Investigating the Effects of Cyclisation on a Protein Tool and a Biocatalyst

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

Heather C. Hayes

Supervisor: Dr Louis Y. P. Luk

2022

”Life, uh, finds a way”

- Dr Ian Malcolm, Jurassic Park (1993)

Abstract

As therapeutics and biocatalysts, polypeptides are attractive alternatives to traditional small molecules and catalysts, respectively. However, the applications of peptides and proteins are often limited by their instabilities towards elevated temperatures, organic solvents and/or digestive enzymes. Cyclisation by intramolecular covalent bond formation between the polypeptide chain C- and N-termini, has been demonstrated to improve stability and enhance activity. As a result, cyclisation could provide a convenient route for expanding the scope of peptide and protein application in research and industrial settings. In this thesis, the effects of cyclisation on the stabilities of two biotechnologically relevant proteins were explored.

TET12 is a tetrahedral shaped artificial CCPO protein cage. Two cyclic variants of TET12 were generated using a split intein-mediated approach and a SpyTag/SpyCatcher-N^{TEV} isopeptide bond cyclisation strategy. The thermal, chemical and proteolytic stabilities of the linear and cyclic TET12 variants were compared. Cyclisation was found to be beneficial for enhancing proteolytic resistance but thermal and chemical stability were not improved. Increased aggregation of cyclic TET12 was observed, resulting from a combination of covalent oligomerisation and protein misfolding. These results indicate that connection the polypeptide termini disrupts the TET12 folding pathway and highlights the importance of selecting an appropriate cyclisation technique.

Improvements to the stability and activity of the PET degrading enzyme *Is*PETase, are required if it is to be utilised for large scale recycling applications. SpyTag/SpyCatcher(-N^{TEV}) was employed to cyclise *Is*PETase in a range of cyclic topologies, including a cyclic monomer, cyclic dimer and catenane. The stabilities and activities of the linear and cyclic *Is*PETase variants were examined. In contrast to previous findings for cyclic protein topologies, the cyclic *Is*PETase variants did not display improved thermal stability compared to the wild type enzyme and were found to be more susceptible to trypsin digest. The dimeric variants of *Is*PETase exhibited enhanced PET degradation under optimal conditions and were also found to be more resistant to agitation-induced inactivation. This demonstrates the need for further investigation into the mechanisms behind protein stabilisation using cyclisation.

Acknowledgements

First and foremost I would like to thank Dr Louis Luk for his invaluable advice, guidance and support throughout this PhD. I am extremely grateful for the opportunity to be a part of his research group and for encouraging me to pursue my own research ideas. Of course, all this would not have been possible without funding from the EPSRC CDT in Catalysis.

I would also like to express my deepest gratitude to all the past and present members of the Luk/ Tsai/ Jin group, with whom it has been a pleasure and an honour to work alongside. Their combined sense of humour and wit have made this journey a memorable one, during the good “thymes”, as well as on tougher days. In particular, I am grateful to Simon for taking the time to train me in the ways of chemical biology and for answering my many questions. Thank you to Tom, Alex N. and Patrick for your insightful comments and suggestions, as well as for lifting the big centrifuge rotor. Also to Victoria, Emily, Luke, Davide C., Adriana and Alex L. for some much needed coffee breaks and pub sessions. Muge and Mia, thank you for the delicious bakes. While on the topic of extra-curricular activities, I wish to thank the CDT cohort of 2017/18 and Ellie for all the laughs and adventures upon arrival in Cardiff. In addition, thanks to Davide Z. and Robin for getting me back into climbing and giving me a welcome distraction from the stresses of thesis writing in the final months of this PhD.

A massive thank you to my family for their unwavering support throughout my life. I always said I didn’t want to do “science” but it seems I have spent quite a lot of time doing just that…

Last but by no way least, an extra special thanks to David for always making me smile (or at least trying), even on my grumpiest days. I look forward to finally being able to spend a lot more time together.

Table of Contents

List of Figures and Schemes

List of Tables

List of Abbreviations

AEP	Asparaginyl endopeptidase
AUC	Analytical ultracentrifugation
BHET	Bis(2-hydroxyethyl) terephthalic acid
BLA	β-lactamase
CCPO	Coiled coil protein origami
CD	Circular dichroism
C_m	Concentration at midpoint of protein unfolding
CPY	Carboxypeptidase Y
CuAAC	Copper(I)-catalysed azide-alkyne cycloaddition
DHFR	Dihydrofolate reductase
DIEA	N,N-diisopropylethylamine
DLS	Dynamic light scattering
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DMT	Dimethyl terephthalate
DNA	Deoxyribonucleic acid
DSC	Differential scanning calorimetry
EG	Ethylene glycol
EPL	Expressed protein ligation
Ext^C	C-extein
Ext^N	N-extein
FRET	Förster resonance energy transfer
GCSF	Granulocyte-colony stimulating factor
GdnHCl	Guanidine hydrochloride
GFP	Green fluorescent protein
HE	Homing endonuclease
HPLC	High performance liquid chromatography
Int^C	C-extein
Int^N	N-extein
IPA	Isopropanol

IPTG	Isopropyl-β-D-1-thioglactopyranoside
*Is*PETase	*Ideonella sakaiensis*
KAHA	α-ketoacid-hydroxylamine
k_{cat}	Catalytic turnover
KIH	Knobs into holes packing
K_m	Michaelis constant
LAD	Ligase-activity determinant
MD	Molecular dynamics
MHET	Mono(2-hydroxyethyl) terephthalic acid
MLA	Marker of ligase activity
mRNA	Messenger RNA
MS	Mass spectrometry
NCL	Native chemical ligation
Ni-NTA	Nickel(II)-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
Npu	*Nostoc punctiforme* PCC73102
OaAEP1	*Oldenlandia affinis* asparaginyl endopeptidase 1b
PAL	Peptide asparaginyl ligase
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PET	Polyethylene terephthalate
Pen	Pencillamine
PMT	Post-translational modification
p-NP	Paranitrophenol
p-NPA	Paranitrophenyl acetate
PTS	Protein *trans*-splicing
QM/MM	Quantum mechanics/molecular mechanics
RNA	Ribonucleic acid
SDM	Site directed mutagenesis
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide electrophoresis

SEC	Size exclusion chromatography
SH3	N-terminal Src homology 3 domain
SICLOPPS	Split intein circular ligation of proteins and peptides
SPAAC	Strain-promoted azide-alkyne cycloaddition
SPPS	Solid phase peptide synthesis
SrtA	Sortase A (from *Staphylococcus aureus)*
Ssp	*Synechocystis* sp. PCC6803
TET12-Int	TET12SN CCPO protein cage cyclised by split intein
TET12SN	Supernegatively charged tetrahedral CCPO protein cage
TET12-Spy	TET12SN CCPO protein cage cyclised by SpyTag/SpyCatcher
TEV	Tobacco etch virus
T_g	Glass transition temperature
T_m	Melting temperature
TPA	Terephthalic acid
tRNA	Transfer RNA
UV	Ultraviolet
V_{max}	Maximum velocity
WT	Wild type

CHAPTER 1:
INTRODUCTION

1.1 BACKGROUND

Polypeptides play vital roles for the correct functioning of cellular life. They are involved in a diverse array of processes including catalysis, signalling and transport, as well as making up structural features such as membranes. Peptides and proteins are mostly linear polymers composed of amino acids connected by amide bonds. Generally, peptides are defined as shorter chains of up to 50 amino acids, while proteins have more than 50 residues. Often, these linear sequences of amino acids fold into more complex three-dimensional structures which determine the function of a particular peptide or protein.[1] Even a small change to the structure, such as the mutation of a single amino acid, can cause major disruption to function, with potentially detrimental consequences.[2]

The Central Dogma of molecular biology describes the production of polypeptides from genetic information, which takes place via two main steps known as transcription and translation (**Figure 1.1**).[3] During transcription, a particular sequence of double stranded DNA is copied into a single strand of messenger RNA (mRNA) by RNA polymerase. The mRNA is then transported from the nucleus into the cytoplasm. There, with the help of a ribosome and transfer RNA (tRNA), it is used to assemble the corresponding polypeptide chain, in the process known as translation.

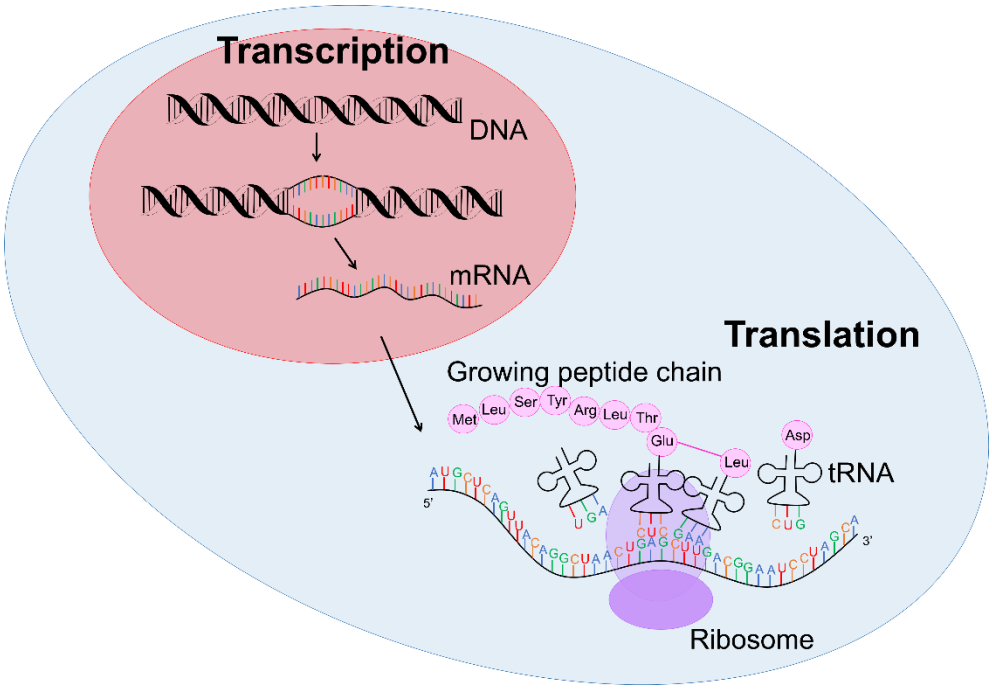
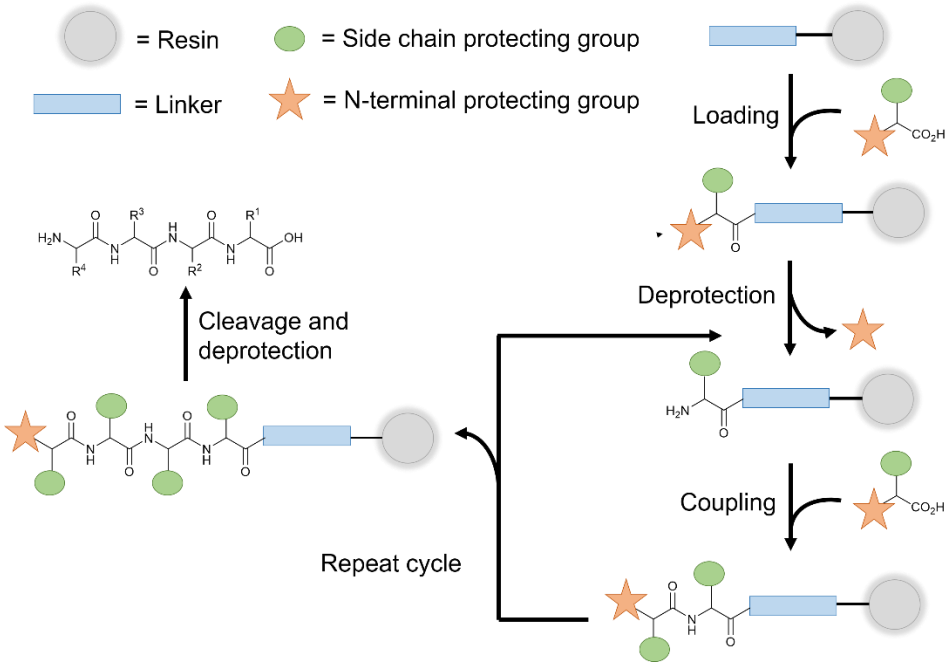


Figure 1.1 The Central Dogma of molecular biology. DNA is transcribed into mRNA, which is then translated into a polypeptide by a ribosome and tRNA.

Polypeptide chains can also be synthesised chemically, by a process known as solid-phase peptide synthesis (SPPS), in which a solid support is used to assemble a peptide chain, by the step-wise amide bond forming reaction of the carboxyl group of one amino acid with the amine group of another (**Scheme 1.2**). It is an effective method for the chemical preparation of polypeptides of up to 40-50 amino acids in length, and can facilitate the incorporation of additional functionalities such as unnatural amino acids or fluorescence tags, which can be challenging by recombinant methods.[4] However, the chemical synthesis of longer polypeptides is often impeded by a low overall yield and purity as a result of the aggregation of long peptide chains leading to truncation, side product formation and epimerisation.



Scheme 1.2 SPPS, where a peptide is chemically assembled by the step-wise attachment of protected amino acids using repeating deprotection and coupling steps before cleavage from the solid support.

Proteins which catalyse reactions in living organisms are known as enzymes. They have long been employed to carry out a wide variety of bioconversions in both research and industrial areas including food production, textiles, agriculture, cosmetics and health, among others.[5] With a growing global population, resulting in a continuously rising demand for energy and resources, the development of alternative technologies and sustainable practices is vital, both from an environmental and economic point of view. As nature's catalysts, enzymes have a key part to play. However, for many applications, enzymes are currently unsuitable due to their limited stability under harsh reaction conditions. Therefore, enhancing enzyme stability e.g. at elevated temperatures, extreme pH or in non-aqueous solvent, would expand the scope of enzyme application.[6,7] In addition, peptide and protein

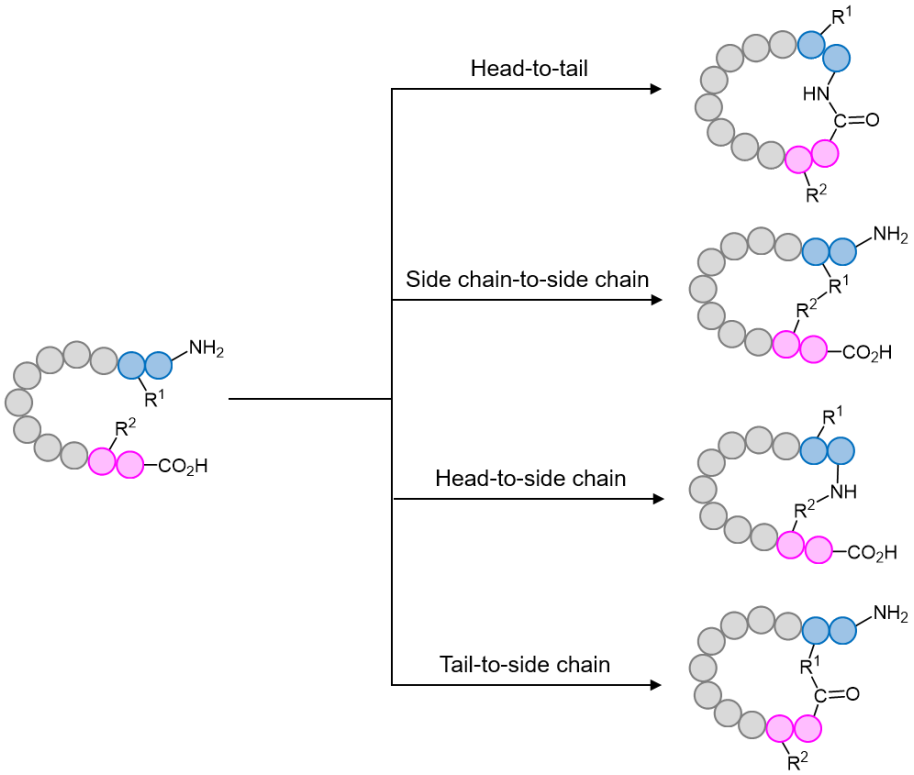
therapeutics, which are becoming increasingly popular due to their high activity, specificity and low toxicity compared to traditional small molecule drugs, could benefit from stabilisation, with the potential to overcome problems such as poor oral bioavailability and enzymatic hydrolysis.[8]

In nature, many polypeptides have been found to be cyclised, whereby the N- and C-termini are connected by a covalent bond.[9] Cyclisation has been shown to confer both peptides and proteins with a number of favourable properties, including stabilisation. For proteins, this increased stability compared to their linear counterparts, originates from the destabilisation of the unfolded state.[10] Upon cyclisation, the number of conformational states accessible to the unfolded polypeptide chain is reduced. Consequently, the entropy is decreased and Gibbs free energy of the unfolded state is increased, assuming there is no offsetting enthalpic cost due to the introduction of strain to the system. Cyclisation has successfully been employed to improve the stability of several industrially relevant enzymes for application outside of their native conditions.[11]

For peptide therapeutics, cyclisation can also lead to improved stability as well as biological activity, enabling enhanced binding of the target molecule due to the reduced conformational flexibility and decreased change in entropy upon binding of the cyclic peptide to the target.[12] In addition, cyclisation can result in improved proteolytic resistance and membrane permeability.[12,13]

1.2 CYCLISATION METHODS

To cyclise a protein or peptide, a covalent bond must be formed between the N- and C-terminal amino acid residues (**Scheme 1.3**). This can be done by directly linking the respective terminal amine and carboxylic acid groups (known as head-to-tail or backbone cyclisation) or by crosslinking side chain groups leaving the termini free (side chain-to-side chain cyclisation). Alternatively, cyclisation can be achieved by connecting either the N- or C- termini with a side chain (head-to-side chain or tail-to-side chain cyclisation, respectively). Over the years, a large variety of methods have been developed for polypeptide ligation, whereby a bond is formed between the functionalities of two separate polypeptide chains. These can be applied in an intramolecular manner to the peptide or protein of interest to result in cyclisation.[14] In the following sections, some of the more commonly encountered cyclisation techniques will be discussed. These are categorised as either chemical, enzymatic or protein tag methods.



Scheme 1.3 Modes of polypeptide cyclisation.

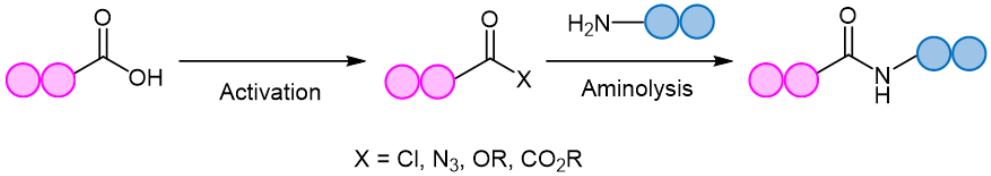
1.2.1 Chemical Cyclisation Methods

1.2.1.1 Direct Amide Bond Formation

The ideal method of amide bond formation would involve the direct condensation of a carboxylic acid with an amine, producing one equivalent of water as the only by-product. However, in reality, this requires the use of either high temperatures or microwave irradiation to avoid the formation of a stable ammonium carboxylate salt.[15,16] Consequently, such harsh conditions are incompatible with amino acid substrates and therefore unsuitable for peptide ligation.

To promote amide bond formation under milder conditions, activation of the carboxylic acid group is necessary (**Scheme 1.4**). This can be achieved by transforming the carboxylic acid –OH into a better leaving group, for example an acyl halide, acyl azide, anhydride or an activated ester, through the use of coupling reagents.[17,18] Consequently, nucleophilic (e.g. Lys, Ser, Thr) and carboxylate (e.g. Asp, Glu) amino acid side chains must be protected to prevent side reactions. As a result, this approach is better suited to peptides that are synthesised in a fully protected form. In addition, carboxylic acid activation often requires high dilution to avoid oligomerisation from intermolecular reaction, as well as the use of additives to enhance the rate of reaction and suppress racemisation at the ligation site.[19] For example, in the cyclisation step of the total synthesis of the antibiotic teixobactin, a

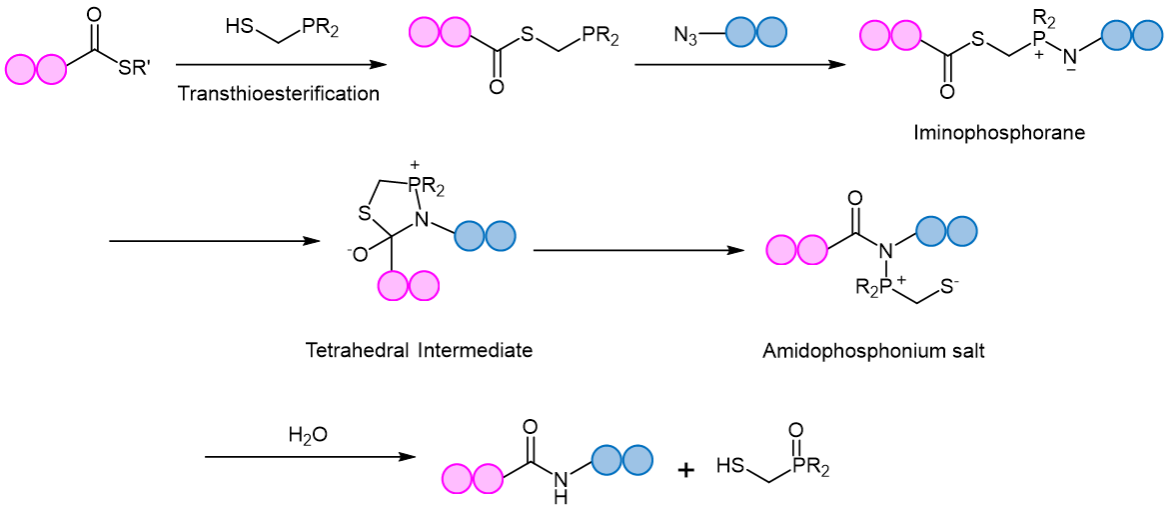
combination of coupling reagents (HOAt/OxymaPure/HATU) with a tertiary amine base (DIEA) were used.[20] Nevertheless, alternative amide bond forming cyclisation techniques are often preferred.



Scheme 1.4 Activation of a carboxylic acid group, before amide bond formation with the amine group.

1.2.1.2 Staudinger Ligation

The Staudinger ligation is based on the Staudinger reaction, which involves the reduction of an azide into an amine using a phosphine.[21] For application as a chemoselective peptide ligation strategy, the Staudinger reaction was modified through the use of a bifunctional phosphinothiol reagent (**Scheme 1.5**).[22] In the first step, the phosphinothiol thiol group undergoes a transthioesterification reaction with a peptide thioester. The phosphino group of the resulting C-terminal phosphinothioester then reacts with a second peptide, bearing an N-terminal azide, to produce a reactive iminophosphorane intermediate. Upon attack of the iminophosphorane nitrogen on the thioester, a tetrahedral intermediate is generated, which collapses into an amidophosphonium salt. This is finally hydrolysed to produce a native amide bond between the peptide fragments and a phosphine oxide side product is released. As no atoms from the phosphinothiol remain in the amide product, it can be considered a traceless ligation. In addition, the reaction can be performed using unprotected peptide fragments as the reaction is chemoselective towards the azide.[23]



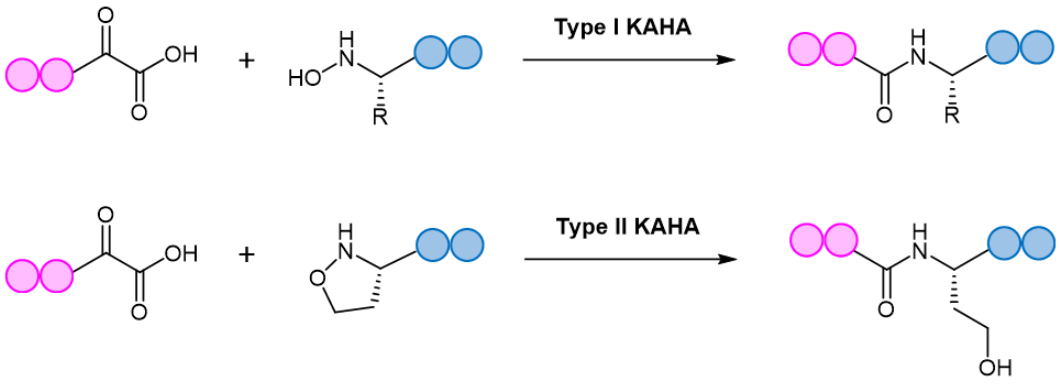
Scheme 1.5 Traceless Staudinger ligation.[23]

The Staudinger ligation was applied in an intramolecular manner for the cyclisation of a synthetic 11 residue model peptide. The cyclisation reaction was initiated by the addition of excess base for the neutralisation of TFA remaining from the peptide deprotection step. The cyclised peptide product was obtained in a moderate yield of around 30%.[24]

Unfortunately, the application of the Staudinger ligation in aqueous solutions is limited due to the low solubility of phosphinothiols and corresponding phosphinothioesters. Although water soluble phosphinothiol reagents have been developed, their use is still limited due to the laborious preparation required.[25] The widespread application of the Staudinger ligation has further been hindered by the requirement for substrates with glycine residues at the ligation site. Due to increased steric strain on the tetrahedral intermediate in presence of bulkier residues, a covalent bond between the oxygen and the oxophilic phosphorous atom is favoured over thiol displacement. However, by increasing the electron density on the phosphorous atom of the phoshinothiol reagent, P-O bond formation is discouraged and an improved yield can be achieved for non-glycyl Staudinger ligation reactions.[26]

1.2.1.3 KAHA Ligation

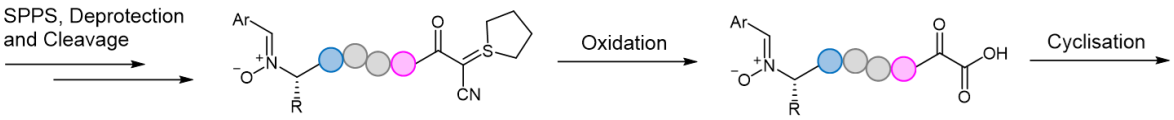
An alternative method for peptide ligation via amide bond formation is α-ketoacid-hydroxylamine (KAHA) ligation.[27] The chemoselective ligation takes place between an N-terminal hydroxylamine and a C-terminal α-ketoacid. Depending on the nature of the hydroxylamine substituent, there are two main mechanistic pathways (**Scheme 1.6**). Type I KAHA uses a free hydroxylamine, while type II KAHA involves the use of an O-substituted hydroxylamine.



Scheme 1.6 Type I KAHA with an N-terminal hydroxylamine. Type II KAHA with a 5-oxaproline.

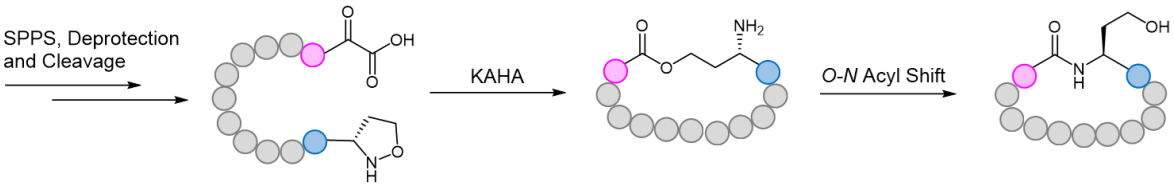
The type I KAHA ligation approach has been used to cyclise a range cyclic peptide natural products (including Gramicidin S, Tyrocidine A, Hymenamide B and Stylostatin A) of varying ring sizes (5-10 residues).[28] The linear precursors were synthesised using SPPS during

which the α-ketoacid and hydroxylamine functionalities were introduced in protected forms as a sulfur ylide and a nitrone, respectively (**Scheme 1.7**). After cleavage from the resin, the sulfur ylide was oxidised and cyclisation proceeded in DMF/H_2O (50:1) in the presence of oxalic acid at 40 °C. Some epimerisation of the α-ketoacid was observed during the oxidation step, but could be separated by HPLC. The purified cyclic products were obtained in an overall yield of < 22% after 48 hours of reaction.



Scheme 1.7 Type I KAHA peptide cyclisation, showing incorporation of the α-ketoacid as a sulfur ylide and the hydroxylamine as a nitrone.

Although demonstrated to be a feasible method for the cyclisation of medium length peptides, type I KAHA is rarely used, mainly due to the instability of the free hydroxylamine in aqueous media, with oxidation often observed. Instead, the type II KAHA ligation method is generally preferred. It uses a water stable O-substituted hydroxylamine, most commonly 5-oxaproline, which is easily prepared and incorporated into the peptide of interest during SPPS. After cleavage from the resin, the α-ketoacid and 5-oxaproline cyclise directly producing a depsipeptide intermediate. Under basic conditions, *O-N* acyl shift then takes place to give the desired cyclic peptide product, with a homoserine residue generated at the ligation site (**Scheme 1.8**).[29]



Scheme 1.8 Type II KAHA peptide cyclisation.

The type II KAHA approach has been used to cyclise a variety of shorter peptides (8 to 20 residues) without requiring the use of protecting groups or coupling reagents.[30] Furthermore, the approach was shown to be compatible with larger and more challenging substrates, as demonstrated by the chemical synthesis of the cyclic antibacterial protein AS-28 (70 amino acid residues).[31] The synthetic protein was found to display similar activities to those of native AS-28 (isolated from *Enterococcus faecalis*) and proved that type II KAHA is a well-suited method for the assembly of hydrophobic polypeptides of varying sizes.

However, even with the introduction of 5-oxaproline, type II KAHA ligation still suffers from a number of drawbacks. In many cases, these include the relatively laborious synthesis of precursors, slow reaction rates, epimerisation at the C-terminal residue and the incorporation of a non-native homoserine residue during ligation.

1.2.1.4 Azide-Alkyne Cycloaddition

1,4-disubstituted triazoles are known to effectively mimic the topology and electronic properties of native *trans*-amide bonds.[32] Although triazoles and amide bonds differ structurally, with an extra atom in the triazole backbone and consequently increased distance between R^1 and R^2 groups, increased dipole moment of the triazole allows N(2) and N(3) to act as hydrogen acceptors. Meanwhile, C(5) can function as a hydrogen donor (like the amide proton).[33]

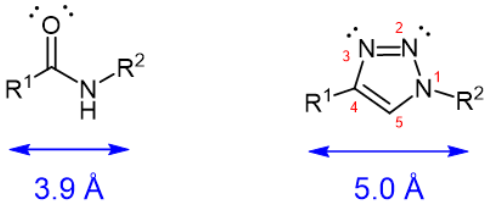
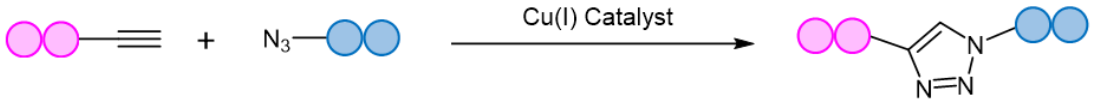


Figure 1.9 Topological and electronic similarities of amides and 1,4-disubstituted triazoles.[33]

In the presence of a copper(I) catalyst, a 1,4-disubstituted 1,2,3-triazole is produced by the 1,3-dipolar cycloaddition reaction of an azide with a terminal alkyne (**Scheme 1.10**).[34,35] This commonly used “click”-type reaction is regioselective, efficient and takes place under mild conditions in a variety of solvents, including water. Therefore it is suitable for peptide ligation.

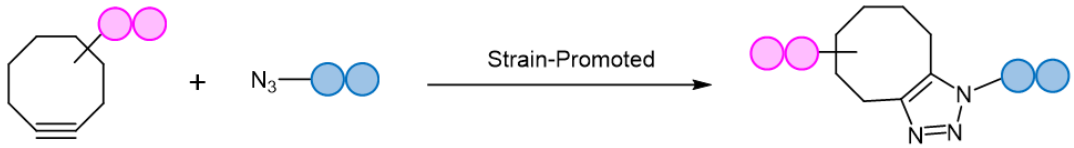


Scheme 1.10 Copper(I) catalysed azide-alkyne cycloadditon.

The copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) has been used for the synthesis of a series of cyclotetrapeptide tyrosinase inhibitor analogues in an efficient and high yielding manner.[32] Previous attempts to synthetically cyclise the Pro-Tyr-Pro-Val peptide had been unsuccessful. Moreover, the inhibitory activity of the triazole containing cyclic products was found to be comparable to that of the natural cyclotetrapeptide isolated from *Lactobacillus helveticus*.

While CuAAC is a seemingly straightforward reaction, the requirement for Cu(I) as the active catalytic species is problematic. In solution, Cu(I) is rapidly oxidised to the more stable Cu(II). Instead, Cu(I) is often generated in situ using Cu(II) and excess reducing agent. This can result in the oxidation of His and Arg residues leading to the requirement for Cu-stabilising ligands to minimise degradation and accelerate the CuAAC reaction. It should also be noted that Cu is toxic to cells, thus limiting in vivo application.[36]

The strain promoted azide-alkyne cycloaddition (SPAAC) was developed to circumvent the need for a copper catalyst and its concomitant complexities. Through the use of a highly strained cyclic alkyne (typically cyclooctyne), cycloaddition with an azide can proceed rapidly in the absence of a catalyst, under physiological conditions (**Scheme 1.11**).[37] However, SPAAC lacks the regiospecificity of CuAAC, and forms a mixture of 1,4-disubstituted products. Furthermore, cyclooctyne reagents are considerably more expensive than their terminal alkyne counterparts, although the development of new synthetic routes are slowly making SPAAC reagents more accessible.[38]



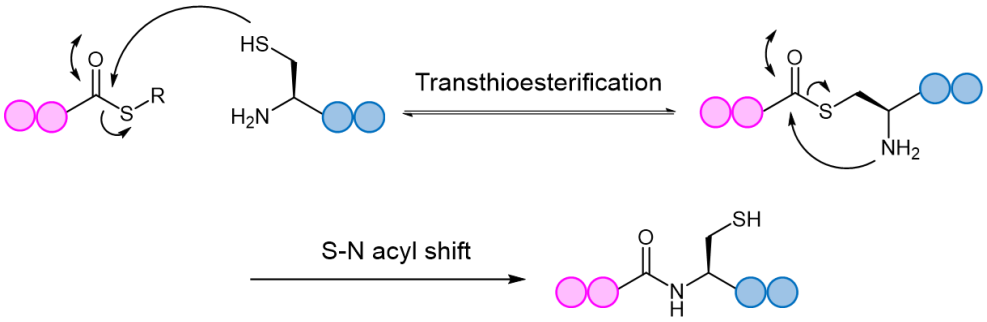
Scheme 1.11 Strain-promoted azide-alkyne cycloaddition.

1.2.1.5 Native Chemical Ligation

The synthesis of longer peptides and proteins by SPPS is low yielding and often impractical. Therefore polypeptide ligation technologies have been developed to enable shorter peptide segments to be connected together to generate larger polypeptide chains. In particular, native chemical ligation (NCL) has been used to synthesise hundreds of protein targets to date, and has played an important role in the advancement of peptide chemistry.[39,40]

The NCL reaction occurs between a peptide bearing a C-terminal thioester and another bearing an N-terminal cysteine (**Scheme 1.12**). In the first step, the nucleophilic thiol group of the cysteine side chain attacks the thioester peptide in a reversible transthioesterification reaction. In the second step, the resulting thioester intermediate undergoes an S-to-N acyl shift to generate the peptide bond and restores the thiol side chain. The reaction proceeds in aqueous conditions at neutral pH. Chaotropic reagents (e.g. guanidine hydrochloride) are used to prevent secondary structure formation and aggregation so that the reaction can be carried out at high concentrations. Reducing agents can also be employed to prevent disulfide bond formation. The reaction is both regio- and chemoselective. Internal cysteine

residues do not have the 1,2-aminothiol group required for ligation. Meanwhile, the side chains of other nucleophilic residues have poor nucleophilicity at pH 7 so will not compete.[40]



Scheme 1.12 NCL mechanism, showing transthioesterification and S-N acyl shift steps, resulting in the formation of an amide bond between C-terminal thioester and N-terminal Cys peptides.

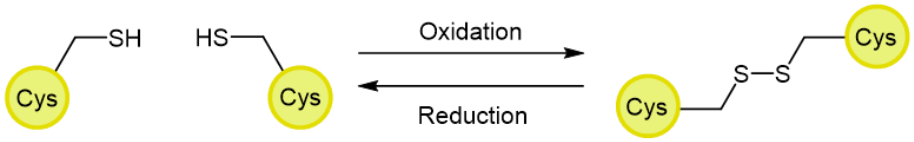
In recent years, a variety of extensions to the NCL methodology have been devised to expand the scope of its application.[41] One of the most commonly employed is post-ligation desulfurisation, whereby cysteine is converted to alanine using a free-radical and metal-free cysteine reduction method.[42] The removal of the cysteine and replacement by a much more abundant amino acid is advantageous, as the desired polypeptide of interest can be generated with no trace of ligation or mutation. The development of other thiol-derived canonical amino acids has further increased the number of sites possible for the ligation reaction to take place.[43]

As well as intermolecular ligation between two separate peptides, intramolecular NCL can also be performed, resulting in the formation of a cyclic peptide. For a range peptides between 5-16 residues in length, cyclisation was observed to take place in a high yield with no oligomer formation through intermolecular reaction, even at millimolar concentrations.[44] Nevertheless, the cyclisation of large proteins by this method is still cumbersome. This is primarily due to the need for their recombinant expression and therefore lack of C-terminal thioester.

1.2.1.6 Disulfide Bond Formation

All the aforementioned chemical methods of polypeptide cyclisation rely on one or more non-native functional groups, which require incorporation into the sequence of interest during synthesis. This usually limits the application of these techniques to shorter peptides accessed through SPPS, though non-native functionalities may be introduced into recombinantly produced proteins using genetic code expansion.[45] On the other hand, amino acid side chains provide access to a range of chemical functionalities, which can be exploited irrespective of polypeptide size and synthetic origin.

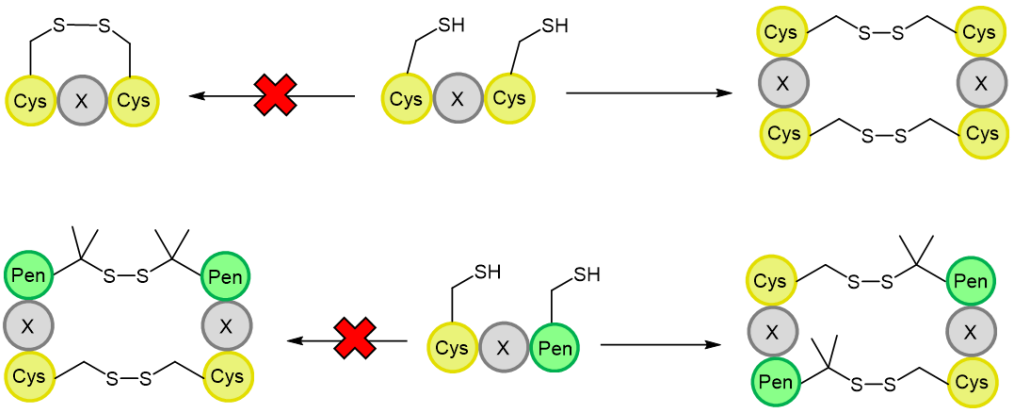
The thiol group of cysteine residues is one of the most reactive amino acid functional groups due to the intrinsic nucleophilicity of the thiolate anion under physiological conditions.[46] The formation of disulfide bonds between cysteine residues plays an important role in the folding, stability and activity of many proteins (**Scheme 1.13**).[47] For example, cyclotides are family of naturally occurring cyclic peptides which contain a characteristic cyclic cysteine knot motif consisting of three disulfide bridges. This unique and topologically complex protein fold confers high stability against thermal, chemical and enzymatic degradation.[48,49]



Scheme 1.13 Disulfide bond formation between two cysteine residues.

Cyclisation can be achieved by direct disulfide bond formation between two cysteine residues, and was used to design cyclic peptides to inhibit the interaction between coactivators and steroid receptors for application in anticancer drugs.[50] Alternatively, a linker molecule of the appropriate size and bearing thiol reactive groups can be employed. Examples of such reagents include benzylbromides, allyhalides, haloacteamides, haloacteones and bromomaleimides, many of which are commercially available and suitable for use in aqueous buffers under mild conditions.[51] For example, a 1,2-bis(bromomethyl)benzene crosslinker was found to be an appropriate choice for the cyclisation of the hormone hepcidin after the screening of multiple linkers.[52]

However, problems can arise when multiple cysteine residues are present, as the number of possible disulfide bond patterns is increased. In cells, enzymes are produced to ensure the correct disulfide bond pattern is generated.[53] Nevertheless, in vitro this can be more difficult to control, though regioselective disulfide bond formation can be encouraged through the use of cysteine protecting groups.[54] Alternatively, the number of possible disulfide patterns can be minimised by using orthogonal disulfide pairing, which relies on the use of a unique sequence of cysteine residues or non-native side chain functionalities (**Scheme 1.14**). For instance, two cysteine residues positioned either side of a single amino acid (i.e. C-X-C, where X is any amino acid), will not form a disulfide bond. Instead the formation of two disulfide bonds between two C-X-C sequences is preferred.[55] In addition, when a pencillamine (Pen) is present (i.e. C-X-Pen), the formation of a mixed disulfide bond between the Cys and Pen is favoured.[56] This is due to the slower rate of thiol-disulfide exchange of Cys-Pen disulfides compared to Cys-Cys disulfides, resulting from steric hindrance imposed by the additional methyl groups adjacent to the Pen thiol group.



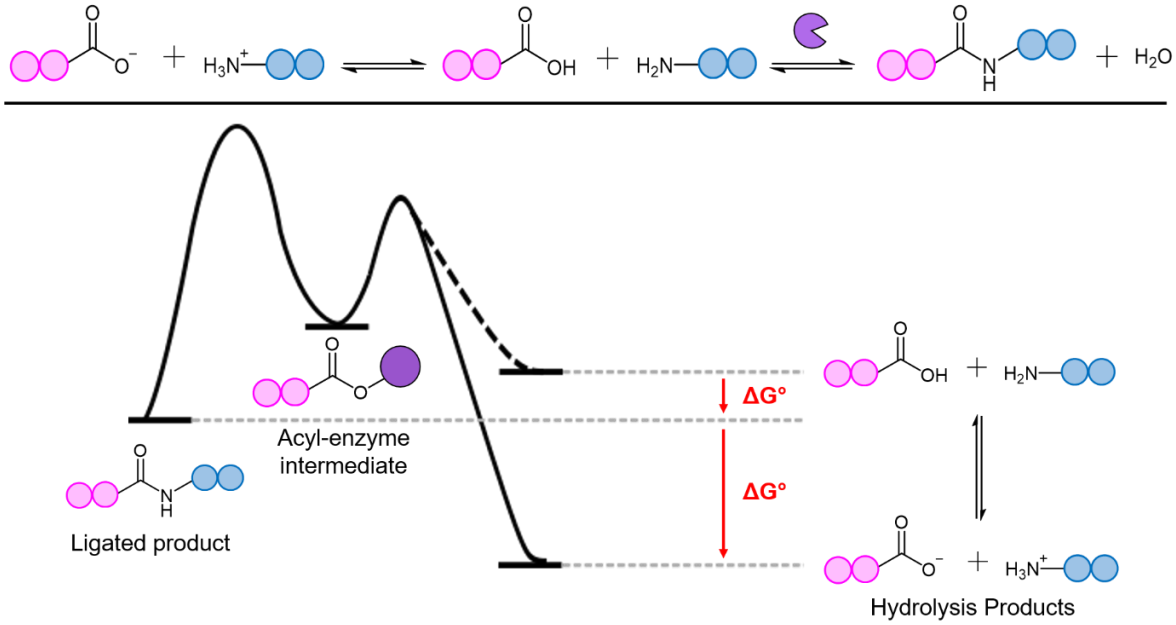
Scheme 1.14 Orthogonal disulfide pairing strategies.

1.2.2 Enzymatic Cyclisation Methods

1.2.2.1 Ligation versus Hydrolysis

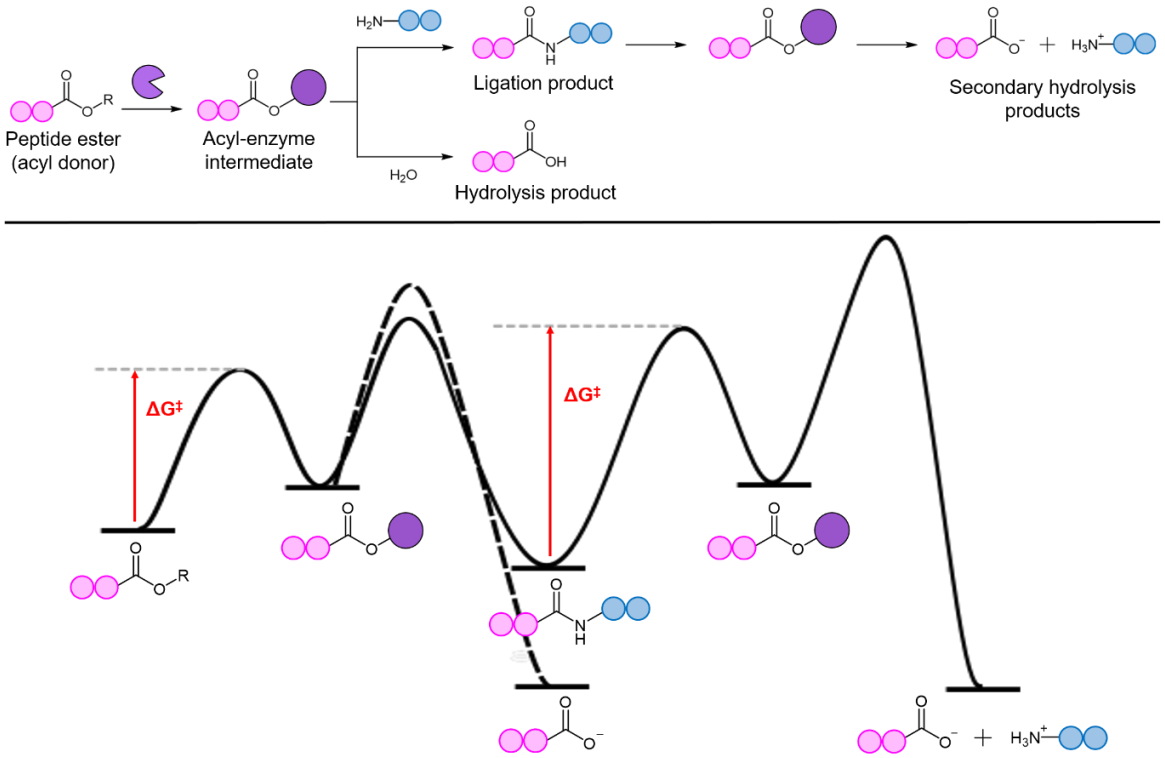
Enzymes are known to catalyse reactions efficiently and selectively under mild conditions. However, only a handful of enzymes have been identified that naturally catalyse peptide bond formation. On the other hand, proteases, enzymes that catalyse the hydrolysis of peptide bonds, are abundant and often well characterised. As such, there has long been interest in using proteases to catalyse the reverse reaction, i.e. peptide bond formation. This can be manipulated either thermodynamically or kinetically.[57]

As enzymes, proteases obey the principle of microscopic reversibility, whereby forward and reverse reactions (i.e. peptide bond hydrolysis and formation) proceed via the same mechanism and through the same reaction intermediates. At equilibrium these reactions will occur at the same rate. The protease only alters the rate at which the equilibrium is attained but not the position. In order to increase the yield of the desired ligation product, the equilibrium must be shifted to favour peptide bond formation (i.e. thermodynamic control) (**Scheme 1.15**).[58,59] Under physiological conditions, the hydrolysis products are more thermodynamically stable and so equilibrium strongly favours hydrolysis. It is the neutral forms of the carboxylic acid and amine substrates that undergo the ligation reaction.[60] Therefore, the ligation/hydrolysis equilibrium can be manipulated by controlling the equilibrium of ionisation. This is most easily done by directly altering the pH of the reaction mixture.[61] Alternatively, the use of an organic co-solvent reduces the acidity of the carboxylic acid group of the acyl donor, increasing the concentration of uncharged reactive substrate and shifting the equilibrium towards peptide bond formation.[62,63] However, not all enzymes are compatible with high concentrations of organic solvent and often exhibit reduced stability and activity when placed in a non-aqueous environment. In addition, peptide substrates and products may be poorly soluble.[64]



Scheme 1.15 Protease (shown in purple) catalysed peptide bond formation under thermodynamic control. Ligation and hydrolysis product yields depend on their relative thermodynamic stabilities (ΔG°).[65]

On the other hand, by applying kinetic control (**Scheme 1.16**), the yield of products generated can be determined by the kinetics of the enzyme, instead of by the thermodynamic stabilities of the substrates and products. This approach requires a (thio)ester as an acyl donor and so requires an enzyme that can form an acyl-enzyme intermediate (i.e. a serine or cysteine protease).[66] Although the hydrolysis of amides and (thio)esters proceed through a similar mechanism, (thio)esters are intrinsically more reactive than amides. As such, the more activated (thio)ester substrate will kinetically outcompete the ligated peptide product (an amide) to form the acyl-enzyme intermediate. This results in reduced hydrolysis and an increased yield of ligation product.[67] However, unwanted hydrolysis of the acyl-enzyme intermediate can still occur, limiting the reaction yield by consuming the substrate.

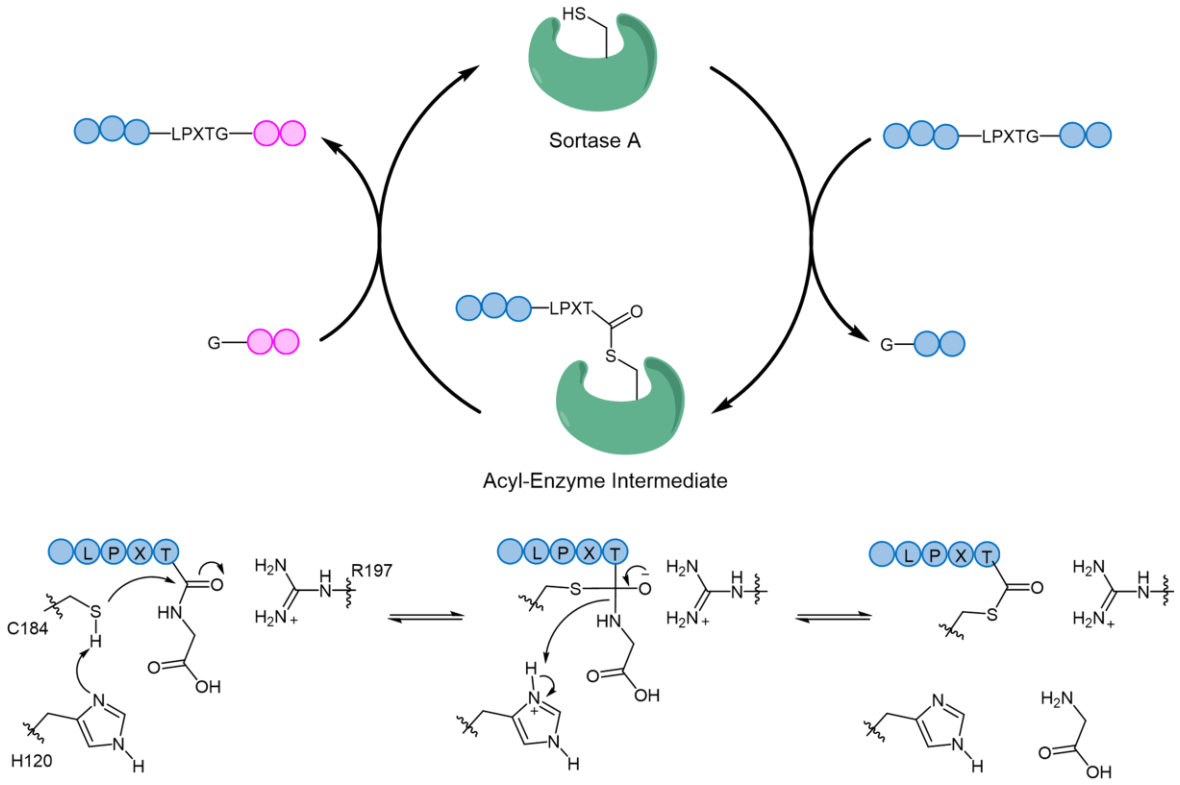


Scheme 1.16 Protease (shown in purple) catalysed peptide bond formation under kinetic control. Relative yield of ligation and hydrolysis products depends on the energy barrier (ΔG‡) for the formation of the acyl-enzyme intermediate, for which peptide ester substrate and ligation product compete. The ligation product can also undergo secondary hydrolysis.[65]

1.2.2.2 Sortase

Sortases are cysteine transpeptidases, which catalyse the transacylation of amines. First discovered in *Staphylococcus aureus* (*S. aureus*), sortases have since been identified in other bacterial species, the majority of which are Gram-positive.[68] Most of these bacterial species contain multiple sortase enzymes (named A – F). Each performs a different role, from the attachment of proteins to the peptidoglycan layer in bacterial cell walls to the assembly of pili.[69]

Sortases recognise a so-called sorting motif. For sortase A from *S. aureus* (SrtA), this was identified as LPXTG, where X is any amino acid, and the only residue in the motif where changes are tolerated.[70,71] The SrtA catalysed reaction has been well-studied, and is currently accepted to occur via a reverse protonation mechanism, catalysed by the active site Cys184, Arg197 and His120 residues (**Scheme 1.17**).[72,73]



Scheme 1.17 (Top) SrtA catalytic cycle. (Bottom) Reverse protonation mechanism of SrtA catalysed ligation involving the catalytic triad Cys184, His120 and Arg197.

SrtA is commonly employed as a means of conjugating synthetic peptides to a larger recombinant protein, a technique that has come to be known as sortase-mediated ligation or sortagging.[74] SrtA has also been shown to be effective for peptide cyclisation, when the sorting motif and the Gly acceptor sequence (preferably two or more Gly) are placed at the C- and N-termini of the same peptide chain, respectively.[75] The efficiency of the cyclisation reaction has been found to depend on both the size and concentration of substrate. [76] It was determined that a substrate of a minimum length of 19 residues (including the sorting motif) was required for cyclisation to take place in preference to intermolecular reaction leading to multimeric products (i.e. dimers and trimers, both linear and cyclised). With regard to concentration, as concentration increased, so did the yield of dimer and trimer products due to increased intermolecular reaction.[76]

SrtA has also been used to cyclise larger recombinant proteins including various cytokines, enhanced green fluorescence protein (GFP) and ubiquitin C-terminal hydrolase L3, demonstrating the versatility of SrtA as a cyclisation tool.[77,78] However, the use of SrtA is not without its drawbacks. Firstly, the ligation reaction is inherently reversible due to the sorting motif remaining in the product. A variety of methods have been developed to circumvent this problem, for example the removal of small glycyl leaving groups using dialysis, deactivation of glycyl leaving groups by diketopiperazine formation, or the use of a

flow-based system.[79–81] Secondly, high enzyme concentrations or long reaction times are often required due to the low catalytic efficiency of wild type SrtA. Using directed protein evolution and high throughput screening approaches, variants with significantly improved catalytic efficiencies have been identified.[82,83] Thirdly, SrtA dependence on Ca^{2+} limits its use in living cells where Ca^{2+} levels are highly regulated. However, mutation of calcium binding residues to match those found in calcium-independent sortases allows SrtA to function in the absence of Ca^{2+}.[84] While StrA remains a valuable tool for protein modification, the use of alternative enzymes could perhaps be considered to be superior for cyclisation reactions with regard to catalytic efficiency and promiscuity.

1.2.2.3 Subtilisin-Derived Ligases

Subtiligase is a double mutant of the serine protease subtilisin BPN' from *Bacillus amyloliquefaciens*.[85] To promote ligase activity and repress hydrolysis, the active site serine was replaced with a cysteine residue (S221C), resulting in a mutant known as thiolsubtilisin. This had two main effects. Firstly, reduced hydrolysis due to the increased reactivity of the thioester acyl-enzyme intermediate towards amine substrates, compared to the ester acyl-enzyme intermediate. Secondly, decreased secondary hydrolysis of the ligated peptide product due to the lower amidase activity of the thiolsubtilisin enzyme.[86] However, the S122C mutation caused steric crowding in the active site leading to reduced acyl-enzyme intermediate formation compared to the wild type subtilisin. Therefore, a second mutation (P225A) was introduced to better accommodate the C-S bond in the active site. The resulting enzyme (i.e. subtiligase) was demonstrated to have improved ligase activity compared to both thiolsubtilisin and the wild type subtilisin.[87] The reaction is chemoselective for N-terminal amines, does not react with lysine ε-amines, can be applied to unprotected peptides/proteins and used under mild, aqueous conditions.[65] Subtiligase has also been used to cyclise peptide esters ranging in length between 12 and 31 residues in a high yield (<88%).[88] Furthermore, subtiligase recognises a broad range of substrates sequences. This is illustrated in **Figure 1.18**.[89,90]

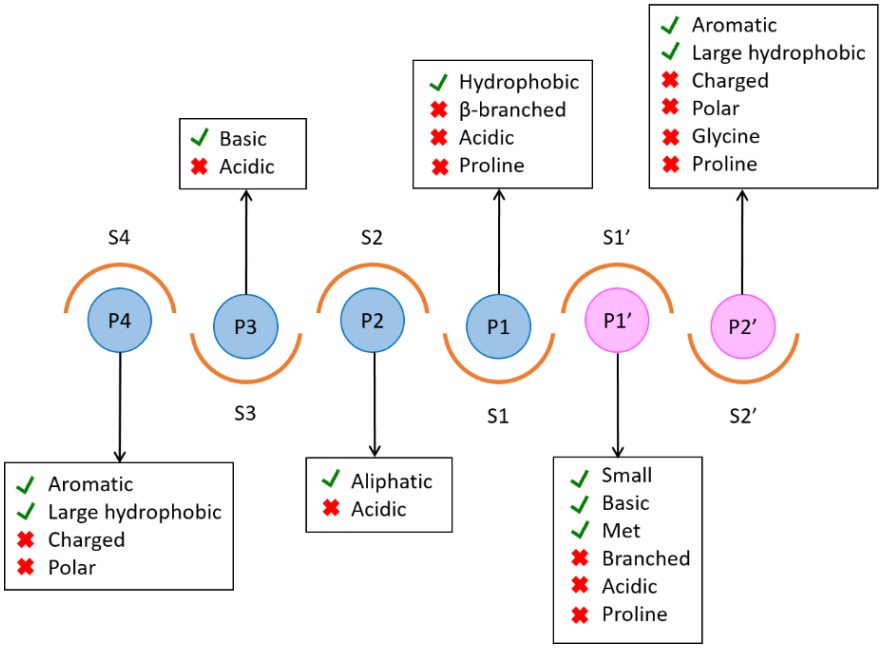


Figure 1.18 Subtiligase substrate specificity, where the P4-P2' positions of the substrate interact with the enzyme's substrate binding cleft (S4-S2').[65]

A major drawback to the use of subtiligase is its dependence on calcium for stabilisation. Therefore, a cation independent variant was created by deleting the calcium binding domain of subtiligase and incorporating an additional 18 stabilising mutations, including a disulfide bridge. The enzyme was named peptiligase.[91] It was shown to efficiently catalyse the ligation of peptides in a high yield with minimal hydrolysis under aqueous conditions, could tolerate the presence of a variety of additives and cosolvents, and catalyse the cyclisation of various peptides. Further modifications to peptiligase generated omniligase-1.[92] This variant is commercially available and has an even broader substrate scope, allowing traceless ligation. As well as cyclising peptides of 14 or more amino acids in a high yield (>90% yield), ominiligase-1 was also demonstrated to cyclise peptides containing polyethylene glycol (PEG) spacers, D-amino acids and isopeptide bonds.[93]

The broad substrate scope of these engineered peptide ligases, in particular peptiligase and omniligase-1, is very advantageous in situations where flexibility is desirable. In cases where more specific modifications are required (e.g. for use in living cells), the use of subtiligase variants may be problematic due to the catalysis of off-target modifications. In addition, the requirement for C-terminal (thio)ester substrates, often produced using chemical synthetic methods, limits application when it comes to recombinantly produced proteins.[94]

1.2.2.4 Asparaginyl Endopeptidases

Asparaginyl endopeptidases (AEPs), also known as legumains, are cysteine proteases that catalyse peptide bond cleavage after an asparagine or an aspartate residue (Asx). However, some AEPs preferentially function as ligases. In plants, these peptide asparaginyl ligases (PALs) catalyse the cyclisation of Asx containing peptides, such as cyclotides, which function as host defences.[95,96]

As with many proteases, AEPs are expressed as inactive zymogens, consisting of a cap domain covering the active site of the core domain (**Figure 1.19**). Under acidic conditions, such as those found in plant vacuoles, the zymogen undergoes an autoactivation process in which the cap domain is cleaved from the core domain, generating the active enzyme.[97]

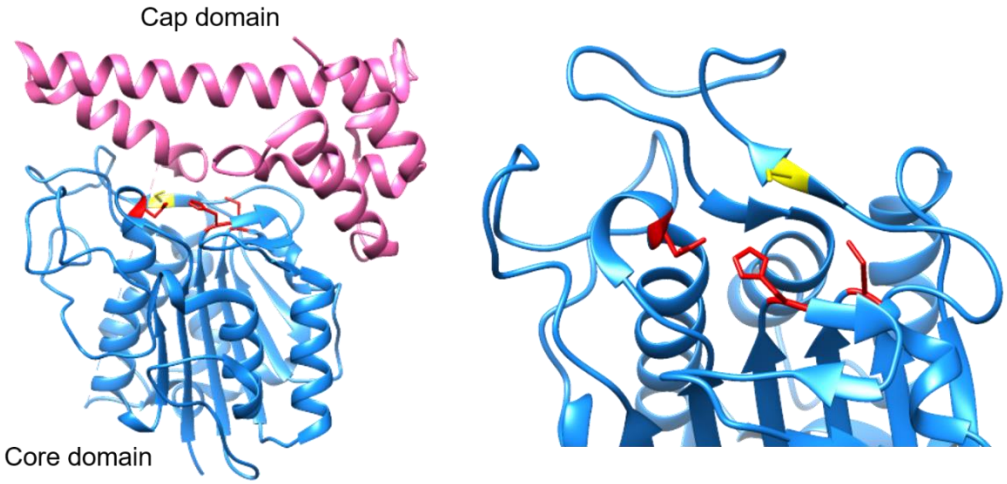


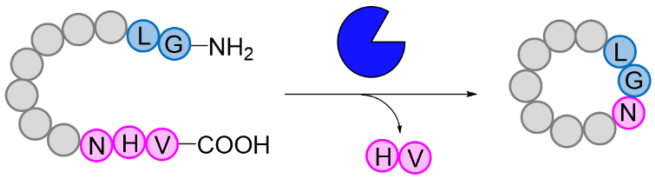
Figure 1.19 OaAEP1 zymogen (PDB 5H0I), the core domain is shown in blue and the cap domain in pink (the linker connecting the two is not shown). Catalytic residues Cys217, His175 and Asn170 are in red, while the gatekeeper Cys247 is highlighted in yellow.

Despite opposite activities, the crystal structures of several AEPs and PALs (in both the zymogen and active enzyme forms) show high structural similarities. So far, a number of sites, both near and distal from the enzyme active site, have been identified to be critical for ligase activity. Among these are a “gatekeeper” residue situated within a so-called ligase-activity determinant (LAD) region, a poly-proline loop and a marker of ligase activity (MLA). The gatekeeper residue and surrounding residues control the access of water or amine nucleophiles to the active site. As a result, ligation is favoured over hydrolysis when these residues are hydrophobic.[98,99] On the other hand, the MLA is located away from the active site and appears to affect only the ligation reaction rate. In ligases the MLA has been found to be either truncated or rich in hydrophobic residues.[100]

Butelase 1 was the first PAL to be reported.[101] It was isolated from the tropical cyclotide producing plant *Clitoria ternatea*. In contrast to its isoform butelase 2, butelase 1 shows a

strong preference for the catalysis of transpeptidation over hydrolysis of Asx containing substrates, making it a highly efficient ligase. It was demonstrated to catalyse the cyclisation of non-native substrate targets from various origins, including the plant cyclotide kalata B1, sunflower trypsin inhibitor, conotoxin MrIA from marine snails, insect antimicrobial peptide thanatin and the antimicrobial peptide histatin-3 from human saliva. Using a low enzyme-to-substrate ratio (1:400), cyclic peptide products were generated in a high yield (>90%, except thanatin), in as little as 12 minutes.[101] In addition, butelase 1 was also demonstrated to cyclise larger proteins such as GFP, interleukin-1 receptor antagonist and somatropin. Again, proteins were cyclised efficiently in a high yield.[102]

On top of high catalytic activity, butelase 1 exhibits broad sequence promiscuity and enables ligation to be carried out in a near traceless manner, leaving only a single Asx residue. At the C-terminal of the substrate, the preferred tripeptide motif for enzyme recognition is Asn-His-Val (NHV). As an Asx specific ligase, butelase 1 also recognises Asp in place of Asn, though cyclisation efficiency is significantly reduced. Meanwhile, at the N-terminal, if the first residue is Gly or if the second residue Leu, any amino acid is accepted at the other of the two positions. Interestingly, substrates composed of D-amino acids are also tolerated, as long as the L-isomer of the Asx residue remains.[103]



Scheme 1.20 Cyclisation catalysed by butelase 1. The C-terminal tripeptide recognition sequence is shown in pink and the N-terminal residues in blue.

Initially, butelase 1 was only available by extraction from the pods of *C. ternatea*, a labour intensive route requiring four chromatographic steps.[104] As a result, the application of this highly efficient ligase was severely limited. Fortunately, in more recent work, the recombinant expression of the zymogen of butelase 1 in both *E. coli* and in yeast (*Pichia pastoris*) was made possible.[105,106] Therefore, the practicality of butelase 1 as a ligation tool has been greatly improved for future use.

OaAEP1 is another PAL. Originally isolated from the plant *Oldenlandia affinis*, it can be readily produced by recombinant expression in *E. coli*.[107] Though this enzyme preferentially functions as a ligase, it does so with a lower catalytic efficiency than butelase 1. Similarly, OaAEP1 recognises and processes a C-terminal Asn containing tripeptide motif. In the native substrate kalata B1, this is Asn-Gly-Leu (NGL). However, Ala and Cys have also

been shown to be suitable residues in place of Gly (i.e. NAL and NCL, respectively) and can improve cyclisation efficiency by up to 30%.[108]

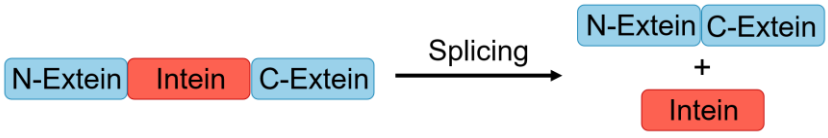
A diverse range of peptide and protein substrates have been cyclised using OaAEP1.[98,107] Usually protein cyclisation is limited to structurally well-defined globular proteins with N- and C- termini in close proximity. However, OaAEP1 was demonstrated to cyclise the malarial vaccine candidate merozoite surface protein 2 (MSP2), which is an intrinsically disordered protein when expressed recombinantly.[109,110] Even the closely related butelase 1 failed to cyclise proteins after heat denaturation.[102] It should also be pointed out, that in larger proteins there is increasing likelihood of multiple Asx recognition sites being present in the structure. This can result in increased off-target processing by the PAL.[109]

So far butelase 1 and OaAEP1 are the most studied and well characterised AEP ligases. However, in recent years there has been a surge in the discovery of new AEPs and PALs, some of these include VyPALs (*Viola yedoensis*), HeAEP3 (*Hybanthus enneaspermus*) and HaAEP1 from (*Helianthus annuus*).[99,100,111] Not only does this advance research into AEPs as valuable ligation tools, but it also enhances the understanding of the AEP and PAL enzymes and the mechanisms by which they catalyse hydrolysis and ligation reactions, respectively.

1.2.3 Protein Tag Cyclisation Methods

1.2.3.1 Intein-Mediated Cyclisation

Inteins are protein domains that undergo protein splicing, a post-translational auto-processing event during which the intein excises itself from the protein and ligates its N- and C- flanking sequences (exteins) by the formation of a native peptide bond (**Scheme 1.21**).

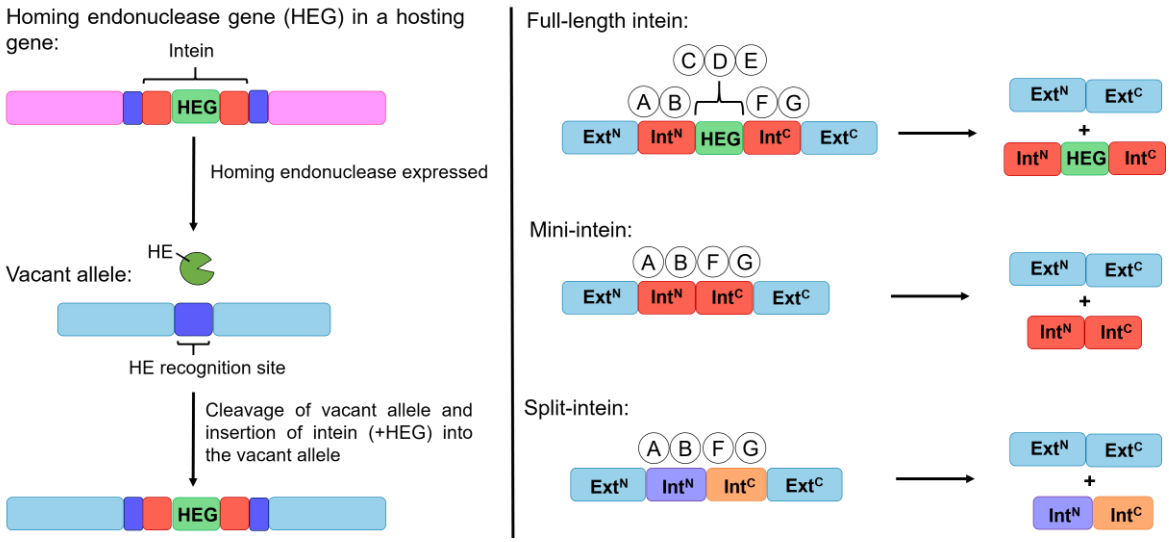


Scheme 1.21 Intein-mediated protein splicing.

Inteins have been identified in unicellular archaea, bacteria and eukaryota domains of life. They have been found to be inserted into a diverse range of host proteins, usually involved in DNA replication, transcription and maintenance.[112] Despite their widespread occurrence in nature, inteins have not yet been shown to have a clear biological role and are considered to be selfish genetic elements as neither the gene nor the protein product appear to contribute anything to the host.[113] However, there is increasing evidence that intein excision

may act as an environmental cue to regulate biological functions during stress, with conditional splicing observed in a number of cases.[114,115]

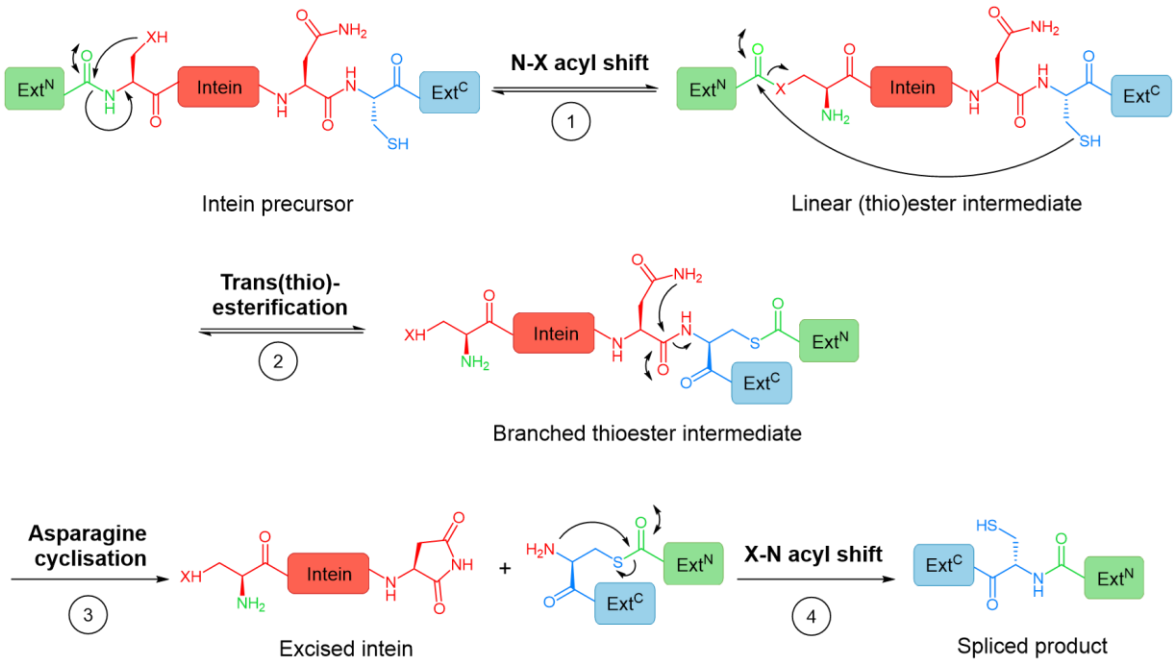
In addition to splicing, many inteins exhibit homing endonuclease (HE) activity. The HE domain is incorporated into the gene of full-length inteins and promotes horizontal propagation of the respective intein into vacant alleles, the same mechanism by which introns are spread. The HE recognises and cleaves specific DNA sequences. When repaired, this results in the intein gene (containing the HE domain) being copied into the cleavage site (**Scheme 1.22**).[116] It is this HE activity that is thought to be responsible for intein invasion and spread in populations.[112] Inteins lacking the HE domain are known as mini-inteins whose protein splicing activity is unaffected by the absence of endonuclease activity.[117]



Scheme 1.22 (Left) Intein homing endonuclease activity. (Right) Types of intein and locations of their conserved motifs.

While most inteins share a low sequence homology (<40%), four conserved motifs (blocks A, B, F and G) have been found in all known inteins and can be used to identify and characterise new inteins.[118] These four motif blocks contain highly conserved amino acid residues essential for protein splicing. Meanwhile, blocks C, D, E and H (in full-length inteins) are involved in endonuclease activity. Generally protein splicing proceeds via the steps outlined below (**Scheme 1.23**).

1. N-O or N-S acyl shift leads to (thio)ester intermediate formation.
2. Trans(thio)esterification between N- and C-exteins results in a branched intermediate.
3. Intein excision proceeds through asparagine (or sometime glutamine[119]) cyclisation.
4. O-N or S-N acyl shift results in peptide bond formation between exteins.

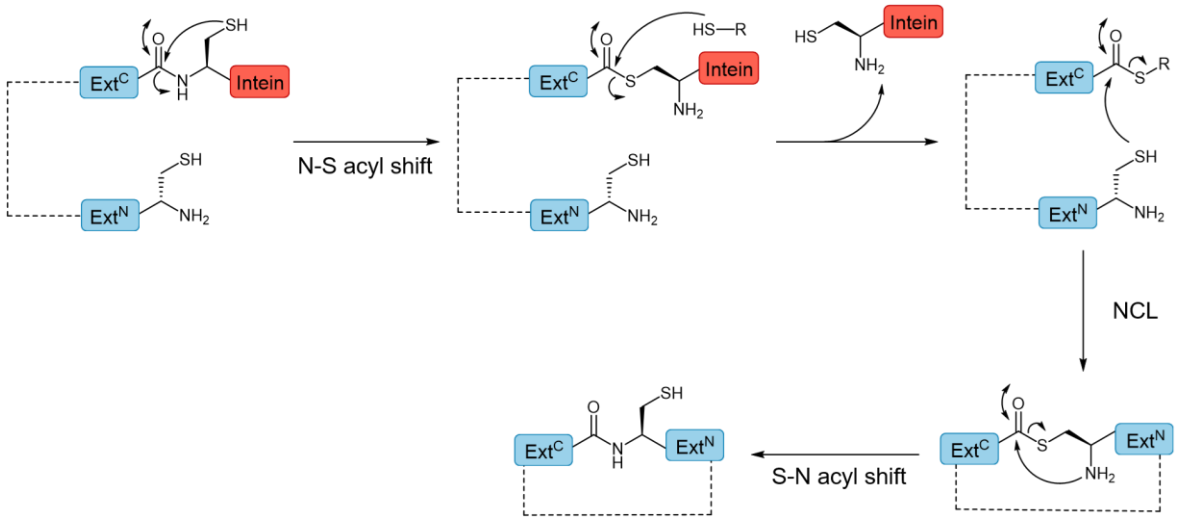


Scheme 1.23 Intein splicing mechanism showing steps 1-4 (X = O/S).

The splicing reaction is often spontaneous, only requiring the correct folding of the intein for the alignment of critical residues, without the need of a cofactor or external energy source. The intein fold is highly conserved, with all intein structures solved by NMR and X-ray crystallography exhibiting a horseshoe-like folding topology, consisting of β-sheets, loops and two short helices.[120] This complex folding topology is responsible for bringing the termini of the extein sequences into close proximity with conserved sequence motifs oriented around the extein termini to carry out each of the protein splicing steps.

1.2.3.1.1 Expressed Protein Ligation

Expressed protein ligation (EPL) is a semisynthetic extension of NCL (see 1.2.1.5) and enables the ligation of larger recombinantly produced proteins to shorter synthetic peptides. In NCL a native peptide bond is produced between a polypeptide bearing a C-terminal thioester and a polypeptide bearing an N-terminal cysteine. However, this imposes a size limitation when thioester groups are produced synthetically. During intein-mediated protein splicing a thioester intermediate is produced. This thioester intermediate can be trapped by mutation of the intein to stop the transfer of the N-extein (Ext^N) to the C-extein (Ext^C) during N-S acyl shift. Subsequent reaction with a synthetic peptide bearing a N-terminal cysteine generates the ligated product (**Scheme 1.24**).[121]

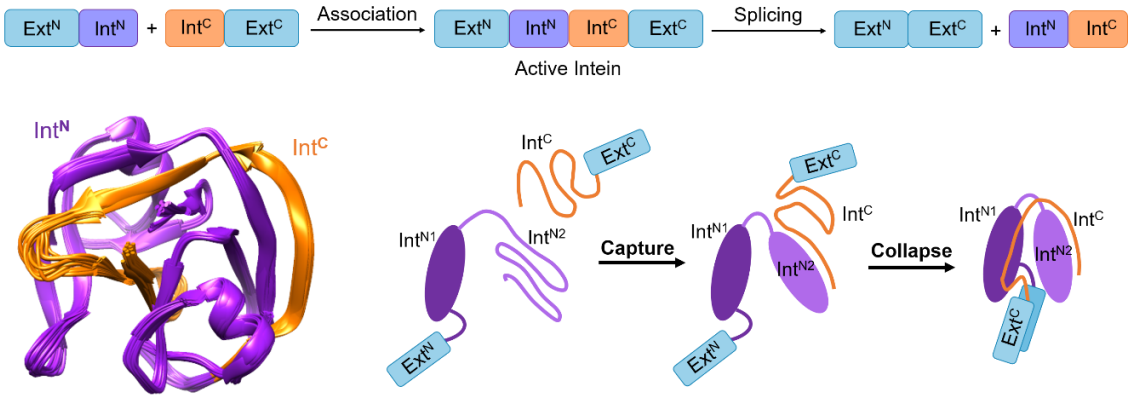


Scheme 1.24 EPL mechanism.

As well as intermolecular ligation, EPL can also be used to carry out intramolecular ligation for the production of cyclic peptides and proteins. For example, EPL was successfully employed to cyclise the N-terminal Src homology 3 domain of c-Crk, as well as β-lactamase (BLA).[122,123]

1.2.3.1.2 Split Inteins

While most inteins exist naturally in the single contiguous polypeptide form (*cis*-splicing inteins), they can also exist in a rarer split form.[114] These so called split inteins are transcribed and translated as two separate polypeptide chains and undergo a protein *trans*-splicing (PTS) mechanism, whereby protein splicing (i.e. steps 1-4) is preceded by the association of the N- (Int^N) and C-intein (Int^C) fragments to form the active intein (**Scheme 1.25**). This intein association step has been found to take place via a "capture and collapse" mechanism, leading to a tight (low nanomolar affinity) and rapid binding of Int^N and Int^C fragments.[124,125] Electrostatic charge complementarity between the anionic and disordered C-terminal half of the Int^N (Int^{N2}) and the cationic and disordered Int^C, initiates the folding ("capture"). The resulting intermediate is then stabilised by hydrophobic interactions between Int^C and the more ordered N-terminal half of the Int^N (Int^{N1}) ("collapse").



Scheme 1.25 (Top) *Trans*-splicing. The split intein fragments first associate to form the active intein before splicing occurs. (Bottom left) Structure of *Npu* DnaE split intein (PDB 2KEQ), N- and C-intein fragments are shown in purple and orange, respectively. (Bottom right) Capture and collapse mechanism of split intein fragment association.

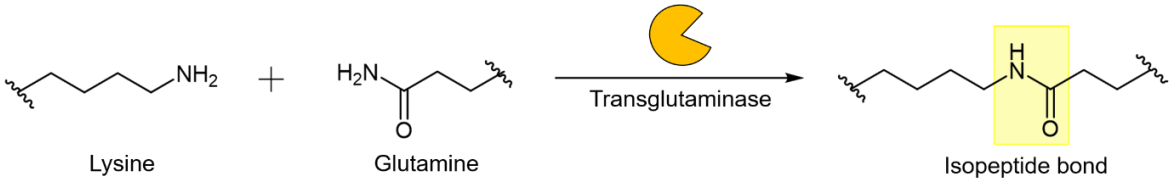
Split inteins are often found in cyanobacteria. One such split intein is *Ssp* DnaE, found in the catalytic subunit (DnaE) of DNA polymerase III from the cyanobacterium *Synechocystis* sp. strain PCC6803 (*Ssp*).[126] It has been used for both in vivo and in vitro applications. Nevertheless, its widespread use in chemical biology is limited by slow kinetics (first order rate constant of 6.6 x 10^{-5} s^{-1} and $t_{1/2}$ of 175 min) and dependence on the extein sequence.[127,128] On the other hand, the well-studied DnaE split intein from the cyanobacterium *Nostoc punctiforme* PCC73102 (*Npu*) exhibits much more favourable characteristics for biotechnological application.[129] In vitro characterisation under optimal conditions showed the *Npu* DnaE split intein to exhibit more efficient splicing activity, with the splicing reaction taking place in minutes instead of hours (first order rate constant of 3.7 x 10^{-2} s^{-1} and $t_{1/2}$ of 19 s).[130] Furthermore, *Npu* DnaE exhibits greater tolerance towards extein sequence variation, although deviation from the native Ext^C sequence can result in a significant reduction in splicing efficiency. Meanwhile, the Ext^N sequence has been shown to be less sensitive to the nature of residues. However, mutation of the native residues can alter linear thioester formation or promote hydrolysis of the thioester intermediate.[131]

Like EPL, PTS is an efficient peptide and protein ligation method. When the Int^C and Int^N fragments are placed at separate ends of the extein sequence of interest, an intramolecular splicing reaction takes place resulting in traceless cyclisation (leaving only a single Cys/Ser residue at the ligation site). From PTS, the split intein circular ligation of proteins and peptide (SICLOPPS) method was developed, in which cyclic peptide libraries consisting of hundreds of millions of members, can be rapidly generated in cells (*E. coli*, yeast and human). These cyclic peptides can then be screened for activity against a specific drug target.[132,133] PTS has also been employed for the cyclisation of proteins. For example,

dihydrofolate reductase (DHFR) and granulocyte-colony stimulating factor (GCSF) were cyclised using *Ssp* and *Npu* DnaE split inteins, respectively.[134,135] Both cyclised variants showing enhanced thermal and proteolytic stability. Nevertheless, it is commonly reported that the introduction of intein sequences can cause protein misfolding. In addition, the relatively large size of the intein can lead to reduced yields from recombinant expression, with the *Npu* DnaE split intein reported to be toxic to *E. coli*.[133]

1.2.3.2 Isopeptide Bond Formation

Isopeptide bonds are chemically stable and protease resistant amide linkages outside the protein main chain formed between amino acid side chains. In nature, their formation is catalysed by transglutaminases, which facilitate reaction usually between the side chain groups of Gln and Lys residues, releasing NH_3 as a by-product (**Scheme 1.26**). Although commercially available, transglutaminases have low sequence specificity which hinders their application in protein cyclisation.[136]



Scheme 1.26 Isopeptide bond formation catalysed by a transglutaminase.

The crystal structure of Spy0128, a major pilin protein from *Streptococcus pyogenes*, showed that an isopeptide bond was present in its structure.[137] By splitting Spy0128 into a 16 residue peptide (Isopeptag) and a larger fragment (Pilin-C), a peptide-protein partner was created which could spontaneously re-associate through amide bond formation with a yield of up to 98% within 24 hours.[138] Furthermore, the reaction was found to be largely independent of temperature, pH and buffer. This led to further investigation into other systems to find a smaller and faster reactive pair.

The CnaB2 domain from a fibronectin binding protein in *S. pyogenes*, was found to contain an isopeptide bond. It was split into a 13 residue peptide (SpyTag) and a 116 residue fragment (SpyCatcher) (**Figure 1.27**).[139] The two parts could spontaneously and efficiently reconstitute in vitro and in vivo via isopeptide bond formation between Asp117 and Lys31, promoted by the carboxy group of the nearby Glu77. Water was released as a by-product.

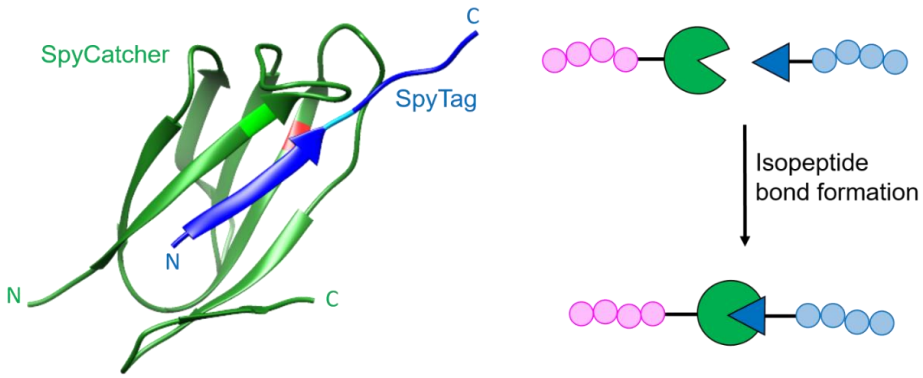
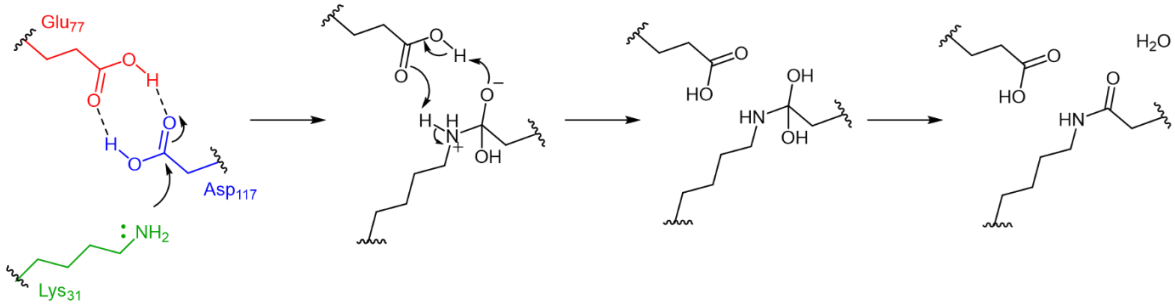


Figure 1.27 (Left) SpyTag/SpyCatcher (PDB 4MLI). SpyTag peptide is shown in blue with the position of the Asp117 highlighted in bright blue. SpyCatcher protein fragment is shown in green with the position of the Lys31 residue in bright green and the catalytic Glu77 in red. (Right) Ligation using SpyTag/SpyCatcher peptide-protein partners.

The accepted mechanism of isopeptide bond formation by SpyTag/SpyCatcher is shown in **Scheme 1.28**. It is based on that of the intact CnaB2 domain which has been studied using QM/MM calculations, NMR and X-ray crystallography.[140–142] In the first step, the Lys31 nucleophilically attacks the Asp117. Next, a neutral tetrahedral intermediate is formed by two concerted proton transfers with Glu77 acting as a proton shuttle. Finally, the intermediate collapses, water is released and an amide bond is formed. The SpyTag/SpyCatcher protein environment facilitates the reaction by bringing the residues into close proximity with the correct orientation. In addition, the Lys31, Asp117 and Glu77 residues are located within a hydrophobic pocket lined with aromatic residues, resulting in the inverse protonation states of the residues being favoured.



Scheme 1.28 The reaction mechanism of isopeptide bond formation between the Lys31 residue (green) of SpyCatcher and the Asp117 residue (blue) of SpyTag, catalysed by Glu77 (red).

The SpyTag/SpyCatcher partners have been used to cyclise a number of proteins including BLA, DHFR and firefly luciferase among others, all of which showed improved stability compared to their linear forms.[143,144] As a method of ligation, the SpyTag/SpyCatcher approach offers several advantages. For example, isopeptide bond formation is high yielding and rapid (70% reconstitution in 5 min, second order rate constant of $1.4 \times 10^3 \pm 43\ M^{-1}\ s^{-1}$) over a range of reaction conditions (4 – 37 °C, pH 5 – 8, with no requirement for

specific anions or cations). The reaction only requires mixing and can take place when SpyTag and SpyCatcher fragments are located at either the C- or N-terminus, or at internal positions of the protein. Furthermore, SpyTag/SpyCatcher complex formation is irreversible and mechanically stable.[139]

The major drawback associated with SpyTag/SpyCatcher is its large size, meaning cyclisation is not traceless. However, efforts have been made to optimise and reduce the size of the scar remaining after ligation (**Figure 1.29**). By removing a total of 32 residues, a minimised SpyCatcherΔN1C2 fragment (reduced from 116 to 84 residues) was produced without adversely effecting activity or structure.[145] Further reduction in size was achieved by circular permutation of SpyCatcher (SpyCatcher-N) and introduction of a TEV protease cleavage site (SpyCatcher-N^{TEV}). This resulted in the reactive Lys being positioned nearer the C-terminus of the construct and allowed the removal of the N-terminal fragment of SpyCatcher by proteolytic digestion.[146]

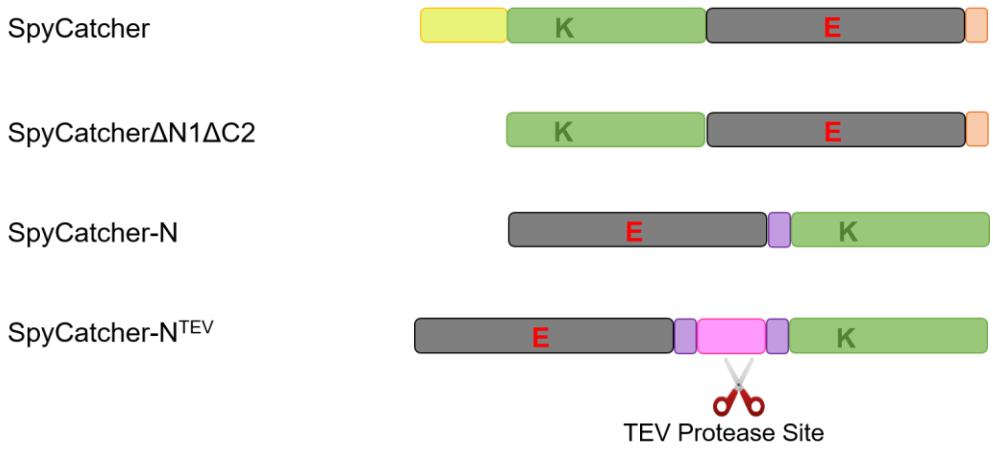


Figure 1.29 SpyCatcher variants for reduction in size. Different regions of the sequence are shown in different colours. The green segment contains the Lys31 residue, the black segment contains the Glu77 residue and the cleavable region in the -N^{TEV} variant is shown in pink. Yellow, orange and purple segments correspond to shorter sequences and linkers.

Similarly, splitting of the D4 Ig-like domain from *Streptococcus pneumonia* produced the SnoopTag/SnoopCatcher system, where isopeptide bond is formed between a Lys and Asn residue.[147] It was successfully employed to cyclise firefly luciferase and was observed to confer improved stability to the cyclised product compared to the linear control.[144]

Table 1.1 Summary of chemical, enzymatic and protein tag cyclisation techniques discussed in Chapter 1.

	Cyclisation Approach	Advantages	Potential Drawbacks
Chemical	Direct amide bond formation	• Reaction can proceed under mild conditions with C-terminal activation • Traceless cyclisation	• Side chain protecting groups required • Additives used to enhance reaction rate and reduce racemisation
	Staudinger ligation	• Chemoselective towards the azide (does not require protecting groups) • Traceless cyclisation	• Unnatural functionalities required in the starting material • Poor solubility of phosphinothiols in aqueous solution • Glycine residues required at the ligation site
	KAHA ligation (type I)	• Chemoselective • Traceless cyclisation	• Unnatural functionalities required in the starting material • Relatively slow reaction • Free hydroxylamine is unstable in aqueous solution
	KAHA ligation (type II)	• Chemoselective • O-substituted hydroxylamine is stable in aqueous solution • Compatible with polypeptides of varying sizes (chemically synthesised)	• Unnatural functionalities required in the starting material • Relatively slow reaction • Non-native homoserine residue formed at the ligation site
	CuAAC	• Regioselective and efficient • Requires mild conditions and can be carried out in aqueous solution	• Requires unnatural functionalities in the starting material • Cu(I) catalyst generated in situ from Cu(II) using excess reducing agent and Cu-stabilising ligands • Cu is toxic to cells • Produces an unnatural functionality in the product
	SPAAC	• Circumvents the need for a Cu catalyst • Rapid reaction under physiological conditions	• Produces a mixture of products • Cyclooctyne reagents are relatively expensive
	NCL	• Reaction proceeds under aqueous conditions at neutral pH • Regio- and chemoselective • Variety of extensions developed to expand scope of application (e.g. desulfurisation)	• Requires a C-terminal thioester • Some extensions require unnatural functionalities in place of the N-terminal cysteine
	Disulfide bond formation	• Cysteine residues can be incorporated chemically or recombinantly • Occurs under physiological conditons	• Correct disulfide bond formation may require the use of protecting groups, orthogonal pairing strategies or stapling reagents (containing unnatural functionalities)
Enzymatic	Sortase (SrtA)	• Reaction is well-studied and understood • Commercially available • Mutants developed for improved activity	• Ligation is reversible (recognition sequence remains in the product) • Relatively low catalytic efficiency (requires long reaction times or high enzyme concentrations) • Ca^{2+} dependent
	Subtiligase	• Chemoselective for N-terminal amines • Well-studied and broad substrate scope • Effectively traceless cyclisation • Variants (peptiligase and omniligase-1) available with improved properties	• Ca^{2+} dependent • Broad substrate scope can lead to off-target modifications • Requires peptide ester substrates (often chemically synthesised)
	AEP	• High catalytic efficiency • Effectively traceless cyclisation • Relatively broad substrate scope • Enzyme preparation can be laborious	• Ligation is reversible (recognition sequence remains in the product) • Short recognition sequence can lead to off-target modifications
Protein Tag	Intein (EPL and split intein)	• Effectively traceless • Overcomes the size limitations of NCL • Can be expressed and purified alongside the protein/peptide of interest	• Can lead to protein misfolding • Can reduce protein yield after recombinant expression
	SpyTag/SpyCatcher (and SnoopTag/SnoopCatcher)	• Wide range of reaction conditions • Efficient and high yielding • Can be expressed and purified alongside the protein/peptide of interest	• Leaves a large scar at the ligation site

1.3 CONSIDERATIONS FOR CYCLISATION

Cyclisation has been demonstrated to be a convenient means for extending the stability and activity of peptides and proteins outside of physiological environments. A wide variety of methods for carrying out cyclisation exist, each has its own strengths and limitations. Some of these were discussed in the section above and are summarised in **Table 1.1**. Clearly, no single approach is ideally suited to the cyclisation of all peptides and proteins. Therefore, some of the key factors to consider when choosing a cyclisation method will be discussed in this section, with the focus mainly on protein cyclisation and its effects on stability.[14]

1.3.1 Distance Between Termini

The function of a protein is highly dependent on its overall three-dimensional structure. Therefore, it is critical that cyclisation does not interfere adversely with the protein's complex folded structure, which would be detrimental to its activity. As a result, it is important to consider the distance between the residues that are to be ligated during cyclisation. If situated too far apart, strain could be introduced leading to distortion. This would offset any stability enhancement brought about through a decreased entropy. Existing structural information, such as a crystal structure, can be employed to check the suitability of cyclisation to the protein of interest and to select the most appropriate residues to take part. Around half of the single domain proteins in the Protein Data Bank (PDB) have contact between N- and C-terminal elements.[148] Moreover, it was found that out of approximately 2000 representative proteins, 31% have their termini within 20 Å and 11% within 15 Å.[149] Hence, a large number of proteins could be amenable to cyclisation.

In the situation where termini are located too distally for direct ligation, linker sequences or bridging reagents could be used to bring the connecting residues into closer proximity.[51,150] GCSF, a cytokine composed of a four-helix bundle, was cyclised using a split intein (*Npu* DnaE).[135] By varying the length of the linker connecting the N- and C-termini of GCSF, it was shown that the structure of the connector could be optimised to promote either enhanced stability or efficiency of split intein cyclisation (**Figure 1.30**). When a longer connector was used (5 residues), thermal stability was improved by almost 13 °C compared to the linear protein, though some unspliced starting material remained. On the other hand, only a 2 °C increase was seen with a shorter linker (2 residues). However, split intein cyclisation efficiency increased with only the cyclised product being observed.

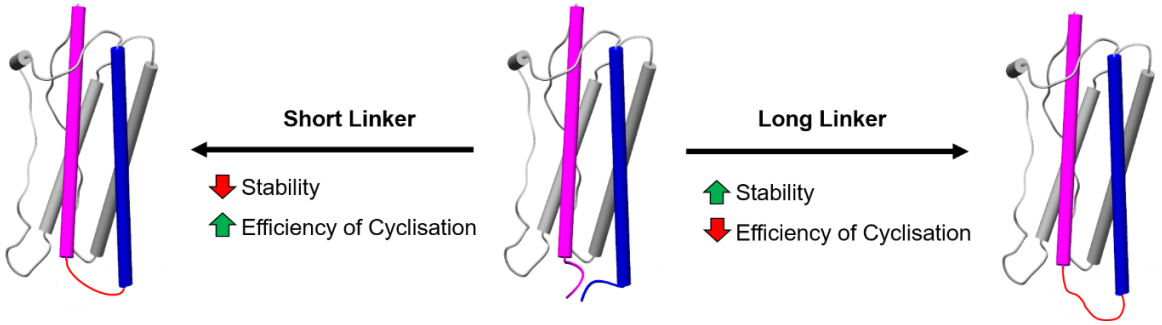


Figure 1.30 Cyclisation of GSCF (PDB 2D9Q) with shorter and longer linker lengths.

Interestingly, the short linker cyclised GCSF became the first cyclised protein to have its crystal structure published (**Figure 1.31**).[151] It confirmed that the termini were connected by a peptide bond and that the introduction of the short connector caused a destabilising structural change to compensate for the stress introduced through cyclisation. Tilting of the C-terminus was believed to favour splicing by blocking the access of water leading to the observed increase in splicing efficiency.

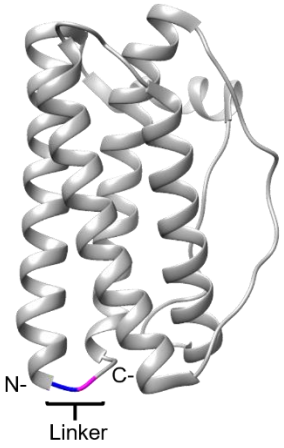
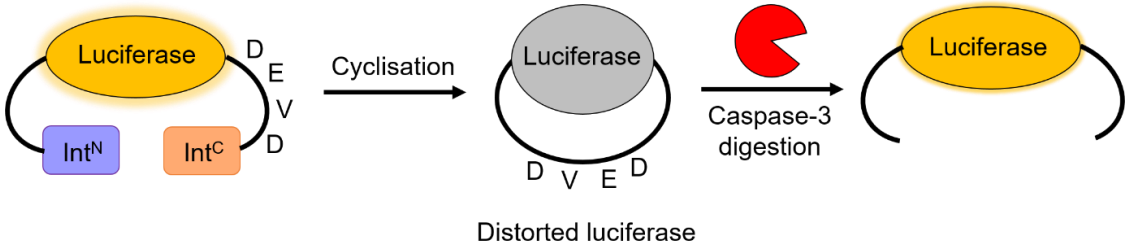


Figure 1.31 Structure of cyclised GSCF (PDB 5GW9). Linker residues between N- and C-termini shown in blue and pink, respectively.

In a further study, molecular dynamic simulation of the long linker cyclised GCSF, revealed there to be an increase in the number of intramolecular hydrogen bonds present in the cyclic structure compared to the linear protein. These additional hydrogen bonds, situated in and adjacent to the linker region, were generated upon cyclisation. Therefore, it was demonstrated that both the decrease in entropy of the unfolded state, as well as the increase in enthalpy of the folded state, were synergistically responsible for the superior heat tolerance and resistance to chemical denaturation of the cyclised GCSF.[150] Furthermore, it highlights the importance of linker optimisation for achieving maximum stabilisation through cyclisation.

Although distortion caused by cyclisation is disadvantageous for improved stability, it can be exploited. For example, it was demonstrated that cyclic luciferase could be used for real-time sensing of the activity of the protease caspase-3 in living mammals.[152] Cyclisation by split intein led to distortion of the luciferase structure and loss of bioluminescence activity. If N- and C-termini were linked by the caspase-3 recognition sequence (DEVD), activity would be restored as a result of digestion by caspase-3 (**Scheme 1.32**).



Scheme 1.32 Cyclisation of luciferase by a split intein leads to distortion and loss of bioluminescence activity. In the presence of caspase-3 the DEVD recognition sequence is cleaved, resulting in restoration of luciferase structure and activity.

1.3.2 Polypeptide Preparation

The methods by which the polypeptide of interest is produced and purified are also important to consider when choosing a cyclisation approach. Short peptides are commonly synthesised using SPPS, followed by purification using reverse-phase HPLC. Larger polypeptides can also be chemically synthesised by ligating smaller fragments, for example using NCL. Nevertheless, it is often preferable to produce larger proteins recombinantly using cellular machinery. Multiple purification steps are then usually required to achieve the desired level of purity. These include affinity, ion exchange and/or size exclusion chromatographies. The general relevance of these preparation techniques to each category of cyclisation approach will be discussed below.

For chemical cyclisation approaches, non-native functionalities and/or non-physiological conditions are invariably required. As such, polypeptides that have been produced chemically are generally better suited to these approaches, as they can be prepared in organic solvent, in a fully protected form. Moreover, the necessary non-native functional groups can be more easily incorporated into the desired positions in the polypeptide chain. These can also be introduced recombinantly using genetic code expansion, though there are more limitations regarding the amino acid substrates that can be incorporated.[153]

Generally, enzyme catalysed cyclisation approaches are applicable to both chemically synthesised and recombinantly produced polypeptides. Despite this, the preparation of the

enzyme itself can be time-consuming and labour intensive, unless it is commercially available. In addition, the presence of multiple recognition sequences in larger proteins can result in off-target modifications resulting in degradation of both the linear starting material and/or cyclised product.[109]

Protein tag cyclisation approaches employ inteins or isopeptide bond forming Tag/Catcher partners which are often composed of sequences of ~100 amino acid residues. As a result, it is most convenient to recombinantly express these long sequences alongside the polypeptide sequence of interest, circumventing the need for any additional synthetic or purification steps. Furthermore, the presence of the long protein tag flanking sequences can facilitate the recombinant production of short peptides which are often produced in low yields by cellular machinery.[154] If using an intein-mediated strategy, these long sequences are excised during the cyclisation process and so do not remain in the cyclic product.[133]

1.3.3 Effect of Cyclisation Technique on Stability

Stability can be defined in different ways, including tolerance towards increased temperatures (thermal stability), non-neutral pH and chemical denaturants (chemical stability), as well as resistance to enzymatic degradation (proteolytic stability). In this section, the effect of cyclisation on the stabilities of some cyclised proteins will be discussed. Interestingly, the cyclisation strategy employed appears to influence the extent of stabilisation achieved.

1.3.3.1 Thermal Stability

The thermal stability of a protein is most commonly reported as the melting temperature (T_m). This is the temperature at which the midpoint of unfolding occurs and can be determined by monitoring the protein secondary structure at increasing temperatures, by using for example, circular dichroism (CD) (**Figure 1.33**).[155] Alternatively, heat induced aggregation can be investigated, by observing the temperature at which the protein is lost from solution. When studying enzymes, thermal inactivation can be a useful measure of its resistance to thermal denaturation.

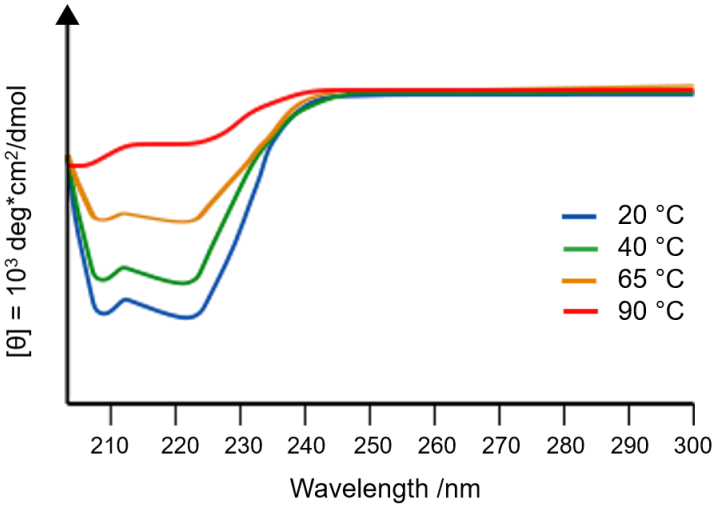


Figure 1.33 Representation of a CD spectrum showing loss of protein secondary structure as temperature is increased from 20 to 90 °C.

A variety of different cyclised proteins have been demonstrated to have improved thermal stability over their linear counterparts using a range of different cyclisation methods. For example, BLA cyclised using EPL was shown to have an increased T_m of 5 °C compared to is linear form.[123] More recently, a larger improvement in thermal stability was observed when BLA was cyclised by isopeptide formation through the use of SpyTag/SpyCatcher.[143] This cyclised BLA could remain in solution even when heated to 100 °C, after which it still retained enzymatic activity. In comparison, the linear form completely aggregated at 55 °C.

A similar resistance to heat induced aggregation was observed for DHFR when cyclised by SpyTag/SpyCatcher. Increased thermal stability was also observed when DHFR was cyclised by split intein ligation and by disulfide bond formation with a dithiothreitol linker, but not to the same extent as with SpyTag/SpyCatcher system.[134,156] This indicates that the extent of stabilisation achieved can vary depending on the cyclisation method employed. Both BLA and DHFR were more resistant to heat treatment upon cyclisation by SpyTag/SpyCatcher than by other methods, such as split intein mediated cyclisation.[143]

The stabilisation effect of isopeptide bond formation was investigated by cyclising luciferase using different Tag/Catcher partners. For both SpyTag/SpyCatcher and SnoopTag/SnoopCatcher, the cyclised luciferase was found to remain catalytically active over a broader range of temperatures compared to the wild type luciferase and linear controls. The SnoopTag/SnoopCatcher luciferase was found to have an increase in T_m of 7.3 °C, while the T_m of the SpyTag/SpyCatcher luciferase was increased by 16.7 °C. With the minimised SpyTag/SpyCatcherΔN1ΔC2 system (see 1.2.3.2) an even larger increase in T_m was observed with a 19.6 °C improvement over the linear wild type.[144,157]

Differential scanning calorimetry (DSC) was employed to investigate the heat induced unfolding of BLA cyclised by SpyTag/SpyCatcher, SnoopTag/SnoopCatcher and Isopeptag/Pilin-C isopeptide bond forming peptide and protein partners.[158] In the melting profiles obtained for the cyclised BLA, two maxima were observed. The T_m of the cyclised BLA corresponded to the lower temperature peak, while that of the peptide-protein partner cyclisation domain corresponded to the higher temperature peak. This was found to be 85 °C, 64.1 °C and 72.9 °C for SpyTag/SpyCatcher, SnoopTag/SnoopCatcher and Isopeptag/Pilin-C, respectively. Therefore, the isopeptide bond forming domains are likely conferring thermal resilience to the cyclised protein by facilitating refolding. This was exploited for the purification of SpyTag/SpyCatcher cyclised phytase, whereby it could be purified from the cell lysate just by heating.[158]

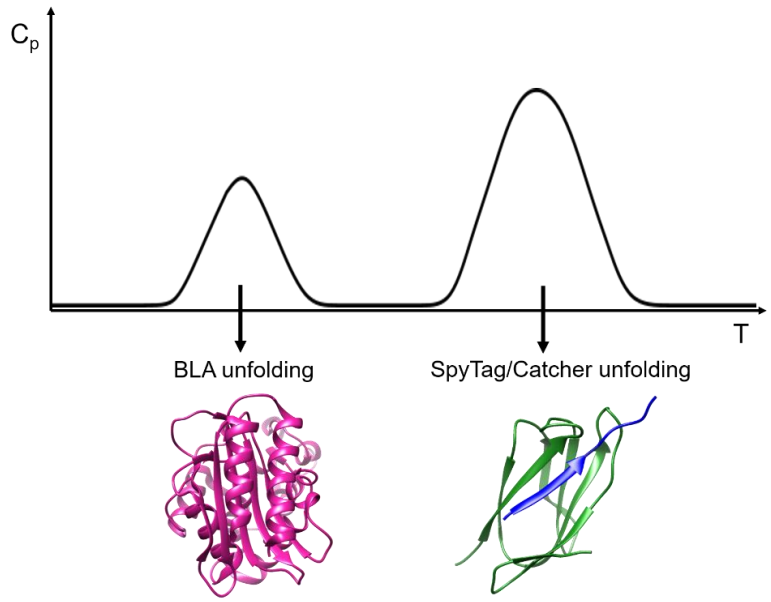


Figure 1.34 DSC profile of BLA-SpyTag/SpyCatcher (BLA, PDB 1ZG4). The lower temperature peak corresponds to BLA unfolding. The higher temperature peak corresponds to SpyTag/SpyCatcher unfolding.

1.3.3.2 Chemical Stability

As well as conferring enhanced thermal stability, cyclisation can also improve the stability of proteins towards the presence of chemical denaturants. Fluorescence and CD analysis of linear and cyclic GFP in varying concentrations of guanidine hydrochloride (GdnHCl) showed the cyclic form to be protected against chemical denaturation.[159] In 7 M GdnHCl, the unfolding rate of the cyclic form was half that of the linear form and residual secondary structure was still detected after 3 hours in high concentrations of GdnHCl.

In some cases, cyclisation leads to improved tolerance to a broader range of pH. Cyclised lichenase was observed to have an optimal pH at 0.6 units lower than its linear counterpart.[160] On the other hand, cyclised mannase (*B. subtilis*) had the same optimal pH for activity as the linear form (pH 6.0). Nevertheless, at an increased pH of 9.0, the cyclised variant had a residual activity 198% higher than that of the uncyclised protein.[161]

1.3.3.3 Proteolytic Stability

Cyclisation can also confer increased resistance to proteolytic degradation,[134] whereby the peptide or protein of interest is broken down by proteases which catalyse the hydrolysis of peptide bonds. In theory, this improved stability originates from the preference of the cyclised polypeptide towards the folded state, making it less likely to sample partially and fully unfolded states in which protease recognition sequences are exposed to the solvent and become accessible to protease attack.[10,162] Furthermore, head-to-tail cyclised polypeptides lack N- and C-termini and so are less susceptible to proteolysis by exopeptidases which catalyse the cleavage of terminal peptide bonds. For example, intein-cyclised BLA was demonstrated to be resistant towards exopeptidase treatment, while the linear form was observed to undergo degradation.[163] However, it is worth mentioning that cyclisation approaches employing side chain functionalities (e.g. disulfide bond formation) leave the termini free. As a result, the cyclic polypeptide of interest may still be susceptible to exopeptidase degradation, which should be considered when choosing a cyclisation approach.[164]

1.3.3.4 Other Effects

Despite the numerous reports of the enhanced thermal, chemical and/or proteolytic stabilities of cyclic proteins, cyclisation does not always result in improvements. Bovine pancreatic trypsin inhibitor was the first cyclised protein to be reported, but no significant stabilisation was observed.[165] As enthalpy as well as entropy can influence protein stability, it is possible that strain introduced through the crosslinking of N- and C-termini offset any favourable entropic effects.

In a more recent study, cyclisation of the N-terminal Src homology 3 (SH3) domain was also found to have little effect on stability. This was thought to be due to distortion of the protein backbone upon connection of the N- and C-termini by EPL.[122] However, in a truncated variant lacking a vital stabilising glutamate residue, backbone cyclisation was able to compensate for the loss of stability caused by the deletion. In addition, both the folding and unfolding rates of the cyclised SH3 domain were observed to increase, with the cyclised

protein being more resistant to denaturation by urea, as measured by fluorescence spectroscopy. An NMR study showed cyclisation to have a minor effect on the three-dimensional structure of the SH3 domain. On the other hand, the backbone dynamics of the cyclised protein were found to be more rigid on the subnanosecond timescale, though rigidity decreased as the connector length was increased.[166]

1.4 SUMMARY, AIMS AND OBJECTIVES

In summary, cyclisation has been demonstrated to confer a variety of peptides and proteins with enhanced properties compared to their linear forms. These include greater tolerance to heat treatment, the presence of chemical denaturants and resistance to proteolytic cleavage. As a result, cyclisation could be a beneficial tool for expanding the scope of polypeptide application in therapeutics and as biocatalysts. Nevertheless, the effects of cyclisation on the protein of interest are in many cases difficult to predict. The cyclisation technique employed can often influence the type and extent of stabilisation achieved. Moreover, the suitability of the cyclisation approach employed to the peptide or protein substrate of interest should be carefully considered, and may require optimisation. Furthermore, cyclisation may provide a useful tool for gaining insight into protein stabilisation and folding. Therefore, the aim of this thesis was to use cyclisation to improve the stabilities of two different biotechnologically relevant proteins. These are outlined as follows.

1. **To enhance the stability of an artificial protein cage (TET12) using cyclisation.** The internal cavities of protein cages have the potential be exploited for numerous biotechnologicial applications, from the delivery of drugs to use as catalytic reaction vessels. For many of these applications stability towards raised temperatures and/or proteolysis would be beneficial. As a result, the effect of cyclisation on the properties of the artificial TET12 protein cage will be investigated. Two different cyclisation approaches will be employed, specifically a head-to-tail split intein-mediated approach, as well as a side chain-to-side chain SpyTag/SpyCatcher-N^{TEV} strategy. The stabilities of the linear and cyclised TET12 variants will be compared. CD will be employed to monitor protein folding in response to increased temperature and concentration of chemical denaturant. Proteolytic stability of the linear and cyclic variants towards exopeptidase treatment will also be investigated. Furthermore, the effect of cyclisation on protein aggregation and folding will be examined.

2. **To improve the stability and activity of a plastic degrading enzyme (*Is*PETase) by topology engineering.** The biodegradation of plastic offers a potentially mild and economical route for tackling plastic pollution. The enzyme *Is*PETase has attracted much attention as promising candidate for the recycling of polyesters. However, improvements to the stability and activity of *Is*PETase are required if it is to be used for industrial scale applications. Therefore, *Is*PETase will be cyclised in a range of topologies using SpyTag/SpyCatcher(-N^{TEV}) isopeptide bond forming partners. The stabilities and activities of the linear and cyclic *Is*PETase variants will be compared. CD and residual activity experiments will be used to analyse thermal stability. The effect of cyclisation on the activity of the enzyme for the degradation of a plastic substrate will also be investigated. Various conditions will be employed, including a range of temperatures and agitation speeds.

CHAPTER 2:

CYCLISATION OF AN ARTIFICAL PROTEIN CAGE

2.1 PREFACE

2.1.1 Protein Folding

Catalysis, signalling and transport are some of the essential tasks performed by proteins for the continued functioning of all living systems. Although diverse in the roles they play, each protein carries out a specific function as determined by its shape.[167] The overall three-dimensional structure of a protein is directly dependent on the sequence of amino acid building blocks from which it is composed. This linear polypeptide chain is often referred to as the protein's primary structure. Considering there are 20 different canonical amino acids, which can be combined in any order, any number of times, to make a polypeptide chain of any length, there is an almost infinite number of possibilities for the sequence of the primary structure.[168]

Each type of amino acid bears a unique side chain functional group. Interactions between these functionalities are responsible for guiding the folding of the linear polypeptide chain and holding it in a specific conformation. Intramolecular hydrogen bonding between the backbone amino and carboxyl groups leads to the formation of secondary structures such as α-helices and β-sheets. An α-helix can be characterised as a right-handed helix with 3.6 residues per turn. Meanwhile, a β-sheet consists of two or more polypeptide segments (β-strands) aligned side-by-side with either a parallel, or more commonly, an antiparallel arrangement. Most proteins contain multiple α-helices and/or β-sheets, along with some rarer secondary structure folds, which themselves interact and pack together to generate the more complex protein tertiary structure.[169] Both long- and short-range interactions between amino acid side chain groups contribute towards the protein's three-dimensional geometry including ionic, hydrogen and disulfide bonds. Furthermore, non-polar residues tend to be buried in the hydrophobic core, while more hydrophilic residues are located at the surface.[1] In some cases, further interactions occur between multiple tertiary structures generating a quaternary structure, made up of different polypeptide subunits.

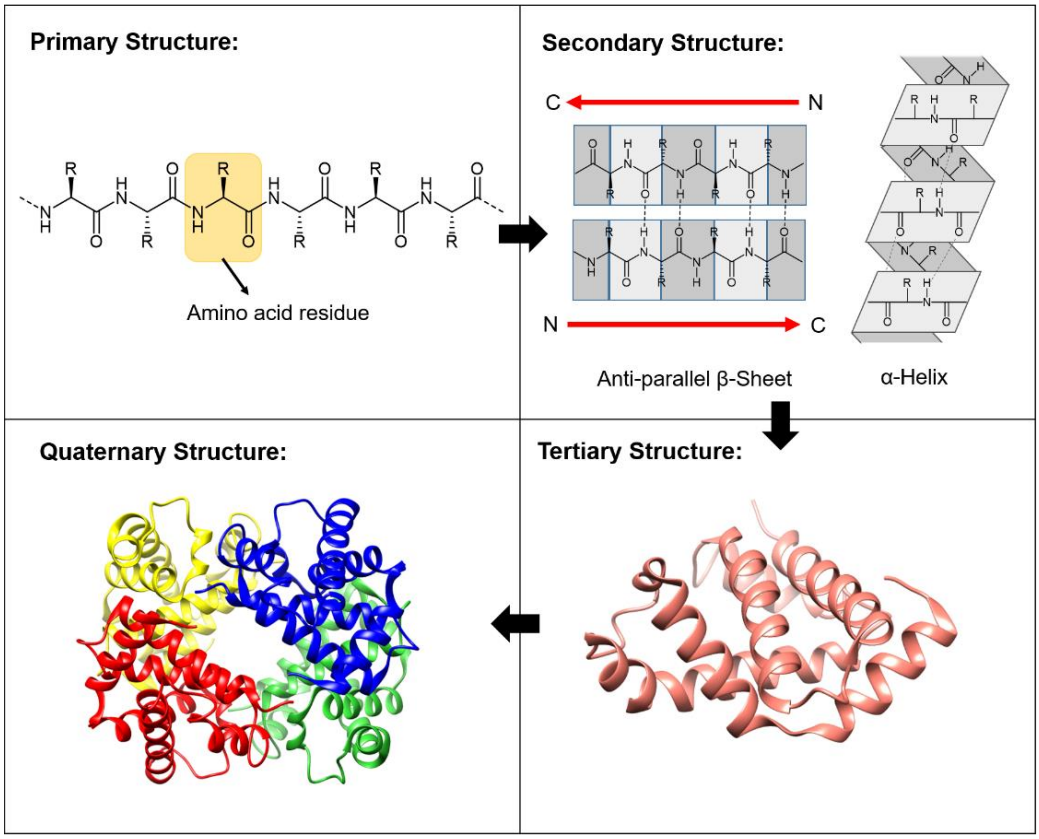


Figure 2.1 The protein structural hierarchy (PDB 4HHB).

The number of primary sequences that are theoretically possible for a polypeptide chain of *n* amino acid residues is enormous (20^n). Accordingly, it would be expected that the number of possible ways a polypeptide chain could fold, would be equally inexhaustible. In reality, evolution has led to the clustering of proteins into families as a result of incremental mutation and selection. Consequently, the proteins within these families are often structurally and functionally similar.[168,170] In short, it is unlikely that all possible protein sequences and folds have been sampled over the course of evolution, and those folds identified from nature may represent only a tiny fraction of those that could exist.[171,172] Consequently, there is much interest in designing new protein folds that have the potential to be tailored for specific roles. Proteins are already attractive materials for many biotechnological and therapeutic applications due to their intrinsic biocompatibility and relatively facile production, as well as providing a variety of chemical functionalities in the form of amino acid side chains. Therefore, the design of new proteins could play an important part in solving many of the challenges with which the world is currently faced.

Although the majority of protein topologies remain unexplored, the design of new folds is challenging. Only a small fraction of sequences code for functional proteins with defined structures. Moreover, predicting the structure of a natural protein based on its amino acid

sequence, without the use of a known structural homolog, is difficult.[173] The advancement of computational protein design has much aided the search for novel protein structures and improved understanding of the sequence-to-structure relationship.[168] Nevertheless, most of the successful protein designs created so far cannot arguably be considered fully *de novo*, as they are based on folds already found in nature, though it is an ingenious method for creating protein structures not observed naturally.[174] Using a modular design approach, complex protein structures can be assembled from simple building blocks, through covalent or non-covalent association.

2.1.2 Coiled Coils

Coiled coils are one of the most widespread protein structure elements found in nature, estimated to be present in around 10% and 5% of all eukaryote and prokaryote proteins, respectively.[175,176] They are commonly found in fibrous structural proteins such as α-keratin, one of the strongest biological materials and a major component of hair, horns and claws.[177] The coiled coil architecture has also been found to participate in many key biological processes, for example the basic leucine zipper (bZIP) domain, a class of transcription factor, uses the coiled coil motif to dimerise two DNA binding regions whose basic residues can then interact with the major groove of the DNA for the regulation of transcription.[178]

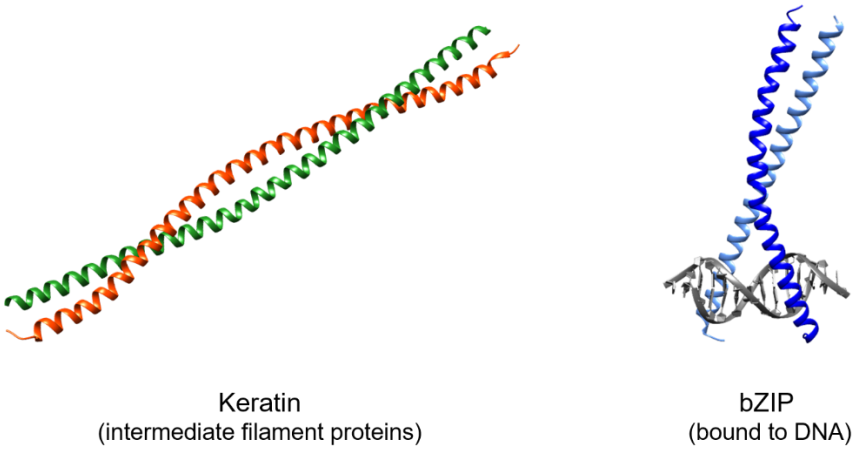


Figure 2.2 Examples of the coiled coil structural motif. (Left) Human intermediate filament proteins keratin 1 and keratin 10 (PDB 6EC0). (Right) Yeast bZIP transcription factor from *Schizosaccharomyces pombe* bound to DNA shown in grey (PDB 1GD2).

Coiled coils can be characterised as two or more amphipathic α-helices wound together with a left-handed superhelical twist. The distinct arrangement of amino acid side chains at the hydrophobic core of the coiled coil bundle is termed “knobs into holes” (KIH) packing, in which a residue from one helix (the knob) packs into a space surrounded by four side chains of the facing helix (the hole).[179,180] This packing arrangement leads to distortion of the component α-helices reducing the number of residues per turn from 3.6 to 3.5. Therefore,

the positioning of side chains repeat every two turns, resulting in a seven residue periodicity, known as a heptad repeat, made up of repeating sequence of hydrophobic (*h*) and polar (*p*) residues, *hpphppp*, conventionally labelled as *abcdefg*. **Figure 2.3** shows the interactions between a parallel and an antiparallel coiled coil dimer. The specificity of the interaction between the two helices arises from a combination of hydrophobic and electrostatic interactions between the residues *a*-*d* and *e*-*g*, respectively. Generally, it is the residues at positions *a* and *d* that influence the oligomeric state of the coiled coil structure, due to differences in the way the hydrophobic side chains are accommodated in the KIH packing arrangement. Meanwhile, the charged residues at *e* and *g* positions contribute towards the coiled coil orientation and whether a homo- or heteromeric assembly is adopted.[181] On the other hand, residues at positions *b*, *c* and *f* do not form interchain interactions as they face away from the dimerisation interface. By modifying these non-interacting residues, the stability of the coiled coil dimer can be tuned without affecting the specificity of the interaction.[182,183] Hence, the coiled coil dimer motif has gained much interest as a simple and versatile building block for the engineering of modular protein structures. Nevertheless, it should be noted that despite being one of the best-understood protein folds, the rules governing the diverse range of possible coiled coil assemblies are by no means straightforward and in some cases are still being elucidated. Some important design rules are detailed in the literature,[181] but from here in, coiled coil dimers will be the focus.

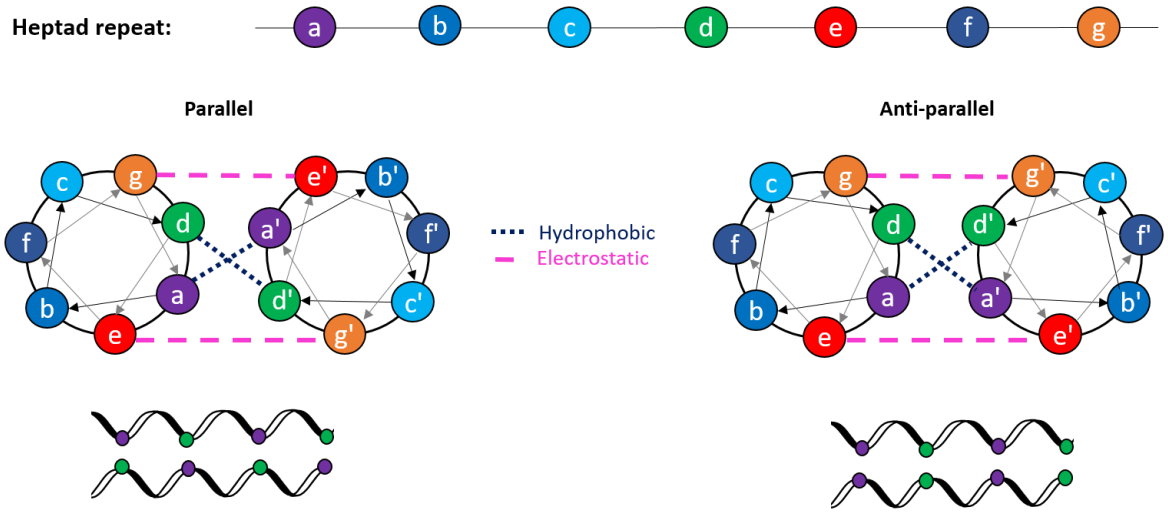


Figure 2.3 Schematic representation of the heptad repeat and coiled-coil structure with either a parallel or antiparallel orientation. Hydrophobic interactions (blue dotted line) occur between residues *a* and *d,* electrostatic interactions (pink dashed line) occur between residues *e* and *g*.

2.1.3 CCPO

The advancement of DNA nanotechnology has allowed the rational design of complex two- and three-dimensional architectures in a reliable and predictable manner.[184] The

straightforward base-pairing complementarity along with the ability to synthesise polynucleotide chains of any length or sequence, makes DNA a versatile molecular scaffold that has potential for application in biosensing and drug delivery systems.[185] As the coiled coil protein motif can also be designed to exhibit pairwise complementarity, it provides an ideal building block for modular protein design through coiled coil protein origami (CCPO). CCPO relies on the simultaneous assembly of multiple coiled coils. As a result, orthogonal coiled coil sets are required which interact specifically with their designated partner and do not cross interact. To achieve this goal, a large free energy difference is required between the desired interaction and that of the most stable off-target association.[183]

The successful design of several orthogonal coiled coil dimer sets, varying in length and orientation, has been achieved through a mixture of rational and computational approaches based on the previously discussed rules governing the pairing specificity and stability of coiled coil interactions. For example, one of the largest sets was created using a predictive algorithm (bCIPA), from which a set of eight parallel heterodimers were constructed from a semi-randomised peptide library.[186] All peptides contained 37 residues, 4 heptad repeats, with the option of Lys or Glu residues at *e* and *g* positions, and either Ile or Asn residues at *a* positions. Along with a fixed Leu amino acid at every *d* position and Ala at other positions, residues were chosen to encourage the formation of coiled coil heterodimers with the desired parallel orientation. Melting temperature (T_m) cut-offs were then employed to increase stringency of the screening. A minimum T_m for desired heterodimers was set to 70 °C, for undesired heterodimers T_m was set to a maximum of 20 °C and a maximum T_m of 10 °C was accepted for homodimers. Experimental characterisation of the coiled coil dimer sets demonstrated the T_m of seven of the eight pairs to be accurate and thus function as designed.

The utility of synthetic orthogonal coiled coil dimer sets for two-dimensional modular protein design was demonstrated by the construction of a tripartite chassis for application as a hub for a Tumbleweed molecular motor (**Figure 2.4**).[187] Six peptides were designed to self-assemble into three coiled coil dimers upon interaction with a specified partner. A single cysteine residue was incorporated at one terminus of each peptide for intermolecular disulfide bond formation between coiled coil dimers, resulting in the desired hub structure.

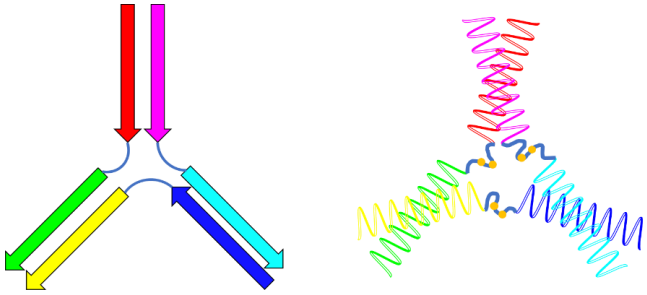
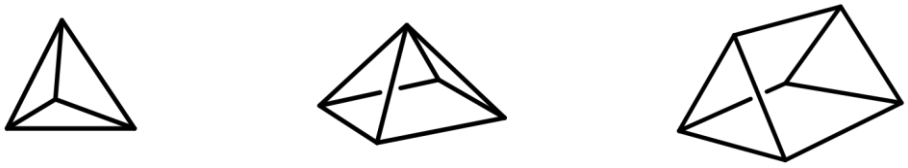


Figure 2.4 Tripartite chassis formed from three sets of coiled coil dimers linked together by disulfide bonds (cysteine residues shown as orange circles).

Nevertheless, the intermolecular assembly of independent coiled coil subunits is limited to symmetrical structures and suffers from concentration dependence, with relatively high concentrations required for the production of stable structures.[188] On the other hand, the self-assembly of a single polypeptide chain, composed of concatenated orthogonal coiled coil segments, allows access to unique topological arrangements. This CCPO approach has led to the design and construction of a number of polyhedron-shaped protein cages, whose folding along a double Eulerian trail is driven by the formation of rigid coiled coil dimers between complementary coiled coil segments placed at appropriate positions along the polypeptide chain.[189] The design of CCPO structures can be automated by encoding design principals into the CCPO design platform (CoCoPOD), which involves the following steps:[183,190]

1. <u>Selection of the target polyhedral structure</u>

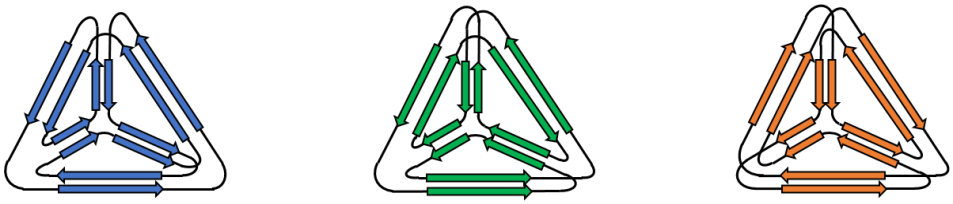
   Coiled coil dimers form the edges of the polyhedral cage. Therefore, until recently, it was assumed that the size and complexity of the cage was limited by the number of available orthogonal building blocks. In fact, the same coiled coil modules can be used multiple times in the same design.[191]



2. <u>Selection of optimal topology</u>

   All possible double Eulerian paths are computed. Multiple paths are possible since the polyhedral topology can be constructed from differing numbers of parallel and antiparallel coiled coil modules. All circular permutations of each topology must also be examined. As Eulerian paths are circular, the path must be made linear for its conversion into an amino acid sequence by an incision at one of the polyhedron

vertices. The circular permutation with the lowest total contact order† is expected to result in smoother folding.[192]



3. Selection of building modules from the orthogonal coiled coil toolbox

   Coiled coil dimer modules differ in stability, charge and length. Through experimental testing several design rules have been uncovered. Firstly, the presence of a SPED N-terminal capping sequence results in the stabilisation of coiled coil dimer modules. Secondly, the positioning of less stable coiled coil partners at the C- and N-terminus should be avoided as they are prone to fraying. Thirdly, variations in the length (5 – 9 residues), charge and polarity of the residues in the flexible linker connecting coiled coil segments have a minor effect on CCPO design success and/or stability. However, small, polar, helix-breaking residues are preferable.

4. Construction of a 3D model and in silico validation of the design

   Modelling enables properties including volume, contact order and solvent-accessible surfaces to be calculated. Folding simulations also allow any folding bottlenecks to be identified.

2.1.4 TET12

Polyhedral CCPO cages that have so far been designed, isolated and characterised, include a tetrahedron, square-based pyramid and a triangular prism.[190] Of these, the simplest and best characterised is the tetrahedron, named TET12. It is composed of 12 coiled coil segments which dimerise to generate the six sides of the polyhedron, while the flexible linker sequences form the vertices (**Figure 2.5**). The internal cavity is approximately 40 nm^3. Although the first generation TET12 CCPO cage was insoluble and required refolding in vitro, destabilisation of the coiled coil dimer modules by the introduction of negatively charged residues at the non-interacting positions (*b*, *c* and *f*), resulted in a

† Contact order is a measure of the proximity of non-adjacent amino acid residues in the protein primary structure which contact in the folded tertiary structure.[192]

supercharged (-47 net negative charge) and soluble second generation TET12 protein cage (TET12SN) which self-assembles under physiological conditions in vivo and in vitro.[190]

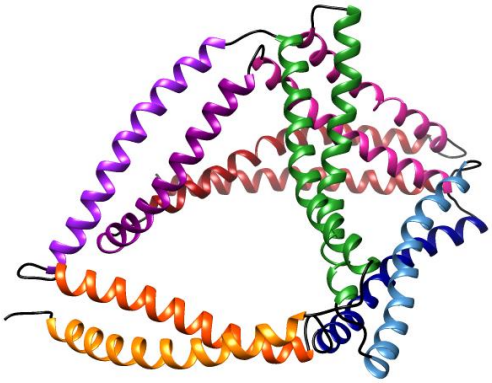


Figure 2.5 TET12 tetrahedral CCPO cage.

Extensive characterisation of the TET12SN CCPO cage was carried out with the correct folding and topology of the protein cage verified by a number of techniques.[189,190] Firstly, the correct pairing of coiled coil modules was confirmed by chemical crosslinking coupled with proteolytic digestion and mass spectrometry. Three different crosslinking reagents were chosen to cover a range of distances relevant for the TET12 protein cage. The crosslinked protein was then subjected to trypsin digest and the mass of the fragments determined. Long range connections detected between non-neighbouring peptide partners indicated that the polypeptide chain folded according to the design.

The polypeptide folding kinetics of the secondary and tertiary structures were determined using stopped-flow circular dichroism and rate of Förster resonance energy transfer (FRET) increase between fluorescently labelled N- and C-termini using cyanine dyes (Cy3 and Cy5), respectively. The secondary structure folding rate (17 s^{-1}) was found to be comparable to that of natural globular proteins of a similar size, despite CCPO cages' lack of hydrophobic core. The tertiary structure folding rate (31 s^{-1}) was also found to be in agreement. Furthermore, the observation of an increased FRET signal, when donor and acceptor chromophores were placed at the polypeptide termini, indicated the correct folding of the TET12 structure as the termini of the self-assembled protein cage are designed to coincide at the same vertex. Hence, FRET should only be observed when termini are in close proximity with one another.

Even though a crystal structure is yet to be obtained for any of the CCPO cages, transmission electron microscopy (TEM) and atomic force microscopy (AFM) characterisation have allowed visualisation of the TET12 structure and provide evidence of

its tetrahedral topology. In both cases, the images obtained showed particles to have a size in agreement with the hydrodynamic diameter of self-assembled TET12 (~7 nm) as determined by dynamic light scattering (DLS) experiments. In addition, particles were observed to have shapes corresponding to different projections of a tetrahedron. Nanogold beads (1.8 nm) coated in Ni-NTA for the attachment to the N-terminal polyhistidine purification tag, enabled location of the terminus of the self-assembled tetrahedron and improved the contrast of the TEM images.

Finally, small angle X-ray scattering (SAXS) analysis confirmed that the structures obtained experimentally were in good agreement with the models created with CoCoPOD. Normalised Kratky plots acquired from the SAXS measurements (often employed as an indicator of the presence of protein flexibility)[193] indicated partial flexibility of the CCPO structure. As the coiled coil elements are rigid, the conformational changes observed, likely arise from the flexible linkers that make up the vertices of the polyhedral cage.

2.1.5 Potential Applications of CCPO Cages

Protein cages are commonly found in nature where they function as containers for the encapsulation of guest molecules such as metal ions (e.g. ferritin), nucleic acids (e.g. viral capsids) or other proteins (e.g. encapsulins).[194] As a result, they have attracted much interest as functional entities for various applications such as drug delivery or catalysis, though much of this research remains in its early stages.[195–197] It should be noted that CCPO protein cages differ from those found in nature.

1. CCPO cages are non-symmetrical, in contrast to natural cages which often adopt spherical or rod-like structures.
2. The internal cavities of CCPO cages are hydrophilic. This is because the hydrophobic residues are located at the coiled coil dimer interface which make up the edges of the cage.
3. The CCPO cavity is relatively exposed, whereas protein cages found in nature rely on pores for the entry and exit of molecules.[198]

Nevertheless, CCPO cages are versatile and customisable, making them appealing structures for a diverse range of applications. In addition, the TET12SN CCPO protein cage, has been demonstrated to successfully self-assemble in vivo in both mammalian cells and living animals without the observation of adverse side effects.[190] As such, CCPO cages could in future be employed as delivery systems for therapeutic drugs, photosensitizers or cytotoxic enzymes. Alternatively, the high customisability of CCPO protein cages makes

them appealing for use as reaction vessels. Indeed, improved efficiency and selectivity of reactions is often observed when carried out in confined spaces as it enables control over stoichiometry and the spatial arrangement of molecules.[199,200]

In summary, CCPO protein cages not only offer the desirable properties that are intrinsic to polypeptides, such as self-assembly, biocompatibility and programmability, but also provide an internal cavity which has the potential to be exploited for a variety of therapeutic and biotechnological applications. For many of these applications, an improved stability would be beneficial, enabling their use outside of physiological conditions. In this chapter, the effect of cyclisation on the tetrahedral CCPO cage TET12SN will be investigated, with the aim of improving the tolerance of the artificial protein cage towards heat, chemical and/or proteolytic treatment. In addition, TET12SN would be the first artificial protein to be cyclised (not including synthetic peptides), and could provide relevant for an improved understanding of protein folding and stabilisation in general.

2.2 RESULTS

2.2.1 TET12SN Preparation

The CCPO nomenclature signifies that TET$12_{1.10}$SN-c_6 is a tetrahedral polyhedron (TET) composed of 12 coiled coil modules. The charged linkers (c) joining these soluble supernegatively charged (SN) modules are 6 amino acid residues in length. The topology and circular permutation of the cage denoted is by subscript (1.10).[190]

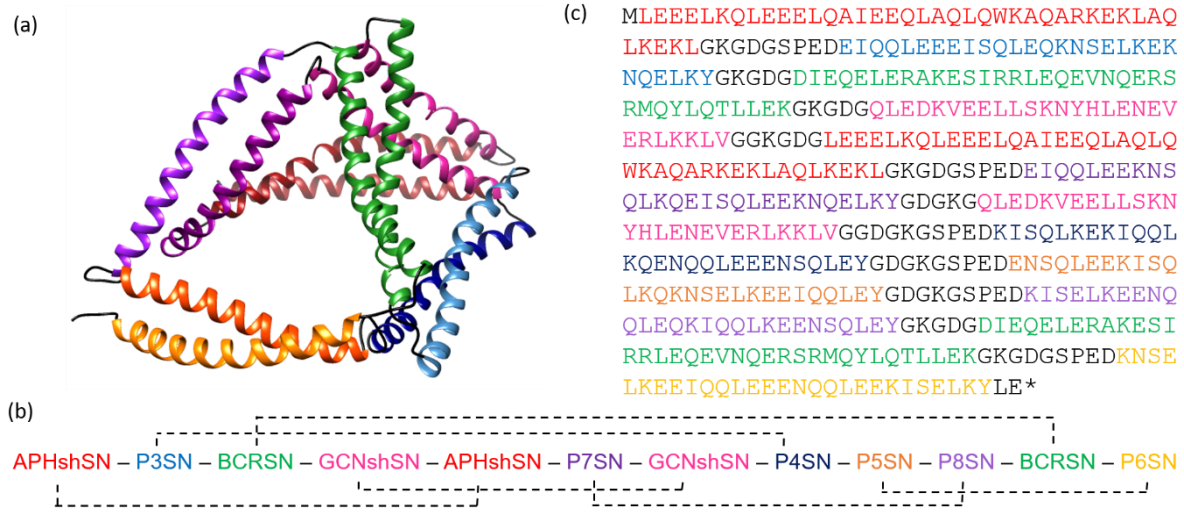


Figure 2.6 (a) The computation model of the tetrahedral CCPO cage TET12SN obtained from the Jerala group. (b) The constituent polypeptide chain of TET12SN composed of concatenated coiled coil fragments joined by short linker sequences. The dashed lines indicate the coiled coil dimer pairs. (c) The amino acid sequence of TET12SN. Amino acids are coloured according to coiled coil sequence they belong, linkers and capping sequences are shown in black.

The plasmid containing the gene for TET12$_{1.10}$SN-c$_6$ (**Figure 2.7**, Appendix 7.1.1), herein referred to as TET12SN, was already available in the laboratory. In the pET-28a(+) plasmid, the TET12SN sequence is located between the T7 promoter and T7 terminator regions, placing its expression under the control of T7 RNA polymerase.

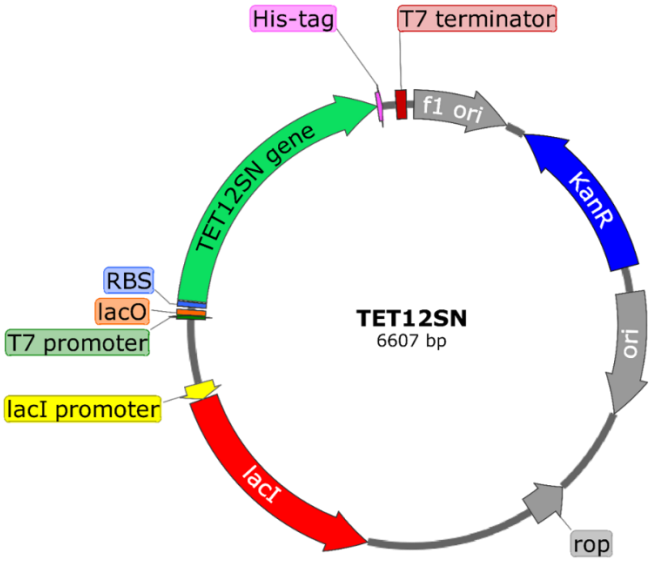


Figure 2.7 The TET12SN pET28a(+) plasmid, showing the lac repressor (*lacI*) gene and promoter, the T7 promoter, the lac operator (*lacO*), the ribosome binding site (RBS), the gene of interest (TET12SN gene), the T7 terminator, the f1 bacteriophage origin of replication (f1 ori), kanamycin resistance gene (KanR), the origin of replication (ori) and the repressor of primer gene (rop).

For the expression of TET12SN, the BL21(DE3) *Escherichia coli* (*E. coli*) strain was employed. After transformation, the cells were grown to an OD_{600} of 0.5-0.6 at 37 °C in LB media. Expression was then induced by adding 0.5 mM ITPG. The culture was grown for an additional 4 hours at 30 °C before the cell pellet was harvested by centrifugation. Next the TET12SN protein product was purified using nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography and analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (**Figure 2.8**).

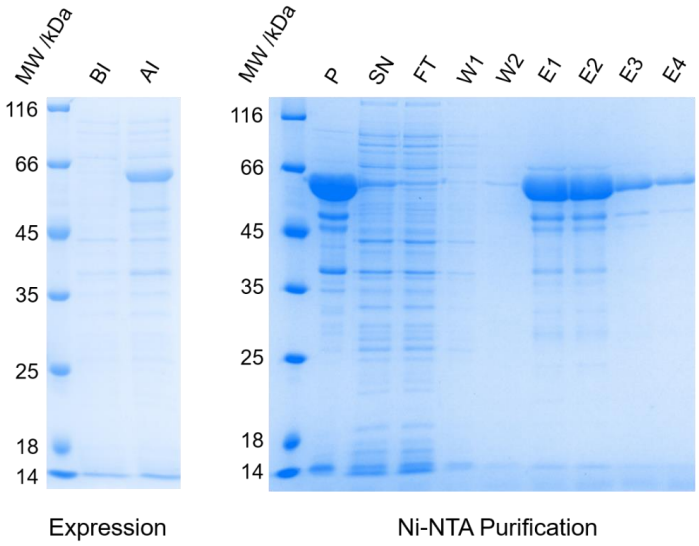


Figure 2.8 (Left) SDS-PAGE of TET12SN expression showing samples taken before and after induction with IPTG (BI and AI, respectively). (Right) SDS-PAGE showing samples taken during TET12SN Ni-NTA purification. P = pellet (i.e. insoluble fraction after centrifugation); SN = supernatant (i.e. soluble fraction after centrifugation); FT = flow through; W = wash; E = elution.

By SDS-PAGE, a primary band was observed at just under 66 kDa. This corresponds to the expected molecular weight for TET12SN (54.3 kDa). In addition, some lower molecular weight contaminants are observed in the eluted fractions (E1-4). These residual impurities were removed by further purification using size exclusion chromatography (SEC). Absorbance at 280 nm was used to track the elution of the TET12SN monomer. The chromatogram is shown in **Figure 2.9** along with the SDS-PAGE analysis for the SEC purification step.

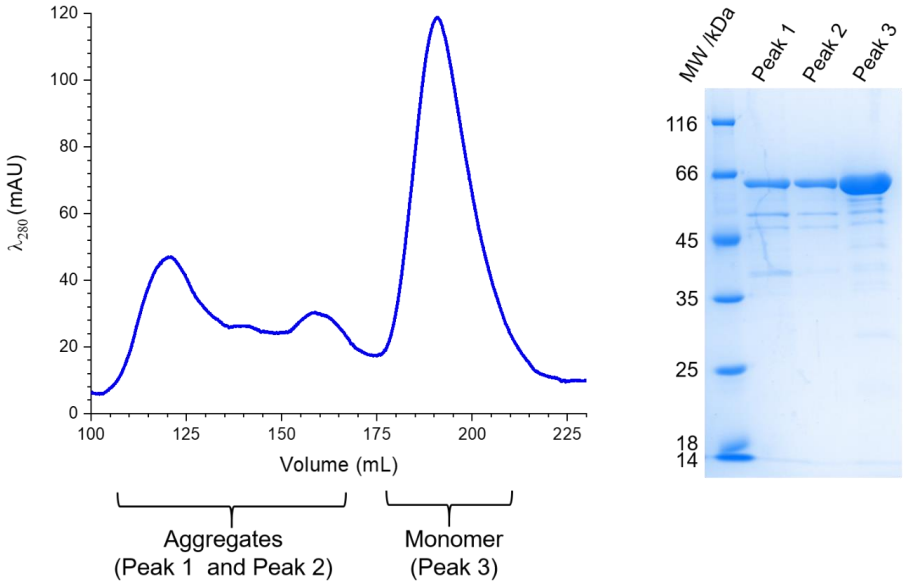


Figure 2.9 (Left) SEC chromatogram of TET12SN. (Right) SDS-PAGE analysis of TET12SN SEC purification.

The SEC chromatogram shows one major peak (Peak 3), as well as two smaller peaks (Peak 1 and 2). SDS-PAGE analysis of fractions belonging to each of these peaks showed the presence of a protein corresponding to the molecular weight of TET12SN. As proteins eluted later are smaller in size, Peak 3 most likely corresponds to the TET12SN monomer, while Peaks 1 and 2 are likely to be non-covalent TET12SN dimers or aggregates resulting from intermolecular coiled coil dimerisations. This contrasts with the published purification of TET12SN, which found the protein to be highly monodisperse with up to 99% eluted in the monomeric fraction.[190] Here, the increased aggregation observed could have resulted from differences in the expression and purification procedures employed. For example, the use of a different expression vector (pET-41a), a different expression strain (NiCo21(DE3)) and a N-terminal His-tag, may all have contributed towards the improved production, solubility and folding of TET12SN reported in the literature. Nevertheless, the fractions containing the eluted TET12SN monomer were collected and concentrated to give an overall yield of 5-10 mg/L. Mass spectrometry (MS) was employed for further analysis of the sample (**Figure 2.10**), with a protein of the expected molecular weight of 54346 Da observed (including the start Met and His-tag). A mass of 54374 Da was also present, 28 Da higher than that expected. This could correspond to TET12SN with a N-formylated Met (fMet), which contains an additional C=O attached to the Met amino group.

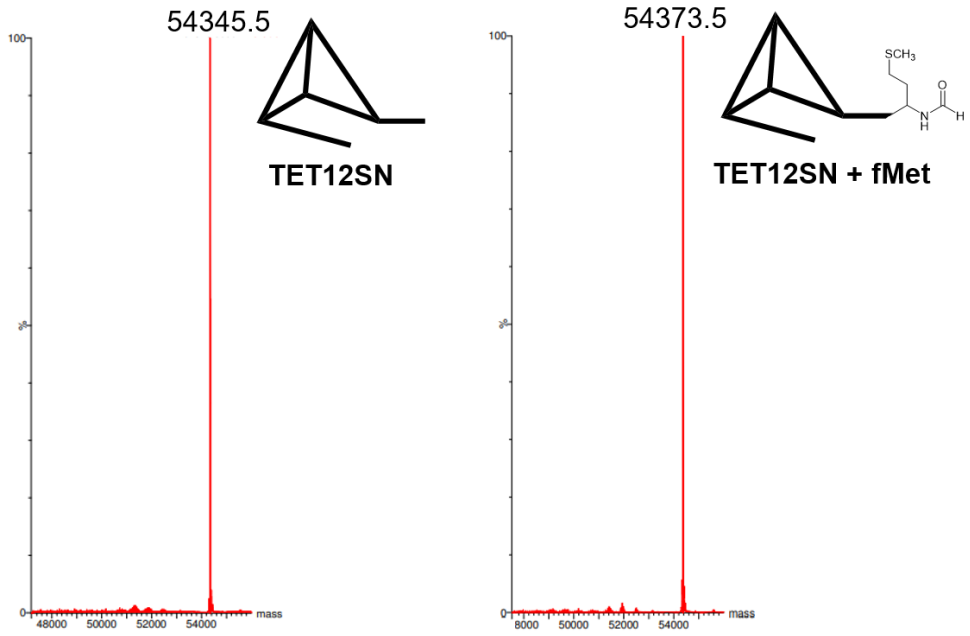
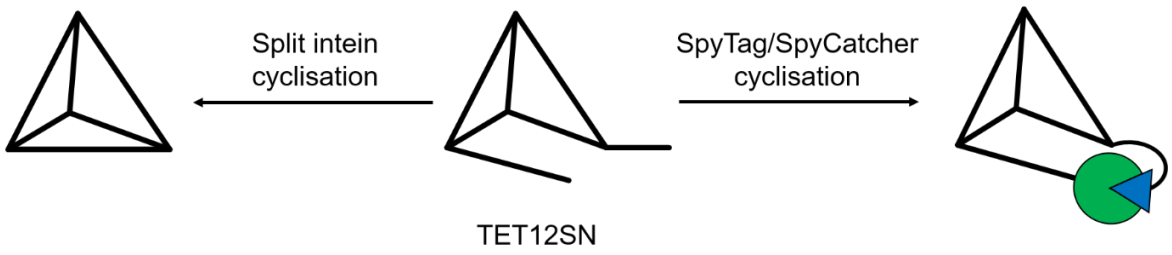


Figure 2.10 Deconvoluted ESI-MS of purified TET12SN (with and without fMet).

2.2.2 TET12SN Cyclisation Approaches

For the cyclisation of TET12SN, two different approaches were employed (**Scheme 2.11**). The first was split intein mediated cyclisation (see section 1.2.3.1.2), whereby the C- and

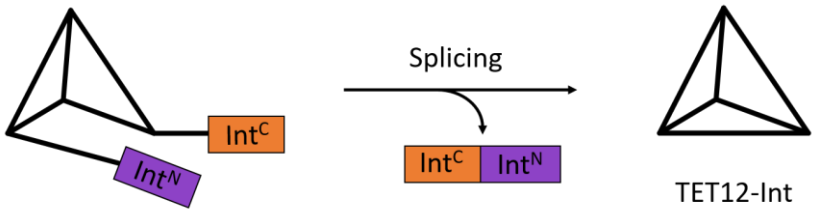
N-termini are directly linked through the formation of a native peptide bond, resulting head-to-tail cyclisation. As only a cysteine residue from the intein sequence remains in the cyclised product, this approach can be regarded as traceless. In the second approach, using SpyTag/SpyCatcher-N^{TEV} (see section 1.2.3.2), an isopeptide bond is formed between the side chain groups of an asparate and a lysine amino acid residue on SpyTag and SpyCatcher-N^{TEV}, respectively. Both SpyTag and SpyCatcher-N^{TEV} sequences remain in the cyclised product resulting in a large scar at the ligation site.



Scheme 2.11 Cyclisation of TET12SN.

2.2.2.1 Split Intein Cyclisation

In the first TET12SN cyclisation approach, the *Npu* DnaE split intein was employed due to its efficient splicing activity and tolerance towards variation in the extein sequence. The Int^C and Int^N fragments were positioned at the N- and C-termini of the TET12SN sequence, respectively (**Scheme 2.12**, Appendix 7.1.1 and 7.2.1). The split intein cyclised TET12SN product will herein be referred to as TET12-Int.



Scheme 2.12 The cyclisation of TET12SN by split intein producing TET12-Int.

The plasmid for TET12-Int was already available in the laboratory. It was expressed and purified as described for TET12SN. As the splicing of the *Npu DnaE* split intein should be rapid and spontaneous, the spliced product should be produced directly after expression without the need for additional incubation time before or after purification. The SDS-PAGE analyses for the purification of TET12-Int are shown in **Figure 2.13**.

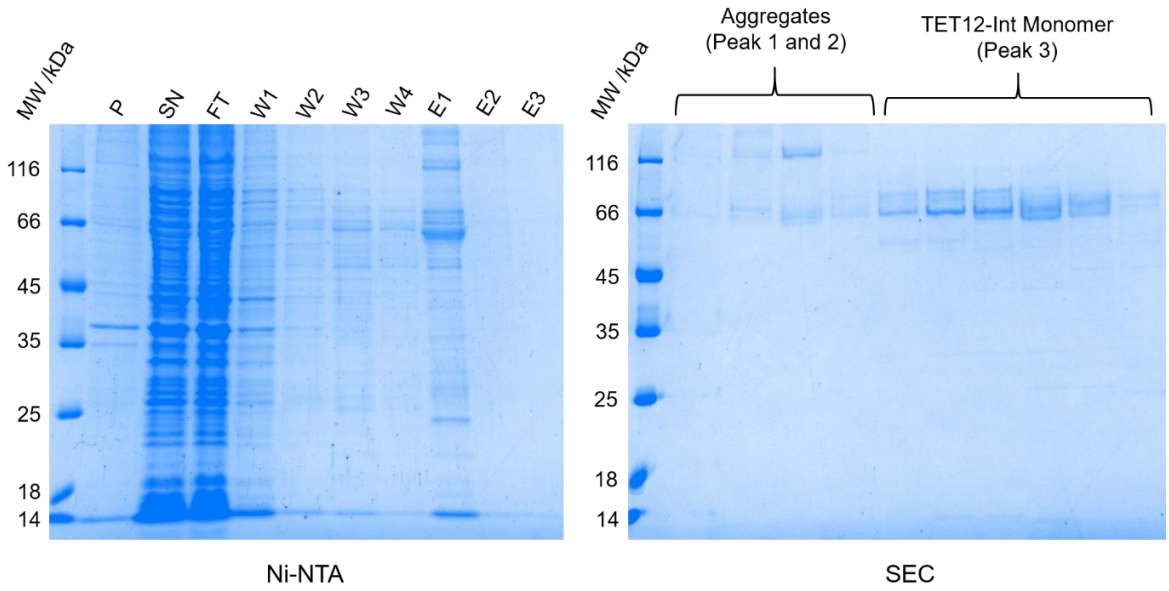


Figure 2.13 SDS-PAGE analysis of TET12-Int purification by Ni-NTA and SEC.

The expected molecular weight for TET12-Int is 55988 Da. A band is observed at around 66 kDa, higher than expected for this protein. However, cyclised proteins commonly have increased or reduced mobility through the polyacrylamide gel compared to their linear counterparts, and hence appear at a higher or lower position than would be predicted by their molecular weight.[77,135,143,144] MS confirmed that a protein with a molecular weight of 55985 Da was present, which corresponds to TET12-Int (**Figure 2.14**). However, some impurities with a similar molecular weight were also observed, perhaps resulting from the post-translational modification of TET12-Int. This could lead to the poor resolution of the SDS-PAGE band, appearing to be composed of multiple thinner bands. Nevertheless, similar observations have been reported for some other cyclised proteins and speculated to arise from either inefficient cyclisation or conformation of the cyclised protein leading to altered electrophoretic mobility.[160,161] Subsequently, the possibility of inefficient splicing was investigated.

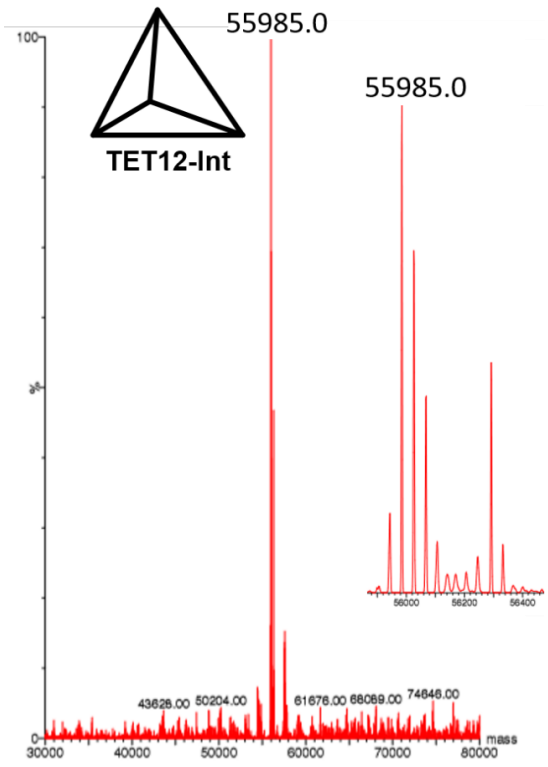
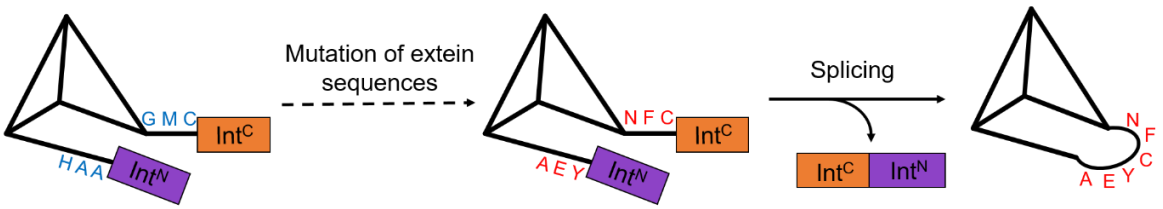


Figure 2.14 Deconvoluted ESI-MS of purified TET12-Int.

While the *Npu* DnaE split intein is known to be relatively tolerant towards the identity of the amino acids in the extein sequence either side of the ligation site (**Scheme 2.15**), deviation from the native extein sequence ($C_{+1}F_{+2}N_{+3}/A_{-3}E_{-2}Y_{-1}$)‡ can still lead to reduced splicing efficiency. Of particular importance are Cys_{+1}, which is a vital nucleophile in the transthioesterification step of the intein splicing mechanism, and Phe_{+2}, whose bulky side chain helps to maintain the intein in a favourable conformation for splicing.[131] In the TET12-Int construct, the amino acid residues $C_{+1}M_{+2}G_{+3}/H_{-3}A_{-2}A_{-1}$ are positioned either side of the ligation site, differing from those in the native sequence. Therefore, in an attempt to improve splicing efficiency, these residues were replaced by those of the native sequence using site directed mutagenesis (SDM), producing TET12-Int^{Nat}.



Scheme 2.15 Diagram of TET12-Int showing the flanking extein sequences.

‡ For consistency with the literature, extein amino acid residues are labelled with respect to their position either side of the intein splicing site (/). The subscript number corresponds to the distance from the splicing site, while the "+" and "-" symbols denote residues at the N- and C-termini of the splicing site, respectively. On the other hand, amino acids found in the intein sequence are numbered separately from those in the extein sequence, without the use of subscript.

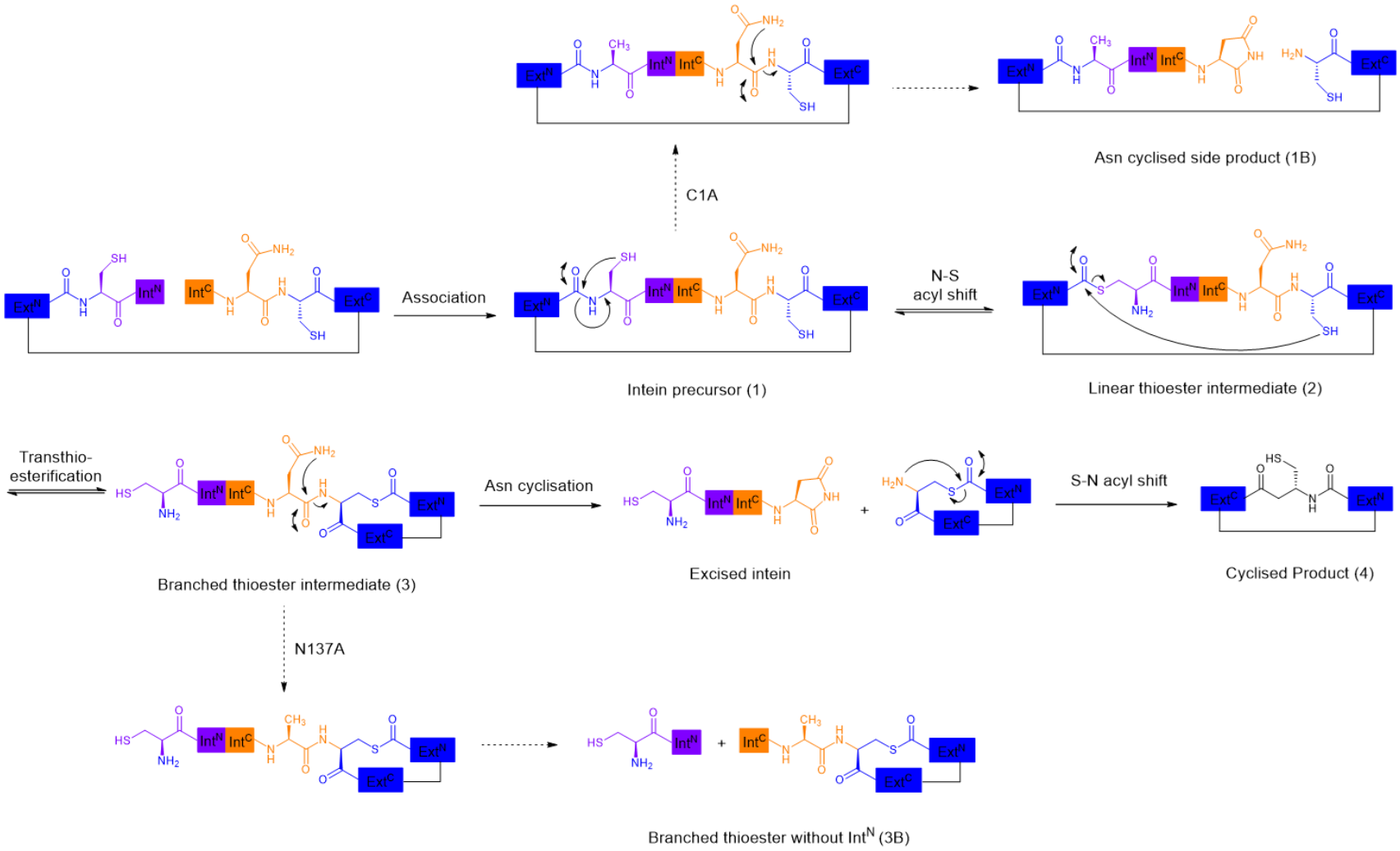
Unfortunately, the presence of the native extein sequence did not improve the resolution of the band observed by SDS-PAGE analysis after purification. It should be noted that both native and non-native flanking extein sequences appear to lead to complete splicing, with no intermediates or unspliced precursor observed by MS. However, Met (and Trp) has been shown to be a suitable residue to take the place of Phe at the +2 position allowing for rapid splicing even in the absence of the native extein sequence.[201]

As mentioned above, inefficient splicing would lead to the observation of unspliced precursor, intermediates and side products. This was demonstrated by creating various TET12-Int mutants using SDM, in which some important amino acid residues involved in the splicing mechanism were mutated. In this way, the linear precursor, thioester intermediates and/or side products would be trapped. Briefly, the contributions of each mutated amino acid residue towards the splicing mechanism (**Scheme 2.16**) will be discussed.

1. C1A: Mutation of Cys to Ala in the Int^N fragment would prevent the formation of the linear thioester intermediate (**2**) as N-S acyl shift would not take place and the linear precursor (**1**) would remain.

2. N137A: Upon cyclisation of the Asn residue in Int^C, the intein is excised and the remaining extein fragment(s) rearrange via S-N acyl shift to form the spliced product (**4**). Replacement of this Asn with an Ala residue would trap the branched thioester intermediate (**3**) so excision would not take place. As a result, the thioester intermediate would be observed.

3. $M_{+2}A$ and H125N: Both the presence of a bulky hydrophobic amino acid at the +2 position of the Ext^C and the His located in the Int^C sequence are important for branched thioester resolution, in which Asn cyclisation leads to excision of the intein.[131,202] Mutation of either of these residues would slow the Asn cyclisation step. As a result, the branched thioester intermediate (**3**) would be observed.



Scheme 2.16 Split intein splicing mechanism (solid arrows). The production of side products as a result of incomplete splicing and the mutations responsible are also shown (dashed arrows).

Figure 2.17 compares the TET12-Int mutants by SDS-PAGE. The unspliced linear TET12-Int has an expected molecular weight of around 72 kDa, the branched thioester intermediate also of 72 kDa, and spliced product of 56 kDa. TET12SN is shown for reference (~54 kDa). MS was employed for further analysis of each sample (**Table 2.1** and Appendix 7.3.1).

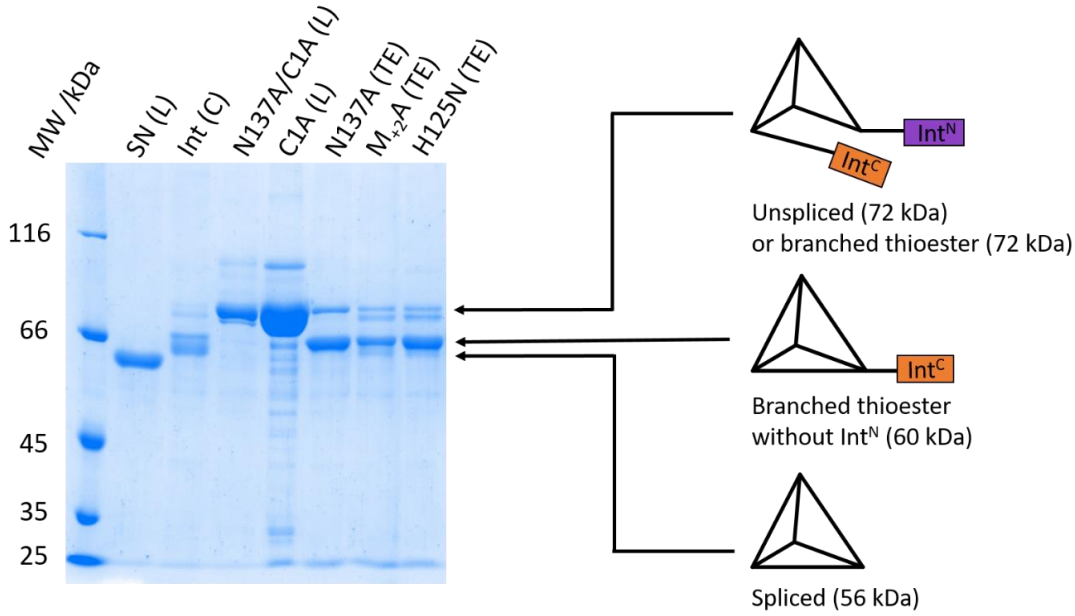


Figure 2.17 SDS-PAGE comparison of the TET12-Int mutants. Their proposed structures are indicated including the linear precursor (L), thioester intermediates (TE) and spliced product (C).

For the C1A mutant, a band at just above 66 kDa is observed on the SDS-PAGE and a molecular weight of 71918 Da confirmed by MS. This corresponds to the linear precursor (**1**) as expected. A side product (**1B**) in which the linear precursor undergoes premature Asn cyclisation may also be observed. However, due to the intramolecular nature of this splicing reaction, the Ext^C fragment would remain attached to the rest of the molecule and so would not be distinguishable from the linear precursor by mass.

For the N137A mutant, the branched thioester intermediate (**3**) was expected (71909 kDa). However, two bands appear on the SDS-PAGE, one above and one below the 66 kDa marker. Analysis of the sample by MS showed these to arise from a protein of 71926 Da and another of 60053 Da. In the branched thioester intermediate, the Int^N fragment is no longer covalently bound to the rest of the molecule. Therefore, when Asn cyclisation is inhibited, the Int^N (11856 Da) can dissociate leaving a 60 kDa fragment (**3B**) consisting of the extein sequence and Int^C. As a result, the two bands observed are the branched thioester intermediate with and without the Int^N fragment.

In an additional mutant, the C1A and N137A mutations were combined, preventing both N-S acyl shift and Asn cyclisation steps from taking place and trapping the unspliced linear

precursor (**1**). By SDS-PAGE only one band is observed for this mutant at just above 66 kDa, which was shown to have a molecular weight of 71876 Da, matching that expected for the linear precursor (71877 Da).

Both the $M_{+2}A$ and H125N mutants show three bands on the SDS-PAGE, two closely spaced above the 66 kDa marker and one larger band below. Similar to the N137A mutant, the lower band is the branched thioester side product without the Int^N fragment (**3B**). This was confirmed by MS of both the $M_{+2}A$ and the H125N mutants, with molecular weights of 60049 Da and 60083 Da, respectively. Proteins of 71909 Da and 71947 Da were found to be present in the $M_{+2}A$ and the H125N samples, respectively. These molecular weights correspond to those of the branched thioester intermediate expected, and are responsible for one of the two higher molecular weight bands on the SDS-PAGE. The other higher molecular weight band could also correspond to a protein of the same molecular weight. Both the linear precursor (**1**) and the linear thioester intermediate (**2**) would be undistinguishable from the branched thioester intermediate (**3**) by molecular weight but their mobility through the polyacrylamide gel could vary slightly leading to the double banding observed. It should also be noted that Asn cyclisation is not completely inhibited in the $M_{+2}A$ and H125N mutants, allowing the formation of spliced product (**4**). However, no spliced product is observed here as a longer incubation time may be needed for splicing of the sub-optimal extein sequence.

Table 2.1 Expected and observed molecular weights for TET12-Int variants. * Molecular weight includes the start Met.

Mutation	Expected Product(s)	Expected Masses /Da	Observed	Observed Masses /Da
None (TET12-Int)	Spliced product (4)	55988	✓	55985
CFN/AEY (TET12-Int^{Nat})	Spliced product (4)	56353	✓	56349 and 56351
C1A (in Int^N)	Intein precursor (1) Cyclised Asn side product (1B)	71920* 71920*	✓ ✓	71918
N137A (in Int^C)	Thioester intermediate (3) Thioester intermediate (without Int^N) (3B)	71909* 60056*	✓ ✓	71926 60053
$M_{+2}A$ (in Ext^C)	Thioester intermediate (3) Thioester intermediate (without Int^N) (3B) Spliced product (4)	71892* 60054* 55928	✓ ✓ x	71909 60049
H125N (in Int^C)	Thioester intermediate (3) Thioester intermediate (Int^N) (3B) Spliced product (4)	71929* 60090* 55988	✓ ✓ x	71947 60083
C1A and N137A	Intein precursor (1)	71877*	✓	71876

As the creation of TET-Int^{Nat} did not improve the band resolution on SDS-PAGE, and no unspliced linear precursor, thioester intermediates or side products were observed, it is unlikely that incomplete or inefficient splicing were responsible for the smeary banding observed on the gel of TET12-Int. Therefore, it was concluded that cyclisation leading to the altered mobility of the protein through the polyacrylamide gel was more likely to be the cause of this observation.

Finally, the yield of TET12-Int after purification (1 mg/L) was considerably reduced compared to that of TET12SN. The SEC chromatogram and SDS-PAGE analysis suggest that the increased aggregation of TET12-Int could be a possible reason for the low yield. **Figure 2.18** shows the chromatogram of the TET12-Int SEC purification, where the peak that likely corresponds to the elution of the monomer (Peak 3), is significantly reduced in size, compared to that of the TET12SN monomer. Instead, Peak 1 is the major peak. Analysis of the SEC fractions by SDS-PAGE showed that a protein of the molecular weight expected for TET12-Int had been eluted, as well as a protein of a higher molecular weight (>116 kDa). Unlike TET12SN aggregates, in which individual polypeptide chains can only interact non-covalently and can be separated during the denaturation step of SDS-PAGE sample preparation, TET12-Int molecules are able to dimerise (and oligomerise) covalently by intermolecular splicing. As a result, the individual polypeptide chains remain connected after denaturation and appear as a band at a higher molecular weight on the gel.

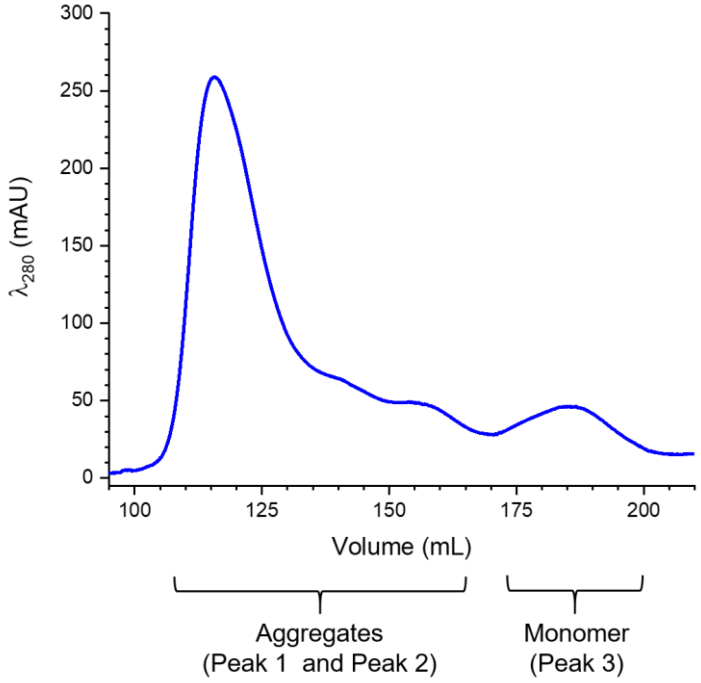
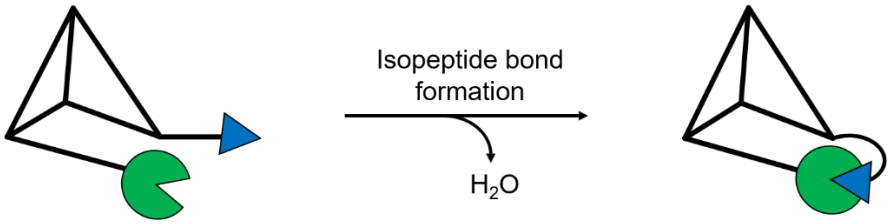


Figure 2.18 SEC chromatogram of TET12-Int.

2.2.2.2 SpyTag/SpyCatcher-N^{TEV} Cyclisation

The cyclisation of TET12SN was also carried out using SpyTag/SpyCatcher-N^{TEV} isopeptide bond forming peptide and protein partners (**Scheme 2.19**). In this TET12-Spy construct, the SpyTag peptide sequence was placed at the C-terminus, while the SpyCatcher-N^{TEV} protein fragment was located at the N-terminus of the TET12SN sequence.



Scheme 2.19 The cyclisation of TET12SN by SpyTag/SpyCatcher-N^{TEV} to generate TET12-Spy.

The TET12-Spy gene (Appendix 7.1.1) was purchased and cloned into a pET-28a(+) vector using Gibson Assembly. The resulting TET12-Spy plasmid was then transformed into chemical competent *E. coli* strain BL21(DE3), expressed and purified (**Figure 2.20**) as described for TET12SN.

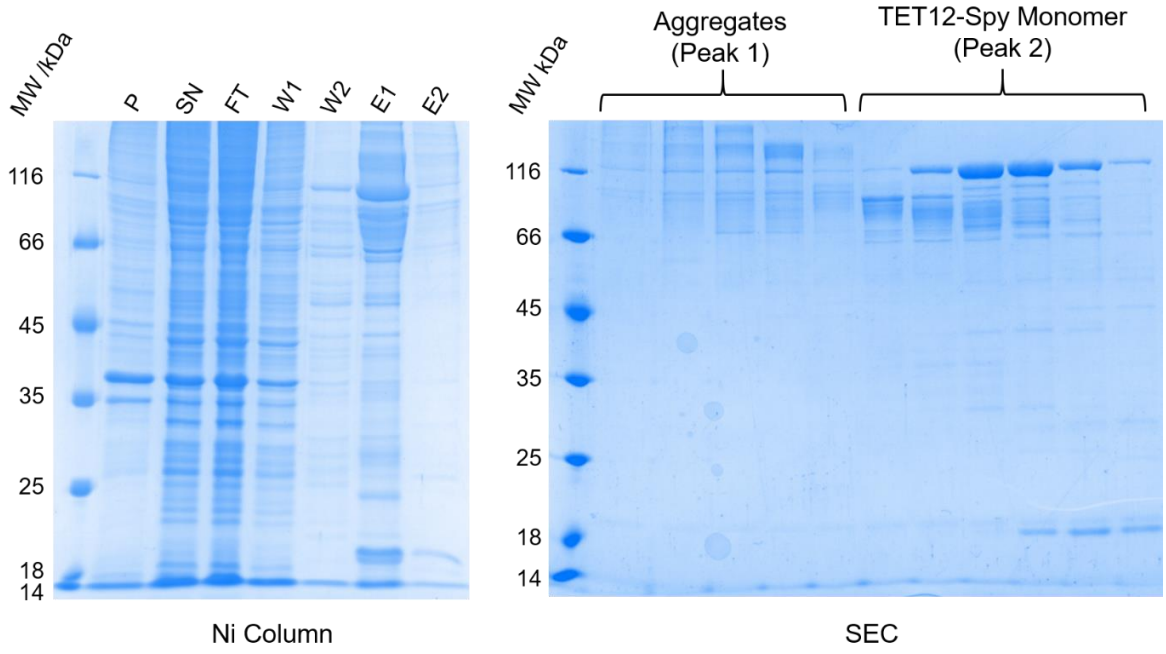


Figure 2.20 SDS-PAGE of TET12-Spy purification by Ni-NTA and SEC.

The SDS-PAGE analysis of the TET12-Spy purification showed similarities to that of TET12-Int. In particular, the primary band corresponding to the cyclised protein is observed at a higher molecular weight than expected (67896 Da). In addition, multiple bands can be seen below this main band resulting in a smeary appearance. Both these observations indicate

cyclisation has taken place, since the conformation of cyclic proteins can lead to their altered mobility through the polyacrylamide gel. MS confirmed the protein to have a molecular weight of 67895 Da (**Figure 2.21**). Some impurities were observed by MS, although at lower molecular weight (around 28.8 and 37.4 Da) than the extra banding on the gel.

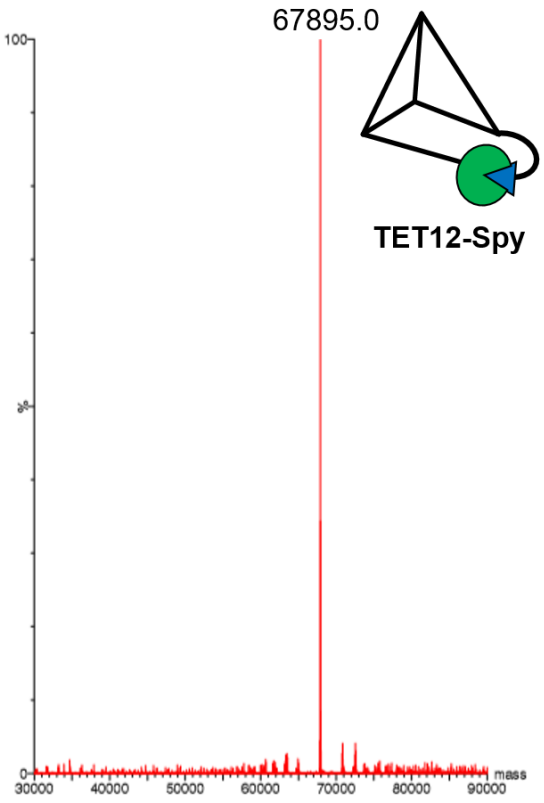


Figure 2.21 Deconvoluted ESI-MS of purified TET12-Spy.

Unlike TET12-Int where a large size difference is observed upon cyclisation, the mass difference between linear and cyclic TET12-Spy is only 18 Da. This corresponds to the loss of one water molecule upon isopeptide bond formation between the SpyTag and SpyCatcher-N^{TEV} sequences which remain in the cyclised molecule. For comparison, the SDS-PAGE analysis of linear and cyclic TET12-Spy are shown in **Figure 2.22**. The linear variant was generated by mutation of the SpyTag Asp amino acid residue (D7A) involved in isopeptide bond formation. Noticeably, the linear TET12-Spy^{D7A} is observed at its expected molecular weight on the SDS-PAGE (~66 kDa), in contrast to the cyclised protein. In addition, the SEC chromatograms of the linear and cyclised TET12-Spy variants are shown. The cyclised protein exhibits increased amounts of aggregation in comparison to its linear counterpart, which is mainly eluted as a monomer. This increased tendency of the cyclised proteins to aggregate may result from disruption to the CCPO folding pathway and/or increased oligomeristion. This will be discussed in more detail later in this chapter (see section 2.3.1.1).

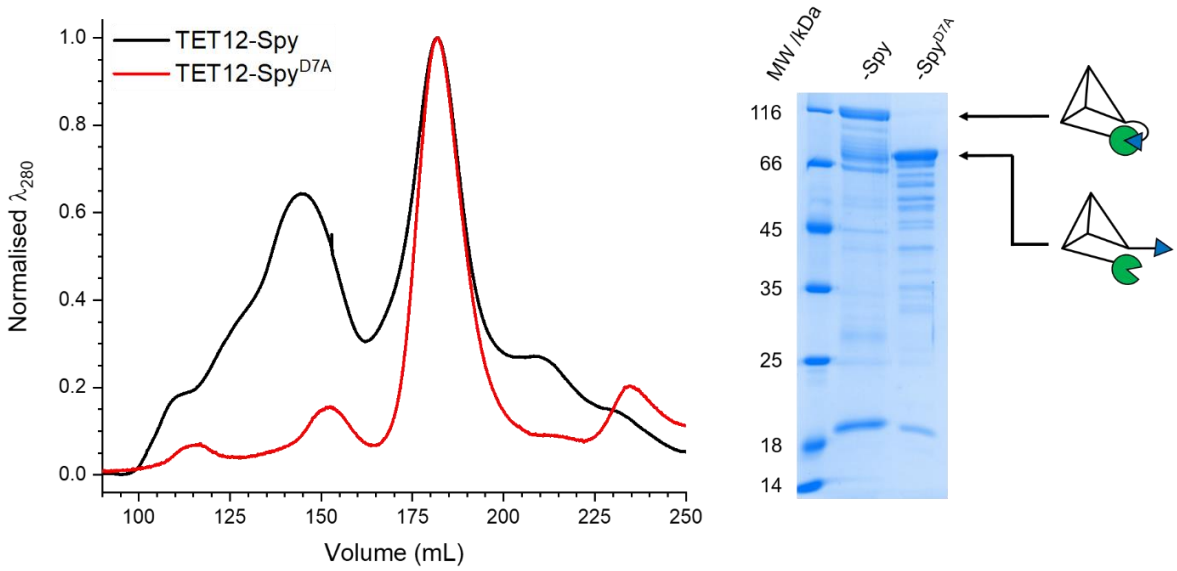


Figure 2.22 Comparison of linear and cyclic TET12-Spy SEC chromatograms and SDS-PAGE.

For the cyclisation of TET12SN, the SpyCatcher-N^{TEV} variant was employed so that the size of the scar left at the ligation site could be reduced after isopeptide bond formation had taken place (see section 1.2.3.2, **Figure 1.29**). The SpyCatcher-N^{TEV} variant is circularly permutated and contains a TEV protease recognition sequence (ENLYFQ/G). Upon protease treatment, around 70 amino acids are cleaved from the SpyCatcher fragment, leading to a reduction in size of about 7.5 kDa.

Various conditions were tested for the TEV protease cleavage of the SpyCatcher-N^{TEV} fragment. The best conditions were found to require incubation of TET12-Spy with TEV protease (1:4.5 molar ratio) overnight at room temperature in the buffer used for SEC. **Figure 2.23** shows the SDS-PAGE comparing the TET12-Spy sample before and after TEV protease treatment. Control samples, in which TEV protease and TET12-Spy were incubated separately under the same conditions, are also shown. In the SDS-PAGE analysis of TET12-Spy after overnight TEV protease treatment, some uncleaved protein was detected although an excess of TEV protease was used. This is indicated by the two bands present at around 116 kDa, the higher and lower bands being the uncleaved and cleaved samples, respectively.

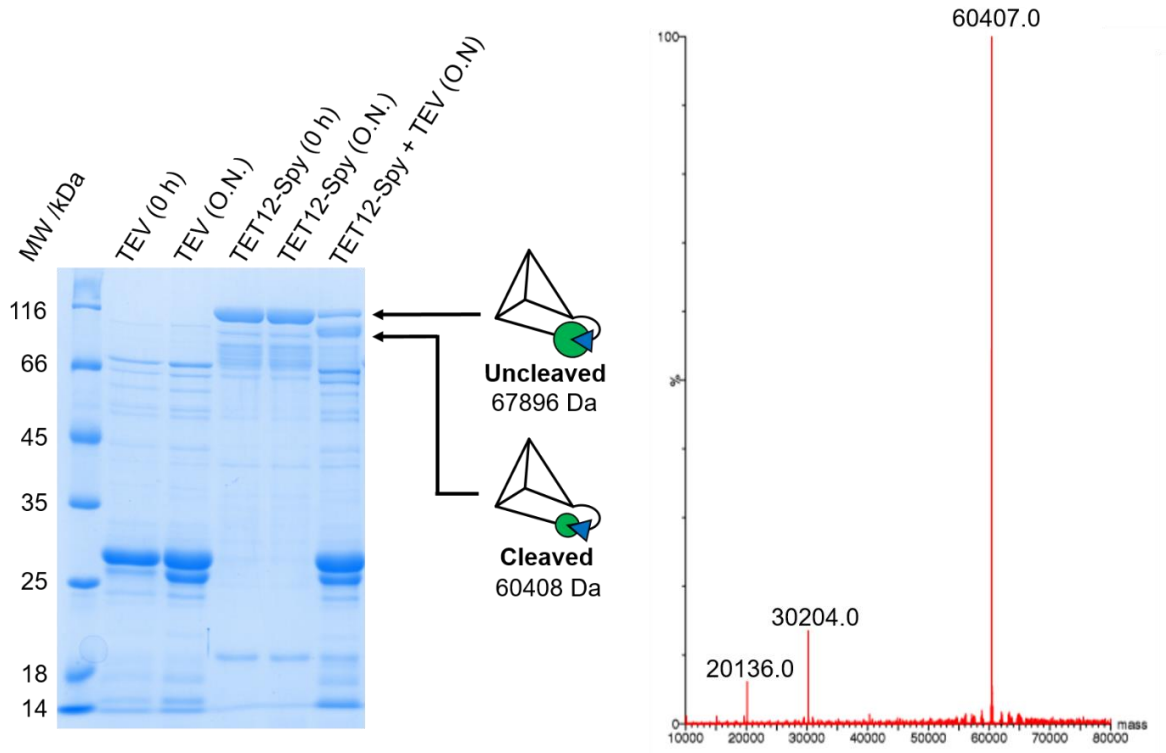


Figure 2.23 SDS-PAGE and ESI-MS analyses of TEV protease (TEVp) cleavage of the TET12-Spy N^{TEV} fragment. Cleavage was carried out at room temperature in 20 mM Tris (pH 7.5), 150 mM NaCl, 10% (v/v) glycerol buffer, using a 1:4.5 molar ratio of TET12-to-TEV protease.

Further analysis of TET12-Spy sample after TEV protease treatment was carried out using MS. Surprisingly, no molecular weight matching that of the uncleaved protein (67896 Da) was observed. Instead, the sample contained a protein of 60407 Da, which was the molecular weight expected for the protein after the cleavage of the N^{TEV} fragment. This discrepancy with the SDS-PAGE could either be due to the extended incubation time of the sample with the TEV protease before MS analysis resulting in completed cleavage or, as discussed previously, the variation in mobility of cyclic proteins through polyacrylamide gels.

Though TEV protease cleavage of the SpyCatcher-N^{TEV} was shown to take place successfully, it was decided that stability experiments should proceed with the full length SpyCatcher-N^{TEV}, as the removal of the N^{TEV} fragment was found to be unviable for a number of reasons. Firstly, a large excess of TEV protease was required for the complete cleavage of the N^{TEV} fragment. The test cleavage reactions were carried out on a small scale (0.44 µM TET12-Spy in 50 µL), and thus N^{TEV} removal with the entire sample would require a large quantity of TEV protease. Secondly, after the cleavage step, removal of the TEV protease would be required. This could be done by carrying out the TEV protease cleavage step after the Ni-NTA purification, followed by SEC to remove the TEV protease

and any other impurities. However, the Ni-NTA elution buffer contains a high concentration of imidazole which would cause inhibition of the TEV protease, thus buffer exchange would be required. Alternatively, two SEC steps could be carried out before and after the cleavage step. In both cases, the extra TEV protease cleavage and removal steps would be time consuming and would reduce an already low yield of TET12-Spy. Finally, the use of the SpyCatcher-N^{TEV} variant, though being kinetically slower than that of the wild type SpyCatcher (9.2 M^{-1} s^{-1} compared to 1.4 x 10^3 M^{-1} s^{-1}, respectively) and disordered when alone solution, still reacts efficiently with SpyTag to form an ordered structure and produce the desired cyclic product.[146] As a result, the presence of the N-terminal fragment in the uncleaved cyclic product would not be expected to deleteriously impact the stability of the protein with respect to the cleaved cyclic product.

2.2.3 Effect of Cyclisation on TET12SN Stability

The stability of a peptide or protein can be defined as resistance towards heat, chemical and/or proteolytic treatment. Enhancing the stability of proteins in one or all of these categories has the potential to expand the scope of their application as pharmaceuticals or biocatalysts.[203] Cyclisation has previously been shown to improve the stability of a number of different industrially relevant proteins (see section 1.3). In this section, the effect of cyclisation on the stability of the artificial protein cage TET12SN will be investigated.

2.2.3.1 TET12SN Proteolytic Stability

The proteolytic degradation of a polypeptide chain is carried out by proteases, which catalyse the hydrolysis of peptide bonds. A peptide or protein that is not broken down in the presence of a protease for an extended amount of time can be classed as proteolytically stable. Cyclisation can lead to improved proteolytic stability due to the absence of free termini which are susceptible to exopeptidase attack. In addition, cyclisation is thought to disfavour protein unfolding by reducing the entropy of the protein unfolded state. As a result, conformational states in which protease recognition sequences are exposed to the solvent are less likely to occur.[10,204]

Carboxypeptidase Y (CPY) is a serine exopeptidase from *Saccharomyces cerevisiae*. It catalyses the release of amino acids from the C-terminus of a polypeptide chain with broad substrate specificity.[205] Here, CPY was employed to compare the proteolytic resistance of linear and cyclised TET12SN variants. Digestion was carried out at 37 °C in buffer containing 20 mM Tris, pH 6.5, 100 mM NaCl, 10% (v/v) glycerol, using a CPY-to-TET12SN molar ratio of 1:3.7. Progress of the digestion was monitored by taking SDS-PAGE samples

at timed intervals for up to 24 hours (**Figure 2.24**). Controls in which CPY and TET12SN variants were incubated separately under the same conditions (not shown), proved the observed degradation resulted from CPY proteolysis and not deterioration of the protein over time.

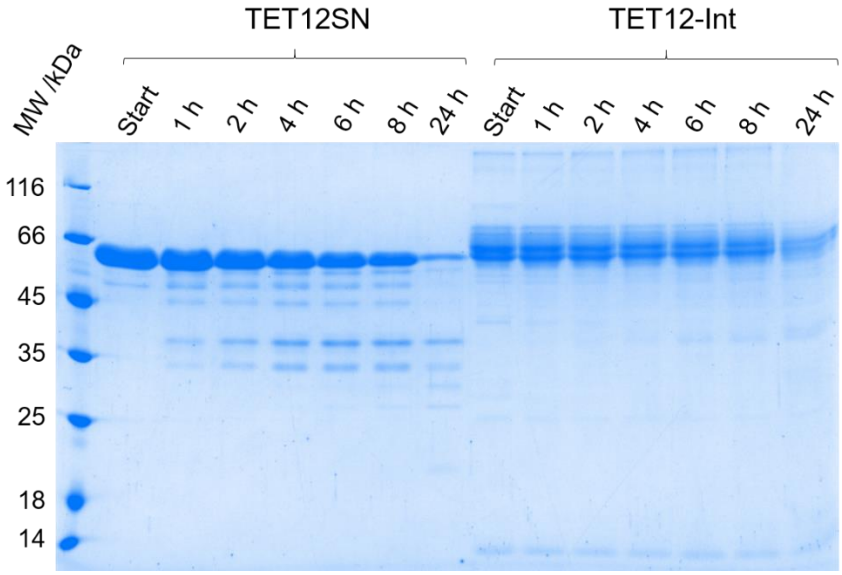


Figure 2.24 SDS-PAGE comparison of the CPY treatment of linear TET12SN and cyclised TET12-Int, carried out at 37 °C in buffer containing 20 mM Tris, pH 6.5, 100 mM NaCl, 10% (v/v) glycerol. A CPY-to-TET12 molar ratio of 1:3.7 was used.

For both TET12SN and TET12-Int, proteolytic degradation is observed over the course of 24 hours with digestion of both proteins incomplete after this time. For the linear TET12SN, protein degradation is clearly observed after 1 hour of incubation with CPY, indicated by the appearance of additional bands at lower molecular weights on the SDS-PAGE. Furthermore, after 24 hours the thickness of the main band at just under 66 kDa, corresponding to undigested (or minimally digested) protein, is reduced. For TET12-Int, some additional bands also appear after 2-4 hours of incubation with CPY, although they are very faint. The size of the undigested TET12-Int band at around 66 kDa also appears to reduce in size over time, most noticeably after 24 hours. This indicates that the cyclic TET12-Int is more stable (but not completely resistant) towards exopeptidase treatment compared to the linear TET12SN. However, due to the blurry nature of the undigested TET12-Int band, the result requires further validation.

2.2.3.2 TET12SN Thermal Stability

In theory, protein cyclisation should lead to improved thermal stability as the number of number of conformational states accessible to the unfolded state would be reduced. As a result, the entropy of the unfolded state would be decreased leading to an increase in Gibbs

free energy. Here, the thermal stabilities of linear and cyclised TET12SN variants were investigated using circular dichroism (CD).

2.2.3.2.1 Circular Dichroism and Melting Temperature

CD is a measure of the difference in the absorption of left- and right-handed components of circularly polarised light. It is a useful technique for the analysis of peptide and protein structural elements, most commonly α-helices and β-sheets, which produce characteristic spectra.[206] In this way, conformational changes in the polypeptide chain can be monitored, including the effect of a certain mutation on a protein's folding or its response to changes in the surrounding environment. By measuring the CD of a protein at increasing temperatures, the midpoint of protein unfolding can be identified. The temperature at which this occurs is known as the melting temperature (T_m) and can be used as a measure of a protein's thermal stability.[155]

Figure 2.25 shows the CD spectrum of TET12SN. TET12SN has a high helical content due to the presence of coiled coil modules which make up the edges of the protein cage. Therefore, the TET12SN CD spectrum is characteristic of α-helices exhibiting minima at 208 and 222 nm. Here, the CD is reported in units of mean residual ellipticity ($[\theta]_{MRE}$):

$$[\theta]_{MRE} = \frac{MRW \times \theta_\lambda}{10 \times l \times c} = deg.cm^2/dmol \qquad (\boldsymbol{Eq.\ 2.1})$$

$$MRW = \frac{M}{N-1} \qquad (\boldsymbol{Eq.2.2})$$

where θ_λ is observed ellipticity at wavelength λ (in mdeg), l is cell pathlength (in cm), c is protein concentration (in g/mL), MRW is mean residue weight, M is molecular weight (in g/mol) and N is number of amino acid residues.

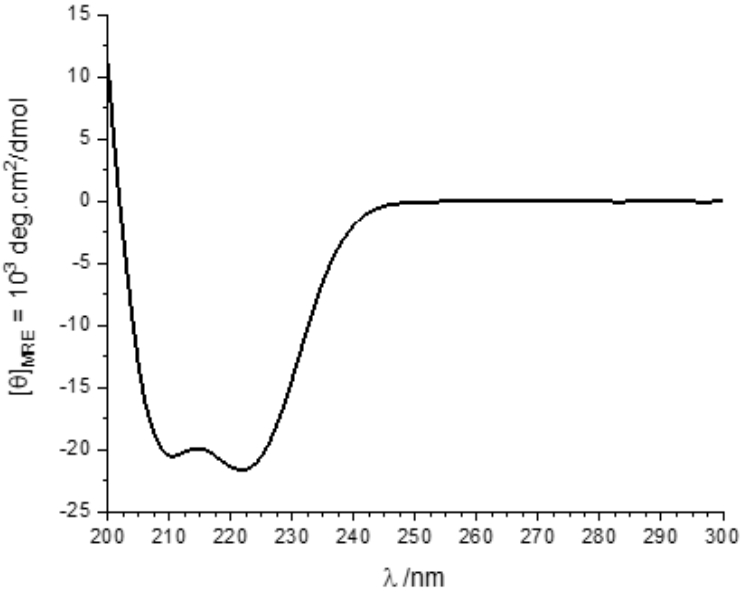
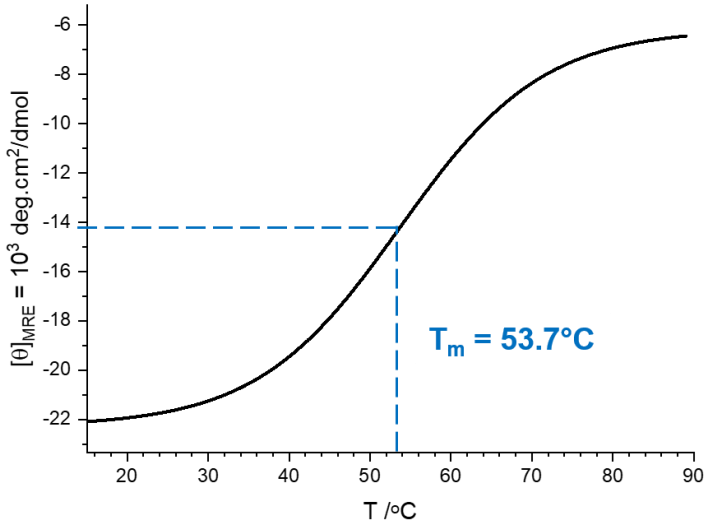


Figure 2.25 The CD spectra of TET12SN between 200 and 300 nm.

The T_m of TET12SN was determined by following unfolding at 222 nm between 15 and 90 °C. The resulting data points were then fitted using the dose-response equation (**Equation 2.3**) from which the midpoint of protein unfolding could then be determined (**Figure 2.26**), i.e. the T_m. For TET12SN, a T_m of 53.7 °C was ascertained, slightly lower than that previously reported by Ljubetič et al. (54.5 °C).[190]



$$(\boldsymbol{Eq.2.3}) \qquad y = A_1 + \frac{A_2 - A_1}{1 + 10^{(\boldsymbol{Logx0} - x)p}}$$

Figure 2.26 The dose-response curve for TET12SN from which T_m is determined (illustrated as a dashed blue line). The dose-response equation is shown where, y is $[\theta]_{MRE}$, A_1 is the bottom y plateau, A_2 is the top y plateau, p is the variable slope factor, x is temperature and the mid-point of unfolding (T_m) is given by logx0.

2.2.3.2.2 Effect of Buffer on Stability

Buffers play an important role in the stabilisation of proteins. Not only do they help to maintain the solution at a constant pH, but additives such as salt and cosolvents (e.g. glycerol) can also confer additional benefits that help to protect the protein against denaturation, prevent protein aggregation and improve solubility.[207] This said, the same additives are also capable of protein destabilisation and in many cases a full understanding is still lacking.[208]

The effect of NaCl concentration and the presence of glycerol on the thermal stability of TET12SN was investigated. CD was used as described above to carry out T_m experiments in a range of 20 mM Tris buffers at pH 7.5 (buffers A-D). In the literature, CCPO cage T_m measurements were made in buffers containing 150 mM NaCl and 10% (v/v) glycerol (buffer A). Higher concentrations of NaCl and the presence of glycerol would be expected to stabilise the protein, resulting in an increased T_m compared to buffers containing a lower NaCl concentration and/or no glycerol. **Figure 2.27** shows the T_m of TET12SN measured in these different buffers.

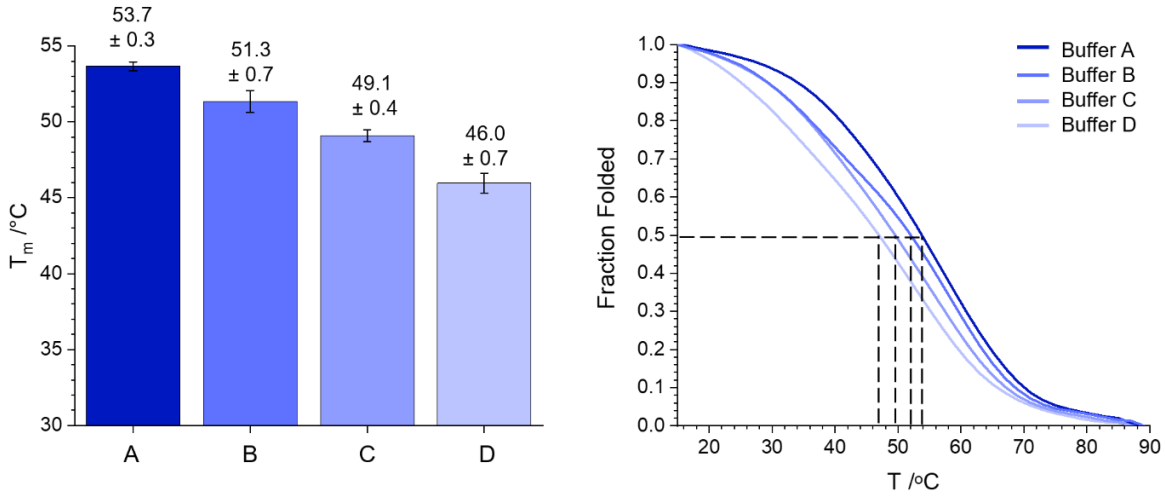


Figure 2.27 (Left) A bar chart comparing the T_m of TET12SN in buffer A (150 mM NaCl, 10% glycerol), B (10 mM NaCl, 10% glycerol), C (150 mM NaCl, 0% glycerol) and D (10 mM NaCl, 0% glycerol). Error bars represent the 95% confidence interval. (Right) Fraction of TET12SN folded with increasing temperature in different buffers, dashed lines indicate the mid-point of unfolding.

As predicted, the absence of glycerol combined with a reduced NaCl concentration (buffer D) resulted in a dramatically decreased thermal stability of TET12SN, with a midpoint of unfolding observed at 46.0 °C. This is almost 8 °C lower than that observed for the most stabilising buffer (buffer A) containing both a higher NaCl concentration and glycerol. In addition, removal of glycerol from the buffer had a greater impact on T_m than the reduction in NaCl concentration (buffers C and B, respectively).

2.2.3.2.3 Linear versus Cyclised TET12SN Melting Temperature

To compare the thermal stabilities of linear and cyclised TET12SN variants, two different buffers were chosen to carry out the T_m experiments (**Figure 2.28**). These were the most stabilising buffer (buffer A) and the least stabilising buffer (buffer D).

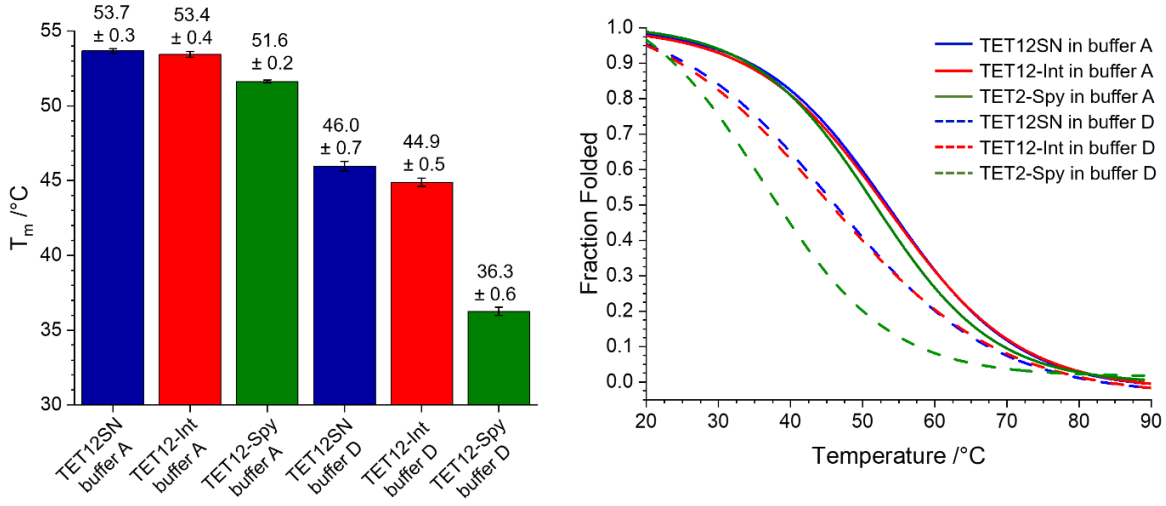


Figure 2.28 Bar chart comparing the T_m of TET12SN, TET12-Int and TET12-Spy in buffer A and D. Error bars represent the 95% confidence interval.

In buffer A, no significant difference was observed between the T_m of the cyclised TET12-Int and linear TET12SN (53.4 and 53.7 °C, respectively). Meanwhile, in buffer D the T_m of TET12-Int was reduced by around 8 °C. In addition, the T_m of TET12-Int in the less stabilising buffer was slightly reduced compared to that of TET12SN (44.9 and 46.0 °C, respectively). This indicates that the protein was not stabilised by split intein cyclisation.

On the other hand, the cyclic TET12-Spy showed a reduced T_m in both buffers compared to TET12SN and TET12-Int. In the more stabilising buffer the T_m was decreased by around 2 °C compared to the linear TET12SN. However, in the less stabilising buffer a T_m of 36.3 °C was observed, almost 10 °C lower than TET12SN in the same buffer and around 15 °C lower compared to TET12-Spy in buffer A. This clearly indicated that cyclisation of TET12SN by SpyTag/SpyCatcher-N^{TEV} resulted in destabilisation. For further confirmation, the T_m of the linear TET12-Spy^{D7A} mutant was measured (**Figure 2.29**). In both buffers T_m was increased compared to the cyclic TET12-Spy variant, although it still exhibited slightly lower thermal stability than the linear TET12SN. Therefore, it is likely that the destabilisation of TET12-Spy was primarily due to strain introduced upon cyclisation, rather than by the presence of the SpyTag/SpyCatcher-N^{TEV} sequences located at the termini of TET12SN.

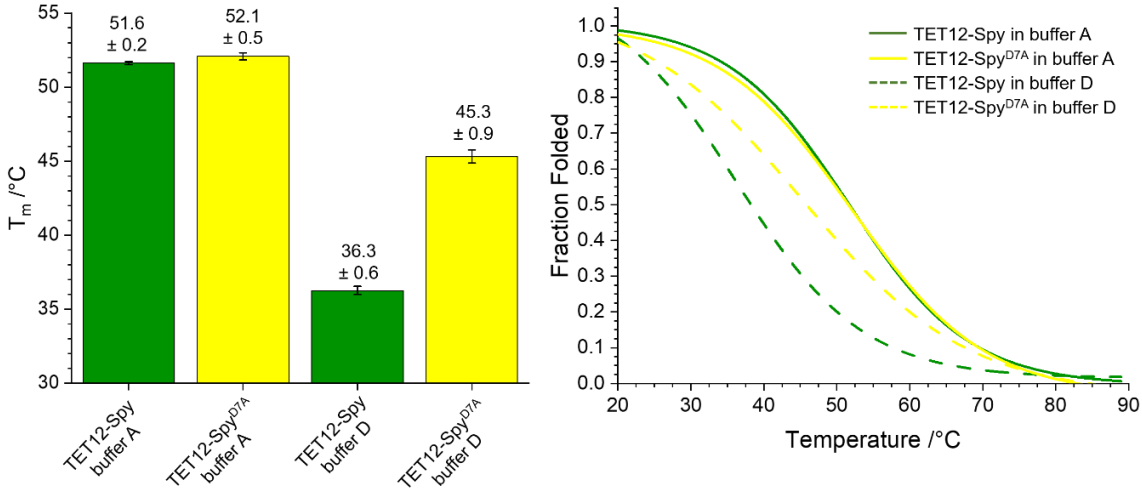
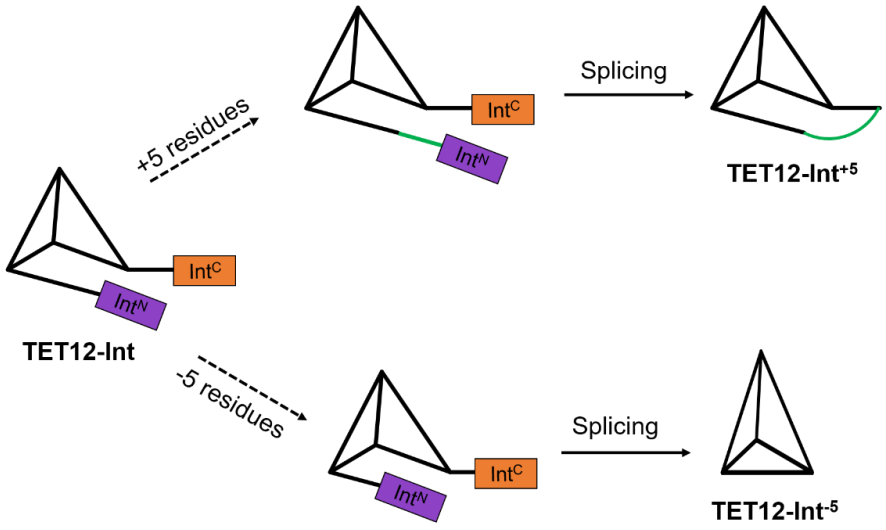


Figure 2.29 Bar chart comparing the T_m of linear and cyclised TET12-Spy variants in buffer A and D. Error bars represent the 95% confidence interval.

2.2.3.2.4 Effect of Linker Length on Melting Temperature

Strain introduced upon cyclisation of a protein can result in destabilisation due to distortion of the tertiary structure. To remove strain, the number of amino acid residues in between the end of the sequence of the protein of interest (e.g. TET12SN) and the start of the cyclisation machinery sequence (e.g. split intein fragments) can be increased (**Scheme 2.30**). This extension of the linker should relieve the strain and result in improved stability.



Scheme 2.30 Modification of the TET12-Int linker length by the addition or removal of 5 amino acid residues.

As the split intein cyclisation of TET12SN did not appear to significantly affect the thermal stability of the protein cage, the linker length of TET12-Int was modified using SDM to check for the presence of strain. Surprisingly, extension the linker by 5 amino acid residues

(TET12-Int^{+5}) resulted in a reduced T_m of around 1 °C in both buffers tested (**Figure 2.31**). Furthermore, shortening the linker by 5 residues (TET12-Int^{-5}), which was expected to result in destabilisation of the protein, in fact had no significant effect on T_m.

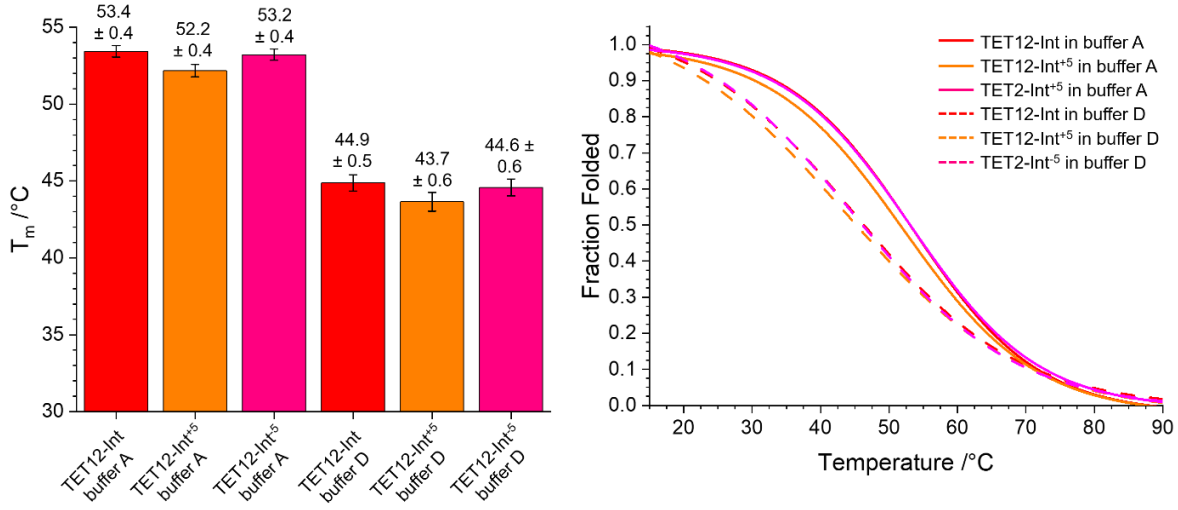


Figure 2.31 Bar chart showing the effect of linker length modification on the T_m of TET12-Int in buffer A and D. Error bars represent the 95% confidence interval.

The length of the linker in TET12-Spy was also investigated. The initial cyclisation of TET12SN using SpyTag/SpyCatcher-N^{TEV} clearly resulted in destabilisation. It was therefore expected that the addition of 5 amino acids (TET12-Spy^{+5}) to the linker would relieve this strain leading to an increased T_m. In buffer D this was the case, with a T_m of 42.3 °C observed, an improvement of 6 °C, although this was still below that of the linear TET12-Spy and TET12SN (**Figure 2.32**). On the other hand, in buffer A, the T_m of TET12-Spy^{+5} was decreased compared to that of both the linear and cyclised TET12-Spy as well as TET12SN.

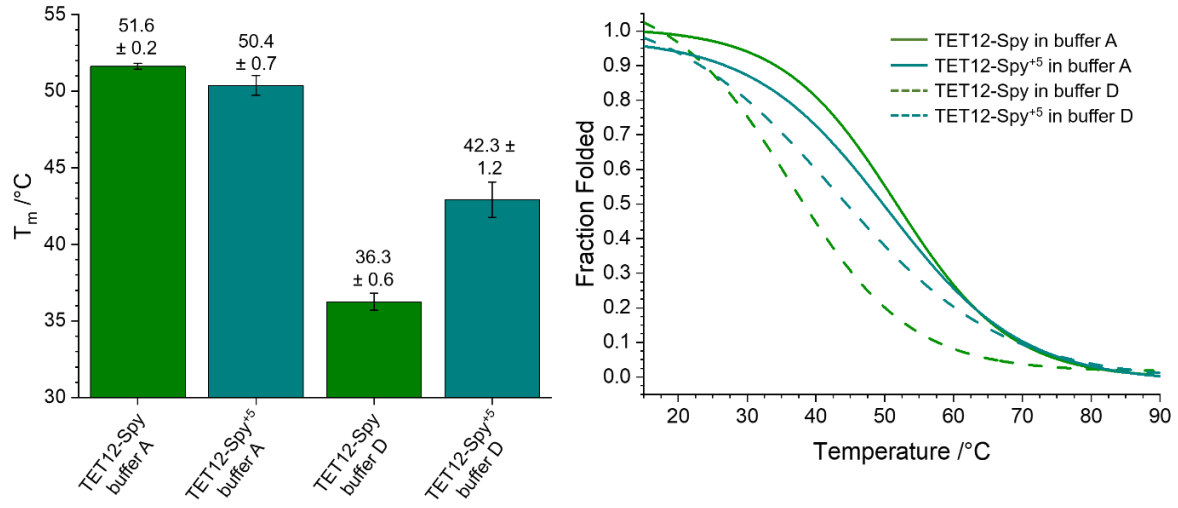


Figure 2.32 Bar chart showing the effect of increasing linker length by 5 residues on the T_m of TET12-Spy in buffer A and D. Error bars represent the 95% confidence interval.

Table 2.2 summarises all T_m values measured for the linear and cyclised TET12SN variants discussed above. Overall, neither the cyclisation of TET12SN by split intein nor SpyTag/SpyCatcher-N^{TEV} resulted in an improvement to the T_m measured using CD. While TET12-Int had a similar T_m to TET12SN, TET12-Spy was destabilised, most likely due to the introduction of strain to the protein. For both TET12-Int and TET12-Spy cyclised variants, the effect of linker length on thermal stability was investigated. Extending the linker by 5 residues appeared to relieve strain introduced by cyclisation using SpyTag/SpyCatcher-N^{TEV}, although there was not an overall improvement to T_m compared to the linear TET12SN. Further recovery of T_m could be possible by a creating an even longer linker. However, the addition and removal of 5 residues to/from the linker of TET12-Int did not dramatically affect the T_m and indicates that even with the removal of strain, cyclisation does not lead to the improved thermal stability of TET12SN.

Table 2.2 The T_m measured using CD for linear and cyclised TET12SN variants.

TET12SN Variant	Linear (L) or Cyclised (C)	T_m /°C		ΔT_m Buffer A vs D /°C	T_m vs TET12SN /°C	
		Buffer A	Buffer D		Buffer A	Buffer D
TET12SN	L	53.7 ± 0.3	46.0 ± 0.7	7.7	-	-
TET12-Int	C	53.4 ± 0.4	44.9 ± 0.5	8.6	-0.2	-1.1
TET12-Int^{+5}	C	52.2 ± 0.4	43.7 ± 0.6	8.5	-1.5	-2.3
TET12-Int^{-5}	C	53.2 ± 0.4	44.6 ± 0.6	8.6	-0.5	-1.4
TET12-Spy	C	51.6 ± 0.2	36.3 ± 0.6	15.4	-2.0	-9.7
TET12-Spy^{D7A}	L	52.1 ± 0.5	45.3 ± 0.9	6.8	-1.6	-0.7
TET12-Spy^{+5}	C	50.4 ± 0.7	42.3 ± 1.2	8.1	-3.3	-3.7

2.2.3.3 TET12SN Chemical Stability

The use of chemical denaturants, such as guanidine hydrochloride and urea, are commonly employed to induce protein unfolding. Accordingly, exposing proteins to increasing concentrations of such denaturants while monitoring changes to their folding, is a convenient method for assessing chemical stability. Here, the chemical stability of TET12SN and TET12-Int variants were compared using CD to follow changes in their secondary structures as the concentration of urea was increased.

All chemical stability measurements were carried out in buffer A, containing 150 mM NaCl and 10% glycerol with a urea concentration ranging from 0 to 8 M. The proteins were incubated at room temperature in the appropriate buffer for 1 hour before CD

measurements were taken at 20 °C. Due to the high absorbance of urea below 217 nm, a wavelength range of 217-260 nm was employed (see Appendix 7.4). To find the concentration of urea at which the mid-point of unfolding (C_m) occurred, a dose-response curve was fitted to the data points taken at 222 nm. **Figure 2.33** shows the C_m obtained for linear TET12SN and cyclised TET12-Int, TET12-Int^{Nat}, TET12-Int^{+5} and TET12-Int^{-5}.

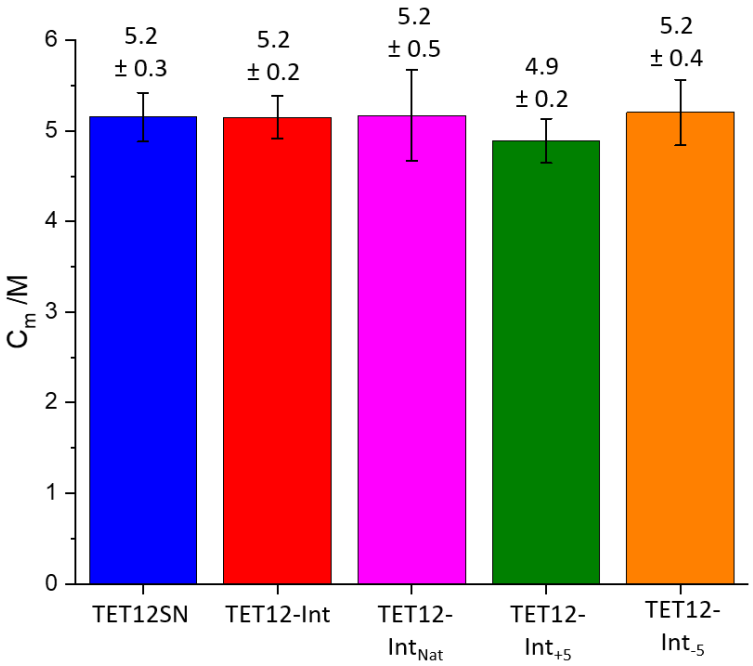


Figure 2.33 Bar chart comparing the C_m in urea of linear and cyclised TET12SN. Error bars represent the 95% confidence interval.

The results obtained for the C_m of TET12SN and TET12-Int variants emulated to those obtained for the T_m of the same proteins. For both TET12SN and TET12-Int, the C_m occurred at 5.2 M urea. The TET12-Int^{Nat} variant also showed no significant difference in C_m compared to TET12SN, indicating that cyclisation by split intein did not affect the chemical stability of the protein cage. In addition, the C_m of the TET12-Int^{+5} and TET12-Int^{-5} variants exhibited a similar trend to their T_m values. Extension of the linker length led to a slight decrease in stability, while shortening the linker had little effect on C_m.

2.2.3.4 TET12SN Stability Summary

The proteolytic, thermal and chemical stabilities of linear and cyclised TET12SN variants were investigated. Encouragingly, cyclised TET12-Int appeared to show improved tolerance towards exopeptidase treatment compared to its linear TET12SN counterpart, presumably due to the absence of free-termini in the cyclised form. On the other hand, TET12-Int showed little difference in both thermal and chemical stability compared to TET12SN. Moreover, the TET12-Spy cyclic variant was thermally destabilised due to cyclisation,

although extension of the linker length did allow partial recovery of T_m. Overall, it is concluded that cyclisation of the CCPO cage TET12SN does not lead to improved thermal and chemical stability, and may actually result in destabilisation.

2.3 DISCUSSION

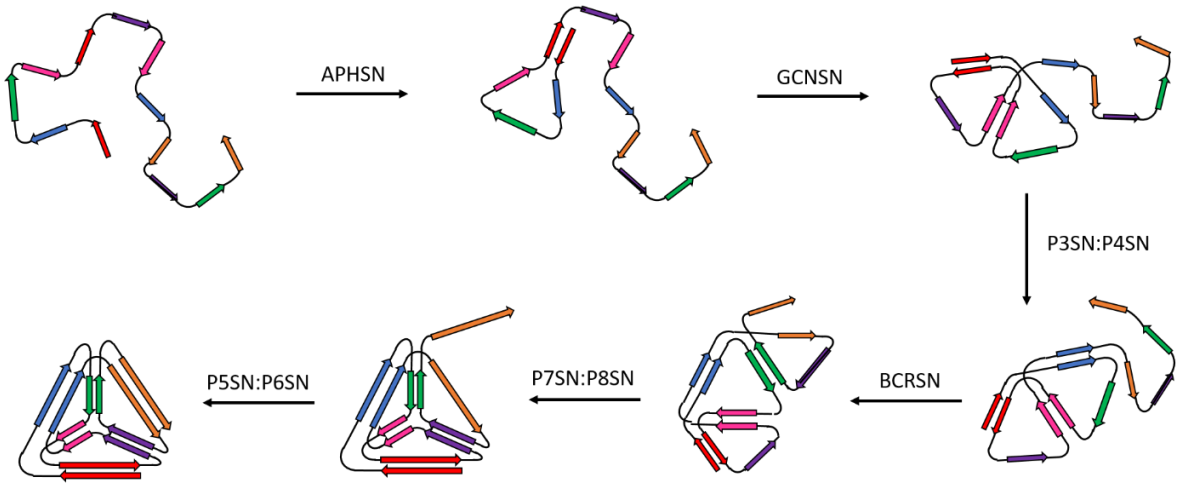
The tetrahedral CCPO cage TET12SN was cyclised using two different approaches. The *Npu* DnaE split intein approach resulted in head-to-tail cyclisation, while the SpyTag/SpyCatcher-N^{TEV} generated a side chain-to-side chain cyclised product. In both cases, the cyclic monomers were successfully isolated, although in a low yield likely due to the formation of aggregates. The proteolytic, thermal and chemical stabilities of linear and cyclised proteins were then compared. Unexpectedly, cyclisation did not appear to result in improved thermal or chemical stability. In this section, the aggregation of the cyclic TET12-Int and TET12-Spy will first be examined in more detail with respect to the folding mechanism of the CCPO cage, followed by a discussion on stability. Some future directions for further research will also be highlighted.

2.3.1 TET12SN Cyclisation Results in Increased Aggregation

CCPO relies on the use of orthogonal coiled coil sets concatenated along a single polypeptide chain. The correct dimerisation of these coiled coil partners is essential for guiding the folding of the polypeptide chain along the intended double Eulerian trail. Any non-specific interactions would lead to misfolding. Hence, the desired polyhedral cage structure would not be produced. Initially, CCPO cages were designed to include each coiled coil building block only once. However, this limits the size and complexity of structures possible. On the contrary, domain repeats are often observed in natural proteins, whose folding is suggested to be mainly determined by chain topology.[209,210] Subsequently, the folding pathway of CCPO structures have been found to be governed by the intra-chain distance of coiled coil dimer forming units. CCPO polyhedra containing multiple copies of the same coiled coil building blocks have been designed and successfully constructed.[191]

Scheme 2.34 shows the stepwise folding mechanism proposed for TET12SN, which was investigated by Aupič et al. using molecular dynamics simulations and FRET.[191] The folding of the tetrahedral CCPO cage proceeds sequentially, with one coiled coil dimer edge forming in each step. The order of the coiled coil dimer assembly depends on the intra-chain distance between the constituting coiled coil forming units. The pair located at (or amongst those with) the shortest intra-chain distance or at the polypeptide chain termini, preferentially assemble in that step. Therefore, folding order is determined by the chain

topology and not the thermodymanic or kinetic stability of the coiled coil dimer forming segments.



Scheme 2.34 TET12SN CCPO cage folding pathway.[191]

Accordingly, when the sequence of coiled coil building blocks was reordered for a suboptimal folding pathway, substantial aggregation was observed.[191] This has been suggested to be due to the formation of misfolded species and the exposure of non-paired coiled coil segments which could then undergo intermolecular interactions.

Here, a large amount of aggregation was observed for the cyclised TET12-Int and TET12-Spy. As the linear mutants of both proteins (N137A/C1A and D7A mutants, respectively) were isolated primarily in their monomeric forms, it suggests the presence of the additional sequences (i.e. split intein or SpyTag/SpyCatcher-N^{TEV} fragments) at the termini of the TET12SN protein cage were not responsible for the increased aggregation. To illustrate, the SEC chromatograms of linear and cyclised TET12-Int and TET12-Spy are shown in **Figure 2.35**. The peaks corresponding to the aggregated protein are clearly larger in the chromatograms of the cyclised proteins, whereas the linear proteins are primarily eluted as monomers. Also notably, more monomeric cyclic TET12SN is produced using the SpyTag/SpyCatcher-N^{TEV} isopeptide bond forming approach compared to the split intein approach, with cyclic TET12-Int eluted mainly as aggregates.

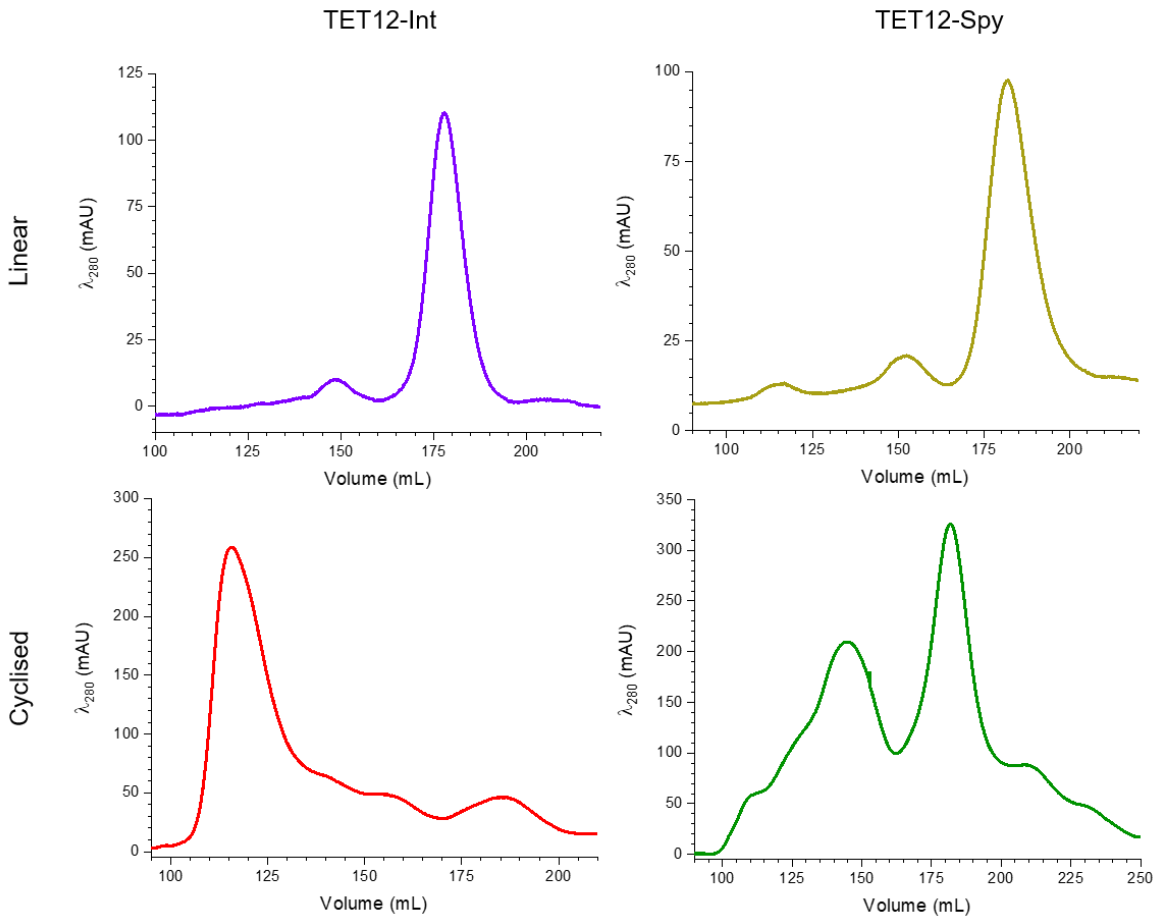


Figure 2.35 SEC chromatograms for linear and cyclised TET12-Int and TET12-Spy.

As discussed above, the folding pathway of TET21SN depends on the intra-chain distance of the coiled coil dimer folding units. Hence, if the cyclisation step was to occur before the folding of the cage was complete, it may interfere with the CCPO folding pathway, particularly as coiled coil building blocks located at the termini preferentially dimerise first, as well as those located more closely together along the polypeptide chain. Therefore, premature cyclisation of the polypeptide chain constituting TET12SN would result in an increase in misfolded species and in turn aggregation.

Upon closer inspection, the rate constants of *Npu* DnaE split intein and SpyTag/SpyCatcher-N^{TEV} are reported to be 3.7 x 10^{-2} s^{-1} and 9.2 M^{-1} s^{-1}, respectively.[130,146] These correspond to a $t_{1/2}$ of 19 s for *Npu* DnaE split intein and 90.5 min for SpyTag/SpyCatcher-N^{TEV} (based on a starting concentration of 20 µM). Meanwhile, the folding kinetics of TET12SN are reported to be comparable to those of natural globular proteins of a similar size, with a rate constant of 31 s^{-1} corresponding to a $t_{1/2}$ of 0.02 s.[190] These values indicate that the folding of the TET12SN cage is faster than native peptide bond formation by the split intein and isopeptide bond formation by SpyTag/SpyCatcher-

N^{TEV}, although this does not rule out interference to the folding mechanism. Furthermore, a slower cyclisation step (i.e. by SpyTag/SpyCatcher-N^{TEV}) would be expected to result in a higher yield of monomeric cyclic TET12SN, as it would allow more time for the complete folding of the protein cage before ligation of the termini. Here, the monomer of cyclic TET12-Spy was obtained in a higher yield than cyclic TET12-Int, further suggesting that increased aggregation could result from premature cyclisation leading to misfolding of the protein cage.

Another reason for the increased quantities of aggregates observed for the TET12-Int and TET12-Spy species, could be the formation of oligomers. This could result from intermolecular covalent bond formation between two or more TET12SN molecules, instead of the desired intramolecular ligation of the termini of a single molecule. Oligomerisation is a common problem encountered during cyclisation, especially at high concentrations when there is an increased likelihood of intermolecular reaction. Though it is impossible to entirely prevent intermolecular interactions, some strategies for reducing the tendency of protein cage aggregation are proposed below.

1. **Slower protein production**, for example by reducing the concentration of inducer and/or reducing induction temperature, could lower the concentration of unreacted Int^{C}/Int^{N} or SpyTag/SpyCatcher-N^{TEV} fragments present at any time after recombinant production in *E. coli*. As a result, the desired intramolecular reaction (i.e. cyclisation) would be favoured over intermolecular reaction (i.e. oligomerisation). Therefore, an improved yield of the cyclised monomer would be obtained.

2. **Reduced rate of cyclisation** could prevent premature cyclisation, ensuring complete folding of the tetrahedral cage before the ligation of the termini. For the split intein approach, this could be done by using a less efficient split intein, for example *Ssp* DnaE or by mutation of one or multiple amino acid residues in the extein sequence and/or intein sequence which are known to be influential in the splicing mechanism. However, inefficient splicing often leads incomplete splicing and the formation of side products (see section 2.2.2.1), resulting in a reduced yield of the desired cyclised product.

3. **Conditional cyclisation**, whereby the spontaneous protein cyclisation step is completely inhibited until after the folding of the TET12SN cage has taken place,

may be more practical for preventing aggregation from either misfolding or oligomerisation. For the split intein approach, the splicing reaction could be triggered using external stimuli such as light,[211–213] small molecules and protease[214] to induce splicing.[215] Similarly, SpyTag/SpyCatcher-N^{TEV} reactivity could be manipulated using a redox-responsive conformational restriction approach.[216]

Of course, all strategies suggested above have their own limitations and are likely to bring their own unforeseen complications to the cyclisation of TET12SN. Chapter 1 describes just a few of the many cyclisation methodologies available, each having benefits and limitations, making a particular cyclisation approach either more or less suitable for a specific application. In short, finding the most suitable approach for the cyclisation of TET12SN would require significant trial and error. Nevertheless, on a more positive note, both the *Npu* DnaE split intein and SpyTag/SpyCatcher-N^{TEV} approaches employed for the cyclisation of TET12SN did produce the desired cyclic TET12-Int and TET12-Spy products, respectively, albeit in a low yield.

2.3.2 TET12SN Cyclisation Did Not Improve Stability

With the exception of proteolytic stability, the cyclised variants of TET12SN did not show improved stability compared to their linear counterparts. While the large quantity of aggregation observed for cyclic TET12-Int and TET12-Spy can be explained by a combination of covalent oligomerisation and misfolding, it does not explain the decreased thermal and chemical stability, assuming the isolated cyclic monomers are indeed correctly folded.

Here, CD was employed to investigate the stability of TET12SN variants towards increasing temperatures and concentrations of urea. These measurements monitored the protein secondary structure, composed mainly of helices which make up the coiled coil dimer edges of the CCPO cage. Hence, the midpoints of unfolding (T_m and C_m) observed for all linear and cyclised TET12SN variants correspond to loss of helical structure, rather than loss of the overall tetrahedral cage structure. It has previously been reported that the unfolding of TET12SN takes place via a multi-step process, with loss of tertiary structure occurring before the unfolding of the secondary structure.[191] As a result, it is likely that cyclisation of the CCPO tetrahedron not only affects the folding pathway, but also the unfolding pathway of the protein cage. Therefore, it could be possible that the rate of tertiary structure unfolding varies been linear and cyclic TET12SN protein cages.

Alternatively, the lack of stabilisation achieved by the cyclisation of the TET12SN protein cage may be more intrinsic of the CCPO structure, which is composed of rigid coiled coil elements joined by flexible peptide linker sequences. In addition, CCPO cages lack a hydrophobic core. As a result, the TET12SN protein cage likely exhibits increased flexibility compared to natural proteins. Hence, the unfolded state of cyclic TET12SN may not be sufficiently destabilised by a reduction in entropy, and so cyclisation would not result in the stabilisation of the protein's folded state.

The folding and unfolding of the TET12SN cage has previously been investigated using FRET, through the use of strategically placed donor and acceptor chromophores.[190,191] When the two chromophores come into close proximity (<10 nm) the excited donor transfers energy to the acceptor by a non-radiative process known as resonance and can be measured by monitoring the acceptor emission intensity. As a result, FRET is a useful technique for determining the spatial proximity of parts of protein molecules, and could be applied to investigate the folding and unfolding of cyclic TET12SN variants. The correct folding of cyclic TET12-Int and TET12-Spy could be demonstrated by placement of the chromophore pair at the termini of the TET12SN sequence (before the split intein or SpyTag/SpyCatcher-N^{TEV} sequences). As the termini of the TET12SN protein cage are designed to coincide at the same vertex, FRET should occur upon the formation of the last coiled dimer in the folding mechanism (P5SN:P6SN) when the termini come into close proximity. Therefore, the observation of FRET would indicate that cyclic TET12SN is folded correctly.

In addition, the donor and acceptor chromophores could be placed on coiled coil building blocks which form a particular edge of the cage. Multiple different labelled variants could be created and used to monitor the unfolding of the cyclised TET12SN protein cages in the presence of increased temperature or concentration of chemical denaturing agent. Upon unfolding, it would be expected that that chromophore pair would become separated and FRET would be lost. The midpoint of unfolding (i.e. T_m or C_m) could then be employed to compare the stability of the linear and cyclised TET12SN tetrahedral tertiary structure. A similar experiment could also be employed to monitor the refolding of the protein after denaturation. If cyclisation was interfering with the CCPO folding mechanism, refolding of denatured TET12SN would not be expected.

Finally, as briefly discussed above, cyclisation may not benefit artificial CCPO proteins in the same way as natural proteins, likely due to their increased flexibility and lack of hydrophobic core. Consequently, strategies based on the stabilisation of the constituent

coiled coil dimer modules are more likely to be successful in enhancing the overall stability of the CCPO cage.[191] Of course, this would need to be done without disrupting the orthogonality of the coiled coil dimer sets.[182] Nevertheless, a trigonal bipyramid CCPO cage (BIP18SN) was recently demonstrated to be sufficiently resilient towards heat treatment to enable the use of a thermal lysis purification procedure.[217]

2.4 CONCLUSIONS

To summarise, cyclisation has received much attention as an effective method for peptide and protein stabilisation. This is evidenced by cyclic peptides found in nature, as well as proteins artificially cyclised in the laboratory, for which a large number of techniques are available. Here, the artificial tetrahedral CCPO cage TET12SN was cyclised, with the aim of improving its stability. Two different cyclisation techniques were employed. Specifically, these were a *Npu* DnaE split intein-mediated cyclisation approach, as well as a SpyTag/SpyCatcher-N^{TEV} intramolecular isopeptide bond forming strategy. Both techniques afforded the desired cyclised products as confirmed by SDS-PAGE and MS analysis. However, only a low yield of the cyclic monomeric products could be isolated as large quantities of aggregates were produced. This was likely due to a combination oligomerisation and misfolding.

Compared to linear TET12SN, cyclic TET12-Int demonstrated improved resistance to exopeptidase treatment. However, neither of the cyclised variants exhibited an improved thermal stability when compared to the linear TET12SN. Extension of the linker length did not result in any further improvement to the T_m. In addition, the chemical stabilities of TET12SN and TET12-Int followed a similar trend to that of thermal stability. As a result, in future, it could be interesting to investigate more closely why cyclisation affects different proteins in different ways. This could enable a more tailored approach for achieving the desired stabilisation effects, as well as provide a better understanding of protein stabilisation in general.

CHAPTER 3:
CYCLISATION OF A PLASTIC DEGRADING ENZYME

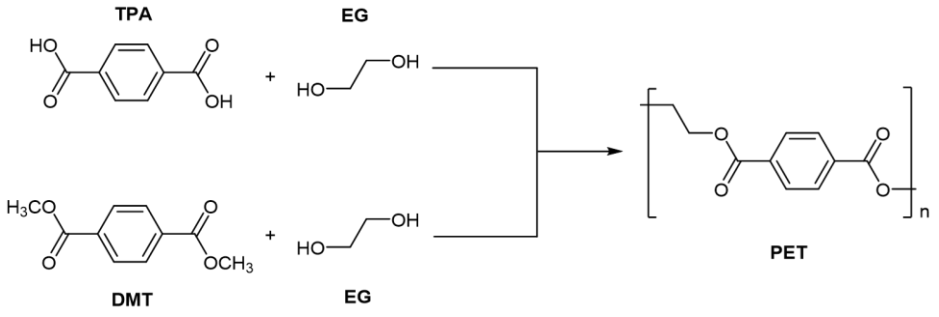
3.1 PREFACE

3.1.1 Background

Many things we take for granted in everyday life would not be possible without plastic. From the clothes we wear and the food we eat, to the keyboards we type on and the floors we walk on, plastic is omnipresent. Durability, light-weight, transparency and low cost, are just a few of the many desirable properties which make plastic the material of choice for an endless variety of applications and have transformed the world we live in, resulting in medical advances and space travel. However, our reliance on plastic is taking its toll on the world around us, causing potentially irreversible damage to other species with which we share this planet. Plastic discarded into landfill is transported into the ocean. Animals become entangled, leading to injury and starvation. Partial deterioration of larger plastic fragments through weathering leads to the formation of microplastics, which are easily transported by the wind and rain to some of the world's remotest places such as Mount Everest and the Artic.[218,219] Furthermore, a newly discovered amphipod from the Mariana Trench was named *Eurythenes plasticus* due to the microplastic fibre found in its gut.[220] Even more alarmingly, plastic particles were recently discovered for the first time in samples of human blood.[221] Currently, the full environmental and health implications of microplastics are unknown, but it is clear drastic changes are urgently required.[222]

3.1.2 Polyethylene Terephthalate

Due to its excellent mechanical and thermal properties, polyethylene terephthalate (PET) is one of the most widely produced plastics worldwide and predominantly employed for making packaging and textiles. It is a semi-crystalline material, composed of ordered crystalline and disordered amorphous phases, the proportion of which determines the rigidity or flexibility of the plastic.[223] Composed of repeating terephthalate (TPA) and ethylene glycol (EG) monomer units, PET is an aromatic polyester obtained either by the esterification of EG with TPA or the transesterification of EG with dimethyl terephthalate (DMT).[224]



Scheme 3.1 Production of PET.

3.1.3 PET Recycling

Though the recycling and reuse of many plastics, including PET, is becoming more widespread, the majority of post-consumer waste plastic is still incinerated (42.6% in Europe, 2018) or sent to landfill (25.0% in Europe, 2018).[225] Currently, PET is recycled either mechanically or chemically.[226] In mechanical recycling, the plastic is ground up and melted into pellets which are used to make new products. It is a relatively simple process with minimal environmental impact but it cannot be used to recycle contaminated PET waste and produces recycled products with deteriorated properties and value compared to virgin PET.[227] On the other hand, chemical recycling methods such as glycolysis, methanolysis and hydrolysis, depolymerises the plastic into its constituent monomers, oligomers or other chemical substances.[228] These products can then be reused for the synthesis of more polymers without compromising the quality of the recycled product. Nevertheless, chemical recycling methods generally require the use of harsh conditions such as high temperatures, raised pressures and/or the use of hazardous reagents. Moreover, the high costs associated with these approaches limit their economic viability. Consequently, the biodegradation PET is an attractive alternative as it would employ milder conditions, reducing energy consumption and its associated costs and pollution. Similar to chemical recycling, the biodegradation of PET would enable the recovery of the plastic monomers which could be reused for the manufacture of plastic without loss of desirable properties or value.[229] Nevertheless, the biodegradation of plastics is not without its challenges. Generally, these arise from the physicochemical characteristics of the plastic.[230,231]

1. Plastics are composed of long polymer chains of high molecular weight. These large structures cannot be transported across cell membranes without first being depolymerised into smaller oligomers or monomers.
2. Many plastics lack favourable functional groups for enzymatic attack, being composed of long carbon chains (e.g. PE, PP, PS, PVC).
3. Polymers containing hydrolysable bonds are more susceptible to enzymatic degradation (e.g. PET and PU), but are often inaccessible due to the close packing and high crystallinity of the plastic.
4. Hydrophobicity of plastics also hinders enzymatic attack and the formation of biofilms.

3.1.4 PET Degrading Enzymes

Enzymes capable of plastic degradation have been isolated from fungi, such as *Fusarium solani pisi* and *Humicola insolens*, and bacteria, particularly from the *Thermobifida* species.[232] Generally, PET degrading enzymes are carboxylic ester hydrolases (EC 3.1.1)

including cutinases, lipases and esterases. These enzymes preferentially act on the amorphous phase of the PET where the polymer chains are more flexible and accessible. Therefore, as the percentage crystallinity of the plastic increases, depolymerisation becomes increasingly difficult.[233] However, pre-treatment of PET samples before they are subjected to enzymatic hydrolysis has been demonstrated to assist degradation by increasing the accessibility of polymer chains through amorphogenesis and/or reduction in particle size.[234] Subsequently, direct comparison between the PET degrading activities of different enzymes is often not possible due to the wide variety of PET substrates employed (i.e. varying crystallinity, orientation of polymer chains and pre-treatments).

Two features that have been identified to be advantageous for the enzymatic degradation of PET are an open hydrophobic active site and thermal stability.[232] An open hydrophobic active site enables access to the surface of the plastic and provides sufficient space and affinity for bulky aromatic substrates (i.e. PET). Meanwhile, thermal stability permits the use of enzymes at elevated temperatures. This is beneficial for PET degradation due to the increased mobility and accessibility of the PET amorphous phase at around 70 °C, whereupon a rigid-to-flexible transition occurs, known as the glass transition temperature (T_g) (**Figure 3.2**).[235]

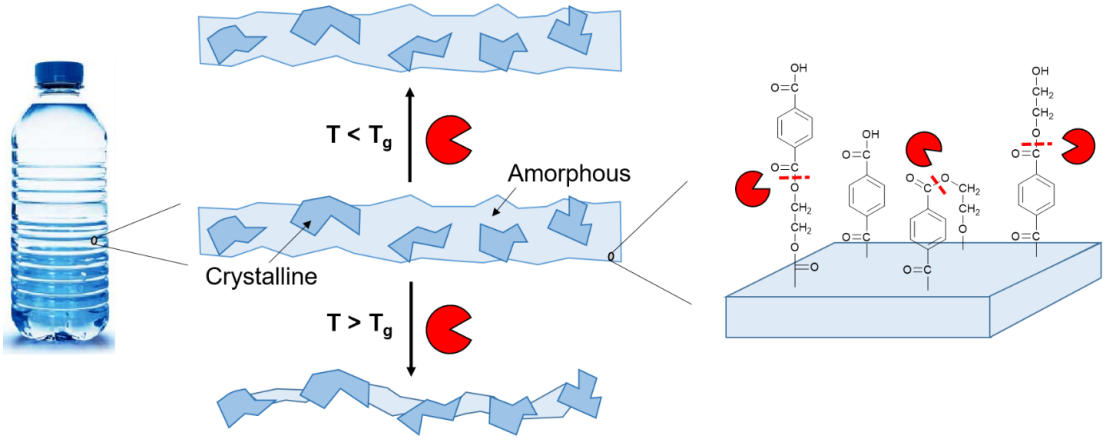


Figure 3.2 Enzymatic attack of the PET amorphous phase.

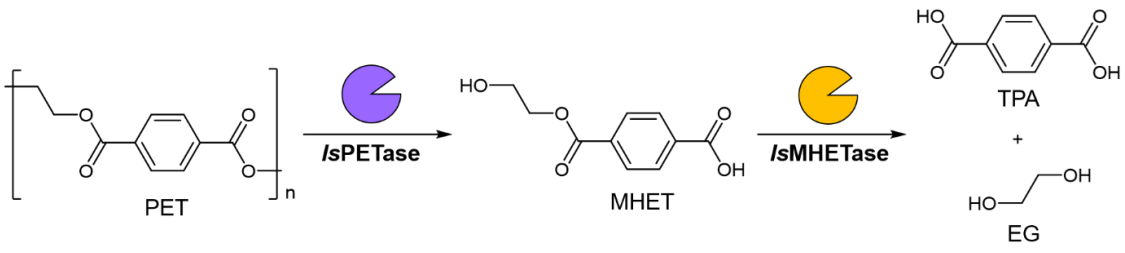
Some promising PET degrading enzymes include TfCut2, LC-cutinase (LCC) and HiC. These enzymes originate from thermophilic sources and hence exhibit thermal stability which is advantageous for PET degradation. Indeed, incubation of a TfCut2 variant (cutinase from *Thermobifida fusca* KW3) with an amorphous PET film at 65 °C for 50 hours, resulted in a weight loss of around 40%.[236] Meanwhile, LCC, a thermally stable PET hydrolase (T_m = 86 °C)[237] isolated from leaf-branch compost by a metagenomic approach, was shown to have optimal PET degrading activity at 50 °C.[238] Using computer-aided engineering, a LCC variant with a T_m of 94 °C was generated which could catalyse

degradation of pre-treated post-consumer waste PET with up to 90% conversion in less than 10 hours at 70 °C.[239] Moreover, complete degradation of a low crystallinity PET film was achieved by HiC, a cutinase from *H. insolens*, after 96 hours at 70 °C.[240] Despite this, HiC exhibited significantly reduced activity for the degradation of high crystallinity PET samples.[241,242]

The high levels of PET degrading activity displayed by these enzymes make them promising candidates for future recycling and bioremediation purposes. However, further improvements to activity, stability and product distribution are still required if they are to meet the desired thresholds for efficient large scale application.[243] In addition, a detailed understanding of the mechanisms surrounding enzymatic PET degradation is often lacking. For the majority of known PET degrading enzymes, plastic is not the native substrate for which they evolved to catalyse ester hydrolysis. Therefore, these enzymes often require engineering to tailor their activities towards PET substrates.[244] Research in this field has been aided by the discovery and characterisation of PETase, isolated from the PET metabolising bacteria *Ideonella sakaiensis* (*Is*PETase). Initial studies demonstrated the preference of the enzyme towards PET substrates (compared to aliphatic esters) and was found to outperform other known PET degrading enzymes under ambient conditions.[245] As a result, *Is*PETase has been the focus of much research and is the subject of the work in this chapter.

3.1.5 *Ideonella sakaiensis* PETase

The bacteria *I. sakaiensis* 201-F6 was discovered in a bottle recycling plant as part of a microbial consortium.[245] It appeared to have evolved to use PET as its major energy and carbon source for which it produces two enzymes (**Figure 3.3**). The first enzyme, *Is*PETase, is secreted by the bacteria onto the surface of the plastic where it breaks down the PET, producing mono(2-hydroxyethyl) terephthalic acid (MHET) as the major product as well as bis(2-hydroxyethyl) terephthalic acid (BHET) and TPA minor products. The second enzyme, named *Is*MHETase, converts the MHET into TPA and EG which are metabolised by the bacteria.



Scheme 3.3 *Is*PETase degrades PET producing MHET which is converted into TPA and EG by *Is*MHETase.

The crystal structure of *Is*PETase (**Figure 3.4**) showed it to belong to the serine hydrolase superfamily of enzymes.[246,247] It possesses an α/β hydrolase fold and a Ser-His-Asp catalytic triad (S160-H237-D206)[††] located within a broad hydrophobic substrate cleft.[247–249] The enzyme also contains two disulfide bonds, one situated at the C-terminus and the other near the active site, which have been shown to stabilise the enzyme.[247]

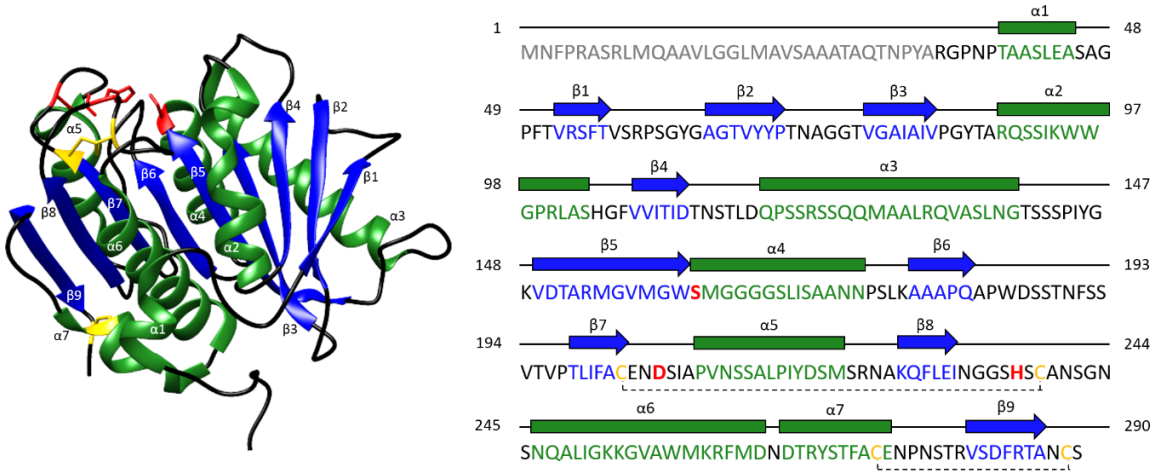
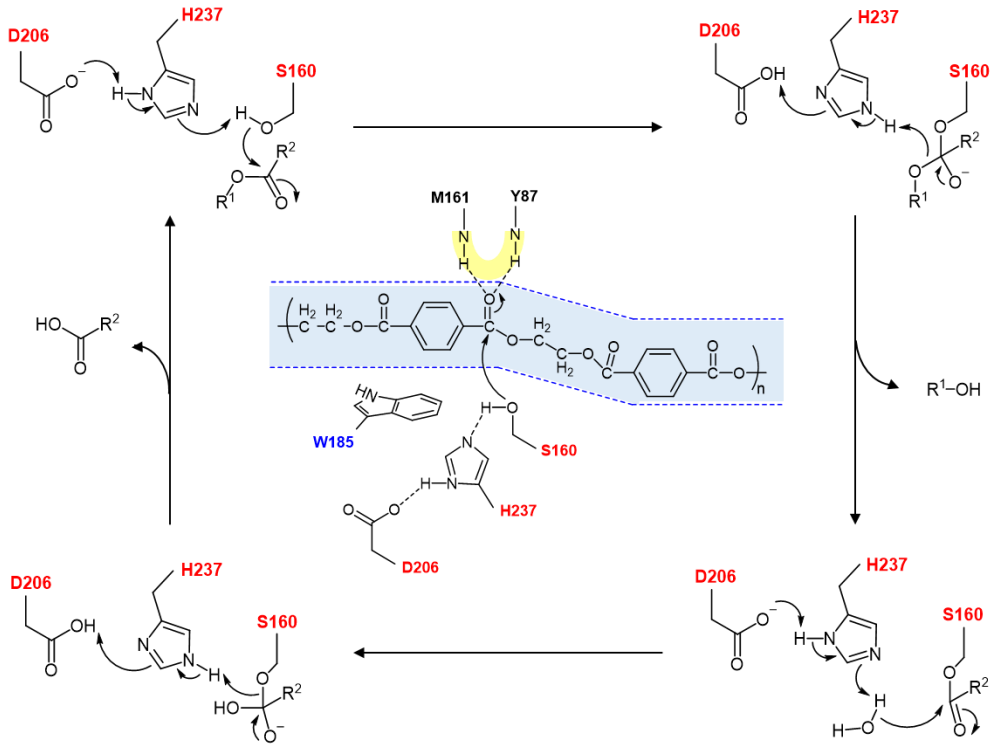


Figure 3.4 The crystal structure (PDB 6EQD) and sequence of *Is*PETase. The α-helices are coloured in green, β-sheets in blue, the signal peptide in grey, disulfide bond Cys in yellow and the catalytic triad in red.

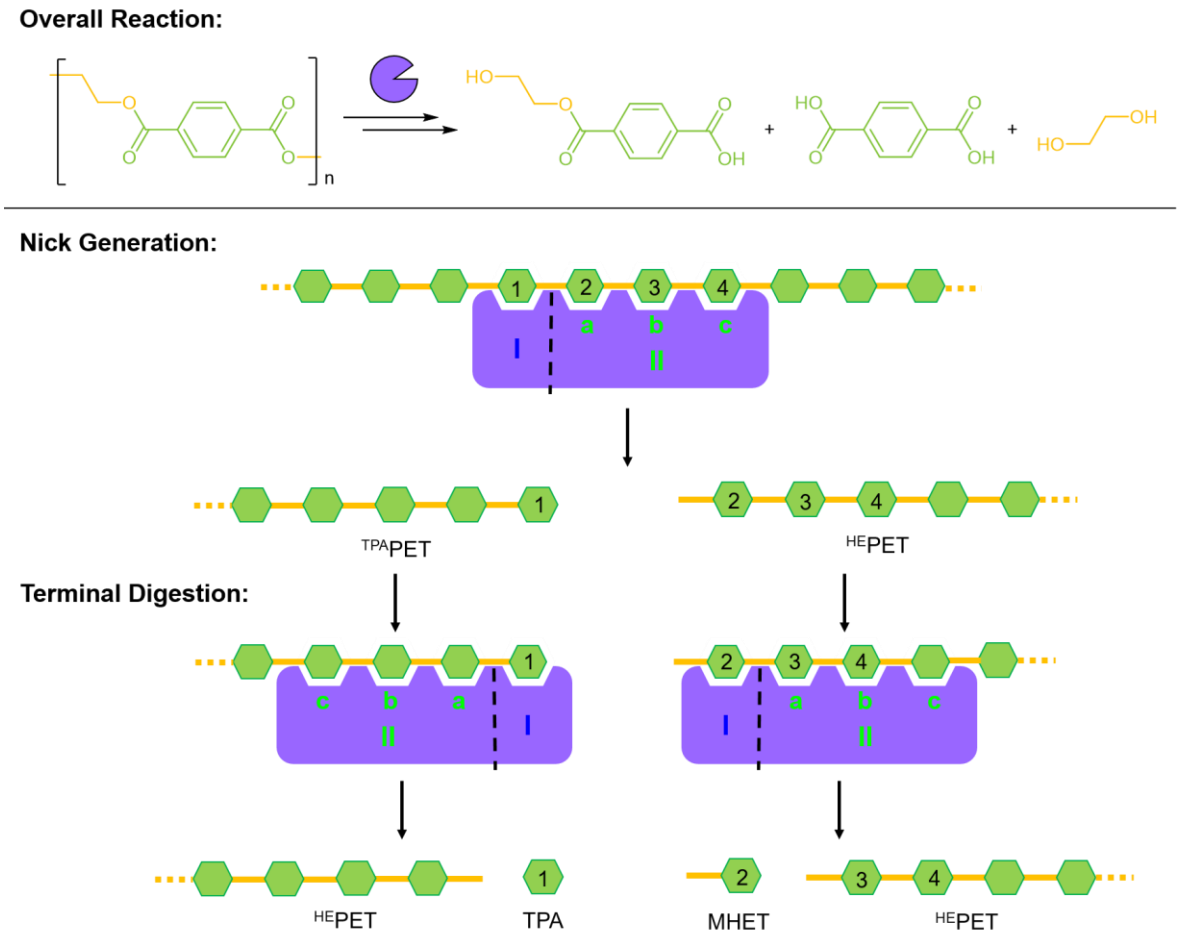
The *Is*PETase catalytic pathway proposed for PET degradation is shown in **Scheme 3.5**.[246] Upon binding of the substrate in the enzyme's shallow cleft, the first benzene ring of the substrate T-stacks with a "wobbling" tryptophan residue (W185). Meanwhile, the adjacent carbonyl group is directed towards the catalytic triad for nucleophilic attack by the activated Ser160. The oxyanion hole (M161 and Y87) stabilises the resulting tetrahedral intermediate. After formation of the acyl enzyme intermediate and subsequent nucleophilic attack by a water molecule, the ester bond is cleaved. The TPA terminated chain is rotated away from its original position by face-to-face stacking with the Trp185 residue, and the product is released from the binding site.

[††] Two different numbering systems are found in the literature for the amino acid sequence of *Is*PETase. Here, the system introduced by Joo et al.[247] is employed where numbering begins at the start Met (i.e. Met1) and includes the N-terminal signal peptide sequence.



Scheme 3.5 Proposed positioning of PET in the *Is*PETase substrate binding cleft and the serine hydrolase catalytic mechanism.

PET degradation by *Is*PETase has also been suggested to occur via a two-step mechanism (**Scheme 3.6**) in which the substrate binding site is divided into subsites I, IIa, IIb and IIc, each accommodating one of four MHET moities.[247] The catalytic Ser160 is located between subsites I and IIa. In the first *nick generation step*, cleavage of the scissile ester bond between subsites I and IIa generates two chains with either a TPA (^{TPA}PET) or a hydroxyl ethyl (HE) terminus (^{HE}PET). These chains are released from subsites I and II, respectively. In the second *terminal digestion step*, the ^{TPA}PET and ^{HE}PET chains are digested. For the ^{HE}PET chain, digestion produces an MHET monomer and another ^{HE}PET chain which is digested further, producing more MHET and more ^{HE}PET chains. Meanwhile, digestion of the ^{TPA}PET chain produces a TPA monomer and a ^{HE}PET chain which then also undergoes further digestion. Alternatively, the ^{TPA}PET and ^{HE}PET chains can be bound to the enzyme in the reverse direction so that one or two MHET moieties are positioned in subsite II, instead of three. Upon digestion, a mixture of monomer and dimer products are generated, which are eventually digested to MHET, TPA and EG. It should be noted that this mechanism, based on the covalent computational docking of an oligomeric substrate, has come under some criticism as it does not consider the conformations and motions of amorphous PET at room temperature, which would not have the free rotational properties required to fit into the substrate binding site as suggested.[250,251]



Scheme 3.6 Two-step mechanism of PET degradation by *Is*PETase. TPA and EG moieties are shown as green hexagons and orange lines, respectively. *Is*PETase is shown in purple.

3.1.6 Thermally Stable *Is*PETase Variants

*Is*PETase was found to degrade a PET film up to 120 times faster than several other known PET degrading enzymes at 30 °C.[245] However, being from thermophilic sources, these other enzymes show optimal PET degrading performances at increased temperatures where the mesophilic *Is*PETase is unable to function. Nevertheless, evidence suggests that improving the tolerance of *Is*PETase towards elevated temperatures (~70 °C) would impart increased PET degradation activity.[232] Therefore, much research has been directed towards improving the thermal stability of *Is*PETase.

Various rational design approaches have been employed to improve the stability and activity of *Is*PETase. It relies upon structural and functional information obtained from the crystal structure of *Is*PETase and its homologues. For example, the active site cleft of *Is*PETase was engineered to be more cutinase-like based on homology modelling with *T. fusca* cutinase.[252] By introducing two mutations (S238F and W159H), a narrower substrate binding cleft was achieved. This double mutant (*Is*PETase$^{S238F/W159H}$) exhibited improved

PET degradation (at 30 °C) compared to the wild type (WT) enzyme, suggested to be due to improved binding affinity of the substrate to the enzyme.

Using a different rational engineering approach the T_m of *Is*PETase was increased by around 9 °C to 57.6 °C.[253] This more thermally stable mutant (*Is*PETase$^{S121E/D186H/R280A}$) was created by extending the substrate binding cleft (R280A), stabilising the β6-β7 connecting loop by alleviating polarity collision (D186H) and introducing a H_2O-mediated hydrogen bond between α3 and α4 (S121E). As a result, the mutant enzyme showed improved activity at 30-40 °C and enabled the enzyme to remain active for longer at 50 °C compared to the WT enzyme. In a similar mutant (*Is*PETase$^{S121E/D186H/S242T/N246D}$), T_m was further improved (60 °C) by mutation of amino acid residues in the substrate binding site (S242T and N246D) for optimisation of the substrate binding affinity.[254] Subsequently, the introduction of a third disulfide bond between β8 and β9 strands of *Is*PETase$^{S121E/D186H/R280A}$ resulted in an even larger improvement to stability, with a T_m of 69.4 °C.[255] This thermally stable variant (TS-PETase) exhibited improved activity at both ambient and elevated temperatures (58 °C).

One of the most thermally stable *Is*PETase variants reported to date was designed via a computational strategy with minimal experimental screening.[256] The computationally redesigned enzyme (DuraPETase) exhibited a dramatically increased T_m of 77 °C and could remain active at 60 °C for up to 3 days. In addition, incubation of a PET film with DuraPETase at 37 °C for 10 days resulted in a 300-fold increase in the concentration of product release compared to the WT enzyme. This improved performance is suggested to originate from the reduction in the flexibility of the enzyme resulting in increased stability and better accommodation of the polymer substrate.

Chemical modification methods have also been used to improve the enzyme performance. For example, a zwitterionic polymer composed of alternating Glu and Lys residues was attached to the C-terminus of *Is*PETase to create a stabilising environment for the protein.[257] While this had little effect on the optimum temperature of the enzyme, the activity of the modified *Is*PETase was improved, likely due to a more compact and less aggregation prone structure. Similar effects were observed when the enzyme's amine groups (N-terminus and Lys residues) were modified through the attachment of polymer units exhibiting either hydrophobic, hydrophilic, positive or negative properties.[258]

Improvements to the stability of *Is*PETase have also been achieved by immobilisation. In two separate studies using different nanoparticle solid supports, the activity of the immobilised enzyme appeared to be prolonged compared to the free enzyme in

solution.[259,260] By attachment to magnetic iron oxide nanoparticles, *Is*PETase$^{S238F/W159H}$ could be recovered magnetically after incubation with PET and reused for 10 cycles, with around 50% activity remaining from cycle 4.[259] Immobilisation of *Is*PETase on flower-like cobalt phosphate nanoparticles resulted in improved tolerance to a broader range of pH and increased the optimal temperature to 45 °C (using a 4-nitrophenyl butyrate substrate).[260] Similarly, *Is*PETase exhibited improved stability and prolonged activity when displayed on the surface of yeast cells (*Pichia pastoris*).[261] In addition, this whole-cell biocatalyst remained active (~50%) after incubation with different additives including 10% methanol, 10% ethanol and 0.1% Triton X-100.

To summarise, the dependence of our society on plastic could have potentially devastating consequences for life on this planet if urgent action is not taken. Though various methods for the recycling of plastics have been developed, they are often environmentally and/or economically unviable. Therefore, the biodegradation of plastics is an attractive alternative. For the biodegradation of PET, there are a number of promising candidates. One of these is the enzyme *Is*PETase from the bacteria *I. sakaiensis*. However, improvements to its activity and stability are required if it is to be utilised for large scale recycling applications. Enhanced stability would be expected to be beneficial for the biodegradation of PET, due to the increased accessibility of the enzyme to the amorphous PET chains at the T_g of the plastic. Furthermore, stabilisation of *Is*PETase towards other industrially relevant conditions, such as non-physiological pH and agitation, would also be favourable for future application. In this chapter, the effect of cyclisation on *Is*PETase will be investigated, with the aim of improving the stability of the enzyme at elevated temperatures. In addition, the effect of cyclic topology on the PET degrading activity of the enzyme will be examined under a range of conditions.

3.2 RESULTS

To investigate the effects of cyclisation on *Is*PETase, three different cyclic *Is*PETase variants were prepared for comparison with the linear WT enzyme. These included a cyclic monomer, a cyclic dimer and a catenane (**Figure 3.7**). In the chapter below, the design, expression and purification of the different variants will first be described, followed by a comparison of stability and activity.

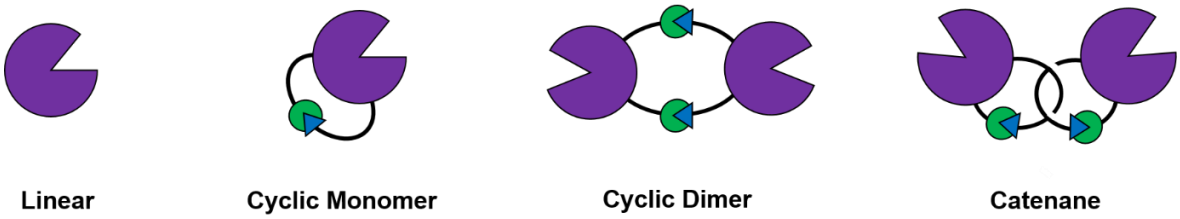


Figure 3.7 Cartoon representations of the linear and cyclic *Is*PETase variants.

3.2.1 Preparation of Linear and Cyclic *Is*PETase Variants

3.2.1.1 Linear *Is*PETase Preparation

The gene (see Appendix 7.1.2) encoding for the linear wild type *Is*PETase (*Is*PETase-WT) was purchased and cloned into a pET-21b(+) vector using Gibson Assembly. The plasmid was then transformed into calcium competent SHuffle T7 Express *E. coli* cells, which are known to promote the formation of intracellular disulfide bonds.[262] Expression was carried out in TB media where cells were grown at 30 °C to an OD_{600} of 0.8-1.0, before induction with a final concentration of 0.5 mM IPTG. The cells were then incubated overnight at 18 °C and harvested using centrifugation. Initially, the native secretion signal peptide sequence (Asp2-Ala27) was included at the N-terminus of the protein (see Appendix 7.2.2), but was later removed to enhance recombinant protein production (**Figure 3.8**).

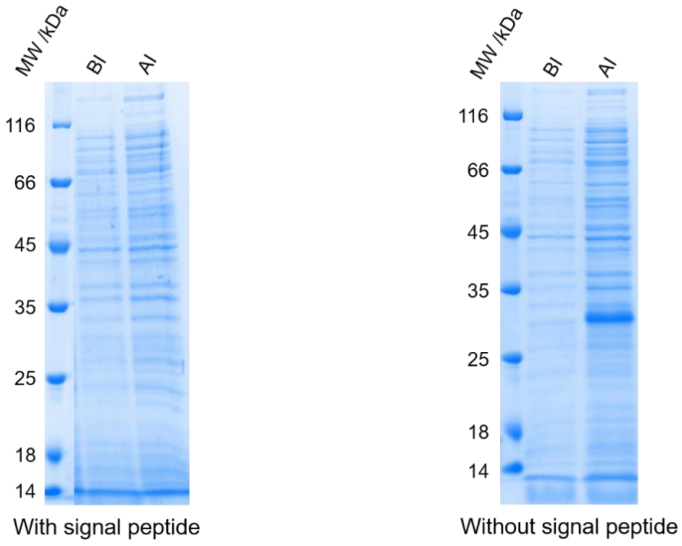


Figure 3.8 SDS-PAGE of *Is*PETase-WT expression with and without the native N-terminal secretion signal peptide, showing samples taken before and after induction with IPTG (BI and AI, respectively).

To purify *Is*PETase-WT, Ni-NTA affinity chromatography was employed, followed by SEC to remove any remaining impurities. The fractions containing the purified protein were combined and concentrated to give an overall yield of ~30 mg/L. **Figure 3.9** shows the SDS-PAGE and SEC chromatogram for the purification. A major band was observed by SDS-

PAGE in between the 25-35 kDa markers, corresponding to *Is*PETase-WT.

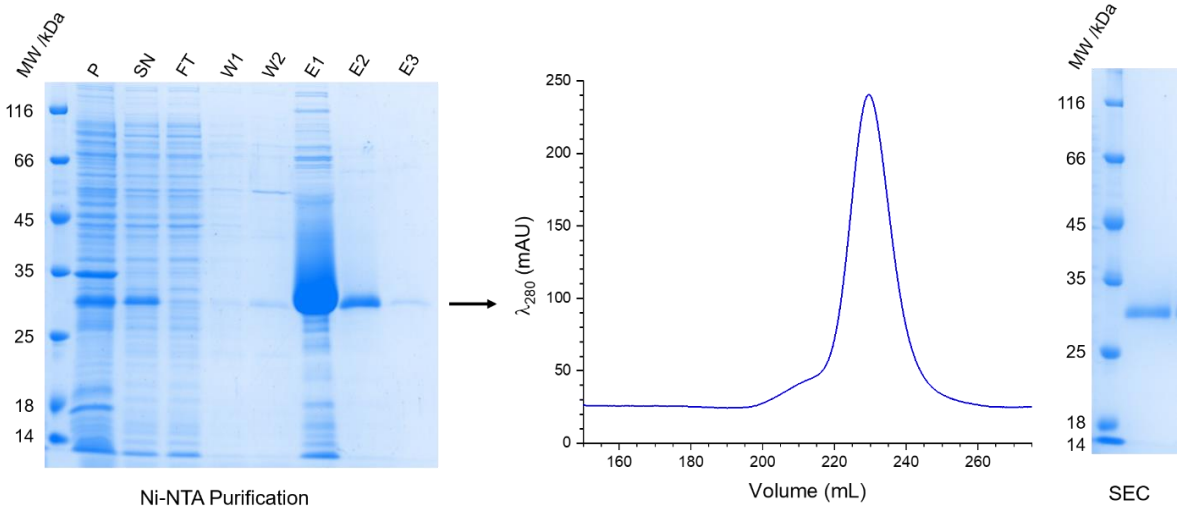


Figure 3.9 *Is*PETase-WT purification. (Left) SDS-PAGE analysis of the Ni-NTA purification, where P = pellet (i.e. insoluble fraction after centrifugation); SN = supernatant (i.e. soluble fraction after centrifugation); FT = flow through; W = wash; E = elution. (Middle) SEC chromatogram. (Right) SDS-PAGE analysis after purification by SEC.

In the SEC chromatogram, the elution of *Is*PETase-WT was observed as one main peak. This was confirmed by MS analysis, with a protein of the expected molecular weight for *Is*PETase-WT (28751 Da) shown to be present (**Figure 3.10**). A small shoulder is observed to the left of the main peak in the chromatogram, resulting from the elution of a protein with a molecular weight of 29010 Da. This is likely a post-translationally modified (PTM) variant of *Is*PETase-WT. By careful selection of SEC fractions, the unmodified protein could be isolated, avoiding a mixture of products.

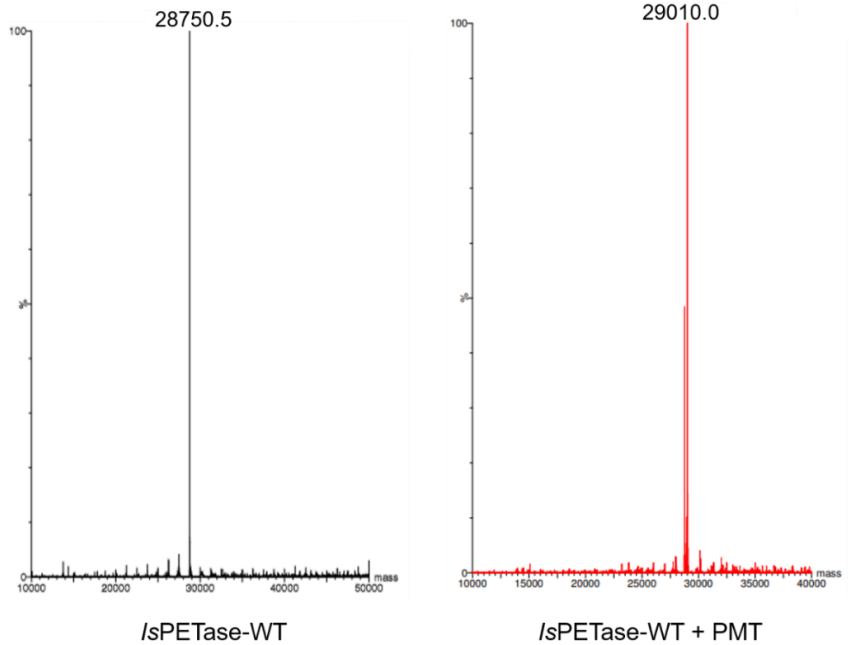


Figure 3.10 Deconvoluted ESI-MS of *Is*PETase-WT (with and without PMT).

3.2.1.2 Cyclic Monomer *Is*PETase Preparation

To cyclise *Is*PETase, SpyTag/SpyCatcher-N^{TEV} (see section 1.2.3.2) was employed due to its ability to undergo rapid isopeptide bond formation under a wide variety of conditions. In the *Is*PETase-Spy construct, the SpyTag and SpyCatcher-N^{TEV} sequences were placed at the C- and N-termini of the *Is*PETase sequence (see Appendix 7.1.2 and 7.2.2), respectively. Expression and purification were carried out as described above for *Is*PETase-WT, with the purified *Is*PETase-Spy obtained in a final yield of ~30 mg/L. By SDS-PAGE the cyclised *Is*PETase is observed as a band at 35 kDa (Peak 3, **Figure 3.11**). Due to the increased mobility of the cyclised protein through the polyacrylamide gel, this is lower than would be anticipated for a protein of an expected molecular weight of 42444 Da (confirmed by MS). A protein of a higher molecular weight was also eluted by SEC (Peak 2). This is likely dimeric *Is*PETase-Spy, produced via intermolecular isopeptide bond formation.

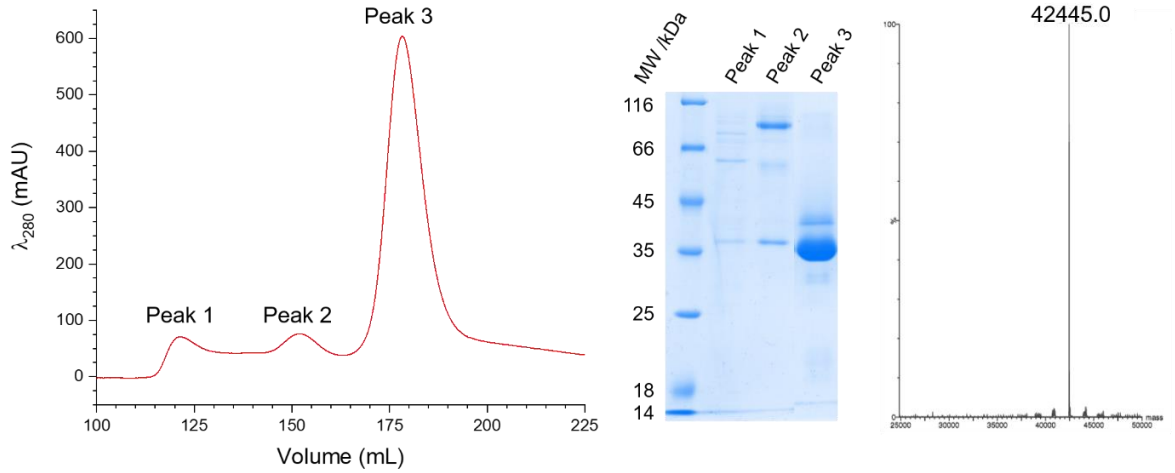


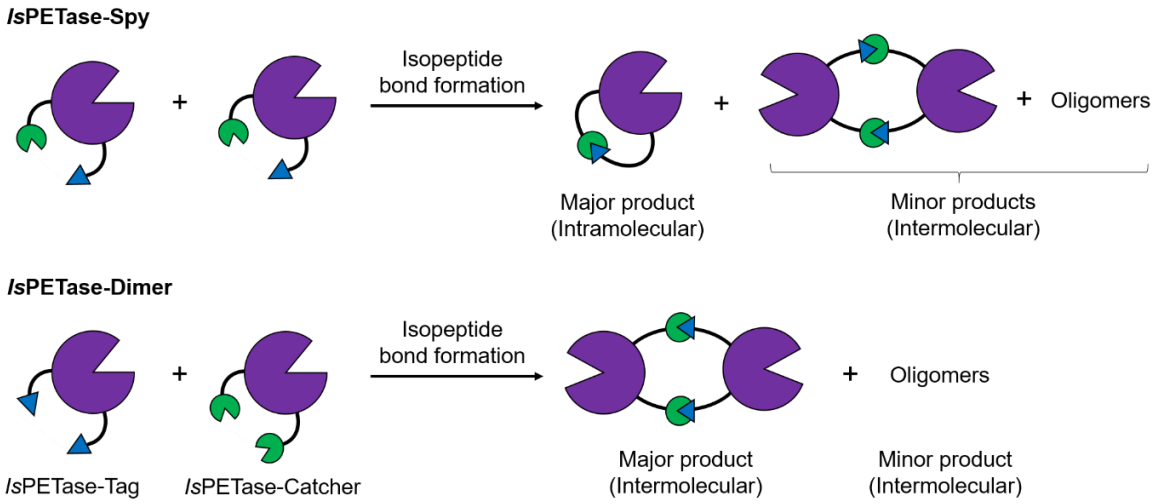
Figure 3.11 SEC chromatogram, SDS-PAGE and deconvoluted ESI-MS of purified *Is*PETase-Spy.

Both the linear WT and cyclic enzymes were obtained in a good yield. Notably, the purification of *Is*PETase-Spy was found to be more straightforward than that of the linear enzyme. For the purification of *Is*PETase-WT, the addition of large quantities of endonuclease to the cell lysate before Ni-NTA purification was required to avoid nucleic acid contamination. In addition, as described above, a small amount of suspected PTM *Is*PETase-WT was generated, although this could be separated to a reasonable extent using SEC. The effect of the PTM on the stability and activity of the enzyme was not tested.

3.2.1.3 Cyclic Dimer *Is*PETase Preparation

In the above preparation of the *Is*PETase cyclic monomer, some cyclic dimer side product was also produced due to intermolecular reaction between SpyTag and SpyCatcher fragments. When producing the *Is*PETase cyclic dimer (*Is*PETase-Dimer) as the desired

product for further characterisation and analysis, intermolecular reaction needed to be promoted over intramolecular reaction. To do this, two different *Is*PETase monomers were generated (**Scheme 3.12**). In one of these monomers SpyTag sequences were incorporated at both termini (*Is*PETase-Tag), while in the other monomer SpyCatcher fragments were located at the termini (*Is*PETase-Catcher). Consequently, the monomers would remain in their unreacted linear forms until mixed with their reactive partners, producing the desired *Is*PETase-Dimer. It should be noted that while SpyCatcher-N^{TEV} was employed in the cyclic monomer construct, the original (116 residue) SpyCatcher sequence (see **Figure 1.29**, section 1.2.3.2) was employed in the design of *Is*PETase-Dimer.



Scheme 3.12 Inter- versus intramolecular isopeptide formation. (Top) Intramolecular isopeptide bond formation leading to the formation of a cyclic monomer as the major product. (Bottom) Intermolecular isopeptide bond formation between *Is*PETase-Tag and *Is*PETase-Catcher producing a cyclic dimer as the major product.

The genes encoding for the *Is*PETase-Tag and *Is*PETase-Catcher sequences were each cloned into pET-28a(+) vectors. The expression and purification of each enzyme was carried out separately using the procedure described above for *Is*PETase-WT. Once both enzymes were purified, different ratios of *Is*PETase-Tag-to-*Is*PETase-Catcher were tested for optimum *Is*PETase-Dimer formation at room temperature for up to an hour (**Figure 3.13**). For all ratios tested, the cyclic dimer was rapidly formed in a high yield, with the most complete reaction observed when *Is*PETase-Tag was used in excess (3T:1C). Therefore, for larger scale reactions, the *Is*PETase-Tag monomer was used in excess to generate the *Is*PETase-Dimer. Moreover, the large size difference between the *Is*PETase-Tag starting material and the *Is*PETase-Dimer product would facilitate removal of the unreacted *Is*PETase-Tag by SEC.

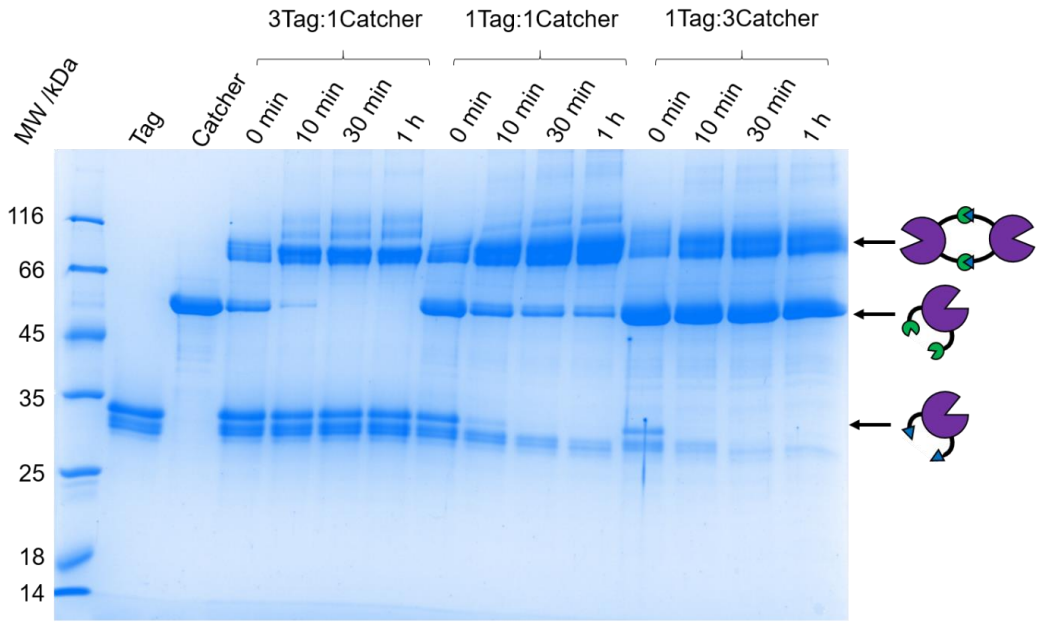


Figure 3.13 SDS-PAGE showing the different ratios of *Is*PETase-Tag-to-*Is*PETase-Catcher tested for *Is*PETase-Dimer formation at room temperature.

To shorten the purification procedure of *Is*PETase-Dimer, the separate SEC steps for each monomer were omitted. After separate Ni-NTA purifications, the *Is*PETase-Tag and –Catcher monomers were directly combined for *Is*PETase-Dimer formation before a single SEC step was carried out, reducing the total number of purification steps from five to three. In the SEC chromatogram (**Figure 3.14**), Peak 3 corresponds to the elution of *Is*PETase-Dimer, while Peak 4 corresponds to the elution of the excess *Is*PETase-Tag monomer.

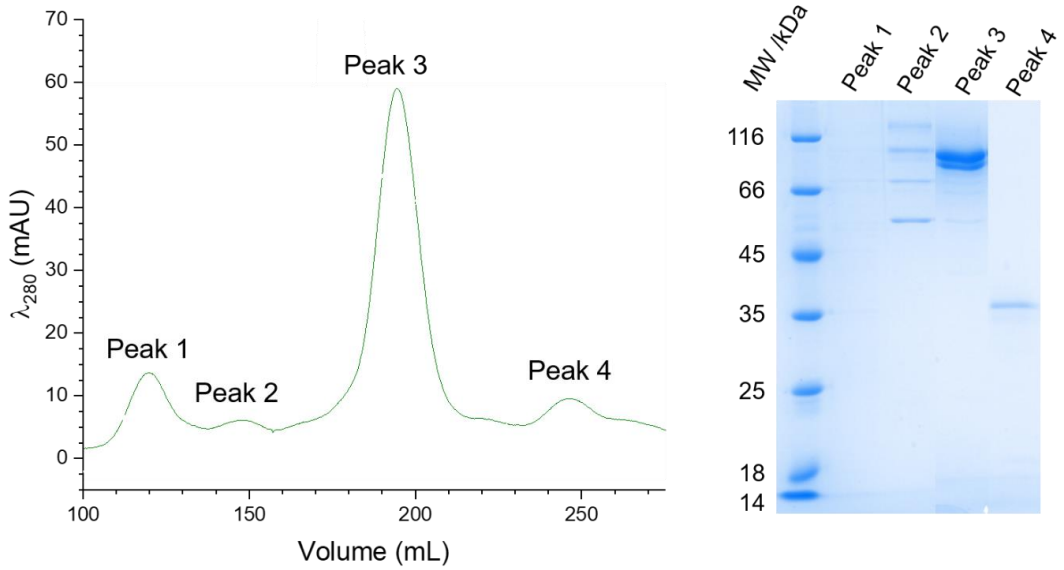
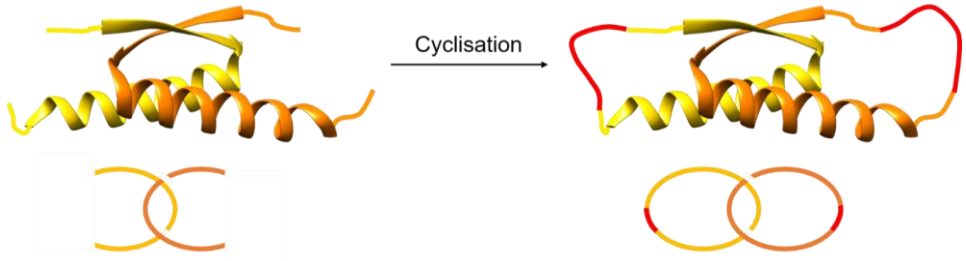


Figure 3.14 SEC chromatogram and SDS-PAGE analysis of *Is*PETase-Dimer.

3.2.1.4 Catenane *Is*PETase Preparation

Molecular catenanes are composed of interlinking rings. In nature protein catenanes have been found to confer thermal stability, most notably in the "capside chainmail" of the bacteriophage HK97, which protects the genetic information even under harsh conditions.[263] Other examples of protein catenanes found in nature include the *Pyrobaculum aerophilum* citrate synthase dimer,[264] bovine mitochondrial peroxiredoxin III[265] and *E. coli* class Ia ribonucleotide reductase.[266] Catenation has also been demonstrated to lead to the improved activity and stability of proteins in the laboratory.[267,268] Though published examples at the time of writing were limited, GFP, DHFR and γ-lactamase catenanes were all reported to have improved properties compared to their linear and cyclic counterparts.[269,270] Stability of protein catenanes is thought to result from a combination of (i) cyclisation of the protein subunits, and (ii) constraint of the relative motion of the subunits by the interlocking of the two chains.[271,272] Therefore, the catenation of *Is*PETase could be beneficial for enhancing plastic degradation.

For the production of the *Is*PETase catenane variant (*Is*PETase-Cat), a dimer mutant (M15E/L19K)[273] of the tetramisation domain of the tumour suppressor protein p53 was employed.‡‡ **Scheme 3.15** shows the two intertwined polypeptide chains of the p53 dimerisation domain. Upon intramolecular cyclisation of the separate polypeptide chains, a catenane is produced.[267]

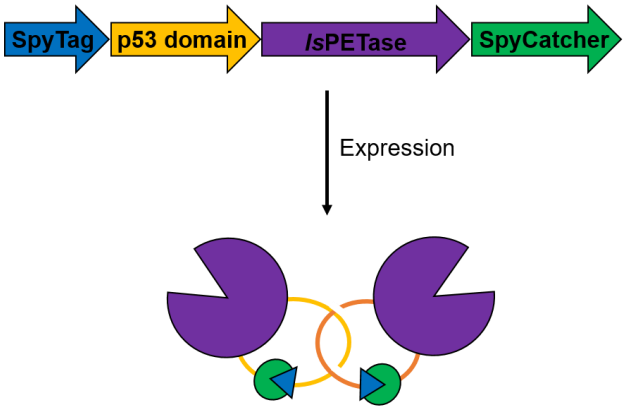


Scheme 3.15 Catenane formation by cyclisation of the p53 dimerisation domain (PDB 4D1L).

To create *Is*PETase-Cat (**Scheme 3.16**), the gene construct encoding for one monomer unit of the desired protein catenane (Appendix 7.1.2) was designed so that the *Is*PETase sequence was located in between the p53 dimerisation domain and SpyCatcher sequences at the N- and C-termini, respectively. The SpyTag sequence was then placed at the N-

‡‡ The tetramisation domain of the tumour suppressor protein p53 is composed of a dimer of polypeptide dimers (i.e. it is composed of a total of four separate polypeptide chains). The mutation specified inhibits interaction between the two sets of dimers, resulting in a species referred to as the "p53 dimerisation domain" (i.e. the two entwined polypeptide chains). The single un-entwined polypeptide chains are referred to as monomer subunits.

terminal of the p53 dimerisation domain. Upon expression in *E. coli*, the catenane was generated after the intertwining of two p53 dimerisation domain polypeptides followed by isopeptide bond formation between SpyTag and SpyCatcher. It should be noted that like in the *Is*PETase-Catcher construct, the original SpyCatcher sequence was employed for the preparation of *Is*PETase-Cat.



Scheme 3.16 Diagrams of the *Is*PETase-Cat gene construct and protein catenane.

The *Is*PETase-Cat variant was expressed and purified as described for *Is*PETase-WT. In the SEC chromatogram the major peak (Peak 3) corresponds to the elution of the protein catenane (**Figure 3.17**). The cyclic monomer was eluted shortly after and is observed as the shoulder (Peak 4) to the right of the main catenane elution peak. Despite the reasonably large difference in molecular weight, the cyclic monomer could not be completely removed by SEC, likely due to association. The desired catenane product has an expected molecular weight of ~98 kDa and is observed by SDS-PAGE as a band at approximately 116 kDa. Meanwhile, the lower molecular weight band at around 45 kDa is believed to be the cyclic monomer side product.

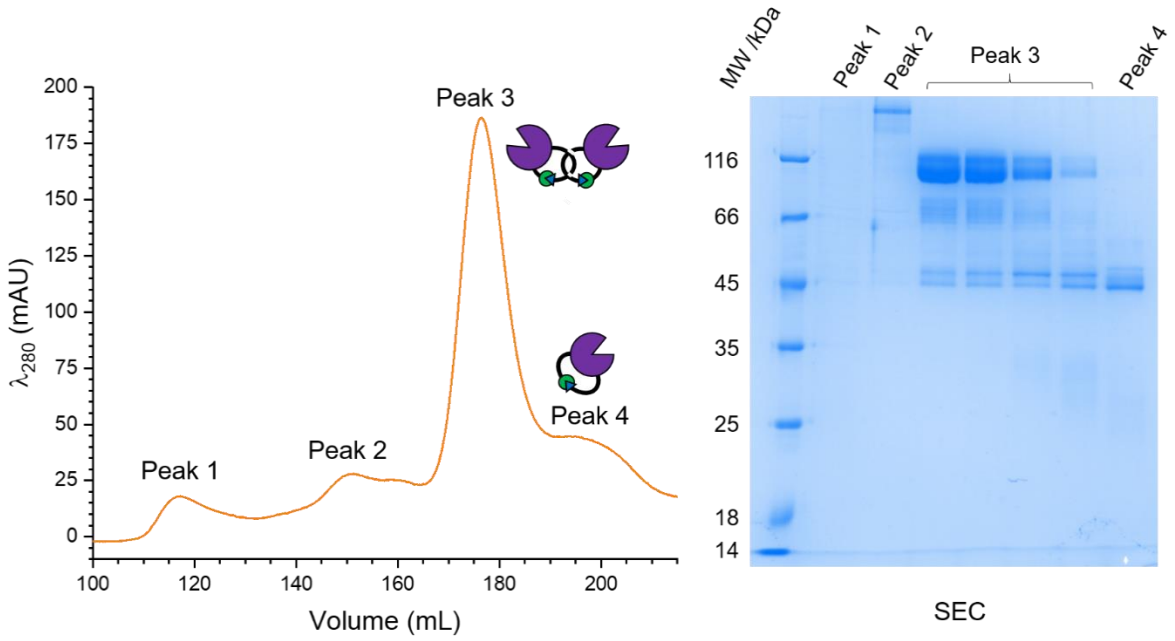


Figure 3.17 SEC chromatogram and SDS-PAGE analysis of *Is*PETase-Cat.

Various modifications to the expression and purification procedures of *Is*PETase-Cat were made in an attempt to reduce the cyclic monomer contamination, including the use of: (i) a slower flow rate during SEC to improve separation of the two proteins, (ii) an imidazole gradient for elution of the His-tagged proteins during Ni-NTA purification, (iii) a lower induction temperature during expression for slower protein production. Unfortunately, significant amounts of cyclic monomer contamination persisted after all of these changes.

The distance between the protein termini has previously been suggested to affect catenane formation efficiency.[268] Therefore, a 5 residue linker was incorporated into the *Is*PETase-Cat construct (*Is*PETase-Cat^{+5}), which was expressed and purified using the same procedures described for the other *Is*PETase variants. Though a significant quantity of cyclic monomer was still produced (**Figure 3.18**), its separation from the catenane by SEC appeared to be slightly improved compared to the *Is*PETase-Cat. As a result, all the stability and activity investigations discussed herein were carried out using the *Is*PETase-Cat^{+5} variant.

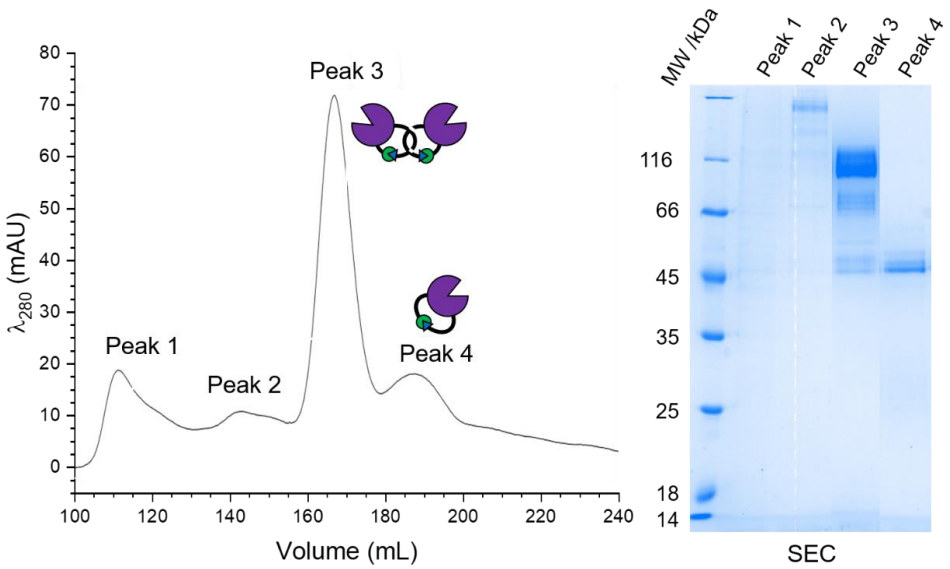


Figure 3.18 SEC chromatogram and SDS-PAGE of *Is*PETase-Cat^{+5}.

TEV protease digestion was carried out to confirm that the *Is*PETase-Cat produced was a catenane, rather than a cyclic dimer of the same molecular weight.[269,270] The TEV protease recognition sequence was incorporated during the design of the *Is*PETase construct in between the *Is*PETase and SpyCatcher sequences. Upon partial digestion by TEV protease, the catenane would be expected to generate linear and cyclic monomers. Meanwhile, the cyclic dimer would produce the linear dimer. Complete digestion of both the catenane and the cyclic dimer would yield the linear monomer. These different digestion products are distinguishable from one another by SDS-PAGE (**Figure 3.19**). Linear and cyclic monomer controls were created by mutation of the Asp7 to Ala in the SpyTag (*Is*PETase-Cat^{D7A}) and Lys19 to Pro in the p53 dimerisation domain (*Is*PETase-Cat^{K19P}). Partial digestion of both *Is*PETase-Cat and -Cat^{+5} were observed to produce the cyclic monomer, indicating that the desired catenane was present.

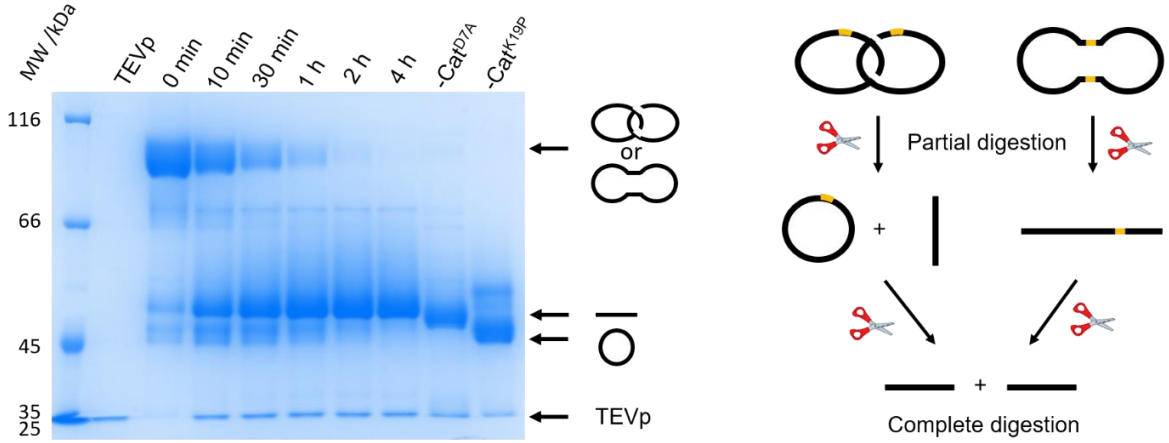


Figure 3.19 (Left) SDS-PAGE of the TEV protease (TEVp) digestion of *Is*PETase-Cat. (Right) A diagram of the expected partial and complete digestion products of a catenane and cyclic dimer.

3.2.2 Effect of Cyclisation on *Is*PETase Stability

After the successful expression and purification of *Is*PETase-WT, -Spy, -Dimer and –Cat^{+5}, the stabilities of the linear and cyclised enzymes were compared. In this section, the resistance of the *Is*PETase variants towards thermal, chemical and proteolytic treatment will be discussed.

3.2.2.1 *Is*PETase Thermal Stability

To compare the thermal stabilities of the *Is*PETase variants, CD was employed to analyse the changes in the proteins' secondary structures with increasing temperature, from which thermal denaturation curves were obtained. All measurements were carried out in buffer containing 50 mM Na_2HPO_4 (pH 8.0) and 100 mM NaCl.

3.2.2.1.1 Secondary Structure Analysis

The secondary structures of the linear and cyclised *Is*PETase enzymes were first analysed at 20 °C. **Figure 3.20** compares the CD spectra obtained for linear *Is*PETase-WT and cyclised *Is*PETase-Spy. A linear control of *Is*PETase-Spy is also shown in which the isopeptide bond forming Asp residue in the SpyTag sequence is mutated to Ala (*Is*PETase-Spy^{D7A}).

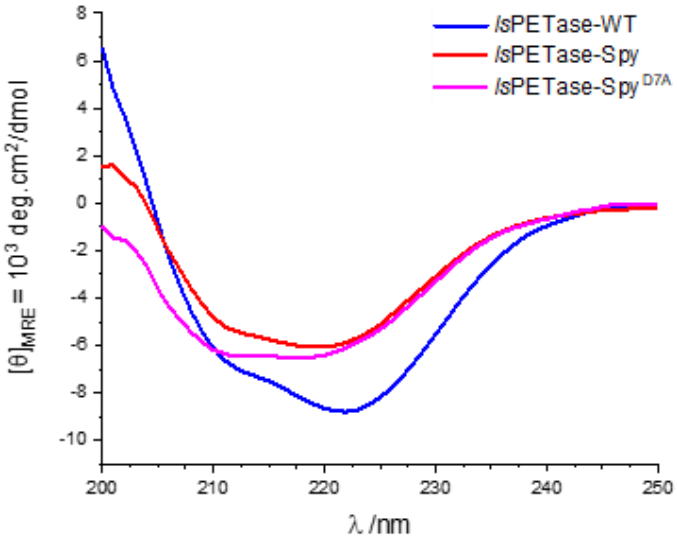


Figure 3.20 CD spectra of *Is*PETase-WT, -Spy and –Spy^{D7A} between 200-250 nm.

*Is*PETase-WT belongs to the α/β-hydrolase superfamily of enzymes, its structure consists of central twisted β-sheets surrounded by α-helices. As a result, the CD spectrum produced by the linear WT enzyme results from a combination of these secondary structural features. Using the crystal structure (PDB 6EQD, see **Figure 3.4**) and the primary structure of *Is*PETase, α-helices and β-sheets were found to make up around 36% and 22% of the secondary structure content, respectively. On the other hand, SpyTag/SpyCatcher(-N^{TEV})

complex does not contain any α-helices, being composed only of β-sheets (PDB 4MLI, see **Figure 1.27**).[145] Therefore, an overall reduction in the protein helical content is observed in the spectra of *Is*PETase-Spy and –Spy^{D7A}, likely due to the presence of the SpyTag and SpyCatcher-N^{TEV} sequences. Similarly, the CD spectrum of the cyclic dimer variant (**Figure 3.21**) suggests that the enzyme has a reduced helical content compared to the linear WT enzyme.

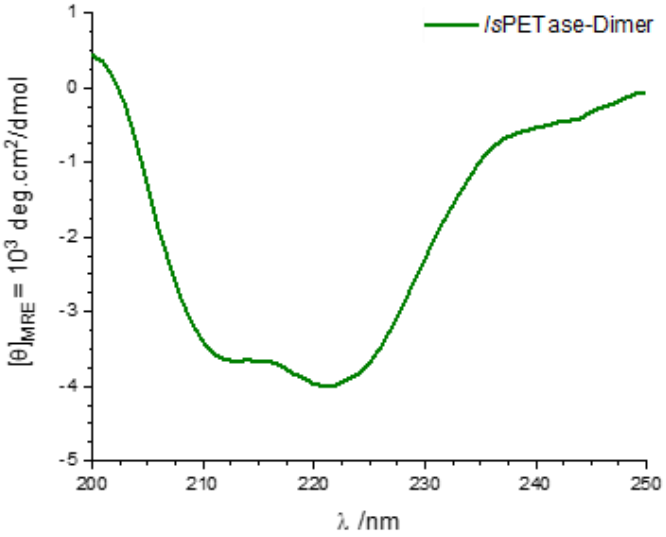


Figure 3.21 CD spectra of *Is*PETase-Dimer between 200-250 nm.

For completeness, the CD spectrum of *Is*PETase-Cat^{+5} is show (**Figure 3.22**). However, as the sample also contains some cyclic monomer contaminant, caution should be exercised in its interpretation as the catenane secondary structure is not the only contributor. It would be expected for *Is*PETase-Cat^{+5} to exhibit a decrease in helical secondary structure content compared to *Is*PETase-WT due to the incorporation of the SpyTag/SpyCatcher sequences. Nevertheless, a slight increase in helical character would be expected compared to *Is*PETase-Spy and –Dimer resulting from the p53 dimerisation domain which is composed of two intertwined helices.

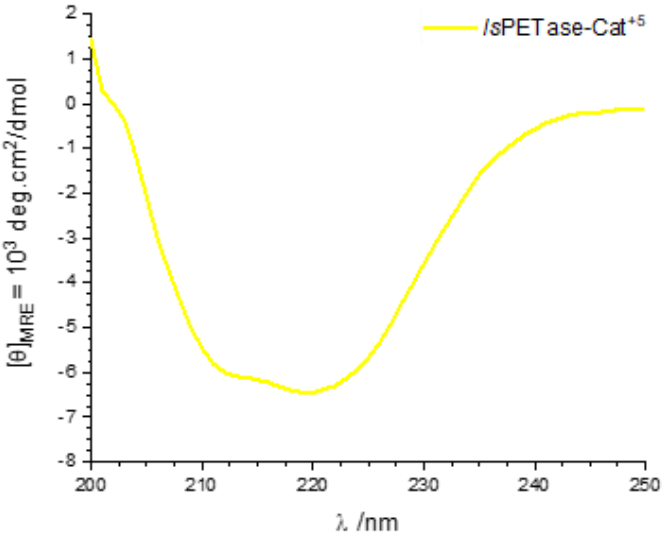


Figure 3.22 CD spectra of *Is*PETase-Cat^{+5} between 200-250 nm.

Overall, the CD data measured at 20 °C suggests that the structure of *Is*PETase is not considerably altered as a result of cyclisation, dimerisation or catenation. It is likely that the additional β-sheet content of the SpyTag/SpyCatcher(-N^{TEV}) machinery is responsible for the majority of the differences observed.

3.2.2.1.2 Melting Temperature

The thermal stabilities of the linear and cyclised *Is*PETase variants were determined by following the unfolding of each protein between 10 and 90 °C using the CD signal at 222 nm. From the thermal denaturation curves obtained, the mid-point of unfolding corresponded to the T_m. **Figure 3.23** shows the T_m measured for *Is*PETase-WT, -Spy and -Spy^{D7A}.

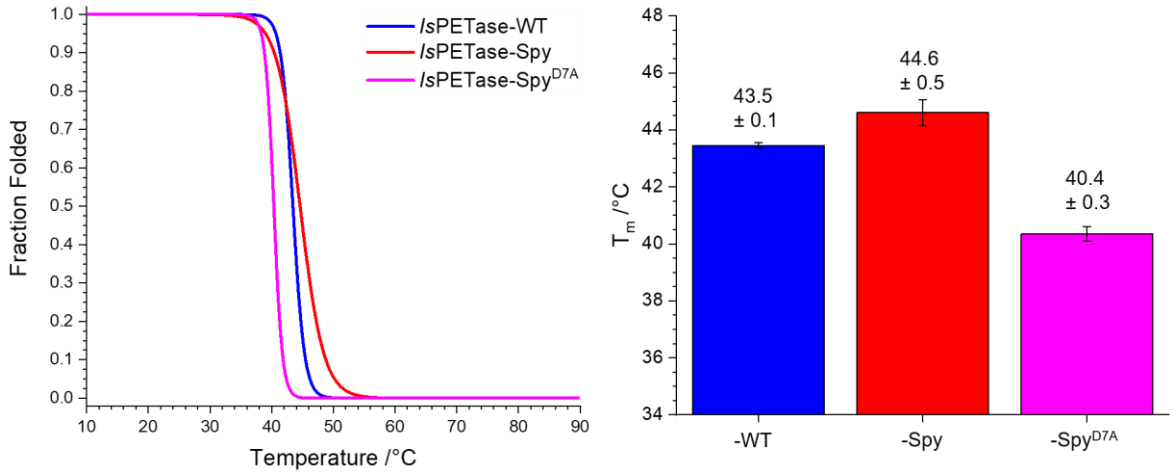


Figure 3.23 (Left) Thermal denaturation curves of *Is*PETase-WT, -Spy and –Spy^{D7A}. (Right) Bar chart comparing T_m. Error bars represent the 95% confidence interval.

For *Is*PETase-WT a T_m of 43.5 °C was obtained. This is slightly lower than values previously reported in the literature (46.8 and 48.8 °C),[247,253] likely due to differences in the experimental techniques employed. For *Is*PETase-Spy, the T_m was found to be 44.6 °C, indicating that thermal stabilisation of *Is*PETase was not achieved by cyclisation. Meanwhile, the linear control was found to be destabilised compared to both the linear WT and cyclised enzymes with a T_m of 40.4 °C.

The length of the linker between the *Is*PETase sequence and SpyTag/SpyCatcher-N^{TEV} was examined. Extension of the linker would be expected to relieve any destabilising strain resulting from distortion of the protein due to cyclisation. The T_m values measured for the cyclised *Is*PETase variants with 5 and 10 additional residues in the linker (*Is*PETase-Spy^{+5} and -Spy^{+10}) were 44.6 ± 0.5 and 44.7 ± 0.4 °C, respectively, showing no significant improvement to thermal stability. On the other hand, reducing the linker length would be expected to introduce strain, leading to a reduction in T_m. However, removal of 5 residues from the linker (*Is*PETase-Spy^{-5}) also did not result in significant change to the T_m of the protein (44.2 ± 0.5 °C).

The SEC chromatogram overlays of *Is*PETase-Spy, -Spy^{+5} and –Spy^{-5} are shown in **Figure 3.24**. The elution of both *Is*PETase-Spy and –Spy^{+5} are shown as single, nearly symmetrical peaks, whereas elution of *Is*PETase-Spy^{-5} is observed as a continuous broad peak. SDS-PAGE analysis of these fractions (*A*-*J*) showed that a protein of the expected molecular weight for *Is*PETase-Spy^{-5} was present.

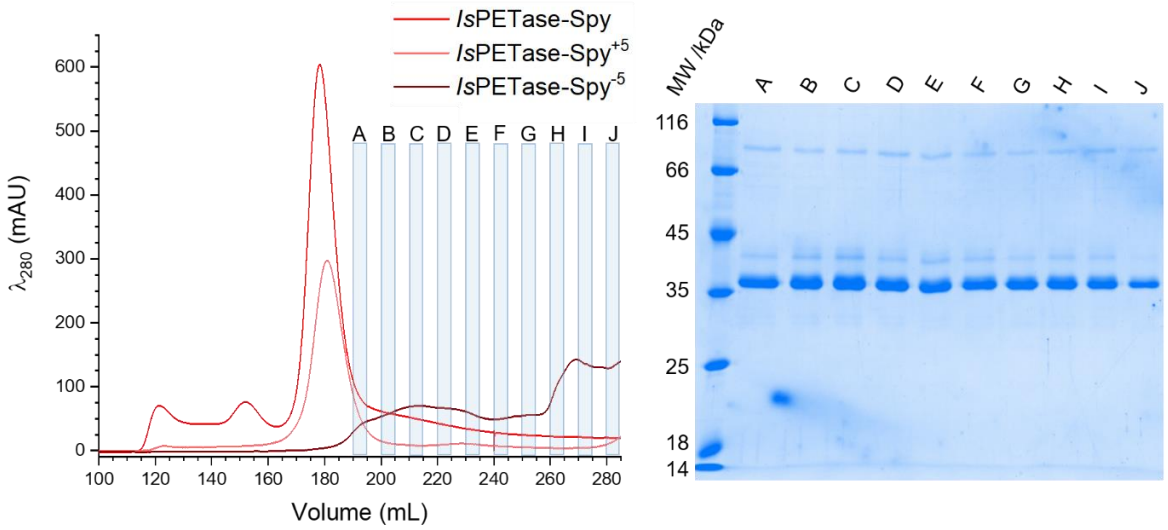


Figure 3.24 (Left) *Is*PETase-Spy SEC chromatogram overlay. The fractions of *Is*PETase-Spy^{-5} collected for further analysis are labelled *A*-*J*. (Right) SDS-PAGE analysis of the SEC fractions *A*-*J*.

A large variation in the folding of *Is*PETase-Spy^{-5} was observed using CD spectroscopic analysis (**Figure 3.25**). For the protein eluted by SEC at a lower retention volume, the CD spectra obtained were similar to those of the other mono-cyclic variants. This indicates that the enzyme eluted in these fractions had a folded secondary structure. Hence, it is possible the tertiary structure may also be intact. However, as elution volume increased, the less secondary structure the proteins exhibited. In particular, the CD spectrum of fraction *I* is characteristic of a random coil, suggesting the presence of misfolded protein. Therefore, the reduction in the linker length of *Is*PETase-Spy likely causes distortion and misfolding, although the resulting destabilisation was not captured by the T_m measurement. This could be because fraction *D* was chosen for T_m analysis, which still exhibits a reasonable extent of folding. It would perhaps be more representative to measure the thermal stability over a range of fractions, or after combining the different fractions.

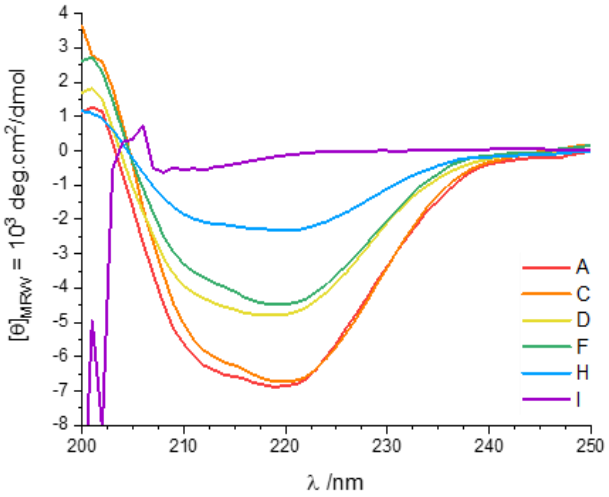


Figure 3.25 CD spectra of *Is*PETase–Spy^{-5} SEC fractions.

Next the T_m of *Is*PETase-Dimer and –Cat^{+5} were determined (**Figure 3.26**). *Is*PETase-Dimer was found to have a T_m of 42.4 °C. Meanwhile, a T_m of 41.4 °C was measured for *Is*PETase-Cat^{+5}, although this value should be interpreted with caution due to the cyclic monomer impurity also present in the sample. Nevertheless, the lack of change was unexpected, particularly for *Is*PETase-Cat^{+5}. The few protein catenanes that have previously reported exhibited improved stability compared to both their linear and cyclic counterparts. [269–271]

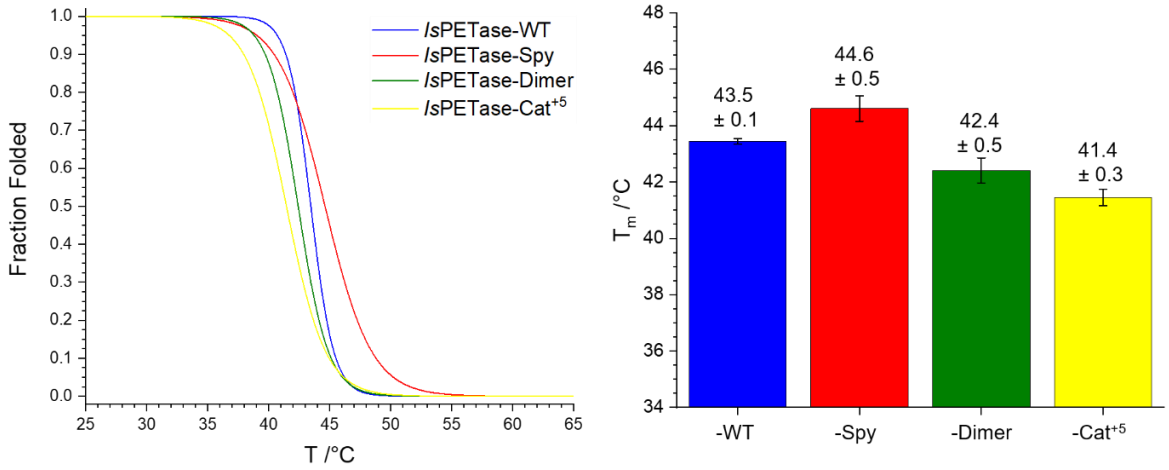


Figure 3.26 (Left) Thermal denaturation curves of *Is*PETase-WT, -Spy and –Dimer and –Cat^{+5}. (Right) Bar chart comparing T_m. Error bars represent the 95% confidence interval.

3.2.2.2 *Is*PETase Proteolytic Stability

To further compare the stabilities of the *Is*PETase variants, trypsin digest was employed. Trypsin is a serine protease which specifically cleaves polypeptides at the C-terminal side of Lys and Arg amino acid residues.[274] It would be expected that the cyclised variants would be less susceptible to proteolytic digestion because in theory, cyclisation restricts the number of conformations accessible to the protein's unfolded state. As a result, it is less likely for protease recognition sequences to be exposed and accessible for protease attack.[275]

Trypsin digest of the *Is*PETase variants was performed at 30, 40 and 50 °C in buffer containing 50 mM Na_2HPO_4 (pH 8.0) and 100 mM NaCl. A 1:100 molar ratio of trypsin-to-*Is*PETase was employed. The progress of the reaction was monitored by taking SDS-PAGE samples at timed intervals for up to 4 hours. **Figure 3.27** shows the SDS-PAGE analyses of the trypsin digest of the linear and cyclic variants at increasing temperature.

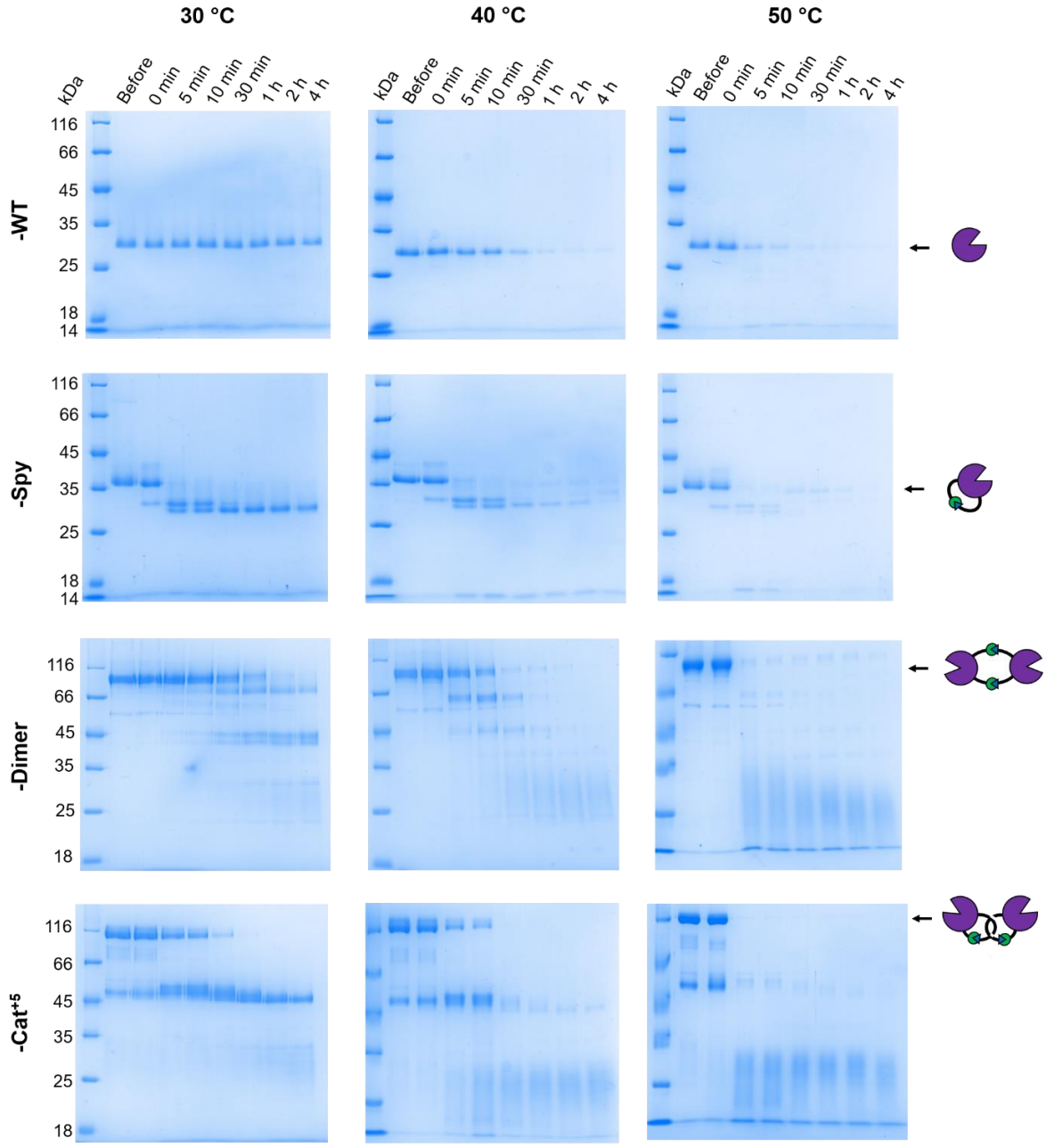


Figure 3.27 SDS-PAGE analysis of trypsin digest of the linear and cyclic *Is*PETase variants at 30, 40 and 50 °C. 50 mM Na_2HPO_4 (pH 8.0) and 100 mM NaCl buffer was used with a 1:100 molar ratio of trypsin-to-*Is*PETase.

Surprisingly, at 30 °C, *Is*PETase-WT appeared to be resistant to trypsin digest, with no observable difference in the size or location of the main band corresponding to the enzyme, even after 4 hours. In addition, no digestion products were detected. On the other hand, digestion of all of the cyclised variants was observed. This indicates that the proteolytic stability of the enzyme was reduced as a result of cyclisation. Nevertheless, the extent of trypsin digestion between the cyclic variants did vary. The most rapid digestion was observed for *Is*PETase-Spy, with complete digestion occurring almost immediately (<5 minutes). *Is*PETase-Dimer was the most resistant cyclic variant to proteolytic digestion,

though it was still almost completely digested after about an hour. *Is*PETase-Cat^{+5} was mostly digested after 30 minutes.

Here, “complete digestion” is defined as the disappearance of the SDS-PAGE band corresponding to the undigested protein. However, the formation of partial digestion products were observed as bands of lower molecular weights. These partial digestion products appear to be relatively resistant to proteolytic digestion at 30 °C. On closer inspection, the primary partial degradation product of *Is*PETase-Spy has a molecular weight similar to that of the linear WT enzyme, which demonstrated resistance to digestion (at 30 °C for <4 h). This could suggest that the *Is*PETase sequence has high proteolytic stability and that the digestion observed is of the SpyTag/SpyCatcher(-N^{TEV}) sequences. However, this contrasts with previous literature reports in which intramolecular isopeptide bonds have been demonstrated to confer increased resistance to proteolytic digestion.[276,277] Furthermore, partial digestion products of the dimeric enzymes are observed at around 45 kDa by SDS-PAGE. This is roughly the molecular weight of one subunit, and could indicate digestion of the SpyTag/SpyCatcher domain while the *Is*PETase sequence remains untouched.

At 40 and 50 °C, it would be expected that proteolytic digestion of the *Is*PETase variants would occur more rapidly due to: (i) the increased activity of trypsin, and/or (ii) the increased unfolding of *Is*PETase whereby protease cleavage sequences are exposed. As predicted, digestion of all variants was observed after shorter amounts of time at 40 °C. This included *Is*PETase-WT, which was digested after roughly 30 minutes. The dimeric enzymes were both digested after around 10 minutes. Meanwhile, *Is*PETase-Spy was still digested almost immediately. Incubation of the enzymes with trypsin at 50 °C, further increased the rate of digestion. The majority of *Is*PETase-Spy, -Dimer and –Cat^{+5} were digested before the 5 minute time point. Meanwhile, the linear WT enzyme was digested after around 10 minutes. The proteolytic stability of *Is*PETase-Spy^{D7A} was also investigated using trypsin digest at 30 °C (**Figure 3.28**). Similar to cyclic *Is*PETase-Spy, complete digestion of the linear control was observed in under 5 minutes. This further suggests that it is the SpyTag/SpyCatcher(-N^{TEV}) sequences that are susceptible to proteolytic digestion, rather than the *Is*PETase fragment.

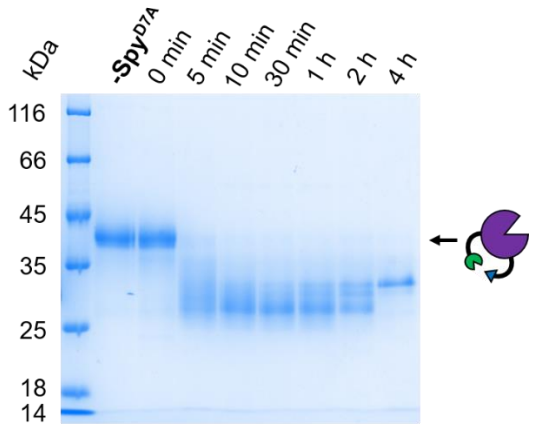


Figure 3.28 SDS-PAGE analysis of trypsin digest of linear *Is*PETase-Spy^{D7A} at 30 °C.

3.2.2.3 *Is*PETase Chemical Stability

The proteolytic stability of a protein can indicate that it is kinetically stable, i.e. that the protein is trapped in a specific conformation due to a high activation energy for unfolding (see section 3.3.2).[278] As a result, proteins under kinetic control generally exhibit slow unfolding, enabling them to resist proteolytic digestion, misfolding aggregation and harsh detergents.[275] Previously, it has been demonstrated that kinetically stable proteins can be identified through their resistance to SDS-induced denaturation, by comparing the migration of boiled and unboiled samples on a polyacrylamide gel.[279,280] Proteins that are at least partially resistant to SDS-induced denaturation exhibit slower migration in the absence of boiling. This is due to reduced SDS binding and lower overall negative charge, compared to those fully bound by SDS. On the other hand, proteins which are denatured by SDS migrate to the same location on the polyacrylamide gel as their boiled counterparts.[279]

As protein cyclisation has been demonstrated to stabilise the native state by restricting the mobility of the protein backbone,[166] the stabilities of the linear and cyclic *Is*PETase variants towards SDS-induced denaturation were investigated. Being thermally labile (T_m = 43.5 °C) *Is*PETase-WT was not expected to exhibit resistance to SDS-induced denaturation, despite being relatively resistant to trypsin digest (at 30 °C). Indeed, a protein can be proteolytically stable without being kinetically stability.[279]

To examine the stability of the *Is*PETase variants towards SDS-induced denaturation, two equivalent protein samples were first prepared using SDS-PAGE loading buffer. One sample was boiled for 10 minutes, while the other was left unheated. The migrations of the two samples were then compared using SDS-PAGE (**Figure 3.29**).

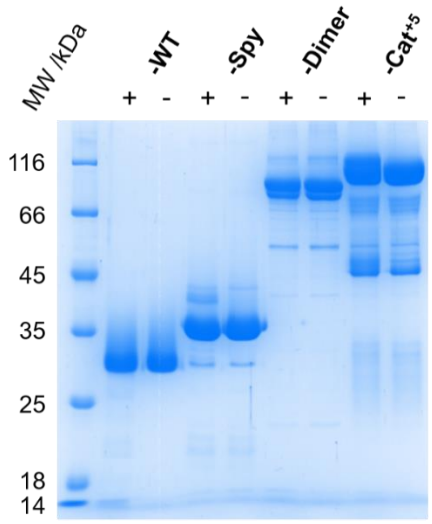


Figure 3.29 Comparison of boiled (+) and unboiled (-) samples of the linear and cyclic *Is*PETase variants.

As expected, the boiled and unboiled samples of *Is*PETase-WT were observed to migrate to the same position on the gel. This indicated that the enzyme was denatured by SDS, and so was not kinetically stable. Similarly, the cyclised *Is*PETase variants showed no significant difference in the distance migrated between the boiled and unboiled samples. It follows that cyclisation does not confer *Is*PETase with improved resistance towards SDS-induced denaturation or improved kinetic stability.

3.2.2.4 *Is*PETase Stability Summary

To investigate thermal stability of the *Is*PETase variants, CD was employed to monitor the unfolding of the proteins' secondary structures as the temperature was increased. All of the variants tested, were found to have similar T_m, indicating that cyclisation did not result in an improvement to the thermal stability of *Is*PETase. Next, the proteolytic stabilities of the *Is*PETase variants were examined. Unexpectedly, *Is*PETase-WT was found to be the most resistant to trypsin digest. In contrast, *Is*PETase-Spy was observed to be digested almost immediately at all the temperatures tested. Meanwhile, *Is*PETase-Dimer and -Cat^{+5} appeared to be less susceptible to protetolysis than the cyclic monomer, but were still digested more rapidly than the linear WT enzyme. This decrease in proteolytic stability could result from the presence of the SpyTag/SpyCatcher(-N^{TEV}) sequences. Finally, the chemical stabilities of the linear and cyclic *Is*PETase variants were investigated using SDS, which led to the unfolding of all *Is*PETase variants regardless of topology. Overall, cyclisation was not found to be beneficial for the thermal, proteolytic or chemical stabilisation of *Is*PETase (**Table 3.1**).

Table 3.1 Summary of stability experiments.

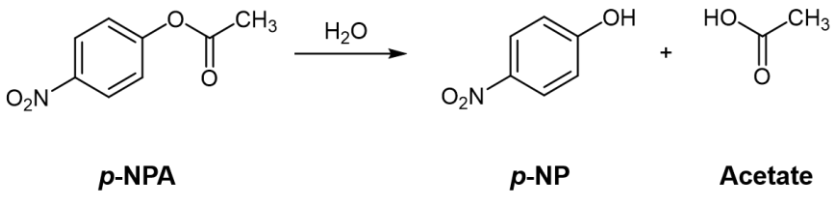
Stability Experiment		-WT	-Spy	-Dimer	-Cat^{+5}	Conclusion
T_m /°C		43.5	44.6	42.4	41.4	Cyclisation did not improve thermal stability
Resistance to trypsin	30 °C	>4 h	<5 min	~1 h	~30 min	Cyclisation reduced proteolytic stability
	40 °C	~30 min	<5 min	~10 min	~10 min	
	50 °C	<5 min	<5 min	<5 min	<5 min	
SDS		Denatured	Denatured	Denatured	Denatured	Cyclisation did not improve resistance to SDS-induced denaturation

3.2.3 Effect of Cyclisation on *Is*PETase Activity

To compare the activities of the linear and cyclised *Is*PETase variants two different activity assays were chosen. In the first assay, the hydrolysis of *p*-nitrophenyl acetate (*p*-NPA) was employed to detect esterase activity and enabled a comparison of the enzymes' residual activities after heat treatment. As the *p*-NPA substrate is not realistic of the insoluble and hydrophobic structure of PET, a second absorbance assay was used to detect the soluble PET degradation products released after treatment of PET by the different enzymes at increasing temperatures.

3.2.3.1 Hydrolysis of *p*-NPA

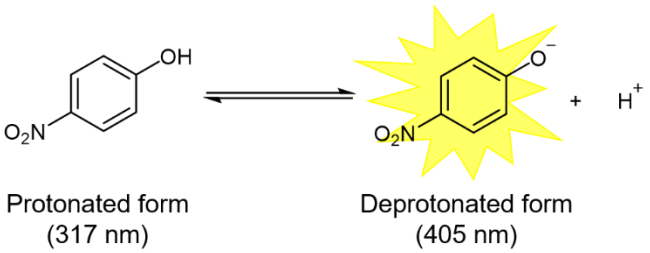
The hydrolysis of *p*-NPA is commonly used to detect the esterase activity of enzymes such as lipases and hydrolases.[281] In addition, it has been employed to investigate the hydrolysis kinetics of *Is*PETase and other PET degrading enzymes.[238,282] Upon the hydrolysis of the *p*-NPA ester, *p*-nitrophenol (*p*-NP) and acetate products are generated (**Scheme 3.30**). The release of the *p*-NP can be measured spectrophotometrically by monitoring absorbance at 405 nm (ε = 18400 M^{-1} cm^{-1}).[238]



Scheme 3.30 *p*-NPA hydrolysis producing *p*-NP.

Two important factors that should be taken into consideration when carrying out enzyme activity assays using *p*-NPA are as follows. Firstly, significant autohydrolysis of the *p*-NPA substrate occurs in aqueous solution. As a result, the rate of non-enzymatic hydrolysis should be subtracted from the measured absorbance value to give the rate of enzymatic hydrolysis. Secondly, *p*-NP absorbance is both pH and temperature dependent, arising from the presence of protonated and deprotonated forms of *p*-NP. The protonated form exhibits

an absorbance maximum at 317 nm, while the deprotonated *p*-NP has a maximum at 405 nm (**Scheme 3.31**).[283] As a result, changes in pH and/or temperature lead to changes in the ratio of protonated versus deprotonated molecules, and therefore the absorbance measured.



Scheme 3.31 Protonated and deprotonated forms of *p*-NP forms and their absorbance maxima.

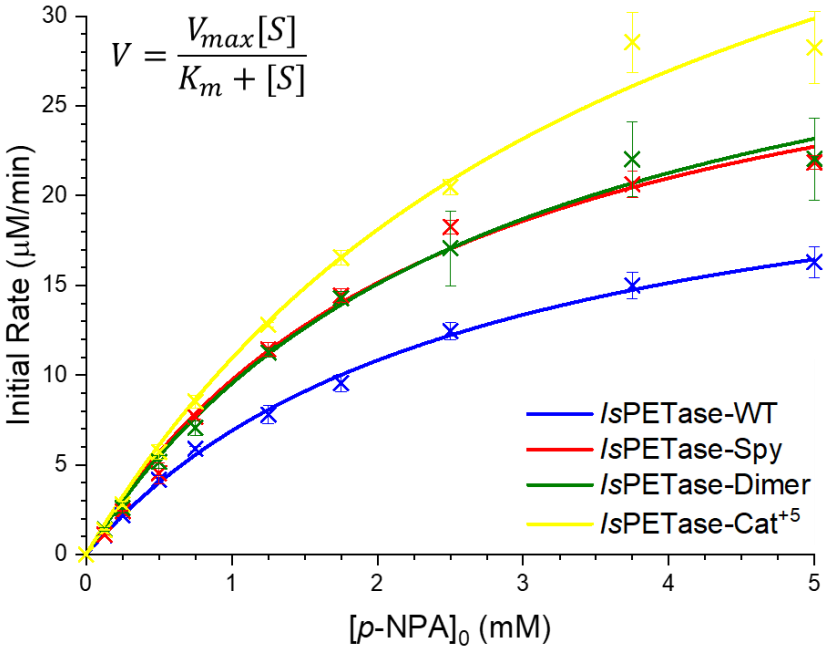
Here, all *p*-NPA activity assays were carried out in a 50 mM Na_2HPO_4, pH 7.5, 100 mM NaCl buffer. The use of a phosphate buffer limited variation in pH with temperature (compared to e.g. Tris), while a pH of 7.5 was selected to enable good enzyme functionality and provided a reasonable absorbance signal at 405 nm. In addition, 5% DMSO containing the *p*-NPA substrate of the appropriate concentration was present in all samples. A reaction temperature of 30 °C was chosen as it is the typical temperature for which *Is*PETase activity is reported.

3.2.3.1.1 *Is*PETase Kinetics

Activity assays for the linear and cyclised *Is*PETase variants were carried out in triplicate using the conditions described above, measuring the increase in absorbance at 405 nm for 300 s. *p*-NPA substrate concentrations ranged from 0 to 5.0 mM. Due to substrate insolubility, the use of higher *p*-NPA concentrations was not possible. Regarding enzyme concentration, 10 nM of the monomeric enzymes (*Is*PETase-WT and –Spy) were employed. Meanwhile, 5 nM of the dimeric enzymes (*Is*PETase-Dimer and –Cat^{+5}) were used, taking into consideration that each molecule contained two active sites. During the calculation of kinetic parameters, the enzyme concentrations of the dimeric species were adjusted to reflect the active site concentration.

To determine the rate of non-enzymatic *p*-NPA hydrolysis, reactions were carried out using the same procedure. Buffer was used to replace the missing volume of enzyme. After the conversion of all the 405 nm absorbance measurements into the concentration of *p*-NP released, the initial rate of non-enzymatic *p*-NPA hydrolysis was calculated and subtracted from the corresponding values measured for the enzyme catalysed *p*-NPA hydrolysis reactions. Enzyme activity was then fitted using the Michaelis-Menten model (**Figure 3.32**). The values obtained from the curve for the maximum reaction rate (V_{max}) and Michaelis

constant (K_m) were used to calculate the turnover number (k_{cat}) and catalytic efficiency (k_{cat}/K_m) for *Is*PETase-WT, -Spy, -Dimer and –Cat^{+5}.



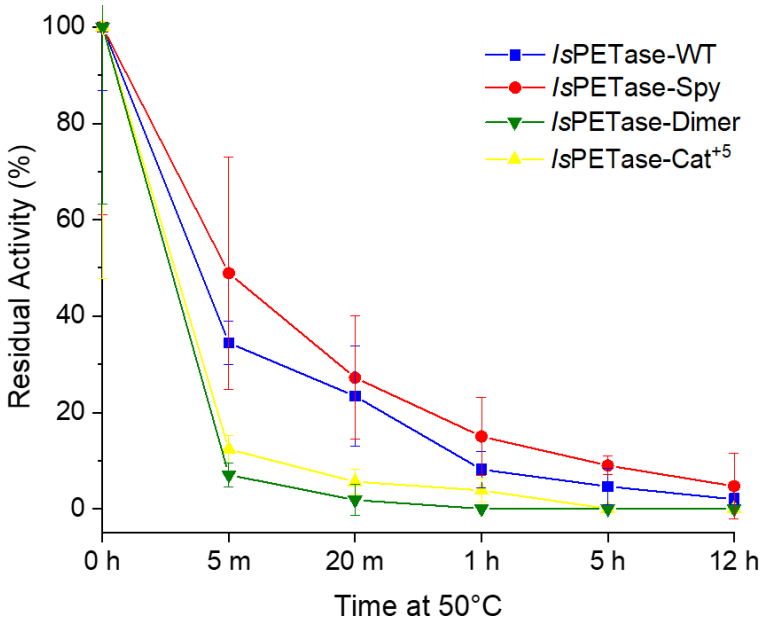
***Is*PETase variant**	**V_{max} (µM/min)**	**k_{cat} * (s^{-1})**	**K_m (mM)**	**k_{cat}/K_m (x10^4 M^{-1} s^{-1})**
-WT	25.2 ± 1.6	48.5 ± 3.2	2.7 ± 0.3	1.83 ± 0.24
-Spy	34.2 ± 4.4	51.3 ± 6.5	2.5 ± 0.7	2.04 ± 0.79
-Dimer	36.1 ± 4.4	56.7 ± 7.0	2.8 ± 0.7	2.03 ± 0.75
-Cat^{+5}	52.7 ± 9.9	68.0 ± 12.8	3.8 ± 1.3	1.78 ± 0.93

Figure 3.32 *Is*PETase-WT, -Spy, -Dimer and –Cat^{+5} kinetics for the hydrolysis of *p*-NPA at 30 °C and pH 7.5, calculated using the Michaelis-Menten model. *k_{cat} = V_{max}/[E], where [E] is the concentration of active sites. Error bars represent the standard deviation from the mean.

As aforementioned, the *p*-NPA substrate could not be used in concentrations above 5.0 mM without precipitation. Therefore, the V_{max} values calculated were extrapolated from the data points measured at lower substrate concentrations upon fitting of the Michaelis-Menten curve. As the K_m values were found to be within the substrate concentration range employed, and the value determined for *Is*PETase-WT was similar to that previously reported in the literature for *p*-NPA hydrolysis (2.6 ± 0.3 mM),[282] the V_{max} values extrapolated for all *Is*PETase variants were assumed to be of a reasonable estimate. The catalytic efficiencies calculated for all *Is*PETase variants was found to be around 2 x 10^4 M^{-1} s^{-1}, suggesting that there was no significant catalytic difference between the four enzymes.

3.2.3.1.2 *Is*PETase Residual Activity

Residual activity is the activity of an enzyme remaining after treatment that causes inhibition. Accordingly, the more tolerant the enzyme is towards the inhibiting conditions, the more residual activity it would exhibit. Here, the linear and cyclic *Is*PETase variants were heated for increasing amounts of time at 50 °C (above the T_m of all variants), before the activity of the enzyme remaining in solution was determined using the hydrolysis of *p*-NPA at 30 °C. **Figure 3.33** shows the residual activities of the *Is*PETase variants as a percentage of their catalytic activities before heating.



***Is*PETase variant**	**Residual activity after heating at 50 °C (%)**				
	5 min	**20 min**	**1 h**	**5 h**	**12 h**
-WT	34 ± 5	23 ± 10	8 ± 4	5 ± 4	2 ± 1
-Spy	49 ± 24	27 ± 13	15 ± 8	9 ± 2	5 ± 7
-Dimer	7 ± 2	2 ± 3	0	0	0
-Cat^{+5}	12 ± 3	6 ± 3	4 ± 3	0	0

Figure 3.33 Percentage residual activities of *Is*PETse-WT, -Spy, -Dimer and –Cat^{+5} after heating for increasing lengths of time at 50 °C. Error bars represent the 95% confidence interval.

The activities of all *Is*PETase variants decreased as the time the enzymes were heated at 50 °C increased. *Is*PETase-WT and –Spy displayed the highest residual activities throughout, with both enzymes retaining partial activity even after heating for 12 hours. Of the two enzymes, the cyclic monomer had the largest mean residual activity at each time point tested, which could suggest improved thermal stability over the linear WT enzyme, though they were not found to be significantly different at the 95% confidence interval. On

the other hand, the activities of both *Is*PETase-Cat^{+5} and -Dimer dropped by around 90% after just 5 minutes of heating, and were completely deactivated after 20 min and 1 hour at 50 °C, respectively. This rapid deactivation by heat treatment indicates destabilisation the cyclic dimer and catenane enzymes relative to *Is*PETase-WT.

While this research was in progress, a patent (WO/2021/145822) for a thermally stabilised cyclic *Is*PETase variant was published in which SpyTag and SpyCatcher were employed to connect the termini.[284] The cyclised enzyme was shown by a *p*-NPA hydrolysis assay to have improved residual activity compared to its linear counterpart. Around 70% activity remained after heating for 10 minutes at 50 °C, and up to 35% activity was retained after heat treatment at 90 °C. This is a more considerable improvement to the thermal stability of *Is*PETase than was observed in this work, where the SpyCatcher-N^{TEV} variant was employed for cyclisation with SpyTag.

To date, the majority of SpyTag/SpyCatcher cyclised proteins reported in the literature used the unmodified SpyCatcher sequence.[143,158,160,285,286] The tightly folded SpyTag/SpyCatcher complex is thought to facilitate protein refolding after thermal denaturation (see section 1.3.3.1).[143] However, modification of the SpyCatcher sequence can affect the thermal stability of the SpyTag/SpyCatcher complex as well as the cyclic protein of interest. For example, the use of the minimised SpyCatcherΔN1C2 variant for the cyclisation of firefly luciferase resulted in a 3 °C increase in T_m over the unmodified SpyCatcher cyclised enzyme.[144] Furthermore, the engineered SpyTag003/SpyCatcher003 complex exhibited a T_m of 95.2 °C upon reconstitution, approximately 10 °C higher than that of the original SpyTag/SpyCatcher complex.[287] This suggests that the stability of the SpyTag/SpyCatcher complex could also be disrupted as a result of SpyCatcher modification, leading to no stabilisation of cyclised protein construct. Considering this, the use of the circularly permuated SpyCatcher-N^{TEV} sequence for the cyclisation of *Is*PETase may be responsible for the lack of improvement to the residual activity of *Is*PETase-Spy observed in this work.

3.2.3.1.3 Heat-Induced Aggregation and Precipitation

After heat treatment and centrifugation, the *Is*PETase samples were further analysed by SDS-PAGE (**Figure 3.34**). For the monomeric samples, the concentration of enzyme remaining in solution clearly decreased with time, with no distinct protein band observable after 12 hours at 50 °C. This is likely due to precipitation of *Is*PETase-WT and –Spy. As all the residual activity experiments were carried out using the same concentration of enzyme, it indicates that enzyme not lost from solution was partially inactivated. Therefore a decrease in residual activity was observed. On the other hand, the dimeric enzymes were

retained in solution for longer than the monomeric samples. This is surprising considering the large drop in residual activity observed for *Is*PETase-Dimer and –Cat^{+5} after relatively short periods of time at 50 °C. It would suggest that the dimeric enzymes were less able to recover an active conformation after heat treatment, even though they did not precipitate so readily.

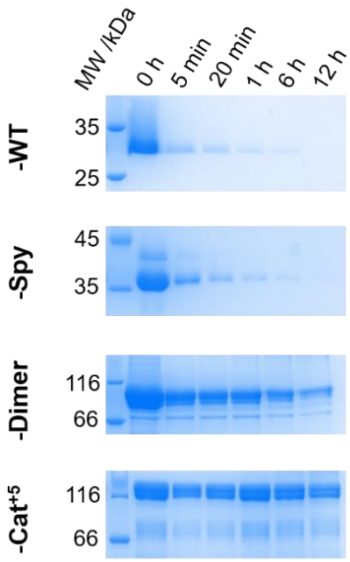


Figure 3.34 SDS-PAGE analysis of *Is*PETase samples after heating at 50 °C.

To investigate how heating affected the structures of the monomeric and dimeric enzymes remaining in solution after heating, SEC was employed. Each enzyme was heated for 5 min at 50 °C, before rapid cooling. The samples were then filtered to remove any precipitated enzyme and loaded onto the column. During the filtering process, it is possible that larger soluble aggregates (>0.22 µm) were also removed. **Figure 3.35** compares the SEC chromatograms of *Is*PETase-WT and -Spy obtained before (i.e. the initial SEC purification after Ni-NTA) and after heating.

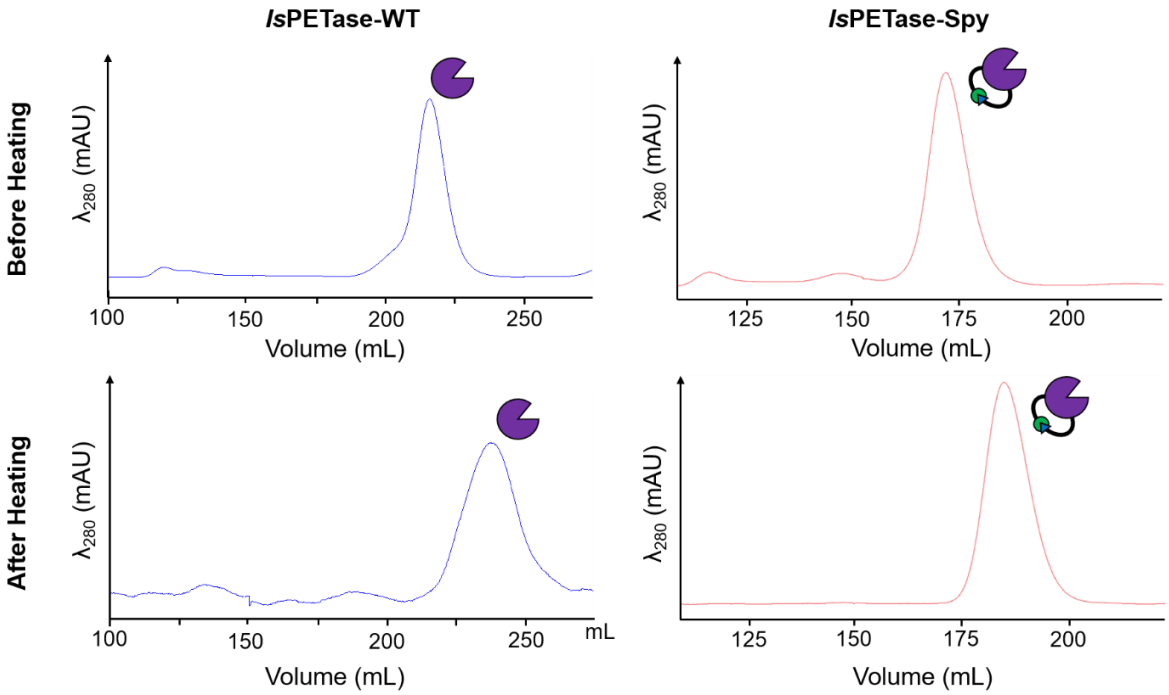


Figure 3.35 SEC chromatograms of *Is*PETase-WT and –Spy before and after heating at 50 °C for 5 minutes.

In the chromatograms of the monomeric enzymes after heat treatment, a single symmetrical peak is observed. This corresponded to the elution of the protein, which was confirmed by SDS-PAGE (not shown). Nevertheless, for both enzymes, elution from the column occurred slightly later than for the fully folded protein (i.e. before heating), with an increased elution volume of roughly 15 mL. Heating would be expected to result in protein unfolding, leading to a less compact structure and a reduced retention volume.[288] However, non-binding interactions can occur between the protein and the resin.[289] Depending on the nature of the interaction, the protein of interest may be eluted earlier or later than expected.[290] As a result, hydrophobic residues exposed in the unfolded protein species could be interacting with the column resin leading to the increased retention observed.

In contrast, the chromatogram of the catenane after heating at 50 °C (**Figure 3.36**) showed one major peak (Peak 1) and two smaller peaks (Peak 2 and 3). SDS-PAGE analysis found that the proteins eluted in Peaks 1 and 2 had molecular weights corresponding to that expected for *Is*PETase-Cat^{+5}. Meanwhile, Peak 3 had a molecular weight corresponding to the cyclic monomer contaminant. Before heating, *Is*PETase-Cat^{+5} was eluted at the same volume as Peak 2. Therefore, the large decrease in the elution volume of *Is*PETase-Cat^{+5} is suggestive of the formation of aggregates, which are eluted in the void volume due to their large size. To determine the nature of intermolecular interactions (i.e. covalent or non-covalent) between *Is*PETase-Cat^{+5} molecules, SDS-PAGE analysis of Peak 1 and 2 was

carried out in the absence of reducing agent in the sample loading buffer. For Peak 2, removal of the reducing agent did not affect the position of the major band. This suggested that no intermolecular disulfide bond formation had occurred between catenane molecules in these fractions. On the other hand, in the absence of reducing agent in the Peak 1 sample, the band at the expected molecular weight of the catenane became fainter and additional bands were observed at the top of the gel. This indicated the presence of aggregates which may have formed through intermolecular disulfide exchange.[291] It is likely that these aggregated enzymes are inactive for PET degradation, due to the unfolding of the active conformation and/or due to the active site becoming inaccessible upon aggregate formation. Regardless, it is likely that rapid aggregate formation correlates with the large decrease in residual activity observed for the dimeric enzymes.

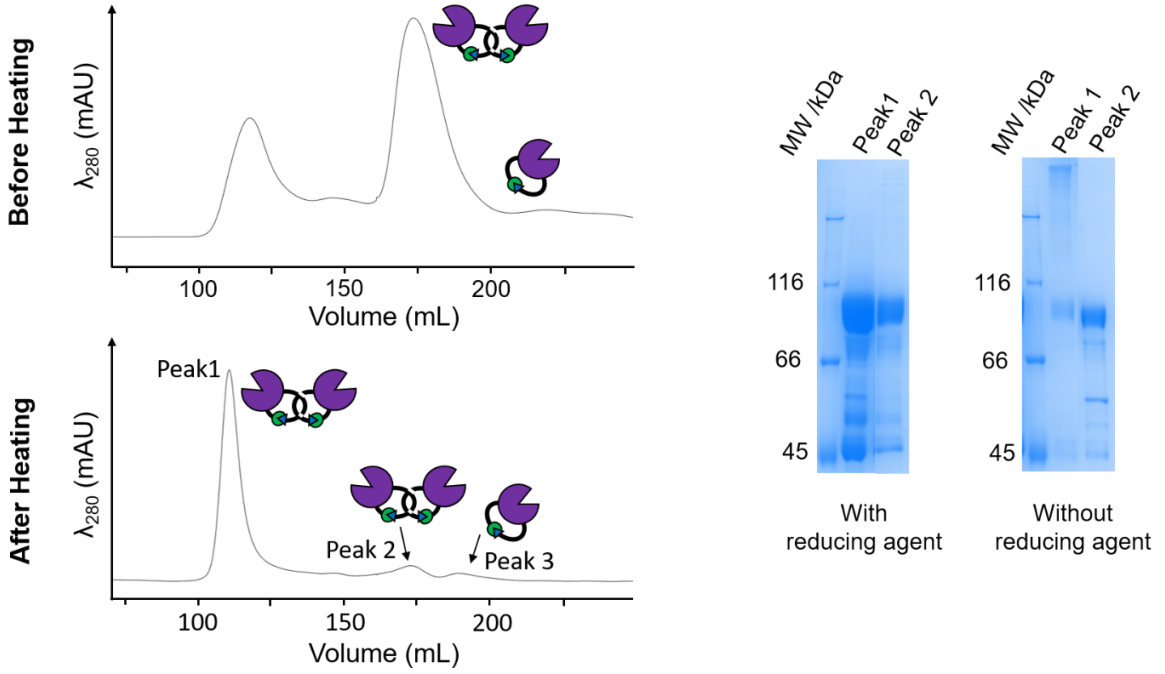
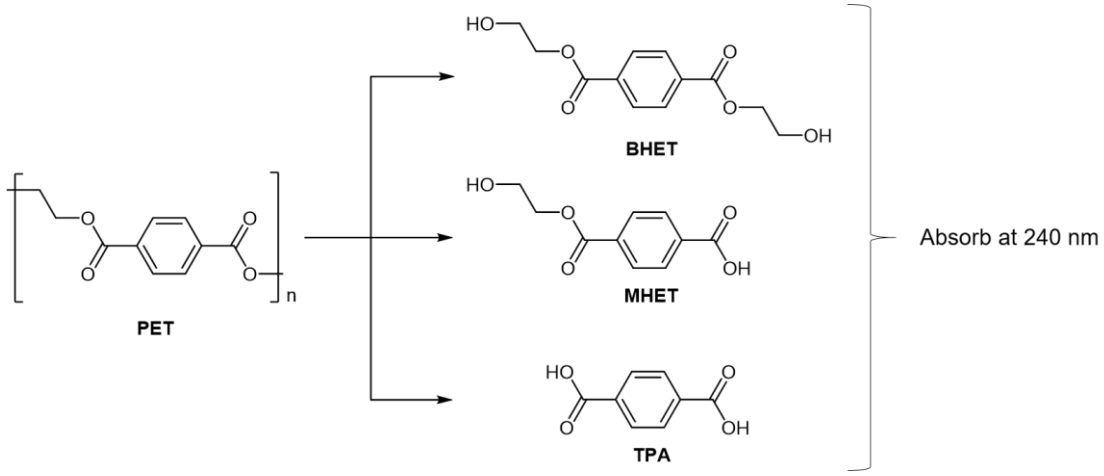


Figure 3.36 (Left) SEC chromatograms of *Is*PETase-Cat^{+5} before and after heating at 50 °C for 5 min. (Right) SDS-PAGE analysis of Peak 1 and 2, with and without β-mercaptoethanol in the SDS-PAGE loading buffer.

3.2.3.2 PET Degradation

The PET degrading activities of the linear and cyclised *Is*PETase variants were directly assessed using a commercially available PET powder substrate. PET hydrolysis generally affords a mixture of products (i.e. different monomers and oligomers of varying length), leading to complicated and/or laborious analyses for quantifying the amount of product released.[292] Here, a bulk absorbance-based procedure was employed to quantify enzymatic PET degradation.[293,294] Vigorous shaking was used to maintain the PET particles in a uniform suspension. Upon centrifugation of the sample, the reaction was quenched and the

supernatant removed for analysis using absorbance in the UV region to detect the release of soluble degradation products (i.e. TPA, MHET, BHET, etc.) (**Scheme 3.37**). These compounds exhibit an absorption maximum at around 240 nm with an identical extinction coefficient (ε_{240} = 13,800 M^{-1} cm^{-1}).[292] As a result, absorbance at 240 nm can be used to measure the overall sum of soluble PET hydrolysis products released, though over- or underestimation of enzyme activity may occur if oligomers are generated in significant quantities.[293] Nevertheless, cyclisation would not be expected to cause major perturbation to the active site of *Is*PETase. Hence, it is reasonable to assume that the mechanism of PET degradation, and general product distribution, of the different variants would remain unaffected. Therefore, direct comparison of the activity of the *Is*PETase variants should be possible using this method.



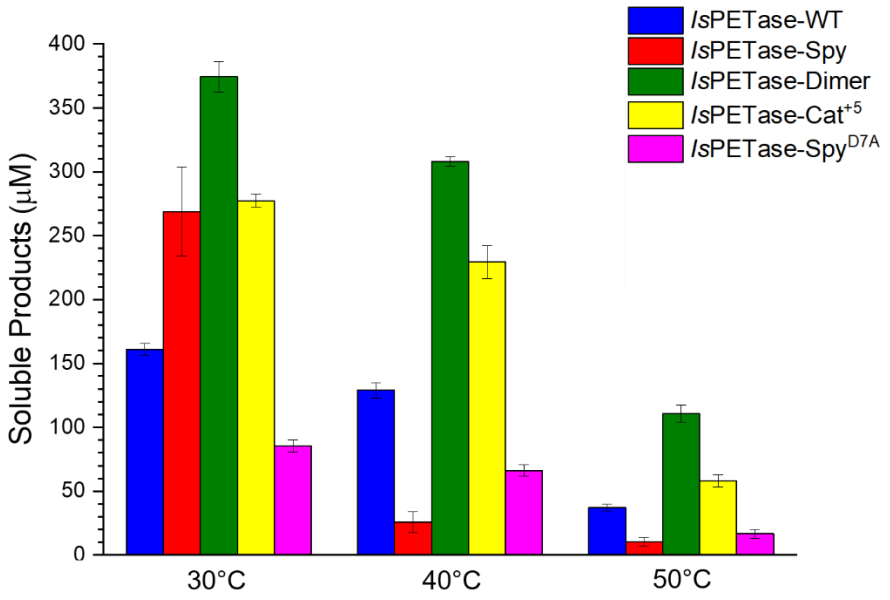
Scheme 3.37 Common soluble degradation products released as a result of PET hydrolysis which can be detected by absorbance at 240 nm.

The PET degradation activity assays were carried out in a 50 mM Na_2HPO_4 (pH 8.0) and 100 mM NaCl buffer, containing 7.5 mg/mL of powdered semi-crystalline PET. 0.1 µM of monomeric *Is*PETase-WT and –Spy were employed, while 0.05 µM of dimeric *Is*PETase-Dimer and –Cat^{+5} were used. Reactions were performed in triplicate at 30, 40 and 50 °C with shaking at 1100 rpm. After centrifugation, the samples were stored on ice and the supernatant analysed by measuring absorbance between 210 and 310 nm.

PET only control samples revealed there to be minimal product released after 24 hours in the absence of the enzyme, though this did increase with temperature. On the other hand, more significant changes to absorbance between 230 – 290 nm were observed for enzyme only controls, with the absorbance in this range increasing with temperature and time. As a result, enzyme contribution to absorbance at 240 nm was subtracted from the values measured for samples containing both PET and enzyme.

3.2.3.2.1 Effect of Temperature on PET Degradation

Figure 3.38 compares the results obtained for the *Is*PETase variants after incubation with PET at either 30, 40 or 50 °C for 24 hours. The absorbance values measured after 24 hours of incubation have been converted into concentration of soluble products.[293]



*Is*PETase Variant	30 °C	40 °C	50 °C
-WT	161 ± 5	129 ± 6	37 ± 3
-Spy	269 ± 35	26 ± 8	10 ± 3
-Dimer	374 ± 12	308 ± 4	111 ± 7
-Cat^{+5}	277 ± 5	229 ± 13	58 ± 5
-Spy^{D7A}	85 ± 5	66 ± 5	17 ± 4

Figure 3.38 Bar chart and table comparing the release of soluble degradation products (in µM) by the different *Is*PETase variants after 24 hour of incubation at 1100 rpm at increasing temperatures. Error bars represent the standard deviation from the mean.

For all four enzymes, the largest soluble degradation product release was measured after 24 hours of incubation at 30 °C. This suggests that the optimal PET degradation temperature of the cyclised variants remained unchanged compared to *Is*PETase-WT. Nevertheless, at this temperature, the cyclised enzymes appeared to outperform the linear WT enzyme (161 µM), with almost double the soluble degradation products generated by *Is*PETase-Spy and -Cat^{+5} (269 and 277 µM, respectively). The largest release of soluble degradation products was observed for *Is*PETase-Dimer (374 µM).

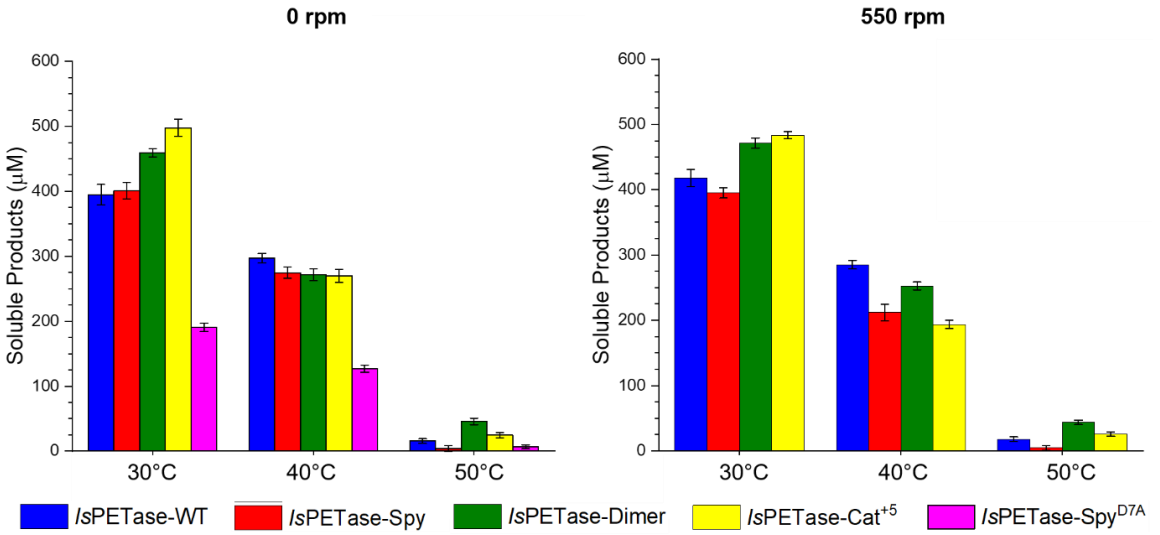
After incubation at 40 °C for 24 hours, the soluble products generated by the four enzymes was reduced compared to at 30 °C. Both the cyclic dimer and catenane variants (308 and 229 µM, respectively) maintained improved activity over the linear WT enzyme (129 µM).

However, a 10-fold drop in soluble degradation products released by the cyclic monomer was observed (26 µM). This could indicate denaturation of the enzyme by the increase in temperature, though this was surprising considering *Is*PETase-Spy had improved residual activity compared to the dimeric variants. A further drop in the production of soluble degradation products was observed for all enzymes after incubation for 24 hours at 50 °C, and follows a similar trend to that at 40 °C. Specifically, *Is*PETase-Dimer and –Cat^{+5} exhibited improved activity compared to *Is*PETase-WT, while the *Is*PETase-Spy was the least active variant (111, 58, 37 and 10 µM, respectively).

The PET degrading activity of the linear *Is*PETase-Spy^{D7A} mutant was also investigated. At 30°C, the amount of soluble degradation products released by the enzyme (85 µM) was lower than both the linear WT and cyclic monomer enzymes. This suggested that the improved activity observed for *Is*PETase-Spy was a result of cyclisation and not just the presence of the SpyTag and SpyCatcher-N^{TEV} sequences. On the contrary, after 24 hours of incubation with the PET substrate at 40 and 50 °C, the amount of soluble PET degradation products released by the linear *Is*PETase-Spy^{D7A} was higher than that of cyclic *Is*PETase-Spy, although still reduced compared to that of linear *Is*PETase-WT. This indicated that *Is*PETase was destabilised as a result of SpyTag/SpyCatcher-N^{TEV} cyclisation at these temperatures. However, *Is*PETase-Dimer and –Cat^{+5} were cyclised using SpyTag/SpyCatcher but exhibited improved activity over the linear WT enzyme at all temperatures tested, implying their activities were not affected in the same way as the cyclic monomer.

3.2.3.2.2 Effect of Agitation on PET Degradation

Agitation including shaking, stirring and vortexing is known to induce protein denaturation and aggregation.[295,296] This is thought to result from the exposure of proteins to the gas-liquid interface, leading to the partial exposure of hydrophobic residues and increasing the chance of aggregation. In addition, cavitation (i.e. the rapid formation and collapse of bubbles) can generate shockwaves, extreme temperatures and pressures.[297] For the PET degradation assay, vigorous shaking (1100 rpm) was employed to maintain the PET particles in suspension. Therefore, it was suspected that agitation was the cause of premature inactivation of the *Is*PETase variants. Previously, it was reported that the highest PET conversion was achieved by *Is*PETase-WT and *Is*PETase$^{S238F/W159H}$ in the absence of agitation.[298] As a result, PET degradation by the *Is*PETase variants was investigated without agitation (0 rpm), as well as with moderate shaking (550 rpm). **Figure 3.39** compares the amount of soluble degradation products generated by the different *Is*PETase variants after 24 hours of incubation at 0 and 550 rpm.



*Is*PETase Variant	0 rpm			550 rpm		
	30 °C	40 °C	50 °C	30 °C	40 °C	50 °C
-WT	395 ± 16	297 ± 7	16 ± 3	418 ± 13	285 ± 7	18 ± 3
-Spy	401 ± 13	274 ± 9	4 ± 4	395 ± 8	212 ± 13	5 ± 3
-Dimer	459 ± 6	271 ± 9	48 ± 5	472 ± 8	253 ± 6	44 ± 3
-Cat^{+5}	497 ± 13	270 ± 10	25 ± 4	483 ± 5	194 ± 6	25 ± 3
-Spy^{D7A}	190 ± 6	127 ± 6	7 ± 3	-	-	-

Figure 3.39 Bar charts and table showing the release of soluble degradation products (in μM) by the different *Is*PETase variants after 24 hours of incubation with 7.5 mg/mL PET at 0 and 550 rpm. Error bars represent the standard deviation from the mean.

In general, at 30 and 40 °C, the amount of soluble degradation products generated by the all *Is*PETase variants was increased in the absence of vigorous agitation. In most cases, stationary incubation of the *Is*PETase variants was slightly more favourable for PET degradation than incubation with moderate shaking. The optimal temperature for PET degradation remained at 30 °C for all variants.

In the absence of vigorous agitation at 30 °C, the largest improvement in activity was observed for *Is*PETase-WT, with an increase of roughly 250 μM soluble degradation products (**Figure 3.40**). A large increase was also observed for *Is*PETase-Cat^{+5}. *Is*PETase-Dimer showed the least improvement in activity, though the dimeric enzymes still generated more soluble degradation products compared to the monomeric enzymes at both 0 and 550 rpm. Again, the linear control was observed to have reduced activity compared to both *Is*PETase-WT and -Spy, suggesting the presence of the SpyTag/SpyCatcher-N^{TEV} sequences alone were not responsible for the increased activity observed.

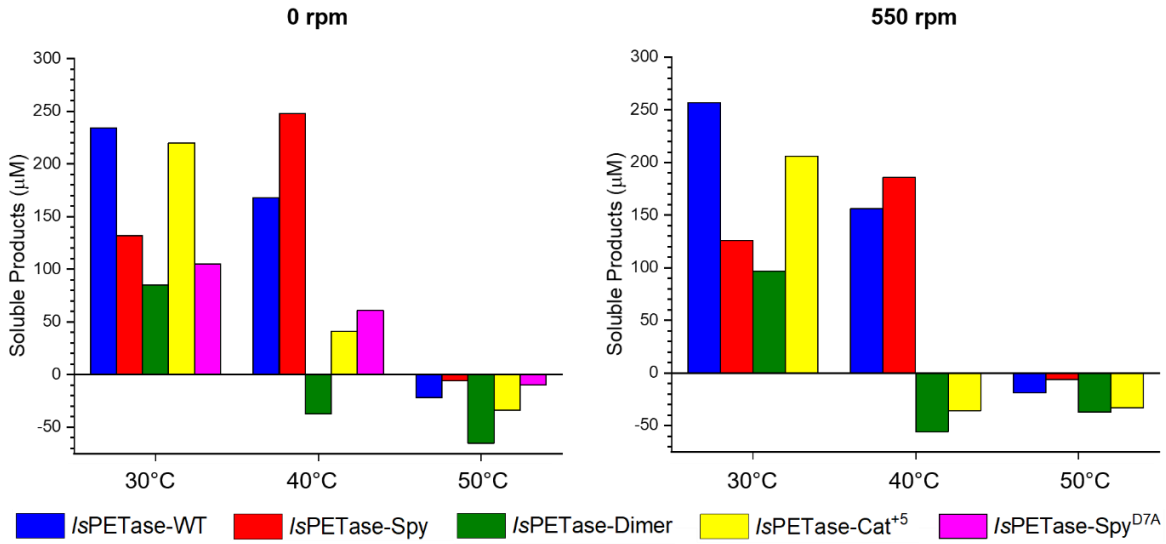


Figure 3.40 The increase or decrease in soluble PET degradation products (in µM) by the *Is*PETase variants at 0 and 550 rpm compared to at 1100 rpm.

After incubation at 40 °C for 24 hours in the absence of vigorous agitation, both the monomeric variants exhibited relatively large increases in the amount of soluble degradation products released. This indicated sensitivity of these enzymes to agitation. In particular, *Is*PETase-Spy generated around 250 µM more degradation products when incubated with PET at 0 rpm, compared to at 1100 rpm. In contrast, the dimeric variants did not benefit as greatly from the reduced agitation. Indeed, *Is*PETase-Dimer was found to generate fewer soluble degradation products (~50 µM) compared to at 1100 rpm. This indicates that the dimeric *Is*PETase variants could be less susceptible to agitation-induced inactivation than the monomeric enzymes, with agitation perhaps being advantageous for enzymatic PET hydrolysis in some cases. Nevertheless, at 40 °C, *Is*PETase-WT exhibited the highest PET degrading activity in the absence of vigorous agitation, outperforming all cyclic variants and the linear control.

For PET degradation performed at 50 °C in the absence of vigorous agitation, minimal soluble degradation product release was observed. Moreover, all the *Is*PETase variants generated fewer soluble degradation products compared to at 50 °C at 1100 rpm. This was most significant for the dimeric variants, with *Is*PETase-Dimer producing around 65 µM less soluble degradation products when incubation was carried out at 0 rpm. Both the raised temperature and faster shaking speed likely contributed to an increased mass transfer and a higher probability of collision between enzyme and substrate. Therefore, the rate of reaction was enhanced in the limited time before enzyme inactivation occurred, increasing the yield of soluble degradation products.

3.2.3.2.3 PET Degradation with Time

To further investigate the effects of temperature on the PET degrading activity of the linear and cyclic *Is*PETase variants, the course of the reaction was monitored by taking samples at selected time points during the 24 hours of incubation with PET. It would be expected that the more stable *Is*PETase variants would remain active for longer, and produce more degradation products overall after 24 hours of incubation. **Figure 3.41** compares the absorbance measured at different time points for *Is*PETase-WT, -Spy, -Dimer and -Cat^{+5} at 30, 40 and 50 °C with incubation in the absence of agitation (0 rpm).

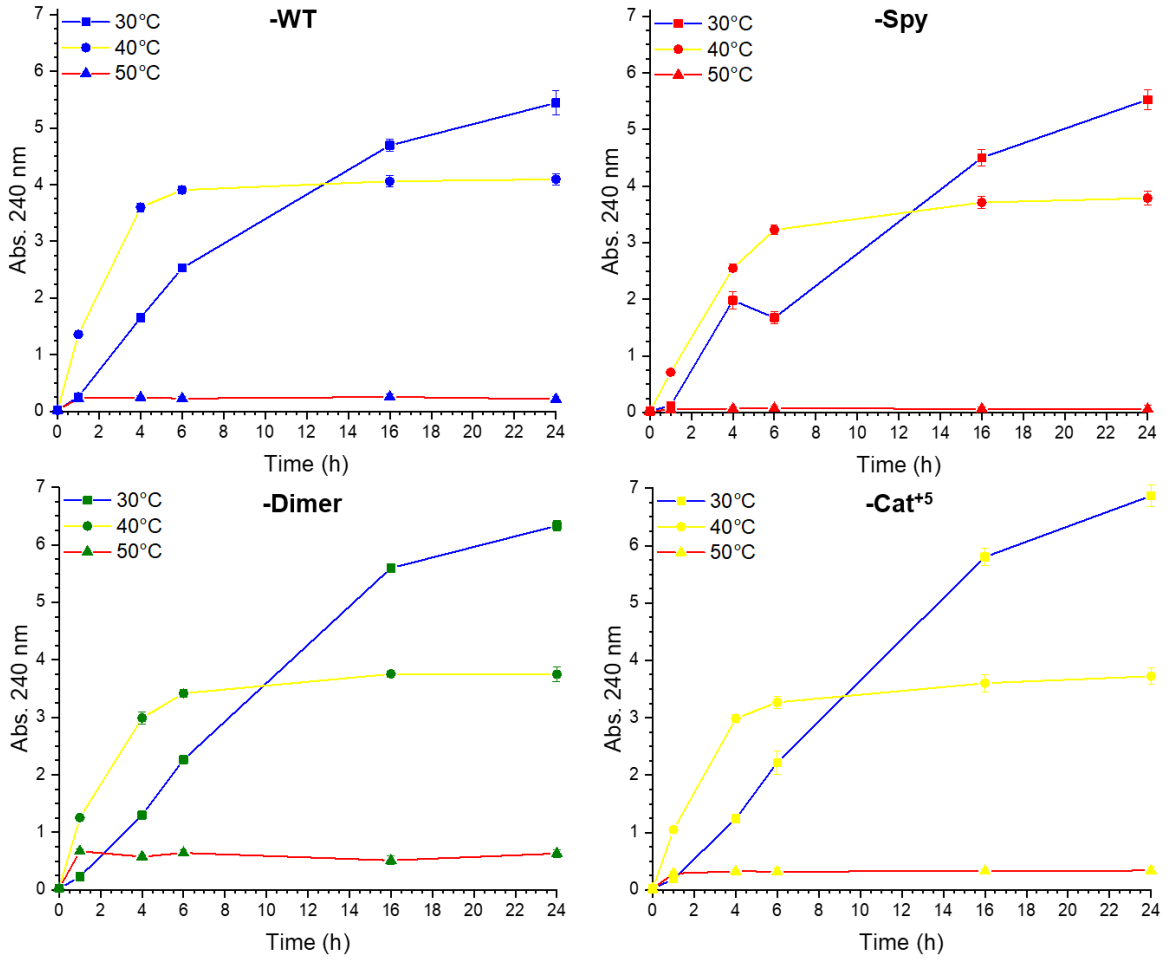


Figure 3.41 Absorbance at 240 nm measured a different time points during the 24 hours of PET incubation at 0 rpm with *Is*PETase-WT, -Spy, -Dimer and –Cat^{+5} at 30 (blue line), 40 (yellow line) and 50°C (red line). Error bars represent the standard deviation of the mean.

At 30 °C, the temperature at which all *Is*PETase variants showed optimal PET degrading activities, the absorbance measured at 240 nm continued to increase up until the 24 hour time point. This indicates that the activity of the enzymes likely extends beyond 24 hours as previously reported for *Is*PETase-WT. However, the start of a plateau was observed for all variants after 16 hours of reaction, suggesting that the production of soluble degradation

products slows after this time. As the temperature was increased, the length of time before the plateau occurred, decreased. At 40 °C, plateauing was observed for all enzymes after around 6 hours of incubation. Meanwhile, at 50 °C, an increase in absorbance at 240 nm was not observed past 1 hour of incubation. This implies that enzyme inactivation occurs more quickly with increasing temperature. It follows that fewer soluble degradation products would be generated overall after 24 hours due to the curtailed activity of the enzymes.

Though the *Is*PETase variants have more prolonged activities at 30 °C, the initial increase in absorbance was observed to be steeper at 40 °C, indicating soluble degradation products were generated more quickly as a result of PET hydrolysis. This could be explained by an increased rate of diffusion, leading to more collisions between enzyme and PET particles with sufficient energy and correct orientation to react. As more enzyme becomes inactivated over time, the rate of PET degradation slows and eventually stops. Therefore, the increased rate of reaction was only temporary.

Next, the effect of agitation on the duration of PET degradation was investigated for the linear and cyclic *Is*PETase variants. It was expected that enzymes that were less tolerant to agitation would be inactivated more quickly as shaking speed was increased. **Figure 3.42** compares the results obtained for PET degradation over time by the linear and cyclic *Is*PETase variants using moderate agitation (550 rpm).

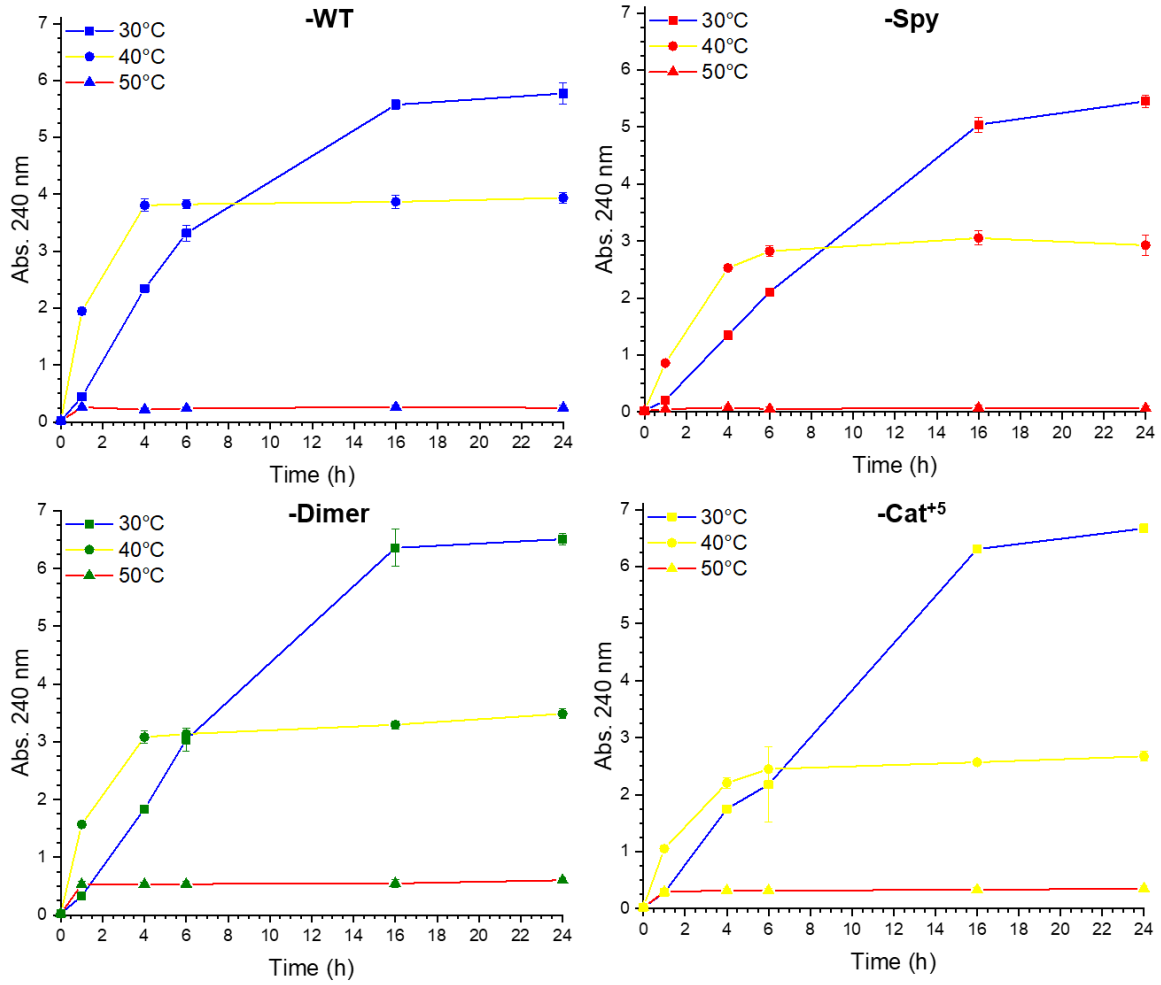


Figure 3.42 Absorbance at 240 nm measured a different time points during the 24 hours of PET incubation at 550 rpm with *Is*PETase-WT, -Spy, -Dimer and –Cat^{+5} at 30 (blue line), 40 (yellow line) and 50°C (red line). Error bars represent the standard deviation of the mean.

At 550 rpm, the *Is*PETase variants exhibited the most prolonged activity at 30 °C. However, the plateau in the absorbance at 240 nm appeared to occur slightly earlier (16-24 hours) than when incubated in the absence of agitation (>24 hours at 0 rpm). Meanwhile, plateaus in the absorbance were observed after 4-6 hours at 40 °C and after 1 hour at 50 °C, similar to at 0 rpm. This indicates that moderate agitation had minimal effect on the duration of enzyme activity for the degradation of PET. Furthermore, a similar quantity of soluble degradation products were generated by all *Is*PETase variants after 24 hours of incubation at 0 and 550 rpm. Nevertheless, the effects of moderate agitation may become more apparent if the duration of the reaction was extended past 24 hours.

Figure 3.43 compares the PET degrading activities of the *Is*PETase variants over time when incubated at increasing temperatures with vigorous agitation (1100 rpm). It would be

expected that enzyme inactivation would occur earlier than at 550 and 0 rpm, resulting in an overall inferior yield of soluble degradation products.

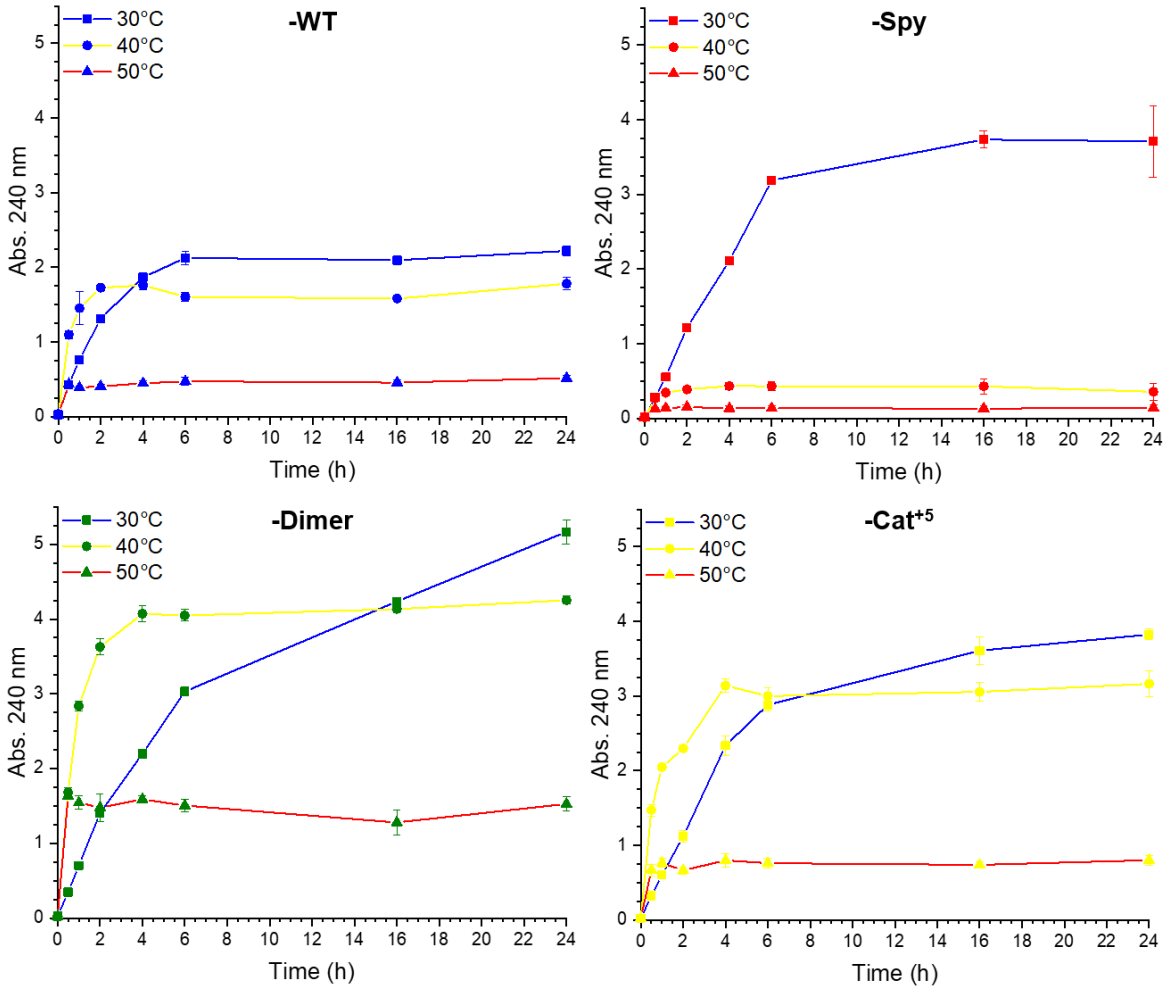


Figure 3.43 Absorbance at 240 nm measured a different time points during the 24 hours of PET incubation at 1100 rpm with *Is*PETase-WT, -Spy, -Dimer and –Cat^{+5} at 30 (blue line), 40 (yellow line) and 50°C (red line). Error bars represent the standard deviation of the mean.

At 30 °C at 1100 rpm, the duration of activity of the monomeric enzymes was substantially reduced compared to at 0 and 550 rpm. For *Is*PETase-WT a plateau in absorbance occurred after just 6 hours, while *Is*PETase-Spy appeared to become inactive between 6-16 hours. On the other hand, vigorous agitation did not curtail the activities of the dimeric *Is*PETase variants as significantly. Indeed, the activity of *Is*PETase-Dimer appeared to extend past 24 hours, with no plateau in absorbance observed. This correlates with the cyclic dimer and catenane variants generating more soluble degradation products under these conditions compared to the monomeric variants. It could suggest that dimerisation helps to protect *Is*PETase from agitation-induced inactivation. Nevertheless, a higher concentration of soluble degradation products was still obtained for all *Is*PETase variants in the absence of agitation at 30 °C.

At 40 °C, *Is*PETase-Spy rapidly underwent inactivation when vigorous agitation was employed, with a plateau in absorbance observed after only 1 hour. This was reflected in the poor yield of soluble degradation products generated by the mono-cyclic enzyme under these conditions. As a result, it is likely that *Is*PETase-Spy is particularly susceptible to agitation-induced inactivation. In comparison, the activities of *Is*PETase-WT, -Dimer and –Cat^{+5} were not curtailed as severely, and produced significantly more soluble degradation products in the presence of vigorous agitation. This further indicates the ability of some topologies to withstand agitation-induced inactivation more than others.

Similar to at 0 and 550 rpm, all the *Is*PETase variants were observed to be rapidly inactivated (<1 hour) at 50 °C when vigorous agitation was employed. This suggests that at this temperature, enzyme inactivation occurred primarily as a result of heat-induced denaturation, irrespective of the shaking speed.

3.2.3.3 *Is*PETase Activity Summary

To investigate the activities of the linear and cyclic *Is*PETase variants, two different assays were employed. Firstly, the hydrolysis of *p*-NPA was used to compare the residual activities of the enzymes. After heat treatment, the residual activities of the cyclised variants was not improved compared to *Is*PETase-WT. Though both monomeric enzymes retained the most activity, they were prone to precipitation upon heating. On the other hand, *Is*PETase-Dimer and –Cat^{+5} formed large soluble aggregates but were more rapidly deactivated compared to *Is*PETase-WT.

The absorbance-based PET degradation assay also suggested that the thermal stability of *Is*PETase was not improved by cyclisation. For all variants, the optimal temperature for PET degradation was consistently found to be 30 °C (**Table 3.2**). Nevertheless, the yield of soluble degradation products generated was found to vary with enzyme topology, temperature and agitation. Though general trends are difficult to discern, dimerisation of *Is*PETase appears to benefit PET degradation most likely by (i) increasing the probability of enzyme-substrate complex formation (see section 3.3.3), and (ii) protecting the enzyme from agitation-induced inactivation.

Table 3.2 Conditions used for maximum observed soluble degradation product release by each of the *Is*PETase variants.

***Is*PETase Variant**	**Max. Soluble Product /μM**	**Conditions**	**Time Active /h**
-WT	420	30 °C, 550 rpm	16-24
-Spy	400	30 °C, 0 rpm	>24
-Dimer	470	30 °C, 550 rpm	16-24
-Cat^{+5}	500	30 °C, 0 rpm	>24
-Spy^{D7A}	190	30 °C, 0 rpm	-

3.3 DISCUSSION

In this chapter, the effect of cyclisation on the stability and activity of *Is*PETase was investigated using SpyTag/SpyCatcher(-N^{TEV}) to connect the termini of the enzyme. A range of cyclic topologies were constructed, specifically, a cyclic monomer (*Is*PETase-Spy), a cyclic dimer (*Is*PETase-Dimer) and a catenane (*Is*PETase-Cat). Though the recombinant expression of *Is*PETase was initially challenging, removal of the signal peptide sequence and optimisation of the expression conditions resulted in the production of the desired proteins in a good yield. The stabilities and activities of the linear and cyclised enzymes were then compared, with some unexpected results obtained. The findings are summarised in the following section. Some directions for future research are also described.

3.3.1 Thermal Stability and Residual Activity

Thermodynamic stability is defined by the difference in Gibbs free energy (ΔG) between the folded and the unfolded state (**Figure 3.44**).[275] According to the *thermodynamic hypothesis* of protein folding, the most stable three-dimensional conformation of a protein under given conditions, is the one in which ΔG of the whole system is at its lowest. In other words, the protein folding process is driven by the tendency to reach the thermodynamically favoured state where ΔG is at a global minimum.[1]

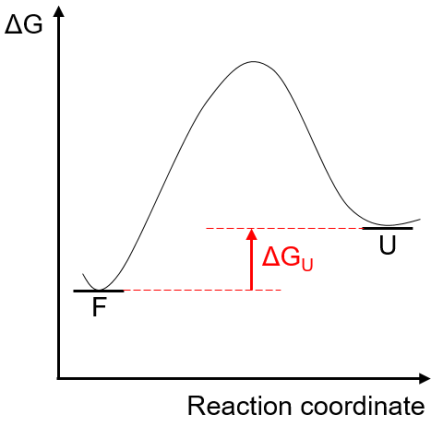


Figure 3.44 Thermodynamic stability of a protein, defined by the difference in the Gibbs free energy between folded and unfolded states.

In this chapter, T_m was used as a measure of *Is*PETase thermal stability, determined from protein thermal denaturation curves measured using CD. This assumes the stability of the protein is under thermodynamic control. However, the thermodynamic information can only be derived indirectly, by applying a folding model. Most commonly this is a two-state approximation, in which protein molecules reversibly populate either the folded (F) or the unfolded state (U), without passing through any intermediate unfolding states.[155,299] It follows that the equilibrium (F ⇌ U) can be described by a temperature dependent equilibrium constant (K_U):

$$(\boldsymbol{Eq.3.1}) \qquad K_U = \frac{[U]}{[F]} = \frac{\alpha_U}{1-\alpha_U} = \frac{1-\alpha_F}{\alpha_F}$$

where [F] and [U] are the concentrations of folded and unfolded proteins, respectively, and α_F and α_U are the fraction of protein folded and unfolded, respectively. Therefore, at T_m, the mid-point of protein unfolding, [F] and [U] would be equal and $K_U = 1$. As K_U is related to ΔG_U according to **Equation 3.2** (where R is the gas constant and T is temperature), it can be seen that ΔG_U would be equal to 0 at T_m.

$$(\boldsymbol{Eq.3.2}) \qquad \Delta G_U = -RTlnK_U$$

Stabilisation of cyclic proteins is thought to originate from the decreased entropy of the protein's unfolded state. The contributions of entropy (ΔS_U) and enthalpy (ΔH_U) towards ΔG_U are shown in **Equation 3.3**.

$$(\boldsymbol{Eq.3.3}) \qquad \Delta G_U = \Delta H_U - T\Delta S_U$$

It can be seen that a decrease in ΔS_U would lead to an increase in ΔG_U which would be thermodynamically unfavourable. As a result, population of the protein's folded state would be favoured. Nevertheless, the T_m of *Is*PETase-WT and –Spy were not considerably different. This could indicate the presence of unfavourable enthalpic contributions to ΔG_U, although the cyclic protein was not shown to be destabilised by strain. Moreover, cyclisation of *Is*PETase has since been demonstrated to enhance the residual activity of the enzyme when SpyCatcher instead of SpyCatcher-N^{TEV} was employed.[284] This indicates that modification of the SpyCatcher sequence through circular permutation may perturb the ability of the SpyTag/SpyCatcher complex to facilitate protein refolding after heat treatment. However, SpyCatcher was used for the preparation of the cyclic dimer and catenane variants of *Is*PETase which were found to have significantly reduced residual activities compared to *Is*PETase-WT. This likely resulted from rapid soluble aggregate formation, but this would need to be confirmed experimentally, as inclusion bodies have been known to exhibit catalytic activity.[300]

Often aggregation begins with structural perturbation to the structure of the protein by an external stress. This leads to protein-protein interactions and the formation of large oligomeric species.[301,302] Temperature is known to influence the aggregation propensity of proteins in a number of ways, including the protein conformational stability, rate of protein diffusion, protein-protein interactions and the solubility of the proteins and their aggregates.[303] Previously, cyclised proteins have been shown to be more resistant towards thermal-induced aggregation in comparison to their linear counterparts, including SpyTag/SpyCatcher cyclised proteins.[143,304] This is thought to originate from the enhanced conformational and colloidal stability of the cyclised protein.[288] The former is determined by ΔG_u, while the latter is the tendency of the folded or unfolded protein molecules to undergo intermolecular association.[305] It is likely that the conformational stability of *Is*PETase was not significantly improved by cyclisation. As a result, thermal unfolding, leading to the exposure of aggregation-prone sequences, would be just as likely for linear and cyclic enzymes. Moreover, the rapid formation of aggregates by the cyclic dimer and catenane variants could suggest that the dimeric variants had reduced colloidal stability compared to the monomeric enzymes, and were more likely to undergo intermolecular interactions upon heating.

In future work, some techniques that could be employed to study the heat-induced denaturation and aggregation of *Is*PETase are briefly outlined below. A better understanding of the complex heat-induced inactivation pathways of the *Is*PETase variants

could provide beneficial for enhancing the thermal stability of *Is*PETase (and other proteins) through cyclisation in future.

1. **Dynamic light scattering (DLS)** is commonly used to determine the size distribution of particles in solution and can be used to detect protein aggregate formation.[306,307] This would be useful for determining the oligomerisation state of the soluble aggregates and monitoring how it changes with temperature and time.[308] In addition, **analytical ultracentrifugation (AUC)** could be employed for the analysis of the solution structures and dispersibilities of *Is*PETase variants in their folded and unfolded structures.[288,309]

2. **Intrinsic tryptophan fluorescence spectroscopy** can be used to monitor changes in the tertiary structure of proteins by selectively probing changes in the intrinsic fluorescence signals of Trp residues, which are known to be sensitive to the polarity of their local environment. Therefore, structural changes in the vicinity of Trp amino acids alter the emission spectra observed.[310,311] As a result, fluorescence spectroscopy could be employed to compare the structures of the linear and cyclic *Is*PETase variants, and to monitor their response to heating. Moreover, it could be employed to investigate the substrate binding of the different variants.[257]

3. **Disulfide bond reduction** (or removal of Cys residues through mutation) could be carried out for further analysis of the effects of cyclisation on the stability and activity of *Is*PETase. The enzyme is known to contain two disulfide bonds, which upon reduction or mutation lead to destabilisation.[246,282] As a result, cyclisation of *Is*PETase could potentially restore the stability of the enzyme, even if it does not lead to an overall increase in stabilisation when disulfide bonds are present. In addition, disulfide exchange was found to occur upon heat-induced aggregation of *Is*PETase-Cat^{+5}. Therefore, it could be interesting to examine how the removal of the disulfide bonds impact on the aggregation pathways of the different *Is*PETase variants.

4. **DSC** measures the change in heat capacity (C_p) of a sample as temperature is increased. As a result, the thermodynamic parameters (i.e. ΔG_U, ΔH_U, ΔS_U) of protein unfolding can be directly derived, without the use of a folding model.[299] Though the apparent T_m of the *Is*PETase variants was obtained by CD, it was not possible to assess the conformational stability of the enzymes. This is because the two-state

folding model assumes protein unfolding to be reversible, though in reality the heat-induced denaturation of *Is*PETase is irreversible.[155] Alternatively, **chemical denaturants**, such as urea and GdnHCl can be employed for reversible protein denaturation. The dependence of protein folding/unfolding on concentration of chemical denaturant can be monitored spectroscopically, from which the conformational stability of a protein can be determined.[312]

3.3.2 Proteolytic and Chemical Stability

The stability of a protein under kinetic control is defined by the activation energy of unfolding ($\Delta G_U^{\ddagger}$) (**Figure 3.45**).[275] Those with large $\Delta G_U^{\ddagger}$, have a high unfolding energy barrier and so in theory would unfold more slowly than a protein with a smaller $\Delta G_U^{\ddagger}$. As a result, kinetically stable proteins are trapped in their native conformations, even if ΔG is favourable for unfolding. Therefore, thermodynamic stability does not entirely protect proteins from irreversible denaturation.

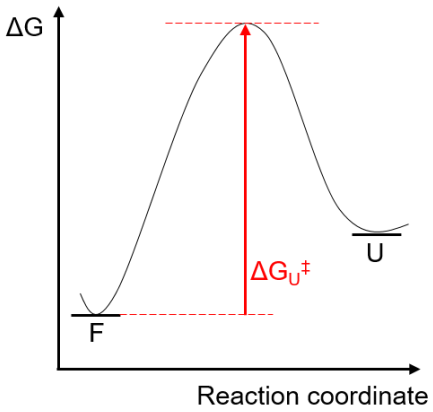


Figure 3.45 Kinetic stability of a protein, defined by the activation energy of unfolding.

Kinetically stable proteins rarely access unfolded and partially unfolded states. Therefore they are often less susceptible to proteolytic cleavage and are resistant to detergent-induced denaturation.[275] Similarly, cyclisation should theoretically disfavour the exposure of protease cleavage sequences to the solvent and has been reported to lead to the rigidification of the protein by restricting the mobility of the backbone.[166]

Using the Expasy PeptideCutter online tool (https://web.expasy.org/peptide_cutter/), the *Is*PETase sequence was found to contain 20 potential cleavage sites, corresponding to the total number of 20 Arg and Lys residues present in the primary sequence. Of these sites, 16 were predicted to have 100% probability of cleavage by trypsin. However, *Is*PETase-WT was found to be relatively resistant to digestion. This could suggest that the WT enzyme has a compact conformation that is resistant to trypsin digest under ambient conditions (30

°C).[313–315] Nevertheless, denaturation in the presence of SDS demonstrated that the proteolytic stability of *Is*PETase was not due to kinetic stability. This is unsurprising considering the rapid inactivation of *Is*PETase at raised temperatures, which would be uncharacteristic of a protein trapped in the native state by a large unfolding energy barrier. In any case, the binding requirements of a detergent and a protease differ. While the sulfate group and alkyl chain of SDS bind to the positively charged and hydrophobic residues, respectively,[316] proteases require access to specific recognition sites within the protein chain. Therefore, a protein may be a poor substrate for proteolytic digestion if it does not contain the required cleavage sequence, or if access to these sequences is restricted.[162]

Cyclisation did not improve the proteolytic stability of *Is*PETase, though this is most likely due to the presence of the SpyTag/SpyCatcher(-N^{TEV}) sequences in the cyclic topologies. In addition, the cyclised *Is*PETase variants were not found to stabilise *Is*PETase towards SDS-induced denaturation. Nevertheless, SDS is strongly denaturing, to which relatively few proteins are even partially resistant. Therefore, in future work, it may be more beneficial to compare the stabilities of the linear and cyclic *Is*PETase variants using a milder detergent such as sodium lauroyl sarcosinate (sarkosyl).[317]

3.3.3 Stability versus PET Degradation

The overall aim of the work in this chapter was to use cyclisation to improve the stability of *Is*PETase to enhance the PET degrading activity of the enzyme at elevated temperatures. Both T_m and residual activity experiments showed that this desired improvement to thermal stability was not achieved. Regardless, the topology of *Is*PETase was still demonstrated to impact the extent of PET degradation achieved. Indeed dimerisation was found to beneficial under several of the conditions tested.

Under optimal conditions (30 °C, 0 rpm), the *Is*PETase-Dimer and –Cat^{+5} variants generated roughly 17 and 26% more soluble degradation products compared to *Is*PETase-WT, respectively. Enzymatic PET degradation relies on the formation of an enzyme-substrate complex, i.e. occupation of the enzyme active site by the PET substrate. Each molecule of *Is*PETase-Dimer and –Cat^{+5} contains two active sites. Therefore, when one active site of the dimeric enzyme is occupied, the other is also in close proximity to the substrate. This could increase the chances of enzyme-substrate complex formation by the second active site, resulting in an increased rate of PET hydrolysis and yield of soluble degradation products. In addition, the differences in activity observed between the two dimeric enzymes could result from different orientations of the active sites relative to one another.[270]

The dimeric topologies of *Is*PETase were found to be more tolerant towards vigorous agitation compared to the monomeric variants. Even at increased temperatures, the dimeric variants generally yielded more soluble degradation products when agitiation was employed. The stability of cyclic proteins towards agitation is an important factor that is often overlooked in the literature, yet agitation is known to induce enzyme inactivation through denaturation and aggregate formation. The tendency of a particular protein towards agitation-induced aggregation is thought to depend on several factors, including the availability of hydrophobic sites, charge distribution, thermal stability and protein-protein interactions.[296] In addition, the solution conditions (e.g. temperature, pH, concentration and additives) influence the aggregation tendency of proteins. Experiments that could be employed for further investigation of the effects of cyclisation on the PET degrading activity of *Is*PETase are described below.

1. **Duration of PET degradation**: Here, PET degradation experiments were carried out over the course of 24 hours. However, enzymatic PET degradation is known to continue after this time. Therefore, the performance of the enzymes should be evaluated over a more extended period to fully understand the extent of PET degradation achieved by the different enzymes.[255]

2. **PET degradation analysis**: Here, a bulk absorbance-based assay was employed to monitor the release of soluble degradation products by the *Is*PETase variants. Though it was relatively facile to perform, the release of oligomeric products can lead to discrepancies between the measured and actual activity of the enzymes.[293] Therefore, in future work, it would be recommended to use HPLC in combination with the bulk absorbance assay to verify the product distribution resulting from enzymatic PET treatment.[255] Nevertheless, this would still only provide information on the accumulation of soluble degradation products. Hence, measuring the sample weight loss and/or the use of scanning electron microscopy would provide a more complete assessment of PET degradation by the different *Is*PETase variants.[292]

3. **pH**: The pH of a solution dictates the surface charge of a protein. As a result, both intra- and intermolecular electrostatic interactions are affected. This can lead to changes in a protein's conformation, stability, as well as its tendency towards aggregation.[303,318] Here, all stability and activity assays were carried out at pH 8.0 (with the exception of the *p*-NPA hydrolysis assay which used pH 7.5). Though the pH dependence of the different variants was not examined, cyclisation has

previously been demonstrated to improve the tolerance of proteins to a broader range of pH.[161,319] This could be beneficial for the biodegradation of PET, as the acidification of the reaction media by the release of large quantities of TPA, may lead to a reduction in enzyme activity.[320]

4. **Effect of surfactants on activity**: The linear and cyclic *Is*PETase variants were found to undergo aggregation upon heating and agitation, resulting in loss of enzyme activity. Surfactants are commonly employed to minimise protein aggregation and so may be beneficial for improving the stability of *Is*PETase towards environmental stresses and prolong catalytic activity.[321,322] Nevertheless, surfactants can instead promote protein unfolding and induce aggregation.[323,324] Therefore, screening of a variety of different surfactants would be required to find the most suitable candidate. Moreover, surfactants have been employed to enhance enzymatic PET degradation by facilitating contact between the enzyme and hydrophobic PET surface.[325,326] Other additives have also been demonstrated to be beneficial for PET degradation, such as DMSO which promotes the solubility of liberated degradation products.[298]

5. **Effect of SpyCatcher variant on cyclic protein stability**: Here the circularly permutated variant of SpyCatcher (SpyCatcher-N^{TEV}) was employed for the preparation of mono-cyclic *Is*PETase. However, the results indicate that modification of the SpyCatcher sequence could affect the stability of the resulting SpyTag/SpyCatcher complex, leading to no significant improvement in residual activity of the cyclised protein. Therefore, it would be useful to directly compare how the choice of SpyCatcher variant impacts the stability of the SpyTag/SpyCatcher complex, as well as the overall stability of the cyclic protein construct.

6. **DuraPETase cyclisation**: DuraPETase is a thermally stable mutant of *Is*PETase, reported to have a T_m above the T_g of PET.[256] As a result, examining the effects of cyclisation on the stability and activity of DuraPETase could provide insightful for understanding the behaviours of the different cyclic topologies in response to heat treatment, proteolysis and agitation.

7. ***Is*PETase:*Is*MHETase dimers**: It is known that the two enzymes have a synergistic relationship, whereby *Is*PETase breaks down PET into soluble intermediates, which are subsequently converted by *Is*MHETase into TPA and EG monomers. In nature,

the monomers are metabolised by *I. sakaiensis*. However, they could instead be utilised for the resynthesis of PET. Previously, an *Is*MHETase:*Is*PETase chimeric enzyme was reported, in which the two enzymes were connected by a flexible linker through their C- and N- termini, respectively.[161] This chimeric construct outperformed the unlinked enzymes in solution. Therefore, cyclisation may enhance the activity of the chimeric enzyme further.

3.4 CONCLUSIONS

To summarise, the biodegradation of plastics is an attractive alternative towards closed-loop recycling, an important step towards tackling pollution and climate change in general. The enzyme *Is*PETase, isolated from the PET metabolising bacterium *I. sakaiensis*, is viewed as a promising candidate for the recycling of polyesters. Despite this, the utility of *Is*PETase-WT is limited by its low stability and activity. Therefore, improvements are required if it is to be employed for future plastic recycling applications. It is generally accepted that enhancing the thermal stability of PET degrading enzymes would result in increased activity. This is because at elevated temperatures (i.e. T_g), the amorphous phase of the plastic undergoes a rigid to flexible transition, resulting in improved access of the enzyme to the polymer chains.

In this chapter, *Is*PETase was cyclised using SpyTag/SpyCatcher(-N^{TEV}). Three cyclic topologies including a cyclic monomer, a cyclic dimer and a catenane were constructed. All enzymes were ultimately obtained in a good yield. Subsequently, the activities and stabilities of these cyclic variants were compared to that of the linear WT enzyme. Cyclisation did not considerably improve the thermal stability of *Is*PETase, with both linear and cyclic enzymes prone to aggregation upon heating. With regard to proteolytic stability, the cyclic variants were all digested more rapidly by trypsin compared to *Is*PETase-WT. Nevertheless, cyclisation did appear to benefit PET degradation under other industrially relevant conditions. The dimeric variants generally produced a higher yield of soluble degradation products, likely because the two active sites are located in close proximity, increasing the likelihood of enzyme-substrate complex formation. Furthermore, the dimeric variants exhibited improved resistant to vigorous agitation, a class of stability that has not been adequately investigated for cyclic proteins. As a result, the stability of cyclic protein topologies towards agitation should be examined in more detail as it a frequently encountered stress during all stages of protein production and application.

On a final note, a great deal of research has been directed towards improving the thermal stability of *Is*PETase. Indeed, a number of variants have already been engineered which exhibit superior stability and PET degrading performance over *Is*PETase-WT and other PET hydrolysing enzymes. Therefore, future research in the general area of enzymatic PET degradation should focus on enhancing other aspects relevant to the large scale application of PET degrading enzymes for recycling purposes. This includes improving enzyme activity on highly crystalline PET substrates, optimisation of large scale enzyme production, plastic pretreatment processes, as well as the recovery of monomers for the efficient resynthesis of PET.

CHAPTER 4:

GENERAL CONCLUSIONS AND FUTURE WORK

4.1 GENERAL CONCLUSIONS

Cyclisation has been shown to confer peptides and proteins with a range of favourable properties, including enhanced biological activity, improved stability towards raised temperatures and resistance to proteolytic degradation. As such, cyclisation has the potential to expand the scope of polypeptide application, for example as therapeutics or as industrial biocatalysts. This is of particular importance considering the urgent need for more sustainable practices and technologies.

In this thesis, the effects of cyclisation on two biotechnologically relevant proteins were examined, with the aim of improving their stabilities, in particular towards heat treatment. Specifically, these proteins were TET12SN, an artificial protein cage (Chapter 2), and *Is*PETase, a plastic degrading enzyme (Chapter 3). The future applications of both proteins could benefit from stabilisation, allowing their use outside of physiological conditions.

For TET12SN, cyclisation was achieved by two different approaches. The first was traceless head-to-tail cyclisation using a split intein, while the second was side chain-to-side chain cyclisation using SpyTag/SpyCatcher-N^{TEV}, which remained in the cyclised product. For *Is*PETase, cyclisation was also performed using SpyTag/SpyCatcher(-N^{TEV}) and a range of cyclic topologies were constructed (i.e. a cyclic monomer, cyclic dimer and catenane). Subsequent analysis of thermal stability found none of the cyclic TET12SN or *Is*PETase variants were more resilient to heat treatment compared to their linear counterparts. Nevertheless, cyclisation led to some other interesting observations, these were as follows.

- **Cyclisation of TET12SN resulted in increased aggregate formation**. This likely occurs due to a combination of oligomerisation and protein misfolding. The former could result from intermolecular covalent bond formation between protein molecules. Meanwhile, the latter is likely a consequence of premature cyclisation leading to interference to the stepwise folding pathway of the protein.

- **The use of SpyCatcher-N^{TEV} may not enable efficient protein refolding after heat treatment.** Here, mono-cylic *Is*PETase was prepared using SpyCatcher-N^{TEV}. After heat treatment, the cyclic enzyme did not display improved residual activity compared to the linear WT enzyme. This constrasts with results from separate work in which SpyTag/SpyCatcher cyclised enzymes (including *Is*PETase) have been demonstated to be more resilient towards raised temperatures. The properties of proteins cyclised using modified SpyCatcher variants have yet to be directly

compared, though it is feasible that alteration of the SpyCatcher sequence could perturb the overall stability of the SpyTag/SpyCatcher complex.

- **SpyTag/SpyCatcher(-N^{TEV}) sequences were susceptible to trypsin digest.** The cyclic *Is*PETase variants and linear control were all more rapidly digested by trypsin than the linear WT enzyme. The presence of the SpyTag and SpyCatcher(-N^{TEV}) sequences in the cyclic variants is believed to be the most probable cause.

- **Dimerisation of *Is*PETase improved resistance to agitation**. All *Is*PETase variants exhibited reduced PET degrading activity in the presence of vigorous agitation, likely resulting from an increased rate of inactivation. However, the cyclic dimer and catenane variants generally retained activity for longer than the monomeric variants and yielded more degradation products overall.

In summary, neither the cyclisation of TET12SN nor *Is*PETase resulted in the desired enhancement to protein stability, despite the numerous reports in the literature of the benefits of polypeptide cyclisation. Regardless, it is hoped that the results presented within this thesis may contribute towards a better understanding of protein folding and stabilisation in general. Furthermore, the improved tolerance of dimeric *Is*PETase towards agitation is believed to be an important finding which merits further investigation.

4.2 FUTURE WORK

Future directions of research with specific regard to the cyclisation and stabilisation of TET12SN and *Is*PETase have been outlined in sections 2.4 and 3.4, respectively. In this section future work relating to the cyclisation of proteins in general will be discussed.

Applicability of cyclisation. Cyclisation is widely reported to improve the stability of proteins. Nevertheless, the extent of stabilisation achieved differs greatly between cyclised proteins, ranging from no measurable stabilisation to increases in T_m of around 20 °C. This suggests that cyclisation may not be suitable for all proteins, though it is often unclear why this may be. Several factors have been reported to influence cyclisation, such as the introduction of strain and the formation of intramolecular hydrogen bonds, however, this is often difficult to predict in advance. The use of X-ray crystal structures and MD simulations has enabled optimisation of cyclisation strategies but have only been applied to a limited number of proteins to date. The majority of protein cyclisation studies reported in the literature have concentrated predominantly on the positive outcomes of cyclisation. As a

result, future research should seek to uncover the mechanisms behind the stabilisation (or destabilisation) exhibited by a broader range of cyclised proteins. This could lead to a more tailored approach for achieving stabilisation through cyclisation, as well as result in an overall improved understanding of protein stabilisation in general.

Tolerance of cyclic proteins to agitation. A large number of environmental stresses can lead to protein aggregation and inactivation. Among these, the thermal-induced aggregation of cyclic proteins has understandably been the most extensively studied. It has been reported that cyclic proteins exhibit improved resistance to aggregation over their linear counterparts due to improved conformational and colloidal stability. On the other hand, agitation-induced aggregation of cyclic proteins has not been considered, despite being one of the most commonly encounter environmental stresses. Therefore, further research to examine the effects of agitation on different cyclic protein topologies should be carried out.

CHAPTER 5:

MATERIALS AND METHODS

5.1 GENERAL MATERIALS AND METHODS

All reagents were purchased from Fisher Scientific®, unless stated otherwise. Custom oligonucleotides were purchased from Merck Sigma Aldrich, and restriction enzymes from Thermo Scientific. Deionised water (dH_2O) was obtained from an Elga® PURELAB Chorus 2 system.

5.1.1 Preparation of Growth Media and Stocks

5.1.1.1 Luria-Bertani (LB) Media

Yeast extract (5.0 g), tryptone (10.0 g) and NaCl (10.0 g) were dissolved in 1 L of deionised water (dH_2O). The solution was sterilised by autoclaving at 121 °C for 20 min and stored at 4 °C.

5.1.1.2 LB Agar

Yeast extract (0.5 g), tryptone (1.0 g), NaCl (1.0 g) and agar (1.5 g) were dissolved in 100 mL of dH_2O and sterilised by autoclaving at 121 °C for 20 min. After cooling to around 50 °C, the appropriate antibiotic was added, poured into sterile petri dish and left to solidify at room temperature.

5.1.1.3 Terrific Broth (TB) Media

Yeast extract (24.0 g) and tryptone (20.0 g) were dissolved in 900 mL of dH_2O with glycerol (4 mL) and sterilised by autoclaving at 121 °C for 20 min. After cooling, 100 mL of sterile phosphate buffer (0.17 M KH_2PO_4 and 0.72 M K_2HPO_4) was added.

5.1.1.4 Antibiotic Stock Solutions (1000X)

Kanamycin (500 mg) and ampillicin (1000 mg) were dissolved in 10 mL of dH_2O and stored at -20 °C.

5.1.1.5 IPTG Inducer Stock Solution

To make a 1 M stock solution, IPTG (2.4 g) (Fluorochem) was dissolved in 10 mL of dH_2O and stored at -20 °C.

5.1.2 Calcium Competent *E. coli* Preparation

5.1.2.1 100 mM $CaCl_2$ Competent Cell Solution

$CaCl_2$ (2.2 g) was dissolved in 200 mL of dH_2O. The solution was sterilised by autoclaving at 121 °C for 20 min and then stored at 4 °C.

5.1.2.2 85 mM $CaCl_2$ and 15% (v/v) Glycerol Competent Cell Solution

$CaCl_2$ (0.9 g) was dissolved in 85 mL of dH_2O and 15 mL of glycerol. The solution was sterilised by autoclaving at 121°C for 20 min and then stored at 4°C.

5.1.2.3 Calcium Competent *E. coli* Preparation Protocol

Calcium competent *E. coli* strains BL21(DE3) and SHuffle® T7 Express were used for gene expression, while DH5α was used for plasmid cloning and propagation. SHuffle® T7 Express cultures were incubated at 30 °C, instead of 37 °C. Glycerol stocks of all three strains were available in the laboratory. An aliquot of the appropriate strain was plated on LB agar and incubated overnight at 37 °C. A single colony was selected and used to inoculate 10 mL of LB media. After growing overnight at 37 °C, 2 mL of the culture was used to inoculate 200 mL of LB media which was incubated at 37 °C until an OD_{600} of 0.5-0.6 was reached. The culture was cooled on ice for 15 min before centrifugation (4000 rpm, 4 °C, 15 min). The supernatant was discarded and the cell pellet resuspended in 50 mL of ice cold $CaCl_2$ (100 mM) solution. The suspension was centrifuged (4000 rpm, 4 °C, 15 min), the supernatant discarded and the pellet resuspended in 5 mL of ice cold $CaCl_2$ (85 mM) solution containing 15% (v/v) glycerol. The cell suspension was divided into 100 µL aliquots in sterile microcentrifuge tubes and stored at -80 °C.

As a control for antibiotic resistance, the newly prepared calcium competent cells were plated onto LB agar containing different antibiotics. In addition, transformation efficiency (TE) was assessed by heat-shock transformation of the calcium competent *E. coli* with pUC19. The number of colony forming units (cfu) were counted and applied to **Equation 5.1**.

$$(\boldsymbol{Eq.5.1}) \qquad TE = \frac{Number\ of\ cfu}{\mu g\ DNA}$$

5.1.2.4 Heat-Shock Transformation

A glycerol stock of the calcium competent cells was thawed on ice for 10 min before the plasmid (1 µL, 10-50 ng/µL) was added and incubated on ice for a further 10 min. Next heat-shock was carried out in a water bath at 42 °C for 45 s. After returning to ice for 1 min, 1 mL of LB media was added to the cells and incubated at 37 °C for 1-1.5 h with constant agitation. The cells (150 µL) were spread onto a LB agar plate containing the appropriate antibiotic and then incubated overnight at 37 °C.

5.1.3 Agarose Gel Electrophoresis

5.1.3.1 Tris-Acetate EDTA (TAE) Buffer (50X)

Tris base (242.0 g), disodium EDTA (18.6 g) and glacial acetic acid (57 mL) were dissolved in 1 L of dH_2O and stored at room temperature. For use, the buffer was diluted to 1X with dH_2O.

5.1.3.2 Agarose Gel Preparation

A 1% (w/v) agarose gel was made by dissolving agarose (0.5 g) in hot 1X TAE buffer (50 mL). SYBR® Safe DNA Gel Stain (5 µL) (Invitrogen) was added and the solution poured into a gel cast with a comb to create sample wells. The gel was allowed to set at room temperature.

5.1.3.3 Agarose Gel Electrophoresis Protocol

The agarose gel was submerged in 1X TAE buffer in a horizontal electrophoresis cell. DNA marker (5 µL) (Gene Ruler 1 kb Plus DNA Ladder, Thermo Scientific) and DNA samples were loaded into separate wells in the gel. Electrophoresis was performed at 120 V for 45 min.

5.1.4 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

5.1.4.1 SDS-PAGE Resolving Buffer (4X)

Tris base (36.34 g, 1.5 M) and SDS (0.8 g, 0.4% w/v) were dissolved in 150 mL dH_2O. The pH was adjusted to 8.8 with HCl and the volume made up to 200 mL.

5.1.4.2 SDS-PAGE Stacking Buffer (4X)

Tris base (12.11 g, 0.5 M) and SDS (0.8 g, 0.4% w/v) were dissolved in 150 mL dH_2O. The pH was adjusted to 6.8 with HCl and the volume made up to 200 mL.

5.1.4.3 SDS-PAGE Sample Loading Buffer (4X)

1 M Tris-HCl (2.5 mL, pH 6.8), SDS (0.8 g), bromophenol blue (20 mg) (Alpha Aesar), glycerol (4 mL) and β-mercaptoethanol (2 mL) (Acros) were mixed. The volume was adjusted to 10 mL with dH_2O and stored at room temperature.

5.1.4.4 SDS-PAGE Running Buffer (10X)

Tris base (30.3 g), glycine (144.1 g) and SDS (10.0 g) were dissolved in 1 L of dH_2O and stored at room temperature. For use the buffer was diluted to 1X with dH_2O.

5.1.4.5 Coomassie Brilliant Blue SDS-PAGE Stain

Coomassie Brilliant Blue G (150 mg) (Apollo Scientific) was dissolved in 2 L dH_2O and left to stir at room temperature before the addition of concentrated HCl (8 mL). The solution was stored at room temperature.

5.1.4.6 Polyacrylamide Gel Preparation

Resolving and stacking gels were prepared using the recipes outlined in **Table 5.1**. The resolving gel was prepared first, poured into gel casts and covered with a layer of isopropanol (IPA). Once polymerised, the IPA was removed. Next, the stacking gel was prepared and poured on top of the resolving gel. A comb was inserted to create sample wells before polymerisation.

Table 5.1 Resolving and stacking gel recipes.

Reagent	Resolving Gel (12%)	Resolving Gel (8%)	Stacking Gel
dH_2O	3.40 mL	4.73 mL	2.90 mL
4X resolving/stacking buffer	2.50 mL	2.50 mL	1.25 mL
30% acrylamide/bis-acrylamide solution	4.00 mL	2.67 mL	0.83 mL
TEMED	10 µL	10 µL	5 µL
10% (w/v) APS	100 µL	100 µL	50 µL

5.1.4.7 SDS-PAGE Sample Preparation

Samples (30 µL) were mixed with 4X sample loading buffer (10 µL), before heating at 95°C for 5 min and centrifugation (13,300 rpm, 1 min).

5.1.4.8 Polyacrylamide Gel Electrophoresis Protocol

The gels were placed into a vertical electrophoresis cell containing 1X SDS-PAGE running buffer. The prepared samples and unstained protein molecular weight marker (ThermoFisher Scientic) were loaded into separate wells in the gel. Electrophoresis was performed at (30 mA) for 45 min. The gel was then stained with Coomassie Blue for visualisation of the protein bands.

5.1.5 Molecular Cloning Methods

5.1.5.1 QG Buffer

The QG buffer containing 20 mM Tris base and 550 mM guanidine thiocyanate (Fluorochem) was made by dissolving both compounds in 40 mL of dH_2O. The pH of the solution was adjusted to 6.5 with HCl and made up to a volume of 50 mL with dH_2O. The solution was stored at room temperature.

5.1.5.2 Polymerase Chain Reaction (PCR)/ Site Directed Mutagenesis (SDM)

PCR was performed using PrimeStar® Max DNA polymerase (TaKaRa). On ice, the template DNA (200 ng) and forward and reverse primers (0.5 µL) were mixed with the DNA polymerase (12.5 µL). The volume was made up to 40 µL using dH_2O. The reaction was performed in a ^{3}PrimeG PCR thermal cycler (Techne) using the reaction conditions shown in **Table 5.2**.

Table 5.2 PCR steps.

Step	Temperature /°C	Time	Number of cycles
Initial Denaturation	95	10 min	1
Denaturation	95	10 s	30
Annealing	55	5 s	
Elongation	72	Variable*	
Final Elongation	72	10 min	1

* Elongation times varied depending on the length of the DNA template to be amplified (5s/kbp).

After the PCR cycle was complete, 10X FastDigest Green Buffer (5 µL) (Thermo Scientific) and the restriction enzyme DpnI (0.5 µL) were added to the PCR product. The mixture was incubated at 37 °C for 0.5-1 h before purification by agarose gel electrophoresis followed by extraction and purification.

5.1.5.3 PCR Product Extraction and Purification

The required DNA fragment, was extracted from the agarose gel using a scalpel. QG buffer was added to the excised piece of gel (300 µL per 100 mg of gel) and incubated at 50 °C until the gel completely dissolved. IPA (100 µL per 100 mg of gel) was added, the solution loaded onto a solid phase DNA extraction spin column and centrifuged (13,300 rpm, 1 min). The flow through was discarded and the column washed with PE buffer (750 µL) (Qiagen). The DNA was eluted using 15-30 µL of dH_2O. Concentration was measured using absorbance at 230 nm on a NanoDrop™ spectrophotometer.

5.1.5.4 Gibson Assembly

Gibson Assembly was carried out using NEBuilder® HiFi DNA Assembly Master Mix (2X), which was mixed with the linearised vector and insert DNA fragments to give a reaction volume of 5 µL. A vector-to-insert molar ratio of 1:5 was employed. The mixture was incubated at 50 °C for 15 min, followed by transformation into calcium competent DH5α and plasmid purification.

5.1.5.5 Plasmid Purification

Plasmid purification was carried out using a Qiagen Miniprep Kit. After transformation, a single colony was selected and used to inoculate 10 mL of LB media (containing the appropriate antibiotic) which was incubated overnight at 37 °C. Subsequently, the culture was centrifuged (4000 rpm, 4 °C, 15 min) and purified following the protocol supplied by the manufacturers. Concentration was determined using absorbance at 230 nm on a NanoDrop™ spectrophotometer. Plasmids were stored at -20 °C (and -80 °C for long term storage).

5.1.6 Estimation of Protein Concentration

To estimate protein concentration, absorbance of the sample at 280 nm (A_{280}) was measured on a NanoDrop™ spectrophotometer (Thermo Scientific). From the absorbance value measured, concentration (c, mol/L) was calculated using the Beer-Lambert law (**Equation 5.2**), where A is absorbance, ε is the molar extinction coefficient (M^{-1} cm^{-1}) and l is pathlength (cm). The molecular weight (MW) and ε_{280} of the protein were determined by inputting the amino acid sequence into the Expasy ProtoParam online tool (https://web.expasy.org/protparam/). Conversion of protein concentration into mg/mL is shown by **Equation 5.3**.

$$A = \varepsilon c l \qquad (\boldsymbol{Eq.\,5.2})$$

$$c\ (mg/mL)\ = c(mol/L) \times MW \qquad (\boldsymbol{Eq.\,5.3})$$

5.1.7 Protein Mass Spectrometry

Protein liquid chromatography-mass spectrometry (LC-MS) was carried out by Analytical Services in the School of Chemistry at Cardiff University. The analysis was performed 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 used

was a Waters Acquity UPLC Protein C4 BEH column (300 Å, 1.7 µm, 2.1 by 100 mm), held at 60 °C. A gradient of H_2O containing 0.1% CHO_2H and acetonitrile containing 0.1% CHO_2H was employed. Data was collected in positive electron spray ionisation mode and analysed using Waters MassLynx software version 4.1. Maximum entropy 1 software was used to generate deconvoluted mass spectra.

5.1.8 Circular Dichroism (CD)

All CD measurements were performed on a Chirascan™ CD spectrophotometer (Applied Photophysics) with a temperature control unit. All buffers were filtered and degassed. Protein samples were diluted with the appropriate buffer, filtered and then the concentration measured using absorbance at 280 nm on a NanoDrop™ spectrophotometer. The protein secondary structure spectra were recorded at 20 °C, between 200 and 300 nm (in 1 nm steps), using a quartz cuvette with a path length of 1 mm (Hellma). The observed ellipticity (in mdeg) was converted into units of mean residue ellipticity ($[\theta]_{MRE}$) (**Equation 2.1**). To follow protein thermal denaturation, the signal at 222 nm was monitored as the temperature of the sample was increased from 10 to 90 °C in 2 °C steps (600 s at each temperature). The mid-point of protein unfolding (T_m) was then determined by the fitting of a dose-response equation (**Equation 2.3**). All CD data analyses were carried out using Origin2019b (OriginLab Corporation).

5.2 CYCLISATION OF TET12SN METHODS

5.2.1 Molecular Cloning of TET12SN Variants

The plasmids containing the genes encoding for TET12SN and TET12-Int were already available in the laboratory. The preparation of all other constructs discussed in Chapter 2 are described below, using molecular cloning methods detailed in section 5.1.5. The sequences of all the product plasmids were confirmed by DNA sequencing (Eurofins Genomics). The DNA and amino acid sequences of the TET12 constructs can be found in Appendix 7.1.1 and 7.2.1.

5.2.1.1 Preparation of TET12-Int Mutants

The primers 1 to 16 listed in **Table 5.3** were employed for the SDM of TET12-Int.

Table 5.3 Primers for TET12-Int SDM, where F and R are forward and reverse primers, respectively.

Product	Primer Number	F/R	Sequence (5' to 3')
TET12-Int^{Nat}	1	F	ATTGCGAGCAACTGCTTTAACTTAGGTAGC
	2	R	CAGTTGCTCGCAATAAAGC
	3	F	CACCACCACCACGCGGCGGAATATTGCCTGAGCTACGAAACCG
	4	R	CGTGGTGGTGGTGGTGGTGCGGACCTGAACC
TET12-Int^{C1A}	5	F	CACCACCACCACGCGGCGGCCCTGAGCTACGAAACCG
	6	R	CGTGGTGGTGGTGGTGGTGCGGACCTGAACC
TET12-Int^{N137A}	7	F	AAACGGCTTTATTGCGAGCGCGTGCATGGGCTTAGG
	8	R	AATAAAGCCGTTTTTCAGCGCGAAGTTGTG
TET12-Int^{M+2A}	9	F	ATTGCGAGCAACTGCGCGGGCTTAGGTAGC
	10	R	CAGTTGCTCGCAATAAAGCCG
TET12-Int^{H125N}	11	F	CATCGGTGTTGAACGTGATAACAACTTCG
	12	R	TTCAACACCGATGTCATACACGTTCTG
TET12-Int^{+5}	13	F	TTATTGCGAGCAACTGCATGGCGGGCGAGCGGCCCGGCTTAGGTAGCGGTCCGCATATGC
	14	R	TGCTCGCAATAAAGCCGTTTTTCAGCGCGAAG
TET12-Int^{-5}	15	F	AATATCTGGAAGCTAGCGGTGGTTCAGGTCCGCACCAC
	16	R	GCTTCCAGATATTTCAGTTCGGAGATTTTCTC

5.2.1.2 Preparation of TET12-Spy

The gene for TET12-Spy was purchased (Life Technologies Ltd) and cloned into a pET-28a(+) vector using Gibson Assembly. The linearised vector was generated by restriction enzyme digest of the TET12SN plasmid using NcoI and XhoI restriction enzyme digest (37 °C, 4 h).

5.2.1.3 Preparation of TET12-Spy Mutants

The primers 17 to 20 listed in **Table 5.4** were employed for the SDM of TET12-Spy.

Table 5.4 Primers for TET12-Spy SDM.

Product	Primer Number	F/R	Sequence (5' to 3')
TET12-Spy^{D7A}	17	F	CCACATCGTGATGGTGGCGGCCTACAAGC
	18	R	CATCACGATGTGGGCTCCGC
TET12-Spy^{+5}	19	F	AGTGGTGGCAGCGGAGGTTCTGGTGGTTCTGCCCACATCG
	20	R	GCTGCCACCACTACCAGGAC

5.2.2 TET12SN Expression and Purification

All TET12 constructs were expressed and purified based on procedures reported by Ljubetič et al.[190] Expression and purification were confirmed by SDS-PAGE and MS analysis.

5.2.2.1 TET12SN Expression

TET12 plasmids were transformed into calcium competent BL21(DE3) cells and grown overnight at 37 °C on LB agar plates containing kanamycin (50 µg/mL). A single colony was used to inoculate 10 mL of LB media containing kanamycin (50 µg/mL) and incubated overnight at 37 °C. The starter culture was diluted into 1 L of LB media containing kanamycin (50 µg/mL) and grown at 37 °C until an OD_{600} of 0.5-0.6 was reached. After induction with 0.5 mM IPTG, the culture was incubated overnight at 16 °C, or at 30 °C for 4 hours. The cell pellet was harvested by centrifugation (20 min, 4 °C, 4000 rpm) and stored at -20 °C.

5.2.2.2 TET12SN Ni-NTA

The cell pellet was resuspended in lysis buffer (50 mM Tris base, 120 mM NaCl, 10 mM imidazole, 0.5 mg/mL lysozyme (Sigma Aldrich) and 100 µM phenylmethylsulfonyl fluoride (Apollo Scientific), pH 8.0), sonicated on ice (5 min, 5 s on, 10 s off) and centrifuged (30 min, 4 °C, 20,000 rpm). The supernatant was filtered through a 0.44 µm filter unit and loaded onto a Ni-NTA gravity flow column previously equilibrated with wash buffer (50 mM Tris base, 120 mM NaCl, 10 mM imidazole (Fluorochem), pH 8.0). After washing with buffers containing 10 mM and 20 mM imidazole, the bound protein was eluted with 250 mM imidazole. The eluted fractions were concentrated using a Vivaspin® centrifugal concentrator.

5.2.2.3 TET12SN SEC

Further purification was carried out by SEC (Superdex200 26/600, GE Healthcare) using a fast protein liquid chromatography (FPLC) system (BioRad NGC™ Chromatography System Quest™ 10 Plus). The column was equilibrated with buffer (20 mM Tris base, 150 mM NaCl, 10% (v/v) glycerol, pH 7.5). After the protein sample was loaded, the column was washed using the same buffer (1 mL/min). Protein elution from the column was monitored by absorbance at 280 nm. Fractions containing the desired protein were collected and concentrated using a Vivaspin® centrifugal concentrator. The concentration of the purified protein was measured using absorbance at 280 nm on a NanoDrop™ spectrophotometer. The protein was stored at 4 °C (and -20 °C for longer term storage).

5.2.2.4 TEV Protease Cleavage of SpyCatcher-N^{TEV}

A stock solution of TEV protease (37 µM) (TEVp) was already available in the laboratory. For the cleavage of SpyCatcher-N^{TEV}, TEVp was mixed with TET12-Spy (0.6 mg/mL in SEC buffer, pH 7.5) with a TEVp-to-TET12-Spy molar ratio of 1:4.5. After incubation overnight at room temperature, the 50 µL reaction was quenched by the addition of 4X SDS-PAGE

loading buffer. TEVp and TET12-Spy were incubated separately under the same conditions as control samples. All samples were then analysed by SDS-PAGE.

5.2.3 Stability Analysis of TET12SN Variants

5.2.3.1 TET12SN Proteolytic Stability

A carboxypeptidase Y (CPY) stock solution (33 µM) was made by dissolving CPY (Sigma Aldrich) in dH_2O. The digestions of TET12SN (0.16 mg/mL) and TET12-Int (0.14 mg/mL) were carried out in SEC buffer (pH 6.5) at 37 °C using a CPY-to-TET12 molar ratio of 1:3.7. The progress of the digestion was monitored by taking SDS-PAGE samples at time intervals up to 24 h. Reactions were quenched upon addition of the 4X SDS-PAGE loading buffer. Control reactions were carried out by incubating CPY, TET12SN and TET12-Int separately under the same conditions.

5.2.3.2 TET12SN Thermal Stability

The thermal stabilities of linear and cyclised TET12SN variants were compared by measuring T_m. After SEC, protein samples were buffer exchanged into the appropriate buffer (buffers A-D, section 2.2.3.2.2) and diluted to 0.1 mg/mL. The CD spectra and thermal denaturation curves of the TET12SN variants were recorded and analysed as described in section 5.1.8.

5.2.3.3 TET12SN Chemical Stability

The chemical stabilities of linear and cyclised TET12SN variants were compared by measuring C_m. After SEC, protein samples were concentrated to 2.0 mg/mL and then diluted to 0.03 mg/mL using buffers containing increasing concentrations of urea (0-8 M) (Sigma). Subsequently, the samples were incubated for 1 hour at room temperature, before CD spectra were measured between 217-260 nm at 20 °C. Chemical denaturation curves were fitted using data points taken at 222 nm using a dose-response equation (**Equation 2.3**).

5.3 CYCLISATION OF *Is*PETASE METHODS

5.3.1 Molecular Cloning of *Is*PETase Variants

The preparation of constructs discussed in Chapter 3 are described below using molecular cloning methods detailed in section 5.1.5. The sequences of all the product plasmids were confirmed by DNA sequencing (Eurofins Genomics). The DNA and amino acid sequences of the *Is*PETase constructs can be found in Appendix 7.1.2 and 7.2.2.

5.3.1.1 Preparation of *Is*PETase-WT

The gene for *Is*PETase was purchased (Life Technologies Ltd) and cloned into a pET-21b(+) vector using Gibson Assembly. The linearised vector was generated by restriction enzyme digest using NdeI and XhoI (37 °C, 4 h). The N-terminal signal peptide sequence (Asn2-Ala27) was removed using Gibson Assembly. The primers 21 and 22 listed in **Table 5.5** were used to generate the insert, while the pET-21b(+) vector was linearised using NdeI and XhoI restriction enzyme digest (37 °C, 4 h).

5.3.1.2 Preparation of *Is*PETase-Spy

Gibson Assembly was used for the cloning of *Is*PETase-Spy. The insert was generated from the *Is*PETase-WT plasmid by PCR of the *Is*PETase gene using primers 23 and 24 listed in **Table 5.5**. Restriction enzyme digest (NdeI and NheI, 37 °C, 4 h) of TET12-Spy was used to generate the linear pET-28a(+) vector containing SpyTag and SpyCatcher-N^{TEV} sequences.

5.3.1.3 Preparation of *Is*PETase-Spy^{D7A}

Site directed mutagenesis was used to generate *Is*PETase-Spy^{D7A} using primers 25 and 26 listed in **Table 5.5**.

5.3.1.4 Preparation of *Is*PETase-Tag

The gene encoding for *Is*PETase-Tag was purchased (Life Technologies Ltd) and cloned into a pET-28a(+) vector using Gibson Assembly. The linearised vector was generated by restriction enzyme digest using NcoI and XhoI (37 °C, 4 h).

5.3.1.5 Preparation of *Is*PETase-Catcher

Gibson Assembly was employed for the cloning of the *Is*PETase-Catcher construct. The gene encoding for *Is*PETase and the N-terminal SpyCatcher sequence was purchased (Life Technologies Ltd). It was cloned into a pET-28a(+) vector containing a second (C-terminal) SpyCatcher sequence (see the first Gibson Assembly product in section 5.3.1.6), linearised by restriction enzyme digest using NcoI and NheI (37 °C, 4 h).

5.3.1.6 Preparation of *Is*PETase-Cat

The *Is*PETase-Cat construct was prepared using two sequential Gibson Assembly steps. In the first step, the gene encoding for the SpyTag, p53 dimer domain and SpyCatcher sequences was purchased (Life Technologies Ltd) and cloned into a pET-28a(+) vector linearised using restriction enzyme digest (NcoI and XhoI, 37 °C, 4 h). In the second step,

the gene encoding for *Is*PETase was inserted between the p53 domain and SpyCatcher sequences. The *Is*PETase insert was generated from *Is*PETase-WT by PCR using primers 27 and 28 (**Table 5.5**). The plasmid from the first Gibson Assembly step was linearised using NdeI and NheI restriction enzyme digest (37 °C, 4 h).

5.3.1.7 Preparation of *Is*PETase-Cat Mutants

Site directed mutagenesis was used to generate *Is*PETase-Cat^{+5}, -Cat^{D7A} and –Cat^{K19P} using primers 29 to 34 listed in **Table 5.5**.

Table 5.5 Primers used for PCR and SDM of the *Is*PETase variants.

Product	Primer Number	F/R	Sequence (5' to 3')
*Is*PETase-WT (no signal peptide)	21	F	CTTTAAGAAGGAGATATACAATGCAGACCAATCCGTATGC
	22	R	AGTGGTGGTGGTGGTGGTGCTCG
*Is*PETase-Spy	23	F	CGTGATTCAGTGGATGGCCATCAGACCAATCCGTATGC
	24	R	TGCCACCACTACCAGGACCGCTAGCCTCGAGGGAACAGTTCG
*Is*PETase-Spy^{D7A}	25	F	ATCGTGATGGTGGCGGCCTACAAGC
	26	R	ACCATCACGATGTGGGCTC
*Is*PETase-Cat	27	F	GGCTCTGGTGGCAGCGGCCATATGCAGACCAATCCGTATGC
	28	R	GGTTTTCGCCAGAGCCGCCGCTAGCGGAACAGTTCGCGGTGC
*Is*PETase-Cat^{+5}	29	F	TCTGGTGGCAGCGGCGGTAGCGGTGGTTCTCATATGCAGACCAATCCG
	30	R	CTGCCACCAGAGCCACCCGGTTC
*Is*PETase-Cat^{D7A}	31	F	TATTGTGATGGTGGCGGCGTATAAACC
	32	R	CACCATCACAATATGC
*Is*PETase-Cat^{K19P}	33	F	GAAGAATTTCGCGAACCGAACGAAGCGCTGGAACTG
	34	R	CGCGAAATTCTTCAAAGCGTTCG

5.3.2 *Is*PETase Expression and Purification

With the exception of *Is*PETase-WT, all *Is*PETase constructs were grown in media or agar containing kanamycin (50 µg/mL). For *Is*PETase-WT ampicillin (100 µg/mL) was used.

5.3.2.1 *Is*PETase Expression

*Is*PETase plasmids were transformed into Shuffle® T7 Express calcium competent cells and grown overnight at 30 °C on agar plates containing the appropriate antibiotic. A single colony was used to inoculate 10 mL of LB media containing antibiotic and incubated overnight at 30 °C. The starter culture was diluted into 1 L of TB media containing antibiotic and grown at 30 °C until an OD_{600} of 0.8-1.0 was reached. After induction with 0.5 mM IPTG,

the culture was incubated overnight at 18 °C. The cell pellet was harvested by centrifugation (20 min, 4 °C, 4000 rpm) and stored at -20 °C.

5.3.2.2 *Is*PETase Ni-NTA

The cell pellet was resuspended in lysis buffer (20 mM Tris base, pH 8.0, 2 mM $MgCl_2$, 20 mM NaCl, 10% glycerol) and incubated on ice with Benzonase® Nuclease (Merck) for 15-30 min. After sonicated on ice (5 min, 5 s on, 10 s off) and centrifugation (30 min, 4 °C, 20,000 rpm), the supernatant was filtered through a 0.44 µm filter unit and loaded onto a Ni-NTA gravity flow column previously equilibrated with wash buffer (50 mM Tris base, 120 mM NaCl, 10 mM imidazole, pH 8.0). After washing with buffers containing 10 mM and 20 mM imidazole, the bound protein was eluted with buffer containing 500 mM imidazole. The eluted fractions were concentrated using a Vivaspin® centrifugal concentrator.

5.3.2.3 *Is*PETase SEC

Further purification was carried out by SEC (Superdex200 26/600 or Superdex75 26/600 , GE Healthcare) using a fast protein liquid chromatography (FPLC) system (BioRad NGC™ Chromatography System Quest™ 10 Plus). The column was equilibrated with buffer (50 mM Na_2HPO_4 (Acros), 100 mM NaCl, pH 7.5 or 8.0). After the protein sample was loaded, the column was washed using the same buffer (1.5 mL/min). Protein elution from the column was monitored by absorbance at 280 nm. Fractions containing the desired protein were collected and concentrated using a Vivaspin® centrifugal concentrator. The concentration of the protein was measured using absorbance at 280 nm on a NanoDrop™ spectrophotometer. The protein was stored at 4 °C.

5.3.3 Stability Analysis of *Is*PETase Variants

5.3.3.1 *Is*PETase Thermal Stability

The thermal stabilities of the linear and cyclised *Is*PETase variants were compared by measuring T_m. After SEC, protein samples were diluted to around 0.1 mg/mL. CD spectra and thermal denaturation curves were recorded in SEC buffer (pH 8.0) as described in section 5.1.8.

5.3.3.2 *Is*PETase Trypsin Digest

A stock solution of pancreatic bovine trypsin (40 µM) (Melford) was made by dissolving 1 mg in 20 mM $CaCl_2$ and 1 mM HCl, this was then diluted with SEC buffer (pH 8.0) to 4 µM. The digestions of the *Is*PETase variants (10 µM) were carried out in SEC buffer (pH 8.0) at 30, 40 and 50 °C using a trypsin-to-*Is*PETase molar ratio of 1:100. The progress of the

digestion was monitored by taking SDS-PAGE samples at time intervals up to 4 h. Reactions were quenched upon addition of the 4X SDS-PAGE loading buffer followed by the addition of 8 M urea (final concentration of 5 M). Samples were analysed by SDS-PAGE.

5.3.3.3 *Is*PETase SDS Denaturation

Two SDS-PAGE samples were prepared by mixing purified *Is*PETase (1 mg/mL, 30 µL) with 4X sample loading buffer (10 µL). One sample was boiled for 10 minutes, the other remained unheated. Both samples were then analysed by SDS-PAGE.

5.3.4 Residual Activity Assay

5.3.4.1 Hydrolysis of *p*-NPA

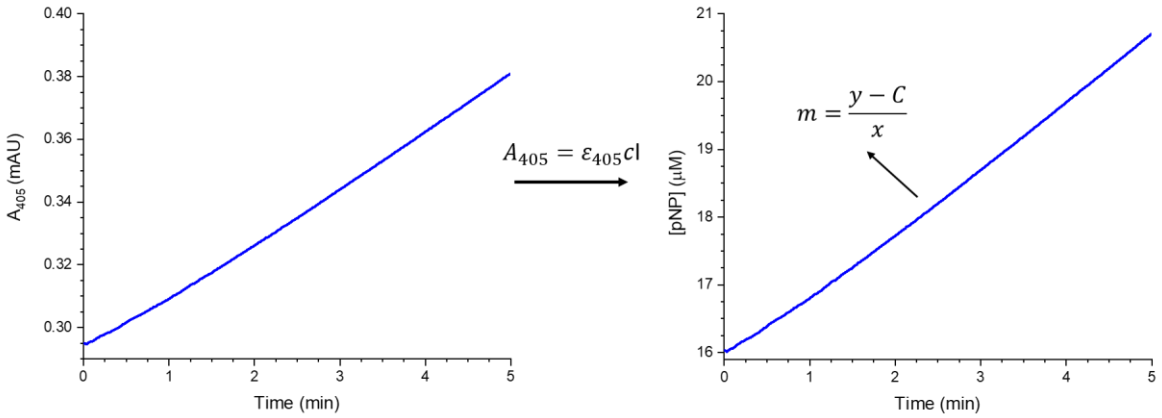
The production of *p*-NP from the hydrolysis of *p*-NPA (Acros) was followed spectrophotometrically using a UV-2600 UV-Vis spectrophotometer (Shimadzu) by monitoring the increase in absorbance at 405 nm for 300 s. All measurements were carried out in 50 mM Na_2HPO_4, pH 7.5, 100 mM buffer with 5% (v/v) DMSO at 30 °C.

A *p*-NPA stock solution was prepared at a concentration of 250 mM in DMSO. The stock solution was then diluted to the appropriate starting concentrations (2.5 - 100 mM) using DMSO. Buffer was pre-incubated at 30 °C in quartz cuvettes (Hellma) for 5-10 min before the addition of *p*-NPA starting solutions (20 µL in 378 µL buffer), giving final substrate concentrations of 0.125 - 5.0 mM. To initiate the enzymatic reaction, 2 µL of enzyme stock solution was added to the substrate and buffer solution, giving an enzyme concentration of 5 (*Is*PETase-Dimer and –Cat^{+5}) or 10 nM (*Is*PETase-WT and -Spy). For non-enzymatic hydrolysis reactions, 2 µL of buffer was added to the buffer and substrate solution. Each measurement was repeated a minimum of three times. Data was analysed as described below using Origin2019b (OriginLab Corporation).

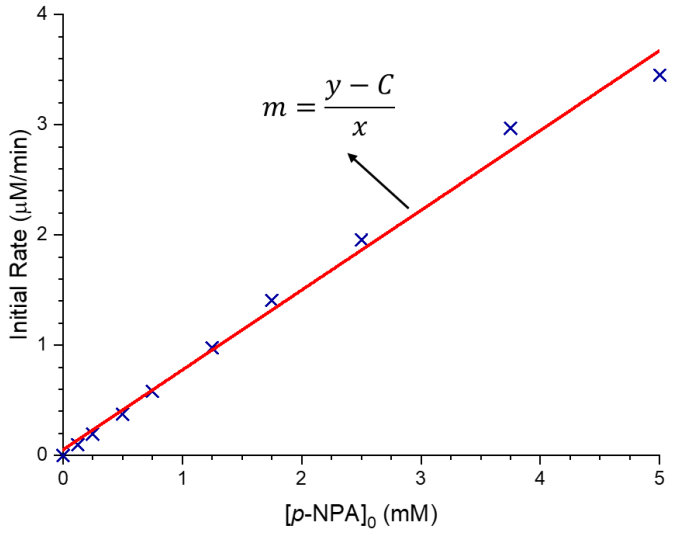
5.3.4.2 Non-Enzymatic Kinetics

The Beer-Lambert law (**Equation 5.2**) was used to convert the absorbance measured at 405 nm (A_{405}) into concentration of *p*-NP produced ([pNP]), with an ε_{405} of 18,400 M^{-1} cm^{-1}. Concentration of *p*-NP product (y, µM) was plotted against time (x, min), and the gradient of the slope (m, µM min^{-1}) was calculated using **Equation 5.4**, where C is the intercept.

$$(\boldsymbol{Eq.\ 5.4}) \qquad y = mx + C$$

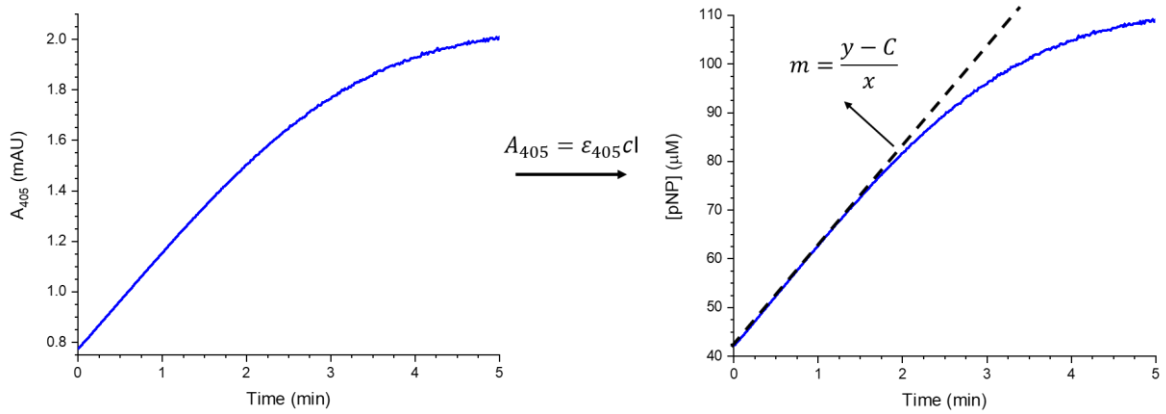


The average gradient for each *p*-NPA starting concentration ([pNPA]$_0$, mM) was plotted against the corresponding [pNPA]$_0$ and fitted using **Equation 5.4**, where slope (m) was the first order rate constant for non-enzymatic *p*-NPA hydrolysis (min^{-1}).



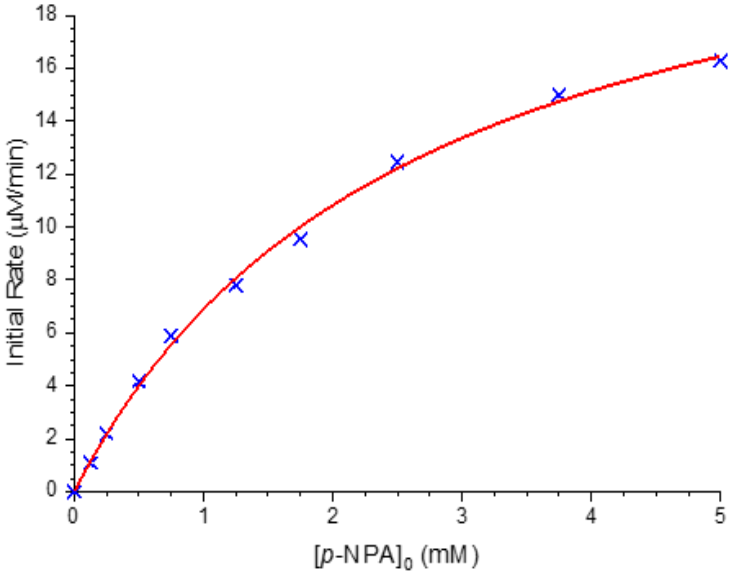
5.3.4.3 Enzyme Kinetics Analysis

The 405 nm absorbance measurements were converted into [pNP] released using **Equation 5.2**, and then plotted against time. The gradient of the initial linear section of the curves was calculated using **Equation 5.4**, and the rate of non-enzymatic hydrolysis subtracted.



The average gradient for each $[pNPA]_0$ was plotted against the corresponding $[pNPA]_0$ and the curve fitted using the Michaelis-Menten model (**Equation 5.5**).

$$(\boldsymbol{Eq.\ 5.5}) \qquad V = \frac{V_{max}[S]}{K_m + [S]}$$



The values obtained for the maximum reaction rate (V_{max}, μM min^{-1}) and Michaelis constant (K_m, mM) were used to calculate the turnover number (k_{cat}, s^{-1}) (**Equation 5.6**) and catalytic efficiency (k_{cat}/K_m, M^{-1} s^{-1}) for the enzymatic hydrolysis of *p*-NPA. [E] is equal to the concentration of active sites and determined using a BCA assay.

$$(\boldsymbol{Eq.5.6}) \qquad k_{cat} = V_{max}/[E]$$

5.3.4.4 Bicinchroninic Acid (BCA) Assay

BCA (Sigma) and Cu_2SO_4 solutions were mixed in a 1-to-49 ratio. 90 µL of the mixture was added to 10 µL of the *Is*PETase sample in a clear non-treated 96-well microplate (Corning®). After incubation at 37 °C for 30 min, the absorbance at 562 nm was measured using a FLUOstar Omega microplate reader (BMG Labtech). Bovine serum albumin (BSA) (Pierce™, Thermo Scientific) was used to generate a standard curve (25 – 500 µg/mL) from which the concentration of the *Is*PETase sample was determined. Measurements were carried out in triplicate.

5.3.4.5 Residual Activity Kinetics

200 µL of enzyme (250 µM of *Is*PETase-WT and –Spy, 125 µM of *Is*PETase-Dimer and –Cat^{+5}) was heated at 50 °C in a TC-512 PCR thermal cycler (Techne) for 5 min - 12 h. After heating, the enzyme was incubated at 4 °C for 10 min before centrifugation to remove any precipitate. 100 µL of the supernatant was transferred into an Eppendorf tube and diluted to the appropriate concentration using buffer. Residual activity was determined using the hydrolysis of *p*-NPA, as described above (section 5.3.4.1). For each sample, the enzyme kinetics were analysed as previously described using the Michaelis-Menten model (**Equation 5.5**). The percentage residual activity was calculated using **Equation 5.7**, where t is the time the sample was heated at 50 °C.

$$(\boldsymbol{Eq.5.7}) \qquad \%\ Residual\ Activity = \frac{[k_{cat}/K_m]_{t=x}}{[k_{cat}/K_m]_{t=0}} \times 100$$

5.3.5 PET Hydrolysis

The activities of the *Is*PETase variants for PET degradation, were investigated according to a modified procedure reported by Arnling Bååth et al.[293] Absorbance spectra were recorded between 200-500 nm using a UV-Vis spectrophotometer and quartz cuvettes.

Semi-crystalline PET powder (Goodfellow) with particle size of <300 µm was suspended in 50 mM Na_2HPO_4 (pH 8.0) and 100 mM NaCl buffer. Reactions were performed in triplicate on a 1 mL scale in Eppendorf tubes, with a PET concentration of 7.5 mg/mL. Enzymes were used in concentrations of 0.1 (*Is*PETase-WT or -Spy) or 0.05 µM (*Is*PETase-Dimer or -Cat^{+5}). For PET only and enzyme only control samples, the appropriate volume of buffer was used to replace the enzyme or the PET suspension, respectively. Samples were incubated in a ThermoMixer™ (Eppendorf) at the required temperature (30, 40 or 50 °C) with shaking (500 or 1100 rpm) or without shaking (0 rpm). At selected time intervals,

samples were quenched by centrifugation (13,300 rpm, 10 min, 4 °C) and the supernatant analysed using absorbance at 240 nm (A_{240}). Measurements were made in triplicate and the A_{240} averaged. The A_{240} measured for the enzyme only control samples were subtracted and then the concentration of soluble degradation products calculated using **Equation 5.2** (ε_{240} = 13,800 M^{-1} cm^{-1}).

CHAPTER 6:
REFERENCES

1. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science (1979)* **181**, 223–230 (1973).

2. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).

3. Cobb, M. 60 years ago, Francis Crick changed the logic of biology. *PLoS Biol.* **15**, e2003243 (2017).

4. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).

5. Abdelraheem, E. M. M., Busch, H., Hanefeld, U. & Tonin, F. Biocatalysis explained: From pharmaceutical to bulk chemical production. *React. Chem. Eng.* **4**, 1878–1894 (2019).

6. Rigoldi, F., Donini, S., Redaelli, A., Parisini, E. & Gautieri, A. Review: Engineering of thermostable enzymes for industrial applications. *APL Bioeng.* **2**, 011501 (2018).

7. Ibrahim, N. E. & Ma, K. Industrial Applications of Thermostable Enzymes from Extremophilic Microorganisms. *Curr. Biochem. Eng.* **4**, 75–98 (2017).

8. Tapeinou, A., Matsoukas, M. T., Simal, C. & Tselios, T. Review cyclic peptides on a merry-go-round; towards drug design. *Biopolymers* **104**, 453–461 (2015).

9. Kaas, Q. & Craik, D. J. Analysis and Classification of Circular Proteins in CyBase. *Peptide Science* **94**, 584–591 (2010).

10. Flory, P. J. Theory of elastic mechanisms in fibrous proteins. *J. Am. Chem. Soc.* **78**, 5222–5235 (1956).

11. Purkayastha, A. & Kang, T. J. Stabilization of Proteins by Covalent Cyclization. *Biotechnol. Bioprocess Eng.* **24**, 702–712 (2019).

12. Zorzi, A., Deyle, K. & Heinis, C. Cyclic peptide therapeutics: past, present and future. *Curr. Opin. Chem. Biol.* **38**, 24–29 (2017).

13. Park, S. E., Sajid, M. I., Parang, K. & Tiwari, R. K. Cyclic cell-penetrating peptides as efficient intracellular drug delivery tools. *Mol. Pharm.* **16**, 3727–3743 (2019).

14. Hayes, H. C., Luk, L. Y. P. & Tsai, Y.-H. Approaches for peptide and protein cyclisation. *Org. Biomol. Chem.* **19**, 3983–4001 (2021).

15. Cossy, J. & Pale-Grosdemange, C. A convenient synthesis of amides from carboxylic acids and primary amines. *Tetrahedron Lett.* **30**, 2771–2774 (1989).

16. Perreux, L., Loupy, A. & Volatron, F. Solvent-free preparation of amides from acids and primary amines under microwave irradiation. *Tetrahedron* **58**, 2155–2162 (2002).

17. Montalbetti, C. A. G. N. & Falque, V. Amide bond formation and peptide coupling. *Tetrahedron* **61**, 10827–10852 (2005).

18. Dunetz, J. R., Magano, J. & Weisenburger, G. A. Large-Scale Applications of Amide Coupling Reagents for the Synthesis of Pharmaceuticals. *Org. Process Res. Dev.* **20**, 140–177 (2016).

19. El-Faham, A. & Albericio, F. Peptide Coupling Reagents, More than a Letter Soup. *Chem. Rev.* **111**, 6557–6602 (2011).

20. Jin, K. *et al.* Total synthesis of Teixobactin. *Nat. Commun.* **7**, 12394 (2016).

21. Staudinger, H. & Meyer, J. Über neue organische Phosphorverbindungen III. Phosphinmethylenderivate und Phosphinimine. *Helv. Chim. Acta* **2**, 635–646 (1919).

22. Nilsson, B. L., Kiessling, L. L. & Raines, R. T. Staudinger ligation: A peptide from a thioester and azide. *Org. Lett.* **2**, 1939–1941 (2000).

23. Soellner, M. B., Nilsson, B. L. & Raines, R. T. Reaction mechanism and kinetics of the traceless Staudinger ligation. *J. Am. Chem. Soc.* **128**, 8820–8828 (2006).

24. Kleineweischede, R. & Hackenberger, C. P. R. Chemoselective peptide cyclization by traceless staudinger ligation. *Angew. Chemi. Int. Ed.* **47**, 5984–5988 (2008).

25. Tam, A., Soellner, M. B. & Raines, R. T. Water-soluble phosphinothiols for traceless Staudinger ligation and integration with expressed protein ligation. *J. Am. Chem. Soc.* **129**, 11421–11430 (2007).

26. Soellner, M. B., Tam, A. & Raines, R. T. Staudinger ligation of peptides at non-glycyl residues. *J. Org. Chem.* **71**, 9824–9830 (2006).

27. Bode, J. W., Fox, R. M. & Baucom, K. D. Chemoselective amide ligations by decarboxylative condensations of N-alkylhydroxylamines and α-ketoacids. *Angew. Chem. Int. Ed.* **45**, 1248–1252 (2006).

28. Fukuzumi, T., Ju, L. & Bode, J. W. Chemoselective cyclization of unprotected linear peptides by α-ketoacid-hydroxylamine amide-ligation. *Org. Biomol. Chem.* **10**, 5837–5844 (2012).

29. Pattabiraman, V. R., Ogunkoya, A. O. & Bode, J. W. Chemical protein synthesis by chemoselective α-ketoacid-hydroxylamine (KAHA) ligations with 5-oxaproline. *Angew. Chem. Int. Ed.* **51**, 5114–5118 (2012).

30. Rohrbacher, F., Deniau, G., Luther, A. & Bode, J. W. Spontaneous head-to-tail cyclization of unprotected linear peptides with the KAHA ligation. *Chem. Sci.* **6**, 4889–4896 (2015).

31. Rohrbacher, F., Zwicky, A. & Bode, J. W. Chemical synthesis of a homoserine-mutant of the antibacterial, head-to-tail cyclized protein AS-48 by α-ketoacid-hydroxylamine (KAHA) ligation. *Chem. Sci.* **8**, 4051–4055 (2017).

32. Bock, V. D., Speijer, D., Hiemstra, H. & van Maarseveen, J. H. 1,2,3-Triazoles as peptide bond isosteres: Synthesis and biological evaluation of cyclotetrapeptide mimics. *Org. Biomol. Chem.* **5**, 971–975 (2007).

33. Bock, V. D., Hiemstra, H. & van Maarseveen, J. H. CuI-Catalyzed Alkyne—Azide "Click" Cycloadditions from a Mechanistic and Synthetic Perspective. *Eur. J. Org.* **37**, (2006).

34. Rostovtsev, V. v., Green, L. G., Fokin, V. v. & Sharpless, K. B. A stepwise huisgen cycloaddition process: Copper(I)-catalyzed regioselective 'ligation' of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **41**, 2596–2599 (2002).

35. Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-Triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **67**, 3057–3064 (2002).

36. Pickens, C. J., Johnson, S. N., Pressnall, M. M., Leon, M. A. & Berkland, C. J. Practical Considerations, Challenges, and Limitations of Bioconjugation via Azide-Alkyne Cycloaddition. *Bioconjug. Chem.* **29**, 686–701 (2018).

37. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046–15047 (2004).

38. Chadwick, R. C., van Gyzen, S., Liogier, S. & Adronov, A. Scalable synthesis of strained cyclooctyne derivatives. *Synthesis (Stuttg)* **46**, 669–677 (2014).

39. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. H. Synthesis of proteins by native chemical ligation. *Science (1979)* **266**, 776–779 (1994).

40. Dawson, P. E. & Kent, S. B. H. Synthesis of Native Proteins by Chemical Ligation. *Annu. Rev. Biochem.* **69**, 923–960 (2000).

41. Conibear, A. C., Watson, E. E., Payne, R. J. & Becker, C. F. W. Native chemical ligation in protein synthesis and semi-synthesis. *Chem. Soc. Rev.* **47**, 9046–9068 (2018).

42. Wan, Q. & Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: A powerful advance in the synthesis of polypeptides and glycopolypeptides. *Ang. Chem. Int. Ed.* **46**, 9248–9252 (2007).

43. Malins, L. R. & Payne, R. J. Synthetic amino acids for applications in peptide ligation-desulfurization chemistry. *Aust. J. Chem.* **68**, 521–537 (2015).

44. Zhang, L. & Tam, J. P. Synthesis and application of unprotected cyclic peptides as building blocks for peptide dendrimers. *J. Am. Chem. Soc.* **119**, 2363–2370 (1997).

45. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).

46. Jacob, C., Giles, G. I., Giles, N. M. & Sies, H. Sulfur and Selenium: The Role of Oxidation State in Protein Structure and Function. *Angew. Chem. Int. Ed.* **42**, 4742–4758 (2003).

47. Hatahet, F. & Ruddock, L. W. Protein disulfide isomerase: A critical evaluation of its function in disulfide bond formation. *Antioxid. Redox Signal.* **11**, 2807–2850 (2009).

48. 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).

49. Zhang, R. Y., Thapa, P., Espiritu, M. J., Menon, V. & Bingham, J. P. From nature to creation: Going around in circles, the art of peptide cyclization. *Bioorg. Med. Chem.* **26**, 1135–1150 (2018).

50. Leduc, A. M. *et al.* Helix-stabilized cyclic peptides as selective inhibitors of steroid receptor - Coactivator interactions. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 11273–11278 (2003).

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

52. Chua, K. *et al.* Small cyclic agonists of iron regulatory hormone hepcidin. *Bioorg. Med. Chem. Lett.* **25**, 4961–4969 (2015).

53. Wang, L., Wang, X. & Wang, C. C. Protein disulfide-isomerase, a folding catalyst and a redox-regulated chaperone. *Free Radic. Biol. Med.* **83**, 305–313 (2015).

54. Postma, T. M. & Albericio, F. Disulfide Formation Strategies in Peptide Synthesis. *Eur. J. Org.* **2014**, 3519–3530 (2014).

55. Wu, C., Leroux, J. C. & Gauthier, M. A. Twin disulfides for orthogonal disulfide pairing and the directed folding of multicyclic peptides. *Nat. Chem.* **4**, 1044–1049 (2012).

56. Zheng, Y., Zhai, L., Zhao, Y. & Wu, C. Orthogonal Cysteine-Penicillamine Disulfide Pairing for Directing the Oxidative Folding of Peptides. *J. Am. Chem. Soc.* **137**, 15094–15097 (2015).

57. Morihara, K. Using proteases in peptide synthesis. *Trends Biotechnol.* **5**, 164–170 (1987).

58. Jakubke, H. -D, Kuhl, P. & Könnecke, A. Basic Principles of Protease-Catalyzed Peptide Bond Formation. *Angew. Chem. Int. Ed.* **24**, 85–93 (1985).

59. Guzmán, F., Barberis, S. & Illanes, A. Peptide synthesis: Chemical or enzymatic. *Electron. J. Biotechnol.* **10**, 279–314 (2007).

60. Martin, R. B. Free energies and equilibria of peptide bond hydrolysis and formation. *Biopolymers* **45**, 351–353 (1998).

61. Blanco, R. M., Alvaro, G. & Guisán, J. M. Enzyme reaction engineering: Design of peptide synthesis by stabilized trypsin. *Enzyme Microb. Technol.* **13**, 573–583 (1991).

62. Barbas, C. F., Matos, J. R., West, J. B. & Wong, C. H. A Search for Peptide Ligase: Cosolvent-Mediated Conversion of Proteases to Esterases for Irreversible Synthesis of Peptides. *J. Am. Chem. Soc.* **110**, 5162–5166 (1988).

63. Homandberg, G. A., berg, Mattis, J. A. & Laskowski, M. Synthesis of Peptide Bonds by Proteinases. Addition of Organic Cosolvents Shifts Peptide Bond Equilibria toward Synthesis. *Biochemistry* **17**, 5220–5227 (1978).

64. Klibanov, A. M. Improving enzymes by using them in organic solvents. *Nature* **409**, 241–246 (2001).

65. Weeks, A. M. & Wells, J. A. Subtiligase-Catalyzed Peptide Ligation. *Chem. Rev.* **120**, 3127–3160 (2019).

66. Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev.* **102**, 4501–4523 (2002).

67. West, J. B. *et al.* Modification of Proteases to Esterases for Peptide Synthesis: Methylchymotrypsin. *J. Am. Chem. Soc.* **110**, 3709–3710 (1988).

68. Mazmanian, S. K., Liu, G., Ton-That, H. & Schneewind, O. Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science (1979)* **285**, 760–763 (1999).

69. Spirig, T., Weiner, E. M. & Clubb, R. T. Sortase enzymes in Gram-positive bacteria. *Mol. Microbiol.* **82**, 1044–1059 (2011).

70. Navarre, W. W. & Schneewind, O. Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in Gram-positive bacteria. *Mol. Microbiol.* **14**, 115–121 (1994).

71. Kruger, R. G. *et al.* Analysis of the Substrate Specificity of the Staphylococcus aureus Sortase Transpeptidase SrtA. *Biochemistry* **43**, 1541–1551 (2004).

72. Suree, N. *et al.* The structure of the Staphylococcus aureus sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. *J. Biol. Chem.* **284**, 24465–24477 (2009).

73. Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O. & Narayana, S. V. L. Crystal structures of Staphylococcus aureus Sortase A and its substrate complex. *J. Biol. Chem.* **279**, 31383–31389 (2004).

74. Mao, H., Hart, S. A., Schink, A. & Pollok, B. A. Sortase-Mediated Protein Ligation: A New Method for Protein Engineering. *J. Am. Chem. Soc.* **126**, 2670–2671 (2004).

75. Huang, X. *et al.* Kinetic mechanism of Staphylococcus aureus sortase SrtA. *Biochemistry* **42**, 11307–11315 (2003).

76. Wu, Z., Guo, X. & Guo, Z. Sortase A-catalyzed peptide cyclization for the synthesis of macrocyclic peptides and glycopeptides. *ChemComm.* **47**, 9218–9220 (2011).

77. Popp, M. W., Dougan, S. K., Chuang, T. Y., Spooner, E. & Ploegh, H. L. Sortase-catalyzed transformations that improve the properties of cytokines. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 3169–3174 (2011).

78. Antos, J. M. *et al.* A straight path to circular proteins. *J. Biol. Chem.* **284**, 16028–16036 (2009).

79. Pritz, S. *et al.* Synthesis of Biologically Active Peptide Nucleic Acid−Peptide Conjugates by Sortase-Mediated Ligation. *J. Org. Chem.* **72**, 3909–3912 (2007).

80. Liu, F., Luo, E. Y., Flora, D. B. & Mezo, A. R. Irreversible sortase a-mediated ligation driven by diketopiperazine formation. *J. Org. Chem.* **79**, 487–492 (2014).

81. Policarpo, R. L. *et al.* Flow-based enzymatic ligation by sortase A. *Angew. Chem. Int. Ed.* **53**, 9203–9208 (2014).

82. 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. U. S. Am.* **108**, 11399–11404 (2011).

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

84. Hirakawa, H., Ishikawa, S. & Nagamune, T. Design of Ca 2+-independent Staphylococcus aureus sortase A mutants. *Biotechnol. Bioeng.* **109**, 2955–2961 (2012).

85. Homandberg, G. A. & Laskowski, M. Enzymatic Resynthesis of the Hydrolyzed Peptide Bond(s) in Ribonuclease S. *Biochemistry* **18**, 586–592 (1979).

86. Nakatsuka, T., Sasaki, T. & Kaiser, E. T. Peptide Segment Coupling Catalyzed by the Semisynthetic Enzyme Thiolsubtilisin. *J. Am. Chem. Soc.* **109**, 3808–3810 (1987).

87. Abrahms, L. *et al.* Engineering Subtilisin and Its Substrates for Efficient Ligation of Peptide Bonds in Aqueous Solution. *Biochemistry* **30**, 4151–4159 (1991).

88. Jackson, D. Y., Burnier, J. P. & Wells, J. A. Enzymatic Cyclization of Linear Peptide Esters Using Subtiligase. *J. Am. Chem. Soc.* **117**, 819–820 (1995).

89. Weeks, A. M. & Wells, J. A. Engineering peptide ligase specificity by proteomic identification of ligation sites. *Nat. Chem. Biol.* **14**, 50–57 (2018).

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

91. Toplak, A., Nuijens, T., Quaedflieg, P. J. L. M., Wu, B. & Janssen, D. B. Peptiligase, an Enzyme for Efficient Chemoenzymatic Peptide Synthesis and Cyclization in Water. *Ad. Synth. Catal.* **358**, 2140–2147 (2016).

92. Schmidt, M. *et al.* Omniligase-1: A Powerful Tool for Peptide Head-to-Tail Cyclization. *Ad. Synth. Catal.* **359**, 2050–2055 (2017).

93. Schmidt, M. *et al.* Efficient Enzymatic Cyclization of Disulfide-Rich Peptides by Using Peptide Ligases. *ChemBioChem* **20**, 1524–1529 (2019).

94. Welker, E. & Scheraga, H. A. Use of benzyl mercaptan for direct preparation of long polypeptide benzylthio esters as substrates of subtiligase. *Biochem. Biophys. Res. Commun.* **254**, 147–151 (1999).

95. de Veer, S. J., Kan, M. W. & Craik, D. J. Cyclotides: From Structure to Function. *Chem. Rev.* **119**, 12375–12421 (2019).

96. Tam, J. P., Chan, N. Y., Liew, H. T., Tan, S. J. & Chen, Y. Peptide asparaginyl ligases—renegade peptide bond makers. *Sci. China Chem.* **63**, 296–307 (2020).

97. Zhao, L. *et al.* Structural analysis of asparaginyl endopeptidase reveals the activation mechanism and a reversible intermediate maturation stage. *Cell Res.* **24**, 344–358 (2014).

98. Yang, R. *et al.* Engineering a Catalytically Efficient Recombinant Protein Ligase. *J. Am. Chem. Soc.* **139**, 5351–5358 (2017).

99. Hemu, X. *et al.* Structural determinants for peptide-bond formation by asparaginyl ligases. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11737–11746 (2019).

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

101. 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).

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

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

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

105. 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).

106. 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).

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

108. 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).

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

110. 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).

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

112. Novikova, O. *et al.* Intein clustering suggests functional importance in different domains of life. *Mol. Biol. Evol.* **33**, 783–799 (2016).

113. Pietrokovski, S. Intein spread and extinction in evolution. *Trends Genet.* **17**, 465–472 (2001).

114. Shah, N. H. & Muir, T. W. Inteins: Nature's gift to protein chemists. *Chem. Sci.* **5**, 446–461 (2014).

115. Pavankumar, T. Inteins: Localized Distribution, Gene Regulation, and Protein Engineering for Biological Applications. *Microorganisms* **6**, 19 (2018).

116. Barzel, A. *et al.* Native homing endonucleases can target conserved genes in humans and in animal models. *Nucleic Acids Res.* **39**, 6646–6659 (2011).

117. Chong, S. & Xu, M. Q. Protein splicing of the Saccharomyces cerevisiae VMA intein without the endonuclease motifs. *J. Biol. Chem.* **272**, 15587–15590 (1997).

118. Gogarten, J. P., Senejani, A. G., Zhaxybayeva, O., Olendzenski, L. & Hilario, E. Inteins: Structure, Function, and Evolution. *Annu. Rev. Microbiol.* **56**, 263–287 (2002).

119. Pietrokovski, S. Identification of a virus intein and a possible variation in the protein-splicing reaction. *Curr. Biol.* **8**, (1998).

120. Eryilma, E., Shah, N. H., Muir, T. W. & Cowburn, D. Structural and dynamical features of Inteins and implications on protein splicing. *J. Biol. Chem.* **289**, 14506–14511 (2014).

121. Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6705–6710 (1998).

122. Camarero, J. A. *et al.* Rescuing a destabilized protein fold through backbone cyclization. *J. Mol. Biol.* **308**, 1045–1062 (2001).

123. Iwai, H. & Plückthun, A. Circular β-lactamase: Stability enhancement by cyclizing the backbone. *FEBS Lett.* **459**, 166–172 (1999).

124. Sorci, M. *et al.* Oriented covalent immobilization of antibodies for measurement of intermolecular binding forces between zipper-like contact surfaces of split inteins. *Anal. Chem.* **85**, 6080–6088 (2013).

125. Shah, N. H., Eryilmaz, E., Cowburn, D. & Muir, T. W. Naturally split inteins assemble through a 'capture and collapse' mechanism. *J. Am. Chem. Soc.* **135**, 18673–18681 (2013).

126. Wu, H., Hu, Z. & Liu, X. Q. Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9226–9231 (1998).

127. Shi, J. & Muir, T. W. Development of a tandem protein trans-splicing system based on native and engineered split inteins. *J. Am. Chem. Soc.* **127**, 6198–6206 (2005).

128. Martin, D. D., Xu, M. Q. & Evans, T. C. Characterization of a naturally occurring trans-splicing intein from Synechocystis sp. PCC6803. *Biochemistry* **40**, 1393–1402 (2001).

129. Iwai, H., Züger, S., Jin, J. & Tam, P. H. Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. *FEBS Lett.* **580**, 1853–1858 (2006).

130. Shah, N. H., Dann, G. P., Vila-Perelló, M., Liu, Z. & Muir, T. W. Ultrafast protein splicing is common among cyanobacterial split inteins: Implications for protein engineering. *J. Am. Chem. Soc.* **134**, 11338–11341 (2012).

131. Shah, N. H., Eryilmaz, E., Cowburn, D. & Muir, T. W. Extein residues play an intimate role in the rate-limiting step of protein trans -splicing. *J. Am. Chem. Soc.* **135**, 5839–5847 (2013).

132. Tavassoli, A. SICLOPPS cyclic peptide libraries in drug discovery. *Curr. Opin. Chem. Biol.* **38**, 30–35 (2017).

133. Townend, J. E. & Tavassoli, A. Traceless Production of Cyclic Peptide Libraries in E. coli. *ACS Chem. Biol.* **11**, 1624–1630 (2016).

134. Scott, C. P., Abel-Santos, E., Wall, M., Wahnon, D. C. & Benkovic, S. J. Production of cyclic peptides and proteins in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13638–13643 (1999).

135. Miyafusa, T. *et al.* Backbone Circularization Coupled with Optimization of Connecting Segment in Effectively Improving the Stability of Granulocyte-Colony Stimulating Factor. *ACS Chem. Biol.* **12**, 2690–2696 (2017).

136. Touati, J., Angelini, A., Hinner, M. J. & Heinis, C. Enzymatic cyclisation of peptides with a transglutaminase. *ChemBioChem* **12**, 38–42 (2011).

137. Hae, J. K., Coulibaly, F., Clow, F., Proft, T. & Baker, E. N. Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. *Science (1979)* **318**, 1625–1628 (2007).

138. Zakeri, B. & Howarth, M. Spontaneous intermolecular amide bond formation between side chains for irreversible peptide targeting. *J. Am. Chem. Soc.* **132**, 4526–4527 (2010).

139. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E690–E697 (2012).

140. Hu, X. *et al.* Autocatalytic intramolecular isopeptide bond formation in Gram-positive bacterial pili: A QM/MM simulation. *J. Am. Chem. Soc.* **133**, 478–485 (2011).

141. Hagan, R. M. *et al.* NMR spectroscopic and theoretical analysis of a spontaneously formed lys-asp isopeptide bond. *Angew. Chem. Int. Ed.* **49**, 8421–8425 (2010).

142. Reddington, S. C. & Howarth, M. Secrets of a covalent interaction for biomaterials and biotechnology: SpyTag and SpyCatcher. *Curr. Opin. Chem. Biol.* **29**, 94–99 (2015).

143. Schoene, C., Fierer, J. O., Bennett, S. P. & Howarth, M. SpyTag/Spycatcher cyclization confers resilience to boiling on a mesophilic enzyme. *Angew. Chem. Int. Ed.* **53**, 6101–6104 (2014).

144. Si, M., Xu, Q., Jiang, L. & Huang, H. Spytag/spycatcher cyclization enhances the thermostability of firefly luciferase. *PLoS One* **11**, e016231 (2016).

145. Li, L., Fierer, J. O., Rapoport, T. A. & Howarth, M. Structural analysis and optimization of the covalent association between SpyCatcher and a peptide tag. *J. Mol. Biol.* **426**, 309–317 (2014).

146. Zhang, X.-J. *et al.* SpyCatcher-N TEV : A Circularly Permuted, Disordered SpyCatcher Variant for Less Trace Ligation. *Bioconjugate Chem.* **29**, 1622–1629 (2018).

147. Veggiani, G. *et al.* Programmable polyproteams built using twin peptide superglues. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1202–1207 (2016).

148. Krishna, M. M. G. & Englander, S. W. The N-terminal to C-terminal motif in protein folding and function. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 1053–1058 (2005).

149. Trabi, M. & Craik, D. J. Circular proteins - No end in sight. *Trends Biochem. Sci.* **27**, 132–138 (2002).

150. Shibuya, R., Miyafusa, T. & Honda, S. Stabilization of backbone-circularized protein is attained by synergistic gains in enthalpy of folded structure and entropy of unfolded structure. *FEBS J.* **287**, 1554–1575 (2019).

151. Miyafusa, T., Shibuya, R. & Honda, S. Structural insights into the backbone-circularized granulocyte colony-stimulating factor containing a short connector. *Biochem. Biophys. Res. Commun.* **500**, 224–228 (2018).

152. Kanno, A., Yamanaka, Y., Hirano, H., Umezawa, Y. & Ozawa, T. Cyclic Luciferase for Real-Time Sensing of Caspase-3 Activities in Living Mammals. *Angewandte Chemie - International Edition* **46**, 7595–7599 (2007).

153. Dien, V. T., Morris, S. E., Karadeema, R. J. & Romesberg, F. E. Expansion of the Genetic Code via Expansion of the Genetic Alphabet. *Curr. Opin. Chem. Biol.* **46**, 196–202 (2018).

154. Wegmuller, S. & Schmid, S. Recombinant Peptide Production in Microbial Cells. *Curr. Org. Chem.* **18**, 1005 —1019 (2014).

155. Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **1**, 2527–2535 (2007).

156. Iwakura, M. & Honda, S. Stability and reversibility of thermal denaturation are greatly improved by limiting terminal flexibility of escherichia coli dihydrofolate reductase. *J. Biochem.* **119**, 414–420 (1996).

157. Xu, Q. *et al.* Catcher/Tag cyclization introduces electrostatic interaction mediated protein-protein interactions to enhance the thermostability of luciferase. *Process Biochem.* **80**, 64–71 (2019).

158. Schoene, C., Bennett, S. P. & Howarth, M. SpyRing interrogation: Analyzing how enzyme resilience can be achieved with phytase and distinct cyclization chemistries. *Sci. Rep.* **6**, 21151 (2016).

159. Iwai, H., Lingel, A. & Plückthun, A. Cyclic Green Fluorescent Protein Produced in Vivo Using an Artificially Split PI-PfuI Intein from Pyrococcus furiosus. *J. Biol. Chem.* **276**, 16548–16554 (2001).

160. Wang, J. *et al.* Enhanced thermal stability of lichenase from Bacillus subtilis 168 by SpyTag/SpyCatcher-mediated spontaneous cyclization. *Biotechnol. Biofuels* **9**, (2016).

161. Gao, D. Y. *et al.* Characterization of Thermostable and Chimeric Enzymes via Isopeptide Bond-Mediated Molecular Cyclization. *J. Agric. Food Chem.* **67**, 6837–6846 (2019).

162. Park, C. & Marqusee, S. Probing the High Energy States in Proteins by Proteolysis. *J. Mol. Biol.* **343**, 1467–1476 (2004).

163. Iwai, H. & Plückthun, A. Circular β-lactamase: Stability enhancement by cyclizing the backbone. *FEBS Lett.* **459**, 166–172 (1999).

164. Satoh, T. *et al.* Synthetic Peptides Derived from the Fourth Domain of CD4 Antagonize CD4 Function and Inhibit T Cell Activation. *Biochem. Biophys. Res. Commun.* **224**, 438–443 (1996).

165. Goldenberg, D. P. & Creighton, T. E. Folding pathway of a circular form of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **179**, 527–545 (1984).

166. Schumann, F. H. *et al.* Changing the topology of protein backbone: The effect of backbone cyclization on the structure and dynamics of a SH3 domain. *Front. Chem.* **3**, (2015).

167. Dill, K. A. & MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science (1979)* **338**, 1042–1046 (2012).

168. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).

169. Koga, N. *et al.* Principles for designing ideal protein structures. *Nature* **491**, 222–229 (2012).

170. Nepomnyachiy, S., Ben-Tal, N. & Kolodny, R. Complex evolutionary footprints revealed in an analysis of reused protein segments of diverse lengths. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 11703–11708 (2017).

171. Taylor, W. R., Chelliah, V., Hollup, S. M., MacDonald, J. T. & Jonassen, I. Probing the 'Dark Matter' of Protein Fold Space. *Structure* **17**, 1244–1252 (2009).

172. Lapenta, F. & Jerala, R. Design of novel protein building modules and modular architectures. *Curr. Opin. Struct. Biol.* **63**, 90–96 (2020).

173. Dorn, M., E Silva, M. B., Buriol, L. S. & Lamb, L. C. Three-dimensional protein structure prediction: Methods and computational strategies. *Comput. Biol. Chem.* **53**, 251–276 (2014).

174. Woolfson, D. N. *et al.* De novo protein design: How do we expand into the universe of possible protein structures? *Curr. Opin. Struct. Biol.* **33**, 16–26 (2015).

175. Odgren, P. R., Harvie, L. W. & Fey, E. G. Phylogenetic occurrence of coiled coil proteins: Implications for tissue structure in metazoa via a coiled coil tissue matrix. *Proteins: Struct. Funct. Genet.* **24**, 467–484 (1996).

176. Gáspári, Z. & Nyitray, L. Coiled coils as possible models of protein structure evolution. *Biomol. Concepts* **2**, 199–210 (2011).

177. McKittrick, J. *et al.* The structure, functions, and mechanical properties of keratin. *J. Mater.* **64**, 449–468 (2012).

178. Bader, A. G. & Vogt, P. K. Leucine Zipper Transcription Factors: bZIP Proteins. in *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine* (eds. Ganten, D. et al.) 104–204 (Springer, Berlin, Heidelberg, 2006).

179. Crick, F. H. C. The packing of α-helices: simple coiled-coils. *Acta Crystallogr.* **6**, 689–697 (1953).

180. Lupas, A. N. & Gruber, M. The structure of α-helical coiled coils. *Adv. Protein Chem.* **70**, 37–78 (2005).

181. Rink, W. M. & Thomas, F. De Novo Designed α-Helical Coiled-Coil Peptides as Scaffolds for Chemical Reactions. *Chem. Eur. J.* **25**, 1665–1677 (2019).

182. Drobnak, I., Gradišar, H., Ljubetič, A., Merljak, E. & Jerala, R. Modulation of Coiled-Coil Dimer Stability through Surface Residues while Preserving Pairing Specificity. *J. Am. Chem. Soc.* **139**, 8229–8236 (2017).

183. Lapenta, F., Aupič, J., Strmšek, Ž. & Jerala, R. Coiled coil protein origami: From modular design principles towards biotechnological applications. *Chem. Soc. Rev.* **47**, 3530–3542 (2018).

184. Saccà, B. & Niemeyer, C. M. DNA origami: The art of folding DNA. *Angew. Chem. Int. Ed.* **51**, 58–66 (2012).

185. Hill, A. C. & Hall, J. High-order structures from nucleic acids for biomedical applications. *Mater. Chem. Front.* **4**, 1074–1088 (2020).

186. Crooks, R. O., Lathbridge, A., Panek, A. S. & Mason, J. M. Computational Prediction and Design for Creating Iteratively Larger Heterospecific Coiled Coil Sets. *Biochemistry* **56**, 1573–1584 (2017).

187. Small, L. S. R. *et al.* Construction of a Chassis for a Tripartite Protein-Based Molecular Motor. *ACS Synth. Biol.* **6**, 1096–1102 (2017).

188. Kočar, V. *et al.* TOPOFOLD, the designed modular biomolecular folds: Polypeptide-based molecular origami nanostructures following the footsteps of DNA. *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.* **7**, 218–237 (2015).

189. Gradišar, H. *et al.* Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat. Chem. Biol.* **9**, 362–366 (2013).

190. Ljubetič, A. *et al.* Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. *Nat. Biotechnol.* **35**, 1094–1101 (2017).

191. Aupič, J. *et al.* Designed folding pathway of modular coiled-coil based proteins. *Nat. Commun.* **12**, 940 (2021).

192. Plaxco, K. W., Simons, K. T. & Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994 (1998).

193. Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**, 2570–2577 (2015).

194. Aumiller, W. M., Uchida, M. & Douglas, T. Protein cage assembly across multiple length scales. *Chem. Soc. Rev.* **47**, 3383–3760 (2018).

195. Zhang, N., Mei, K., Guan, P., Hu, X. & Zhao, Y. Protein-Based Artificial Nanosystems in Cancer Therapy. *Small* **16**, 1907256 (2020).

196. Zhang, Y., Ardejani, M. S. & Orner, B. P. Design and Applications of Protein-Cage-Based Nanomaterials. *Asian J. Chem.* **11**, 2814–2828 (2016).

197. Rother, M., Nussbaumer, M. G., Renggli, K. & Bruns, N. Protein cages and synthetic polymers: A fruitful symbiosis for drug delivery applications, bionanotechnology and materials science. *Chem. Soc. Rev.* **45**, 6213–6249 (2016).

198. Renggli, K. *et al.* Selective and responsive nanoreactors. *Adv. Funct. Mater.* **21**, 1241–1259 (2011).

199. Yoshizawa, M., Klosterman, J. K. & Fujita, M. Functional molecular flasks: new properties and reactions within discrete, self-assembled hosts. *Angew. Chem. Int. Ed. Engl.* **48**, 3418–3438 (2009).

200. Fang, Y. *et al.* Catalytic reactions within the cavity of coordination cages. *Chem. Soc. Rev.* **48**, 4707–4730 (2019).

201. Cheriyan, M., Pedamallu, C. S., Tori, K. & Perler, F. Faster protein splicing with the nostoc punctiforme DnaE intein using non-native extein residues. *J. Biol. Chem.* **288**, 6202–6211 (2013).

202. Stevens, A. J. *et al.* A promiscuous split intein with expanded protein engineering applications. *Proc. Natl. Acad. Sci.* **114**, 8538–8543 (2017).

203. Liu, Q., Xun, G. & Feng, Y. The state-of-the-art strategies of protein engineering for enzyme stabilization. *Biotechnol. Adv.* **37**, 530–537 (2019).

204. Park, C. & Marqusee, S. Probing the High Energy States in Proteins by Proteolysis. *J. Mol. Biol.* **343**, 1467–1476 (2004).

205. Yu, X. *et al.* High-level expression and characterization of carboxypeptidase Y from Saccharomyces cerevisiae in Pichia pastoris GS115. *Biotechnol. Lett.* **37**, (2015).

206. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751**, 119–139 (2005).

207. Zbacnik, T. J. *et al.* Role of Buffers in Protein Formulations. *J. Pharma. Sci.* **106**, 713–733 (2017).

208. Rani, A. & Venkatesu, P. Changing relations between proteins and osmolytes: a choice of nature. *Phys. Chem. Chem. Phys.* **20**, 20315--20333 (2018).

209. Shank, E. A., Cecconi, C., Dill, J. W., Marqusee, S. & Bustamante, C. The folding cooperativity of a protein is controlled by its chain topology. *Nature* **465**, 637–641 (2010).

210. Best, R. B., Hummer, G. & Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci.* **110**, 17874–17879 (2013).

211. Jones, D. C., Mistry, I. N. & Tavassoli, A. Post-translational control of protein function with light using a LOV-intein fusion protein. *Mol. BioSyst.* **12**, (2016).

212. Ren, W., Ji, A. & Ai, H. Light Activation of Protein Splicing with a Photocaged Fast Intein. *J. Am. Chem. Soc.* **137**, 2155–2158 (2015).

213. Alford, S. C., O'Sullivan, C., Obst, J., Christie, J. & Howard, P. L. Conditional protein splicing of a-sarcin in live cells. *Mol Biosyst* **10**, 831–837 (2014).

214. Gramespacher, J. A., Stevens, A. J., Nguyen, Duy, P., Chin, J. W. & Muir, T. W. Intein Zymogens: Conditional Assembly and Splicing of Split Inteins via Targeted Proteolysis. *J. Am. Chem. Soc.* **139**, 8074–8077 (2017).

215. di Ventura, B. & Mootz, H. D. Switchable inteins for conditional protein splicing. *Biol. Chem.* **400**, 467–475 (2019).

216. Wu, W.-H., Wei, J. & Zhang, W.-B. Controlling SpyTag/SpyCatcher Reactivity via Redox-Gated Conformational Restriction. *ACS Macro Lett.* **7**, 1388–1393 (2018).

217. Lapenta, F. *et al.* Self-assembly and regulation of protein cages from pre-organised coiled-coil modules. *Nat. Commun.* **12**, 939 (2021).

218. Napper, I. E. *et al.* Reaching New Heights in Plastic Pollution— Preliminary Findings of Microplastics on Mount Everest. *One Earth* **3**, 621–630 (2020).

219. Peeken, I. *et al.* Arctic sea ice is an important temporal sink and means of transport for microplastic. *Nat. Commun.* **9**, 1505 (2018).

220. Weston, J. J., Carrillo-Barragan, P., Linley, T. D., Reid, W. D. K. & Jamieson, A. J. New species of Eurythenes from hadal depths of the Mariana Trench, Pacific Ocean (Crustacea: Amphipoda). *Zootaxa* **4748**, 163–181 (2020).

221. Leslie, H. A. *et al.* Discovery and quantification of plastic particle pollution in human blood. *Environ. Int.* **163**, 107199 (2022).

222. Zhang, K. *et al.* Understanding plastic degradation and microplastic formation in the environment: A review. *Environ. Pollut.* **274**, 116554 (2021).

223. Thomsen, T. B., Hunt, C. J. & Meyer, A. S. Influence of substrate crystallinity and glass transition temperature on enzymatic degradation of polyethylene terephthalate (PET). *N Biotechnol* **69**, 28–35 (2022).

224. Dutt, K. & Soni, R. K. A review on synthesis of value added products from polyethylene terephthalate (PET) waste. *Polymer Science Series B* **55**, 430–452 (2013).

225. Plastics – the Facts 2020. *Plastics Europe* https://plasticseurope.org/wp-content/uploads/2021/09/Plastics_the_facts-WEB-2020_versionJun21_final.pdf.

226. Raheem, A. B. *et al.* Current developments in chemical recycling of post-consumer polyethylene terephthalate wastes for new materials production: A review. *J. Clean. Prod.* **225**, 1052–1064 (2019).

227. Malik, N., Kumar, P., Shrivastava, S. & Ghosh, S. B. An overview on PET waste recycling for application in packaging. *Int. J. Plast. Technol.* **21**, 1–24 (2017).

228. Shojaei, B., Abtahi, M. & Najafi, M. Chemical recycling of PET: A stepping-stone towardsustainability. *Polym. Adv.Technol.* **31**, 2912–2938 (2020).

229. Danso, D., Chow, J. & Streita, W. R. Plastics: Environmental and biotechnological perspectives on microbial degradation. *Appl. Environ. Microbiol.* **85**, 19 (2019).

230. Wei, R. & Zimmermann, W. Microbial enzymes for the recycling of recalcitrant petroleum-based plastics: how far are we? *Microb. Biotechnol.* **10**, 1308–1322 (2017).

231. Ali, S. S. *et al.* Plastic wastes biodegradation: Mechanisms, challenges and future prospects. *Sci.Total Environ.* **780**, 1465902 (2021).

232. Kawai, F., Kawabata, T. & Oda, M. Current State and Perspectives Related to the Polyethylene Terephthalate Hydrolases Available for Biorecycling. *ACS Sustain. Chem. Eng.* **8**, 8894–8908 (2020).

233. Kawai, F., Kawabata, T. & Oda, M. Current knowledge on enzymatic PET degradation and its possible application to waste stream management and other fields. *Appl. Microbiol. Biotechnol.* **103**, 4253–4268 (2019).

234. Carniel, A., de Abrue Waldow, V. & Machado de Castro, A. A comprehensive and critical review on key elements to implement enzymatic PET depolymerization for recycling purposes. *Biotechnol. Adv.* **52**, 107811 (2021).

235. Alves, N. M., Mano, J. F., Balaguer, E., Meseguer Dueñas, J. M. & Gómez Ribelles, J. L. Glass transition and structural relaxation in semi-crystalline poly(ethylene terephthalate): A DSC study. *Polymer (Guildf)* **43**, 4111–4122 (2002).

236. Wei, R. *et al.* Engineered bacterial polyester hydrolases efficiently degrade polyethylene terephthalate due to relieved product inhibition. *Biotechnol. Bioeng.* **113**, 1658–1665 (2016).

237. Sulaiman, S., You, D. J., Kanaya, E., Koga, Y. & Kanaya, S. Crystal structure and thermodynamic and kinetic stability of metagenome-derived LC-cutinase. *Biochemistry* **53**, 1858–1869 (2014).

238. Sulaiman, S. *et al.* Isolation of a novel cutinase homolog with polyethylene terephthalate-degrading activity from leaf-branch compost by using a metagenomic approach. *Appl. Environ. Microbiol.* **78**, 1556–1562 (2012).

239. Tournier, V. *et al.* An engineered PET depolymerase to break down and recycle plastic bottles. *Nature* **580**, 216–219 (2020).

240. Ronkvist, Å. M., Xie, W., Lu, W. & Gross, R. A. Cutinase-catalyzed hydrolysis of poly(ethylene terephthalate). *Macromolecules* **42**, 5128–5138 (2009).

241. Castro, A. M. d. *et al.* High-fold improvement of assorted post-consumer poly(ethylene terephthalate) (PET) packages hydrolysis using Humicola insolens cutinase as a single biocatalyst. *Process Biochem.* **81**, 85–91 (2019).

242. Quartinello, F. *et al.* Synergistic chemo-enzymatic hydrolysis of poly (ethylene terephthalate) from textile waste. *Microb. Biotechnol.* **10**, 1376–1383 (2017).

243. Ellis, L. D. *et al.* Chemical and biological catalysis for plastics recycling and upcycling. *Nat Catal* **4**, 539–556 (2021).

244. Magalhães, R. P., Cunha, J. M. & Sousa, S. F. Perspectives on the Role of Enzymatic Biocatalysis for the Degradation of Plastic PET. *Int J Mol Sci* **22**, 11257 (2021).

245. Yoshida, S. *et al.* A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science (1979)* **351**, 1196–1199 (2016).

246. Han, X. *et al.* Structural insight into catalytic mechanism of PET hydrolase. *Nat. Commun.* **8**, 2106 (2017).

247. Joo, S. *et al.* Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. *Nat. Commun.* **9**, 382 (2018).

248. Silva, C. *et al.* Engineered Thermobifida fusca cutinase with increased activity on polyester substrates. *Biotechnol. J.* **6**, 1230–9 (2011).

249. Roth, C. *et al.* Structural and functional studies on a thermostable polyethylene terephthalate degrading hydrolase from Thermobifida fusca. *Appl. Microbiol. Biotechnol.* **98**, 7815–7823 (2014).

250. Wei, R. *et al.* Conformational fitting of a flexible oligomeric substrate does not explain the enzymatic PET degradation. *Nat.Commun.* **10**, 5581 (2019).

251. Seo, H. *et al.* Reply to “Conformational fitting of a flexible oligomeric substrate does not explain the enzymatic PET degradation”. *Nat. Commun.* **10**, 5582 (2019).

252. Austin, H. P. *et al.* Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4350–E4357 (2018).

253. Son, H. F. *et al.* Rational Protein Engineering of Thermo-Stable PETase from Ideonella sakaiensis for Highly Efficient PET Degradation. *ACS Catal.* **9**, 3519–3526 (2019).

254. Son, H. F. *et al.* Structural bioinformatics-based protein engineering of thermo-stable PETase from Ideonella sakaiensis. *Enzyme Microb. Technol.* **141**, 109656 (2020).

255. Zhong-Johnson, E. Z. L., Voigt, C. A. & Sinskey, A. J. An absorbance method for analysis of enzymatic degradation kinetics of poly(ethylene terephthalate) films. *Sci. Rep.* **11**, 928 (2021).

256. Cui, Y. *et al.* Computational Redesign of a PETase for Plastic Biodegradation under Ambient Condition by the GRAPE Strategy. *ACS Catal.* **11**, 1340–1350 (2021).

257. Chen, K., Hu, Y., Dong, X. & Sun, Y. Molecular Insights into the Enhanced Performance of EKylated PETase Toward PET Degradation. *ACS Catal.* **11**, 7358–7370 (2021).

258. Chen, K., Quan, M., Dong, X., Shi, Q. & Sun, Y. Low modification of PETase enhances its activity toward degrading PET: Effect of conjugate monomer property. *Biochem. Eng. J.* **175**, 108151 (2021).

259. Schwaminger, S. *et al.* Immobilization of PETase enzymes on magnetic iron oxide nanoparticles for the decomposition of microplastic PET. *Nanoscale Adv.* **3**, 4395–4399 (2021).

260. Jia, Y. *et al.* Nano-immobilization of PETase enzyme for enhanced polyethylene terephthalate biodegradation. *Biochem. Eng. J.* **176**, 108205 (2021).

261. Chen, Z. *et al.* Efficient biodegradation of highly crystallized polyethylene terephthalatethrough cell surface display of bacterial PETase. *Sci. Total Environ.* **709**, 136138 (2020).

262. Lobstein, J. *et al.* SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Factories* **11**, 56 (2012).

263. Wikoff, W. R. *et al.* Topologically Linked Protein Rings in the Bacteriophage HK97 Capsid. *Science (1979)* **289**, 2129–2133 (2000).

264. Boutz, D. R., Cascio, D., Whitelegge, J., Perry, L. J. & Yeates, T. O. Discovery of a Thermophilic Protein Complex Stabilized by Topologically Interlinked Chains. *J. Mol. Biol.* **368**, 1332–1344 (2007).

265. Cao, Z., Roszak, A. W., Gourlay, L. J., Lindsay, J. G. & Isaacs, N. W. Bovine Mitochondrial Peroxiredoxin III Forms a Two-Ring Catenane. *Structure* **13**, 1661–1664 (2005).

266. Zimanyi, C. M. *et al.* Tangled Up in Knots: Structures of Inactivated Forms of E. coli Class Ia Ribonucleotide Reductase. *Structure* **20**, 1374–1383 (2012).

267. Yan, L. Z. & Dawson, P. E. Design and synthesis of a protein catenane. *Angew. Chem. Int. Ed.* **40**, 3625–3627 (2001).

268. Wang, X.-W. & Zhang, W.-B. Cellular Synthesis of Protein Catenanes. *Angew. Chem. Int. Ed.* **55**, 3442–3446 (2016).

269. Wang, X.-W. & Zhang, W.-B. Protein Catenation Enhances Both the Stability and Activity of Folded Structural Domains. *Angew. Chem. Int. Ed.* **56**, 13985–13989 (2017).

270. Zheng, Q. *et al.* Topology engineering via protein catenane construction to strengthen an industrial biocatalyst. *J. Biotech.* **325**, 271–279 (2021).

271. Blankenship, J. W. & Dawson, P. E. Thermodynamics of a Designed Protein Catenane. *J. Mol. Biol.* **327**, 537–548 (2003).

272. Zhou, H.-X. Loops, Linkages, Rings, Catenanes, Cages, and Crowders: Entropy-Based Strategies for Stabilizing Proteins. *Acc. Chem. Res.* **37**, 123–130 (2004).

273. Davison, T. S. *et al.* Structure and Functionality of a Designed p53 Dimer. *J. Mol. Biol.* **307**, 605–617 (2001).

274. Olsen, J. v., Ong, S.-E. & Mann, M. Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. *Mol. Cell. Proteom.* **3**, 608–614 (2004).

275. Colón, W. *et al.* Biological Roles of Protein Kinetic Stability. *Biochemistry* **56**, 6179–6186 (2017).

276. Kang, H. J. & Baker, E. N. Intramolecular Isopeptide Bonds Give Thermodynamic and Proteolytic Stability to the Major Pilin Protein of Streptococcus pyogenes. *J. Biol. Chem.* **284**, 20729 –20737 (2009).

277. Walden, M., Crow, A., Nelson, M. D. & Banfield, M. J. Intramolecular isopeptide but not internal thioester bonds confer proteolytic and significant thermal stability to the S. pyogenes pilus adhesin Spy0125. *Proteins: Struct. Funct. Genet.* **82**, 517–527 (2014).

278. Sanchez-Ruiz, J. M. Protein kinetic stability. *Biophys. Chem.* **148**, 1–15 (2010).

279. Manning, M. & Colón, W. Structural Basis of Protein Kinetic Stability: Resistance to Sodium Dodecyl Sulfate Suggests a Central Role for Rigidity and a Bias Toward beta-Sheet Structure. *Biochemistry* **43**, 11248–11254 (2004).

280. Xia, K. *et al.* Quantifying the Kinetic Stability of Hyperstable Proteins via Time Dependent SDS Trapping. *Biochemistry* **51**, 100–107 (2012).

281. Syedd-Leon, R., Sandoval-Barrantes, M., Trimino-Vasquez, H., Villegas-Penaranda, L. R. & Rodriguez-Rodriguez, G. Revisiting the fundamentals of p-nitrophenol analysis for its application in the quantification of lipases activity. A graphical update. *Uniciencia* **34**, 31–43 (2020).

282. Fecker, T. *et al.* Active Site Flexibility as a Hallmark for Efficient PET Degradation by I. sakaiensis PETase. *Biophys. J.* **114**, 1302–1312 (2018).

283. Peng, Y., Fu, S., Liu, H. & Lucia, L. A. Accurately determining esterase activity via the isobetic point of p-nitrophenol. *Bioresources* **11**, 10099–10111 (2016).

284. Ghadessy, F. J. & Sana, B. Thermostable PETase Enzyme. Patent WO/2021/145822 (2021).

285. Xu, C., Xu, Q., Huang, H. & Jiang, L. Enhancing the stability of trehalose synthase via spytag/spycatcher cyclization to improve its performance in industrial biocatalysts. *Biosci Biotechnol Biochem* **82**, 1473–1479 (2018).

286. Waldhauer, M. C. *et al.* Backbone circularization of Bacillus subtilis family 11 xylanase increases its thermostability and its resistance against aggregation. *Mol Biosyst* **11**, 3231–3243 (2015).

287. Keeble, A. H. *et al.* Approaching infinite affinity through engineering of peptide–protein interaction. *Proc. Natl. Acad. Sci.* **116**, 26523–26533 (2019).

288. Shibuya, R., Miyafusa, T., Imamura, H., Ooishi, A. & Honda, S. Effect of backbone circularization on colloidal stability: Compaction of unfolded structures improves aggregation resistance of granulocyte colony-stimulating factor. *Int. J. Pharm.* **605**, 120774 (2021).

289. Arakawa, T., Ejima, D., Li, T. & Philo, J. S. The critical role of mobile phase composition in size exclusion chromatography of protein pharmaceuticals. *J. Pharm. Sci.* **99**, 1674–1692 (2010).

290. Tantipolphan, R. *et al.* Elution behavior of insulin on high-performance size exclusion chromatography at neutral pH. *J. Pharma. Biomed. Anal.* **52**, 195–202 (2010).

291. Cabra, V., Vázquez-Contreras, E., Moreno, A. & Arreguin-Espinosa, R. The effect of sulfhydryl groups and disulphide linkage in the thermal aggregation of Z19 α-zein. *Biochim. Biophys. Acta* **1784**, 1028–1036 (2008).

292. Pirillo, V., Pollegioni, L. & Molla, G. Analytical methods for the investigation of enzyme catalyzed degradation of polyethylene terephthalate. *FEBS J.* **288**, 4730–4745 (2021).

293. Arnling Bååth, J., Borch, K. & Westh, P. A suspension-based assay and comparative detection methods for characterization of polyethylene terephthalate hydrolases. *Anal. Biochem.* **607**, 113873 (2020).

294. Heyde, S. A. H., Arnling Bååth, J., Westh, P., Nørholm, M. H. H. & Jensen, K. Surface display as a functional screening platform for detecting enzymes active on PET. *Microb. Cell Factories* **20**, 93 (2021).

295. Bai, G., Bee, J. S., Biddlecombe, J. G., Chen, Q. & Leach, W. T. Computational fluid dynamics (CFD) insights into agitation stress methods in biopharmaceutical development. *Int. J. Pharm.* **423**, 264–280 (2012).

296. Jayaraman, M., Buck, P. M., Ignatius, A. A., King, K. R. & Wang, W. Agitation-induced aggregation and subvisible particulate formation in model proteins. *Eur. J. Pharm. Biopharm.* **87**, 299–309 (2014).

297. Torisu, T., Maruno, T., Hamaji, Y., Ohkubo, T. & Uchiyama, S. Synergistic Effect of Cavitation and Agitation on Protein Aggregation. *J. Pharm. Sci.* **106**, 521–529 (2017).

298. Erickson, E. *et al.* Comparative Performance of PETase as a Function of Reaction Conditions, Substrate Properties, and Product Accumulation. *ChemSusChem2* **15**, e2021019 (2022).

299. Seelig, J. & Schönfeld, H.-J. Thermal protein unfolding by differential scanning calorimetry and circular dichroism spectroscopy Two-state model versus sequential unfolding. *Q. Rev. Biophys.* **49**, e9 (2016).

300. Jäger, V. D. *et al.* Catalytically-active inclusion bodies for biotechnology—general concepts, optimization, and application. *Appl. Microbiol. Biotechnol.* **104**, 7313–7329 (2020).

301. Alam, P., Siddiqi, K., Chturvedi, S. K. & Khan, R. H. Protein aggregation: From background to inhibition strategies. *Int. J. Biol. Macromol.* **103**, 208–219 (2017).

302. Wang, W. & Roberts, C. J. Protein aggregation – Mechanisms, detection, and control. *Int. J. Pharm.* **550**, 251–268 (2018).

303. Wang, W., Nema, S. & Teagarden, D. Protein aggregation—Pathways and influencing factors. *Int. J. Pharm.* **390**, 89–99 (2010).

304. Sun, X. B. *et al.* SpyTag/SpyCatcher molecular cyclization confers protein stability and resilience to aggregation. *N. Biotechnol.* **49**, 28–36 (2019).

305. Chi, E. Y. *et al.* Roles of conformational stability and colloidal stability in the aggregation of recombinant human granulocyte colony-stimulating factor. *Protein Sci.* **12**, 903–912 (2003).

306. Al-Ghobashy, M. A., Mostafa, M. M., Abed, H. S., Fathalla, F. A. & .Y., Salem. M. Correlation between Dynamic Light Scattering and Size Exclusion High Performance Liquid Chromatography for Monitoring the Effect of pH on Stability of Biopharmaceuticals. *J. Chromatogr. B* **1060**, 1–9 (2017).

307. Stetefeld, J., McKenna, S. A. & Patel, P. R. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys. Rev.* **8**, 409–427 (2016).

308. Panyukov, Y., Yudin, I., Drachev, V., Dobrov, E. & Kurganov, B. The study of amorphous aggregation of tobacco mosaic virus coat protein by dynamic light scattering. *Biophys. Chem.* **127**, 9–18 (2007).

309. Edwards, G. B., Muthurajan, U. M., Bowerman, S. & Luger, K. Analytical Ultracentrifugation (AUC): An Overview of the Application ofvFluorescence and Absorbance AUC to the Study of Biological Macromolecules. *Curr. Protoc. Mol. Biol.* **133**, e131 (2020).

310. Vivian, J. T. & Callis, P. R. Mechanisms of Tryptophan Fluorescence Shifts in Proteins. *Biophys. J.* **80**, 2093–2109 (2001).

311. Ghisaidoobe, A. B. T. & Chung, S. J. Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Förster Resonance Energy Transfer Techniques. *Int. J. Mol. Sci.* **15**, 22518–22538 (2014).

312. Shirdel, S. A. & Khalifeh, K. Thermodynamics of protein folding: methodology, data analysis and interpretation of data. *Eur. Biophys. J.* **48**, 305–316 (2019).

313. Danko, S. *et al.* ADP-insensitive phosphoenzyme intermediate of sarcoplasmic reticulum Ca2+-ATPase has a compact conformation resistant to proteinase K, V8 protease and trypsin. *FEBS Lett.* **489**, 277–282 (2001).

314. Riahi-Madvar, A. & Hosseinkhani, S. Design and characterization of novel trypsin-resistant firefly luciferases by site-directed mutagenesis. *Protein Eng. Des. Sel.* **22**, 655–663 (2009).

315. Lin, Y. *et al.* Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins. *Anal. Biochem.* **377**, 259–266 (2008).

316. Hansen, J. H. *et al.* Stable Intermediates Determine Proteins' Primary Unfolding Sites in the Presence of Surfactants. *Biopolymers* **51**, 221–231 (2009).

317. Thibeault, J. *et al.* Sarkosyl: A milder detergent than SDS for identifying proteins with moderately high hyperstability using gel electrophoresis. *Anal. Biochem.* **571**, 21–24 (2019).

318. Chi, E. Y., Krishnan, S., Randolf, T. W. & Carpenter, E. J. Physical Stability of Proteins in Aqueous Solution: Mechanism and Driving Forces in Nonnative Protein Aggregation. *Pharm. Res.* **20**, 1325–1336 (2003).

319. Wang, J. *et al.* Enhanced thermal stability of lichenase from Bacillus subtilis 168 by SpyTag/SpyCatcher-mediated spontaneous cyclization. *Biotechnol Biofuels* **9**, (2016).

320. Maurya, A., Bhattacharya, A. & Khare, S. K. Enzymatic Remediation of Polyethylene Terephthalate (PET)–Based Polymers for Effective Management of Plastic Wastes: An Overview. *Front. Bioeng. Biotechnol.* **8**, 602325 (2020).

321. Otzen, D. Protein–surfactant interactions: A tale of many states. *Biochim. Biophys. Acta* **1814**, 562–591 (2011).

322. Khan, T. A., Mahler, H.-C. & Kishore, R. S. K. Key interactions of surfactants in therapeutic protein formulations: A review. *Eur. J. Pharm. Biopharm.* **97**, 60–67 (2015).

323. Hong, T. *et al.* Aggregation of hen egg white proteins with additives during agitation. *LWT - Food Sci. Technol.* **146**, 111378 (2021).

324. Abbas, S. A., Sharma, N., Patapoff, T. W. & Kalonia, D. S. Opposite Effects of Polyols on Antibody Aggregation: Thermal Versus Mechanical Stresses. *Pharm. Res.* **29**, 683–694 (2012).

325. Furukawa, M., Kawakami, N., Oda, K. & Miyamoto, K. Acceleration of Enzymatic Degradation of Poly(ethyleneterephthalate) by Surface Coating with Anionic Surfactants. *ChemSusChem* **11**, 4018–4025 (2018).

326. Furukawa, M., Kawakami, N., Tomizawa, A. & Miyamoto, K. Efficient Degradation of Poly(ethylene terephthalate) with Thermobifida fusca Cutinase Exhibiting Improved Catalytic Activity Generated using Mutagenesis and Additive-based Approaches. *Sci. Rep.* **9**, 16038 (2019).

CHAPTER 7:
APPENDICES

7.1 DNA SEQUENCES

7.1.1 TET12SN DNA Sequences

TET12SN (5’ - 3’)

ATGCTGGAAGAAGAACTGAAACAGCTGGAAGAGGAATTACAGGCAATTGAAGAACAGCTGGCACAG
CTGCAGTGGAAAGCACAGGCACGTAAAGAAAAACTGGCCCAGCTGAAAGAAAAATTAGGTAAAGGT
GATGGTAGTCCCGAAGATGAAATTCAGCAGCTTGAAGAAGAAATTAGCCAGCTGGAACAGAAGAAT
AGCGAACTGAAAGAGAAAAACCAAGAGCTGAAATATGGTAAAGGCGACGGTGATATTGAACAAGAA
CTGGAACGTGCAAAAGAAAGCATTCGTCGTCTGGAACAAGAAGTGAATCAAGAACGTAGCCGTATG
CAGTATCTGCAGACCCTGCTGGAAAAAGGCAAAGGTGACGGTCAACTGGAAGATAAAGTGGAAGAA
CTGCTGAGCAAAAACTATCACCTGGAAAATGAAGTTGAGCGCCTGAAAAAACTGGTTGGTGGCAAA
GGTGATGGCCTGGAAGAAGAGTTAAAACAGTTAGAGGAAGAACTTCAGGCCATCGAGGAACAGTTA
GCCCAGTTACAATGGAAAGCCCAAGCTAGAAAAGAGAAACTTGCGCAGTTAAAAGAAAAGCTTGGC
AAAGGCGACGGCAGTCCGGAAGATGAGATCCAACAGTTAGAAGAAAAAAACAGCCAGCTGAAACAA
GAGATTAGTCAGCTTGAGGAAAAGAATCAAGAGCTTAAGTATGGCGACGGTAAAGGCCAATTAGAG
GATAAAGTTGAGGAACTTCTGTCCAAGAATTACCATCTGGAAAACGAGGTGGAACGTCTGAAAAAG
TTAGTTGGTGGTGATGGAAAAGGTTCACCTGAAGATAAAATCTCCCAGCTTAAAGAGAAGATCCAG
CAACTGAAGCAAGAAAACCAGCAATTGGAGGAAGAAAATTCTCAGCTGGAATACGGTGACGGCAAA
GGTAGCCCTGAGGATGAAAATAGTCAATTAGAAGAGAAAATTTCGCAGCTGAAGCAAAAGAACTCC
GAACTTAAAGAAGAGATACAACAGCTTGAATATGGTGATGGTAAGGGTAGTCCAGAGGATAAAATC
AGTGAGCTGAAAGAGGAAAATCAGCAGTTGGAACAGAAAATTCAACAATTGAAAGAAGAAAACTCC
CAGTTAGAGTACGGTAAAGGTGACGGCGACATCGAGCAAGAATTAGAGCGTGCCAAAGAGTCAATC
CGTCGCTTAGAACAAGAGGTTAACCAAGAGCGTTCACGCATGCAGTATTTACAGACACTGTTAGAG
AAAGGCAAAGGTGATGGTTCTCCGGAAGATAAAAACTCGGAGTTGAAAGAAGAGATTCAGCAATTA
GAAGAAGAGAACCAACAACTGGAAGAGAAAATCTCCGAACTGAAATATCTGGAACACCACCACCAC
CACCACCACCACTGA

TET12-Int (5’ - 3’)

ATGATCAAGATTGCGACCCGTAAGTACCTGGGCAAACAGAACGTGTATGACATCGGTGTTGAACGT
GATCACAACTTCGCGCTGAAAAACGGCTTTATTGCGAGCAACTGCATGGGCTTAGGTAGCGGTCCG
CATATGCTGGAAGAAGAACTGAAACAGCTGGAAGAGGAATTACAGGCAATTGAAGAACAGCTGGCA
CAGCTGCAGTGGAAAGCACAGGCACGTAAAGAAAAACTGGCCCAGCTGAAAGAAAAATTAGGTAAA
GGTGATGGTAGTCCCGAAGATGAAATTCAGCAGCTTGAAGAAGAAATTAGCCAGCTGGAACAGAAG
AATAGCGAACTGAAAGAGAAAAACCAAGAGCTGAAATATGGTAAAGGCGACGGTGATATTGAACAA
GAACTGGAACGTGCAAAAGAAAGCATTCGTCGTCTGGAACAAGAAGTGAATCAAGAACGTAGCCGT
ATGCAGTATCTGCAGACCCTGCTGGAAAAAGGCAAAGGTGACGGTCAACTGGAAGATAAAGTGGAA

GAACTGCTGAGCAAAAACTATCACCTGGAAAATGAAGTTGAGCGCCTGAAAAAACTGGTTGGTGGC
AAAGGTGATGGCCTGGAAGAAGAGTTAAAACAGTTAGAGGAAGAACTTCAGGCCATCGAGGAACAG
TTAGCCCAGTTACAATGGAAAGCCCAAGCTAGAAAAGAGAAACTTGCGCAGTTAAAAGAAAAGCTT
GGCAAAGGCGACGGCAGTCCGGAAGATGAGATCCAACAGTTAGAAGAAAAAAACAGCCAGCTGAAA
CAAGAGATTAGTCAGCTTGAGGAAAAGAATCAAGAGCTTAAGTATGGCGACGGTAAAGGCCAATTA
GAGGATAAAGTTGAGGAACTTCTGTCCAAGAATTACCATCTGGAAAACGAGGTGGAACGTCTGAAA
AAGTTAGTTGGTGGTGATGGAAAAGGTTCACCTGAAGATAAAATCTCCCAGCTTAAAGAGAAGATC
CAGCAACTGAAGCAAGAAAACCAGCAATTGGAGGAAGAAAATTCTCAGCTGGAATACGGTGACGGC
AAAGGTAGCCCTGAGGATGAAAATAGTCAATTAGAAGAGAAAATTTCGCAGCTGAAGCAAAAGAAC
TCCGAACTTAAAGAAGAGATACAACAGCTTGAATATGGTGATGGTAAGGGTAGTCCAGAGGATAAA
ATCAGTGAGCTGAAAGAGGAAAATCAGCAGTTGGAACAGAAAATTCAACAATTGAAAGAAGAAAAC
TCCCAGTTAGAGTACGGTAAAGGTGACGGCGACATCGAGCAAGAATTAGAGCGTGCCAAAGAGTCA
ATCCGTCGCTTAGAACAAGAGGTTAACCAAGAGCGTTCACGCATGCAGTATTTACAGACACTGTTA
GAGAAAGGCAAAGGTGATGGTTCTCCGGAAGATAAAAACTCGGAGTTGAAAGAAGAGATTCAGCAA
TTAGAAGAAGAGAACCAACAACTGGAAGAGAAAATCTCCGAACTGAAATATCTGGAAGCTAGCGGT
CCTGGTAATGGTTTAGGTTCAGGTCCGCACCACCACCACCACCACGCGGCGTGCCTGAGCTACGAA
ACCGAGATTCTGACCGTGGAATATGGCCTGCTGCCGATCGGCAAGATTGTTGAGAAACGTATCGAA
TGCACCGTGTACAGCGTTGACAACAACGGCAACATTTATACCCAGCCGGTGGCGCAATGGCACGAT
CGTGGCGAACAAGAGGTTTTCGAGTACTGCCTGGAAGACGGTAGCCTGATCCGTGCGACCAAGGAC
CACAAATTCATGACCGTGGATGGTCAGATGCTGCCGATCGACGAGATTTTTGAACGTGAGCTGGAC
CTGATGCGTGTTGATAACCTGCCGAACTGA

TET12-Spy (5’ – 3’)

ATGGGTAAAACTATTAGTACATGGATTTCAGATGGACAAGTGAAAGATTTCTACCTGTATCCAGGA
AAATATACATTTGTCGAAACCGCAGCACCAGACGGTTATGAGGTAGCAACTGCTATTACCTTTACA
GTTAATGAGCAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGGCAGCGAAAACCTGTAC
TTCCAGGGTGGCAGCGGAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTGATGAGGACGGC
AAAGAGTTAGCTGGTGCAACTATGGAGTTGCGTGATTCAGTGGATGGCCATATGCTGGAAGAAGAA
CTGAAACAGCTGGAAGAGGAATTACAGGCAATTGAAGAACAGCTGGCACAGCTGCAGTGGAAAGCA
CAGGCACGTAAAGAAAAACTGGCCCAGCTGAAAGAAAAATTAGGTAAAGGTGATGGTAGTCCCGAA
GATGAAATTCAGCAGCTTGAAGAAGAAATTAGCCAGCTGGAACAGAAGAATAGCGAACTGAAAGAG
AAAAACCAAGAGCTGAAATATGGTAAAGGCGACGGTGATATTGAACAAGAACTGGAACGTGCAAAA
GAAAGCATTCGTCGTCTGGAACAAGAAGTGAATCAAGAACGTAGCCGTATGCAGTATCTGCAGACC
CTGCTGGAAAAAGGCAAAGGTGACGGTCAACTGGAAGATAAAGTGGAAGAACTGCTGAGCAAAAAC
TATCACCTGGAAAATGAAGTTGAGCGCCTGAAAAAACTGGTTGGTGGCAAAGGTGATGGCCTGGAA
GAAGAGTTAAAACAGTTAGAGGAAGAACTTCAGGCCATCGAGGAACAGTTAGCCCAGTTACAATGG
AAAGCCCAAGCTAGAAAAGAGAAACTTGCGCAGTTAAAAGAAAAGCTTGGCAAAGGCGACGGCAGT

CCGGAAGATGAGATCCAACAGTTAGAAGAAAAAAACAGCCAGCTGAAACAAGAGATTAGTCAGCTT
GAGGAAAAGAATCAAGAGCTTAAGTATGGCGACGGTAAAGGCCAATTAGAGGATAAAGTTGAGGAA
CTTCTGTCCAAGAATTACCATCTGGAAAACGAGGTGGAACGTCTGAAAAAGTTAGTTGGTGGTGAT
GGAAAAGGTTCACCTGAAGATAAAATCTCCCAGCTTAAAGAGAAGATCCAGCAACTGAAGCAAGAA
AACCAGCAATTGGAGGAAGAAAATTCTCAGCTGGAATACGGTGACGGCAAAGGTAGCCCTGAGGAT
GAAAATAGTCAATTAGAAGAGAAAATTTCGCAGCTGAAGCAAAAGAACTCCGAACTTAAAGAAGAG
ATACAACAGCTTGAATATGGTGATGGTAAGGGTAGTCCAGAGGATAAAATCAGTGAGCTGAAAGAG
GAAAATCAGCAGTTGGAACAGAAAATTCAACAATTGAAAGAAGAAAACTCCCAGTTAGAGTACGGT
AAAGGTGACGGCGACATCGAGCAAGAATTAGAGCGTGCCAAAGAGTCAATCCGTCGCTTAGAACAA
GAGGTTAACCAAGAGCGTTCACGCATGCAGTATTTACAGACACTGTTAGAGAAAGGCAAAGGTGAT
GGTTCTCCGGAAGATAAAAACTCGGAGTTGAAAGAAGAGATTCAGCAATTAGAAGAAGAGAACCAA
CAACTGGAAGAGAAAATCTCCGAACTGAAATATCTGGAAGCTAGCGGTCCTGGTAGTGGTGGCAGC
GGAGCCCACATCGTGATGGTGGACGCCTACAAGCCGACGAAGGGTTCAGGTCCGCACCATCACCAT
CACCATTAA

7.1.2 *Is*PETase DNA Sequences

*Is*PETase-WT (including signal peptide) (5’ – 3’)

ATGAACTTCCCCCGTGCCTCGCGCCTTATGCAGGCTGCTGTGCTGGGCGGCCTTATGGCCGTTTCC
GCAGCGGCCACCGCGCAGACCAATCCGTATGCGCGCGGCCCCAACCCTACCGCCGCCTCGTTGGAA
GCCAGCGCGGGACCCTTTACCGTTCGTAGCTTTACCGTTAGCCGTCCGTCCGGATATGGTGCAGGG
ACCGTCTATTACCCAACCAATGCAGGCGGCACCGTTGGCGCGATTGCAATCGTCCCCGGGTACACC
GCGCGTCAAAGCAGCATTAAGTGGTGGGGTCCGCGCTTAGCTAGCCATGGCTTTGTGGTTATTACC
ATCGATACGAACAGCACTCTAGACCAGCCCAGCAGCCGTAGCTCGCAACAGATGGCCGCGCTTCGT
CAAGTTGCGAGCTTGAACGGGACCAGCAGTAGCCCGATTTACGGAAAGGTCGATACTGCCCGCATG
GGTGTGATGGGCTGGTCAATGGGGGGCGGCGGTTCACTTATTAGCGCCGCGAACAACCCGAGTTTA
AAAGCAGCGGCACCGCAGGCGCCATGGGACTCTTCAACCAACTTCAGCAGTGTTACCGTGCCGACG
CTGATTTTCGCGTGCGAGAATGATAGCATTGCACCGGTGAACAGCAGCGCGCTGCCGATTTATGAT
AGCATGTCCCGCAACGCAAAACAGTTTCTGGAAATTAACGGCGGTAGCCACTCTTGTGCCAACTCT
GGGAACAGCAACCAGGCACTGATCGGAAAAAAAGGGGTTGCATGGATGAAACGATTCATGGATAAT
GACACCCGTTACTCAACCTTCGCCTGTGAGAATCCCAACAGCACACGCGTGTCGGATTTTCGCACC
GCGAACTGTTCCCTCGAGCACCACCATCACCACCACTGA

7.1.2.2 IsPETase-WT (no signal peptide) (5’ – 3’)

ATGCAGACCAATCCGTATGCGCGCGGCCCCAACCCTACCGCCGCCTCGTTGGAAGCCAGCGCGGGA
CCCTTTACCGTTCGTAGCTTTACCGTTAGCCGTCCGTCCGGATATGGTGCAGGGACCGTCTATTAC
CCAACCAATGCAGGCGGCACCGTTGGCGCGATTGCAATCGTCCCCGGGTACACCGCGCGTCAAAGC
AGCATTAAGTGGTGGGGTCCGCGCTTAGCTAGCCATGGCTTTGTGGTTATTACCATCGATACGAAC

AGCACTCTAGACCAGCCCAGCAGCCGTAGCTCGCAACAGATGGCCGCGCTTCGTCAAGTTGCGAGC
TTGAACGGGACCAGCAGTAGCCCGATTTACGGAAAGGTCGATACTGCCCGCATGGGTGTGATGGGC
TGGTCAATGGGGGGCGGCGGTTCACTTATTAGCGCCGCGAACAACCCGAGTTTAAAAGCAGCGGCA
CCGCAGGCGCCATGGGACTCTTCAACCAACTTCAGCAGTGTTACCGTGCCGACGCTGATTTTCGCG
TGCGAGAATGATAGCATTGCACCGGTGAACAGCAGCGCGCTGCCGATTTATGATAGCATGTCCCGC
AACGCAAAACAGTTTCTGGAAATTAACGGCGGTAGCCACTCTTGTGCCAACTCTGGGAACAGCAAC
CAGGCACTGATCGGAAAAAAAGGGGTTGCATGGATGAAACGATTCATGGATAATGACACCCGTTAC
TCAACCTTCGCCTGTGAGAATCCCAACAGCACACGCGTGTCGGATTTTCGCACCGCGAACTGTTCC
CTCGAGCACCACCATCACCACCACTGA

*Is*PETase-Spy (5’ – 3’)

ATGGCGATGGGTAAAACTATTAGTACATGGATTTCAGATGGACAAGTGAAAGATTTCTACCTGTAT
CCAGGAAAATATACATTTGTCGAAACCGCAGCACCAGACGGTTATGAGGTAGCAACTGCTATTACC
TTTACAGTTAATGAGCAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGGCAGCGAAAAC
CTGTACTTCCAGGGTGGCAGCGGAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTGATGAG
GACGGCAAAGAGTTAGCTGGTGCAACTATGGAGTTGCGTGATTCAGTGGATGGCCATCAGACCAAT
CCGTATGCGCGCGGCCCCAACCCTACCGCCGCCTCGTTGGAAGCCAGCGCGGGACCCTTTACCGTT
CGTAGCTTTACCGTTAGCCGTCCGTCCGGATATGGTGCAGGGACCGTCTATTACCCAACCAATGCA
GGCGGCACCGTTGGCGCGATTGCAATCGTCCCCGGGTACACCGCGCGTCAAAGCAGCATTAAGTGG
TGGGGTCCGCGCTTAGCTAGCCATGGCTTTGTGGTTATTACCATCGATACGAACAGCACTCTAGAC
CAGCCCAGCAGCCGTAGCTCGCAACAGATGGCCGCGCTTCGTCAAGTTGCGAGCTTGAACGGGACC
AGCAGTAGCCCGATTTACGGAAAGGTCGATACTGCCCGCATGGGTGTGATGGGCTGGTCAATGGGG
GGCGGCGGTTCACTTATTAGCGCCGCGAACAACCCGAGTTTAAAAGCAGCGGCACCGCAGGCGCCA
TGGGACTCTTCAACCAACTTCAGCAGTGTTACCGTGCCGACGCTGATTTTCGCGTGCGAGAATGAT
AGCATTGCACCGGTGAACAGCAGCGCGCTGCCGATTTATGATAGCATGTCCCGCAACGCAAAACAG
TTTCTGGAAATTAACGGCGGTAGCCACTCTTGTGCCAACTCTGGGAACAGCAACCAGGCACTGATC
GGAAAAAAAGGGGTTGCATGGATGAAACGATTCATGGATAATGACACCCGTTACTCAACCTTCGCC
TGTGAGAATCCCAACAGCACACGCGTGTCGGATTTTCGCACCGCGAACTGTTCCCTCGAGGCTAGC
GGTCCTGGTAGTGGTGGCAGCGGAGCCCACATCGTGATGGTGGACGCCTACAAGCCGACGAAGGGT
TCAGGTCCGCACCATCACCATCACCATTAA

*Is*PETase-Tag (5’ – 3’)

ATGGGCCATCATCATCATCATCATGGAAGCTTGGCGCATATTGTGATGGTGGATGCGTATAAACCG
ACCAAAGGTAGCGGCGCTAGCCAGACCAATCCGTATGCGCGCGGCCCCAACCCTACCGCCGCCTCG
TTGGAAGCCAGCGCGGGACCCTTTACCGTTCGTAGCTTTACCGTTAGCCGTCCGTCCGGATATGGT
GCAGGGACCGTCTATTACCCAACCAATGCAGGCGGCACCGTTGGCGCGATTGCAATCGTCCCCGGG
TACACCGCGCGTCAAAGCAGCATTAAGTGGTGGGGTCCGCGCTTAGCAAGCCATGGCTTTGTGGTT

ATTACCATCGATACGAACAGCACTCTAGACCAGCCCAGCAGCCGTAGCTCGCAACAGATGGCCGCG
CTTCGTCAAGTTGCGAGCTTGAACGGGACCAGCAGTAGCCCGATTTACGGAAAGGTCGATACTGCC
CGCATGGGTGTGATGGGCTGGTCAATGGGGGGCGGCGGTTCACTTATTAGCGCCGCGAACAACCCG
AGTTTAAAAGCAGCGGCACCGCAGGCGCCATGGGACTCTTCAACCAACTTCAGCAGTGTTACCGTG
CCGACGCTGATTTTCGCGTGCGAGAATGATAGCATTGCACCGGTGAACAGCAGCGCGCTGCCGATT
TATGATAGCATGTCCCGCAACGCAAAACAGTTTCTGGAAATTAACGGCGGTAGCCACTCTTGTGCC
AACTCTGGGAACAGCAACCAGGCACTGATCGGAAAAAAAGGGGTTGCATGGATGAAACGATTCATG
GATAATGACACCCGTTACTCAACCTTCGCCTGTGAGAATCCCAACAGCACACGCGTGTCGGATTTT
CGCACCGCGAACTGTTCCCTCGAGGGTAGCGGTGGCAGCGGAGCCCACATCGTGATGGTGGACGCC
TACAAGCCGACGAAGTGA

*Is*PETase-Catcher (5’ – 3’)

ATGGGCCATCATCATCATCATCATGGAAGCTTGGGCGCGATGGTGGATACCCTGAGCGGCCTGAGC
AGCGAACAGGGCCAGAGCGGCGATATGACCATTGAAGAAGATAGCGCGACCCATATTAAATTTAGC
AAACGCGATGAAGATGGCAAAGAACTGGCGGGCGCGACGATGGAACTGCGCGATAGCAGCGGCAAA
ACCATTAGCACCTGGATTAGCGATGGCCAGGTGAAAGATTTTTATCTGTATCCGGGCAAATATACC
TTTGTGGAAACCGCGGCGCCGGATGGCTATGAAGTGGCGACCGCGATTACCTTTACCGTGAACGAA
CAGGGCCAGGTGACCGTGAACGGCAAAGCGACCAAAGGCGATGCGCATATTGATGGTAGCGGCGCT
AGCCAGACCAATCCGTATGCGCGCGGCCCCAACCCTACCGCCGCCTCGTTGGAAGCCAGCGCGGGA
CCCTTTACCGTTCGTAGCTTTACCGTTAGCCGTCCGTCCGGATATGGTGCAGGGACCGTCTATTAC
CCAACCAATGCAGGCGGCACCGTTGGCGCGATTGCAATCGTCCCCGGGTACACCGCGCGTCAAAGC
AGCATTAAGTGGTGGGGTCCGCGCTTAGCCAGCCATGGCTTTGTGGTTATTACCATCGATACGAAC
AGCACTCTAGACCAGCCCAGCAGCCGTAGCTCGCAACAGATGGCCGCGCTTCGTCAAGTTGCGAGC
TTGAACGGGACCAGCAGTAGCCCGATTTACGGAAAGGTCGATACTGCCCGCATGGGTGTGATGGGC
TGGTCAATGGGGGGCGGCGGTTCACTTATTAGCGCCGCGAACAACCCGAGTTTAAAAGCAGCGGCA
CCGCAGGCGCCATGGGACTCTTCAACCAACTTCAGCAGTGTTACCGTGCCGACGCTGATTTTCGCG
TGCGAGAATGATAGCATTGCACCGGTGAACAGCAGCGCGCTGCCGATTTATGATAGCATGTCCCGC
AACGCAAAACAGTTTCTGGAAATTAACGGCGGTAGCCACTCTTGTGCCAACTCTGGGAACAGCAAC
CAGGCACTGATCGGAAAAAAAGGGGTTGCATGGATGAAACGATTCATGGATAATGACACCCGTTAC
TCAACCTTCGCCTGTGAGAATCCCAACAGCACACGCGTGTCGGATTTTCGCACCGCGAACTGTTCC
GCTAGCGGCGGCTCTGGCGAAAACCTGTATTTTCAGGGCGCGATGGTGGATACCCTGAGCGGCCTG
AGCAGCGAACAGGGCCAGAGCGGCGATATGACCATTGAAGAAGATAGCGCGACCCATATTAAATTT
AGCAAACGCGATGAAGATGGCAAAGAACTGGCGGGCGCGACGATGGAACTGCGCGATAGCAGCGGC
AAAACCATTAGCACCTGGATTAGCGATGGCCAGGTGAAAGATTTTTATCTGTATCCGGGCAAATAT
ACCTTTGTGGAAACCGCGGCGCCGGATGGCTATGAAGTGGCGACCGCGATTACCTTTACCGTGAAC
GAACAGGGCCAGGTGACCGTGAACGGCAAAGCGACCAAAGGCGATGCGCATATTGATTGA

_Is_PETase-Cat

ATGGGCCATCATCATCATCATCATGGAAGCTTGGCGCATATTGTGATGGTGGATGCGTATAAACCG
ACCAAAGGCAGCGCGGCCGCAGGCAGCGGCGAATATTTTACCCTGCAGATTCGTGGCCGCGAACGC
TTTGAAGAATTTCGCGAAAAAAACGAAGCGCTGGAACTGAAAGATGCGCAGGCGGGCAAAGAACCG
GGTGGCTCTGGTGGCAGCGGCCATATGCAGACCAATCCGTATGCGCGCGGCCCCAACCCTACCGCC
GCCTCGTTGGAAGCCAGCGCGGGACCCTTTACCGTTCGTAGCTTTACCGTTAGCCGTCCGTCCGGA
TATGGTGCAGGGACCGTCTATTACCCAACCAATGCAGGCGGCACCGTTGGCGCGATTGCAATCGTC
CCCGGGTACACCGCGCGTCAAAGCAGCATTAAGTGGTGGGGTCCGCGCTTAGCTAGCCATGGCTTT
GTGGTTATTACCATCGATACGAACAGCACTCTAGACCAGCCCAGCAGCCGTAGCTCGCAACAGATG
GCCGCGCTTCGTCAAGTTGCGAGCTTGAACGGGACCAGCAGTAGCCCGATTTACGGAAAGGTCGAT
ACTGCCCGCATGGGTGTGATGGGCTGGTCAATGGGGGGCGGCGGTTCACTTATTAGCGCCGCGAAC
AACCCGAGTTTAAAAGCAGCGGCACCGCAGGCGCCATGGGACTCTTCAACCAACTTCAGCAGTGTT
ACCGTGCCGACGCTGATTTTCGCGTGCGAGAATGATAGCATTGCACCGGTGAACAGCAGCGCGCTG
CCGATTTATGATAGCATGTCCCGCAACGCAAAACAGTTTCTGGAAATTAACGGCGGTAGCCACTCT
TGTGCCAACTCTGGGAACAGCAACCAGGCACTGATCGGAAAAAAAGGGGTTGCATGGATGAAACGA
TTCATGGATAATGACACCCGTTACTCAACCTTCGCCTGTGAGAATCCCAACAGCACACGCGTGTCG
GATTTTCGCACCGCGAACTGTTCCGCTAGCGGCGGCTCTGGCGAAAACCTGTATTTTCAGGGCGCG
ATGGTGGATACCCTGAGCGGCCTGAGCAGCGAACAGGGCCAGAGCGGCGATATGACCATTGAAGAA
GATAGCGCGACCCATATTAAATTTAGCAAACGCGATGAAGATGGCAAAGAACTGGCGGGCGCGACG
ATGGAACTGCGCGATAGCAGCGGCAAAACCATTAGCACCTGGATTAGCGATGGCCAGGTGAAAGAT
TTTTATCTGTATCCGGGCAAATATACCTTTGTGGAAACCGCGGCGCCGGATGGCTATGAAGTGGCG
ACCGCGATTACCTTTACCGTGAACGAACAGGGCCAGGTGACCGTGAACGGCAAAGCGACCAAAGGC
GATGCGCATATTGATTGA

7.2 AMINO ACID SEQUENCES

7.2.1 TET12SN Amino Acid Sequences

Colour code: TET12 sequence; Linker; His-tag; ^{C}Npu DnaE; ^{N}Npu DnaE; SpyCatcher-N^{TEV}; SpyTag; TEV Recognition Site; (mutations shown are shown in bold and underlined for insertions/substitutions and as asterisks for deletions)

TET12SN

(M)LEEELKQLEEELQAIEEQLAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEEISQLEQ
KNSELKEKNQELKYGKGDGDIEQELERAKESIRRLEQEVNQERSRMQYLQTLLEKGKGDGQLEDKV
EELLSKNYHLENEVERLKKLVGGKGDGLEEELKQLEEELQAIEEQLAQLQWKAQARKEKLAQLKEK
LGKGDGSPEDEIQQLEEKNSQLKQEISQLEEKNQELKYGDGKGQLEDKVEELLSKNYHLENEVERL
KKLVGGDGKGSPEDKISQLKEKIQQLKQENQQLEEENSQLEYGDGKGSPEDENSQLEEKISQLKQK

NSELKEEIQQLEYGDGKGSPEDKISELKEENQQLEQKIQQLKEENSQLEYGKGDGDIEQELERAKE
SIRRLEQEVNQERSRMQYLQTLLEKGKGDGSPEDKNSELKEEIQQLEEENQQLEEKISELKYLEHH
HHHHHH*

TET12-Int

(M)IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASNCMGLGSGPHMLEEELKQLEEELQAIEEQ
LAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEEISQLEQKNSELKEKNQELKYGKGDGDI
EQELERAKESIRRLEQEVNQERSRMQYLQTLLEKGKGDGQLEDKVEELLSKNYHLENEVERLKKLV
GGKGDGLEEELKQLEEELQAIEEQLAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEKNSQ
LKQEISQLEEKNQELKYGDGKGQLEDKVEELLSKNYHLENEVERLKKLVGGDGKGSPEDKISQLKE
KIQQLKQENQQLEEENSQLEYGDGKGSPEDENSQLEEKISQLKQKNSELKEEIQQLEYGDGKGSPE
DKISELKEENQQLEQKIQQLKEENSQLEYGKGDGDIEQELERAKESIRRLEQEVNQERSRMQYLQT
LLEKGKGDGSPEDKNSELKEEIQQLEEENQQLEEKISELKYLEASGPGNGLGSGPHHHHHHAACLS
YETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRAT
KDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN*

TET12-Int^{Nat}

(M)IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN**CFN**LGSGPHMLEEELKQLEEELQAIEEQ
LAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEEISQLEQKNSELKEKNQELKYGKGDGDI
EQELERAKESIRRLEQEVNQERSRMQYLQTLLEKGKGDGQLEDKVEELLSKNYHLENEVERLKKLV
GGKGDGLEEELKQLEEELQAIEEQLAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEKNSQ
LKQEISQLEEKNQELKYGDGKGQLEDKVEELLSKNYHLENEVERLKKLVGGDGKGSPEDKISQLKE
KIQQLKQENQQLEEENSQLEYGDGKGSPEDENSQLEEKISQLKQKNSELKEEIQQLEYGDGKGSPE
DKISELKEENQQLEQKIQQLKEENSQLEYGKGDGDIEQELERAKESIRRLEQEVNQERSRMQYLQT
LLEKGKGDGSPEDKNSELKEEIQQLEEENQQLEEKISELKYLEASGPGNGLGSGPHHHHHHA**AEY**C
LSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIR
ATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN*

TET12-Int$^{C1A/N137A}$

(M)IKIATRKYLGKQNVYDIGVERDHNFALKNGFIAS**A**CMGLGSGPHMLEEELKQLEEELQAIEEQ
LAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEEISQLEQKNSELKEKNQELKYGKGDGDI
EQELERAKESIRRLEQEVNQERSRMQYLQTLLEKGKGDGQLEDKVEELLSKNYHLENEVERLKKLV
GGKGDGLEEELKQLEEELQAIEEQLAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEKNSQ
LKQEISQLEEKNQELKYGDGKGQLEDKVEELLSKNYHLENEVERLKKLVGGDGKGSPEDKISQLKE
KIQQLKQENQQLEEENSQLEYGDGKGSPEDENSQLEEKISQLKQKNSELKEEIQQLEYGDGKGSPE

DKISELKEENQQLEQKIQQLKEENSQLEYGKGDGDIEQELERAKESIRRLEQEVNQERSRMQYLQT
LLEKGKGDGSPEDKNSELKEEIQQLEEENQQLEEKISELKYLEASGPGNGLGSGPHHHHHHAAALS
YETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRAT
KDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN*

TET12-Int^{M+2A}

(M)IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASNCAGLGSGPHMLEEELKQLEEELQAIEEQ
LAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEEISQLEQKNSELKEKNQELKYGKGDGDI
EQELERAKESIRRLEQEVNQERSRMQYLQTLLEKGKGDGQLEDKVEELLSKNYHLENEVERLKKLV
GGKGDGLEEELKQLEEELQAIEEQLAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEKNSQ
LKQEISQLEEKNQELKYGDGKGQLEDKVEELLSKNYHLENEVERLKKLVGGDGKGSPEDKISQLKE
KIQQLKQENQQLEEENSQLEYGDGKGSPEDENSQLEEKISQLKQKNSELKEEIQQLEYGDGKGSPE
DKISELKEENQQLEQKIQQLKEENSQLEYGKGDGDIEQELERAKESIRRLEQEVNQERSRMQYLQT
LLEKGKGDGSPEDKNSELKEEIQQLEEENQQLEEKISELKYLEASGPGNGLGSGPHHHHHHAACLS
YETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRAT
KDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN*

TET12-Int^{H125N}

(M)IKIATRKYLGKQNVYDIGVERDNNFALKNGFIASNCMGLGSGPHMLEEELKQLEEELQAIEEQ
LAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEEISQLEQKNSELKEKNQELKYGKGDGDI
EQELERAKESIRRLEQEVNQERSRMQYLQTLLEKGKGDGQLEDKVEELLSKNYHLENEVERLKKLV
GGKGDGLEEELKQLEEELQAIEEQLAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEKNSQ
LKQEISQLEEKNQELKYGDGKGQLEDKVEELLSKNYHLENEVERLKKLVGGDGKGSPEDKISQLKE
KIQQLKQENQQLEEENSQLEYGDGKGSPEDENSQLEEKISQLKQKNSELKEEIQQLEYGDGKGSPE
DKISELKEENQQLEQKIQQLKEENSQLEYGKGDGDIEQELERAKESIRRLEQEVNQERSRMQYLQT
LLEKGKGDGSPEDKNSELKEEIQQLEEENQQLEEKISELKYLEASGPGNGLGSGPHHHHHHAACLS
YETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRAT
KDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN*

TET12-Int^{+5}

(M)IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASNCMGAGSGPLGSGPHMLEEELKQLEEELQ
AIEEQLAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEEISQLEQKNSELKEKNQELKYGK
GDGDIEQELERAKESIRRLEQEVNQERSRMQYLQTLLEKGKGDGQLEDKVEELLSKNYHLENEVER
LKKLVGGKGDGLEEELKQLEEELQAIEEQLAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLE
EKNSQLKQEISQLEEKNQELKYGDGKGQLEDKVEELLSKNYHLENEVERLKKLVGGDGKGSPEDKI

SQLKEKIQQLKQENQQLEEENSQLEYGDGKGSPEDENSQLEEKISQLKQKNSELKEEIQQLEYGDG
KGSPEDKISELKEENQQLEQKIQQLKEENSQLEYGKGDGDIEQELERAKESIRRLEQEVNQERSRM
QYLQTLLEKGKGDGSPEDKNSELKEEIQQLEEENQQLEEKISELKYLEASGPGNGLGSGPHHHHHH
AACLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGS
LIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN*

TET12-Int^{-5}

(M)IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASNCMGLGSGPHMLEEELKQLEEELQAIEEQ
LAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEEISQLEQKNSELKEKNQELKYGKGDGDI
EQELERAKESIRRLEQEVNQERSRMQYLQTLLEKGKGDGQLEDKVEELLSKNYHLENEVERLKKLV
GGKGDGLEEELKQLEEELQAIEEQLAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEKNSQ
LKQEISQLEEKNQELKYGDGKGQLEDKVEELLSKNYHLENEVERLKKLVGGDGKGSPEDKISQLKE
KIQQLKQENQQLEEENSQLEYGDGKGSPEDENSQLEEKISQLKQKNSELKEEIQQLEYGDGKGSPE
DKISELKEENQQLEQKIQQLKEENSQLEYGKGDGDIEQELERAKESIRRLEQEVNQERSRMQYLQT
LLEKGKGDGSPEDKNSELKEEIQQLEEENQQLEEKISELKYLEASG*****GSGPHHHHHHAACLS
YETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRAT
KDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN*

TET12-Spy

(M)AMGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGGS
ENLYFQGGSGEDSATHIKFSKRDEDGKELAGATMELRDSVDGHMLEEELKQLEEELQAIEEQLAQL
QWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEEISQLEQKNSELKEKNQELKYGKGDGDIEQEL
ERAKESIRRLEQEVNQERSRMQYLQTLLEKGKGDGQLEDKVEELLSKNYHLENEVERLKKLVGGKG
DGLEEELKQLEEELQAIEEQLAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEKNSQLKQE
ISQLEEKNQELKYGDGKGQLEDKVEELLSKNYHLENEVERLKKLVGGDGKGSPEDKISQLKEKIQQ
LKQENQQLEEENSQLEYGDGKGSPEDENSQLEEKISQLKQKNSELKEEIQQLEYGDGKGSPEDKIS
ELKEENQQLEQKIQQLKEENSQLEYGKGDGDIEQELERAKESIRRLEQEVNQERSRMQYLQTLLEK
GKGDGSPEDKNSELKEEIQQLEEENQQLEEKISELKYLEASGPGSGGSGAHIVMVDAYKPTKGSGP
HHHHHH*

TET12-Spy^{D7A}

(M)AMGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGGS
ENLYFQGGSGEDSATHIKFSKRDEDGKELAGATMELRDSVDGHMLEEELKQLEEELQAIEEQLAQL
QWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEEISQLEQKNSELKEKNQELKYGKGDGDIEQEL
ERAKESIRRLEQEVNQERSRMQYLQTLLEKGKGDGQLEDKVEELLSKNYHLENEVERLKKLVGGKG

DGLEEELKQLEEELQAIEEQLAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEKNSQLKQE
ISQLEEKNQELKYGDGKGQLEDKVEELLSKNYHLENEVERLKKLVGGDGKGSPEDKISQLKEKIQQ
LKQENQQLEEENSQLEYGDGKGSPEDENSQLEEKISQLKQKNSELKEEIQQLEYGDGKGSPEDKIS
ELKEENQQLEQKIQQLKEENSQLEYGKGDGDIEQELERAKESIRRLEQEVNQERSRMQYLQTLLEK
GKGDGSPEDKNSELKEEIQQLEEENQQLEEKISELKYLEASGPGSGGSGAHIVMVAAYKPTKGSGP
HHHHHH*

TET12-Spy^{+5}

(M)AMGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGGS
ENLYFQGGSGEDSATHIKFSKRDEDGKELAGATMELRDSVDGHMLEEELKQLEEELQAIEEQLAQL
QWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEEISQLEQKNSELKEKNQELKYGKGDGDIEQEL
ERAKESIRRLEQEVNQERSRMQYLQTLLEKGKGDGQLEDKVEELLSKNYHLENEVERLKKLVGGKG
DGLEEELKQLEEELQAIEEQLAQLQWKAQARKEKLAQLKEKLGKGDGSPEDEIQQLEEKNSQLKQE
ISQLEEKNQELKYGDGKGQLEDKVEELLSKNYHLENEVERLKKLVGGDGKGSPEDKISQLKEKIQQ
LKQENQQLEEENSQLEYGDGKGSPEDENSQLEEKISQLKQKNSELKEEIQQLEYGDGKGSPEDKIS
ELKEENQQLEQKIQQLKEENSQLEYGKGDGDIEQELERAKESIRRLEQEVNQERSRMQYLQTLLEK
GKGDGSPEDKNSELKEEIQQLEEENQQLEEKISELKYLEASGPGSGGSG**GSGGS**AHIVMVDAYKPT
KGSGPHHHHHH*

7.2.2 *Is*PETase Amino Acid Sequences

Colour code: *Is*PETase sequence; Linker; His-tag; SpyCatcher(-N^{TEV}); SpyTag; p53 dimerisation domain; TEV Recognition Site; (mutations shown are shown in bold and underlined for insertions/substitutions and as asterisks for deletions)

_Is_PETase-WT (including signal peptide)

(M)NFPRASRLMQAAVLGGLMAVSAAATAQTNPYARGPNPTAASLEASAGPFTVRSFTVSRPSGYG
AGTVYYPTNAGGTVGAIAIVPGYTARQSSIKWWGPRLASHGFVVITIDTNSTLDQPSSRSSQQMAA
LRQVASLNGTSSSPIYGKVDTARMGVMGWSMGGGGSLISAANNPSLKAAAPQAPWDSSTNFSSVTV
PTLIFACENDSIAPVNSSALPIYDSMSRNAKQFLEINGGSHSCANSGNSNQALIGKKGVAWMKRFM
DNDTRYSTFACENPNSTRVSDFRTANCSLEHHHHHH*

_Is_PETase-WT (no signal peptide)

(M)QTNPYARGPNPTAASLEASAGPFTVRSFTVSRPSGYGAGTVYYPTNAGGTVGAIAIVPGYTAR
QSSIKWWGPRLASHGFVVITIDTNSTLDQPSSRSSQQMAALRQVASLNGTSSSPIYGKVDTARMGV
MGWSMGGGGSLISAANNPSLKAAAPQAPWDSSTNFSSVTVPTLIFACENDSIAPVNSSALPIYDSM

SRNAKQFLEINGGSHSCANSGNSNQALIGKKGVAWMKRFMDNDTRYSTFACENPNSTRVSDFRTAN
CSLEHHHHHH*

*Is*PETase-Spy

(M)AMGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGGS
ENLYFQGGSGEDSATHIKFSKRDEDGKELAGATMELRDSVDGHQTNPYARGPNPTAASLEASAGPF
TVRSFTVSRPSGYGAGTVYYPTNAGGTVGAIAIVPGYTARQSSIKWWGPRLASHGFVVITIDTNST
LDQPSSRSSQQMAALRQVASLNGTSSSPIYGKVDTARMGVMGWSMGGGGSLISAANNPSLKAAAPQ
APWDSSTNFSSVTVPTLIFACENDSIAPVNSSALPIYDSMSRNAKQFLEINGGSHSCANSGNSNQA
LIGKKGVAWMKRFMDNDTRYSTFACENPNSTRVSDFRTANCSLEASGPGSGGSGAHIVMVDAYKPT
KGSGPHHHHHH*

*Is*PETase-Spy^{D7A}

(M)AMGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGGS
ENLYFQGGSGEDSATHIKFSKRDEDGKELAGATMELRDSVDGHQTNPYARGPNPTAASLEASAGPF
TVRSFTVSRPSGYGAGTVYYPTNAGGTVGAIAIVPGYTARQSSIKWWGPRLASHGFVVITIDTNST
LDQPSSRSSQQMAALRQVASLNGTSSSPIYGKVDTARMGVMGWSMGGGGSLISAANNPSLKAAAPQ
APWDSSTNFSSVTVPTLIFACENDSIAPVNSSALPIYDSMSRNAKQFLEINGGSHSCANSGNSNQA
LIGKKGVAWMKRFMDNDTRYSTFACENPNSTRVSDFRTANCSLEASGPGSGGSGAHIVMV**A**AYKPT
KGSGPHHHHHH*

*Is*PETase-Tag

(M)GHHHHHHGSLAHIVMVDAYKPTKGSGASQTNPYARGPNPTAASLEASAGPFTVRSFTVSRPSG
YGAGTVYYPTNAGGTVGAIAIVPGYTARQSSIKWWGPRLASHGFVVITIDTNSTLDQPSSRSSQQM
AALRQVASLNGTSSSPIYGKVDTARMGVMGWSMGGGGSLISAANNPSLKAAAPQAPWDSSTNFSSV
TVPTLIFACENDSIAPVNSSALPIYDSMSRNAKQFLEINGGSHSCANSGNSNQALIGKKGVAWMKR
FMDNDTRYSTFACENPNSTRVSDFRTANCSLEGSGGSGAHIVMVDAYKPTK*

*Is*PETase-Catcher

(M)GHHHHHHGSLGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDSS
GKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIDGS
GASQTNPYARGPNPTAASLEASAGPFTVRSFTVSRPSGYGAGTVYYPTNAGGTVGAIAIVPGYTAR
QSSIKWWGPRLASHGFVVITIDTNSTLDQPSSRSSQQMAALRQVASLNGTSSSPIYGKVDTARMGV
MGWSMGGGGSLISAANNPSLKAAAPQAPWDSSTNFSSVTVPTLIFACENDSIAPVNSSALPIYDSM

SRNAKQFLEINGGSHSCANSGNSNQALIGKKGVAWMKRFMDNDTRYSTFACENPNSTRVSDFRTAN
CSASGGSGENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDS
SGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHID*

_Is_PETase-Cat

(M)GHHHHHHGSLAHIVMVDAYKPTKGSAAAGSGEYFTLQIRGRERFEEFREKNEALELKDAQAGK
EPGGSGGSGHMQTNPYARGPNPTAASLEASAGPFTVRSFTVSRPSGYGAGTVYYPTNAGGTVGAIA
IVPGYTARQSSIKWWGPRLASHGFVVITIDTNSTLDQPSSRSSQQMAALRQVASLNGTSSSPIYGK
VDTARMGVMGWSMGGGGSLISAANNPSLKAAAPQAPWDSSTNFSSVTVPTLIFACENDSIAPVNSS
ALPIYDSMSRNAKQFLEINGGSHSCANSGNSNQALIGKKGVAWMKRFMDNDTRYSTFACENPNSTR
VSDFRTANCSASGGSGENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAG
ATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKAT
KGDAHID*

_Is_PETase-Cat^{+5}

(M)GHHHHHHGSLAHIVMVDAYKPTKGSAAAGSGEYFTLQIRGRERFEEFREKNEALELKDAQAGK
EPGGSGGSG**GSGGS**HMQTNPYARGPNPTAASLEASAGPFTVRSFTVSRPSGYGAGTVYYPTNAGGT
VGAIAIVPGYTARQSSIKWWGPRLASHGFVVITIDTNSTLDQPSSRSSQQMAALRQVASLNGTSSS
PIYGKVDTARMGVMGWSMGGGGSLISAANNPSLKAAAPQAPWDSSTNFSSVTVPTLIFACENDSIA
PVNSSALPIYDSMSRNAKQFLEINGGSHSCANSGNSNQALIGKKGVAWMKRFMDNDTRYSTFACEN
PNSTRVSDFRTANCSASGGSGENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDG
KELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTV
NGKATKGDAHID*

_Is_PETase-Cat^{D7A}

(M)GHHHHHHGSLAHIVMV**A**AYKPTKGSAAAGSGEYFTLQIRGRERFEEFREKNEALELKDAQAGK
EPGGSGGSGHMQTNPYARGPNPTAASLEASAGPFTVRSFTVSRPSGYGAGTVYYPTNAGGTVGAIA
IVPGYTARQSSIKWWGPRLASHGFVVITIDTNSTLDQPSSRSSQQMAALRQVASLNGTSSSPIYGK
VDTARMGVMGWSMGGGGSLISAANNPSLKAAAPQAPWDSSTNFSSVTVPTLIFACENDSIAPVNSS
ALPIYDSMSRNAKQFLEINGGSHSCANSGNSNQALIGKKGVAWMKRFMDNDTRYSTFACENPNSTR
VSDFRTANCSASGGSGENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAG
ATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKAT
KGDAHID*

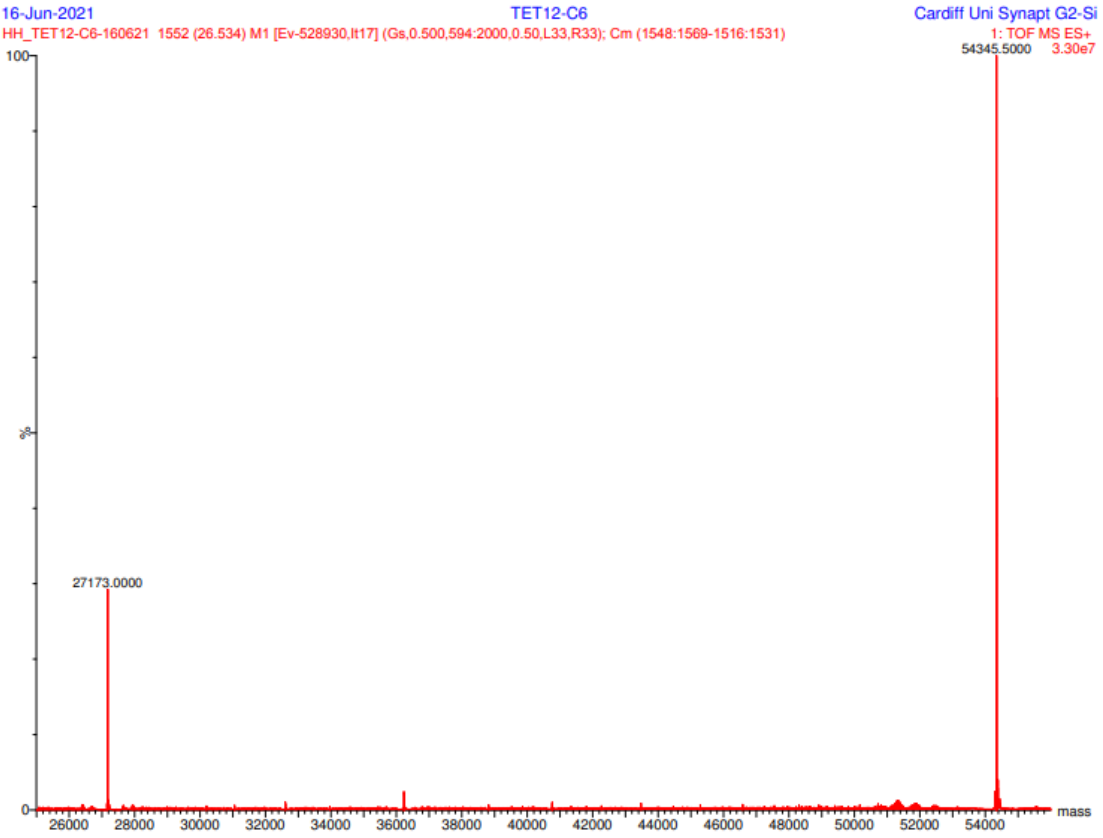
_Is_PETase-Cat^{K19P}

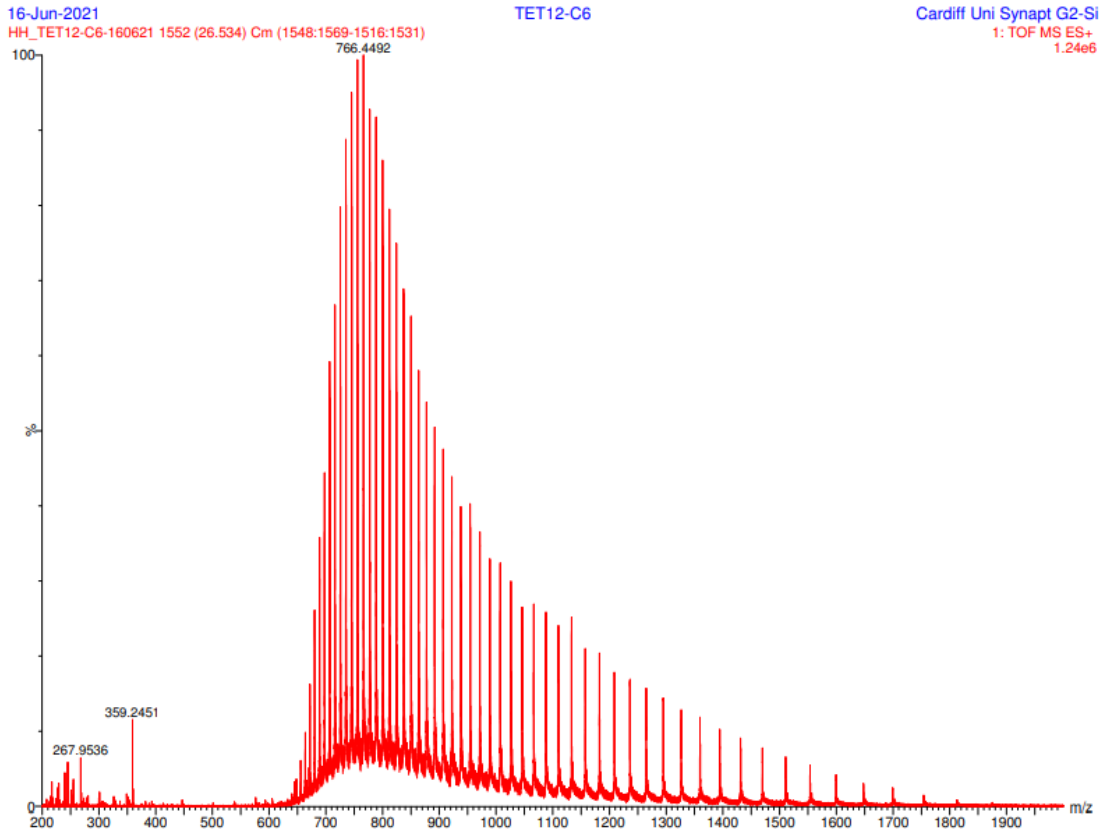
(M)GHHHHHHGSLAHIVMVDAYKPTKGSAAAGSGEYFTLQIRGRERFEEFRE**P**NEALELKDAQAGK
EPGGSGGSGHMQTNPYARGPNPTAASLEASAGPFTVRSFTVSRPSGYGAGTVYYPTNAGGTVGAIA
IVPGYTARQSSIKWWGPRLASHGFVVITIDTNSTLDQPSSRSSQQMAALRQVASLNGTSSSPIYGK
VDTARMGVMGWSMGGGGSLISAANNPSLKAAAPQAPWDSSTNFSSVTVPTLIFACENDSIAPVNSS
ALPIYDSMSRNAKQFLEINGGSHSCANSGNSNQALIGKKGVAWMKRFMDNDTRYSTFACENPNSTR
VSDFRTANCSASGGSGENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAG
ATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKAT
KGDAHID*

7.3 PROTEIN MASS SPECTRA

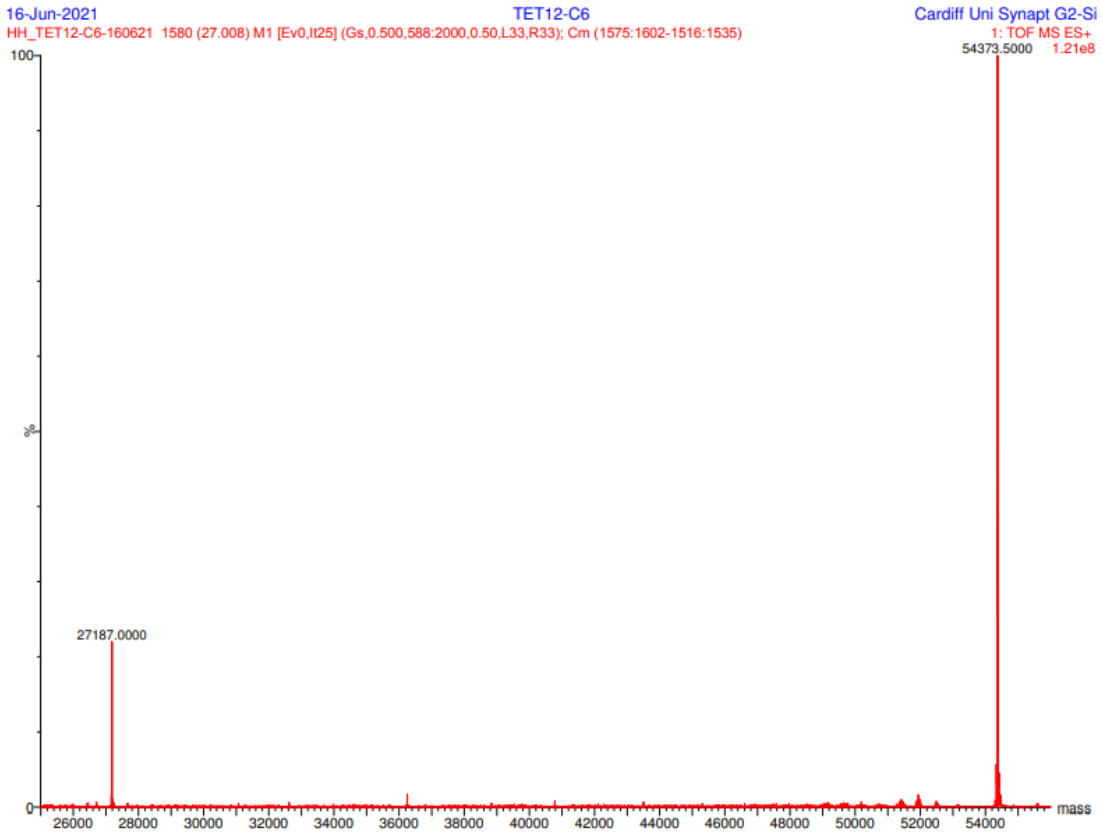
7.3.1 TET12 Protein Mass Spectra

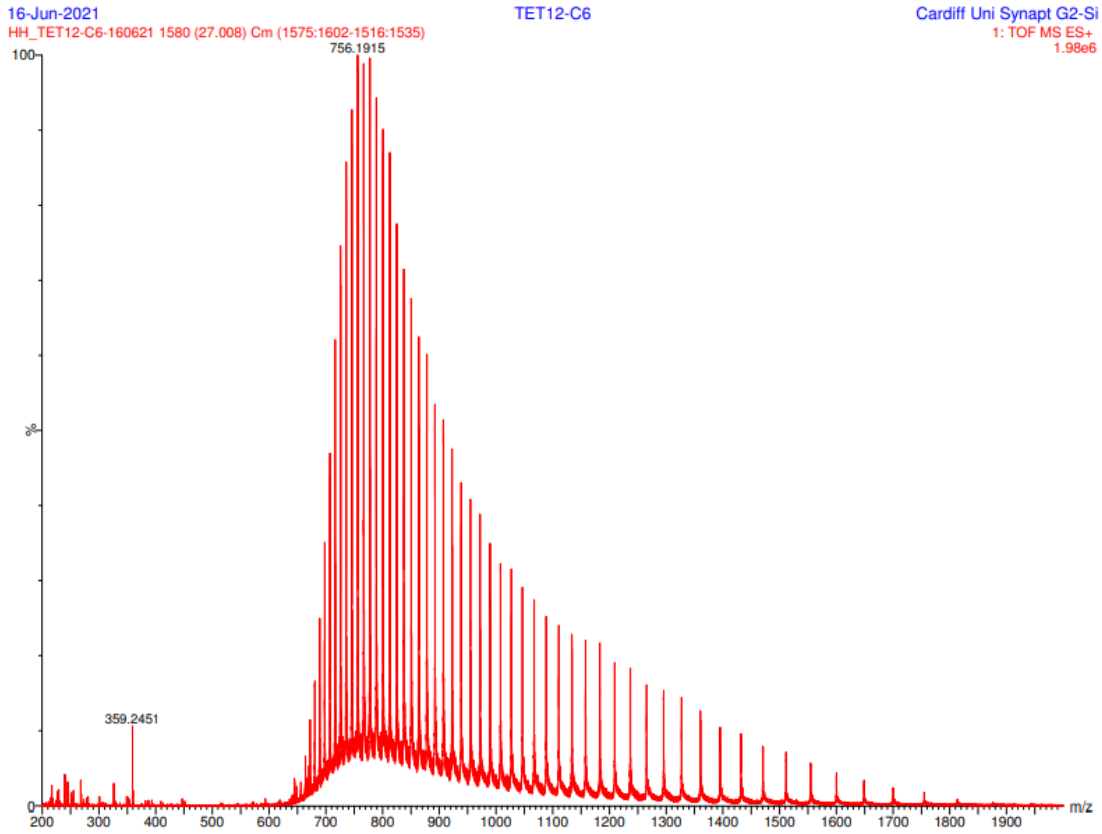
TET12SN



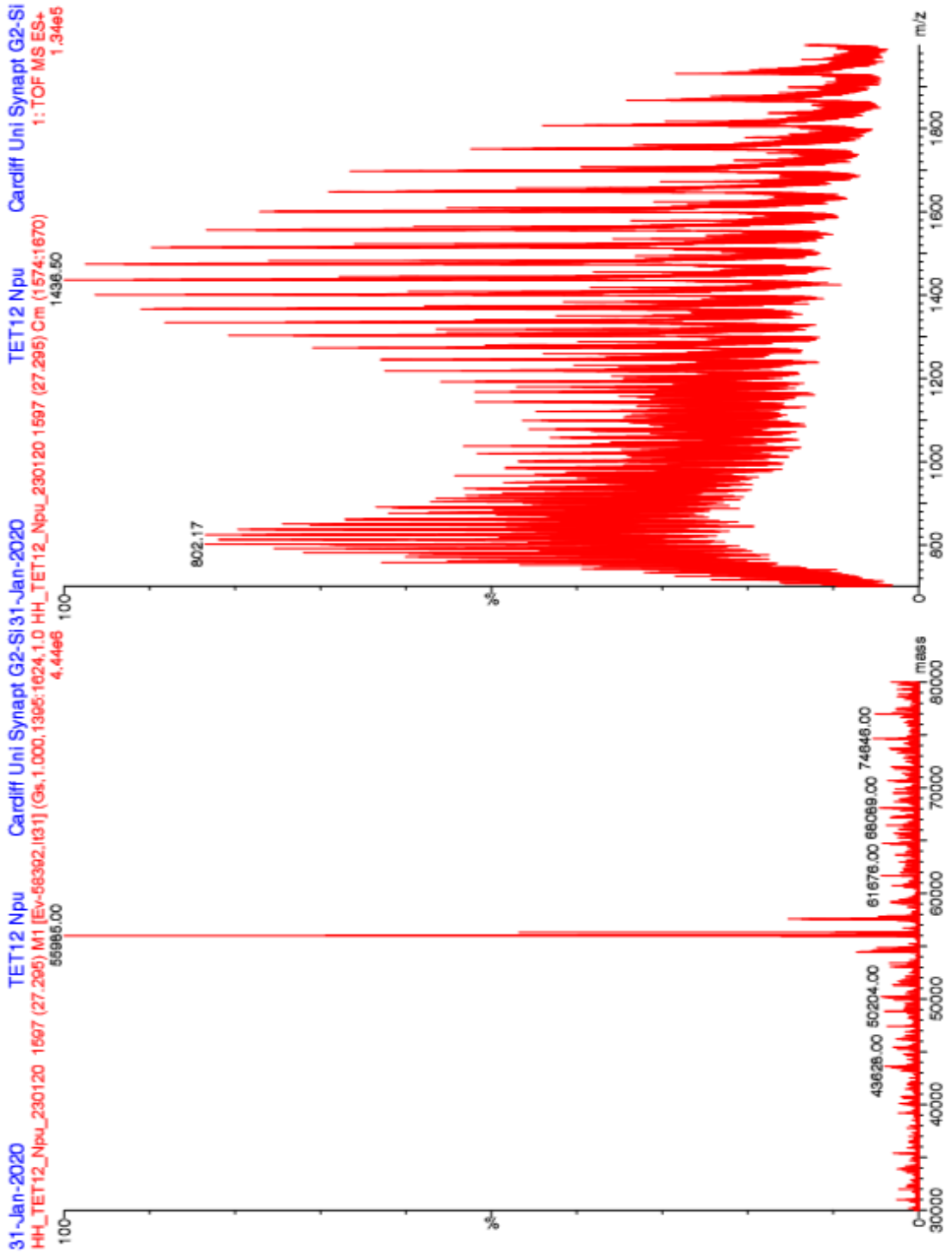


TET12SN (fMet)

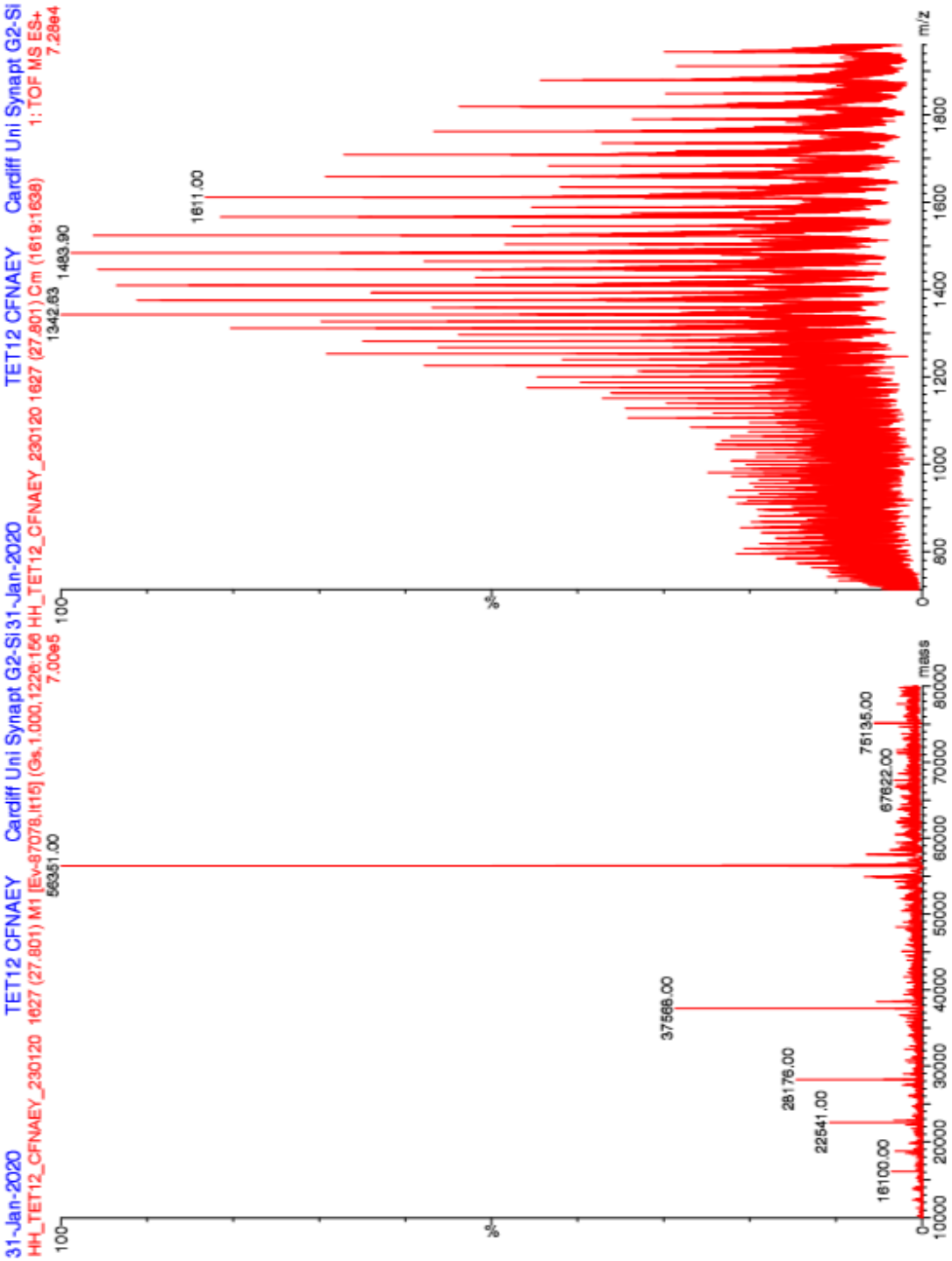




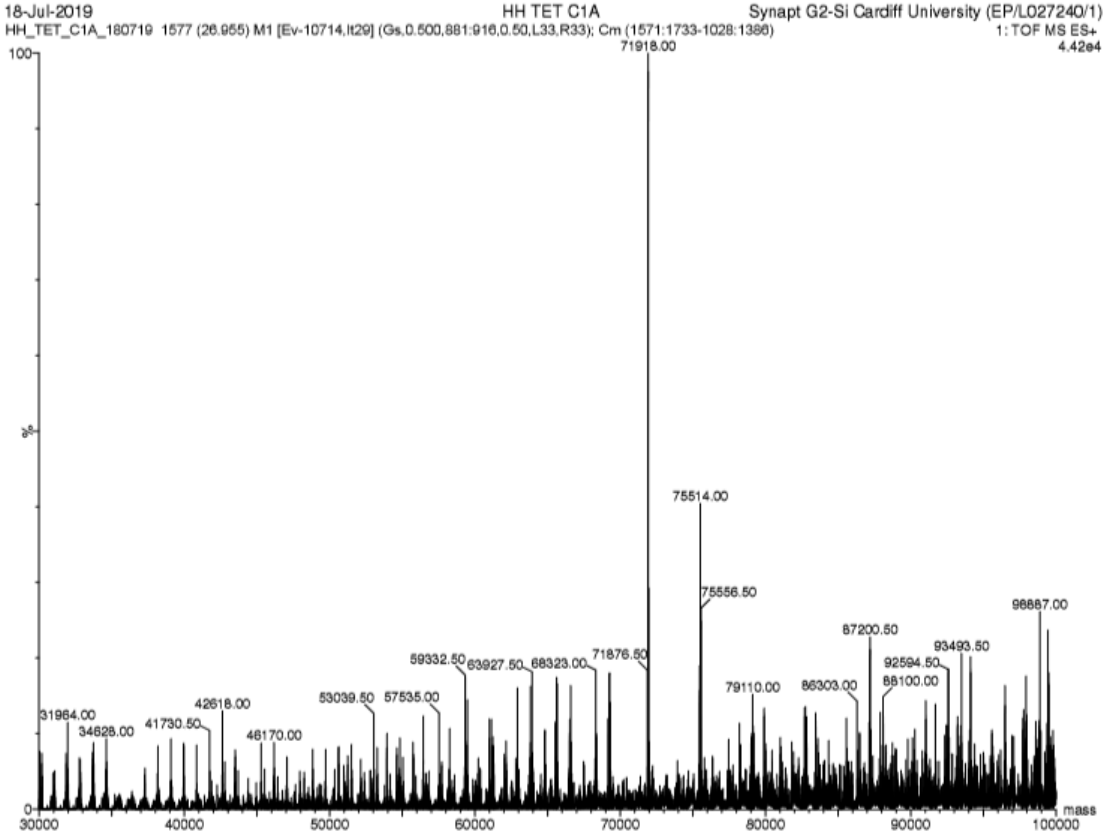
TET12-Int

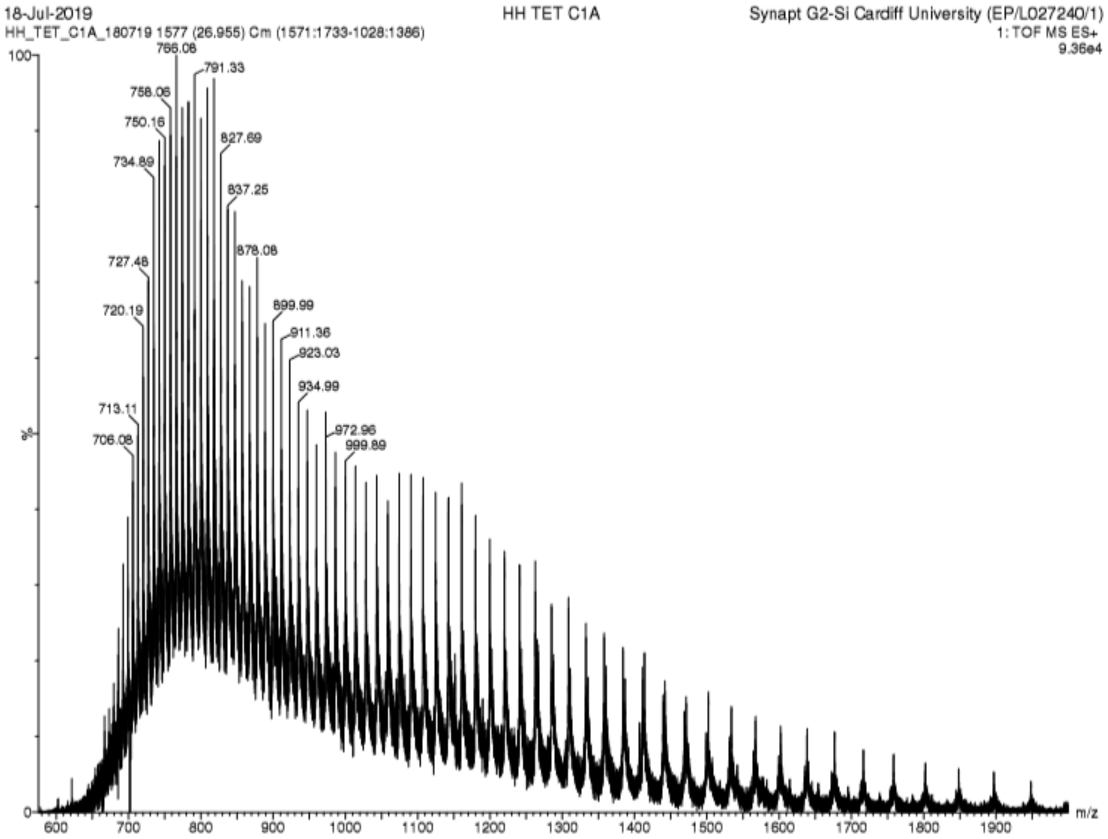

31-Jan-2020
HH_TET12_Npu_230120 1597 (27.295) M1 [Ev-58392.I31] (Gs,1.000,1395:1624,1.0 HH_TET12_Npu_230120 1597 (27.295) Cm (1574:1670)
TET12 Npu
Cardiff Uni Synapt G2-Si
31-Jan-2020
TET12 Npu
Cardiff Uni Synapt G2-Si
1: TOF MS ES+
1.34e5
55995.00
4.44e6
1436.50
802.17
61676.00
68069.00
74646.00
50204.00
43626.00
mass
m/z
%

TET12-Int^{Nat}

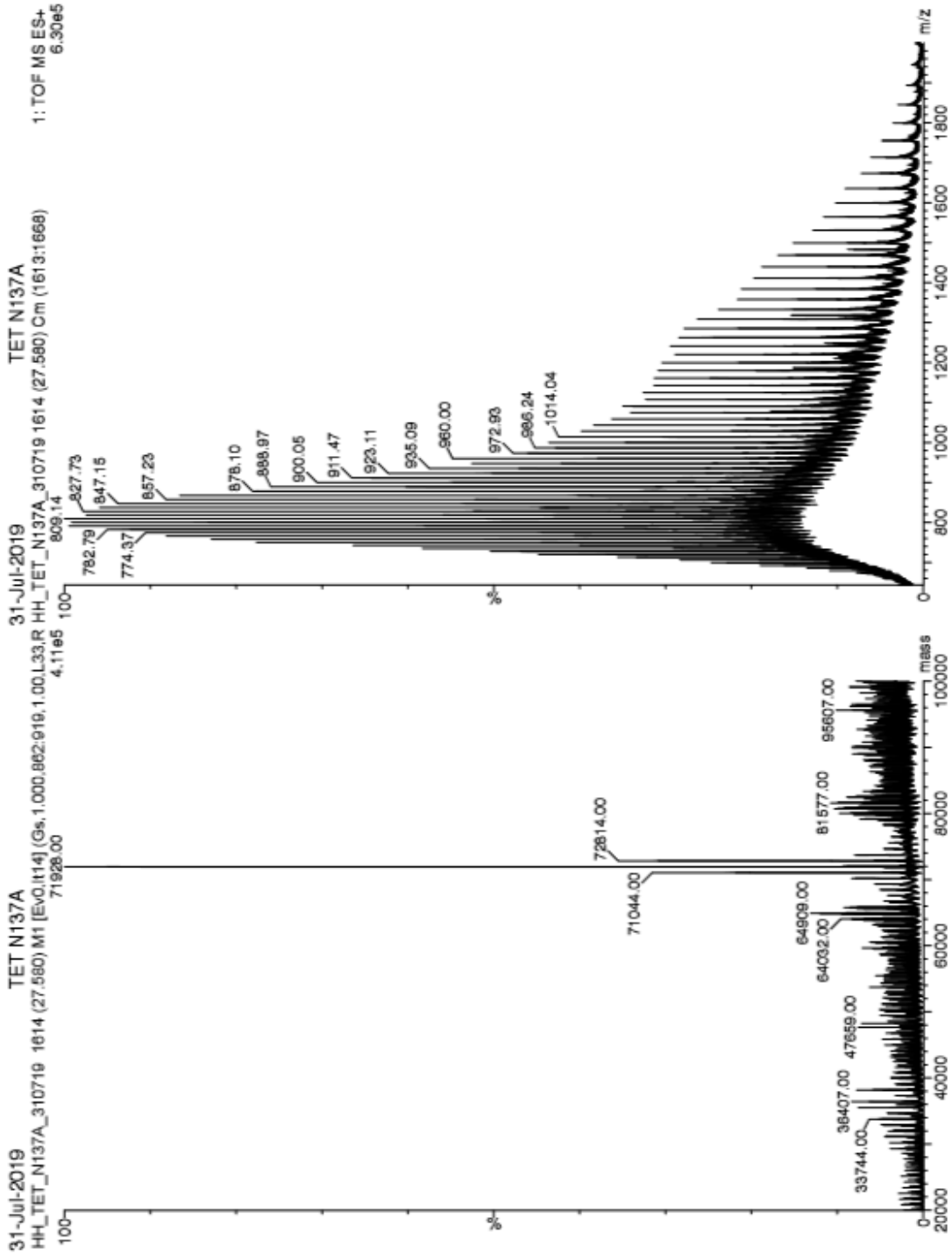


TET12-Int^{C1A}

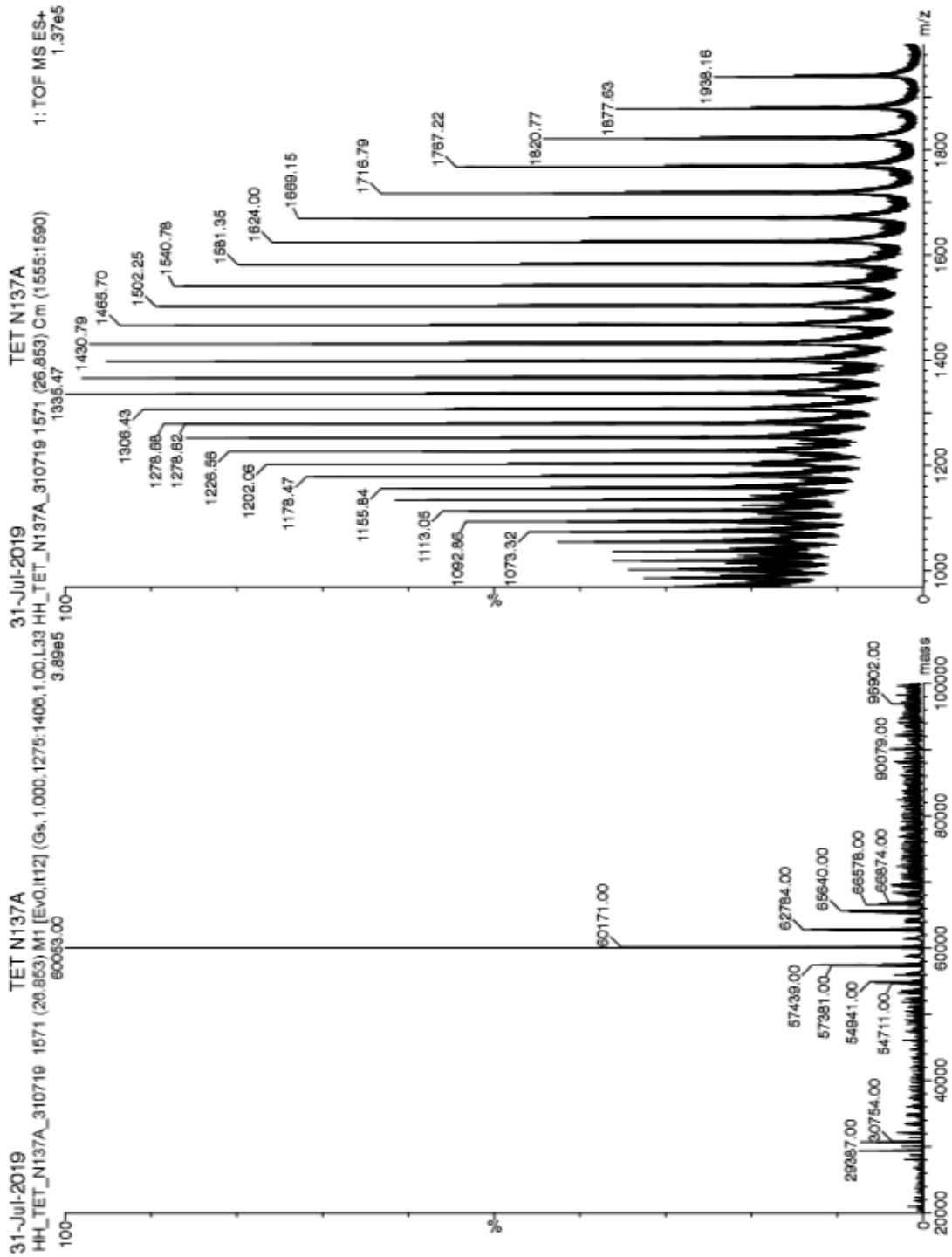




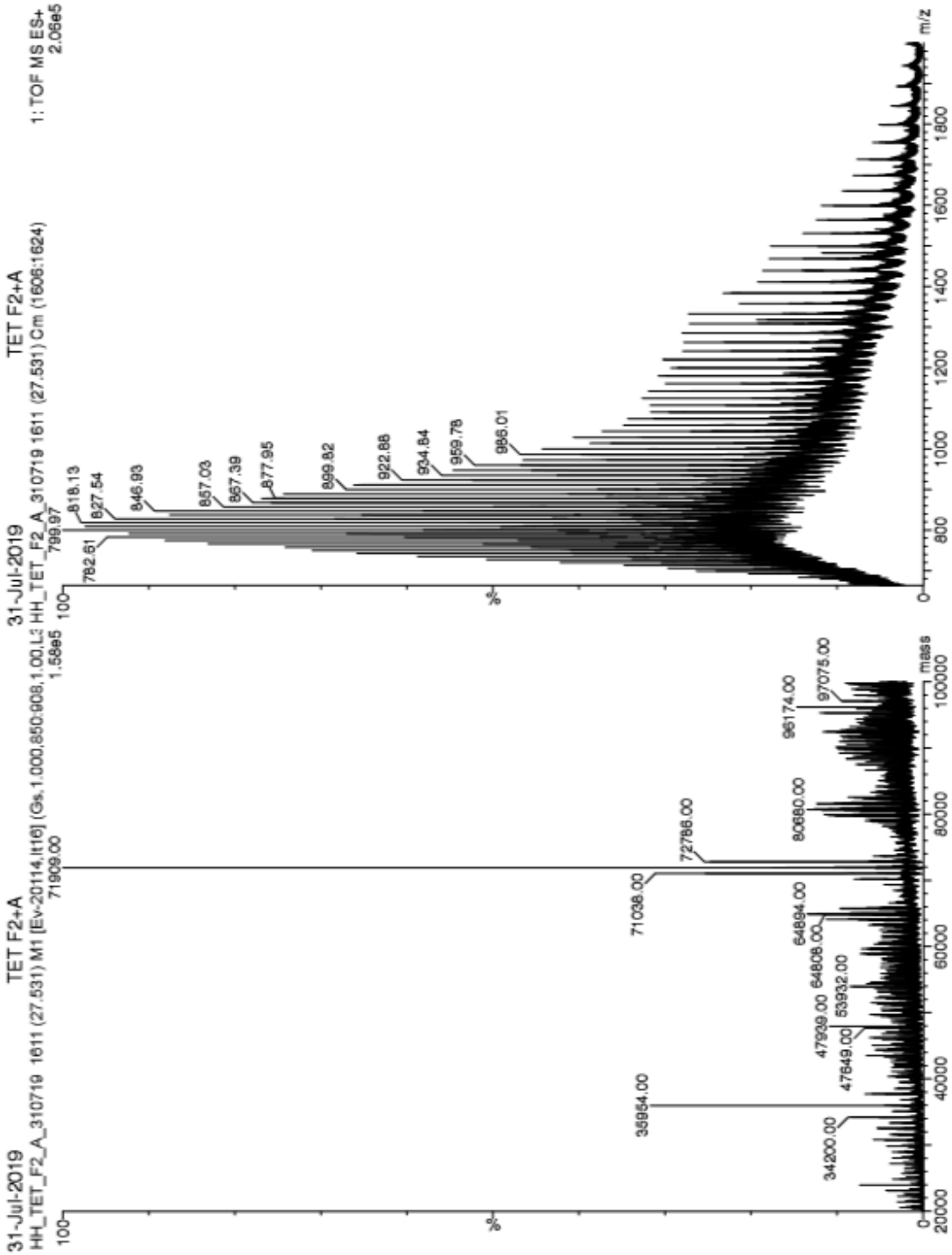
TET12-Int^{N137A} (Thioester intermediate)



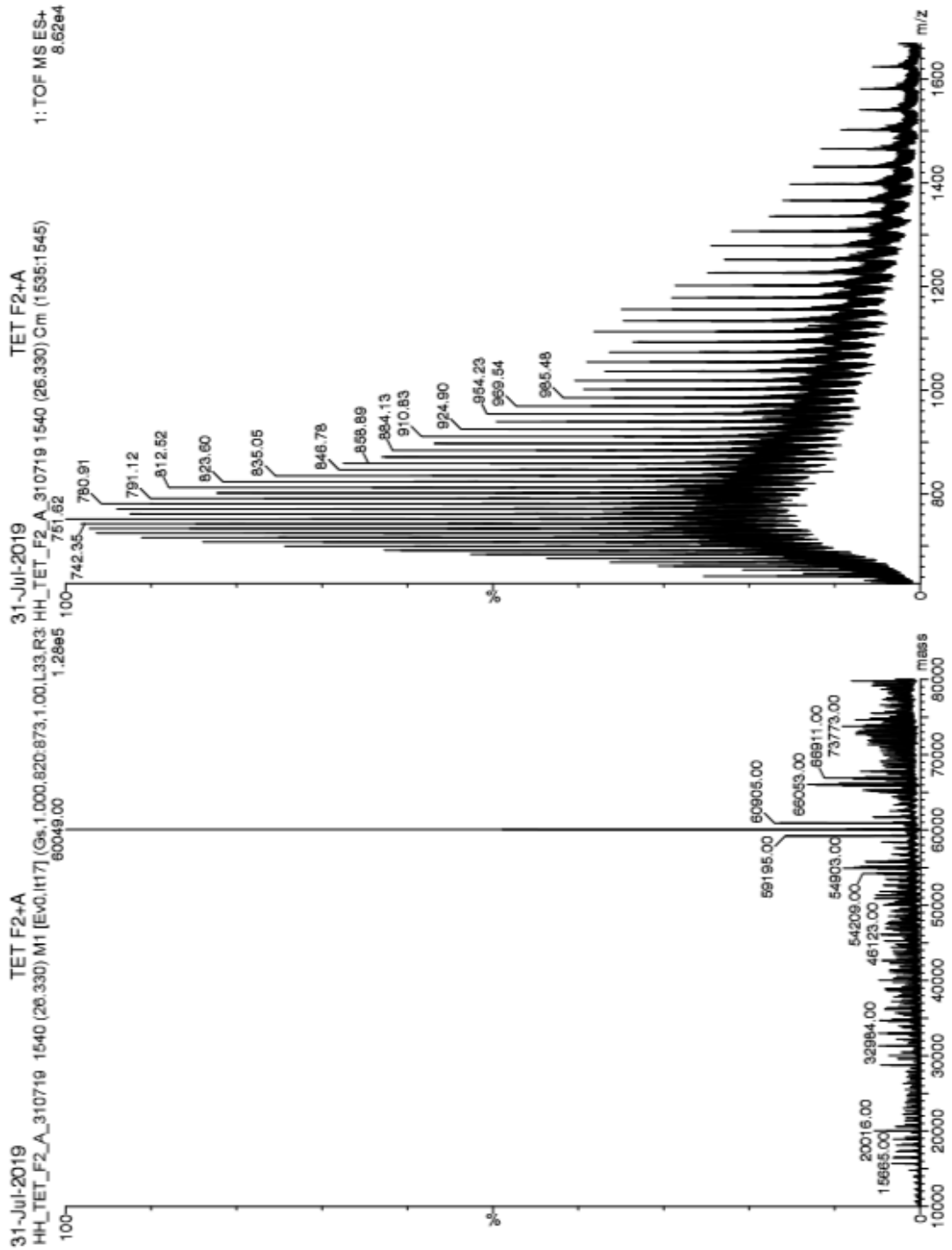
TET12-Int^{N137A} (Thioester intermediate without Int^{N})



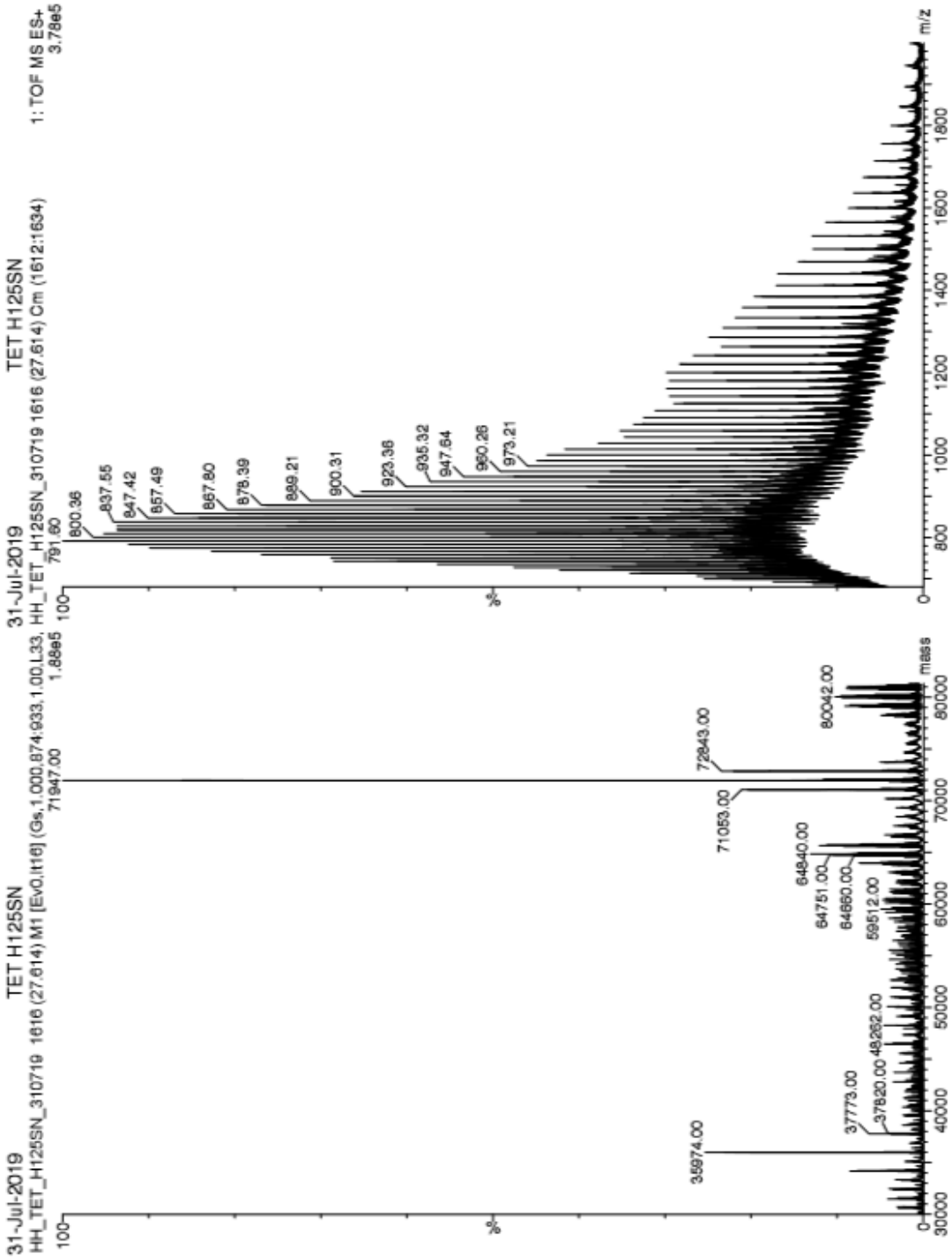
TET12-Int^{M+2A} (Thioester intermediate)



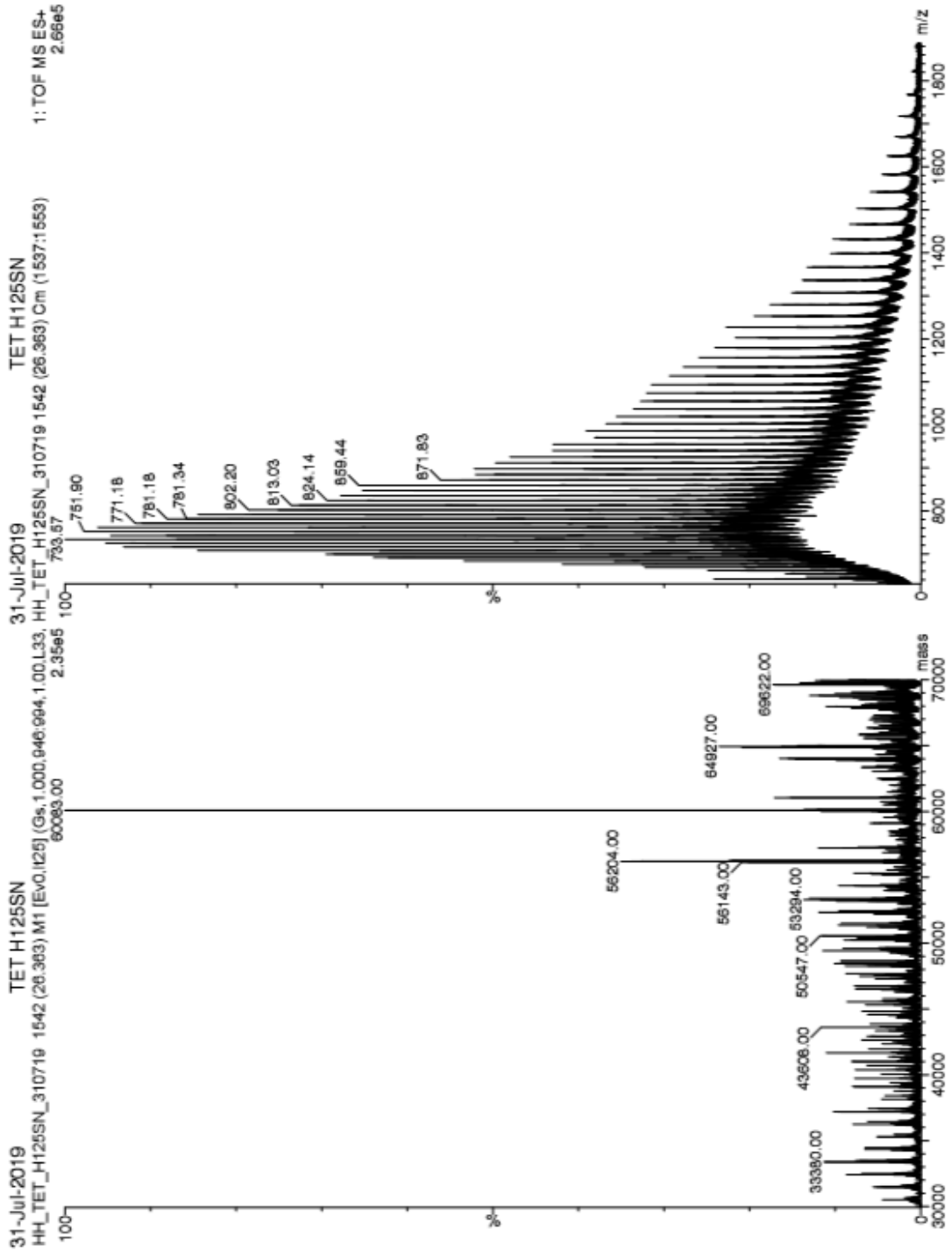
TET12-Int^{M+2A} (Thioester Intermediate without Int^{N})



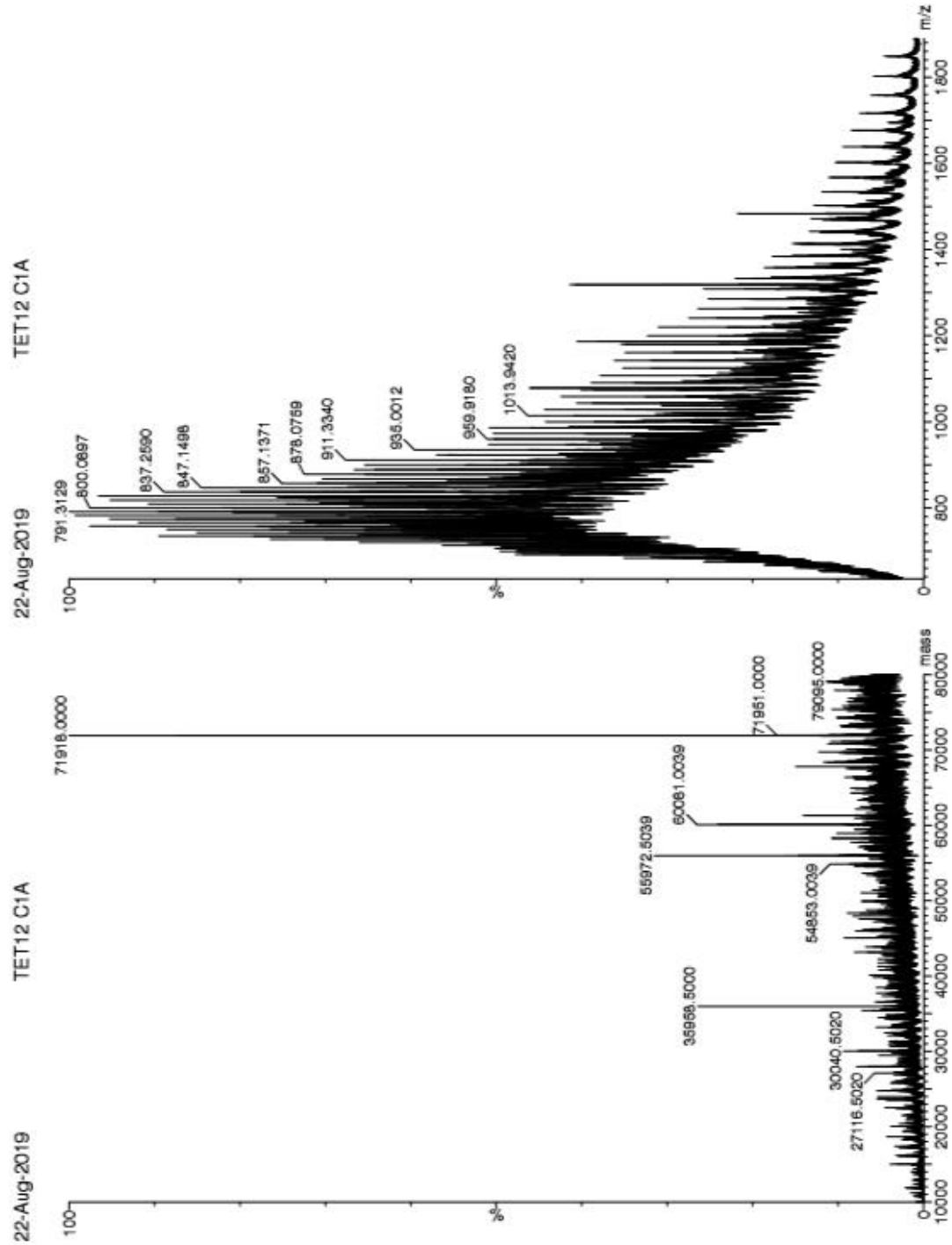
TET12-Int^{H125N} (Thioester intermediate)



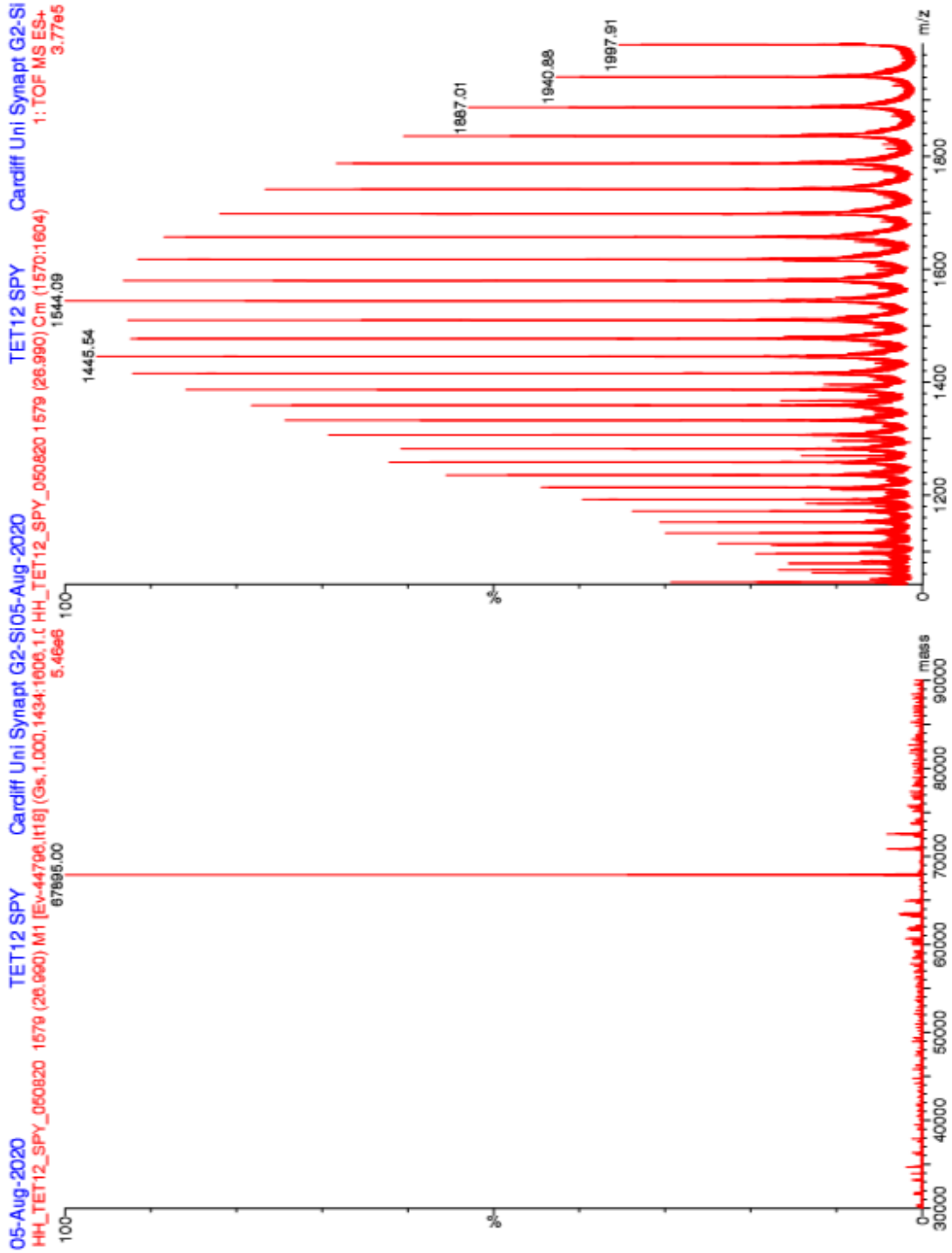
TET12-Int^{H125N} (Thioester intermediate without Int^{N})



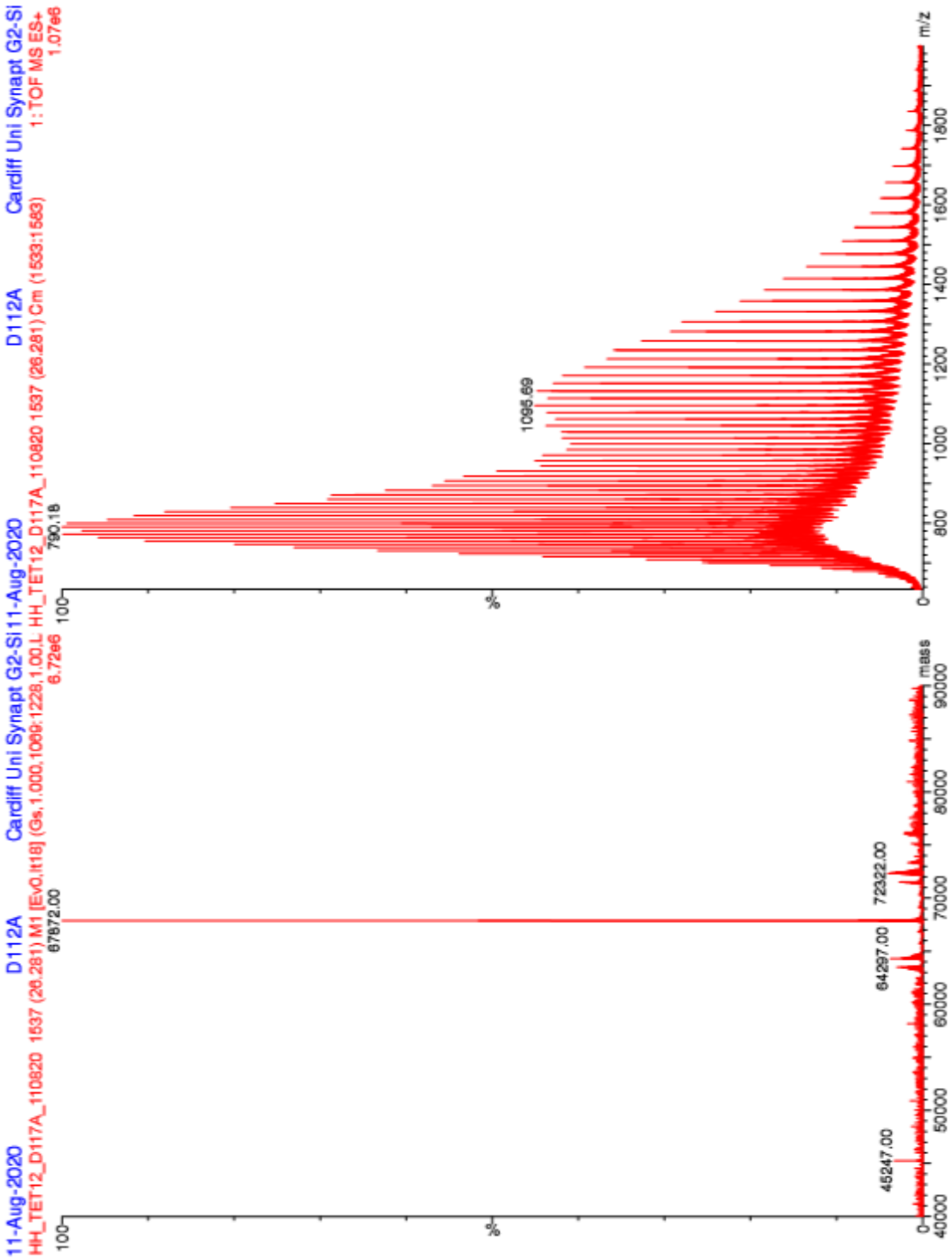
TET12-Int$^{C1A/N137A}$



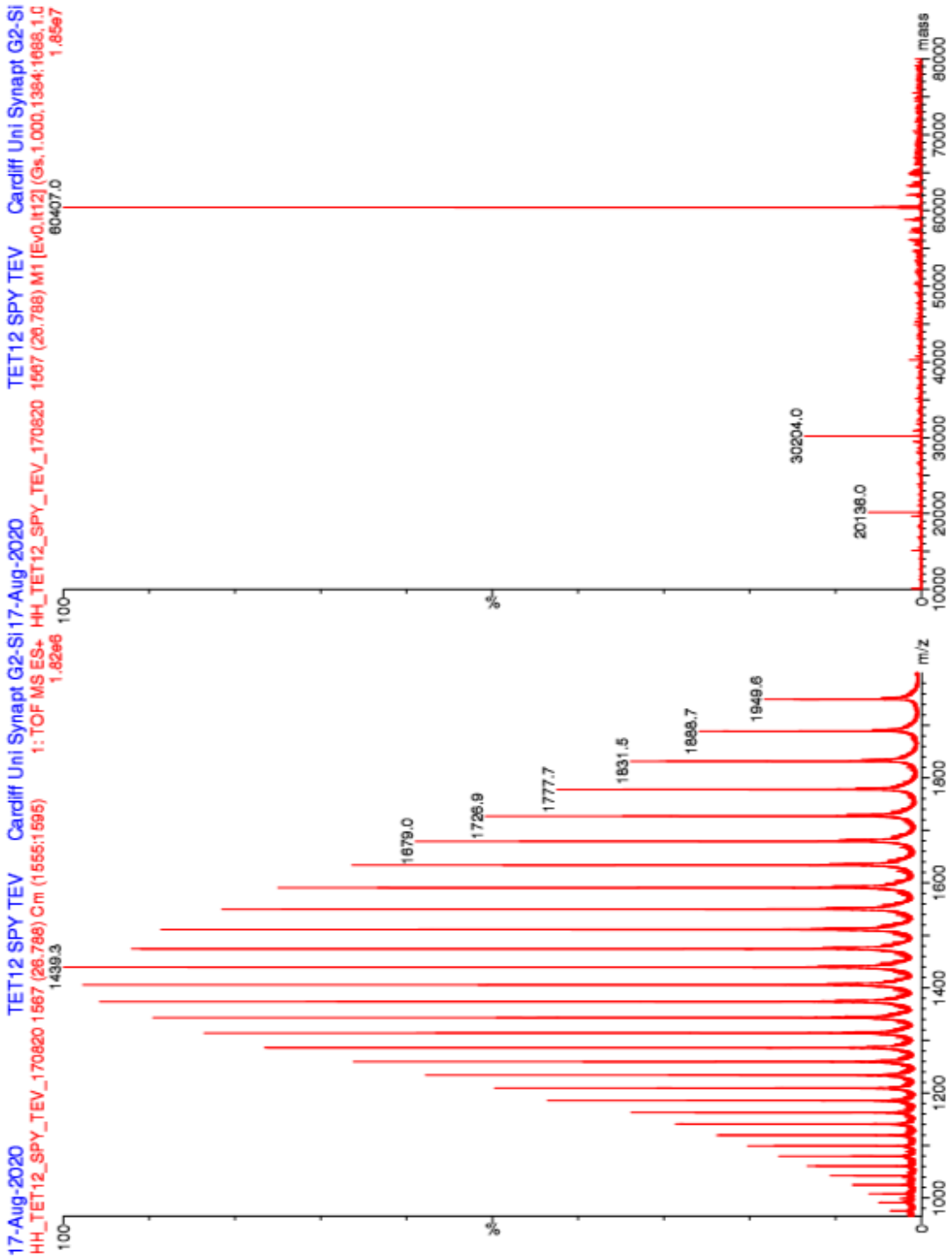
TET12-Spy


Cardiff Uni Synapt G2-Si
1: TOF MS ES+
3.77e5
TET12 SPY
(26.990) Cm (1570:1604)
1544.09
1445.54
1887.01
1940.88
1997.91
m/z
05-Aug-2020
HH_TET12_SPY_050820 1579 (26.990) M1 [Ev-44796,It18] (Gs,1.000,1434:1606,1.0 HH_TET12_SPY_050820 1579 (26.990) Cm (1570:1604)
5.46e6
87695.00
mass

TET12-Spy^{D7A}

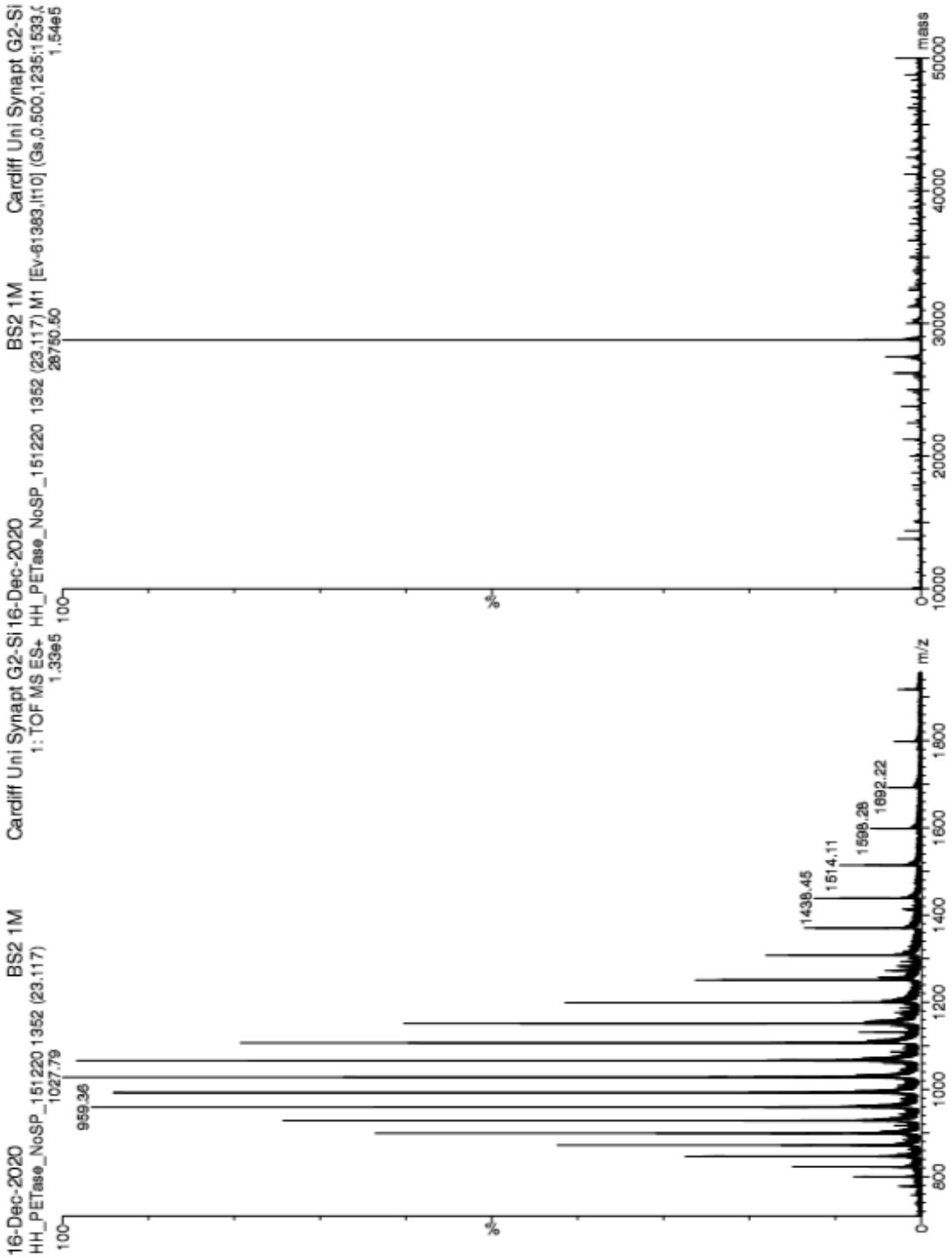


TET12-Spy (Catcher-N^{TEV} cleavage)

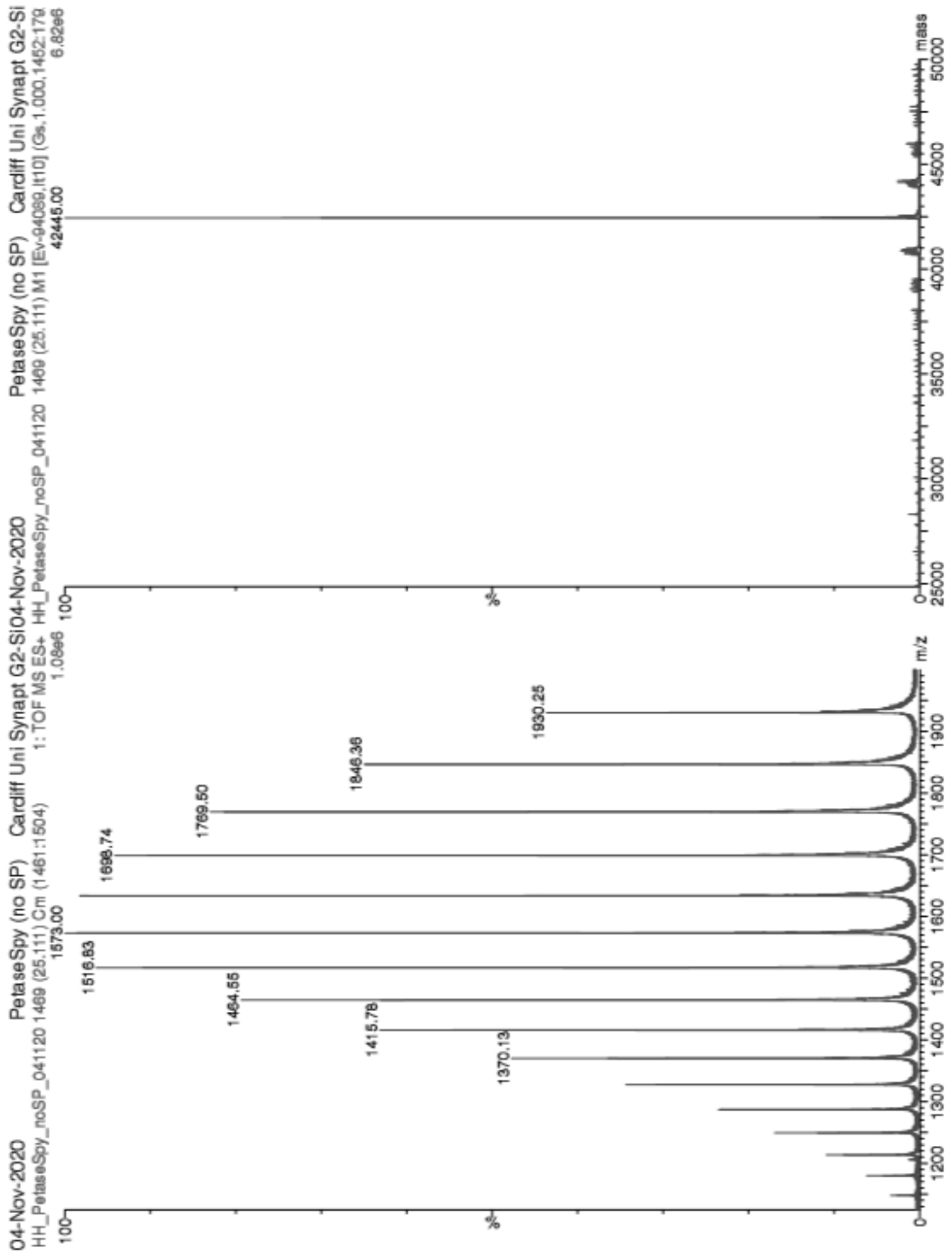


7.3.2 *Is*PETase Protein Mass Spectra

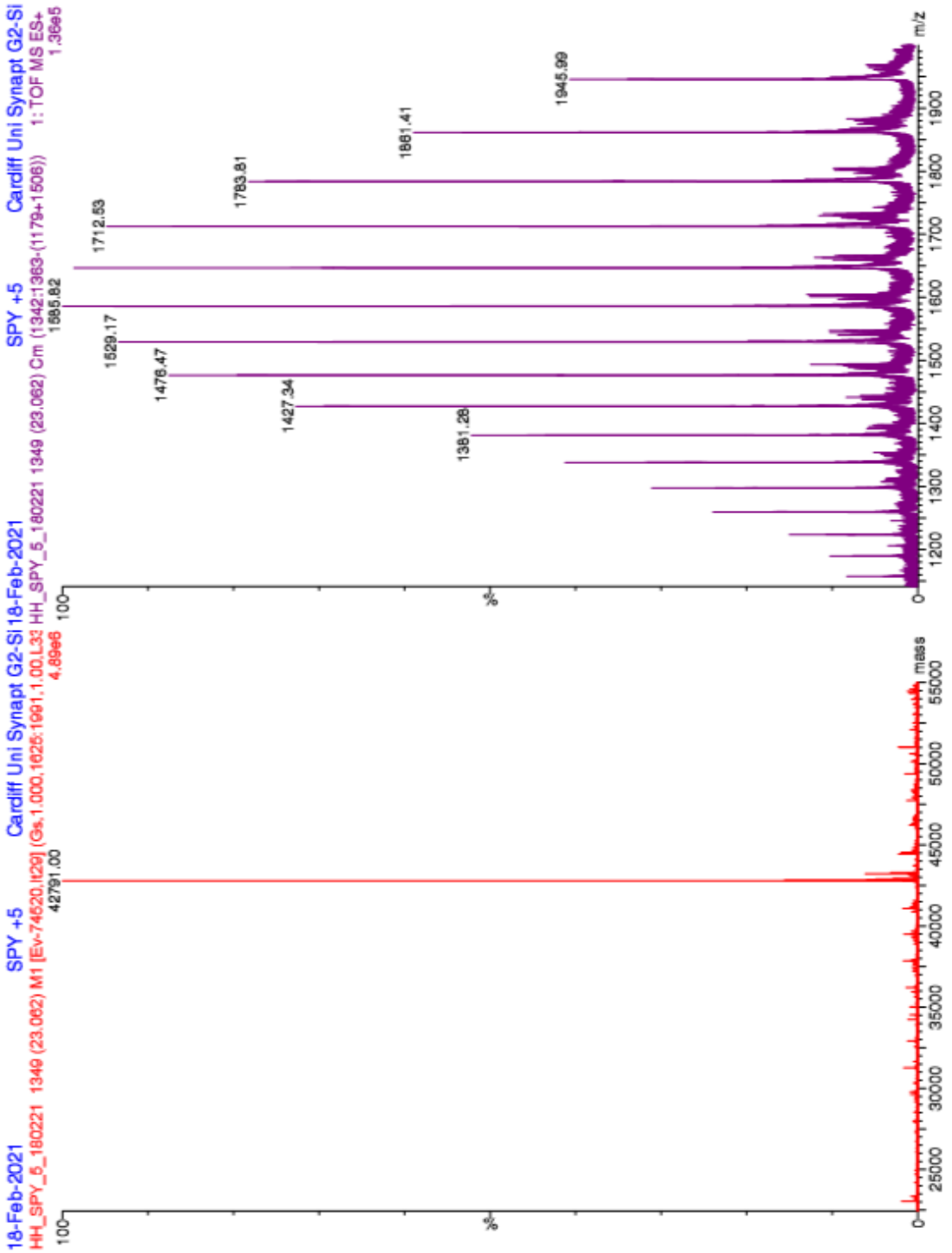
*Is*PETase-WT



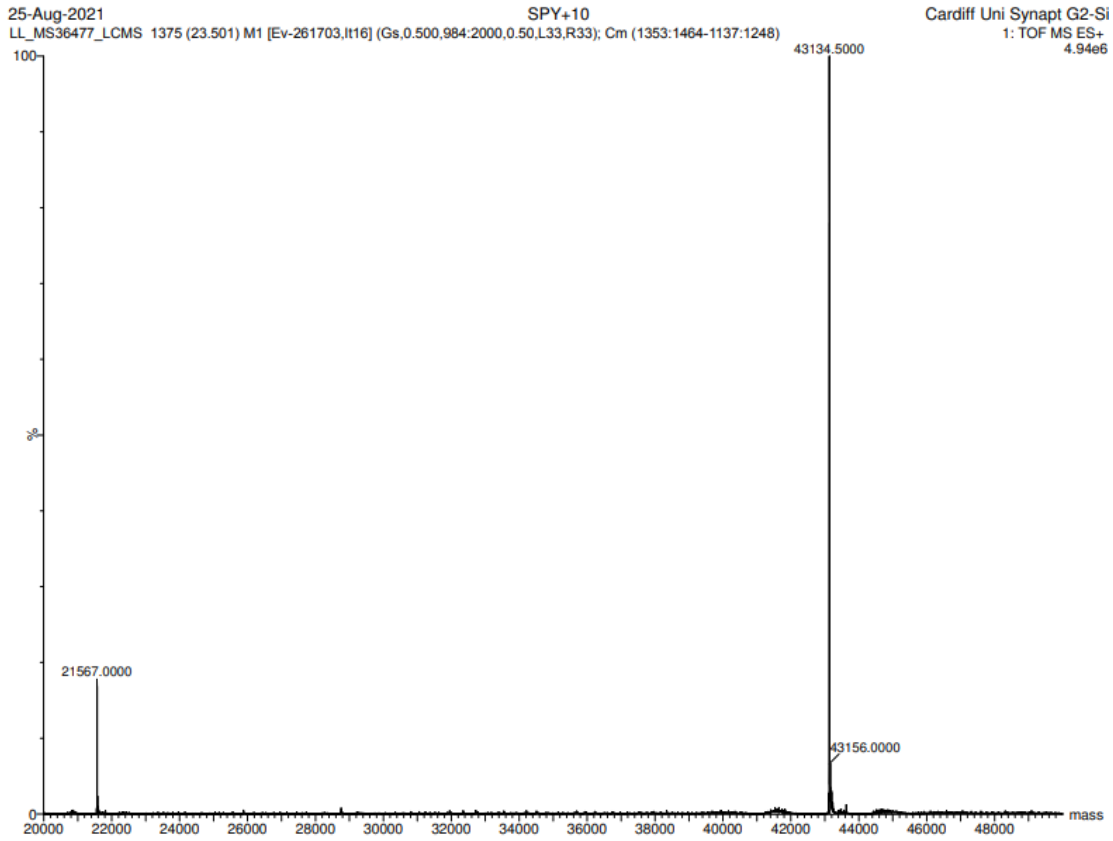
***Is*PETase-Spy**

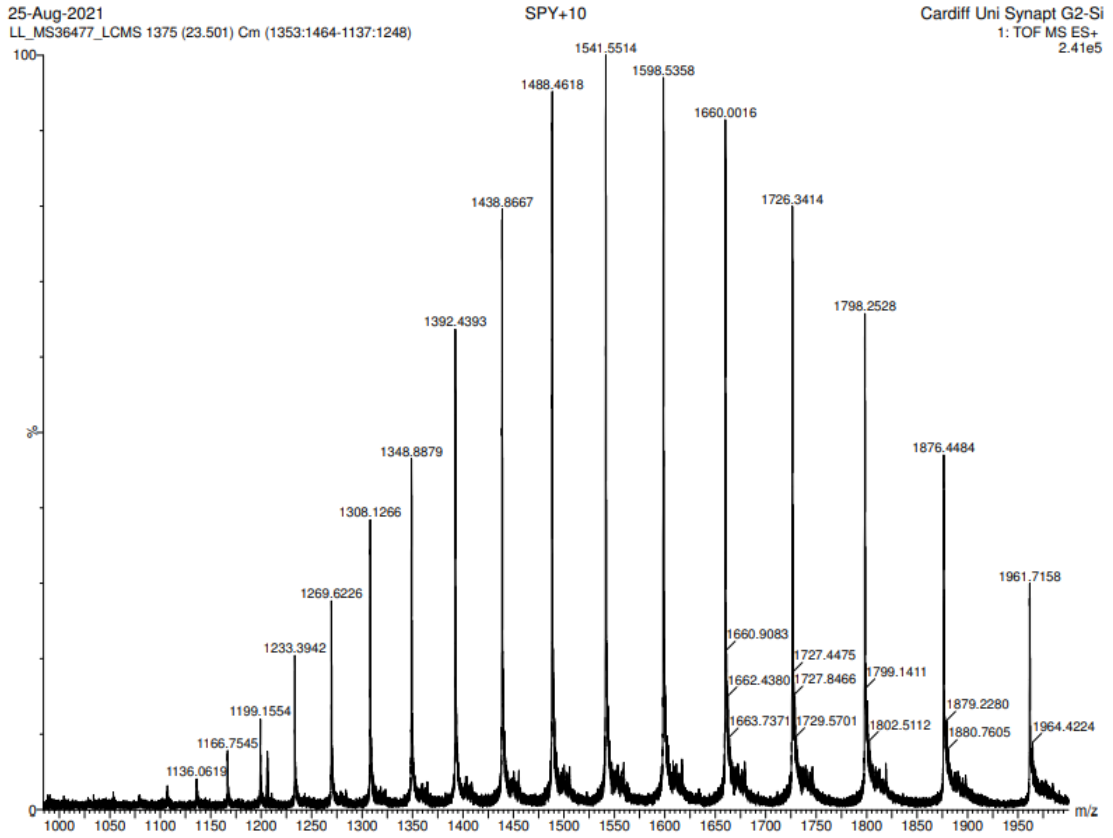


***Is*PETase-Spy^{+5}**

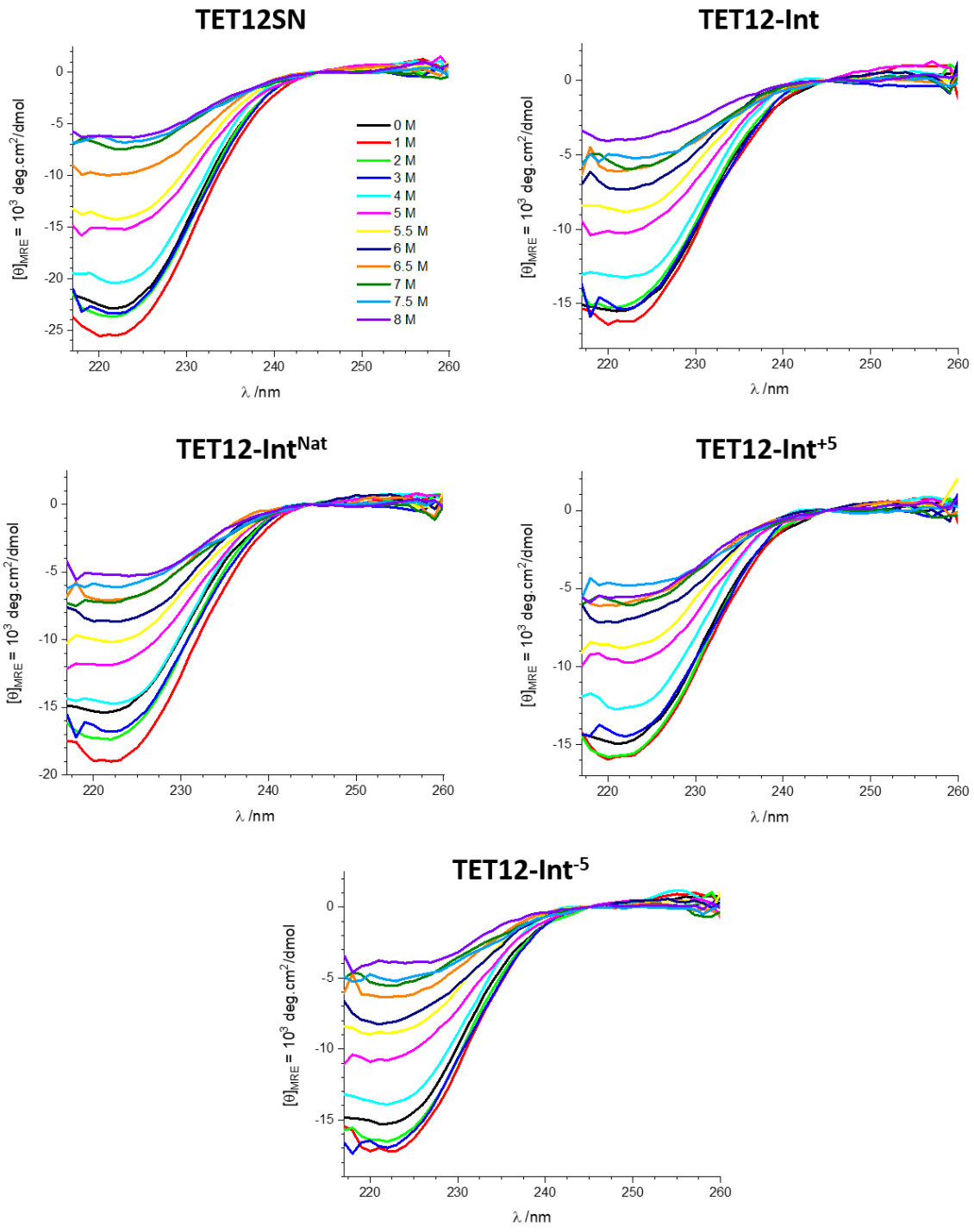


*Is*PETase-Spy^{+10}



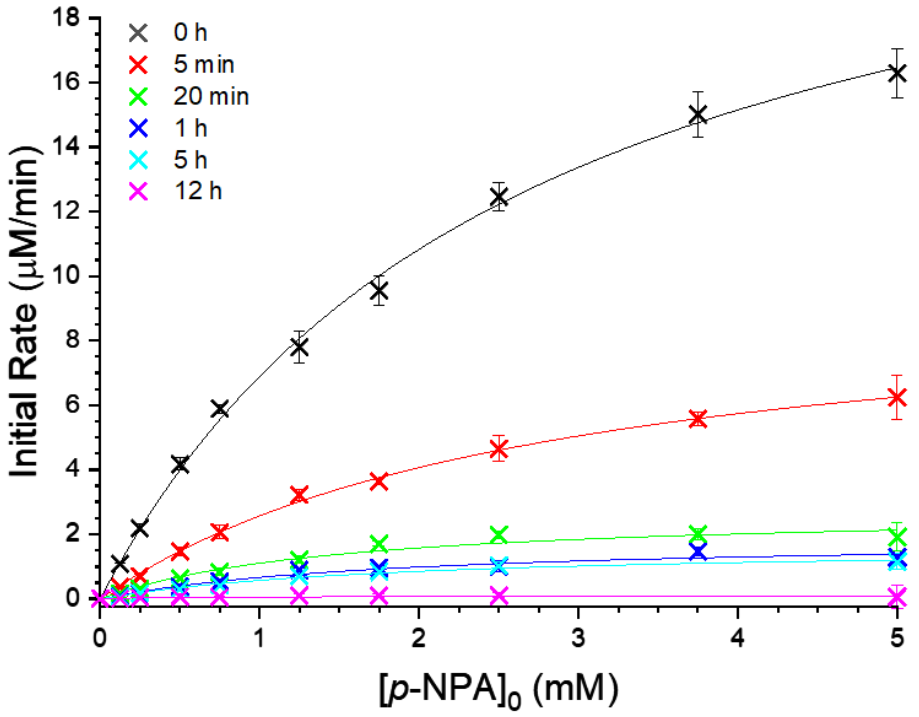


7.4 UREA DENATURATION OF TET12



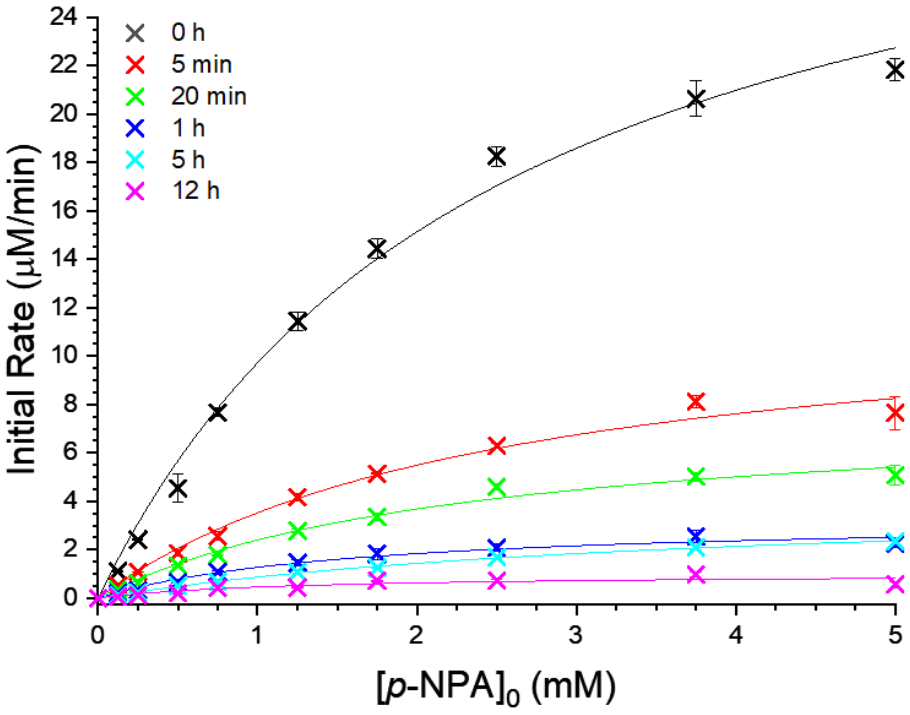
7.5 RESIDUAL ACTIVITY KINETICS

*Is*PETase-WT



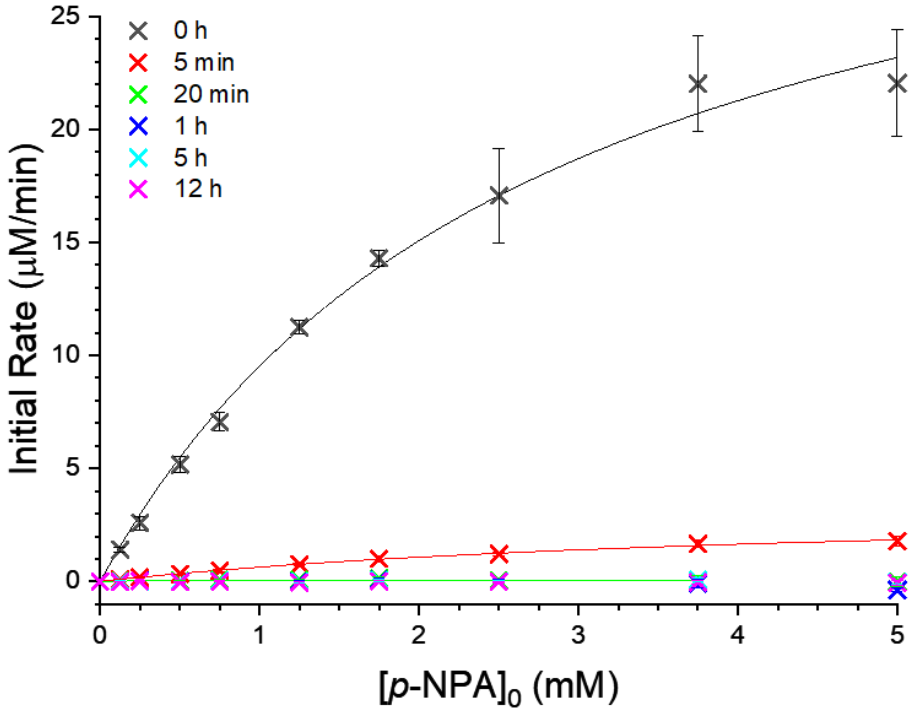
Time at 50°C	V_{max} (μM/min)	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (x10^4 M^{-1} s^{-1})
0 min	25.2 ± 1.6	2.6 ± 0.3	48.5 ± 3.2	1.83 ± 0.24
5 min	9.7 ± 0.6	2.8 ± 0.4	25.2 ± 1.7	0.63 ± 0.08
20 min	2.8 ± 0.5	1.5 ± 0.7	12.6 ± 2.5	0.43 ± 0.19
1 hour	1.9 ± 0.4	1.8 ± 0.9	4.9 ± 1.1	0.15 ± 0.07
5 hours	1.7 ± 0.3	1.8 ± 0.6	6.9 ± 1.1	0.08 ± 0.07
12 hours	0.1 ± 0.0	0.3 ± 0.4	1.2 ± 0.4	0.04 ± 0.02

*Is*PETase-Spy



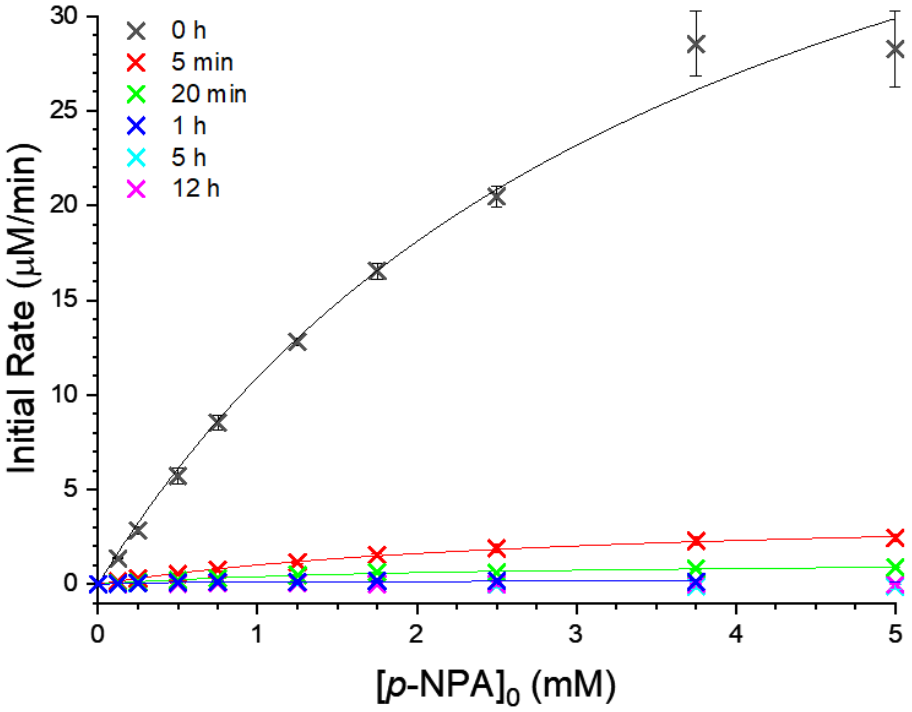
Time at 50°C	V_{max} (μM/min)	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (x10^4 M^{-1} s^{-1})
0 min	34.2 ± 4.4	2.5 ± 0.7	51.3 ± 6.5	2.04 ± 0.79
5 min	12.4 ± 2.0	2.5 ± 0.8	25.2 ± 4.1	1.00 ± 0.49
20 min	7.9 ± 1.2	2.3 ± 0.7	12.6 ± 1.9	0.56 ± 0.26
1 hour	3.3 ± 0.5	1.6 ± 0.6	4.9 ± 0.7	0.31 ± 0.16
5 hours	4.1 ± 0.3	3.8 ± 0.5	6.9 ± 0.5	0.18 ± 0.04
12 hours	1.0 ± 0.4	1.2 ± 1.3	1.2 ± 0.5	0.10 ± 0.14

*Is*PETase-Dimer



Time at 50°C	V_{max} (µM/min)	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (x10^4 M^{-1} s^{-1})
0 min	36.1 ± 4.4	2.8 ± 0.7	56.7 ± 7.0	2.03 ± 0.75
5 min	3.5 ± 0.5	4.5 ± 1.0	6.4 ± 0.8	0.14 ± 0.05
20 min	0.1 ± 0.0	0.6 ± 0.6	0.2 ± 0.1	0.04 ± 0.07
1 hour	-	-	-	-
5 hours	-	-	-	-
12 hours	-	-	-	-

***Is*PETase-Cat^{+5}**

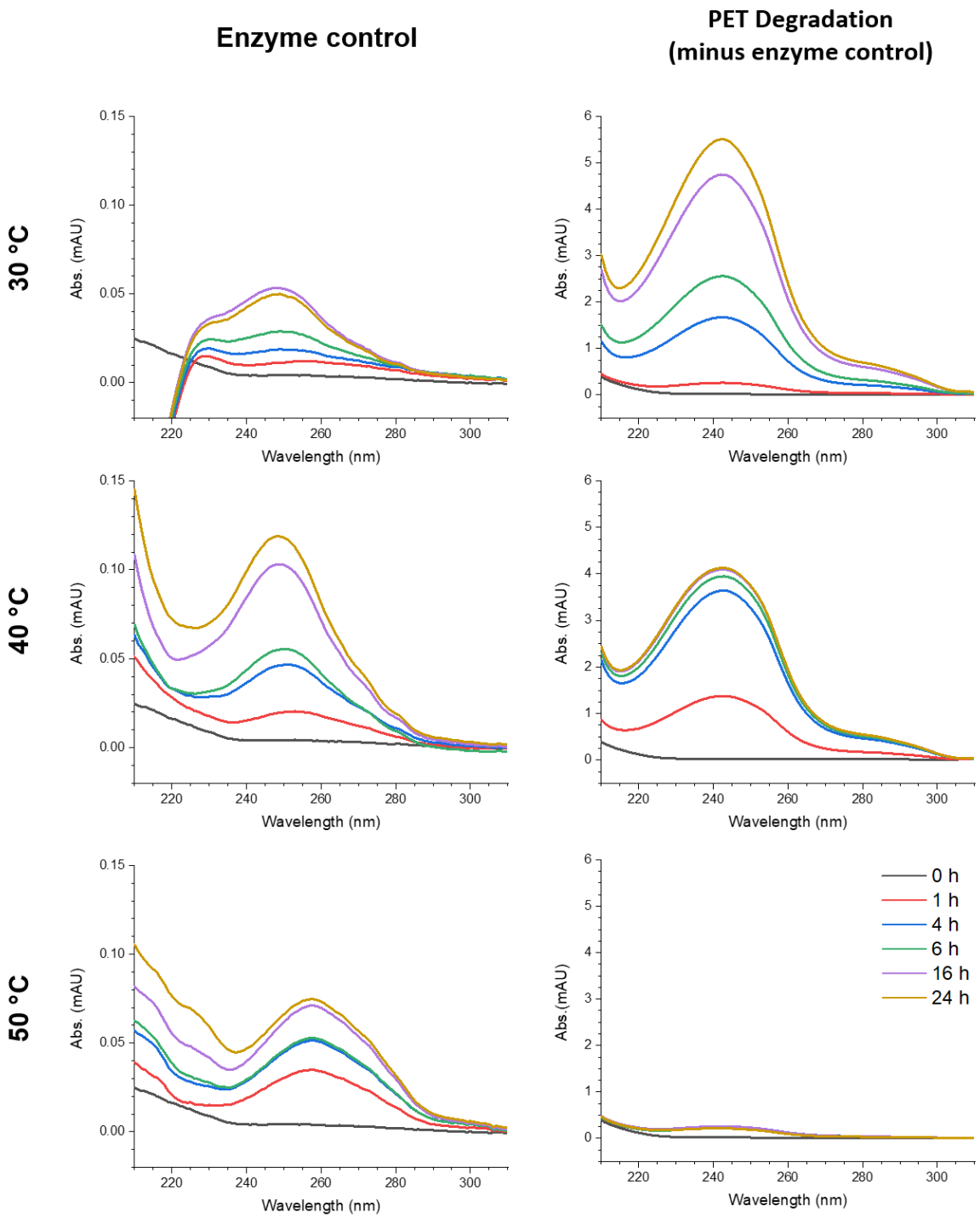


Time at 50°C	V_{max} (μM/min)	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (x10^4 M^{-1} s^{-1})
0 min	52.7 ± 9.9	3.8 ± 1.3	68.0 ± 12.8	1.78 ± 0.93
5 min	4.0 ± 0.3	2.9 ± 0.5	6.4 ± 0.5	0.22 ± 0.05
20 min	1.3 ± 1.2	2.2 ± 0.7	2.2 ± 0.3	0.10 ± 0.05
1 hour	0.2 ± 0.0	0.5 ± 0.2	0.3 ± 0.1	0.07 ± 0.04
5 hours	-	-	-	-
12 hours	-	-	-	-

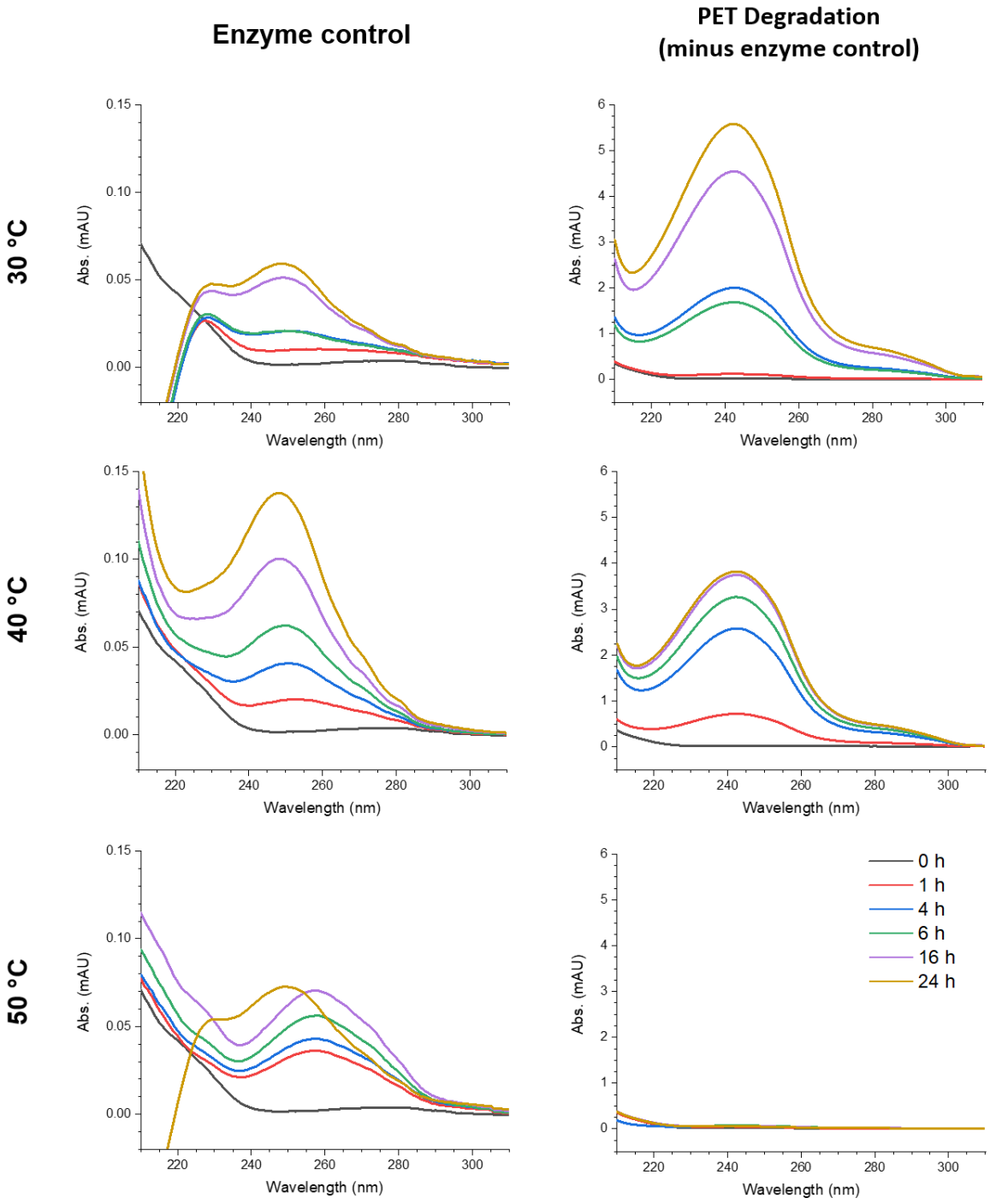
7.6 PET DEGRADATION ABSORBANCE SPECTRA

7.6.1 PET Degradation at 0 rpm

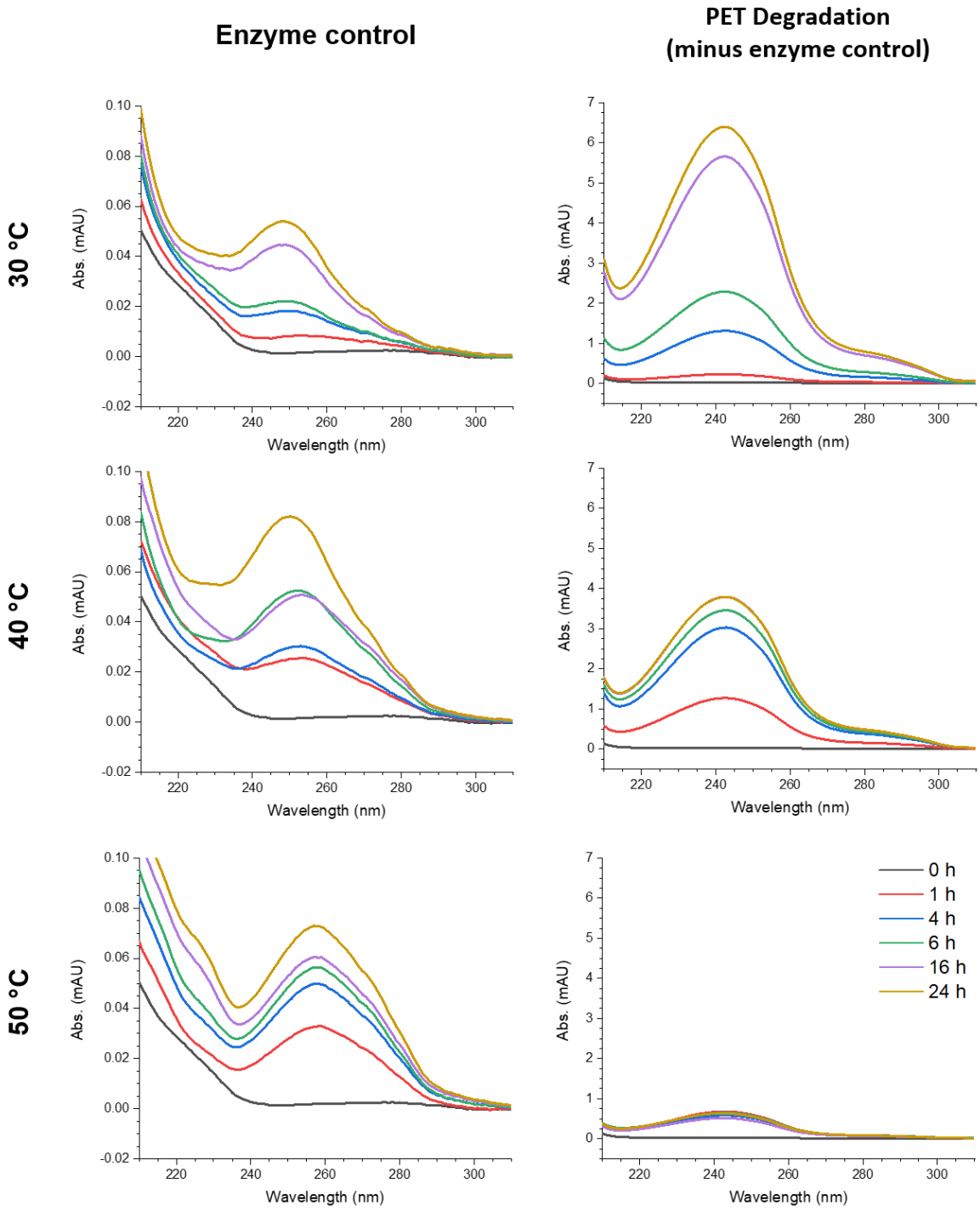
*Is*PETase-WT (0 rpm)



_Is_PETase-Spy (0 rpm)



***Is*PETase-Dimer (0 rpm)**



_Is_PETase-Cat^{+5} (0 rpm)

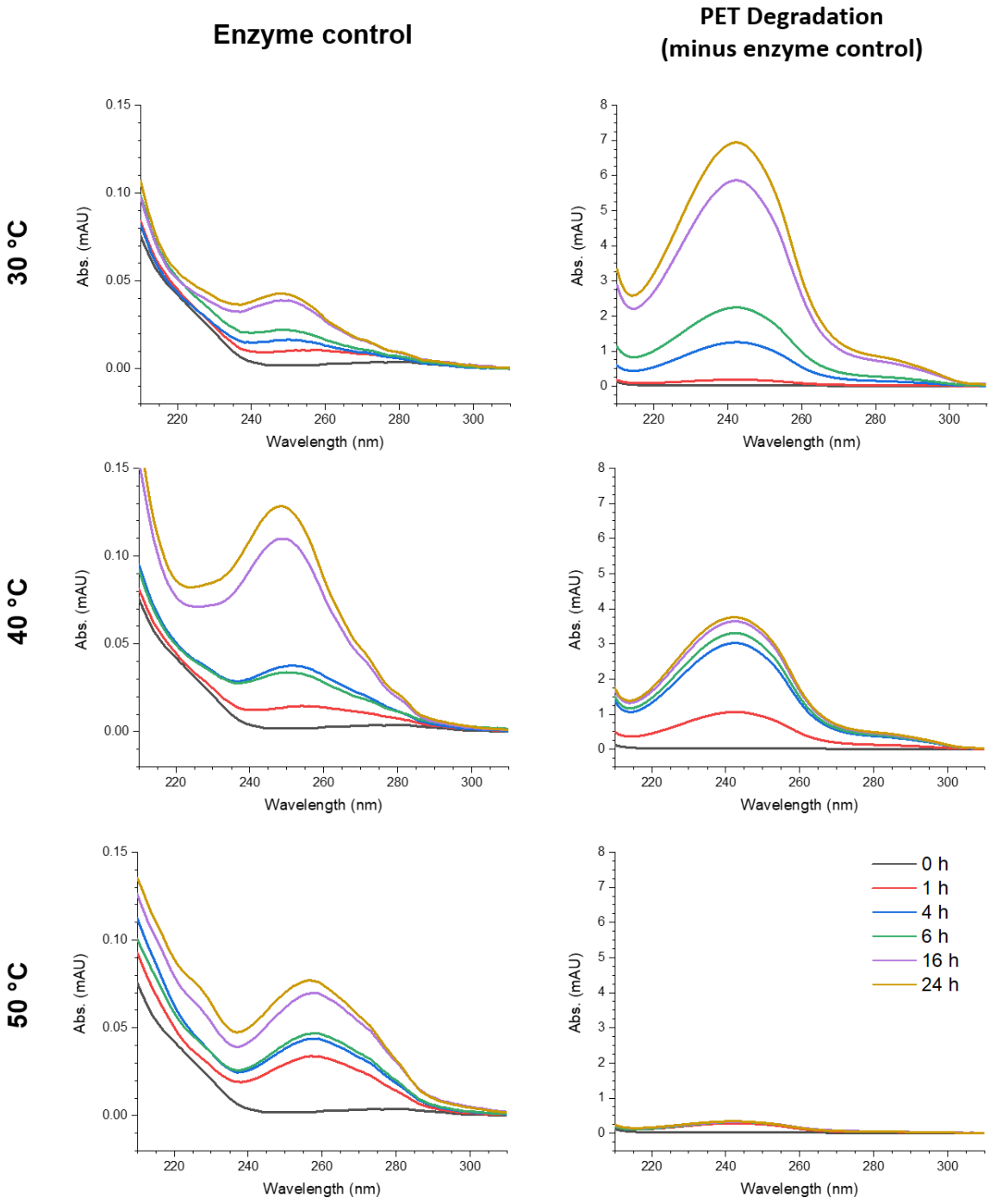


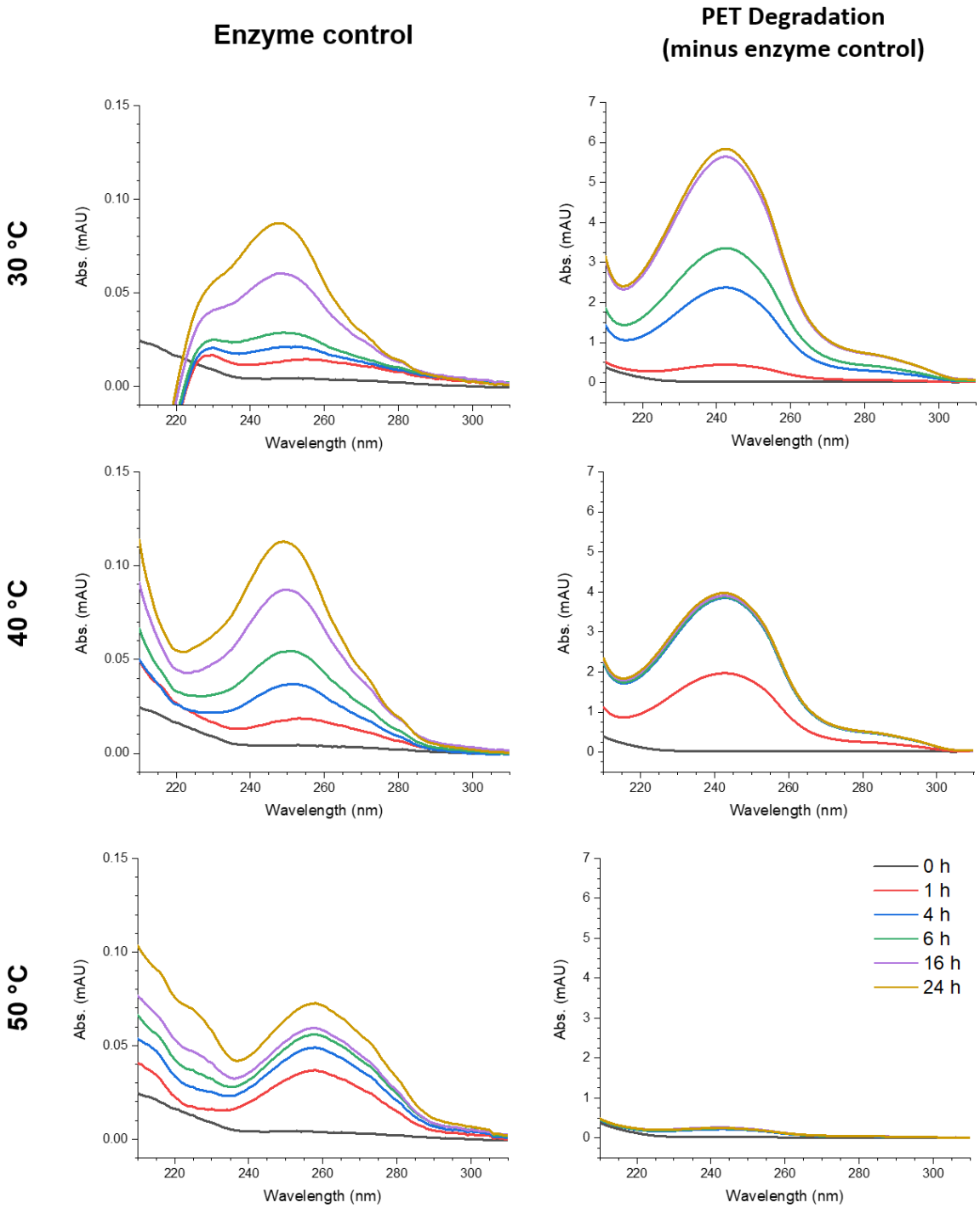
Table of Absorbance Values Measured at 240 nm (0 rpm)

T /°C	*Is*PETase variant	Time / h					
		0	1	4	6	16	24
30	-WT	0.02 ± 0.00	0.26 ± 0.01	1.65 ± 0.07	2.53 ± 0.04	4.69 ± 0.10	5.44 ± 0.22
	-Spy	0.02 ± 0.00	0.12 ± 0.01	1.98 ± 0.16	1.67 ± 0.11	4.50 ± 0.14	5.52 ± 0.17
	-Dimer	0.02 ± 0.00	0.23 ± 0.02	1.30 ± 0.05	2.26 ± 0.06	5.60 ± 0.05	6.34 ± 0.09
	-Cat^{+5}	0.02 ± 0.00	0.19 ± 0.02	1.24 ± 0.06	2.22 ± 0.20	5.80 ± 0.15	6.87 ± 0.19
	-Spy^{D7A}	-	-	-	-	-	2.63 ± 0.09
40	-WT	0.02 ± 0.00	1.36 ± 0.04	3.60 ± 0.08	3.91 ± 0.07	4.06 ± 0.10	4.10 ± 0.10
	-Spy	0.02 ± 0.00	0.71 ± 0.05	2.55 ± 0.07	3.23 ± 0.09	3.71 ± 0.10	3.79 ± 0.12
	-Dimer	0.02 ± 0.00	1.26 ± 0.03	3.00 ± 0.10	3.42 ± 0.06	3.75 ± 0.05	3.75 ± 0.13
	-Cat^{+5}	0.02 ± 0.00	1.05 ± 0.04	3.00 ± 0.06	3.27 ± 0.11	3.60 ± 0.15	3.72 ± 0.14
	-Spy^{D7A}	-	-	-	-	-	1.75 ± 0.08
50	-WT	0.02 ± 0.00	0.23 ± 0.02	0.25 ± 0.03	0.22 ± 0.04	0.25 ± 0.05	0.22 ± 0.05
	-Spy	0.02 ± 0.00	0.05 ± 0.02	0.07 ± 0.03	0.07 ± 0.03	0.06 ± 0.04	0.06 ± 0.06
	-Dimer	0.02 ± 0.00	0.67 ± 0.04	0.58 ± 0.03	0.64 ± 0.05	0.51 ± 0.08	0.63 ± 0.07
	-Cat^{+5}	0.02 ± 0.00	0.29 ± 0.03	0.32 ± 0.03	0.32 ± 0.05	0.34 ± 0.06	0.34 ± 0.06
	-Spy^{D7A}	-	-	-	-	-	0.09 ± 0.04

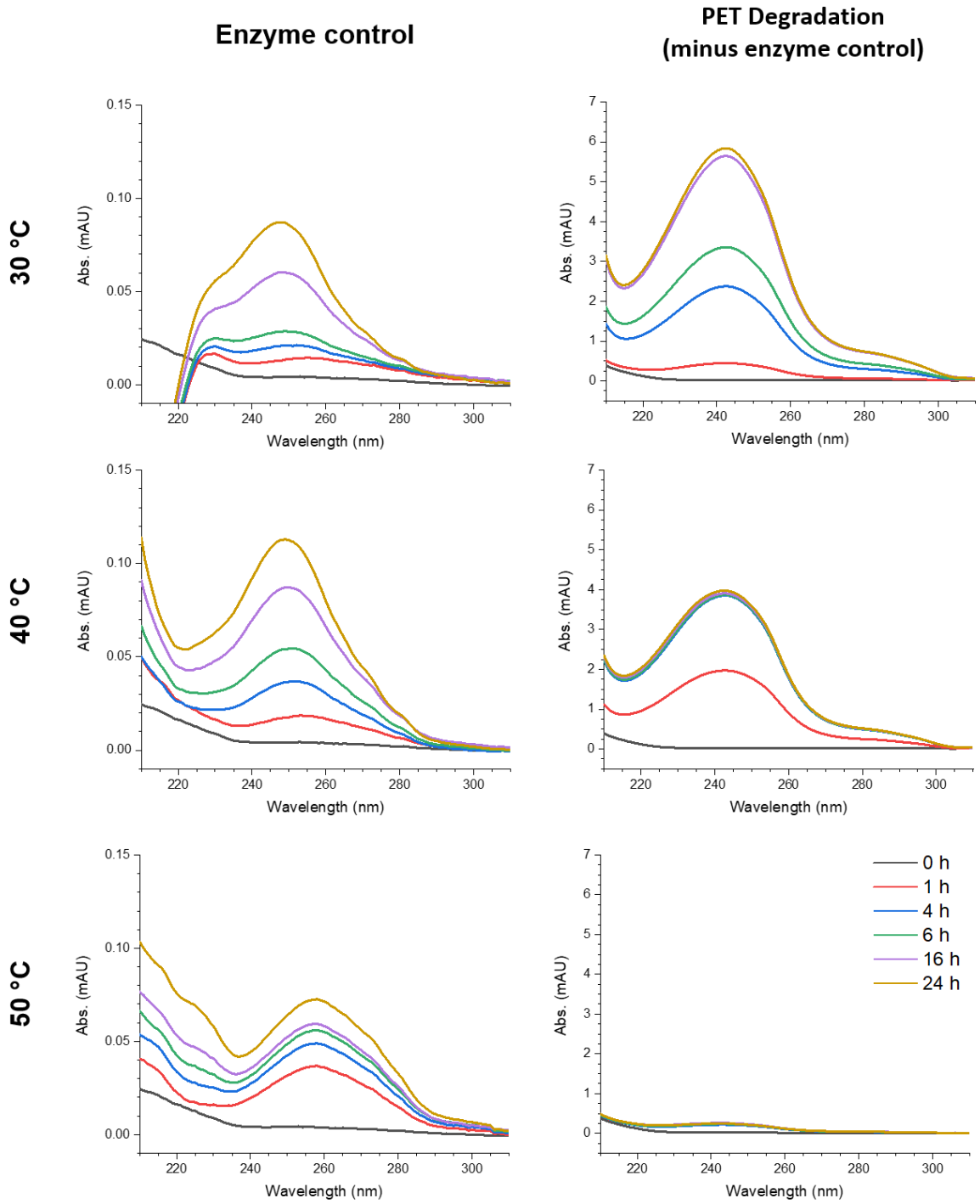
(Values given are those after averaging and the subtraction of the A_{240} values obtained from the enzyme only controls.)

7.6.2 PET Degradation at 550 rpm

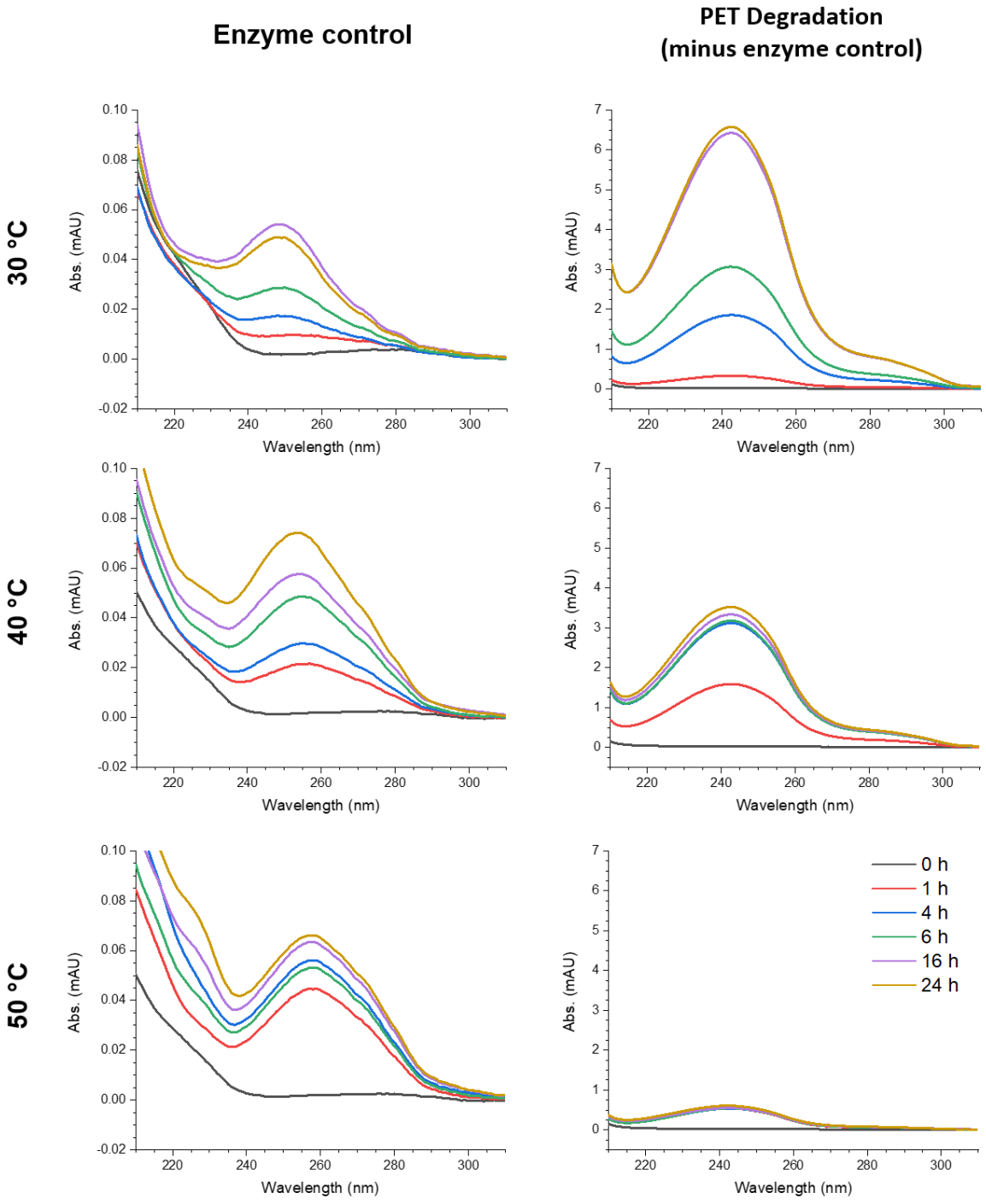
*Is*PETase-WT (550 rpm)



***Is*PETase-Spy (550 rpm)**



*Is*PETase-Dimer (550 rpm)



***Is*PETase-Cat^{+5} (550 rpm)**

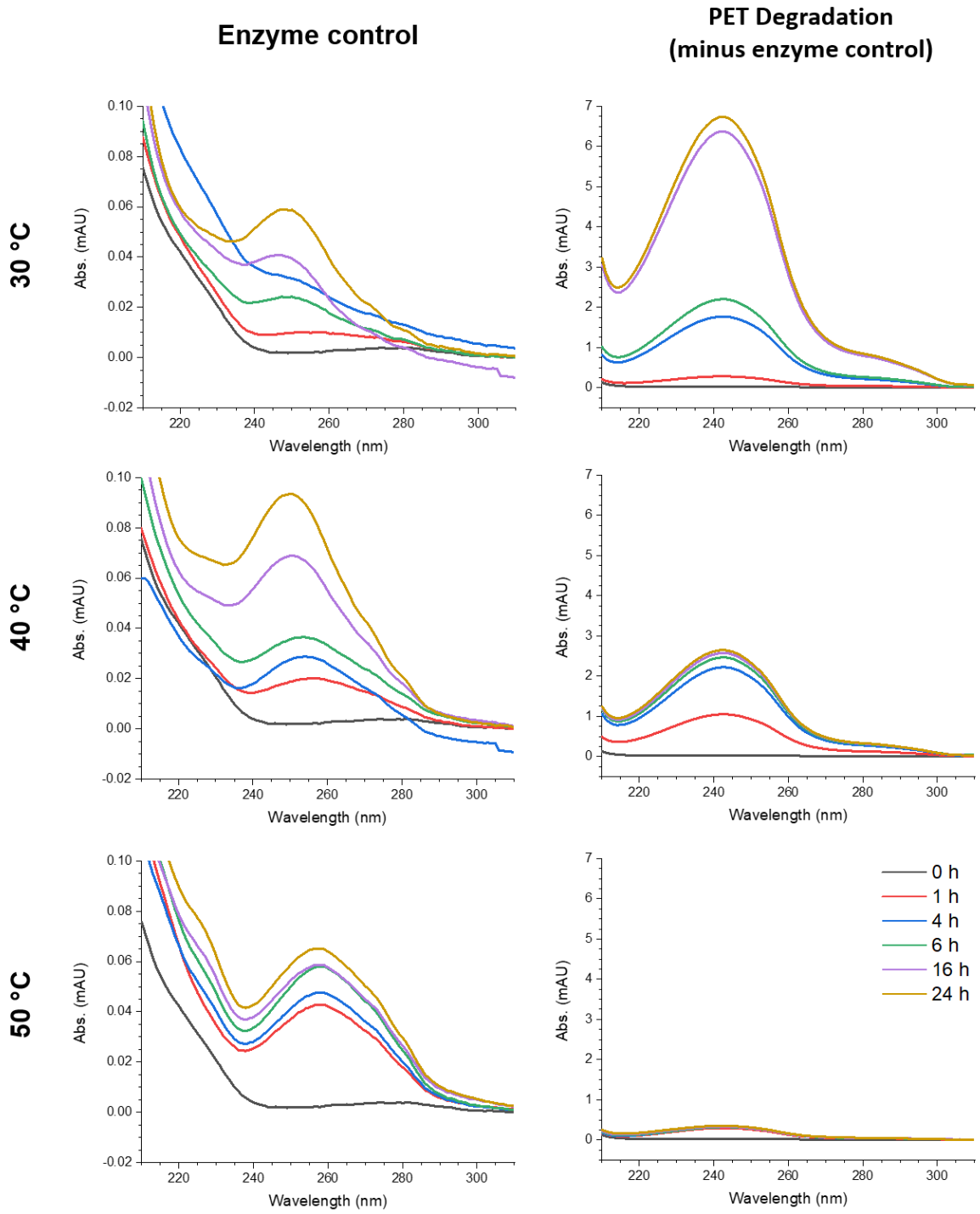


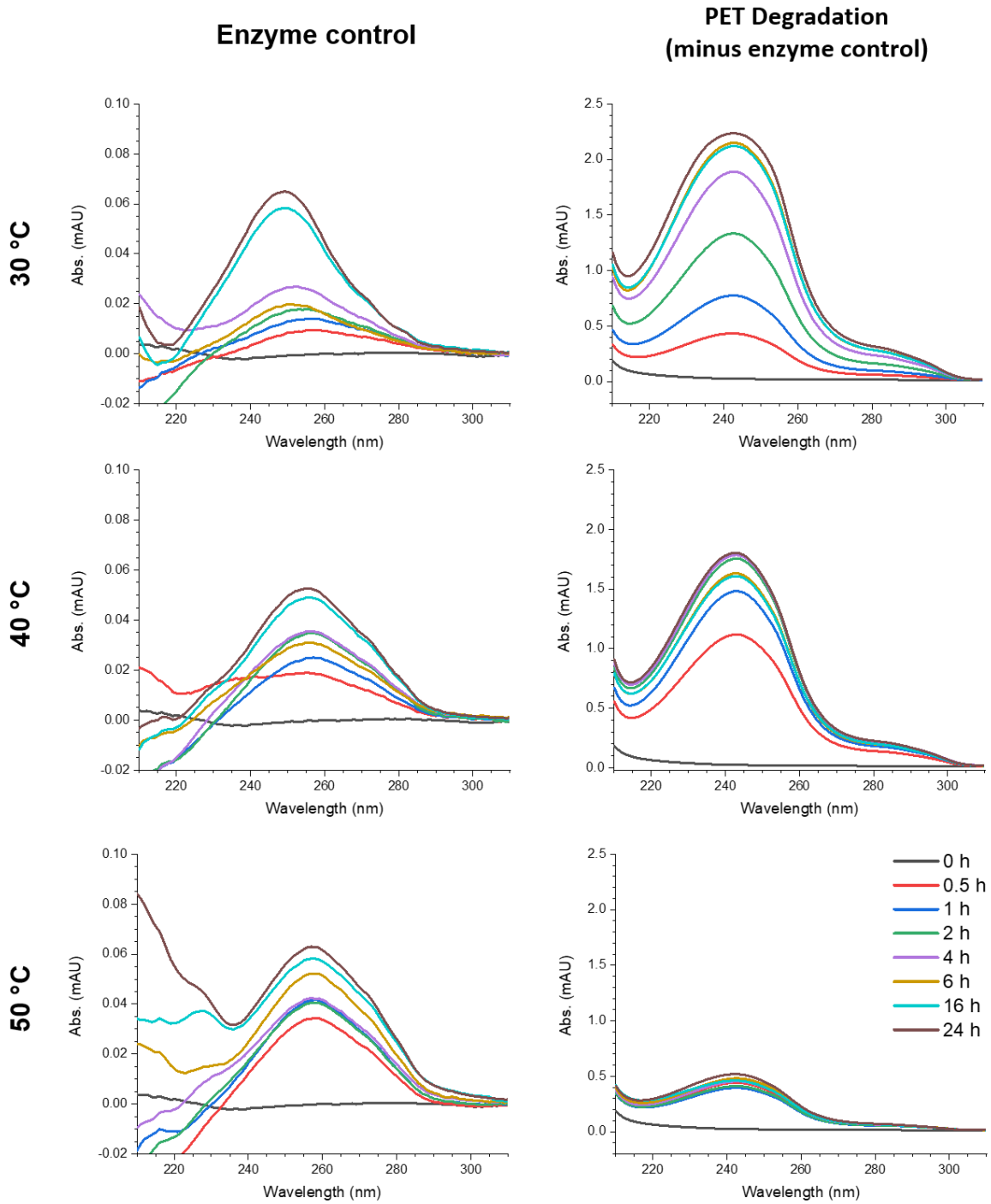
Table of Absorbance Values Measured at 240 nm (550 rpm)

T /°C	*Is*PETase variant	Time / h					
		0	1	4	6	16	24
30	-WT	0.02 ± 0.00	0.44 ± 0.02	2.35 ± 0.06	3.32 ± 0.14	5.58 ± 0.09	5.77 ± 0.18
	-Spy	0.02 ± 0.00	0.21 ± 0.02	1.35 ± 0.08	2.11 ± 0.07	5.04 ± 0.13	5.45 ± 0.11
	-Dimer	0.02 ± 0.00	0.33 ± 0.01	1.84 ± 0.02	3.04 ± 0.19	6.36 ± 0.32	6.51 ± 0.10
	-Cat^{+5}	0.02 ± 0.00	0.28 ± 0.02	1.75 ± 0.04	2.18 ± 0.66	6.31 ± 0.07	6.67 ± 0.08
40	-WT	0.02 ± 0.00	1.95 ± 0.07	3.81 ± 0.11	3.83 ± 0.08	3.87 ± 0.12	3.94 ± 0.09
	-Spy	0.02 ± 0.00	0.86 ± 0.03	2.53 ± 0.04	2.82 ± 0.10	3.05 ± 0.12	2.93 ± 0.18
	-Dimer	0.02 ± 0.00	1.57 ± 0.05	3.08 ± 0.10	3.14 ± 0.05	3.30 ± 0.07	3.49 ± 0.08
	-Cat^{+5}	0.02 ± 0.00	1.05 ± 0.05	2.21 ± 0.09	2.45 ± 0.03	2.57 ± 0.06	2.67 ± 0.09
50	-WT	0.02 ± 0.00	0.26 ± 0.02	0.22 ± 0.03	0.24 ± 0.03	0.26 ± 0.04	0.25 ± 0.05
	-Spy	0.02 ± 0.00	0.05 ± 0.02	0.08 ± 0.03	0.06 ± 0.03	0.07 ± 0.04	0.06 ± 0.05
	-Dimer	0.02 ± 0.00	0.53 ± 0.06	0.53 ± 0.04	0.54 ± 0.05	0.55 ± 0.07	0.60 ± 0.05
	-Cat^{+5}	0.02 ± 0.00	0.29 ± 0.03	0.32 ± 0.03	0.32 ± 0.04	0.33 ± 0.04	0.35 ± 0.05

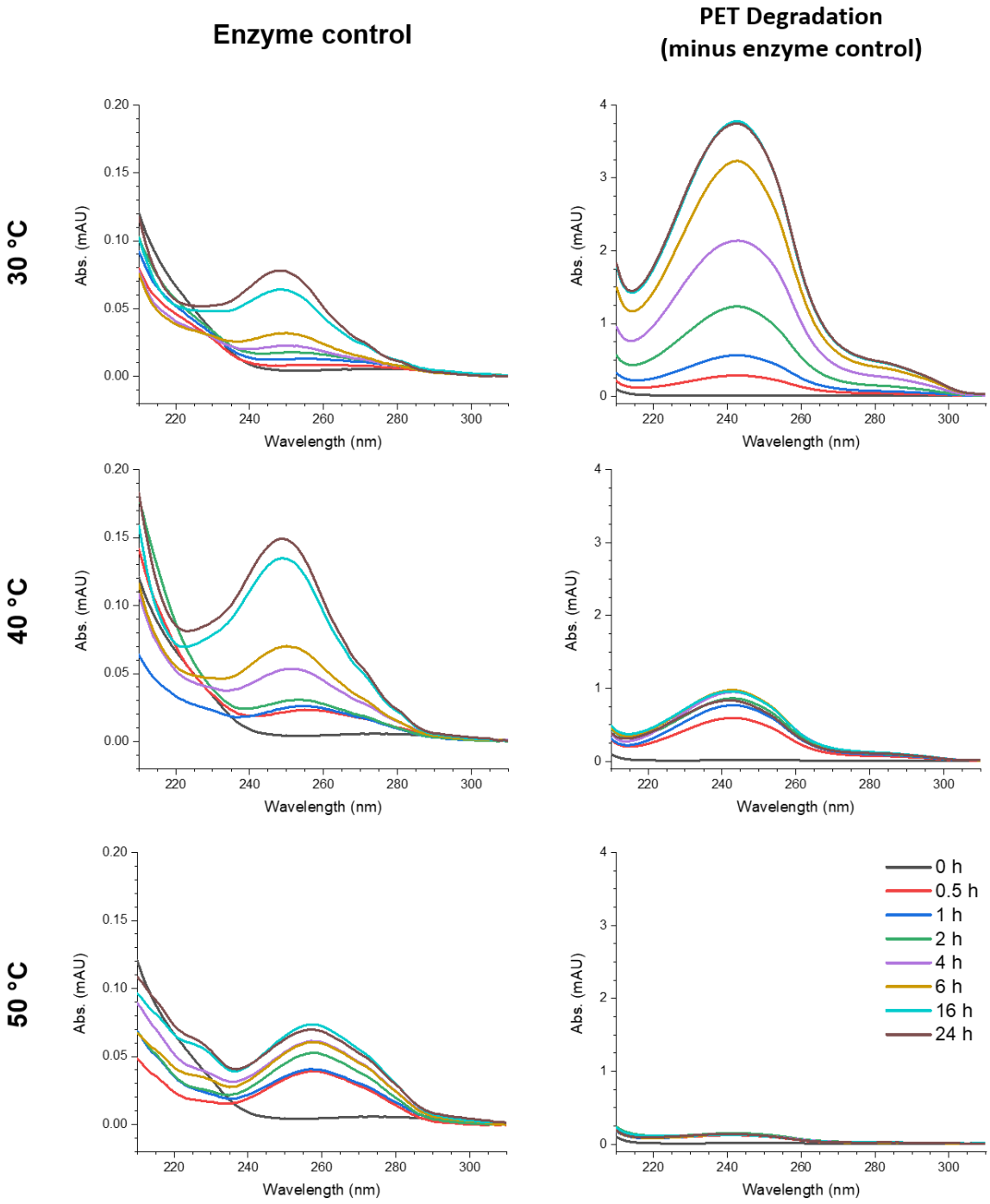
(Values given are those after averaging and the subtraction of the A_{240} values obtained from the enzyme only controls.)

7.6.3 PET Degradation at 1100 rpm

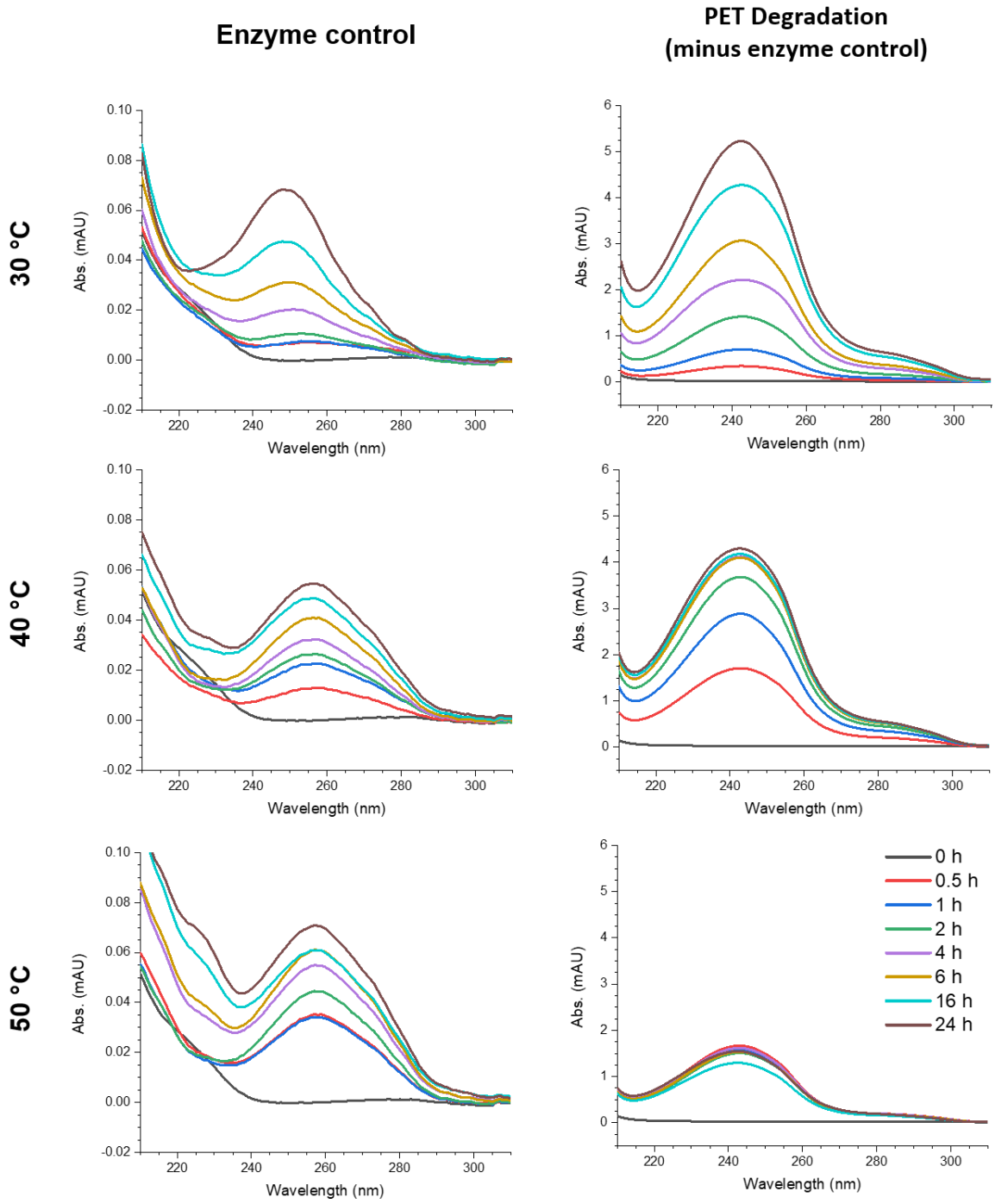
*Is*PETase-WT (1100 rpm)



***Is*PETase-Spy (1100 rpm)**



***Is*PETase-Dimer (1100 rpm)**



_Is_PETase-Cat^{+5} (1100 rpm)

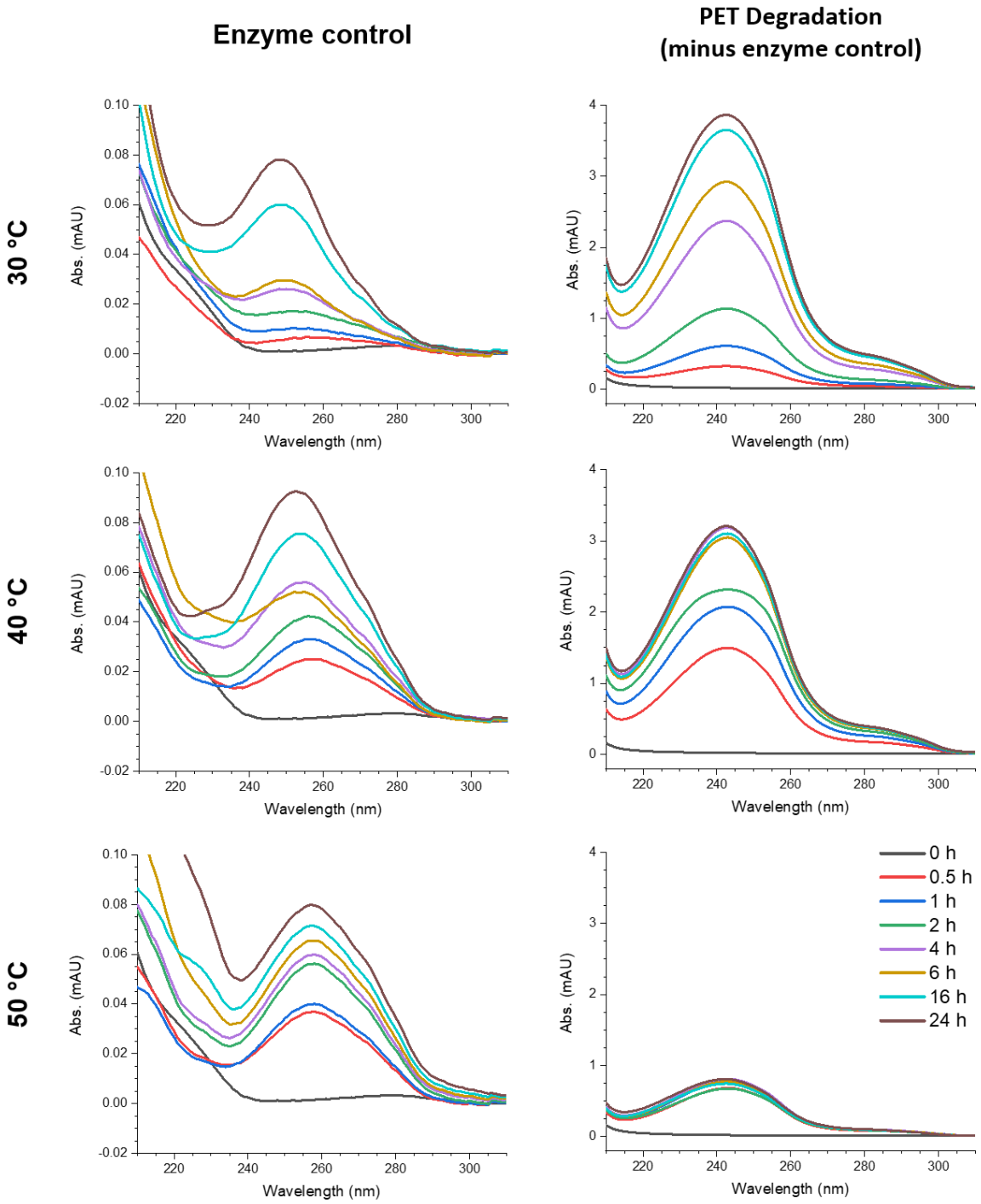
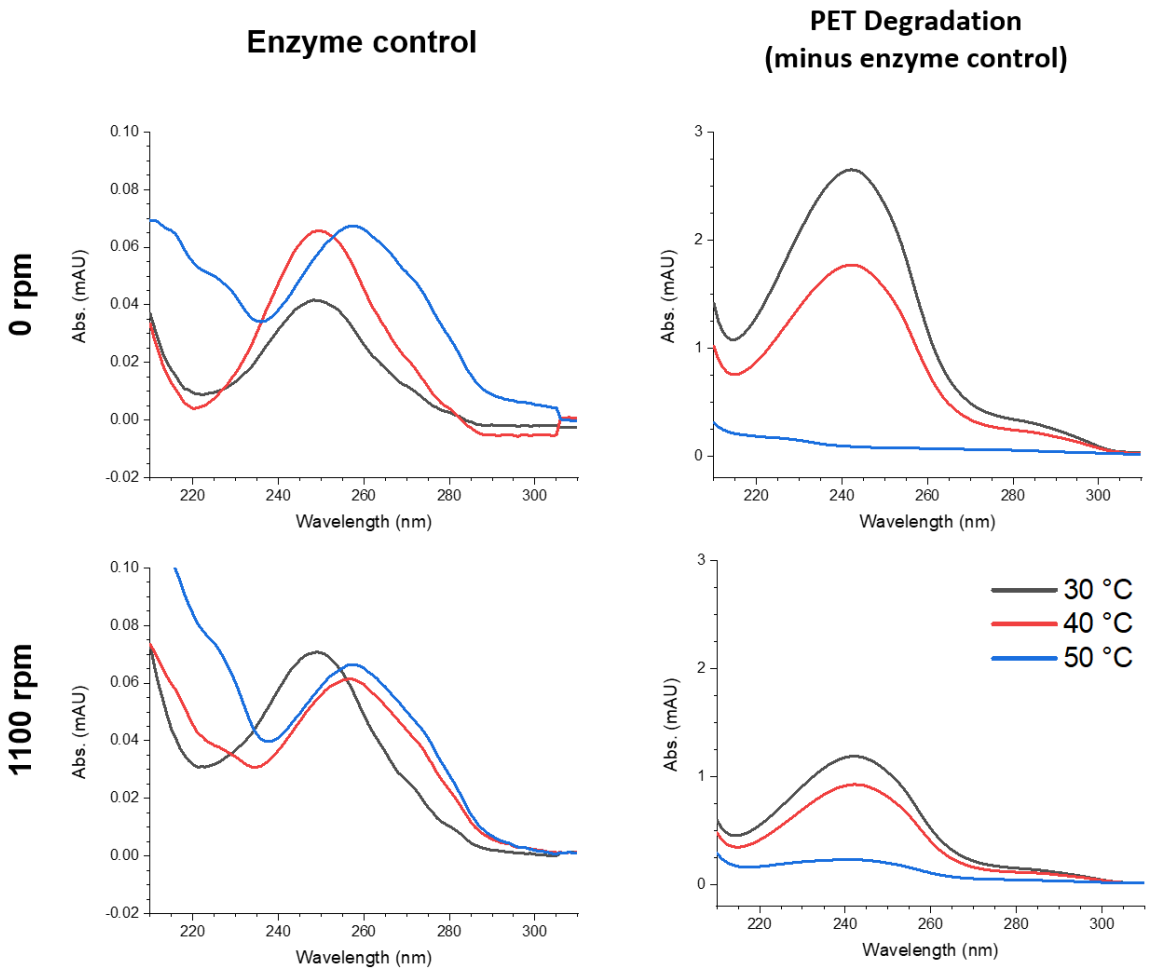


Table of Absorbance Values Measured at 240 nm (1100 rpm)

T /°C	*Is*PETase variant	Time / h							
		0	**0.5**	**1**	**2**	**4**	**6**	**16**	**24**
30	-WT	0.03 ± 0.00	0.43 ± 0.01	0.76 ± 0.01	1.31 ± 0.04	1.87 ± 0.05	2.13 ± 0.09	2.10 ± 0.06	2.22 ± 0.06
	-Spy	0.02 ± 0.01	0.28 ± 0.01	0.56 ± 0.02	1.22 ± 0.06	2.11 ± 0.04	3.19 ± 0.04	3.74 ± 0.12	3.71 ± 0.48
	-Dimer	0.03 ± 0.00	0.35 ± 0.03	0.70 ± 0.05	1.40 ± 0.03	2.20 ± 0.04	3.03 ± 0.06	4.23 ± 0.05	5.17 ± 0.16
	-Cat^{+5}	0.02 ± 0.00	0.32 ± 0.01	0.60 ± 0.02	1.12 ± 0.08	2.34 ± 0.13	2.88 ± 0.08	3.61 ± 0.18	3.83 ± 0.07
	-Spy^{D7A}	-	-	-	-	-	-	-	1.18 ± 0.06
40	-WT	0.03 ± 0.00	1.10 ± 0.06	1.45 ± 0.23	1.73 ± 0.02	1.76 ± 0.06	1.61 ± 0.06	1.58 ± 0.03	1.78 ± 0.08
	-Spy	0.02 ± 0.01	0.59 ± 0.03	0.76 ± 0.04	0.85 ± 0.07	0.93 ± 0.05	0.96 ± 0.07	0.95 ± 0.12	0.83 ± 0.13
	-Dimer	0.03 ± 0.00	1.68 ± 0.06	2.84 ± 0.07	3.63 ± 0.11	4.07 ± 0.10	4.05 ± 0.08	4.13 ± 0.04	4.25 ± 0.05
	-Cat^{+5}	0.02 ± 0.00	1.47 ± 0.08	2.05 ± 0.04	2.30 ± 0.03	3.14 ± 0.09	3.00 ± 0.11	3.06 ± 0.13	3.16 ± 0.18
	-Spy^{D7A}	-	-	-	-	-	-	-	0.91 ± 0.06
50	-WT	0.03 ± 0.00	0.44 ± 0.02	0.39 ± 0.02	0.41 ± 0.02	0.45 ± 0.03	0.47 ± 0.05	0.46 ± 0.03	0.51 ± 0.04
	-Spy	0.02 ± 0.01	0.13 ± 0.02	0.14 ± 0.02	0.15 ± 0.03	0.14 ± 0.04	0.14 ± 0.03	0.13 ± 0.04	0.14 ± 0.04
	-Dimer	0.03 ± 0.00	1.64 ± 0.02	1.54 ± 0.09	1.48 ± 0.18	1.59 ± 0.05	1.51 ± 0.09	1.28 ± 0.17	1.53 ± 0.09
	-Cat^{+5}	0.02 ± 0.00	0.67 ± 0.08	0.76 ± 0.06	0.66 ± 0.05	0.80 ± 0.09	0.76 ± 0.07	0.74 ± 0.04	0.80 ± 0.06
	-Spy^{D7A}	-	-	-	-	-	-	-	0.23 ± 0.05

(Values given are those after averaging and the subtraction of the A_{240} values obtained from the enzyme only controls.)

7.6.4 *Is*PETase-Spy^{D7A} PET Degradation (24 h)



7.6.5 PET Only Controls (24 h)