequences highlight OaAEP1-C247A (herein referred to as OaAEP1) as a potential biocatalytic tool for site specific protein labelling. While OaAEP1 has been utilised for the ligation of proteins and peptides, current methods to prepare recombinant OaAEP1 are labour intensive. Catalytically active OaAEP1 is obtained through multiple purification steps and an incubation period at low pH. Thus, the gene construct for the production of recombinant OaAEP1 is modified to separate the pro-domain from the core domain. Direct production of the active AEP bypasses the low pH activation step involved in previously reported procedures. This has enabled the development of a simplified method for enzyme preparation. The yield and activity of the active enzyme were measured and compared against OaAEP1 prepared with the procedure reported in the literature.

[I] This chapter includes the work and reproduces some of figures published in Chem. Sci., 2020, **11**, 5881. (DOI: 10.1039/D0SC02023K). Authors of this article are T. M. Simon Tang, Davide Cardella, Alexander J. Lander, Xuefei Li, Jorge S. Escudero, Yu-Hsuan Tsai and Louis Y. P. Luk. TMST, the author of this thesis, is the only first author of this article.

2.2 Introduction

Biocatalytic approaches are potentially safer and more sustainable compared to chemical approaches which commonly employ organic solvents and high reaction temperatures. However, the adoption of enzymatic methods is often limited by the inconvenience of enzyme preparation. OaAEP1 from *O. affinis* is one such enzyme that can benefit from a streamlined recombinant production procedure. The development of simplified methods to prepare recombinant OaAEP1 may facilitate a greater uptake of this biocatalyst for protein bioconjugation. Current recombinant methods to prepare OaAEP1 mimics the biosynthesis of AEPs observed *in planta* (see Chapter 1.5.1). Recombinant OaAEP1 is produced as a zymogen, with a pro-domain which silences the catalytic activity.[14] The pro-domain has been suggested to confer stability to the catalytic core domain of AEPs at physiological pH,[11] as mature AEPs are typically located in organelles with acidic environments.[2,9,11,13,14] Three chromatographic steps are required prior to the activation of OaAEP1, which is an incubation in acidic conditions (pH 4.0) to trigger the proteolytic cleavage of the pro-domain. Since the presence of impurities could result in the aggregation and precipitation of OaAEP1 at low pH, the laborious purification steps are necessary to obtain the pro-enzyme in high purity.

The protocol for the preparation of OaAEP1, its expected expression yield and catalytic activity were replicated from the literature.[14,22] In addition, the effect of pH on the peptide cyclisation activity of OaAEP1 was investigated, which indicated OaAEP1 was active and stable at neutral and physiological pH. This presented an opportunity to develop a method for the preparation of OaAEP1 which does not require activation in acidic conditions. Firstly, the simplified approach avoids the activation step thereby avoiding the incubation period which is time-consuming. Secondly, this also negates the risk of protein aggregation and precipitation at low pH.

To explore activation-free methods of OaAEP1 preparation, two new gene constructs were designed to separate the gene encoding for the pro-domain from the gene encoding for catalytic core domain. Recombinant gene expression was performed in *E. coli*, the activity and yield of the purified OaAEP1 was analysed.

2.3 Results and discussion

2.3.1 Preparation of recombinant OaAEP1

The codon-optimised construct for recombinant gene expression by *E. coli* encoding for the His$_6$-ubiquitin-OaAEP1 fusion protein (fOaAEP1) was purchased and inserted into a pET-28b vector. Gene expression was induced by isopropyl β-D-1-thiogalactopyranoside (IPTG) in *E. coli* strain BL21 (DE3). The enzyme was isolated by IMAC, followed by SAX and SEC. Samples taken during the purification process were analysed by SDS-polyacrylamide gel electrophoresis (PAGE). Two dominant bands around 59 kDa and 50 kDa were observed, corresponding to the fOaAEP1 and a truncated OaAEP1 (tOaAEP1) where His$_6$-ubiquitin was lost, respectively (figure 2.1). Notably, tOaAEP1 was retained by Ni^{2+}-nitrilotriacetic acid (NTA) during IMAC despite the loss of the His$_6$-tag. OaAEP1b and other AEPs have been shown to adopt a dimeric quaternary state *via* interactions between the pro-domains.[11,14] The retention of tOaAEP1 by Ni^{2+}-NTA may be rationalised by its dimer interaction with the full-length recombinant AEP to facilitate binding to the stationary phase (figure 2.1 B). An additional band positioned above the 66 kDa maker was assigned as an unknown impurity which could not be removed by chromatographic purification (figure 2.1 B). This contamination was also observed in the procedures reported in the literature.[14] The protein contaminant was no longer detected after the activation step in acidic conditions.
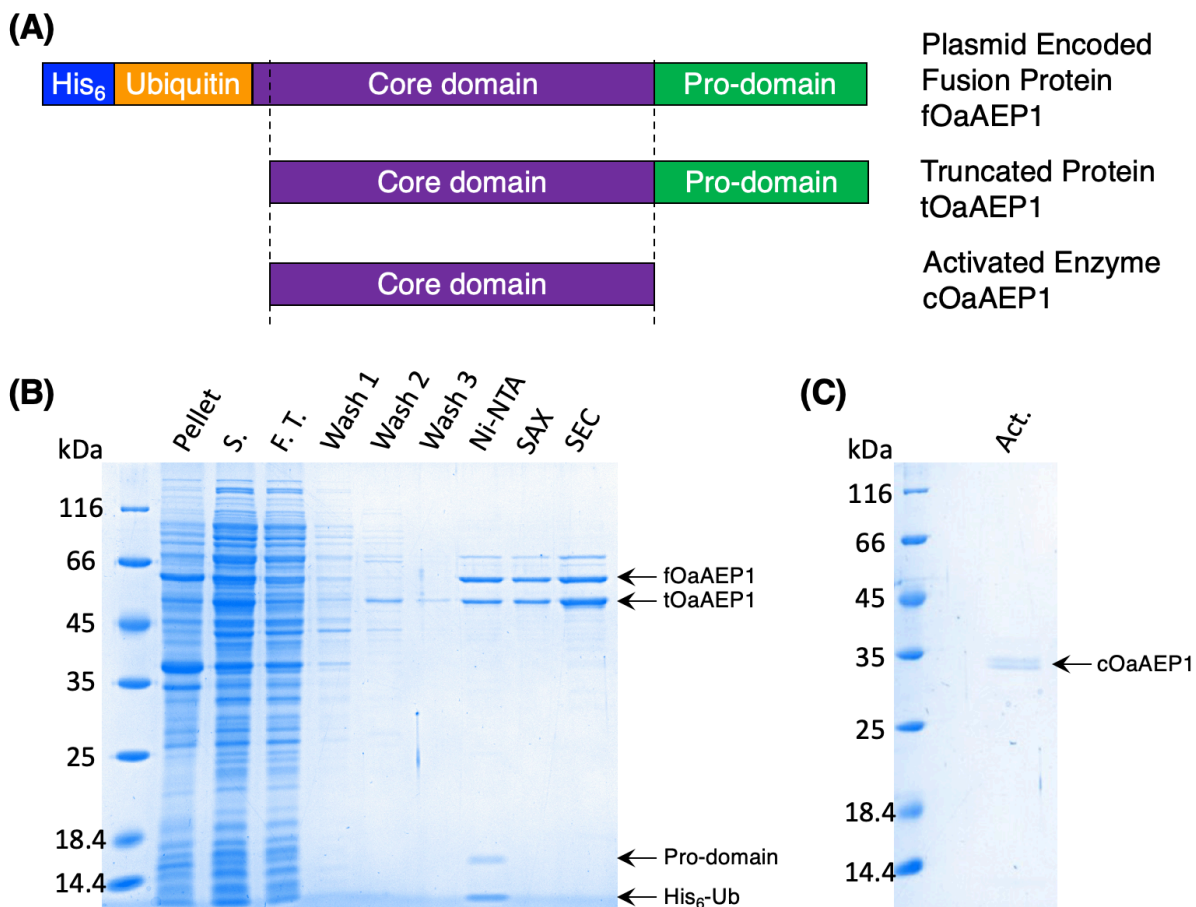
Figure 2.1 Purification and activation of recombinant OaAEP1. **(A)** Schematic representation of AEP fusion protein, truncated protein and acid induced activated enzyme employed in this study. **(B)** SDS-PAGE analysis of fractions obtained during purification and a sample of the activated enzyme. SDS-PAGE lanes: insoluble fraction of lysate (pellet), soluble fraction of lysate (S.), flow through from Ni^{2+}-NTA (F. T.), wash fractions from Ni^{2+}-NTA (wash 1-3), protein elution from Ni^{2+}-NTA (Ni-NTA), protein elution from strong anion exchange chromatography (SAX), protein elution from size exclusion chromatography (SEC). **(C)** SDS-PAGE analysis of the activated enzyme (Act.).

Activation of the zymogen was triggered by acidic and reducing environments, resulting in the auto-proteolytic removal of the N-terminal His_6-ubiquitin domains and the C-terminal pro-domain. A chelating agent, ethylenediaminetetraacetic acid (EDTA, 1 mM), and a reducing agent, tris(2-carboxyethyl) phosphine (TCEP, 0.5 mM), were added to fractions from SEC containing the desired AEP zymogen. EDTA was added to prevent interactions between the catalytic cysteine residue and any trace metal contaminants that may be present in the solution. The protein solution was then acidified to pH 4.0 using acetic acid (AcOH) and incubated at room temperature for 16

h. Several potential cleavage sites for OaAEP1b activation were previously identified via digestion and tandem mass spectrometry (MS/MS) experiments.[22] Samples were analysed by SDS-PAGE after activation, where two bands located closely below the 35 kDa marker were observed (figure 2.1 B). The presence of another protein species may be due to incomplete truncation to the final OaAEP1 core domain species (cOaAEP1). On the other hand, electrospray ionisation mass spectrometry (ESI-MS) analysis detected only one protein species, which correspond to the mass of the catalytically active cOaAEP1 with plausible N and C terminal AEP auto-processing sequences (figure 2.2). The cleavage sites for activation identified in this work was in agreement with the previous study by Harris *et al*. using OaAEP1b.[22] Protein concentration was estimated by UV absorbance at 280 nm wavelength, measured on a nanodrop instrument. An estimated yield of 2 mg of activated recombinant enzyme was obtained from *E. coli* cultured in one litre of LB media. The observed deconvoluted mass of 30264 Da is 20 Da lower than the calculated mass of cOaAEP1 (30284 Da). This mass difference was attributed to a disulfide bridge formation (-2 Da) and a condensation of an Asp side chain in the formation of an aspartimide motif (-18 Da).[25,35,38]

(A)

```
  1 MGMAHHHHHHMQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQ 50
 51 QRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGARDGDYLHLPSEVS 100
101 RFFRPQETNDDHGEDSVGTRWAVLIAGSKGYANYRHQAGVCHAYQILKRG 150
151 GLKDENIVVFMYDDIAYNESNPRPGVIINSPHGSDVYAGVPKDYTGEEVN 200
201 AKNFLAAILGNKSAITGGSGKVVDSGPNDHIFIYYTDHGAAGVIGMPSKP 250
251 YLYADELNDALKKKHASGTYKSLVFYLEACESGSMFEGILPEDLNIYALT 300
301 STNTTESSWAYYCPAQENPPPPEYNVCLGDLFSVAWLEDSDVQNSWYETL 350
351 NQQYHHVDKRISHASHATQYGNLKLGEEGLFVYMGSNPANDNYTSLDGNA 400
401 LTPSSIVVNQRDADLLHLWEKFRKAPEGSARKEEAQTQIFKAMSHRVHID 450
451 SSIKLIGKLLFGIEKCTEILNAVRPAGQPLVDDWACLRSLVGTFETHCGS 500
501 LSEYGMRHTRTIANICNAGISEEQMAEAASQACASIP                537
```

(B)
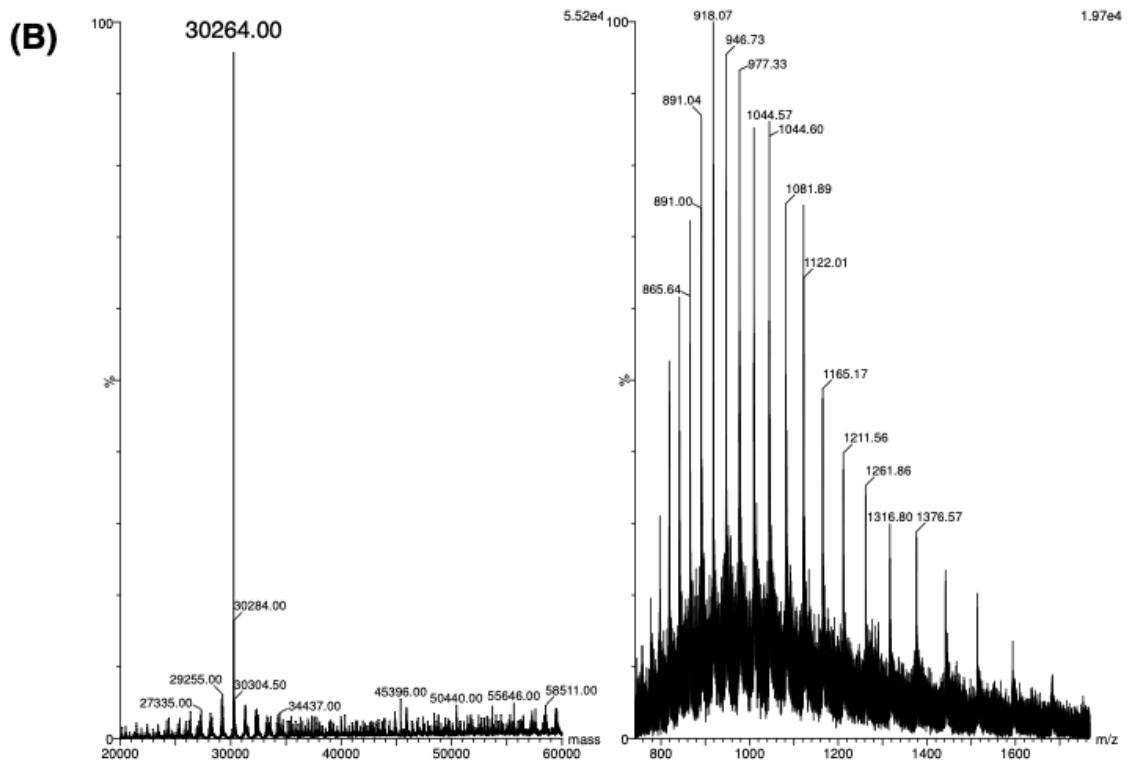
Figure 2.2 Acid-induced activation of OaAEP1 to yield the catalytically active enzyme, cOaAEP1. **(A)** Amino acid sequence of fOaAEP1, with His_6 tag (blue), ubiquitin (orange), core domain (purple) and pro-domain (green). Black triangles denote cleavage sites during protein activation to afford cOaAEP1. **(B)** Mass corresponding to the cOaAEP1 was detected by ESI-MS. Deconvoluted mass (left) and mass spectrum (right).

2.3.2 Kinetics studies of recombinant OaAEP1 for peptide cyclisation

To evaluate enzymatic activity, peptide cyclisation assays were conducted to establish the kinetic parameters of OaAEP1 for the pH range from 4.5 to 7.5. The cyclisation of a 15-amino acid peptide (GLPVSTKPVATRNGL) was adopted as a model reaction (reaction 2.1, figure 2.3 A), where OaAEP1 should mediate the cleavage and ligation of the putative recognition sequence Asn-Gly-Leu and N-terminal Gly-Leu. The same substrate was previously employed to examine the catalytic activity of OaAEP1 at 37 °C and pH 6.0 (k_{cat}/K_M = 3.42 × 10^4 M^{-1}·s^{-1}).[14] Here, the assay was performed at room temperature (20 °C), which is a more representative reaction condition of the potential *in vitro* applications of AEP catalysis. The reactions were monitored by liquid chromatography-mass spectrometry (LC-MS), where the yield was estimated from the *UV* (210 nm) chromatogram with the area under the peak corresponding to the cyclic product relative to the linear substrate. Cyclic peptide product was observed at all tested pH values, with optimal catalytic turnover constant (k_{cat}) and catalytic efficiency (k_{cat}/K_M) observed at pH 5.0 (figure 2.3). A bell-shaped curve was observed for the plot of k_{cat} against pH, with a maximum k_{cat} of 9.15 s^{-1} observed at pH 5.0 (figure 2.3 B and C). The Michaelis constant (K_M) remains largely unchanged. Hence, a bell-shaped curve was also observed for the plot of k_{cat}/K_M against pH. The k_{cat}/K_M values started from 0.36 × 10^4 M^{-1}·s^{-1} at pH 4.5, reaching a maximum of 3.03 × 10^4 M^{-1}·s^{-1} at pH 5.0 and subsequently decreased to 1.04 × 10^4 M^{-1}·s^{-1} at pH 7.5 (figure 2.3 B and D). The kinetic parameters at pH 6.0 for the peptide cyclisation reaction shown here were in general agreement with the literature values (k_{cat}/K_M = 3.42 × 10^4 M^{-1}·s^{-1}).[14] The decrease in k_{cat}/K_M observed here was attributed to the lowered reaction temperature. The optimal catalytic performance at pH 5.0 reflects the native localisation of OaAEP1 in the vacuole.

Proteins are suited to specific pH environments and deviations from the acceptable pH range often result in denaturation and precipitation. Given the potential application of OaAEP1 in protein bioconjugation, broad pH tolerance of this enzyme is a notable advantage. In particular, the k_{cat} and k_{cat}/K_M values at pH 7.5 (1.53 s^{-1} and 1.04 × 10^4 M^{-1}·s^{-1}, respectively) are comparable to other transpeptidases reported in the literature.[46–48] For example, an engineered variant of sortase A was reported to

catalyse peptide ligation with a k_{cat}/K_M value of 2.80×10^4 $M^{-1} \cdot s^{-1}$ at 22.5 °C and pH 7.5.[46]
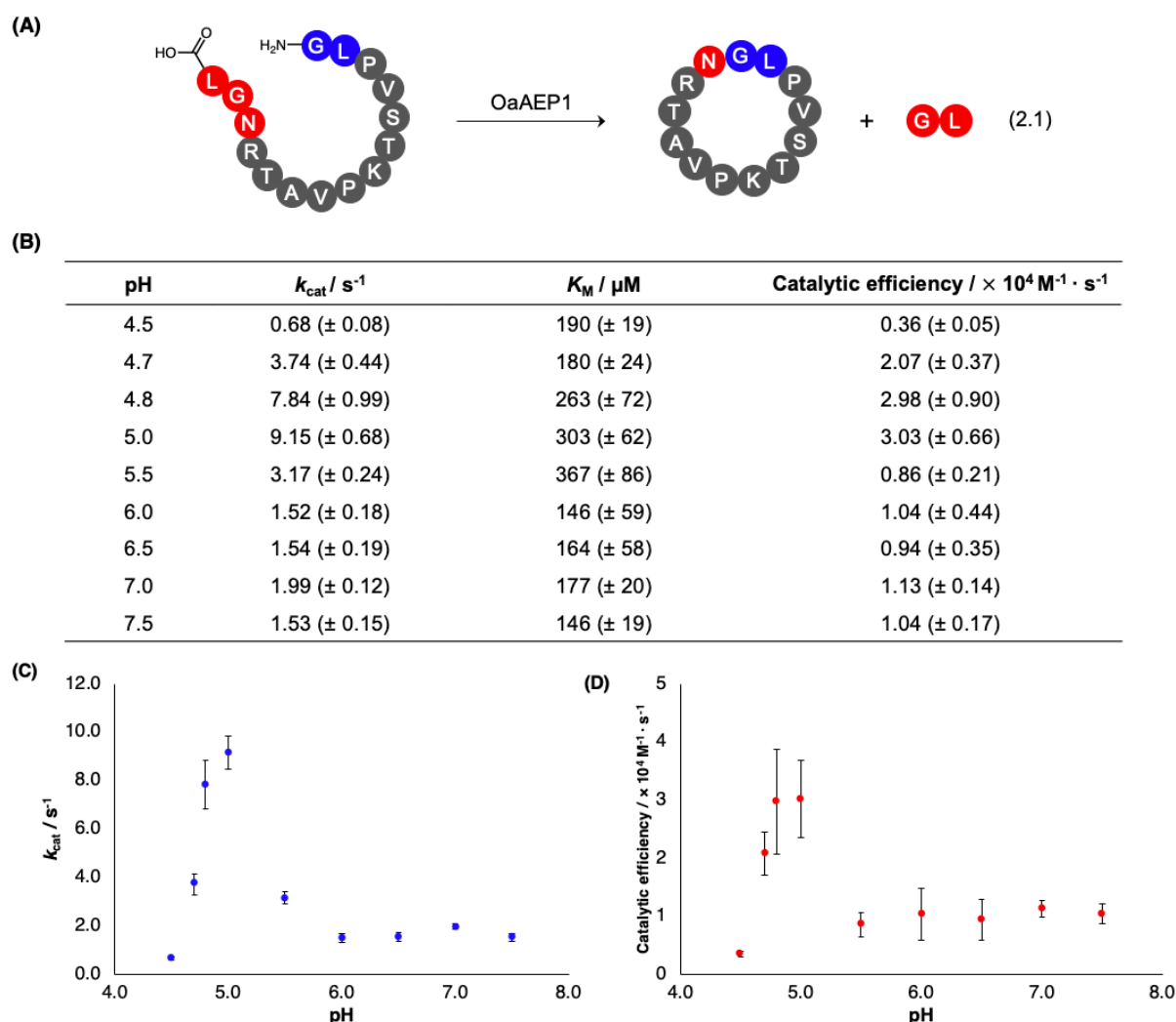
Figure 2.3 Kinetic parameters for OaAEP1 catalysed peptide cyclisation. **(A)** Scheme of the model peptide cyclisation reaction employed. **(B)** Steady-state kinetic parameters of OaAEP1 **(C)** Turnover number (k_{cat}) and **(D)** catalytic efficiency (k_{cat}/K_M) of peptide cyclisation at pH 4.5-7.4, error bars represent the 95% confidence interval of the mean value from triplicate experiments performed at 20 °C.

2.3.3 Simplified approach to recombinant OaAEP1

Recombinant OaAEP1 was produced as a zymogen as the catalytic core domain of AEPs, such as AtLEGγ, was suggested to be unstable at neutral pH.[11] However, the pH-rate profile described here has shown that OaAEP1 remains active and stable at neutral and physiological pH (figure 2.3). Consequently, strategies to bypass the activation step during the preparation of OaAEP1 were explored. The sequence

identity of the AEP catalytic core-domain and inhibitory pro-domain have been identified in previous studies and confirmed here by ESI-MS (figure 2.2).[22] To investigate whether the catalytic core-domain could be prepared without the pro-domain, the gene encoding for fOaAEP1 was truncated by polymerase chain reaction (PCR) to remove the DNA sequence corresponding to the pro-domain. Expression of the modified gene would produce His_6-ubiquitin-cOaAEP1 fusion protein (sOaAEP1), thus bypassing the acid-induced activation step during the preparation of cOaAEP1.

The absence of the pro-domain was found to impair protein solubility, as *E. coli* strain BL21 (DE3) was unable to produce the protein of interest in the soluble form. The structure of cOaAEP1 presents a disulfide-forming cystine pair as well as the catalytic cysteine residue. Incorrect formation of disulfide bonds can lead to misfolding and aggregation. SHuffle® T7 express is an *E. coli* strain which constitutively expresses a gene for a disulfide bond isomerase (DsbC). The isomerase promotes disulfide bond formation in the cytoplasm to correct mis-oxidised cystine pairs. Therefore, *E. coli* strain SHuffle® T7 express was employed for the production of soluble sOaAEP1.

The protein of interest was purified by IMAC followed by SEC. SDS-PAGE analysis of the isolated protein showed a band around 43 kDa (figure 2.4 B), which corresponds to the calculated molecular weight of sOaAEP1.
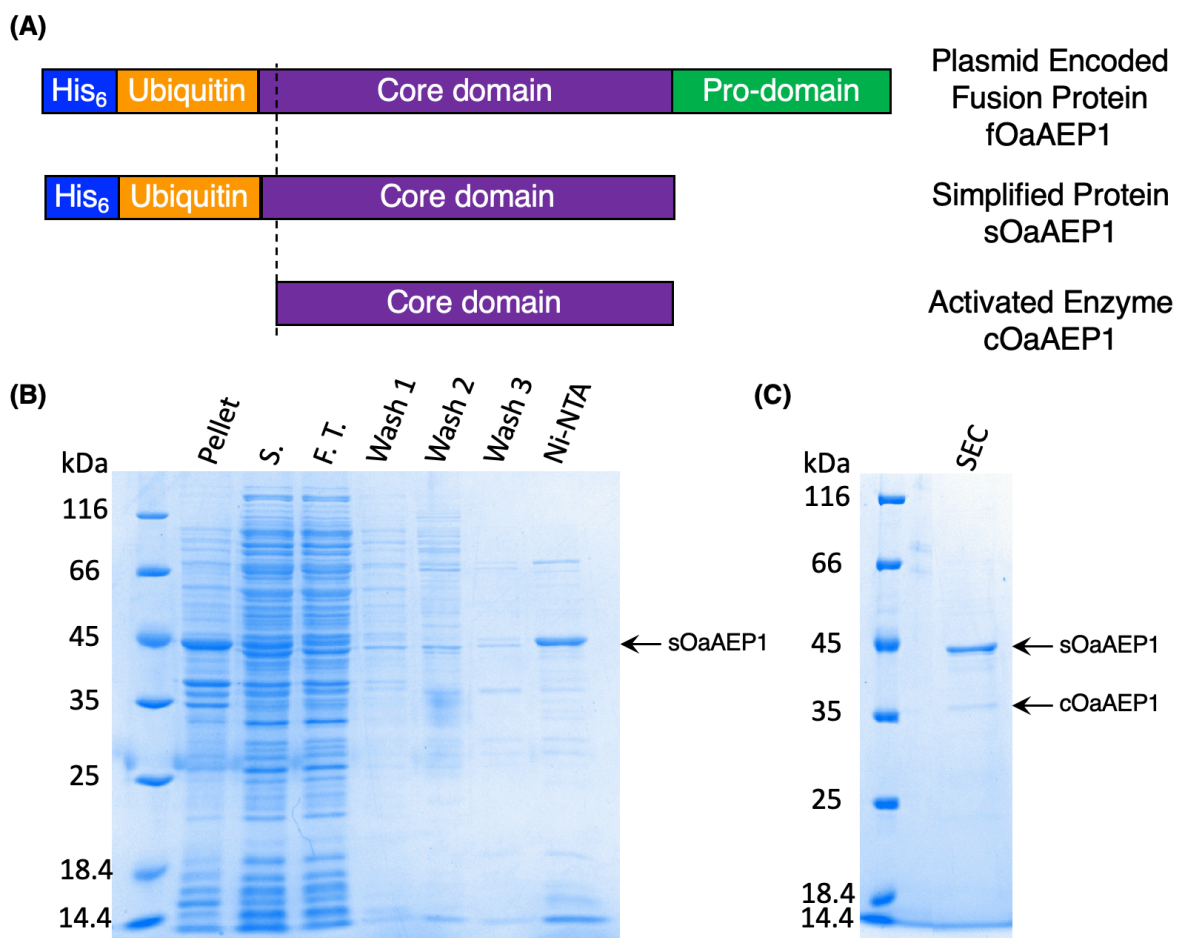
Figure 2.4 Purification of recombinant sOaAEP1. **(A)** Schematic representation of AEP fusion protein, simplified protein and acid induced activated enzyme employed in this study. **(B)** SDS-PAGE analysis of fractions obtained during purification of sOaAEP1 by IMAC. SDS-PAGE lanes: insoluble fraction of lysate (pellet), soluble fraction of lysate (S.), flow through from Ni^{2+}-NTA (F. T.), wash fractions from Ni^{2+}-NTA (wash 1-3), protein elution from Ni^{2+}-NTA (Ni-NTA), **(C)** SDS-PAGE analysis of fractions obtained during purification of sOaAEP1 by size exclusion chromatography (SEC).

Analysis by ESI-MS detected a protein species with a mass of 33135 Da which indicated sOaAEP1 was further truncated. The loss of His_6-ubiquitin from the N-terminus is characteristic of ubiquitin tagged AEP constructs.[14] While a similar proteolytic processing at the N-terminus was also observed during the acid-induced activation of OaAEP1 (figure 2.1), the location of the truncation site within the sOaAEP1 was shifted towards the N-terminus (figure 2.2 A *vs.* 2.5 A). The relatively low signal to noise ratio indicates that there may be other truncated species present (figure 2.5). Mirroring the ESI-MS of cOaAEP1 shown in figure 2.2, the observed mass was 20 Da lower than the calculated mass (figure 2.5 B). The difference was attributed

to the formation of a disulfide bond and an aspartimide motif. Protein concentration was estimated by UV absorbance at 280 nm wavelength, measured on a nanodrop instrument. The yield of the purified active enzyme (2 mg/L) was the same as the original approach described above in Chapter 2.3.1.

(A)
```
  1 MGMAHHHHHHMQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQ 50
 51 QRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGARDGDYLHLPSEVS 100
101 RFFRPQETNDDHGEDSVGTRWAVLIAGSKGYANYRHQAGVCHAYQILKRG 150
151 GLKDENIVVFMYDDIAYNESNPRPGVIINSPHGSDVYAGVPKDYTGEEVN 200
201 AKNFLAAILGNKSAITGGSGKVVDSGPNDHIFIYYTDHGAAGVIGMPSKP 250
251 YLYADELNDALKKKHASGTYKSLVFYLEACESGSMFEGILPEDLNIYALT 300
301 STNTTESSWAYYCPAQENPPPPEYNVCLGDLFSVAWLEDSDVQNSWYETL 350
351 NQQYHHVDKRISHASHATQYGNLKLGEEGLFVYMGSNPAND          391
```

Figure 2.5 Identification of sOaAEP1. **(A)** Amino acid sequence of sOaAEP1, with His$_6$ tag (blue), ubiquitin (orange) and core domain (purple). Black triangle denotes truncation site. **(B)** Mass corresponding to the cOaAEP1 was detected by ESI-MS. Deconvoluted mass (left) and mass spectrum (right).

2.3.4 Peptide cyclisation kinetics of sOaAEP1

In order to confirm the activity of the enzyme prepared with the newly developed protocol, its kinetic parameters were examined using the same peptide cyclisation assay described above (reaction 2.1, figure 2.3 A). No significant difference in enzyme activity was observed at pH 6.0 (Table 2.1) suggesting that the active OaAEP1

prepared *via* the simplified method displayed comparable catalytic activities to the recombinant enzyme obtained through the original approach. Thus, a streamlined method to prepare the catalytically active OaAEP1 core domain which bypasses acid-induced activation is described here for the first time. While the original protocol involved three chromatographic steps and one activation step which required 2 to 3 days to perform, the two purification steps in this newly established method can be performed in one day. These findings suggest the pro-domain may not be needed in the context of using recombinant AEPs for *in vitro* biocatalysis. Due to the relatively low expression yield for this protein, the concentrations of active AEP present in the cytoplasm may not be sufficient to induce toxicity.

Table 2.1 Kinetic parameters of OaAEP1 catalysed peptide cyclisation

Method	k_{cat} / s^{-1}	K_M / μM	k_{cat}/K_M / × 10^4 M^{-1}·s^{-1}
Original	1.52 (± 0.18)	146 (± 59)	1.04 (± 0.44)
sOaAEP1	1.54 (± 0.16)	201 (± 59)	0.79 (± 0.29)

Assay carried out in triplicate at room temperature using 50 mM MES buffer (pH 6.0), 100 mM NaCl, 1 mM EDTA, 0.5 mM TCEP. Data presented as a mean with 95% confidence interval enclosed in brackets (n = 3)

2.3.5 Introduction of a non-covalently bound pro-domain during gene expression

In the absence of the pro-domain, significant quantities of the OaAEP1 core domain were detected in inclusion bodies and precipitates (figure 2.4), which lowered the yield from recombinant expression. The pro-domain of AEPs were previously reported to offer stability and enhance solubility of the enzyme.[11] To ascertain the functional relevance of the pro-domain during the production of OaAEP1, the gene encoding for the pro-domain was reintroduced in a separate open reading frame under the same T7 promoter to enable the production of a split enzyme (figure 2.6).
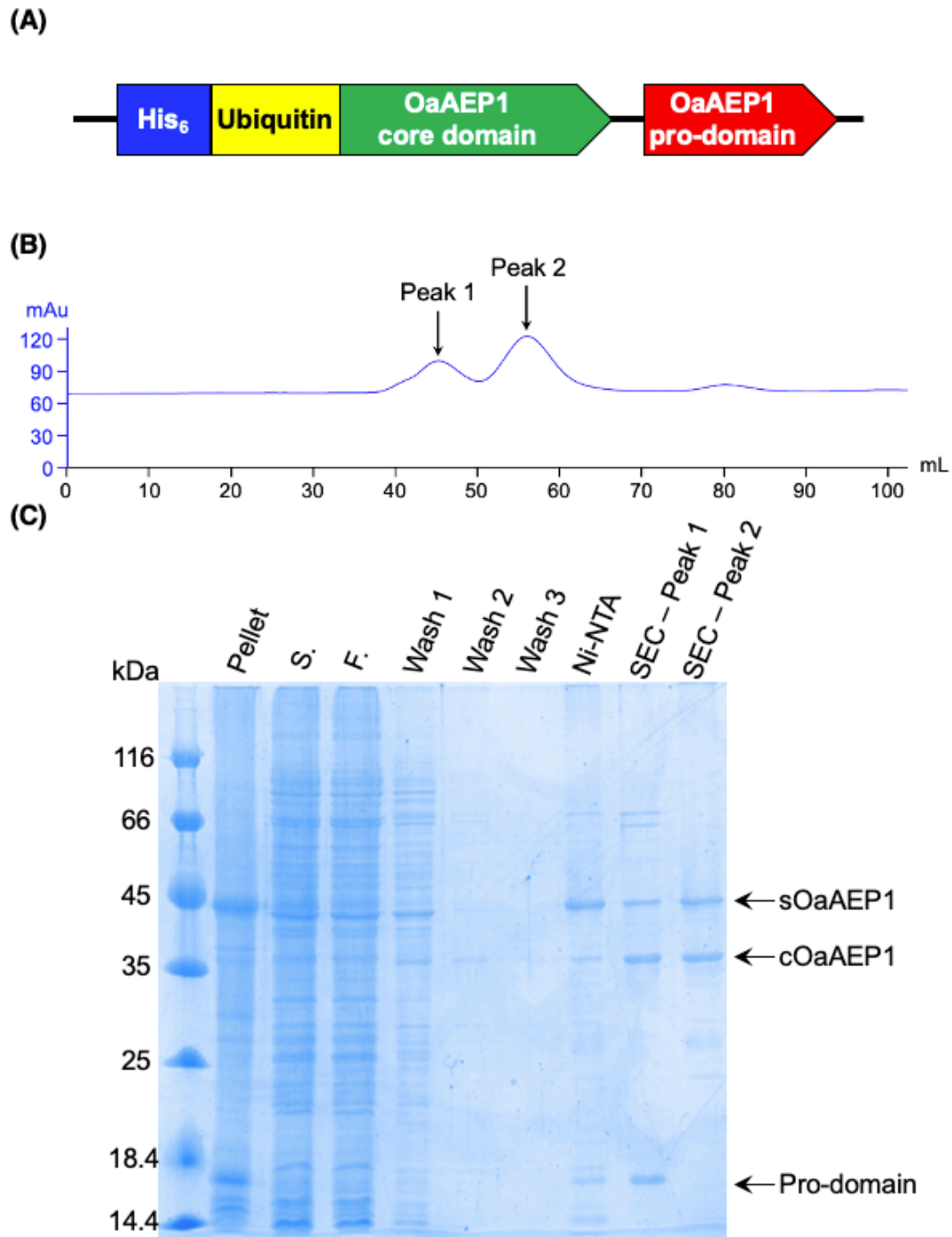
Figure 2.6 Purification of split-OaAEP1. **(A)** Schematic representation of the gene construct for the split AEP. A 23-base pair ribosome binding site was positioned upstream of each gene, both genes were under control of a single T7 promoter. **(B)** Size exclusion chromatogram showing UV absorbance at 280 nm. **(C)** SDS-PAGE analysis of fractions obtained during purification of sOaAEP1. SDS-PAGE lanes: insoluble fraction of lysate (pellet), soluble fraction of lysate (S.), flow through from Ni^{2+}-NTA (F. T.), wash fractions from Ni^{2+}-NTA (wash 1-3), protein elution from Ni^{2+}-NTA (Ni-NTA), protein samples from size exclusion chromatography corresponding to peak 1 and 2 (SEC – peak 1 and 2)

The proteins of interest were observed in the soluble fraction when recombinant gene expression was performed with *E. coli* strain BL21 (DE3) (figure 2.6 C). Notably, sOaAEP1 was insoluble in previous expression attempts by the BL21 (DE3) strain. This suggest the pro-domain can function as a chaperone to assist with the folding and improve solubility of the catalytic core domain without being covalently bound. The proteins were isolated by IMAC and SEC. Samples collected throughout the purification process were analysed by SDS-PAGE. Characteristic bands corresponding to sOaAEP1 and cOaAEP1 were observed. A faint band matching the size of the pro-domain was also observed below the 18.4 kDa marker (figure 2.6 C). Two protein species were isolated during purification by SEC (peak 1 and peak 2, figure 2.6 B). Analysis by SDS-PAGE indicated the presence of the catalytically active AEP species in both samples. However, the AEP pro-domain and a faint band around 59 kDa was also detected in peak 1. The presence of the pro-domain suggests the association of the AEP core and pro-domains in its quaternary structure. The non-covalent two-chain state was previously identified in AtLEGγ, an AEP from *A. thaliana*, as an activation-intermediate where the non-covalently bound pro-domain confers stability to the core domain at neutral pH.[11] The higher molecular weight protein detected in peak 1 corresponded to the size of the zymogenic fOaAEP1; likely to have arisen from the ligation activity of OaAEP1 which reconstituted the split protein into the zymogen. The catalytically active species, sOaAEP1 and cOaAEP1, were isolated in peak 2. Enzyme activity was validated using the same peptide cyclisation reaction described previously (reaction 2.1, figure 2.3 A). Protein concentration was estimated by UV absorbance at 280 nm wavelength, measured on a nanodrop instrument. The yield of the active enzyme was around 1.6 mg from 1 L of LB culture, which is slightly lower than the other AEP preparation strategies described above. However, re-introduction of the AEP pro-domain as a separate protein enabled the production of soluble sOaAEP1 from *E. coli* strain BL21 (DE3).

2.4   Conclusion

An engineered OaAEP1 variant (C247A) was recombinantly expressed, purified and activated according to previously published procedures.[14,22] The sequence identity of the activated enzyme was confirmed by ESI-MS and consistent with previous reports.[22] The peptide cyclisation activity of the enzyme was established *via* steady

state kinetic studies, which also agreed with literature values.[14] For the first time, the effect of pH on the steady state kinetic parameters of OaAEP1 was characterised. Optimal enzyme activity was observed at pH 5.0. Although the activity of the enzyme is significantly reduced at neutral and physiological pH, its reaction kinetics are comparable to other transpeptidases reported in the literature.[46–48]

To improve the accessibility of recombinant OaAEP1 for *in vitro* applications, strategies to simplify the preparation of active OaAEP1 were explored. Having shown that OaAEP1 remains active and stable at physiological pH, a shortened gene construct was designed to enable the production of OaAEP1 without the activity regulating pro-domain. OaAEP1 without the pro-domain was found to be insoluble when recombinant gene expression was performed in *E. coli* strain BL21 (DE3). Soluble protein was obtained using SHuffle® T7 express, an *E. coli* strain which promotes disulfide bond formation in the cytoplasm *via* constitutive expression of a gene for a disulfide bond isomerase. The novel gene construct enabled the preparation of active OaAEP1 with a simplified purification process, in similar yield and catalytic activity (figure 2.7). These findings suggest the pro-domain of OaAEP1 may not be an essential feature for recombinant expression. The newly established method enables the preparation of catalytically active recombinant OaAEP1 with reduced time, effort and material costs.
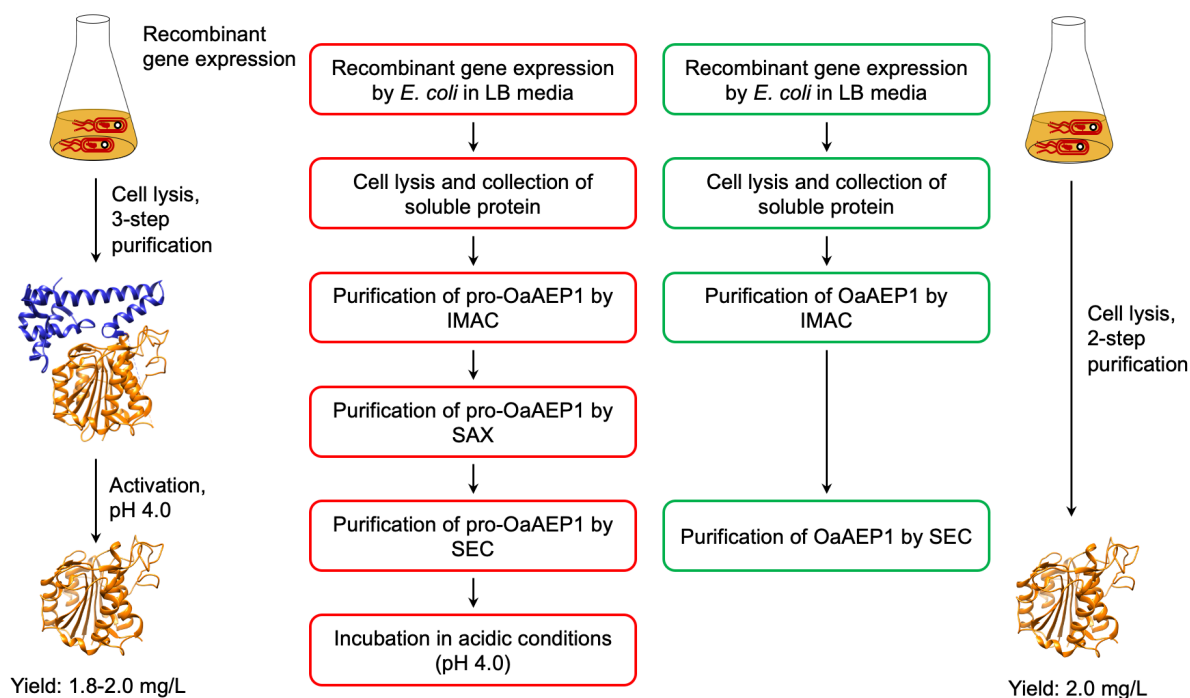
Figure 2.7 Comparison of procedures for recombinant OaAEP1. The simplified approach (green box) described in this Chapter bypasses purification by SAX and activation at pH 4.0.

To evaluate the functional relevance of the AEP pro-domain, a split OaAEP1 construct was prepared. The catalytic core and regulatory pro-domains were co-produced as separate protein modules. The presence of the pro-domain was found to improve the solubility of the AEP core domain, and thus suggests potential functions as a protein chaperone. The co-elution of the two protein domains from SEC demonstrated that the split OaAEP1 adopts a plant-specific two chain state, where the core and pro-domains were non-covalently bound. This was previously observed in AtLEGγ, an AEP from *A. thaliana*.[11] In addition, traces of fOaAEP1 were detected, which confirms previous reports suggesting the pro-domain can be re-ligated to the core domain at neutral pH.[39] Due to the presence of the pro-domain in SEC fractions containing the active enzyme, the isolated yield of the active OaAEP1 from recombinant gene expression was slightly lower than the other methods described here (1.6 *vs.* 2.0 mg/L). Nevertheless, further investigation towards the preparation of recombinant AEP from a split protein construct could offer applicable value. Providing a strategy to interrupt the non-covalent assembly can be established, the pro-domain may be utilised as a protein chaperone to improve the yield of recombinant AEP production.

3 A chemo-enzymatic approach to efficient intermolecular ligation by asparaginyl endopeptidase [II]

3.1 Preface

The ligase activities of AEPs are ideal for peptide and protein labelling. However, reversible ligation limits the efficiency of AEP-mediated intermolecular ligation. The reverse reaction can be supressed by using a large excess of the labelling substrate, but this is uneconomical. To a varying degree of success, a number of other approaches such as thiodepsipeptides and nicked proteins have been reported to address this issue but have also brought their respective limitations. Chemical approaches which label N-terminal cysteine residues are relatively simple to perform with cheap reagents such as 2-formylphenylboronic acid (FPBA). However, the availability of N-terminal cysteines inherently limits the potential applications. Here, FPBA is employed as a scavenger to address the reversibility issue of AEP catalysis. The chemo-enzymatic labelling system was created to exploit the non-specific substrate recognition of AEPs at P1' with the facile chemical reaction between N-terminal cysteine and FPBA. In this approach, OaAEP1-C247A (herein referred to as OaAEP1) is used to ligate polypeptides with an Asn–Cys–Leu recognition sequence with counterparts possessing a N-terminal Gly–Leu. The by-product peptide Cys–Leu is chemically sequestered by FPBA to produce an inert thiazolidine derivative. Consequently, the AEP ligation is driven forward to product formation with a lower ratio of labelled peptide to protein substrate. Through screening of reaction conditions, factors affecting the protease activity of AEPs, resulting in hydrolytic side reactions were identified. The optimised chemo-enzymatic system generates ligated products in the model reaction in excellent yields. The versatility of this AEP-ligation/FPBA-coupling system was further demonstrated by site-specifically labelling of the N- and C-termini of various proteins.

3.2   Introduction

There have been considerable advances in the methodology of protein bioconjugation.[80,91,143] With an expanding library of commercially available labelling reagents, chemical methods are relatively simple to perform.[144–149] However, the efficiency and selectivity of these reactions are often limited by overabundance, or limited availability of specific residues among proteins. For example, reactive aldehyde species such as FPBA reacts rapidly and selectively with N-terminal cysteine residues (figure 3.1).[150,151] N-terminal cysteine residues are susceptible to post translational modification *in vivo*, and thus proteins bearing an N-terminal cysteine residue are neither common, nor simple to prepare.[90,91] In contrast, specificity and reactivity under mild conditions are major advantages of protein-based approaches which utilise enzymes or inteins (protein splicing domain).[92,93,104,152,153]
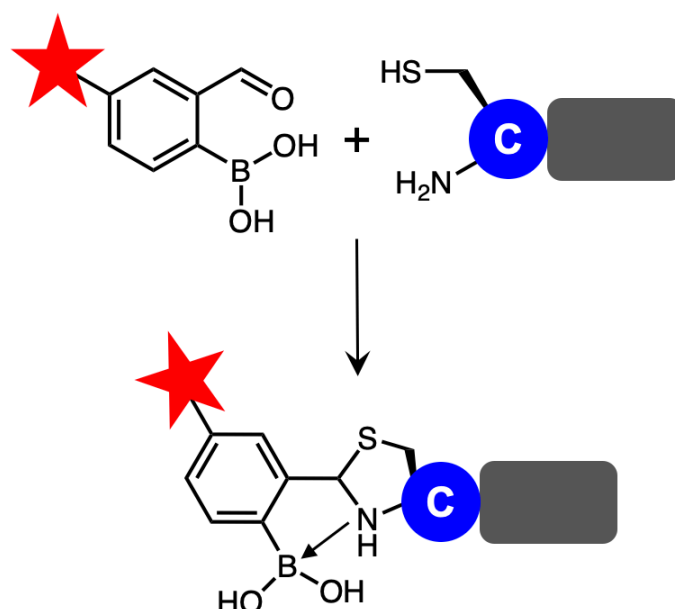
Figure 3.1 FPBA conjugates to the 1,2-aminothiol group of N-terminal cysteine to form a thiazolidine motif. The aldehyde species can be used to modify the N-terminus of proteins and peptides.

Several AEPs have been employed to facilitate protein bioconjugation (see Chapter 1.4.2). Similar to transpeptidases such as sortase A, the AEP-mediated intermolecular ligation is reversible.[37,42,69] Current approaches to overcome this limitation were outlined in Chapter 1.5.2. These include the use of excess labelling substrates, alternative natural and unnatural substrate recognition motifs and large protein domains.[42,43,97] Nevertheless, these methods require labelling substrates in large excess or the inclusion of large self-assembling protein domains. Thus, strategies which are more broadly applicable and economical require exploration. Recently, a

strategy which employed divalent Ni^{2+} ions to prevent the reverse reaction by OaAEP1 was reported.[100] The concept of quenching the reactivity of the nucleophilic by-product from the AEP reaction mirrors the chemo-enzymatic approach described in this chapter. The two methods of sequestering the by-product peptide complement each other and broaden the scope of OaAEP1 applications in protein labelling.

Non-specific substrate recognition at the P1' position is a common trait among AEPs, and previous studies have demonstrated Cys at P1' can be recognized by the wild-type OaAEP1b.[22,39,69] A thorough examination of the substrate recognition by OaAEP1 towards canonical amino acids at the P1' and P2' positions was performed. Prompted by the substrate recognition of OaAEP1, a strategy to enable irreversible AEP-mediated intermolecular ligation was developed (figure 3.2). By employing Asn–Cys–Leu as the recognition sequence for OaAEP1, a by-product that carries a N-terminal cysteine (Cys–Leu) would be generated by the enzymatic forward reaction. The 1,2-aminothiol functionality of N-terminal cysteine contains two nucleophilic centres, which can react with aldehydes to form a thiazolidine.[147,150,151] Having sequestered the reactive α-amine, the by-product peptide (Cys-Leu) would not be able to participate in the reverse enzymatic reaction.
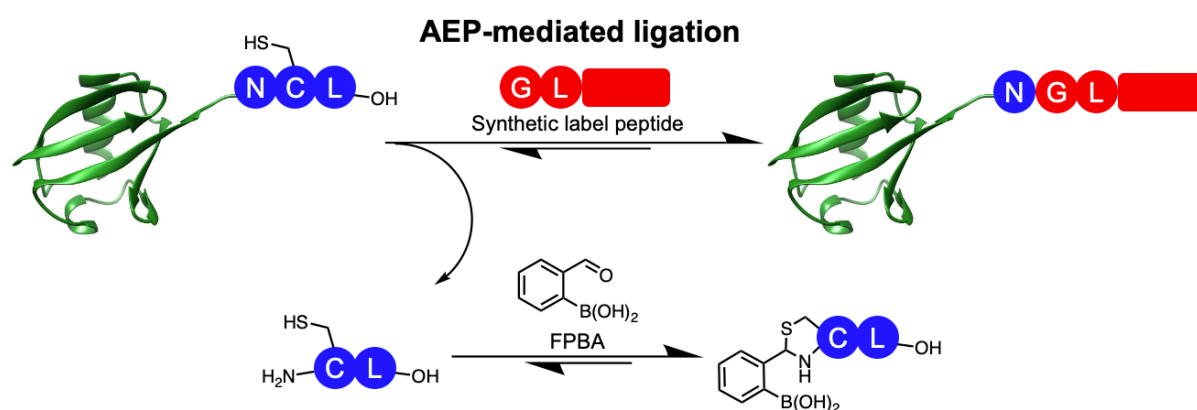
Figure 3.2 A chemo-enzymatic approach to protein bioconjugation which employs FPBA, a cheap and simple chemical reagent, to enhance an enzymatic ligation. Asn-Cys-Leu was used as the AEP recognition sequence. The N-terminal cysteine bearing by-product was sequestered by FPBA. Consequently, the protein labelling reaction equilibrium was driven towards product formation.

The chemo-enzymatic protein labelling strategy employs FPBA to quench the reactivity of the by-product, Cys-Leu (figure 3.2). FPBA was reported to be an electrophile that reacts with N-terminal cysteine peptides forming stable thiazolidine adducts in neutral and aqueous conditions. This reaction proceeds with good selectivity and efficiency with biomolecular rate constant up to 5.5×10^3 $M^{-1} \cdot s^{-1}$.[150,151] To develop this chemo-enzymatic system, the effects of pH on the reaction kinetics of FPBA towards a N-terminal cysteine peptide were investigated. Subsequently, FPBA was employed as a scavenger for the ligation of model peptides. Reaction conditions such as pH, amount of FPBA added, temperature and reaction time were optimised using ligation reactions with model peptides. The addition of FPBA as a scavenger enables peptide ligation to proceed in excellent yields (up to 95%) with a lower ratio of nucleophilic label peptide (1.2 equivalents). The method was also applied to labelling the N- or C-terminus of proteins of different sequences.

3.3   Results and discussion

3.3.1   Substrate specificity of recombinant OaAEP1

The recognition and specificity towards substrate peptide sequences by various AEPs were discussed earlier (see Chapter 1.3.2). To obtain a comprehensive understanding of the substrate specificity of OaAEP1, peptide ligation assays were performed to assess activity towards substrates bearing different amino acid residues C-terminal to Asn (i.e. P1' and P2'). A model ligation reaction between peptides with the sequence of CFRAN**X**L (where X at P1' position is any of the 20 canonical amino acids, 50 µM) and GLGGIR (250 µM, 5 equivalents) was performed at pH 5.0 by adding 0.1 µM of enzyme (1:500 enzyme to substrate ratio, Figure 3.3 A). The reaction mixtures were incubated for 1 hour at 20 ºC then quenched and analysed by HPLC-MS. The ligated peptide product was observed in most cases, indicating a relaxed specificity at the P1' position; OaAEP1 was able to hydrolyse the peptide bond between asparagine and all the canonical amino acids except proline (Figure 3.2 A). Relatively low conversion was observed when the P1' residue was a hydrophobic β-branched amino acid such as Ile or Val. In contrast, substrate peptides bearing Thr, a more polar β-branched residue, demonstrated good conversion (Figure 3.2 A). Subsequently, the assay was repeated using peptides with the sequence of CFRANG**X** as the substrate. Specificity at the P2' position was found to be more restricted, where large hydrophobic residues such as

Phe, Ile, Leu, Met and Trp were accepted (figure 3.2 B). These outcomes are in agreement with the literature reported substrate specificity of OaAEP1b, the corresponding wild-type to the enzyme employed here.[22,39] Non-specific P1' and relatively strict P2' substrate recognition can also be observed among other AEPs such as butelase 1 from *C. ternatea*.[23] Importantly, it confirmed the recognition sequence Asn-Cys-Leu can be hydrolysed by OaAEP1.
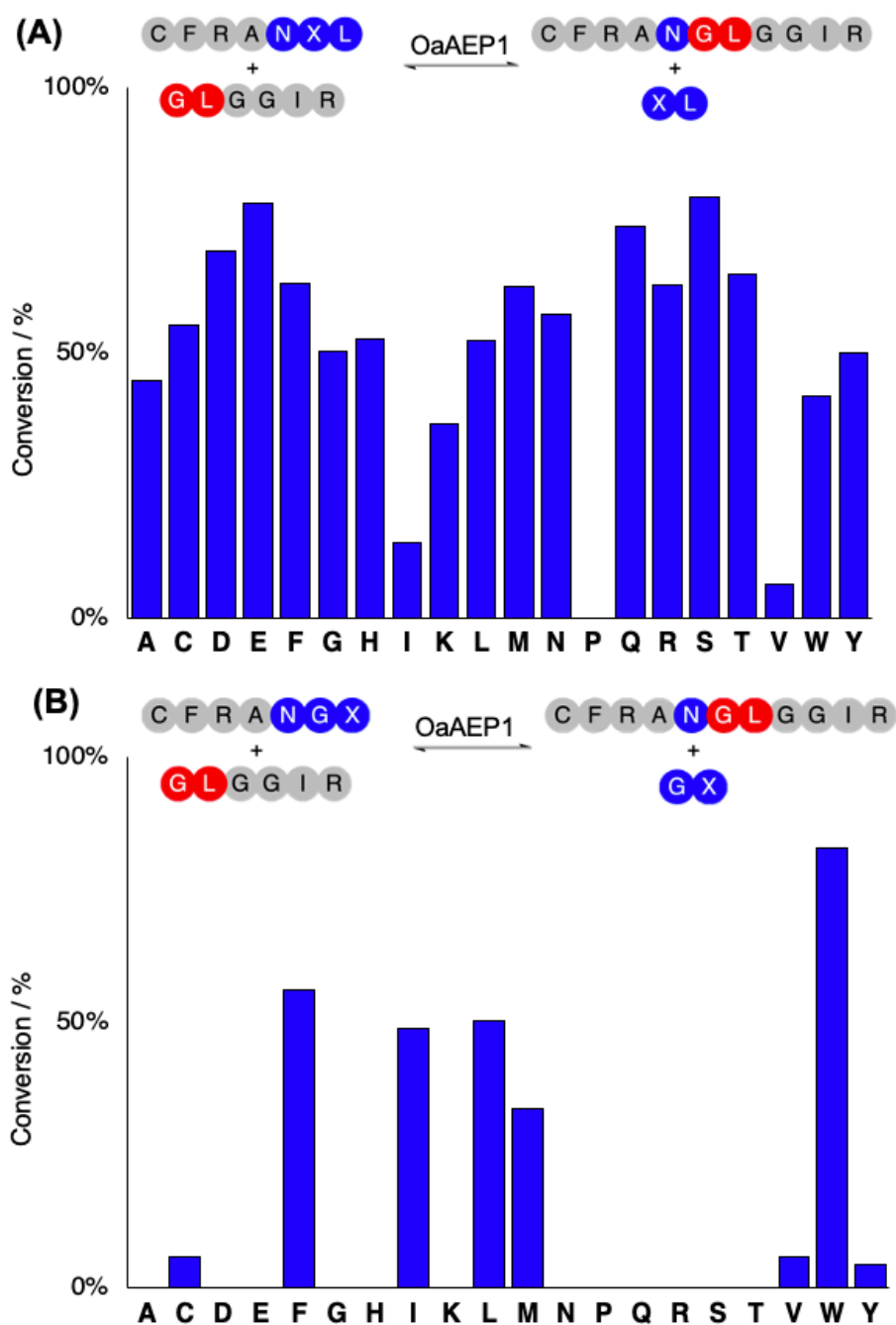
Figure 3.2 Activity profile for peptide ligation catalysed by OaAEP1. (**A**) The ligation of CFRANXL to GLGGIR and (**B**) the profile using CFRANGX, where X is any of the 20 canonical amino acids.

3.3.2 Kinetic characterisation of FPBA conjugation

To drive the AEP-mediated ligation to completion by quenching the reactivity of the peptide by-product (Cys–Leu) using the electrophile FPBA, the two reactions must be kinetically compatible for the chemo-enzymatic cascade to proceed effectively. Consequently, the reaction kinetics of FPBA conjugation was investigated. Adopting an assay from the literature, the thiazolidine formation reaction between FPBA and a N-terminal cysteine peptide (CFRANGL) was monitored by UV spectroscopy.[150,151] The FPBA reaction took place at all examined pH's with second order rate constants ranging from 0.7 at pH 5.0 to 6.0×10^3 $M^{-1} \cdot s^{-1}$ at pH 7.0 (figure 3.3).

The second order reaction kinetics of FPBA conjugation shown here was comparable to the kinetic parameters of OaAEP1 catalysis described in Chapter 2 (figure 2.3 *vs.* figure 3.3), where the k_{cat}/K_M ranged from 25×10^3 $M^{-1} \cdot s^{-1}$ at pH 5.0 to 11×10^3 $M^{-1} \cdot s^{-1}$ at pH 7.4 (see Chapter 2.3.2, figure 2.3).
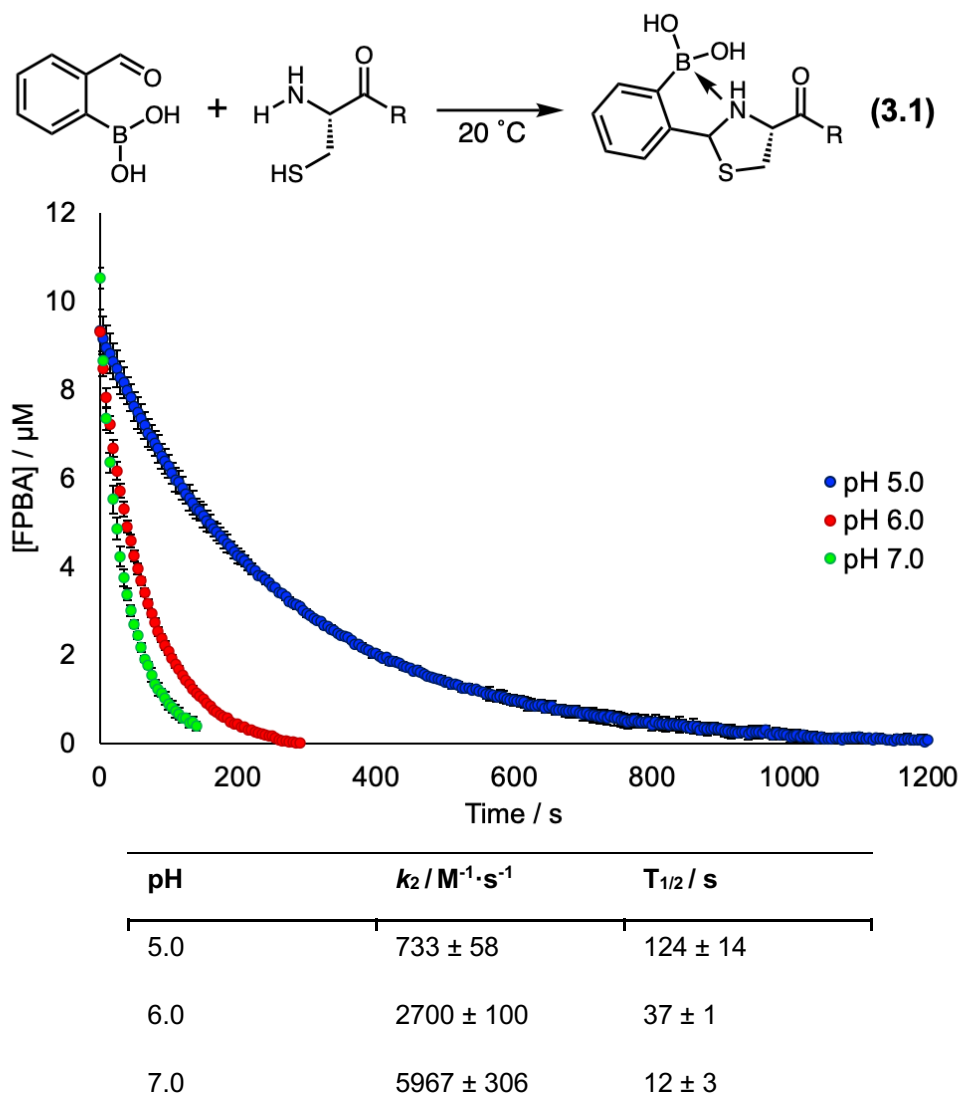
Figure 3.3 Second order kinetic studies of N-terminal cysteine peptide coupling to FPBA. Peptide sequence: CFRANGL. **(Top)** Reaction scheme of N-terminal cysteine peptide coupling to FPBA with graph showing consumption of FPBA estimated by the UV absorbance at 254 nm. **(Bottom)** Kinetic parameters for the reaction

Previous studies have proposed a mechanistic rationale for the rapid and selective reactivity between FPBA and the 1,2-aminothiol group of N-terminal cysteines (figure 3.4).[150,151] Key features include a reversible formation of a Schiff base intermediate, and coordination of the imine nitrogen lone pair to the empty π-orbital of the boronic acid. The conformation adopted as a result of the N-to-B coordination accelerates the nucleophilic attack from the thiol side chain in the formation of the thiazolidine motif. In agreement with the proposed mechanism, this study found that the rate of FPBA

consumption by the N-terminal cysteine peptide decreased when the reaction pH was lowered from pH 7.0 to 5.0. This is likely due to the protonation of the α-amine in acidic environments, which hindered its nucleophilic attack towards the aldehyde group of FPBA.
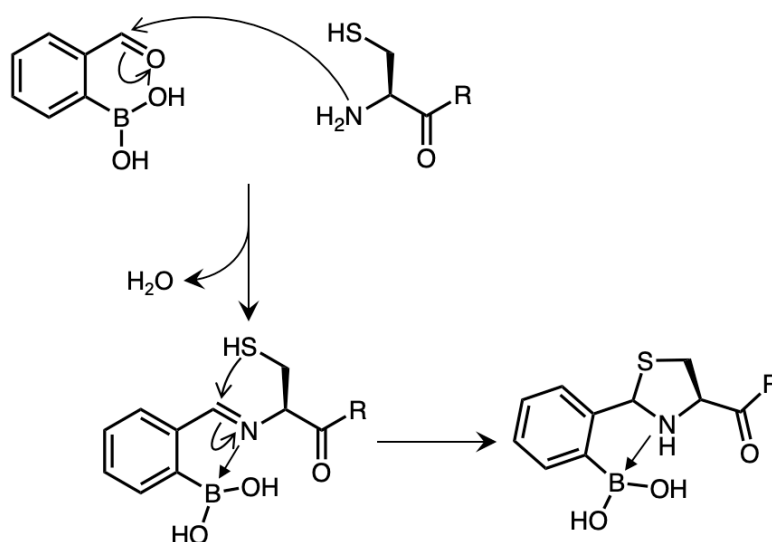
Figure 3.4 Proposed mechanism of FPBA conjugation to the 1,2-aminothiol group of N-terminal cysteine to form a thiazolidine motif. The reaction is accelerated by the coordination of the nitrogen lone pair from the imine to the boron π-orbital of the boronic acid.

3.3.3 Intermolecular peptide ligation mediated by OaAEP1

The effect of FPBA on the OaAEP1 catalysed peptide ligation was investigated using LFRANCLK and GLGGIR as model substrates and HPLC to monitor reaction progress (reaction 3.2, table 3.1). An extra K was added to the C-terminus of the substrate peptide (LFRANCLK) to improve separation during analysis by HPLC. When the substrate peptides were employed in near equimolar quantities (1-1.2 molar equivalents of GLGGIR relative to LFRANCLK), conversions were limited to around 50%, which indicated the reaction was reversible and remain in equilibrium (table 3.1, entries 2, 5 and 6). On the other hand, inclusion of FPBA in reaction mixtures under the same conditions resulted in conversions up to 95% (entry 2 and 6, table 3.1). To highlight the role of the Cys residue at the P1' position in this chemo-enzymatic system, a control experiment was performed using a cysteine-free peptide, LFRAN**A**LK, in place of LFRAN**C**LK. No significant difference in conversion was observed by adding FPBA (table 3.1, entries 7 and 8). Under the optimised conditions, the model reaction

proceeded at pH 5.7 and 20 ºC with a 94% conversion and no observable hydrolysis in 4 h using 1.2 equivalent of the nucleophilic peptide (GLGGIR) and 2 equivalents of FPBA (entry 6, table 3.1).

Table 3.1 Model reaction catalysed by OaAEP1 in the absence and presence of FPBA[a]
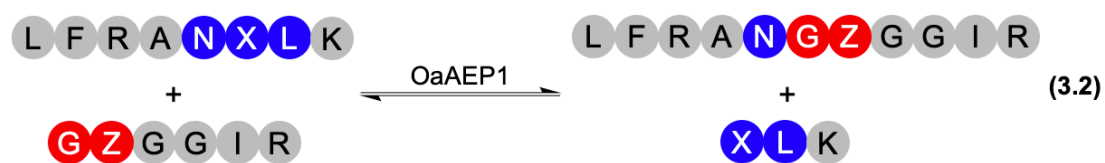
Entry	pH	X	Z	GZGGIR[b] (equiv.)	Time (h)	(-) FPBA	(+) FPBA
						Conversion	
1	5.0[c]	Cys	Leu	1.2	3	53[e]	89[e]
2	5.2[c]	Cys	Leu	1.2	3	54[f]	95[f]
3	5.5[d]	Cys	Leu	1.2	3	39[f]	85[f]
4	5.5[d]	Cys	Leu	1.2	4	38[f]	90[f]
5	5.7[d]	Cys	Leu	1.2	3	51[g]	78[g]
6	5.7[d]	Cys	Leu	1.2	4	50[g]	94[g]
7	5.7[d]	Ala	Leu	1.2	3	67[g]	61[g]
8	5.7[d]	Ala	Leu	1.2	4	70[g]	67[g]
9	5.7[d]	Cys	Val	1.0	4	24[f]	45[f]
10	5.7[d]	Cys	Val	1.5	4	33[f]	55[f]
11	5.7[d]	Cys	Val	2.0	4	32[f]	66[f]
12	5.7[d]	Cys	Val	5.0	4	65[g]	>95[g]
13	5.7[d]	Cys	Val	10.0	4	>95[g]	>95[g]
14	5.7[d]	Cys	Val	20.0	4	>95[g]	>95[g]

[a] All reactions were carried out in triplicate with OaAEP1 (0.3 μM), LFRANXLK (300 μM) and GZGGIR (300–6000 μM) at 20 °C. [b] Equivalents of labelling peptide used relative to the peptide substrate, LFRANXLK. [c] 50 mM NaOAc buffer with 50 mM NaCl, 1 mM EDTA, 0.5 mM TCEP. [d] 50 mM MES buffer with 50 mM NaCl, 1 mM EDTA, 0.5 mM TCEP. [e] 10–15%. [f] 5–10%. [g] <5% of the undesired hydrolysis product observed.

Although the addition of FPBA improves conversion, peptide ligations mediated by OaAEP1 were found to proceed slightly slower in the presence of FPBA compared to the control reaction without FPBA (Chapter 8.3, table 8.1, entries 1-4, 80-85). This side-effect may be attributed to the reversible interactions of FPBA with either the enzyme or the α-amine of the ligating peptide. Nevertheless, this effect does not significantly impact the overall conversion and the improvements from the scavenger effects outweigh the kinetic impediment.

Hydrolysis of the starting peptide LFRANCLK was observed in several test reactions, resulting in the undesired side product of LFRAN. The formation of the hydrolysis product is minor in most of the tested conditions compared to the ligation product. The extent of peptide hydrolysis was shown to be influenced by reaction pH, temperature and time. Lower reaction temperature also appeared to minimise hydrolysis (Chapter 8.3, table 8.1, entries 9, 18 *vs.* 46-48, 51 *vs.* 92-94, 96). In contrast, prolonged reaction time was found to favour hydrolysis (Chapter 8.3, table 8.1, entries 37, 58 and 59). These results indicate that although the hydrolysis mechanism is kinetically inferior to ligation, the undesired reaction is irreversible as the hydrolysed peptide (LFRAN) no longer possesses an OaAEP1 recognition sequence.

The extent of peptide hydrolysis by OaAEP1 also depended upon the availability of a suitable peptide to resolve the enzyme-substrate intermediate. Hydrolysis was diminished when an excess of the ligating peptide (GLGGIR) was used (table 3.1, entries 9–14). Furthermore, more hydrolysis was observed at lower pH conditions where more of the α-amino group of peptides are likely to become protonated, and thus less nucleophilic and rendered unsuitable for ligation (Chapter 8.3, table 8.1, entries 57, 63, 67, 73, 79, 83). Similar observations have been reported from studies in AEPs from other plant species.[11,25]

It has been reported that hydrolysis can be avoided by using polypeptides that carry Gly–Val at the P1''-P2'' positions.[69] To examine if this concept can be applied in tandem to the chemo-enzymatic strategy described here, an alternative peptide GVGGIR was tested in the model ligation reaction (table 3.1, entries 9–14).

Independent of FPBA addition, excellent conversion (>95%) was observed using 10–20 equivalents of peptides (table 3.1, entries 13, 14). Using lower equivalents of GVGGIR (1-5 equivalents), conversion was substantially lower in the absence of FPBA, and the addition of FPBA was found to improve conversion by around 1.5 to 2-fold (table 3.1, entries 9-12). However, product conversion when using GVGGIR was notably lower than when GLGGIR was used (table 3.1, entries 6 *vs.* 10). Reduced conversion when using GVGGIR at low molar equivalents was likely due to the poorer compatibility of Gly-Val at P1''-P2'', resulting in hydrolysis of LFRANCLK to LFRAN. The findings in this study suggest although Asn-Gly-Val in the product peptide may be resistant to AEP hydrolysis, a large excess of Gly-Val bearing label peptide is required to achieve good conversions.

3.3.4 OaAEP1-mediated protein bioconjugation

The newly developed AEP/FPBA approach was employed for site-specific C-terminal labelling. The protein substrate was enhanced green fluorescent protein (eGFP) modified at the C-terminus with a linker containing Asn–Cys–Leu. A biotinylated peptide GLGGZ (where Z is biotinylated lysine) was employed as the labelling substrate. Performed in triplicate and estimated by LC-MS analysis, conversions up to 92% (mean = 86%, n = 3) for the C-terminally labelled protein were observed when using two equivalents of the biotin label (table 3.2 and figure 3.5). Similar findings were observed in the labelling of other monomeric and multimeric proteins. 85% and 80% of β-lactamase and ubiquitin, respectively, were modified at the C-terminus under the same reaction condition (figure 3.5).

Since OaAEP1 catalysis have been shown to function in a broad pH range (Chapter 2.3.2),[39] the effectiveness of the protein labelling system at neutral pH was also examined. The protein substrate employed for this investigation was an engineered lumazine synthase from *A. aeolicus*, AaLS-13. The protein forms a macromolecular complex composed of 360 protein subunits which is prone to precipitate at pH below 7.0.[130,140,154] Using the same reagent stoichiometry and reaction temperature as above at neutral pH condition, up to 75% of the AaLS-13 subunits were found to be labelled (figure 3.5).

Table 3.2 C-terminus labelling of eGFP with biotinylated peptide using OaAEP1 in the presence and absence of FPBA

Entry	GLGGZ[a] (equiv.)	Conversion[b] (%)	
		(-) FPBA	(+) FPBA
1	1.0	43	63
2	1.2	58	66
3	1.5	64	75
4	2.0	71	86

[a] Equivalents of labelling peptide used relative to the protein substrate eGFP. An Asp-to-Ala mutation was needed to avoid an off-target AEP side-reaction. [b] <5% of the undesired hydrolysis product observed.
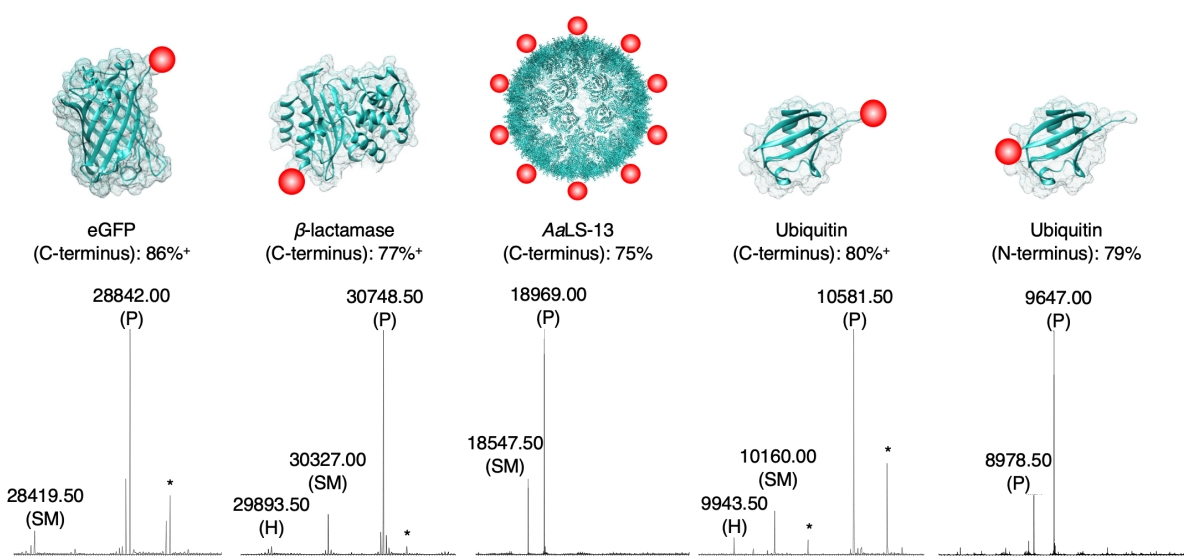
Figure 3.5 Protein modification. Proteins (100 µM) modified with a short linker and the recognition sequence (Asn–Cys–Leu) for OaAEP1 were incubated with biotin labelled peptide (200 µM), OaAEP1 (0.25 µM) and FPBA (200 µM). The reactions were quenched with 1.0 M HCl then analysed by UPLC-MS. SM refers to starting material and P to product. (*) Denotes peaks correlating to α-N-gluconoylation of recombinant protein starting material and products (+178 Da).[155,156] (+) Denotes conversion averaged from duplicate experiments. Asp-to-Ala mutation was introduced to eGFP in order to avoid undesired hydrolytic reaction. The corresponding chromatograms and full mass spectra of the UPLC-MS analysis are reported in Chapter 8.4 Other than the species reported above, there is no evidence of hydrolysed peptides or formation of any other by-products.

The utility of the chemo-enzymatic method for site-specific N-terminus labelling was also investigated. A ubiquitin construct bearing an N-terminal tobacco etch virus protease (TEVp) cleavage sequence was prepared. Cleavage by TEVp revealed a Gly–Leu sequence at the N-terminus. A biotin-labelled peptide bearing the recognition sequence Asn–Cys–Leu at the C-terminus (biotin-ATRNCL) was employed as the labelling substrate. 79% of ubiquitin was labelled at the N-terminus using two equivalents of the biotinylated label and FPBA (figure 3.5).

3.3.5 OaAEP1 activity at recognition motifs within the protein backbone

The combination of a relatively non-specific and short recognition sequence results in a greater chance to encounter OaAEP1 recognisable sites within the backbone of protein substrates. An Asp235Ala mutation at the solvent-exposed site of eGFP (11 residues from the C-terminus) was needed to prevent off-target reactivity during protein labelling (figure 3.6). However, it was noted that all the other proteins used here contained internal Asn and Asp residues which are neither hydrolysed nor modified by OaAEP1. The ability for the enzyme to target compatible recognition sequences within the protein backbone was likely to be dependent on their accessibility to binding by the AEP.

	Calculated MW	Amino acid sequence
Starting Material (SM)	28484.16 28352.96 (-Met)	MHHHHHHMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGD ATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ HDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE LKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIR HNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPN EKRDHMVLLEFVTAAGITLGMDELYKGGSGNCL
Desired Product (P)	28906.66 28886.66 (GFP)	MHHHHHHMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGD ATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ HDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE LKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIR HNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPN EKRDHMVLLEFVTAAGITLGMDELYKGGSGNGLGGZ
Side Product (SP)	28000.69 27980.69 (GFP)	MHHHHHHMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGD ATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ HDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE LKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIR HNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPN EKRDHMVLLEFVTAAGITLGMDGLGGZ

Figure 3.6 Bioconjugation of biotin labelled peptide to the C-terminus of eGFP. OaAEP1 mediates the ligation reaction at a recognition site within the native protein backbone (highlighted in red) instead of the intended site. Label peptide sequence: GLGGZ (Z = biotinylated lysine)

3.4 Conclusion

The work presented in this chapter combines the advantages of chemical and enzymatic labelling, creating a bioconjugation system with features that could not be achieved by either method alone. Due to substrate specificity and reactivity under mild and biocompatible conditions, transpeptidases such as sortase A and AEPs are appealing tools for bioconjugation.[27,45,69,92–94,99,105] However, their reactions are reversible, and thus a large labelling agent to protein substrate ratio is needed to achieve high conversion. The chemo-enzymatic method described here, offers a versatile approach to perform bioconjugation with low label-to-protein ratios and is applicable to both N and C-termini. Through the optimisation of the model peptide reaction, key features which underpin the AEP ligase *vs.* peptidase activity were uncovered. Parameters to minimise undesired substrate and product hydrolysis by OaAEP1 were identified. This chemo-enzymatic method requires a C-terminal recognition sequence containing a cysteine residue. While some of the proteins used here, such as eGFP, β-lactamase and AaLS13, contain cysteine residues, the introduction of an extra cysteine residue may not be tolerated by all proteins.[157,158] In particular, proteins which contain a significant number of disulfide bonds such as antibodies and nano-bodies.[157,158] Nevertheless, provided that the AEP-ligation/FPBA-coupling method lowers the ratio of label to protein substrate, it is particularly applicable when expensive or non-commercially available labels, such as radioactive or isotopic probes, are used.[43,97,109,110] It should be noted that the use of FPBA is not strictly limited to OaAEP1. The concept of using scavengers in enzymatic reactions present potential utilities in other enzymatic reactions. The development of this technology complements the use of existing transpeptidases such as sortase A, as the differences in substrate specificity may be fully exploited in combination to develop orthogonal ligation strategies for the production of complex, multiply labelled proteins.[69,105] In summary, the pairing of enzymatic transpeptidation with well-established chemical reactions offers a versatile and efficient approach to the preparation of tailored protein constructs.

4 Encapsulation of an asparaginyl endopeptidase in a nano-scale protein container

4.1 Preface

The combination of a relatively short, non-specific substrate recognition and protein modifying activities of AEPs can lead to detrimental consequences *in vivo*. In plants, active AEPs commonly accumulate in the vacuole. Spatial confinement within organelles provides control over the destructive reactivity of the enzyme. Vacuole collapse, resulting in the release of active AEPs into the cytoplasm has been associated with programmed cell death. Subcellular organisation by membrane bound organelles are not present in prokaryotic host cells such as *E. coli*. Consequently, the accumulation of active AEPs in the cytoplasm is likely to be toxic and preclude the development of *in vivo* applications such as whole cell catalysis. AaLS-13 is an engineered variant of lumazine synthase which assembles into an icosahedral protein cage. The protein compartment has been employed to host a variety of protein guest molecules, including active enzymes such as TEVp. The encapsulation of OaAEP1-C247A (herein referred to as OaAEP1) within the AaLS-13 protein cage represents an artificial organelle with which to mimic to the spatial confinement of AEPs observed *in planta*. The OaAEP1/AaLS-13 complex may enable further applicational developments of AEP *in vivo*. To generate a functional encapsulated OaAEP1 system, two reported methods of loading protein cargos into AaLS-13 capsids are explored. Consequently, the construction of an AaLS-13/OaAEP1 complex which is capable of facilitating peptide ligation is described. Protein encapsulation was confirmed by negative stain TEM. Employing a neutral and a positively charged substrate peptide, activity of the encapsulated enzyme is analysed and compared against the free enzyme.

4.2 Introduction

The engineered LS from *A. aeolicus*, AaLS-13, is a cage forming protein which is capable of hosting active enzymes within its negatively supercharged cavity.[130,140–142] Electrostatic interaction-driven cargo molecule uptake by AaLS-13 provides a general and predictable strategy for encapsulation. This property was previously exploited to create a substrate sorting system for the encapsulated enzyme. Azuma *et al.* showed that TEVp bearing a GFP(+36) tag can be sequestered by AaLS-13 to create a proteolytic nanoreactor.[142] Notably, the anionic environment in the cavity of the protein cage was found to impart selectivity towards substrates with an overall net positive charge (figure 4.1 A).

(A) Substrate selective nanoreactor
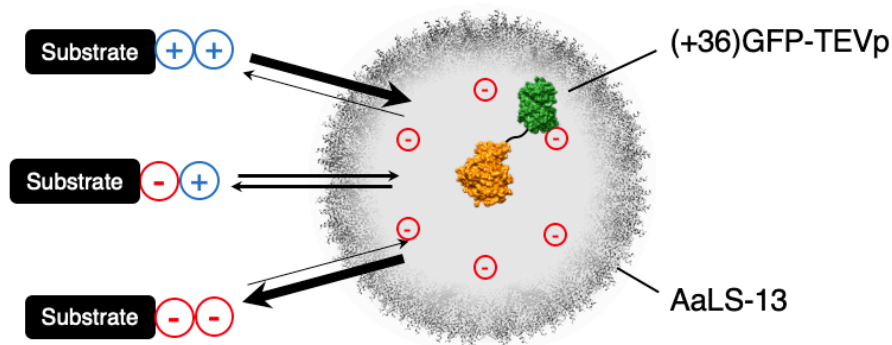
(B) Two-component "patchwork" protein capsid
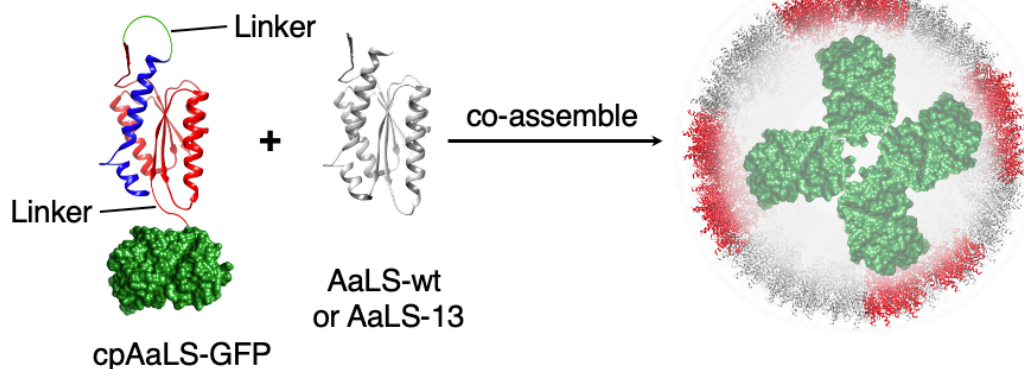
Figure 4.1 Engineered variants of AaLS enable selective incorporation and uptake of cargo molecules. **(A)** AaLS-13 protein capsid encapsulates TEVp fused to positively supercharged GFP. The anionic microenvironment within the cavity of the protein host enables selective uptake of cationic substrates.[142] **(B)** The N and C termini of cpAaLS are positioned on the inner surface of capsid forming units, which allow protein cargo to be encapsulated by genetic fusion. Co-assembly of cpAaLS with other AaLS variants affords mixed "patchwork" protein capsids.[131]

Further perturbation to the AaLS gene construct by circular permutation was performed to explore the determinants of protein capsid assembly and alternative modes of cargo loading.[131] Circularly permutated AaLS (cpAaLS) was produced by the addition of an artificial linker to connect the native N and C-terminus of wild type AaLS (AaLS-wt), and new termini were repositioned to the interior of the protein capsid. The rearranged construct allowed encapsulation of polypeptides that were genetically fused to cpAaLS subunits. Since circular permutation shuffles the polypeptide sequence of a protein with relatively small effect on three-dimensional structure overall, cpAaLS has been shown to co-assemble with the wild-type and other engineered AaLS variants (figure 4.1 B). Patchwork LS capsids resulting from the co-assembly of cpAaLS and AaLS-13 would allow both genetic fusion and electrostatic interactions as orthogonal strategies to internalisation of cargo molecules. The substrate sorting behaviour of AaLS-13 protein capsids provides a protein-based simulation to membrane bound organelles such as the vacuole. Consequently, the utility of AaLS-13 capsid as a mimic for AEP containing organelles was explored.

4.3   Results and discussion

4.3.1   Preparation recombinant AaLS-13 and GFP(+36)-TEVp/AaLS-13 complex

The reported findings of TEVp encapsulation by AaLS-13 were replicated by employing procedures and plasmids available in the literature.[141,142] The codon-optimised gene encoding for AaLS-13 was expressed by *E. coli* strain BL21 (DE3). The protein of interest was isolated *via* IMAC and SEC. Samples taken from the bacterial culture and purification process were analysed by SDS-PAGE. The band corresponding to the isolated protein of interest was positioned above the 18.4 kDa marker, which does not agree with the calculated weight of AaLS-13 (17.7 kDa) (figure 4.2 A). The negatively charged nature of AaLS-13 may interfere with the binding of SDS, resulting in a slower migration through the poly-acrylamide gel. Nevertheless, this outcome was in agreement with the results observed in previous studies described in the literature.[141,142] The 17.7 kDa protein was eluted with an exceptionally short retention time during SEC, which indicated the presence of a multimeric protein assembly (figure 4.2 B). Formation of AaLS-13 protein capsids were confirmed by

negative stain TEM, where ring-shaped particles with diameters around 40 nm were observed (figure 4.2 C).
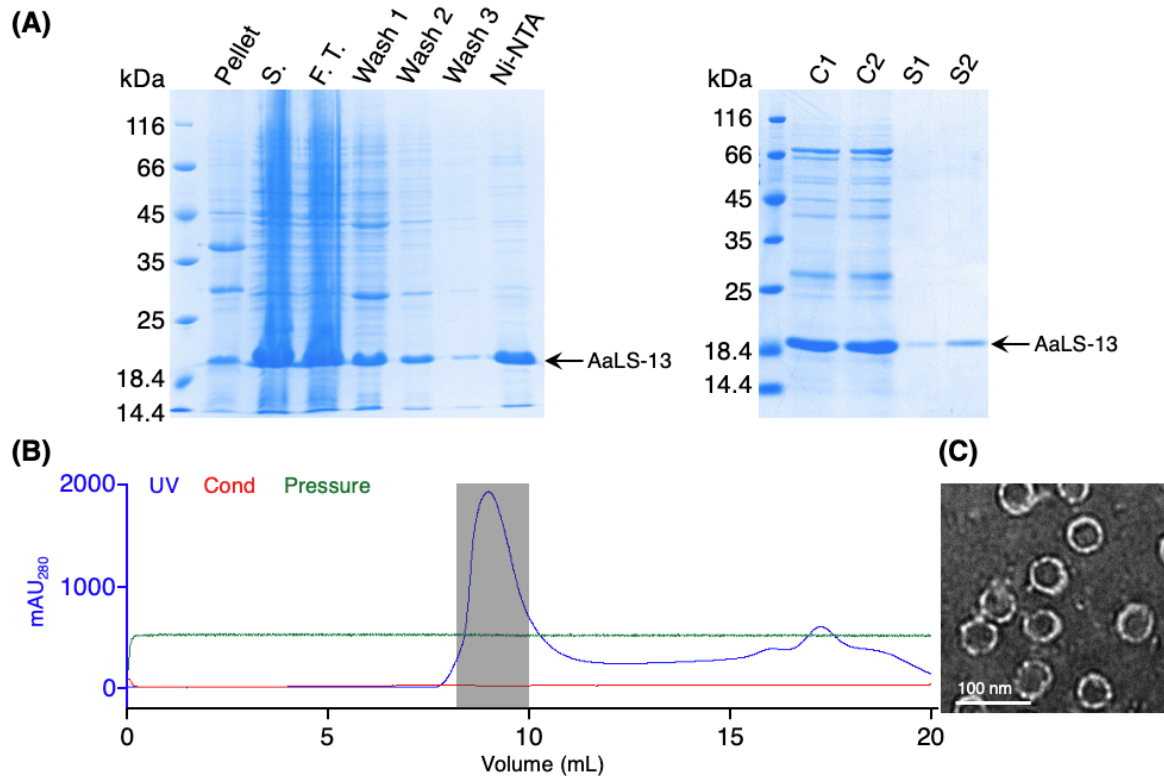
Figure 4.2 Purification of recombinant AaLS-13. **(A)** SDS-PAGE analysis of fraction obtained during purification. Arrows indicate bands corresponding to the protein of interest, AaLS-13. SDS-PAGE lanes: insoluble fraction of lysate (pellet), soluble fraction of lysate (S.), flow through from Ni^{2+}-NTA (F. T.), wash fractions from Ni^{2+}-NTA (wash 1-3), protein elution from Ni^{2+}-NTA (Ni-NTA), concentrated protein solution before size exclusion chromatography (C1 and C2), protein elution from SEC (S1 and S2). **(B)** Size exclusion chromatogram for AaLS-13. Grey shaded area corresponds to AaLS-13 capsids, which were collected and analysed by SDS-PAGE. **(C)** TEM image of the AaLS-13 cages. Scale bar = 100 nm

The fusion protein GFP(+36)-TEVp was also prepared according to the published protocols.[141,142] Encapsulation of GFP(+36)-TEVp was achieved by mixing the purified fusion protein and AaLS-13 capsids in aqueous buffer. With several negatively charged residues on the internal surface of AaLS-13, the lumen of the protein capsids has an anionic microenvironment. Consequently, attractive Coulombic forces enabled the uptake of the cationic fusion protein, GFP(+36)-TEVp, by the protein cage. The protein complexes were isolated by SEC and characterised by SDS-PAGE followed

by TEM (figure 4.3). Filled circular particles imaged by TEM indicate the presence of a cargo species in the lumen of the AaLS-13 capsids.
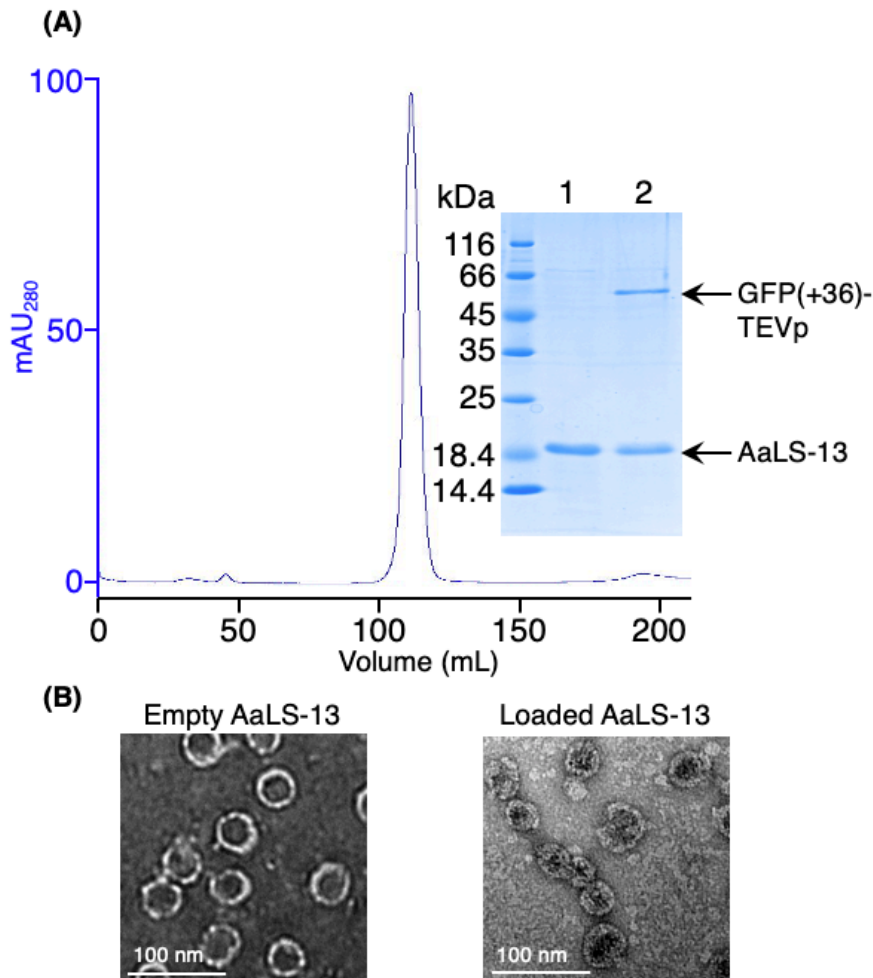
Figure 4.3 AaLS-13 capsids containing GFP(+36)-TEVp. **(A)** Size exclusion chromatogram for the mixture of GFP(+36)-TEVp and AaLS-13. SDS-PAGE analysis of AaLS-13 (lane 1) and GFP(+36)-TEVp/AaLS-13 complex isolated by SEC (lane 2). **(B)** TEM images of empty and loaded AaLS-13 capsids. Scale bar = 100 nm.

4.3.2 Peptidase activity of the encapsulated GFP(+36)-TEVp

While anionic microenvironment in the cavity of AaLS-13 facilitates the encapsulation of proteins fused to the positively charged GFP(+36), it also presented an effective substrate sorting mechanism. The activity of the encapsulated enzyme and the substrate sorting ability of the protein capsid were verified using a fluorescence assay. The substrate peptide, TEV-N, contains the canonical TEVp cleavage sequence, 2-

aminobenzoyl (Abz) fluorophore at the N-terminus and a dinitrophenol (Dnp) quencher on the ε-amine of a lysine residue near the C-terminus (figure 4.4 A). Fluorescence from the Abz group is quenched by Dnp *via* Förster resonance energy transfer (FRET) in the substrate peptide. Separation of the FRET donor and acceptor chromophores by TEVp-mediated peptide hydrolysis results in an increase in fluorescence intensity. A second peptide, TEV-K, bearing a positively charged Lys_6 tag at the C-terminus should favour its uptake by the AaLS-13 capsid, and thus access to the encapsulated enzyme.

The extent of peptide hydrolysis, and hence enzyme activity was estimated by monitoring the fluorescent intensity at regular time intervals. Protease activity was observed for the encapsulated and free GFP(+36)-TEVp fusion protein against both neutral and cationic substrate peptides (figure 4.4 B). The fluorescence intensity indicated both the encapsulated and the free GFP(+36)-TEVp turnover TEV-K at a similar rate. Lower fluorescence intensities were observed when the encapsulated enzyme was added to TEV-N, which indicate lower catalytic activity compared to the free enzyme (figure 4.4 B). In agreement with the findings reported in the literature,[142] encapsulation of TEVp within the AaLS-13 capsid reduces enzyme activity towards the neutral peptide TEV-N, resulting in selectivity towards the positively tagged peptide TEV-K. The protein cage demonstrated proteosome-like behaviour which selectively enables access to suitably tagged targets.
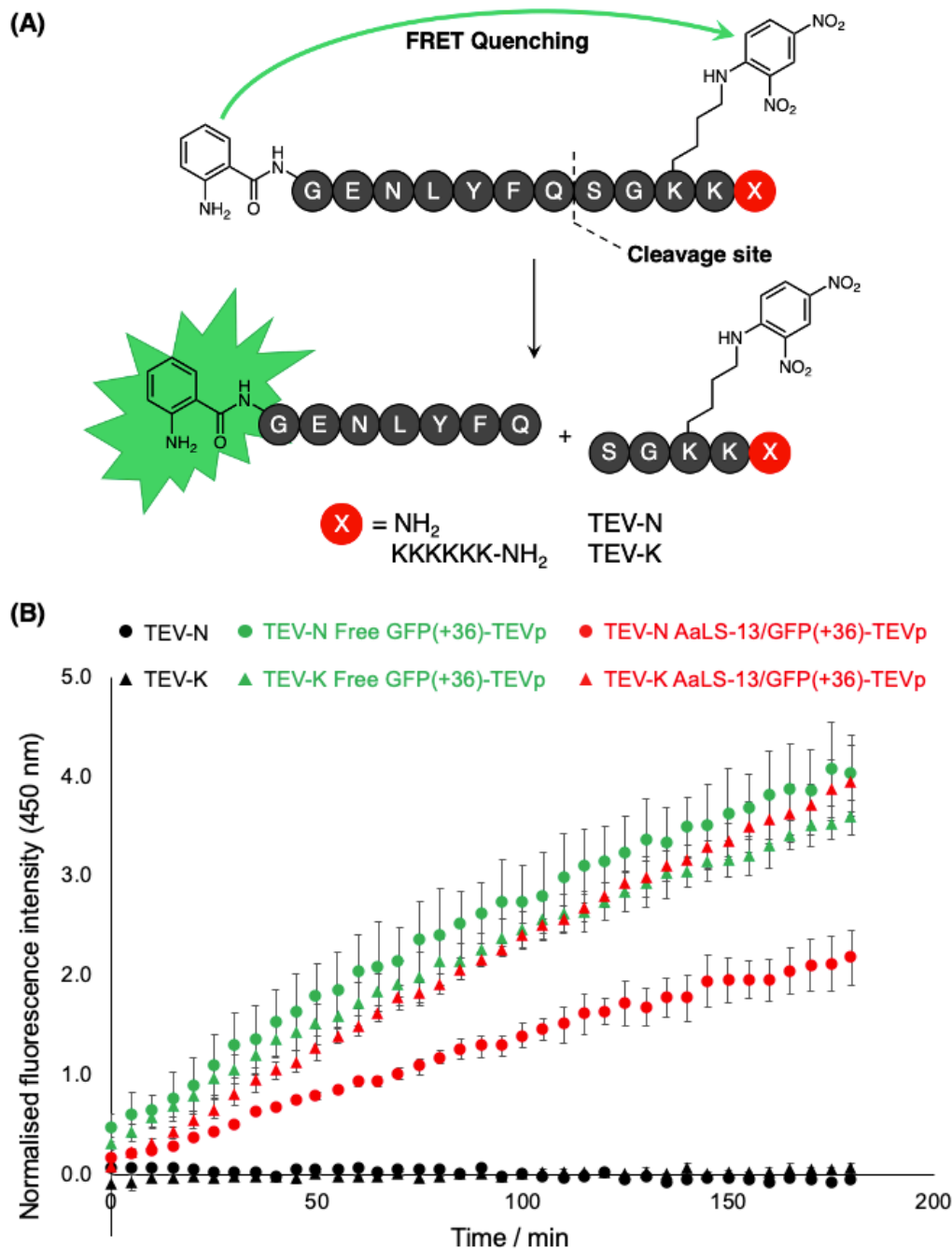
Figure 4.4 Activity assay for free and encapsulated GFP(+36)-TEVp towards TEV-N and TEV-K. **(A)** Schematic of TEVp-mediated peptide cleavage resulting in the separation of FRET donor and acceptor pair, which leads to an increase in fluorescence. Position X is either C-terminal amide or Lys_6 for the neutral (TEV-N) and cationic (TEV-K) substrates, respectively. **(B)** Fluorescent output of reaction mixtures with free or encapsulated GFP(+36)-TEVp towards TEV-N or TEV-K. The reactions were performed in triplicates at 25 ºC and pH 7.4. Reaction mixtures were excited at 350 nm and fluorescence measured at 450 nm. Data reported are the mean and normalised to the fluorescence intensity of the respective substrate peptides in the reaction buffer. Error bars represent the standard deviation from the mean.

4.3.3   Recombinant fusion protein of GFP(+36) and OaAEP1

Having validated the substrate selectivity of AaLS-13 encapsulated GFP(+36)-TEVp, the enzyme encapsulation system was adapted for OaAEP1. The gene encoding for His_6-GFP(+36) was appended upstream and in frame with the gene corresponding to OaAEP1. The newly assembled gene enabled the production of His_6-GFP(+36)-OaAEP1 by *E. coli*. The His_6 tag enables purification by IMAC and GFP(+36) provides an overall net positive charge, which facilitates loading into AaLS-13 capsids. Recombinant gene expression was induced by IPTG in *E. coli* strain BL21 (DE3) and the protein of interest was purified by IMAC followed by SEC. The protein containing fractions after each step were analysed by SDS-PAGE. Upon IPTG-induced gene expression, a band corresponding to the full length GFP(+36)-OaAEP1 fusion protein (79.6 kDa) was observed. However, a significant degree of undesired protein truncation was detected throughout purification. While a faint band positioned slightly above the 66 kDa marker remained, more distinctive bands were identified around 25 and 30 kDa. Two proteins were isolated in the subsequent purification by SEC. Unexpectedly, the protein eluted in the first peak (figure 4.5 B, peak 1) was shown to be the smaller protein identified at 25 kDa by SDS-PAGE and the larger protein was eluted later (figure 4.5 B, peak 2). The short retention time of lower molecular weight protein indicates a large hydrodynamic radius, which may be attributed to its quaternary structure *via* non-covalent interactions.
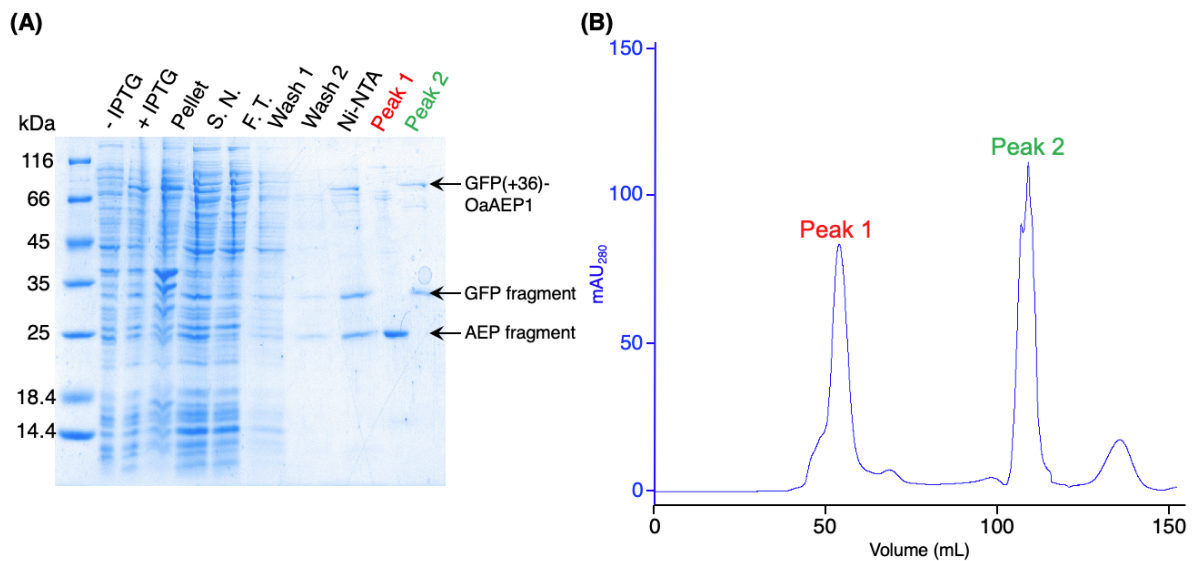
Figure 4.5 Purification of recombinant GFP(+36)-OaAEP1. **(A)** SDS-PAGE analysis samples obtained from the *E. coli* culture and protein purification. Arrows indicate bands corresponding to the protein of interest, GFP(+36)-OaAEP1 and proposed truncation fragments. SDS-PAGE lanes: *E. coli* culture before addition of IPTG (- IPTG), *E. coli* culture after incubation with IPTG (+ IPTG), insoluble fraction of lysate (pellet), soluble fraction of lysate (S.), flow through from Ni^{2+}-NTA (F. T.), wash fractions from Ni^{2+}-NTA (wash 1 and 2), protein elution from Ni^{2+}-NTA (Ni-NTA), protein elution from SEC (Peak 1 and 2), colour coded corresponding to peaks shown in chromatogram. **(B)** Size exclusion chromatogram for GFP(+36)-OaAEP1. Peak 1 and 2 correspond to fractions which were collected and analysed by SDS-PAGE.

To establish whether protein truncation can be avoided by changing the position of the supercharged protein, a new gene was designed to reverse the order of GFP(+36) and OaAEP1. In this alternative construct, the core domain of OaAEP1 was positioned at the N-terminus followed by GFP(+36) and a His_6 tag at the C-terminus. A tripeptide spacer (Pro-Gly-Ser) connects the AEP and GFP protein domains to afford the OaAEP1-GFP(+36) fusion protein. IPTG-induced gene expression was performed using *E. coli* strain BL21 (DE3) and IMAC was performed to isolate the protein of interest from the cell lysate. Similar to the previous protein construct, a faint band corresponding to the desired fusion protein (63 kDa) was observed (figure 4.6). However, smaller proteins were present in the protein elution from Ni^{2+}-NTA, which indicate protein truncation (figure 4.6). In particular, both OaAEP1 fusion protein constructs yielded a similar band positioned below the 35 kDa marker, which could be the same truncated protein species (figure 4.5 A and 4.6).
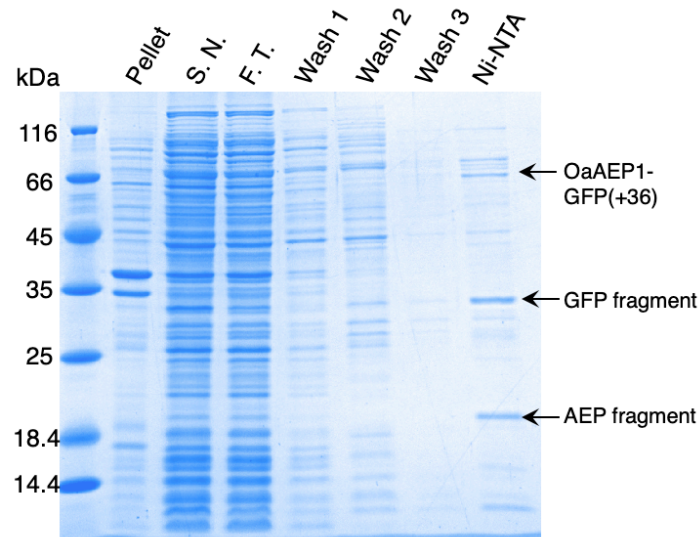
Figure 4.6 Analysis of recombinant OaAEP1-GFP(+36) by SDS-PAGE. Samples obtained from IMAC. Arrows indicate bands corresponding to the protein of interest, OaAEP1-GFP(+36), and proposed truncation fragments. SDS-PAGE lanes: insoluble fraction of lysate (pellet), soluble fraction of lysate (S.), flow through from Ni^{2+}-NTA (F. T.), wash fractions from Ni^{2+}-NTA (wash 1-3), protein elution from Ni^{2+}-NTA (Ni-NTA)

Short, positively charged peptide tags such as Arg_{10} have previously been employed as an alternative strategy to load cargo proteins into AaLS-13 capsids.[130] However, this was not tested, because short peptide tags were predicted to be ineffective in creating a net positive charge overall. The catalytic core domain of OaAEP1 contains a large number of negatively charged residues with an estimated pI of 5.11. (ProtParam, Expasy)

4.3.4 Stability of GFP(+36) in the presence of OaAEP1

An analysis of the peptide sequence of GFP(+36) revealed the presence of AEP-compatible recognition sites. Therefore, the undesired protein truncation may be associated with AEP activity, indicating that GFP(+36) may not be a compatible appendage to OaAEP1. To examine the stability of GFP(+36) in the presence of OaAEP1, the cationic fluorescent protein was prepared by IPTG-induced gene expression in *E. coli* strain BL21 (DE3). The protein was purified by IMAC, followed by SEC (figure 4.7 A and B).
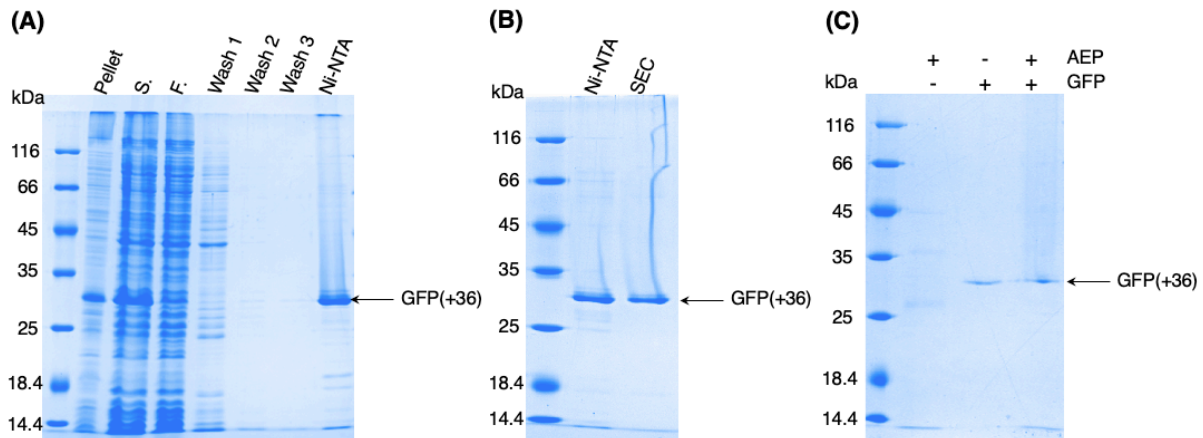
Figure 4.7 Analysis of recombinant GFP(+36) by SDS-PAGE. **(A)** Samples obtained from IMAC. SDS-PAGE lanes: insoluble fraction of lysate (pellet), soluble fraction of lysate (S.), flow through from Ni^{2+}-NTA (F. T.), wash fractions from Ni^{2+}-NTA (wash 1-3), protein elution from Ni^{2+}-NTA (Ni-NTA) **(B)** Protein samples before (Ni-NTA) and after size exclusion chromatography (SEC). **(C)** GFP(+36) stability towards OaAEP1 in 50 mM MES buffer (pH 6.0) with 50 mM NaCl, 1 mM EDTA and 0.5 mM TCEP. The reaction mixtures had a total volume of 100 μL, with final concentrations of GFP(+36) at 4 μM and OaAEP1 at either 2 or 0 μM. The mixture was incubated at 20 °C for 22 h. Arrows indicate bands corresponding to the protein of interest, GFP(+36).

No degradation of GFP(+36) (4 μM) was detected by SDS-PAGE after incubation with OaAEP1 (2 μM) overnight (22 h) at 20 ºC (figure 4.7 C). The lack of AEP-mediated hydrolysis on proteins bearing AEP-compatible recognition sequences was previously observed during the labelling of proteins such as β-lactamase and ubiquitin (see Chapter 3.3.4). This result suggests the locations of OaAEP1 recognition sites within the GFP(+36) backbone are not accessible to the enzyme. Nevertheless, this only suggests that GFP(+36) alone is resistant to the protease activity of OaAEP1. The proteins are held in close proximity in the fusion protein, and thus increase the effective concentration. OaAEP1 activity towards GFP(+36) may be observed at higher concentrations. Moreover, there may be subtle differences in the structure of the fusion protein which can reveal the OaAEP1 recognition sites for reaction. The cause of the truncations observed in the GFP(+36)-OaAEP1 fusion protein is currently unclear, further investigations, such as fragment analysis by MS, are required.

4.3.5 Circularly permutated AaLS facilitate OaAEP1 loading into AaLS-13 capsids

Given the production of fusion proteins with GFP(+36) and OaAEP1 was unsuccessful, another encapsulation approach was explored. A circularly permutated variant of AaLS (cpAaLS) was previously described, where the N and C-termini were moved from the exterior surface to the interior of the protein cage (figure 4.8 A).[131] The location of the newly introduced protein termini allows introduction of guest proteins by genetic fusion. Co-production of cpAaLS and AaLS-13 in *E. coli* were shown to afford patchwork protein co-assemblies with the fused cargo protein incorporated in the cavity of the protein capsid (figure 4.1).[131]

To apply the patchwork AaLS cage system for the encapsulation of OaAEP1, the gene for the core domain of OaAEP1 was appended downstream of the gene encoding for cpAaLS. The fusion gene was controlled by a tetracycline-induced gene expression system, which offered orthogonal expression control to the IPTG-induced gene expression for AaLS-13. When cpAaLS-OaAEP1 and AaLS-13 were co-produced in *E. coli* strain BL21 (DE3), cpAaLS-OaAEP1 was found to aggregate in inclusion bodies. Protein aggregation was potentially caused by mis-oxidation of the cystine pair present in OaAEP1. SHuffle® T7 express, a *E. coli* strain which promotes disulfide bond formation in the cytoplasm, was used to produce cpAaLS-OaAEP1 in the soluble form. A two-step purification procedure with IMAC followed by SEC was employed to isolate the protein assemblies. Although cpAaLS-OaAEP1 does not have a His$_6$-tag, the fusion protein was retained by Ni^{2+}-NTA (figure 4.8 C). The retention of cpAaLS-OaAEP1 by Ni^{2+}-NTA can be rationalised by its association with the His$_6$-tagged AaLS-13. When the protein mixture was further purified by SEC, the short retention time (8.0 mL) is indicative of AaLS capsid formation (figure 4.8 B). Analysis of the purified protein by SDS-PAGE revealed the presence of two proteins corresponding to the size of AaLS-13 and cpAaLS-OaAEP1. Finally, the formation of cpAaLS-OaAEP1/AaLS-13 complex was supported by the negative stain TEM analysis, where filled circular particles with diameters around 40 nm were observed (figure 4.8 D).
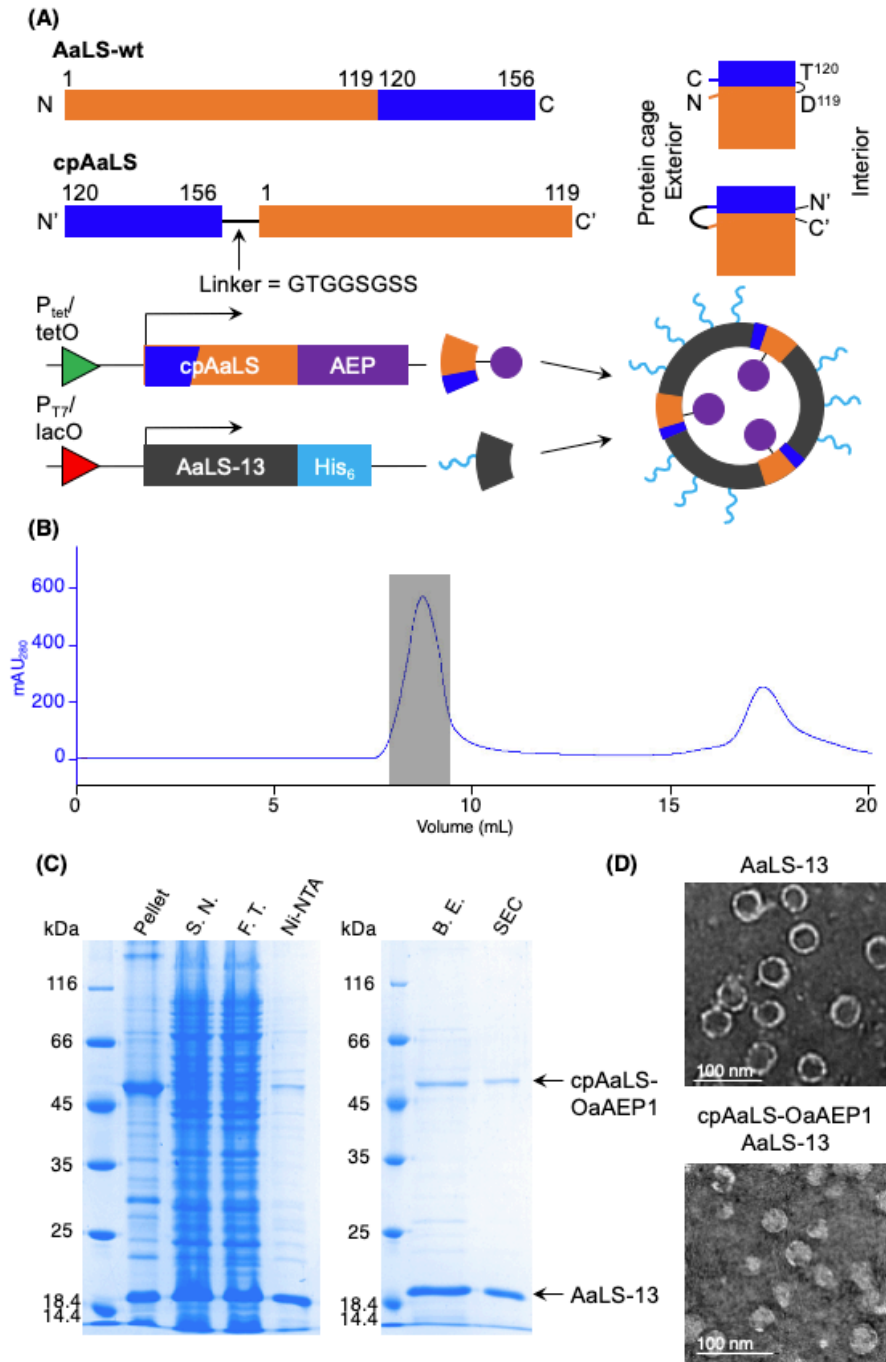
Figure 4.8 Production of AaLS-13 protein capsids containing OaAEP1. **(A)** Topological description of cp*Aa*LS with respect to *Aa*LS-wt and schematic detailing the co-production of cpAaLS-OaAEP1 and AaLS-13, which assembles as a patchwork protein cage. P_{tet}, tetracycline promoter; tetO, tetracycline operon; P_{T7}, T7 promoter; lacO, lactose operon. **(B)** Size exclusion chromatogram for the protein mixture. Grey shaded area corresponds to AaLS capsids, which were collected. **(C)** SDS-PAGE analysis of fraction obtained during purification. Arrows indicate bands corresponding to the protein of interest, cpAaLS-OaAEP1 and AaLS-13. SDS-PAGE lanes: insoluble fraction of lysate (pellet), soluble fraction of lysate (S.), flow through from Ni^{2+}-NTA (F. T.), protein elution from Ni^{2+}-NTA (Ni-NTA), imidazole removed by buffer exchange (B. E.), protein elution from SEC (SEC). **(D)** TEM images of empty and loaded AaLS-13 capsids. Scale bar = 100 nm.

4.3.6 Stability of patchwork protein cage in acidic conditions

While the activity OaAEP1 is known to be optimal in acidic conditions (see Chapter 2.3.2), AaLS-13 capsid were unstable and prone to precipitation at pH < 7.0.[130,140,154] Consequently, the stability of the patchwork cage complex (cpAaLS-OaAEP1/AaLS-13) was investigated. The protein cage was prepared according to the procedure described previously (Chapter 4.3.5), pH was lowered by buffer exchange through either SEC or rapid dilution and concentration after purification. Large multimeric protein capsids were obtained when purified at pH 7.0 and 7.4, indicated by the characteristic absorbance at 280 nm (A_{280}) eluting after 8 mL of buffer. A significantly reduced absorbance at the same retention time was observed when SEC was performed using a pH 6.5 buffer, which indicate the absence of large protein assemblies (figure 4.9 A). When acidic buffers were introduced by dilution and concentration, precipitation was observed. Analysis by SDS-PAGE indicated that the proteins of interest were only present in the insoluble precipitate (figure 4.9 B). The findings reported here confirmed that AaLS protein cages are unstable in acidic conditions. Fortunately, OaAEP1 has a broad pH tolerance and remains active at neutral pH conditions. Subsequent studies in the activity of the encapsulated AEP were performed at pH 7.0 or above.
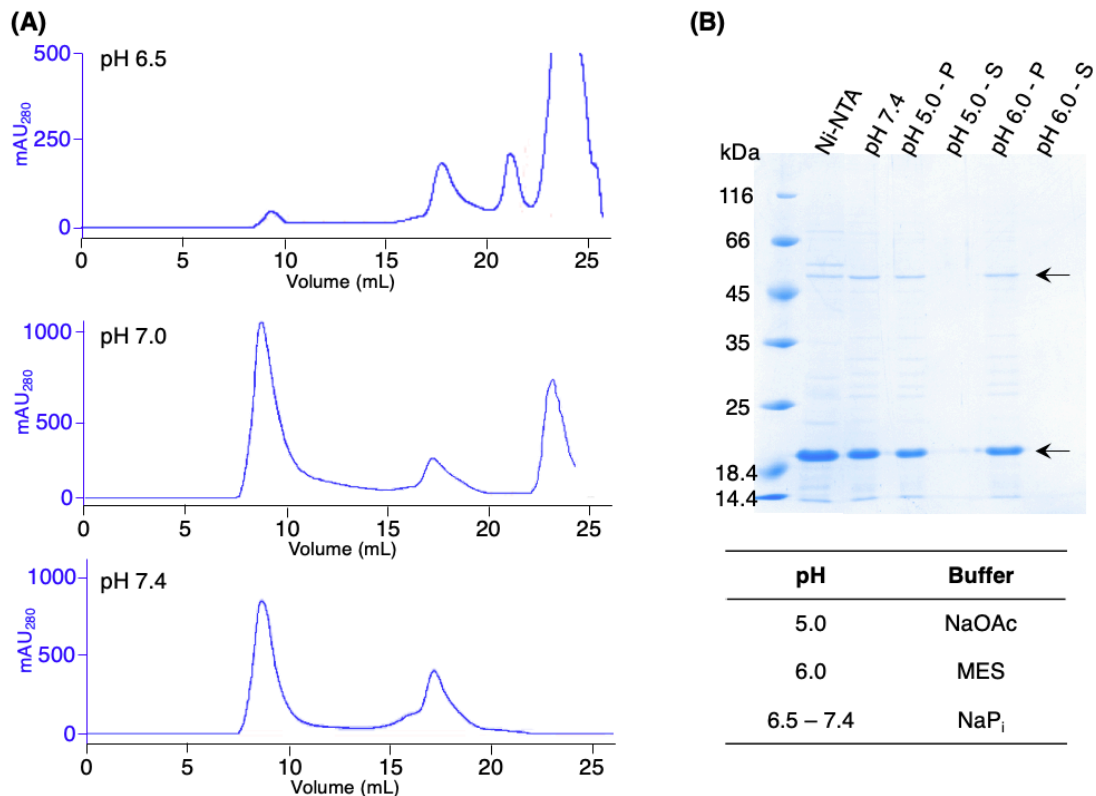
Figure 4.9 Patchwork AaLS protein capsids were unstable in acidic conditions. **(A)** Size exclusion chromatogram for SEC equilibrated and run at pH 6.5-7.4. Characteristic peaks around 8-9 mL indicate AaLS capsid formation. **(B)** AaLS capsids obtained from SEC (pH 7.4) were acidified by buffer exchange. The resultant solution and precipitates were analysed by SDS-PAGE. Arrows indicate bands corresponding to the protein of interest, cpAaLS-OaAEP1 and AaLS-13. SDS-PAGE lanes: protein elution from Ni^{2+}-NTA (Ni-NTA), protein elution from SEC (pH 7.4), precipitate after buffer exchanged to pH 5.0 (pH 5.0 – P), soluble fraction after buffer exchanged to pH 5.0 (pH 5.0 – S), precipitate after buffer exchanged to pH 6.0 (pH 6.0 – P), soluble fraction after buffer exchanged to pH 6.0 (pH 6.0 – S)

4.3.7 Enzyme activity of the patchwork protein cage containing OaAEP1

With the encapsulated AEP in hand, the activity of the enzyme was assessed using the peptide cyclisation reaction described in Chapter 2 (reaction 2.1, figure 2.3). As expected, the free OaAEP1 readily converted linear substrate (GLPVSTKPVATRNGL, pep-M) into the cyclic product peptide (figure 4.10). In contrast, when the protein complex was incubated with the substrate peptide, only a small peak corresponding to the cyclic product peptide was observed in the HPLC chromatogram monitoring UV$_{210}$ absorbance (figure 4.10). The low conversion of the linear substrate to the cyclic product by the encapsulated OaAEP1 here concur with the encapsulated TEVp

system and suggest encapsulation by AaLS-13 does indeed prevent access to the untagged peptide substrate.
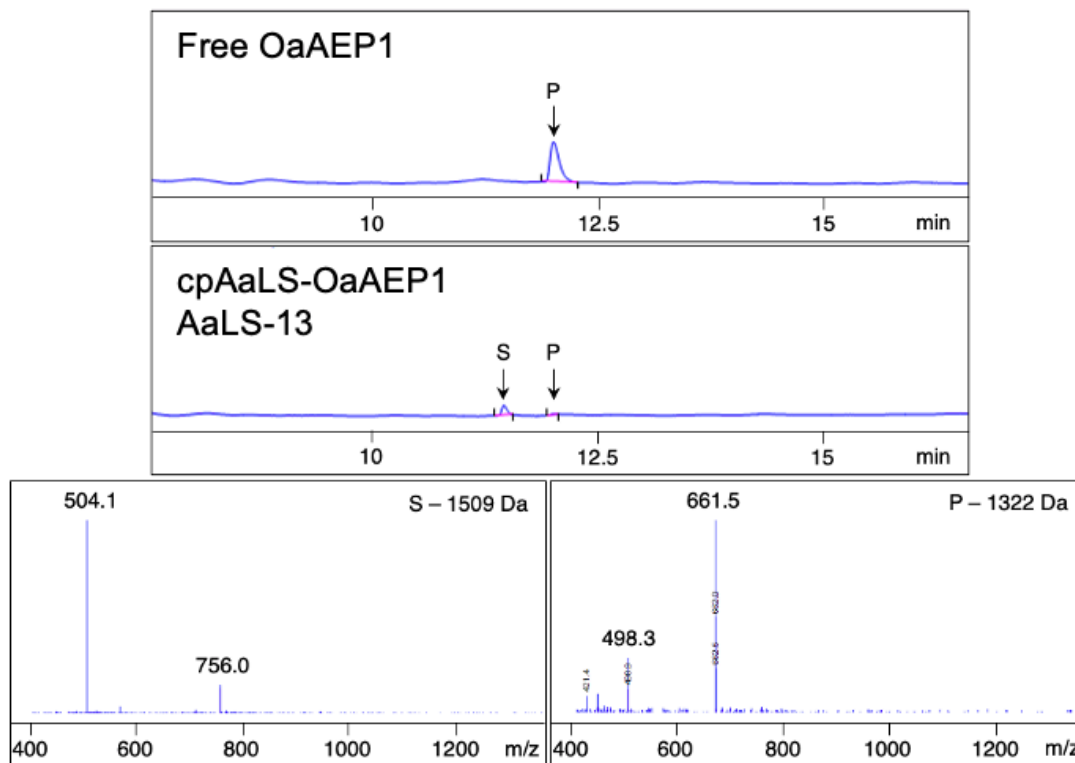
Figure 4.10 Peptide cyclisation activity of the encapsulated OaAEP1 towards pep-M assessed by HPLC-MS. **(Top)** Chromatogram monitoring UV absorbance at 210 nm. **(Bottom)** Mass spectrums of the annotated peaks from cpAaLS-OaAEP1/AaLS13. S = starting material (GLPVSTKPVATRNGL, 1509 Da). P = product (*cyclic*-GLPVSTKPVATRN, 1332 Da). For the reactions, 500 µM of the substrate peptide were incubated with OaAEP1 (0.05 µM) or cpAaLS-OaAEP1 (0.1 µM) at 20 ºC and pH 7.0. Full conversion to the cyclic product was observed after 1 h when the free OaAEP1 was used, whereas traces of cyclic product were observed only after 20 h with the encapsulated enzyme.

A Lys$_6$-tagged peptide (GLPVSTEPVATENCLGKKKKKK, pep-K) was employed to examine the activity of the encapsulated enzyme towards positively charged substrates. Given the negatively charged interior of the AaLS-13 protein cage, Coulombic attraction was expected to enable the uptake of the cationic pep-K, and thus access to the encapsulated OaAEP1. However, peptide cyclisation activity by cpAaLS-OaAEP1 was poor. No desired cyclic peptide product was observed in the UV chromatogram at 210 nm. However, a weak signal with the target mass was detected when the sample was analysed by HPLC-MS using single ion monitoring (figure 4.11). The observed signal was too weak to provide a reliable estimation for product conversion. Reaction conditions such as reaction temperature, pH and ionic

strength were explored for pep-K, but no significant increases in conversion were found.
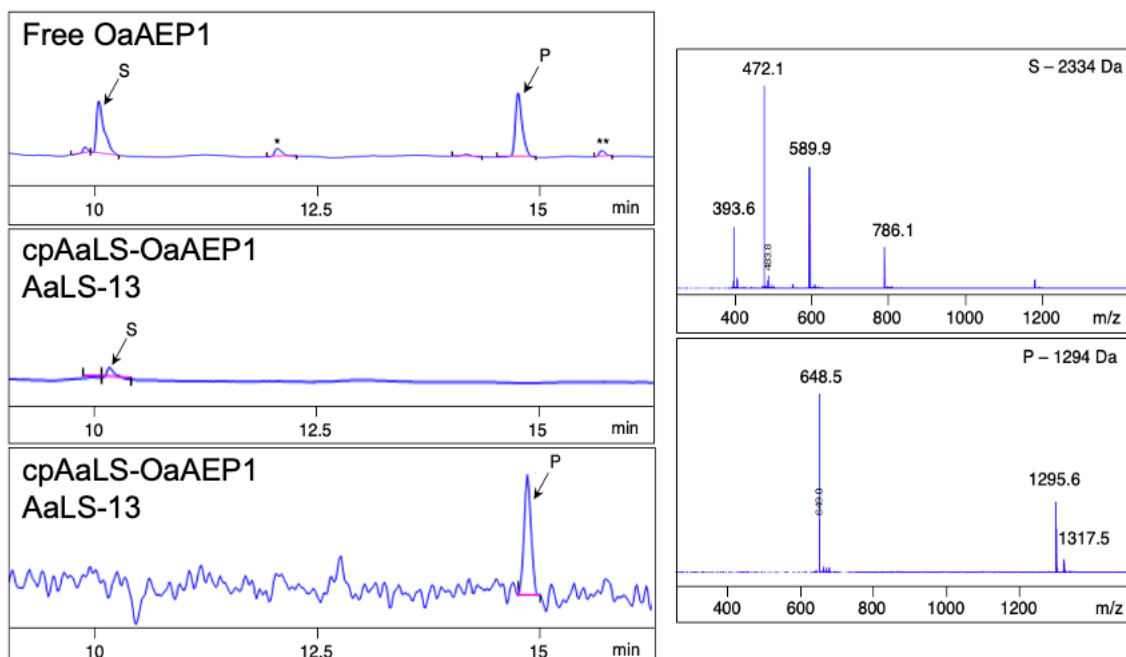
Figure 4.11 Cyclisation activity of the encapsulated OaAEP1 towards pep-K assessed by HPLC-MS. **(Top and middle left)** Chromatogram monitoring UV absorbance at 210 nm. **(Bottom left)** Chromatogram with single ion monitoring for the target product mass. **(Right)** Mass spectrums of the annotated peaks from free OaAEP1. S = starting material (GLPVSTEPVATENCLGKKKKKK, 2334 Da). P = product (*cyclic*-GLPVSTEPVATEN, 1294 Da). Intermolecular peptide ligation and cyclic dimer products were annotated by * and **, respectively. For the reactions, 500 µM of the substrate peptide were incubated with 0.05 µM of OaAEP1 or 0.1 µM of cpAaLS-OaAEP1 at 20 ºC and pH 7.0. Conversion to the cyclic product was observed after 1 h when the free OaAEP1 was used, whereas traces of cyclic product were observed only after 20 h with the encapsulated enzyme.

A potential cause of poor AEP activity is the amino acid sequence of pep-K. In addition to the Lys_6 tag, there are other differences between pep-K and pep-M. The difference in recognition sequences between pep-K (Asn-**Cys**-Leu) and pep-M (Asn-**Gly**-Leu) should have minimal affect to AEP catalysis, as the substrate recognition by OaAEP1 at P1' was shown to be non-specific in Chapter 3 (figure 3.2 A). On the other hand, the presence of Glu at the P2 position (residue preceding Asn) for pep-K instead of Arg in pep-M may have a more profound effect. Although AEP substrate recognition has been widely reported to be three residues in length (P1-P1'-P2') starting with Asn,[22,23] some studies have suggested residues beyond the putative recognition sequence may impact activity.[22,27] Substitution at the P2 position from Arg to Ala was

reported to impair the activity of OaAEP1b, the corresponding wild type enzyme to the AEP employed here.[22] Therefore, the activity of the encapsulated OaAEP1 was re-examined using a different set of substrate peptides.

In addition, stopped assay monitored by HPLC are low throughput and relatively insensitive. Therefore, the catalytic activity of the encapsulated AEP was also examined using a fluorescence assay. The substrate peptides employed contain a FRET pair which is similar to those adopted for the TEVp activity assay described above (see Chapter 5.2.2). The peptide, FRET-N, contains the OaAEP1 recognition sequence, Abz fluorophore at the N-terminus and a Dnp quencher on the ε-amine of a lysine residue near the C-terminus (figure 4.12 A). A large excess (10 equivalents) of a tripeptide, Gly-Leu-Pro, was added to the assay mixture to enable peptide ligation. The transpeptidation activity of OaAEP1 would result in the separation of the FRET donor and acceptor chromophores leading to an increase in fluorescence intensity (figure 4.12 A). Addition of a Lys_6 tag to the C-terminus of FRET-N gives rise to FRET-K, which was employed to examine the activity of the encapsulated AEP system towards positively charged substrates.
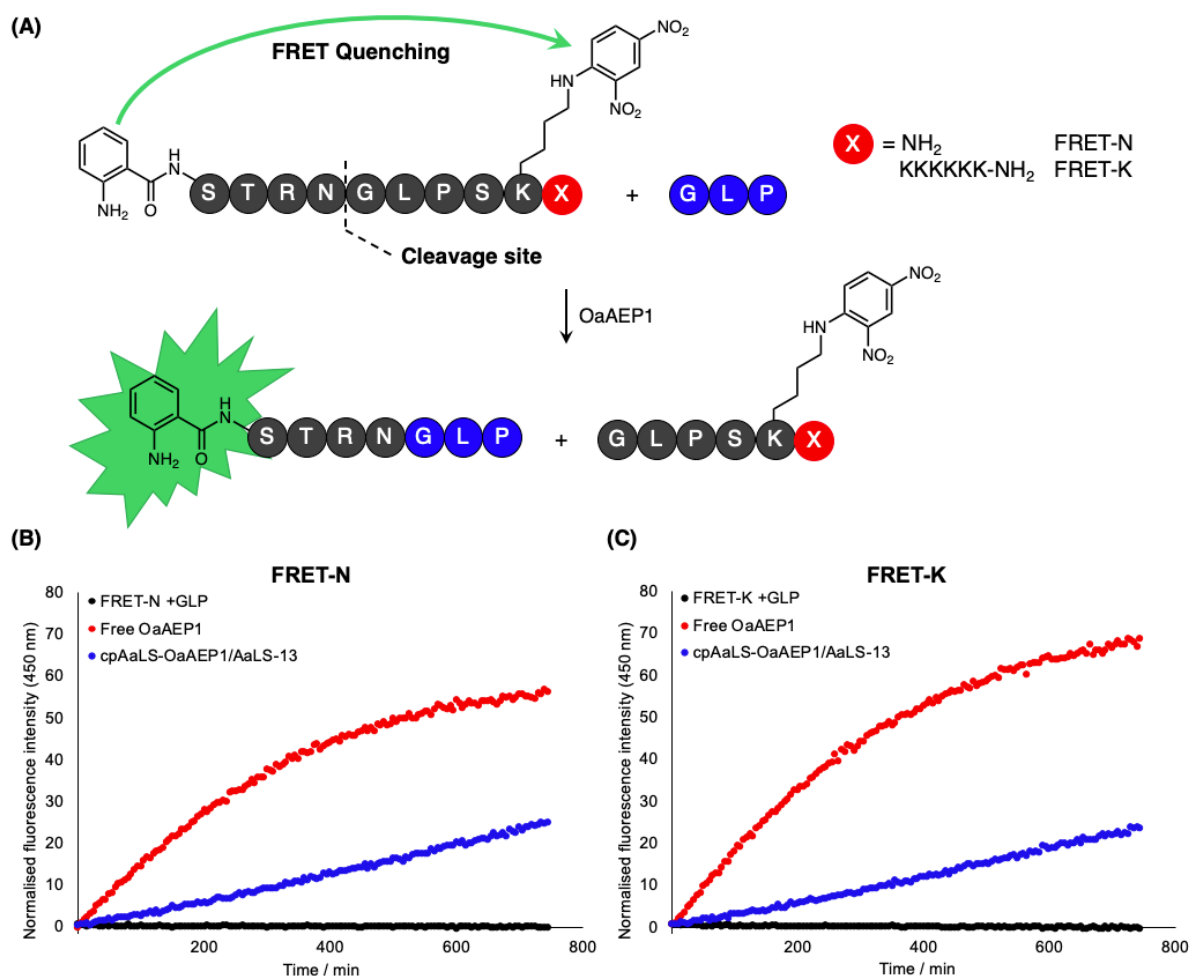
Figure 4.12 Activity assay by fluorescence for free and encapsulated OaAEP1 towards FRET-N. **(A)** Schematic of AEP-mediated transpeptidation resulting in the separation of FRET donor and acceptor pair, which leads to an increase in fluorescence. Position X can be either C-terminal amide or hexa-lysine for the neutral (FRET-N) and cationic (FRET-K) substrates respectively. **(B)** Normalised fluorescent output of reaction mixtures with free OaAEP1 (0.05 µM) or cpAaLS-OaAEP1/AaLS-13 (1.3 µM) towards FRET-N, and **(C)** FRET-K. The reactions were performed at 25 ºC and pH 7.0. Reaction mixtures were excited at 350 nm and fluorescence measured at 450 nm. Data reported are normalised to the fluorescence intensity of the substrate peptides in the reaction buffer.

An increase in fluorescence intensity was observed when FRET-N and FRET-K were incubated with the cpAaLS-OaAEP1/AaLS-13 complex and the tripeptide GLP (figure 4.12 B and C). This indicates that the encapsulated enzyme is active. The observed activity may be attributed to the relatively high enzyme concentration employed (1.3 µM *vs.* 0.1 µM in the HPLC assay). Nevertheless, the transpeptidation activity of cpAaLS-OaAEP1 was found to be significantly lower than the free enzyme, as shown by the difference in fluorescence output (figure 4.12 B and C). A 1:100 ratio between

cpAaLS-OaAEP1 and AaLS-13 was assumed for all patchwork cage systems described here. However, the actual ratio of cpAaLS to AaLS-13 is unknown. Errors in the estimation of enzyme concentration is a potential reason for the observed decrease in activity. The rate of increase in fluorescence intensity were similar when comparing the activity of cpAaLS-OaAEP1 towards either substrate (figure 4.12 B and C), which suggest a lack of substrate sorting by the patchwork AaLS protein compartment.

4.3.8 Enzyme activity of the patchwork protein cage containing TEVp

Substrate sorting by AaLS-13 was observed when TEVp was encapsulated by fusion to GFP(+36).[142] While AaLS-13 should provide similar sorting properties to the patchwork AaLS system, no data regarding the substrate sorting ability of the patchwork protein container were available. To ascertain the selective uptake of charged substrate molecules by the cpAaLS/AaLS-13 patchwork cage system, OaAEP1 was substituted for TEVp, an enzyme with proven activity in AaLS-13 protein capsids.

The gene encoding for TEVp was appended downstream of the gene for cpAaLS. Echoing the preparation of the encapsulated AEP system, expression of the cpAaLS-TEVp gene was controlled by a tetracycline-induced expression system. Co-production of cpAaLS-TEVp and AaLS-13 was performed by IPTG and tetracycline induced gene expression in *E. coli* strain BL21 (DE3). A two-step purification procedure with IMAC followed by SEC was employed to isolate the protein assemblies (figure 4.13).
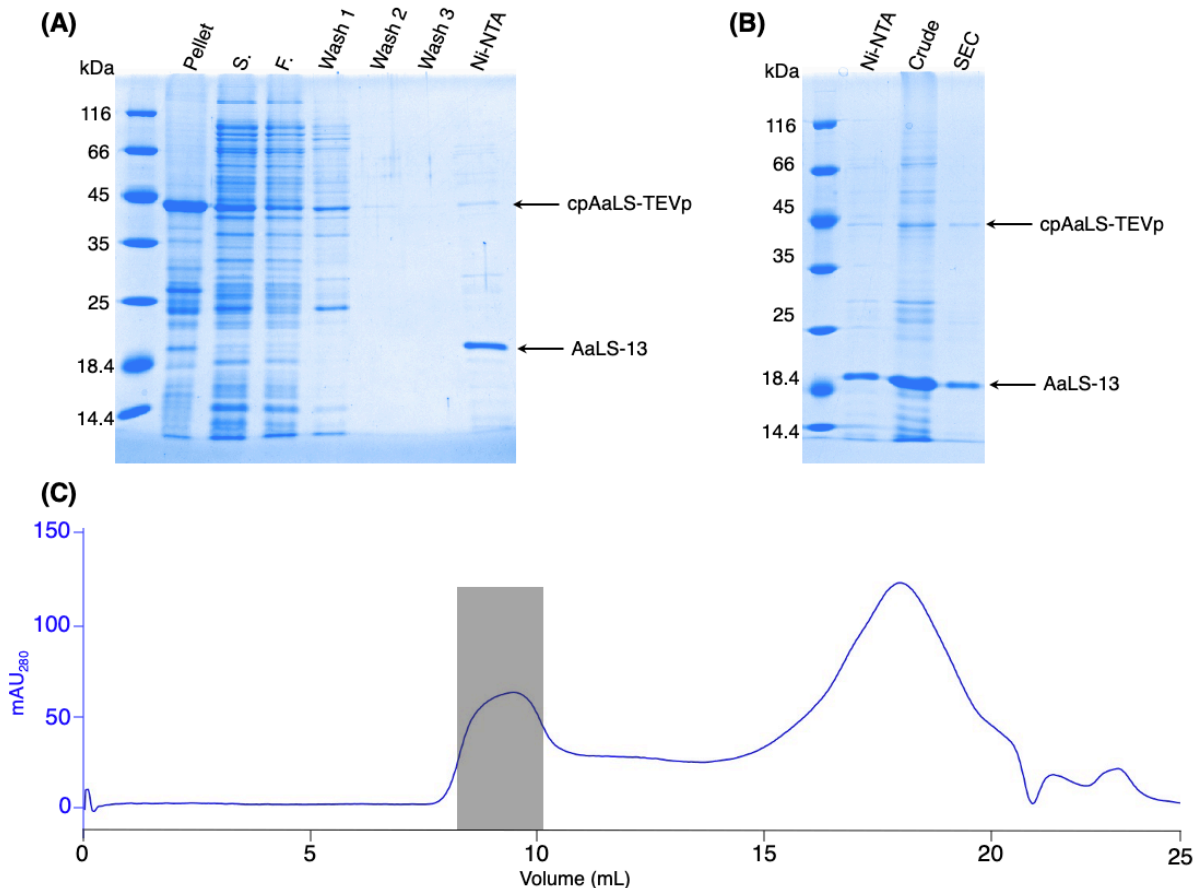
Figure 4.13 Preparation of cpAaLS/AaLS-13 protein capsids containing TEVp. **(A)** Samples obtained from IMAC. SDS-PAGE lanes: insoluble fraction of lysate (pellet), soluble fraction of lysate (S.), flow through from Ni²⁺-NTA (F. T.), wash fractions from Ni²⁺-NTA (wash 1-3), protein elution from Ni²⁺-NTA (Ni-NTA) **(B)** Protein samples before (Ni-NTA), before (Crude) and after size exclusion chromatography (SEC). Arrows indicate bands corresponding to the proteins of interest. **(C)** Size exclusion chromatogram for the protein mixture. Grey shaded area corresponds cpAaLS-TEVp/AaLS-13 capsids, which were collected.

Fluorescent peptides TEV-N and TEV-K were employed to assay the peptide cleavage activity of the cpAaLS/AaLS-13 encapsulated TEVp. Enzyme activity was estimated by monitoring the fluorescent intensity at regular time intervals. Protease activity was observed for the encapsulated and free TEVp against both neutral and cationic substrate peptides (figure 4.14). The slower rate of increase in fluorescence intensity indicate the encapsulated TEVp turnover both substrates at a slightly lower rate compared to the free enzyme (figure 4.14). However, the unknown ratio of cpAaLS and AaLS-13 within the patchwork protein cage system prevents an accurate estimation of enzyme concentration. Similar to the patchwork OaAEP1 system, a

1:100 ratio between cpAaLS-TEVp and AaLS-13 was assumed. Therefore, the apparent decrease in activity may be due to an overestimation of the TEVp concentration in the encapsulated system. The comparable enzyme activities towards both TEV-N and TEV-K indicate a lack of substrate sorting by the cpAaLS/AaLS-13 protein capsid.
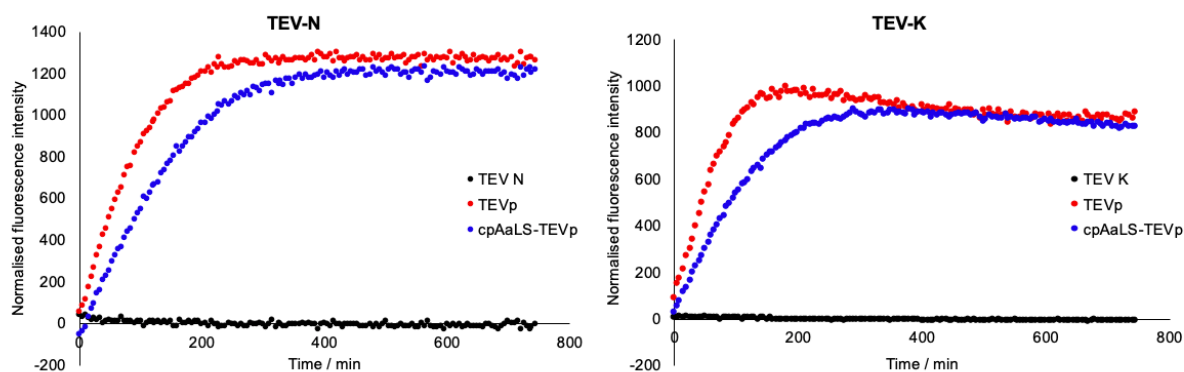
Figure 4.14 Activity of cpAaLS-TEVp/AaLS-13 patchwork protein cage. Normalised fluorescent output of reaction mixtures with free TEVp (100 nM) or cpAaLS-TEVp/AaLS-13 towards TEV-N and TEV-K (both 100 μM). The reactions were performed at 25 ºC and pH 7.4. Reaction mixtures were excited at 350 nm and fluorescence measured at 450 nm. Data reported are normalised to the fluorescence intensity of the substrate peptides in the reaction buffer.

4.4 Conclusion

Multimeric protein assemblies give rise to subcellular architectures that bear resemblance to membrane bound organelles.[113,159] Lumazine synthase isolated from *A. aeolicus* is an example of a cage forming protein.[130] The engineered variant, AaLS-13, assembles with a negatively charged interior which offers a simple strategy for loading guest molecules.[140,141] Moreover, the electrostatic nature of AaLS-13 protein compartments provide a substrate sorting effect, thus enhancing the substrate selectivity of encapsulated enzymes.[142] The AaLS-13 protein compartment presented an opportunity to host OaAEP1. A multimeric AaLS-13 assembly containing OaAEP1 represents an artificial organelle, mimicking the plant vacuole which imparts additional substrate specificity to the enclosed enzyme.

The substrate sorting nanoreactor system has been previously demonstrated using TEVp,[142] and replicated here. Unfortunately, significant protein truncation was observed when the reported method of enzyme encapsulation was applied to OaAEP1. The positively charged protein tag, GFP(+36), was not truncated when incubated with OaAEP1. This evidence suggests that the observed truncation of the fusion protein may not be associated with the protease activity of OaAEP1. However, the effective concentrations of OaAEP1 and GFP(+36), and the tertiary structures of the fusion protein may differ from individual proteins. Consequently, further investigations, such as fragment analysis by MS, are required to identify the cause of the truncation.

An alternative strategy was adopted to encapsulate OaAEP1 using another engineered variant of AaLS which has been circularly permutated (cpAaLS). Genetic fusion of cpAaLS to OaAEP1 and co-production with AaLS-13 in *E. coli* enabled the formation of patchwork AaLS protein compartments containing OaAEP1. Initial examination of the encapsulated enzyme for peptide cyclisation activity showed a significant reduction compared to the free OaAEP1. While this is expected for pep-M due to the substrate sorting effect of the AaLS-13 protein cage, poor activity towards pep-K was confounding. A rationale for the lack of activity towards pep-K may be the presence of an unsuitable amino acid residue at the P2 position. The activity of the encapsulated AEP complex was re-examined using internally quenched fluorescent peptides with a modified substrate recognition sequence (FRET-N and FRET-K). The fluorescence-based assay enables reaction monitoring in real time, offering significant advantage over the HPLC-based stop assay which is substantially more time-consuming.

The encapsulated enzyme was shown to be active towards both FRET-N and FRET-K in the fluorescence-based assay. However, cpAaLS-OaAEP1 processed the substrate peptides at significantly lower rate when compared to the free enzyme. Weak AEP activity may be due to low incorporation of cpAaLS-OaAEP1 in the patchwork AaLS protein container, thus resulting in an overestimation of enzyme concentration. Similar rates of increase in fluorescence were observed for both peptide substrates, indicating a lack of substrate sorting by the patchwork protein

container. Another enzyme, TEVp, was used as a control to verify the cpAaLS/AaLS-13 system for the encapsulation of active enzymes. Similar peptidase activities were observed between the cpAaLS-TEVp/AaLS-13 and the free TEVp construct. This finding indicated the AaLS encapsulation system did not limit the activity of the encapsulated enzyme, and thus precludes encapsulation as a reason for the lowered activity. However, the cpAaLS-TEVp/AaLS-13 system also demonstrated comparable activities towards both neutral and positively charged substrates, implying that the patchwork protein cage may not offer the same substrate sorting capacity demonstrated by AaLS-13 protein capsids. Consequently, further investigations such as steady-state kinetics studies are required to quantify any substrate preference by the encapsulated enzymes in the patchwork protein compartments.

5 General conclusions and future work

Protein bioconjugation is an essential element of biochemical research and critical to the growing trend of protein therapeutics.[80,145,146] Thus, there is an ongoing demand for efficient and site-specific conjugation methods.[80,91–93,143,144,147–149,160] Given proteins are the intended substrates, reactivity in relatively mild conditions is also desirable. The characteristics of ligase-type AEPs present these enzymes as an ideal tool to facilitate the assembly of protein conjugates. Using OaAEP1-C247A (referred to as OaAEP1) as a model AEP, this thesis aimed to address some key challenges which limit the application of AEPs:

1. **To establish a simplified protocol for the preparation of recombinant AEPs.** The preparation of recombinant AEPs involves multiple purification steps and an incubation step at low pH,[14,22,39] which are labour intensive and time consuming. This issue was addressed in **Chapter 2** by removing the gene encoding for the pro-domain. OaAEP1 was prepared according to a procedure reported in the literature.[14] The catalytically active core domain was analysed by ESI-MS and the pH tolerance of OaAEP1 was investigated by examination of peptide cyclisation kinetics at different pH values. Subsequently, new gene constructs encoding for OaAEP1 without the pro-domain were designed to enable a streamlined purification protocol. The functional role of the pro-domain was also investigated by introducing non-covalently bound pro-domain of OaAEP1 during recombinant expression. A simplified protocol for the preparation of active OaAEP1 was described. Improved access to active AEP would facilitate greater uptake of the enzyme for *in vitro* applications.

2. **To develop a strategy to address the reversibility of AEP-driven intermolecular peptide ligation**. Intermolecular peptide ligations catalysed by AEPs are reversible, and thus require a large excess of labelling peptide substrate to achieve high product conversion.[37,42,69,98,105] This limitation was addressed through the development of a chemo-enzymatic strategy described in **Chapter 3**. Substrate recognition by OaAEP1 at the P1' and P2' positions was examined using a peptide ligation assay. Utilising the non-specific substrate recognition at P1' by OaAEP1, the enzymatic ligation was coupled to

a N-terminal cysteine specific reaction with FPBA. Factors which promote protease activity of AEPs, resulting in hydrolytic side reactions, were identified and discussed. The chemo-enzymatic method was shown to enable peptide and protein ligation to proceed with high conversions while using relatively low equivalents of labelling peptide. This method is particularly applicable when using expensive or non-commercially available labels, such as radioactive or isotopic probes.[43,97,109,110]

3. **Explore the applications of AaLS-13 protein capsids as an artificial organelle to sequester and regulate the activity of activated AEPs.** Activated AEPs are under strict spatial control in nature. Localised in the vacuole of plant cells, protease activity of the AEPs are associated with senescence and programmed cell death.[2,9,11,13,14] Therefore, the accumulation of active AEPs in the cytoplasm of recombinant hosts such as *E. coli* is likely to be toxic and limits the development of *in vivo* application of AEPs such as whole cell catalysis. Encapsulation of the enzyme to mimic the native localisation of AEPs within organelles present a potential solution to this limitation. The encapsulation of OaAEP1 within AaLS-13 protein capsids was described in **Chapter 4**. Two methods of cargo loading into AaLS-13 hosts were explored. The preparation of a patchwork AaLS protein capsid which contained OaAEP1 was described. The activity and substrate selectivity of the encapsulated enzyme was examined. Although the encapsulated OaAEP1 was shown to process peptide substrates, the activity was lower than the free enzyme. Moreover, the cpAaLS/AaLS-13 protein container did not appear to offer the desired substrate sorting capacity. Consequently, further investigations are required to develop an encapsulated AEP system which resembles an organelle.

5.1 Future work

Access to AEPs for *in vitro* applications. A simplified procedure for the preparation of OaAEP1 was established, and thus addressed one of the key limitations to the adoption of AEPs for biocatalytic applications. Nevertheless, the yield of active enzyme remains low (2 mg/L). Avenues to improve the production of recombinant

AEPs would provide applicational value. The preparation of split-OaAEP1 and the potential role of the pro-domain as a protein chaperone is one route for further exploration. A second way to improve the yield of recombinant AEP would be to explore alternative recombinant expression hosts such as yeast (*P. pastoris*).[70]

A chemo-enzymatic method using AEPs in protein bioconjugation. The chemo-enzymatic protein labelling strategy outlined in Chapter 3 offers a proof-of-concept (POC). To progress from a POC to an applicable and reliable methodology, the chemo-enzymatic approach should be tested against a larger, more diverse library of proteins and labels to identify further limitations. For example, compatibility with cysteine-rich proteins such as antibodies can be investigated. The use of FPBA as a scavenger offer an economical and versatile solution to the reversible OaAEP1 ligation. Reversibility is a concurrent in many enzymatic methods for protein bioconjugation. Thus, the use of FPBA as a scavenger may be applicable to other enzymatic methods, providing the enzymatic reaction generates a by-product bearing a N-terminal cysteine. The reactivity of the aldehyde from FPBA was critical to the success of the method described here. However, reactive aldehydes may have limited selectivity and applications, such as in cellular contexts or towards glycosylated peptide labels.[89] Therefore, alternative scavenger systems should be explored to further expand the chemo-enzymatic approach to AEP catalysis. For example, the use of divalent Ni^{2+} ions to sequester the nucleophilic by-product peptide and improve ligation by OaAEP1 was recently reported.[100]

Encapsulation of AEPs to create a peptide-processing artificial organelle. The preparation of an encapsulated OaAEP1 protein complex was described. However, enzyme activity was significantly lower than the free enzyme. This is likely to be caused by an overestimation of enzyme concentration as the ratio of cpAaLS and AaLS-13 within the patchwork protein cage system was unclear. Therefore, a strategy to measure enzyme concentration, and thus the incorporation rate of cpAaLS into the patchwork protein capsid would be critical for further research. For example, the enzyme could be labelled with a hemagglutanin (HA) tag to enable quantitative analysis by western blot. Although the encapsulated OaAEP1 demonstrates activity

towards the chosen peptides, selectivity towards the substrate with a positively charged tag was not observed. Substrate sorting was also not detected with cpAaLS-TEVp, which indicated that the patchwork protein capsid may not exhibit the same substrate selective properties of capsids consisting of only AaLS-13. Protein encapsulation may require further validation. For example, specific labelling of the cargo AEP with a fluorophore could enable visualisation by super-resolution microscopy. Investigations towards the reaction kinetics of the encapsulated enzymes towards will be required to quantify the substrate sorting capacity of the patchwork protein container. Since AaLS-13 protein capsids have been shown to facilitate substrate sorting, enabling the production of a positively tagged OaAEP1 without truncation presents another direction for further investigation.

To further build upon the findings reported in Chapter 4, investigations can proceed in several directions. The ideal reaction conditions for OaAEP1 is at pH 5.0, which was shown to be incompatible with AaLS-13. Consequently, replacing OaAEP1 with another AEP may be a beneficial to create a catalytically active encapsulated AEP system. AEPs such as butelase 1 and MCoAEP2 have been reported to exhibit optimal activity at pH 6.5 and 6.0, repectively.[23,27] Preference towards higher pH conditions could enable greater catalytic activity when encapsulated within the AaLS-13 host at neutral pH.

A potentially critical rationale for the lack of activity exhibited by the AaLS-13/cpAaLS-OaAEP1 protein complex could be the incorrect folding of the AEP when fused to cpAaLS. Analysis by SDS-PAGE has shown the present of both target proteins, and particles resembling loaded AaLS-13 capsids were observed by TEM. However, the folded structure of the catalytic AEP protein remains unknown. Thus, structural biology characterisation represents another option for further research.

Applications to other AEPs. Finally, OaAEP1-C247A was employed for the all of the experiments outlined in this thesis. Applicational research and developments have mostly focused on OaAEP1 (both wild type and the C247A variant) from *O. affinis,*[14,22,39,43,97] and butelase 1 from *C. ternatea*.[23,37,42,98,99] Investigations towards

the application of other AEPs, such as MCoAEP2 and VyPAL2,[4,27] with subtle differences in substrate recognition and optimal pH would complement the current toolbox of enzymes for protein bioconjugation.

6 Materials and methods

6.1 Materials

All reagents and antibiotics were purchased from Fisher Scientific®, unless stated otherwise.

Deionised water (dH_2O) was obtained from an Elga® PURELAB Chorus 2 system.

2-formylphenylboronic acid (FPBA) and Tris(2-carboxyethyl)phosphine (TCEP) were purchased from Fluorochem.

Peptide substrates were purchased from ISCA Biochemicals, Genscript or synthesised by members of the Luk research group at Cardiff University (Davide Cardella, Alexander J. Lander, Xuefei Li and Jorge S. Escudero)

Calcium chloride and isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from Melford®.

30% acrylamide/bis-acrylamide solution, N,N,N',N'- tetramethylenediamine (TEMED) and SYBR safe was purchase from Merck Sigma-Aldrich

Custom oligonucleotides were purchased from Merck Sigma Aldrich. Buffers, nucleotide triphosphate, and PrimeSTAR® HS DNA polymerase were purchased from Takara. All restriction enzymes were purchased from ThermoFisher Scientific. NEBuilder® HiFi DNA assembly master mix was purchased from New England Biolabs.

The codon-optimised gene for expression in *E. coli*, encoding a protein composed of a N-terminal His_6 tag, the 76 amino-acid residues of human ubiquitin and residues 24 – 474 of OaAEP1-C247A was inserted into the coding region (NcoI-NdeI) of the pET-28b (+) vector (Genscript).

Plasmids containing genes encoding for AaLS-13, cpAaLS and GFP(+36)-TEVp were obtained courtesy of the generous donation from Professor Don Hilvert and the Hilvert group at ETH, Zurich.

Copper sample grids (carbon film 400 mesh) for TEM were purchased from TAAB.

6.2   Preparation of growth medium, solutions and buffers

6.2.1   Luria-Bertani (LB) liquid medium

Tryptone (10 g), NaCl (10 g) and yeast extract (5.0 g) were dissolved in 1 L of deionised water (dH$_2$O). The solution was sterilised by autoclaving at 121 °C for 20 min. The sterile solution was stored at 4 °C.

6.2.2   Luria-Bertani (LB) solid medium

Tryptone (1.0 g), NaCl (1.0 g), yeast extract (0.5 g) and agar (1.5 g) were dissolved in 100 mL dH$_2$O. The suspension was sterilised by autoclaving at 121 °C for 20 min, then cooled to 40°C before the addition of appropriate antibiotic solutions (*e.g.* kanamycin). The resultant solution was poured into sterile petri dishes. The medium was allowed to cool and solidify at room temperature, then stored at 4 °C.

6.2.3   1000X Antibiotic stock solutions

Three antibiotics (chloramphenicol, ampicillin and kanamycin) were used. Chloramphenicol solution was prepared by dissolving 250 mg of the antibiotic in 10 mL of ethanol. Ampicillin (1000 mg) and kanamycin (500 mg) were dissolved in 10 mL of dH$_2$O. The solutions were stored at -20 °C.

6.2.4   IPTG inducer stock solution

To make a 1 M stock solution, IPTG (2.4 g) was dissolved in 10 mL of dH$_2$O. The solution was stored at -20 °C.

6.2.5 Tetracycline inducer stock solution

To make a 0.1 mg/mL stock solution, tetracycline (10 mg) was dissolved in 1 mL of ethanol, then diluted with ethanol to a final concentration of 0.1 mg/mL. The solution was stored at -20 °C.

6.2.6 Calcium competent cell solution 1

Dissolve calcium chloride (2.2 g) in 200 mL of dH_2O. The solution was sterilised by autoclaving at 121 °C for 20 min, then stored at 4 °C.

6.2.7 Calcium competent cell solution 2

Dissolve calcium chloride (0.9 g) in 85 mL of dH_2O, and 15 mL of glycerol. The solution was sterilised by autoclaving at 121 °C for 20 min, then stored at 4 °C.

6.2.8 4X SDS stacking buffer

Tris free base (6.0 g, 0.5 M) was dissolved in 80 mL of dH_2O. The pH was brought to 6.8 with diluted HCl. Sodium-dodecyl sulphate (SDS) (1.0 g) was add to the solution and the volume was adjusted to 100 mL with dH_2O.

6.2.9 4X SDS resolving buffer

Tris free base (27.23 g, 1.5 M) was dissolved in 80 mL of dH_2O. The pH was brought to 8.8 with diluted HCl, SDS (1.5 g) was added to the solution and the volume was adjusted to 150 mL with dH_2O.

6.2.10 4X SDS sample loading buffer

4X SDS stacking buffer (1.8 mL) (section 7.2.8), glycerol (2.0 mL), 10% (w/v) bromophenol blue (0.2 mL) and β-mercaptoethanol (1.0 mL) were mixed and stored at room temperature.

6.2.11 10X SDS running buffer

Tris free base (30.3 g), glycine (144.0 g) and SDS (10.0 g) were dissolved in 1 L of dH$_2$O.

6.2.12 50X TAE buffer

Dissolve Tris free base (242.0 g) and disodium EDTA (18.6 g) in 800 mL of dH$_2$O followed by glacial acetic acid (57.1 mL). Top up to 1 L with dH$_2$O.

6.3   General Methods

6.3.1   Calcium competent *E. coli* preparation

Calcium competent *E. coli* strains BL21 (DE3), Shuffle® T7 express and DH5α were prepared using glycerol stocks available to the research group. For Shuffle® T7 express, cultures were incubated at 30 °C, instead of 37 °C. Glycerol stocks were plated on LB agar and incubated at 37 °C, overnight (16-20 h). A single colony of *E. coli* was selected to inoculate 5 mL of LB medium, which was then incubated at 37 °C, overnight (16-20 h). 200 mL of medium was inoculated with 2 mL of the overnight culture, then incubated at 37 °C. Growth and cell density of the *E. coli* culture was estimated by monitoring the optical density at 600 nm wavelength (OD$_{600}$). When OD$_{600}$ reached 0.6-0.7, the culture was cooled on ice for 10 min, then centrifuged at 3600 g, 4 °C for 15 min. The resulting supernatant was discarded, the *E. coli* containing pellet was resuspended in 40 mL of calcium competent cell solution 1 (section 6.2.6), then incubated on ice for a further 15 min. The suspension was centrifuged (same as above), the supernatant was discarded, and the cell pellet was suspended in 4 mL of calcium competent cell solution 2 (section 6.2.7). The suspended cells were split into 100 µL aliquots in Eppendorf® microcentrifuge tubes and stored at -80 °C.

Newly prepared calcium competent *E. coli* were spread on LB agar plates with antibiotics (chloramphenicol, ampicillin or kanamycin) to control for antibiotic resistance. Heat-shock transformation with pUC19 was also performed to assess transformation efficiency.

6.3.2 Heat-shock transformation

Plasmid (1 µL) was added to a glycerol stock of calcium competent *E. coli*, then incubated on ice for 10 min. To stimulate plasmid uptake, the cells were subject to heat-shock in a water bath at 42 °C for 45 s, then immediately returned to ice for 2 min. 0.5 mL of LB media was added to the heat-shocked *E. coli*, then incubated at 37 °C (or 30 °C for Shuffle® T7 express) for 1 h. 100 µL of the *E. coli* culture was then spread on LB agar containing the corresponding antibiotic for selection, and incubated at 37 °C (or 30 °C) overnight (16-20 h)

6.3.3 SDS-PAGE analysis

6.3.3.1 Gel preparation

Resolving gel was prepared according to table 6.1. The mixture was poured into gel casters and covered with a layer of isopropanol (IPA).

Table 6.1 preparation of SDS resolving gel		
	12%	**15%**
dH$_2$O	3.40 mL	2.40 mL
4X SDS resolving buffer	2.50 mL	2.50 mL
30% acrylamide/bis-acrylamide solution	4.00 mL	5.00 mL
10% (w/v) ammonium persulfate (APS)	100 µL	100 µL
TEMED	10 µL	10 µL

Upon polymerisation, IPA was removed. The stacking gel (4%) was prepared according to table 6.2. The mixture was poured into gel casters on top of the polymerised resolving gel. A comb was inserted prior to setting to create 15 sample wells.

Table 6.2 preparation of SDS stacking gel	
dH$_2$O	2.90 mL
4X SDS stacking buffer	1.25 mL
30% acrylamide/bis-acrylamide solution	0.83 mL

10% (w/v) APS	50 µL
TEMED	5 µL

6.3.3.2 SDS-PAGE sample preparation

Samples (30 µL) were mixed with 4X sample loading buffer (section 6.2.10) (10 µL), then heated at 95 °C for 4 min.

6.3.3.3 Electrophoresis protocol

SDS-PAGE gel was set in a gel running tank containing 1X SDS running buffer (see section 6.2.11, diluted from 10X with dH$_2$O) was loaded with 5 µL of unstained protein molecular weight marker and protein samples prepared according to section 6.3.3.2 (5-10 µL). Electrophoresis was performed with a constant current (30 mA) for 45 min. The gel was then removed and stained with Coomassie Blue to visualise the protein bands.

6.3.4 Agarose gel electrophoresis

6.3.4.1 Gel preparation

Agarose was dissolved in hot 1X TAE buffer (section 6.2.12, diluted from 50X with dH$_2$O) to obtain a 1% (w/v) agarose solution. SYBR safe nucleic acid stain (5 µL) was added, then the solution was poured into a gel cast with a sample comb inserted. The gel was allowed to cool and set at room temperature.

6.3.4.2 Electrophoresis protocol

The agarose gel was submerged in 1X TAE buffer. 5 µL of gene ruler 1kb plus and DNA samples were loaded into sample wells. Electrophoresis was performed with a constant voltage (110 V) for 35 min.

6.3.5 Polymerase chain reaction

Polymerase chain reactions were performed according to PrimeSTAR DNA polymerase product manual using a 3PrimeG PCR Thermocycler (Techne) with PCR conditions described in table 6.3.

Table 6.3 Thermocycler conditions for DNA amplification by PCR			
Step	**Temperature / ºC**	**Duration / min:sec**	**Cycle repeat**
Initial denaturation	95	10:00	1
DNA melting	95	00:10	30
Primer annealing	Variable*	00:05	
DNA elongation	72	Variable**	
Final elongation	72	10:00	1
Final hold	4	Hold	1

* Annealing temperatures were selected according to the melting temperature of partially annealed primer predicted by SnapGene

** Elongation times were calculated according to the length of the target DNA to be amplified and allowing an elongation rate of 1 kb/min.

6.3.6 Plasmid preparation

Relevant plasmids were introduced to DH5α *E. coli* by heat-shock transformation (section 6.3.2). A single colony was selected from the agar plate to inoculate 5 mL of LB medium with the appropriate antibiotic, then incubated at 37 °C overnight (16-20 h). *E. coli* was harvested by 3600 g, 4 °C for 15 min, the supernatant was discarded. The target plasmid was extracted and purified using QIAprep® spin miniprep kit (QIAGEN) following the protocol from the supplier. DNA concentration was estimated using a nanodrop instrument. Purified plasmid solutions were stored at -20 °C.

6.3.7 Recombinant gene expression

Relevant plasmids were introduced to *E. coli* strains BL21 (DE3) or Shuffle® T7 express *E. coli* by heat-shock transformation (section 6.3.2). A single colony was selected from the agar plate to inoculate 10 mL of LB medium with the appropriate antibiotic, then incubated at 37 °C (or 30 °C for Shuffle® T7 express) overnight (16-20 h). The overnight culture was used to inoculate LB media with the appropriate

antibiotic at a 1:100 ratio (e.g. 10 mL into 1 L). The inoculated LB media was incubated at 37 °C (or 30 °C) until the appropriate OD_{600} value was reached (0.6-0.7 unless stated otherwise). Subsequently, IPTG and/or tetracycline were added to induce gene expression. The cultures containing inducers were incubated further (temperature and time to be specified for each protein). *E. coli* was harvested by centrifugation at 5000 g and 4 °C for 20 min, the supernatant was discarded, and the pellet was stored at - 20 °C.

6.3.8 Fast protein liquid chromatography (FPLC)

With the exception of gravity flow columns used for Ni^{2+}-NTA, protein purifications by column chromatography (SAX and SEC) were performed using an ÄKTA purifier or ÄKTA start systems with stationary phase columns from GE Healthcare.

6.3.9 Estimation of protein concentration

Protein concentrations were estimated by measuring the absorbance at 280 nm wavelength (A_{280}) using a Nanodrop UV/Vis spectroscopy (Thermo Scientific), with absorbance baseline correction at 350 nm. The corresponding extinction coefficients at 280 nm (ε_{280}) and molecular weight (MW) were computed by inputting the amino acid sequence into Expasy-ProtParam (https://web.expasy.org/protparam/). Concentration (c) in mol/L and mg/mL were calculated using equations 1 and 2, respectively.

$$A_{280} = \varepsilon_{280} \times c \times l \qquad \textbf{(1)}$$

$$c \text{ (in mg/mL)} = c \text{ (in mol/L)} \times MW \qquad \textbf{(2)}$$

6.3.10 LCMS method

LCMS was performed on an Agilent Infinity II HPLC system coupled to an Agilent 6120 single quadrupole mass spectrometer, using an Eclipse plus (100 x 4.6 mm) C18 column (Agilent) with a linear gradient from 5% to 35% acetonitrile (0.1% formic acid) over 15 min. Data was collected in positive electrospray ionisation mode.

6.3.11 Analytical HPLC method

HPLC was performed on an Agilent Infinity II HPLC system, using a Gemini (250 x 10 mm, 110 Å) C18 column (Phenomonex) held at 40 ºC, with a linear gradient from 5% to 30% acetonitrile (0.1% formic acid) over 15 min. Detection was performed at a wavelength of 210 nm.

6.3.12 Protein ultraperformance liquid chromatography-MS

UPLC-MS analysis of proteins were performed by the Chemistry Analytical Services at Cardiff University on a Waters Synapt G2-Si quadrupole time-of-flight mass spectrometer coupled to a Waters Acquity H-Class ultraperformance liquid chromatography (UPLC) system. The column was an Acquity UPLC protein BEH C4 (300 Å, 1.7 µm by 2.1 mm by 100 mm) operated in reverse phase and held at 60 °C. The gradient employed was 95% A to 35% A over 50 min, where A is water with 0.1% HCO_2H and B is acetonitrile with 0.1% HCO_2H. Spectra were collected in positive ionization mode and analysed using Waters MassLynx software version 4.1. Deconvolution of protein charged states was obtained using the maximum entropy 1 processing software.

6.3.13 UV-Vis spectroscopy

Samples were prepared in 115F-QS, quartz micro-cuvettes (Hellma) and measured using a UV-2600 UV-Vis spectrometer (Shimadzu) from 220 nm to 360 nm.

6.3.14 Fluorescence spectroscopy

Samples were prepared in Nunc microWell 96-well optical-bottom plates with polymer base (Thermo Scientific). Measurements were made using a FLUOstar Omega microplate reader (BMG Labtech) with 10 s of shaking at 150 RPM before the first measurement at 1500 and 2000 gain, 10 flashes per well. Excitation and emission filters were 350-10 and 450-10 nm, respectively.

6.4 Molecular cloning

6.4.1 Primer sequences

The primers listed here were used for the cloning of gene constructs described in sections 6.4.2 to 6.4.12, where the primers are reference by the names given in this table.

Primer name	Sequence (5' to 3')
ST7FI	GCGGCTCCGGCGGCTCGAGCGCGCGTGACGGTGATTATCTGC
ST7RI	AGCCGGATCAGCTGACTAGATTACGGGATGCTCGCGCACG
ST7FV	CGTGCGCGAGCATCCCGTAATCTAGTCAGCTGATCCGGCTGCTAAC
ST7RV	AGATAATCACCGTCACGCGCGCTCGAGCCGCCGGAGC
ST9FN	TTGTTTAACTTTAAGAAGGAGATATACATATGCACCACCACC
ST9RN	ACGGAGCTCGAATTCGGATCTTATAAGCAATTACCGCTACCACCCTTGTACAGCTCGTCCATGCCGAG
ST14F	ATAATTGGATCCAGCGTGGGTACCCGTTGG
ST14R	AATTAACTCGAGTTACGGGATGCTCGCGCACG
ST17F	TAACTCGAGTAAGCGGCGAACGATG
ST17R	ACTCGAGTTAGTCGTTCGCCGGGTTGC
ST19F-OaAEP1	AACTTTAAGAAGGAGATATACCATGGCGCGTGACGGTGATTATCTGCAC
ST19R-OaAEP1	GGATCCCGGGTCGTTCGCCGGGTTGC
ST19F-GFP	GGCAGCAACCCGGCGAACGACCCGGGATCCATGGCTAGCAAAGGTGAACGTCTGTTTCG
ST19R-GFP	TGGTGGTGGTGGTGCTCGAGCTTGTAGCGTTCGTCGCGTCC
ST21F	GGTCTGAGCGGTAGCGGTAGCCAGATCTTCGTTAAAACCCTGACCGGC
ST21R	CTGGCTACCGCTACCGCTCAGACCCTGGAAGTACAGGTTCTCACC
ST22F	GCGGTAGCGGTAGCCATATGGTGAGCAAGGGCGAGGAG
ST22R	TGTCGACGGAGCTCGAATTCTTAGGATCCCTTGTACAGCTCGTCCATGCCG
GFP D235A For	CGGCATGGCCGAGCTGTACAAGGG

GFP D235A Rev	CAGCTCGGCCATGCCGAGAGTGATC
ST23F	TGCACCACCACCACCACCACCAGATCTTCGTTAAAACCCTGACCGGC
ST23R	AAGCAATTACCGCTACCACCGGATCCGCCACCACGCAGACGCAG
ST23F G76P	GTGGTGGCCCGTCCGGTGGTAGCGGTAATTGC
ST23R G76P	CACCGGACGGGCCACCACGCAGACGC
ST24F	TGAGCGGTAGCGGTAGCCATATGGGTGCAGATCTGGCAG
ST24R	GACGGAGCTCGAATTCTTAGGATCCTGCCAGAACACCG
ST25F	TGCACCACCACCACCACCACATGGGTGCAGATCTGGCAG
ST25R	AAGCAATTACCGCTACCACCGGATCCTGCCAGAACACCGGCAAC
ST26F	TGAGCGGTAGCGGTAGCCATATGGAAATCTACGAAGGTAAACTAACTGCTGAAG
ST26R	GACGGAGCTCGAATTCTTAGGATCCCTCGAGTCGGAGAGACTTGAATAAGTTTGC
ST27F	TGCACCACCACCACCACCACATGGAAATCTACGAAGGTAAACTAACTGCTGAAG
ST27R	AAGCAATTACCGCTACCACCGGATCCCTCGAGTCGGAGAGACTTGAATAAGTTTGC
ST28F-Cap	TTTGTTTAACTTTAAGAAGGAGATATCATATGAACTATACCAGCCTGGATGGTAACGC
ST28R-Cap	CTCGAGTGCGGCCGCAAGCTTTTACGGGATGCTCGCGCACG
ST28FV	AAGCTTGCGGCCGCACTCGAGGAGATCCGGCTGCTAACAAAGCC
ST28RV	TCTCCTTCTTAAAGTTAAACAAAATTATTTCGGATCCTTAGTCGTTCGCCGGGTTGC
ST29R	CTCGAGTGCGGCCGCAAGCTTTTAGTCGTTCGCCGGGTTGC
ST34F	ACAAGGTACCCGGTTAATCTAGTCAGCTGATCCGGCTGCTAAC
ST34R	ACCGGGTACCTTGTAGCGTTCG
ST35F	ATAATTGGATCCGGAGAAAGCTTGTTTAAGGGGCCGC
ST35R	CCGTGCTCGAGTTAATTCATGAGTTGAGTCGCTTCCTTAACTGGC

6.4.2 OaAEP1-C247A core domain

The plasmid pET28b-His-Ub-OaAEP1 core (ST29) was used for preparing N-terminal His$_6$ tagged OaAEP1 by IPTG induction. It was constructed by removing the pro-

domain from the full length His-Ub-OaAEP1 construct using primers ST28FV and ST29R. The PCR product was digested by DpnI (ThermoFisher) then transformed into DH5α *E. coli* cells. Sequences of the product plasmids (ST29) were confirmed by DNA Sanger sequencing performed by Eurofins Genomics.

6.4.3 Split OaAEP1-C247A

The plasmid pET28b-split-OaAEP1 (ST28) was used for preparing N-terminal His$_6$ tagged OaAEP1-C247A and a separate pro-domain by IPTG induction. The gene encoding for His-Ub-OaAEP1 (core domain) was amplified using primers ST28FV and ST28RV. The gene encoding for the AEP pro-domain and an upstream ribosome binding site were amplified using primers ST28F-Cap and ST28R-Cap. The PCR products were assembled by Gibson Assembly then transformed into DH5α *E. coli*. Sequences of the product plasmids (ST28) were confirmed by DNA Sanger sequencing performed by Eurofins Genomics.

6.4.4 Ubiquitin

The plasmid pET28b-His-TEV-linker-Ub (ST21) was used for the IPTG-inducible production of the N-terminal His$_6$ tagged ubiquitin and was constructed by subcloning the gene encoding human ubiquitin (Genscript) into pET28b using primers ST21F and ST21R via Gibson assembly. For ubiquitin bearing the C-terminal recognition sequence (NCL), the gene encoding for ubiquitin was amplified from ST21 using primers ST23F and ST23R, then subcloned into a pET28b derived vector by Gibson assembly to generate the plasmid pET28b-His-Ub-NCL (ST23). Finally, a Gly to Pro site directed mutagenesis in the C-terminal linker region was carried out using primers ST23F G76P and ST23R G76P to prevent undesired cleavage during recombinant expression. Sequences of all plasmids were confirmed by DNA Sanger sequencing performed by Eurofins Genomics.

6.4.5 Enhanced green fluorescent protein

The plasmid pET28b-eGFP-NCL (ST9N) was used for the IPTG-inducible production of the N-terminal His$_6$ tagged enhanced green fluorescent protein (eGFP) and was constructed by subcloning the gene encoding eGFP (already available in the group)

using primers ST9FN and ST9RN, into a pET28b derived vector by Gibson assembly. For eGFP bearing the N-terminal His_6 tag followed by TEV and OaAEP1 recognition sequences, the gene encoding for His_6-TEV-eGFP (ST22) was amplified from ST9N using primers ST22F and ST22R, then subcloned into a pET28b derived vector by Gibson assembly to generate the plasmid pET28b-His-TEV-eGFP. Finally, Asp to Ala site directed mutagenesis were carried out on both ST9N and ST22 using primers GFP D235A For and GFP D235A Rev to prevent undesired cleavage during recombinant expression. Sequences of all plasmids were confirmed by DNA Sanger sequencing performed by Eurofins Genomics.

6.4.6 **Mycobacterium tuberculosis β-lactamase**

The plasmid pET28b-His-TEV-BlaC (ST24) was used for the IPTG-inducible production of the N-terminal His_6 tagged beta-lactamase (BlaC) and was constructed by subcloning the gene encoding BlaC (already available in the group) using primers ST24F and ST24R, into a pET28b derived vector by Gibson assembly. For BlaC bearing the N-terminal His_6 tag and a C-terminal OaAEP1 recognition sequence (NCL), the gene encoding for BlaC was amplified using primers ST25F and ST25R, then subcloned into a pET28b derived vector by Gibson assembly to generate the plasmid pET28b-His-BlaC-NCL (ST25). Sequences of all plasmids were confirmed by DNA Sanger sequencing performed by Eurofins Genomics.

6.4.7 *Aquifex aeolicus* **lumazine synthase-13 (AaLS-13)**

The plasmid pET28b-His-TEV-AaLS13 (ST26) was used for the IPTG-inducible production of the N-terminal His_6 tagged AaLS13 and was constructed by subcloning the gene encoding AaLS13 (donation from Professor Don Hilvert and the Hilvert group at ETH, Zurich) using primers ST26F and ST26R, into a pET28b derived vector by Gibson assembly. For AaLS13 bearing the N-terminal His_6 tag and a C-terminal *Oa*AEP1 recognition sequence (NCL), the gene encoding for AaLS13 was amplified using primers ST27F and ST27R, then subcloned into a pET28b derived vector by Gibson assembly to generate the plasmid pET28b-His-AaLS13-NCL (ST27). Sequences of all plasmids were confirmed by DNA Sanger sequencing performed by Eurofins Genomics.

6.4.8 His$_6$-GFP(+36)-OaAEP1

The plasmid pACYC-His-GFP(+36)-OaAEP1 (ST7) was used for the IPTG-inducible production of the N-terminal His$_6$ tagged GFP(+36)-OaAEP1 and was constructed by subcloning the gene encoding OaAEP1-C247A (using primers ST7FI and ST7RI), into a pACYC vector derived from the GFP(+36)-TEVp plasmid (donation from Professor Don Hilvert and the Hilvert group at ETH, Zurich) which was amplified by PCR, using primers ST7FV and ST7RV, by Gibson assembly. The Sequence of the newly assembled plasmids was confirmed by DNA Sanger sequencing performed by Eurofins Genomics.

6.4.9 OaAEP1-GFP(+36)-His$_6$

The plasmid pET28b-OaAEP1-GFP(+36)-His (ST19) was used for the IPTG-inducible production of the C-terminal His$_6$ tagged OaAEP1-GFP(+36). It was constructed by subcloning the gene encoding OaAEP1-C247A (primers ST19F-OaAEP1 and ST19R-OaAEP1), and gene encoding GFP(+36) (donation from Professor Don Hilvert and the Hilvert group at ETH, Zurich) using primers ST19F-GFP and ST19R-GFP into a pET28b vector by Gibson assembly. The Sequence of the newly assembled plasmids was confirmed by DNA Sanger sequencing performed by Eurofins Genomics.

6.4.10 His$_6$-GFP(+36)

The gene encoding for GFP(+36) was obtained from a donation from Professor Don Hilvert and the Hilvert group at ETH, Zurich. The plasmid pET28b-His-GFP(+36) (ST34) was used for preparing N-terminal His$_6$ tagged GFP(+36) by IPTG induction. It was constructed by PCR amplification from ST7 using primers ST34F and ST34R. The PCR product was digested by DpnI (ThermoFisher) then transformed into DH5α *E. coli*. Sequences of the product plasmids (ST34) were confirmed by DNA Sanger sequencing performed by Eurofins Genomics.

6.4.11 cpAaLS-OaAEP1

The plasmid pACTet-cpAaLS-OaAEP1 (ST17) was used for the tetracycline-inducible production of the cpAaLS-OaAEP1. The plasmid was constructed in two steps. Initially, the gene encoding for the full length OaAEP1-C247A was subcloned (primers ST14F and ST14R) into a pACTet vector in frame with the gene encoding for cpAaLS (donation from Professor Don Hilvert and the Hilvert group at ETH, Zurich) by Gibson assembly to create ST14. Then, the DNA sequence encoding for the pro-domain of OaAEP1 was deleted from ST14 using primers ST17F and ST17R. The Sequence of all plasmids were confirmed by DNA Sanger sequencing performed by Eurofins Genomics.

6.4.12 cpAaLS-TEVp

The genes encoding for cpAaLS and TEVp were obtained from a donation from Professor Don Hilvert and the Hilvert group at ETH, Zurich. The plasmid pACTet-cpAaLS-TEVp (ST35) was used for the tetracycline-inducible production of the cpAaLS-TEVp. The TEVp gene was amplified from pACYC-His6-GFP(+36)-TEVp by PCR using primers ST35F and ST35R. The PCR was treated with restriction enzymes BamHI and XhoI then inserted into the ST17 (also treated with BamHI and XhoI) by sticky ends T4 DNA ligation. The Sequence of all plasmids were confirmed by DNA Sanger sequencing performed by Eurofins Genomics.

6.5   Protein purification

6.5.1 OaAEP1-C247A

Recombinant expression of OaAEP1 in *E. coli* strain BL21(DE3), was performed following the protocol reported previously.[14] Liquid culture was prepared according to section 6.3.7, gene expression was induced by adding IPTG (0.4 mM) at 16 °C for 18 h.

The cell pellet from a 1 L culture was suspended in a lysis buffer containing 50 mM Tris-HCl (pH 7.4) 150 mM NaCl, lysozyme (0.1 mg/mL), DNase I (5 µg/mL), RNase A (5 µg/mL) and PMSF (35 µg/mL). After lysis by sonication, and clearance by

centrifugation (27000 g, 4 °C, 15 min), the supernatant was loaded onto 4 mL of Ni-NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The column was washed with lysis buffer containing 10 mM and 20 mM imidazole (80 mL each), the remaining bound protein was eluted with lysis buffer containing 250 mM imidazole. Ni-NTA elution fractions containing the protein were diluted and further purified by ion-exchange chromatography using 25 ml of Q-sepharose (GE Healthcare). Over 4 column volumes, bound proteins were eluted using a continuous salt gradient of 0–50% of buffer B (20 mM Bis-Tris propane (pH 7.0), 2 M NaCl). Finally, the protein was purified through a SEC column (superdex 200, GE Healthcare) that had been pre-equilibrated in 20 mM sodium phosphate (pH 7.4), 300 mM NaCl. Fractions containing the desired protein were combined, concentrated using an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO) (Merck Millipore) then stored at -80 °C until activation. Protein concentration was estimated by nanodrop UV-Vis according to section 6.3.9.

6.5.2 OaAEP1-C247A activation

To activate OaAEP1, the concentrated protein stock was diluted with 50 mM sodium acetate (pH 4.0), 50 mM NaCl, 1 mM EDTA, 0.5 mM TCEP-HCl buffer to 1.5 µM. Then, the activation mixture was dialysed at 20 °C (sodium acetate buffer as described above). After 16 h, protein concentration was measure by nanodrop. The buffer of the activated protein solution was then exchanged by dialysis at 4 °C into the desired reaction buffer. Protein precipitation during activation allowed removal of the contaminating proteins by centrifugation (4000g, 4 °C, 20 min). The mature AEP can be further purified by ion exchange chromatography according to protocol reported previously.[39] Protein concentration was estimated by nanodrop UV-Vis according to section 6.3.9.

6.5.3 OaAEP1-C247A core domain

Recombinant gene expression was performed in *E. coli* strain Shuffle® T7 express. Liquid culture was prepared according to section 6.3.7, and gene expression induced by adding IPTG (0.2 mM) at 16 °C for 18 h.

The cell pellet from a 0.5 L culture was suspended in 20 mL of lysis buffer containing 50 mM sodium phosphate (pH 8.0) 300 mM NaCl, lysozyme (0.1 mg/mL), DNase I (5 µg/mL), RNase A (5 µg/mL) and PMSF (35 µg/mL). After lysis by sonication and clearance by centrifugation (27000 g, 4 °C, 15 min), the supernatant was loaded onto 2 mL of Ni-NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The column was washed three times with lysis buffer containing 10 mM imidazole (15 mL each), and the remaining bound protein was eluted with 15 mL of lysis buffer containing 300 mM imidazole. Ni-NTA elution was concentrated using an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO) (Merck Millipore), then further purified by size exclusion chromatography (superdex 75, GE Healthcare) that had been pre-equilibrated in 50 mM MES (pH 6.0), 50 mM NaCl, 1 mM EDTA and 0.5 mM TCEP. Fractions containing desired protein were combined, concentrated using an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO) (Merck Millipore) then stored at -80 °C with 5% (v/v) glycerol. Protein concentration was estimated by nanodrop UV-Vis according to section 6.3.9.

6.5.4 Split OaAEP1

Recombinant gene expression was performed in *E. coli* strain BL21(DE3). Liquid culture was prepared according to section 6.3.7, gene expression induced by adding IPTG (0.2 mM) at 16 °C for 18 h.

The cell pellet from a 0.5 L culture was suspended in 20 mL of lysis buffer containing 50 mM sodium phosphate (pH 8.0) 300 mM NaCl, lysozyme (0.1 mg/mL), DNase I (5 µg/mL). After lysis by sonication and clearance by centrifugation (27000 g, 4 °C, 15 min), the supernatant was loaded onto 2 mL of Ni-NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The column was washed three times with lysis buffer containing 10 mM imidazole (15 mL each), and the remaining bound protein was eluted with 15 mL of lysis buffer containing 300 mM imidazole. Ni-NTA elution was concentrated using an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO) (Merck Millipore), then further purified by size exclusion chromatography (superdex 75, GE Healthcare) that had been pre-equilibrated in 50 mM sodium phosphate (pH 7.4), 200 mM NaCl, 1 mM EDTA. Fractions containing desired protein were combined, concentrated using an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO) (Merck

Millipore) then stored at -80 °C with 5% (v/v) glycerol. Protein concentration was estimated by nanodrop UV-Vis according to section 6.3.9.

6.5.5 Ubiquitin

Recombinant expression was performed in *E. coli* strain BL21(DE3). Liquid culture was prepared according to section 6.3.7, gene expression induced by adding IPTG (0.1 mM) at 20 °C for 18 h.

The cell pellet from a 1 L culture was suspended in 20 mL of lysis buffer containing 50 mM sodium phosphate (pH 8.0) 300 mM NaCl, lysozyme (0.1 mg/mL), DNase I (5 µg/mL), RNase A (5 µg/mL) and PMSF (35 µg/mL). After lysis by sonication, and clearance by centrifugation (27000 g, 4 °C, 15 min), the supernatant was loaded onto 4 mL of Ni-NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The column was washed three times with lysis buffer containing 10 mM imidazole (40 mL each), the remaining bound protein was eluted with 20 mL of lysis buffer containing 300 mM imidazole. The eluant from the Ni-NTA chromatography was concentrated using an Amicon Ultra-15 centrifugal filter unit (3 kDa MWCO) (Merck Millipore), then further purified by size exclusion chromatography (superdex 75, GE Healthcare) that had been pre-equilibrated in 50 mM MES (pH 6.0), 150 mM NaCl, 1 mM EDTA, 0.5 mM TCEP. Fractions containing desired protein were combined, concentrated using an Amicon Ultra-15 centrifugal filter unit (3 kDa MWCO) (Merck Millipore) then stored at -80 °C. For protein constructs bearing TEV cleavage site (ST21), Ni-NTA elution fractions containing the desired protein were combined, buffer exchanged into 50 mM sodium phophate (pH 7.0), 150 mM NaCl, 1mM EDTA, 1 mM TCEP-HCl using an Amicon Ultra-15 centrifugal filter unit (3 kDa MWCO) (Merck Millipore). The protein was concentrated to around 5 mg/mL, then treated with TEV protease (1 mol%) overnight at 20 °C. The processed ubiquitin was isolated from the TEV cleavage reaction mixture by passing over 4 mL of Ni-NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The flow through containing the desired protein was concentrated using an Amicon Ultra-15 centrifugal filter unit (3 kDa MWCO) (Merck Millipore), then purified by size exclusion chromatography as described above. Fractions containing desired protein were combined, concentrated using an Amicon Ultra-15 centrifugal

filter unit (3 kDa MWCO) (Merck Millipore) then stored at -80 °C. Protein concentration was estimated by nanodrop UV-Vis according to section 6.3.9.

6.5.6 Enhanced green fluorescent protein

Recombinant expression was performed in *E. coli* strain BL21(DE3). Liquid culture was prepared according to section 6.3.7, gene expression induced by adding IPTG (0.1 mM) at 20 °C for 18 h.

The cell pellet from a 1 L culture was suspended in a lysis buffer containing 50 mM sodium phosphate (pH 8.0) 300 mM NaCl, lysozyme (0.1 mg/mL), DNase I (5 µg/mL), RNase A (5 µg/mL) and PMSF (35 µg/mL). After lysis by sonication, and clearance by centrifugation (27000 g, 4 °C, 15 min), the supernatant was loaded onto 4 mL of Ni-NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The column was washed three times with lysis buffer containing 10 mM imidazole (80 mL each), the remaining bound protein was eluted with lysis buffer containing 300 mM imidazole. Ni-NTA elution fractions containing the desired protein were combined, buffer exchanged into 50 mM MES (pH 6.0) 150 mM NaCl, 1 mM EDTA, 1 mM TCEP-HCl using an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO) (Merck Millipore). The concentrated protein was then purified through a SEC column (superdex 75, GE Healthcare) that had been pre-equilibrated in 50 mM MES (pH 6.0) 150 mM NaCl, 1 mM EDTA, 1 mM TCEP-HCl. Fractions containing desired protein were combined, concentrated using an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO) (Merck Millipore) then stored at -80 °C. Protein concentration was estimated by nanodrop UV-Vis according to section 6.3.9.

6.5.7 Mycobacterium tuberculosis β-lactamase

Recombinant expression was performed in *E. coli* strain BL21(DE3). Liquid culture was prepared according to section 6.3.7, gene expression induced by adding IPTG (0.1 mM) at 20 °C for 18 h.

The cell pellet from a 1 L culture was suspended in a lysis buffer containing 50 mM sodium phosphate (pH 8.0) 300 mM NaCl, lysozyme (0.1 mg/mL), DNase I (5 µg/mL),

RNase A (5 µg/mL) and PMSF (35 µg/mL). After lysis by sonication, and clearance by centrifugation (27000 g, 4 °C, 15 min), the supernatant was loaded onto 4 mL of Ni-NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The column was washed three times with lysis buffer containing 10 mM imidazole (80 mL each), the remaining bound protein was eluted with lysis buffer containing 300 mM imidazole. Ni-NTA elution fractions containing the desired protein were combined, buffer exchanged into 50 mM MES (pH 6.0) 150 mM NaCl, 1 mM EDTA, 1 mM TCEP-HCl using an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO) (Merck Millipore). The concentrated protein was then purified through a SEC column (superdex 75, GE Healthcare) that had been pre-equilibrated in 50 mM MES (pH 6.0) 150 mM NaCl, 1 mM EDTA, 1 mM TCEP-HCl. Fractions containing desired protein were combined, concentrated using an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO) (Merck Millipore) then stored at -80 °C. Protein concentration was estimated by nanodrop UV-Vis according to section 6.3.9.

6.5.8 AaLS-13

AaLS-13 and subsequent variants bearing N or C terminal recognition sequences for OaAEP1 reactions were produced by recombinant expression performed in *E. coli* strain BL21(DE3). Liquid culture was prepared according to section 6.3.7, gene expression induced by adding IPTG (0.1 mM) at 25 °C for 18 h.

The cell pellet from a 1 L culture was suspended in a lysis buffer containing 50 mM sodium phosphate (pH 8.0) 300 mM NaCl, lysozyme (0.1 mg/mL), DNase I (5 µg/mL), RNase A (5 µg/mL) and PMSF (35 µg/mL), then incubated at room temperature for 1 hour. After lysis by sonication, and clearance by centrifugation (27000 g, 25 °C, 15 min), the supernatant was loaded onto 4 mL of Ni-NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The column was washed three times with lysis buffer containing 10 mM imidazole (80 mL each), the remaining bound protein was eluted with lysis buffer containing 300 mM imidazole. Ni-NTA elution fractions containing the desired protein were combined, then concentrated and buffer exchanged into 50 mM sodium phosphate (pH 7.0) 200 mM NaCl, 5 mM EDTA using an Amicon Ultra-15 centrifugal filter unit (30 kDa MWCO) (Merck Millipore). The NaCl concentration was then increase to 600 mM by adding 5 M NaCl. The protein was incubated at room

temperature for 3 days to allow capsid formation. The protein was then purified through a SEC column (superpose 6 GL increase 16/600, GE Healthcare) that had been pre-equilibrated in 50 mM sodium phosphate (pH 7.0) 150 mM NaCl, 1 mM EDTA, 1 mM TCEP-HCl. Fractions containing desired protein were combined, concentrated using an Amicon Ultra-15 centrifugal filter unit (30 kDa MWCO) (Merck Millipore) then stored at 20 °C. Protein concentration was estimated by nanodrop UV-Vis according to section 6.3.9.

6.5.9 GFP(+36)-TEVp fusion protein

The fusion protein GFP(+36)-TEVp was prepared according to published procedures.[141] Recombinant expression was performed in *E. coli* strain BL21(DE3). Liquid culture was prepared according to section 6.3.7, gene expression induced by adding IPTG (0.1 mM) at 18 °C for 18 h.

The cell pellet from a 1 L culture was suspended in a lysis buffer containing 50 mM sodium phosphate (pH 7.4), 2 M NaCl, lysozyme (0.1 mg/mL), DNase I (5 µg/mL), RNase A (5 µg/mL) and PMSF (35 µg/mL). After lysis by sonication, and clearance by centrifugation (27000 g, 4 °C, 15 min), the supernatant was loaded onto 4 mL of Ni-NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The column was washed three times with lysis buffer containing 10 mM imidazole (80 mL each), the remaining bound protein was eluted with lysis buffer containing 300 mM imidazole. Ni-NTA elution fractions containing the desired protein were combined and concentrated using an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO) (Merck Millipore). The concentrated protein was then purified through a SEC column (superdex 200, GE Healthcare) that had been pre-equilibrated in sodium phosphate (pH 7.4), 2 M NaCl. Fractions containing desired protein were combined, concentrated using an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO) (Merck Millipore) then stored at -80 °C. Protein concentration was estimated by UV-Vis measuring the absorbance at 488 nm ($\varepsilon_{488} = 36600$ M^{-1}cm^{-1}).

6.5.10 His$_6$-GFP(+36)-OaAEP1(C247A)

Recombinant expression was performed in *E. coli* strain BL21(DE3). Liquid culture was prepared according to section 6.3.7, gene expression induced by adding IPTG (0.4 mM) at 16 °C for 18 h.

The cell pellet from a 1 L culture was suspended in a lysis buffer containing 50 mM sodium phosphate (pH 7.4), 2 M NaCl, lysozyme (0.1 mg/mL), DNase I (5 µg/mL), RNase A (5 µg/mL) and PMSF (35 µg/mL). After lysis by sonication, and clearance by centrifugation (27000 g, 4 °C, 15 min), the supernatant was loaded onto 2 mL of Ni-NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The column was washed three times with lysis buffer containing 10 mM imidazole (20 mL each), the remaining bound protein was eluted with lysis buffer containing 300 mM imidazole. Ni-NTA elution fractions containing the desired protein were combined and concentrated using an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO) (Merck Millipore). The concentrated protein was then purified through a SEC column (superdex 200, GE Healthcare) that had been pre-equilibrated in sodium phosphate (pH 7.4), 2 M NaCl. Protein containing fractions were analysed by SDS-PAGE.

6.5.11 OaAEP1(C247A)-GFP(+36)-His$_6$

Recombinant expression was performed in *E. coli* strain BL21(DE3). Liquid culture was prepared according to section 6.3.7, gene expression induced by adding IPTG (0.4 mM) at 16 °C for 18 h.

The cell pellet from a 1 L culture was suspended in a lysis buffer containing 50 mM sodium phosphate (pH 7.4), 2 M NaCl, lysozyme (0.1 mg/mL), DNase I (5 µg/mL), RNase A (5 µg/mL) and PMSF (35 µg/mL). After lysis by sonication, and clearance by centrifugation (27000 g, 4 °C, 15 min), the supernatant was loaded onto 2 mL of Ni-NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The column was washed three times with lysis buffer containing 10 mM imidazole (20 mL each), the remaining bound protein was eluted with lysis buffer containing 300 mM imidazole. Ni-NTA elution fractions were analysed by SDS-PAGE.

6.5.12 His$_6$-GFP(+36)

Recombinant expression was performed in *E. coli* strain BL21(DE3). Liquid culture was prepared according to section 6.3.7, gene expression induced by adding IPTG (0.1 mM) at 16 °C for 20 h.

The cell pellet from a 500 mL culture was suspended in a lysis buffer containing 50 mM sodium phosphate (pH 7.4), 2 M NaCl, lysozyme (0.1 mg/mL), DNase I (5 µg/mL). After lysis by sonication, and clearance by centrifugation (27000 g, 4 °C, 15 min), the supernatant was loaded onto 2 mL of Ni-NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The column was washed three times with lysis buffer containing 10 mM imidazole (15 mL each), the remaining bound protein was eluted with lysis buffer containing 300 mM imidazole. Ni-NTA elution fractions containing the desired protein were concentrated using an Amicon Ultra-15 centrifugal filter unit (10 kDa MWCO) (Merck Millipore). The concentrated protein was then purified through a SEC column (superdex 75, GE Healthcare) that had been pre-equilibrated in sodium phosphate (pH 7.4), 200 mM NaCl, 5 mM EDTA. Protein containing fractions were analysed by SDS-PAGE. Protein concentration was estimated by UV-Vis measuring the absorbance at 488 nm (ε_{488} = 36600 M^{-1}cm^{-1}).

6.5.13 cpAaLS-OaAEP1/AaLS-13 patchwork protein cage

Recombinant expression was performed in *E. coli* strain Shuffle® T7 express, which was doubly transformed with a pACYC plasmid containing the codon optimised gene for cpAaLS-OaAEP1and a pMG plasmid containing the codon optimised gene for AaLS-13. Liquid culture was prepared according to section 6.3.7, gene expression induced by adding IPTG (0.1 mM) and tetracycline (0.1 µg/mL) at 16 °C for 18 h.

The cell pellet from a 1 L culture was suspended in a lysis buffer containing 50 mM sodium phosphate (pH 8.0) 300 mM NaCl, lysozyme (0.1 mg/mL), DNase I (5 µg/mL), RNase A (5 µg/mL) and PMSF (35 µg/mL). After lysis by sonication, and clearance by centrifugation (27000 g, 4 °C, 15 min), the supernatant was loaded onto 2 mL of Ni-

NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The column was washed three times with lysis buffer containing 10 mM imidazole (20 mL each), the remaining bound protein was eluted with lysis buffer containing 300 mM imidazole. Ni-NTA elution fractions containing the desired protein were combined, then concentrated and buffer exchanged into 50 mM sodium phosphate (pH 7.0) 200 mM NaCl, 5 mM EDTA using an Amicon Ultra-15 centrifugal filter unit (30 kDa MWCO) (Merck Millipore). The NaCl concentration was then increased to 600 mM by adding 5 M NaCl. The protein was then purified through a SEC column (superpose 6 GL increase 16/600, GE Healthcare) that had been pre-equilibrated in 50 mM sodium phosphate (pH 7.0) 150 mM NaCl, 1 mM EDTA, 1 mM TCEP-HCl. Fractions containing desired protein were combined, then analysed by SDS-PAGE and TEM. Protein concentration was estimated by nanodrop UV-Vis according to section 6.3.9. Total protein concentration was calculated using the extinction coefficient of AaLS-13. Enzyme concentration was then calculated by assuming a 1:100 enzyme to cage protein ratio.

6.5.14 cpAaLS-TEVp/AaLS-13 patchwork protein cage

Recombinant expression was performed in *E. coli* strain BL21(DE3), which was doubly transformed with a pACYC plasmid containing the codon optimised gene for cpAaLS-TEVp and a pMG plasmid containing the codon optimised gene for AaLS-13. Liquid culture was prepared according to section 6.3.7, gene expression induced by adding IPTG (0.1 mM) and tetracycline (0.1 µg/mL) at 16 °C for 18 h.

The cell pellet from a 0.5 L culture was suspended in a lysis buffer containing 50 mM sodium phosphate (pH 8.0) 300 mM NaCl, lysozyme (0.1 mg/mL), DNase I (5 µg/mL). After lysis by sonication, and clearance by centrifugation (27000 g, 4 °C, 15 min), the supernatant was loaded onto 2 mL of Ni-NTA resin (Bio-Rad) in a gravity flow column (Bio-Rad). The column was washed three times with lysis buffer containing 10 mM imidazole (15 mL each), the remaining bound protein was eluted with lysis buffer containing 300 mM imidazole. Ni-NTA elution fractions containing the desired protein were combined, then concentrated and buffer exchanged into 50 mM sodium phosphate (pH 7.4) 200 mM NaCl, 5 mM EDTA using an Amicon Ultra-15 centrifugal filter unit (30 kDa MWCO) (Merck Millipore). The NaCl concentration was then

increased to 600 mM by adding 5 M NaCl. The protein was then purified through a SEC column (superpose 6 GL increase 16/600, GE Healthcare) that had been pre-equilibrated in 50 mM sodium phosphate (pH 7.0) 200 mM NaCl, 5 mM EDTA. Fractions containing desired protein were combined, then analysed by SDS-PAGE. Protein concentration was estimated by nanodrop UV-Vis according to section 6.3.9. Total protein concentration was calculated using the extinction coefficient of AaLS-13. Enzyme concentration was then calculated by assuming a 1:100 enzyme to cage protein ratio.

6.6 Peptide cyclisation kinetic assay

Cyclisation assays were performed in 50 µL reaction mixtures containing either 50 mM NaOAc (pH 4.5 – 5.5, 50 mM NaCl, 1 mM EDTA), or 50 mM MES (pH 5.5 – 6.0, 50 mM NaCl, 1 mM EDTA), or 20 mM Na_2HPO_4 (pH 6.0 – 7.4, 100 mM NaCl, 1 mM EDTA), supplemented with the AEP (2 to 20 nM) and peptide substrate (GLPVSTKPVATRNGL, 10 to 2000 µM). Each reaction was performed in triplicate at 20 °C and quenched after 1 h by adding 5 µL of 1 M HCl solution. The quenched reaction mixtures were analysed by LCMS (Section 6.3.10). Reaction velocities were calculated by converting the HPLC peak areas of the 210 nm UV chromatogram corresponding to the remaining linear precursors and the cyclised products into concentrations. The identity of each HPLC peak was confirmed by MS. The data was fitted to equation 3 using GraphPad Prism 9.0.0 (GraphPad) to estimate the kinetic parameters (k_{cat} and K_M).

$$v = \frac{V_{max} [S]}{K_M + [S]} \qquad \textbf{(3)}$$

6.7 Substrate scope assay for OaAEP1-C247A

Peptide ligation assays were performed in 50 µL reaction mixtures containing either 50 mM NaOAc (pH 5.0), 50 mM NaCl, 1 mM EDTA, OaAEP1-C247A (50 nM), peptide 1 (GLGGIR, 250 µM), peptide 2 (CFRANXL or CFRANGX, 50 µM). Each reaction was performed at 20 °C and quenched after 1 h by adding 5 µl of 1 M HCl solution. The quenched reaction mixtures were analysed by LCMS (Section 6.3.10). Reaction yields were calculated by converting the HPLC peak areas of the UV chromatogram at 210 nm wavelength corresponding to the desired peptide products into concentrations. A

calibration curve was done using the product peptide (CFRANGLGGIR), which was synthesised by SPPS. The identity of each HPLC peak was confirmed by MS.

6.8 Reaction kinetics of N-terminal cysteine peptide coupling to 2-formylphenyl boronic acid

Peptide CFRANGL (10 µM) and FPBA (10 µM) were dissolved in either 50 mM NaOAc (pH 5.0) or 20 mM Na_2HPO_4 (pH 6.0 or 7.0) buffer with 100 mM NaCl, 1 mM EDTA. The reaction progress was monitored with Shimadzu UV/Vis at 20 °C for 20 min at 5 s intervals. Reaction yields and kinetics were calculated by absorbance at 254 nm.[151] Reaction was performed in triplicate.

6.9 Peptide ligation assay

Substrates peptides were synthesized by SPPS by A. J. Lander, D. Cardella and Dr. X. Li. In 50 µL, peptide 1, LFRANCLK (200-400 µM), peptide 2, GLGGIR (200-480 µM), 2-formylphenylboronic acid (200-800 µM) and OaAEP1 (0.2-0.6 µM) were dissolved in buffer (pH 4.5-5.0, 50 mM NaOAc, 50 mM NaCl, 1 mM EDTA and 0.5 mM TCEP, pH 5.2 – 6.5 50 mM MES buffer with 50 mM NaCl, 1 mM EDTA and 0.5 mM TCEP). The reactions were incubated at either 4 °C, 20 °C or 37 °C for up to 18 hours, then quenched with 5 µL of 1 M HCl. As a control, the reaction was carried out in 50 mM MES (pH 5.7) at 20 °C with LFRANCLK (300 µM), GLGGIR (360 µM), FPBA (600 µM) and OaAEP1 (0.3 µM). All reactions were performed in triplicate. Quenched reaction mixtures were analysed by LCMS (Section 6.3.10) and analytical HPLC (Section 6.3.11). The yields were calculated by converting the HPLC peak areas of the 210 nm UV chromatogram corresponding to the desired ligated products into concentrations. A calibration curve was done using the product peptide (LFRANGLGGIR), which was synthesised by SPPS. The identity of each HPLC peak was confirmed by MS.

6.10 Protein ligation assay

For protein conjugation with biotin labelled peptide, in 100 µL, substrate protein (100 µM), biotin peptide (200 µM), FPBA (200 µM) and OaAEP1 (0.25 µM) were added to 50 mM MES buffer (pH 6.0), 50 mM NaCl, 1 mM EDTA, 0.5 mM TCEP buffer. The reactions were incubated at 20 °C for 16 hours, then quenched with 10 µL of 1 M HCl.

For protein conjugation of AaLS13 with biotin labelled peptide, in 100 μL, substrate protein (100 μM), biotin peptide (200 μM), FPBA (200 μM) and OaAEP1 (0.5 μM) were added to 50 mM NaP$_i$ buffer (pH 7.0), 150 mM NaCl, 1 mM EDTA, 0.5 mM TCEP buffer. The reactions were incubated at 20 °C for 18 hours, then quenched with 10 μL of 1 M HCl. Quenched reaction mixtures were analysed by protein UPLC-MS (Section 6.3.12). The mass spectrum was taken as the average across the whole chromatographic peak and total ion count ratio between the protein starting material and the desired biotin labelled protein was used to estimate the reaction yield.

6.11 Loading of GFP(+36)-TEVP into empty AaLS-13 capsids

Purified AaLS-13 capsids (final concentration: 100 μM) in 50 mM sodium phosphate buffer (pH 8.0), 200 mM NaCl and 5mM EDTA were mixed with GFP(+36)-TEVp at several ratios (5, 10, 20 and 30 μM). The mixture was incubated at room temperature (20 °C), overnight (19 h), then analysed by SEC (superpose 6 GL increase 16/600, GE Healthcare). Protein capsid containing fractions were collected and analysed by SDS-PAGE and TEM.

6.12 Negative stain transmission electron microscopy

Glow discharged copper grids, 400 mesh, covered with carbon (TAAB) were incubated face down on a 10 μL droplet of purified protein samples (0.05 – 0.15 mg/mL) for 1 min. The grids were then washed three times with 10 μL droplets of H$_2$O, followed by two incubations on 10 μL droplets of uranyl acetate for 20 s.

Imaging was carried out with a JOEL 2100-JEM transmission electron microscope, operated at 200 kV. Images were processed using the Gatan software.

6.13 Fluorescent peptide assay for GFP(+36)-TEVp activity

AaLS-13/TEVp complex was prepared according to published procedures.[142] AaLS-13 cage (180 μM, respect to monomer) and GFP(+36)-TEVp (2 μM) were mixed in 50 mM sodium phosphate buffer (pH 8.0), 200 mM NaCl, 1 mM DTT and 1 mM EDTA, then incubated at room temperature (20 °C) for 2 hours. The protein mixture was used for the activity measurements without any further purification.

FRET substrate peptides dissolved in 50 mM sodium phosphate buffer (pH 7.4) containing 1 mM DTT, 1 mM EDTA and 1% dimethyl sulfoxide (DMSO) were mixed with a 2 µM enzyme stock solution in a 96-well plate. The reaction mixtures had a total volume of 100 µL, with final concentrations of substrates and enzymes at 100 µM and 40 nM, respectively. All reactions and controls were prepared in triplicates on the same 96-well plate. Time-dependent fluorescence of the reaction mixtures were monitored at 25 °C with a 350-10 nm excitation filter and an 450-10 nm emission filter on a FLUOstar Omega microplate reader (BMG Labtech).

6.14 GFP(+36) stability assay

The purified protein GFP(+36) was incubated with or without OaAEP1 in 50 mM MES buffer (pH 6.0) with 50 mM NaCl, 1 mM EDTA and 0.5 mM TCEP. The reaction mixtures had a total volume of 100 µL, with final concentrations of GFP(+36) at 4 µM and OaAEP1 at either 2 or 0 µM. The mixture was incubated at 20 °C for 22 h then analysed by SDS-PAGE.

6.15 Fluorescent peptide assay for cp*Aa*LS-TEVp activity

Stock solutions of the FRET substrate peptides (10 mM) were prepared by dissolving in dimethyl sulfoxide (DMSO). Peptides were added to free or encapsulated enzymes in 50 mM sodium phosphate buffer (pH 7.4) containing 200 mM NaCl, 5 mM EDTA and 1% dimethyl sulfoxide (DMSO) in a 96-well plate. The reaction mixtures had a total volume of 100 µL, with final concentrations of substrates and enzymes at 100 µM and 100 nM, respectively. Time-dependent fluorescence of the reaction mixtures were monitored at 25 °C with a 350-10 nm excitation filter and an 450-10 nm emission filter on a FLUOstar Omega microplate reader (BMG Labtech).

6.16 HPLC assay for cp*Aa*LS-OaAEP1(C247A) activity

Substrate peptides pep-M or pep-K (400 µM), were incubated with the free or encapsulated OaAEP1 (0.3–2 µM) in 50 mM NaPi buffer (pH 7.0) with 200 mM NaCl, 5 mM EDTA and 1 mM TCEP. Total volume of the reaction mixture was 100 µL. The reactions were incubated at 20 °C for 20 hours, then quenched with 10 µL of 1 M HCl.

Quenched reaction mixtures were analysed by LCMS (Section 6.3.10). The identity of each HPLC peak was confirmed by MS.

6.17 Fluorescent peptide assay for cp*Aa*LS-OaAEP1(C247A) activity

Stock solutions of the FRET substrate peptides (10 mM) and GLP (100 mM) were prepared by dissolving in dH_2O. Peptides were added to free or encapsulated OaAEP1 (0.05 µM and 0.3 µM, respectively) in 50 mM sodium phosphate buffer (pH 7.0) containing 200 mM NaCl, 5 mM EDTA in a 96-well plate. The reaction mixtures had a total volume of 100 µL, with final concentrations of FRET substrates, GLP and enzymes at 400 µM, 4 mM and 400 nM, respectively. Time-dependent fluorescence of the reaction mixtures were monitored at 25 °C with a 350-10 nm excitation filter and an 450-10 nm emission filter on a FLUOstar Omega microplate reader (BMG Labtech).

6.18 Peptide sequences for activity assay of encapsulated enzymes

Abz = 2-aminobenzoyl, Dnp = 2,4-dinitrophenyl

TEV-N: Abz-GENLYFQSGK(Dnp)K-NH_2

TEV-K: Abz-GENLYFQSGK(Dnp)KKKKKKK-NH_2

pep-N: H-GLPVSTKPVATRNGL-NH_2

pep-K: H-GLPVSTEPVATENCLGKKKKKKK-NH_2

FRET-N: Abz-STRNGLPSK(Dnp)-NH_2

FRET-K: Abz-STRNGLPSK(Dnp)KKKKKK-NH_2

7 References

1.   Hara-Nishimura, I., Takeuchi, Y. & Nishimura, M. Molecular characterization of a vacuolar processing enzyme related to a putative cysteine proteinase of Schistosoma mansoni. *Plant Cell* **5**, 1651 (1993).
2.   Chen, J.-M., Dando, P. M., Stevens, R. A. E., Fortunato, M. & Barrett, A. J. Cloning and expression of mouse legumain, a lysosomal endopeptidase. *Biochem. J.* **335**, 111–117 (1998).
3.   Dall, E. & Brandstetter, H. Structure and function of legumain in health and disease. *Biochimie* **122**, 126–150 (2016).
4.   Hemu, X. *et al.* Structural determinants for peptide-bond formation by asparaginyl ligases. *Proc. Natl. Acad. Sci.* **116**, 11737–11746 (2019).
5.   Chen, J.-M., Rawlings, N. D., Stevens, R. A. E. & Barrett, A. J. Identification of the active site of legumain links it to caspases, clostripain and gingipains in a new clan of cysteine endopeptidases. *FEBS Lett.* **441**, 361–365 (1998).
6.   Barrett, A. J. & Rawlings, N. D. Evolutionary Lines of Cysteine Peptidases. *Biol. Chem.* **382**, 727–734 (2001).
7.   Elsässer, B. *et al.* Distinct Roles of Catalytic Cysteine and Histidine in the Protease and Ligase Mechanisms of Human Legumain As Revealed by DFT-Based QM/MM Simulations. *ACS Catal.* **7**, 5585–5593 (2017).
8.   Chen, J.-M. *et al.* Cloning, Isolation, and Characterization of Mammalian Legumain, an Asparaginyl Endopeptidase. *J. Biol. Chem.* **272**, 8090–8098 (1997).
9.   Hayashi, Y. *et al.* A proteinase-storing body that prepares for cell death or stresses in the epidermal cells of Arabidopsis. *Plant Cell Physiol.* **42**, 894–9 (2001).
10.   Takeda, O. *et al.* Isolation and Analysis of cDNA Encoding a Precursor of Canavalia ensiformis Asparaginyl Endopeptidase (Legumain). *J. Biochem.* **116**, 541–546 (1994).
11.   Zauner, F. B. *et al.* Crystal Structure of Plant Legumain Reveals a Unique Two-Chain State with pH-Dependent Activity Regulation. *Plant Cell* **30**, 686–699 (2018).
12.   Otegui, M. S., Herder, R., Schulze, J., Jung, R. & Staehelin, L. A. The Proteolytic Processing of Seed Storage Proteins in Arabidopsis Embryo Cells Starts in the Multivesicular Bodies. *Plant Cell* **18**, 2567–2581 (2006).
13.   Hara-Nishimura, I. & Hatsugai, N. The role of vacuole in plant cell death. *Cell Death Differ.* **18**, 1298 (2011).
14.   Yang, R. L. *et al.* Engineering a Catalytically Efficient Recombinant Protein Ligase. *J. Am. Chem. Soc.* **139**, 5351–5358 (2017).
15.   Yamada, K., Shimada, T., Nishimura, M. & Hara-Nishimura, I. A VPE family supporting various vacuolar functions in plants. *Physiol. Plant.* **123**, 369–375 (2005).
16.   Hara-Nishimura, I., Inoue, K. & Nishimura, M. A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms. *FEBS Lett.* **294**, 89–93 (1991).
17.   Hatsugai, N., Yamada, K., Goto-Yamada, S. & Hara-Nishimura, I. Vacuolar processing enzyme in plant programmed cell death. *Front. Plant Sci.* **6**, 234 (2015).
18.   Manoury, B. *et al.* An asparaginyl endopeptidase processes a microbial antigen for class II MHC presentation. *Nature* **396**, 695–699 (1998).
19.   Manoury, B. *et al.* Destructive processing by asparagine endopeptidase limits presentation of a dominant T cell epitope in MBP. *Nat. Immunol.* **3**, 169–174

(2002).

20. Sepulveda, F. E. *et al.* Critical Role for Asparagine Endopeptidase in Endocytic Toll-like Receptor Signaling in Dendritic Cells. *Immunity* **31**, 737–748 (2009).

21. Saska, I. *et al.* An asparaginyl endopeptidase mediates in vivo protein backbone cyclization. *J. Biol. Chem.* **282**, 29721–8 (2007).

22. Harris, K. S. *et al.* Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat. Commun.* **6**, 10199 (2015).

23. Nguyen, G. K. T. *et al.* Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat. Chem. Biol.* **10**, 732–738 (2014).

24. Bernath-Levin, K. *et al.* Peptide Macrocyclization by a Bifunctional Endoprotease. *Chem. Biol.* **22**, 571–582 (2015).

25. Haywood, J. *et al.* Structural basis of ribosomal peptide macrocyclization in plants. *Elife* **7**, e32955 (2018).

26. Jackson, M. A. *et al.* Molecular basis for the production of cyclic peptides by plant asparaginyl endopeptidases. *Nat. Commun.* **9**, 2411 (2018).

27. Du, J. *et al.* A bifunctional asparaginyl endopeptidase efficiently catalyzes both cleavage and cyclization of cyclic trypsin inhibitors. *Nat. Commun.* **11**, 1575 (2020).

28. Plan, M. R. R., Saska, I., Cagauan, A. G. & Craik, D. J. Backbone Cyclised Peptides from Plants Show Molluscicidal Activity against the Rice Pest Pomacea canaliculata (Golden Apple Snail). *J. Agric. Food Chem.* **56**, 5237–5241 (2008).

29. Colgrave, M. L. *et al.* Cyclotides: Natural, Circular Plant Peptides that Possess Significant Activity against Gastrointestinal Nematode Parasites of Sheep. *Biochemistry* **47**, 5581–5589 (2008).

30. Colgrave, M. L. & Craik, D. J. Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot. *Biochemistry* **43**, 5965–75 (2004).

31. Jennings, C., West, J., Waine, C., Craik, D. & Anderson, M. Biosynthesis and Insecticidal Properties of Plant Cyclotides: The Cyclic Knotted Proteins from Oldenlandia affinis. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10614–10619 (2001).

32. Jennings, C. V *et al.* Isolation, Solution Structure, and Insecticidal Activity of Kalata B2, a Circular Protein with a Twist: Do Möbius Strips Exist in Nature?,. *Biochemistry* **44**, 851–860 (2005).

33. Poth, A. G., Colgrave, M. L., Lyons, R. E., Daly, N. L. & Craik, D. J. Discovery of an unusual biosynthetic origin for circular proteins in legumes. *Proc. Natl. Acad. Sci.* **108**, 10127–10132 (2011).

34. Mylne, J. S. *et al.* Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. *Nat. Chem. Biol.* **7**, 257–259 (2011).

35. Dall, E. & Brandstetter, H. Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. *Proc. Natl. Acad. Sci.* **110**, 10940–10945 (2013).

36. James, A. M. *et al.* The macrocyclizing protease butelase 1 remains autocatalytic and reveals the structural basis for ligase activity. *Plant J.* **98**, 988–999 (2019).

37. Nguyen, G. K. T. *et al.* Butelase-mediated cyclization and ligation of peptides and proteins. *Nat. Protoc.* **11**, 1977–1988 (2016).

38. Dall, E., Fegg, J. C., Briza, P. & Brandstetter, H. Structure and Mechanism of an Aspartimide-Dependent Peptide Ligase in Human Legumain. *Angew. Chemie Int. Ed.* **54**, 2917–2921 (2015).

39. Harris, K. S. *et al.* A suite of kinetically superior AEP ligases can cyclise an intrinsically disordered protein. *Sci. Rep.* **9**, 10820 (2019).

40. Schechter, I. & Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–62 (1967).

41. Nguyen, G. K. T., Hemu, X., Quek, J.-P. & Tam, J. P. Butelase-Mediated Macrocyclization of d-Amino-Acid-Containing Peptides. *Angew. Chemie Int. Ed.* **55**, 12802–12806 (2016).

42. Nguyen, G. K. T., Cao, Y., Wang, W., Liu, C. F. & Tam, J. P. Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1 and Thiodepsipeptide. *Angew. Chemie Int. Ed.* **54**, 15694–15698 (2015).

43. Mikula, K. M., Tascón, I., Tommila, J. J. & Iwaï, H. Segmental isotopic labeling of a single-domain globular protein without any refolding step by an asparaginyl endopeptidase. *FEBS Lett.* **591**, 1285–1294 (2017).

44. Hemu, X., Zhang, X. & Tam, J. P. Ligase-Controlled Cyclo-oligomerization of Peptides. *Org. Lett.* **21**, 2029–2032 (2019).

45. Nguyen, G. K. T. *et al.* Butelase 1: A Versatile Ligase for Peptide and Protein Macrocyclization. *J. Am. Chem. Soc.* **137**, 15398–15401 (2015).

46. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci.* **108**, 11399–11404 (2011).

47. Chen, L. *et al.* Improved variants of SrtA for site-specific conjugation on antibodies and proteins with high efficiency. *Sci. Rep.* **6**, 31899 (2016).

48. Zou, Z. *et al.* Directed sortase A evolution for efficient site-specific bioconjugations in organic co-solvents. *Chem. Commun.* **54**, 11467–11470 (2018).

49. Sabari, B. R., Zhang, D., Allis, C. D. & Zhao, Y. Metabolic regulation of gene expression through histone acylations. *Nat. Rev. Mol. Cell Biol.* **18**, 90–101 (2017).

50. Guccione, E. & Richard, S. The regulation, functions and clinical relevance of arginine methylation. *Nat. Rev. Mol. Cell Biol.* **20**, 642–657 (2019).

51. Chatterjee, C. & Muir, T. W. Chemical Approaches for Studying Histone Modifications. *J. Biol. Chem.* **285**, 11045–11050 (2010).

52. Schjoldager, K. T.-B. G. *et al.* O -Glycosylation Modulates Proprotein Convertase Activation of Angiopoietin-like Protein 3. *J. Biol. Chem.* **285**, 36293–36303 (2010).

53. Hoeg-Jensen, T., Havelund, S., Nielsen, P. K. & Markussen, J. Reversible insulin self-assembly under carbohydrate control. *J. Am. Chem. Soc.* **127**, 6158–9 (2005).

54. Kudirka, R. *et al.* Generating Site-Specifically Modified Proteins via a Versatile and Stable Nucleophilic Carbon Ligation. *Chem. Biol.* **22**, 293–298 (2015).

55. Chau, C. H., Steeg, P. S. & Figg, W. D. Antibody-drug conjugates for cancer. *Lancet* **394**, 793–804 (2019).

56. Thompson, R. E. & Muir, T. W. Chemoenzymatic Semisynthesis of Proteins. *Chem. Rev.* **120**, 3051–3126 (2020).

57. Wong, L. S., Khan, F. & Micklefield, J. Selective Covalent Protein Immobilization: Strategies and Applications. *Chem. Rev.* **109**, 4025–4053 (2009).

58. Basso, A. & Serban, S. Industrial applications of immobilized enzymes—A review. *Mol. Catal.* **479**, 110607 (2019).

59. Antos, J. M. *et al.* A Straight Path to Circular Proteins. *J. Biol. Chem.* **284**,

16028–16036 (2009).

60. Craik, D. J., Daly, N. L., Bond, T. & Waine, C. Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J. Mol. Biol.* **294**, 1327–1336 (1999).

61. Dang, B., Kubota, T., Mandal, K., Bezanilla, F. & Kent, S. B. H. Native Chemical Ligation at Asx-Cys, Glx-Cys: Chemical Synthesis and High-Resolution X-ray Structure of ShK Toxin by Racemic Protein Crystallography. *J. Am. Chem. Soc.* **135**, 11911–11919 (2013).

62. Zheng, J. S., Tang, S., Qi, Y. K., Wang, Z. P. & Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc.* **8**, 2483–2495 (2013).

63. Minato, Y., Ueda, T., Machiyama, A., Shimada, I. & Iwaï, H. Segmental isotopic labeling of a 140 kDa dimeric multi-domain protein CheA from Escherichia coli by expressed protein ligation and protein trans-splicing. *J. Biomol. NMR* **53**, 191–207 (2012).

64. Aranko, A. S., Oeemig, J. S. & Iwaï, H. Structural basis for protein trans - splicing by a bacterial intein-like domain - protein ligation without nucleophilic side chains. *FEBS J.* **280**, 3256–3269 (2013).

65. Kale, S. S. *et al.* Cyclization of peptides with two chemical bridges affords large scaffold diversities. *Nat. Chem.* **10**, 715–723 (2018).

66. Ludewig, H. *et al.* Characterization of the Fast and Promiscuous Macrocyclase from Plant PCY1 Enables the Use of Simple Substrates. *ACS Chem. Biol.* **13**, 801–811 (2018).

67. Barber, C. J. S. *et al.* The Two-step Biosynthesis of Cyclic Peptides from Linear Precursors in a Member of the Plant Family Caryophyllaceae Involves Cyclization by a Serine Protease-like Enzyme. *J. Biol. Chem.* **288**, 12500–12510 (2013).

68. Hemu, X., To, J., Zhang, X. & Tam, J. P. Immobilized Peptide Asparaginyl Ligases Enhance Stability and Facilitate Macrocyclization and Site-Specific Ligation. *J. Org. Chem.* **85**, 1504–1512 (2020).

69. Rehm, F. B. H. *et al.* Site-Specific Sequential Protein Labeling Catalyzed by a Single Recombinant Ligase. *J. Am. Chem. Soc.* **141**, 17388–17393 (2019).

70. Pi, N. *et al.* Recombinant Butelase-Mediated Cyclization of the p53-Binding Domain of the Oncoprotein MdmX-Stabilized Protein Conformation as a Promising Model for Structural Investigation. *Biochemistry* **58**, 3005–3015 (2019).

71. Wang, C. K. & Craik, D. J. Designing macrocyclic disulfide-rich peptides for biotechnological applications. *Nat. Chem. Biol.* **14**, 417–427 (2018).

72. Chan, L. Y., Craik, D. J. & Daly, N. L. Dual-targeting anti-angiogenic cyclic peptides as potential drug leads for cancer therapy. *Sci. Rep.* **6**, 35347 (2016).

73. Adda, C. G. *et al.* Plasmodium falciparum merozoite surface protein 2 is unstructured and forms amyloid-like fibrils. *Mol. Biochem. Parasitol.* **166**, 159–171 (2009).

74. Zhang, X. *et al.* Solution Conformation, Backbone Dynamics and Lipid Interactions of the Intrinsically Unstructured Malaria Surface Protein MSP2. *J. Mol. Biol.* **379**, 105–121 (2008).

75. Adda, C. G. *et al.* Antigenic Characterization of an Intrinsically Unstructured Protein, Plasmodium falciparum Merozoite Surface Protein 2. *Infect. Immun.* **80**, 4177–4185 (2012).

76. Vakhrushev, S. Y. *et al.* Enhanced mass spectrometric mapping of the human

GalNAc-type O-glycoproteome with simplecells. *Mol. Cell. Proteomics* **12**, 932–944 (2013).

77.   Tran, D. T. *et al.* Multiple members of the UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase family are essential for viability in Drosophila. *J. Biol. Chem.* **287**, 5243–52 (2012).

78.   Rape, M. Ubiquitylation at the crossroads of development and disease. *Nat. Rev. Mol. Cell Biol.* **19**, 59–70 (2018).

79.   Choi, J. *et al.* Engineering Orthogonal Polypeptide GalNAc-Transferase and UDP-Sugar Pairs. *J. Am. Chem. Soc.* **141**, 13442–13453 (2019).

80.   Gunnoo, S. B. & Madder, A. Bioconjugation – using selective chemistry to enhance the properties of proteins and peptides as therapeutics and carriers. *Org. Biomol. Chem.* **14**, 8002–8013 (2016).

81.   Jackson, D. Y. *et al.* A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. *Science (80-.).* **266**, 243–247 (1994).

82.   Chen, I., Howarth, M., Lin, W. & Ting, A. Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* **2**, 99–104 (2005).

83.   Slavoff, S. A., Chen, I., Choi, Y.-A. & Ting, A. Y. Expanding the Substrate Tolerance of Biotin Ligase through Exploration of Enzymes from Diverse Species. *J. Am. Chem. Soc.* **130**, 1160–1162 (2008).

84.   Kulkarni, C., Lo, M., Fraseur, J. G., Tirrell, D. A. & Kinzer-Ursem, T. L. Bioorthogonal Chemoenzymatic Functionalization of Calmodulin for Bioconjugation Applications. *Bioconjug. Chem.* **26**, 2153–2160 (2015).

85.   Grünewald, J. *et al.* Efficient Preparation of Site-Specific Antibody–Drug Conjugates Using Phosphopantetheinyl Transferases. *Bioconjug. Chem.* **26**, 2554–2562 (2015).

86.   Viswanathan, R., Labadie, G. R. & Poulter, C. D. Regioselective covalent immobilization of catalytically active glutathione S-transferase on glass slides. *Bioconjug. Chem.* **24**, 571–577 (2013).

87.   Carrico, I. S., Carlson, B. L. & Bertozzi, C. R. Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* **3**, 321–322 (2007).

88.   Agarwal, P., van der Weijden, J., Sletten, E. M., Rabuka, D. & Bertozzi, C. R. A Pictet-Spengler ligation for protein chemical modification. *Proc. Natl. Acad. Sci.* **110**, 46–51 (2013).

89.   Hudak, J. E., Yu, H. H. & Bertozzi, C. R. Protein glycoengineering enabled by the versatile synthesis of aminooxy glycans and the genetically encoded aldehyde tag. *J. Am. Chem. Soc.* **133**, 16127–16135 (2011).

90.   Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **5**, 4740 (2014).

91.   Stephanopoulos, N. & Francis, M. B. Choosing an effective protein bioconjugation strategy. *Nat. Chem. Biol.* **7**, 876 (2011).

92.   Schmidt, M., Toplak, A., Quaedflieg, P. J. L. M. & Nuijens, T. Enzyme-mediated ligation technologies for peptides and proteins. *Curr. Opin. Chem. Biol.* **38**, 1–7 (2017).

93.   Zhang, Y., Park, K.-Y., Suazo, K. F. & Distefano, M. D. Recent progress in enzymatic protein labelling techniques and their applications. *Chem. Soc. Rev.* **47**, 9106–9136 (2018).

94.   Antos, J. M. *et al.* Site-Specific Protein Labeling via Sortase-Mediated Transpeptidation. *Curr. Protoc. Protein Sci.* **89**, 15.3.1-15.3.19 (2017).

95. Yoshihara, H. A. I., Mahrus, S. & Wells, J. A. Tags for labeling protein N-termini with subtiligase for proteomics. *Bioorg. Med. Chem. Lett.* **18**, 6000–6003 (2008).

96. Chang, T. K., Jackson, D. Y., Burnier, J. P. & Wells, J. A. Subtiligase: a tool for semisynthesis of proteins. *Proc. Natl. Acad. Sci.* **91**, 12544–12548 (1994).

97. Mikula, K. M., Krumwiede, L., Plückthun, A. & Iwaï, H. Segmental isotopic labeling by asparaginyl endopeptidase-mediated protein ligation. *J. Biomol. NMR* **71**, 225–235 (2018).

98. Cao, Y., Nguyen, G. K. T., Tam, J. P. & Liu, C.-F. Butelase-mediated synthesis of protein thioesters and its application for tandem chemoenzymatic ligation. *Chem. Commun.* **51**, 17289–17292 (2015).

99. Bi, X. *et al.* Enzymatic Engineering of Live Bacterial Cell Surfaces Using Butelase 1. *Angew. Chemie Int. Ed.* **56**, 7822–7825 (2017).

100. Rehm, F. B. H., Tyler, T. J., Yap, K., Durek, T. & Craik, D. J. Improved Asparaginyl-Ligase-Catalyzed Transpeptidation via Selective Nucleophile Quenching. *Angew. Chemie Int. Ed.* **60**, 4004–4008 (2021).

101. Lin, M. T. *et al.* A rapid and robust method for selective isotope labeling of proteins. *Methods* **55**, 370–378 (2011).

102. Marley, J., Lu, M. & Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR* **20**, 71–75 (2001).

103. Arrowsmith, C. H. & Yu-Sung, W. NMR of large (s> 25 kDa) proteins and protein complexes. *Prog. Nucl. Magn. Reson. Spectrosc.* **32**, 277–286 (1998).

104. Luk, L. Y. P. *et al.* Chemical Ligation and Isotope Labeling to Locate Dynamic Effects during Catalysis by Dihydrofolate Reductase. *Angew. Chemie Int. Ed.* **54**, 9016–20 (2015).

105. Harmand, T. J. *et al.* One-Pot Dual Labeling of IgG 1 and Preparation of C-to-C Fusion Proteins Through a Combination of Sortase A and Butelase 1. *Bioconjug. Chem.* **29**, 3245–3249 (2018).

106. Williamson, D. J., Webb, M. E. & Turnbull, W. B. Depsipeptide substrates for sortase-mediated N-terminal protein ligation. *Nat. Protoc.* **9**, 253–262 (2014).

107. Meyer, C., Liebscher, S. & Bordusa, F. Selective Coupling of Click Anchors to Proteins via Trypsiligase. *Bioconjug. Chem.* **27**, 47–53 (2016).

108. Liebscher, S. *et al.* N-Terminal Protein Modification by Substrate-Activated Reverse Proteolysis. *Angew. Chemie Int. Ed.* **53**, 3024–3028 (2014).

109. Marega, R., Prasetyanto, E. A., Michiels, C., De Cola, L. & Bonifazi, D. Fast Targeting and Cancer Cell Uptake of Luminescent Antibody-Nanozeolite Bioconjugates. *Small* **12**, 5431–5441 (2016).

110. Karmani, L. *et al.* Biodistribution of 125I-labeled anti-endoglin antibody using SPECT/CT imaging: Impact of in vivo deiodination on tumor accumulation in mice. *Nucl. Med. Biol.* **43**, 415–423 (2016).

111. Alfano, M. & Cavazza, C. Structure, function, and biosynthesis of nickel-dependent enzymes. *Protein Sci.* **29**, 1071–1089 (2020).

112. de Reuse, H., Vinella, D. & Cavazza, C. Common themes and unique proteins for the uptake and trafficking of nickel, a metal essential for the virulence of Helicobacter pylori. *Front. Cell. Infect. Microbiol.* **3**, 94 (2013).

113. Kim, E. Y. & Tullman-Ercek, D. Engineering nanoscale protein compartments for synthetic organelles. *Curr. Opin. Biotechnol.* **24**, 627–632 (2013).

114. Huttanus, H. M. & Feng, X. Compartmentalized metabolic engineering for biochemical and biofuel production. *Biotechnol. J.* **12**, (2017).

115. Menon, G., Okeke, C. & Krishnan, J. Modelling compartmentalization towards

elucidation and engineering of spatial organization in biochemical pathways. *Sci. Rep.* **7**, 12057 (2017).

116. Bobik, T. A. Polyhedral organelles compartmenting bacterial metabolic processes. *Appl. Microbiol. Biotechnol.* **70**, 517–525 (2006).
117. Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. & Sutter, M. Bacterial microcompartments. *Nat. Rev. Microbiol.* **16**, 277–290 (2018).
118. Bobik, T. A., Lehman, B. P. & Yeates, T. O. Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. *Mol. Microbiol.* **98**, 193–207 (2015).
119. Sharma, J. & Douglas, T. Tuning the catalytic properties of P22 nanoreactors through compositional control. *Nanoscale* **12**, 336–346 (2020).
120. Patterson, D. P., Schwarz, B., Waters, R. S., Gedeon, T. & Douglas, T. Encapsulation of an Enzyme Cascade within the Bacteriophage P22 Virus-Like Particle. *ACS Chem. Biol.* **9**, 359–365 (2014).
121. Liu, X. & Theil, E. C. Ferritins: dynamic management of biological iron and oxygen chemistry. *Acc. Chem. Res.* **38**, 167–175 (2005).
122. Kerfeld, C. A. & Melnicki, M. R. Assembly, function and evolution of cyanobacterial carboxysomes. *Curr. Opin. Plant Biol.* **31**, 66–75 (2016).
123. Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C. & Shively, J. M. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat. Rev. Microbiol.* **6**, 681–691 (2008).
124. Morgunova, E. *et al.* Crystal Structure of Lumazine Synthase from Mycobacterium tuberculosis as a Target for Rational Drug Design: Binding Mode of a New Class of Purinetrione Inhibitors,. *Biochemistry* **44**, 2746–2758 (2005).
125. Zhang, X., Meining, W., Fischer, M., Bacher, A. & Ladenstein, R. X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 A resolution: determinants of thermostability revealed from structural comparisons. *J. Mol. Biol.* **306**, 1099–1114 (2001).
126. Ladenstein, R., Fischer, M. & Bacher, A. The lumazine synthase/riboflavin synthase complex: shapes and functions of a highly variable enzyme system. *FEBS J.* **280**, 2537–2563 (2013).
127. Chowdhury, C. *et al.* Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc. Natl. Acad. Sci.* **112**, 2990 LP – 2995 (2015).
128. Price, G. D. & Badger, M. R. Isolation and Characterization of High CO_2-Requiring-Mutants of the Cyanobacterium Synechococcus PCC7942 : Two Phenotypes that Accumulate Inorganic Carbon but Are Apparently Unable to Generate CO_2 within the Carboxysome. *Plant Physiol.* **91**, 514–25 (1989).
129. Price, G. D. & Badger, M. R. Expression of Human Carbonic Anhydrase in the Cyanobacterium Synechococcus PCC7942 Creates a High CO_2-Requiring Phenotype. *Plant Physiol.* **91**, 505–513 (1989).
130. Wörsdörfer, B., Woycechowsky, K. J. & Hilvert, D. Directed Evolution of a Protein Container. *Science (80-.).* **331**, 589–592 (2011).
131. Azuma, Y., Herger, M. & Hilvert, D. Diversification of Protein Cage Structure Using Circularly Permuted Subunits. *J. Am. Chem. Soc.* **140**, 558–561 (2018).
132. Terasaka, N., Azuma, Y. & Hilvert, D. Laboratory evolution of virus-like nucleocapsids from nonviral protein cages. *Proc. Natl. Acad. Sci.* **115**, 5432 LP – 5437 (2018).

133. Benner, N. L. *et al.* Vault Nanoparticles: Chemical Modifications for Imaging and Enhanced Delivery. *ACS Nano* **11**, 872–881 (2017).
134. Bacher, A. *et al.* Riboflavin synthases of Bacillus subtilis. Purification and properties. *J. Biol. Chem.* **255**, 632–637 (1980).
135. Fischer, M. & Bacher, A. Biosynthesis of vitamin B2: Structure and mechanism of riboflavin synthase. *Arch. Biochem. Biophys.* **474**, 252–265 (2008).
136. Kis, K. & Bacher, A. Substrate Channeling in the Lumazine Synthase/Riboflavin Synthase Complex of Bacillus subtilis. *J. Biol. Chem.* **270**, 16788–16795 (1995).
137. Azuma, Y., Zschoche, R. & Hilvert, D. The C-terminal peptide of Aquifex aeolicus riboflavin synthase directs encapsulation of native and foreign guests by a cage-forming lumazine synthase. *J. Biol. Chem.* **292**, 10321–10327 (2017).
138. Sutter, M. *et al.* Structure of a Synthetic β-Carboxysome Shell. *Plant Physiol.* **181**, 1050–1058 (2019).
139. Wörsdörfer, B., Pianowski, Z. & Hilvert, D. Efficient in Vitro Encapsulation of Protein Cargo by an Engineered Protein Container. *J. Am. Chem. Soc.* **134**, 909–911 (2012).
140. Zschoche, R. & Hilvert, D. Diffusion-Limited Cargo Loading of an Engineered Protein Container. *J. Am. Chem. Soc.* **137**, 16121–16132 (2015).
141. Azuma, Y., Zschoche, R., Tinzl, M. & Hilvert, D. Quantitative Packaging of Active Enzymes into a Protein Cage. *Angew. Chemie Int. Ed.* **55**, 1531–1534 (2016).
142. Azuma, Y., Bader, D. L. V & Hilvert, D. Substrate Sorting by a Supercharged Nanoreactor. *J. Am. Chem. Soc.* **140**, 860–863 (2018).
143. deGruyter, J. N., Malins, L. R. & Baran, P. S. Residue-Specific Peptide Modification: A Chemist's Guide. *Biochemistry* **56**, 3863–3873 (2017).
144. Lang, K. & Chin, J. W. Bioorthogonal Reactions for Labeling Proteins. *ACS Chem. Biol.* **9**, 16–20 (2014).
145. Knudsen, L. B. *et al.* Potent Derivatives of Glucagon-like Peptide-1 with Pharmacokinetic Properties Suitable for Once Daily Administration. *J. Med. Chem.* **43**, 1664–1669 (2000).
146. Orlova, A. *et al.* Site-specific radiometal labeling and improved biodistribution using ABY-027, a novel HER2-targeting Affibody molecule–albumin-binding domain fusion protein. *J. Nucl. Med.* **54**, 961–968 (2013).
147. MacDonald, J. I., Munch, H. K., Moore, T. & Francis, M. B. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **11**, 326–331 (2015).
148. Spears, R. J. *et al.* Site-selective C–C modification of proteins at neutral pH using organocatalyst-mediated cross aldol ligations. *Chem. Sci.* **9**, 5585–5593 (2018).
149. Bloom, S. *et al.* Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. *Nat. Chem.* **10**, 205–211 (2017).
150. Faustino, H., Silva, M. J. S. A., Veiros, L. F., Bernardes, G. J. L. & Gois, P. M. P. Iminoboronates are efficient intermediates for selective, rapid and reversible N-terminal cysteine functionalisation. *Chem. Sci.* **7**, 5052–5058 (2016).
151. Bandyopadhyay, A., Cambray, S. & Gao, J. Fast and selective labeling of N-terminal cysteines at neutral pH via thiazolidino boronate formation. *Chem. Sci.* **7**, 4589–4593 (2016).
152. Shah, N. H. & Muir, T. W. Inteins: nature's gift to protein chemists. *Chem. Sci.*

5, 446–461 (2014).

153. Pinto, F., Thornton, E. L. & Wang, B. An expanded library of orthogonal split inteins enables modular multi-peptide assemblies. *Nat. Commun.* **11**, 1529 (2020).

154. Beck, T., Tetter, S., Künzle, M. & Hilvert, D. Construction of Matryoshka-Type Structures from Supercharged Protein Nanocages. *Angew. Chemie Int. Ed.* **54**, 937–940 (2015).

155. Martos-Maldonado, M. C. *et al.* Selective N-terminal acylation of peptides and proteins with a Gly-His tag sequence. *Nat. Commun.* **9**, 3307 (2018).

156. Geoghegan, K. F. *et al.* Spontaneous α-N-6-Phosphogluconoylation of a "His Tag" inEscherichia coli:The Cause of Extra Mass of 258 or 178 Da in Fusion Proteins. *Anal. Biochem.* **267**, 169–184 (1999).

157. Cumming, R. C. *et al.* Protein Disulfide Bond Formation in the Cytoplasm during Oxidative Stress. *J. Biol. Chem.* **279**, 21749–21758 (2004).

158. Futami, J. *et al.* Evaluation of irreversible protein thermal inactivation caused by breakage of disulphide bonds using methanethiosulphonate. *Sci. Rep.* **7**, 12471 (2017).

159. Li, H., Zheng, G. & Zhu, S. Construction of an organelle-like nanodevice via supramolecular self-assembly for robust biocatalysts. *Microb. Cell Fact.* **17**, 26 (2018).

160. Rosen, C. B. & Francis, M. B. Targeting the N terminus for site-selective protein modification. *Nat. Chem. Biol.* **13**, 697–705 (2017).

8 Appendices

8.1 DNA sequences for recombinant proteins

8.1.1 Full length zymogenic His₆-ubiquitin-OaAEP1(C247A) fusion protein (5' – 3'):

atgggcatggcgcaccaccaccaccaccacatgcagatcttcgttaaaaccctgaccggcaa

gaccattaccctggaagtggaaccgagcgacaccatcgagaacgtgaaagcgaagatccaag

acaaagaaggtattccgccggatcagcaacgtctgatttttgcgggcaagcagctggaggac

ggtcgtaccctgagcgattacaacatccaaaaagaaagcaccctgcatctggtgctgcgtct

gcgtggtggcgcgcgtgacggtgattatctgcacctgccgagcgaggtgtctcgtttctttc

gtccgcaggagaccaacgacgatcacggcgaagacagcgtgggtacccgttgggcggttctg

attgcgggtagcaagggctacgcgaactatcgtcatcaggcgggcgtgtgccatgcgtacca

aatcctgaaacgtggtggcctgaaggatgagaacatcgtggttttcatgtacgacgatattg

cgtataacgaaagcaacccgcgtccgggtgttatcattaacagcccgcacggcagcgatgtg

tatgcgggtgttccgaaagattataccggcgaggaagtgaacgcgaagaacttcctggcggc

gatcctgggtaacaaaagcgcgattaccggtggcagcggcaaggtggttgacagcggtccga

acgatcacatctttatttactataccgaccacggtgcggcgggcgttatcggtatgccgagc

aagccgtacctgtatgcggacgagctgaacgatgcgctgaagaaaaagcacgcgagcggtac

ctacaaaagcctggtgttctatctggaggcgtgcgaaagcggcagcatgtttgagggtatcc

tgccggaagatctgaacatttacgcgctgaccagcaccaacaccaccgaaagcagctgggcc

tactattgcccggcgcaggagaacccgccgccgccggaatataacgtgtgcctgggcgacct

gttcagcgttgcgtggctggaggacagcgatgtgcagaacagctggtacgaaaccctgaacc

agcaatatcaccacgttgataagcgtattagccacgcgagccacgcgacccaatacggtaac

ctgaaactgggcgaggaaggtctgtttgtgtacatgggcagcaacccggcgaacgacaacta

taccagcctggatggtaacgcgctgaccccgagcagcatcgtggttaaccagcgtgacgcgg

atctgctgcacctgtgggagaaattccgtaaggcgccggaaggcagcgcgcgtaaagaggaa

gcgcagacccaaatttttaaggcgatgagccaccgtgttcacatcgacagcagcatcaaact

gattggcaagctgctgttcggtatcgagaaatgcaccgaaattctgaacgcggttcgtccgg

cgggtcaaccgctggttgacgattgggcgtgcctgcgtagcctggtgggtacctttgagacc

cactgcggcagcctgagcgaatacggtatgcgtcacacccgtaccatcgcgaacatttgcaa

cgcgggtattagcgaggaacagatggcggaagcggcgagccaagcgtgcgcgagcatcccgt

aa

8.1.2 Simplified His₆-ubiquitin-OaAEP1(C247A) fusion protein (5' – 3'):

atgggcatggcgcaccaccaccaccaccacatgcagatcttcgttaaaaccctgaccggcaa

gaccattaccctggaagtggaaccgagcgacaccatcgagaacgtgaaagcgaagatccaag

acaaagaaggtattccgccggatcagcaacgtctgatttttgcgggcaagcagctggaggac

ggtcgtaccctgagcgattacaacatccaaaaagaaagcaccctgcatctggtgctgcgtct

gcgtggtggcgcgcgtgacggtgattatctgcacctgccgagcgaggtgtctcgtttcttc
gtccgcaggagaccaacgacgatcacggcgaagacagcgtgggtacccgttgggcggttctg
attgcgggtagcaagggctacgcgaactatcgtcatcaggcgggcgtgtgccatgcgtacca
aatcctgaaacgtggtggcctgaaggatgagaacatcgtggttttcatgtacgacgatattg
cgtataacgaaagcaacccgcgtccgggtgttatcattaacagcccgcacggcagcgatgtg
tatgcgggtgttccgaaagattataccggcgaggaagtgaacgcgaagaacttcctggcggc
gatcctgggtaacaaaagcgcgattaccggtggcagcggcaaggtggttgacagcggtccga
acgatcacatctttatttactataccgaccacggtgcggcgggcgttatcggtatgccgagc
aagccgtacctgtatgcggacgagctgaacgatgcgctgaagaaaaagcacgcgagcggtac
ctacaaaagcctggtgttctatctggaggcgtgcgaaagcggcagcatgtttgagggtatcc
tgccggaagatctgaacatttacgcgctgaccagcaccaacaccaccgaaagcagctgggcc
tactattgcccggcgcaggagaacccgccgccgccggaatataacgtgtgcctgggcgacct
gttcagcgttgcgtggctggaggacagcgatgtgcagaacagctggtacgaaaccctgaacc
agcaatatcaccacgttgataagcgtattagccacgcgagccacgcgacccaatacggtaac
ctgaaactgggcgaggaaggtctgtttgtgtacatgggcagcaacccggcgaacgactaa

8.1.3 Split His$_6$-ubiquitin-*Oa*AEP1-C247A fusion protein (5' – 3'):

atgggcatggcgcaccaccaccaccaccacatgcagatcttcgttaaaaccctgaccggcaa
gaccattaccctggaagtggaaccgagcgacaccatcgagaacgtgaaagcgaagatccaag
acaaagaaggtattccgccggatcagcaacgtctgatttttgcgggcaagcagctggaggac
ggtcgtaccctgagcgattacaacatccaaaaagaaagcaccctgcatctggtgctgcgtct
gcgtggtggcgcgcgtgacggtgattatctgcacctgccgagcgaggtgtctcgtttcttc
gtccgcaggagaccaacgacgatcacggcgaagacagcgtgggtacccgttgggcggttctg
attgcgggtagcaagggctacgcgaactatcgtcatcaggcgggcgtgtgccatgcgtacca
aatcctgaaacgtggtggcctgaaggatgagaacatcgtggttttcatgtacgacgatattg
cgtataacgaaagcaacccgcgtccgggtgttatcattaacagcccgcacggcagcgatgtg
tatgcgggtgttccgaaagattataccggcgaggaagtgaacgcgaagaacttcctggcggc
gatcctgggtaacaaaagcgcgattaccggtggcagcggcaaggtggttgacagcggtccga
acgatcacatctttatttactataccgaccacggtgcggcgggcgttatcggtatgccgagc
aagccgtacctgtatgcggacgagctgaacgatgcgctgaagaaaaagcacgcgagcggtac
ctacaaaagcctggtgttctatctggaggcgtgcgaaagcggcagcatgtttgagggtatcc
tgccggaagatctgaacatttacgcgctgaccagcaccaacaccaccgaaagcagctgggcc
tactattgcccggcgcaggagaacccgccgccgccggaatataacgtgtgcctgggcgacct
gttcagcgttgcgtggctggaggacagcgatgtgcagaacagctggtacgaaaccctgaacc
agcaatatcaccacgttgataagcgtattagccacgcgagccacgcgacccaatacggtaac
ctgaaactgggcgaggaaggtctgtttgtgtacatgggcagcaacccggcgaacgactaagg

atccgaaataattttgtttaactttaagaaggagatatcatatgaactataccagcctggat
ggtaacgcgctgaccccgagcagcatcgtggttaaccagcgtgacgcggatctgctgcacct
gtgggagaaattccgtaaggcgccggaaggcagcgcgcgtaaagaggaagcgcagacccaaa
tttttaaggcgatgagccaccgtgttcacatcgacagcagcatcaaactgattggcaagctg
ctgttcggtatcgagaaatgcaccgaaattctgaacgcggttcgtccggcgggtcaaccgct
ggttgacgattgggcgtgcctgcgtagcctggtgggtacctttgagacccactgcggcagcc
tgagcgaatacggtatgcgtcacacccgtaccatcgcgaacatttgcaacgcgggtattagc
gaggaacagatggcggaagcggcgagccaagcgtgcgcgagcatcccgtaa

8.1.4 eGFP-NCL

atgcaccaccaccaccaccacatggtgagcaagggcgaggagctgttcaccggggtggtgcc
catcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcg
agggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgccc
gtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccc
cgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagc
gcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggc
gacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcct
ggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcaga
agaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctc
gccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaacca
ctacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcc
tgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggtggt
agcggtaattgcttataa

8.1.5 eGFP-NCL(D235A)

atgcaccaccaccaccaccacatggtgagcaagggcgaggagctgttcaccggggtggtgcc
catcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcg
agggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgccc
gtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccc
cgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagc
gcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggc
gacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcct
ggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcaga
agaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctc
gccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaacca
ctacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcc

tgctggagttcgtgaccgccgccgggatcactctcggcatggccgagctgtacaagggtggt

agcggtaattgcttataa

8.1.6 β-lactamase-NCL

atgcaccaccaccaccaccaccacatgggtgcagatctggcagatcgttttgcagaactgga

acgtcgttatgatgcacgtctgggtgtttatgttccggcaaccggcaccaccgcagcaattg

aatatcgtgcagatgaacgttttgcattttgcagcacctttaaagcaccgctggttgcagcc

gttctgcatcagaatccgctgacacatctggataaactgattacctatacctccgatgatat

tcgtagcattagtccggttgcacagcagcatgttcagaccggtatgaccattggtcagctgt

gtgatgcagcaattcgttatagtgatggcaccgcagccaatctgctgctggcagatttaggt

ggtcctggtggtggtacagcagcctttaccggttatctgcgtagcctgggtgataccgttag

ccgtctggatgcagaagaaccggaactgaatcgtgatccgcctggtgatgaacgtgatacca

ccacaccgcatgccattgcactggttctgcagcagctggttctgggtaatgcactgcctccg

gataaacgtgcactgctgaccgattggatggcacgtaataccaccggtgccaaacgtattcg

tgcaggttttccggcagattggaaagttattgataaaaccggtacgggtgattatggtcgtg

caaatgatattgcagttgtttggagcccgaccggtgttccgtatgttgttgcagttatgagc

gatcgtgccggtggtggctatgatgccgaaccgcgtgaagcactgctggcggaagcagcaac

ctgtgttgccggtgttctggcaggatccggtggtagcggtaattgcttataa

8.1.7 *Aa*LS-13-NCL

atgcaccaccaccaccaccacatggaaatctacgaaggtaaactaactgctgaaggccttcg

tttcggtatcgtagcatcacgtttttaatcatgctcttgtcggccgtctggtggagggtgcga

ttgattgcatagtccgtcatggcggccgtgaagaagacattactctggtttgtgttccaggc

tcatgggaaataccggttgctgcgggtgaactggcgcgtaaagaggacattgatgctgttat

cgcaattggcgttctcatcgaagggggcagagccacatttcgattatatcgcctctgaagttt

caaaaggcctcgcgaacctttcattagaactacgtaaacctatcagcttcggtgatattaca

gatgacgaattggaagaggctatcgagtgcgccggcacagaacacggcaacaaaggttggga

agcagcgctttctgccattgaaatggcaaacttattcaagtctctccgactcgagggatccg

gtggtagcggtaattgcttataa

8.1.8 Ubiquitin-NCL

atgcaccaccaccaccaccaccagatcttcgttaaaaccctgaccggcaagaccattaccct

ggaagtggaaccgagcgacaccatcgagaacgtgaaagcgaagatccaagacaaagaaggta

ttccgccggatcagcaacgtctgatttttgcgggcaagcagctggaggacggtcgtaccctg

agcgattacaacatccaaaaagaaagcaccctgcatctggtgctgcgtctgcgtggtggcgg

atccggtggtagcggtaattgcttataa

8.1.9 GL-ubiquitin

Atgggccatcaccaccaccatcatggtgagaacctgtacttccagggtctgagcggtagcgg
tagccagatcttcgttaaaaccctgaccggcaagaccattaccctggaagtggaaccgagcg
acaccatcgagaacgtgaaagcgaagatccaagacaaagaaggtattccgccggatcagcaa
cgtctgattttttgcgggcaagcagctggaggacggtcgtaccctgagcgattacaacatcca
aaaagaaagcaccctgcatctggtgctgcgtctgcgtggtggctga

8.1.10 AaLS-13 (5' – 3'):

atggaaatctacgaaggtaaactaactgctgaaggccttcgtttcggtatcgtagcatcacg
ttttaatcatgctcttgtcggccgtctggtggagggtgcgattgattgcatagtccgtcatg
gcggccgtgaagaagacattactctggtttgtgttccaggctcatgggaaataccggttgct
gcgggtgaactggcgcgtaaagaggacattgatgctgttatcgcaattggcgttctcatcga
agggggcagagccacatttcgattatatcgcctctgaagtttcaaaaggcctcgcgaacctttt
cattagaactacgtaaacctatcagcttcggtgatattacagatgacgaattggaagaggct
atcgagtgcgccggcacagaacacggcaacaaaggttgggaagcagcgctttctgccattga
aatggcaaacttattcaagtctctccgactcgagcaccaccaccaccaccactaa

8.1.11 GFP(+36)-TEVp (5' – 3'):

atgcaccatcatcaccaccacggatccggtatggctagcaaaggtgaacgtctgtttcgtgg
taaagtaccgatcttagtggaattaaagggcgacgtgaacggtcataaatttagcgtgcgcg
gcaaaggcaaaggtgacgctacccgtggtaaattgaccctgaagtttatttgcacaacaggc
aaattacccgttccgtggcccaccttagtgaccaccctgacctatggcgttcagtgcttcag
tcgttaccctaaacacatgaaacgtcacgattttttcaaatcagccatgcctaaaggatatg
ttcaagagcgtacaatcagcttcaagaaggatggcaaatataaaacgcgtgcggaagtgaaa
tttgaaggccgcacattagtaaatcgtatcaaactgaaaggtcgtgacttcaaagaaaaagg
caacatttttaggccataaactgcgttataactttaattctcataaggtgtatattacggccg
ataaacgcaagaatggtatcaaggcaaaattcaaaattcgccataacgtgaaagacggcagc
gttcaattagcggatcattatcaacaaaacacgccgattggtcgcgggcctgtactgttacc
tcgcaaccactacctgagcacccgttctaaactgagcaaagatccgaaagaaaaacgcgatc
acatggttctgttagaattcgtgaccgctgcaggcattaagcacggacgcgacgaacgctac
aaggtacccggtggcagcggggggcagcggcggttcgggcggctccggcggctcgagcggaga
aagcttgtttaaggggccgcgtgattacaacccgatatcgagcaccatttgtcatttgacga
atgaatctgatgggcacacaacatcgttgtatggtattggatttggtcccttcatcattaca
aacaagcacttgtttagaagaaataatggaacactgttggtccaatcactacatggtgtatt
caaggtcaagaacaccacgactttgcaacaacacctcattgatgggagggacatgataatta
ttcgcatgcctaaggatttcccaccatttcctcaaaagctgaaatttagagagccacaaagg

gaagagcgcatatgtcttgtgacaaccaacttccaaactaagagcatgtctagcatggtgtc
agacactagttgcacattcccttcatctgatggcatattctggaagcattggattcaaacca
aggatgggcagtgtggcagtccattagtatcaactagagatgggttcattgttggtatacac
tcagcatcgaatttcaccaacacaaacaattatttcacaagcgtgccgaaaaacttcatgga
attgttgacaaatcaggaggcgcagcagtgggttagtggttggcgattaaatgctgactcag
tattgtggggggccataaagttttcatggtgaaacctgaagagcctttcagccagttaag
gaagcgactcaactcatgaattaa

8.1.12 GFP(+36)-OaAEP1(C247A) (5' – 3'):

atgcaccatcatcaccaccacggatccggtatggctagcaaaggtgaacgtctgtttcgtgg
taaagtaccgatcttagtggaattaaagggcgacgtgaacggtcataaatttagcgtgcgcg
gcaaaggcaaaggtgacgctacccgtggtaaattgaccctgaagtttatttgcacaacaggc
aaattacccgttccgtggcccaccttagtgaccaccctgacctatggcgttcagtgcttcag
tcgttaccctaaacacatgaaacgtcacgattttttcaaatcagccatgcctaaaggatatg
ttcaagagcgtacaatcagcttcaagaaggatggcaaatataaaacgcgtgcggaagtgaaa
tttgaaggccgcacattagtaaatcgtatcaaactgaaaggtcgtgacttcaaagaaaaagg
caacattttaggccataaactgcgttataactttaattctcataaggtgtatattacggccg
ataaacgcaagaatggtatcaaggcaaaattcaaaattcgccataacgtgaaagacggcagc
gttcaattagcggatcattatcaacaaaacacgccgattggtcgcgggcctgtactgttacc
tcgcaaccactacctgagcacccgttctaaactgagcaaagatccgaaagaaaaacgcgatc
acatggttctgttagaattcgtgaccgctgcaggcattaagcacggacgcgacgaacgctac
aaggtacccggtggcagcggggggcagcggcggttcgggcggctccggcggctcgagcgcgcg
tgacggtgattatctgcacctgccgagcgaggtgtctcgtttctttcgtccgcaggagacca
acgacgatcacggcgaagacagcgtgggtacccgttgggcggttctgattgcgggtagcaag
ggctacgcgaactatcgtcatcaggcgggcgtgtgccatgcgtaccaaatcctgaaacgtgg
tggcctgaaggatgagaacatcgtggttttcatgtacgacgatattgcgtataacgaaagca
acccgcgtccgggtgttatcattaacagcccgcacggcagcgatgtgtatgcgggtgttccg
aaagattataccggcgaggaagtgaacgcgaagaacttcctggcggcgatcctgggtaacaa
aagcgcgattaccggtggcagcggcaaggtggttgacagcggtccgaacgatcacatcttta
tttactataccgaccacggtgcggcgggcgttatcggtatgccgagcaagccgtacctgtat
gcggacgagctgaacgatgcgctgaagaaaaagcacgcgagcggtacctacaaaagcctggt
gttctatctggaggcgtgcgaaagcggcagcatgtttgagggtatcctgccggaagatctga
acatttacgcgctgaccagcaccaacaccaccgaaagcagctgggcctactattgcccggcg
caggagaacccgccgccgccggaatataacgtgtgcctgggcgacctgttcagcgttgcgtg
gctggaggacagcgatgtgcagaacagctggtacgaaaccctgaaccagcaatatcaccacg
ttgataagcgtattagccacgcgagccacgcgacccaatacggtaacctgaaactgggcgag

gaaggtctgtttgtgtacatgggcagcaacccggcgaacgacaactataccagcctggatgg

taacgcgctgaccccgagcagcatcgtggttaaccagcgtgacgcggatctgctgcacctgt

gggagaaattccgtaaggcgccggaaggcagcgcgcgtaaagaggaagcgcagacccaaatt

tttaaggcgatgagccaccgtgttcacatcgacagcagcatcaaactgattggcaagctgct

gttcggtatcgagaaatgcaccgaaattctgaacgcggttcgtccggcgggtcaaccgctgg

ttgacgattgggcgtgcctgcgtagcctggtgggtacctttgagacccactgcggcagcctg

agcgaatacggtatgcgtcacacccgtaccatcgcgaacatttgcaacgcgggtattagcga

ggaacagatggcggaagcggcgagccaagcgtgcgcgagcatcccgtaa

8.1.13 DNA sequence encoding for OaAEP1(C247A)-GFP(+36) (5' – 3'):

atggcgcgtgacggtgattatctgcacctgccgagcgaggtgtctcgtttctttcgtccgca

ggagaccaacgacgatcacggcgaagacagcgtgggtacccgttgggcggttctgattgcgg

gtagcaagggctacgcgaactatcgtcatcaggcgggcgtgtgccatgcgtaccaaatcctg

aaacgtggtggcctgaaggatgagaacatcgtggttttcatgtacgacgatattgcgtataa

cgaaagcaacccgcgtccgggtgttatcattaacagcccgcacggcagcgatgtgtatgcgg

gtgttccgaaagattataccggcgaggaagtgaacgcgaagaacttcctggcggcgatcctg

ggtaacaaaagcgcgattaccggtggcagcggcaaggtggttgacagcggtccgaacgatca

catctttatttactataccgaccacggtgcggcgggcgttatcggtatgccgagcaagccgt

acctgtatgcggacgagctgaacgatgcgctgaagaaaaagcacgcgagcggtacctacaaa

agcctggtgttctatctggaggcgtgcgaaagcggcagcatgtttgagggtatcctgccgga

agatctgaacatttacgcgctgaccagcaccaacaccaccgaaagcagctgggcctactatt

gcccggcgcaggagaacccgccgccgccggaatataacgtgtgcctgggcgacctgttcagc

gttgcgtggctggaggacagcgatgtgcagaacagctggtacgaaaccctgaaccagcaata

tcaccacgttgataagcgtattagccacgcgagccacgcgacccaatacggtaacctgaaac

tgggcgaggaaggtctgtttgtgtacatgggcagcaacccggcgaacgacccgggatccatg

gctagcaaaggtgaacgtctgtttcgtggtaaagtaccgatcttagtggaattaaagggcga

cgtgaacggtcataaatttagcgtgcgcggcaaaggcaaaggtgacgctacccgtggtaaat

tgaccctgaagtttatttgcacaacaggcaaattacccgttccgtggcccaccttagtgacc

accctgacctatggcgttcagtgcttcagtcgttaccctaaacacatgaaacgtcacgattt

tttcaaatcagccatgcctaaaggatatgttcaagagcgtacaatcagcttcaagaaggatg

gcaaatataaaacgcgtgcggaagtgaaatttgaaggccgcacattagtaaatcgtatcaaa

ctgaaaggtcgtgacttcaaagaaaaaggcaacatttttaggccataaactgcgttataactt

taattctcataaggtgtatattacggccgataaacgcaagaatggtatcaaggcaaaattca

aaattcgccataacgtgaaagacggcagcgttcaattagcggatcattatcaacaaaacacg

ccgattggtcgcgggcctgtactgttacctcgcaaccactacctgagcacccgttctaaact

gagcaaagatccgaaagaaaaacgcgatcacatggttctgttagaattcgtgaccgctgcag

gcattaagcacggacgcgacgaacgctacaagctcgagcaccaccaccaccaccactga

8.1.14 DNA sequence encoding for GFP(+36) (5' – 3'):

atgcaccatcatcaccaccacggatccggtatggctagcaaaggtgaacgtctgtttcgtgg

taaagtaccgatcttagtggaattaaagggcgacgtgaacggtcataaatttagcgtgcgcg

gcaaaggcaaaggtgacgctacccgtggtaaattgaccctgaagtttatttgcacaacaggc

aaattacccgttccgtggcccaccttagtgaccaccctgacctatggcgttcagtgcttcag

tcgttaccctaaacacatgaaacgtcacgattttttcaaatcagccatgcctaaaggatatg

ttcaagagcgtacaatcagcttcaagaaggatggcaaatataaaacgcgtgcggaagtgaaa

tttgaaggccgcacattagtaaatcgtatcaaactgaaaggtcgtgacttcaaagaaaaagg

caacattttaggccataaactgcgttataactttaattctcataaggtgtatattacggccg

ataaacgcaagaatggtatcaaggcaaaattcaaaattcgccataacgtgaaagacggcagc

gttcaattagcggatcattatcaacaaaacacgccgattggtcgcgggcctgtactgttacc

tcgcaaccactacctgagcacccgttctaaactgagcaaagatccgaaagaaaaacgcgatc

acatggttctgttagaattcgtgaccgctgcaggcattaagcacggacgcgacgaacgctac

aaggtacccggttaa

8.1.15 DNA sequence encoding for cpAaLS-OaAEP1(C247A) (5' – 3'):

atgaccttggaacaggctatcgagcgcgccggcacaaaacacggcaacaaaggttgggaagc

agcgctttctgccattgaaatggcaaacttattcaagtctctccgaggtaccggtggctcgg

ggagctcgatggaaatctacgaaggtaaactaactgctgaaggccttcgtttcggtatcgta

gcatcacgtttttaatcatgctcttgtcgaccgtctggtggagggtgcaattgattgcatagt

ccgtcatggcggccgtgaagaagacattactctggttcgtgttccaggctcatgggaaatac

cggttgctgcgggtgaactggcgcgtaaagaggacattgatgctgttatcgcaattggcgtt

ctcatcagaggcgcaacgccacatttcgattatatcgcctctgaagtttcaaaaggcctcgc

gaacctttcattagaactacgtaaacctatcaccttcggtgttattacagctgacgccggtg

gggcgggcggatccagcgtgggtacccgttgggcggttctgattgcgggtagcaagggctac

gcgaactatcgtcatcaggcgggcgtgtgccatgcgtaccaaatcctgaaacgtggtggcct

gaaggatgagaacatcgtggttttcatgtacgacgatattgcgtataacgaaagcaacccgc

gtccgggtgttatcattaacagcccgcacggcagcgatgtgtatgcgggtgttccgaaagat

tataccggcgaggaagtgaacgcgaagaacttcctggcggcgatcctgggtaacaaagcgc

gattaccggtggcagcggcaaggtggttgacagcggtccgaacgatcacatctttatttact

ataccgaccacggtgcggcgggcgttatcggtatgccgagcaagccgtacctgtatgcggac

gagctgaacgatgcgctgaagaaaaagcacgcgagcggtacctacaaaagcctggtgttcta

tctggaggcgtgcgaaagcggcagcatgtttgagggtatcctgccggaagatctgaacattt

<div align="right">143</div>

acgcgctgaccagcaccaacaccaccgaaagcagctgggcctactattgcccggcgcaggag
aacccgccgccgccggaatataacgtgtgcctgggcgacctgttcagcgttgcgtggctgga
ggacagcgatgtgcagaacagctggtacgaaaccctgaaccagcaatatcaccacgttgata
agcgtattagccacgcgagccacgcgacccaatacggtaacctgaaactgggcgaggaaggt
ctgtttgtgtacatgggcagcaacccggcgaacgactaa

8.1.16 DNA sequence encoding for cpAaLS-TEVp (5' – 3'):

atgaccttggaacaggctatcgagcgcgccggcacaaaacacggcaacaaaggttgggaagc
agcgctttctgccattgaaatggcaaacttattcaagtctctccgaggtaccggtggctcgg
ggagctcgatggaaatctacgaaggtaaactaactgctgaaggccttcgtttcggtatcgta
gcatcacgttttaatcatgctcttgtcgaccgtctggtggagggtgcaattgattgcatagt
ccgtcatggcggccgtgaagaagacattactctggttcgtgttccaggctcatgggaaatac
cggttgctgcgggtgaactggcgcgtaaagaggacattgatgctgttatcgcaattggcgtt
ctcatcagaggcgcaacgccacatttcgattatatcgcctctgaagtttcaaaaggcctcgc
gaacctttcattagaactacgtaaacctatcaccttcggtgttattacagctgacgccggtg
gggcgggcggatccggagaaagcttgtttaaggggccgcgtgattacaacccgatatcgagc
accatttgtcatttgacgaatgaatctgatgggcacacaacatcgttgtatggtattggatt
tggtcccttcatcattacaaacaagcacttgtttagaagaaataatggaacactgttggtcc
aatcactacatggtgtattcaaggtcaagaacaccacgactttgcaacaacacctcattgat
gggagggacatgataattattcgcatgcctaaggatttcccaccatttcctcaaaagctgaa
atttagagagccacaaagggaagagcgcatatgtcttgtgacaaccaacttccaaactaaga
gcatgtctagcatggtgtcagacactagttgcacattcccttcatctgatggcatattctgg
aagcattggattcaaaccaaggatgggcagtgtggcagtccattagtatcaactagagatgg
gttcattgttggtatacactcagcatcgaatttcaccaacacaaacaattatttcacaagcg
tgccgaaaaacttcatggaattgttgacaaatcaggaggcgcagcagtgggttagtggttgg
cgattaaatgctgactcagtattgtggggggccataaagttttcatggtgaaacctgaaga
gccttttcagccagttaaggaagcgactcaactcatgaattaa

8.2 Kinetic studies of OaAEP1-C247A peptide cyclisation
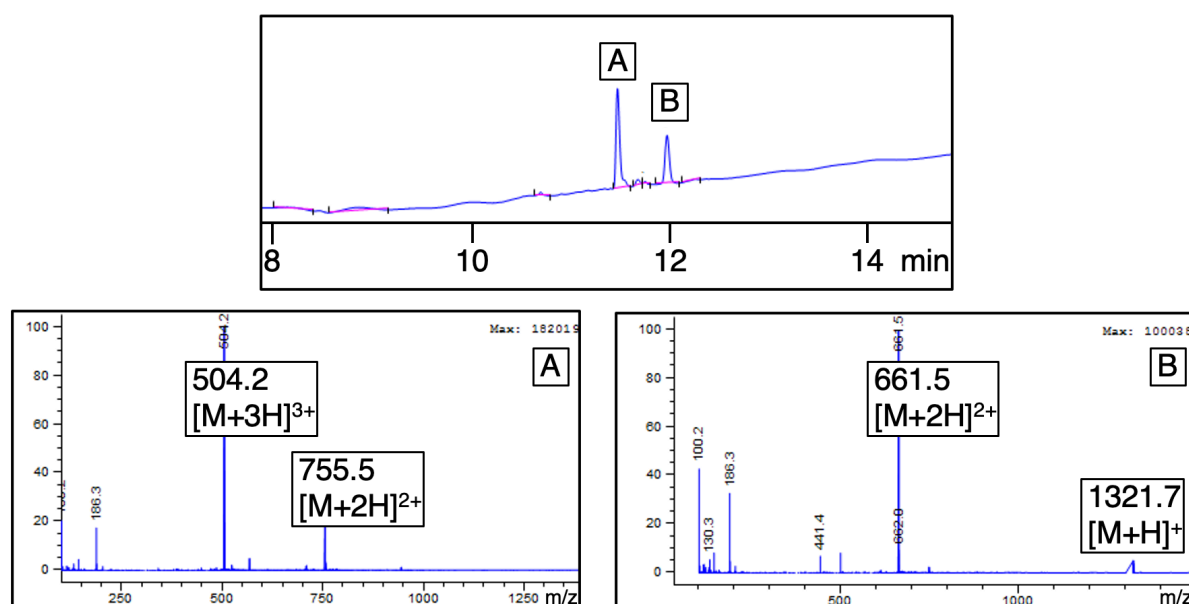
Figure 8.1 Representative HPLC chromatogram showing UV absorbance at 210 nm for the peptide cyclisation reaction by OaAEP1 (Top). GLPVSTRPVATRNGL (300 µM) was incubated with OaAEP1 (0.02 µM) for 1-hour in 50 mM NaOAc (pH 5.0), 50 mM NaCl, 1 mM EDTA. Reaction yields were obtained by integration of the UV absorbance peak area at 210 nm corresponding to the cyclic product peptide. (Bottom) Mass spectrum of peaks A and B, corresponding to the linear starting material and cyclic product, respectively. Calculated mass of GLPVSTRPVATRNGL: 1508.87 Da, cyclic-GLPVSTKPVATRN: 1320.75 Da
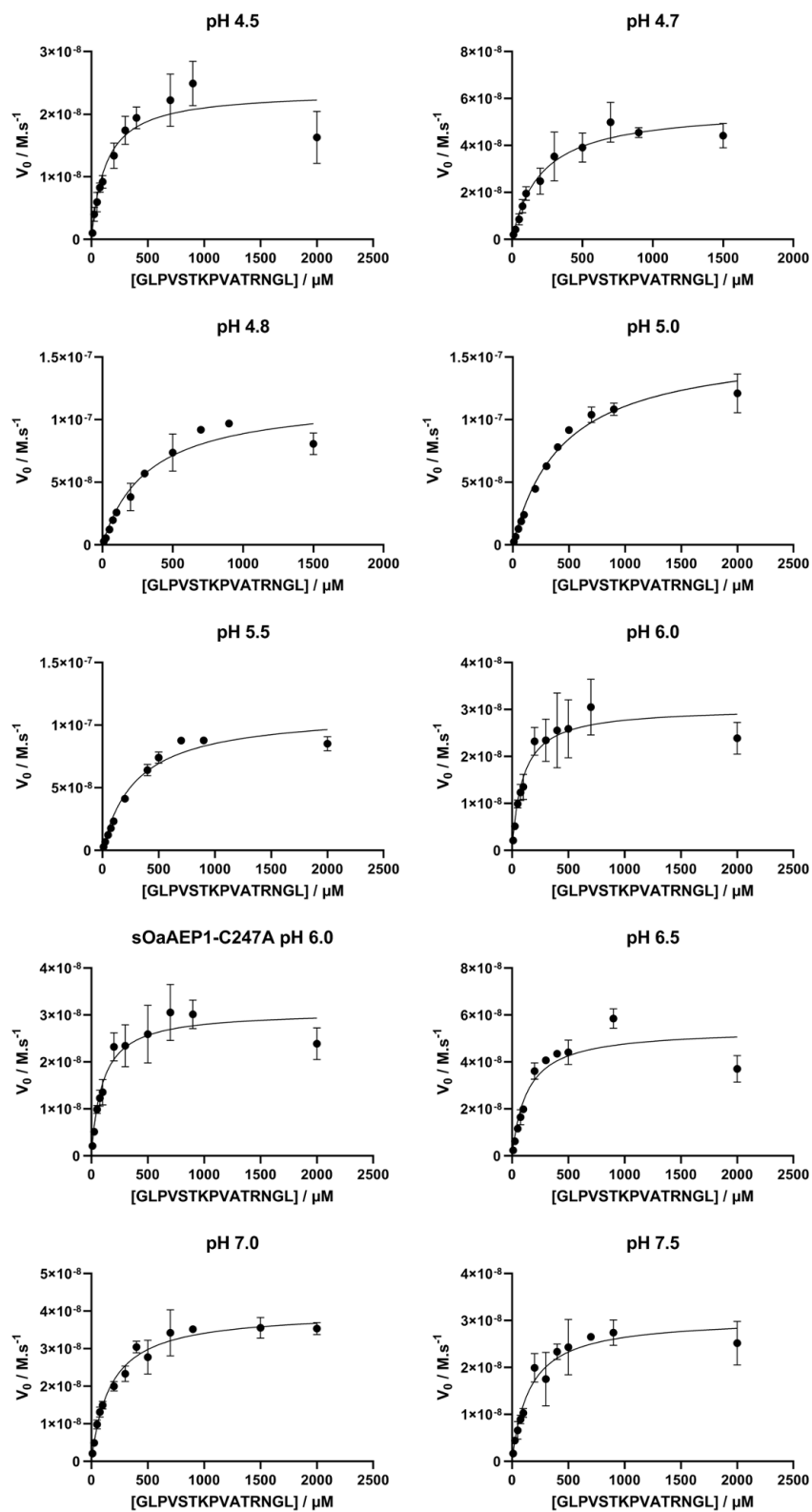
Figure 8.2 Curve fitting for the Michaelis-Menten kinetics studies of OaAEP1.

8.3 Model peptide ligation for the optimisation of AEP/FPBA approach

(A)

LFRANCLK
+
GLGGIR

$\xrightarrow[\text{20 °C}]{\text{OaAEP1-C247A}\ \text{[0.1 mol%]}}$

LFRANGLGGIR
+
CLK

(B)

Area: 469.8
GLGGIR

Area: 2246.3
LFRANCLK

Area: 1284.4
LFRANGLGGIR

(C)

Area: 274.6
GLGGIR

Area: 765.9
LFRANCLK

Area: 538.9
LFRANGLGGIR

(D)

Area: 42.3
LFRAN

Area: 137.5
GLGGIR

Area: 273.1
LFRANCLK

Area: 4030.5
FPBA

Area: 962.7
LFRANGLGGIR
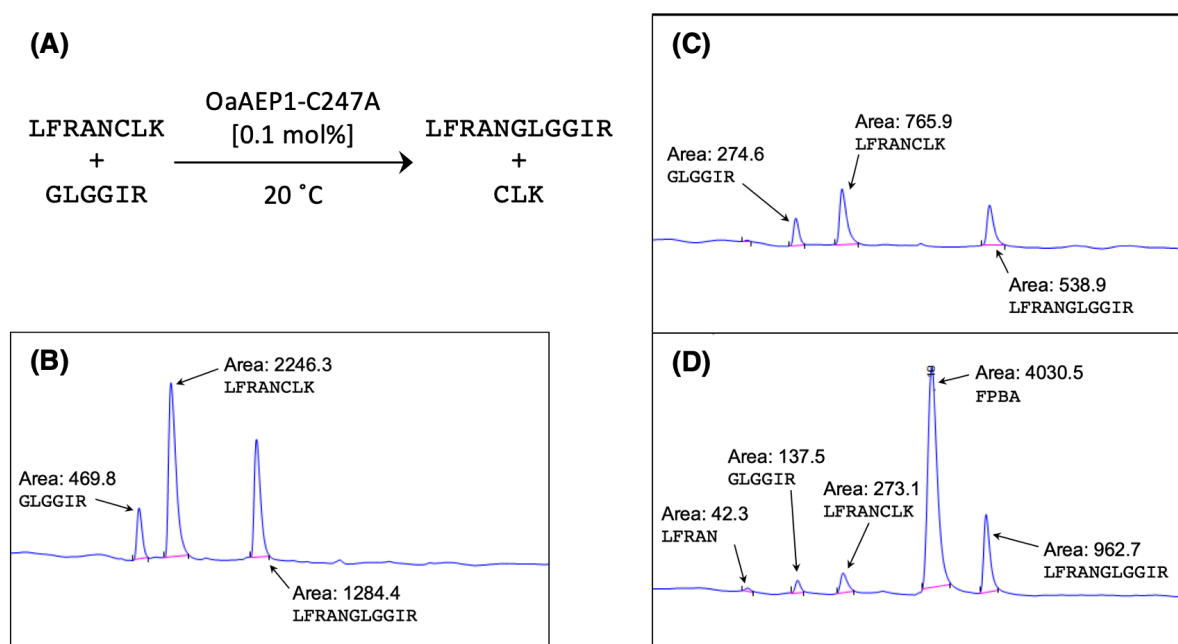
Figure 8.3 Representative HPLC chromatogram showing UV absorbance at 210 nm for the peptide ligation model reaction. (**A**) Peptide ligation reaction mediated by OaAEP1-C247A (**B**) No enzyme control. Standards of peptides LFRANCLK (300 µM), GLGGIR (300 µM) and LFRANGLGGIR (300 µM) synthesized by SPPS. (**C**) OaAEP1-C247A mediated ligation without FPBA additive. LFRANCLK (300 µM), GLGGIR (360 µM) were incubated with OaAEP1-C247A (0.3 µM) for 4 hours. (**D**) OaAEP1 mediated ligation with FPBA additive. LFRANCLK (300 µM), GLGGIR (360 µM) and FPBA (600 µM) were incubated with OaAEP1-C247A (0.3 µM) for 4 hours. Reaction yields were obtained by integration of the UV absorbance peak area at 210 nm corresponding to the product peptide LFRANGLGGIR in comparison with a calibration curve.

Table 8.1 Optimisation of model peptide ligation, where the ligation of LFRANCLK to GLGGIR is catalysed by OaAEP1 in the presence or absence of FPBA

Entry	Temperature °C	pH	Time h	[OaAEP1] μM	[FPBA] μM	[LFRANCLK] μM	[GLGGIR] μM	Yield %	Hydrolysis %
1	4	4.5[a]	2	0.4	0	400	480	27	3
2	4	4.5[a]	2	0.4	800	400	480	3	1
3	4	4.5[a]	3	0.4	0	400	480	33	8
4	4	4.5[a]	3	0.4	800	400	480	8	1
5	4	5.0[a]	1	0.1	400	200	200	23	1
6	4	5.0[a]	1	0.1	400	200	240	24	1
7	4	5.0[a]	1	0.1	400	200	400	34	1
8	4	5.0[a]	1	0.1	600	200	200	17	0
9	4	5.0[a]	1	0.2	200	200	200	25	13
10	4	5.0[a]	1	0.2	400	200	200	49	3
11	4	5.0[a]	1	0.4	400	200	200	55	11
12	4	5.0[a]	1	0.4	400	400	400	47	12
13	4	5.0[a]	1	0.6	400	200	200	46	50
14	4	5.0[a]	2	0.1	400	200	200	35	3
15	4	5.0[a]	2	0.1	400	200	240	37	1
16	4	5.0[a]	2	0.1	400	200	400	42	1
17	4	5.0[a]	2	0.1	600	200	200	31	1
18	4	5.0[a]	2	0.2	200	200	200	51	23
19	4	5.0[a]	2	0.2	400	200	200	64	9
20	4	5.0[a]	2	0.4	400	200	200	65	21
21	4	5.0[a]	2	0.4	400	400	400	66	25
22	4	5.0[a]	2	0.6	400	200	200	33	66
23	4	5.0[a]	3	0.1	400	200	200	43	4
24	4	5.0[a]	3	0.1	400	200	240	46	2
25	4	5.0[a]	3	0.1	400	200	400	49	1
26	4	5.0[a]	3	0.1	600	200	200	42	1
27	4	5.0[a]	3	0.2	400	200	200	60	13
28	4	5.0[a]	3	0.4	400	200	200	64	31
29	4	5.0[a]	3	0.6	400	200	200	27	72
30	4	5.0[a]	4	0.1	400	200	200	47	5
31	4	5.0[a]	4	0.1	400	200	240	50	2
32	4	5.0[a]	4	0.1	400	200	400	53	1
33	4	5.0[a]	4	0.1	600	200	200	44	1
34	4	5.0[a]	4	0.2	400	200	200	64	20
35	4	5.0[a]	4	0.4	400	200	200	57	38
36	4	5.0[a]	4	0.6	400	200	200	13	87
37	4	5.0[a]	18	0.2	200	200	200	43	35
38	20	4.5[a]	2	0.2	0	200	200	35	8
39	20	4.5[a]	2	0.2	200	200	200	49	9

40	20	4.5[a]	2	0.4	0	400	480	39	14
41	20	4.5[a]	2	0.4	800	400	480	58	17
42	20	4.5[a]	3	0.4	0	400	480	38	25
43	20	4.5[a]	3	0.4	800	400	480	69	19
44	20	4.7[a]	2	0.2	0	200	200	36	8
45	20	4.7[a]	2	0.2	200	200	200	49	5
46	20	5.0[a]	0.2	0.2	200	200	200	15	14
47	20	5.0[a]	0.5	0.2	200	200	200	38	19
48	20	5.0[a]	1	0.2	200	200	200	48	20
49	20	5.0[a]	1	0.4	400	400	400	62	18
50	20	5.0[a]	2	0.2	0	200	200	33	21
51	20	5.0[a]	2	0.2	200	200	200	58	32
52	20	5.0[a]	2	0.2	400	200	200	67	27
53	20	5.0[a]	2	0.3	0	300	360	42	7
54	20	5.0[a]	2	0.3	600	300	360	75	9
55	20	5.0[a]	2	0.4	400	400	400	70	23
56	20	5.0[a]	3	0.3	0	300	360	53	13
57	20	5.0[a]	3	0.3	600	300	360	89	10
58	20	5.0[a]	18	0.4	400	400	400	25	73
59	20	5.0[a]	18	0.2	200	200	200	18	78
60	20	5.2[a]	2	0.3	0	300	360	52	3
61	20	5.2[a]	2	0.3	600	300	360	73	4
62	20	5.2[a]	3	0.3	0	300	360	54	3
63	20	5.2[a]	3	0.3	600	300	360	95	3
64	20	5.5[a]	2	0.3	0	300	360	48	1
65	20	5.5[a]	2	0.3	600	300	360	69	0
66	20	5.5[a]	3	0.3	0	300	360	49	2
67	20	5.5[a]	3	0.3	600	300	360	85	1
68	20	5.5[a]	4	0.3	0	300	360	48	3
69	20	5.5[a]	4	0.3	600	300	360	90	5
70	20	5.7[b]	2	0.3	0	300	360	51	1
71	20	5.7[b]	2	0.3	600	300	360	57	0
72	20	5.7[b]	3	0.3	0	300	360	51	1
73	20	5.7[b]	3	0.3	600	300	360	78	2
74	20	5.7[b]	4	0.3	0	300	360	50	2
75	20	5.7[b]	4	0.3	600	300	360	94	3
76	20	5.9[b]	2	0.3	0	300	360	58	0
77	20	5.9[b]	2	0.3	600	300	360	28	0
78	20	5.9[b]	3	0.3	0	300	360	56	1
79	20	5.9[b]	3	0.3	600	300	360	66	1
80	20	6.0[b]	2	0.3	0	300	360	53	0
81	20	6.0[b]	2	0.3	600	300	360	32	0
82	20	6.0[b]	3	0.3	0	300	360	54	0

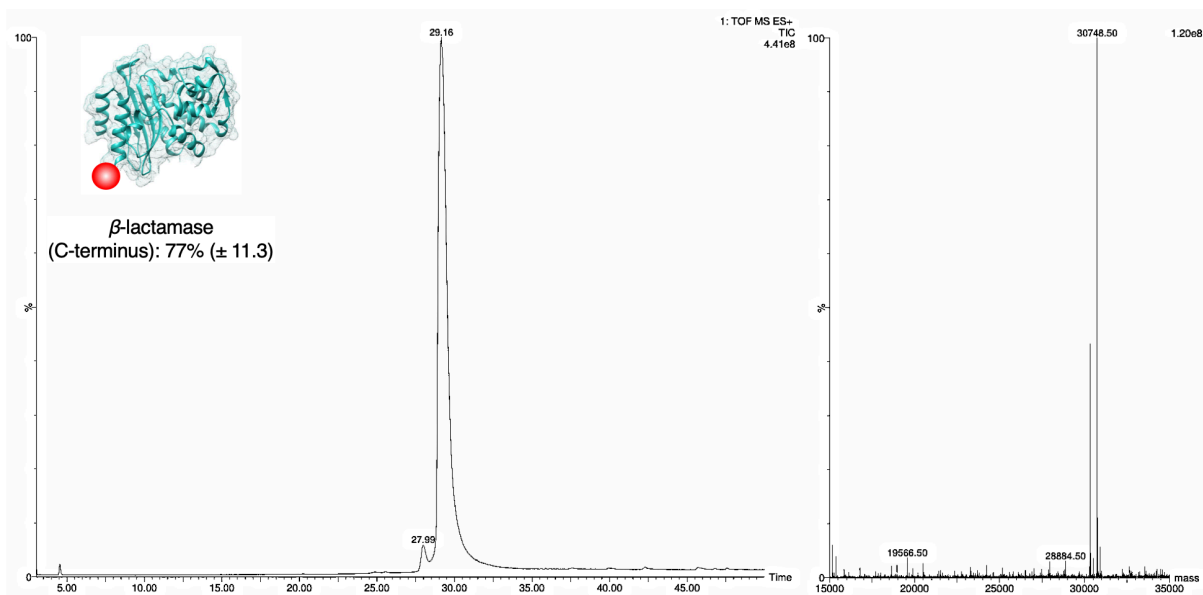| 83	20	6.0[b]	3	0.3	600	300	360	55	0
84	20	6.0[b]	4	0.3	0	300	360	56	1
85	20	6.0[b]	4	0.3	600	300	360	77	0
86	20	6.5[b]	2	0.3	0	300	360	19	0
87	20	6.5[b]	2	0.3	600	300	360	15	0
88	20	6.5[b]	3	0.3	0	300	360	57	0
89	20	6.5[b]	3	0.3	600	300	360	20	0
90	20	6.5[b]	4	0.3	0	300	360	55	0
91	20	6.5[b]	4	0.3	600	300	360	22	1
92	37	5.0[a]	0.2	0.2	200	200	200	17	13
93	37	5.0[a]	0.5	0.2	200	200	200	44	23
94	37	5.0[a]	1	0.2	200	200	200	48	44
95	37	5.0[a]	2	0.2	0	200	200	30	51
96	37	5.0[a]	2	0.2	200	200	200	49	43
97	37	6.5[b]	2	0.3	0	300	360	33	0
98	37	6.5[b]	2	0.3	600	300	360	53	0
99	37	6.5[b]	3	0.3	0	300	360	53	0
100	37	6.5[b]	3	0.3	600	300	360	82	0
101	37	6.5[b]	4	0.3	0	300	360	55	0
102	37	6.5[b]	4	0.3	600	300	360	81	0

[a] 50 mM NaOAc buffer with 50 mM NaCl, 1 mM EDTA, 0.5 mM TCEP [b] 50 mM MES buffer with 50 mM NaCl, 1 mM EDTA, 0.5 mM TCEP

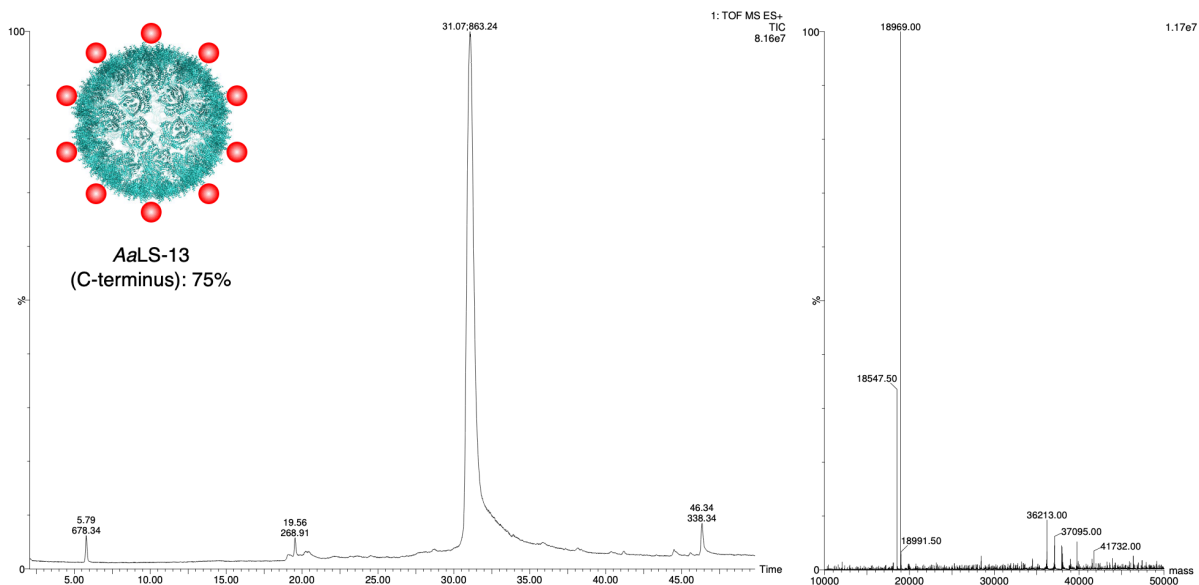## 8.4 Protein labelling using the AEP/FPBA approach

	Calculated MW	Amino acid sequence
Starting Material (SM)	28440.15 28420.15 (GFP)	MHHHHHHMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGD ATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ HDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE LKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIR HNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPN EKRDHMVLLEFVTAAGITLGMAELYKGGSGNCL
Desired Product (P)	28862.65 28842.65 (GFP)	MHHHHHHMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGD ATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ HDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE LKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIR HNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPN EKRDHMVLLEFVTAAGITLGMAELYKGGSGNGLGGZ
Hydrolysed side Product (SP)	28223.85 28203.85 (GFP)	MHHHHHHMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGD ATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ HDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE LKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIR HNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPN EKRDHMVLLEFVTAAGITLGMAELYKGGSGN

Figure 8.4 Bioconjugation of biotin labeled peptide to the C-terminus of eGFP. Chromatogram and mass spectra of the UPLC-MS analysis. Label peptide sequence: GLGGZ (Z = biotinylated lysine).

Figure 8.5 Bioconjugation of biotin labeled peptide to the C-terminus of BlaC. Chromatogram and mass spectra of the UPLC-MS analysis. Label peptide sequence: GLGGZ (Z = biotinylated lysine).
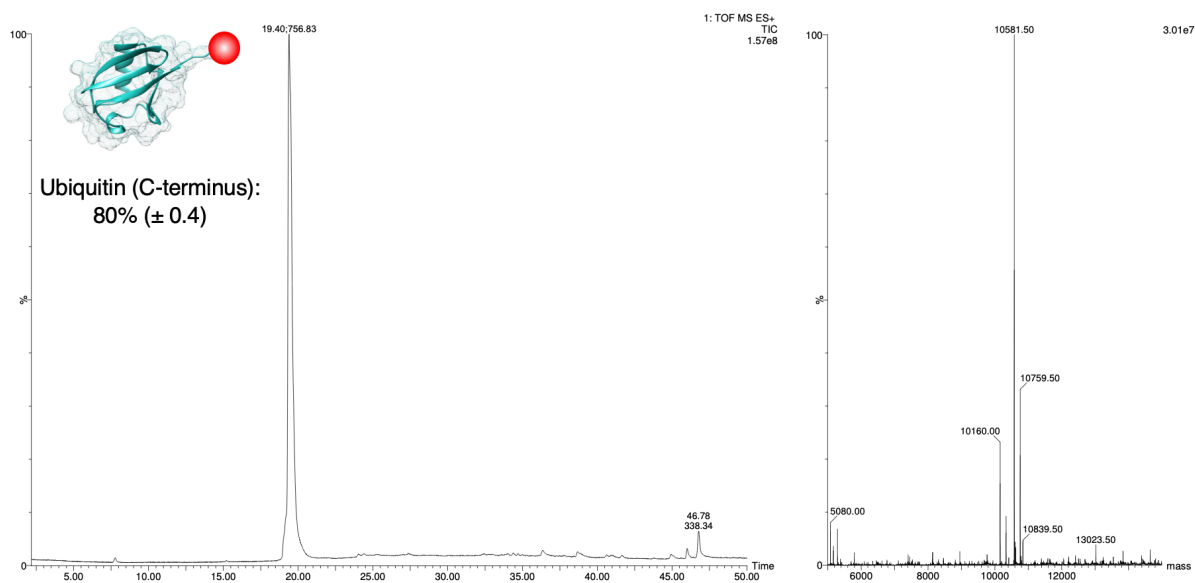
	Calculated MW	Amino acid sequence
Starting Material (SM)	30327.05	MHHHHHHHMGADLADRFAELERRYDARLGVYVPATGTTAAIEYR ADERFAFCSTFKAPLVAAVLHQNPLTHLDKLITYTSDDIRSISP VAQQHVQTGMTIGQLCDAAIRYSDGTAANLLLADLGGPGGGTAA FTGYLRSLGDTVSRLDAEEPELNRDPPGDERDTTTPHAIALVLQ QLVLGNALPPDKRALLTDWMARNTTGAKRIRAGFPADWKVIDKT GTGDYGRANDIAVVWSPTGVPYVVAVMSDRAGGGYDAEPREALL AEAATCVAGVLAGSGGSGNCL
Desired Product (P)	30749.56	MHHHHHHHMGADLADRFAELERRYDARLGVYVPATGTTAAIEYR ADERFAFCSTFKAPLVAAVLHQNPLTHLDKLITYTSDDIRSISP VAQQHVQTGMTIGQLCDAAIRYSDGTAANLLLADLGGPGGGTAA FTGYLRSLGDTVSRLDAEEPELNRDPPGDERDTTTPHAIALVLQ QLVLGNALPPDKRALLTDWMARNTTGAKRIRAGFPADWKVIDKT GTGDYGRANDIAVVWSPTGVPYVVAVMSDRAGGGYDAEPREALL AEAATCVAGVLAGSGGSGNGLGGZ
Hydrolysed side Product (SP)	30110.76	MHHHHHHHMGADLADRFAELERRYDARLGVYVPATGTTAAIEYR ADERFAFCSTFKAPLVAAVLHQNPLTHLDKLITYTSDDIRSISP VAQQHVQTGMTIGQLCDAAIRYSDGTAANLLLADLGGPGGGTAA FTGYLRSLGDTVSRLDAEEPELNRDPPGDERDTTTPHAIALVLQ QLVLGNALPPDKRALLTDWMARNTTGAKRIRAGFPADWKVIDKT GTGDYGRANDIAVVWSPTGVPYVVAVMSDRAGGGYDAEPREALL AEAATCVAGVLAGSGGSGN

AaLS-13
(C-terminus): 75%

	Calculated MW	Amino acid sequence
Starting Material (SM)	18547.97	MHHHHHHMEIYEGKLTAEGLRFGIVASRFNHALVGRLVEGAIDC IVRHGGREEDITLVCVPGSWEIPVAAGELARKEDIDAVIAIGVL IEGAEPHFDYIASEVSKGLANLSLELRKPISFGDITDDELEEAI ECAGTEHGNKGWEAALSAIEMANLFKSLRLEGSGGSGNCL
Desired Product (P)	18970.47	MHHHHHHMEIYEGKLTAEGLRFGIVASRFNHALVGRLVEGAIDC IVRHGGREEDITLVCVPGSWEIPVAAGELARKEDIDAVIAIGVL IEGAEPHFDYIASEVSKGLANLSLELRKPISFGDITDDELEEAI ECAGTEHGNKGWEAALSAIEMANLFKSLRLEGSGGSGNGLGGZ
Hydrolysed side Product (SP)	18331.67	MHHHHHHMEIYEGKLTAEGLRFGIVASRFNHALVGRLVEGAIDC IVRHGGREEDITLVCVPGSWEIPVAAGELARKEDIDAVIAIGVL IEGAEPHFDYIASEVSKGLANLSLELRKPISFGDITDDELEEAI ECAGTEHGNKGWEAALSAIEMANLFKSLRLEGSGGSGN

Figure 8.6 Bioconjugation of biotin labeled peptide to the C-terminus of AaLS13. Chromatogram and mass spectra of the UPLC-MS analysis. Label peptide sequence: GLGGZ (Z = biotinylated lysine).
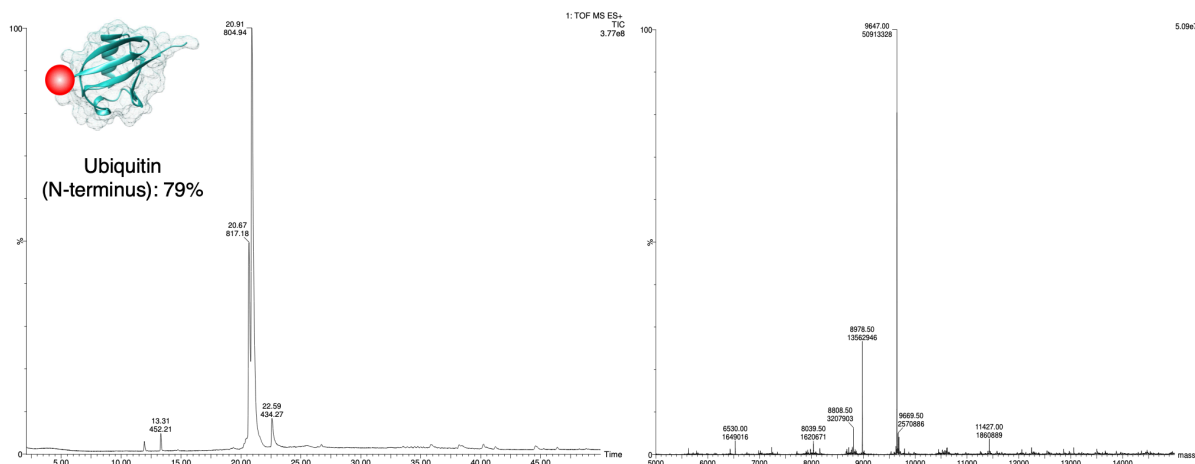
Ubiquitin (C-terminus): 80% (± 0.4)

	Calculated MW	Amino acid sequence
Starting Material (SM)	10160.52	MHHHHHHQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPP DQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGPSGGSG NCL
Desired Product (P)	10583.02	MHHHHHHQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPP DQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGPSGGSG NGLGGZ
Hydrolysed side Product (SP)	9944.22	MHHHHHHQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPP DQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGPSGGSG N

Figure 8.7 Bioconjugation of biotin labeled peptide to the C-terminus of ubiquitin. Chromatogram and mass spectra of the UPLC-MS analysis. Label peptide sequence: GLGGZ (Z = biotinylated lysine).

	Calculated MW	Amino acid sequence
Starting Material (SM)	8979.20	GLSGSGSQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPP DQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG
Desired Product (P)	9647.99	XTRNGLSGSGSQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKE GIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG

Figure 8.8 Bioconjugation of biotin labeled peptide to the N-terminus of ubiquitin. Chromatogram and mass spectra of the UPLC-MS analysis. Label peptide sequence: XTRNCL (X = biotinylated alanine).

Table 8.2 C-terminus labeling of proteins with biotinylated peptide using OaAEP1 in the absence of 2-formyl phenylboronic acid (FPBA)

Protein	GLGGZ (equiv.)	Yield (%)	Hydrolysis (%)
eGFP	1.0	43	4
eGFP	1.2	58	1
eGFP	1.5	64	5
eGFP	1.7	72	2
eGFP	2.0	71	5
Ubiquitin	2.0	60	6
β Lactamase	2.0	56	0

eGFP-NCL (100 µM), GLGGZ (100-200 µM), OaAEP1 (0.25 µM), 2-FPBA (200 µM) in 50 mM MES buffer (pH 6.0), 50 mM NaCl, 1 mM EDTA, at 20 °C

8.5 Preparation of recombinant GFP(+36)-TEVp
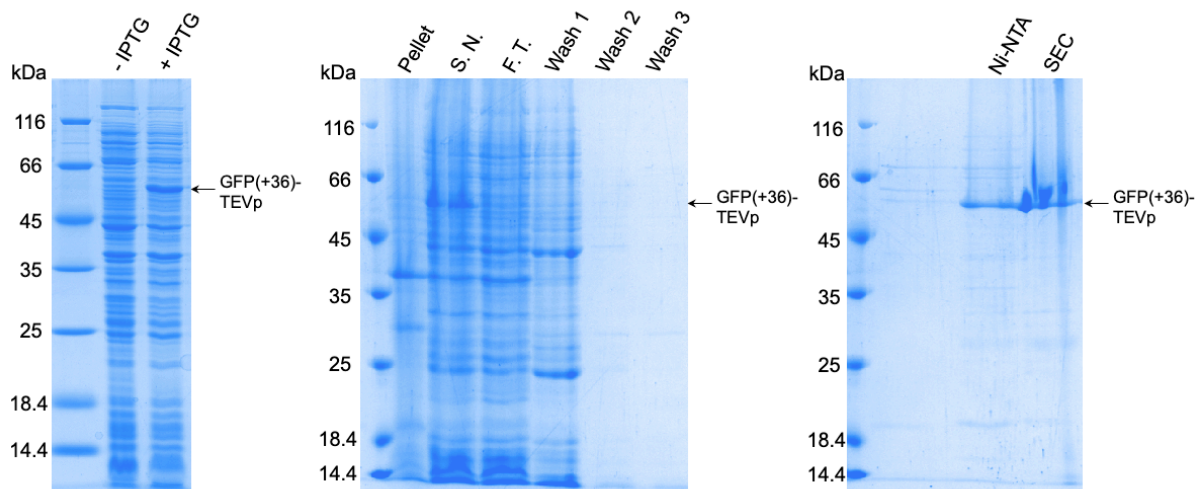
Figure 8.9 SDS-PAGE analysis of fraction obtained during preparation of GFP(+36)-TEVp. Arrows indicate bands corresponding to the protein of interest, GFP(+36)-TEVp. SDS-PAGE lanes: *E. coli* culture before addition of IPTG (- IPTG), *E. coli* culture after incubation with IPTG (+ IPTG), insoluble fraction of lysate (pellet), soluble fraction of lysate (S.), flow through from Ni^{2+}-NTA (F. T.), wash fractions from Ni^{2+}-NTA (wash 1-3), protein elution from Ni^{2+}-NTA (Ni-NTA), protein elution from SEC (SEC).