THE UNIVERSITY *of* EDINBURGH

Creation of an artificial Stetterase through the design, synthesis and installation of an organocatalyst into a protein scaffold.

Alice MacAulay

A Thesis Submitted for the Degree of Doctor of Philosophy

The University of Edinburgh

2023

Lay Summary

Traditional methods to make the chemicals which we use in everyday household products, cosmetics and pharmaceuticals often use environmentally damaging solvents, reagents and harsh conditions. The use of enzymes offer an alternative route to synthesise these chemicals. Enzymes occur naturally in all living things, and function in benign condition for example most of the enzymes in our bodies function at 37 °C in aqueous solutions. Enzymes perform very selective reactions to give rise to specific molecules. However, the enzymes found in nature are not able to naturally synthesise all the complex chemicals used in modern day applications.

We can manipulate the structure of enzymes by altering the DNA which encodes them. This allows us to tune them to perform the reactions we require and change the molecules they accept. It is more difficult to add new reactivities to enzymes, especially to perform reactions which do not usually occur in nature. Adding additional chemical modifications to enzymes can help then perform new reactions. In this thesis we create a chemically modified “artificial enzyme” which can perform a specific carbon- carbon bond forming reaction under benign conditions.

Abstract

The application of biocatalysis in industrial synthesis continues to rise at pace, driven by a demand for sustainable synthetic methods. As a result, chemo-enzymatic cascades which merge the use of chemocatalysis and biocatalysis are of growing interest. Whilst the majority of enzymatic reactions take place in water, chemocatalysis is typically performed in organic solvents making solvent-compatibility an issue. One option to overcome this is the use of protein hosted organocatalytic reactions. Proteins are inherently compatible with aqueous solvents, but their internal environment contains hydrophobic pockets and functional groups suited to organocatalysis. Progress in protein engineering has enabled us to design, evolve and select structures which can incorporate and exploit the reactivity of non-proteinogenic components. Whilst modification of existing cofactors has increased organocatalytic reactivity and reaction scope, in some cases modification is not compatible with the native protein host. In this thesis we look to use an alternative protein scaffold covalently functionalised with an organocatalyst. We also consider how the chiral environment of the protein could be utilised to enable an enantioselective reaction.

In this work we select, express and purify several protein scaffolds with cysteine (Cys) residues at selected positions to allow for functionalisation with an organocatalyst. For some of the scaffolds (e.g. human steroid carrier protein, hSCP) Cys-containing variants already exist. Whilst for others (e.g. Thermus thermophilus SCP, TTSCP), the placement of Cys residues was guided by structural analysis, docking (AutoDock Vina) and modelling of the constructs using AlphaFold. We functionalised our chosen scaffold with N-heterocyclic carbenes (NHCs), a large group of organocatalysts which have been inspired by the natural cofactor thiamine pyrophosphate (TPP). NHCs catalyses a wide number of reactions including C-C bond formation and other reactions that are not known in nature. We synthesised novel NHCs with appropriate handles to enable bioconjugation to specific positions in a protein scaffold. We went on to screen and select conditions for functionalisation of the protein scaffolds and identify the best-functionalised scaffolds to test as catalysts. We also investigated the use of genetic code expansion (GCE) for the incorporation of an NHC unnatural amino acids (UAAs) into a protein. To do this we synthesised novel NHC based UAAs and screen existing orthogonal translation systems.

To test the catalytic activity of our functionalised protein scaffolds (hSCP and TTSCP) we used an intramolecular Stetter reaction as a well-studied model reaction. The Stetter reaction uses a nucleophilic catalyst to catalyse C-C bond formation between an aldehyde and an α,β-

unsaturated carbonyl forming a 1,4 dicarbonyl product. This reaction is of interest as the synthesis of enantiopure 1,4 dicarbonyl compounds remains a challenging transformation in synthetic chemistry. Stetter reactions are mostly undertaken in organic solvents with only one, aqueous Stetter catalyst reported to date, which is not enantioselective. The TPP-dependant Stetterase enzymes such as PigD and MenD can produce chiral products in water, however, these too have limitations in their substrate scope and soluble recombinant expression. By using a protein hosted organocatalyst we have the opportunity to incorporate NHCs with different core structures and enable more diverse chemistries. Here we show SCP scaffolds covalently functionalised with an NHC catalyse an intramolecular Stetter reaction with modest yields and % e.e. We demonstrate that our functionalised proteins operate under ambient conditions with low catalyst loading. Furthermore, we demonstrate that activity can be increased >20 fold by altering the protein scaffold. This is the first example of an artificial Stetterase constructed from an inactive protein scaffold and a synthetic organocatalyst. This breakthrough discovery lays the foundations to optimise the catalytic activity and improve the enantioselectivity. Furthermore, this proof-of-concept study paves the way for this methodology to be applied to other NHC/protein scaffold combinations.

Acknowledgments

First of I would like to thank my supervisor, Prof. Dominic Campopiano, who supported me throughout my time in Edinburgh. He was always enthusiastic about the project and provided endless ideas of new avenues to explore. Thanks must also go to Amanda Jarvis who helped me enormously throughout my time in Edinburgh. First of all for introducing me to the SCP scaffold and sharing numerous plasmids with us but also for providing useful insights into the project be it tips on synthetic chemistry, design of artificial enzymes or genetic code expansion. I would also like to thank EastBio for funding this project and the University of Edinburgh for hosting me and supporting me through the pandemic. Finally, I would like to thank EnginZyme for hosting me for my PIPs placement and reigniting my love for biocatalysis.

I would like to thank all the instrument staff at the University. Thank you to Faye, Logan and Kelly for their help with my MS analysis, to Bob for his help with the instruments in the ASICS lab and to the Juraj and the rest of the NMR staff.

I would like to thank the Campopiano group, both past and present members. In particular Rhona who was there with me from day one until the end. You will always be my go-to analytic guru! To those who were here when I started; Gary, Alexis, Silvia, Cath, Ben, Shona, Michael, thank you for welcoming me to the lab and for taking the time to show me around and introduce me to all the equipment. Zenam, thank you for bringing your lively spirit to the group and sparking my love of plants. Kimberley, Mariyah, Peter, Lisa, thank you for all the lunch break chats whether they were serious science chats, silly gossips, or a bit of a vent. Sam and Gustavo, it's been a pleasure to have you as post-docs during the last couple of years. I also want to thank my students Peter and Luc for your enthusiasm and drive. You were a pleasure to work with I wish you all the best of luck in the future.

To the Jarvis group, I know I was never officially part of your group but thank you for always welcoming me both in your group meetings and lab when necessary. I always knew there would be friendly faces in there if I need them. Thank you Eva for your SCP and molecular biology tips. A huge thank you must go to Richard, you were so helpful and patient with me as I began my journey into synthesis. I would not be where I am now without you.

Thank you to my PhD cohort, in particular my EastBio buddy Dan for always being there for me through the good and the bad. To those of you who joined our lockdown zoom coffee breaks and helped to motivate and keep me sane during those difficult months. I have made

some amazing friends in Edinburgh both at the university and beyond, I want to thank you all for the wonderful times and support.

Most of all I would like to thank my partner Lewis for supporting me through everything. Listening to me excitedly trying to explain good results and consoling me during the tough times. I'm sorry for talking so much science to you but I'm also super proud that you now seem to be able to explain my PhD almost as well as I can. Thank you for believing in me even when I didn't think I could do. I am so glad you supported us moving to Edinburgh and have been my adventure buddy for exploring the Scottish Highlands. Finally, I would like to thank my family for encouraging me to pursue my dreams. My dad for bringing out the scientist in me, my siblings for showing me what they can achieve and making me want to try to keep up and to my mum for loving me unconditionally and reminding me there is sometimes more to life than academic success. I am always grateful for everything you do for me and all the help and support you have given me.

Declaration

I, Alice MacAulay, declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree or professional qualification. Except where otherwise stated by reference or acknowledgment, the work presented is entirely my own.

Alice MacAulay

The University of Edinburgh

2023

Abbreviations

AOS	Alpha oxo-amine synthetase
BAL	Benzaldehyde lyase
BFD	Benzylformate decarboxylase
CD	Circular dichroism
DAD	Diode array detector
DEAE	Diethylaminoethyl
DMSO	Dimethyl sulfoxide
DSB	Disulfide bond
ESI	Electrospray ionisation
e.e.	Enantiomeric excess
GCE	Genetic code expansion
HPLC	High pressure liquid chromatography
IMAC	Immobilised metal affinity chromatography
IPTG	Iso-propyl-β-D-1-thiogalactopyranoside
LB	Lysogeny broth
LCMS	Liquid chromatography mass spectroscopy
LMWM	Low molecular weight marker
MBnThz	N-Benzyl-5-(2-maleimidoethyl)-4-methyl-thiazolium bromide
MS	Mass spectrometry
NHC	N-heterocyclic carbene
PDC	Pyruvate Decarboxylase
RT	Room Temperature
SCP	Steroid carrier Protein
SEC	Size exclusion chromatography
T_m	Melting temperature
TPP	Thiamine Pyrophosphate
TPS	Thiamine phosphate synthase
TIM	Triosephosphate isomerase
UAA	Unnatural amino acid
UV-Vis	UV-visible

Table of contents

1 Introduction

1.1 Biocatalysis

Biocatalysis utilises natural biological catalysts, enzymes or whole cells, to undertake chemical transformations[1]. The ability of enzymes to perform reactions under mild conditions with high regio-, stereo- and chemo-specificity[2] makes them particularly attractive as society looks for greener catalysts. Enzymes, in contrast to many chemical syntheses, do not generally rely on precious metals, protecting groups, toxic reagents, harsh conditions, solvents or generate substantial amounts of unwanted byproducts[3]. Commercial applications of enzymes date back to the late 19th century when extracts from natural sources such as plants and animals were used to perform chemical tranformations[1, 4]. Advances in recombinant protein production led to the availability of cheaper, higher purity enzymes from the 1980s[5, 6]. Now, driven by a demand for sustainable synthetic methods, enzyme design and engineering are playing an increasingly important role in synthesis with enzymes being designed to fit the specifications of a process instead of processes working around the enzymes limitations[3,4, 7]. This progress has been referred to as the "waves" of biocatalysis, with the field currently in the 4th wave making use of cascades and artificial enzymes[4, 6] (Figure 1).

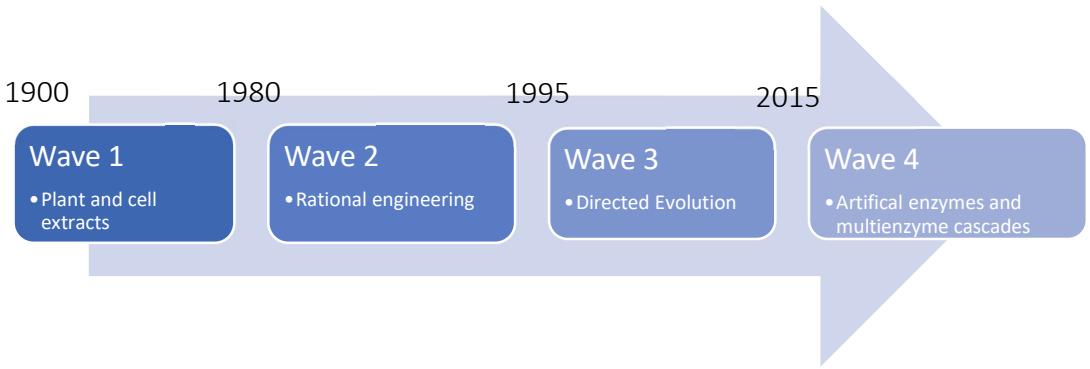


Figure 1: The four waves of biocatalysis described by Bornscheuer [4,6]

De-novo enzymes, new enzymes designed by scientists, are not yet incorporated into industrial processes although we are starting to see success in their initial design and evolution which is discussed further in the next section[7, 8]. However, we can find examples of biocatalytic cascades being applied industrially for several applications including the synthesis of high value pharmaceuticals and fine chemical as well as low priced bulk chemicals used in food and feed, household care and other products[2, 9-11]. Recent work, such as the development of a biocatalytic cascade for the manufacture of the anti-viral drug islatravir highlights the benefits of using enzymes in chemical synthesis[11]. This biocatalytic

cascade uses 9 enzymes, 5 of which have been engineered, to perform the synthesis of islatravir in half the number of steps of previous synthetic routes[11]. Biocatalytic enzyme cascades utilise multiple enzymes in succession, joining their specific functionalities to create a pathway towards the desired product[9]. Cascades can be created *in vivo*, with lysates, purified or immobilised enzymes[12, 13]. Using a cascade can prevent the need for product purification between steps, help overcome unfavourable equilibria, reduce product inhibition, recycle cofactors, and enable the removal of toxic/unstable intermediates[9-11, 14]. For example, incorporated into the cascade for the synthesis of islatravir is acetate kinase (AcK) which, although it does not directly synthesise islatravir, plays an important role in recycling the cofactor adenosine triphosphate (ATP)[11]. Despite the creation of these cascades and the new to nature enzymes, biocatalysts are still not capable of catalysing all the reactions we have access to from synthetic methods. As a result, chemo-enzymatic cascades, which merge the use of chemocatalysis and biocatalysis, are of growing interest.

1.1.1 Enzyme design and engineering

Enzyme design and engineering can be used to optimise properties of natural enzymes, such as their stability, activity, or selectivity, to fit industrial applications[15]. There are several methods for engineering enzymes including random and rational approaches [16, 17](Figure 2). The method chosen will depend on the information available about the enzyme and existing screening methods for the reaction catalysed. For rational design, knowledge of the enzyme being evolved and its structure-function relationship are required[17, 18]. Computational modelling programmes/tools can be used along with sequence alignment and structural analysis to identify key residues[17]. Then site directed mutagenesis can be used to mutate specific residues, or synthetic genes/clones with the desired mutations can be bought. Screening of the mutant/s is then undertaken to see if the desired properties have been obtained.

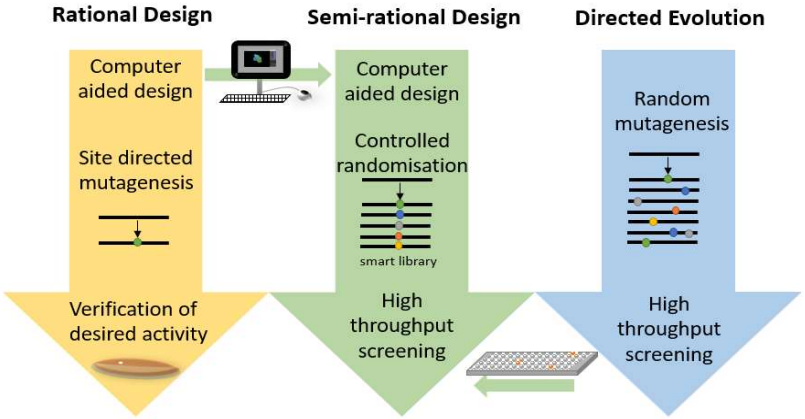


Figure 2: Techniques for engineering proteins highlighting how rational design and directed evolution contribute to semi-rational approaches.

On the other hand, without any structural or mechanistic knowledge of an enzyme, random mutagenesis techniques, most commonly directed evolution, can be used to improve its functionality. Here, iterative rounds of random mutagenesis, to generate diversity, are undertaken using techniques such as error prone PCR (epPCR) or DNA shuffling[15]. Subsequently, a high throughput screen is used to identify mutants with the desired properties[16, 19]. This mimics evolution in nature but uses artificial selection and, to accelerate the process, focuses on the expression of individual genes in fast-growing microorganisms in the laboratory[3, 15].

Unfortunately, each of these methods has its drawbacks. For rational design, as Francis Arnold points out, '*our ability to predict protein sequences or even just changes to a sequence which reliably give rise to a whole new finely tuned catalytic activities is rudimentary at best'*[3]. For random mutagenesis, screening huge numbers of mutants can be an issue. Although the size of the gene libraries produced by random mutagenesis is technically limited by the transformation efficiency of *E.coli*[20], libraries are still large and several rounds of evolution may be needed, as functional changes can be small[16]. A fast, cheap, and reliable high throughput (HTP) screen is therefore required to identify mutants with the desired properties[16, 19]. Currently micro-titre plates are the most widely used screening methods, however, ultra-high throughput methods including flow cytometry and microfluidic screens, which usually rely on fluorescence, can now analyse up to 10^6 mutants per hour[21]. Despite this, smaller libraries make directed evolution more manageable. As a result, random and rational techniques are often used in combination resulting in semi-rational design[16]. Here a portion of the enzyme, such as the substrate-binding site, is identified to mutate by using sequence alignments, structures, and biochemical data. Site saturation mutagenesis can be performed on selected residues using degenerate primers to create small but smart libraries that make screening more managable[16, 22]. Often NNK (N= any , K=G or T) primers are used, due to the degeneracy of the genetic code, these create half the number of mutants but still give all 20 amino acids[23].

Recently a trend towards the creation of *de novo* enzymes and enzymes with new to nature functionality has emerged[15, 17, 24-28]. The natural promiscuity of enzymes, which enables them to accept non-natural substrates and catalyse alternative reactions, can be enhanced by directed evolution to realise new functionalities as has been demonstrated with cytochrome P450 enzymes[3, 29, 30]. Evolving existing proteins has so far been the most successful method of creating new enzyme functionalities[24]. In contrast, computationally designed artificial enzymes have showed low-rate enhancements compared to natural enzymes[24-26, 31, 32]. However, the field of *de novo* protein design is still relatively young and is progressing at a rapidly increasing speed[33]. *De novo* enzymes are often deigned around a theoretical minimal active site or "theozyme" comprising only the transition state and catalytic residues[28, 33] (Figure 3). For example the successful design of a Kemp elimiase[32], Diels-Alderase[26], Retro Aldolase[34] and a Morita Baylis Hillmanase[25] have been achieved by use of the inside-out methodology in Rosetta. This method incorporates the "theozyme" into numerous natural protein folds and optimizes the surrounding residues to design an artificial enzyme[35]. The low initial activities of *de novo* enzymes can be substantially improved by directed evolution as demonstrated for artificial aldolases, enzymes designed for kemp eliminations and other *de novo* enzymes[28, 36-38]. Studies of the evolutionary trajectories of these *de novo* enzymes and the application of new computational techniques can provide insight into how to improve enzyme design[28, 39, 40].

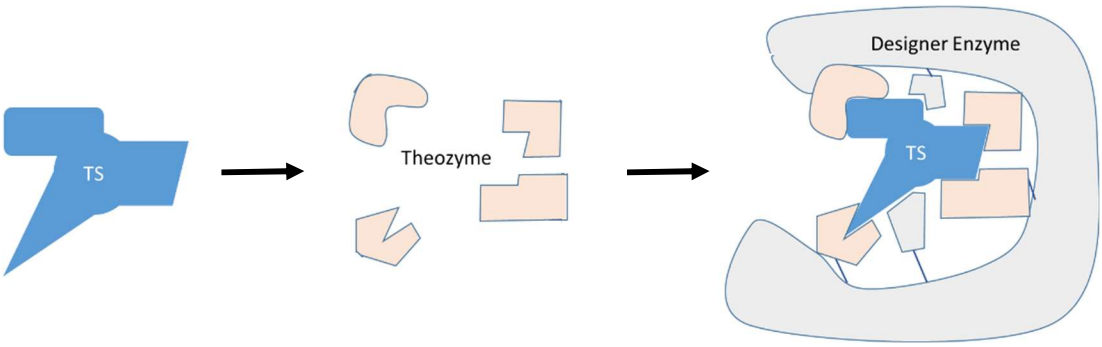


Figure 3: Computational protein design often starts by building a theozyme of catalytic and stabilising residues around a transition state. A protein is then constructed around this theozyme to make a designer enzyme.

Whilst there are 20 natural proteinogenic amino acids which make up the proteins in all living things, their functionality does not allow them to perform the full scope of reactions available to synthetic chemists. In order to carry out all the naturally required functions, nature already increases the relatively limited structural and functional diversity of these 20 amino acids by utilising co-factors and performing post-translational modifications[41]. However, further diversity is required for enzymes to perform certain new to nature reactions. This can be achieved by the incorporation of non-proteinogenic components using posttranslational

modification (see section 1.3.2), the incorporation of unnatural amino acids (UAAs, see section 1.3.3) or the use of artificial cofactors. These molecules introduce novel functional groups allowing for new catalytic mechanisms inaccessible with only the proteinogenic amino acids[7, 27, 42, 43]

1.1.2 Thiamine dependant enzymes

To expand their catalytic repertoire, enzymes utilise cofactors, which contain specific functional groups and properties[2, 44]. These may be small molecule organic enzyme cofactors, of which there are 27, or one of several commonly used metal ions[2, 44]. Together an enzyme–cofactor complex can perform reactions that could not be accomplished by either the enzyme or cofactor alone[45]. Whilst the cofactor may bring new activity, the protein can help to control the regio-, diastereo- or enantio-selectivity of the transformation as well as regulating or increasing the reactivity of the cofactor[46]. A single cofactor can perform a variety of chemically diverse reactions by interacting with different proteins[46].

Cofactors, usually required in stoichiometric amounts, are generally expensive which can make industrial synthesis using cofactor dependant enzymes un-economical without cofactor recycling[12]. Whilst certain cofactors such at thiamine pyrophosphate (TPP) and pyridoxal-5'-phosphate (PLP) are essentially self-regenerating[47] and *in vivo* catalysis can enable the natural regeneration of cofactors within the cell, many biocatalytic pathways need to incorporate additional components for cofactor regeneration[48]. Cofactor engineering strategies, which alter cofactor supply or change an enzymes cofactor preference, for example to accept cofactor mimics, can promote product formation, reduce cost, and increase orthogonality[49-51].

Thiamine, also known as vitamin B1, is a water-soluble vitamin with essential co-enzymatic and non-co-enzymatic roles[52]. There are several phosphorylated forms of thiamine in the cell as well as free thiamine. Of these, TPP is the most abundant form accounting for 70-90% of total thiamine[52](Figure 4). TPP acts as a cofactor in TPP dependant enzymes which perform a broad range of reactions including several with central metabolic roles[53]. TPP dependant enzymes make and break bonds between carbon and oxygen, sulfur, hydrogen, nitrogen and carbon [54]. Here we will focus on TPP dependant carboligases, which form or cleave C–C bonds between oxidized carbon centres.

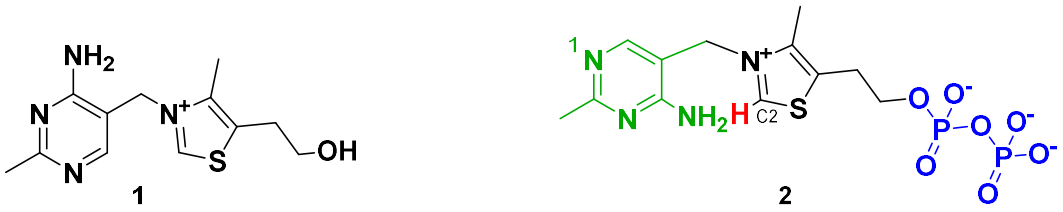


Figure 4: The chemical structures of thiamine and the enzyme cofactor, thiamine pyrophosphate (TPP) showing the thiazolium C2-proton (red), pyrophosphate (PPi, blue) and pyrimidine ring (PYR) with the N1 position labelled, (green)

The structure of thiamine (Figure 4) leads to its catalytic properties. It consists of an aminopyrimidine connected to a substituted thiazole via a methylene bridge and contains a quaternary positive nitrogen crucial for its chemical functionality[52]. In 1957 Breslow observed the reversible deprotonation of the thiazolium ring at the C2 position by monitoring the C2-H exchange with deuterium[55]. Deprotonation generates the thiazolium ylide which is in resonance with the nucleophilic carbene **(3)** that is stabilized by adjacent heteroatoms donating π-electrons and withdrawing σ-electrons[46] (Figure 5). Breslow went on to propose the mechanism of thiamine catalysis shown in Figure 5[56]. Here, the nucleophilic carbene **(3)** undergoes addition to a carbonyl group forming a tetrahedral intermediate **(5)** with the leaving group (most commonly H or CO_2) perpendicular to the aromatic thiazolium plane [24, 57]. The thiazolium ring enables substrate cleavage due to its electron withdrawing nature yielding an enaminol, known as the "Breslow intermediate" **(6)** [46]. As such, the carbonyl substrate is turned into a nucleophilic species which can react with an electrophile such as the carbonyl group of another aldehyde **(7)** in a benzoin condensation reaction (Figure 5a) or an α,β-unsaturated carbonyl in a Stetter reaction (Figure 5b)[24, 58]. Elimination of the products (**9** and **11**) allows regeneration of the carbene catalyst[58]. Both these reactions can form valuable chiral products with the benzoin condensation resulting in the formation of a new chiral centre, whilst in the Stetter reaction, the use of a prochiral Michael acceptor results in the formation of two new adjacent stereocenters[59].

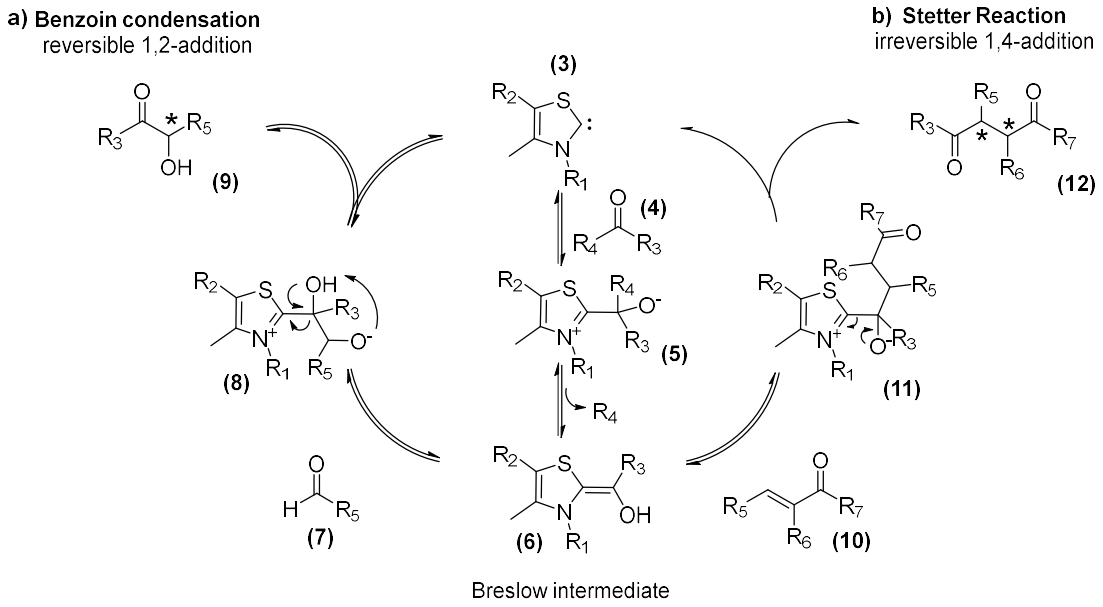


*Figure 5: Schematic representation of the reaction mechanism for **a)** the benzoin condensation and **b)** the Stetter reaction.*

It has been demonstrated that TPP contains the functionality required for catalysis and can perform functions of thiamine dependant enzymes on its own. As such it has inspired the field of N-heterocyclic carbene (NHC) catalysts described in the next section[60]. However, when bound to a TPP dependant enzyme the activity of the TPP cofactor for catalysis increases significantly and the enantioselectivity of the reaction enhanced. For example, yeast pyruvate decarboxylase[60] causes an increase in catalysis by a factor of $3x10^{12}$.

TPP dependant enzymes show a high sequence diversity typically with less than 20 % amino acid sequence identity[54]. However, they all contain both pyrimidine (PYR) and pyrophosphate (PP) binding domains which are highly structurally similar form a conserved TPP-binding fold which holds TPP in the active site[61]. Oligomerization also plays an important role with TPP dependent enzymes forming either dimers or tetramers and all functioning as a dimer of active sites[54]. Despite this conserved fold, TPP dependant enzymes can be divided into 9 superfamilies' with different architectures[61]. The PYR and PP domains are encoded by either single or separate genes with the final protein composition varying depending on the order of the two domains on the gene and if the interface between the PYR and PP domain occurs within same protein monomer (intra-monomer), or different monomers (inter-monomer)[61]. The PYR domain interacts with the pyrimidine ring of the cofactor, whilst the PP domain binds the diphosphate tail[61]. These domains non-covalently bind the TPP cofactor and hold it in its active V-shaped conformation, placing the NH_2 of PYR in close proximity to the C-2 of the thiazole.

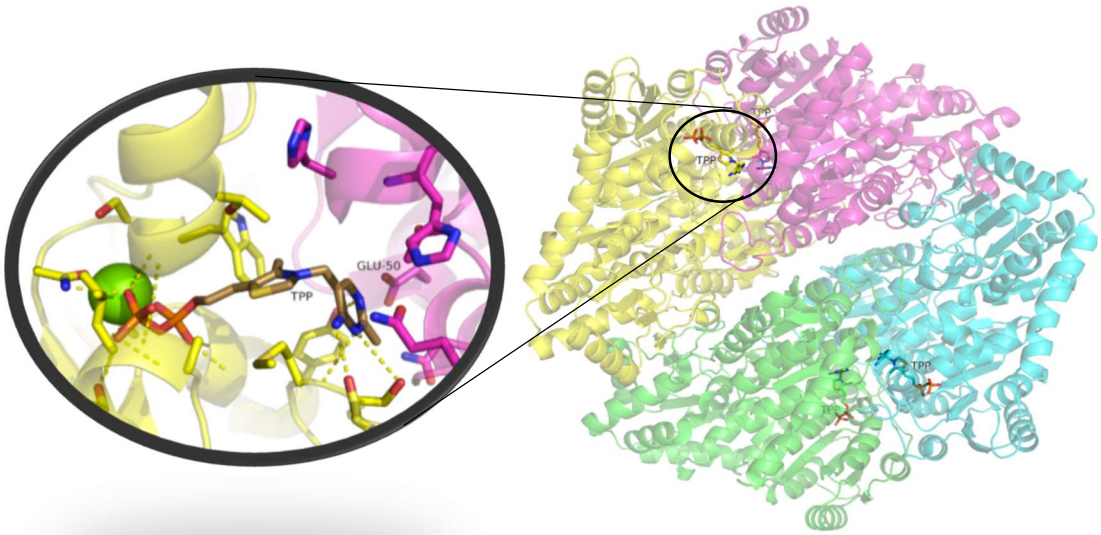


Figure 6: The structure of pfBAL PDB: 2AGO with the 4 TPP molecules highlighted and a magnification of one of the TPP binding sites showing the cofactor interacting with residues of two monomers as they key Glu50 residue.

Figure 6 highlights the structure of the well characterised *Pseudomonas fluorescens* benzaldehyde lyase (pfBAL, PDB: 2AGO), which forms a tetramer acting as a dimer of active sites[62]. pfBAL holds the TPP between two in separate domains, suspending a comparatively mobile thiazolium ring between them[62]. The PYR domain also contains a conserved Glu residue, essential for TPP activation by protonating the N1 position (Figure 4, Figure 6) of TPP which facilitates tautomerisation of the aminopyridine[57]. Structural analysis indicates that a catalytic water then facilitates the critical proton transfer between the thiazolium C2 and imine 4' nitrogen[63]. Thus, deprotonation is accelerated due to the effective lowering of the pKa at the C2 position caused by these active site features. Modelling and structural studies of enzymatic Breslow intermediates in TPP dependant enzymes have suggested that their enhanced catalytic power is also contributed to by their maintenance of a tetrahedral carbanion in the transition state, which minimises kinetic barriers compared to if the planar enamine was formed[57]. Other sites on TPP dependant enzymes have been shown to promote decarboxylation, protonation of the Breslow intermediate and deprotonation of the alcoholic proton[46]. Variations in TPP dependant enzymes can be seen in the additional domains which are present either between the PYR and PP domains or at the N- or C- terminus which perform various roles such as positioning of substrates[61]. In some cases these additional domains provide conduits for electron flow to enable processes such as oxidation to be incorporated[64]. Whilst in most TPP dependant enzymes active sites are comprised of homodimers, a unique split gene transketolase has recently been discovered[65]. This $\alpha_2\beta_2$ heterotetrameric enzyme is of particular interest for industrial applications. Not only is this

transketolase thermophilic and tolerant of organic solvents up to 50 % its split gene structure provides an opportunity for enzyme engineering [57, 65].

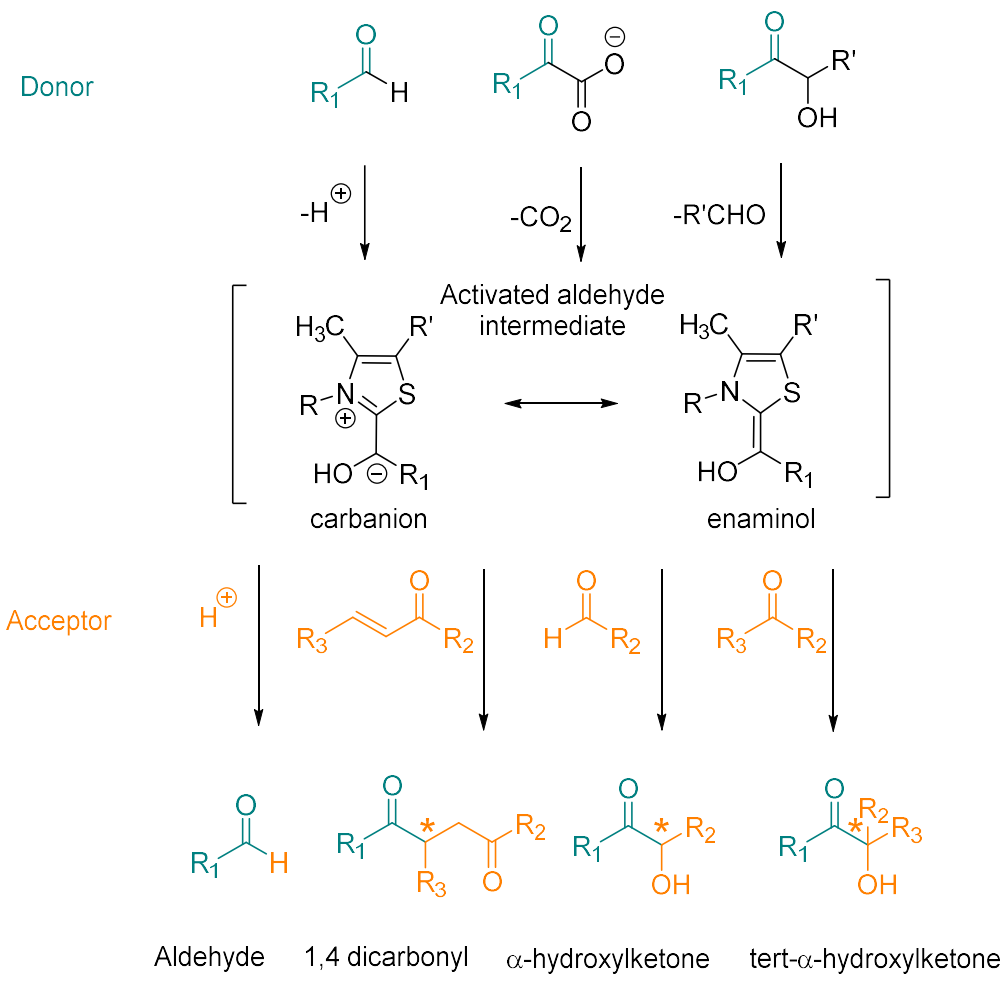


Figure 7: The reaction scope of TPP dependant enzymes is defined by to their ability to accept different donor and acceptor substrates. The donor substrates bind covalently to C2 of the TPP ylide. Donor aldehydes undergo deprotonation whilst α-keto acids are decarboxylated to form the activated aldehyde intermediate bound to TPP. This nucleophilic intermediate undergoes addition with an acceptor electrophile such as the carbonyl group of another aldehyde to form the product. Adapted from Hailes et al.[66]

Figure 7 highlights the donor acceptor pairs utilised in a range of carboligation reactions catalysed by TPP dependant enzymes. The potential to catalyse the formation of mixed carboligation products from donor and acceptor substrates with high chemo- and stereo-selectivity makes TPP dependant enzyme of synthetic interest and favourable over small molecules. For example in an intermolecular cross benzoin reaction, the potential to form eight different products (four enantiomeric pairs) makes using a small molecule catalyst challenging but the enzymes pfBAL and *Pseudomonas putida* benzoylformate decarboxylase (BFD) H281A perform cross couplings with high (90 %) selectivity[24, 67]. To take advantage of the selectivity of TPP enzymes engineering has been widely undertaken to expand their substrate scope, enhance their activity and further improve their enantioselectivity[66].

Directed evolution using colorimetric high throughput screening has made it possible to easily screen multiple variants[68, 69]. The stereospecific carbon-carbon bond-forming capability of TPP enzymes makes them promising industrial biocatalysts for the synthesis of high-value compounds. For example, transketolases have already been used in large-scale biocatalytic reactions for the production of unusual sugars by engineering them to accept a wider substrate range increases their utility[69].

The discovery of TPP dependant Stetterases such as MenD[70] and PigD[71] from the menaquinone and prodigiosin biosynthetic pathways, respectively, offer a biocatalytic solution for the synthesis of 1,4 dicarbonyls[72]. This reaction remains a challenge for synthetic chemistry due to the innate polarity mismatch of carbonyl fragments when forming even-numbered dicarbonyls[73]. As of yet the handful of natural "Stetterases" have not been applied for industrial synthesis of these compounds as they display a limited substrate scope when compared with other TPP-dependent enzymes, which when combined with difficulties in producing soluble, recombinant protein, has impeded their application in biocatalysis[74, 75]. The possibility to repurpose other TPP-dependant enzymes for Stetter reactions was recently highlighted by Chen *et al*.[76]. Their paper examined wild type TPP-containing enzymes (PfBAL, MenD, a PDC and a BFD) and used molecular dynamics (MD) simulations to select PfBAL as a potential Stetterase. The natural promiscuity of PfBAL was then exploited to catalyse an intramolecular reaction with 10 out of 13 substrates forming the corresponding Stetter products with 60-99% yield and e.r.'s ranging from 89:11 to 99:1[76].

1.2 N-heterocyclic carbenes

Synthetic chemists have been inspired by cofactors to create biomimetic catalysts which show structural and functional similarity to natural cofactors. Examples include PLP inspired N-quaternized catalysts for amine production[77] and thiamine inspired N-heterocyclic carbenes (NHCs) for serval chemical bond synthesises[78]. Here we look at NHCs a class of organocatalysts consisting of a neutral compound containing a divalent carbon atom with a six-electron valence shell in an nitrogen heterocycle which closely represent TPP[79]. NHCs can be used directly as organocatalysts, which we will focus on in this project, but it should be noted that they also play important roles in transition-metal and main group chemistry which is covered by several reviews[79-81]

The development of the field of NHC organocatalysis began 80 years ago when Ukai performed the first synthetic benzoin condensation reaction catalysed by thiazolium salts in ethanol[82, 83]. Following this, Breslow's seminal work on the mechanistic evaluation of

thiamine highlighted that is was in fact the deprotonation of the C2 position which formed the active NHC catalyst [56]. Soon after this, asymmetric thiazolium organocatalysts were discovered[84, 85]. Later, Stetter expanded the reaction scope of these thiazolium salts and really began to lay down the fundamentals of the field of NHC organocatalysis[82, 86]. Over the subsequent decades, interest in NHCs grew with the isolation of stable carbenes in 1990s leading to a "renaissance" of NHCs during which Teles *et al.* displayed that imidazolium and triazolium compounds could also catalyse benzoin condensaitons[87-89]. In 1995 Enders and co-workers isolated and characterised the first stable triazolinylide[90](Figure 8). Subsequently, chiral variants of triazolium-based carbene salts have reported by several research groups[91-94]. These NHC-triazolium catalysts now dominate NHC organocatalysis due to their more acidic 1,2,4-triazolium scaffold[95].

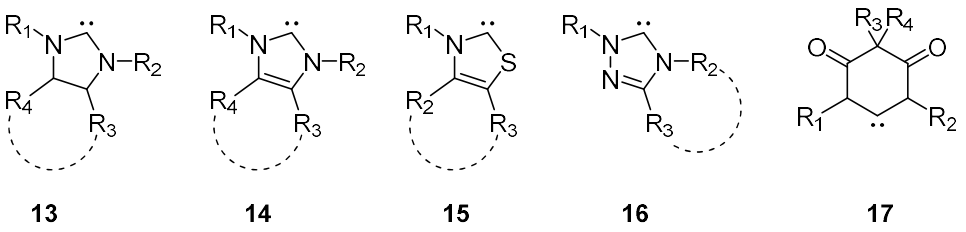


*Figure 8: Structures of some of the different types of NHCs **13.** Imidazolidinylidene **14**. Imidazolylidene **15.** Thiazolylidene **16.** Triazolylidene **17.** N,N'-Diamidocarbene.*

Figure 8 highlights some of the range of NHC structures available today with various N-heterocyclic cores. Of the different types of NHC core, five-membered rings are most common[96]. Important structural features of NHCs include the backbone, nitrogen, heteroatoms, ring size and N-substituents[79]. The backbone of the NHC can provide electronic stabilisation from its aromaticity and the substituents connected to if influence the carbenes electronics[96]. However, the class of N-heterocycle has a larger influence on the electronics with the nitrogen heteroatoms providing stabilisation of the NHCs singlet ground state due to their π donating and σ-withdrawing effect[79, 96]. The ring geometry also plays an important role in the sterics and electronics of the NHC. For example, cyclic structures force the carbene into a bent, more sp^2 like, structure favouring the singlet ground-state. Furthermore increased ring sizes lead to greater steric shielding of the carbene and effecting the electronics as the geometry change alters the nature and degree of heteroatom stabilisation[79]. Finally the N-substituents, in addition having the potential to direct asymmetric reactions, can provide kinetic stabilisation from their steric bulk[79]. Examples of these effects can be seen in the inherently different steric demands than observed in

thiazolylidenes compared to triazolylidenes which contain an additional substituent on the trivalent nitrogen atom[97]. Meanwhile, the carbene environment in thiazolylidenes is highly asymmetric due to the poorly π donating unsubstituted sulfur atom[97].

A wide variety of NHCs have been synthesised by utilising previous work on heterocycles to make cationic heterocyclic azolium salts which can then be deprotonated in situ to generate the corresponding carbene[79, 85, 98]. This method of carbene generation highlights the importance of the acidity of the pre-catalysts. Several studies report experimental and computational pK_a values of NHC precursors in DMSO or H_2O[99-101]. The general trend in acidities shows imidazoliums to be less acidic (pKa~ 20–24 in water), than thiazoliums and triazoliums (pKa~ 16–19 in water), which reflects the trends seen in the parent azoles[85]. The pK_a values are altered by the substituents with N-aryl substituents having a noteworthy impact on the acidity of triazoliums significantly reducing the pK_a[85]. The application of NHCs in organocatalysis centres around their umpolung (polarity inverting) activity although their

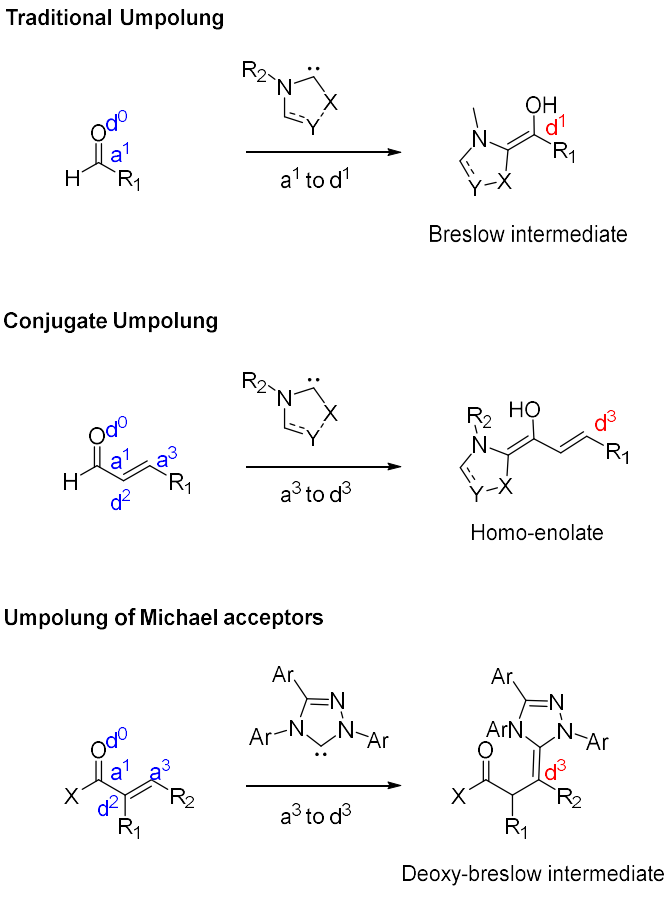


Figure 9: Umpolung reactions of NHCs include traditional a^1 to d^1 umpolung reactions such as benzoin condensations, conjugate umpolung reactions of α,β unsaturated aldehydes where the nucleophilic properties of the Breslow intermediate are transferred to the β position generating the d^3 synthon and finally the umpolung of Michael acceptors where proton transfer follows the conjugate addition of the NHC to Michael acceptor and a deoxy-Breslow intermediate is formed [104].

non-umpolung applications have been of growing interest due to their applicability to range asymmetric reactions[102]. NHCs have several modes of umpolung catalysts including traditional umpolung reactions, conjugate umpolung reactions and the umpolung of Michael acceptors (Figure 9)[103, 104]. These different modes of action enable NHC organocatalyst to perform reactions that are not known to be catalysed by TPP dependant enzymes in nature. However, even using traditional umpolung methods small molecule NHCs can still perform a broader range of reactions than TPP dependant enzymes for example the aza-benzoin reaction[46, 95].

For this project we will focus on traditional umpolung reactions (benzoin condensations and Stetter reactions), these reactions described in section 1.1.2 proceed via addition of an aldehyde followed by proton transfers to form the Breslow intermediate thus reversing the polarity of the aldehyde. The Breslow intermediate then attacks a second equivalent of aldehyde or a Michael acceptor to form a 1,2 or 1,4 functionalised product, respectively. Small molecule NHC catalysts have been designed to try to enable chemo and stereo selective reactions. For example, intermolecular cross-acyloin condensations give rise to a potential mixture of four enantiomeric pairs (Figure 10) with additional challenges arising from the

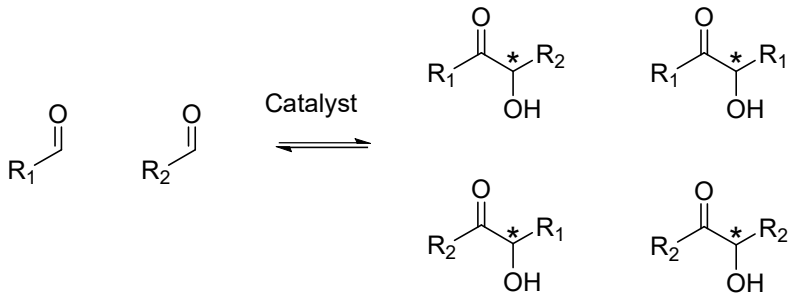


Figure 10: The cross benzoin reaction leads to a mixture of four enantiomeric pairs with the enrichment of one product over the other being influenced by the structural features of both catalyst and substrates as well as tuning of reaction conditions.

reversibility of the reaction and the formation of aldol products by enolizable aldehydes [97]. Chemoselectivity is often difficult to control as, whilst the Breslow intermediate should preferentially be formed from the most electrophilic aldehyde the self-condensation product of said aldehyde is also likely to be formed as it should preferentially undergo the nucleophilic attack of the Breslow intermediate. The careful tuning of reaction conditions and the structural features of both catalyst and substrates has enabled chemoselective intermolecular cross-acyloin condensations[97]. However, biocatalysts can perform highly selective cross benzoin reactions where the regio- and stereo- chemistry of the product can be controlled solely by the enzyme used[67, 105].

Setter reactions can be used for the synthesis of 1,4 bifunctional compounds, for example γ-ketonitriles, γ-diketones, and also γ-ketoesters[106].These are valuable products ,the synthesis of which is challenging due to the innate polarity mismatch of carbonyl fragments when forming even-numbered dicarbonyls[73]. This reaction was first reported in 1973[107] with the first thiazolium-based NHC catalysed version reported 1976[86]. Since then, the development of chiral NHC catalysts for enantioselective Stetter reactions has led to a rapidly growing pool of catalysts, covered by several reviews[73, 79, 80, 95, 108]. In this project we will focus on an intramolecular Stetter reaction which was reported by Ciganekin 1995[109] with the first enantioselective version of this reaction being reported a year later by Enders and collegues[91]. Examples of enantioselective intramolecular reactions are shown in Figure 11 and highlight the variety of products available for these reactions including cyclopentanones[93] **21** and tricyclic spirocarbocyles[110] **23**. We aim to build an artificial Stetterase to catalyse the cyclization of a salicylaldehyde-derived substrate **18** as this reaction has grown to be a benchmark for comparing catalyst efficiency in the intramolecular Stetter reaction[59]. One challenge that will need to be overcome is that this reaction appears to proceed best under inert conditions in non-polar solvents[59, 93]. These findings have been attributed to premature catalyst decomposition and the promotion of minimal charge separation in the transition state, placing the carbonyl oxygen in the proximity of the triazolium carbene carbon atom[59, 93].

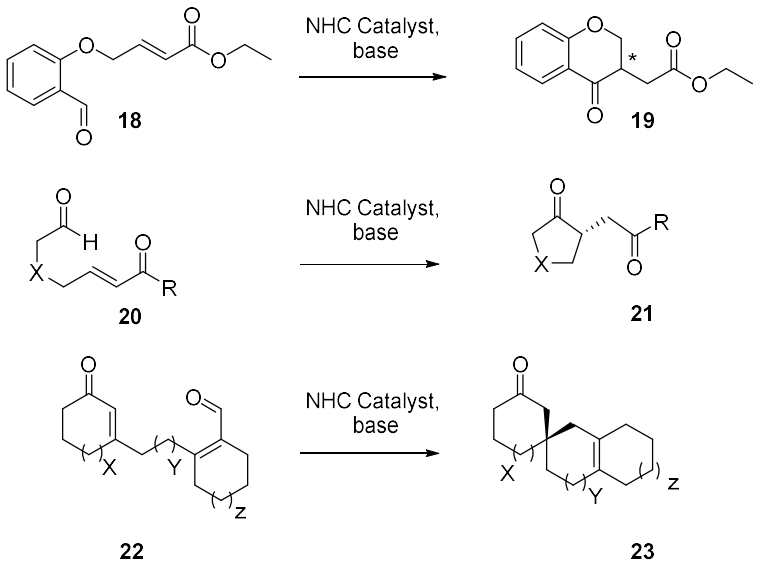


Figure 11: A range of intramolecular Stetter reactions catalysed by chiral NHC organocatalysts.

Despite the above findings an aqueous intermolecular Stetter reaction was reported by Debiais *et al.* with up to 90 % conversion at 75˚C[111] (Figure 12). This reaction was catalysed

by a novel NHC catalyst **24** which formed organic droplets in the aqueous reaction media. Although this reaction did not display stereoselectivity it provides a good benchmark for an aqueous Stetter reaction. To our knowledge no other aqueous Stetter reactions have been reported despite those conducted by Stetterases discussed in section 1.1.2 which we will also use as benchmarks when designing our artificial Stetterase. Other findings from previous work such as the proposed a stereochemical model of the Breslow intermediate of the intramolecular Stetter reaction from Rovis and coworkers could be implemented in the design and evolution of our artificial Stetterase[93].

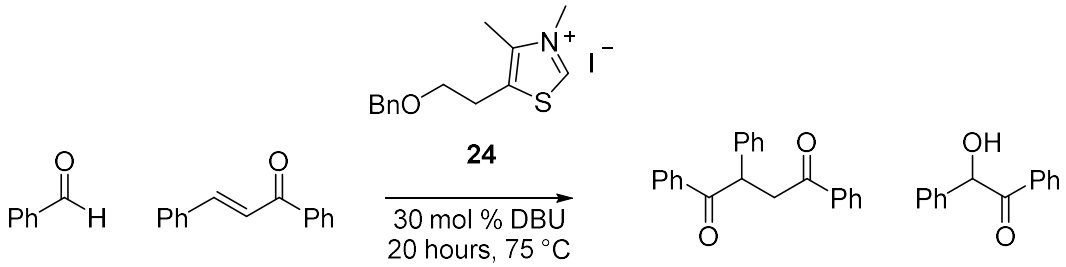


*Figure 12: A novel NHC catalyst **24** reported to catalyse an aqueous Stetter reaction between chalcone and benzaldehyde.*

Other interesting studies which are important to consider in the design of our artificial Stetterase are the use of both polymers[112] and peptides[113] to perform NHC catalysed reactions. Whilst the polymers were used to catalyse the benzoin condensation in aqueous media up to 2300−3300 times faster than the coenzyme mimics alone, they were not assessed for their stereoselectivity. The peptide-based reactions were performed in organic solvents to give moderate yields and up to 81% enantiomeric excess (e.e.). Interestingly the use of a protected amino acid gave a higher e.e. than the use of a tripeptide with a crystal structure suggesting the compact arrangement of the catalyst may be an important feature in maintaining a defined stereoenvironment and directing the stereoselectivity [113].

1.3 Protein hosted organocatalytic reactions

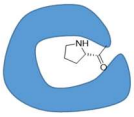
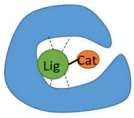
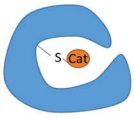
As noted in the previous sections, whilst biocatalysts can dictate the regio-, diasterio- or enantio-selectivity of chemical transformations, despite ongoing developments they still cannot perform all the reactions that can be accessed through chemical synthesis. Efforts to benefit from the advantages of both techniques such as the construction of chemoenzymatic cascades still present their own challenges. For example, many organocatalysts are unable to function under biocompatible conditions whilst biocatalysts often do not tolerate high levels of organic solvents. Although some exceptions exist, such as organocatalysts which function under aqueous conditions [114, 115] and enzyme engineered to be solvent tolerant[116], in general creative solutions are required to combine chemo- and bio- catalysis[117]. The current drive towards sustainable synthesis leads us to focus on enabling chemocatalysis in water; an abundantly available, environmentally friendly, cheap solvent.

Several methods exist to improve the biocompatibility of organocatalysts including the use of polymers, DNA, RNA or proteins to host the reaction thus sequestering the catalyst from the aqueous reaction media[118]. Protein hosted organocatalytic reactions have several benefits, including recombinant expression, aqueous stability, presence of chiral hydrophobic cavities reaction and finally the ability to tune their properties via mutagenesis. The protein's 3D structure can help dictate the regio-, diasterio- or enantio- selectivity of chemical transformations whilst the hydrophobic pockets provide a microenvironment for the organocatalyst that mimics the conditions of organic solvents[118]. Protein engineering enables further tuning of this environment to position required functional groups to stabilise the transition state (TS) or orientate the substrate to assist in chiral catalysis.

The concept of pairing chemical catalysts and genetically encoded tuneable biocatalysts has received great attention in recent years and shows promise in being able to deliver synthetically useful tools[7, 118-120]. However, the field of artificial enzyme design has mostly focused on artificial metalloenzymes (ArMs), in which the enzyme is engineered to bind a catalytic metal cofactor in one of several ways including non-covalent binding, covalent linkages and coordinating residues such as UAAs [24, 27, 119]. We aim to incorporate an organocatalyst instead of a metal to create an artificial organocatalytic enzyme. To date there are several examples of protein hosted organocatalytic enzymes which have been created using a variety of methods highlighted in Table 1. These include the use of proteins bearing an N-terminal proline residue, the functionalisation of supramolecular scaffolds, *de novo*

enzyme design, covalent protein modification and genetic code expansion (GCE) the final two of these we will focus on for this project.

Table 1: An overview of modification techniques used for the creation of protein hosted organocatalyst adapted from [118].

	a) Enzyme bearing N-terminal proline	b) Supramolecular scaffolds	c) De novo enzyme design	d) Covalent modification	e) Genetic code expansion
					
Examples	Conjugate additions[121] Aldol condensations[122]	Conjugate addition[123, 124], decarboxylative Michael addition[125],	Retro-aldol reaction[34], Henry reaction[126] Moritis Baylis Hillman reaction[25], Conjugate addition[127], Formolase[243]	C–C bond formation[128] Reductive amination[129]	Ester hydrolysis[38], Oxime/hydrazone conjugation[43, 130]
Advantages	Easy to generate recombinant protein with N-ter Proline. Computational modelling and chemical modification are not required	Enable quick screening of many different catalytic moieties	Precise designs enable the creation of novel active sites, thus minimising screening time	Enables screening of different catalytic moieties due to the relative ease of preparation	Many catalytically active amino acids are available. Chemical protein modification is not required
Challenges	Limited organocatalysis at the N-terminal position only	Limited to proteins with high affinity ligand(s)	Mechanistic and structural knowledge needed; knowledge in computational chemistry needed	Site-specific labelling can be challenging	Recombinant expression can be challenging leading to low protein yields

One of the most common organocatalysts which has been incorporated into protein scaffolds is proline (and related amines) due to its innate biocompatibility. Whilst proline alone can be used as an organocatalyst the incorporation of proline at the N-terminus of a protein can place the catalyst in a chiral hydrophobic environment surrounded by other functional residues to aid catalysis. L-Proline is an organocatalytic amino acid which contains a secondary amine and can generate iminium and enamine intermediates with carbonyl groups due to the nucleophilicity of the nitrogen[115]. Usually when proline is employed as catalyst in water relatively high catalyst loadings are needed (~10–20%) and poor enantioselectivities are observed[131]. A well-studied example of a protein with an N-terminal proline residue, 4-oxalocrotonate tautomerase (4-OT, Figure 13), helps overcome these limitations. The proline residue naturally functions as a general base to catalyse the tautomerisation of a dienol. However, studies have shown that 4-OT is a promiscuous enzyme which can catalyse iminium and enamine-based reactions including aldol reactions and conjugate additions[121, 132]. The hydrophobic active site environment lowers the pKa of proline's secondary amine to 6.4 compared to 10.5 for the free amino acid[133]. Supramolecular

methods using a streptavidin biotin system have also been exploited for secondary amine based organocatalysis[123, 124]. This enables amines which are not genetically encodable to be incorporated and enables screening of many catalysts within a short timeframe[123].

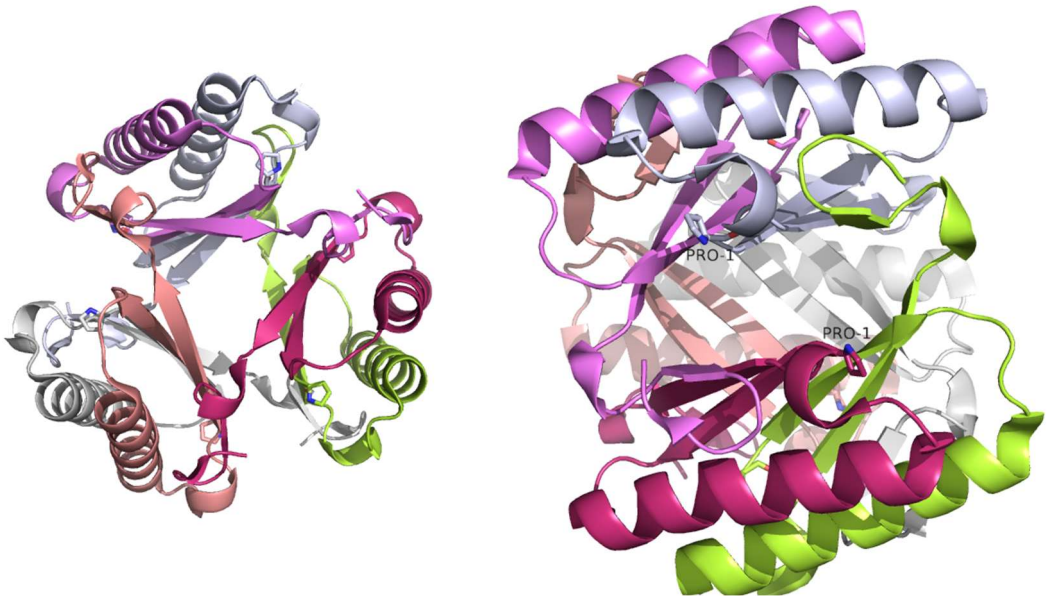


Figure 13: The structure of 4-OT hexamer (PDB: 4X19). A side on view shows the position of two of the catalytic N-terminal prolines within the hydrophobic environment of the protein.

Another interesting examples of protein hosted organocatalysis is the *de novo* design of an MBHase[25]. Whilst the initial designed reactivity of this enzyme, created by incorporating a nucleophilic histidine residue, was very low, directed evolution significantly increased activity by incorporating a key arginine reside to act as a base[134]. This enzyme was also reengineered into a hydrolase using a methyl histidine UAA incorporated by GCE[38]. The incorporation of another organocatalytic UAA, p-aminophenylalanine (pAF), has also been used to perform iminium-based catalysis further highlighting the potential of GCE[130]. We believe despite some challenges, GCE would be an interesting option for incorporating NHC functionality into our protein. Whilst UAAs containing NHC functionality have been reported and incorporated into peptides for catalysis, as of yet they have not been incorporated into a protein using in GCE, therefore development of a suitable orthogonal translation system is required[113]. For more information on GCE, see Section 1.3.3.

Computational enzyme design of a BAL has been used to create a formolase (FLS) enzyme which catalyses the carboligation of three one carbon units into a single three carbon unit[243]. This reaction, not know to occur in nature is catalysed thiazolium salts in organic chemistry. Thus utilising BAL enzyme as a scaffold which contains the TPP cofactor and computationally re-designing the benzaldehyde binding site was a rational approach for creating the FLS

enzyme. The initial design was further improved by both computationally guided mutagenesis and mutations introduced by error prone PCR. Whilst this enzyme catalysed reaction is promising approach for converting one carbon molecules into central metabolites allowing for carbon sequestration and energy storage its activity remains low. Further research has produced several FLS variants, expanded the range of carboligation reactions and create synthetic enzyme cascades[244, 245]. Whilst these approaches show the applications of TPP catalysed carboligations and how they can be evolved for new substrates and increased activities they are currently limited to the use of the TPP cofactor with its thiazolium core to catalyse benzoin type condensations.

Another method of relevance to our work, which aims to create an artificial Stetterase from a protein hosted NHC organocatalyst, is the covalent modification of protein with NHCs. Early studies by Breslow showed that NHC-modified, cyclodextrin protein mimics catalysed benzoin condensations[135, 136]. In 1993 reports by Suckling moved on to using the protease papain covalently modified with an NHC (*N*-benzyl-2-bromomethylthiazolium bromide) to catalyse a benzoin condensation of 6-oxoheptanal[128, 137]. The methods available at the time and the fact that papain was obtained from papaya extract rather than recombinant enzyme expression made it difficult to fully explore the mechanistic details of this important proof of concept work. In particular, papain contains seven cysteine residues and although six of these are likely occupied in disulfide bond formation, they could still become functionalised. Although they proved that the protein no longer bound to a thiol resin, no data was provided to suggest how many of the seven cysteine residues are functionalised. Resultantly, catalysis could be occurring at several positions around the protein. Additionally, it should be noted that the reaction led to a mixture of intra and inter-molecular products with a 28% yield of the intramolecular benzoin product after 160 hours. Recent advances in computational techniques now enable superior modelling of active sites[138], more accurate docking of substrates[139, 140] and the design of *de novo* protein scaffolds[141], which will enable us to build on this work.

1.3.1 Protein Scaffolds

The choice of protein scaffold is essential in the design of an artificial enzyme with scaffolds being derived from *de novo* design, natural enzymes, or proteins with hydrophobic cores[27]. The *de novo* design of an artificial enzyme is an alluring goal which can result in the *in-silico* design of enzymes capable of catalysing new to nature reactions[27]. *De novo* design is particularly useful for design of an artificial enzyme to catalyse a specific reaction but is less

useful when a versatile scaffold that can be applied to various reactions and substrates is desired[27]. For our goal of creating an artificial organocatalytic protein by modifying a protein scaffold with an NHC we come across the difficulties associated with the modelling of cofactors and other non-proteinogenic components as many tools do not working reliably with these additional componets[142].

Examples of natural enzymes being used to make artificial proteins mainly centre around ArMs[27]. The heme proteins are a widely exploited class of enzymes which have been used as protein scaffolds for modified cofactors[27]. Examples of this include the switching of the iron atom for non-native metals such as rhodium, iridium, ruthenium, and cobalt to create artificial metalloenzymes with altered catalytic activity, stereoselectivity, and/or oxygen tolerance[143]. Whilst this method of altering a natural protein uses its native cofactor and binding site in some cases the presence of a specific active site may limit the substrate functionalities and reaction scope that can be achieved or even lead to unwanted side products formation[27].

Proteins with hydrophobic cores but no native activity can help overcome the issues highlighted above, however, deciding on the location of the new active site requires a deeper insight into the scaffold structure. The choice of scaffold needs to be carefully matched to the reaction to select a binding pocket that is neither too wide nor too small[27]. One of the most widely exploited proteins is streptavidin which has been used extensively by the Ward group to create an array of ArMs[144, 145]. The tetrameric streptavidin protein binds four biotin molecules per tetramer via one of the strongest non-covalent protein ligand interactions found in nature[27]. Of particular interest to us was the work of the Luk group which developed a stereoselective secondary amine organocatalysis using streptavidin-biotin protein scaffold and used it to catalyse the Michael addition of nitromethane to cinnamaldehyde [124]. Other proteins with a variety of cavity sizes have been used to create ArMs such as; LmrR which contains a flat pore at the dimer interface[146], the monomeric human steroid carrier protein (hSCP) which contains a 9 Å by 18 Å hydrophobic tunnel[147] and the β barrel protein nitrobindin which naturally binds a heme cofactor. These examples show the variety of protein scaffolds which have been applied for the design of ArMs. It is important to note that although scaffold choice is essential to enable initial activity, rational design and evolution can be used to improve the activity artificial enzymes.

1.3.2 Protein covalent modification

Another consideration when building an artificial enzyme is how non-canonical catalytic features will be incorporated into the protein. One method to add functionality into a protein is to modify the purified protein covalently. There are many protein modification methods from general non-selective methods to chemo and site-specific methods (Figure 14). We are interested in creating a singularly functionalised protein and therefore plan to use a chemoselective, site specific technique. Classical methods for protein functionalisation tend to use endogenous amino acids, whilst modern methods aim to improve selectivity, efficiency and generality, often by using genetic manipulation to incorporate chemical handles at specific sites[148].

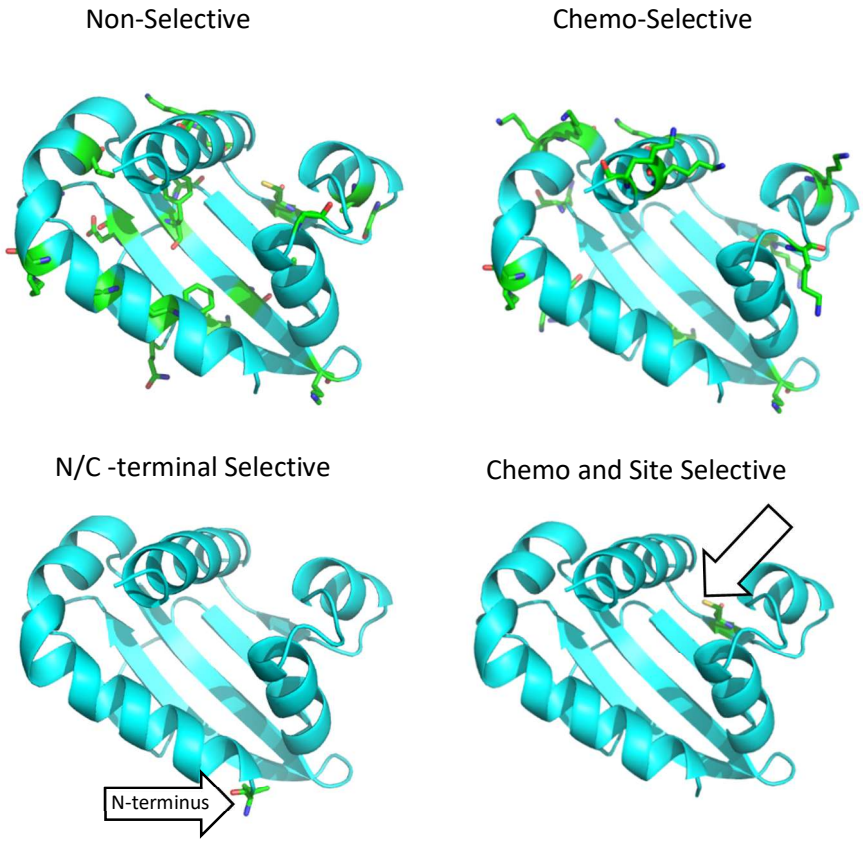


Figure 14: Protein modification techniques.

Functionalisation of natural amino acids offers a straightforward approach to modify proteins. The selectivity of the modification is determined by both the reactivity and relative abundance the target residue[149]. Nucleophilic amino acids provide a good target for functionalisation, for example Lys is commonly targeted for protein modification due to the number of available biocompatible chemical reactions which target its nucleophilic primary amine[148, 149]. However, lysine's high abundance can lead to a mixture of modified proteins,

although modern methods can improve selectivity of Lys modifications based on the amino acids enviroment[148, 150]. A protein's N-terminal primary amine also displays unique activity being the only alpha-amine in each protein, an adjacent amide bond lowers its pKa to 6.0–8.0[151]. However, functionalisation of the N-terminal means that, unless the protein folds to embed the N-terminus inside the protein, the added functionality may be solvent exposed and not surrounded by a chiral environment, which would not be beneficial for our purpose[151]. C-terminal modifications suffer from the same issue in addition to having troubles with site selectivity due to the large amount of carboxylates present in proteins. Methods to combat this and selectively functionalise the C-terminus include the use of flavin photocatalysts to perform photoredox decarboxylative alkylation[152]. Other residues which offer an opportunity for site-selective protein modifications due to their low natural abundance in proteins are aromatic residues including Tyr, Trp, Phe and His in addition to Cys[149]. With the exception of Tyr, the modification of one aromatic residue over another is challenging[149]. Selective modification of Tyr can be achieved with careful control of the reaction's pH during diazonium couplings and alkylation via π-allylpalladium complexes[149]. The conditions required for these reactions are not compatible with all proteins as low pHs are required for diazonium coupling and the use of metals can cause biocompatibility issues.

As such we have chosen to focus on the functionalisation of the thiol of a Cys residue. Cys is a popular choice of residue to functionalise due to its high nucleophilicity with control of pH, allowing selective modification over other nucleophilic residues, for example His and Lys. Common methods of cysteine functionalisation include cysteine disulfide exchange, alkylation with a suitable electrophile such as iodoacetamide and conjugate addition to a maleimide Michael acceptor (Figure 15). Currently, maleimides are the most commonly used Cys-selective modifier due to their high reactivity and chemoselectivity[153]. A maleimide

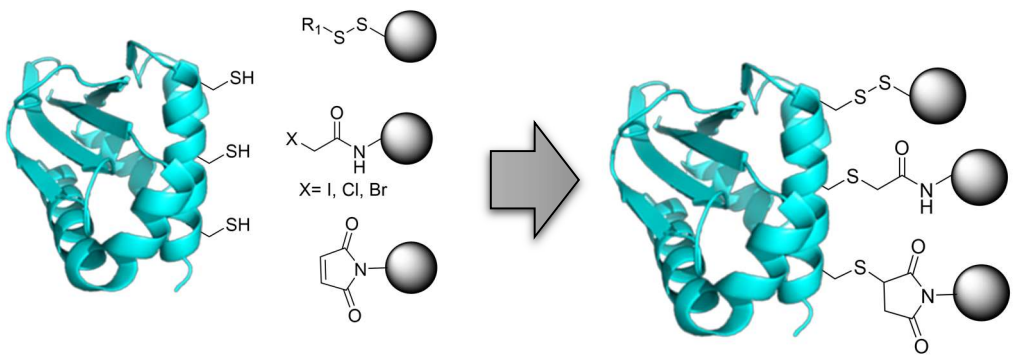


Figure 15: Classical methods for cysteine modification. From top to bottom the figure shows: cysteine disulfide exchange, alkylation, and conjugate addition to a maleimide Michael acceptor.

double bond readily reacts with thiols to form a covalent thioether linkage in a thiol-Michael addition. The reactivity of the thiol is maintained even at low pH due to its low pKa. To avoid the hydrolysis of the electrophile, modification of Cys with a maleimide is usually performed under physiological reaction conditions (pH 6.5–7.5). This also avoids side reactions with Lys and the N-terminus as well as ring-opening reaction with adjacent amines which occur more frequently at higher pH[153]. Newer methods which aim to improve on maleimide for Cys functionalisation are constantly being developed, including sophisticated photochemical activation and alkyl radical addition onto Cys-derived dehydroalanine [148, 154]. When more than one Cys is present in a protein, functionalisation methods can lack selectivity. This can be overcome by mutation of additional Cys residues to Ser. Non-catalytic Cys residues not involved in forming disulfide bonds can often be mutated to Ser with little effect due to their similar properties[155].

1.3.3 Genetic Code Expansion (GCE)

As mentioned previously GCE can be used for the incorporation of UAAs to enable site specific modification via click reactions, however it can also be used to directly add additional functionality to a protein. One way this diversity has been achieved is by altering the genetic-code machinery to enable more than 200 unnatural amino acids (UAA) to be incorporated into proteins[156]. These UAAs can be used to increase activity, stability and create novel functionality[157]. Applications to date include the creation of artificial metalloenzymes, enzymes with photoresponsive elements and incorporation of bioorthogonal handles to name a few[24, 42, 43]. Recently the Roelfes group presented an *in vivo* cascade using an artificial enzyme containing the UAA p-aminophenylalanine (pAF) as an artificial organocatalytic residue that reacts with carbonyl compounds through an iminium ion mechanism to promote reactions that have no equivalent in nature[130].

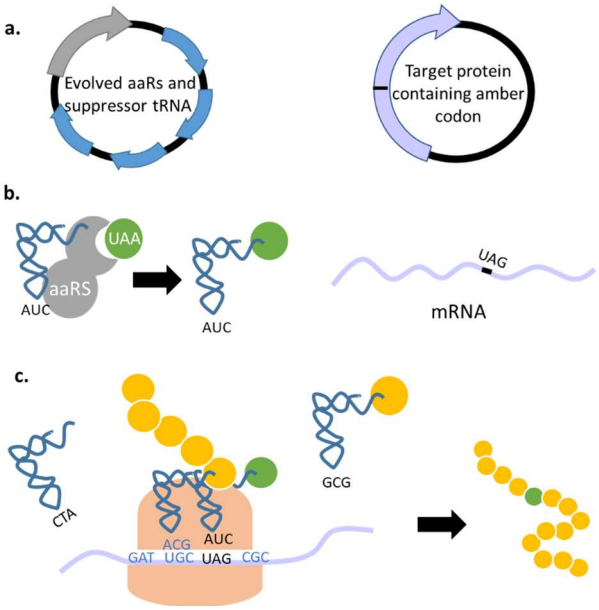


Figure 16: A schematic representation of genetic code expansion by stop ccodon expression. ***a)*** *Plasmids encoding evolved aaRS (grey), suppressor tRNA (blue) and target protein containing amber stop codon (purple) are transformed into E. coli.* ***b)*** *aaRS aminoacylates suppressor tRNA with UAA and target protein mRNA is transcribed.* ***c)*** *Ribosome translates mRNA incorporating the UAA in response to the amber stop codon.*

For this project we will focus on incorporation of UAAs in *E. coli* via stop codon suppression (SCS). However, it should be noted that there are other *in vivo* and *in vitro* methods for UAA incorporation which have been successfully applied to several organisms[156]. SCS uses an orthogonal aminoacyltRNA synthetase (aaRS)/tRNA pair (o-pair), which interact exclusively with each other, avoiding cross-reactions with tRNA and aaRSs in the host organism[158] (Figure 16). The aaRS has been genetically engineered to activate a specific UAA and pair it to its cognate suppressor tRNA which recognises a specific codon, most commonly the amber stop codon, UAG[42, 159]. When used in combination with a synthetic gene containing the amber stop codon the o-pair enables the desired UAA to be incorporated at a specific site in the

recombinant protein[42]. Consideration of the requirements of the ribosome, size and structure of the aaRS active site, competition of releases factors or other tRNAs with the suppressor and cell permeation of the UAA must be considered when incorporating a new UAA[160]. Bacterial strains optimised for amber suppression such as C321.ΔA have been created to overcome competition with the suppressor and prevent truncated gene product formation. This cell line has undergone whole genome recoding to remove all amber codons and has had the gene (*prfA*) for release factor 1, which competes for binding to the amber codon, removed[161].

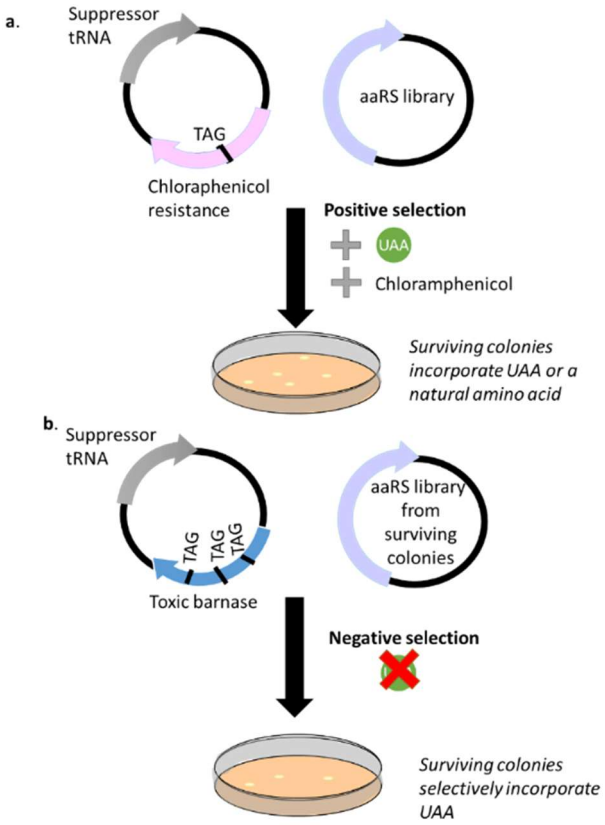


Figure 17: Schematic of the double sieve selection method for aaRS evolution. ***a)*** *Suppressor tRNA, chloramphenicol resistance gene containing the amber stop codon and an aaRS library are expressed in the presence of a UAA and chloramphenicol. Only colonies which successfully incorporate an amino acid (natural or unnatural) in response to the amber stop codon survive.* ***b)*** *Suppressor tRNA, a toxic barnase gene containing 3 copies of the amber stop codon and an aaRS library purified from colonies which grew under positive selection conditions are expressed in the absence of the UAA. Only colonies which selectively incorporate the UAA and no other amino acids in response to the amber stop codon survive.*

Two of the most commonly used o-pairs are the tyrRS/$tRNA_{CUA}$ pair from *methanocaldococcus jannaschi* and pylRS/$tRNA_{CUA}$ pair form *methanosarina mazei/ methanosarcina barkeri*[41]. Pyrrolysine (Pyl) can be considered the 22nd amino acid, and in the natural genetic codes of methanogenic archaea and some bacteria it is assigned to the UAG codon[162]. Many other orthogonal pairs exist, with some only recently being discovered;

however, few naturally recognise the amber stop codon and modifying the anticodon affects aminoacylation in most cases[163]. New aaRSs can help increase productivity of UAA incorporation and bring structural insights to enable recognition of new UAAs as demonstrated by Seki *et al*.[162]. Furthermore, the discovery or creation of multiple o-pairs which do not cross react can enable the incorporation of multiple UAAs into one protein[164].

AaRSs can be evolved to accept a desired UAA whilst excluding endogenous amino acids. The most common technique uses a double sieve selection method which links the viability of *E. coli to* the activity of the aaRS[165] (Figure 17). Firstly, positive selection is used to screen aaRS active site mutant libraries for the ability to incorporate the UAA in response to an amber codon at a permissive site in an essential protein, for example an antibiotic resistance gene such as chloramphenicol[165]. Next, negative selection is used to remove aaRSs that incorporate endogenous amino acids into a lethal protein, commonly the toxic barnase gene is used, in the absence of the UAA[165]. Evolved aaRSs may accept multiple UAAs as whilst they are evolved under selection pressure against endogenous amino acids, no selection pressure against other UAAs is applied[165, 166]. This cross reactivity can save time evolving new aaRSs and, as only the UAA to be incorporated is supplied in the culture media, only the desired product is produced.

Most UAAs are synthesised chemically and then added to the culture media, however alternative options such biosynthesis by the host may be more economical[156, 167]. The chemical synthesis of UAAs is challenging as the stereocentre at the α-carbon must be considered, this results in a synthesis several steps long requiring protecting groups and expensive toxic heavy metal catalysts[156, 168]. The expense of the UAA thus limits the incorporation of chemically synthesized UAAs to small-scale use[156]. Additionally, UAAs, particularly those that are large, charged or significantly deviate from canonical amino acids, are inefficiently taken up by the cell[169]. To improve the feasibility of scale up, the expense of the UAA must be reduced and uptake efficiency increase. Methods including the use of methyl esters and periplasmic binding proteins have been shown to improve cellular uptake of UAAs, however, these are not yet widely used[169-171]. Alternatively, cell-free methods avoid the problem of cellular uptake and have additional benefits including allowing manipulation of the translation system without risking cell viability, use of a specific tRNA composition and addition of chemically charged tRNAs[161]. Unfortunately, cell free methods using UAAs are not currently suited to scale-up due to cost, but improvements in UAA incorporation efficiency and the protein synthesis yields are likely to change this[172].

Synthesis of UAAs inside the cell *via* biocatalytic methods offers the advantage of not only utilising mild conditions, being chemo-, regio- and enantio- selective and requiring fewer steps but it also removes the need to purify the product or consider the efficiency of cellular uptake[156]*. In vivo* biocatalytic synthesis of UAAs has been performed in several instances[173-176]. Methods for the biocatalytic synthesis of UAAs including TAs, AADH and sidechain modification are covered in recent reviews[156, 177, 178]. Advances in this area, such as the implementation of a three-enzyme cascade to overcome the weaknesses of using either TAs or AADHs on their own[179] show promise for the future of economically viable UAA synthesis. Currently however, biocatalytic synthesis is it is not yet widely applied due to drawbacks including the complexity of the enzymatic cascade required, as well as the limited substrate scope and poor stability enzymes involved[156].

1.4 Aims

This project aims to create an artificial Stetterase by covalent linkage of a TPP inspired NHC to a protein a scaffold. Two approaches will be taken to achieve this goal, firstly we will aim to covalently modify a purified protein scaffold with an NHC and secondly, we will attempt to incorporate an NHC based UAA into a protein via genetic code expansion. As a large proportion of the structure of TPP enzymes is dedicated to binding and activating the cofactor, using a covalently attached thiazole derived NHC in its place opens the possibility for significantly different enzyme structures to be used for C-C bond formation.

We will synthesise novel NHCs with handles to enable bioconjugation to proteins containing a cysteine residue. Potential protein scaffolds will be investigated to assess their suitability to host an NHC organocatalysts and the substrates for an intramolecular Stetter reaction. Mutations will be inserted as required to enable for functionalisation before the enzymes are expressed and purified. Finally, functionalisation conditions will be determined to obtain pure mono-functionalised protein.

NHC containing UAAs for GCE will be synthesised and current OTSs will be screened for their ability to incorporate these UAAs. Once UAA incorporation has been achieved these proteins will be expressed and purified to assess their catalytic properties.

The NHC functionalised enzymes will be screened for the ability to catalyse an intramolecular Stetter reaction. The overall aim is to improve stereoselectivity and catalytic efficiency by semi-rational design of the best protein-scaffold/NHC pairs.

2 Synthesis of NHC functionalised protein scaffolds

Synthetic N-heterocyclic carbenes (NHCs) catalyse numerous reactions, some of which have no enzymatic equivalent in nature, such as aza-benzoin reactions (Figure 18). However, in some cases, enabling chemo, regio and stereo selectivity is difficult with small molecule NHC catalysts as it requires the careful tuning of reaction conditions and the synthesis of complex chiral NHCs. Meanwhile, the cofactor thiamine pyrophosphate (TPP), acts as natures NHC organocatalyst in TPP-dependant enzymes to catalyse a range of enantioselective C-C bond forming reactions. Although TPP-dependant enzymes catalyse a wide range of reactions they still have some limitations and do not work for all substrates despite engineering efforts. Attempts to combine the benefits of the selectivity of enzymes with the broad substrate and reaction scope of small molecule NHCs have led to NHCs being inserted into dextran rings[136], polymers[112] and even attached to papain[128]. These methods provide a hydrophobic and/or chiral environment for the NHC whilst being less substrate specific than natural TPP enzymes. We were particularly inspired by the use of papain to create an artificial enzyme by functionalisation with *N*-benzyl-4-bromomethylthiazolium bromide. However, this work performed in 1993 displays limitations (discussed in the introduction). We aim to harness recent advances in computational techniques which enable better modelling of active sites[138] and more accurate docking of substrates[139, 140] to create artificial Stetterase enzymes by functionalising protein scaffolds with NHC precursors.

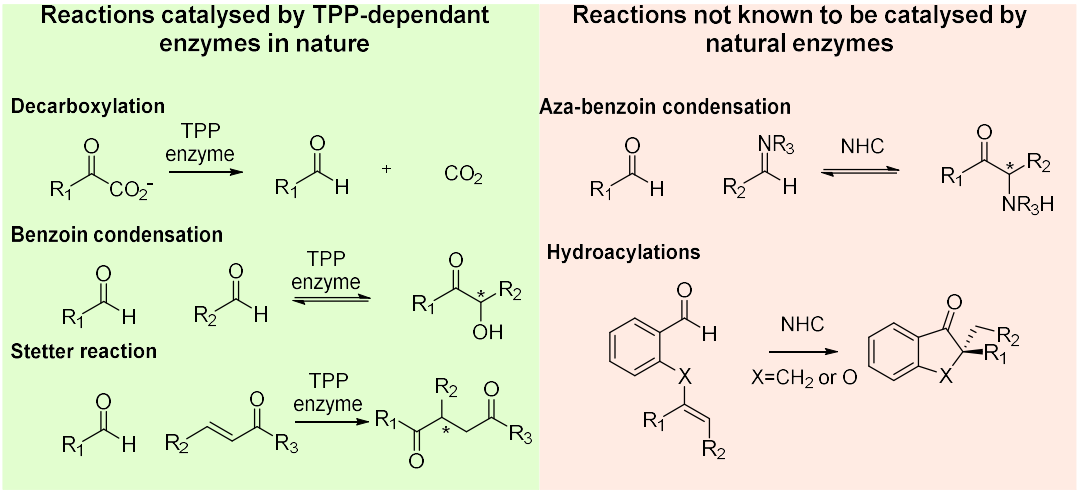


Figure 18: Examples of reactions catalysed by TPP dependant enzymes in nature and reaction which can be catalysed by NHCs but are not known to be catalysed by natural enzymes.

The ideal method to create an artificial Stetterase would be to use *de novo* enzyme design to build an enzyme around the transition state of the reaction. However, there is a large computational cost associated with undertaking such a project, particularly due to challenges associated with the modelling of cofactors and non-proteinogenic components[142]. Instead, we opted to utilise readily available proteins combined with basic modelling to allow us to create protein scaffolds which could be functionalised with NHCs precursors with appropriate handles. Out of the methods discussed for covalent functionalisation of proteins in the introduction, we chose to focus on the use of halo- and maleimide- linked NHC precursors to functionalise a cysteine residue on our protein scaffold.

2.1 Aims

The goal of this section of the thesis was to firstly synthesise a range of NHCs containing functional groups that enable the selective modification of proteins. Next, we aimed to identify, mutate, express, and purify recombinant protein scaffolds which would make a suitable host for covalent functionalisation with an NHC organocatalyst. Finally, we sought to selectively functionalise our purified protein scaffolds with our novel NHCs to obtained mono-functionalised protein scaffold.

2.2 Synthesis of NHC precursors

2.2.1 Bromo thiazolium salts

Inspired by Suckling[128, 137], who functionalised papain with bromo-thiazolium salts (5-bromomethyl-N-methylthiazolium bromide and 5-bromomethyl-N-benzylthiazolium bromide), we first looked to synthesise N-methyl-5-bromomethyl thiazolium bromide. Starting from a cheap hydroxyethyl thiazole **25** we employed established methods for the halogenation of the alcohol and N-alkylation of the thiazole previously used to make coenzyme mimics[112].

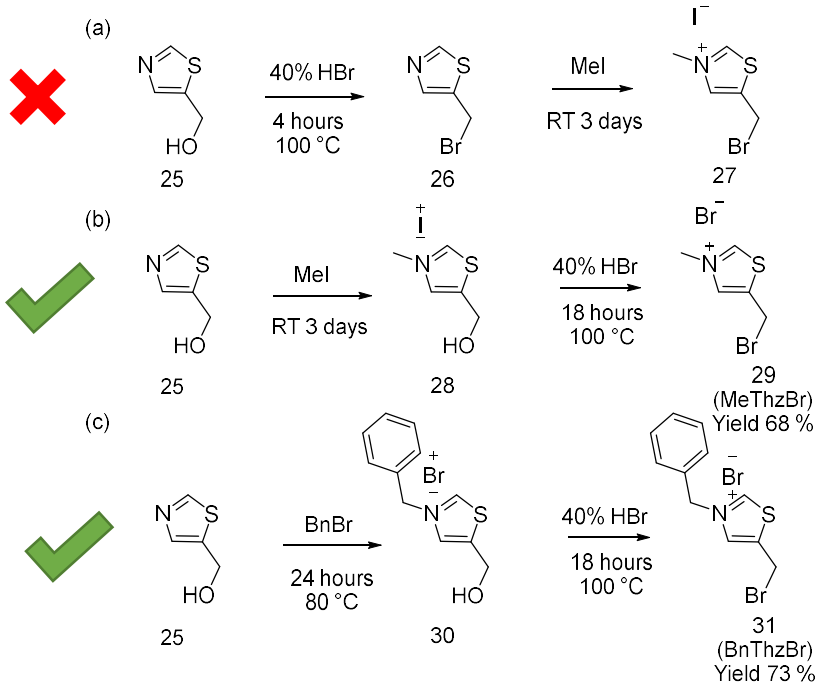


*Figure 19: Synthesis of Bromothiazolium salts via bromination and alkylation of hydroxylmethylthiazole **25.** 1a highlights that bromination prior to alkylation did not yield the desired product whilst 1b and c illustrate that the desired product could be obtained by first alkylating then brominating the intermediate.*

Initial attempts to form the bromothiazolium salt (Figure 19a) were unsuccessful, however performing the methylation first followed by bromination gave product **29** MeThzBr in 68% yield (Figure 19b, Appendix 1.6.1). This same method was applied to make the benzylated product **31** BnThzBr in 73% yield (Figure 19c, Appendix 1.6.2). Both products were hygroscopic forming amorphous orange solids. Our catalysts were functionalised with the bromomethyl group on the 5 position compared to the 4 position used by Suckling[128]. We believe this should somewhat lower the pKa of the C2 proton once the catalyst is installed in a protein as Wang *et al*[99] reported a small decrease in pKa when the 5 position of a thiazole is alkylated compared to the 4 position.

2.2.2 Maleimide linked thiazolium salts

We opted to synthesise two maleimide linked thiazolium salts as the use of a maleimide handle for protein functionalisation rather than a halogen should provide a more stable substrate. We prepared novel thiazole based NHCs, MMeThz **37** and MBnThz **38** via a convergent 5 step synthesis using inexpensive starting materials (Figure 20, Yields MMeThz 9%, MBnThz 14%, Appendix 1.6.4-5). The first step of this synthesis involves the conversion of the alcohol on **32** to a halide through a substitution reaction with thionyl chloride. Meanwhile a reversible Diels Alder reaction is used to protect the maleimide functionality. The products of these first two steps can then be combined in a substitution reaction which, following deprotection and purification, yields the maleimide thiazole **36**. Finally, N-alkylation of this thiazole yields our thiazolium salts **37** and **38**. MMeThz **37** was formed as a yellow powder whilst the benzylated product MBnThz **38** was harder to obtain as it formed an amorphous sticky off-white solid which was more difficult to isolate.

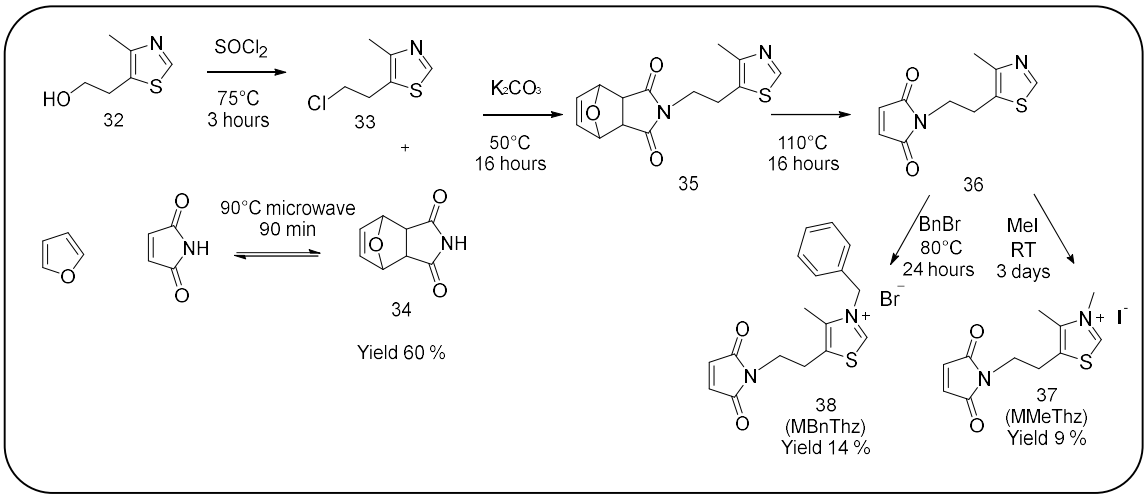


*Figure 20: The five-step synthesis of maleimide linked thiazolium salts MMeThz **37** and MBnThz **38**.*

2.2.3 Functionalised Triazolium salts

In order to expand our research and investigate protein functionalisation with some of the other NHC backbones, we aimed to synthesise a maleimide linked 1,2,4 triazolium salt. 1,2,4 triazolium salts typically exhibit lower pKas than thiazolium salts and are more widely used for catalysis. However, to our knowledge no triazolium salts with a suitable handle for protein modification have been reported in literature although several synthetic routes to hydroxyl functionalised triazole have been reported[180]. Our initial synthetic route aimed to synthesise a halogenated triazolium salt. Whilst several halogenated triazole are reported in literature, the halogen is adjacent to an sp^2 centre which is unsuitable for substitution reactions with cysteine. Using a published method, we synthesised NHC precursor **42**[181]. Unfortunately, our yield was very low due to poor cyclisation in last step of the synthesis

(1.6.17, yield 75 mg). Attempts at bromination of **42** were unsuccessful with the reaction turning black. However alternative methods such as activation with thionyl chloride should be investigated.

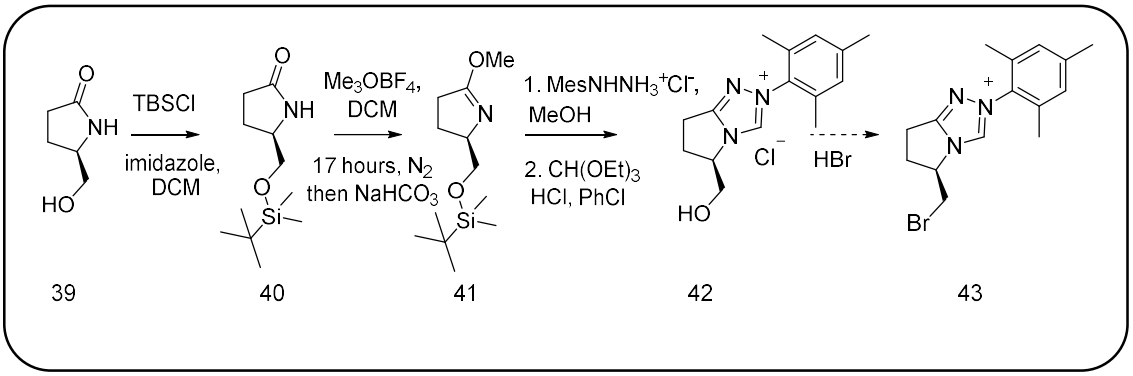


*Figure 21: Proposed synthesis route to the bromotriazolum salt **43** using the reported synthetic route to **42**[181].*

For the synthesis of a maleimide linked triazole and we chose to use a method similar to that used for the synthesis of our thiazolium salts (Figure 22). The maleimide linked triazolium salt MBnTri **47** was synthesised in a 5-step process with a yield of < 15 %. A linker was added to the protected maleimide **34** using 1,3 dibromopropane to give **44**. Next the triazole was N-benzylated using benzyl bromide to yield **45** which underwent a substation with **44** to give the protected triazolium salt **46** verified by ^{1}H-^{15}N HMBC NMR analysis. After purification degradation of the furan protecting group was observed during this step. Deprotection yielded the product **47** as an amorphous yellow solid, however impurities were observed in the NMR analysis but the amount of product was too small for further purification.

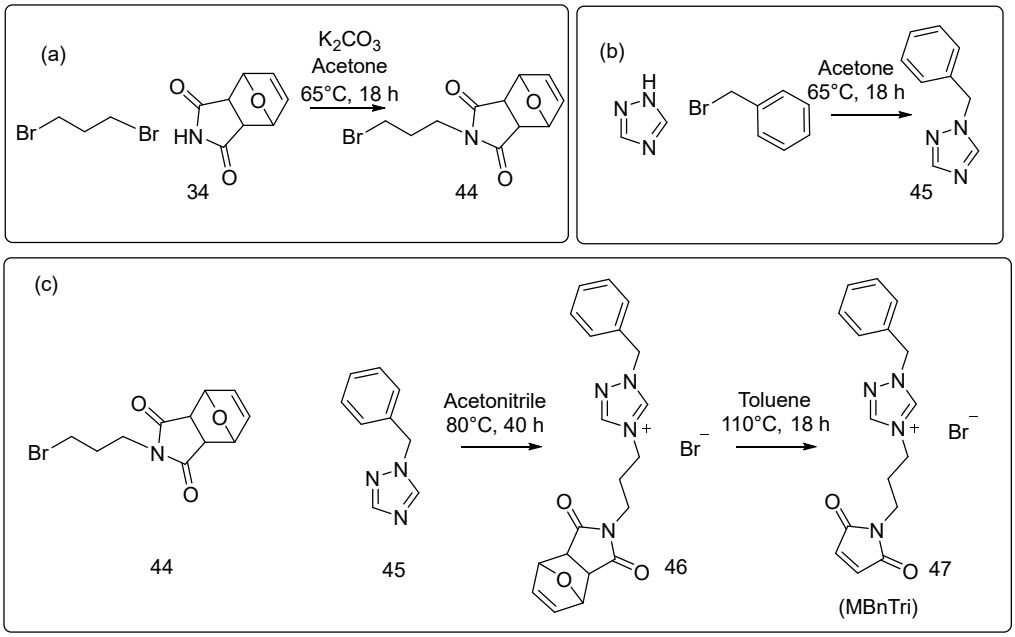


*Figure 22: Synthesis of MBnTri **47** by combing maleimide precursor **44** with the triazole fragment **45.***

Whilst we have synthesised a triazole with a functional handle for bioconjugation to a protein, further synthetic work to produce a library of triazolium salts and lower the C3 proton pKa would be beneficial for this project. It would also be beneficial to scale up the synthesis of **47** to obtain a pure product.

2.3 Enzymes screened

Our ideal Stetterase enzyme scaffold would contain a hydrophobic cavity to accommodate an NHC precursor and the substrate, no native enzymatic activity, good recombinant expression in *E.coli* and good stability; preferably a thermophilic protein stable at pH required for reaction. For single site selective functionalisation of a cysteine residue, it is preferable if the proteins contain no cysteine residues as this allows engineering of a single cysteine at a desired position. Alternatively, an amber stop codon can be encoded to use for GCE (Chapter 4). We selected enzymes to screen from those already available in our lab and looking at structures available in the protein data bank (PDB). We chose to use a variety of different scaffolds to discover what worked best, including testing some proteins which did not meet all our desired criteria. This allowed us to gain an insight into the most important features of an enzyme scaffold.

2.3.1 *Th*AOS

The first protein scaffold we investigated was thermophilic alpha oxoamine synthestase *(Th*AOS), a member of the α-Oxoamine synthase (AOS) enzyme family. These are PLP-dependant enzymes which catalyse the irreversible, Claisen-like condensation of an L-α-amino-acid with an acyl-CoA thioester. *Th*AOS was first reported in 2007 during identification of thermophilic α-Aminooxononanoate Synthase (AONS) analogues for use in a biocatalytic synthesis of biotin[182]. A specific biotin-based assay was used to verify the gene product was active for the condensation of L-Ala with pimeloyl-CoA and an assay using Ellman's reagent revealed that ThAOS had a uniquely broad substrate scope[182]. The enzyme was found to be highly stable with optimal operating conditions of 70 °C, pH 6. Further studies of other thermophilic AOS enzymes found that thermophilicity and promiscuity were not fundamentally linked, making ThAOS unique [183]. Our group determined the crystal structure of *Th*AOS[184] and has investigated a library of ThAOS mutants. Of particular interest for this work, were some of the valine 79 mutants, V79A and V79S, which increased the thermostability and expanded the substrate scope of ThAOS. Interestingly V79S also showed a significant increase in activity (to be published).

2.3.1.1 Structural analysis

In solution AOSs are dimeric and display a large amount of buried surface area at the dimer interface[184]. Each monomer comprises of three main structural domains, firstly an α-helical N-terminal domain which embeds in the opposite dimer. Secondly the catalytic domain which forms the bulk of the enzyme and contains the two active sites that sit at the interface

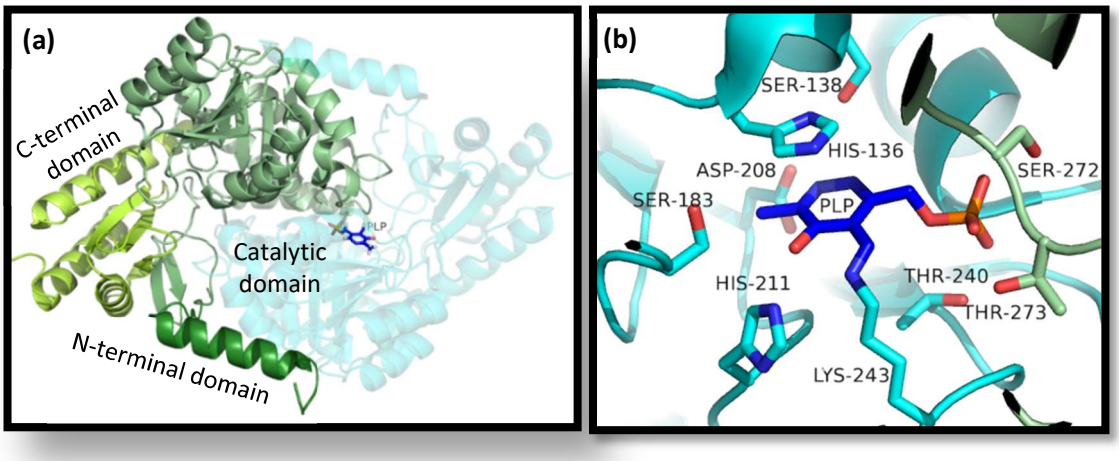


Figure 23: The structure of ThAOS (pdb 7POA) a) Highlights how the two monomers (blue and green) form a dimer and shows the three domains which make up ThAOS; C-terminal domain (lime green), Catalytic domain (green), N-terminal domain (dark green). b) Shows the active site of ThAOS with PLP cofactor covalently linked to lysine 243 forming an internal aldimine. It highlights His136, Asp208 and His211 which interact with PLP and residues which form the phosphate binding cup; Thr240, Ser272' and Thr273'.

between the two monomers. Figure 23 highlights the residues which make up this active site and shows the PLP cofactor covalently linked to Lys243 forming an internal aldimine. Other key residues for PLP binding include His136 which forms a π-stacking interaction with the pyridine ring, and His211 which chelates the C3 hydroxyl. Furthermore, the acidic sidechain of Asp208 maintains the protonated state of the pyridinium ring of PLP. This domain also contains a phosphate binding cup made up from a network of polar residues of both monomers comprising Thr240, Ser272' and Thr273'. The third C-terminal domain sits away from the active site and largely does not interact with the other monomer.

We aim to utilise the functionality available in this active site pocket to deprotonate our thiazolium salt and bind substrates for an intramolecular Stetter reaction. The volume of the cavity according to analysis using CastP is 3379.5 Å^3 for the pocket containing the chain B catalytic lysine and 2855.8 Å^3 for chain A [185] suggesting some asymmetry in the dimer. Both

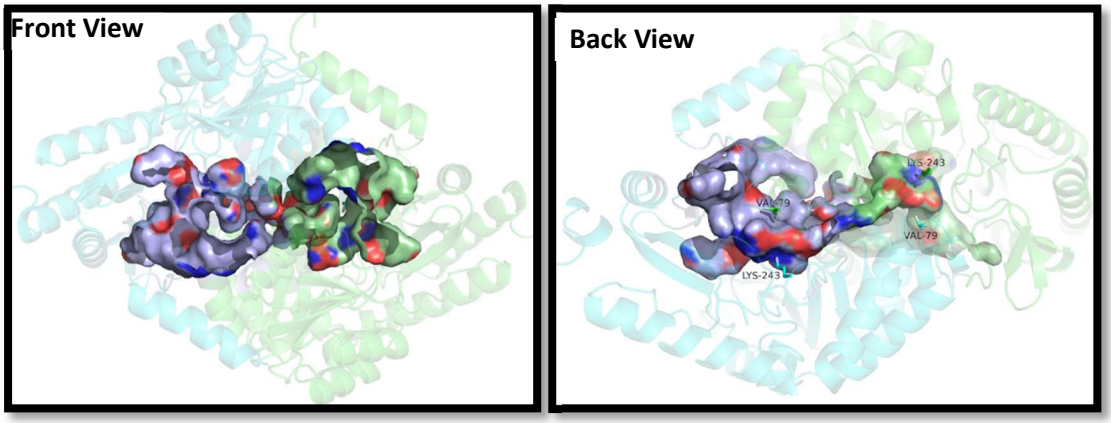


Figure 24: CastP analysis of ThAOS highlights the cavity containing the active site lysine has 5 openings and its volume varies between the two (cavities green and blue). The back view of the cavities shows the positioning of the catalytic lysine 243 and valine 79.

cavities have 5 openings with the largest being seen on the front face, whilst a view from the back shows the catalytic lysine residue and V79. Importantly, for our uses *Th*AOS does not contain any cysteine residues which allows us to design and engineer our own cysteine residue to which a thiazolium salt can be covalently attached.

To install our NHC precursor the catalytic lysine residue K243 was mutated to give K243C. The aim of this mutation was to not only provide a handle for functionalisaiton but also to inactivate the native catalytic activity of the enzyme by preventing PLP binding and formation the internal aldimine. Swiss model[186] was used to build structures of ThAOS K243C and ThAOS

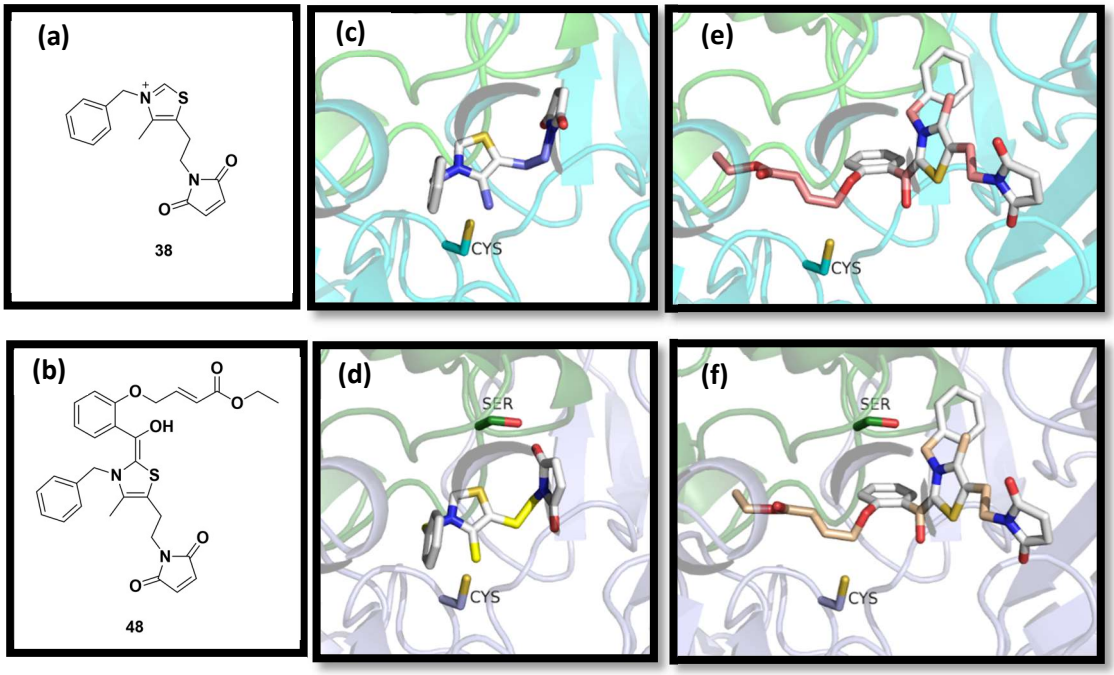


*Figure 25: Structures of **a)** MBnThz **38** and **b)** the Breslow intermediate of MBnThz **48** and the intramolecular Stetter starting material. Poses generated by docking **38** and **48** into models of ThAOS mutants using AutoDock Vina Docking. **c)** Pose 1 of MBnThz docked into ThAOS K243C, **d)** Pose 1 of MBnThz docked into ThAOS V79S K243C. **e)** Pose 1 of Breslow intermediate docked into ThAOS K243C, **f)** Pose 1 of Breslow intermediate docked into ThAOS V79S K243C.*

V79S K243C. To establish if our bulkiest NHC precursor MBnThz **38** would fit into the ThAOS binding site, docking studies were undertaken using AutoDock Vina[140]. These highlighted favourable binding affinities in several conformations for MBnThz, however the maleimide was not orientated towards the cysteine residue (Figure 25). The Breslow intermediate **48** of MBnThz **38** with our chosen substrate **18** for the intramolecular Stetter reaction was also docked to ensure there was sufficient room for its in-situ formation and to see how this altered the orientation of the maleimide. Again, favourable binding affinities were observed but the maleimide was not in the proximity of the cysteine residue (Figure 25). We decided to proceed with making this K243C mutation as an initial test with the view to come back and asses the viability of other positions.

2.3.1.2 Cloning, Expression and Purification

Mutagenesis primers were designed for the K243C mutation using the method described by Liu and Naismith[187]. The plasmids pETHisTev ThAOS, pETHisTev ThAOS V79A and pETHisTev ThAOS V79S were used as templates for site-directed mutagenesis and the plasmids were amplified with the primers incorporating the mutations into the gene. The ThAOS V79A K243C mutation was unsuccessful with no colonies being obtained when transforming the PCR products. Sequencing analysis of the other constructs confirmed the desired mutations had been made.

Previously determined expression and purification conditions were used for ThAOS and ThAOS V79S whilst a series of expression tests of the ThAOS K243C mutant were carried out

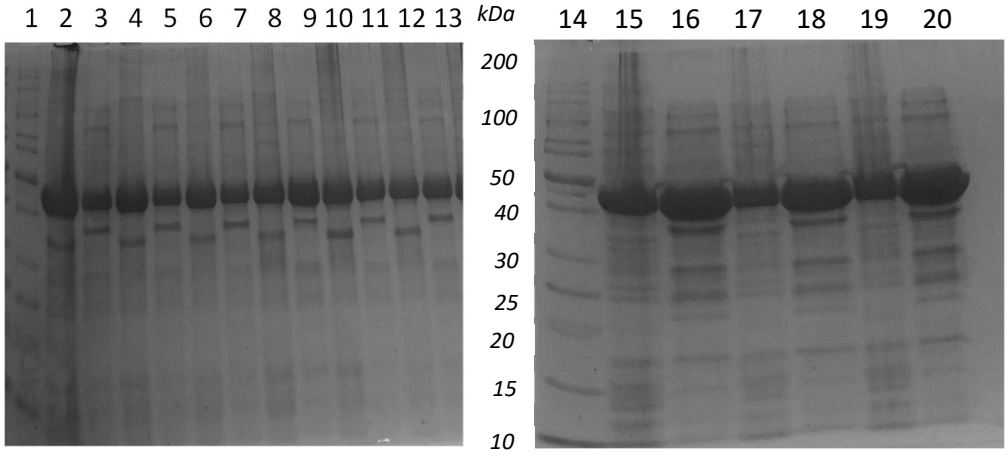


Figure 26: Expression tests of ThAOS K243C 1. PAGERULER MWM, 2 Insoluble 0.1 mM IPTG 37°C 3 hours, 3 Soluble 0.1 mM IPTG 37°C 3 hours,4 Insoluble 0.5 mM IPTG 37°C 3 hours, 5 Soluble 0.5 mM IPTG 37°C 3 hours, 6 Insoluble 1 mM IPTG 37°C 3 hours,7 Soluble 1 mM IPTG 37°C 3 hours, 8 Insoluble 0.1 mM IPTG 30°C 18 hours, 9 Soluble 0.1 mM IPTG 30°C 18 hours, 10 Insoluble 0.5 mM IPTG 30°C 18 hours, 11 Soluble 0.5 mM IPTG 30°C 18 hours, 12 Insoluble 1 mM IPTG 30°C 18 hours, 13 Soluble 1 mM IPTG 30°C 18 hours, 14 PAGERULER MWM, 15 Insoluble 0.1 mM IPTG 16°C 18 hours, 16 Soluble 0.1 mM IPTG 16°C 18 hours, 17 Insoluble 0.5 mM IPTG 16°C 18 hours, 18 Soluble 0.5 mM IPTG 16°C 18 hours, 19 Insoluble 1 mM IPTG 16°C 18 hours,20 Soluble 1 mM IPTG 16°C 18 hours .

to optimise conditions for enzyme production in *E. coli* BL21 (DE3) cells. Similarly to the WT enzyme, ThAOS K243C is a highly expressed protein with all the conditions tested yielding a large amount of soluble protein.

Although SDS PAGE gel analysis of the expression test appear to show a large amount of insoluble protein, this may be due to incomplete lysis and in this case is not a concern as a plenty of soluble protein is also present (Figure 26). The expression tests suggested that the best expression conditions were induction with 0.1 mM IPTG before growth at 16 °C for 18 hours, the same conditions used for expression of WT ThAOS.

Table 2: A summary of the yields and observed masses found by LC MS of ThAOS mutants.

Protein	*Yield*	*Expected mass with N-Met cleavage*	*Observed mass*
WT ThAOS	66 mg/L	46384.02	-
V79S	50 mg/L	46371.97	-
K243C	75 mg/L	46358.99	46357.8
K243C/V79S	97 mg/L	46346.93	46347.4

Using these conditions, the ThAOS mutants were expressed and purified using IMAC in yields of 50-100 mg/mL (Table 2). Purification was performed via immobilised metal affinity chromatography (IMAC) with protein containing fractions being identified by absorbance at 280 nm. Figure 8 shows a representative SDS PAGE analysis of the purified ThAOS mutants with analysis of the full purification of the THAOS V79S mutants. All mutants produced a single band corresponding to the proteins molecular weight (Figure 27). Size exclusion chromatography (SEC) gave a single peak with the retention time of (79 mL) which, when compared to a calibration curve, falls between the expected elution of a monomer (74 mL)

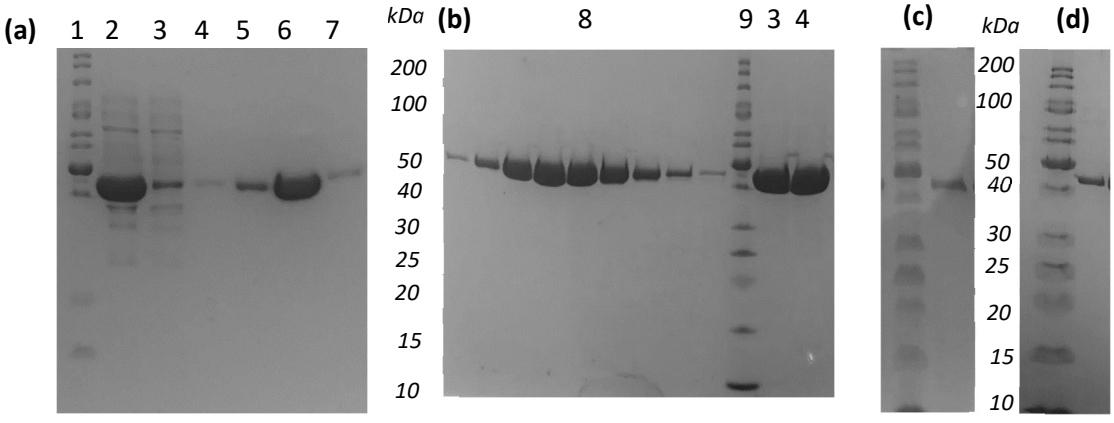


Figure 27: (a) SDS PAGE analysis of IMAC purification of ThAOS V79S with lanes showing 1 PAGERULER Molecular weight marker 2. Lysate, 3. Flow through, 4-7. Elution fractions containing protein. (b)ThAOSV79S SEC protein containing fractions (labelled 8), 9. PAGERULER Molecular weight marker 10. Purified ThAOS V79S, 11. Purified WT ThAOS, (c) Purified ThAOS K243C, (d) Purified ThAOS V79S K234C.

and dimer (82 mL). Although the use of a calibration curve indicates the estimated mass of the protein, deviations from the expected elution volume can occur if the protein is not perfectly globular[188]. Therefore, the comparison to ThAOS V79S which forms an active dimer (Figure 28) reinforced our belief that the K243C mutants still form dimers. Further characterisation by LC-MS analysis revealed the expected mass of mutants (Figure 29, Table 2). Interestingly the K243C mutants maintained a yellow colour upon purification, all be it a less intense colour than the WT enzyme. This yellow colour usually indicates the presence of PLP in the active site of a PLP enzyme and suggests our mutants still bind PLP. This is likely due to the fact that, of the residues which interact with PLP, only the catalytic lysine residue was mutated. To completely eliminate PLP binding we could attempt mutation of other residues, such as His211 known to interact with the phenolic oxygen of PLP and shown to be essential for PLP orientation[189].

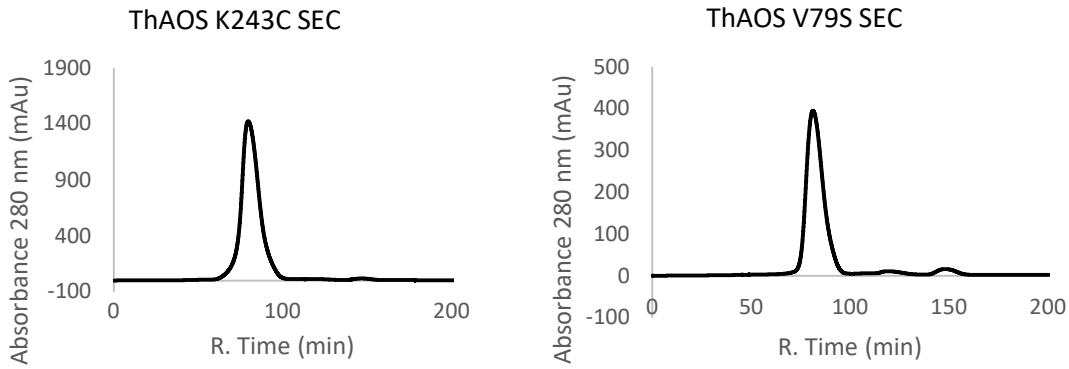


Figure 28: Plots of the absorbance profiles at 280 nm of the SEC of ThAOS K243C and THAOS V79S using an S200 gel filtration column indicate the formation of a dimer.

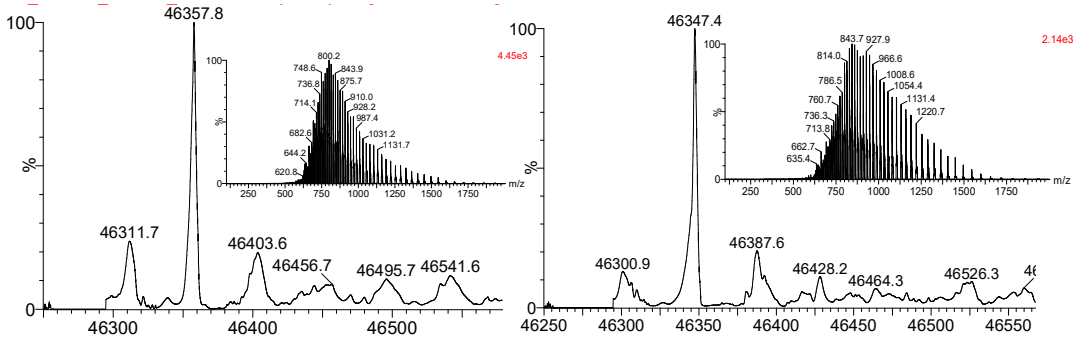


Figure 29: LC-MS analysis of ThAOS mutants shows the expected masses were obtained a) ThAOS K243C, b) ThAOS V79S K243C.

2.3.2 Steroid carrier protein (SCP) scaffolds

Steroid carrier proteins (SCP) are found in many different species and contain diversely sized hydrophobic pockets[190]. Here we will focus on SCP-2 and SCP-2 like proteins, also known as nonspecific lipid transfer proteins which are ubiquitous intracellular ~13 kDa proteins[191]. The physiological roles of SCP-2 remain unknown although vertebrate SCP-2 has been implicated in a wide range of lipid-related functions *in vitro*, with tunnels in the protein binding fatty acids[191, 192]. The human steroid carrier protein 2-like domain (hSCP, Figure 30) from human peroxisomal multifunctional enzyme type 2[193] has been successfully used to create several artificial enzymes for reactions as diverse as hydroformylation[194, 195] and both photocatalytic[196] and transition metal oxidation[147].

2.3.2.1 Structural Analysis

The structure of hSCP reveals an 18 Å long and 9 Å wide hydrophobic tunnel which should be sufficient to accommodate a maleimide-linked NHC and the intramolecular Stetter substrate (Figure 30). The wild type hSCP contains no cysteine residues but several single cysteine mutants of hSCP have been reported[197] and amber stop codon mutants have also been created (all kindly gifted to us by the Jarvis group, University of Edinburgh). Crystal structures of hSCP V83C, A100C and Q111C have also been determined[197]. Figure 31 highlights the position of the cysteine mutants that were investigated in this project and shows additional mutations used to aid stabilisation or in the case of F94H to act as a base. These cysteine mutants give us the opportunity to assess the effect of placing our organocatalyst at various positions inside the hydrophobic tunnel.

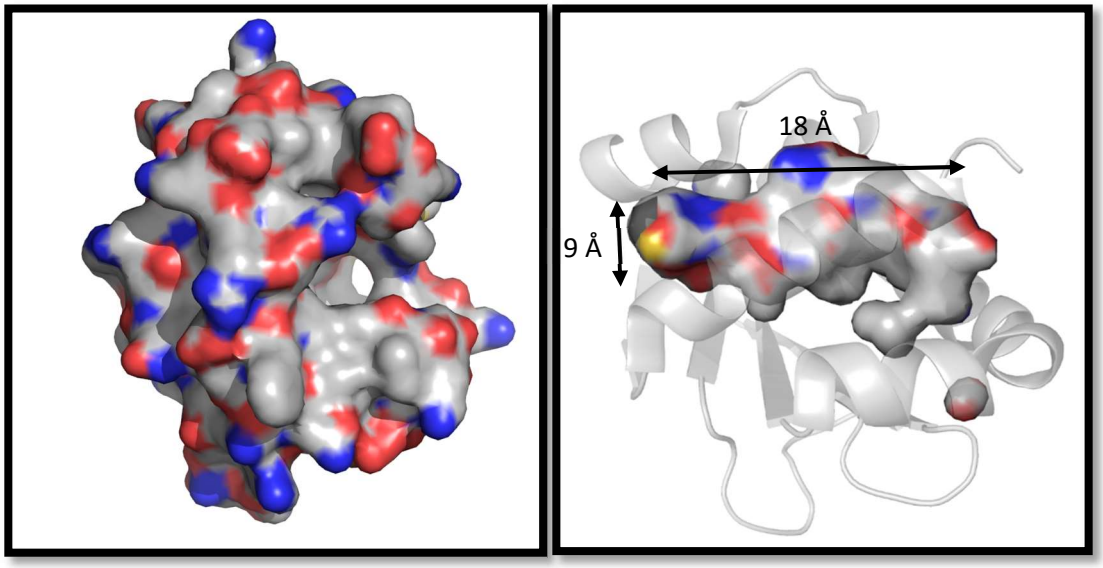


Figure 30: The structure of hSCP A100C (pdb 6Z1W) shows a hydrophobic tunnel through the protein of 18 Å long and 9 Å wide.

a)

```
         10          20
MEGGKLQSTF VFEEIGRRLK

         30          40
DIGPEVVKKV NAVFEWHITK

         50          60
GGNIGAKWTI DLKSGSGKVY

         70          80
QGPAKGAADT TIILSDEDFM

         90          100
EVVLGKLDPQ KAFFSGRLKA

        110
RGNIMLSQKL QMILKDYAK
```

b)

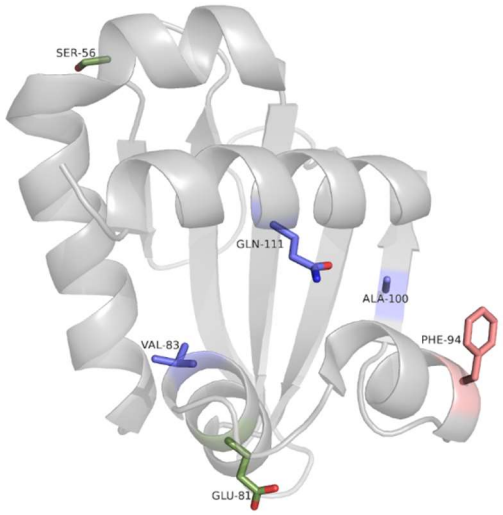


Figure 31: a) The sequence of hSCP and b) the structure of hSCP highlighting where mutations have been introduced. Blue- Val83, Ala100 and Gln111 have been mutated to cysteine. Green- Glu81 and Ser56 have been used for stabilising mutations. Pink – Phe94 has been mutated to histidine to act as a base.

The SCP from *thermus thermophillus* (TTSCP, PDB:2CX7)[190] was identified as another potential scaffold. *Thermus thermophilus* is a Gram-negative, anaerobic, extreme thermophile used a model organism for synthetic biology. As such TTSCP has desirable properties including good thermostability and an acidic PI (4.98). Structures of TTSCP show it does not contain a tunnel but a cavity with an opening at just one end[190]. Interestingly, in the asymmetric crystal unit both an open and closed conformation is observed. Previous work highlights TTSCP likely exists as a monomer in equilibrium between open and closed conformations[190]. CastP analysis of TTSCP reveals the cavity to be much smaller than that of hSCP (468.98 Å^3 for TTSCP vs 1,076.4 Å^3 for hSCP)[185]. Unlike hSCP, the TTSCP protein contains two native cysteine residues (Cys13 and Cys63) which form a disulfide bond (DSB), but no further cysteine residues. Disulfide bonds are often found in thermophilic proteins and can contribute to thermostability. Unfortunately, the presence of the disulfide has the potential to lead to unselective functionalisation when using a bromo or maleimide to functionalise a protein. Therefore, we considered the possibility of mutating Cys13 and Cys63 to alternative residues to allow selective functionalisation of the TTSCP scaffolds. We chose to first identify potential locations cysteine mutants and then remove the native cysteine residues in case their removal had a detrimental effect on stability. To explore other difference between these homologues sequence and structural alignments were performed. The sequence alignment of TTSCP and hSCP created using EMBOSS needle[198] showed 22% of the residues are conserved between TTSCP and hSCP (Figure 33). Meanwhile a structural alignment

performed using Pymol indicates a. RMSD of 4.625 Å and highlights that whilst the two SCPs contain a similar fold they have some structural differences (Figure 33).

Alignment of hSCP (1IKT) and TTSCP (2CX7)

Identity: 33/148 (22.3%), Similarity: 56/148 (37.8%), Gaps: 46/148 (31.1%)

a)

```
 1 MEGGKLQSTFVFEEIG-----RRLKDIGPEVVKKVNAVFEWHITKGGNIG     45
   ||        :|.|..     |:|.:  .|..:|..:.:|      |::.
 1 ME--------LFTEAWAQAYCRKLNE--SEAYRKAASTWE------GSLA     34

46 AKWTIDLKSGSGK-------VYQGPAKGA--------ADTTIILSDEDFM     80
   .....|.|:|..|       ::.|..:||        ||..|......:.
35 LAVRPDPKAGFPKGVAVVLDLWHGACRGAKAVEGEAEADFVIEADLATWQ     84

81 EVVLGKLDPQKAFFSGRLK-ARGNI-------MLSQKLQMILKDYAKL    120
   ||:.|:|:|..|...|.|: .:|.|       ..:|:|..:.::.|
85 EVLEGRLEPLSALMRGLLELKKGTIAALAPYAQAAQELVKVAREVA--    130
```

b)

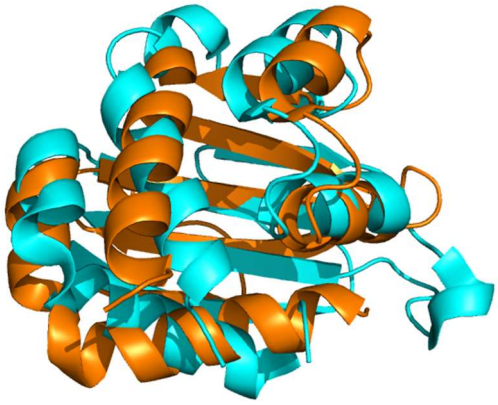
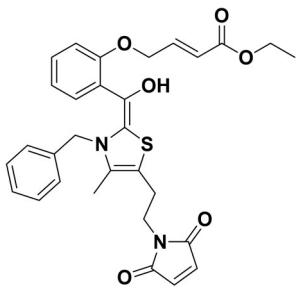


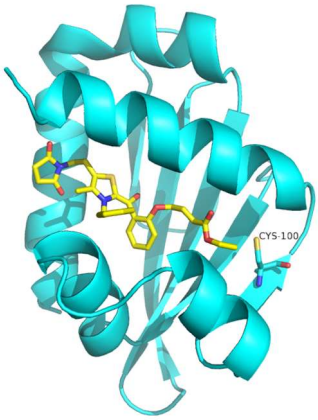
Figure 33: a) Sequence alignment of hSCP and TTSCP performed with emboss needle hSCP (top), TTSCP (bottom) b) Structural alignment of hSCP (1IKT, orange) and TTSCP (2CX7 chain A, blue) performed with Pymol.

a)



39

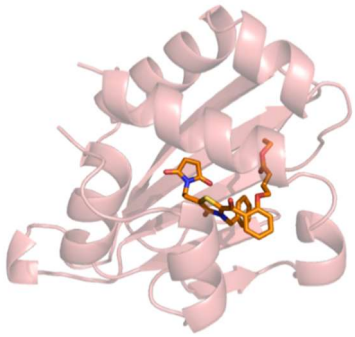
b)



mode	affinity (kcal/mol)	dist from best mode rmsd l.b.	rmsd u.b.
1	-9.1	0.000	0.000
2	-8.7	1.434	2.154
3	-8.5	2.123	3.483
4	-8.0	2.109	3.324
5	-7.9	2.276	3.973
6	-7.8	2.307	3.956
7	-7.7	2.622	4.552
8	-7.6	3.167	9.551
9	-7.6	2.221	4.156

Writing output ... done.

c)



mode	affinity (kcal/mol)	dist from best mode rmsd l.b.	rmsd u.b.
1	5.6	0.000	0.000
2	5.8	0.130	1.421
3	6.8	2.277	6.702
4	6.8	3.051	8.679
5	8.0	2.395	8.427
6	8.6	2.431	7.519

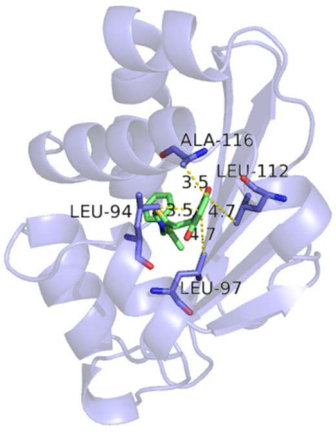
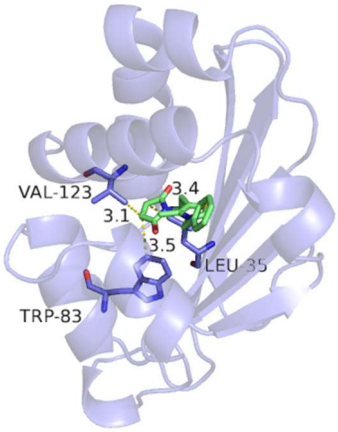
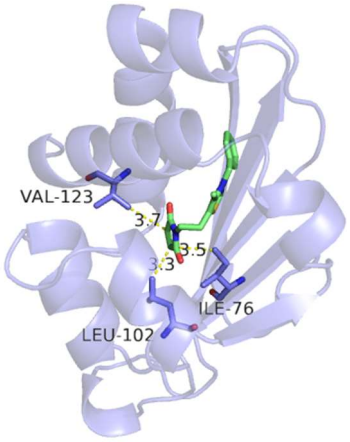
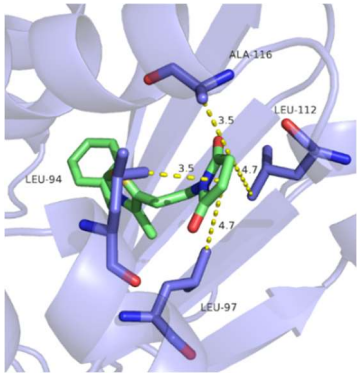
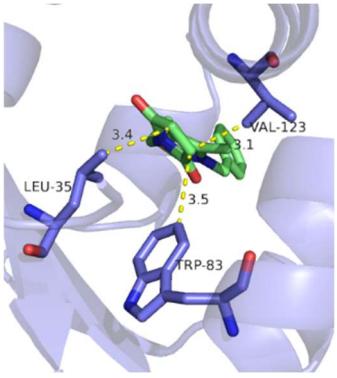
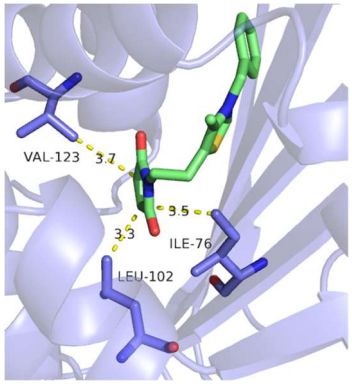
Writing output ... done.

Figure 32: a) Structure of the predicted Breslow intermediate (48) formed from (38) and (18). b) Pose of the docking of the Breslow intermediate into hSCP A100C (6Z1W). c) Pose of the docking of the Breslow intermediate into TTSCP (2CX7, chain A).

AutoDock Vina[140] was used to generate feasible poses of the Breslow intermediate **48** formed from MBnThz **38** our bulkiest NHC precursor and the intramolecular Stetter starting material **18** in the cavity of scaffold TTSCP (PDB: 2CX7 chain A) and hSCP A100C (PDB: 6Z1W) for comparison (Figure 32). The calculated affinities are merely a guide to how well the ligands are accommodated since the Breslow intermediate forms in situ after the covalent attachment of MBnThz **38** onto a cysteine residue altering the exact position. The docking score for 6Z1W gives a negative free energy of binding indicating favourable binding, however none of the poses show the maleimide near A100C. Although slightly positive affinities were obtained with 2CX7, the poses obtained indicate that the smaller cavity size can still accommodate the Breslow intermediate (Figure 32).

Whilst cysteine mutants were available for the hSCP, we needed to assess the best positions to introduce cysteine mutations into the TTSCP structure. To aid us with this decision MBnThz **38** was docked into TTSCP (2CX7, chain A). Several poses gave similar positions but poses 1, 2, and 8 showed three unique positions (Table 3). From these poses residues within 5 Å of

Table 3: Three distinct poses generated by docking of 1 into TTSCP (2CX7 chain A) For each pose, residues within 5 Å of the maleimide are highlighted.

Pose 1	Pose 2	Pose 8
		
affinity -5.2 kcal/mol	affinity -5.1 kcal/mol	affinity -4.0 kcal/mol
		

the maleimide were identified. Initially nine residues were identified: L35, I76 W83, L94, L97 L102, L112, A116 and V123. Preference was given to residues which faced into the cavity (L35, I76, W83, L97, L102 and L112) over those that were solvent exposed (L94, A116 and V123) which may allow the maleimide thiazole to bind outside the cavity. We chose to not mutate residues L35 and I76 since the model suggested they formed part of the beta sheet which we did not want to disrupt. Therefore, we were left with three residues W83, L102 and L112 which were chosen to mutate to cysteine for initial studies.

Models of the chosen mutants were created using AlphaFold2[138] to assess how they altered the cavity size. Although the protein is dynamic with variations seen between the NMR structure and two chains of the crystal structure, we hoped analysis of these structure may provide some insight into how the mutations affect the cavity. Our analysis using CASTp[185] suggested the L102C and L112C mutants increase the cavity volume and size of opening beyond that seen in any of the WT TTSCP structures. Interestingly we also observed that in the W83C model, the cysteine residue became buried and no longer lined the cavity. This in combination with the fact that the W83C structure showed a smaller cavity size, suggests that this mutant may not be well suited to covalent modification with our NHC precursor.

*Table 4: Analysis using CASTp revealed different sized cavities in each SCP construct. # Pose 1 of NMR structure used to cavity analysis. * denotes data derived from AlphaFold2 model.*

Protein	**PDB code**	**Cavity internal area (Å^2)**	**Cavity volume (Å^3)**	**Cavity opening (Å^2)**
Human SCP_2L	1IKT	688.6	1,076.4	119.8
Human SCP_2L A100C	6Z1W	891.2	1,320.2	229.9
TT_SCP (chain A)	2CX7	432.6	468.8	22.4
TT_SCP (chain B)	2CX7	562.9	595.5	0.0
TT_SCP (NMR)#	1WFR	526.6	563.9	0.0
*TT_SCP W83C**	Model	412.7	448.9	0.0
*TT_SCP L102C**	Model	653.3	629.9	46.8
*TT_SCP ΔDSB L102C**	Model	645.0	672.4	69.3
*TT_SCP L112C**	Model	627.2	745.3	77.3

The Breslow intermediate **48** was docked into the mutant structures using AutoDock Vina[140] (Table 4). Although the affinities were poor and, in many cases, slightly positive, these poses indicated that there would be enough space within the cavity of the mutants to contain the

Breslow intermediate. We hypothesize that the weak, non-covalent binding of the Breslow intermediate is overcome by covalent modification of the cysteine residue by the maleimide thiazole. Again, these studies suggested W83C was the least promising of our chosen mutants displaying the least favourable binding affinities. Whilst all models provided feasible poses for the Breslow intermediate, in most cases the maleimide was not in the proximity of the engineered cysteine residue. The exception was L102C pose 7 where the maleimide is <5 Å from Cys102 (Figure 34). Resultantly we chose to generate a model of the L102C with the Cys13 and Cys63 residues mutated to alanine to create the TTSCP ΔDSB L102C scaffold. The AlphaFold2 model of this scaffold did not suggest the removal of the disulphide bond greatly altered the structure, however upon expression and purification we aimed to assess this experimentally.

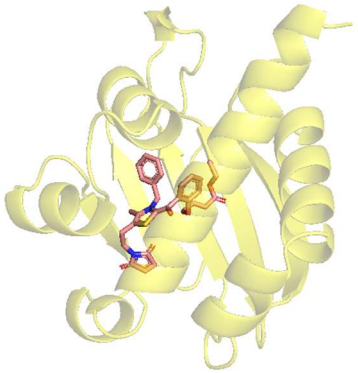
a)

Table 5: Affinities and structures of pose 1 generated by docking of the Breslow intermediate (8) into AlphaFold models of each of the TTSCP cysteine mutants.

W83C

mode	affinity (kcal/mol)	dist from best mode rmsd l.b.	rmsd u.b.
1	7.4	0.000	0.000
2	7.7	1.463	2.443
3	7.8	1.445	2.548
4	9.5	2.400	8.251
5	10.2	1.730	2.891

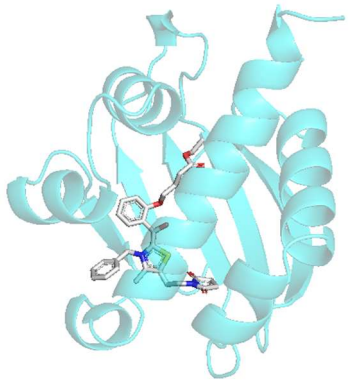
Writing output ... done.



L102C

mode	affinity (kcal/mol)	dist from best mode rmsd l.b.	rmsd u.b.
1	-2.0	0.000	0.000
2	-2.0	0.430	1.479
3	-1.7	1.335	2.108
4	-1.3	2.402	8.742
5	0.1	4.068	8.584
6	0.2	1.564	2.294
7	0.2	2.830	9.466
8	0.6	1.940	3.042
9	0.8	3.574	8.767

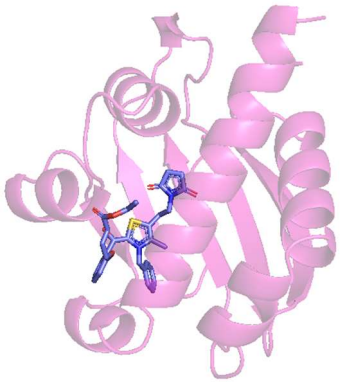
Writing output ... done.



L112C

mode	affinity (kcal/mol)	dist from best mode rmsd l.b.	rmsd u.b.
1	-6.4	0.000	0.000
2	-5.6	2.408	7.823
3	-5.5	2.234	7.627
4	-5.2	1.251	2.097
5	-4.7	2.395	7.410
6	-3.4	2.868	7.694
7	-3.3	2.662	6.124
8	-3.2	2.320	5.841
9	-2.9	2.533	7.813

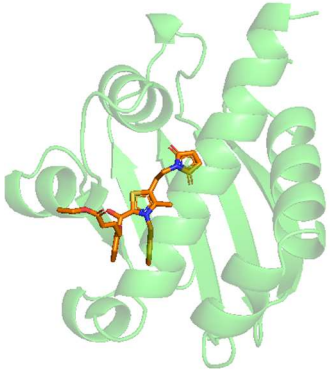
Writing output ... done.



ΔDSB L102C

mode	affinity (kcal/mol)	dist from best mode rmsd l.b.	rmsd u.b.
1	-0.5	0.000	0.000
2	-0.4	0.915	1.565
3	-0.2	0.939	1.982
4	-0.1	1.950	3.870
5	1.5	3.074	5.401
6	1.7	1.151	1.648
7	1.8	2.085	3.901
8	2.8	2.138	3.614
9	3.4	2.117	3.640

Writing output ... done.



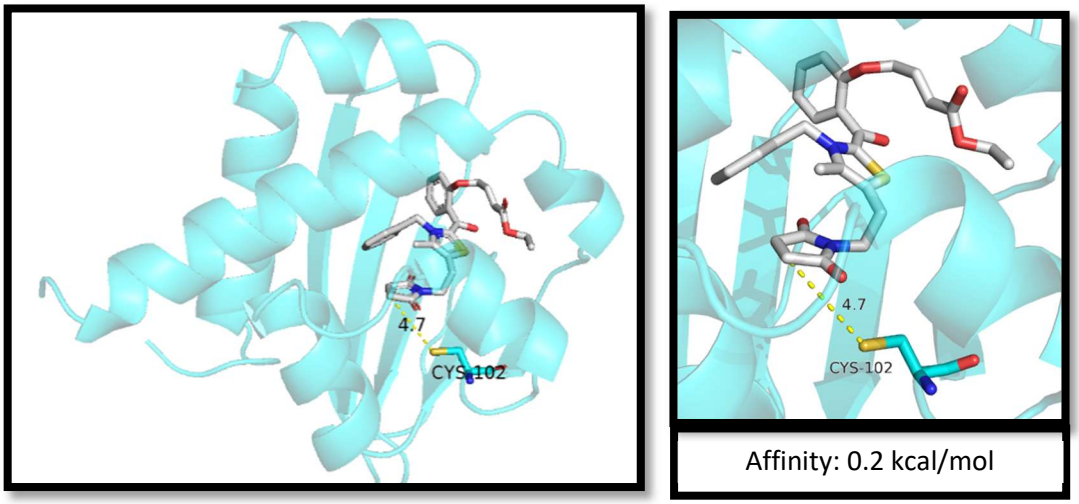


*Figure 34: Pose 7 of the Breslow intermediate of **1** and **2** docked into TTSCP L102C AlphaFold model shows plausible pose with maleimide 4.7 Å away from Cys102. Docking score gives a very slightly positive affinity of 0.2 kcal/mol.*

2.3.2.2 Cloning, Expression and Purification

hSCP mutants in pET 28a vectors were kindly gifted to us by Amanda Jarvis (University of Edinburgh). The wild type TTSCP, TTSCP W83C and TTSCP L112C constructs were prepared by Eva Klemencic (University of Edinburgh). These TTSCP genes were codon optimised for *E. coli* expression and inserted into the pET28b vector. Due to the high GC content of TTSCP, mutagenesis was particularly challenging and therefore L102C mutants were ordered from GenScript, which were codon optimised for expression in *E. coli* K12 and for lower GC content to allow future mutagenesis/engineering.

hSCP and TTSCP mutants were expressed in *E.coli* and purified using IMAC. The hexa-histidine tag (His tag) was cleaved using Tev protease and the cleaved protein was isolated by removing the His tag with a second round of IMAC. Pure proteins of the expected mass were observed by SDS PAGE analysis (Figure 35). Proteins were all expressed in good yields and the mass of each mutant was verified by LC-MS (Appendix 1.2).

Table 6: SCP purification yields and masses of mutants obtained by LC MS analysis.

Protein	**Yield (mg/L)**	**Expected mass (Da)**	**Observed mass (Da)**
V83C	27	13376.61	13376.7
V83C Stable	22	13403.68	13403.8
A100C	11	13404.66	13404.7
A100C F94H	48	13394.63	13394.8
Q111C	32	13347.61	13347.8
TTSCP	54	14865.04	14863.0 (accounts for DSB)
TTSCP W83C	28	14781.9	14780.3 (accounts for DSB) 15046.2 (Stearoylation)
TTSCP L102C	49	14855.02	14853.6 (accounts for DSB)
TTSCP L112C	55	14855.02	14853.4 (accounts for DSB)
TTSCP ΔDSB L102C	24	14790.90	14790.9

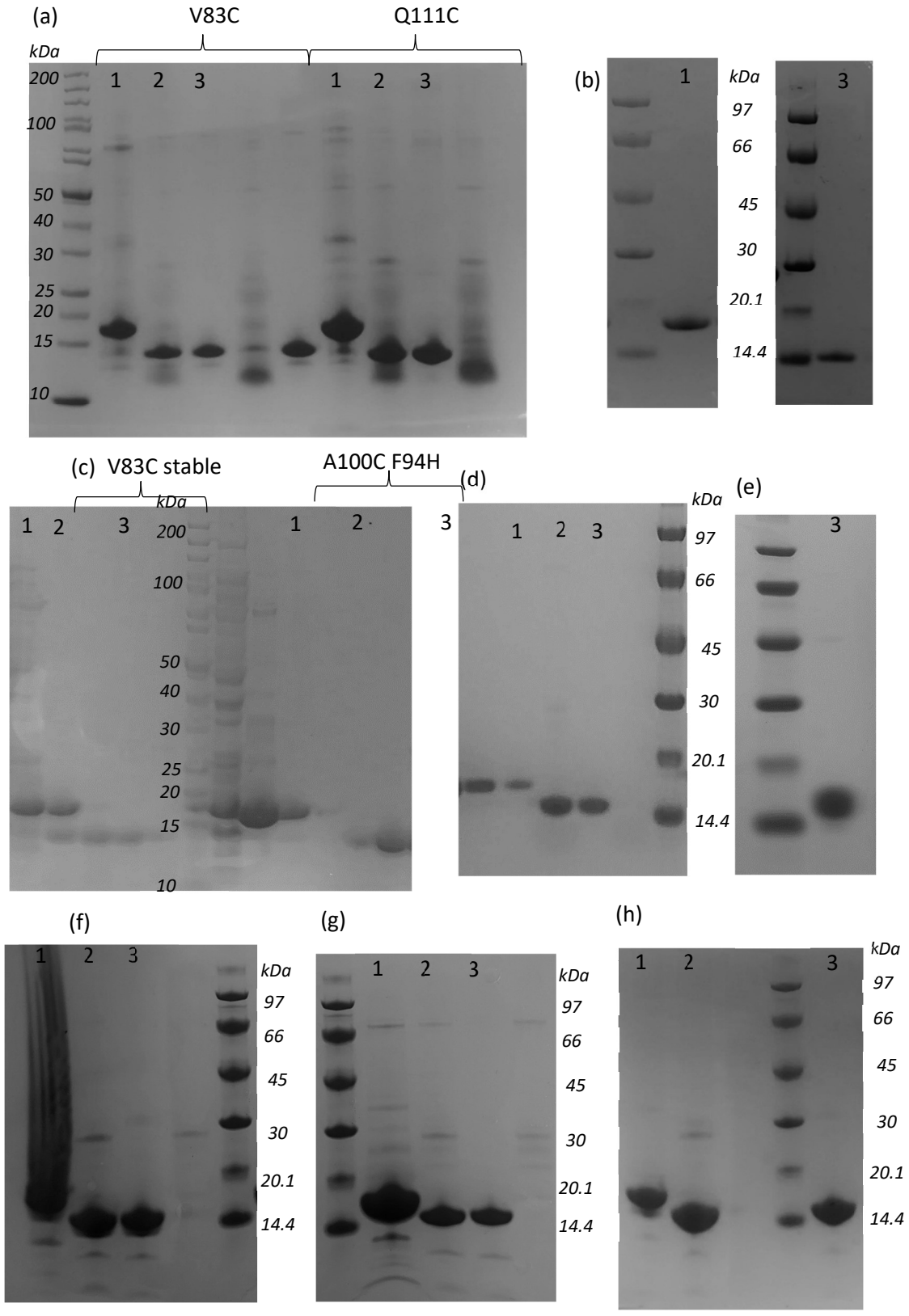


Figure 35: SDS-PAGE analysis of purified SCPs. Lanes contained HisTag protein are labelled 1, TEV protease cleaved proteins lanes are labelled 2 and the lanes with the final purified cleaved protein are labelled 3. a) hSCP V83C and hSCP Q111C b) hSCPA100C c) hSCP V83C Stable and A100C F94H, d) TTSCP e) TTSCP W83C f) TTSCP L102C g) TTSCP ΔDSB L02C h) TTSCP L112C.

2.3.2.3 TTSCP Circular Dichroism

To assess the effect of the removal of the disulfide bond from TTSCP L102C we used circular dichroism (CD). A Chirascan VX circular dichroism instrument was used to collect spectral data for TTSCP, TTSCP L102C, TTSCP ΔDSB L102C under identical conditions, 0.15 mg/mL protein in 10 mM phosphate buffer at 25 °C from 190 to 260 nm (Figure 36). Whilst the three plots have the same shape there is a greater range in the delta epsilon values of TTSCP ΔDSB L102. This data was processed using the DichroWeb server[199], with the CDSSTR algorithm[200] using reference dataset 7[201], to estimate the secondary structure of the proteins (Table 7). Whilst no significant structural changes were observed between the constructs unfortunately the data is not supported by the crystal structure of TTSCP (2CX7), which shows a different distribution of structural features. This highlights some of the limitations of CD analysis, for example its sensitivity to the protein concentration and its dependence on the reference data set used[202]. Therefore, to further analyse the secondary structure of the TTSCP mutants, crystal structures should be obtained.

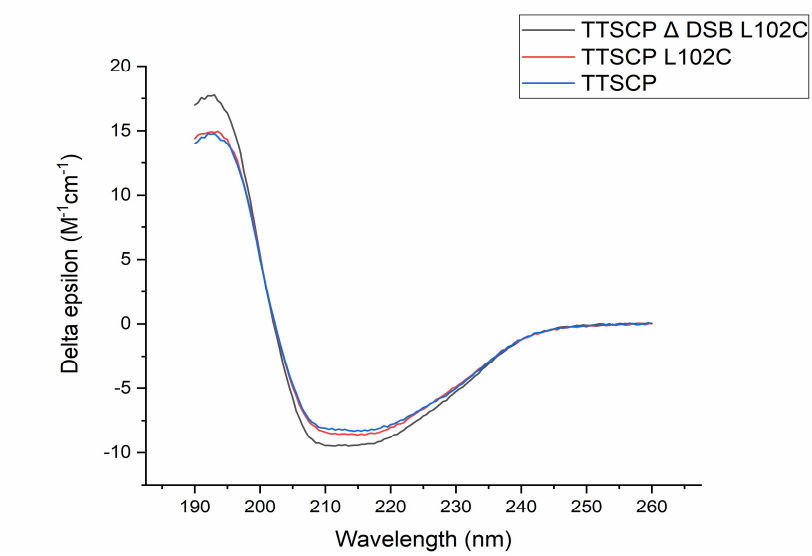


Figure 36: A plot of CD analysis of TTSCP, TTSCP L102C and TTSCP ΔDSB L102C showing the change in delta epsilon with wavelength.

Table 7: The Secondary structure of TTSCP and its L102C mutants, generated by processing CD data using the DichroWeb server, compared to the secondary structure extracted from the crystal structure of TTSCP pdb:2CX7 (grey).

Protein	α- helix	β-sheet	Other
2CX7 crystal structure	39	26	35
TTSCP	84	7	9
TTSCP L102C	79	11	11
TTSCP ΔDSB L102C	86	7	6

To determine the melting temperature of the protein, spectral data was collected from 25 °C to 100 °C in 5 °C increments. The protein was then cooled to 25 °C to assess refolding. Single wavelength melting curves were obtained by plotting the measured values of the 194 nm, 208 nm, 218 nm, 223 nm peaks of each spectrum against temperature (Figure 37). Data for the WT TTSCP was collected by Eva Klemencic (University of Edinburgh). The data shows the T_m of TTSCP and TTSCP L102C has not been reached therefore T_m >95 °C whilst the data for TTSCP ΔDSB L102C shows the T_m to be 79 °C with analysis of the change in delta epsilon at 223 nm, 212 nm, 208 nm, and 194 nm all in agreement. This T_m is similar to that of human

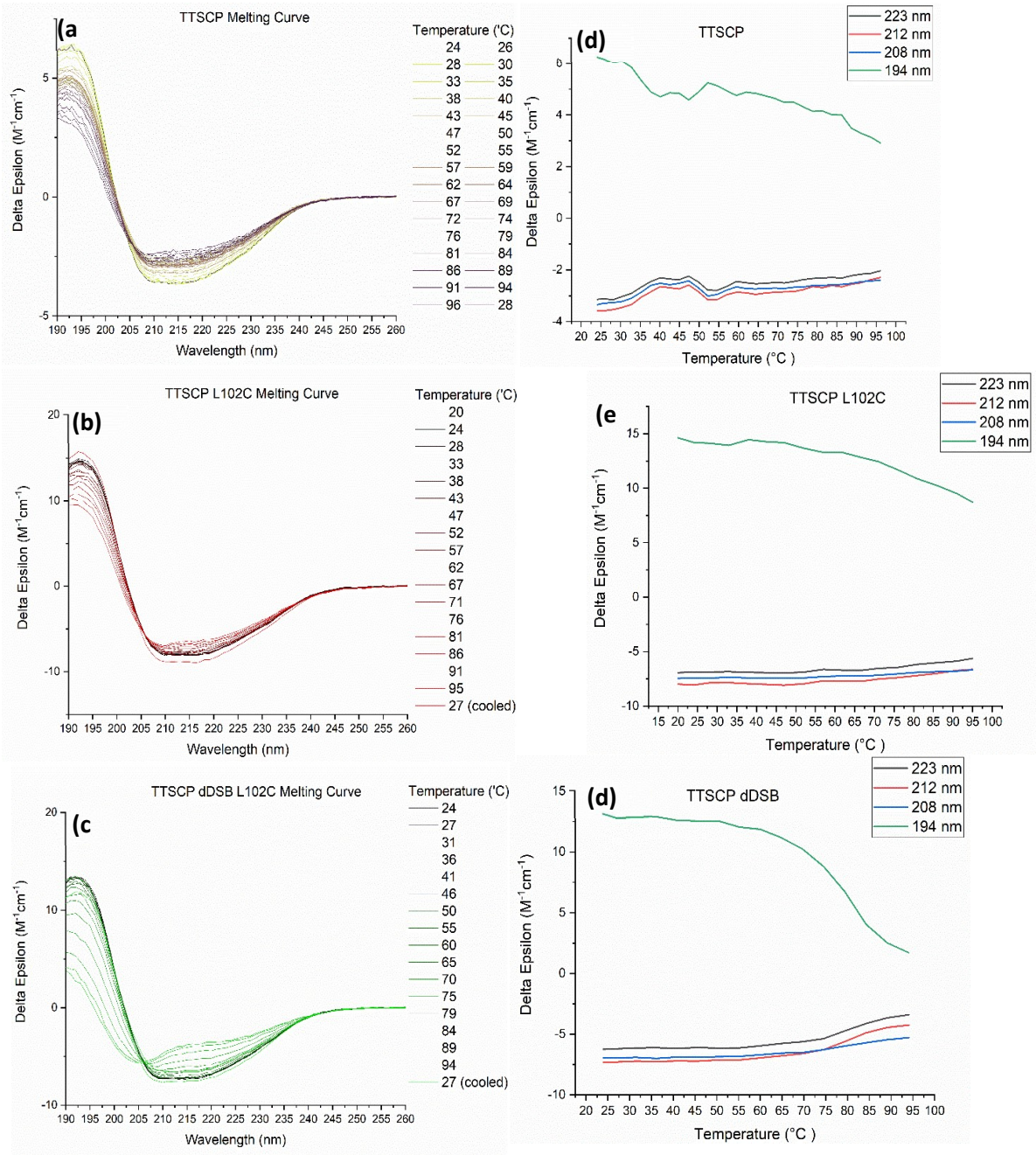


Figure 37: Melting curves of TTSCP mutants obtained by CD (a-c) and secondary plots of the delta epsilon at 223, 212, 208 and 194 nm to allow determination of the melting temperature.

SCP-2L reported in the literature (T_m 80 °C)[203]. These promising results indicate removal of the DSB from TTSCP does not completely destabilise the structure or cause the protein to be in an unfolded state.

2.3.3 Thiamine Phosphate Synthase

For our final two scaffolds we chose to use proteins which naturally bind TPP but do not use it for catalysis. This way we could utilise a monomer with a cavity evolved to accommodate a thiazole. As such we chose to investigate the use of thiamine phosphate synthase (TPS), an α/β barrel protein with a triosephosphate isomerase (TIM) barrel protein fold, which catalyses the formation of thiamine phosphate (TP) from 4-amino-5-(hydroxymethyl)-2-methylpyrimidine pyrophosphate (HMP-PP) and 5-(hydroxyethyl)-4-methylthiazole phosphate (Thz-P, Figure 38). The TIM barrel fold is one of the most widespread protein folds which is found in 10% of enzymes, usually forming the catalytic domain[204]. These TIM barrel proteins utilise a broad range of cofactors which leads to huge functional diversity[204]. Interestingly TIM barrel proteins have also been used for the design of other artificial enzymes due to the binding pocket providing a large number of positions pointing into the cavity which can be manipulated both for use as catalytic residues and for binding the transition-state [32, 205].

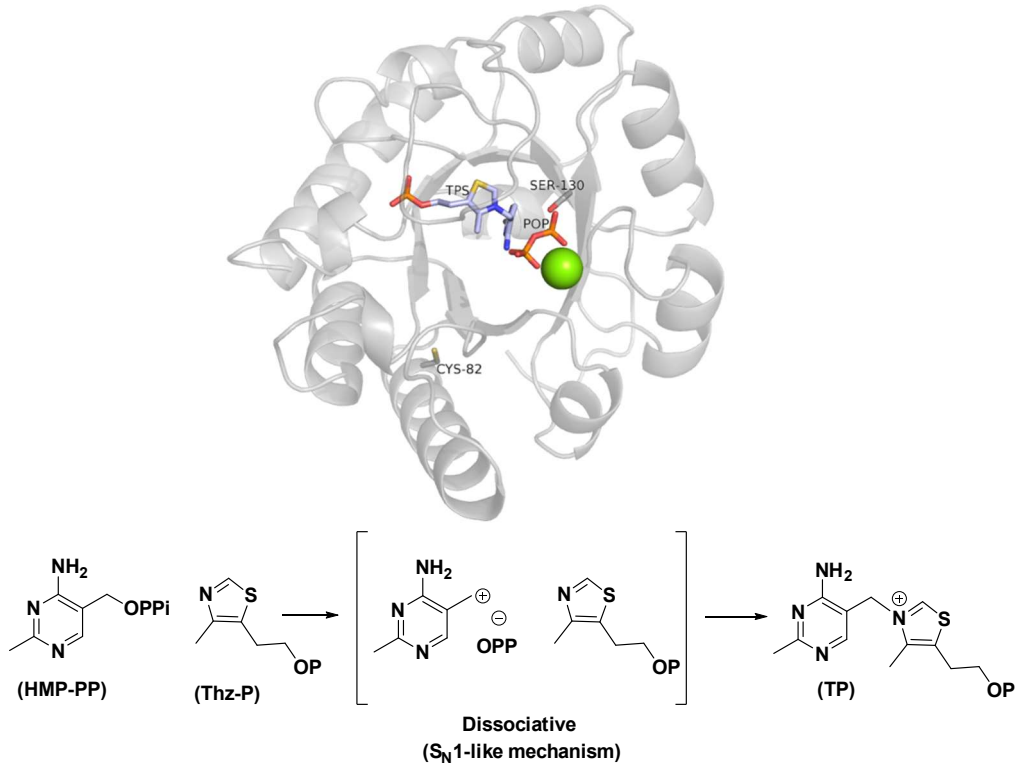


Figure 38: Structure of Bacillus subtilis TPS (PDB: 2TPS) and the proposed mechanism highlighting the dissociative S_N1-like mechanism which forms a carbocation intermediate.

Although TPSs can form multimers they are functional monomers with TPS being the first example of a thiamine binding protein using a single structural domain for binding thiamine[206]. Several crystal structures of *Bacillus subtilis* TPS have been determined and used to elucidate the binding site and catalytic mechanism (Figure 38)[206, 207]. These structures highlight a catalytic serine (S130 in *B. subtilis*) residue which activates the pyrophosphate as a leaving group. Creation of an S130A mutation confirmed the importance of this residue and inactivated the protein with a $>10^4$-fold decrease in activity[206, 207]. Begley *et al* confirmed the reaction mechanism was a dissociative S_N1 like mechanism proceeding via a carbocation intermediate (Figure 38)[208]. Other conserved features of TPSs include a pyrophosphate binding site, a solvent accessible magnesium binding site containing two aspartate residues and a thiamine binding site located near the hydrophobic core of the barrel which holds thiamine phosphate in a "V" conformation and forms hydrogen bonding interactions with its phosphate moeity[206].

TPS is encoded by the ThiE gene, present in many organisms. To choose our protein scaffold we investigated some of the many sequences available, some of which are bi-functional and also catalyse the phosphorylation of 4-methyl-5-hydroxyethylthiazole such as the *Candida glabrata* enzyme[209]. We were particularly interested in the ThiE gene from *Pyrococcus furiosus* (pfThiE), for which a crystal structure 1XI3 is available, as this is a thermophilic microorganism. We performed a sequence alignment using Clustal Omega and visualised it with ESPript to determine if pfThiE contained conserved features (Figure 39)[210-212]. We were pleased to observe that pfThiE contained all of the conserved features for binding TP, showing 33% identity and 53% similarity to *B. subtilis* TPS. In addition, pfThiE did not containing any cysteine residues which is beneficial for protein modification. Interestingly all the other strutures contain a conserved cysteine residue at position 59 (position 82 in *B. subtillis*, blue triangle Figure 39). The pfThiE structure, 1XI3, does not contain any ligands but overlay with the *B. subtillis* struture (2TPS) shows simular architeture, positioning of catalyitc serine residue and allows us to identify poteintial binding sites. CASTp analysis[185] highlights a large cavity (volume 1XI3: 1149.6 Å^3, 2TPS: 1229.6 Å^3) inside the TIM barrel with a flexible loop forming a lid, although this loop is not resolved in the pfThiE crystal structure (1XI3)(Figure 40). This loop, often referred to as the C-loop, covers the active site, opening to allow the entrance of substrates and release of products[206]. Interestingly the addition of a "lid" to a nitrobindin based ArM to attempt to prevent solvent exposure of the cofactor was shown to be beneficial[213]. We hope this C-loop will be similarly useful in maintaining the environment of the active site in our artificial enzyme.

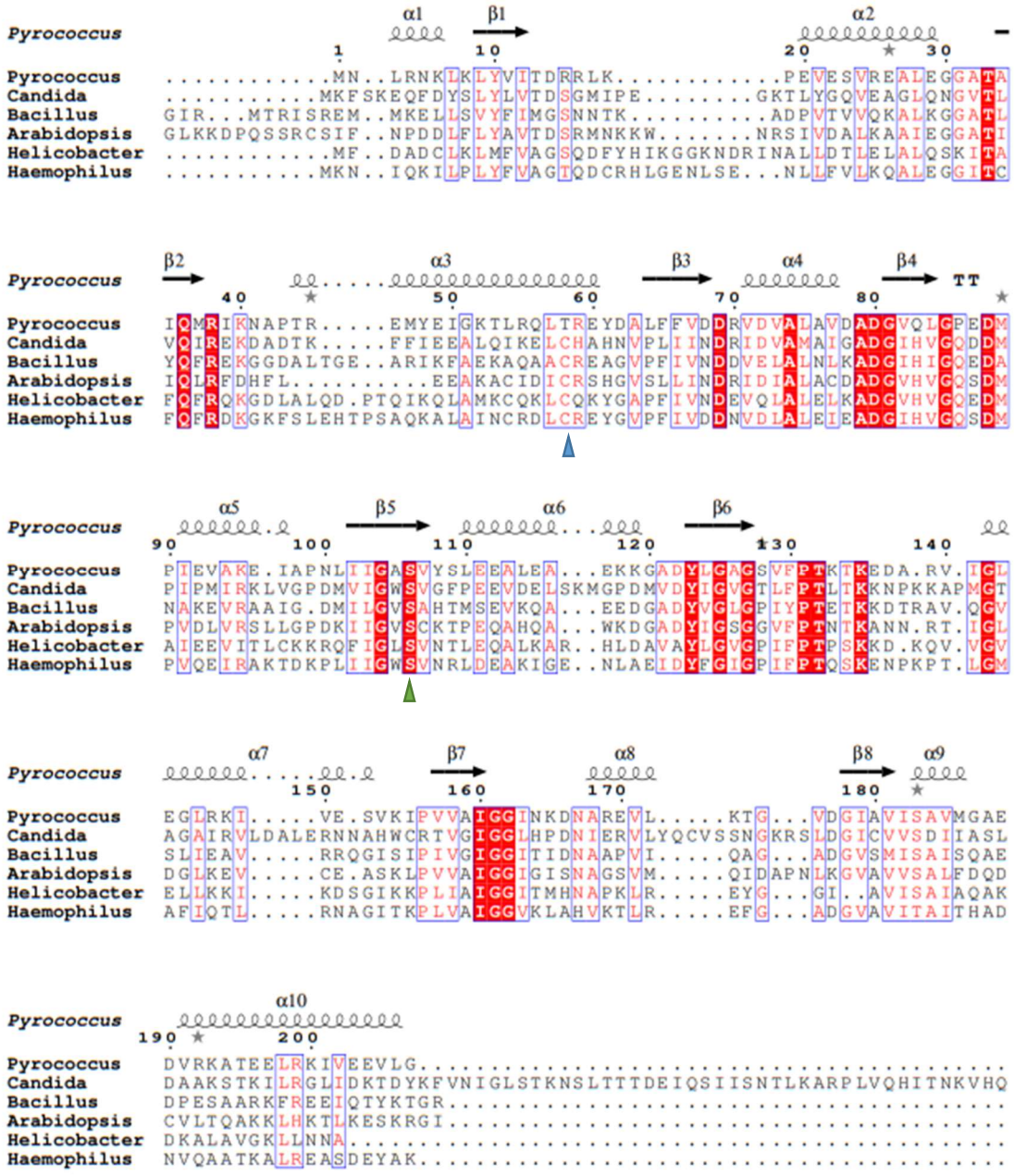


Figure 39: Sequence alignment of ThiE genes from Pyrococcus furiosus, Candida glabrata, Bacillus subtilis, Arabidopsis thaliana, Helicobacter pylori and Haemophilus influenzae created using Clustal Omega and visualised it with ESPript . Conserved residues are highlighted in red with a green arrow marking the catalytic cysteine and a blue arrow marking a cysteine residue present in all but pfThiE.

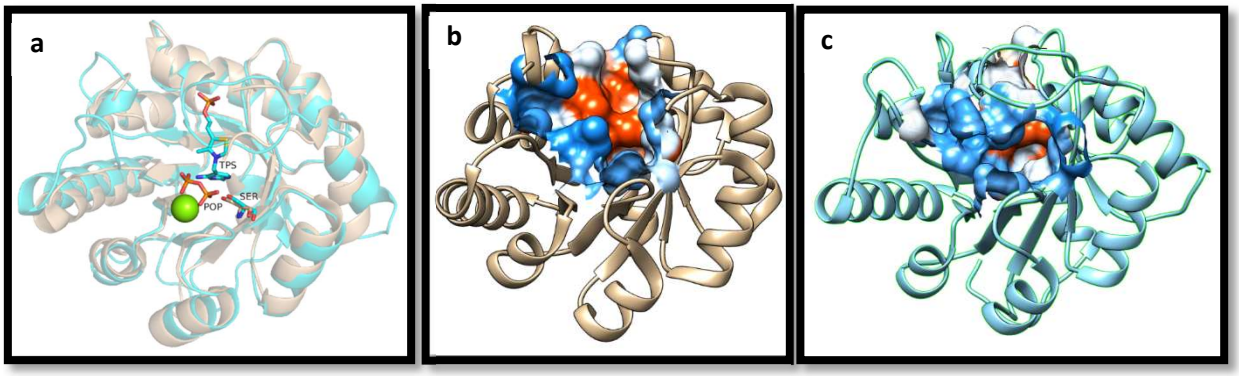


Figure 40: a) Structural overlay of pfThiE (1XI3, brown) and B. subtilis (2TPS, blue) shows the similarity between the two structures and how the ligands from 2TPS transpose into the cavity of 1XI3 b) Rendering of the internal surface of the pfThiE 1XI3 cavity. c) Rendering of the internal cavity of B. subtilis 2TPS which highlights how the disordered loop forms a "lid" over the cavity.

2.3.3.1 Selection of residues for modification

To identify suitable sites to mutate to a cysteine residue and allow modification of pfThiE, docking studies using AutoDock Vina[140] with MBnThz and the Breslow intermediate **48** discussed in pervious sections were performed. The most favourable position for MBnThz was in the centre of the cavity with the maleimide located in the proximity of the magnesium binding site (Figure 41). We opted to analyse the poses of the Breslow intermediate to choose location of our cysteine mutations as this allowed the substrate to occupy the hydrophobic tunnel. From the first three poses of the Breslow intermediate, all of which displayed favourable binding affinities for the Breslow intermediate, several potential residues were identified.

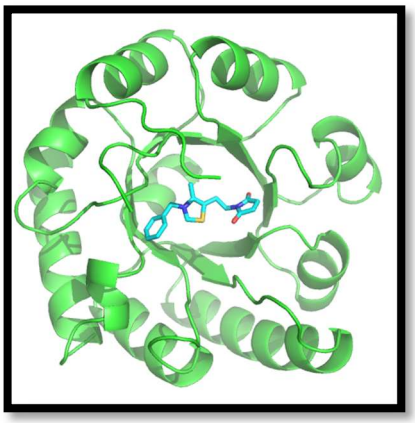
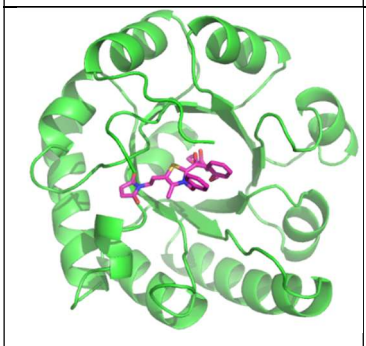
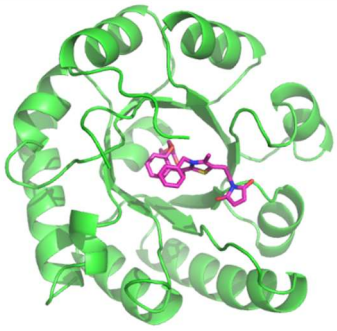
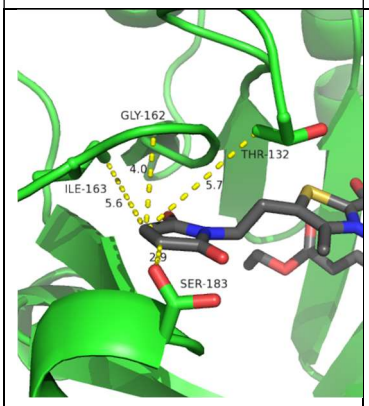
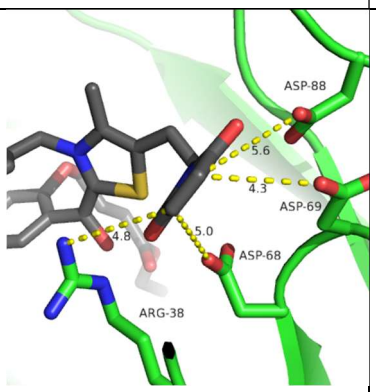
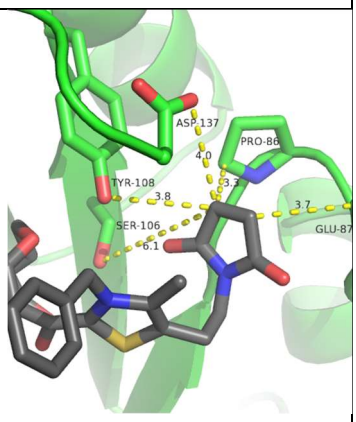


Figure 41: Pose 1 of the docking of MBnThz into pfThiE shows the ligand sits in the centre of the cavity.

In pose 1, the maleimide sits in the section of the thiamine phosphate binding cup which usually forms hydrogen bonding interactions with the phosphate of TP (Table 8). From these residues we chose to mutate T132 to Cys. Although it is not the closest to the maleimide, it sits at the start of the flexible loop which we hope would allow easy functionalisation. Pose 2 positioned the maleimide in the magnesium binding site. From this we decided to mutate

D69 as it was the closest to the maleimide. We also hoped this would disrupt magnesium binding which would be detrimental to the enzyme's normal function, reducing the possibility that we would observe unwanted side reactions. Finally, we had considered mutating the catalytic serine residue to inactive the enzyme. As pose 3 placed this residue in the proximity of the maleimide we decided to make the S106C mutation for our final cysteine mutant.

Table 8: Analysis of the top 3 poses generated by docking the Breslow intermediate into pfThiE highlights residues in the proximity of the maleimide as potential positions for mutagenesis.

Pose1	Pose2	Pose3
		
		

2.3.3.2 *Cloning, Expression and purification*

The synthetic gene of pfThiE inserted into a pET28a vector between NdeI and XhoI restriction sites was ordered from GenScript to enable expression of N-terminally histidine tagged recombinant pfThiE. Mutagenesis primers were designed for the D69C, S106C and T132C mutations using the method described by Liu and Naismith[187]. Using pET28a PfThiE as a template for site directed mutagenesis, the whole plasmid was amplified with the primers incorporating the mutations into the gene. The introduction of the D69C, S106C and T132C mutations were confirmed by sequencing analysis.

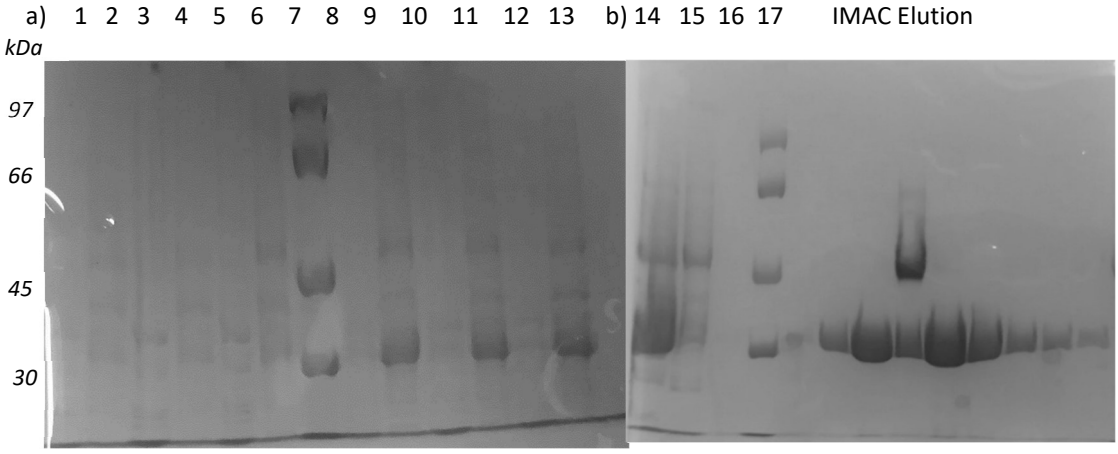


Figure 43: SDS PAGE analysis of WT pfThiE expression tests and IMAC purification. a) shows expression tests at 20 °C (Lanes 1-6) and 30 °C (lanes 8-13) showing insoluble proteins (lane 1, 3, 5, 8, 10, 12) and soluble lysate (lanes 2, 4, 6, 9, 11, 13) after 18 hours induction with IPTG (0.1 mM, 0.5 mM and 1 mM left to right).b) lane 14- 17 show pfThiE BL21 DE3 expressed cell lysate, IMAC flow through, IMAC wash , LMWM and IMAC elution respectively. An impurity of approx. 45 kDa can be observed in IMAC elution which was not removed in the wash step.

The plasmids were transformed into BL21 (DE3) cells and expression tests were undertaken of the WT pfThiE. The protein expressed well at 30 °C with 0.1 mM IPTG showing the largest band of overexpressed protein (Figure 43). Large scale expression and purification via IMAC resulted in the co-purification of pfThiE with an impurity of approx. 45 kDa. This impurity could be avoided by discarding the contaminated fraction; however, we discovered that heat treating the protein prior to purification effectively removed this contaminant and enabled higher protein yields (Figure 43). Heat treatment at 80 °C for 30 mins removed almost all impurities from the protein, however analysis by LCMS only identified protein of the expected mass after further purifications via IMAC, likely due to other components of the lysate remaining after heat purification.

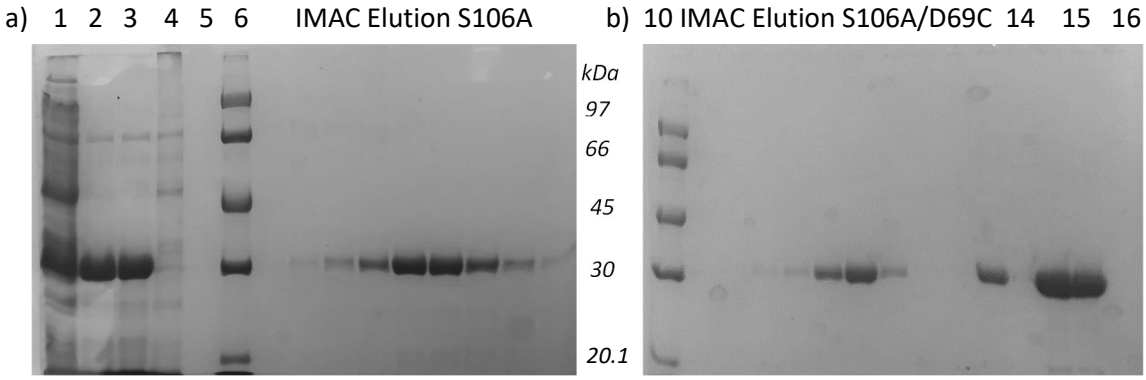


Figure 42: SDS PAGE analysis of the purification of pfThiE mutants a) pfThiE S106C 1. Lysate, 2. heat purified 80 °C 20 min, 3. heat purified 80 °C 30 min 4. IMAC flow through, 5. IMAC wash, 6. LMWM, subsequent lanes show fractions from the IMAC elution which pooled to give purified protein b) 10. LMWM, subsequent lanes show fractions from the IMAC elution of pfThiE S106A/D69C which pooled to give purified protein, 14. Purified pfThiE T132C 15. Purified pfThiE S106C. 16. Purified pfThiE D69C

Table 9: A summary of the expected and observed masses of pfThiE mutants analysed by LCMS

	Expected mass (Da)	Expected mass, N-Met Cleavage (Da)	Observed masses (Da)	Tev Cleaved Expected mass (Da)	Cleaved mass observed (Da)
WT	24995.61	24864.42	24864.7	-	-
S106A	24979.61	24848.42	24848.9	-	-
S106C	25011.67	24880.48	25001.7 (+121 adduct)	-	-
D69C	24983.66	24852.47	24852.9	-	-
T132C	24997.64	24866.45	24866.3 and 24947 (+80, phosphorylation)	22786.33	22786.1
S106A/D69C	24967.66	24836.47	24837.4	22756.34	22756.8

Purification of the three mutants using heat treatment followed by IMAC appeared to produce pure protein of approximately the expected mass when analysed via SDS-PAGE (Figure 42). However, LC MS analysis of the proteins led to unexpected masses being observed with only the D69C mutant displaying the expected mass. WT pfThiE showed an adduct of 177 Da whilst the T132C mutant revealed an adduct of 80 Da which was attributed to protein phosphorylation (Figure 44,Table 9). Phosphorylation is a common post translational modification but due to that fact we only observed this adduct in the T132C mutant it may be occurring due to the engineered cysteine residue. We were happy to see that upon cleavage of the His tag using TEV protease we no longer observed this adduct. The most unusual adduct we observed was on the S106C mutant which showed an adduct of +122 Da from the expected mass. We attributed this to covalent addition of the native reaction intermediate to the protein (Figure 38, Figure 44). It is possible that the cysteine acts as a nucleophile and forms a thioether with the carbocation intermediate. Although we did not add any of the substrate to our cells, TPSs are purified with substrates bound[208]. The carbocation would form in the correct position to undergo addition to the cysteine as this is in the position that the catalytic serine would normally be located. This unwanted adduct is problematic as, if it is forming on the cysteine residue, it will prevent functionalisation of this protein scaffold with an NHC precursor. To test this hypothesis, we attempted to functionalise the protein with iodoacetamide. Whilst the adduct was not removed, the small amount of enzyme which did not display the adduct became functionalised (Figure 44). The presence of the native substrates of pfThiE in purified protein concerned us as we did not want these to interfere with our reactions. Therefore, we made two additional mutants, S106A and the double mutant D69C/S106A. Primers were designed and site directed mutagenesis was undertaken on the pET28a pfThiE and pET28a pfTHiE D69C plasmid to create pfThiE S106A mutant and the D69C/S106A. Both mutants expressed and purified well, giving the expected masses when analysed by LCMS (Figure 44 Table 9)

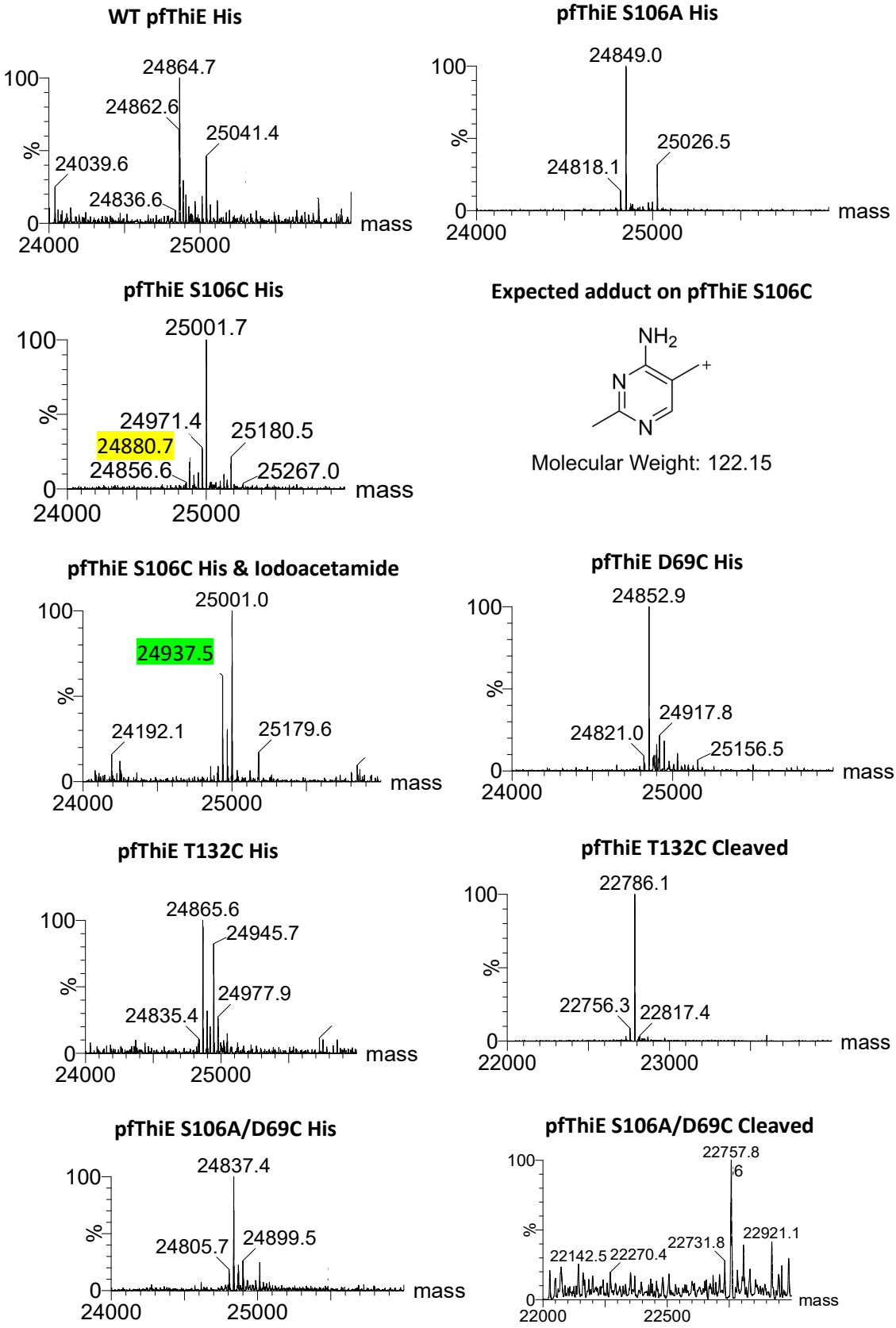


Figure 44: LCMS analysis of pfThiE mutants. Deconvoluted spectra show adducts present for most of proteins. In particular analysis S106C shows most of the protein is + 121 kDa larger than expected with only a small amount of protein, highlighted yellow at the expected mass of 24880 Da. The structure of the expected adduct, the carbocation intermediate of the pfThiE catalysed reaction is shown. LCMS analysis of the S106C mutant after Incubation with iodoacetimide highlights majority of the protein remains at the same mass but the peak indicating a mass of 24880 has disappeared whilst a new mass of 24937 is observed (green). Tev Cleavage of pfThiE T132C and S106A/D69C shows the expected protein mass.

2.3.4 Thiaminase I

Like pfThiE, thiaminase naturally binds TPP but does not use it for catalysis. There are two classes of thiaminases which are found in a wide variety of organisms, including plants, fish, and bacteria[214]. They are responsible for the breakdown of thiamine by catalysing the replacement of the thiazole moiety of thiamine with a variety of nucleophiles (thiaminase I) or specifically water (thiaminase II) [215]. We chose to focus on thiaminase I from *Bacillus thiaminolyticus* which was first characterised by Begley *et al.* and has an available crystal structure (PDB: 3THI) [215, 216].

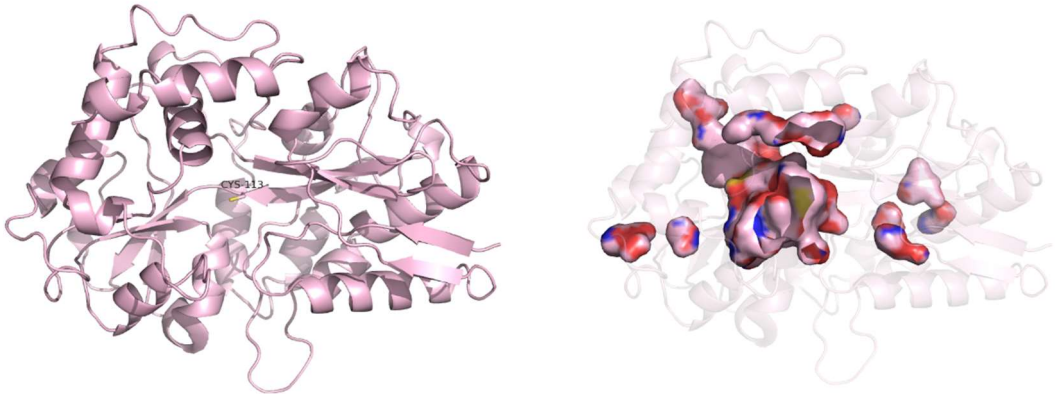


Figure 45: The strucutre of thiaminase 1(PDB: 3THI) highlighting the catalytic cysteine 112 . b) shows the internal surface of the groove betweeen the two domains of 3THI where catalysis takes place.

The thiaminase-I monomer is comprised of two globular domains (N- and C- domains) each with an α/β toplogy[216] (Figure 45a). The two domains are separated by a deep groove, 15 Å deep, 12 Å wide, and 15 Å long, which forms the active site[216] (Figure 45b). Inside this cleft is the only cysteine residue in *B. thiaminolyticus* thiaminase I, Cys113, which has been identified as the active site nucleophile[215]. A C143S mutant of the *Clostridium botulinum* thiaminase I was used to obtain a thiamine bound crystal structure which revealed further information about the active site architecture, highlighting residues involved in binding and catalysis[214].

We ordered a histidine tagged synthetic gene of *B. thiaminolyticus* thiaminase I with the hope of functionalising the active site cysteine residue of this scaffold. Expression tests were performed using LB and PB media at 20 and 30 °C with varying concentrations of IPTG to induce expression. A band of overexpressed protein was observed around the expected molecular weight for our histidine tagged thiaminase 1 (47.6 kDa Figure 47). Interestingly in the expression tests with LB, there appeared to be two proteins formed. With PB, a broad band was formed which could also be composed of two different proteins. It has previously been reported that thiaminase 1 degrades when cultured for extended periods, leading to an ~ 1 kDa decrease in molecular weight[217]. We hypothesised this may be why we observed multiple bands during SDS-PAGE analysis.

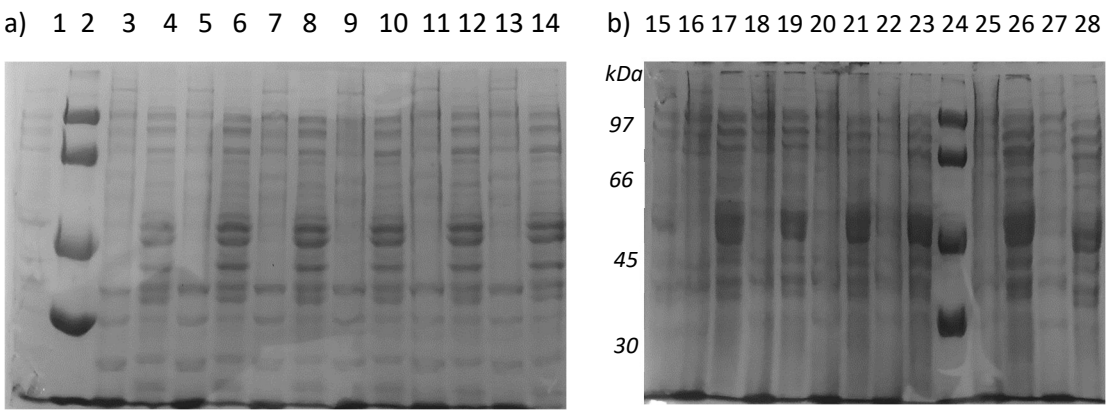


Figure 47: Thiaminase 1 expression tests in BL21 DE3 cells at 20 °C (lanes 3-8 and 16-21) and 30 °C (lanes 9-14 and 22-28) for 18 hours in a) LB b) PB. Lanes 2 and 24 contain LMWM lanes 1 and 15 show uninduced cells whilst lanes 3, 5, 7, 9, 11, 13, 16, 18, 20, 22, 25 and 27 shows insoluble proteins and lanes 4 ,6, 8, 10, 12, 14, 17, 19, 21, 23, 26 and 28 show soluble proteins.

Unfortunately, purification of thiaminase 1 proved difficult with the histidine tagged protein not binding to the HisTrap (Figure 46). Purification by ammonium sulphate precipitation followed by ion exchange chromatography yielded a small amount of pure protein and other samples of less pure protein. Mass spec analysis of this protein revealed that the molecular weight matched that of cleavage of a putative N-terminal signal sequence at three contiguous residues[215]. Re-cloning of the protein into a vector with a C-terminal His-tag would allow increased yields of purified protein. However, the protein would still display a range of molecular weights depending on the point of signal sequence cleavage. Therefore, we decided to move on with our other protein scaffolds.

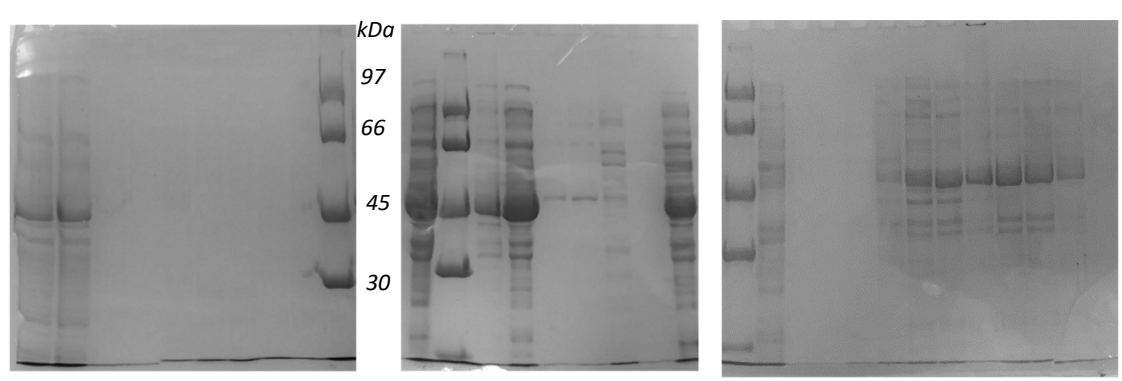


Figure 46: SDS PAGE analysis of the purification of Thiaminase 1. a) HT showed protein did not bind to column as FT (lane 2) was same as load (lane 1) whilst no protein was observed in the elution. b) Analysis of IEX chromatography of Thiaminase 1 on a DEAE column. Lane 1. IEX load (ysate after ammonium sulphate ppt) 2. LMWM, 6-11. IEX elution, 12. IEX flow through. c) SDS page analysis of a second IEX column loaded with the flow through of the first column. 13. LMWM, 14. Flow through.

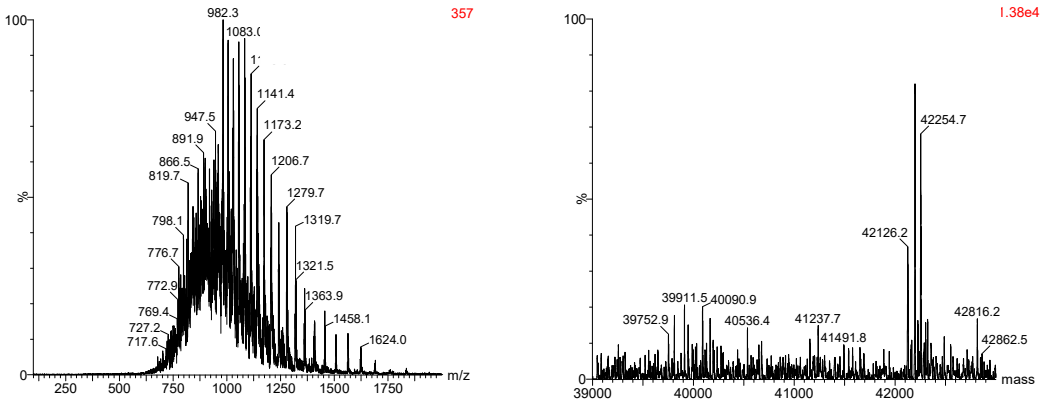


Figure 48: LCMS analysis of recombinantly expressed thiaminase I shows fragmentation to form three distinct peptides corresponding to the mass expected from cleavage at three consecutive residues which follow the signal sequence.

2.4 Functionalisation of proteins

To analyse functionalisation of the protein LC MS was employed (materials and methods). We determined the degree (%) of functionalisation by comparing ion intensities of unmodified with modified forms. Raw ESI-MS data of the ion series is available in the appendix to allow comparison. Whilst this method is not 100 % accurate, Kelleher *et al.* show ionisation effects on different proteoforms (i.e. different modified forms of the same protein) are minimal and so direct ratios are relevant[218]. We attempted to use commercially available papain as a control to replicate the work of Suckling[128]. However due to papain being isolated commercially from a papaya extract it appeared to contain impurities and it did not give defined peak on the LC-MS. Furthermore, papain contains seven cysteine residues, and although six of these are occupied in forming disulfide bonds, functionalisation can still occur at multiple sites. This can lead to several different species, which further dilutes the MS sample. As such, we decided to proceed with our recombinant protein scaffolds, most of which contained a single cysteine residue, so that functionalisation could be analysed by LC MS.

2.4.1 Functionalisation of hSCPA100C with Bromothiazoles

We chose to use hSCPA100C for initial functionalisation trials as this mutant is well studied and a published crystal structure is available (pdb 6Z1W). Functionalisation of hSCPA100C with our bromothiazoles (**29** and **31**) was attempted using conditions similar to those described by Suckling[128]: Protein in 10 mM sodium phosphate buffer pH 7 was incubated with 10 eq of bioconjugation reagent (in 1:1 water/DMSO) overnight at 4 °C. Subsequent additions of bioconjugation reagent were not carried out in this instance as we wanted to assess if functionalisation occurred before optimising conditions to take the reaction to completion. Samples were buffer exchanged to remove excess bioconjugation reagent using centrifugal filtration. Initial LC-MS analysis indicated that functionalisation did not occur with the expected functionalised mass. Nevertheless, upon functionalisation, an adduct was observed using both methyl and benzyl thiazolium bromides. Unexpected mass shifts of +117 for functionalisation with BnThzBr reaction and +41 and +82 (+ 41 twice) for the MeThzBr reaction were observed with less unfunctionalised protein remaining in the reaction with the methyl derivative (Table 10, Figure 49).

*Table 10: Expected and observed masses for the functionalisation of hSCP A100C with bromothiazolium salts **29** and **31**.*

	Expected mass (Da)	Observed mass (Da)	Mass shift (Da)
Unfunctionalised	13404.66	13404.74	-
Functionalisation with MeThzBr 29	13641.73	13404.71, 13445.42, 13486.24	+40.76 +81.58
Functionalisation with BnThzBr 31	13718.05	13404.84 13521.86	+117.20

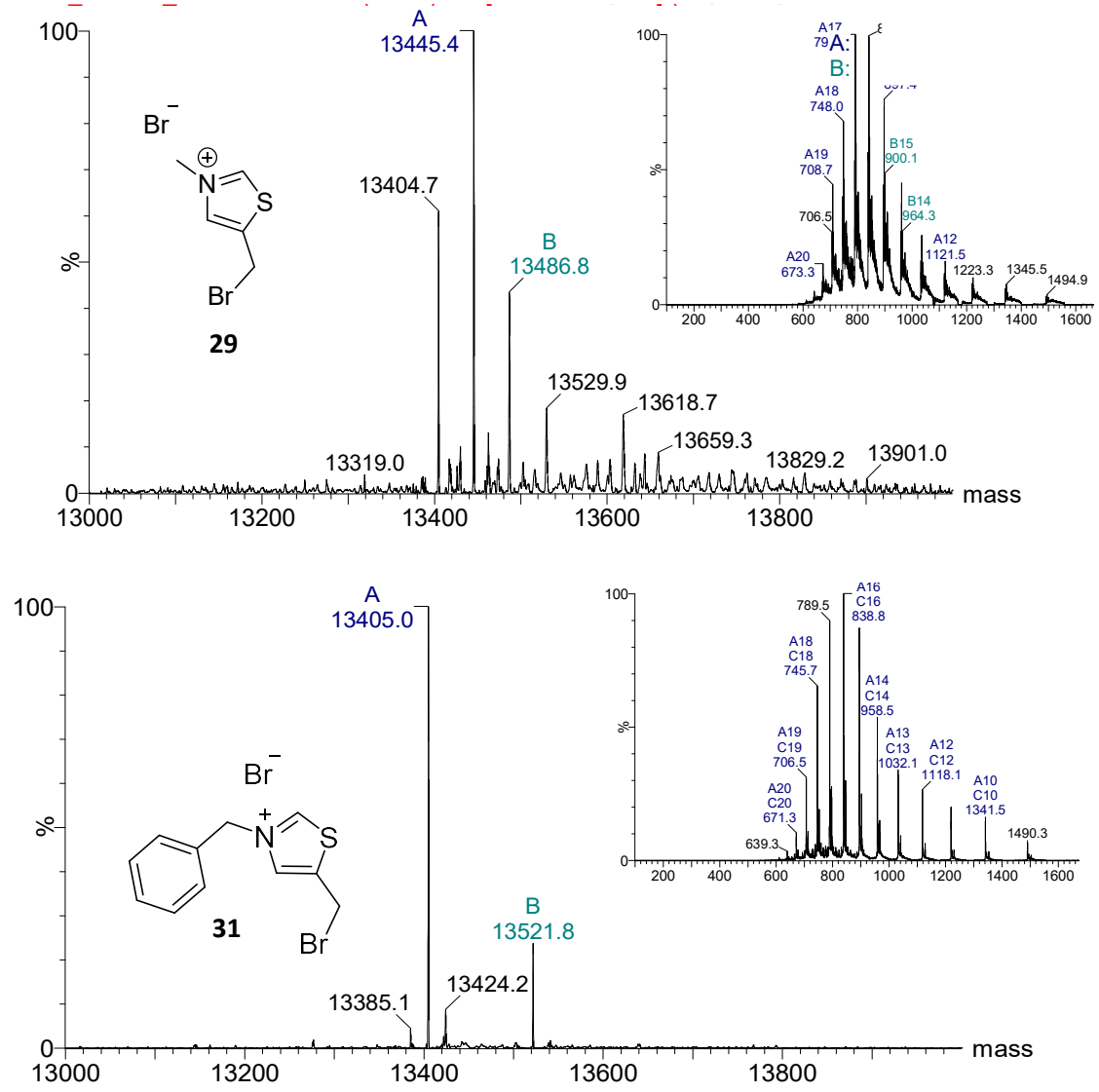


*Figure 49: LCMS analysis of hSCP A100C functionalisation with bromothiazolium salts **29** and **31.** The structures of **29** and **31** are shown along with both the ion series and deconvoluted spectra of the protein obtained from the functionalisation reaction. Masses were observed for the unfunctionalised protein as well as protein adducts of + 41 for **29** and +117 for **31**.*

Interestingly the difference in mass between the two adducts corresponds to the mass difference observed between the methyl and benzyl derivatives. After analysing possible fragmentation patterns of the bromothiazoles which could lead to adducts with the observed mass shifts, we hypothesised that the thiazoles have degraded (Figure 50). Further investigations are required to determine where on the protein this modification is occurring and at what point during the functionalisation thiazole degradation is occurring. A trypsin digest could also be employed to determine at what position in the protein modification is occurring, particularly as **29** appears to be functionalising at least two separate positions. MS analysis of **29** and **31** before they are employed for protein functionalisation does not show this degradation pattern although loss of Br and the benzyl group were observed (Appendix). For our aim of using the inserted thiazolium salt as an NHC catalyst, this method is clearly not viable as an intact heterocycle has not been attached to our protein scaffold. Unfortunately, there is the possibility this could occur with any NHC we attach to a protein, but we decided to first assess other functionalisation methods, specifically the use of a maleimide linked thiazolium salt before trying to resolve this degradation.

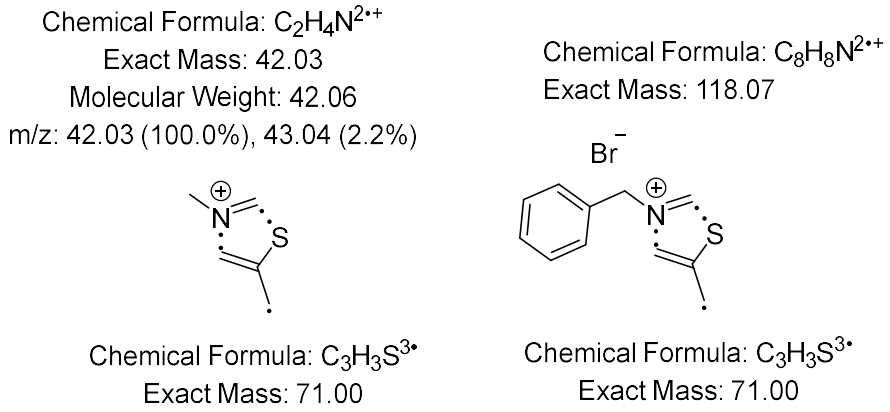


*Figure 50: Fragmentation analysis of **29** and **31** suggest fragmentation of both thiazole in the same position leads to the observed mass differences seen by LCMS.*

2.4.2 Functionalisation with maleimide linked thiazoles

Functionalisation of cysteine residues with maleimides occurs best at or slightly below neutral pH[219] (Figure 51). Higher pH, above pH 7.5, can lead to non-specific functionalisation of free primary amines such as lysine. Therefore, conditions were screened on a small scale (50 µL reactions) to determine the optimal conditions for functionalisation of each protein scaffold before scaling up functionalisation under the optimised conditions to obtain enough functionalised protein for catalysis.

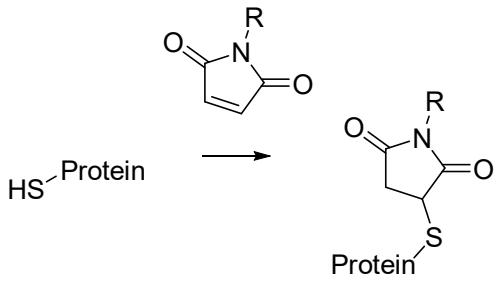


Figure 51: General scheme for functionalisation of a cysteine containing protein with a maleimide

2.4.2.1 hSCP mutants

Initial functionalisation trials using maleimide thiazole were undertaken with hSCP A100C. Successful functionalisation was achieved with both MMeThz and MBnThz in 20 mM Mes pH

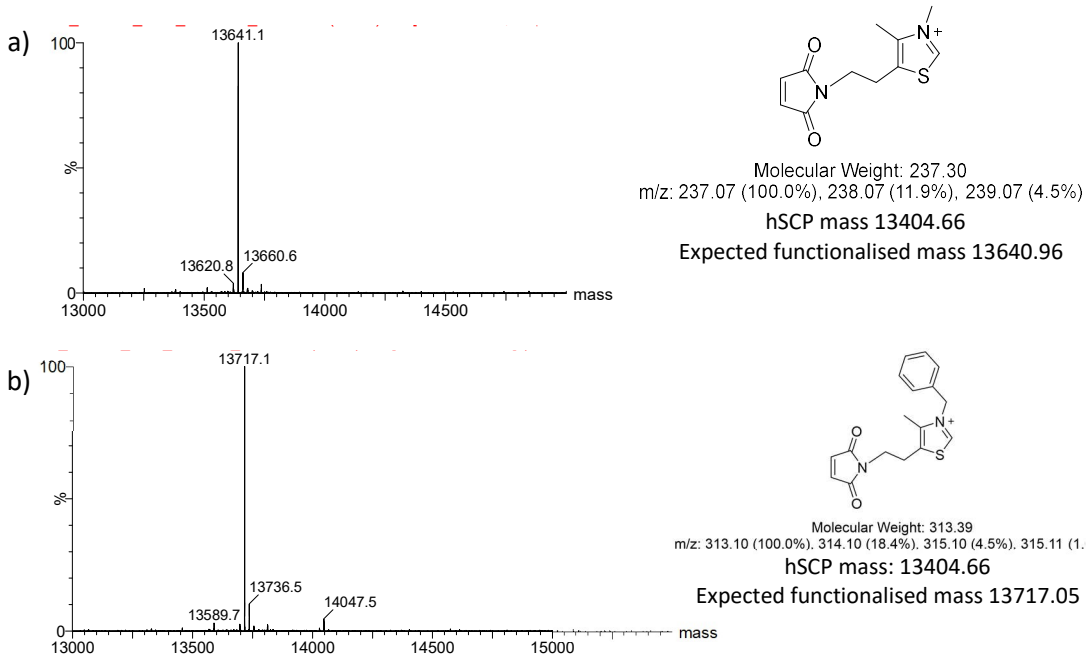


Figure 52: LCMS analysis of functionalisation of hSCP A100C with a) MMeThz and b) MBnThz shows full conversion to mono-functionalised protein.

6, 10 equiv maleimide thiazole for 2 h at RT (Figure 52). Initial catalysis results (Chapter 3) lead us to move forward with MBnThz for functionalisation of the other protein scaffolds.

All hSCP mutants were functionalised with MBnThz under the same conditions (20 mM Mes pH 6, 10 equiv MBnThz, 2 h, RT) with a majority of mono-functionalised product being observed in all cases (LCMS Appendix). A small amount of di functionalisation and maleimide ring opening via hydrolysis was observed which could alter the position of the organocatalyst within the protein scaffold.

*Table 11: hSCP functionalisation*the mass of the functionalised product as seen by MS is always -1 due to the positive charge on the protein adduct.*

	Expected Mass (Da)	Experimental Mass (Da)	Expected Mass Functionalised with MBnThz (Da)	Mass of functionalised product (Da)
hSCP V83C	13376.61	13376.7	13690.00	13689.0 (82 % monofunctionalisation) 13708.6(15 % monofunctionalisation + ring opening) 14019.1 (3 % difuncitonalisation + ring opening)
hSCP A100C	13404.66	13404.7	13718.05	13717.1 (95 % monofunctionalisation) 13736.5 (3 % monofunctionalisation + ring opening) 14047.5 (2 % difuncitonalisation, ring opening)
hSCP Q111C	13347.61	13347.8	13661.00	13660.2 13680.2 (10 %) 13990.9 (3 %)
hSCP V83C Stable	13403.68	13403.8	13717.07	13716.2 13735.7 (10% ring opening)
hSCP A100C F94H	13394.63	13394.8	13708.02	13707.0 13725.9 (12% ring opening) 14037.4 (5% difuncitonalisation + ring opening)

2.4.2.2 TTSCP mutants

It was thought that the smaller cavity size and presence of a disulfide bond in TTSCP made functionalisation of these mutants more difficult. Conditions for functionalisation of each TTSCP mutant needed to be optimised due to large amounts of precipitation and over functionalisation being observed. Some TTSCP mutants were particularly difficult to functionalise with unfunctionalised protein remaining even in "forcing" conditions. When forcing conditions were used, multi-functionalised products were also observed with these increasing above the amount of mono functionalised enzyme. However, after the buffer exchange process it was observed that concentration of the di- and tri- functionalised products decreased significantly. We hypothesize this is due to their instability which leads to precipitation.

2.4.2.2.1 TTSCP W83C Scaffold

We first attempted to functionalise TTSCP W83C using the conditions which were successful for hSCP mutants. Unfortunately, a large amount of protein precipitation was observed and after two hours, minimal functionalisation had occurred. Extending the incubation time to 5 hours still did not lead to significant functionalisation with approximately 15% mono and 15% tri-functionalisation observed. (Figure 53). Our modelling and analysis of TTSCP W83C (Chapter 2) indicated that the cysteine residue in this mutant could be buried, which would explain the difficulty we experienced with functionalisation. Resultantly we decided to discontinue efforts to functionalise this mutant and moved onto the other TTSCP constructs.

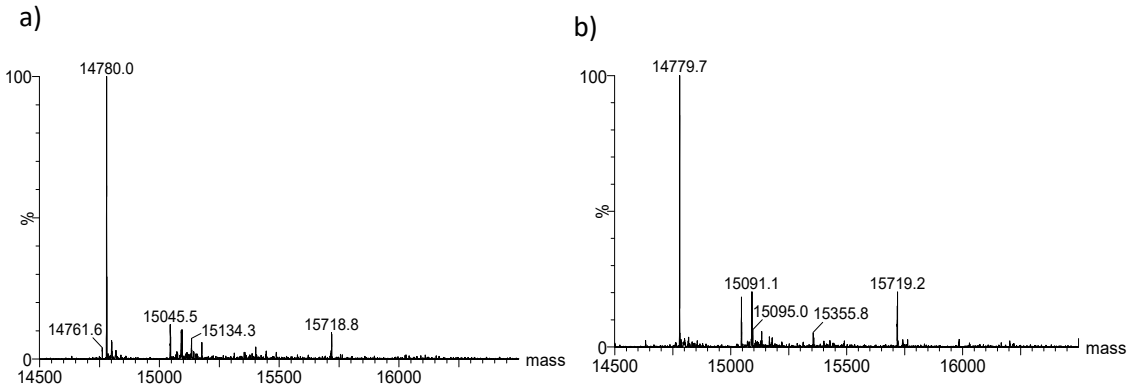


Figure 53: LC MS analysis of TTSCP W83C functionalisation trials in 20 mM Mes pH 6, 10 equiv MBnThz at RT for a) 2 hours b) 5 hours.

2.4.2.2.2 TTSCP L102C Scaffold

The optimal conditions for hSCP functionalisation with MBnThz were also tested with the TTSCP L102C scaffold. After two hours a significant amount of functionalisation had occurred resulting in a sample containing approximately 40 % unfunctionalised protein, 45 % mono-functionalised protein, 15 % di-functionalised protein and traces of tri-functionalised protein (Figure 54). Unfortunately, extending the incubation time did not lead to an increase in the yield of mono-functionalised protein, instead increasing the amount of over-functionalisation (Figure 54).

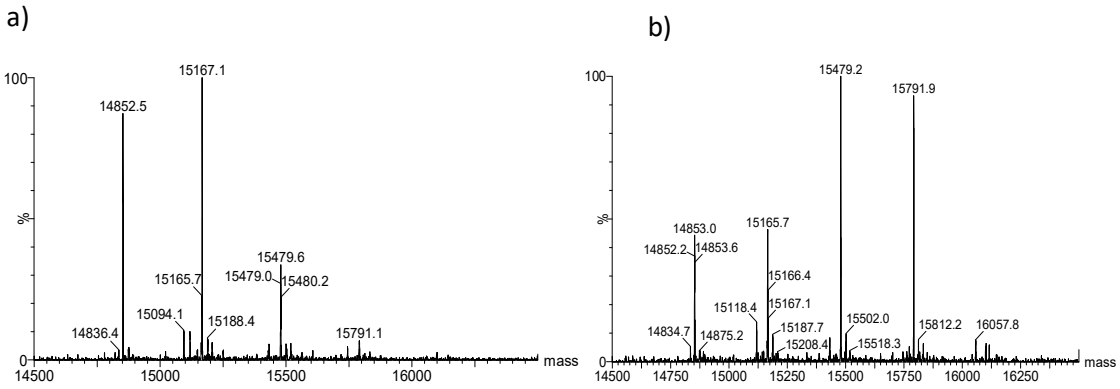


Figure 54: LC MS analysis of TTSCP L102C functionalisation trials in 20 mM Mes pH 6, 10 equiv MBnThz at RT for a) 2 hours b) 5 hours.

Further tests were undertaken at pH 7 and with different equivalents of MBnThz and incubation times. It proved difficult to achieve a significant amount of mono-functionalisation whilst minimising the multi-functionalised products. For example, whilst the use of 10 eq of MBnThz for 2 hours lead to no unfunctionalised protein remaining significant amounts of over-functionalisation was observed (Figure 55a). On the other hand, over-functionalisation was minimised to <10 % by using 2 eq MBnThz for 30 min however the majority of the protein remained unfunctionalised (Figure 55b). We decided to scale up our reaction using 5 equiv for 1.5 hours at pH 7 as less precipitate was observed in this reaction than the pH 6, 2-hour reaction. Interestingly, upon scale up and removal of excess MBnThz by desalting, we observed that the peaks corresponding to the over-functionalised proteins reduced in intensity (Figure 55c and d). We hypothesised that this was due to precipitation of protein. Although small scale samples were centrifuged prior to analysis, the desalting process likely allowed more time for precipitation and more thorough removal of unstable protein.

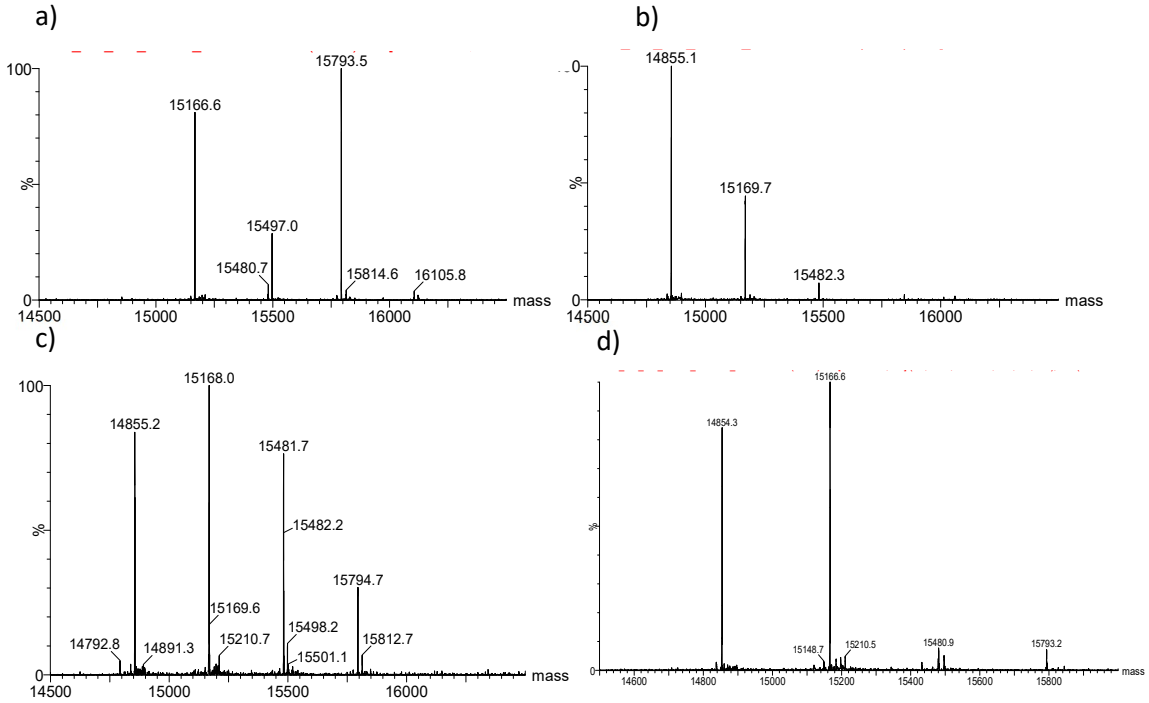


Figure 55: LC MS analysis of optimisation of conditions for TTSCP L102C functionalisation with MBnThz in 20 mM Mops pH 7 a) 10 eq of MBnThz for 2 hours b) 2 eq MBnThz for 30 min c) 5 equiv for 1.5 hours at pH 7 d) Analysis of 2.5 mL scale functionalisation of TTSCP L102C with 5 equiv for 1.5 hours at pH 7 after desalting into 50 mM HEPES pH 8.0.

2.4.2.2.3 TTSCP ΔDSB L102C Scaffold

The optimal conditions for the functionalisation L102C with MBnThz were also tested with TTSCP ΔDSB L102 (20 mM Mes pH 7, 5 equivilents of MBnThz, 1.5 hours at RT). It was gratifying to observe functionalisation of this protein lead to 75 % mono-functionalised protein with minimal amounts of precipitation and over-functionalisation observed (Appendix 1.3.9).

2.4.2.2.4 TTSCP L112C Scaffold

An initial test of histidine tagged TTSCP L112C showed that functionalisation for 1 hour with 10 equiv of MBnThz in 20 mM Mes pH7 formed the majority of mono-functionalised protein. For the cleaved protein, reducing the incubation time to 30 min yielded improved results with 85% mono-functionalised protein being obtained (Appendix 1.3.8).

2.4.2.2.5 Summary of TTSCP scaffold modification

Table 12: Summarises the expected and observed masses after functionalisation of each SCP mutant and summaries the reaction conditions used

	Expected Mass (Da)	**Experimental Mass (Da)**	**Functionalisation conditions**	**Expected Mass Functionalised with MBnThz (Da)**	**Mass of functionalised product (Da)**
TTSCP	14865.04	14863.0 (accounts for DSB)		-	-
TTSCP W83C	14781.9	14780.3 (accounts for DSB) 15046.2 (Stearoylation)	20 mM Mes pH6, 10 equiv MBnThz, 5 h, RT	15094.90	14779.7 (80% unfunctionalised) 15045.8 (Stearoylation unfunctionalised) 15091.0 (~5% potential mono-functionalisation) 15719.0 (~10% tri-functionalised)
TTSCP L102C	14855.02	14853.6 (accounts for DSB)	20 mM Mes pH7, 5 equiv MBnThz, 1.5 h, RT	15168.02	14854.3 (40% unfunctionalised) 15166.6 (50% mono-functionalisation accounts for DSB still present) 15480.9 (5% di-functionalised) 15793.2 (5% tri-functionalised)
TTSCP L112C	14855.02	14853.4 (accounts for DSB)	20 mM Mes pH7, 10 equiv MBnThz, 0.5 h, RT	15168.02	15165.3 (85% mono-functionalisation - accounts for DSB still present,) 15479.5 (10% di-functionalised) 15792.2 (5% tri-functionalised)
TTSCP ΔDSB L102C	14790.90	14790.9	20 mM Mes pH7, 5 equiv MBnThz, 1.5 h RT	15103.90	14792.2 (25% unfunctionalised) 15104.7 (75% mono-functionalisation)

2.4.2.3 ThAOS

To functionalise ThAOS mutants with MBnThz, an initial screen of conditions was undertaken using the same conditions as for hSCP mutants. Results from the initial screen were not successful with little to no functionalised protein being observed. Therefore, reactions with maleimide and iodoacetamine were performed to assess if the cysteine was amenable to functionalisation and to prevent waste of our unique maleimide thiazole. Functionalisation with these small molecules resulted in a higher percentage of functionalised protein and less precipitation although unfunctionalised protein still remained. The highest percentage of mono-functionalisation (~60 % functionalisation) was observed in 50 mM HEPES pH8 using 10 eq Maleimide, 4 °C for 18 hours (Figure 56).

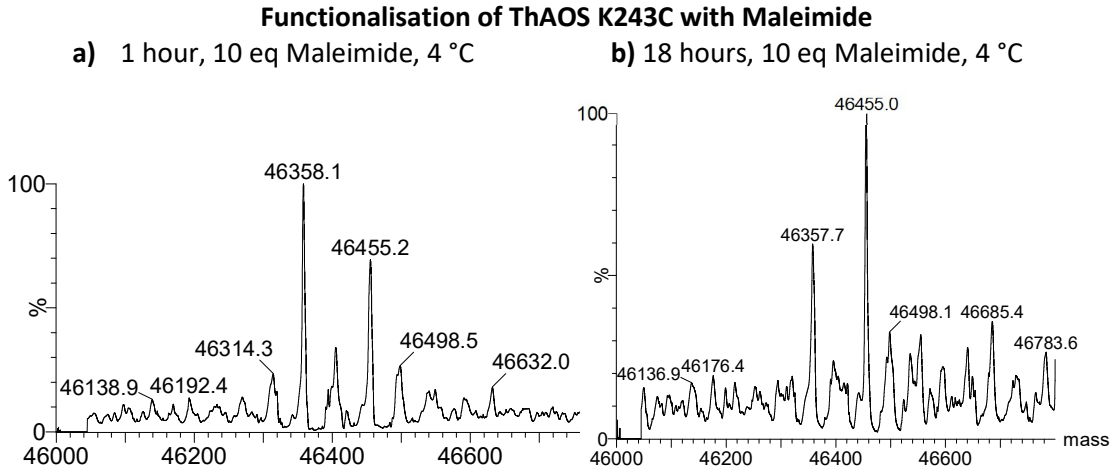


Figure 56: LCMS analysis of functionalisation of ThAOS K243C with 10 equivalents of maleimide in 50 mM HEPES pH8 for a) 1 hour and b) 18 hours.

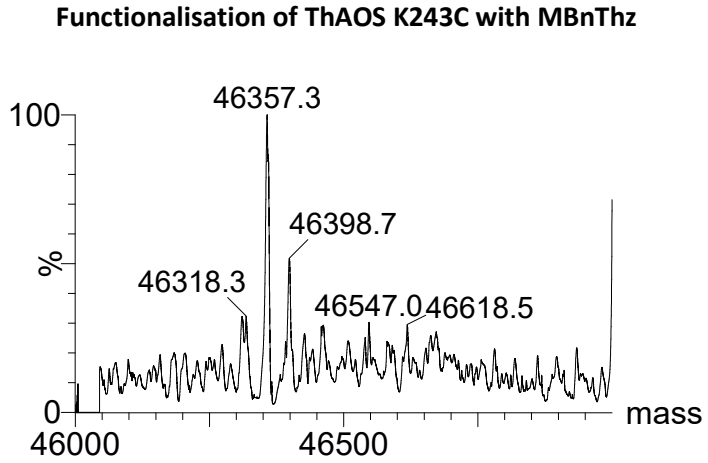


Figure 57: LCMS analysis of the functionalisation of ThAOS K243C with 10 equivalents of MBnThz in 50 mM HEPES pH8 for 18 hours does not show the expected functionalised mass of 46671 Da.

These conditions did not translate when adding MBnThz, with no functionalised protein above background level being obtained under the same conditions (Figure 57). We theorised

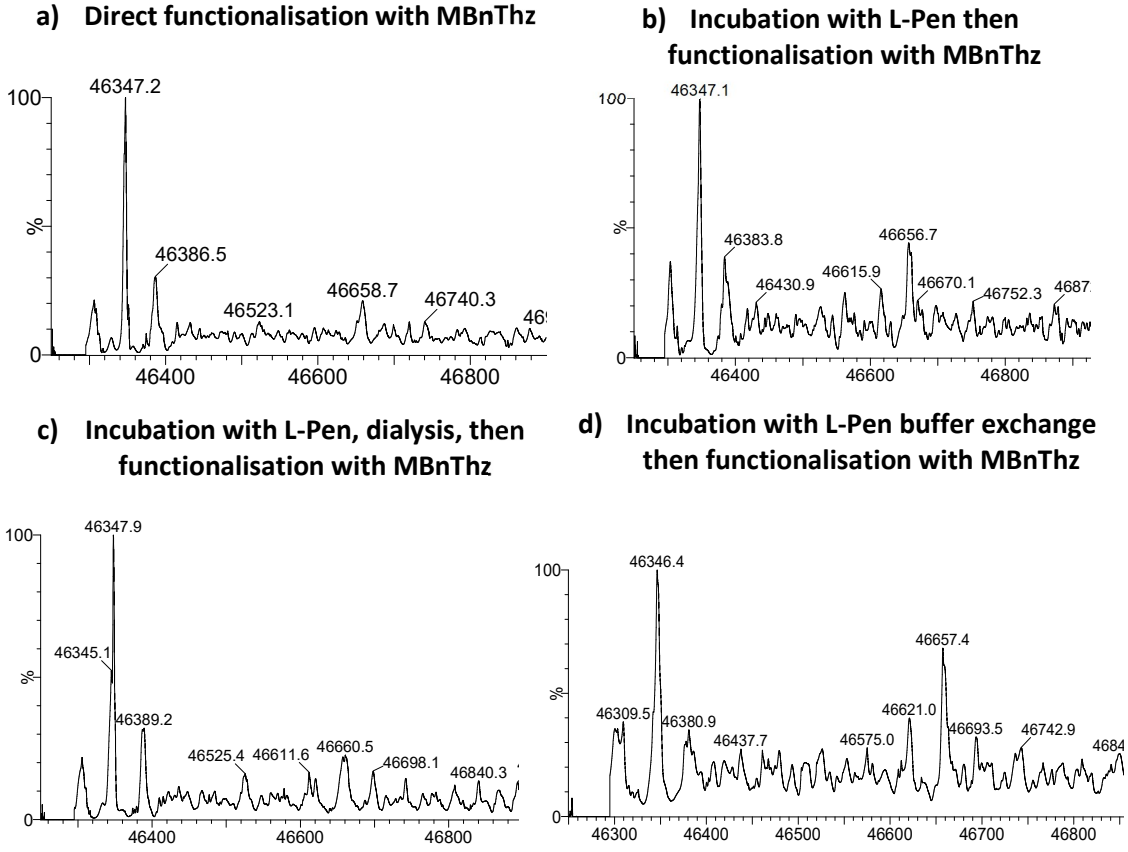


Figure 58: LCMS analysis of the functionalisation of ThAOS K243C V79S with 10 equivalents of MBnThz for 18 hours at 4 °C in 50 mM HEPES pH 8 comparing different methods for removal of PLP. a) Direct functionalisation with MBnThz, b) Incubation of the protein with 1 mM L-Pen for 1 hour followed by functionalisation with MBnThz c) Incubation of the protein with 1 mM L-Pen for 1 hour followed by dialysis then functionalisation with MBnThz d)) Incubation of the protein with 1 mM L-Pen for 1 hour followed by buffer exchange using a pd 10 column then functionalisation with MBnThz.

that residual PLP in the binding site may be preventing access of the maleimide to the cysteine residue. Addition of L-penicillamine (1mM) to ThAOS K243C V79S (100 µM) was performed and incubated at RT for 1 hour. Three different work-up methods were tested to remove excess L-penicillamine and PLP, buffer exchange, dialysis, and no work up. The yellow colour of the enzyme disappeared during both dialysis and buffer exchange, suggesting PLP had been removed from the enzyme. The resultant proteins were functionalised with 10 equivalents of MBnThz for 18 hours at RT. Some improvement was observed when compared with direct functionalisation of ThAOS K243C V79S, but full functionalisation was still not achieved (Figure 58). It was decided to move forward with partially functionalised enzyme to assess the catalytic activity of this scaffold with the aspiration of assessing alternative functionalisation methods in the future.

2.4.2.4 *PfThiE*

For functionalisation of pfThiE we initially focused on the histidine-tagged D69C scaffold. We were able to functionalise this protein in 20 mM Mes pH 7 using 10 equivalents MBnThz for 2 hours, however we did not observed masses which corresponded only with the mass of the protein plus addition of MBnThz (313 Da) (Figure 59). A large amount of maleimide ring opening was observed (25183 adduct of +18 kDa) and another adduct of +49 kDa was observed. Concern about these adducts led us to focus on our two Tev cleaved mutants, T132C, D69C/S106A. Under the same conditions used for D69C, these residues were accessible and enabled functionalisation of the protein without any additional adducts being observed.

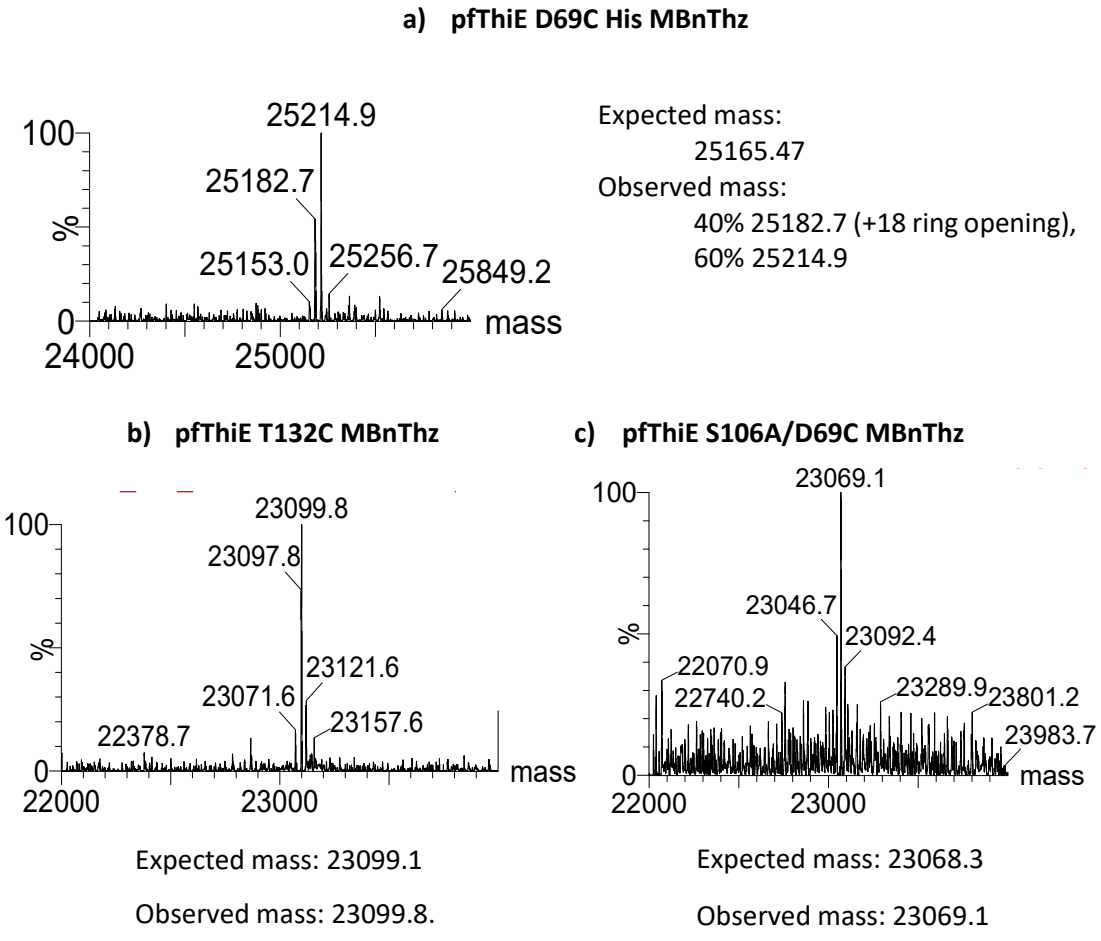


Figure 59: LC MS analysis of the functionalisation of pfThiE mutants. Deconvoluted spectra show the observed masses which were compared with the expected masses of functionalisation with MBnThz. Whilst His tagged D69C showed mass shifts which corresponed with maleimide ring opening and other adducts, the Tev cleaved mutants showed mases close to the expected values.

2.4.3 Functionalisation with Maleimide triazoles

Both TTSCP L102C and TTSCP ΔDSB L102C were functionalised with MBnTri (20 mM Mes pH 7, 10 equivilents of MBnTri, for 1.5 hours at RT). MBnTri was not in a pure form due to low yields during synthesis therefore the concentration of the functionalisation reagent was likely below 10 equiv. The proteins showed significant amounts of precipitate upon incubation with MBnTri, however, almost complete functionalised protein was obtained for both TTSCP L102C and TTSCP ΔDSB L102C (Figure 60). This was promising as due to the impurity of MBnTri we were unsure if functionalisation would be possible however the expected mass shifts for protein functionalisation with MBnTri were observed. Interestingly functionalisation with this MBnTri was better than that observed with the thiazolium analogue MBnThz. This could be attributed to the structural and functional differences between the two molecules. It would be interesting to determine the crystal structures of the functionalised proteins to how each molecule sits inside the protein's cavity.

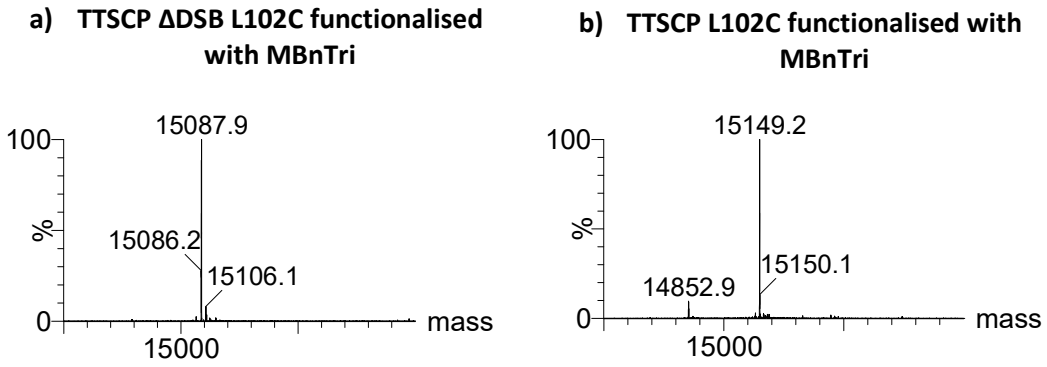


Figure 60: LC MS analysis of TTSCP L102C and TTSCP ΔDSB L102C functionalised with 20 mM Mes pH 7, 10 equivilents of MBnTri, for 1.5 hours at RT showed the expected mass of the functionalised proteins. Expected mass TTSCP ΔDSB L102C: 15088.2 ,Da TTSCP L102C: 15150.9 Da,

2.5 Conclusions and future work

In this chapter, we synthesised five NHC precursors suitable for the selective functionalisation of cysteine containing proteins. Of these the maleimide linked NHC precursors were the most promising for functionalisation of protein scaffolds via a maleimide thiol linkage whilst halogenated thiazolium salts degraded during protein functionalisation.

Several protein scaffolds were investigated including: two proteins which bind thiamine but do not use it as a cofactor pfThiE and Thiaminase 1, the PLP dependant enzyme ThAOS, a protein scaffold commonly used for the creation of ArMs (hSCP) and its thermophilic homologue TTSCP. Cysteine mutants of hSCP were already available and these proved to be the most easily functionalised scaffolds we tested. All hSCP mutants produced >80 % mono-functionalised protein with MBnThz. On the other hand, the native disulfide bond in TTSCP led to difficulties selectively functionalising this scaffold. Removal of the disulfide bond by mutation of Cys13 and Cys63 to alanine overcame this complication and lead to improved functionalisation. For example, TTSCP ΔDSB L102C functionalised with MBnThz yielded 75 % mono-functionalised protein. Importantly, the effect of removing the DSB was assessed and found to not significantly destabilise the protein, with the T_m of the ΔDSB mutant being measured as 79 °C as measured by CD. These findings highlight the value of a naturally inactive protein scaffold which contains only a single cysteine residue to allow for selective functionalisation.

Cysteine residues were substituted into the protein scaffolds other than hSCP using site directed mutagenesis. The choice of the cysteine residues position was guided by ligand docking of the Breslow intermediate and the functionalisation reagent to suggest favourable positions. Some of the selected positions proved difficult to functionalise than others, highlighting the importance of accessibility of the inserted cysteine residue. In particular, we were unable to functionalise TTSCP W83C, a protein for which modelling and analysis of the protein's cavities using CASTp suggested the inserted cysteine residue becomes buried inside the structure rather than lining the cavity.

When choosing the position to insert a cysteine residue, the active site of the enzymatic protein scaffold was the focus as it provides a functional cavity for our organocatalyst to be position in. However native catalytic residues were mutated to cysteine in pfThiE S106C and ThAOS K243C in an attempt to inactivate any natural catalytic activity of the protein's scaffolds. It was observed that, despite mutation of catalytic residues, both pfThiE and ThAOS co-purified with their native substrates/cofactors which led to difficulties in functionalisation.

In the case of pfThiE these were overcome by an inactivating S106A mutation. For ThAOS, attempts were made to remove the native PLP cofactor from ThAOS by the use of L-penicillamine. However, full functionalisation was still not achieved. These results highlight how proteins with native catalytic activity are not well suited as protein scaffold to host organocatalytic reactions.

More sophisticated methods for the design of the protein scaffold would help overcome issues associated with the scaffold's native activity. In particular the pfThiE scaffold was chosen due to the previous success of de-novo enzyme design using TIM barrels. It would be interesting to compare our mutated pfThiE scaffold to a computationally designed TIM barrel which specifically accommodates our Breslow intermediate TS. It should also be considered that computationally making the covalent modification to our protein scaffold and then docking in the substrate may help achieve scaffolds that are more active. Unfortunately, the addition of non-canonical components and cofactors to *de novo* designs increases the complexity of the computations required and is not compatible with all design tools. However, the Baker group recently published a preprint article introducing RoseTTAFold All-Atom which is able to model proteins with covalent modifications which could be applied for the design of future scaffolds in this project [220].

3 Catalysis of an Intramolecular Stetter Reaction

3.1 Aims

The aim of this chapter was to assess the ability of our NHC functionalised protein scaffolds to catalyse an intramolecular Stetter reaction under aqueous buffered conditions. To achieve this, both the starting material and product were synthesised chemically to allow analytical methods to be developed. These methods were then used to screen and optimise reaction conditions for our NHC functionalised protein scaffolds to best catalyse the reaction.

3.2 Model Intramolecular Stetter Reactions

We chose to use a well-characterised intramolecular Stetter reaction for the cyclization of a salicylaldehyde-derived substrate **18** (Figure 61), as this reaction has grown to be a benchmark for comparing catalyst efficiency in the intramolecular Stetter reaction[59]. Previously this reaction has been used to assess the catalytic ability of chiral NHC catalysts, thiazolylidine based UAAs, peptides and enzymes[59, 76, 113]. We chose to focus on an intramolecular reaction, as we only need to consider accommodating one substrate in the active site. However, we hope our findings can be expanded to other NHC catalysed

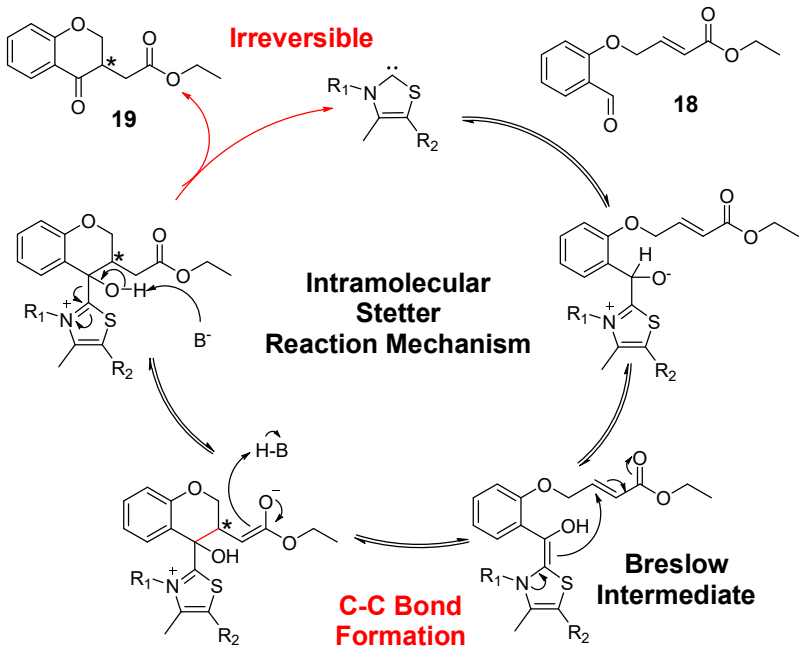


Figure 61: An overview of the intramolecular Stetter reaction mechanism highlighting the Breslow intermediate and key steps including the formation of the C-C bond (in red) and the irreversible conversion of

reactions. The Stetter reaction is of particular interest as the reaction produces a 1,4 dicarbonyl compound, the synthesis of which remains a challenge in synthetic chemistry due to the innate polarity mismatch of carbonyl fragments when forming even-numbered dicarbonyls[73]. Although numerous intramolecular Stetter reactions are possible, we chose to use this benchmark reaction as a proof of concept for using NHC functionalised protein scaffold for catalysis[59]

3.2.1 Intramolecular Stetter reaction in organic solvents

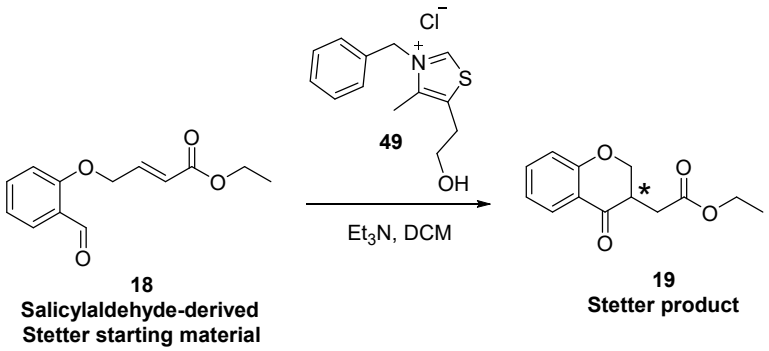


*Figure 62: Intramolecular Stetter reaction of salicylaldehyde-derived Stetter starting material **18** catalysed by the NHC thiamine analogue **49** to give the Stetter product **19.** Reaction conditions: 2.5 mmol **18,** 0.5 mM **49,** 0.5 mM Triethylamine, DCM, 50°C, 16 hours with stirring.*

The salicylaldehyde-derived starting material **18** which contains aldehyde and an α/β-unsaturated ester was synthesised according to literature methods[76, 221] to yield a yellow solid (Figure 62, Appendix 1.6.15). The NHC-catalysed intramolecular Stetter reaction was performed using the commercially available thiazolium salt, 3-benzyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride **49** in DCM with 1 equivalent of triethyl amine to deprotonate the NHC precursor (Figure 62). The Stetter product **19** was purified, characterised and used for development and calibration of analytical HPLC methods.

An HPLC method was developed using the product obtained from the organic reaction. Monitoring the aromatic in the starting material and product allowed us to use the absorbance at 254 nm to give the cleanest spectra. A calibration curve was created for both product and starting material. The racemic product was also used to create a chiral HPLC method to separate the two enantiomers. Initially extraction into organic solvent was used and a normal phase chiral HPLC method was created in accordance with methods reported in literature. However, extraction into organic solvents creates the opportunity for losses to occur and samples require a different work up to those analysed via the C18 column therefore a chiral reverse phase method was also created allowing promising samples to be analysed for their enantioselectivity.

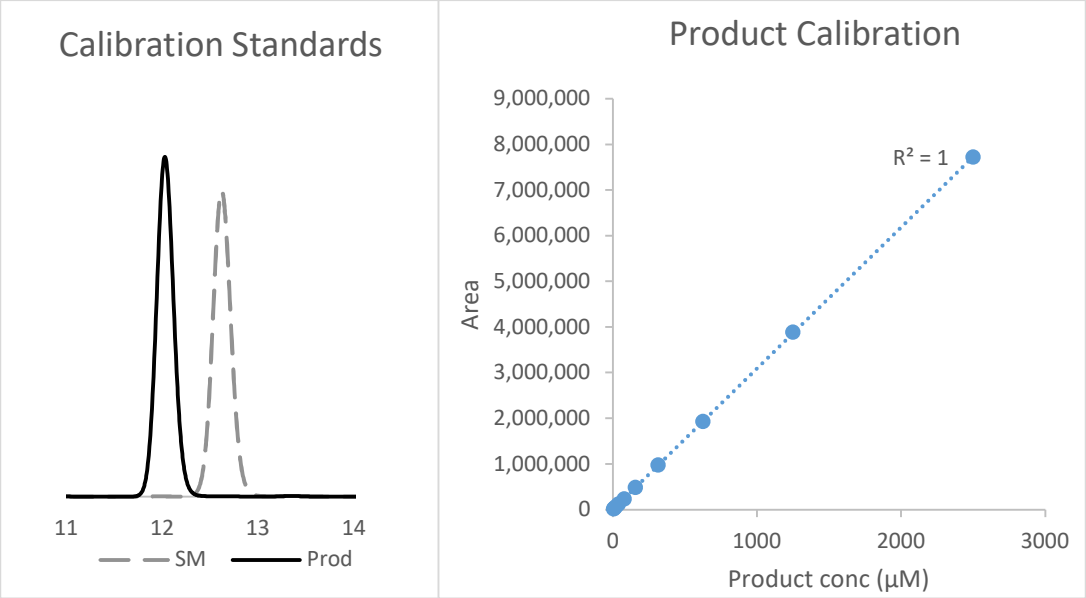


*Figure 63 Calibration standards of **18** (intramolecular Stetter starting material, SM grey dashed line) and **19** (intramolecular Stetter product, black line) and the calibration curve plotted for 3.*

3.2.2 Intramolecular Stetter reaction in aqueous buffer

As a benchmark for our artificial Stetterase we synthesised and tested the aqueous Stetter catalyst **50** reported by Debiais *et al*[111]. The catalyst **50** was synthesised in two steps with a 51 % yield to give a viscous brown oil (Figure 64). The catalyst was tested using the reported conditions for the aqueous intermolecular Stetter between chalcone and benzaldehyde (Figure 64). Unreacted chalcone was still present after 20 hours and yields were much lower than those obtained in the paper[111]. After purification via flash column chromatography, we obtained a 10% yield (32 mg) of the diketone product was obtained compared to 72% achieved by Debiais *et al*[111]. We found the oily catalyst difficult to work with and the two

phases of the system required vigorous mixing to form the droplet displayed in literature. We speculated that the reaction set up, glassware and velocity of stirring could easily affect the yield.

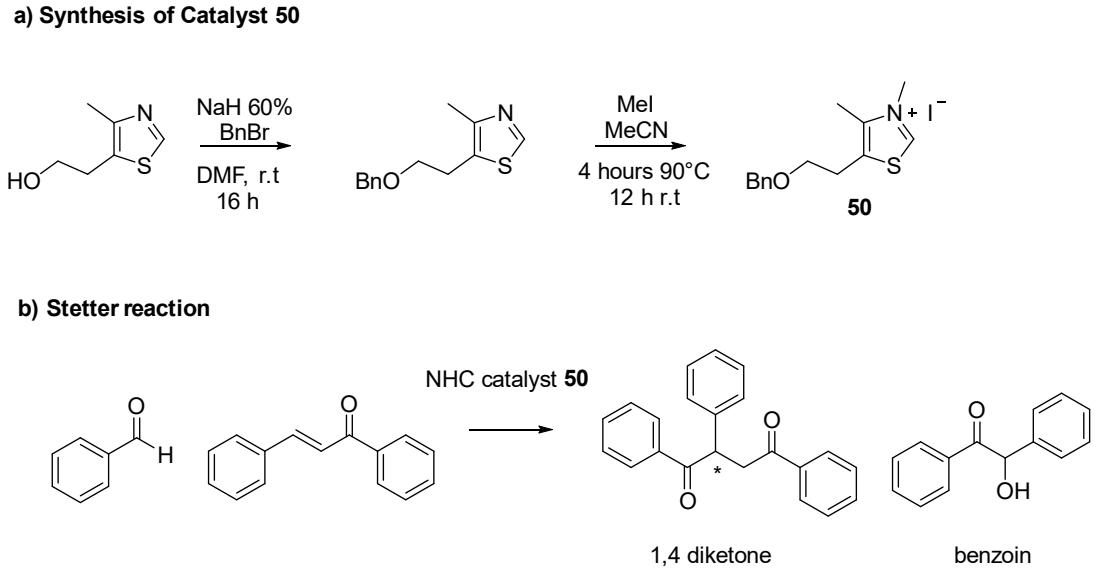


*Figure 64: a) The two-step synthesis of aqueous Stetter catalyst **50.** b) The intermolecular Stetter reaction between benzaldehyde and chalcone catalysed by NHC **50** produces a 1,4 diketone and benzoin.*

To directly compare this catalyst to our functionalised enzymes we dissolved the catalyst in DMSO and used it to set up an intramolecular Stetter reaction using starting material **18** as the starting material on a 250 µL scale. We assessed the catalyst both with and without the base DBU at 4 mol % catalyst as well as testing a 100 mol % catalyst (Figure 65). At 4 mol % catalyst no product formation was observed above background however using 2.5 mM catalyst 85.1 µM ±4.0 product **19** was observed via HPLC. Interestingly, when DBU was used

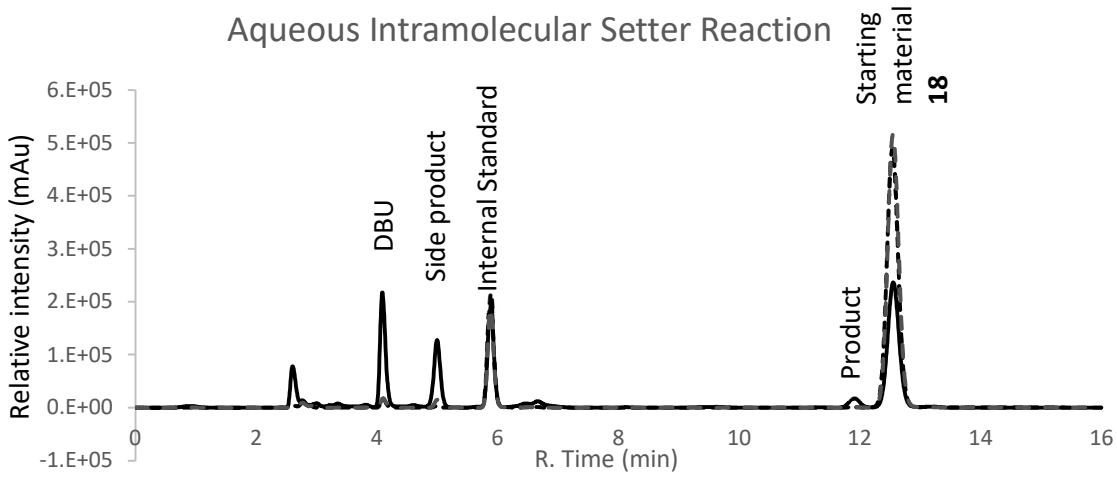


*Figure 65: HPLC analysis of an aqueous intramolecular Stetter reaction catalysed by **50** monitored by UV absorbance at 254 nm. Reaction conditions: 100 µM or 2.5 mM Catalyst **50**, 2.5 mM starting material **18,** 20% DMSO in water,18 hours, 30 °C, 250 rpm. DBU was added in equimolar amounts to the catalyst when used. Reactions were quenched with an equal volume of acetonitrile containing 1 mM coumarin as an internal standard.*

in the reaction, a side product also formed with a retention time of 5 mins. We hypothesised this may be due to base catalysed ester hydrolysis. The low conversion could be attributed to several factors including the use of 20% DMSO to dissolve the catalyst and starting material. Additionally, no organic droplets were observed due to DMSO solubilising the catalysts, and shaking being used to mix the reaction rather than mechanical stirring, which is a more vigorous method.

3.2.3 Enzymatic Intramolecular Stetter reaction

As discussed in the introduction, in 2021 the enzyme pfBAL was identified as having Stetterase activity and has been shown to catalyse an enantioselective intramolecular Stetter reaction[76]. We repeated this work to use pfBAL as a benchmark to compare our NHC-functionalised proteins to.

3.2.3.1 Expression and purification of pfBAL

pfBAL was expressed and purified under conditions reported by Chen *et al.*[76] (Figure 66). The lysate, which was used for catalysis experiments, is shown in the first lane of the SDS-PAGE analysis and displays a band of the expected mass for the over expressed pfBAL (60 kDa, Figure 66). Although we used the same expression conditions and volume of lysis buffer per g of wet cell pellet as mentioned by Chen *et al.* the exact composition of a lysate may vary between labs. Purification of pfBAL using IMAC followed by size exclusion chromatography

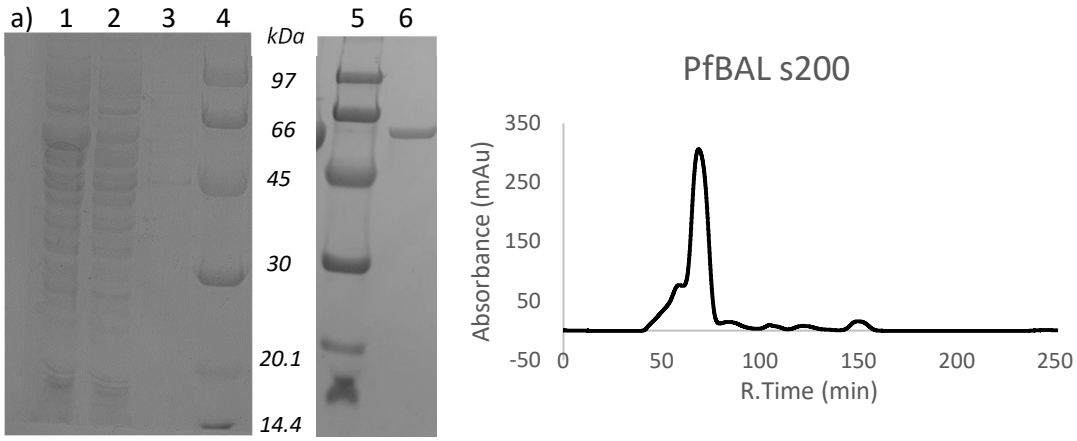


Figure 66: a) SDS PAGE analysis of the purification of PfBAL showing samples in the following lanes 1. Lysate 2. IMAC flow through 3. Wash 4 LMWM 5. LMWM 6. Purified pfBAL. B) A trace of the elution of pfBAL from an s200 size exclusion column.

(SEC) produced 119 mg/L of purified protein. Purified enzyme of the expected molecular weight, 60 kDa was obtained and by comparing the retention time of pfBAL during SEC to a calibration curve, we verified that the dimer was formed (found 69 mL predicted from calibration 71 mL) (Figure 66).

3.2.3.2 Intramolecular Stetter reaction catalysed by pfBAL

We aimed to replicate the results obtained by Chen *et al.* therefore intramolecular Stetter reactions of **18** with both purified enzyme and lysate were performed to compare the performance of the two. Our initial reactions did not go to completion so different concentrations of starting material were screened (Figure 67). We were concerned that despite replicating the expression conditions highlighted in the paper, our lysate may differ compared to that used in the paper. During our reactions we also observed the formation of side products that were not mentioned by the authors, and resultantly despite all the starting material being consumed in the lysate reaction we were only able to obtain ~500 µM of product when 2.5 mM of starting material was used. Although Chen *et al.* do not define how HPLC yields were calculated (neither depletion of starting material nor accumulation of product) upon scale up they obtain an 80% yield. Therefore, we aimed to determine why our yields were lower. Appearance of side product peaks corresponded with a decrease in starting material/product concentration. Therefore, we believed these were side reactions of the starting material/product with the enzyme or other components of the reaction which were reducing our reaction yield.

The most prominent side product formed during the reaction with the pfBAL lysate appeared at 5.1 min (**SP1**) and corresponded with the side product we observed in our aqueous Stetter reaction with **50**. A second side product formed at 7.0 min (**SP2**) and additional trace impurities were observed. Using purified pfBAL as the catalyst, the prominent side products was **SP2** at 7.0 mins whilst trace amounts of **SP1** were observed. To assess when these side products began to accumulate and if running the reaction for a shorter time could increases product yield, time dependant studies were undertaken (Figure 68). Whilst the pure pfBAL reaction was complete in 8 hours, the lysate reaction required a longer reaction time. It was also observed that after the reaction was complete, the product began to slightly deplete at subsequent time points whilst the concentration of side products increased with reaction time. Meanwhile control reactions with the starting material showed a minimal amount of **SP1** formed after 24 hours. This supported our suggestion that **SP1** is likely formed by degradation of **18**. Due to its absorbance at 254 nm, it is likely that this degradation does not alter the chromophore and as previous hypothesised, it may be ester hydrolysis which could well be increased by components of the lysate.

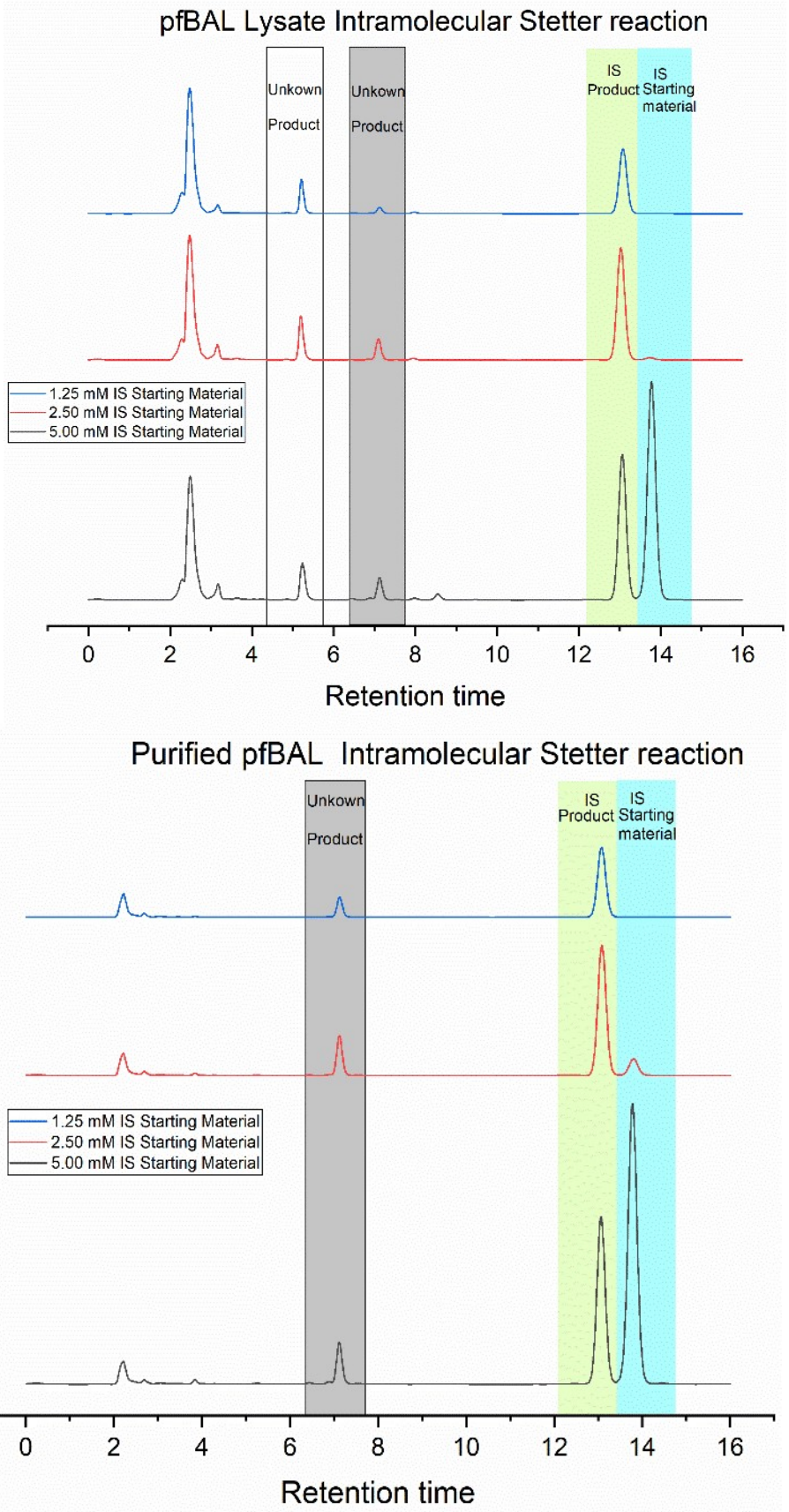


Figure 67: HPLC analysis of PfBAL reactions show the formation of two unexpected side products at 5.1 min and 7 min. Reaction conditions: 150 µL pfBAL lysate or 35 µM purified pfBAL, 50 mM Kphos pH 7, 20% DMSO, 2.5 mM MgSO4, 0.15 mM TPP, 1.25 - 5 mM starting material ***18,*** *30 °C, 250 rpm.*

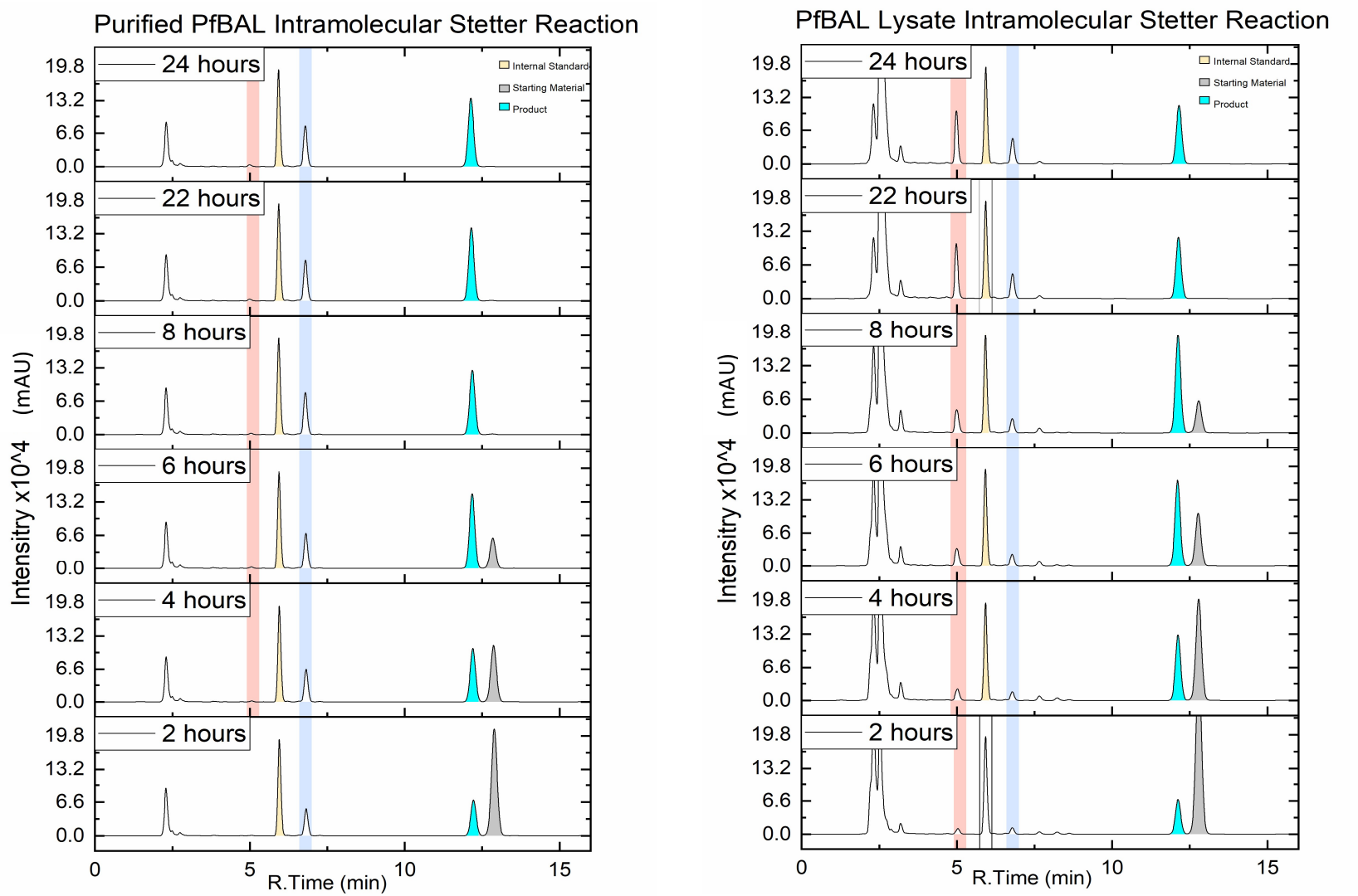


*Figure 68: HPLC analysis of a time dependant intramolecular Stetter reaction catalysed by pfBAL lysate and purified pfBAL. The formation of side products SP1 and SP2 are highlighted in pink and blue respectively. The peak corresponding to the product (cyan) can be seen to increase up to 8 hours whilst the peak corresponding to the starting material (grey) decreases with time. Reaction conditions: 150 µL pfBAL lysate or 35 µM purified pfBAL, 50 mM Kphos pH 7, 20% DMSO, 2.5 mM $MgSO_4$, 0.15 mM TPP, 2.5 mM **18,** 30 °C, 250 rpm.*

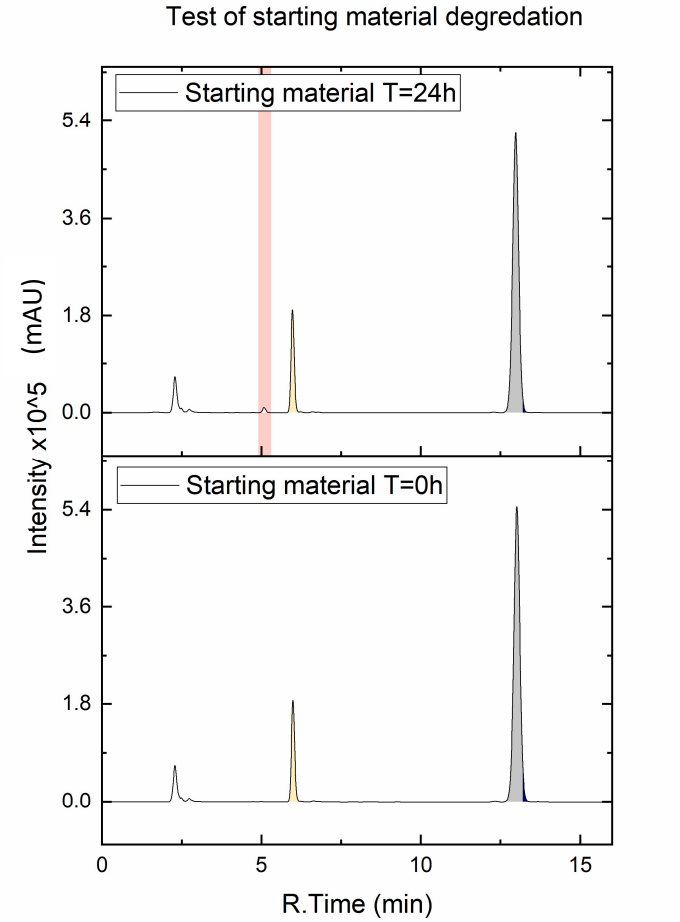


*Figure 70: HPLC analysis of the control reaction of **18** displayed a small peak appearing 5.1 mins (pink) corresponding to SP1 after 24 hours. Reaction conditions: 50 mM Kphos pH 7, 20% DMSO, 2.5 mM $MgSO_4$, 0.15 mM TPP, 2.5 mM **18,** 30 °C, 250 rpm, 18 hours.*

Having determined that the formation of side products was leading to our decreased yields we went on to analyse the stereoselectivity of the products formed in our reactions. Performing reactions with pfBAL under similar conditions to those reported by Chen *et al* we obtained an e.r. 74: 26, similar to their e.r. of 71:29[76] with 20% DMSO.

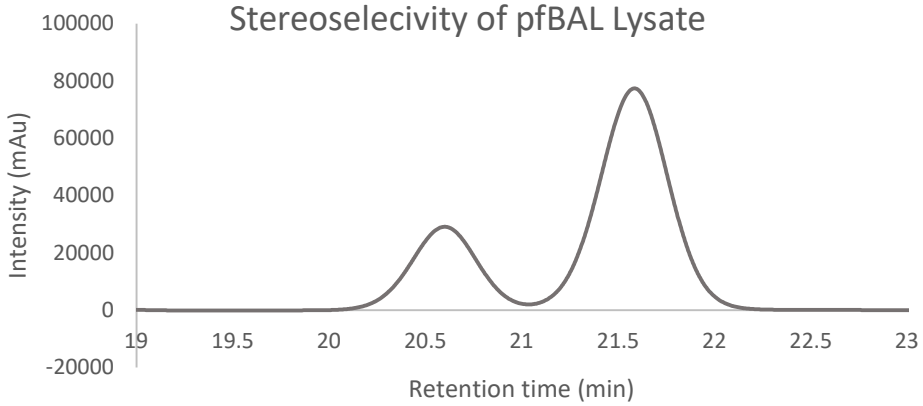


*Figure 69: Chiral HPLC analysis of the intramolecular Stetter reaction performed with pfBAL lysate and TTSCP L102C functionalised with **1**. 20.77 (R–enantiomer), 21.74 (S-enantiomer). Reaction conditions: 50 mM Kphos pH 7, 20% DMSO, 150 µL pfBAL lysate, 2.5 mM $MgSO_4$, 0.15 mM TPP, 2.5 mM 2, 18 hours, 30 °C, 250 rpm*

3.2.4 Analysis of side products

Side product formation was observed in several of the reactions mentioned above. We hypothesised that ester hydrolysis may be occurring of both starting material **18** and product **19**. To test this hypothesis an HPLC fitted with a diode array detector (DAD) was employed to obtain full UV spectra of the starting material, product and the two main side products observed at 5.1 mins (**SP1**) and 7 min (**SP2**). We observed a slight difference between the spectra of the starting material and product (Figure 71). The spectra of **SP1** were identical to that seen for the starting material **18** whilst **SP2** showed a unique spectra. This suggest the chromophore in the **SP1** is the same as that seen in the starting material. Whilst this is consistent with ester hydrolysis, another possibility supported by literature findings is the isomerisation of the starting material **18** which produces side products in the intramolecular Stetter reaction[222] (Figure 72). The spectral differences observed for **SP2** led us to consider possibility for the product to undergo keto-enol tautomerisation which could alter the

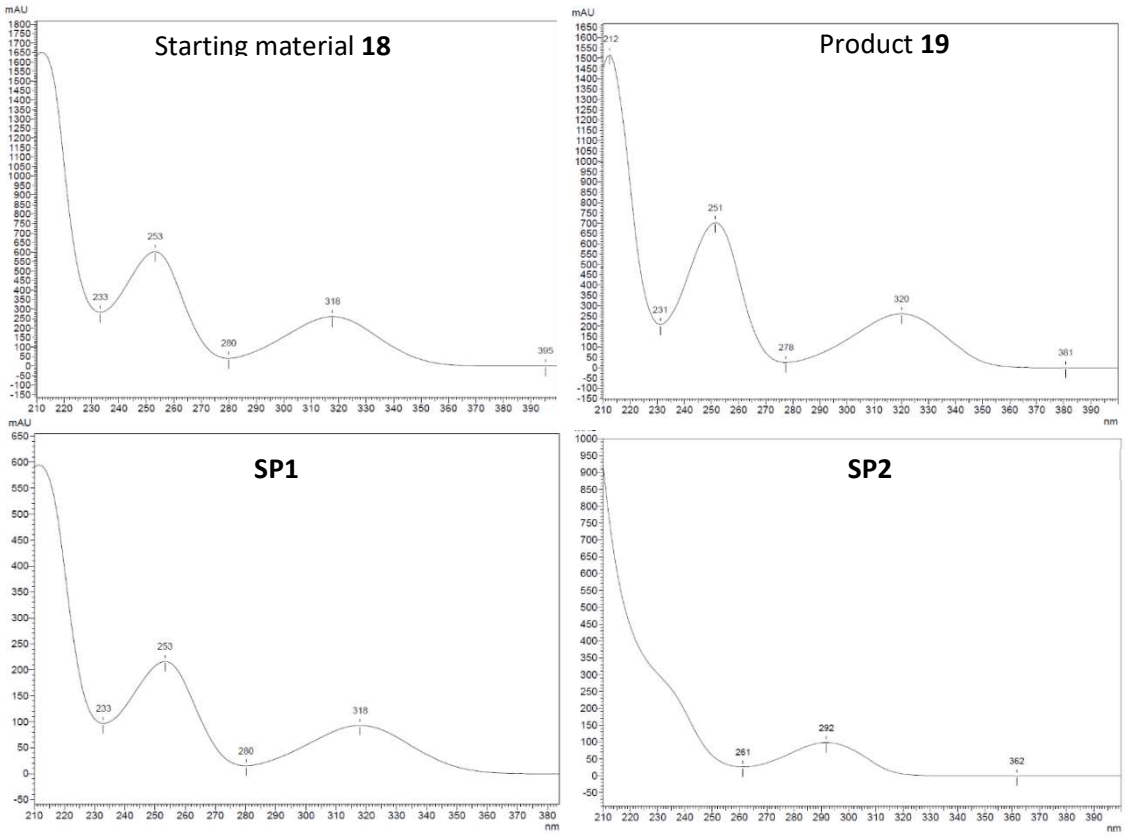


*Figure 71: Full UV spectra of starting material **18,** product **19, SP1** and **SP2** obtained by HPLC DAD analysis. Starting material **18** and **SP1** both show maxima at 253 nm and 318 nm whilst minima occur at 233 nm and 280 nm. The maxima and minima are shifted by 2 nm in the product spectra. **SP2** shows a distinct spectra with a maxima at 292 nm and a minima at 261 nm.*

chromophore by forming a conjugated system. However, HPLC conditions should prevent this being observed as a separate product.

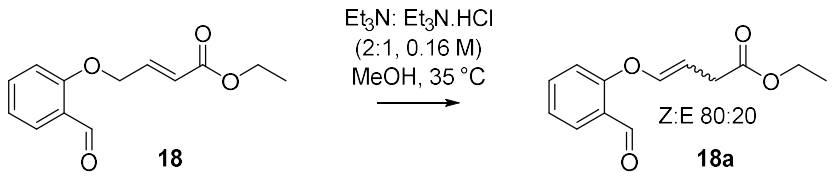
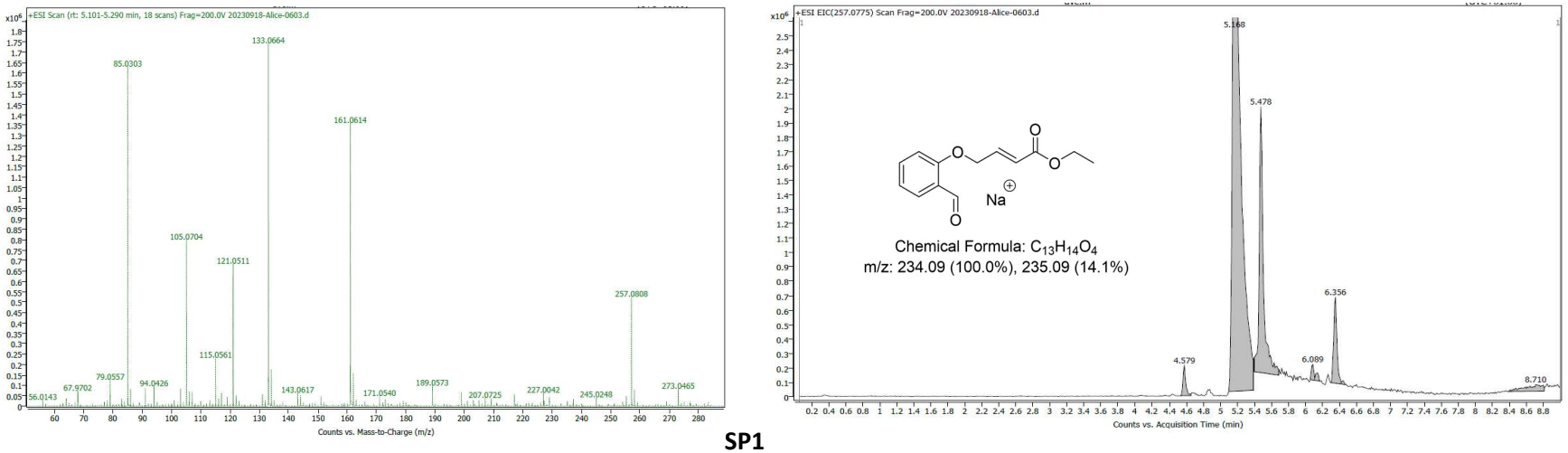


*Figure 72: Isomerisation of **18** observed to form side products during the intramolecular Stetter reaction by Collett et al.*

We aimed to gain further insights into the identity of the side products by performing LC-MS analysis (Figure 73, Figure 74). This revealed that whilst the mass for the starting material structure (and corresponding isomer) was present in the starting material peak (Figure 73, retention time 5.1 min, $[C_{13}H_{14}O_4Na]^+$, $[M+H]^+$ requires *m/z* =257.078). Meanwhile only fragments of the starting material were observed in **SP1** peak (Figure 73, retention time 4.06 min, main fragment observed *m/z* = 121.05 $[C_7H_5O_2]$ corresponding to the aromatic fragment after cleavage of the ether). LC-MS analysis of **SP2** was inconclusive with larger masses being observed. To confirm the structure of these side products a scale up and purification is required to enable full characterisation.

Starting material 18



SP1

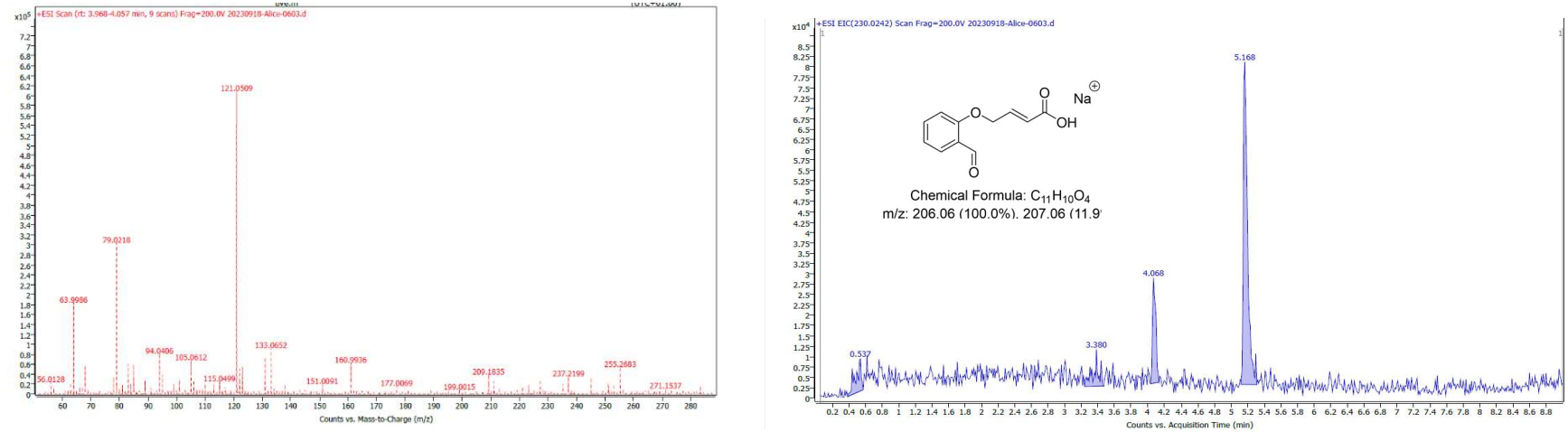


Figure 73: LCMS analysis of SM and SP1 highlights the difference in masses observed at the retention time of each. Extracted ion chromatograms of the expected mass of the starting material indicate it is not present in the peak corresponding to SP1 (retention time 3.968-4.073 min) whilst the peak for the acid is present in both.

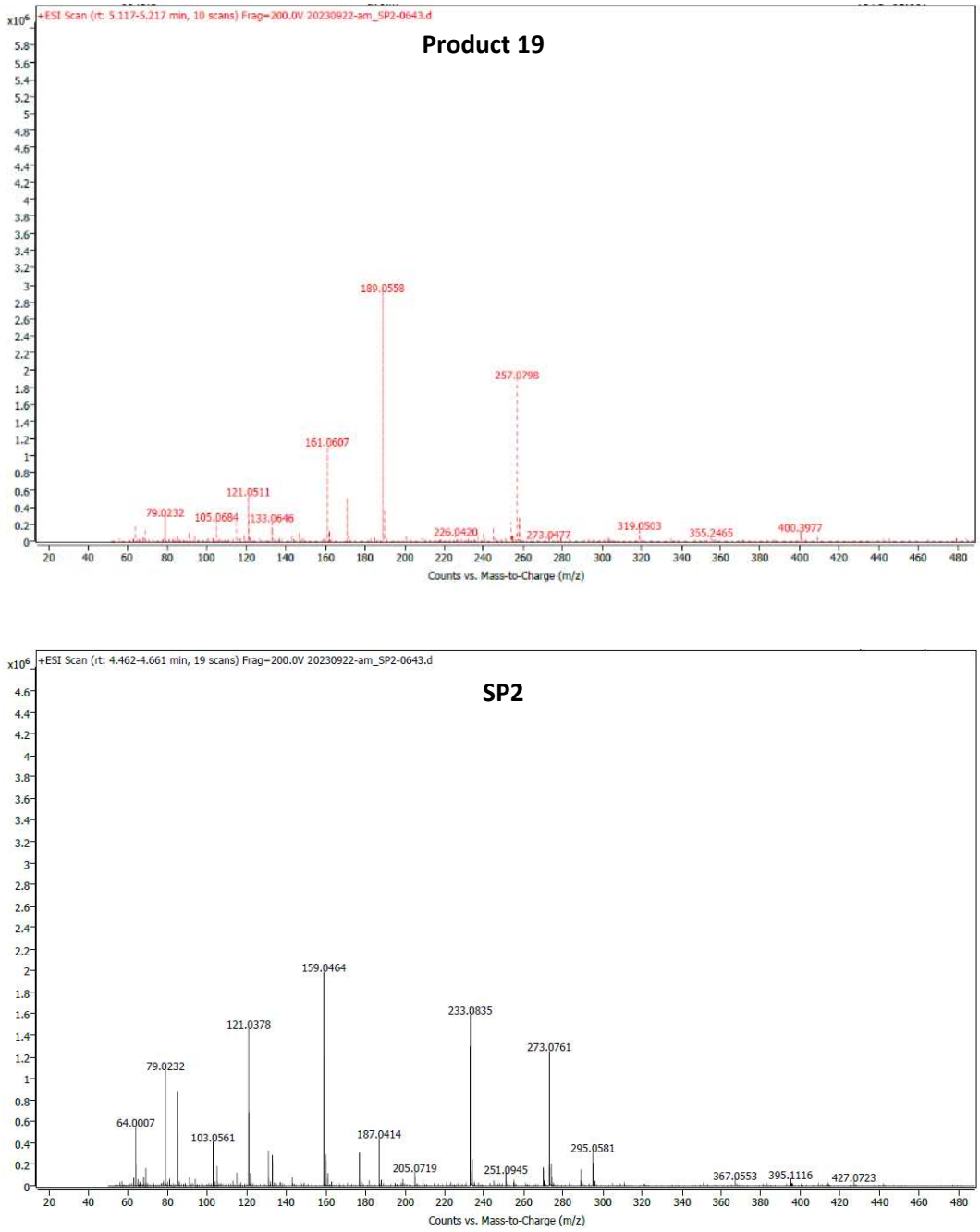


*Figure 74: LCMS analysis of product **19** and **SP2** displayed different fragmentation patterns.*

3.2.5 Intramolecular Stetter reaction using novel NHCs

Using the developed HPLC method that separates the starting material **18** and product **19**, we screened our synthesised thiazoles from Chapter 2 to assess their activity for the Intramolecular Stetter reaction in the absence of a protein. For compatibility with our RP C18 HPLC method, the organic reactions were carried out in methanol using 6 equivalents of triethylamine as a base whilst aqueous reactions were carried out in water again using triethylamine as a base. To ensure we observed product formation, high concentrations (100 mol %) of catalysts were used. It is important to note that tuning the base and solvent/cosolvent used can significantly increase the yields for NHC catalysed reactions, with different catalysts performing better under different reaction conditions[103, 223]. The aim of this screen was not to obtain maximum yields for each catalyst but to check the functionality of our catalysts. We also aimed to determine if any of the catalysts could function under aqueous conditions. Whilst all our novel NHC catalysed the reaction to some extent in methanol, none of them were as efficient as the catalyst **49** (a commonly used thiamine analogue where the pyrimidine is replaced by a benzyl group)**.** Interestingly, MBnThz was the best aqueous catalyst; however, we hope insertion into a protein will improve its catalytic activity. It should also be noted that the impurity of MBnTri may lead to it being less catalytically active. This NHC precursor should be revisited in the future.

*Table 13: HPLC yields of NHC catalysed intramolecular Stetter reactions. Reaction conditions: 2.5 mM starting material **18**, .5 mM Catalyst, 15 mM Et_3N in methanol or water with 10% DMSO. 30 °C, 18 hours, 250 rpm. *Impure catalyst likely lower concentration than reported.*

	Organic HPLC yield (mM)	**Aqueous HPLC yield (mM)**
Thiamine HCl	1.31±0.07	0.03±0.01
Catalyst 49	1.54±0.04	0.01±0.01
MBnTri *	0.02±0.01	-
BnThzBr	0.14±0.01	0.04±0.01
MeThzBr	0.27 ±0.02	0.01±0.01
MMeThz	0.53±0.03	-
MBnThz	1.43±0.10	0.15±0.02

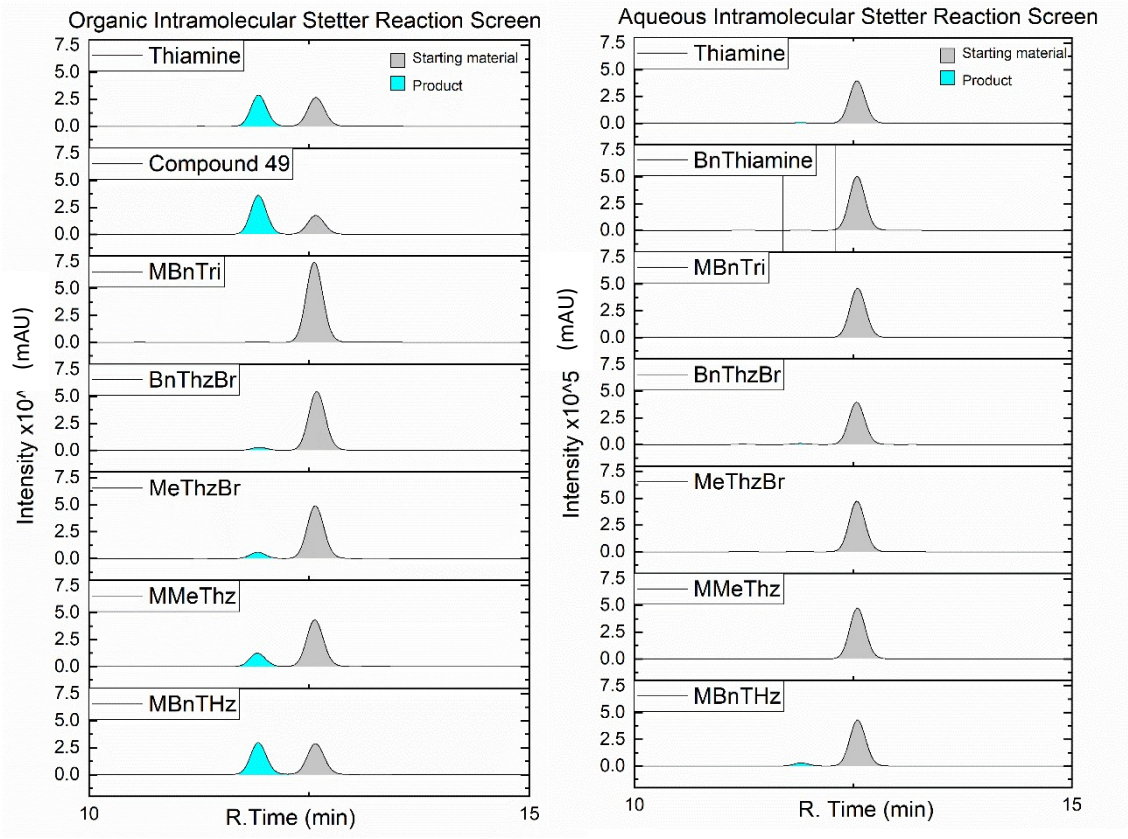


*Figure 75: HPLC analysis of product formation (cyan) during the Intramolecular Stetter reaction using our novel NHC precursors. Reaction conditions: 2.5 mM starting material **18**, 2.5 mM catalyst, 15 mM Et_3N in methanol or water with 10% DMSO. 30 °C, 18 hours, 250 rpm.*

3.3 Reactions of our Artificial Stetterases

With a method in hand for analysis of enzymatic intramolecular Stetter reactions, we then analysed the ability of our NHC functionalised protein scaffolds to catalyse this reaction (Figure 76). Reactions were carried out on a small scale as some of the functionalised protein scaffolds were difficult to synthesise in large quantities due to issues discussed in Chapter 2. Reactions were analysed via HPLC using a standard curve to calculate concentration of product formed. It should be noted that whilst maleimide thiazole functionalised pfThiE mutants were characterised in Chapter 2, they will not be discussed in this chapter as they did not display any catalytic activity.

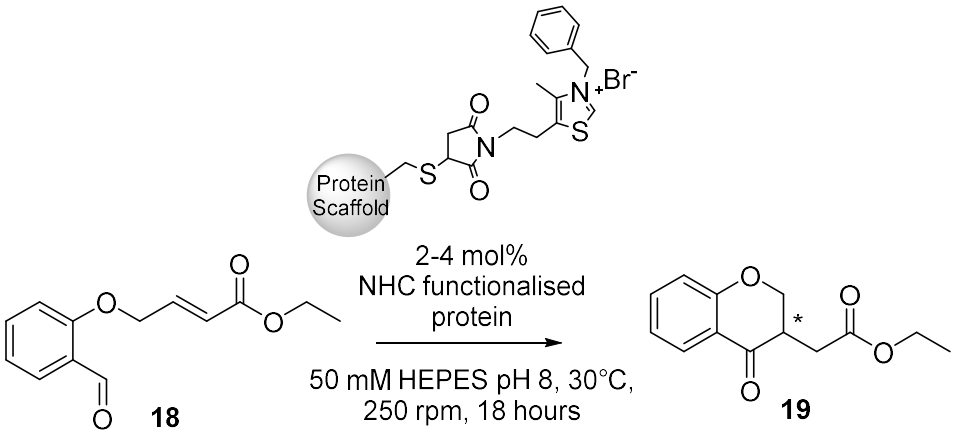


Figure 76: The Intramolecular Stetter reaction catalysed by different thiazolium functionalised SCP scaffolds

3.3.1 Maleimide thiazole functionalised ThAOS

As discussed in the previous chapter ThAOS proved particularly difficult to functionalise and resultantly only partially functionalised proteins were applied for catalysis. Control reactions undertaken with unfunctionalised protein showed that although the ThAOS protein scaffolds did not have any background Stetterase activity, they catalysed the formation of **SP1** (). The different ThAOS mutants led to varying amounts of side product with V79S K243C forming the largest amount. As the concentration of side product increased, the concentration of starting material decreased suggesting that the ThAOS mutants display some promiscuous activity with **18**. Analysing the structure, we hypothesised that a catalytic triad such as those seen in cysteine proteases may have been formed by introduction of the K243C mutation, although the distances and angles between residues are not optimal for hydrogen bonding[224] (Figure 78). This theory does not explain why the K243C mutant does not cause hydrolysis of

the starting material whilst the double mutant V79S K243C does. The V79 position is away from the PLP binding site and usually influences substrate scope. Upon mutation to serine, it is too distant for the histidine residues to participate in forming a catalytic triad, however it may allow better access for the substrate. Further studies are required to confirm why this mutant degrades the starting material.

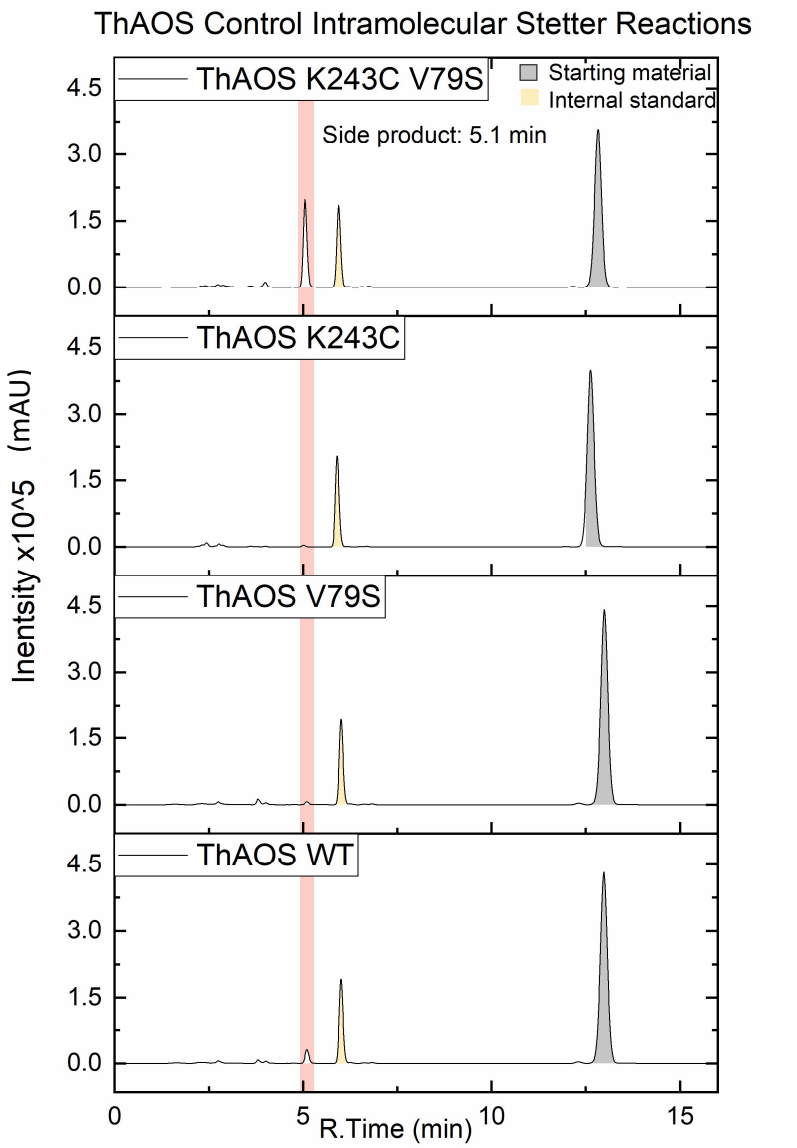


*Figure 77: HPLC traces from control the intramolecular Stetter reactions of 18 using unfunctionalised ThAOS mutants. Pink highlights the formation of **SP1**, yellow- internal standard and grey- starting material 18. Reaction conditions 2.5 mM 18, 100 µM enzyme, 50 mM HEPES pH8, 10% DMSO, 30 °C, 18 hours, 250 rpm.*

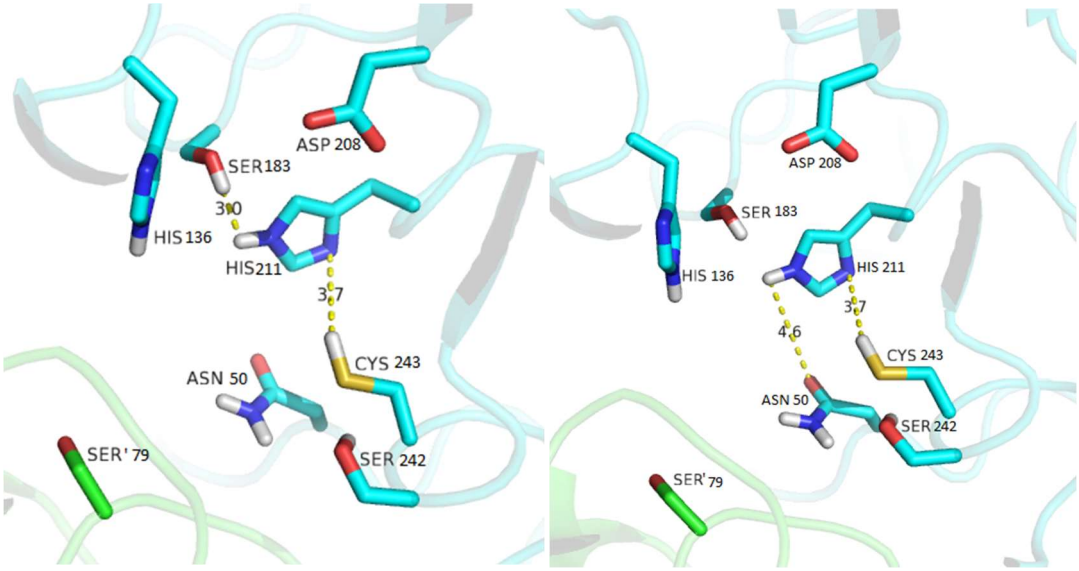


Figure 78: Hypothetical catalytic triads in a model of ThAOS K243C V79S created using Swiss model highlighting the distance between residues which could form a catalytic triad.

We theorised that functionalisation of the cysteine residue may decrease the formation of side product observed as the cysteine residue would no-longer be free to participate in side reactions. Therefore, we tested our MBnThz functionalised SCP scaffolds for their ability to catalyse the intramolecular Stetter reaction. We were encouraged to see that the functionalised ThAOS scaffolds catalysed the reaction, but unsurprisingly low yields were observed (Table 14). Reactions were performed in 50 mM HEPES pH 8, the same pH used for purification of ThAOS. Whilst product formation was observed a more prominent peak was observed for the formation of a side product at 5 min (**SP1,** Figure 79). As with the control reactions, this was much more prevalent in reactions catalysed by the double mutant K243C, V79S.

*Table 14: A summary of the HPLC yields from the intramolecular Stetter reaction of **18** catalysed by MBnThz functionalised ThAOS mutants from figure 58 and 59 chapter 2 with differences between K243C V79S mutants due to the functionalisation method. a) Direct functionalisation with MBnThz, b) Incubation with L-Pen then functionalisation with MBnThz, c) Incubation with L-Pen, dialysis, then functionalisation with MBnThz, d) Incubation with L-Pen buffer exchange then functionalisation with MBnThz. Reaction conditions 2.5 mM **18**, 100 µM – 200 µM functionalised enzyme, 50 mM HEPES pH8, 10% DMSO, 30 °C, 18 hours, 250 rpm*

Protein	Concentration (µM)	Functionalised	Product conc. (µM)
ThAOS K243C	200	< 5%	96
ThAOS K243C V79S a	100	<10 %	57
ThAOS K243C V79S b	100	28%	71
ThAOS K243C V79S c	100	<10%	49
ThAOS K243C V79S d	100	37%	56

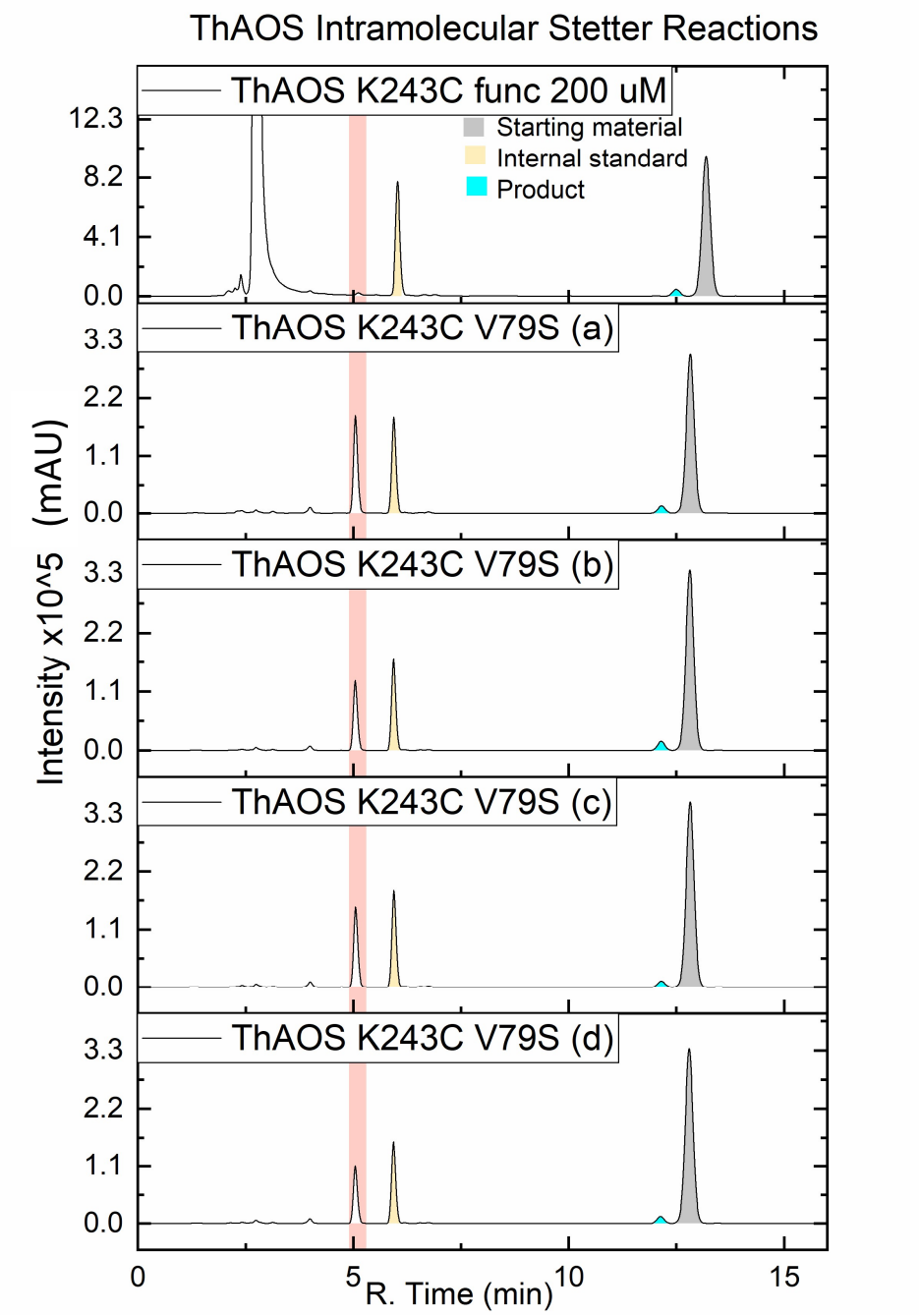


*Figure 79: HPLC traces from the intramolecular Stetter reaction of **18** catalysed by MBnThz functionalised ThAOS mutants from figure 58 and 59 chapter 2. Pink highlights the formation of **SP1,** yellow- internal standard, cyan- product **19**, grey- starting material **18.** Reaction conditions 2.5 mM **18**, 100 µM – 200 µM functionalised enzyme, 50 mM HEPES pH8, 10% DMSO, 30 °C, 18 hours, 250 rpm*

Encouraged by these results which showed for the first time that a protein functionalised with a thiazolium salt could be used to create an artificial Stetterase we decided to move on to test our other more easily functionalised scaffolds.

3.3.2 Maleimide Thiazole functionalised hSCPs

The Stetterase activity of the modified SCP scaffolds was assessed. Initially reactions were carried out at pH 6 with hSCP A100C as hSCP is stable under these conditions. However, no activity was observed so further experiments were untaken with functionalised hSCP A100C in 50 mM KPhos pH 7 using imidazole and DBU as bases. With MMeThz functionalised protein only trace amounts of product, barely above background level, were observed (Figure 80). Meanwhile, reactions with MBnThz functionalised SCPA100C displayed a small peak corresponding to the formation of the intramolecular Stetter reaction product (Figure 80). We noted that in our reactions, including control reactions, the addition of DBU led to the formation of a side products at 5 min, 7.1 min and 7.8-8 min, which correlated with a decrease in the concentration of **18** (Figure 80, Figure 81). The addition of base did not improve the catalysis of the functionalised SCP so further reactions were undertaken in aqueous buffer without the addition of bases to minimise the formation of unwanted side products.

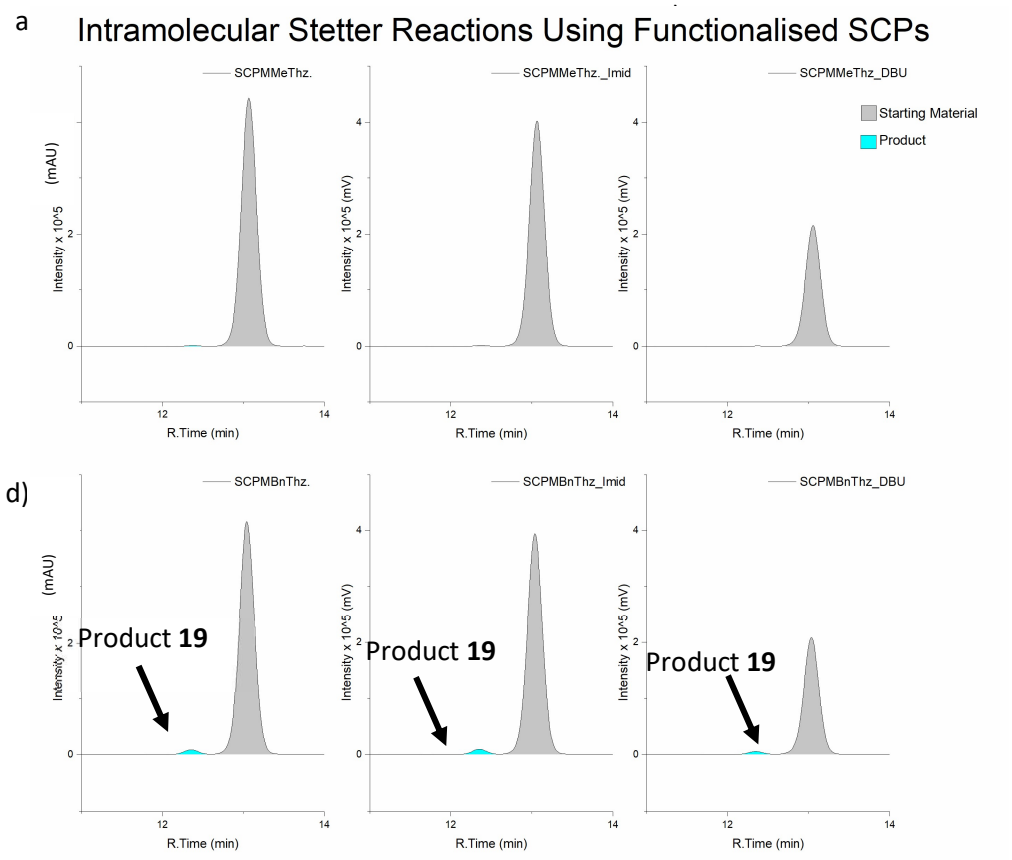


*Figure 80: The intramolecular Stetter reaction of **18** catalysed by functionalised hSCP A00C showing peaks corresponding to starting material **18** at 12.9 min (grey) and product **19** at 12.2 min (cyan). a-c) SCP A100C MMeThz, d-f) hSCP A100C MBnThz. Reaction conditions 2.5 mM **18**, 60 µM functionalised enzyme, 50 mM Kphos pH7, 10% DMSO, 30 °C, 18 hours, 250 rpm, with 20 mM imidazole (b and e) or 1.2 mM DBU (c and f)*

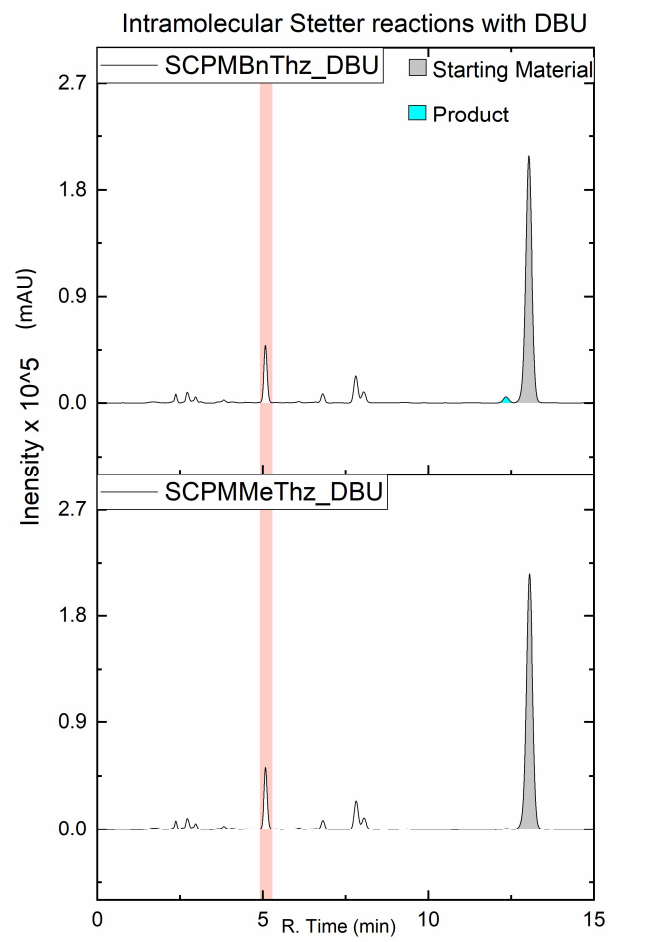


*Figure 81: HPLC analysis of the intramolecular Stetter reaction of **18** catalysed by functionalised hSCP A00C functionalised with MMeThz, and MBnThz highlighting the formation of a **SP1** at 5 mins as well as additional peaks at 7.1 and 7.8-8 mins. Reaction conditions 2.5 mM **18**, 100 µM functionalised enzyme, 50 mM Kphos pH7, 1.2 mM DBU ,10% DMSO, 30 °C, 18 hours, 250 rpm.*

We next considered if the buffer pH was suitable for enabling the deprotonation of the thiazolium salt. The pKa of our thiazolium salt MBnThz is likely to be similar to that of the core NHC structure ~18[99] and therefore we did not expect it to be dissociated under the reaction conditions. However, features of the protein scaffold may aid in deprotonation. Without a crystal structure for each functionalised hSCP cysteine mutant, we could not be sure of the position of the organocatalyst. Depending on the hSCP cysteine mutant used, the organocatalyst could be solvent-exposed or positioned differently inside the hydrophobic tunnel surrounded by protein residues. Therefore, we anticipated each of the hSCP would display different catalytic properties and optimal conditions. For example, SCP A100C F94H, initially designed for another project, contains a histidine residue intended to act as a base, which we hoped may be beneficial for our reaction. A pH screen was undertaken on all the functionalised hSCP scaffolds. Unfortunately, we were unable to test buffers above pH 8 as upon buffer exchange, a large amount of precipitation was observed. Subsequently, not enough protein could obtain enough to undertake reactions at this pH. This observed

precipitation at higher pH is understandable in accordance with the hSCP pI of 9.24. We were delighted to see product formation improved as the pH increased (Figure 82, Table 15), however increasing amounts of precipitate were also observed. All hSCP mutants displayed the same trend although less ppt was observed in hSCP V83C stable.

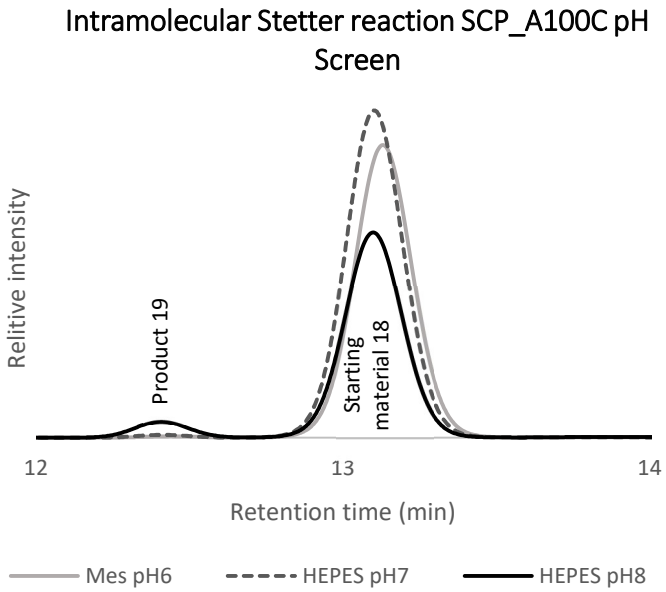
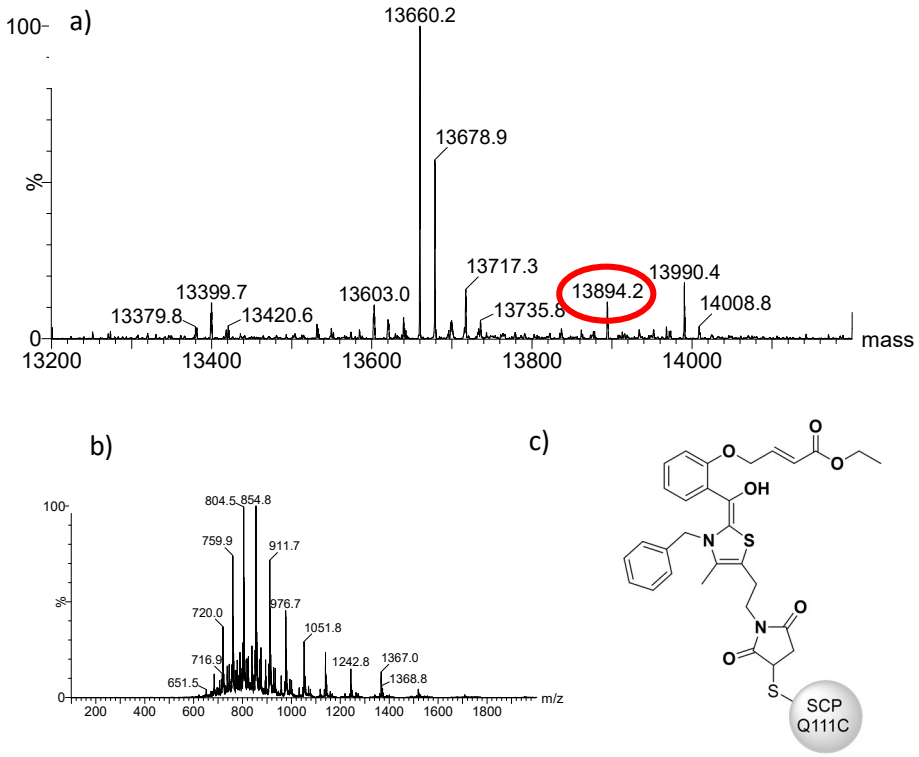


*Figure 82: HPLC traces of the intramolecular Stetter reaction of **18** carried out with hSCPA100C_MBnThz under a range of pHs.(20 mM Mes pH6, grey, 50mM HEPES pH7, grey dashed, 50 mM HEPES pH8 black) The peak corresponding to product **19** at 12.5 mins is observed only in the reaction performed at pH 8.*

*Table 15: Product concentrations from the pH screen of the Intramolecular Stetter reaction of 18 catalysed by MBnThz functionalised hSCP mutants. Reaction conditions 2.5 mM **18**, 100 µM functionalised SCP, 10% DMSO, 30 °C, 18 hours, 250 rpm in 20 mM Mes pH6, 50mM HEPES pH7 or 50 mM HEPES pH8*

Protein Scaffold	Product conc.(µM)
V83Cstable pH 8	141
A100C pH 8	129
A100C F94H Ph8	126
Q111C pH 8	112
V83C pH 8	92
A100C pH 7	72
Q111C pH 7	42
V83Cstable pH 7	38
V83Cstable pH 6	32
V83C pH 7	30
A100C pH 7	23
A100C F94H pH 6	21
Q111C pH 6	20
A100C pH 6	13
V83C pH 6	10
Control no enzyme	6

Despite a slight increase in activity at pH 8, yields were still low. We hypothesised that the reaction may be stalling at the Breslow intermediate. Therefore, we analysed a sample reaction of hSCP Q111C MBnThz after incubation for 1 hour at 30 °C using LC-MS and observed a minor peak in the mass spec corresponding to the mass of the adduct with the Breslow intermediate (mass 13894.2, Figure 83). This was an interesting observation that shows the Breslow intermediate is being formed, nevertheless there is not a significant build-up of this adduct so it does not appear that the reaction is stalling here.



Unfunctionalised protein mass: 13347.61
Mass upon functionalisation with MBnThz: 13661.00
Expected mass upon addition of Breslow intermediated: 13894.21

*Figure 83: a) Deconvoluted spectra of the protein LC MS analysis of the intramolecular Stetter reaction of **18** catalysed by hSCP Q11C MBnThz. A adduct corresponding to the mass of the protein attached to the Breslow intermediate was observed (red circle). Reaction conditions: 2.5 mM **18**, 100 µM functionalised enzyme, 10% DMSO, 50 mM HEPES pH8, 30 °C, 1 hours, 250 rpm b) Ion series of the LCMS analysis. c) Structure of the expected adduct formed when the intramolecular starting material **18** forms a Breslow intermediate with MBnThz functionalised SCP Q11C.*

Finally, we assessed the enantioselectivity of the MBnThz functionalised hSCP mutants. We were excited to observe that some enantioselectivities was observed for reactions carried out with all of the functionalised hSCPs. Interestingly the e.e. (enantiomeric excess) appeared to vary slightly between constructs indicating that the protein environment may be influencing the reaction. However due to the relatively low e.e. we hypothesised that the reaction may not be occurring inside the cavity or that the cavity site is large enough to

allowing multiple conformations leading to a lack of control of enantioselectivity. Obtaining protein crystal structures of the functionalised hSCP scaffold would be a useful next step to aid the improvement of these scaffolds by site directed mutagenesis. The other alternative would be to perfume random mutagenesis and high throughput screening however this would be particularly difficult for this project as protein functionalisation is tricky to perform in high throughput, and functionalisation conditions can vary for each construct.

*Table 16: Enantioselectivities of MBnThz functionalised hSCP mutants calculated from chiral HPLC analysis of the intramolecular Stetter reaction of **18.** Reaction conditions: 2.5 mM **18**, 100 μM functionalised enzyme, 50 mM HEPES pH8 10% DMSO, 30 °C, 18 hours, 250 rpm*

SCP Scaffold	e.e
V83Cstable	-3.1
A100C	-1.9
A100C F94H	0.6
Q111C	-4.8
V83C	-6.3

3.3.3 Maleimide thiazole functionalised TTSCPs

The Stetterase activity of the three MBnThz modified scaffolds (TTSCP L112C, TTSCP L102C and TTSCP ΔDSB L102C) were analysed using the most successful conditions from the functionalised hSCP constructs. We were optimistic these TTSCP scaffolds could enable us to overcome issues we observed with the precipitation of the hSCP scaffold due to their increased thermostability and acidic pI. We were disappointed to observe that the easiest scaffold to functionalise, TTSCP L112C, displayed no catalytic activity above background, although this was useful in serving as a negative control (Table 17, Figure 84). We were pleased to observe the TTSCP ΔDSB L102C scaffold displayed a 2-fold increase in activity over the previous hSCP A100C scaffold (2.2 turnovers, 8 % yield, Table 17, Figure 84). To our surprise, when we used the disulfide-containing scaffold TTSCP L102C, the activity increased ~ 20-fold, considering the degree of functionalisation (22 turnovers, 23% conversion, Figure 84, Table 17). Interestingly this most catalytically active scaffold was the most difficult to selectively functionalise (~50% functionalisation) and as such was used at half the concentration (50 μM, 2 mol %) compared to the other catalysts. Yields were comparable to those achieved with functionalised thiazolylalanine peptides in organic solvents[113] but still fell short of those achieved by PfBAL.[76] Importantly, control reactions with non-specifically functionalised TTSCP (functionalisation conditions: 20 mM Mes pH7, 25 equiv MBnThz, 5 h, RT) and unfuntionalised TTSCP L102C did not lead to product formation above background

levels suggesting neither the attachment of MBnThz to the disulphide cysteine residues nor catalysis by the L102C residue leads to product formation.

*Table 17: Activity of functionalised SCP scaffold for the intramolecular Stetter reaction (2.5 mM **2**, 50 mM HEPES pH8, 30 °C, 250rpm, 18 hours) [a] Functionalisation percentage determined by LC-MS [b] HPLC yield*

Protein	Concentration (µM)	Mono-functionalisation[a] (%)	Conc. of product (µM)	Yield[b] of 19 (%)	TTN
MBnThz (control)	100	-		<1	<0.1
TTSCP	100	0 (control)		<1	<0.1
TTSCP (non-specific functionalisation)	100	20 %		<1	<0.1
TTSCP L102C	100	0 (control)		<1	<0.1
TTSCP W83C	100	<5		-	-
TTSCP L112C	100	85		<1	<0.1
TTSCP ΔDSB L102C	100	90	196± 5	7.8	2.2
TTSCP L102C	50	50	557±10	23	22

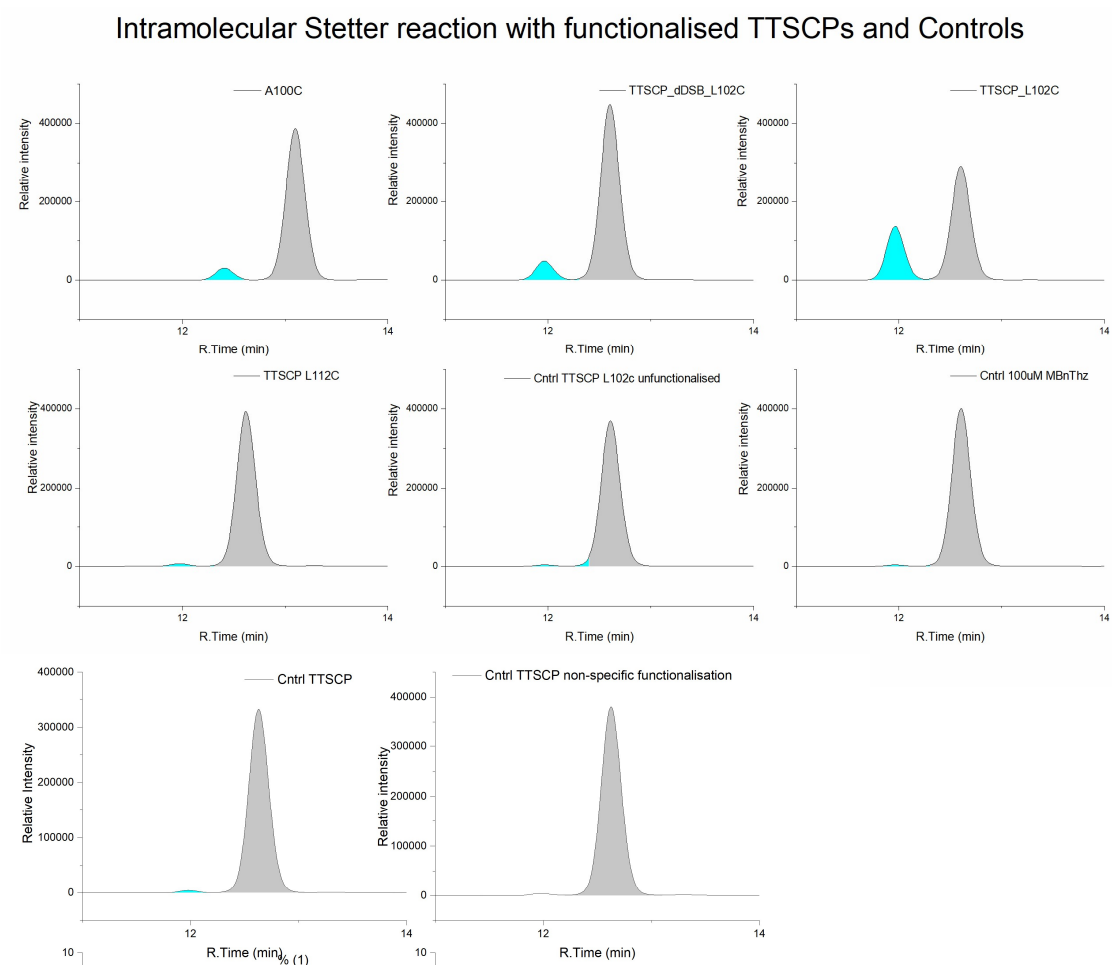


*Figure 84: HPLC traces showing the formation of product 19 at 12.02 min (cyan) and residual starting material 18 at 12.61 min (grey) in the intramolecular Stetter reaction catalysed by maleimide thiazole functionalised SCPs. Reaction conditions: 2.5 mM **18**, 100 µM functionalised enzyme, 50 mM HEPES pH 8, 10% DMSO, 30 °C, 18 hours, 250 rpm*

With an aim to increase the yield of product, further experiments were performed to optimise the conditions for the intramolecular Stetter reaction with MBnThz functionalised TTSCP L102C and TTSCP ΔDSB L102C. A new batch of functionalised enzyme was synthesised; however, protein functionalisation was difficult to repeat exactly, which was believed to be caused by fluctuations in room temperature. After the optimised functionalisation time, the protein was still less than half functionalised, so it was incubated overnight, resulting in over functionalisation. Therefore, the proteins used for these reactions were 47 % mono-functionalised TTSCP L102C and 66 % mono-functionalised TTSCP ΔDSB L102C. These proteins were used to catalyse the intramolecular Stetter reaction in 100 mM CAPS buffer pH 9.5 to determine if a higher pH was beneficial to the reaction. Interestingly using the same batch of functionalised protein under otherwise identical conditions, less product was formed at pH 9.5 (35.8 ±0.3 μM) compared to pH 8 (52.2± 1.7 μM). Next, a time dependant study was undertaken to assess if the reaction had reached completion after 18 hours. Although the reactions did not go to completion, a maximum concentration of product was recorded after 16 hours before its concentration began to decrease. The data suggests that the reaction stalls at this point and the product starts to slowly degrade. Precipitation was not observed in the reactions therefore further analysis is required to determine the reason for the reaction stopping before all the starting material is consumed. For example, protein LCMS should be performed to assess if the functionalised protein is still present after 16 hours and if any adducts can be observed.

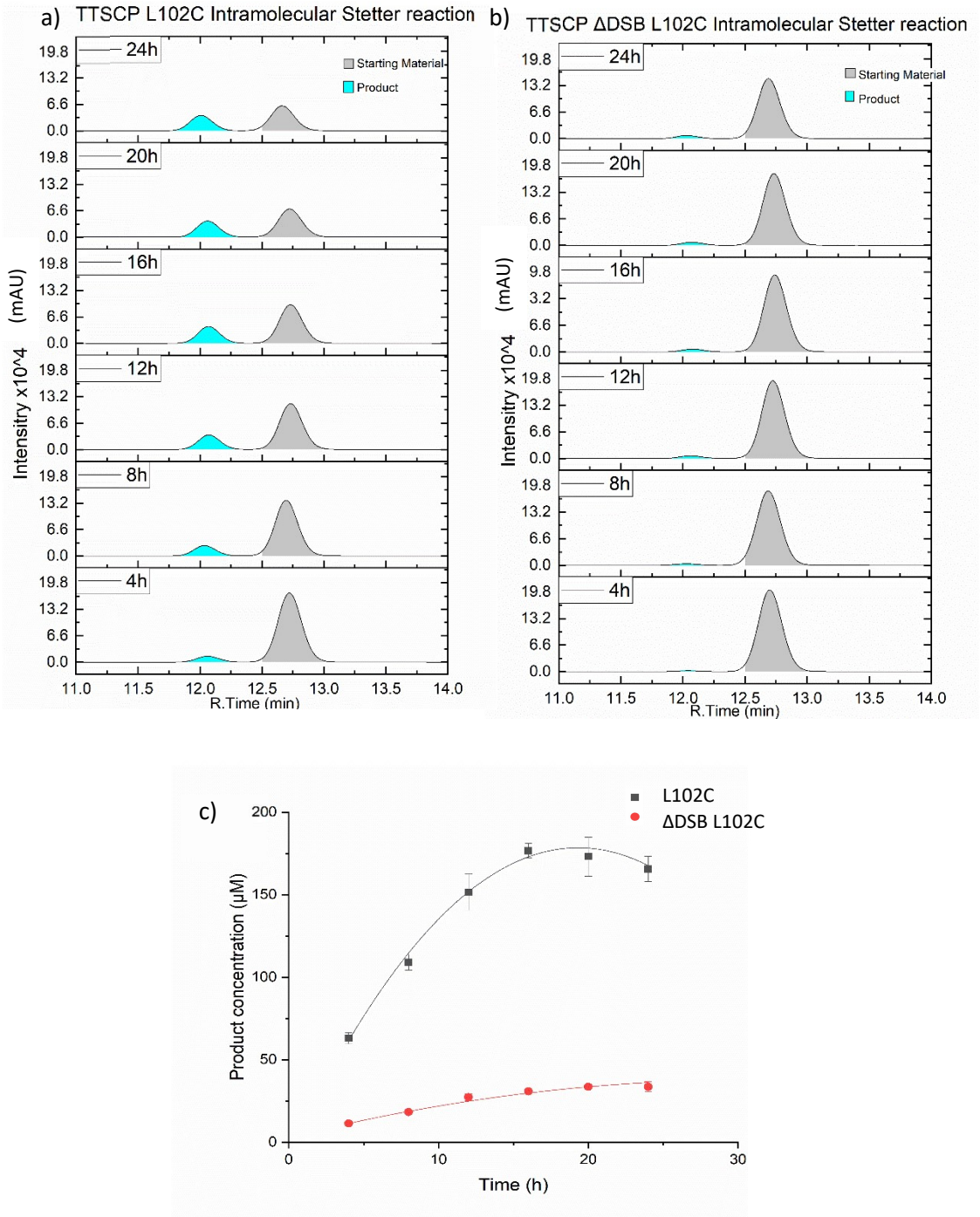


*Figure 85: HPLC traces of starting material (grey) and product (cyan) formation during a time course assay of the intramolecular Stetter reaction of **18** catalysed by a) TTSCP L102C MBnThz and b) TTSCP ΔDSB L102C MBnThz. c) highlights the change in product concentration over time for each functionalised enzyme. It shows a maximum concentration of product 19 is formed between 16-20 hours with TTSCP L102C MBnThz before product concentration begins to decrease. Reaction conditions: 1 mM **18**, 100 μM functionalised enzyme, 50 mM HEPES pH 8, 10% DMSO, 30 °C, 250 rpm*

Investigations into the effect of temperature on the intramolecular Stetter reaction catalysed by TTSCP L102C MBnThz showed higher temperatures lead a lower yield of product and to the formation of side products which had been previously observed (Figure 86).

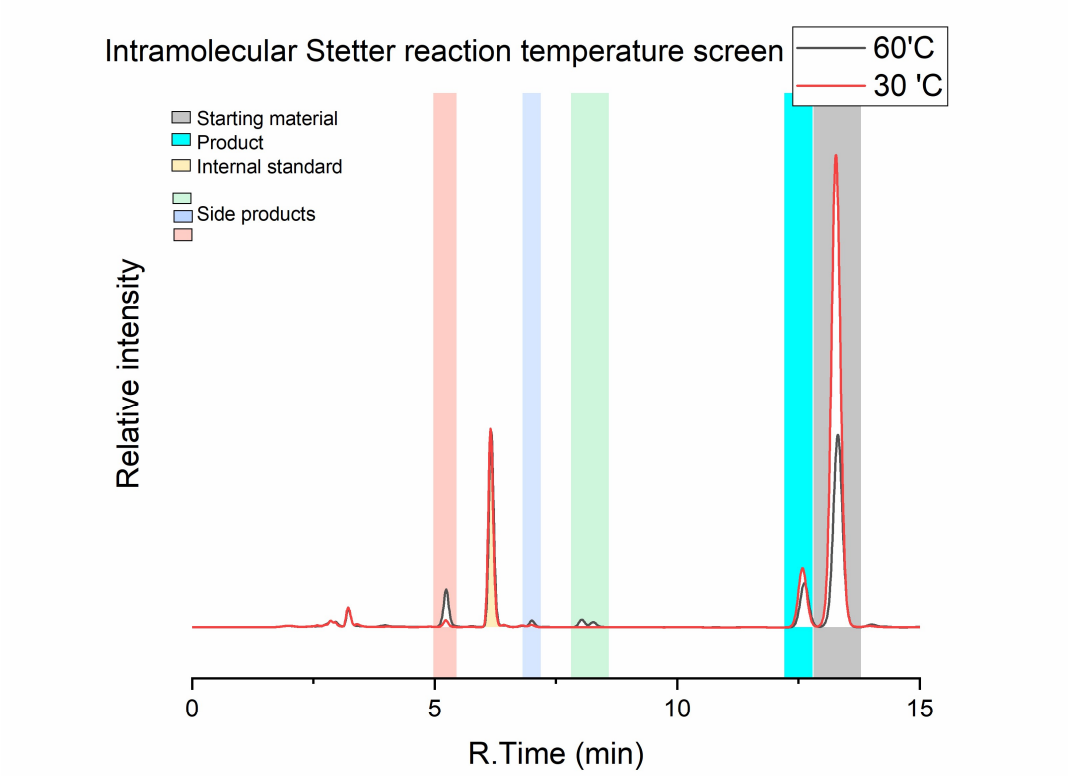


*Figure 86: HPLC traces showing the formation of product 19 (cyan) and residual starting material 18 (grey) in the intramolecular Stetter reaction catalysed by MBnThz functionalised TTSCP L102C at 30 °C (red) and 60 °C (black). Side products formed at 60 °C are highlighted in pink, light blue and green. Reaction conditions: 2.5 mM **18**, 100 µM functionalised enzyme, 50 mM HEPES pH 8, 10% DMSO, 18 hours, 250 rpm*

In order to understand why this TTSCP L102C mutant was catalytically superior, particularly compared to the TTSCP ΔDSB L102C variant, we reviewed our CD, modelling and docking performed in chapter 2. The CD analysis indicated that the TTSCP ΔDSB L102C mutant, although slightly less stable than the L102C mutant, retains a T_m of 79 °C and therefore should be stable throughout the reaction. It is worth noting that although CASTp analysis of our AlphaFold models showed different cavity sizes for our mutants, open and closed conformations of the TTSCP scaffold are observed in the two chains present in the crystal structure (PDB:2CX7), whilst the NMR structure (PDB:1WFR) only displays a closed conformation. From our modelling alone, it is difficult to determine why the two functionalised scaffolds display such different catalytic activities; however, it seems likely to be linked to both conformational changes and protein dynamics, rather than the reduced stability resulting from removal of the disulfide bond. However structural analysis of TTSCP L112C does help us to understand why this mutant is not catalytic (Figure 87).

Functionalisation of this position may result in the cavity being blocked or the functionalisation reagent sticking out into the solvent.

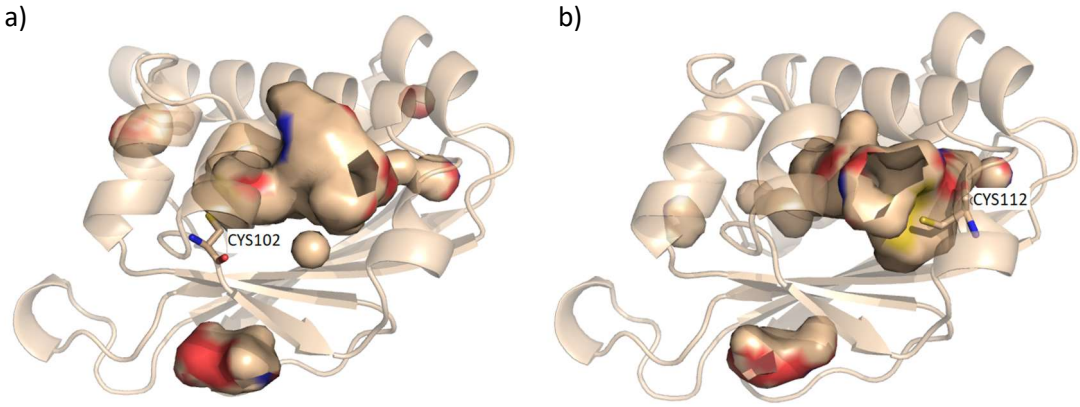


Figure 87: Comparison of AlphaFold structures of a) TTSCP L102C and b) TTSCP L112C shows the different positions of the cysteine residues with Cys 112 sitting right at the entrance to the TTSCP cavity.

Importantly, we also examined the stereoselectivity of the reaction catalysed by the MBnThz modified TTSCP L102C. This catalyst displayed a modest, but clear e.e 5.5 % ±0.5 of product **18** (Figure 88). This exciting observation allows comparison with the PfBAL enzyme that catalyses a stereoselective intramolecular Stetter reaction, which converts **18** to **19** with 98% e.e under optimised conditions. Interestingly our functionalised biocatalyst displays a preference for the formation of the (*R*)-enantiomer, whilst PfBAL catalysed formation of the (*S*)-enantiomer. An important next step is determination of the structure of this novel

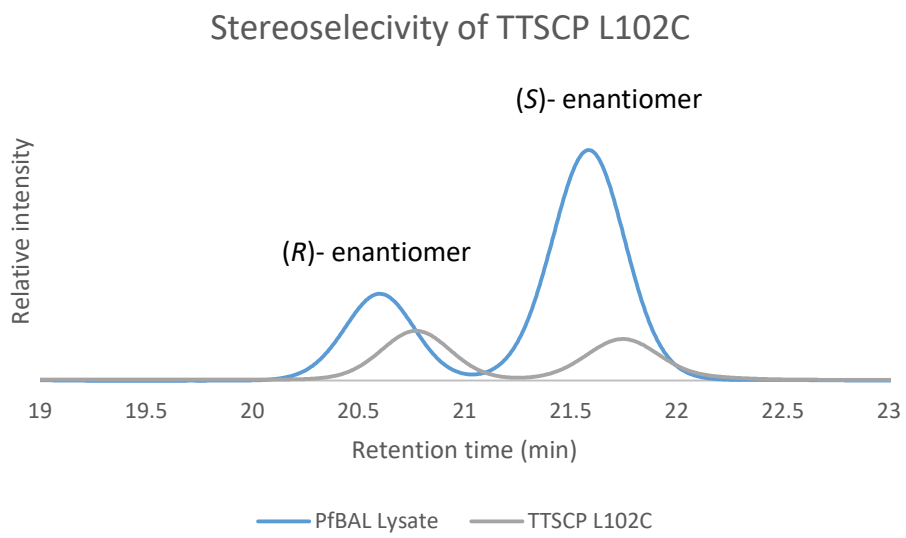


*Figure 88: Chiral HPLC analysis of the stereoselectivity of intramolecular Stetter reaction catalysed by TTSCP L102C compared to pfBAL. The (R)-enantiomer elutes first at 20.5-20.8 min and the (S)-enantiomer elutes at 21.5-21.8 min. The Reaction conditions: 2.5 mM starting material **18**, 100 µM functionalised TTSCP L102C, 50 mM HEPES pH 8, 10% DMSO, 30 °C, 18 hours, 250 rpm.*

catalyst to identify where the NHC sits within the cavity. This will enable analysis of which residues may be involved in catalysis and enable semi-rational engineering to improve activity and stereoselectivity.

3.3.4 Maleimide Triazole functionalised TTSCPs

When functionalised with MBnTri we were disappointed to see that we did not obtain any product from the intramolecular Stetter reaction of **18** (Figure 89). Minimal side products were formed during the reaction. Whilst the structural difference between MBnTri and MBnThz could account for the difference in activities we do also need to take into account the purity of MBnTri. Although the functionalised protein appeared to be functionalised only with the expected mass for MBnTri we cannot be sure of structure of the catalyst without further analysis. However, other factors such as the longer linker between the N-heterocycle and the maleimide in MBnTri (3 carbons vs 2 carbons in MBnThz) will alter how the NHC precursor is positioned in the active site. Docking studies of the Breslow intermediate of MBnTri with **18,** highlight that Cys 102 is not in the proximity of the maleimide in any of the

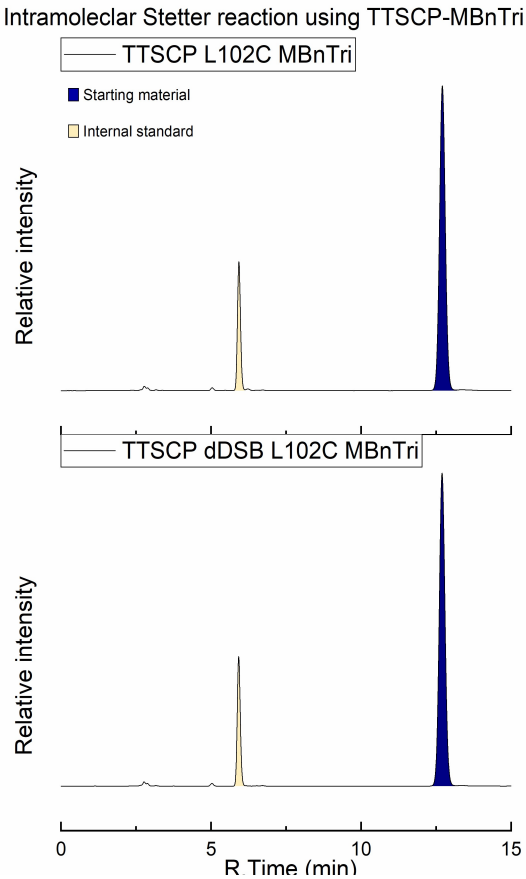


*Figure 89: HPLC analysis of the Intramolecular Stetter reaction catalysed by MBnTri functionalised TTSCP L102C (top) and MBnTri functionalised TTSCP ΔDSB L102C (bottom) did not show any product formation expected to elute at 12.02 min whilst the starting material **18** at 12.61 mins remains unreacted. . Reaction conditions: 2.5 mM **18**, 100 µM functionalised enzyme, 50 mM HEPES pH 8, 10% DMSO, 30 °C, 18 hours, 250 rpm*

four docked conformations (Figure 90). This may prevent the Breslow intermediate from forming when MBnTri is inside the TTSCP cavity; ideally TTSCP cysteine mutants should be designed specifically for each organocatalyst with a range of maleimide linked triazolium salts being synthesise and screen in combination with our protein scaffolds.

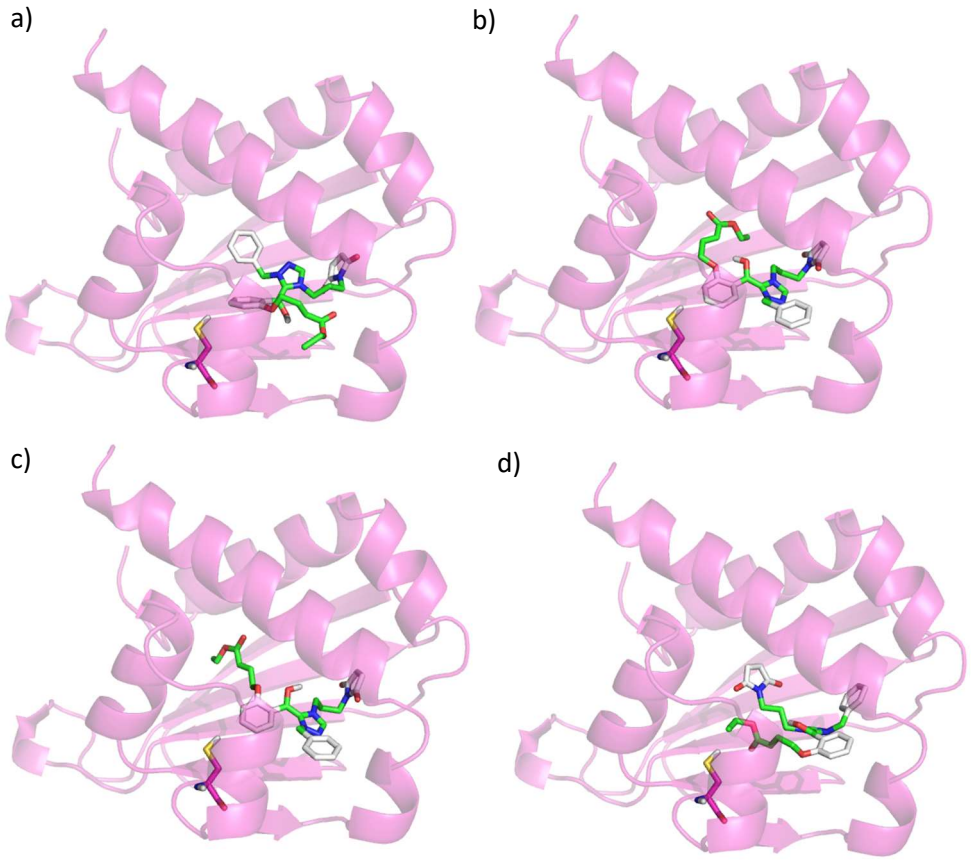


*Figure 90: Docking of the Breslow intermediated of MBnTri with **18** (green) into an alpha fold model of TTSCP L102C using AutoDock Vina. Residue Cys 102 (shown as pink sticks) is not in close proximity to the maleimide of MBnTri in any of the 4 poses (a-d)*

3.3.5 Screening alternative NHC catalysed reactions

Although our functionalised scaffolds were designed to catalyse an intramolecular Stetter reaction, we wanted to assess their reaction scope. We hypothesised that both intermolecular Stetter reactions and benzoin condensations may be catalysed by our functionalised protein scaffolds. Interestingly, no activity was observed when using functionalised TTSCPs for the benzoin condensation and intermolecular Stetter reaction between benzoin and chalcone. Whilst conditions were not optimised for this reaction we were surprised by this lack of reactivity. It may suggest the cavity of TTSCP is not well suited to these substrates. Other scaffolds should also be assessed for their ability to catalyse these reactions as each hydrophobic cavity is a distinct size and shape. Docking studies should be untaken to assess if any of our other functionalised protein scaffold would suit these reactions.

3.4 Conclusions and future work

We set out to combine the power of organocatalysis with the selectivity of biocatalysts and pave the way for creating tuneable biohybrid catalysts for organic synthesis[7]. To that end, we created the first artificial Stetterase by bioconjugation of an NHC to an inert protein scaffold. We highlight how important the choice of scaffold is and how even without applying protein design/evolution, scaffold choice can lead to a ~20-fold increase in catalytic activity. The successful demonstration that the TTSCP L102C MBnThz pairing is catalytically active in aqueous conditions paves the way for the screening of different protein/NHC combinations that can catalyse a range of chemical reactions. Furthermore, it is hoped that this concept could be used to harness the diverse chemistry and catalytic repertoire of NHC organocatalysts once they are incorporated into the chiral environment of a protein. Whilst our triazolium functionalised scaffold was not catalytic for the intramolecular Stetter reaction of **18**, it should be noted that we believe a scaffold designed specifically for this organocatalyst would show functionality.

Whilst the intramolecular Stetter reaction of **18** is a benchmark for assessing chiral NHC catalyst, we encountered issues with the formation of side products. Further work to fully characterise these side products should be undertaken, including scaling up of reactions to isolate side products in large enough quantities NMR studies. Investigations into alternative Stetter and benzoin reactions should also be pursued.

Structural studies and enzyme engineering should now be undertaken to improve the activity and enantioselectivity of this catalyst. It is appreciated that there are technical challenges associated with selective and efficient attachment of the NHC to the selected scaffold. One way to overcome this is to use genetic code expansion to incorporate an NHC based amino acid as an alternative to bioconjugation which we investigate in the next chapter. A preprint article recently published by the Baker group presents a new tool for the incorporation of non-proteinogenic components[220]. RoseTTAFold All-Atom is a deep network capable of modelling full biological assemblies including proteins with covalent modifications[220]. This would be a particularly useful method to employ for this project as it would enable better predictions of how the organocatalyst is orientated in the functionalised protein, thus aiding in the design of organocatalyst/protein scaffold pairs for specific reactions.

4 Genetic Code Expansion

Due to precipitation and non-specific functionalisation of the TTSCP scaffold observed during the previous chapter, we considered alternative ways to covalently incorporate an NHC into a protein scaffold. One option is to use genetic code expansion (GCE), described in the introduction, which can add new functional motifs to proteins by incorporating novel unnatural amino acids (UAAs). GCE uses and orthogonal translation system (OTS) combined with the cellular machinery to selectively functionalise a protein at a specific position. This method of NHC incorporation would also allow for easier directed evolution of the NHC functionalised protein. Here we aim to use GCE via stop codon suppression to incorporate a UAA with the catalytic functionality of an NHC.

It has already been shown that UAA based amino acids can catalyse intramolecular Stetter reactions in organic solvents. Mennen *et al.*[113] used functionalised thiazolylalanines, both as protected amino acids and incorporated into synthetic peptides prepared by solid phase peptide synthesis (SPPS) to perform intramolecular Stetter reactions using salicylic acid derived substrates. However, they did not try to incorporate these UAAs into a protein and to our knowledge no other NHC based UAAs have been used for catalysis or incorporated into a protein. In this chapter we investigate and screen existing OTSs for their ability to incorporate NHC based UAAs into proteins. We investigated systems based on both the tyrosyl tRNA/aminoacyl-tRNA synthetase pair and the pyrolysyl tRNA/aminoacyl-tRNA synthetase pair. The features in each of these systems could be beneficial depending on if a longer or shorter distance between the protein backbone and the NHC functional group is required.

One of the systems we investigate is the pEVOL system, this was first reported in 2010 for the incorporation of UAAs into proteins in *E. coli*[225]. It consists of a single vector containing two copies of the *Methanocaldococcus jannaschii* aminoacyl-tRNA synthetase(s) (aaRS) under control of a dual promotor system and an optimised suppressor tRNA pair ($tRNA^{opt}_{CUA}$) with a modified T-stem[225]. This system has been shown to outcompete previously reported systems: pBK[226] which requires two plasmids as the aaRS is on a different plasmid to the suppressor tRNA and pSUP[227] a single plasmid system with a different construction to pEVOL which doesn't use the optimised amber suppressor. The use of mutant aaRS libraries and rounds of positive and negative selection have led to many variants of the pEVOL system. We focus on two of these, pEVOL AZF[228] which incorporates p-azido-l-phenylalanine a widely

used photocrosslinking agent and pEVOL NMH[229] which selectively incorporated 3-N-methylhistidine (3NMH). pEVOL NMH was created by inserting two copies of an evolved pyrrolysyl-tRNA synthetase into the pEVOL backbone[229, 230]. We believe this mutant is particularly promising for incorporation of a thiazolium based UAA as, since our work began, it has also been shown to incorporate 5-thiazoylalanine(5ThzA)[231].

We also investigate pTECHchpylRS IPYE[232], pULTRA CNF[233] and pBK-ThzKRS[234]. The pTECHchpylRS IPYE is another single plasmid system consisting of a chimeric *Methanosarcina spp.* pyrrolysyl-tRNA synthetase (PylRS) with activating mutations V31I, T56P, H62Y, A100E (IPYE) and the pyrrolysyl tRNA[232]. The mutations impart a 45-fold increase in enzymatic efficiency compared to the parent enzyme[232]. It was shown to incorporate commercially available Boc-lysine as an alternative to pyrrolysine, suggesting it has some promiscuity[232]. pULTRA CNF is a tyrosyl tRNA/aminoacyl-tRNA synthetase based single plasmid system evolved for incorporation of para-cyanophenylalanine an infrared reporter of protein structure[233]. Finally, we also chose to investigate a two plasmid system pBK-ThzKRS[234] in combination with pGFP$_{150}$TAG pylT which was evolved for the incorporation of N$^{\varepsilon}$-L-Thiaprolyl-L-lysine (ThzK). We were optimistic that this construct could be a good candidate for incorporation of an NHC based UAA with a thiazole backbone due to the similarity in structure (Figure 91).

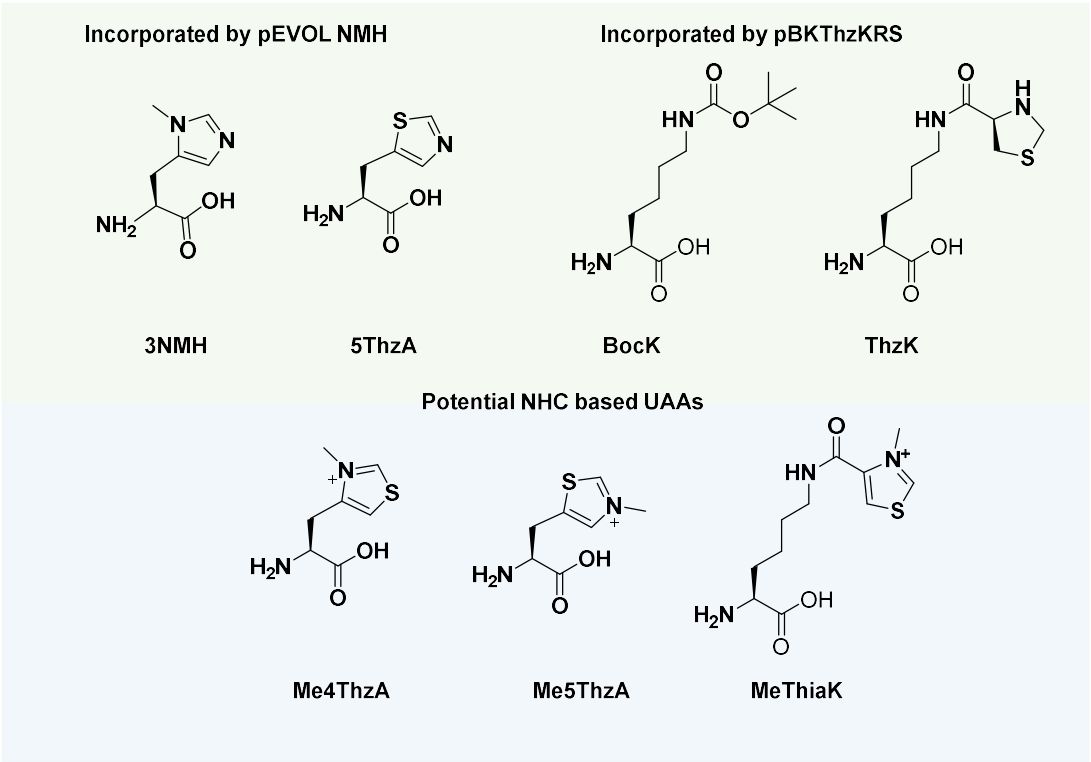


Figure 91: UAAs incorporated by pEVOL NMH and pBKThzKRS and the potential NHC based UAAs they inspire

In order to test these OTSs we were first required to synthesise NHC based UAAs. Ideally, we would use an enzymatic approach to synthesise the UAAs *in vivo.* This involves identifying and engineering several enzymes. For example, we could make use of enzymes such as TvaG[235], carnosine methyl transferase (CARNMT)[236, 237] and O-acetylserine sulfhydrylase[238] (Figure 92). TvaG is the first enzyme known to catalyse the bis-N,N-dimethylation of histidine as a PTM, creating an NHC *in vivo*, however it only accepts a cyclic peptide as a subtrate[235]. Interestingly results indicate, TvaG may favour singularly methylated substrate over histidine which could be useful for engineering efforts[235]. CARNMT's natural reaction is the methylation of anserine, however, it has also been shown to accept histidine as a substrate (Figure 92). Alternatively, if we were interested in the synthesis of triazole based UAAs, O-acetylserine sulfhydrylase[238] has been shown to be able to synthesise these from the triazole and acetylserine (Figure 92). Further functionalisation of these triazole UAAs would then be required to produce the NHC. As it can be seen, a lot of work is required to build these into successful cascades for the synthesis of NHC based UAAs with enzyme engineering or combination with chemical processes to give the required UAA. Therefore, this something to consider in the future but for now to allow us to screen NHC based amino acids without a known biocatalytic synthesis route we synthesised them chemically.

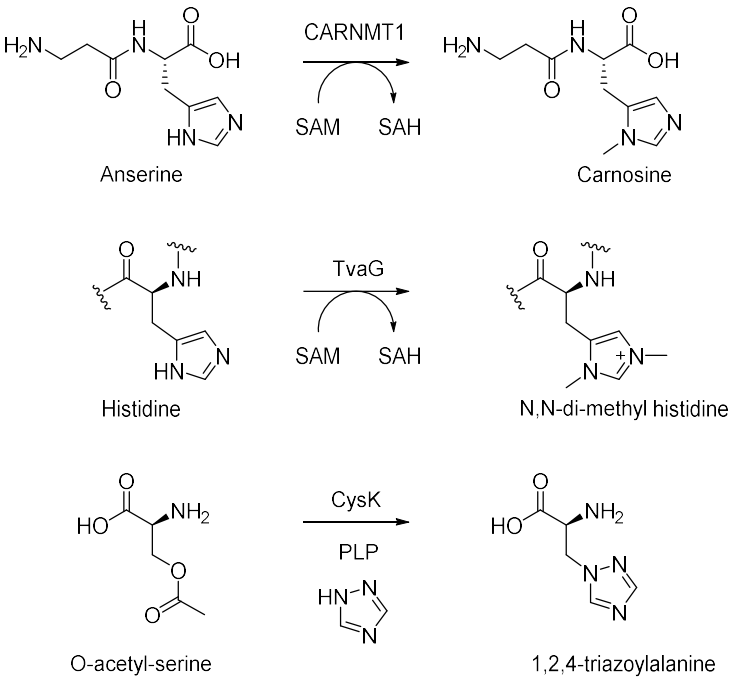


Figure 92: Reactions of enzymes which could be used for the synthesis of NHC based UAAs. ***a)*** *anserine to carnosine by CARNMT-1* ***b)*** *di-methylation of a histidine containing cyclic peptide by TvaG* ***c)*** *synthesis of 1,2,4 triazoyl amino acid by CysK*

4.1 Aims

The aim of this part of the project was to synthesise NHC based unnatural amino acids and incorporate them into a protein via genetic code expansion. To achieve this goal, 3 different NHC based UAAs were synthesised and existing OTSs were screened to determine if they incorporated the UAA into a protein.

4.2 Synthesis of thiazole based UAAs

4.2.1 Synthesis of 3-methyl-4-alaninethiazolium chloride

The UAA 3-methyl-4-L-alaninethiazolium chloride (L-Me4ThzA, **53**) was synthesised from commercially available N-boc-L-4-thiazolylalanine (**51**) in a two-step process with a 12% yield (Figure 93). Products were analysed by NMR spectroscopy (Appendix 1.6.8). A labile proton of the thiazolium C2 was observed. Exchange at this position is important as the pKa of this proton is essential for activity of the UAA in formation of the nucleophilic carbene[56]. ^{2}H NMR spectroscopic analysis was used to confirm proton exchange with the deuterated solvent at this position. The NMR data was unable to clarify that methylation was occurring on N therefore crystals of the UAA were grown by the slow evaporation solvent. Gary Nicol at the University of Edinburgh determined the structure by single crystal x-ray diffraction. The structure obtained (Figure 94) confirmed N-methylation was occurring and L-Me4ThzA had been successfully synthesised. Further titration experiments using NMR spectroscopy could be performed to determine the exact pKa at this position for comparison with thiamine.

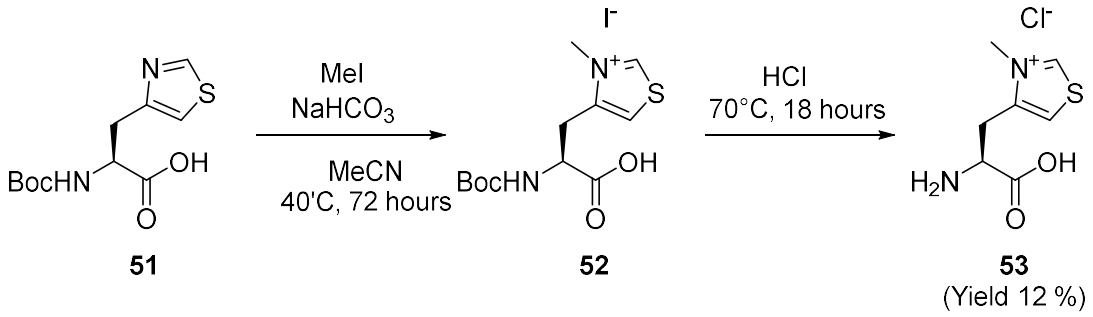


Figure 93: The 2-step synthesis of 3-methyl-4-L-alaninethiazolium chloride 53 gave a 12% overall yield.

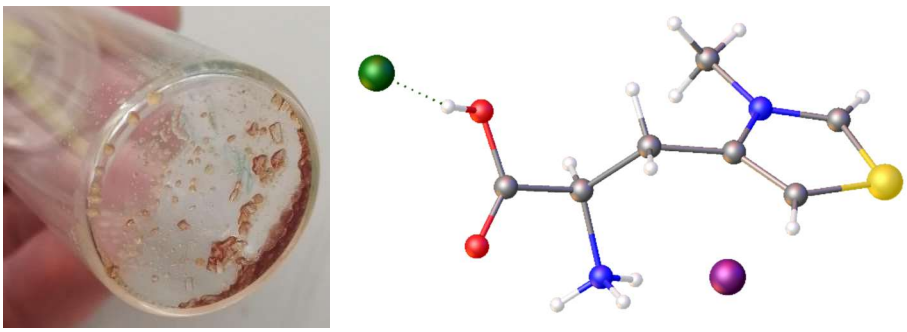


*Figure 94: a) Crystals of **53** grown by slow evaporation of methanol b) Crystal structure of **53** determined by Gary Nicol University of Edinburgh using X-ray diffraction.*

Synthesis of the racemic version of Me4ThzA was also undertaken due to the expense of the enantiopure starting materials. When performing GCE only the desired enantiomer of the UAA should be incorporated into the protein due to the ribosome and the aaRS not accepting the D-amino acid. Therefore, as long as the UAA is not toxic to the cells, addition of excess racemic UAA is not detrimental to incorporation. The cheap production of UAAs is essential, as they must be added in excess during GCE. A 4-step synthetic route with a 20 % yield was used for the synthesis of Me4ThzA (Figure 95). Products were analysed via NMR (Appendix 1.6.10) and again exchange at the C2 position was observed.

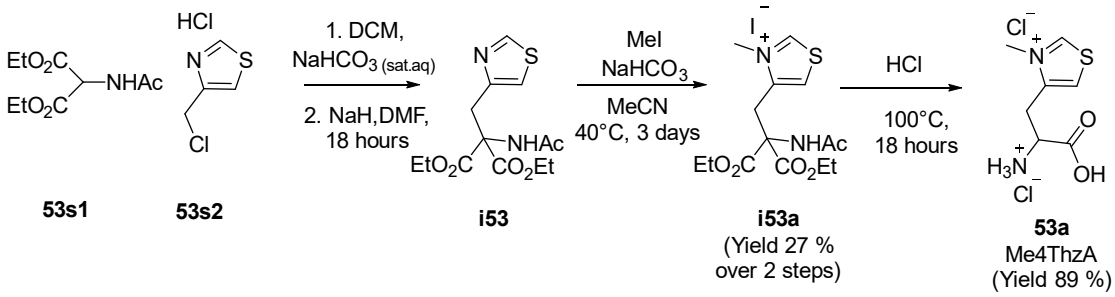


Figure 95: The 4-step synthesis of racemic Me4ThzA from commercially available starting materials diethylacetamidomalonate 53s1 and 4-Chloromethyl thiazole hydrochloride 53s2 with a 20% overall yield.

4.2.2 Synthesis of 3-methyl-5-alaninethiazolium chloride

During the course of our work it was shown by Hilvert *et al.* that 5ThzA could be successfully incorporated into a protein by GCE using pEVOL NMH[231]. Therefore, we set out to synthesise the methylated version of this UAA. Here we chose to synthesise just the racemic version of the UAA. Starting from cheap readily available starting materials, we synthesised Me5ThzA in four steps with an 8 % overall yield. Products were analysed via NMR (Appendix 1.6.9). Intermediate **i54** was also deprotected to give 5ThzA **54** in 39% yield, which could be used in our GCE experiments.

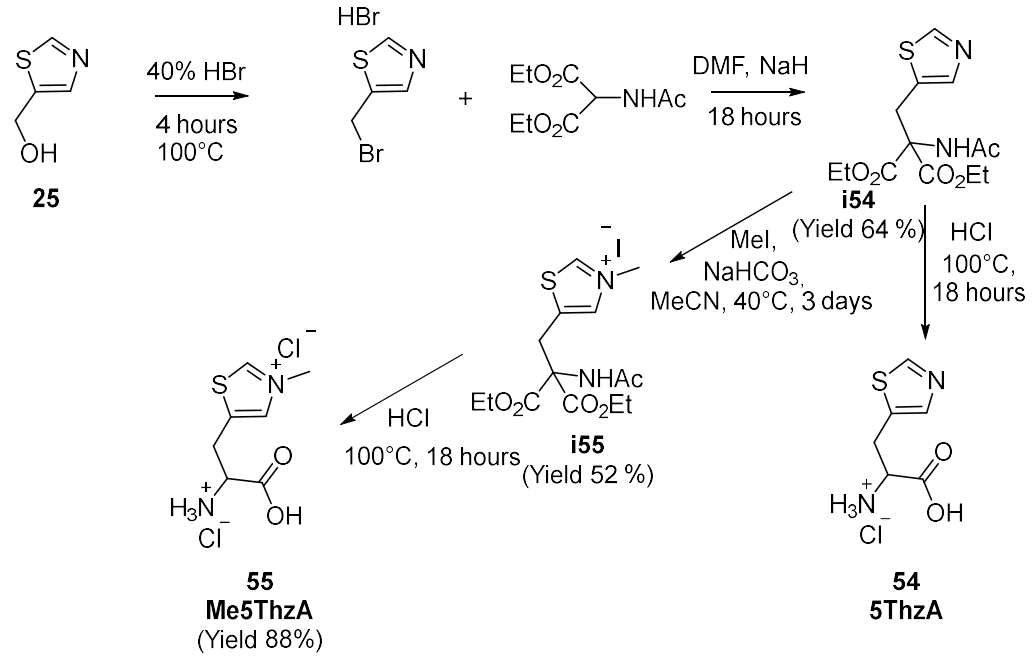


Figure 96: The 4-step synthesis of racemic 5ThzA and Me5ThzA from 5-hydroxymethylthiazole 25

4.2.3 Synthesis of Nε- ((3-methyl-4-thiazolyl) carbonyl) lysine chloride

In addition to the alanine based UAAs, pyrrolysine based UAAs were synthesised to make use of other pylRS based OTSs. We hypothesised that in some cases a longer distance between the protein backbone and the catalytic thiazole may prove beneficial for catalysis. The UAA Nε-((3-methyl-4-thiazolyl)carbonyl)lysine chloride (MeThiaK, **60**) was synthesised from commercially available Boc-lysine in a four-step process (Figure 97). First 4-thiazolecarboxylic was activated with thionyl chloride prior to the addition of Boc-lysine to form BocThiaK, **56** in 75% yield. From here several synthetic routes were followed to yield thiazole based UAAs and their corresponding methyl esters (ThiaK **57**, MeThiaK **60**, and MeThiaKOMe **60a**) for use in GCE experiments. This allowed us to test incorporation of the thiazole based UAA as well as the N-methylated counterpart to assess if the thiazole sidechain was compatible with the OTSs we screened. Initial attempts at methylation using MeI with DIPEA led to only trace amounts of N-methylation although formation of the methyl ester was observed by NMR analysis of the crude product. However, using a microwave reactor enabled N-methylation and yielded a sticky hygroscopic yellow product, **59**. Deprotection via two different methods yielded our N-methylthiaolium based UAA **60** and its corresponding methyl ester **60a** as sticky yellow products. The methyl ester versions of the UAAs were synthesised as it has been reported that the use of methyl ester protected UAAs can enhance protein yields[170].

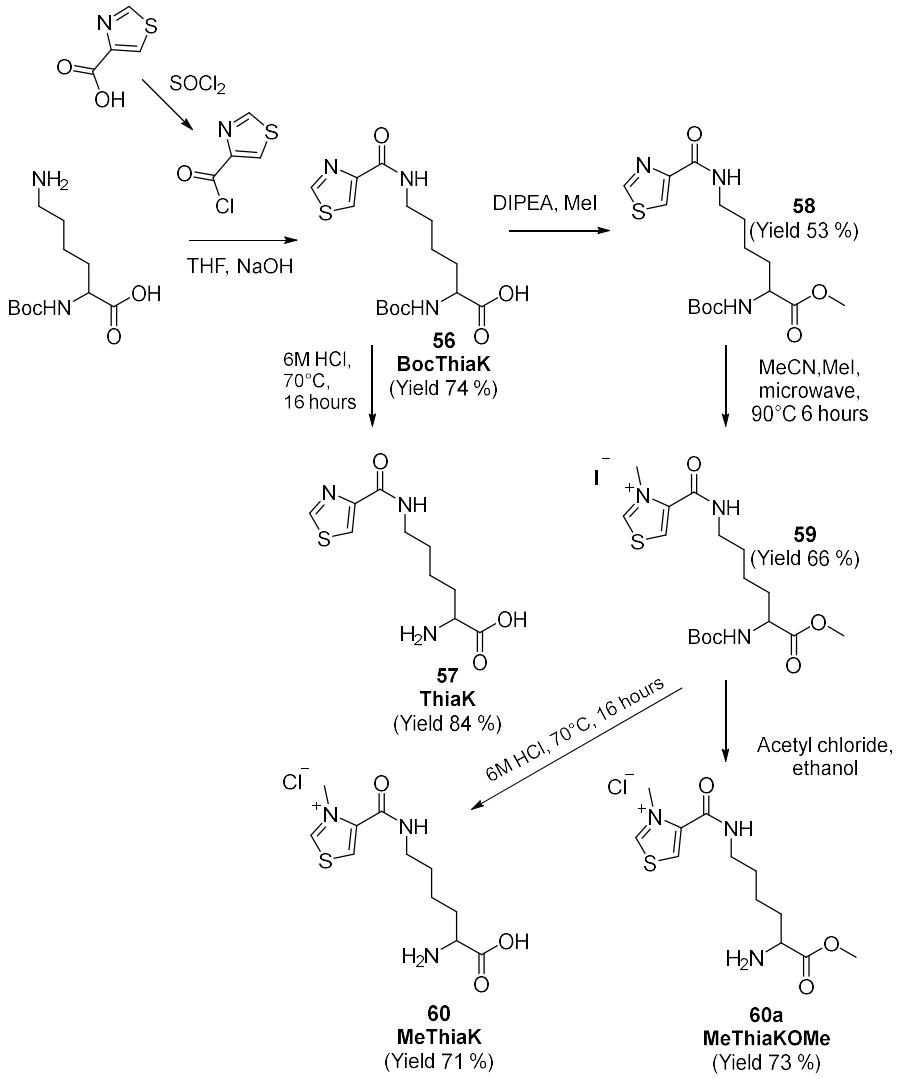


Figure 97: An overview of the synthesis of ThiaK based UAA. Starting from Boc-lysine and 4-thiazole carboxylic acid a divergent synthesis method was used to produce ThiaK, MeThiaK and its corresponding methyl ester MeThiKOMe

4.3 Activity screen of UAAs

Derivatives of Me4ThzA have been shown to be catalytically active for intramolecular Stetter reactions[113]. To verify the activity of our UAAs, both intramolecular Stetter and Benzoin condensation reactions were performed. As most wild type proteins are not particularly tolerant of organic solvents and function best in aqueous media, once the UAAs are incorporated into a protein, reactions will be carried out in aqueous solution. Therefore, we sought to determine if the UAAs were capable of catalysing reactions in aqueous reaction mixtures without being sequestered inside a protein. The thiazolium based UAAs must be deprotonated to their ylide form, therefore the choice of base used in the reaction is

important. In this section we do not aim to optimise the reactions of the free UAAs merely to prove they are capable of acting as NHCs for catalysis.

4.3.1 Benzoin condensation

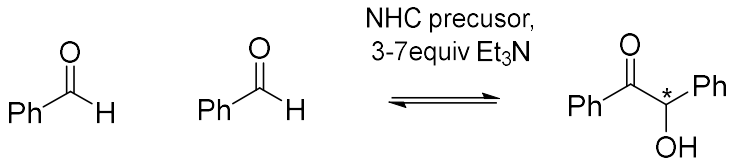


Figure 98: The benzoin condensation of benzaldehyde can be catalysed by thiamine HCl and Me4ThzA in the presence of 3-7 equivalents of the base triethylamine

The benzoin condensation can be catalysed by thiamine hydrochloride in combination with a base to aid deprotonation. We assessed the ability of both Thiamine and Me4ThzA **53a** to catalyse the benzoin condensation in methanol with varying equivalents of base (**Error! Reference source not found.**). Benzoin formation was observed with 5-7 equivalents of triethylamine to enable deprotonation (Figure 99). However, product was not seen when the reaction was run in aqueous buffered conditions at pH7-9 or in water with triethyl amine (data not shown). The quantity of base required to deprotonate the Me4ThzA was more than is required for thiamine, with the best results being observed with 7 equivalents of base.

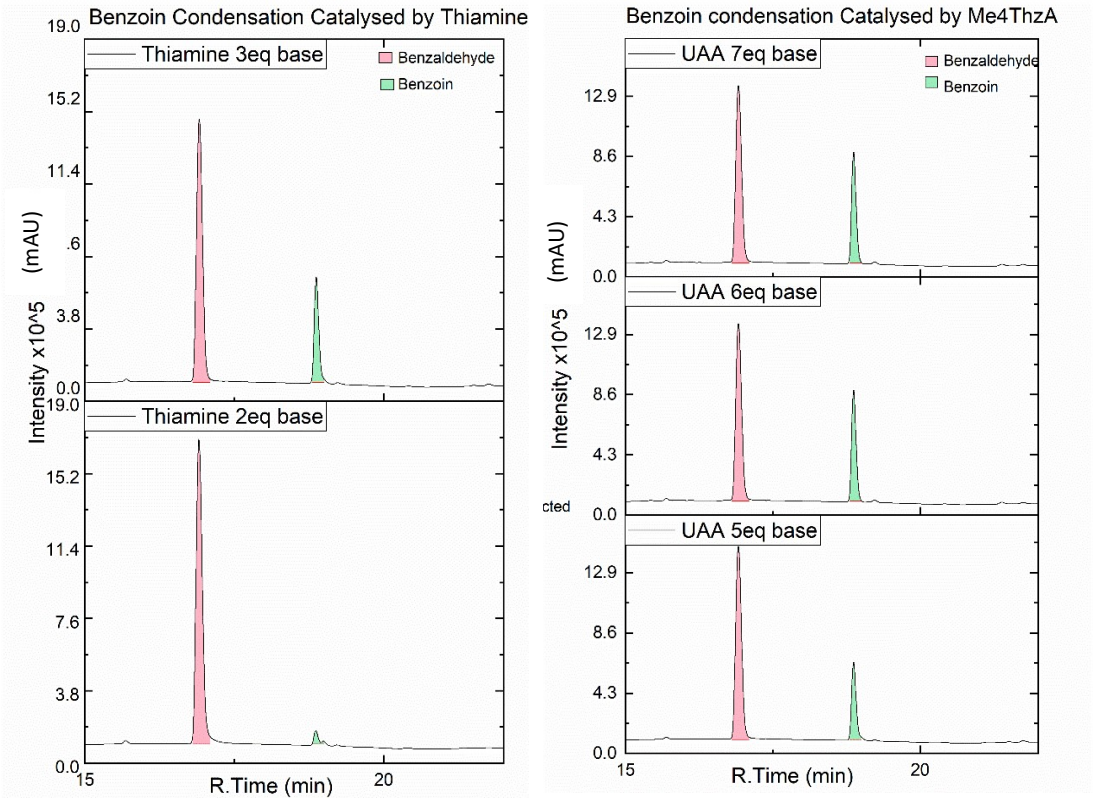


Figure 99: HPLC analysis of the benzoin condensation catalysed by Thiamine and Me4ThzA (53a). Reaction conditions:24 mM Benzaldehyde, 24 mM Thiamine HCl or Me4ThzA, 2-7 equiv Et3N, 24 hours 50 °C, 250 rpm.

4.3.2 Intramolecular Stetter reaction

The ability of the UAAs to catalyse the intramolecular Stetter reaction of **18** in methanol was assessed. It was observed that both the alanine based UAA (Me5ThzA, **55**) and the pyrrolysine based UAA (MeThiaKOMe, **60a**) catalysed the reaction in methanol when 6 equivalents of base were used. Aqueous buffered reactions did not show any product formation (data not shown). Interestingly MeThiaKOMe showed more product formation (0.13 mM $\pm$ 0.002) than Me5ThzA, which only formed traces of product (0.03 mM $\pm$ 0.01, Figure 100). In future work it would be interesting to investigate the effect of the ester on the pKa of the C2 proton by comparing these results with their counter parts MeThiaK **60** and Me5ThzAOMe which was not synthesised.

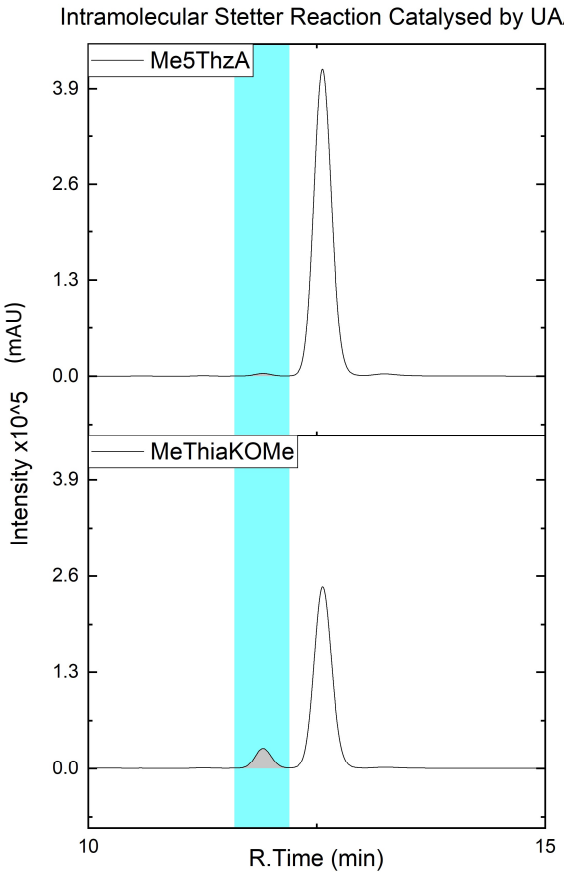


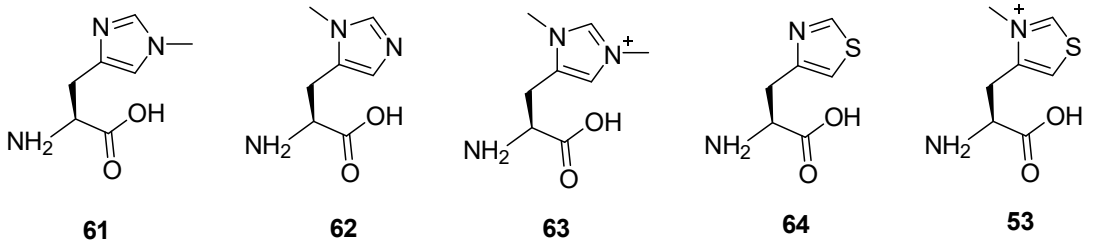
*Figure 100: Intramolecular Stetter reaction of **18** in methanol catalysed by Me5ThzA and MeThiaKOMe. Both reactions show the formation of product **19** (highlighted blue). Reaction conditions: 2.5 mM **18,** 2.5 mM UAA, 15 mM triethylamine, 20% DMSO in methanol, 30 °C, 18 hours, 250 rpm.*

These results satisfied us that our synthesised UAAs are capable of catalysing an intramolecular Stetter reaction, all be it with low yields under the tested conditions. Therefore, we moved on to trying to incorporate them into a protein using genetic code expansion.

4.4 Test of current OTS's for genetic code expansion

4.4.1 OTS's for incorporation of Histidine analogues

Three different plasmids containing orthogonal translation systems (OTS), pEVOL, pCNF or pEVOL_NMH were tested for the ability to incorporate five unnatural amino acids: 1-methylhistidine **61,** 3- methylhistidine **62,** dimethylhistidine **63,** L-4-thiazolylalanine **64,** and 3-methyl-L-4-thiazolylalanine **53,** into SCPM112TAG. Of these UAAs only **62** and **53** are NHC precursors, however, we were interested in assessing the promiscuity of the OTS with the idea to evolve them further for our substrates if required. Expressed protein was purified from 10 mL cultures and analysed via SDS-PAGE (Figure 101). The pEVOL based OTS did not incorporate any of the tested UAAs. Although pCNF showed a band on the SDS PAGE analysis of the correct size for SCP for all UAAs this is due to leaky expression. The band corresponding to SCP expression is of equal size for all conditions including the control. This agrees with other data which suggests pCNF incorporates phenylalanine <20% in response to the amber codon[165]. pEVOL NMH successfully incorporated **62,** the substrate it was evolved to accept[230]. Disappointingly, none of the aaRSs were compatible with **61, 63, 64** or **53**. However, recently it was reported that 5ThzA could be incorporated by the pEVOL NMH OTS[231].



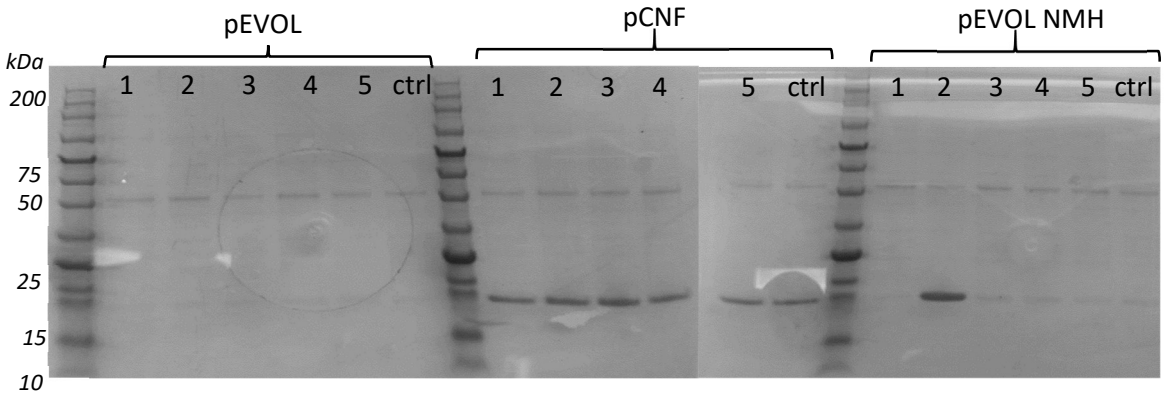


Figure 101: Structures of UAAs tested for incorporation. SDS PAGE analysis of UAA incorporation test using the orthogonal translation systems pEVOL, pCNF and pMeHis. Lanes labelled 1-5 correspond to the UAAs 61-64 and 53 respectively being added to each reaction. Lanes labelled ctrl indicate no UAA added to the reaction.

Therefore, we synthesised and tested the incorporation of 5ThzA and Me5ThzA with this OTS and SCPM112TAG using the same method. As reported 5ThzA was successfully incorporated into our protein as verified via SDS PAGE and LC-MS however to our disappointment Me5ThzA was not (Figure 102).

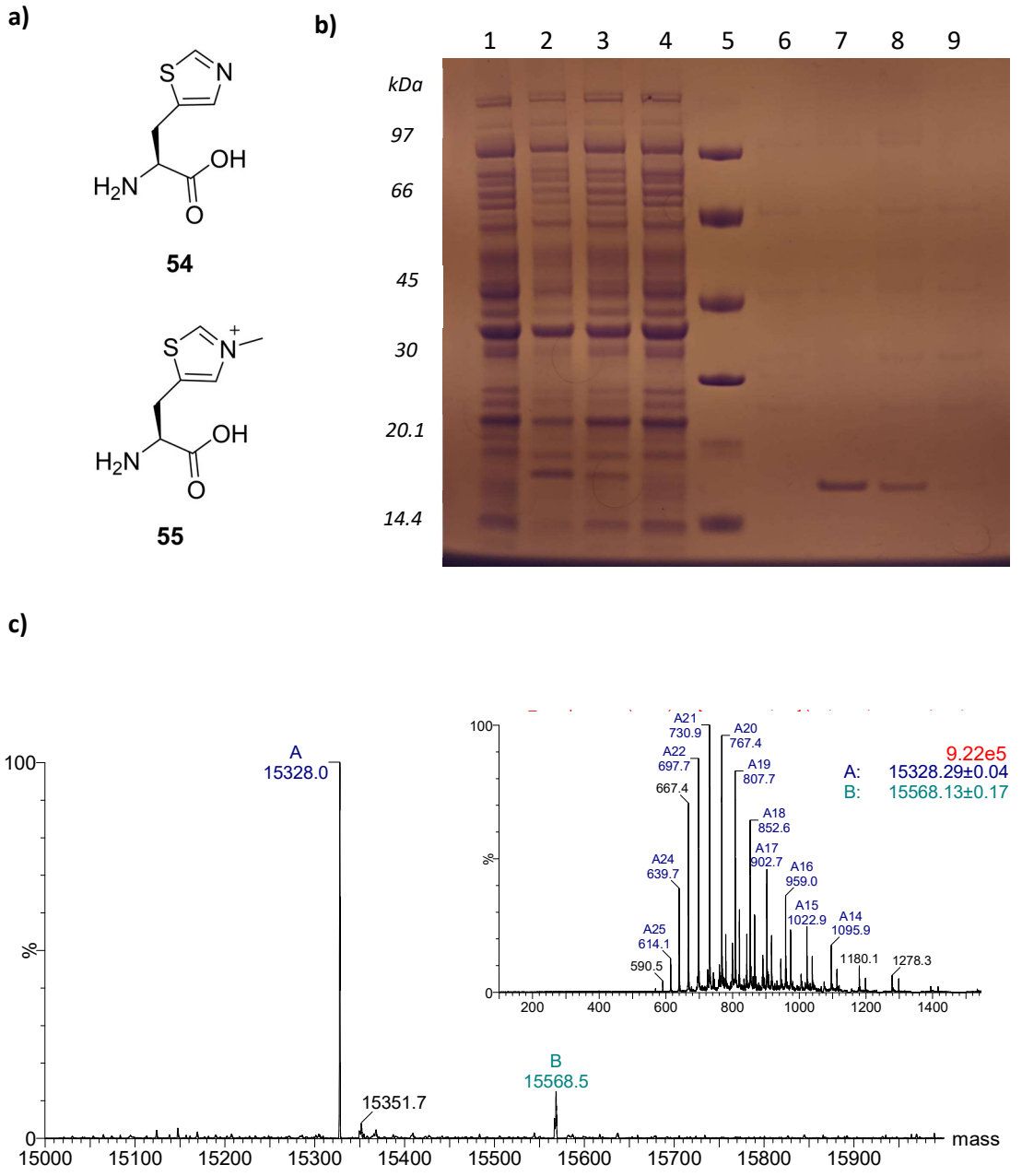


*Figure 102: **a)** Structures of 5ThzA and Me5ThzA, **b)** A gel showing the results of attempted incorporation of 5ThzA and Me5ThzA into SCPM112TAG using the OTS pEVOL NMH . Lanes show 1. Control no UAA lysate 2. 5 mM 3NMH lysate, 3. 3 mM 5ThzA lysate, 4. 10 mM Me5ThzA lysate, 5. LMWM, 6. Cntrl pure, 7. 5 mM 3NMH pure, 8. 3 mM 5ThzA pure, 9. 10 mM Me5ThzA pure. **C)** LCMS (ESI) analysis showing correct incorporation of 5ThzA into SCPM112TAG.Predicted mass 15327.8, Mass found 15328.0l*

4.4.2 pBK-ThzKRS for incorporation of MeThzK

The pBK-ThzKRS system has been used to incorporate thiazole like UAAs[234]. This system differs to the pEVOL based system in that the pylT is on a second plasmid with the reporter gene. This is useful for evolution and screening of aaRSs, but it is less convenient for using the OTS to efficiently incorporate an UAA into the protein of interest, as that plasmid for recombinant expression of the protein must then also encode pylT. Additionally, the system is not as efficient as pEVOL so lower levels of protein production are observed however using a reporter protein such as green fluorescent protein (GFP) incorporation can be easily monitored.

Plasmids pBK-ThzKRS and pylT-GFP (kindly provided by Jason Chin, University of Cambridge) were co-transformed into BL21 (DE3) cells with the appropriate antibiotics. These cultures were particularly slow growing so antibiotic concentrations were reduced to 80% to try to improve growth. Sufficient growth occurred to test incorporation of UAAs, BocKOMe, ThiaK and MeThiaKOMe. After 22-hour induction at 30°C, fluorescence was recorded again then cultures were centrifuged, and mini purification was performed. The bright green/yellow colour of the pelleted cells clearly indicated synthesis of full length GFP (Figure 103) and SDS

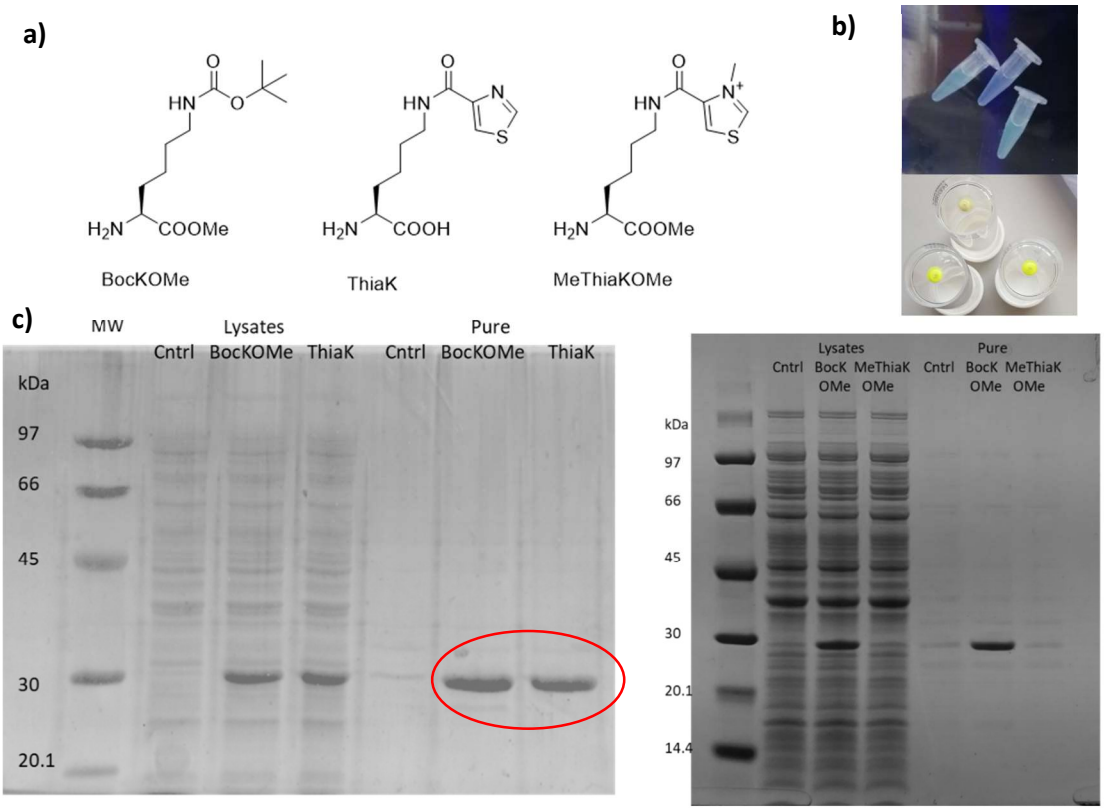


Figure 103: a) Structures of UAAs tested, b) Pictures of green/yellow-coloured cultures and pellets c) gels pBK-ThzKRS eGFP test of UAA incorporation showing bands corresponding to the incorporation of BocKOMe and ThiaK in both lysates and purified samples (circled red).

PAGE analysis of the purified protein confirmed these results with cells expressed with BocKOMe and ThiaK displaying bands of the expected molecular weight of eGFP (eGFPY151BocK, expected mass: 27941.50 Da, eGFPY151ThiaK, expected mass: 27951.30 Da). Again, to our disappointment the desired protein was not expressed in the presence of our thiazolium based UAA, MeThiaKOMe. The use of GFP containing an amber stop codon allows fluorescence to be used to screen for UAA incorporation. To further verify these results and ensure purification was not the lowering our limit of detection, we retested the incorporation of MeThiaKOMe. Fluorescence readings at 4 hours and 22 hours after induction indicated UAA incorporation of BocKOMe whilst the culture containing MeThiaKOMe did not show any increased fluorescence above background levels (Figure 104).

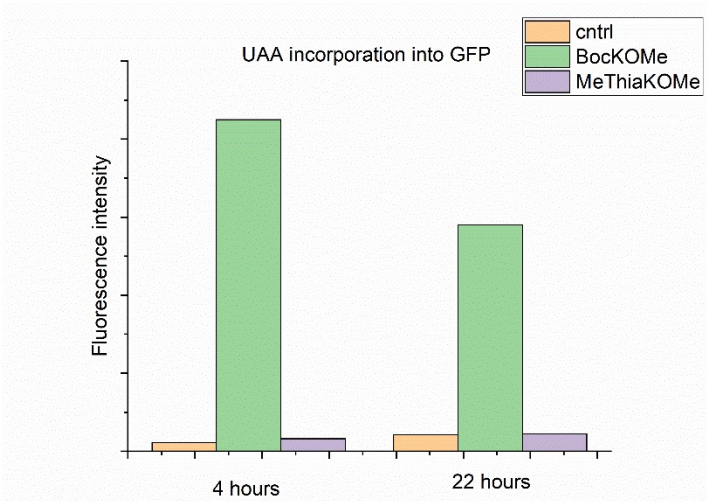


Figure 104: Fluorescence intensities of BL21 DE3, pBK-ThzKRS, pylTT-GFP expression tests after addition of UAAs and inductions for 4 hours or 22 hours at 30 °C. No UAA was added to the control sample.

To fully exclude the possibility of MeThiaK incorporation we analysed our samples by LCMS. The samples analysed by LCMS were dilute and appeared to contain quite a few small peptides but the correct masses for incorporated UAAs BocKOMe and ThiaK were observed however, the expected mass was not observed for MeThiaKOMe incorporation (Figure 105).

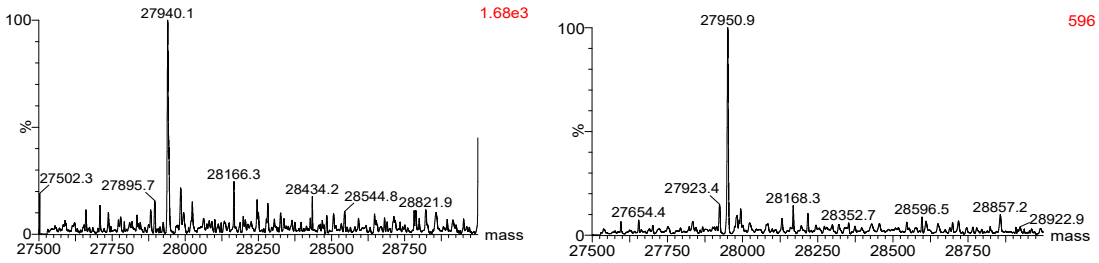


Figure 105: LC-MS of GFP with UAAs incorporated a) BocK, observed mass 27940.1 Da, b) ThiaK observed mass 27950.9 Da.

The other OTSs were also screened for incorporation of MeThiaK and ThiaK. Unsurprisingly no bands were visible on the gel for incorporation of MeThiaK or ThiaK using pEVOL, pEVOL NMH, pEVOL AZF, likely due to the pocket size of these proteins not being large enough to accommodate the structures (data not shown).

4.4.3 Test with the pTECHchpylRS with eGFPY151TAG reporter

To use the power of the eGFP reporter with the Me4ThzA UAA, mutagenesis was undertaken to create eGFP Y151TAG which was used in combination with pTECHchpylRS IPYE. In my hands pTECHchpylRS IPYE with eGFP Y151TAG did not express protein as well as pEVOL systems with low fluorescence counts being obtained. Fluorescence readings were used to verify incorporation of BocK whilst MeThiaAla did not show incorporation (Figure 106).

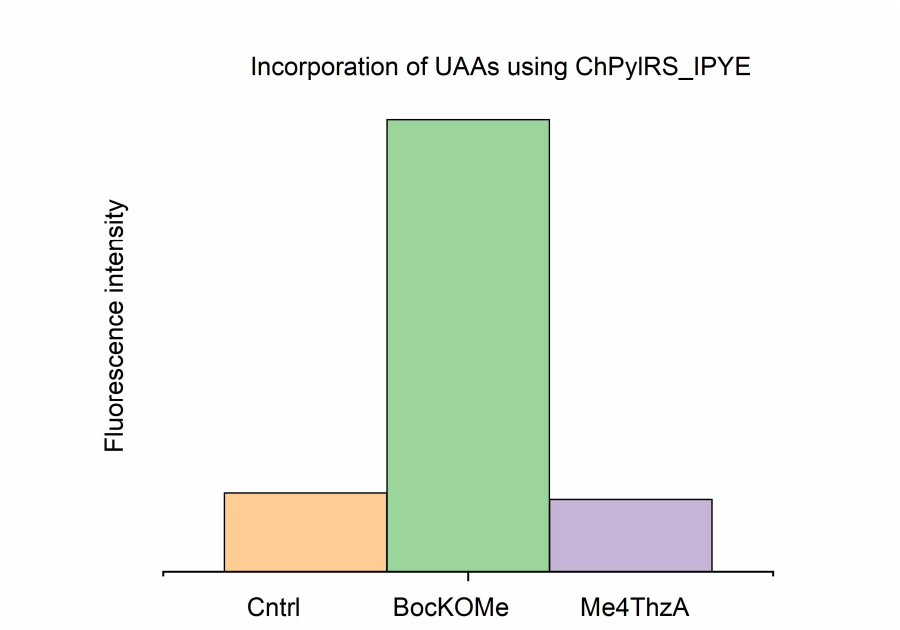


Figure 106 Incorporation of UAAs into eGFPY151TAG during expression tests using the pTECHchpylRS IPYE OTS measured by the fluorescence intensity after 22 hours of induction at 30 °C.

4.4.4 Structural analysis of aaRS

After verifying none of the OTSs available incorporated our thiazolium based UAAs (Table 18) we had to make a decision as to if we should carry out a directed evolution campaign. As this process can be time intensive, we wanted to make sure it was the aaRS which was preventing incorporation not other factors such as the UAA not entering the cell, being degraded, or stalling the ribosome. Therefore, structural analysis and docking studies using AutoDock Vina[140] were performed to assess if the UAAs fit into the active site of the aaRSs.

Table 18: Summary of results showing which UAAs were incorporated by each OTS.

	From TyrRS		From PylRS		
UAA	**pEVOL**	**pUltra_CNF**	**pEVOL NMH**	**pTECHchpylRS IPYE**	**pBK-ThzKRS + pylTGFP150TAG**
Me-4-ThiaK(OMe)	X	X		X	X
4ThiaK(OMe)	X	X		✓	✓
BocK(OMe)				✓	✓
1-NMH	X	X	X		
3-NMH	X	X	✓		
Di-NMH	X	X	X		
4ThzA	X	X	X		
Me4ThzA	X	X	X	X	
5ThzA			✓		
Me5ThzA			X		

Docking of the ThzA based UAAs into an AlphaFold model of N-methyl histidine synthetase (NMHRS) using AutoDock Vina was inconclusive leading to several poses where the UAAs docked in positions that would not lead to formation of the adenylate[138, 140]. Whist the affinities obtained for the most feasible poses for each UAA were similar (Figure 107). To try to get a better idea of the position of the side chain of each UUA, docking of the adenylate of each UAA was undertaken. Again, this showed no significant differences in docking affinities between methylated and no methylated UAAs (Figure 108). The orientation of the amine and carbonyl groups differed from that of the native substrates in pylRS[239](Figure 108a).The orientation of the UAA must be correct for substrate recognition and formation of the key UAA adenylate intermediate.

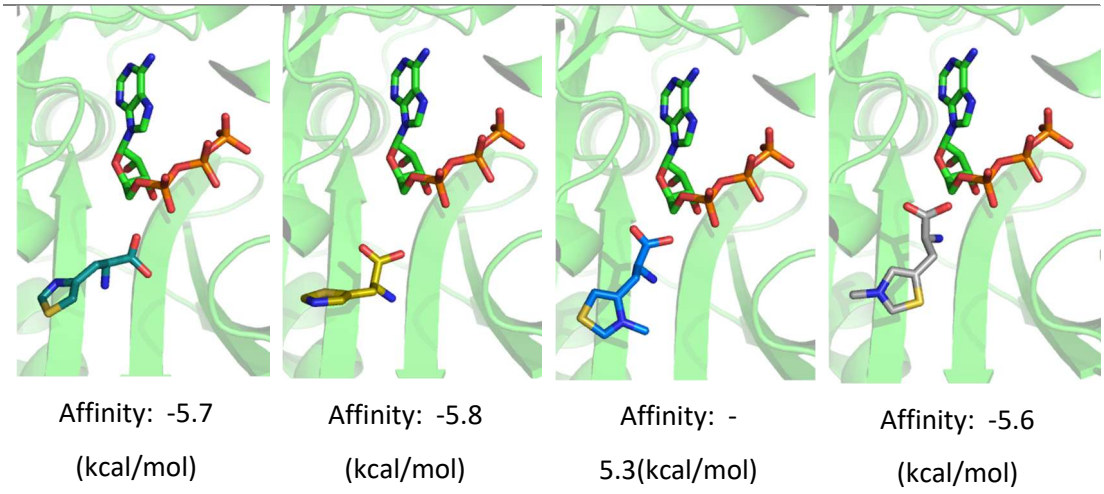


Figure 107: Feasible poses produced by the docking of UAAs into an AlphaFold model of NMHRS produced similar binding affinities. a) 4ThzA, b)5ThzA c) Me4ThzA, d) Me5ThzA.

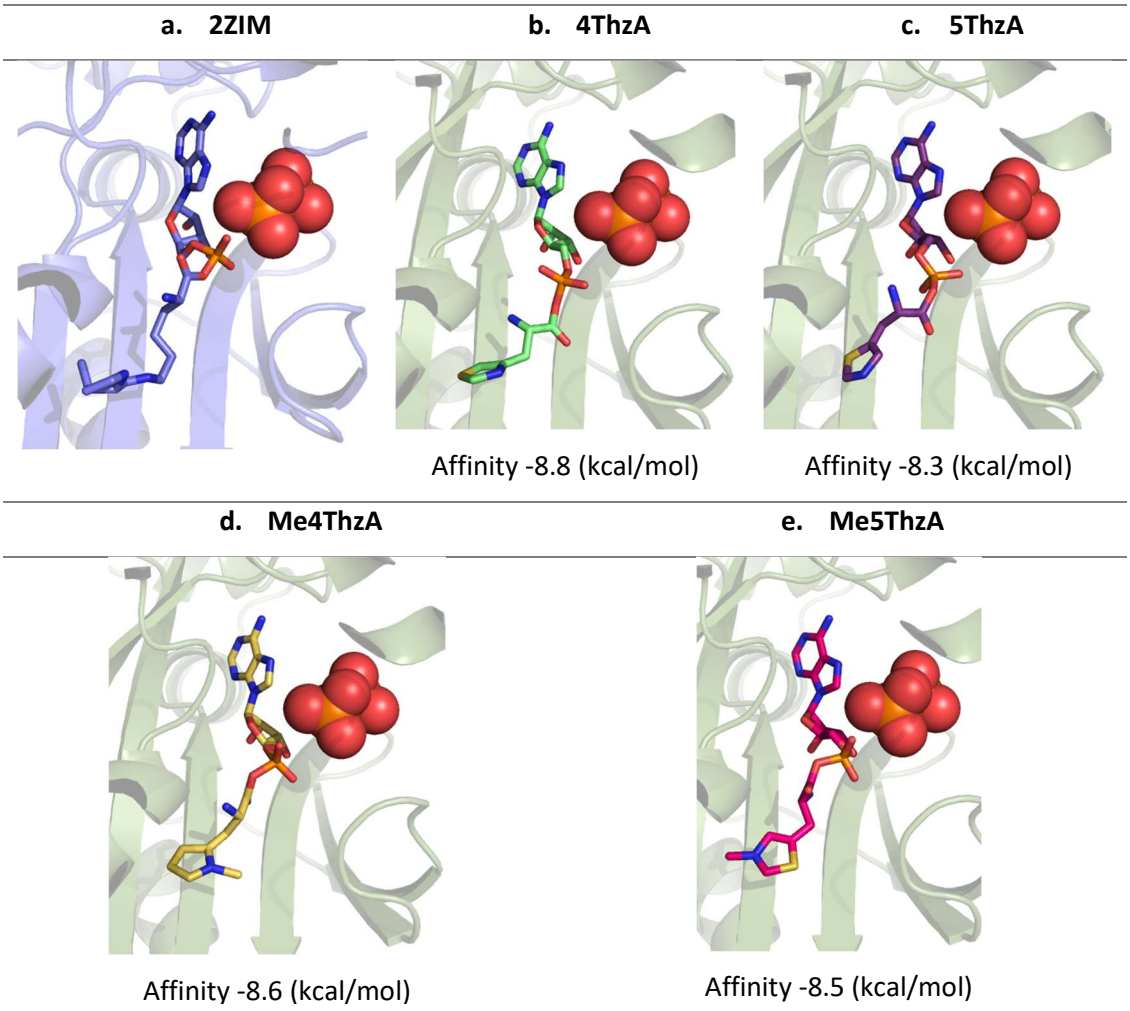


Figure 108: a) The pdb structure of the native pylRS (PDB: 2ZIM) highlighting the ligand and the pyrophosphate (red spheres). b-e) Result of the docking of the adenylate of each UAA into an AlphaFold model of NMHRS with a diphosphate using AutoDockVIna. b) 4ThzA adenylate, c)5ThzA adenylate, d) Me4ThzA adenylate, e) Me5ThzA adenylate. Similar binding affinities for each adenylate but difference can be seen between the position of the amine and the carbonyl in each structure and the ligand in the native crystal structure 2ZIM.

Side-chain movements associated with Pyl recognition and positioning during amino acid adenylation have previously been reported[239]. In particular N346 and Y384 (numbering used from *M. mazei* PylRS however these synthetases were created from the homologous M. *barkeri* PylRS) play key roles in substrate recognition. Upon formation of the intermediate, N346 moves to hydrogen bond with the secondary carbonyl and to the primary amino group via a water mediated contact[239]. Meanwhile Y384 makes hydrogen-bonding interactions with both the pyrrole nitrogen and the α-amino group of the adenylate[239]. Whilst neither of these residues have been mutated in either NMHRS or ThzKRS they will not be able to play the same role in substrate recognition with modified substrates. Interestingly, NMHRS appears not to tolerate changes of the substrate at the 1 position of the heterocycle as it does not accept 1-MeHis, Me5ThzA or 4ThzA. Whilst a methylation clearly adds steric bulk, differences between the thiazole and imidazole heterocyclic side chains could explain these observations for 4ThzA. Firstly the heteroatom influences the aromaticity of the heterocycle with thiazole containing more aromatic character than imidazole[240] however this does not seem to cause an issue for 5ThzA. Therefore, we need to consider the fact that the thiazole contains a lone pair perpendicular to the pi system in an sp^2 orbital (Figure 109). Interestingly, since this work

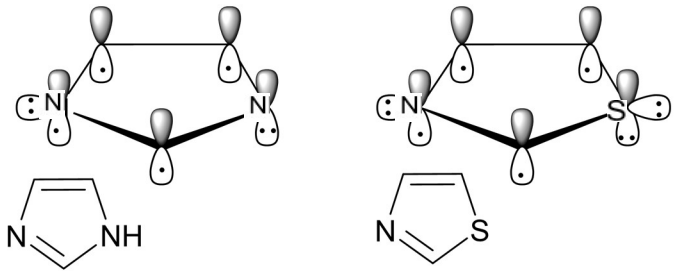


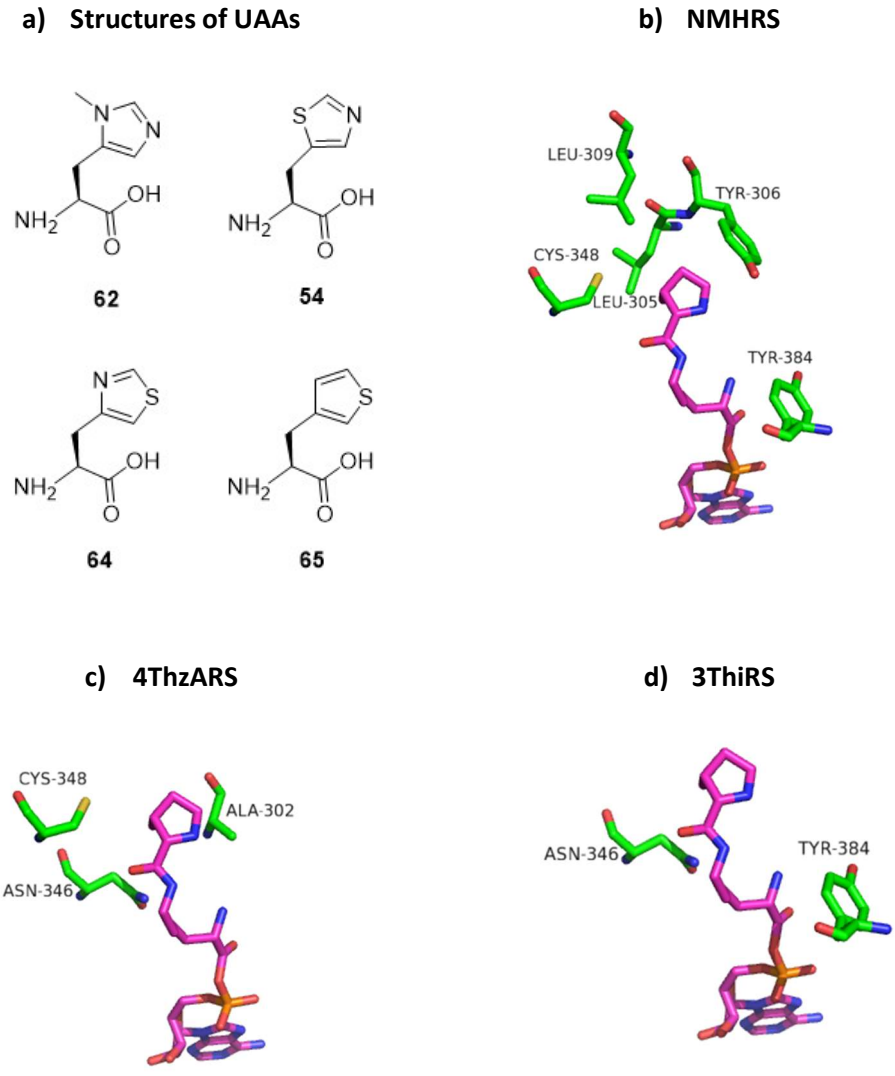
Figure 109: Orbitals of imidazole and thiazole showing the lone pair in the Sp^2 orbital of the thiazole.

was undertaken aaRSs which accept other N-heterocyclic UAAs, 4ThzA and 3ThiA have been created by the Hilvert group[231].

Construct	**Mutations in reported synthetases**								
NMH		L305I	Y306F	L309G		C348F			Y384F
4ThzA	A302Q				N346S	C348W			
3ThiA					N346Q				Y384F

Figure 110 shows the positions of these mutations around the adenylate intermediate of pyrrolysine from the crystal structure of 2ZIM. Interestingly these aaRSs both have N346 mutated. This mutation appears to be to key to allow S at the 1 position in the heterocycle. Remarkably the only other mutation in 3ThiARS, Y384F increases tRNA aminoacylation

independent of substrate further highlighting the role of N346 in altering the substrate scope of pylRS. Instead, the NHMRS is mutated at the nearby C346 to an aromatic residue, Phe which is useful for pi-pi interactions and makes active site smaller to better suit NMH compared to pyrrolysine. It would be interesting to assess the effects of the alternate mutations and the ability of these synthetases to accept our NHC precursor UAAs, Me5ThA and Me4ThzA, particularly as they show a slightly higher level of leaky expression[231]



Construct	Mutations in reported synthetases								
NMH		L305I	Y306F	L309G		C348F			Y384F
4ThzA	A302Q				N346S	C348W			
3ThiA					N346Q				Y384F

Figure 110: a) Structures of UAAs accepted by modified pylRSs, NMH (62) and 5ThzA (54) are accepted by NMH RS whilst 4ThzA is accepted by 4ThzARs and 3ThiA is accepted by 3ThiRS. b-d) The locations of residues which are mutated in pylRS (3vqx) to from each aaRSs overlayed on the adenylate of the native pylRS substrate pyrrolysine.

4.5 Conclusion and Future work

In conclusion we synthesised three thiazolium-based UAAs and their uncharged thiazole precursors. The catalytic activity of these thiazolium based UAA for the benzoin condensation of benzaldehyde and the model Intramolecular Stetter reaction used throughout this thesis was verified. The uncharged, non-catalytic UAAs were successfully incorporated into proteins using existing OTS's. This is the first time that ThiaK has been synthesised and in incorporated into a protein. However, we were unable to incorporate the charged thiazolium based UAAs. Preliminary docking studies suggest the methylated UAA adenylates can still fit in the binding site and revealed similar docking scores to their counterparts. Further modelling and molecular dynamic simulations should be performed may give a better insight into the side chains movements through the adenylating process as these have not been considered in our studies. Additionally, investigation of newly reported synthetases which incorporate 4ThzA and 3ThiA should be assessed for their ability to accept our thiazolium based UAAs. Our analysis suggests the N346 mutation which these synthetases contain may be beneficial for incorporation of the methylated substrates.

Alternative reasons for why the UAAs cannot be incorporated into the protein should also be considered. This requires the examination of how well the translation machinery in the cell accepts the UAAs of interest. Investigations with chemically synthesised UAA-tRNA pairs should be undertaken to help determine at what point our system is failing. If it is only the aaRS that is the issue these chemically synthesised pairs may be incorporated into proteins. However, there are many further stumbling blocks to be considered. One obvious feature of our UAAs is their positive charge, a feature which has been reported to cause ribosome stalling in several studies. Meanwhile this is not a universal phenomenon, and it is likely that this stalling can be attributed to consecutive lysine codons which cause ribosome sliding. If stalling is not the issue, there are still other reasons the translation machinery may not be compatible with our UAAs. In vitro GCE combined with engineering approaches to modify ribosomes, EF-Tu, and tRNAs could be used to help determine at which point our system is failing[241]. Another factor to consider is that the UAAs may not be entering the cell. Here the use of a cell free system could be of advantage. Evolution of the aaRS should not be undertaken until it can be verified that the UAA is getting into cell and that UAA-tRNA pair is accepted by the translational machinery.

In conclusion we speculate that the thiazolium based UAAs are not incorporated into proteins by GCE due to their positive charge causing incompatibility with the translational machinery.

However, a directed evolution campaign to develop an aaRS specifically to incorporate this UAA has not been undertaken. After investigating the above avenues, a directed evolution campaign of pylRS should be undertaken to create synthetases especially for these thiazolium based UAAs. If unsuccessful, other approaches including the use of PTMs to form NHC *in situ* should also be considered. In particular, TvaG which catalyses the bis-N-dimethylation of histidine as a PTM, should be investigated. However, this enzyme would require engineering to enable it to accept non-cyclic substrates.

5 Overall Conclusions & Outlook

The overall aim of this project was to create the first artificial Setterase by the insertion of an organocatalyst into a protein scaffold. This goal was achieved with several protein NHC precursor combinations being successfully synthesised and shown to exhibit Stetterase activity. The TTSCP L102C MBnThz scaffold-organocatalyst combination was found to be the most effective catalyst for an intramolecular Stetter reaction using a salicylicaldehyde derived substrate **18**. The choice of protein scaffold lead to a ~20-fold increase in catalytic activity. The next step for this protein scaffold-organocatalyst combination is to apply semi – rational design to create mutant libraries of the scaffold with the aim to improve its catalytic activity, stereoselectivity, reaction and substrate scope. Whilst the current e.e.s observed for the intramolecular Stetter reaction catalytsed by this protein scaffold-organocatalyst combination we are optimistic that a directed evolution campaign would lead to a more selective catalyst.

Our investigation into different protein scaffold to host our organocatalyst highlighted the difficulties associated with cross reactivities when repurposing natural enzymes. Whilst we found a thermophilic homologue of the commonly used SCP scaffold could be applied to our reactions the next steps would involve designing de-novo a scaffold specifically for this reaction. Advances in computational techniques for the incorporation of small molecules and non-proteinogenic components make this a realistic next step to pursue.

During this project several novel organic molecules were synthesised; both NHC precursors for the functionalisation of protein scaffold and NHC-based UAAs. Whilst we successfully applied the maleimide linked NHC precursors for the functionalisation of protein scaffold and catalysis, it would be beneficial to synthesise a more diverse range of NHC precursors to test as catalyst for a range of protein hosted reactions. The NHC-based UAAs displayed catalytic activity prior to incorporation into an enzyme. However the use of GCE with a range of existing OTS's was unable to incorporate them into a protein. Further work in this area to assess if methods such as cell free GCE could be applied to these UAA should be undertaken. It is key to determine the point at which these novel UAAs are failing to be incorporated during translation. If incompatibilities with the ribosome exist due to the positive charge on these UAAs screen existing OTS's or attempts to evolve new aaRSs will be futile. However alternative option exist such as post translational modification of thiazole based UAAs which we successfully demonstrated could be incorporated into proteins using existing OTS's.

To summaries we successfully created the first artificial Stetterase by functionalisation of a TTSCP scaffold with a novel maleimide linked thiazole. This artificial enzyme is a good starting point for directed evolution studies and paves the way for the investigation of a range of protein-NHC combinations to be developed for novel reactions.

6 Materials and Methods

6.1 Materials and equipment/reagents

All chemicals, media and reagents were procured from the following suppliers: Sigma-Aldrich, Thermo Fisher Scientific and Bio-Rad. Competent cells and microbiology components were purchased from New England Biolabs (NEB), Novagen, Agilent Technologies, and Merck Chemicals. Synthetic gene constructs were purchased from GenScript and primers were purchased from Sigma-Aldrich. Chromatography columns were purchased from GE Healthcare.

6.1.1 Competent *E. coli* Cell Lines

A variety of competent *E. coli* cell lines were used for plasmid amplification, storage, and expression of recombinant protein (Table 19).

Table 19 Competent E.coli cell lines

Cell Line	Application
BL21 (DE3) (Novagen)	Expression host
BL21 Gold (DE3) (Novagen)	High transformation efficiency, Expression host
DH10B (Thermo scientific)	High transformation efficiency, Expression host
C2987 (NEB)	High transformation efficiency, Plasmid storage
DH5α (Invitrogen)	Plasmid storage
Rosetta 2 (DE3) (Novagen)	Expression host for rare codons

6.1.2 Antibiotics

Antibiotic solutions were prepared in deionised water, filtered through a 0.22 µm filter and stored at -20 °C. The appropriate antibiotic was added to cool growth media after sterilisation for growth of cell lines harbouring plasmids.

Table 20 Antibiotics used and their respective stock and working concentrations

Antibiotic	Stock Concentration	Working Concentration
Ampicillin	100 mg mL^{-1}	100 µg mL^{-1}
Kanamycin	30 mg mL^{-1}	30 µg mL^{-1}
Gentamycin	20 mg mL^{-1}	20 µg mL^{-1}
Spectinomycin	100 mg mL^{-1}	100 µg mL^{-1}
Chloramphenicol*	34 mg mL^{-1}	34 µg mL^{-1}
Tetracyline	25 mg mL^{-1}	25 µg mL^{-1}

*Prepared in ethanol absolute

6.1.3 Plasmids

The plasmids used throughout this project for recombinant protein expression and genetic code expansion are listed in Table 21. Where required recombinant proteins contained hexa-histidine tags to enable purification. In many cases a Tev cleavage site follows this tag to enable cleavage of the His-Tag from the recombinant protein after cleavage. Table 21 also highlights the induction method for each plasmid, IPTG was used at concentrations between 0.1-1.0 mM and Arabinose was used at 0.04% for all plasmids other than pBK-ThzKRS for which it was used at 0.2%

Table 21 Plasmids used in this project and their respective resistances, tags and induction methods

Name	Resistance	Tag	Induction method
p28_hSCP_A100C **p28_hSCP_V83C** **p28_hSCP_Q111C** **p28_hSCP_A100C** **p28_hSCP_F94H** **p28_hSCP_V83C Stable**	Kanamycin	6X His Tev N-terminal	IPTG
p28_TTSCP_ W83C **p28_TTSCP_ L112C**	Kanamycin	6X His Tev N-terminal	IPTG
p28_CO_TTSCP L102C **p28_CO_TTSCP ΔDSB L102C**	Kanamycin	6X His Tev N-terminal	IPTG
pNHisTev_ThAOS **pNHisTev_ThAOS_V79S**	Kanamycin	6X His Tev N-terminal	IPTG
p28_Thiaminase 1	Kanamycin	6X His Tev N-terminal	IPTG
p28_pfThiE	Kanamycin	6X His Tev N-terminal	IPTG
p22b_pfBAL	Ampicillin	6X His Tev N-terminal	IPTG
pRK793 (MBP_TevProtease)	Ampicillin	N-terminal MBP Tev cleavage site which is self cleaved. 6X His N-terminal	IPTG
pEVOL NMH	Chloramphenicol	-	Arabinose
pEVOL	Chloramphenicol	-	Arabinose
pEVOL AZF	Chloramphenicol	-	Arabinose
pULTRA CNF	Spectinomycin	-	IPTG
pBK-ThzKRS	Ampicillin	-	Arabinose
pTECHchpylRS IPYE	Chloramphenicol	-	IPTG
SCPM112TAG	Kanamycin	6X His C-terminal	IPTG
pylT-GFP	Tetracycline	6X His (GFP)	Arabinose
p20b_eGFPY151TAG	Ampicillin	6X His	IPTG

6.1.4 Growth Media

Media components were dissolved in deionised water and autoclaved for 20 min at 120°C. Components that could not be autoclaved were sterile filtered through a 0.22 µm filter then added to the other components once cooled.

Table 22 Media composition

Media	Components Autoclaved	Final composition
LB	Tryptone (10 g L^{-1}), yeast extract (5 g L^{-1}), NaCl (10 g L^{-1})	-
PB	**Media Base**: Tryptone (20 g L^{-1}), yeast extract (10 g L^{-1}), NaCl (5 g L^{-1}), K_2HPO_4 (8.7 g L^{-1}), **D-Glucose 10X**: 50 g L^{-1}	Tryptone (20 g L^{-1}), yeast extract (10 g L^{-1}), NaCl (5 g L^{-1}), K_2HPO_4 (8.7 g L^{-1}), D-glucose (5 g L^{-1})
TB	Tryptone (12 g L^{-1}), yeast extract (24 g L^{-1}), glycerol (0.4% v/v), KH_2PO_4 (2.31 g L^{-1}), K_2HPO_4 (12.54 g L^{-1}).	
LB Agar	Tryptone (10 g L^{-1}), yeast extract (5 g L^{-1}), NaCl (10 g L^{-1}), agar (15 g L^{-1})	
Super Optimal Broth (SOC)	Tryptone (10 g L^{-1}), yeast extract (5 g L^{-1}), NaCl (10 g L^{-1}), KCl (2.5 mM), $MgCl_2$ (10 mM), glucose (2% w/v)	

6.1.5 Buffers

Buffer components were dissolved in dH_2O and their pH was adjusted to the required value with hydrochloric acid (HCl) or sodium hydroxide (NaOH). Buffers were filtered through a 0.22 µM filter prior to use. Phosphate buffer saline (PBS, pH = 7.4) was prepared by dissolving 5 tablets (Sigma-Aldrich) in 1 L dH_2O to give a solution consisting of KH_2PO_4 (0.24 g/L), Na_2HPO_4 (1.44 g/L), KCL (0.2 g/L) and NaCl (8.0 g/L). Stock solutions of benzamidine (0.120 g in 10 mL) were prepared in ethanol and stock solution of PMSF (0.174 g in 10 mL) were prepared in methanol before addition to the buffer.

Table 23 Buffers used in this project and their compositions.

Buffer	Name	Components
1A	Lysis Buffer THAOS/pfThiE/Thase1	HEPES (20 mM), NaCl (150 mM), glycerol (5% v/v), pH 8.0
1B	Wash Buffer THAOS/pfThiE/Thase1	HEPES (20 mM), NaCl (150 mM), 20 mM Imidazole, glycerol (5% v/v), pH 8.0
1C	Elution Buffer THAOS/pfThiE/Thase1	HEPES (20 mM), NaCl (150 mM), glycerol (5% v/v), imidazole (350 mM) pH 8.0
1D	Gel Filtration Buffer THAOS/pfThiE/Thase1	HEPES (20 mM), NaCl (150 mM), glycerol (5% v/v), pH 8.0

2A	Lysis Buffer SCP	30 mL, 50 mM Tris-HCl, 20 mM imidazole, 150 mM NaCl, 0.5 mM benzamidine, pH 8 Lysozyme 1 mg/mL
2B	Wash Buffer SCP	30 mM Tris-HCl, 20 mM Imidazole, 150 mM NaCl, pH 8
2C	Elution Buffer SCP	30 mM Tris-HCl, 330 mM Imidazole, 150 mM NaCl, pH 8
2D	Dialysis Buffer SCP	30 mM Tris-HCl, 10 mM Imidazole, 150 mM NaCl, pH 8
3A	Lysis Buffer PfBAL	50 mM MOPs pH 7.0
3B	Wash Buffer PfBAL	50 mM Kphos, 300 mM NaCl, 5 mM imidazole, pH 7.4
3C	Elution Buffer PfBAL	50 mM Kphos, 300 mM NaCl, 200 mM imidazole, pH 7.4
3D	Dialysis Buffer PfBAL	50 mM MOPs pH 7.0
4A	Lysis Buffer Tev	PBS, NaCl (300 mM), imidazole (10 mM), PMSF (1 mM), benzamidine (1 mM)
4B	Elution buffer Tev	PBS, NaCl (300 mM), imidazole (500 mM), PMSF (1 mM), benzamidine (1 mM)
4C	Tev self cleavage buffer	50mM TrisHCl pH 8.0, 300mM NaCl, 1mM PMSF
4D	Tev storage buffer	50mM TrisHCl pH 8.0, 300mM NaCl, 10% glycerol
5A	Reaction Buffer	50 mM HEPES pH 8.0

6.1.6 PCR primers

Lyophilised DNA primers were ordered from Sigma-Aldrich. Primers were resuspended in dH_2O and diluted to 10 µM prior to use.

Table 24 Primers used for site directed mutagenesis.

Target	**Primer Sequences 5'-> 3'**
ThAOS K234C	For: CCCTGAGCTGCGCGTGGGCGGGTATCGG Rev: GCCCACGCGCAGCTCAGGGTCGCCACTTG
pfThiE D69A	For: CTTCGTCGATGCTAGGGTAGATGTAGCTTTAGC Rev: CTACCCTAGCATCGACGAAGAAAAGGGC
pfThiE D69C	For: CTTTTCTTCGTCGATTGTAGGGTAGATGTAGCTTTAGCTG Rev: CTACATCT ACCCTACAATCGACGAAGAAAAGGGCG
pfThiE S106A	For: AGGAGCAGCTGTATACTCCCTTGAAGAG Rev: GAGTATACAGCTGCTCCTATAATTAGATTGGGTGC
pfThiE S106C	For: AGGAGCATGTGTATACTCCCTTGAAGAGG Rev: GGAGTATACACATGCTCCTATAATTAGATTGGGTGC
pfThiE T132A	For: CGTCTTTCCAGCTAAGACAAAGGAAGATGCGAG Rev: GTCTTAGCTGGAAAG ACGGAGCCAGCTCC
pfThiE T132C	For: CCGTCTTTCCATGTAAGACAAAGGAAGAT GCGAGAG Rev: GTCTTACATGGAAAG ACGGAGCCAGCTCCTAAG
eGFPY151TAG	For: CAACGTCTAGATCATGGCCGACAAGCAGAAGAACGGC Rev: GCCATGATCTAGACGTTGTGGCTGTTGTAGTTGTACTCC

6.2 Methods

6.2.1 General Experimental Conditions

Centrifugation: during protein purification was performed using a Heraeus multifuge X3R (thermoscientific). Other centrifugation as performed using the benchtop microstar 17 (VWR) at room temperature.

Protein storage: Proteins were flash frozen in liquid nitrogen after purification and stored at –80 °C until then thawed on ice for 30 min before use.

NMR: ^{1}H, ^{2}H and ^{13}C NMR spectra were obtained in deuterated solvent using the Bruker Ascend 500 MHz, or Bruker Ascend 600 MHz spectrometer. Data for ^{1}H NMR are reported in the conventional form: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad), coupling constant (Hz), and integration Spectra were referenced to residual solvent signals. ($CDCl_3$ 7.26 ppm, CH_3OD 3.31ppm, $(CD_3)_2SO$ 2.50 ppm) and the coupling constants are reported to the nearest 0.1 Hz. 13C NMR spectra were obtained at 126 Hz or 151 Hz using the spectrometers mentioned above and referenced from solvent peaks ($CDCl_3$ 77.16 ppm, CH_3OD 49.00 ppm, $(CD_3)_2SO$ 39.52 ppm). 2D NMR spectra including COSY, HSQC and HMBC were used for proton signal assignments.

Mass Spectrometry (MS) Mass spectrometry analysis of small molecules was completed using a Bruker microTOF instrument with electrospray ionisation (ESI+). Protein LC ESI-MS was completed using a Synapt G2-Si Q-TOF (Waters) instrument with electrospray ionisation (ESI+).

HPLC: Both chiral and C18 reverse phase HPLC methods were carried out using a Shimadzu instruments HPLC fitted with an autosampler (SIL-20A HT), pump (LC-20AT), a UV/visible detector (SPD-20A) and a column oven (CTO-20A). For reverse phase analysis a Luna 5u C18 RP-Phenomenex column (100 Å, 250 x 4.6 mm) was used and for chiral analysis a CHIRALPAK® IB N-5 analytical column (5 µm, 250 × 4.6 mm) column was used.

6.2.2 Site directed mutagenesis

6.2.2.1 Extension overlap PCR

Table 25 Typical thermocyler program for overlap extension PCR

Step	Temp	Time
Denaturation	98°C	30 sec
30 cycles of:		
Denaturation	98°C	10 sec
Annealing	70°C	10 sec
Extension	72°C	195 sec (30sec/kb)
Final extension	72°C	170 sec

Table 26 Reaction components for typical PCR reaction

Component	Volume (µL)	Final Conc.
Nuclease-free water	to 50 µL	-
5X Phusion HF Buffer	10	1X
10 mM dNTPs	1	200 µM
10 µM Forward Primer	2.5	0.5 µM
10 µM Reverse Primer	2.5	0.5 µM
Template DNA	1	< 250 ng
Phusion DNA Polymerase	1	2.0 units

Lyophilised DNA primers (Sigma-Aldrich) were diluted to create 10 µM stocks using sterile filtered deionised H_20. Reactions were set up using with components shown in Table 26. NEB T_m calculator was used to determine the optimal annealing temperature of the primers. A typical reaction cycle (Table 25) was performed on a T100 Thermal Cycler (BioRad).

Agrose gel (1% w/v) was prepared by mixing agarose (0.5 g) with TAE buffer (50 mL) and microwaving until dissolved. After the solution had cooled to approximately 50°C GelRed™ (3 µL) was added. Gel was poured into a cast and allowed set for at least 30 min at room temp. Gel was immersed in TAE buffer and samples (10 µL sample mixed with 2 µL 6 X loading dye (NEB)) were loaded along with 100 bp or 1 kb DNA ladder (5 µL, NEB). The gel was run at a constant voltage (100 V) and the bands were visualised with UV light to determine presence of the PCR product.

DpnI digest was performed by incubating 20 µL of PCR product with 0.5 µL of DpnI at 37°C for 2 hours. 2 µL of digested PCR product was transformed into C2987 cells and plated on LB agar plates with appropriate antibiotic. Single colonies were picked and grown up in 10 mL of media for mini prep and sent for sequence analysis (IDTA) to determine correct product formation.

6.2.3 Protein Expression

6.2.3.1 Transformation of competent cells

6.2.3.1.1 General procedure

Competent cells (50 µL) were thawed on ice after removal from storage at -80°C. Once defrosted, plasmid DNA (2-5 µL) was added to the cells and placed on ice for 30 min. Heat-shock was performed for 35 sec at 42°C before cooling on ice for 2 min. SOC media (200 µL) was added and mixture was incubated at 37°C for 1 hour with shaking (250 rpm) at. Cells were plated on LB agar plates containing the required antibiotic and incubated overnight at 37°C.

6.2.3.1.2 Co-transformation

For co-transformation of two plasmids, which was required for genetic code expansion, the general protocol was followed with the following modifications: 1 µL of each plasmid was added prior to incubation on ice and cells were plated on agar plates containing both the required antibiotics.

6.2.3.2 DNA isolation - Mini Prep

GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific) was used to isolate plasmid DNA from an overnight liquid culture (10 mL) inoculated with the desired colonies. The manufacturers protocol was followed with the exception that sterile filtered dH_2O was used to elute the plasmid DNA from the GeneJET spin column.

6.2.3.3 Test Expression

Small scale test expressions were undertaken to determine optimal conditions for protein expression. A starter culture of LB (10 mL) containing the appropriate antibiotic was inoculated with a single colony of *E.coli* cells harbouring the desired expression plasmid and grown overnight 37°C 200 rpm. The starter culture was diluted 1:100 in several sterile tubes and grown to OD_{600}:~0.6 before induction with IPTG (varying conc.) and incubated at 16, 20, 30 or 37 °C for 4- 24 hours. Either a sample (20 µL) was analysed directly via SDS page or cells were harvested by centrifugation at 3500 rpm for 10 min and cell lysis was performed to enable analyse soluble and insoluble protein expression. For lysis, cultures were centrifuged (10 min, 13000 rpm) and the supernatant was discarded. Cell pellets from 10 mL test expressions were resuspended in BugBuster™ Protein extraction reagent (500 µL) and incubated at room temperature for 30 mins . Lysate (50 µL) was centrifuged (13000 rpm 10 min) to give soluble and insoluble fractions. Pellet was resuspended in 1XSB (50 µL) sup was suspended in 2XSB (50 µL). 15 µL of each sample was analysed via SDS PAGE.

6.2.3.4 UAA incorporation screening

6.2.3.4.1 SCPM112TAG

BL21 (DE3) cells were co-transformed with the SCPM112TAG plasmid and a plasmid for expression of an orthogonal translation system (pEVOL, pEVOL NMH, pEVOL AZF or pULTRA CNF). To test for incorporation of an UAA, starter cultures containing LB (10 mL) and the appropriate antibiotics for each plasmid were prepared in 50 mL flacon tubes an incubated at 37°C for 18 hours, 180 rpm. Starter culture (2 mL) was used to inoculate LB (100 mL) containing the appropriate antibiotic and grown at 37°C until they reached OD_{600}:0.6. The culture was split into to 10 mL aliquots and the desired UAA (1 mM) was added to each culture at OD_{600}:0.6. Expression was induced by adding IPTG (0.5 mM) and the appropriate inducer for the orthogonal translation system (see Table 4). Cultures were incubated overnight at 30°C, 180 rpm before being subjected to mini purification. During initial studies with histidine analogues these were kindly prepared by Richard Brewster, University of Edinburgh, subsequently UAAs were synthesised according to the methods provided in section 5.3.

6.2.3.4.2 PylT-GFP with pBK-ThzKRS

Plasmids pBK-ThzKRS and pylT-GFP (kindly provided by Jason Chin, University of Cambridge) were co-transformed into BL21 DE3 cells with the appropriate antibiotics. To test for incorporation of an UAA an LB culture (10 mL) containing ampicillin (100 µg mL^{-1}) and tetracycline (25 µg mL^{-1}) was inoculated from a single colony. Due to the slow growth of the bacteria under these conditions, culture was incubated for 48 hours at 37 °C 180 rpm. This culture was used to inoculate new cultures (1 mL into 9 mL LB) with 80% antibiotic concentrations used. Cultures were grown for 24 hours OD= 0.62. UAA (1 mM) was added and incubated 30 min before expression was induced with 0.2% arabinose and IPTG (0.1 mM) and incubated at 37 °C, 180 rpm. Fluorescence was checked after 4 hours and UAA incorporation could be seen. Cultures were centrifuged after 24-hour induction at 37°C, fluorescence readings were taken and a mini purification was performed.

6.2.3.4.3 eGFP151TAG with chPylRSIPYE

Plasmids pTECHchPylRSIPYE[232] (purchased from Addgene) and p20b_eGFPY151TAG were co-transformed into BL21 DE3 cells with the appropriate antibiotics (Table 20). To test for UAA incorporation an LB culture (10 mL) containing the appropriate antibiotics was inoculated from a single colony. The culture was incubated for 18 hours at 37 °C, 180 rpm then used to new cultures (100 µL into 10 mL LB). Cultures were grown for 2-4 hours until OD_{600} = 0.5 –

0.6. UAA (1mM) was added and incubated 30 min before expression was induced with IPTG (0.5 mM) and incubated at 30°C, 18 hours, 180 rpm. Cultures were centrifuged after 18-hour induction at 37°C, fluorescence readings were taken, and a mini purification was performed.

6.2.3.5 Fluorescence measurements

A sample (200 µL) was taken from the culture to be analysed and centrifuged (10 min, 13000rpm). Supernatant was discarded and cells were resuspended in PBS (200 µL) and analysed for fluorescence using a BioTek Synergy HT plate reader with the following settings, excitation 485 nm (bandwidth 20 nm) emission 530 nm (bandwidth 25 nm)

6.2.3.6 Mini purification

Cultures were centrifuged for 12 min at 4000 rpm. Cell pellets were resuspended in PBS (1 mL) and transferred to a 1.5 mL Eppendorf tube. Cells were centrifuged for 5 mins, 13000 rpm. Supernatant was discarded and BugBuster™ Protein extraction reagent (0.5 mL) was used to resuspend pellet. Solution was mixed for 30 mins on a rotary plate before centrifuging for 5 mins 13000 rpm. Supernatant (500 µL) was added to a tube containing Ni resin (30 µL) equilibrated with buffer 2B and mixed for 15 mins on a rotating plate. Tubes were centrifuged 1 min, 1000 rpm then supernatant was removed. Buffer 2B (1 mL) was added and samples were centrifuged again 1 min, 1000 rpm. Supernatant was removed and discarded. Elution buffer 2C (200 µL) was added and samples were mixed for 15 mins then centrifuged 1 min, 1000 rpm. Supernatant containing purified HisTagged protein was transferred to a fresh tube. Sample (20 µL) was mixed with 4XSB (6 µL) boiled, centrifuged, and analysed via SDS PAGE.

6.2.3.7 Large Scale Growth and expression

A single colony was used to inoculate media (100 mL) containing the appropriate antibiotic and gown at 37°C, 200 rpm for 18 hours. This pre-culture (20 mL) was used to inoculate LB media (1 L) supplemented with the appropriate and grown at 37°C, 200 rpm to an OD_{600nm} ~ 0.6. Temperature was lowered to the required induction temperature and protein expression was induced by the addition of a specific concentration of the required inducer (Table 27). Cultures were incubated at 200 rpm under conditions described in Table 27. Cells were then harvested by centrifugation (8,000 x g 10 min, 4°C). Pelleted cells were resuspended and washed with PBS buffer (40 mL per 1 L culture). Cells were pelleted again (8,000 g, 30 min, 4°C), supernatant discarded, and the pelleted cells were frozen and stored at -20°C until required.

Table 27 Expression conditions used of each recombinant protein. All mutants of the same base protein were expressed under the same conditions

	Expression host	**Media**	**Induction Temperature (°C)**	**Inducer**	**Induction time**
ThAOS	BL21 (DE3)	LB	16	0.1 mM IPTG	18 hours
hSCPs	Rosetta 2 (DE3)	PB	16	0.2 mM IPTG	18 hours
TTSCP	BL21 (DE3)	LB	30	0.5 mM IPTG	18 hours
pfThiE	BL21 (DE3)	LB	30	0.1 mM IPTG	18 hours
Thiminase 1	BL21 (DE3)	LB	30	0.1 mM IPTG	18 hours
pfBAL	BL21 (DE3)	TB	20	1 mM IPTG	18 hours
Tev	Rosetta 2 (DE3)	LB	20	0.4 mM IPTG	18 hours

6.2.4 Protein Purification

All protein purification steps were completed on AKTA Go or AKTA Start (Cytiva Lifesciences) using a F9-R fraction collector.

6.2.4.1 Cell Lysis

6.2.4.1.1 General procedure

Cell pellets were defrosted and resuspended in the required of lysis buffer (Table 23, 8 mL per g). Cell lysates were sonicated (10 min, 90 % power with 30 s pulses). Lysate was clarified by centrifugation (12 000 xg 45 min, 4°C) and supernatant was filtered (0.45 µm filter Millipore, Merck).

6.2.4.1.2 SCP constructs

Cell pellets were defrosted and resuspended in lysis buffer **3A** Lysozyme (20 mg), DNase I (1 mg), and $MgCl_2$ (1 mL, 1 M) were added to the lysate and incubated at 4°C for 1 h. Cell lysates were sonicated (2 min, 90 % power with 10 s pulses) then clarified by centrifugation (12 000 xg , 45 min, 4°C) and supernatant was filtered (0.45 µm filter Millipore, Merck),.

6.2.4.2 IMAC purification

A HisTrap™ FF column (1 mL or 5 mL, GE Healthcare Life Science) was attached to an ÄKTA start (GE Healthcare), washed with azide (0.01%) and equilibrated in buffer for 10 CV (0.3 MPa, 1.0 mL/min or 5.0 mL/min). Sample was loaded via the sample line equilibrated in buffer. Column was washed with wash buffer (15-20 CV) to remove unbound protein. The protein was eluted by increasing the concentration of elution buffer. Fractions (3 mL) were collected during the elution. Fractions with peaks indicating increases absorbance (A280) on

the chromatogram were collected and analysed by SDS-PAGE. The HisTrap™ FF column was washed by buffer A, azide 0.1% and 20% ethanol for storage.

Table 28 Purification conditions for Histidine tagged proteins using a 5 mL HisTrap™ FF column

	Loading Buffer	**Elution buffer**	**Elution Method**
ThAOS	1B	1C	Gradient 0-100% B 20CV
hSCPs	2B	2C	6 CV 60% 2C
TTSCP	2B	2C	6 CV 60% 2C
pfThiE	1B	1C	Gradient 0-100% B 20CV
Thiminase 1	1B	1C	Gradient 0-100% B 20CV
pfBAL	3B	3C	Gradient 0-100% B 20CV
Tev	4A	4B	Gradient 0-100% B 10CV

6.2.4.3 Tobacco Etch Virus (TEV) cleavage

TEV Protease was expressed and purified in house from an N-terminal hexa-histidine MBP tagged construct using the conditions described in Table 27. Protein was purified via IMAC before dialysis in the self-cleavage buffer **4C** for 2 hours followed by dialysis into storage buffer **4D** for 2 hours. Purity was analysed by SDS- PAGE (Figure 111) and aliquots were flash frozen, stored at -80°C and thawed on ice immediately prior to use

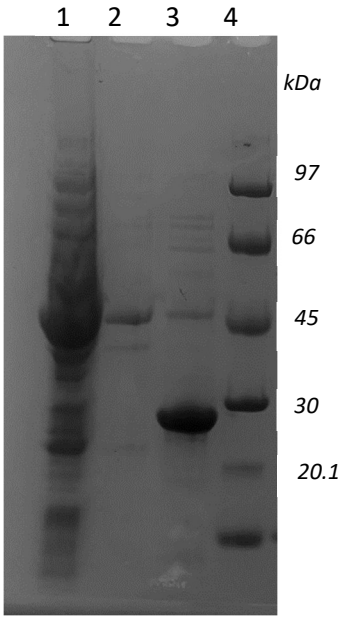


Figure 111 Purification of Tev Protease.1) Lysate, 2) Flow through, 3) Purified Tev,4) LMWM

DTT (1 mM) and EDTA (0.5 mM) were added to the protein (conc 1-1.5 mg/mL) and the His-tag was cleaved by the addition of TEV-protease (0.014 equivalents) and incubation at room temperature for 6 h. After the TEV cleavage, the protein solution was subjected to 5 mL HisTrap FF column equilibrated with wash buffer at a constant flow of 5 mL/min. The flow-through that contained purified protein with His-tag cleaved was collected and purity was analysed via SDS-PAGE.

6.2.4.4 Size Exclusion Chromatography

A HiPrep™ 16/600 superdex™ S200 size exclusion column (120 mL) was attached to an ÄKTA purifier (GE Healthcare). The column was equilibrated with buffer for 140 minutes (0.3 MPa, 1.0 mL/min). The injection loop was washed with both water and buffer. The concentrated protein sample (1-10 mL) was injected into the loop and loaded onto the pre-equilibrated

column. The protein was eluted with buffer for 140 minutes and collected in 3 mL fractions. The fractions corresponding to peaks on the chromatogram were collected and analysed by SDS-PAGE. Fractions containing pure protein were combined and concentrated.

6.2.4.5 Ammonium sulphate precipitation of Thiaminase 1

Cell pellet (2.4g) was resuspended in buffer (50 mM Tris pH7.5, 2 mM DTT, 2mM EDTA, 29 mL). Cells were lysed by the standard protocol and centrifuged to remove cell debris. The sample was stirred on ice and ammonium sulfate (8.95 g, 50% solution) was added over 10 mins. The sample was stirred for 1 hour before centrifugation (12000 rpm, 15 min). The supernatant was transferred into a fresh vessel and ammonium sulfate (4.4 g, 70% final solution) was added over 10 mins. The sample was stirred on ice for 1 hour before centrifugation (12000 rpm, 15 min). The supernatant and pellet (50-70 % ammonium sulfate precipitate) were separated. The pellet was resuspended in buffer (50 mM Tris pH7.5, 2 mM DTT, 2 mM EDTA, 20 mL).

6.2.4.6 Ion exchange chromatography

A HiTrap DEAE FF 1mL (weak anion exchange column) was equilibrated in 10 CV buffer (50 mM Tris pH7.5, 2 mM DTT, 2 mM EDTA, 20 mL). The protein sample was loaded at 1 mL/min. The column was washed with 10 CV of buffer (50 mM Tris pH7.5, 2 mM DTT, 2mM EDTA,) until the UV absorbance returned to baseline level. The protein was eluted by increasing the concentration of elution buffer (50 mM Tris pH7.5, 2 mM DTT, 2mM EDTA, 2M NaCl) and collected in 3 mL fractions. The elution profile was as follows: gradient 0-10 % elution buffer over 5 mL, isocratic elution at 10% for 5 mL, gradient 10-15 % elution buffer over 5 mL, isocratic elution at 15% for 10 mL, gradient 15-100 % elution buffer over 5 mL. Protein fractions were analysed by SDS-PAGE to determine which fractions contained the desired protein.

6.2.5 Protein Characterisation

6.2.5.1 SDS-PAGE

A typical SDS gel (6 mL) was made up of 4.5 mL 12 % running gel and 1.5 mL 6% stacking gel. Samples were denatured by the addition of 2x Sample buffer (SB) and boiled at 90-100°C for 10 mins. Samples were centrifuged at 11000 rpm for 30 sec before being loaded onto the gel. Low molecular weight maker (5 µL, GE Healthcare) was used to determine the MW of each band. The gel was run for ~60 minutes at 180 volts in TGS buffer (BIO-RAD). Protein bands were observed by staining the gel either with InstantBlue™ (sigma) for 10 min-1 hour or incubating overnight with Coomassie stain (50% v/v H_2O; 40% v/v methanol; 10% v/v acetic acid; 0.25% w/v Coomassie brilliant blue R250) before de-staining with (50% v/v H_2O; 40% v/v methanol; 10% v/v acetic acid) until excess blue was removed.

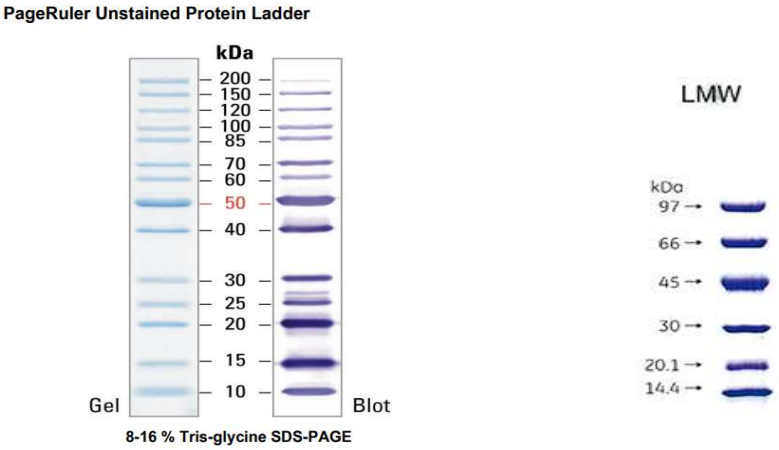


Figure 112 The two protein ladders used for calibration of protein analysis by SDS PAGE gel were PageRuler Unstained Protein Ladder (Thermo Scientific, left) and Calibration Kit Low Molecular Weight for Electrophoresis (Cytiva, right)

6.2.5.2 Nu-PAGE SDS-PAGE

To separate small proteins (SCP constructs) NuPAGE 4-12 % Bis-Tris Gels with NuPAGE MES SDS Running Buffer (Life Technologies) were used. Samples were run for ~30 minutes at 180 volts. Protein bands were observed by staining the gel either with InstantBlue™ (sigma) for 10 min-1 hour or incubating overnight with Coomassie stain.

6.2.5.3 Protein concentrations

Initially protein concentrations were measured by absorbance at 280 nm using the nanovue spectrophotometer. However, the extinction coefficient of the protein changes upon functionalisation protein concentrations prior to functionalisation and after functionalisation were measured using a Bradford assay.

6.2.5.4 nanovue spectrophotmeter (GE)

Protein concentrations were determined by measuring 3 µL of protein for absorbance A280 after the instrument was blanked with the corresponding buffer. The extinction coefficients corrected for each mutant were used to convert the A280 to the protein concentration.

6.2.5.5 Bradford assay

Protein concentration was estimated using the Bradford assay. This method is based on the change in colour of the Coomassie Brilliant Blue G-250 dye from brown to blue under acidic conditions, which can be used to measure the presence of basic amino acids (lysine, arginine and histidine). A calibration curve was created using varying concentrations of bovine serum albumin (BSA) in dH_2O (0.25 mg/mL, 0.5 mg/mL, 1 mg/mL and 1.4 mg/mL). Each BSA solution (50 µL) was added to Bradford reagent (1.5 mL, Sigma-Aldrich) and the absorbance was measured at 595 nm. The purified protein of interest (50 µL) was added to Bradford reagent (1.5 mL) and incubated at room temperature for 15 min. The absorbance at 595 nm was measured and the concentration was estimated from the calibration curve. Protein with concentrations greater than 1.3 mg/mL were diluted to be within the bounds of the calibration.

6.2.5.6 Protein Liquid Chromatography – Mass Spectrometry (LC-MS)

Protein was diluted to a concentration of 5-10 µM and centrifuged for 10 min at 17000 xg. 5 µL of the supernatant was used for LC ESI-MS on the Synapt G2-Si Q-TOF (Waters) instrument with Phenomenex C4 3.6 µM LC column coupled to an electrospray ionisation ESI source. The LC gradient started at 5% acetonitrile and 95% water with 0.1% formic acid and ran to 95% acetonitrile over 7 min.

6.2.5.7 Circular dichroism (CD)

CD data was collected on a Chirascan VX circular dichroism instrument, using a quartz "bottle" cell with a pathlength of 0.005 cm (Starna Scientific Ltd). The purified protein was buffer exchanged into 10 mM phosphate buffer, pH 7 using a PD-10 desalting column (Cytiva) and the concentration of the protein was adjusted to between 0.15 and 0.3 mg/mL. A baseline reading with 10 mM phosphate buffer, pH 7 was obtained. To determine the secondary structure of each protein, data was collected for 0.15 mg/mL protein in duplicate at 25 °C from 260 to 190 nm in 0.5-nm steps. Data was processed using the DichroWeb server[199], with the CDSSTR algorithm[200] using reference dataset 7[201], to estimate the secondary structure of the proteins. To determine the melting temperature (T_m) of each protein, data was collected and over the temperature range from 25 °C to 100 °C in 5 °C

increments, with measurements at each temperature following a 60 s equilibration period. The spectral run was cooled to 25 °C after heating to 100 °C to assess refolding. The concentrations used for each protein for T_m analysis were 0.3 mg/mL TTSCP, 0.15 mg/mL TTSCP L102C and 0.2 mg/mL TTSCP ΔDSB L102C.

6.2.6 Protein functionalisation

Proteins were prepared for functionalisation by buffer exchange into the required buffer (Table 29) using a pd10 column (Cytiva) equilibrated in the functionalisation buffer or by using a Vivaspin 500 centrifugal concentrator, MWCO 10 kDa (Sartorius) to exchange to buffer with functionalisation buffer (3 X 500 µL). Proteins were diluted to (100 µM) then functionalised by incubation with the functionalisation regent.

Stocks (1 mM) of each functionalisation reagent were prepared in 50:50 DMSO/water (MMeThz, MeThzBr and BnThzBr) or water (MBnThz and MBnTri) and frozen at -20 °C until required.

6.2.6.1 Screening of conditions

Functionalisation reagents (1-50 equiv) were added to each protein sample (50 µL, 100 µM) and incubated for 30 mins to 24 hours. Samples were centrifuged to remove any precipitate, diluted to 5 µM and analysed by LCMS.

6.2.6.2 Functionalisation to prepare catalysts

Reactions (1 - 5 mL) were set up under the conditions specified in Table 29. After the required incubation time, reactions were centrifuged to remove any precipitate and excess functionalisation reagent was removed via buffer exchanging using a pd10 column equilibrated with reaction buffer (50 mM HEPES, pH 8.0). The concentration of the functionalised protein was determined via Bradford assay and when required protein was concentrated using Vivaspin 6 centrifugal concentrator, MWCO 10 kDa (Sartorius). Functionalised proteins were analysed via LCMS to assess the level of functionalisation.

Table 29 Optimised protein functionalisation conditions

Protein	**Functionalisation reagent**	**Functionalisation conditions**
hSCP A100C	10 equiv MThzBr	10 mM sodium phosphate buffer pH 7 , 18 h, 4 °C
hSCP A100C	10 equiv BnThzBr	10 mM sodium phosphate buffer pH 7 , 18 h, 4 °C
ThAOS K234C	10 equivMaleimide	50 mM HEPES pH8, 18 h, 4 °C

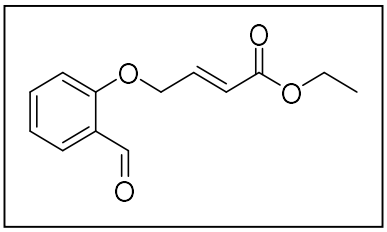
ThAOS K234C	10 equiv MBnThz	50 mM HEPES pH8, 18 h, 4 °C
ThAOS K243C V79S	10 equiv MBnThz	Buffer 1A, pH8, 18 h, RT
pfThiE (all mutants)	10 equiv MBnThz	20 mM Mes pH 7, 2 h, RT
hSCP A100C	10 equiv MMeThz	20 mM Mes pH 6.0, 2 h, RT
hSCP (All mutants)	10 equiv MBnThz	20 mM Mes pH 6.0, 2 h, RT
TTSCP V83C	10 equiv MBnThz	20 mM Mes pH 6, , 5 h, RT
TTSCP L102C	5 equiv MBnThz	20 mM Mes pH 7, 1.5 h,RT
TTSCP L112C	10 equiv MBnThz	20 mM Mes pH 7, 0.5 h, RT
TTSCP ΔDSB L102C	5 equiv MBnThz	20 mM Mes pH 7, 1.5 h, RT
TTSCP L102C	10 equiv MBnTri	20 mM Mes pH 7, 1.5 h, RT
TTSCP ΔDSB L102C	10 equiv MBnTri	20 mM Mes pH 7, 1.5 h, RT

6.2.6.3 ThAOS PLP removal and functionalisation

L-Penicillamine (75 µL, 10 mM) was added ThAOS K234C V79S (675 µL, 111 µM in Buffer 1A) and incubated for 1 hour at RT. The sample was split into 3 X 250 µL samples. One sample was buffer exchanged using a Vivaspin 500 centrifugal concentrator, MWCO 30 kDa (Sartorius) into buffer 1A (3 X 500 µL). The second sample was dialysed in 15 mL Buffer 1A using Slide-A-Lyzer™ Dialysis Cassettes, 10K MWCO, 0.5 mL (Thermo Scientific). The third sample was used without any work up and a 4th sample of ThAOS K234C V79S (100 µM) was prepared. All samples were functionalised with MBnThz (10 equiv, 18 hours at RT). After incubation samples were centrifuged then exchanged using a Vivaspin 500 centrifugal concentrator, MWCO 30 kDa (Sartorius) into buffer 5A (3 X 500 µL), analysed via LCMS and used for intramolecular Stetter reactions.

6.3 Synthesis

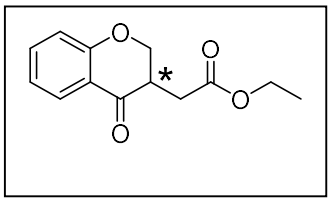
6.3.1 Intramolecular Stetter starting material **18**



Compound 18 (4-(2-Formyl-phenoxy)-but-2-enoic acid ethyl ester): Salicylic aldehyde (1.7 mL, 16 mmol, 1 equiv) and K_2CO_3 (3.3 g, 24 mmol, 1.5 equiv) were dissolved in 30 mL of acetone, and ethyl 4-bromocrotonate (2.35 mL, 17 mmol, 1.05 equiv) was added into the mixture then stirred overnight. The reaction was quenched with water (10 mL) and the organic layer was extracted with ethyl acetate (30 mL X 3). The organic layer was dried over MgSO4 and concentrated *in vacuo*. The product was purified by flash chromatography (hexane/ethyl

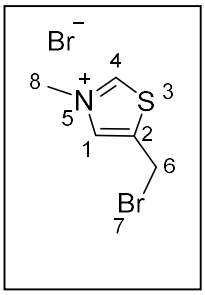
acetate).[1]H NMR (500 MHz, $CDCl_3$) δ 10.57 (s, 1H), 7.88 (dd, *J* = 7.7, 1.8 Hz, 1H), 7.56 (ddd, *J* = 8.5, 7.3, 1.9 Hz, 1H), 7.15 – 7.07 (m, 2H), 6.95 (dd, *J* = 8.4, 0.9 Hz, 1H), 6.23 (dt, *J* = 15.8, 2.1 Hz, 1H), 4.84 (dd, *J* = 4.1, 2.1 Hz, 2H), 4.24 (q, *J* = 7.1 Hz, 2H), 1.32 (t, *J* = 7.1 Hz, 3H).[13]C NMR (126 MHz, $CDCl_3$) δ 189.28, 165.81, 160.20, 141.14, 135.91, 128.86, 125.17, 122.58, 121.46, 112.55, 66.89, 60.74, 14.22. Exact mass calcd for [C13H14O4Na]+ requires *m/z* =257.0789, found *m/z* = 257.0786 (ESI). NMR and MS data are in accordance with reported values[76]

6.3.2 Intramolecular Stetter product 19



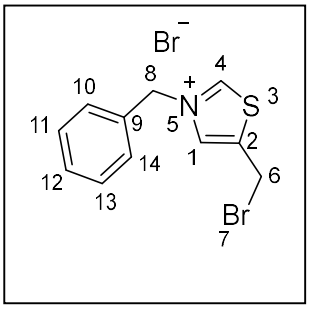
Compound 19: Compound 18 (2.5 mmol, 585 mg, 1 equiv) and 3-benzyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride (0.5 mmol, 134.5 mg, 0.2 equiv) were dissolved in DCM (30 mL). Triethylamine (0.5 mmol, 70 µL, 0.2 equiv) was added and the reaction was stirred at 50°C for 16 hours. Reaction was cooled and solvent was removed under reduced pressure. Product was purified via flash column chromatography (5:1 hexane ethylacetate). [1]H NMR (500 MHz, $CDCl_3$) δ 7.90 (dd, *J* = 7.9, 1.8 Hz, 1H), 7.51 – 7.47 (m, 1H), 7.05 – 6.96 (m, 2H), 4.61 (dd, *J* = 11.2, 5.3 Hz, 1H), 4.31 (t, *J* = 11.9, 11.3 Hz, 1H), 4.24 – 4.17 (m, 2H), 3.34 (ddt, *J* = 12.1, 8.2, 5.1 Hz, 1H), 2.94 (dd, *J* = 16.9, 4.8 Hz, 1H), 2.43 (dd, *J* = 16.9, 8.1 Hz, 1H), 1.30 (t, *J* = 7.1 Hz, 3H).[13]C NMR (126 MHz, $CDCl_3$) δ 192.61, 171.36, 161.75, 136.01, 127.40, 121.53, 120.52, 117.84, 70.26, 60.98, 42.54, 30.38, 14.18. Exact mass calcd for [C13H14O4Na]+ requires *m/z* =257.0789, found *m/z* = 257.0784 (ESI). NMR and MS data are in accordance with reported values[76]

6.3.3 *N*-methyl-5-bromomethylthiazolium bromide 29 (MThzBr)



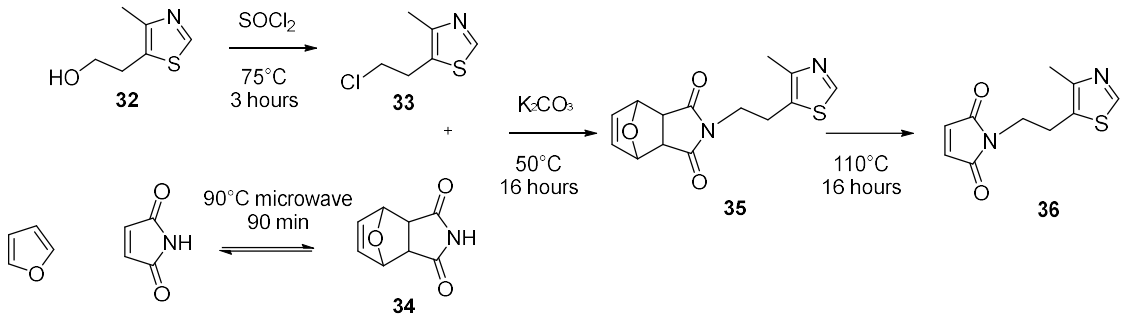
Compound 29 5-hydroxymethylthiazole (0.5 mmol, 57 mg, 1 equiv) was dissolved in acetonitrile (5 mL). Methyl iodide (5 mmol, 312 µL, 10 equiv) was added and the solution was stirred for 4 days. Residual methyl iodide and acetonitrile were removed under reduced pressure. The sample was dissolved in DCM and purified via column chromatography 9:1 DCM: methanol. 40% HBr solution aq. (1 mL) was added to the intermediate **28**, 5-hydroxymethyl-N-methyl-thiazolium iodide (0.35 mmol, 90 mg) and stirred for at 100°C overnight. The solvent was evaporated under reduced pressure give 5-bromomethyl-N-methylthiazolium bromide Yield: 68 %, 93 mg. [1]H NMR (500 MHz, MeOD) δ 10.08 (s, $1H^4$), 8.48 (s, $1H^1$), 4.98 (s, $2H^6$), 4.26 (s, $3H^8$).[13]C NMR (126 MHz, MeOD) δ 159.68 ($1C^4$), 142.00 ($1C^2$), 136.85 ($1C^1$), 41.39 ($1C^8$), 19.73 ($1C^6$). Exact mass calcd for [C5H7N1S1Br1] *m/z* = 191.94771 Found *m/z* = 191.9481 (ESI) Error = +/- 2.03 ppm

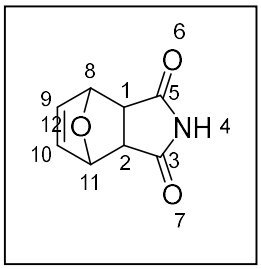
6.3.4 *N*-benzyl-5-bromomethylthiazolium bromide **31** (BnThzBr)



Compound 31 5-hydroxymethylthiazole hydrogen bromide (0.25 mmol, 64 mg, 1 eq) was dissolved in acetonitrile (5 mL) and benzylbromide (2 mmol, 237 µL, 8 eq) was added and the solution stirred for 24 hours at 80°C. The product was concentrated under reduced pressure in fume hood and recrystallized from hot methanol. The recrystallized intermediate **30,** 5-hydroxymethyl-N-benzyl-thiazoliumbromide (0.35 mmol, 90 mg) was added to 40% HBr solution aq. (1 mL) and stirred for at 100°C overnight. The solvent was evaporated under reduced pressure in a fume hood to give 5-bromomethyl-N-methylthiazolium bromide in a 73% yield, 63 mg. Exact mass calcd for [C11H11Br1N1S1] *m/z* = 267.97901 Found mass *m/z* = 267.9801 (ESI) Error = +/- 4.07 ppm.^{1}H NMR (500 MHz, MeOD) δ 10.24 (s, 1H^{4}), 8.55 (s, 1H^{1}), 7.57 – 7.45 (m, 5H$^{10-14}$), 5.78 (s, 2H^{8}), 4.95 (s, 2H^{6}).^{13}C NMR (126 MHz, MeOD) δ 158.98 (1C^{4}), 142.93(1C^{2}), 135.46 (1C^{1}), 132.46 (1C^{9}), 129.63 (1C^{12}), 129.31-128.88 (4C^{10,11,12,14} 129.31, 128.88) 58.70 (1C^{8}), 19.70 (1C^{6}).

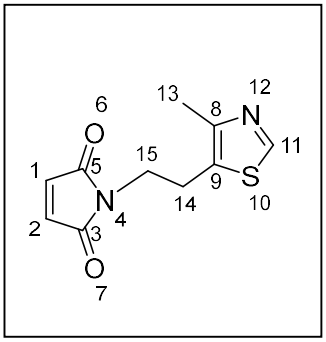
6.3.5 5-(2-maleimidoethyl)-4-methyl-thiazole **36**





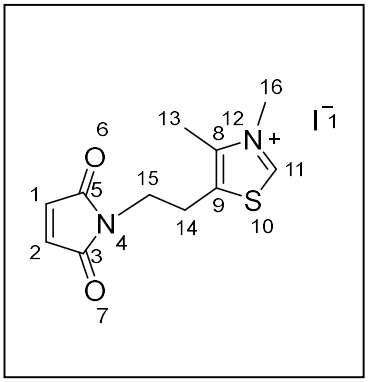
Compound 34: Maleimide (10 mmol, 970 mg, 1 eq) was dissolved in dH_2O (11.4 mL) in a 35 mL microwave tube. Furan (17 mmol, 1.24 mL, 1.7 eq) was added. The reaction was heated to 90 °C in a microwave reactor for 90 min. The reaction was cooled, filtered and product **5** was washed with dH_2O then dried under vacuum (white powder, 60% yield 1.0 g). ^{1}H NMR (500 MHz, Methanol-d_4) δ 6.55 (t, *J* = 0.9 Hz, 2H^{9,10}), 5.18 (t, *J* = 1.0 Hz, 2H^{8,9}), 2.92 (s, 2H^{1,2}). ^{13}C NMR (126 MHz, Methanol-d_4) δ 178.43 (2C^{3,5}), 136.22 (2C^{9,10}), 80.81 (2C^{8,11}), 48.69 (2C^{1,2}). Mass calcd for $[C_8H_7O_4N_1O_3Na]^+$ ESI $[M+]^+$ requires *m/z* =188.0318 ,found *m/z* = 188.0315 (ESI). In agreement with published data [242]

Intermediate 33. 2-(4-methylthiazol-5-yl)ethanol (1.20 mL , 10 mmol) was added to thionyl chloride (10 mL) and the mixture was stirred at reflux for 3 hours. Thionyl chloride was removed by rotary evaporation and the product was washed with toluene and dried under vacuum. The product was used without further purification.



Compound 36 (5-(2-maleimideethyl)-4-methyl-thiazole): Compound **34** (3 mmol, 500 mg, 1 eq) and potassium carbonate (15 mmol, 2.1 g, 5 equiv) were dissolved in DMF (10 mL) under argon and heated to 50 °C for 1 hour. Crude intermediate **33** (3.6 mmol, 581 mg, 1.2 equiv) was added and the reaction was stirred at 50 °C for 18 hours. The reaction was cooled, filtered and DMF was removed under reduced pressure. The crude product was dissolved in ethyl acetate (50 mL) and washed with water (3X 10 mL). The organic extract was dried with $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was dissolved in ethyl acetate, filtered through a silica pad and concentrated under reduced pressure. The colourless oil **6** was dissolved in toluene 10 mL and refluxed at 110 °C for 18 hours. Toluene was removed under reduced pressure to give **7** in 242 mg 35% yield. ^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.60 (s, 1H^{11}), 6.72 (s, 2H^{1,2}), 3.77 (t, *J* = 7.2, 2H^{15}), 3.11 (t, *J* = 7.2 Hz, 2H^{14}), 2.41 (s, 3H^{13}).^{13}C NMR (126 MHz, $CDCl_3$) δ 170.3 (2C^{3,5}), 149.9 (2C^{8,11}), 134.2 (2C^{1,2}), 126.6 (1C^{9}), 38.4 (C^{15}), 25.0 (C^{14}), 14.8 (C^{13}). Mass calcd for $[C_{10}H_{10}N_2O_2S]^+$ requires ESI $[M+H]^+$ *m/z*= 223.0536, found *m/z* = 223.0535 (ESI) .MP 83-84 °C. IR : 3068.95, 1437.70, 1410.56, 1367.71, 1150.59, 902.05, 829.91

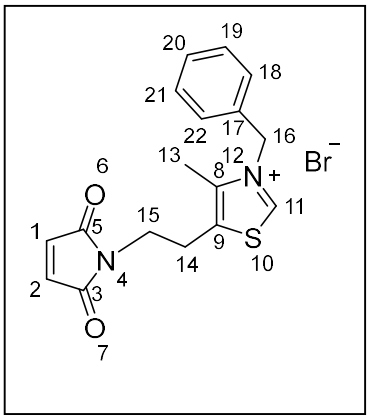
6.3.6 N-methyl-5-(2-maleimidoethyl)-4-methyl-thiazolium bromide 37 (MMeThz)



Compound 37 (N-methyl-5-(2-maleimidoethyl)-4-methyl-thiazolium bromide): Compound 36 (0.5 mmol, 111 mg, 1 eq) and methyl iodide (5 mmol, 311 uL, 10 eq) were dissolved in acetonitrile 5 mL and refluxed at 40 °C for 4 days. Acetonitrile was removed under reduced pressure and the product was triturated in diethyl ether and dried under vacuum to give the product in 78 % yield, 142 mg.

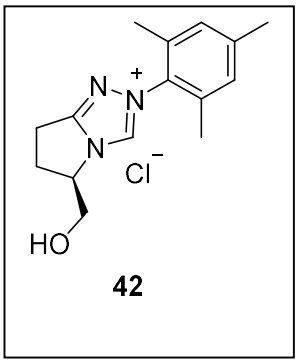
^{1}H NMR (500 MHz, Chloroform-*d*) δ 11.07 (s, 1H^{11}), 6.77 (s, 2H^{1,2}), 4.37 (d, *J* = 0.7 Hz, 3H^{16}), 3.84 (t, *J* = 6.9 Hz, 2H^{15}), 3.22 (t, *J* = 6.9 Hz, 2H^{14}), 2.53 (s, 3H^{13}).^{13}C NMR (126 MHz, $CDCl_3$) δ 170.21 (2C^{5,3}), 158.26(1C^{11}), 143.49 (1C^{8}), 134.49 (2C^{1,2}), 133.35 (1C^{9}), 41.72 (1C^{15}), 37.24(1C^{16}), 26.13 (1C^{14}), 12.20 (1C^{13}).Mass calcd for [C11H13N2O2S1] *m/z* = 237.06922, Found mass *m/z* = 237.0698 (ESI), Error = +/- 2.45 ppm

6.3.7 N-benzyl-5-(2-maleimidoethyl)-4-methyl-thiazolium bromide **38** (MBnThz)



Compound 38 (N-Benzyl-5-(2-maleimidoethyl)-4-methyl-thiazolium bromide): Compound 36 (1 mmol, 222 mg, 1 eq) and benzylbromide (2 mmol, 238 uL, 2 eq) were dissolved in acetonitrile (5 mL) and refluxed at 80°C for 24 hours. Acetonitrile was removed under reduced pressure. The product was triturated in diethyl ether and isolated via filtration. (Sticky/amorphous off-white solid) 282 mg 72% yield. ^{1}H NMR (500 MHz, Chloroform-*d*) δ 11.64 (s, 1H, C11), 7.45 – 7.36 (m, 3H^{18,20,22}), 7.36 – 7.30 (m, 2H^{19,21}), 6.69 (s, 2H^{1,2}), 6.13 (s, 2H^{16}), 3.81 (t, *J* = 6.7 Hz, 2H^{15}), 3.19 (t, *J* = 6.7 Hz, 2H^{14}), 2.39 (s, 3H^{13}). ^{13}C NMR (126 MHz, $CDCl_3$) δ 170.1 (2C^{3,5}), 159.2(1C^{11}), 143.4 (1C^{8}), 134.3 (1C^{1}), 134.0 (1C^{2}), 131.6 (1C^{9}), 129.5(3C^{22,20,18}), 129.4(2C^{21,19}), 128.1(1C^{9}), 57.6 (1C^{15}), 37.4 (1C^{16}), 26.1 (1C^{14}), 12.3 (1C^{13}). Mass calcd for $[C_{17}H_{17}N_2O_2S]^+$ requires ESI $[M+H]^+$ *m/z* = 313.1010, found *m/z* = 313.1009. MP 179-180°C.IR: 3033.24, 1497.70, 1373.42, 1333.43, 1158.45, 951.33, 913.47, 834.91, 692.78

6.3.8 Compound 42



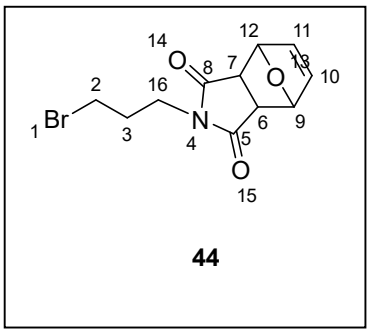
42

D-pyroglutaminol (3.00 g, 26.05 mmol, 1 eq.), TBS-Cl (4.71 g, 31.26 mmol, 1.2 eq.) and imidazole (3.54 g, 52 mmol, 2 eq.) were added to a flask under nitrogen atmosphere, dissolved in 60 mL of dry DCM (0.4 M), and left reacting at room temperature overnight. The crude was diluted to 75 mL and was washed three times with deionized water (3X 20 mL) and once with brine (20 mL). The crude product was added to a flame-dried 1-L round-bottomed flask was charged with a magnetic stir bar, DCM (150 mL) and trimethyloxonium tetrafluoroborate (3.5 g, 23.5 mmol, 1.1 equiv) were added. The solution was stirred at ambient temperature under an atmosphere of N_2 (g) for 17 h. The solution was cooled to 0 °C and quenched by the portion-wise addition of sat aq NaHCO3 (150 mL) over 1.5 h followed by stirring for 1 h. The biphasic solution was transferred to a separating funnel, the organic phase separated, dried over Na_2SO_4, filtered, and concentrated to afford the crude ether (intermediate 41) which was used without further purification (4.05 g). ^{1}H NMR (500 MHz, chloroform-*d*) δ 3.97 (dddt, *J* = 8.1, 5.9, 4.3, 1.8 Hz, 1H), 3.84 – 3.74 (m, 4H), 3.59 – 3.46 (m, 1H), 2.57 – 2.46 (m, 1H), 2.40 (ddd, *J* = 16.4, 10.4, 5.8 Hz, 1H), 2.10 (dddd, *J* = 13.0, 10.4, 8.2, 6.7 Hz, 1H), 1.95 (dddd, *J* = 12.9, 10.3, 5.9, 4.5 Hz, 1H), 0.96 – 0.85 (m, 9H), 0.06 (d, *J* = 1.2 Hz, 6H).

A flame-dried 500-mL round-bottomed flask was charged with the crude ether (1.61 g, 5.3 mmol, 1.0 equiv), a magnetic stir bar, 2- mesitylhydrazinium chloride (1 g, 5.3 mmol, 1.0 equiv) and MeOH (30 mL) was stirred at for 1 h 50 °C under an atmosphere of Argon. The reaction was allowed to cool to ambient temperature and the concentrated under reduced pressure to afford the hydrazone as a crude orange solid. The crude material was suspended in EtOAc and placed in a sonicating bath. The orange precipitate was removed by suction filtration and the filtrate was concentrated under reduced pressure to afford the crude hydrazine (0.82 mg, 3.3 mmol) which was used immediately without purification.

Triethylorthoformate (4.5 mL, 33 mmol, 10 equiv), chlorobenzene (3.3. mL) and anhydrous HCl (4 M in 1,4-dioxane, 0.83 mL, 3.3 mmol, 1.0 equiv) were added to the crude hydrazone. The reaction was heated to 100 °C for 50 min under argon. The brown solution cooled to ambient temperature and concentrated under reduced pressure (800 mg). A portion (250 mg) of the crude brown solid was purified by flash column chromatography (DCM: 10% MeOH with 7M ammonia) to give the desired triazole. (Yield 75 mg, 2.4 %) ^{1}H NMR (500 MHz, chloroform-*d*) δ 7.94 (s, 1H), 6.94 (d, 2H), 5.39 (s, 1H), 3.94 (tt, *J* = 8.1, 3.7 Hz, 1H), 3.78 (dd, *J* = 11.6, 3.3 Hz, 1H), 3.58 (dd, *J* = 11.7, 5.2 Hz, 1H), 2.79 (ddd, *J* = 16.8, 9.6, 7.3 Hz, 1H), 2.66 (ddd, *J* = 16.7, 10.1, 5.4 Hz, 1H), 2.30 (d, *J* = 1.4 Hz, 6H), 2.24 (s, 3H), 2.18 (dtd, *J* = 10.3, 5.3, 2.6 Hz, 1H), 1.91 (dddd, *J* = 13.3, 9.5, 5.5, 3.9 Hz, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 164.11, 158.12, 138.84, 136.58, 136.41, 136.27, 129.47, 129.26, 66.21, 59.86, 29.43, 23.94, 20.98, 18.16, 17.88. In agreement with published data[181]

6.3.9 MBnTri



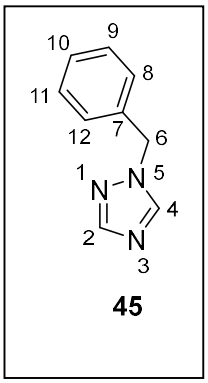
44

Compound 44: Furan protected maleimide (2.27 mmol, 375 mg, 1 eq) and potassium carbonate (9.1 mmol, 1.25 g, 4 eq) were dissolved in acetone (10 mL) and refluxed at 65°C for 10 minutes. Potassium iodide (0.45 mmol, 70 mg, 0.2 eq) and 1,3-dibromopropane (2.5 mmol, 628 mg, 1.1 eq) were added and the reaction was stirred overnight at 65°C until completion.

The reaction was cooled, diluted with ethylacetate (15 mL) and washed with water (3 X 10 mL) and brine (1x 10 mL). The organic layer was dried with $MgSO_4$, filtered and solvent was removed under reduced pressure to afford the crude product. The product was further purified via column chromatography in hexane: ethyl acetate 2:3. (Yield 843 mg *product still contains water as shown by NMR) ^{1}H NMR (500 MHz, chloroform-*d*) δ 6.54 (t, *J* = 1.0 Hz,

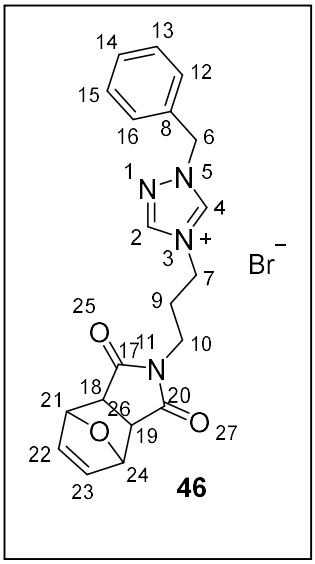
2H[10,11]), 5.29 (t, *J* = 0.9 Hz, 2H[9,12]), 3.66 (t, *J* = 6.9 Hz, 2H[16]), 3.36 (t, *J* = 6.7 Hz, 2H[2]), 2.87 (s, 2H[6,7]), 2.17 (p, *J* = 6.8 Hz, 2H[3]).[13]C NMR (126 MHz, CDCl3) δ 176.13 (2C[5,8]), 136.53 (2C[10,11]), 80.99 (2C[9,12]), 47.42(2C[6,7]), 37.63 (1C[16]), 30.65 (1C[3]), 29.60 (1C[2]).

Compound 45



45

Sodium 1,2,3 triazole (364 mg, 1 equiv, 4 mmol) was added to acetone (10 mL). Benzylbromide (476 µL, 1 equiv, 4 mmol) was added and the reaction was heated to 65°C and stirred for 18 hours. The reaction was cooled to ambient temperature, filtered to remove white ppt and the filtrate was concentrated under reduced pressure to yield a yellow oil. The product was purified by flash column chromatography (DCM: 5 % MeOH) to give a yellow powder. (Yield 54%, 348 mg) [1]H NMR (500 MHz, Chloroform-*d*) δ 8.07 (s, 1H[4]), 8.00 (s, 1H[2]), 7.46 – 7.34 (m, 3H[9,10,11]), 7.31 – 7.27 (m, 2H[8,12] obscured by $CDCl_3$), 5.37 (s, 2H[6]). [13]C NMR (126 MHz, $CDCl_3$) δ 152.19 (1C[2]), 143.05 (1C[4]), 134.56 (1C[7]), 129.10 (2C[9,11]), 128.70 (1C[10]), 128.04 (2C[8,12]), 53.64 (1C[6]).

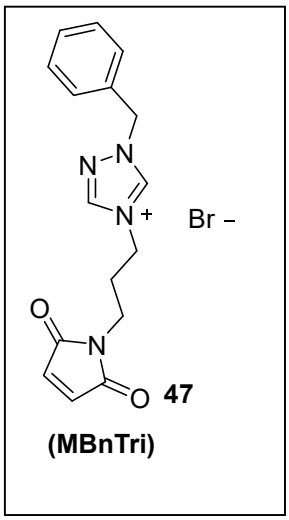


46

Compound 46

Compound 45 (160 mg, 1 mmol, 1 equiv) was dissolved in acetonitrile (10 mL). Compound 45 (538 mg, 2 mmol, 2 equiv) was added and the reaction was refluxed at 80 °C for 40 hours with stirring. The solvent was removed under reduced pressure and the product was purified by flash column chromatography (DCM: 5-10% gradient MeOH) Yield 32 mg. [1]H-[15]N HMBC was used to help assign the spectra. The product appears to be present in two forms of the triazole which appears to be due to some of the furan protecting group being degraded [1]H NMR (500 MHz, Chloroform-*d*) δ 11.14 (s, 0.5H), 11.11 (s, 0.5H), 9.05 (s, 0.5H), 8.99 (s, 0.5H), 7.58 (ddt, *J* = 8.1, 5.2, 2.3 Hz, 2H), 7.46 – 7.36 (m, 3H), 6.74 (s, 1H), 6.50 (t, *J* = 0.9 Hz, 1H), 5.68 (d, *J* = 12.6 Hz, 2H), 5.22 (t, *J* = 1.0 Hz, 1H), 4.57 (dt, *J* = 20.2, 7.0 Hz, 2H), 3.63 – 3.57 (m, 1H), 3.57 – 3.52 (m, 1H), 3.06 (s, 1H), 2.38 (tdd, *J* = 11.2, 7.5, 4.7 Hz, 2H).[13]C NMR (126 MHz, $CDCl_3$) δ 176.96, 171.01, 144.16, 144.13, 142.64, 142.56, 136.44, 134.48, 131.46, 131.44, 129.80, 129.61, 129.56, 129.50, 129.38, 80.93, 56.69, 56.68, 52.30, 50.77, 48.00, 46.49, 46.39, 34.99, 34.03, 29.35, 28.36.

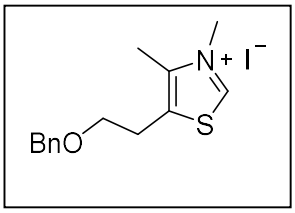
Compound 47



Compound 46 (25 mg) was dissolved in toluene (5 mL) and refluxed at 110 °C for 18 hours with stirring. The solvent was removed under reduced pressure. 16 mg of product was obtained but NMR analysis of this product showed it was not pure. However due to the low yield the crude product was used to functionalise the TTSCP and the expected mass shift of MBnTri was obtained.

^{1}H NMR (500 MHz, DMSO-d_6) δ 10.23 (d, *J* = 14.8 Hz, 1H), 7.47 – 7.40 (m, 7H), 7.07 (s, 2H), 5.64 (d, *J* = 1.8 Hz, 3H), 4.31 (t, *J* = 7.1 Hz, 1H), 4.28 – 4.22 (m, 2H), 3.49 (t, *J* = 6.4 Hz, 2H), 3.32 (s, 5H), 3.17 (t, *J* = 4.8 Hz, 1H), 2.17 – 2.10 (m, 2H), 2.09 (s, 1H).

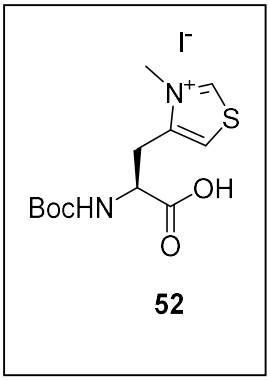
6.3.10 Synthesis of aqueous Stetter catalyst compound 50



Compound 50 To a suspension of NaH 60% (726 mg, 18.16 mmol) in mineral oil in anhydrous DMF (21 mL), at 0 °C, were added dropwise benzyl bromide (2.16 mL, 18.16 mmol) followed by a solution of 2-(4-methylthiazol-5-yl)ethanol (1.67 mL, 13.97 mmol) in anhydrous DMF (10 mL). The reaction mixture was stirred during 16 h at room temperature and ethyl acetate (20 mL) was added. The organic layer was washed with sat. aq. NH4Cl (3 × 10 mL), brine (3 × 10 mL) and water (3 × 10 mL), dried over Na_2SO_4, filtered and the solvent was removed under reduced pressure.

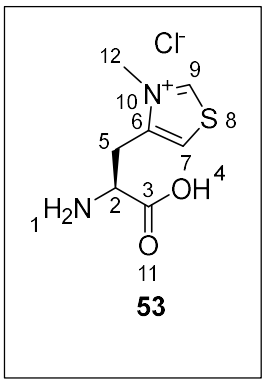
A solution of iodomethane (747 µL, 12 mmol) in acetonitrile (2 mL) was added to product of first step, 4-methyl-5-[2-(phenylmethoxy)ethyl]-thiazole (2.79 g, 12 mmol). The reaction mixture was stirred at 90 °C in microwave reactor for 4 h then at room temperature for 12 h. The solvent was then removed under reduced pressure. The product was washed with diethyl ether (10 mL) and dried under reduced pressure to give the product as a brown oil (Yield 2.6 g, 51%) ^{1}H NMR (500 MHz, Chloroform-*d*) δ 10.64 (s, 1H), 7.33 – 7.27 (m, 3H), 7.25 – 7.20 (m, 2H), 4.49 (s, 2H), 4.26 (s, 3H), 3.68 (t, *J* = 5.5 Hz, 2H), 3.09 (t, *J* = 5.5 Hz, 2H), 2.47 (s, 3H). In agreement in published data[111]

6.3.11 3-methyl-L-4-thiazolylalanine (L-Me4ThzA)



52

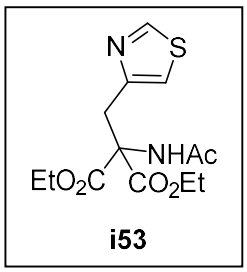
Intermediate 52 N-Boc-3-methyl-4-L-alaninethiazolium iodide: MeI (0.5 mL, 8 x10^{-3} moles) was added to a mixture of N-boc-L-4-thiazolylalanine (136 mg, 5 x10^{-4} moles), $NaHCO_3$ (126 mg, 1.5 x10^{-3} moles) and MeCN (4 mL) and stirred at 40°C for 3 days. The reaction was filtered under gravity and washed with MeCN (10 mL). The mixture was concentrated under reduced pressure and the crude product was purified by column chromatography with 19:1 DCM/MeOH. The product was eluted with 9:1 DCM/MeOH and concentrated under reduced pressure to produce a yellow solid in 36 % yield (73 mg). NMR analysis was performed in MeOD to confirm N-Boc-3-methyl-4-L-alaninethiazolium iodide formation. NMR shows the methyl ester is also formed due to a peak at **δ** 3.98.^{1}H NMR (500 MHz, MeOD) δ 10.3 (s, 1H), 8.2 (s, 1H), 4.49 (s, 3H), 3.98 (s, 3H), 3.79-3.75 (m, 1H), 3.52 (m,2H) 1.57 (s, 9H) ppm.



53

Compound 53: HCl (5 mL, 6 M) was added to intermediate **52** (73 mg, 1.8 x10^{-4} moles) at 70°C with stirring for 18h. The reaction was concentrated under reduced pressure then freeze dried to give a 32% yield (13 mg) of a yellow solid. (12% yield over two steps). ^{1}H NMR (500 MHz, D_2O) δ 10.0 (s, 1H^{9}) 8.2 (s, 1H^{7}), 4.62 (t, *J*=7.5 Hz, 1H^{2}), 4.28 (s, 3H^{12}), 3.8-3.6 (m, 2H^{5}) ppm.^{2}H NMR (500 MHz, D_2O) δ 10.0 (s, 1H) ppm.^{13}C NMR (126 MHz, D_2O) δ 169.9 (1C^{3}), 159.9 (1C^{9}), 143.8 (1C^{6}), 124.1 (1C^{7}), 51.0 (1C^{2}), 40.5 (1C^{12}), 27.4 (1C^{5}) ppm.

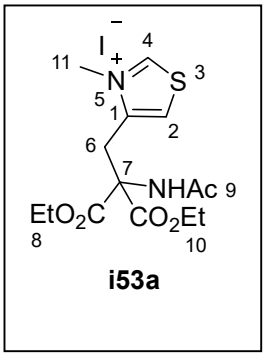
6.3.12 3-methyl-4-alaninethiazolium iodide (Me4ThzA)



i53

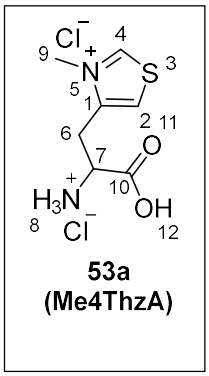
Intermediate i53 4-Chloromethyl thiazole hydrochloride (0.71 g, 6.6 mmol) was dissolved DCM (5 mL). A saturated solution of sodium bicarbonate (5 mL) was added. Insoluble compound was still present so additional DCM (2 mL), and saturated sodium carbonate solution (2 mL) was added. Organic phase was separated, dried with MgSO4, filtered and concentrated under reduced pressure. DMF (10 mL) was added to NaH (246 mg, 6.6 mmol: 60% in mineral oil) was at 0 °C under argon. Diethylacetamidomalonate (1.43 g, 6.6 mmol) was dissolved in DMF (10 mL) then added dropwise to NaH solution and stirred for 30 min. Neutralised 4-Chloromethyl thiazole was dissolved in DMF (10 mL) and added to the solution, which was stirred overnight at room temperature. DMF was removed

under reduced pressure and the diethyl (4-thiazolyl) methylacetamidomalonate product was extracted from brine (20 mL) with EtOAc (3 x 20 mL).The organic phase was dried with MgSO4, filtered and concentrated in vacuo. The product was purified by flash column chromatography (1:1 hexane ethyl acetate, RF = 0.4) and concentrated under reduced pressure Some residual diethylacetamidomalonate was observed in the NMR ^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.75 (d, *J* = 2.0 Hz, 1H), 7.12 – 7.08 (m, 1H), 6.74 (s, 1H), 4.36 – 4.25 (m, 5H), 3.92 – 3.89 (m, 2H), 2.10 (s, 1H), 1.99 (s, 3H), 1.34 (s, 1H), 1.31 (t, *J* = 7.1 Hz, 6H).



i53a

Intermediate i53a MeI (0.5 mL, 8 mmol) was added to Diethyl (4-thiazolyl) methylacetamidomalonate (intermediate i53,) dissolved in MeCN (4 mL) and stirred at 40°C for 3 days. Reaction was removed from heat and diethyl ether was added to see if the product crashes out. MeCN and diethyl ether were removed in vacuo. The crude product was purified by flash column chromatography with 9:1 DCM/MeOH and concentrated under reduced pressure to produce a yellow solid (850 mg, 27% yield over 2 steps). ^{1}H NMR (500 MHz, Chloroform-*d*) δ 10.45 (d, *J* = 2.6 Hz, 1H^{4}), 8.13 (d, *J* = 2.6 Hz, 1H^{2}), 7.46 (s, 1H^{amide}), 4.37 (qd, *J* = 7.1, 1.6 Hz, 4H^{8,10} (CH_2)), 4.24 (s, 3H^{11}), 3.93 (s, 2H^{6}), 2.19 (s, 3H^{9}), 1.31 (t, *J* = 7.1 Hz, 6H$^{(8,10}$ (CH_3)). ^{13}C NMR (126 MHz, DMSO) δ 170.60, 166.77, 160.90, 144.22, 124.89, 66.04, 63.07, 30.48, 22.56, 14.20.

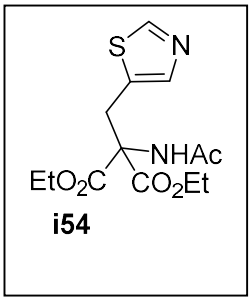


53a
(Me4ThzA)

Compound 53a Intermediate **i53a** (850 mg) was mixed with HCl (10 mL, 6M, aq.) and heated to 100°C for overnight. HCl was removed under reduced pressure (very sticky product dried under vacuum).Yield 89 % (414 mg)

^{1}H NMR (500 MHz, D_2O) δ 9.89 (d, *J* = 2.6 Hz, 1H^{9}), 8.08 (d, *J* = 2.8 Hz, 1H^{7}), 4.39 (t, *J* = 7.2 Hz, 1H^{2}), 4.16 (s, 3H^{12}), 3.62 (dd, *J* = 16.3, 7.6 Hz, 1H^{6a}), 3.51 (dd, *J* = 16.3, 6.9 Hz, 1H^{6b}).^{13}C NMR (126 MHz, D_2O) δ 170.43 (1C^{3}), 159.77 (1C^{9}), 144.03 (1C^{6}), 123.80 (1C^{7}), 123.71 (1C^{2}), 40.15 (1C^{12}), 27.48 (1C^{5}).

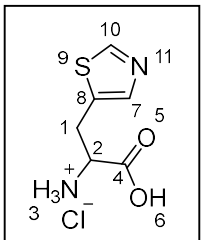
6.3.13 5-thiazolylalanine5ThzA



i54

5-hydroxymethyl thiazole (10 mmol, 1.15 g) was added to 40% HBr solution aq. (10 mL) and stirred for 4 hours at 100 °C. The solvent was evaporated and the resulting 5-Bromomethyl thiazole hydrogen bromide (2.6 g, 10 mmol) was dissolved in DCM (5 mL) and washed with saturated sodium bicarbonate aq. (5 mL). The organic extract was dried with MgSO4, filtered and concentrated *in vacuo*.

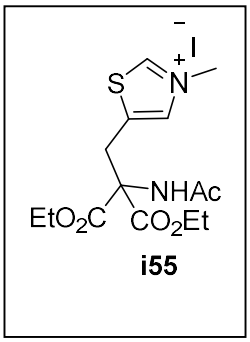
DMF (10 mL) was added to NaH (372 mg, 10 mmol: 60% in mineral oil) at 0 °C under argon. Diethylacetamidomalonate (2.17 g, 10 mmol) was dissolved in DMF (10 mL) added dropwise and stirred for 30 min. Neutralised 5-bromomethyl thiazole was dissolved in DMF (10 mL) and added dropwise to the solution, which was stirred overnight. DMF was removed *in vacuo* and the diethyl (5-thiazolyl) methylacetamidomalonate product extracted from brine (20 mL) with EtOAc (3 x 30 mL). The organic phase was dried over $MgSO_4$, filtered and concentrated *in vacuo*. The product was purified by flash column chromatography (1:1 hexane EtOAc Rf= 0.25) however DMF remained in the sample after drying. Yield ~64% ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.72, 7.98, 7.54, 7.26, 6.80, 4.21, 4.09, 3.91, 2.92, 2.83, 2.06, 2.05, 2.00, 1.23.



54 (5ThzA)

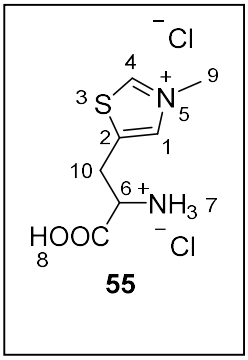
Compound 54 (5ThzA) Intermediate **i54** (240 mg, 0.52 mmol) was mixed with HCl (5 mL, 6M, aq.) and heated to 100°C for 18 hours. HCl was removed under reduced pressure. NMR shows residual DMSO in addition to the MeOH peak as the sample was dissolved in DMSO which could not be fully removed (500 MHz, Methanol-d_4) δ 9.50 (s, 1H^{10}), 8.09 (s, 1H^{7}), 4.44 (t, J = 6.0 Hz, 1H^{2}), 3.77 – 3.59 (m, 2H^{1}).

3-methyl-5-alaninethiazolium iodide (Me5ThzA)



i55

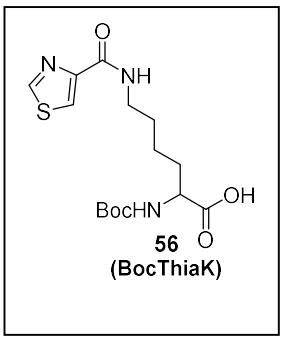
Intermediate i55 Diethyl (5-thiazolyl) methylacetamidomalonate **i54** (1 mmol, 314 mg, 1 eq) was dissolved in Acetonitrile (4 mL). Methyl Iodide (10 mmol, 622 µL ,10 eq) was added and the solution mixed in a round bottomed flask with a glass stopper for 4 days. Acetonitrile and excess MeI was removed under vacuum. Sample was dissolved in DCM and purified via silica column run with 9:1 DCM/MeOH. DMF remained after methylation. (~52% yield) ^{1}H NMR (500 MHz, MeOD) δ 8.22 (1H), 7.99 (DMF), 4.23 (7H), 3.92 (2H), 2.99 (DMF), 2.86(DMF), 2.08 (3H), 2.05(solvent), 1.27 (6H).



55

Compound 55 (Me5ThzA) Intermediate **i55** (62 mg, 0.24 mmol) was mixed with HCl (3 mL, 6M, aq.) and heated to 100°C for 18 hours. HCl was removed under reduced pressure. NMR shows residual DMSO in addition to the MeOH peak as the sample was dissolved in DMSO which could not be fully removed. 88% Yield (55 mg, 0.21 mmol). ^{1}H NMR (500 MHz, Methanol-d_4) δ 10.05 (s, 1H^{4}), 8.39 (s, 1H^{4}), 4.50 (t, 1H^{6}), 4.29 (s, 3H^{9}), 3.80 – 3.73 (m, 1H^{10a}), 3.65 (ddt, *J* = 15.7, 6.0, 1.0 Hz, 1H^{10b}). ^{13}C NMR (126 MHz, MeOD) δ 168.53 (1C^{8}), 159.48 (1C^{4}), 137.89 (1C^{2}), 137.26(1C^{1}), 52.35 (1C^{6}), 33.97(1C^{7}), 26.98 (1C^{10}).

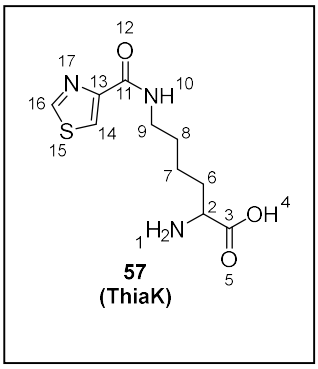
6.3.14 ThiaLys based UAAs



56
(BocThiaK)

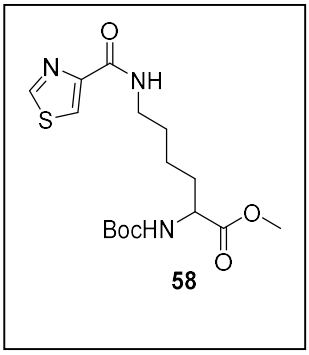
Intermediate 56: 4-Thiazolecarboxylic acid (1.29 g, 10 mmol) was refluxed with $SOCl_2$ (10 mL) for 90 min. Analysis via C^{13}NMR showed carbonylchloride had formed (AM-025-CP). Boc-L-lysine (1.87 g, 7.5 mmol) was dissolved in THF (9 mL) followed by 1M NaOH (19 mL). Crude 4-Thiazolecarbonyl chloride was dissolved in THF (10 mL) added dropwise (pH was checked to ensure it was basic). The reaction was stirred overnight at room temp then washed with Et_2O (37.5 mL). The aqueous phase was acidified to pH 4.0 with 1 M HCl and extracted with AcOEt (3 X 90 mL). The combined organic phases were dried over Na_2SO_4, filtered and solvent was removed under reduced pressure. A viscous oil formed and was analysed via H^{1}NMR (AM-025-UAACP). The product was purified by flash column chromatography (DCM, 10% methanol, 0.1% acetic acid) and concentrated under reduced pressure to give the product (2.01 g, 74% yield). ^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.78 (d, *J* = 2.1 Hz, 1H), 8.24 (d, *J* = 2.2 Hz, 1H), 5.31 (s, 1H), 3.50 (dt, *J* = 17.8, 6.9 Hz, 2H), 1.94 (d, *J* = 11.7 Hz, 1H), 1.79 (dddd, *J* = 13.8, 9.6, 7.9, 5.9 Hz, 1H), 1.72 – 1.63 (m, 2H), 1.59 – 1.48 (m, 2H), 1.46 (s, 9H).

Compound 57:



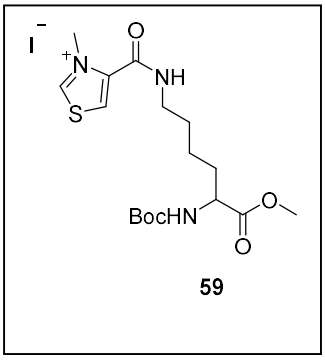
Intermediate **56** 300 mg was dissolved in aqueous HCl (10 mL, 6 M) 70 °C with stirring for 16 hours. Solvent and residual HCl were removed under reduced pressure to give the product in 84 % yield (206 mg). ^{1}H NMR (500 MHz, Methanol-d_4) δ 9.08 (d, *J* = 2.0 Hz, 1H^{16}), 8.29 (d, *J* = 2.0 Hz, 1H^{14}), 4.00 (t, *J* = 6.3 Hz, 1H^{2}), 3.46 (t, *J* = 7.0 Hz, 2H^{9}), 2.07 – 1.90 (m, 2H^{6}), 1.72 (p, *J* = 7.3 Hz, 2H^{8}), 1.63 – 1.49 (m, 2H^{7}). ^{13}C NMR (126 MHz, MeOD) δ 170.36 (1C^{3}), 161.93 (1C^{11}), 154.34(1C^{16}), 150.10 (1C^{13}), 123.48 (1C^{14}), 52.40 (1C^{2}), 38.44 (1C^{9}), 29.75(1C^{6}), 28.70(1C^{8}), 21.93 (1C^{7}).

Intermediate 58



Compound **56** (1.07g, 3 mmol, 1 equiv) was dissolved in MeCN 24 mL and DIPEA (1 mL, 6 mmol, 2 equiv) was added. MeI, (3 mL, 48 mmol, 16 equiv) was added and the reaction was heated to 40 °C with stirring for 3 days. MeCN was removed under reduced pressure to give a yellow solid (likely DIPEA salt Yield 2.83 g). To remove the DIPEA salt the sample was dissolved in DCM 30 mL then extracted with 1M HCl (3X 25 mL) followed by saturated sodium bicarbonate (25 mL). The organic phase was dried over $MgSO_4$, filtered and concentrated under reduced pressure to give a yellow oil (Yield 53 %, 589 mg) ^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.77 (d, *J* = 2.2 Hz, 1H), 8.19 (d, *J* = 2.2 Hz, 1H), 7.46 (s, 1H), 5.32 (s, 1H), 3.75 (s, 3H), 3.53 – 3.43 (m, 2H), 1.76 – 1.69 (m, 1H), 1.69 – 1.59 (m, 2H), 1.45 (s, 12H).

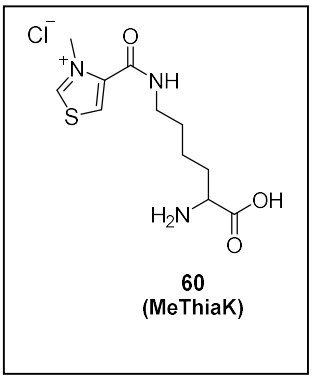
Intermediate 59



Intermediate **58** (589 mg, 1.58 mmol) was dissolved in Acetonitrile (2.4 mL). MeI (2.4 mL, 0.8 mL). was added the reaction was heated to 90 °C for 3 hours in a microwave reactor. Crude analysis showed the reaction had not gone to completion. An additional 800 µL MeI was added and heated to 90 °C for 3 hours. Residual MeI and solvent were removed under reduced pressure. Initially the product was purified by flash column chromatography in DCM 10% MeOH, but the separation was poor so as the project progressed the product was purified by automated flash column chromatography with a

Biotage® Sfär C18 Duo. The product was concentrated under reduced pressure Yield 66 %, 340 mg. ^{1}H NMR (500 MHz, Methanol-d_4) δ 10.13 (d, *J* = 2.7, 0.8 Hz, 1H), 8.65 – 8.59 (m, 1H), 4.40 (s, 3H), 4.16 – 4.09 (m, 1H), 3.73 (s, 3H), 3.42 (td, *J* = 7.0, 2.3 Hz, 2H), 1.85 (dd, *J* = 14.3, 7.2 Hz, 1H), 1.69 (dddd, *J* = 18.1, 10.9, 9.0, 4.8 Hz, 4H), 1.55 – 1.48 (m, 2H), 1.45 (s, 9H).

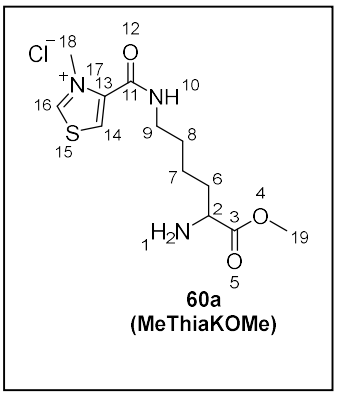
Compound 60



60
(MeThiaK)

Compound **59** 290 mg was dissolved in aqueous HCl (10 mL, 6 M) 70 °C with stirring for 16 hours. Solvent and residual HCl were removed under reduced pressure to give the product in 71 % yield NMR analysis highlighted impurities in the final product carried through from poor column conditions used to synthesise 58. ^{1}H NMR (601 MHz, DMSO-d_6) δ 10.21 (d, *J* = 2.5 Hz, 1H), 9.38 (t, *J* = 5.8 Hz, 1H), 8.77 (d, *J* = 2.5 Hz, 1H), 8.38 (d, *J* = 5.6 Hz, 3H), 7.93 (s, 1H), 4.34 (s, 1H), 4.29 (s, 3H), 3.93 – 3.83 (m, 3H), 3.27 (d, *J* = 7.3 Hz, 5H), 3.18 (d, *J* = 11.2 Hz, 3H), 2.82 (s, 1H), 2.80 – 2.75 (m, 1H), 2.57 (t, *J* = 5.0 Hz, 2H), 2.37 (q, *J* = 6.1 Hz, 1H), 1.89 – 1.74 (m, 4H), 1.64 – 1.52 (m, 4H), 1.49 (dtd, *J* = 17.9, 11.0, 9.1, 5.3 Hz, 2H), 1.45 – 1.34 (m, 2H).

Compound 60a



60a
(MeThiaKOMe)

Ethanol (7 eq, 1.8 mmol, 106 µL) was mixed with DCM (3 mL). Acetyl chloride (7 eq, 1.8 mmol, 128 µL) was mixed with DCM and added to the ethanol solution. Methylthiazoyl-Boc-L-lysine-OMe (1 eq, 0.26 mmol, 100 mg) was dissolved in DCM (2 mL) at 0° C and the acetyl chloride solution was added over 5 min. The solution was warmed to room temp and left to stir for 5 h. Solid precipitate was filtered and washed with DCM (5 mL) and dried under vacuum to give an off white amorphous solid (Yield 73 %, 62 mg) ^{1}H NMR (500 MHz, DMSO-d_6) δ 10.21 (d, *J* = 2.5 Hz, 1H^{16}), 9.39 (t, *J* = 5.6 Hz, 1H^{10}), 8.77 (d, *J* = 2.5 Hz, 1H^{14}), 8.56 (d, *J* = 5.4 Hz, 2H^{1}), 5.76 (DCM), 4.29 (s, 3H^{18}), 4.02 (q, *J* = 5.8 Hz, 1H^{2}), 3.76 (s, 3H^{19}), 3.27 (q, *J* = 6.5 Hz, 2H^{9}), 1.90 – 1.77 (m, 2H^{8}), 1.57 (dq, *J* = 14.0, 6.6 Hz, 2H^{7}), 1.51 – 1.41 (m, 1H^{6b}), 1.42 – 1.31 (m, 1H^{6a}). 13C NMR (126 MHz, DMSO) δ 170.44 (1C^{3}), 163.15(1C^{16}), 157.17 (1C^{11}), 141.05 (1C^{13}), 128.90 (1C^{14}), 55.38 (1C^{3}), 53.28 (1C^{19}), 52.27 (1C^{2}), 42.71 (1C^{18}), 30.02 (1C^{8}), 28.38(1C^{7}), 22.07(1C^{6}).

6.4 Protein structural prediction and Ligand Docking

Strutural predictions were created using ColabFold via AlphaFold2 using MMseqs2[138] and relaxed using amber. Docking was performed with AutoDock Vina (v1.1.2)[139, 140]. Ligands and receptors were prepared in AutoDockTools. Grid boxes were chosen by visual insection. CASTp 3.0[185] was used to analyse the cavities in each structure. Chem3D and Biovia discovery studio visualizer were used to prepare ligands before converting them to PDBQT files using AutoDock tools. Visual inspection of outputs was performed using Pymol (v2.4.6).

6.5 HPLC Reactions

6.5.1 Intramolecular Stetter reaction HPLC.

To analyse the concentration of product formed the samples were run on a Luna 5u C18 RP-Phenomenex column (100 Å, 250 x 4.6 mm) using an isocratic elution at 45% MeCN 0.1% TFA at 1 mL/min over 16 min, 30 °C. A UV/visible detector (SPD-20A) was used to monitor the reactions at 254 nm and 210 nm. Data is reported at 254 nm.

To analyse the enantio-selectivity of the reaction, the samples were run on CHIRALPAK® IB N-5 analytical column (5 µm, 250 × 4.6 mm) using an isocratic method for 30 mins, 40% MeCN, 60% 50 mM Kphos pH2 at 1 mL/min. A UV/visible detector (SPD-20A) was used to monitor the reactions at 254 nm and 210 nm. Data is reported at 254 nm.

6.5.2 Intramolecular Stetter reactions using UAAs and functionalisation reagents

Stock solutions of Starting material **18** (25 mM in DMSO) and each catalyst (25 mM in MeOH) were prepared. Reactions (200 µL) were set up using 2.5 mM starting material **18**, 2.5 mM catalyst, 15 mM Et_3N in methanol or water. Reaction plate was sealed and incubated at 30 °C, 18 hours, 250 rpm. Reactions were quenched with 200 µL MeCN containing 1 mM Coumarin as an internal standard, centrifuges 10 min 1000 rpm then analysed via HPLC.

6.5.3 Benzoin condensation with UAAs

Thiamine HCl or Me4ThzA (24 mM) were dissolved in each of the following solvents/buffers (500 µL) methanol, water, 50 mM HEPES pH 7, 50 mM HEPES pH 8 or 50 mM CAPS pH 9. 2-7 equiv Et_3N was added followed by Benzaldehyde (24 mM). The reactions were incubated at 30 °C 250 rpm for 24 hours. Reactions were quenched with an equal volume of MeCN with 0.1 % TFA and analysed by HPLC using a Luna 5u C18 RP-Phenomenex column (100 Å, 250 x 4.6 mm) and a gradient from 40- 60 % MeCN 0.1% TFA, flow rate 1 mL/min over 20 mins.

6.5.4 Intramolecular Stetter reactions using PfBAL

Purified PfBAL (35 µM) or pfBAL lysate (600 µL prepared from 3.5 mg WCW/mL buffer) were added to $MgSO_4$ (2.5 mM), TPP (0.15 mM) and starting material **18** (2.5 mM). Stock solutions of each reaction component were prepared in 50 mM Kphos pH7 with the exception of **18** which was prepared as a 10X stock in DMSO. Reactions were topped up to 1 mL 50 mM Kphos pH7 and incubated from 2-24 hours at 30 °C, 250 rpm. Reactions were quenched with an equal volume of MeCN and analysed via HPLC using the method described on 6.5.1.

6.5.5 Aqueous Stetter reaction methods

Intermolecular Stetter reaction

Catalyst **50** (0.3 mmol, 112.5 mg) was added to a round bottom flask with water (600 µL). DBU (0.3 mmol, 45 µL) was added followed by benzadlehyde (1.2 mmol, 122 µL) and chalcone (1 mmol). The solution was stirred 20 h at 50°C then concentrated under reduced pressure. The crude product was purified by flash chromatograph (hexane/ethyl acetate 9:1).

Intramolecular Stetter reaction

HPLC analysis of an aqueous intramolecular Stetter reaction catalysed by **50** was carried out on a 200 µL scale in water catalyst **50** (a. 100 µM b. 2.5 mM from 10 X stock in DMSO), starting material **18** (2.5 mM from 10 X stock in DMSO), reactions were run both with and without DBU (equimolar to catalyst) for 18 hours at 30 °C, 250 rpm. Reactions were quenched with an equal volume of acetonitrile containing 1 mM coumarin as an internal standard and analysed by HPLC using the method described on 6.5.1.

6.5.6 Intramolecular Stetter reactions using functionalised protein

Protein was buffer exchanged into reaction buffer (50 mM HEPES pH8) using a pd10 column (Cytiva). Starting material **18** (10 µL, 25 mM in DMSO) and functionalised protein (final concentration 100 µM) were added to a 1.5 mL Eppendorf tube. Reaction buffer was added to a final volume of 100 µL. Each reaction was set up in triplicate. Reactions were incubated in a thermoshaker (Grant Instruments™ PCMT) at 30°C, 250 rpm for 18 hours as standard conditions (incubation times were varied for time dependant studies). Reactions were quenched with 100 µL MeCN containing 0.1% TFA and 1 mM Coumarin as an internal standard. Reactions were centrifuged to remove precipitated protein and supernatant was analysed by HPLC. To analyse the concentration of product formed the samples were run on a Luna 5u C18 RP-Phenomenex column (100 Å, 250 x 4.6 mm) using an isocratic elution at 45% MeCN 0.1% TFA at 1 mL/min over 16 min, 30 °C. To analyse the enantio selectivity of the

reaction, the samples were run on CHIRALPAK® IB N-5 analytical column (5 µm, 250 × 4.6 mm) using an isocratic method for 30 mins, 40% MeCN, 60% 50 mM Kphos pH2 at 1 mL/min.

Reaction screening with different buffers followed the same procedure as above but alternative reaction buffers: 50 mM Mes pH 6, 50 mM HEPES pH8, 50 mM CAPS pH9 were used.

References

1. Sheldon, R. A.; Pereira, P. C., Biocatalysis engineering: the big picture. *Chem. Soc. Rev.* **2017,** *46* (10), 2678-2691.
2. Sheldon, R. A.; Brady, D.; Bode, M. L., The Hitchhiker's guide to biocatalysis: recent advances in the use of enzymes in organic synthesis. *Chem. Sci.* **2020,** *11* (10), 2587-2605.
3. Arnold, F. H., Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem. Int. Ed. Engl.* **2018,** *57* (16), 4143-4148.
4. Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K., Engineering the third wave of biocatalysis. *Nature* **2012,** *485* (7397), 185-94.
5. Hammer, S. C.; Knight, A. M.; Arnold, F. H., Design and evolution of enzymes for non-natural chemistry. *Current Opinion in Green and Sustainable Chemistry* **2017,** *7*, 23-30.
6. Bornscheuer, U. T., The fourth wave of biocatalysis is approaching. *Philos Trans A Math Phys Eng Sci* **2018,** *376* (2110).
7. Lovelock, S. L.; Crawshaw, R.; Basler, S.; Levy, C.; Baker, D.; Hilvert, D.; Green, A. P., The road to fully programmable protein catalysis. *Nature* **2022,** *606* (7912), 49-58.
8. Woolfson, D. N., A Brief History of De Novo Protein Design: Minimal, Rational, and Computational. *J. Mol. Biol.* **2021,** *433* (20), 167160.
9. Cutlan, R.; De Rose, S.; Isupov, M. N.; Littlechild, J. A.; Harmer, N. J., Using enzyme cascades in biocatalysis: Highlight on transaminases and carboxylic acid reductases. *Biochim Biophys Acta Proteins Proteom* **2020,** *1868* (2), 140322.
10. Ruales-Salcedo, A. V.; Higuita, J. C.; Fontalvo, J.; Woodley, J. M., Design of enzymatic cascade processes for the production of low-priced chemicals. *Z Naturforsch C J Biosci* **2019,** *74* (3-4), 77-84.
11. Huffman, M. A.; Fryszkowska, A.; Alvizo, O.; Borra-Garske, M.; Campos, K. R.; Canada, K. A.; Devine, P. N.; Duan, D.; Forstater, J. H.; Grosser, S. T.; Halsey, H. M.; Hughes, G. J.; Jo, J.; Joyce, L. A.; Kolev, J. N.; Liang, J.; Maloney, K. M.; Mann, B. F.; Marshall, N. M.; McLaughlin, M.; Moore, J. C.; Murphy, G. S.; Nawrat, C. C.; Nazor, J.; Novick, S.; Patel, N. R.; Rodriguez-Granillo, A.; Robaire, S. A.; Sherer, E. C.; Truppo, M. D.; Whittaker, A. M.; Verma, D.; Xiao, L.; Xu, Y.; Yang, H., Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science* **2019,** *366* (6470), 1255-1259.
12. Hartley, C. J.; Williams, C. C.; Scoble, J. A.; Churches, Q. I.; North, A.; French, N. G.; Nebl, T.; Coia, G.; Warden, A. C.; Simpson, G.; Frazer, A. R.; Jensen, C. N.; Turner, N. J.; Scott, C., Engineered enzymes that retain and regenerate their cofactors enable continuous-flow biocatalysis. *Nature Catalysis* **2019,** *2* (11), 1006-1015.
13. Quin, M. B.; Wallin, K. K.; Zhang, G.; Schmidt-Dannert, C., Spatial organization of multi-enzyme biocatalytic cascades. *Org. Biomol. Chem.* **2017,** *15* (20), 4260-4271.
14. Finnigan, W.; Thomas, A.; Cromar, H.; Gough, B.; Snajdrova, R.; Adams, J. P.; Littlechild, J. A.; Harmer, N. J., Characterization of Carboxylic Acid Reductases as Enzymes in the Toolbox for Synthetic Chemistry. *ChemCatChem* **2017,** *9* (6), 1005-1017.
15. Vornholt, T.; Jeschek, M., The Quest for Xenobiotic Enzymes: From New Enzymes for Chemistry to a Novel Chemistry of Life. *ChemBioChem* **2020,** *21* (16), 2241-2249.
16. Wiltschi, B.; Cernava, T.; Dennig, A.; Galindo Casas, M.; Geier, M.; Gruber, S.; Haberbauer, M.; Heidinger, P.; Herrero Acero, E.; Kratzer, R.; Luley-Goedl, C.; Muller, C. A.; Pitzer, J.; Ribitsch, D.; Sauer, M.; Schmolzer, K.; Schnitzhofer, W.; Sensen, C. W.; Soh, J.; Steiner, K.; Winkler, C. K.; Winkler, M.; Wriessnegger, T., Enzymes revolutionize the bioproduction of value-added compounds: From enzyme discovery to special applications. *Biotechnol. Adv.* **2020,** *40*, 107520.
17. Korendovych, I. V., Rational and Semirational Protein Design. *Methods Mol. Biol.* **2018,** *1685*, 15-23.

18. Ali, M.; Ishqi, H. M.; Husain, Q., Enzyme engineering: Reshaping the biocatalytic functions. *Biotechnol. Bioeng.* **2020,** *117* (6), 1877-1894.
19. Cheng, F.; Zhu, L.; Schwaneberg, U., Directed evolution 2.0: improving and deciphering enzyme properties. *Chem. Commun. (Camb.)* **2015,** *51* (48), 9760-72.
20. Baumann, T.; Hauf, M.; Richter, F.; Albers, S.; Moglich, A.; Ignatova, Z.; Budisa, N., Computational Aminoacyl-tRNA Synthetase Library Design for Photocaged Tyrosine. *Int. J. Mol. Sci.* **2019,** *20* (9), 2343.
21. Markel, U.; Essani, K. D.; Besirlioglu, V.; Schiffels, J.; Streit, W. R.; Schwaneberg, U., Advances in ultrahigh-throughput screening for directed enzyme evolution. *Chem. Soc. Rev.* **2020,** *49* (1), 233-262.
22. Lutz, S., Beyond directed evolution--semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **2010,** *21* (6), 734-43.
23. Andrews, F. H.; McLeish, M. J., Using site-saturation mutagenesis to explore mechanism and substrate specificity in thiamin diphosphate-dependent enzymes. *FEBS J.* **2013,** *280* (24), 6395-411.
24. Chen, K.; Arnold, F. H., Engineering new catalytic activities in enzymes. *Nature Catalysis* **2020,** *3* (3), 203-213.
25. Bjelic, S.; Nivon, L. G.; Celebi-Olcum, N.; Kiss, G.; Rosewall, C. F.; Lovick, H. M.; Ingalls, E. L.; Gallaher, J. L.; Seetharaman, J.; Lew, S.; Montelione, G. T.; Hunt, J. F.; Michael, F. E.; Houk, K. N.; Baker, D., Computational design of enone-binding proteins with catalytic activity for the Morita-Baylis-Hillman reaction. *ACS Chem. Biol.* **2013,** *8* (4), 749-57.
26. Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D., Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **2010,** *329* (5989), 309-13.
27. Hanreich, S.; Bonandi, E.; Drienovska, I., Design of Artificial Enzymes: Insights into Protein Scaffolds. *ChemBioChem* **2023,** *24* (6), e202200566.
28. Bunzel, H. A.; Anderson, J. L. R.; Mulholland, A. J., Designing better enzymes: Insights from directed evolution. *Curr. Opin. Struct. Biol.* **2021,** *67*, 212-218.
29. Jia, Z. J.; Gao, S.; Arnold, F. H., Enzymatic Primary Amination of Benzylic and Allylic C(sp(3))-H Bonds. *J. Am. Chem. Soc.* **2020,** *142* (23), 10279-10283.
30. Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H., Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **2013,** *339* (6117), 307-10.
31. Schafer, J. W.; Zoi, I.; Antoniou, D.; Schwartz, S. D., Optimization of the Turnover in Artificial Enzymes via Directed Evolution Results in the Coupling of Protein Dynamics to Chemistry. *J. Am. Chem. Soc.* **2019,** *141* (26), 10431-10439.
32. Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D., Kemp elimination catalysts by computational enzyme design. *Nature* **2008,** *453* (7192), 190-5.
33. Korendovych, I. V.; DeGrado, W. F., De novo protein design, a retrospective. *Q. Rev. Biophys.* **2020,** *53*, e3.
34. Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Rothlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F., 3rd; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D., De novo computational design of retro-aldol enzymes. *Science* **2008,** *319* (5868), 1387-91.
35. Cui, Y. L.; Sun, J. Y.; Wu, B., Computational enzyme redesign: large jumps in function. *Trends in Chemistry* **2022,** *4* (5), 409-419.

36. Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R.; Grutter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D., Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **2013,** *503* (7476), 418-21.
37. Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P. R.; Baker, D.; Griffiths, A. D.; Hilvert, D., Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **2017,** *9* (1), 50-56.
38. Burke, A. J.; Lovelock, S. L.; Frese, A.; Crawshaw, R.; Ortmayer, M.; Dunstan, M.; Levy, C.; Green, A. P., Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature* **2019,** *570* (7760), 219-223.
39. Watson, J. L.; Juergens, D.; Bennett, N. R.; Trippe, B. L.; Yim, J.; Eisenach, H. E.; Ahern, W.; Borst, A. J.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Hanikel, N.; Pellock, S. J.; Courbet, A.; Sheffler, W.; Wang, J.; Venkatesh, P.; Sappington, I.; Torres, S. V.; Lauko, A.; De Bortoli, V.; Mathieu, E.; Ovchinnikov, S.; Barzilay, R.; Jaakkola, T. S.; DiMaio, F.; Baek, M.; Baker, D., De novo design of protein structure and function with RFdiffusion. *Nature* **2023,** *620* (7976), 1089-1100.
40. Broom, A.; Rakotoharisoa, R. V.; Thompson, M. C.; Zarifi, N.; Nguyen, E.; Mukhametzhanov, N.; Liu, L.; Fraser, J. S.; Chica, R. A., Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico. *Nat Commun* **2020,** *11* (1), 4808.
41. Drienovská, I.; Roelfes, G., Expanding the enzyme universe with genetically encoded unnatural amino acids. *Nature Catalysis* **2020,** *3* (3), 193-202.
42. Zhao, J.; Burke, A. J.; Green, A. P., Enzymes with noncanonical amino acids. *Curr. Opin. Chem. Biol.* **2020,** *55*, 136-144.
43. Drienovska, I.; Mayer, C.; Dulson, C.; Roelfes, G., A designer enzyme for hydrazone and oxime formation featuring an unnatural catalytic aniline residue. *Nat. Chem.* **2018,** *10* (9), 946-952.
44. Fischer, J. D.; Holliday, G. L.; Thornton, J. M., The CoFactor database: organic cofactors in enzyme catalysis. *Bioinformatics* **2010,** *26* (19), 2496-7.
45. Richter, M., Functional diversity of organic molecule enzyme cofactors. *Nat. Prod. Rep.* **2013,** *30* (10), 1324-45.
46. Prier, C. K.; Arnold, F. H., Chemomimetic biocatalysis: exploiting the synthetic potential of cofactor-dependent enzymes to create new catalysts. *J. Am. Chem. Soc.* **2015,** *137* (44), 13992-4006.
47. Chenault, H. K.; Simon, E. S.; Whitesides, G. M., Cofactor regeneration for enzyme-catalysed synthesis. *Biotechnol. Genet. Eng. Rev.* **1988,** *6* (1), 221-70.
48. Kara, S.; Schrittwieser, J. H.; Hollmann, F.; Ansorge-Schumacher, M. B., Recent trends and novel concepts in cofactor-dependent biotransformations. *Appl. Microbiol. Biotechnol.* **2014,** *98* (4), 1517-29.
49. Wang, M.; Chen, B.; Fang, Y.; Tan, T., Cofactor engineering for more efficient production of chemicals and biofuels. *Biotechnol. Adv.* **2017,** *35* (8), 1032-1039.
50. Black, W. B.; Zhang, L.; Mak, W. S.; Maxel, S.; Cui, Y.; King, E.; Fong, B.; Sanchez Martinez, A.; Siegel, J. B.; Li, H., Engineering a nicotinamide mononucleotide redox cofactor system for biocatalysis. *Nat. Chem. Biol.* **2020,** *16* (1), 87-94.
51. Martinoli, C.; Dudek, H. M.; Orru, R.; Edmondson, D. E.; Fraaije, M. W.; Mattevi, A., Beyond the Protein Matrix: Probing Cofactor Variants in a Baeyer-Villiger Oxygenation Reaction. *ACS Catal* **2013,** *3* (12), 3058-3062.
52. Tylicki, A.; Lotowski, Z.; Siemieniuk, M.; Ratkiewicz, A., Thiamine and selected thiamine antivitamins - biological activity and methods of synthesis. *Biosci. Rep.* **2018,** *38* (1), BSR20171148.

53. Bunik, V. I.; Tylicki, A.; Lukashev, N. V., Thiamin diphosphate-dependent enzymes: from enzymology to metabolic regulation, drug design and disease models. *FEBS J.* **2013,** *280* (24), 6412-42.
54. Frank, R. A.; Leeper, F. J.; Luisi, B. F., Structure, mechanism and catalytic duality of thiamine-dependent enzymes. *Cell. Mol. Life Sci.* **2007,** *64* (7-8), 892-905.
55. Clapés, P., Enzymatic CC Bond Formation. In *Organic Synthesis Using Biocatalysis*, Goswami, A.; Stewart, J. D., Eds. Academic Press: 2016; pp 285-337.
56. Breslow, R., On the Mechanism of Thiamine Action. IV.1 Evidence from Studies on Model Systems. *J. Am. Chem. Soc.* **2002,** *80* (14), 3719-3726.
57. Prajapati, S.; Rabe von Pappenheim, F.; Tittmann, K., Frontiers in the enzymology of thiamin diphosphate-dependent enzymes. *Curr. Opin. Struct. Biol.* **2022,** *76*, 102441.
58. Enders, D.; Niemeier, O.; Henseler, A., Organocatalysis by N-heterocyclic carbenes. *Chem. Rev.* **2007,** *107* (12), 5606-55.
59. de Alaniz, J. R.; Rovis, T., The Catalytic Asymmetric Intramolecular Stetter Reaction. *Synlett* **2009,** *8* (2009), 1189-1207.
60. Alvarez, F. J.; Ermer, J.; Hubner, G.; Schellenberger, A.; Schowen, R. L., Catalytic Power of Pyruvate Decarboxylase - Rate-Limiting Events and Microscopic Rate Constants from Primary Carbon and Secondary Hydrogen Isotope Effects. *J. Am. Chem. Soc.* **1991,** *113* (22), 8402-8409.
61. Vogel, C.; Pleiss, J., The modular structure of ThDP-dependent enzymes. *Proteins* **2014,** *82* (10), 2523-37.
62. Mosbacher, T. G.; Mueller, M.; Schulz, G. E., Structure and mechanism of the ThDP-dependent benzaldehyde lyase from Pseudomonas fluorescens. *FEBS J.* **2005,** *272* (23), 6067-76.
63. Medina, F. E.; Prejanò, M., Water Molecules Allow the Intramolecular Activation of the Thiamine Di-Phosphate Cofactor in Human Transketolase: Mechanistic Insights into a Famous Proposal. *Acs Catalysis* **2021,** *11* (7), 4136-4145.
64. Duggleby, R. G., Domain relationships in thiamine diphosphate-dependent enzymes. *Acc. Chem. Res.* **2006,** *39* (8), 550-7.
65. James, P.; Isupov, M. N.; De Rose, S. A.; Sayer, C.; Cole, I. S.; Littlechild, J. A., A 'Split-Gene' Transketolase From the Hyper-Thermophilic Bacterium Carboxydothermus hydrogenoformans: Structure and Biochemical Characterization. *Front. Microbiol.* **2020,** *11*, 592353.
66. Hailes, H. C.; Rother, D.; Muller, M.; Westphal, R.; Ward, J. M.; Pleiss, J.; Vogel, C.; Pohl, M., Engineering stereoselectivity of ThDP-dependent enzymes. *FEBS J.* **2013,** *280* (24), 6374-94.
67. Dunkelmann, P.; Kolter-Jung, D.; Nitsche, A.; Demir, A. S.; Siegert, P.; Lingen, B.; Baumann, M.; Pohl, M.; Muller, M., Development of a donor-acceptor concept for enzymatic cross-coupling reactions of aldehydes: the first asymmetric cross-benzoin condensation. *J. Am. Chem. Soc.* **2002,** *124* (41), 12084-5.
68. Lingen, B.; Grotzinger, J.; Kolter, D.; Kula, M. R.; Pohl, M., Improving the carboligase activity of benzoylformate decarboxylase from Pseudomonas putida by a combination of directed evolution and site-directed mutagenesis. *Protein Eng.* **2002,** *15* (7), 585-93.
69. Yu, H.; Hernandez Lopez, R. I.; Steadman, D.; Mendez-Sanchez, D.; Higson, S.; Cazares-Korner, A.; Sheppard, T. D.; Ward, J. M.; Hailes, H. C.; Dalby, P. A., Engineering transketolase to accept both unnatural donor and acceptor substrates and produce alpha-hydroxyketones. *FEBS J.* **2020,** *287* (9), 1758-1776.
70. Bhasin, M.; Billinsky, J. L.; Palmer, D. R., Steady-state kinetics and molecular evolution of Escherichia coli MenD [(1R,6R)-2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase], an anomalous thiamin diphosphate-dependent decarboxylase-carboligase. *Biochemistry* **2003,** *42* (46), 13496-504.

71. Dresen, C.; Richter, M.; Pohl, M.; Ludeke, S.; Muller, M., The enzymatic asymmetric conjugate umpolung reaction. *Angew. Chem. Int. Ed. Engl.* **2010,** *49* (37), 6600-3.
72. Muller, M.; Sprenger, G. A.; Pohl, M., CC bond formation using ThDP-dependent lyases. *Curr. Opin. Chem. Biol.* **2013,** *17* (2), 261-70.
73. Lemmerer, M.; Schupp, M.; Kaiser, D.; Maulide, N., Synthetic approaches to 1,4-dicarbonyl compounds. *Nature Synthesis* **2022,** *1* (12), 923-935.
74. Kasparyan, E.; Richter, M.; Dresen, C.; Walter, L. S.; Fuchs, G.; Leeper, F. J.; Wacker, T.; Andrade, S. L.; Kolter, G.; Pohl, M.; Muller, M., Asymmetric Stetter reactions catalyzed by thiamine diphosphate-dependent enzymes. *Appl. Microbiol. Biotechnol.* **2014,** *98* (23), 9681-90.
75. Beigi, M.; Waltzer, S.; Zarei, M.; Muller, M., New Stetter reactions catalyzed by thiamine diphosphate dependent MenD from E. coli. *J. Biotechnol.* **2014,** *191*, 64-8.
76. Chen, X.; Wang, Z.; Lou, Y.; Peng, Y.; Zhu, Q.; Xu, J.; Wu, Q., Intramolecular Stereoselective Stetter Reaction Catalyzed by Benzaldehyde Lyase. *Angew. Chem. Int. Ed. Engl.* **2021,** *60* (17), 9326-9329.
77. Chen, J.; Gong, X.; Li, J.; Li, Y.; Ma, J.; Hou, C.; Zhao, G.; Yuan, W.; Zhao, B., Carbonyl catalysis enables a biomimetic asymmetric Mannich reaction. *Science* **2018,** *360* (6396), 1438-1442.
78. Chen, J.; Huang, Y., Asymmetric catalysis with N-heterocyclic carbenes as non-covalent chiral templates. *Nat Commun* **2014,** *5* (1), 3437.
79. Hopkinson, M. N.; Richter, C.; Schedler, M.; Glorius, F., An overview of N-heterocyclic carbenes. *Nature* **2014,** *510* (7506), 485-96.
80. Janssen-Muller, D.; Schlepphorst, C.; Glorius, F., Privileged chiral N-heterocyclic carbene ligands for asymmetric transition-metal catalysis. *Chem. Soc. Rev.* **2017,** *46* (16), 4845-4854.
81. Doddi, A.; Peters, M.; Tamm, M., N-Heterocyclic Carbene Adducts of Main Group Elements and Their Use as Ligands in Transition Metal Chemistry. *Chem. Rev.* **2019,** *119* (12), 6994-7112.
82. Holloczki, O., The Mechanism of N-Heterocyclic Carbene Organocatalysis through a Magnifying Glass. *Chemistry (Easton)* **2020,** *26* (22), 4885-4894.
83. Ukai, T.; Tanaka, R.; Dokawa, T. J. J. P. S. J., A new catalyst for acyloin condensation. **1943,** *63*, 296-300.
84. Sheehan, J. C.; Hunneman, D. H., Homogeneous Asymmetric Catalysis. *J. Am. Chem. Soc.* **1966,** *88* (15), 3666-+.
85. Flanigan, D. M.; Romanov-Michailidis, F.; White, N. A.; Rovis, T., Organocatalytic Reactions Enabled by N-Heterocyclic Carbenes. *Chem. Rev.* **2015,** *115* (17), 9307-87.
86. Stetter, H., Catalyzed Addition of Aldehydes to Activated Double Bonds—A New Synthetic Approach. *Angewandte Chemie International Edition in English* **2003,** *15* (11), 639-647.
87. Arduengo, A. J.; Harlow, R. L.; Kline, M., A Stable Crystalline Carbene. *J. Am. Chem. Soc.* **1991,** *113* (1), 361-363.
88. Regitz, M., Nucleophilic carbenes: An incredible renaissance. *Angewandte Chemie-International Edition* **1996,** *35* (7), 725-728.
89. Henrique Teles, J.; Melder, J. P.; Ebel, K.; Schneider, R.; Gehrer, E.; Harder, W.; Brode, S.; Enders, D.; Breuer, K.; Raabe, G., The Chemistry of Stable Carbenes. Part 2. Benzoin-type condensations of formaldehyde catalyzed by stable carbenes. *Helv. Chim. Acta* **2004,** *79* (1), 61-83.
90. Enders, D.; Breuer, K.; Raabe, G.; Runsink, J.; Teles, J. H.; Melder, J. P.; Ebel, K.; Brode, S., Preparation, Structure, and Reactivity of 1,3,4-Triphenyl-4,5-Dihydro-1h-1,2,4-Triazol-5-Ylidene, a New Stable Carbene. *Angewandte Chemie-International Edition in English* **1995,** *34* (9), 1021-1023.

91. Enders, D.; Breuer, K.; Runsink, J.; Teles, J. H., The First Asymmetric Intramolecular Stetter Reaction. Preliminary Communication. *Helv. Chim. Acta* **2004,** *79* (7), 1899-1902.
92. Enders, D.; Han, J.; Henseler, A., Asymmetric intermolecular Stetter reactions catalyzed by a novel triazolium derived N-heterocyclic carbene. *Chem. Commun. (Camb.)* **2008,** (34), 3989-91.
93. de Alaniz, J. R.; Kerr, M. S.; Moore, J. L.; Rovis, T., Scope of the asymmetric intramolecular stetter reaction catalyzed by chiral nucleophilic triazolinylidene carbenes. *J. Org. Chem.* **2008,** *73* (6), 2033-40.
94. Knight, R. L.; Leeper, F. J., Comparison of chiral thiazolium and triazolium salts as asymmetric catalysts for the benzoin condensation. *Journal of the Chemical Society-Perkin Transactions 1* **1998,** (12), 1891-1893.
95. Sharma, D.; Chatterjee, R.; Dhayalan, V.; Dandela, R., Recent Advances in Enantioselective Organocatalytic Reactions Enabled by N-Heterocyclic Carbenes (NHCs) Containing Triazolium Motifs. *Synthesis-Stuttgart* **2022,** *54* (19), 4129-4166.
96. Nelson, D. J.; Nolan, S. P., Quantifying and understanding the electronic properties of N-heterocyclic carbenes. *Chem. Soc. Rev.* **2013,** *42* (16), 6723-53.
97. Gaggero, N.; Pandini, S., Advances in chemoselective intermolecular cross-benzoin-type condensation reactions. *Org. Biomol. Chem.* **2017,** *15* (33), 6867-6887.
98. Benhamou, L.; Chardon, E.; Lavigne, G.; Bellemin-Laponnaz, S.; Cesar, V., Synthetic routes to N-heterocyclic carbene precursors. *Chem. Rev.* **2011,** *111* (4), 2705-33.
99. Wang, Z.; Xue, X. S.; Fu, Y.; Ji, P., Comprehensive Basicity Scales for N-Heterocyclic Carbenes in DMSO: Implications on the Stabilities of N-Heterocyclic Carbene and CO(2) Adducts. *Chem. Asian J.* **2020,** *15* (1), 169-181.
100. Higgins, E. M.; Sherwood, J. A.; Lindsay, A. G.; Armstrong, J.; Massey, R. S.; Alder, R. W.; O'Donoghue, A. C., pKas of the conjugate acids of N-heterocyclic carbenes in water. *Chem. Commun. (Camb.)* **2011,** *47* (5), 1559-61.
101. Wang, N.; Xu, J.; Lee, J. K., The importance of N-heterocyclic carbene basicity in organocatalysis. *Org. Biomol. Chem.* **2018,** *16* (37), 8230-8244.
102. Zhao, C. G.; Blaszczyk, S. A.; Wang, J. M., Asymmetric reactions of N-heterocyclic carbene (NHC)-based chiral acyl azoliums and azolium enolates. *Green Synthesis and Catalysis* **2021,** *2* (2), 198-215.
103. Yan, J.; Sun, R.; Shi, K.; Li, K.; Yang, L.; Zhong, G., N-Heterocyclic Carbene-Catalyzed Asymmetric Benzoin Reaction in Water. *J. Org. Chem.* **2018,** *83* (14), 7547-7552.
104. Bugaut, X.; Glorius, F., Organocatalytic umpolung: N-heterocyclic carbenes and beyond. *Chem. Soc. Rev.* **2012,** *41* (9), 3511-22.
105. Beigi, M.; Gauchenova, E.; Walter, L.; Waltzer, S.; Bonina, F.; Stillger, T.; Rother, D.; Pohl, M.; Muller, M., Regio- and Stereoselective Aliphatic-Aromatic Cross-Benzoin Reaction: Enzymatic Divergent Catalysis. *Chemistry (Easton)* **2016,** *22* (39), 13999-14005.
106. Heravi, M. M.; Zadsirjan, V., Recent advances in application of the Stetter reaction in heterocyclic chemistry. In *Adv. Heterocycl. Chem.*, Scriven, E. F. V.; Ramsden, C. A., Eds. Academic Press: 2023; Vol. 141, pp 127-178.
107. Stetter, H.; Schreckenberg, M., A New Method for Addition of Aldehydes to Activated Double Bonds. *Angewandte Chemie International Edition in English* **2003,** *12* (1), 81-81.
108. Heravi, M. M.; Zadsirjan, V.; Kafshdarzadeh, K.; Amiri, Z., Recent Advances in Stetter Reaction and Related Chemistry: An update. *Asian Journal of Organic Chemistry* **2020,** *9* (12), 1999-2034.
109. Ciganek, E., Esters of 2,3-Dihydro-3-Oxobenzofuran-2-Acetic Acid and 3,4-Dihydro-4-Oxo-2h-1-Benzopyran-3-Acetic Acid by Intramolecular Stetter Reactions. *Synthesis-Stuttgart* **1995,** *1995* (10), 1311-&.
110. Hsu, D. S.; Liang, S. P., NHC-Mediated Synthesis of Tricyclic Spirocarbocycles via an Intramolecular Stetter Reaction of Cyclic Enal-Enones. *J. Org. Chem.* **2020,** *85* (2), 1270-1278.

111. Debiais, M.; Hamoud, A.; Drain, R.; Barthelemy, P.; Desvergnes, V., Bio-inspired NHC-organocatalyzed Stetter reaction in aqueous conditions. *RSC Adv* **2020,** *10* (67), 40709-40718.
112. Zhao, H.; Foss, F. W., Jr.; Breslow, R., Artificial enzymes with thiazolium and imidazolium coenzyme mimics. *J. Am. Chem. Soc.* **2008,** *130* (38), 12590-1.
113. Mennen, S. M.; Blank, J. T.; Tran-Dube, M. B.; Imbriglio, J. E.; Miller, S. J., A peptide-catalyzed asymmetric Stetter reaction. *Chem. Commun. (Camb.)* **2005,** (2), 195-7.
114. Notz, W.; Tanaka, F.; Barbas, C. F., 3rd, Enamine-based organocatalysis with proline and diamines: the development of direct catalytic asymmetric Aldol, Mannich, Michael, and Diels-alder reactions. *Acc. Chem. Res.* **2004,** *37* (8), 580-91.
115. List, B.; Lerner, R. A.; Barbas, C. F., Proline-catalyzed direct asymmetric aldol reactions. *J. Am. Chem. Soc.* **2000,** *122* (10), 2395-2396.
116. Cui, H. Y.; Stadtmüller, T. H. J.; Jiang, Q. J.; Jaeger, K. E.; Schwaneberg, U.; Davari, M. D., How to Engineer Organic Solvent Resistant Enzymes: Insights from Combined Molecular Dynamics and Directed Evolution Study. *Chemcatchem* **2020,** *12* (16), 4073-4083.
117. Groger, H.; Gallou, F.; Lipshutz, B. H., Where Chemocatalysis Meets Biocatalysis: In Water. *Chem. Rev.* **2023,** *123* (9), 5262-5296.
118. Nodling, A. R.; Santi, N.; Williams, T. L.; Tsai, Y. H.; Luk, L. Y. P., Enabling protein-hosted organocatalytic transformations. *RSC Adv* **2020,** *10* (27), 16147-16161.
119. Jarvis, A. G., Designer metalloenzymes for synthetic biology: Enzyme hybrids for catalysis. *Curr. Opin. Chem. Biol.* **2020,** *58*, 63-71.
120. Wilson, M. E.; Whitesides, G. M., Conversion of a Protein to a Homogeneous Asymmetric Hydrogenation Catalyst by Site-Specific Modification with a Diphosphinerhodium(I) Moiety. *J. Am. Chem. Soc.* **1978,** *100* (1), 306-307.
121. Rahimi, M.; Geertsema, E. M.; Miao, Y.; van der Meer, J. Y.; van den Bosch, T.; de Haan, P.; Zandvoort, E.; Poelarends, G. J., Inter- and intramolecular aldol reactions promiscuously catalyzed by a proline-based tautomerase. *Org. Biomol. Chem.* **2017,** *15* (13), 2809-2816.
122. Zandvoort, E.; Geertsema, E. M.; Baas, B. J.; Quax, W. J.; Poelarends, G. J., Bridging between organocatalysis and biocatalysis: asymmetric addition of acetaldehyde to beta-nitrostyrenes catalyzed by a promiscuous proline-based tautomerase. *Angew. Chem. Int. Ed. Engl.* **2012,** *51* (5), 1240-3.
123. Nodling, A. R.; Santi, N.; Castillo, R.; Lipka-Lloyd, M.; Jin, Y.; Morrill, L. C.; Swiderek, K.; Moliner, V.; Luk, L. Y. P., The role of streptavidin and its variants in catalysis by biotinylated secondary amines. *Org. Biomol. Chem.* **2021,** *19* (47), 10424-10431.
124. Nodling, A. R.; Swiderek, K.; Castillo, R.; Hall, J. W.; Angelastro, A.; Morrill, L. C.; Jin, Y.; Tsai, Y. H.; Moliner, V.; Luk, L. Y. P., Reactivity and Selectivity of Iminium Organocatalysis Improved by a Protein Host. *Angew. Chem. Int. Ed. Engl.* **2018,** *57* (38), 12478-12482.
125. Pellizzoni, M. M.; Schwizer, F.; Wood, C. W.; Sabatino, V.; Cotelle, Y.; Matile, S.; Woolfson, D. N.; Ward, T. R., Chimeric Streptavidins as Host Proteins for Artificial Metalloenzymes. *Acs Catalysis* **2018,** *8* (2), 1476-1484.
126. Garrabou, X.; Macdonald, D. S.; Hilvert, D., Chemoselective Henry Condensations Catalyzed by Artificial Carboligases. *Chemistry (Easton)* **2017,** *23* (25), 6001-6003.
127. Garrabou, X.; Macdonald, D. S.; Wicky, B. I. M.; Hilvert, D., Stereodivergent Evolution of Artificial Enzymes for the Michael Reaction. *Angew. Chem. Int. Ed. Engl.* **2018,** *57* (19), 5288-5291.
128. Suckling, C. J.; Zhu, L. M., Carbon-Carbon Bond Formation Mediated by Papain Chemically Modified by Thiazolium Salts. *Bioorg. Med. Chem. Lett.* **1993,** *3* (4), 531-534.
129. Wu, Z. P.; Hilvert, D., Selenosubtilisin as a Glutathione-Peroxidase Mimic. *J. Am. Chem. Soc.* **1990,** *112* (14), 5647-5648.

130. Ofori Atta, L.; Zhou, Z.; Roelfes, G., In Vivo Biocatalytic Cascades Featuring an Artificial-Enzyme-Catalysed New-to-Nature Reaction. *Angew. Chem. Int. Ed. Engl.* **2023,** *62* (1), e202214191.
131. Vachan, B. S.; Karuppasamy, M.; Vinoth, P.; Vivek Kumar, S.; Perumal, S.; Sridharan, V.; Menéndez, J. C., Proline and its Derivatives as Organocatalysts for Multi- Component Reactions in Aqueous Media: Synergic Pathways to the Green Synthesis of Heterocycles. *Adv. Synth. Catal.* **2019,** *362* (1), 87-110.
132. Guo, C.; Saifuddin, M.; Saravanan, T.; Sharifi, M.; Poelarends, G. J., Biocatalytic Asymmetric Michael Additions of Nitromethane to alpha,beta-Unsaturated Aldehydes via Enzyme-bound Iminium Ion Intermediates. *ACS Catal* **2019,** *9* (5), 4369-4373.
133. Lukesch, M. S.; Pavkov-Keller, T.; Gruber, K.; Zangger, K.; Wiltschi, B., Substituting the catalytic proline of 4-oxalocrotonate tautomerase with non-canonical analogues reveals a finely tuned catalytic system. *Sci. Rep.* **2019,** *9* (1), 2697.
134. Crawshaw, R.; Crossley, A. E.; Johannissen, L.; Burke, A. J.; Hay, S.; Levy, C.; Baker, D.; Lovelock, S. L.; Green, A. P., Engineering an efficient and enantioselective enzyme for the Morita-Baylis-Hillman reaction. *Nat. Chem.* **2022,** *14* (3), 313-320.
135. Hilvert, D.; Breslow, R., Functionalized Cyclodextrins as Holoenzyme Mimics of Thiamine-Dependent Enzymes. *Bioorg. Chem.* **1984,** *12* (3), 206-220.
136. Breslow, R.; Kool, E., A γ-cyclodextrin thiazolium salt holoenzyme mimic for the benzoin condensation. *Tetrahedron Lett.* **1988,** *29* (14), 1635-1638.
137. Aitken, D. J.; Alijah, R.; Onyiriuka, S. O.; Suckling, C. J.; Wood, H. C. S.; Zhu, L. M., Approaches to Chemically Modified Enzymes as Synthetic Catalysts. *Journal of the Chemical Society-Perkin Transactions 1* **1993,** (5), 597-608.
138. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021,** *596* (7873), 583-589.
139. Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S., AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021,** *61* (8), 3891-3898.
140. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31* (2), 455-61.
141. Huang, P. S.; Boyken, S. E.; Baker, D., The coming of age of de novo protein design. *Nature* **2016,** *537* (7620), 320-7.
142. Korbeld, K. T.; Furst, M., Curse and Blessing of Non-Proteinogenic Parts in Computational Enzyme Engineering. *ChemBioChem* **2023,** *24* (12), e202300192.
143. Sreenilayam, G.; Moore, E. J.; Steck, V.; Fasan, R., Metal Substitution Modulates the Reactivity and Extends the Reaction Scope of Myoglobin Carbene Transfer Catalysts. *Adv. Synth. Catal.* **2017,** *359* (12), 2076-2089.
144. Liang, A. D.; Serrano-Plana, J.; Peterson, R. L.; Ward, T. R., Artificial Metalloenzymes Based on the Biotin-Streptavidin Technology: Enzymatic Cascades and Directed Evolution. *Acc. Chem. Res.* **2019,** *52* (3), 585-595.
145. Heinisch, T.; Ward, T. R., Artificial Metalloenzymes Based on the Biotin-Streptavidin Technology: Challenges and Opportunities. *Acc. Chem. Res.* **2016,** *49* (9), 1711-21.
146. Bos, J.; Fusetti, F.; Driessen, A. J.; Roelfes, G., Enantioselective artificial metalloenzymes by creation of a novel active site at the protein dimer interface. *Angew. Chem. Int. Ed. Engl.* **2012,** *51* (30), 7472-5.

147. Deuss, P. J.; Popa, G.; Slawin, A. M. Z.; Laan, W.; Kamer, P. C. J., Artificial Copper Enzymes for Asymmetric Diels-Alder Reactions. *Chemcatchem* **2013,** *5* (5), 1184-1191.
148. Boutureira, O.; Bernardes, G. J., Advances in chemical protein modification. *Chem. Rev.* **2015,** *115* (5), 2174-95.
149. Naowarojna, N.; Cheng, R.; Lopez, J.; Wong, C.; Qiao, L.; Liu, P., Chemical modifications of proteins and their applications in metalloenzyme studies. *Synth Syst Biotechnol* **2021,** *6* (1), 32-49.
150. Haque, M.; Forte, N.; Baker, J. R., Site-selective lysine conjugation methods and applications towards antibody-drug conjugates. *Chem. Commun. (Camb.)* **2021,** *57* (82), 10689-10702.
151. De Rosa, L.; Di Stasi, R.; Romanelli, A.; D'Andrea, L. D., Exploiting Protein N-Terminus for Site-Specific Bioconjugation. *Molecules* **2021,** *26* (12).
152. Bloom, S.; Liu, C.; Kolmel, D. K.; Qiao, J. X.; Zhang, Y.; Poss, M. A.; Ewing, W. R.; MacMillan, D. W. C., Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. *Nat. Chem.* **2018,** *10* (2), 205-211.
153. Reddy, N. C.; Kumar, M.; Molla, R.; Rai, V., Chemical methods for modification of proteins. *Org. Biomol. Chem.* **2020,** *18* (25), 4669-4691.
154. Griffiths, R. C.; Smith, F. R.; Li, D.; Wyatt, J.; Rogers, D. M.; Long, J. E.; Cusin, L. M. L.; Tighe, P. J.; Layfield, R.; Hirst, J. D.; Muller, M. M.; Mitchell, N. J., Cysteine-Selective Modification of Peptides and Proteins via Desulfurative C-C Bond Formation. *Chemistry (Easton)* **2023,** *29* (16), e202202503.
155. Pavlin, M.; Qasem, Z.; Sameach, H.; Gevorkyan-Airapetov, L.; Ritacco, I.; Ruthstein, S.; Magistrato, A., Unraveling the Impact of Cysteine-to-Serine Mutations on the Structural and Functional Properties of Cu(I)-Binding Proteins. *Int. J. Mol. Sci.* **2019,** *20* (14).
156. Biava, H. D., Tackling Achilles' Heel in Synthetic Biology: Pairing Intracellular Synthesis of Noncanonical Amino Acids with Genetic-Code Expansion to Foster Biotechnological Applications. *ChemBioChem* **2020,** *21* (9), 1265-1273.
157. Sandoval, B. A.; Hyster, T. K., Emerging strategies for expanding the toolbox of enzymes in biocatalysis. *Curr. Opin. Chem. Biol.* **2020,** *55*, 45-51.
158. Melnikov, S. V.; Soll, D., Aminoacyl-tRNA Synthetases and tRNAs for an Expanded Genetic Code: What Makes them Orthogonal? *Int. J. Mol. Sci.* **2019,** *20* (8), 1929.
159. Bullwinkle, T. J.; Ibba, M., Emergence and evolution. *Top. Curr. Chem.* **2014,** *344*, 43-87.
160. Neumann, H., Rewiring translation - Genetic code expansion and its applications. *FEBS Lett.* **2012,** *586* (15), 2057-64.
161. Arranz-Gibert, P.; Vanderschuren, K.; Isaacs, F. J., Next-generation genetic code expansion. *Curr. Opin. Chem. Biol.* **2018,** *46*, 203-211.
162. Seki, E.; Yanagisawa, T.; Kuratani, M.; Sakamoto, K.; Yokoyama, S., Fully Productive Cell-Free Genetic Code Expansion by Structure-Based Engineering of Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetase. *ACS Synth Biol* **2020,** *9* (4), 718-732.
163. Cervettini, D.; Tang, S.; Fried, S. D.; Willis, J. C. W.; Funke, L. F. H.; Colwell, L. J.; Chin, J. W., Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase-tRNA pairs. *Nat. Biotechnol.* **2020,** *38* (8), 989-999.
164. Meineke, B.; Heimgartner, J.; Lafranchi, L.; Elsasser, S. J., Methanomethylophilus alvus Mx1201 Provides Basis for Mutual Orthogonal Pyrrolysyl tRNA/Aminoacyl-tRNA Synthetase Pairs in Mammalian Cells. *ACS Chem. Biol.* **2018,** *13* (11), 3087-3096.
165. Young, D. D.; Young, T. S.; Jahnz, M.; Ahmad, I.; Spraggon, G.; Schultz, P. G., An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **2011,** *50* (11), 1894-900.

166. Stokes, A. L.; Miyake-Stoner, S. J.; Peeler, J. C.; Nguyen, D. P.; Hammer, R. P.; Mehl, R. A., Enhancing the utility of unnatural amino acid synthetases by manipulating broad substrate specificity. *Mol. Biosyst.* **2009,** *5* (9), 1032-8.
167. Young, D. D.; Schultz, P. G., Playing with the Molecules of Life. *ACS Chem. Biol.* **2018,** *13* (4), 854-870.
168. Almhjell, P. J.; Boville, C. E.; Arnold, F. H., Engineering enzymes for noncanonical amino acid synthesis. *Chem. Soc. Rev.* **2018,** *47* (24), 8980-8997.
169. Takimoto, J. K.; Xiang, Z.; Kang, J. Y.; Wang, L., Esterification of an unnatural amino acid structurally deviating from canonical amino acids promotes its uptake and incorporation into proteins in mammalian cells. *ChemBioChem* **2010,** *11* (16), 2268-72.
170. Zhou, H.; Cheung, J. W.; Carpenter, T.; Jones, S. K., Jr.; Luong, N. H.; Tran, N. C.; Jacobs, S. E.; Galbada Liyanage, S. A.; Cropp, T. A.; Yin, J., Enhancing the incorporation of lysine derivatives into proteins with methylester forms of unnatural amino acids. *Bioorg. Med. Chem. Lett.* **2020,** *30* (2), 126876.
171. Ko, W.; Kumar, R.; Kim, S.; Lee, H. S., Construction of Bacterial Cells with an Active Transport System for Unnatural Amino Acids. *ACS Synth Biol* **2019,** *8* (5), 1195-1203.
172. Gao, W.; Cho, E.; Liu, Y.; Lu, Y., Advances and Challenges in Cell-Free Incorporation of Unnatural Amino Acids Into Proteins. *Front. Pharmacol.* **2019,** *10* (611), 611.
173. Jung, J.-E.; Lee, S. Y.; Park, H.; Cha, H.; Ko, W.; Sachin, K.; Kim, D. W.; Chi, D. Y.; Lee, H. S., Genetic incorporation of unnatural amino acids biosynthesized from α-keto acids by an aminotransferase. *Chemical Science* **2014,** *5* (5), 1881-1885.
174. Ma, Y.; Biava, H.; Contestabile, R.; Budisa, N.; di Salvo, M. L., Coupling bioorthogonal chemistries with artificial metabolism: intracellular biosynthesis of azidohomoalanine and its incorporation into recombinant proteins. *Molecules* **2014,** *19* (1), 1004-22.
175. Kim, S.; Sung, B. H.; Kim, S. C.; Lee, H. S., Genetic incorporation of l-dihydroxyphenylalanine (DOPA) biosynthesized by a tyrosine phenol-lyase. *Chem. Commun. (Camb.)* **2018,** *54* (24), 3002-3005.
176. Mehl, R. A.; Anderson, J. C.; Santoro, S. W.; Wang, L.; Martin, A. B.; King, D. S.; Horn, D. M.; Schultz, P. G., Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **2003,** *125* (4), 935-9.
177. Narancic, T.; Almahboub, S. A.; O'Connor, K. E., Unnatural amino acids: production and biotechnological potential. *World J. Microbiol. Biotechnol.* **2019,** *35* (4), 67.
178. Xue, Y. P.; Cao, C. H.; Zheng, Y. G., Enzymatic asymmetric synthesis of chiral amino acids. *Chem. Soc. Rev.* **2018,** *47* (4), 1516-1561.
179. Zhou, H. S.; Meng, L. J.; Yin, X. J.; Liu, Y. Y.; Xu, G.; Wu, J. P.; Wu, M. B.; Yang, L. R., Artificial Biocatalytic Cascade with Three Enzymes in One Pot for Asymmetric Synthesis of Chiral Unnatural Amino Acids. *Eur. J. Org. Chem.* **2019,** *2019* (38), 6470-6477.
180. Dhayalan, V.; Mal, K.; Milo, A., Practical Synthesis of Chiral N-Heterocyclic Carbene Triazolium Salts Containing a Hydroxy Functional Handle. *Synthesis-Stuttgart* **2019,** *51* (14), 2845-2864.
181. Zheng, P.; Gondo, C. A.; Bode, J. W., Late-stage diversification of chiral N-heterocyclic-carbene precatalysts for enantioselective homoenolate additions. *Chem. Asian J.* **2011,** *6* (2), 614-20.
182. Kubota, T.; Shimono, J.; Kanameda, C.; Izumi, Y., The first thermophilic alpha-oxoamine synthase family enzyme that has activities of 2-amino-3-ketobutyrate CoA ligase and 7-keto-8-aminopelargonic acid synthase: cloning and overexpression of the gene from an extreme thermophile, Thermus thermophilus, and characterization of its gene product. *Biosci. Biotechnol. Biochem.* **2007,** *71* (12), 3033-40.
183. Kubota, T.; Izumi, Y., Detection and Characterization of a Thermophilic Biotin Biosynthetic Enzyme, 7-Keto-8-aminopelargonic Acid Synthase, from Various Thermophiles. *Biosci., Biotechnol., Biochem.* **2012,** *76* (4), 685-690.

184. Ashley, B.; Basle, A.; Sajjad, M.; El Ashram, A.; Kelis, P.; Marles-Wright, J.; Campopiano, D. J., Versatile Chemo-Biocatalytic Cascade Driven by a Thermophilic and Irreversible C-C Bond-Forming alpha-Oxoamine Synthase. *ACS Sustain Chem Eng* **2023,** *11* (21), 7997-8002.
185. Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J., CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **2018,** *46* (W1), W363-W367.
186. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T., SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018,** *46* (W1), W296-W303.
187. Liu, H.; Naismith, J. H., An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **2008,** *8* (1), 91.
188. Some, D.; Amartely, H.; Tsadok, A.; Lebendiker, M., Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS). *J Vis Exp* **2019,** (148), e59615.
189. Turbeville, T. D.; Zhang, J.; Hunter, G. A.; Ferreira, G. C., Histidine 282 in 5-aminolevulinate synthase affects substrate binding and catalysis. *Biochemistry* **2007,** *46* (20), 5972-81.
190. Goroncy, A. K.; Murayama, K.; Shirouzu, M.; Kuramitsu, S.; Kigawa, T.; Yokoyama, S., NMR and X-ray structures of the putative sterol carrier protein 2 from Thermus thermophilus HB8 show conformational changes. *J Struct Funct Genomics* **2010,** *11* (4), 247-56.
191. Edqvist, J.; Blomqvist, K., Fusion and fission, the evolution of sterol carrier protein-2. *J. Mol. Evol.* **2006,** *62* (3), 292-306.
192. Gallegos, A. M.; Atshaves, B. P.; Storey, S. M.; Starodub, O.; Petrescu, A. D.; Huang, H.; McIntosh, A. L.; Martin, G. G.; Chao, H.; Kier, A. B.; Schroeder, F., Gene structure, intracellular localization, and functional roles of sterol carrier protein-2. *Prog. Lipid Res.* **2001,** *40* (6), 498-563.
193. Haapalainen, A. M.; van Aalten, D. M.; Merilainen, G.; Jalonen, J. E.; Pirila, P.; Wierenga, R. K.; Hiltunen, J. K.; Glumoff, T., Crystal structure of the liganded SCP-2-like domain of human peroxisomal multifunctional enzyme type 2 at 1.75 A resolution. *J. Mol. Biol.* **2001,** *313* (5), 1127-38.
194. Deuss, P. J.; Popa, G.; Botting, C. H.; Laan, W.; Kamer, P. C., Highly efficient and site-selective phosphane modification of proteins through hydrazone linkage: development of artificial metalloenzymes. *Angew. Chem. Int. Ed. Engl.* **2010,** *49* (31), 5315-7.
195. Jarvis, A. G.; Obrecht, L.; Deuss, P. J.; Laan, W.; Gibson, E. K.; Wells, P. P.; Kamer, P. C. J., Enzyme Activity by Design: An Artificial Rhodium Hydroformylase for Linear Aldehydes. *Angew. Chem. Int. Ed. Engl.* **2017,** *56* (44), 13596-13600.
196. Kuckhoff, T.; Brewster, R. C.; Ferguson, C. T. J.; Jarvis, A. G., Reactivity Tuning of Metal-Free Artificial Photoenzymes through Binding Site Specific Bioconjugation. *Eur. J. Org. Chem.* **2023,** *26* (13), e202201412.
197. Doble, M. V.; Obrecht, L.; Joosten, H. J.; Lee, M.; Rozeboom, H. J.; Branigan, E.; Naismith, J. H.; Janssen, D. B.; Jarvis, A. G.; Kamer, P. C. J., Engineering Thermostability in Artificial Metalloenzymes to Increase Catalytic Activity. *Acs Catalysis* **2021,** *11* (6), 3620-3627.
198. Madeira, F.; Pearce, M.; Tivey, A. R. N.; Basutkar, P.; Lee, J.; Edbali, O.; Madhusoodanan, N.; Kolesnikov, A.; Lopez, R., Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **2022,** *50* (W1), W276-W279.
199. Whitmore, L.; Wallace, B. A., DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **2004,** *32* (Web Server issue), W668-73.

200. Sreerama, N.; Woody, R. W., Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **2000,** *287* (2), 252-60.
201. Lees, J. G.; Miles, A. J.; Wien, F.; Wallace, B. A., A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics* **2006,** *22* (16), 1955-62.
202. Nagy, G.; Grubmuller, H., How accurate is circular dichroism-based model validation? *Eur. Biophys. J.* **2020,** *49* (6), 497-510.
203. Deuss, P. J., Modified Proteins as Tuneable Transition Metal Catalysts. *University of St Andrews* **2011**.
204. Goldman, A. D.; Beatty, J. T.; Landweber, L. F., The TIM Barrel Architecture Facilitated the Early Evolution of Protein-Mediated Metabolism. *J. Mol. Evol.* **2016,** *82* (1), 17-26.
205. Macdonald, D. S.; Garrabou, X.; Klaus, C.; Verez, R.; Mori, T.; Hilvert, D., Engineered Artificial Carboligases Facilitate Regioselective Preparation of Enantioenriched Aldol Adducts. *J. Am. Chem. Soc.* **2020,** *142* (23), 10250-10254.
206. Chiu, H. J.; Reddick, J. J.; Begley, T. P.; Ealick, S. E., Crystal structure of thiamin phosphate synthase from Bacillus subtilis at 1.25 A resolution. *Biochemistry* **1999,** *38* (20), 6460-70.
207. Peapus, D. H.; Chiu, H. J.; Campobasso, N.; Reddick, J. J.; Begley, T. P.; Ealick, S. E., Structural characterization of the enzyme-substrate, enzyme-intermediate, and enzyme-product complexes of thiamin phosphate synthase. *Biochemistry* **2001,** *40* (34), 10103-14.
208. Reddick, J. J.; Nicewonger, R.; Begley, T. P., Mechanistic studies on thiamin phosphate synthase: evidence for a dissociative mechanism. *Biochemistry* **2001,** *40* (34), 10095-102.
209. Paul, D.; Chatterjee, A.; Begley, T. P.; Ealick, S. E., Domain organization in Candida glabrata THI6, a bifunctional enzyme required for thiamin biosynthesis in eukaryotes. *Biochemistry* **2010,** *49* (45), 9922-34.
210. Robert, X.; Gouet, P., Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **2014,** *42* (Web Server issue), W320-4.
211. Goujon, M.; McWilliam, H.; Li, W. Z.; Valentin, F.; Squizzato, S.; Paern, J.; Lopez, R., A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **2010,** *38* (suppl_2), W695-W699.
212. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; Thompson, J. D.; Higgins, D. G., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011,** *7* (1), 539.
213. Kato, S.; Onoda, A.; Schwaneberg, U.; Hayashi, T., Evolutionary Engineering of a Cp*Rh(III) Complex-Linked Artificial Metalloenzyme with a Chimeric beta-Barrel Protein Scaffold. *J. Am. Chem. Soc.* **2023,** *145* (15), 8285-90.
214. Sikowitz, M. D.; Shome, B.; Zhang, Y.; Begley, T. P.; Ealick, S. E., Structure of a Clostridium botulinum C143S thiaminase I/thiamin complex reveals active site architecture. *Biochemistry* **2013,** *52* (44), 7830-9.
215. Costello, C. A.; Kelleher, N. L.; Abe, M.; McLafferty, F. W.; Begley, T. P., Mechanistic studies on thiaminase I. Overexpression and identification of the active site nucleophile. *J. Biol. Chem.* **1996,** *271* (7), 3445-52.
216. Campobasso, N.; Costello, C. A.; Kinsland, C.; Begley, T. P.; Ealick, S. E., Crystal structure of thiaminase-I from Bacillus thiaminolyticus at 2.0 A resolution. *Biochemistry* **1998,** *37* (45), 15981-9.
217. Abe, M.; Ito, S.; Kimoto, M.; Hayashi, R.; Nishimune, T., Molecular studies on thiaminase I. *Biochim. Biophys. Acta* **1987,** *909* (3), 213-21.

218. Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L., Quantitative analysis of modified proteins and their positional isomers by tandem mass spectrometry: human histone H4. *Anal. Chem.* **2006,** *78* (13), 4271-80.
219. Gober, I. N.; Riemen, A. J.; Villain, M., Sequence sensitivity and pH dependence of maleimide conjugated N-terminal cysteine peptides to thiazine rearrangement. *J. Pept. Sci.* **2021,** *27* (7), e3323.
220. Krishna, R.; Wang, J.; Ahern, W.; Sturmfels, P.; Venkatesh, P.; Kalvet, I.; Lee, G. R.; Morey-Burrows, F. S.; Anishchenko, I.; Humphreys, I. R.; McHugh, R.; Vafeados, D.; Li, X.; Sutherland, G. A.; Hitchcock, A.; Hunter, C. N.; Baek, M.; DiMaio, F.; Baker, D., Generalized Biomolecular Modeling and Design with RoseTTAFold All-Atom. **2023**, 2023.10.09.561603.
221. Mantelingu, K.; Lin, Y.; Seidel, D., Intramolecular [3 + 2]-cycloadditions of azomethine ylides derived from secondary amines via redox-neutral C-H functionalization. *Org. Lett.* **2014,** *16* (22), 5910-3.
222. Collett, C. J.; Young, C. M.; Massey, R. S.; O'Donoghue, A. C.; Smith, A. D., Kinetic and Structure-Activity Studies of the Triazolium Ion- Catalyzed Intramolecular Stetter Reaction. *Eur. J. Org. Chem.* **2021,** *2021* (26), 3670-3675.
223. Rezazadeh Khalkhali, M.; Wilde, M. M. D.; Gravel, M., Enantioselective Stetter Reactions Catalyzed by Bis(amino)cyclopropenylidenes: Important Role for Water as an Additive. *Org. Lett.* **2021,** *23* (1), 155-159.
224. Gupta, V.; Prakash, N. A.; Lakshmi, V.; Boopathy, R.; Jeyakanthan, J.; Velmurugan, D.; Sekar, K., Recognition of active and inactive catalytic triads: A template based approach. *Int. J. Biol. Macromol.* **2010,** *46* (3), 317-23.
225. Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G., An enhanced system for unnatural amino acid mutagenesis in E. coli. *J. Mol. Biol.* **2010,** *395* (2), 361-74.
226. Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G., Addition of the keto functional group to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100* (1), 56-61.
227. Ryu, Y.; Schultz, P. G., Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nat. Methods* **2006,** *3* (4), 263-5.
228. Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G., Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J. Am. Chem. Soc.* **2002,** *124* (31), 9026-7.
229. Green, A. P.; Hayashi, T.; Mittl, P. R.; Hilvert, D., A Chemically Programmed Proximal Ligand Enhances the Catalytic Properties of a Heme Enzyme. *J. Am. Chem. Soc.* **2016,** *138* (35), 11344-52.
230. Xiao, H.; Peters, F. B.; Yang, P. Y.; Reed, S.; Chittuluru, J. R.; Schultz, P. G., Genetic incorporation of histidine derivatives using an engineered pyrrolysyl-tRNA synthetase. *ACS Chem. Biol.* **2014,** *9* (5), 1092-6.
231. Pott, M.; Tinzl, M.; Hayashi, T.; Ota, Y.; Dunkelmann, D.; Mittl, P. R. E.; Hilvert, D., Noncanonical Heme Ligands Steer Carbene Transfer Reactivity in an Artificial Metalloenzyme*. *Angew. Chem. Int. Ed. Engl.* **2021,** *60* (27), 15063-15068.
232. Bryson, D. I.; Fan, C.; Guo, L. T.; Miller, C.; Soll, D.; Liu, D. R., Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **2017,** *13* (12), 1253-1260.
233. Schultz, K. C.; Supekova, L.; Ryu, Y.; Xie, J.; Perera, R.; Schultz, P. G., A genetically encoded infrared probe. *J. Am. Chem. Soc.* **2006,** *128* (43), 13984-5.
234. Nguyen, D. P.; Elliott, T.; Holt, M.; Muir, T. W.; Chin, J. W., Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. *J. Am. Chem. Soc.* **2011,** *133* (30), 11418-21.
235. Hu, L.; Qiao, Y.; Liu, J.; Zheng, C.; Wang, X.; Sun, P.; Gu, Y.; Liu, W., Characterization of Histidine Functionalization and Its Timing in the Biosynthesis of Ribosomally Synthesized and Posttranslationally Modified Thioamitides. *J. Am. Chem. Soc.* **2022,** *144* (10), 4431-4438.

236. Liu, X.; Wu, J.; Sun, Y.; Xie, W., Substrate Recognition Mechanism of the Putative Yeast Carnosine N-methyltransferase. *ACS Chem. Biol.* **2017,** *12* (8), 2164-2171.
237. Cao, R.; Zhang, X.; Liu, X.; Li, Y.; Li, H., Molecular basis for histidine N1 position-specific methylation by CARNMT1. *Cell Res.* **2018,** *28* (4), 494-496.
238. Maier, T. H., Semisynthetic production of unnatural L-alpha-amino acids by metabolic engineering of the cysteine-biosynthetic pathway. *Nat. Biotechnol.* **2003,** *21* (4), 422-7.
239. Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Soll, D.; Steitz, T. A., Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (27), 11268-73.
240. Horner, K. E.; Karadakov, P. B., Shielding in and around Oxazole, Imidazole, and Thiazole: How Does the Second Heteroatom Affect Aromaticity and Bonding? *J. Org. Chem.* **2015,** *80* (14), 7150-7.
241. Katoh, T.; Suga, H., In Vitro Genetic Code Reprogramming for the Expansion of Usable Noncanonical Amino Acids. *Annu. Rev. Biochem.* **2022,** *91* (1), 221-243.
242. Gorecki, L.; Andrys, R.; Schmidt, M.; Kucera, T.; Psotka, M.; Svobodova, B.; Hrabcova, V.; Hepnarova, V.; Bzonek, P.; Jun, D.; Kuca, K.; Korabecny, J.; Musilek, K., Cysteine-Targeted Insecticides against A. gambiae Acetylcholinesterase Are Neither Selective nor Reversible Inhibitors. *ACS Med. Chem. Lett.* **2020,** *11* (1), 65-71.
243. J. B. Siegel, A. L. Smith, S. Poust, A. J. Wargacki, A. Bar-Even, C. Louw, B. W. Shen, C. B. Eiben, H. M. Tran, E. Noor, J. L. Gallaher, J. Bale, Y. Yoshikuni, M. H. Gelb, J. D. Keasling, B. L. Stoddard, M. E. Lidstrom and D. Baker, Computational protein design enables a novel one-carbon assimilation pathway, *Proc. Natl. Acad. Sci. U. S. A.*,**2015**, *112,* 3704-3709.
244. S. Güner, V. Wegat, A. Pick and V. Sieber, Design of a synthetic enzyme cascade for the in vitro fixation of a C1 carbon source to a functional C4 sugar, *Green Chem*., **2021**, *23*, 6583-6590.
245. L. Zhang, R. Singh, S. D, Z. Guo, J. Li, F. Chen, Y. He, X. Guan, Y. C. Kang and J.-K. Lee, An artificial synthetic pathway for acetoin, 2,3-butanediol, and 2-butanol production from ethanol using cell free multi-enzyme catalysis, *Green Chem*., **2018**, *20,* 230-242.

1 Appendix

1.1 Protein Sequences

1.1.1 ThAOS

Gene Sequence

ATGTCGTACTACCATCACCATCACCATCACGATTACGACATCCCAACGACCGAAAACCTGTATTTTCA
GGGCGCCATGGGCAGCCTGGATCTGCGTGCGCGTGTGCGTGAAGAACTGGAGCGTCTGAAGCGTG
AAGGTCTGTATATTAGCCCGAAAGTGCTGGAAGCGCCGCAGGAACCGGTGACCCGTGTTGAAGGC
CGTGAGGTGGTTAACCTGGCGAGCAACAACTACCTGGGTTTTGCGAACCACCCGTATCTGAAGGAA
AAAGCGCGTCAATACCTGGAGAAATGGGGTGCGGGTAGCGGTGCGGTGCGTACCATCGCGGGCA
CCTTCACCTATCACGTTGAACTGGAGGAAGCGCTGGCGCGTTTTAAAGGTACCGAGAGCGCGCTGG
TGCTGCAGAGCGGTTTCACCGCGAACCAAGGCGTTCTGGGTGCGCTGCTGAAGGAAGGCGACGTG
GTTTTTAGCGATGAGCTGAACCACGCGAGCATCATTGACGGTCTGCGTCTGACCAAAGCGACCCGT
CTGGTGTTCCGTCACGCGGATGTTGCGCACCTGGAGGAACTGCTGAAGGCGCACGACACCGATGG
TCTGAAACTGATTGTGACCGACGGCGTTTTTAGCATGGACGGTGATATCGCGCCGCTGGATAAGAT
TGTGCCGCTGGCGAAGAAATACAAAGCGGTGGTTTATGTGGACGATGCGCACGGCAGCGGTGTTC
TGGGCGAAAAGGGCAAAGGTACCGTGCACCACTTCGGTTTTCACCAGGACCCGGATGTGGTTCAA
GTGGCGACCCTGAGCAAAGCGTGGGCGGGTATCGGTGGCTACGCGGCGGGTGCGCGTGAGCTGA
AGGACCTGCTGATTAACAAAGCGCGTCCGTTCCTGTTTAGCACCAGCCACCCGCCGGCGGTGGTTG
GTGCGCTGCTGGGTGCGCTGGAACTGATCGAGAAGGAACCGGAGCGTGTGGAACGTCTGTGGGA
GAACACCCGTTATTTCAAACGTGAGCTGGCGCGTCTGGGCTACGATACCCTGGGTAGCCAGACCCC
GATCACCCCGGTTCTGTTCGGTGAAGCGCCGCTGGCGTTTGAGGCGAGCCGTCTGCTGCTGGAGGA
AGGCGTGTTCGCGGTTGGCATTGGTTTTCCGACCGTGCCGCGTGGTAAAGCGCGTATCCGTAACAT
TGTTACCGCGGCGCACACCAAAGAGATGCTGGACAAGGCGCTGGAGGCGTATGAAAAGGTTGGCA
AACGTCTGGGTATTATCCGCTAA

Protein Sequence

MSYYHHHHHHDYDIPTTENLYFQGA**M**GSLDLRARVREELERLKREGLYISPKVLEAPQEPVTRVEGREV
VNLASNNYLGFANHPYLKEKARQYLEKWGAGSGAVRTIAGTFTYHVELEEALARFKGTESALVLQSGFT
ANQGVLGALLKEGDVVFSDELNHASIIDGLRLTKATRLVFRHADVAHLEELLKAHDTDGLKLIVTDGVFS
MDGDIAPLDKIVPLAKKYKAVVYVDDAHGSGVLGEKGKGTVHHFGFHQDPDVVQVATLS**K**AWAGIGG
YAAGARELKDLLINKARPFLFSTSHPPAVVGALLGALELIEKEPERVERLWENTRYFKRELARLGYDTLGSQ
TPITPVLFGEAPLAFEASRLLLEEGVFAVGIGFPTVPRGKARIRNIVTAAHTKEMLDKALEAYEKVGKRLGII
R

The N-terminal affinity Tag is marked in red and residue M1 is bold and underlined. The V79S mutant was created previously (Ben Ashley, University of Edinburgh) The K243C mutant (underlined above) was created using primers described in materials and methods.

1.1.2 hSCP

Gene Sequence

ATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCAACGACCGAAAACCTGTATTTTCA
GGGCGCCATGGAGGGAGGGAAGCTTCAGAGTACCTTTGTATTTGAGGAAATAGGACGCCGCCTAA
AGGATATTGGGCCTGAGGTGGTGAAGAAAGTAAATGCTGTATTTGAGTGGCATATAACCAAAGGC
GGAAATATTGGGGCTAAGTGGACTATTGACCTGAAAAGTGGTTCTGGAAAAGTGTACCAAGGCCCT
GCAAAAGGTGCTGCTGATACAACAATCATACTTTCAGATGAAGATTTCATGGAGGTGGTCCTGGGC

AAGCTTGACCCTCAGAAGGCATTCTTTAGTGGCAGGCTGAAGTGCAGAGGGAACATCATGCTGAGC
CAGAAACTTCAGATGATTCTTAAAGACTATGCCAAGCTCTGA

Protein Sequence

MSYYHHHHHHDYDIPTTENLYFQGA**M**EGGKLQSTFVFEEIGRRLKDIGPEVVKKVNAVFEWHITKGGNI
GAKWTIDLKSGSGKVYQGPAKGAADTTIILSDEDFMEVVLGKLDPQKAFFSGRLKARGNIMLSQKLQMI
LKDYAKL

The N-terminal affinity Tag is marked in red and residue M1 is bold and underlined. The following cysteine mutants (underlined above) were previously created from this protein: hSCP V83C, hSCP A100C, hSCP Q111C, hSCP A100C F94H, hSCP V83C stable (E81K, S56D)[147, 194].

1.1.3 TTSCP

The SCP gene from *Thermus Thermophilus* was ordered from IDT and codon optimised for *E. coli* expression with the 5' and 3'-terminus for pET28b vector insertion via Gibson assembly incorporating a C-terminal His tag. Gibson assembly was performed by Eva Klemencic, University of Edinburgh, using HiFi DNA assembly Master Mix at 20 µl.

Gene Sequence

ATGGAGCTTTTCACCGAGGCCTGGGCCCAGGCGTACTGCCGGAAGCTGAACGAGAGCGAGGCCTA
CCGCAAGGCGGCGAGCACCTGGGAGGGCTCCCTGGCCCTCGCGGTGCGCCCGGACCCCAAGGCGG
GGTTCCCCAAGGGGGTGGCCGTGGTCCTGGACCTCTGGCACGGGGCCTGCCGGGGGGCGAAGGC
GGTGGAGGGGGAGGCGGAGGCGGACTTCGTCATTGAGGCCGACCTCGCCACCTGGCAGGAGGTG
CTGGAGGGACGCCTCGAGCCCCTAAGCGCCCTCATGCGGGGACTTTTGGAGCTCAAGAAGGGCAC
CATCGCCGCCCTCGCCCCTTACGCCCAGGCGGCCCAGGAGCTCGTCAAAGTGGCCCGGGAGGTGG
CAGAAAACCTGTATTTTCAGGGCCTCGAGCACCACCACCACCACCACTGA

Protein Sequence

MELFTEAWAQAY**C**RKLNESEAYRKAASTWEGSLALAVRPDPKAGFPKGVAVVLDLWHGA**C**RGAKAVE
GEAEADFVIEADLATWQEVLEGRLEPLSALMRGLLELKKGTIAALAPYAQAAQELVKVAREVA**ENLYFQ
G**LEHHHHHH*

Residues Cys13 and Cys60 are bold underlined. The residues at the C-terminus are the affinity tag and TEV protease sequence (ENLYFQ/G). The C-terminal sequence in red is removed.

1.1.4 TTSCP W83C

This construct was created by Eva Klemencic, University of Edinburgh, using site directed mutagenesis of TTSCP using the following primers:

TT_W83C_Fw (5'-TGCCaaCCTGTCAGGAGGTGC-3')

TT_W83C_Rv (5'-GGTCGGCCTCAATGACGAAG-3').

Gene Sequence

ATGGAGCTTTTCACCGAGGCCTGGGCCCAGGCGTACTGCCGGAAGCTGAACGAGAGCGAGGCCTA
CCGCAAGGCGGCGAGCACCTGGGAGGGCTCCCTGGCCCTCGCGGTGCGCCCGGACCCCAAGGCGG
GGTTCCCCAAGGGGGTGGCCGTGGTCCTGGACCTCTGGCACGGGGCCTGCCGGGGGGCGAAGGC
GGTGGAGGGGGAGGCGGAGGCGGACTTCGTCATTGAGGCCGACCTCGCCACCTGTCAGGAGGTG
CTGGAGGGACGCCTCGAGCCCCTAAGCGCCCTCATGCGGGGACTTTTGGAGCTCAAGAAGGGCAC

CATCGCCGCCCTCGCCCCTTACGCCCAGGCGGCCCAGGAGCTCGTCAAAGTGGCCCGGGAGGTGG
CAGAAAACCTGTATTTTCAGGGCCTCGAGCACCACCACCACCACCACTGA

Protein Sequence

MELFTEAWAQAY**C**RKLNESEAYRKAASTWEGSLALAVRPDPKAGFPKGVAVVLDLWHGA**C**RGAKAVE
GEAEADFVIEADLAT**C**QEVLEGRLEPLSALMRGLLELKKGTIAALAPYAQAAQELVKVAREVAENLYFQG
LEHHHHHH*

Residues Cys13, Cys60 and Cys83 are bold underlined. The residues at the C-terminus are the affinity tag and TEV protease sequence (ENLYFQ/G). The C-terminal sequence in red is removed.

1.1.5 TTSCP_L102C and TTSCP ΔDSB L102C

Due to the high GC content of the TTSCP genes hindering mutagenesis attempts, synthetic gene cloned into pET-28a(+) between NcoI/XhoI restriction sites was purchased from IDT. These constructs were codon optimised for expression in *E.coli* K12 and further optimised for lower GC content

1.1.6 TTSCP_L102C

Gene Sequence

ATGGAATTGTTTACAGAAGCATGGGCACAGGCATATTGTCGCAAGCTGAATGAATCTGAGGCATAC
CGCAAGGCGGCATCGACTTGGGAAGGGTCACTGGCGCTTGCGGTTCGTCCCGATCCTAAAGCCGG
GTTTCCAAAAGGTGTTGCGGTCGTCCTGGATTTATGGCACGGCGCGTGCCGCGGCGCCAAAGCCGT
GGAAGGGGAAGCCGAGGCTGATTTTGTCATTGAAGCCGATTTAGCAACATGGCAGGAGGTTTTAG
AGGGGCGCCTGGAACCATTGTCAGCCCTGATGCGCGGCTTGTGCGAACTTAAGAAGGGCACAATC
GCGGCCTTGGCTCCATACGCCCAAGCGGCTCAGGAGCTGGTAAAAGTGGCCCGCGAGGTGGCCGA
AAACCTTTATTTCCAGGGACTCGAGCACCACCACCACCACCACTAG

Protein Sequence

MELFTEAWAQAY**C**RKLNESEAYRKAASTWEGSLALAVRPDPKAGFPKGVAVVLDLWHGA**C**RGAKAVE
GEAEADFVIEADLATWQEVLEGRLEPLSALMRGL**C**ELKKGTIAALAPYAQAAQELVKVAREVAENLYFQ
GLEHHHHHH*

Residues Cys13, Cys60 and Cys102 are bold underlined. The residues at the C-terminus are the affinity tag and TEV protease sequence (ENLYFQ/G). The C-terminal sequence in red is removed.

1.1.7 TTSCP ΔDSB L102C

Gene Sequence

ATGGAATTGTTTACAGAAGCATGGGCACAGGCATACGCTCGCAAGCTGAATGAATCTGAGGCATAC
CGCAAGGCGGCATCGACTTGGGAAGGGTCACTGGCGCTTGCGGTTCGTCCCGATCCTAAAGCCGG
GTTTCCAAAAGGTGTTGCGGTCGTCCTGGATTTATGGCACGGCGCGGCACGCGGCGCCAAAGCCGT
GGAAGGGGAAGCCGAGGCTGATTTTGTCATTGAAGCCGATTTAGCAACATGGCAGGAGGTTTTAG
AGGGGCGCCTGGAACCATTGTCAGCCCTGATGCGCGGCTTGTGCGAACTTAAGAAGGGCACAATC
GCGGCCTTGGCTCCATACGCCCAAGCGGCTCAGGAGCTGGTAAAAGTGGCCCGCGAGGTGGCCGA
AAACCTTTATTTCCAGGGACTCGAGCACCACCACCACCACCACTGA

Protein Sequence

MELFTEAWAQAYARKLNESEAYRKAASTWEGSLALAVRPDPKAGFPKGVAVVLDLWHGAARGAKAV
EGEAEADFVIEADLATWQEVLEGRLEPLSALMRGL**C**ELKKGTIAALAPYAQAAQELVKVAREVAENLYFQ
GLEHHHHHH*

Residue Cys102 is bold underlined. The residues at the C-terminus are the affinity tag and TEV protease sequence (ENLYFQ/G). The C-terminal sequence in red is removed.

1.1.8 TTSCP L112C

The gene for this construct was ordered from IDT and the clone was created by Eva Klemencic in the same way as the WT TTSCP using Gibson assembly.

Gene Sequence

ATGGAGCTTTTCACCGAGGCCTGGGCCCAGGCGTACTGCCGGAAGCTGAACGAGAGCGAGGCCTA
CCGCAAGGCGGCGAGCACCTGGGAGGGCTCCCTGGCCCTCGCGGTGCGCCCGGACCCCAAGGCGG
GGTTCCCCAAGGGGGTGGCCGTGGTCCTGGACCTCTGGCACGGGGCCTGCCGGGGGGCGAAGGC
GGTGGAGGGGGAGGCGGAGGCGGACTTCGTCATTGAGGCCGACCTCGCCACCTGGCAGGAGGTG
CTGGAGGGACGCCTCGAGCCCCTAAGCGCCCTCATGCGGGGACTTTTGGAGCTCAAGAAGGGCAC
CATCGCCGCCTGCGCCCCTTACGCCCAGGCGGCCCAGGAGCTCGTCAAAGTGGCCCGGGAGGTGG
CAGAAAACCTGTATTTTCAGGGCCTCGAGCACCACCACCACCACCACTGA

Protein Sequence

MELFTEAWAQAY**C**RKLNESEAYRKAASTWEGSLALAVRPDPKAGFPKGVAVVLDLWHGA**C**RGAKAVE
GEAEADFVIEADLATWQEVLEGRLEPLSALMRGLLELKKGTIAA**C**APYAQAAQELVKVAREVAENLYFQ
GLEHHHHHH*

Residues Cys13, Cys60 and Cys112 are bold underlined. The residues at the C-terminus are the affinity tag and TEV protease sequence (ENLYFQ/G). The C-terminal sequence in red is removed.

1.1.9 Thiaminase1

Gene Sequence

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCGAGAATCTTTATTTTCAGGGCCATATGT
CAAAGGTAAAAGGCTTCATTTACAAACCACTGATGGTGATGCTGGCGTTGCTGCTCGTCGTTGTCAG
TCCGGCCGGAGCCGGGGCGGCCCATTCCGATGCGTCTTCGGATATTACGCTGAAGGTGGCGATCTA
CCCGTACGTTCCGGATCCTGCCCGTTTCCAGGCGGCCGTTCTTGATCAGTGGCAGCGGCAGGAGCC
TGGCGTCAAGCTGGAGTTTACGGACTGGGATTCCTACTCCGCGGATCCGCCGGACGATTTGGATGT
GTTCGTCCTGGATTCGATTTTTTTAAGCCATTTTGTCGATGCGGGGTATCTCCTGCCCTTCGGCAGCC
AAGATATCGACCAGGCGGAGGATGTGCTCCCTTTTGCTCTTCAAGGGGCGAAGCGCAACGGCGAG
GTCTACGGTCTGCCGCAAATTTTATGCACGAACCTGCTTTTTTACCGTAAAGGCGATTTGAAGATAG
GGCAAGTCGACAATATTTATGAACTGTACAAAAAAATAGGAACCAGCCACTCCGAGCAGATTCCGC
CTCCCCAGAATAAGGGTTTGCTAATTAATATGGCCGGCGGGACGACGAAGGCAAGCATGTATTTGG
AGGCGCTTATCGATGTGACCGGCCAGTATACGGAATATGACCTTCTTCCGCCGCTGGATCCCCTAAA
TGACAAGGTCATTCGCGGCTTGCGCCTGTTAATCAATATGGCGGGGGAGAAGCCGTCGCAGTATGT
TCCCGAGGATGGCGACGCTTATGTAAGGGCGTCGTGGTTCGCGCAAGGGAGCGGCAGAGCCTTCA
TCGGCTACAGCGAGTCGATGATGCGCATGGGCGACTATGCGGAACAAGTCCGGTTCAAGCCGATTT
CCTCGTCAGCCGGGCAGGACATTCCTCTGTTCTACAGCGATGTCGTGAGCGTGAACTCCAAGACGG
CCCATCCGGAGCTGGCCAAAAAACTGGCGAATGTCATGGCTTCCGCGGATACGGTGGAGCAAGCT
CTGCGCCCGCAGGCCGACGGCCAATATCCTCAGTATCTGTTGCCTGCGCGGCATCAGGTCTACGAA
GCATTGATGCAGGATTATCCGATTTATTCCGAATTGGCCCAGATCGTGAACAAGCCGTCAAACCGG

GTGTTCCGGCTCGGACCTGAAGTCCGGACATGGCTAAAAGACGCCAAGCAAGTCTTGCCCGAAGC
GCTCGGCTTGACGGATGTATCGAGCCTGGCGAGCTGA

Protein sequence

MGSSHHHHHHSSGENLYFQGH**M**SKVKGFIYKPLMVMLALLLVVVSPAGAG|A|A|HSDASSDITLKVA
IYPYVPDPARFQAAVLDQWQRQEPGVKLEFTDWDSYSADPPDDLDVFVLDSIFLSHFVDAGYLLPFGSQ
DIDQAEDVLPFALQGAKRNGEVYGLPQILCTNLLFYRKGDLKIGQVDNIYELYKKIGTSHSEQIPPPQNKG
LLINMAGGTTKASMYLEALIDVTGQYTEYDLLPPLDPLNDKVIRGLRLLINMAGEKPSQYVPEDGDAYVR
ASWFAQGSGRAFIGYSESMMRMGDYAEQVRFKPISSSAGQDIPLFYSDVVSVNSKTAHPELAKKLANV
MASADTVEQALRPQADGQYPQYLLPARHQVYEALMQDYPIYSELAQIVNKPSNRVFRLGPEVRTWLKD
AKQVLPEALGLTDVSSLAS-

The residues at the N-terminus are the affinity tag and TEV protease sequence (ENLYFQ/G). The N-terminal sequence in red can be removed using TEV protease. The vertical lines | denote the sites where the putative N-terminal signal sequence is cleaved.

1.1.10 PfThiE

Gene Sequence

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCGAGAATCTTTATTTTCAGGGCCATATGA
ACCTAAGGAACAAGCTGAAGCTCTACGTTATAACAGACAGAAGACTCAAGCCCGAAGTTGAGAGT
GTAAGGGAAGCCCTGGAAGGAGGGGCTACTGCAATTCAGATGAGGATAAAGAACGCCCCAACAA
GAGAAATGTATGAGATTGGAAAGACTTTAAGACAATTAACCAGGGAATACGACGCCCTTTTCTTCG
TCGATGATAGGGTAGATGTAGCTTTAGCTGTCGATGCCGATGGAGTTCAGCTGGGGCCCGAGGAC
ATGCCAATAGAAGTAGCTAAGGAAATAGCACCCAATCTAATTATAGGAGCATGTGTATACTCCCTTG
AAGAGGCCCTGGAAGCTGAGAAAAAGGGAGCTGACTACTTAGGAGCTGGCTCCGTCTTTCCAACTA
AGACAAAGGAAGATGCGAGAGTTATAGGCCTAGAGGGATTGAGAAAGATAGTTGAGAGCGTTAA
AATTCCCGTAGTTGCAATAGGAGGAATAAATAAAGATAATGCGAGAGAAGTTCTCAAGACTGGAGT
TGACGGAATAGCCGTAATATCCGCTGTTATGGGTGCTGAAGATGTTAGAAAGGCTACAGAGGAGTT
AAGAAAAATCGTAGAGGAGGTTTTAGGATGA

Protein sequence

MGSSHHHHHHSSGENLYFQGH**M**NLRNKLKLYVITDRRLKPEVESVREALEGGATAIQMRIKNAPTREM
YEIGKTLRQLTREYDALFFVDDRVDVALAVDADGVQLGPEDMPIEVAKEIAPNLIIGACVYSLEEALEAEK
KGADYLGAGSVFPTKTKEDARVIGLEGLRKIVESVKIPVVAIGGINKDNAREVLKTGVDGIAVISAVMGAE
DVRKATEELRKIVEEVLG*

1.1.11 PfBAL

Gene Sequence

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCGAGAATCTTTATTTTCAGGGCCATATG
GCGATGATTACAGGCGGCGAACTGGTTGTTCGCACCCTAATAAAGGCTGGGGTCGAACATCT
GTTCGGCCTGCACGGCGCGCATATCGATACGATTTTTCAAGCCTGTCTCGATCATGATGTGCCGATC
ATCGACACCCGCCATGAGGCCGCCGCAGGGCATGCGGCCGAGGGCTATGCCCGCGCTGGCGCCAA
GCTGGGCGTGGCGCTGGTCACGGCGGGCGGGGGATTTACCAATGCGGTCACGCCCATTGCCAACG
CTTGGCTGGATCGCACGCCGGTGCTCTTCCTCACCGGATCGGGCGCGCTGCGTGATGATGAAACCA
ACACGTTGCAGGCGGGGATTGATCAGGTCGCCATGGCGGCGCCCATTACCAAATGGGCGCATCGG
GTGATGGCAACCGAGCATATCCCACGGCTGGTGATGCAGGCGATCCGCGCCGCGTTGAGCGCGCC
ACGCGGGCCGGTGTTGCTGGATCTGCCGTGGGATATTCTGATGAACCAGATTGATGAGGATAGCGT
CATTATCCCCGATCTGGTCTTGTCCGCGCATGGGGCCAGACCCGACCCTGCCGATCTGGATCAGGCT
CTCGCGCTTTTGCGCAAGGCGGAGCGGCCGGTCATCGTGCTCGGCTCAGAAGCCTCGCGGACAGC
GCGCAAGACGGCGCTTAGCGCCTTCGTGGCGGCGACTGGCGTGCCGGTGTTTGCCGATTATGAAG
GGCTAAGCATGCTCTCGGGGCTGCCCGATGCTATGCGGGGCGGGCTGGTGCAAAACCTCTATTCTT

TTGCCAAAGCCGATGCCGCGCCAGATCTCGTGCTGATGCTGGGGGCGCGCTTTGGCCTTAACACCG
GGCATGGATCTGGGCAGTTGATCCCCCATAGCGCGCAGGTCATTCAGGTCGACCCTGATGCCTGCG
AGCTGGGACGCCTGCAGGGCATCGCTCTGGGCATTGTGGCCGATGTGGGTGGGACCATCGAGGCT
TTGGCGCAGGCCACCGCGCAAGATGCGGCTTGGCCGGATCGCGGCGACTGGTGCGCCAAAGTGAC
GGATCTGGCGCAAGAGCGCTATGCCAGCATCGCTGCGAAATCGAGCAGCGAGCATGCGCTCCACC
CCTTTCACGCCTCGCAGGTCATTGCCAAACACGTCGATGCAGGGGTGACGGTGGTAGCGGATGGTG
CGCTGACCTATCTCTGGCTGTCCGAAGTGATGAGCCGCGTGAAACCCGGCGGTTTTCTCTGCCACG
GCTATCTAGGCTCGATGGGCGTGGGCTTCGGCACGGCGCTGGGCGCGCAAGTGGCCGATCTTGAA
GCAGGCCGCCGCACGATCCTTGTGACCGGCGATGGCTCGGTGGGCTATAGCATCGGTGAATTTGAT
ACGCTGGTGCGCAAACAATTGCCGCTGATCGTCATCATCATGAACAACCAAAGCTGGGGGGCGACA
TTGCATTTCCAGCAATTGGCCGTCGGCCCCAATCGCGTGACGGGCACCCGTTTGGAAAATGGCTCCT
ATCACGGGGTGGCCGCCGCCTTTGGCGCGGATGGCTATCATGTCGACAGTGTGGAGAGCTTTTCTG
CGGCTCTGGCCCAAGCGCTCGCCCATAATCGCCCCGCCTGCATCAATGTCGCGGTCGCGCTCGATCC
GATCCCGCCCGAAGAACTCATTCTGATCGGCATGGACCCCTTCGCACTCGAGTGA

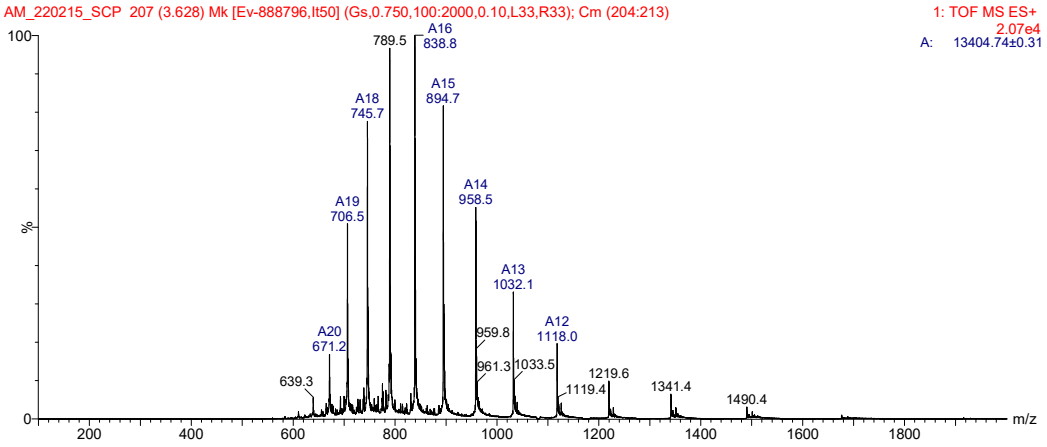
Protein sequence

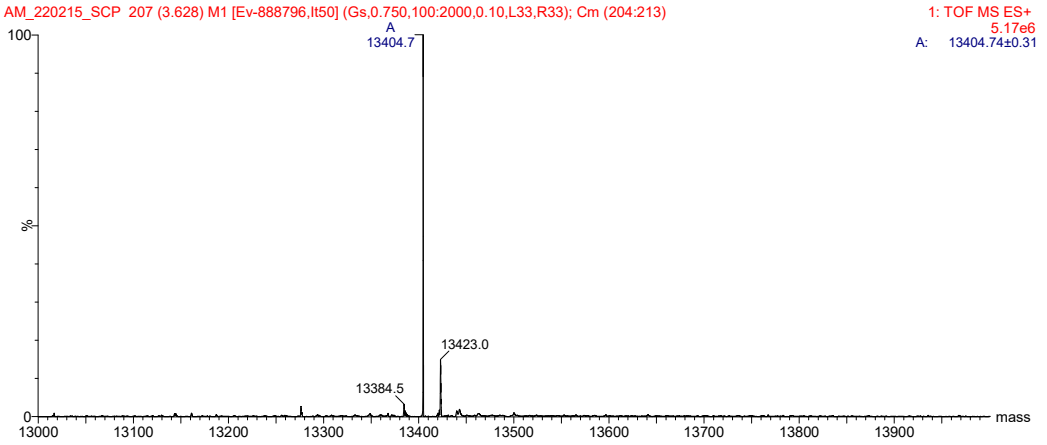
MGSSHHHHHHSSGENLYFQGH**M**AMITGGELVVRTLIKAGVEHLFGLHGAHIDTIFQACLDHDVPIIDTR
HEAAAGHAAEGYARAGAKLGVALVTAGGGFTNAVTPIANAWLDRTPVLFLTGSGALRDDETNTLQAGI
DQVAMAAPITKWAHRVMATEHIPRLVMQAIRAALSAPRGPVLLDLPWDILMNQIDEDSVIIPDLVLSA
HGARPDPADLDQALALLRKAERPVIVLGSEASRTARKTALSAFVAATGVPVFADYEGLSMLSGLPDAMR
GGLVQNLYSFAKADAAPDLVLMLGARFGLNTGHGSGQLIPHSAQVIQVDPDACELGRLQGIALGIVAD
VGGTIEALAQATAQDAAWPDRGDWCAKVTDLAQERYASIAAKSSSEHALHPFHASQVIAKHVDAGVT
VVADGALTYLWLSEVMSRVKPGGFLCHGYLGSMGVGFGTALGAQVADLEAGRRTILVTGDGSVGYSIG
EFDTLVRKQLPLIVIIMNNQSWGATLHFQQLAVGPNRVTGTRLENGSYHGVAAAFGADGYHVDSVESF
SAALAQALAHNRPACINVAVALDPIPPEELILIGMDPFALE

1.2 Protein LC-MS

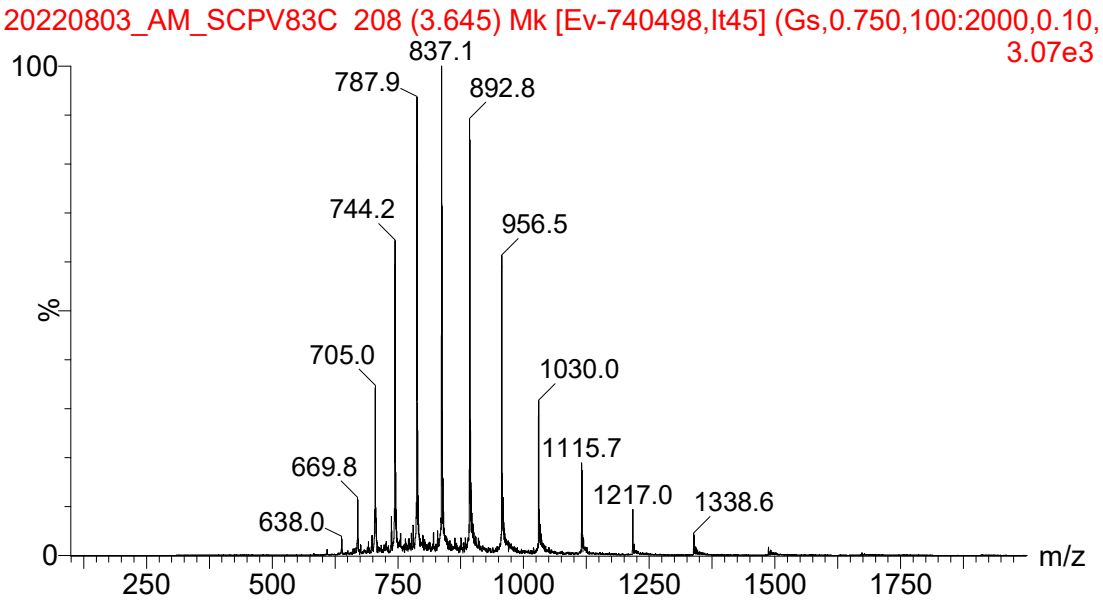
1.2.1 hSCP A100C

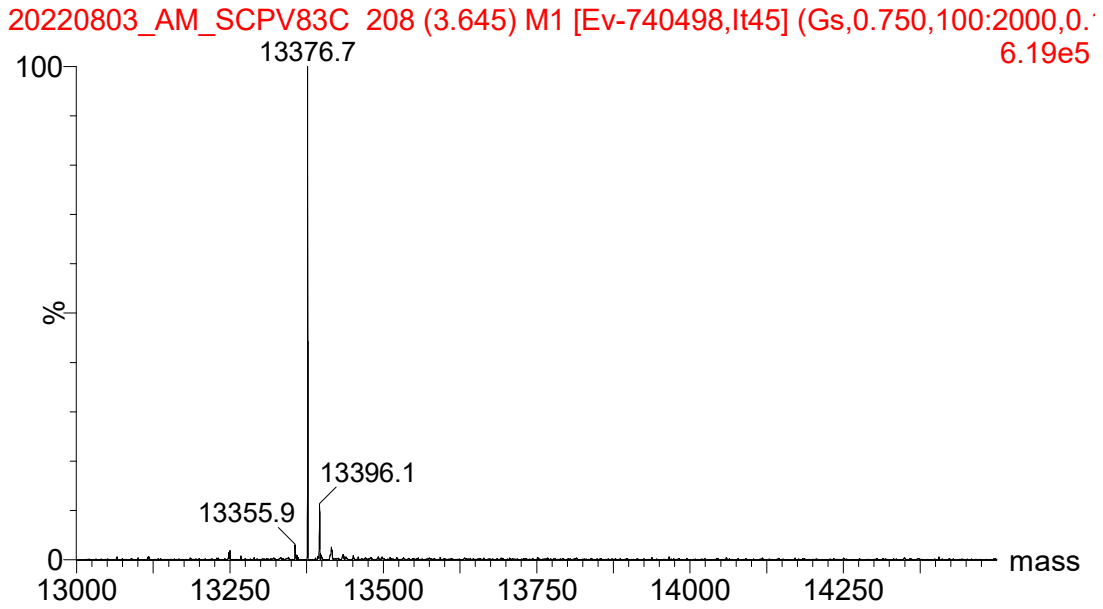
Expected mass: 13404.66 Mass Found: 13404.74



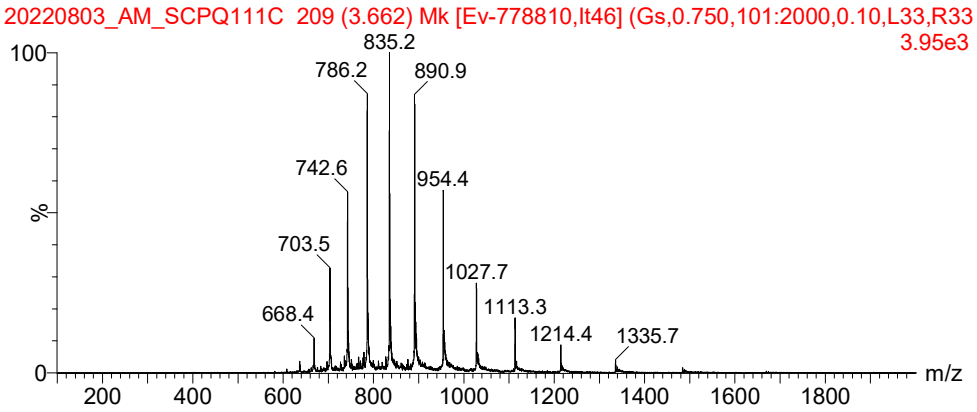


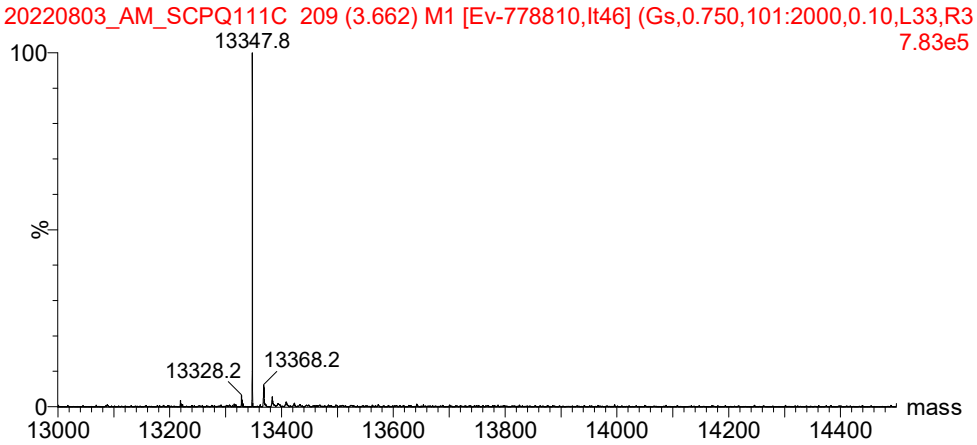
1.2.2 hSCP V83C



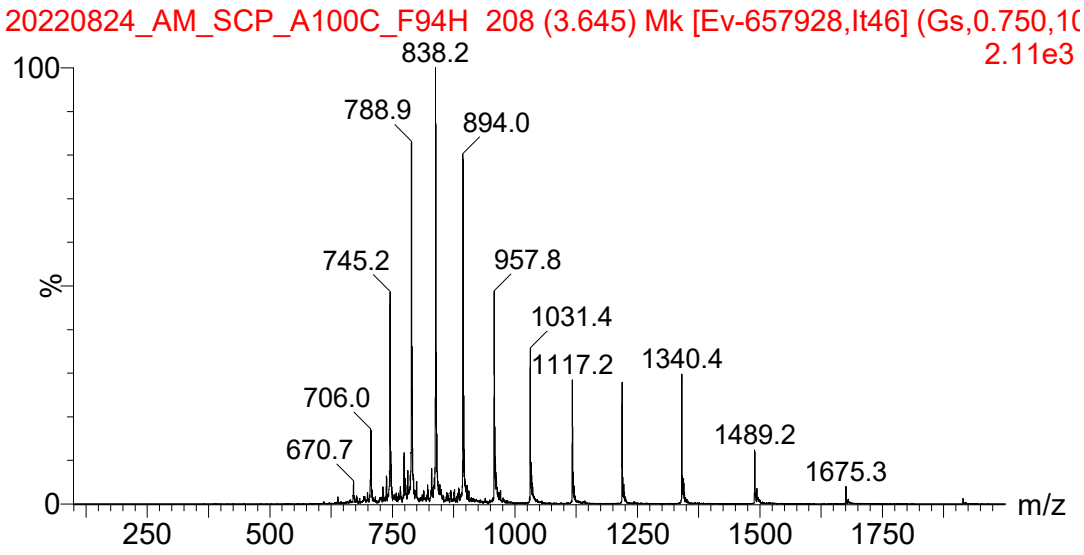


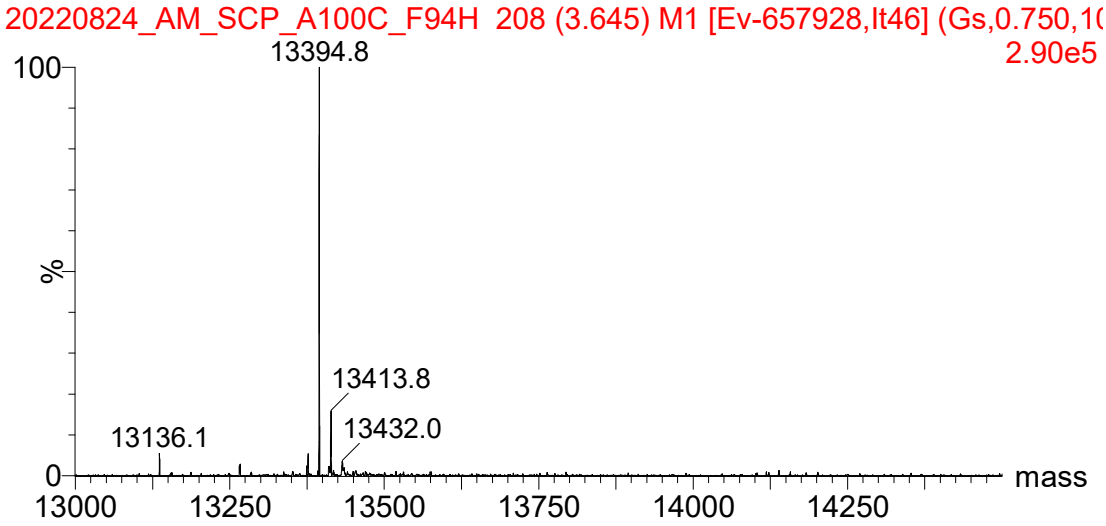
1.2.3 hSCP Q111C



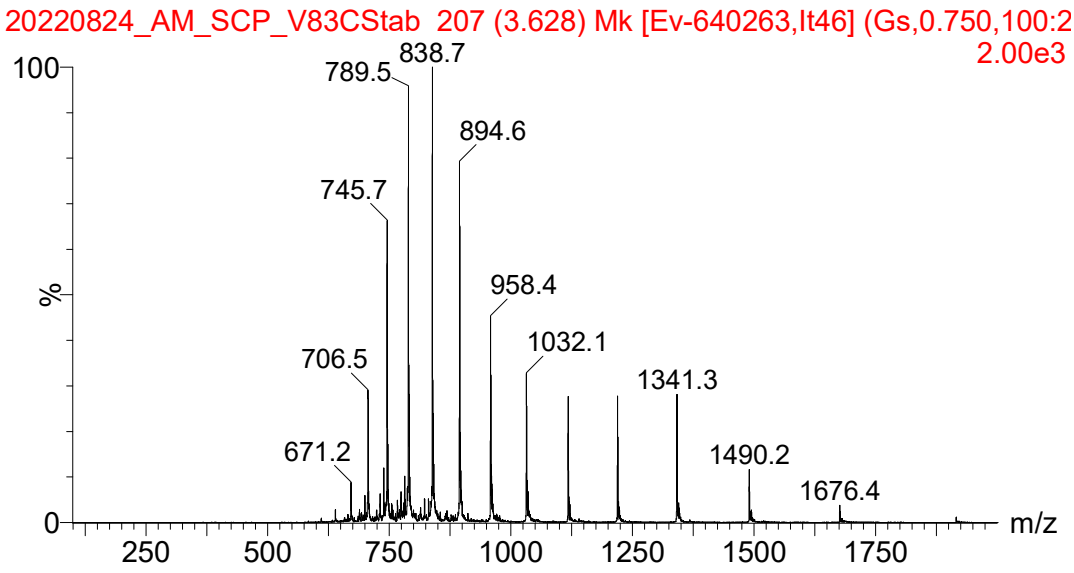


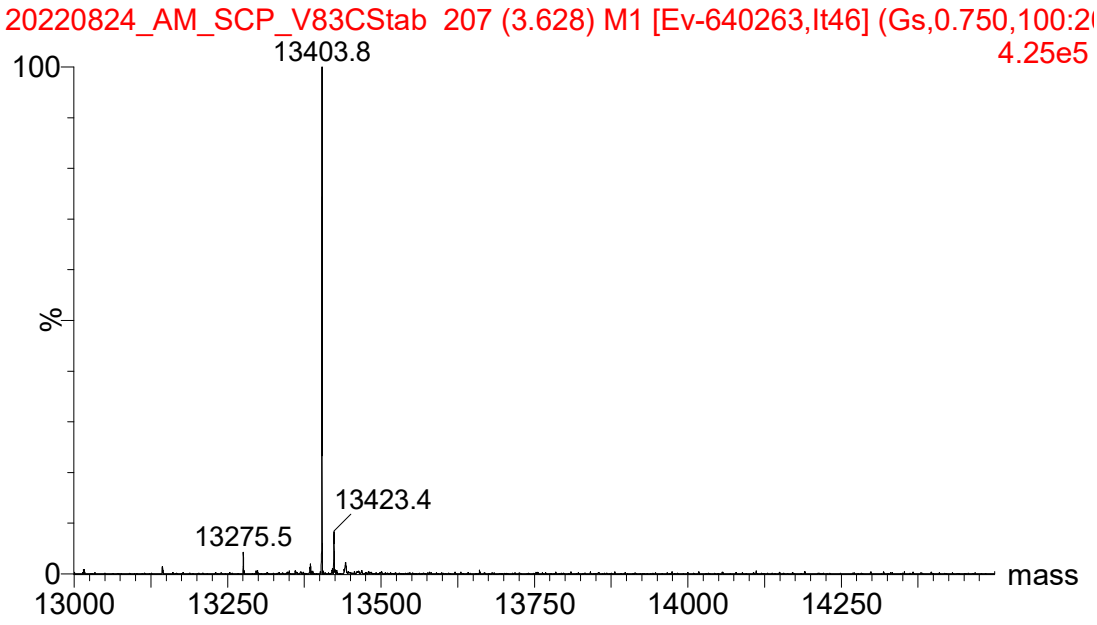
1.2.4 hSCP A100C F94H





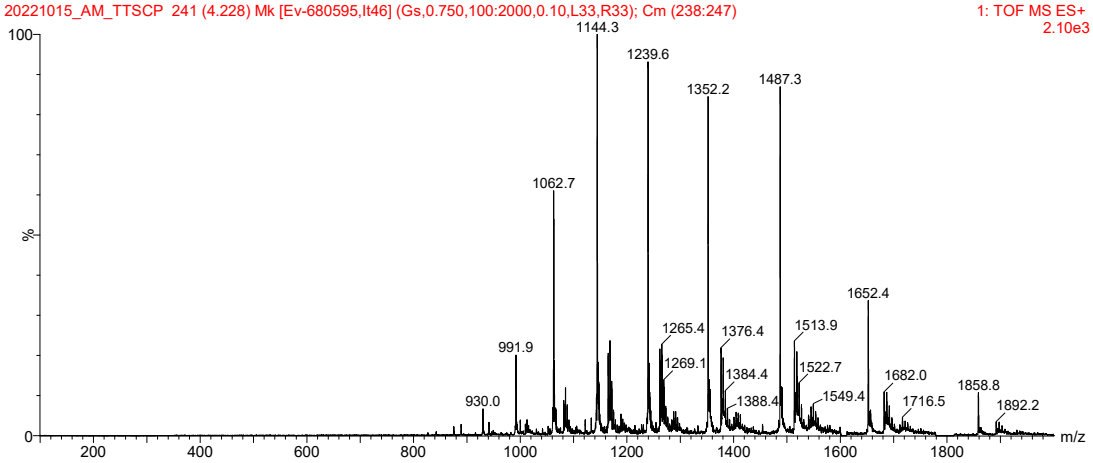
1.2.5 hSCP V83C Stable

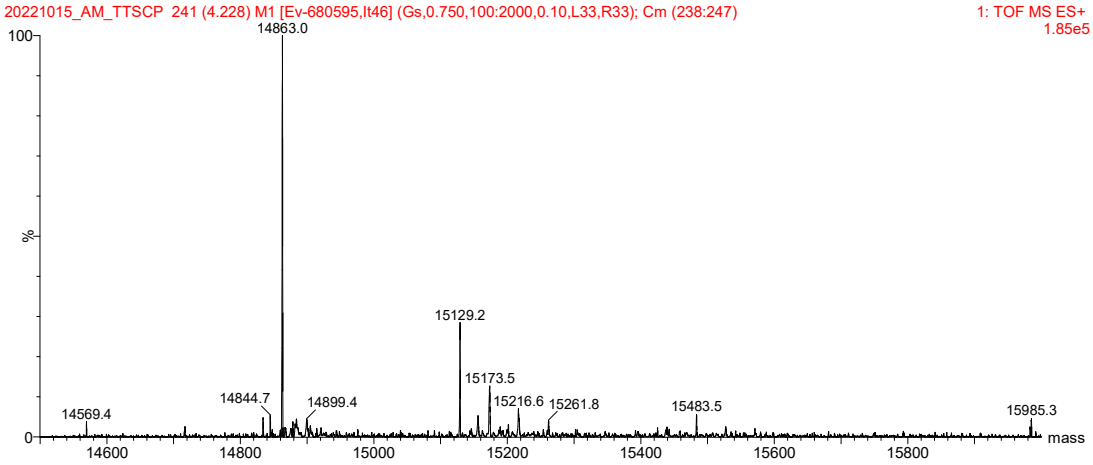




1.2.7 TTSCP

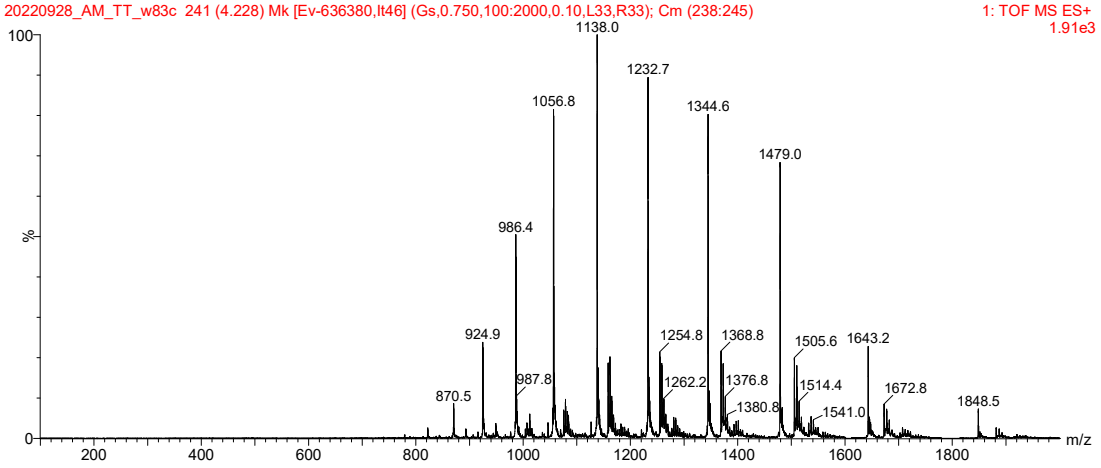
Expected mass: 14865.04 Mass Found 14863.0 (accounts for disulphide bond formation)

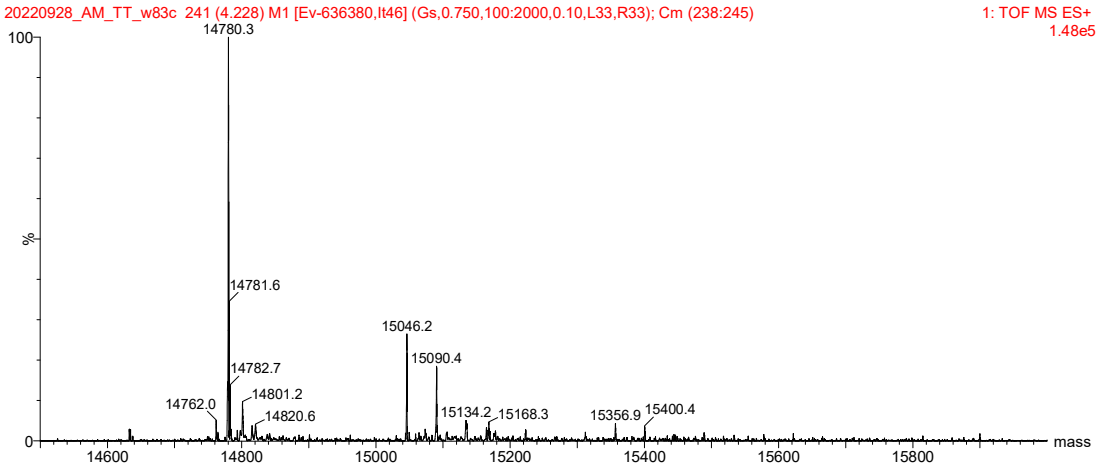




1.2.8 TTSCP W83C

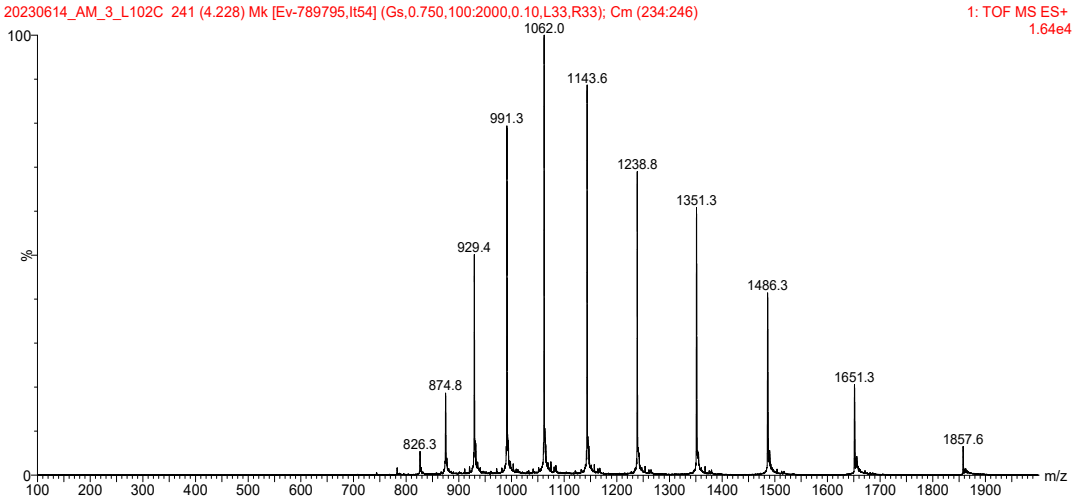
Expected mass: 14781.9 Mass Found 14780.3 (accounts for disulphide bond formation)

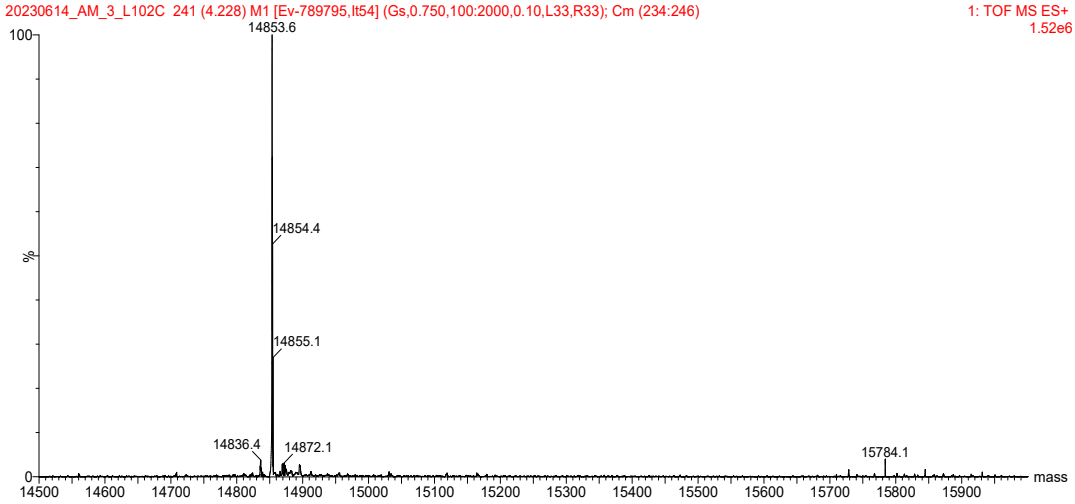




1.2.9 TTSCP L102C

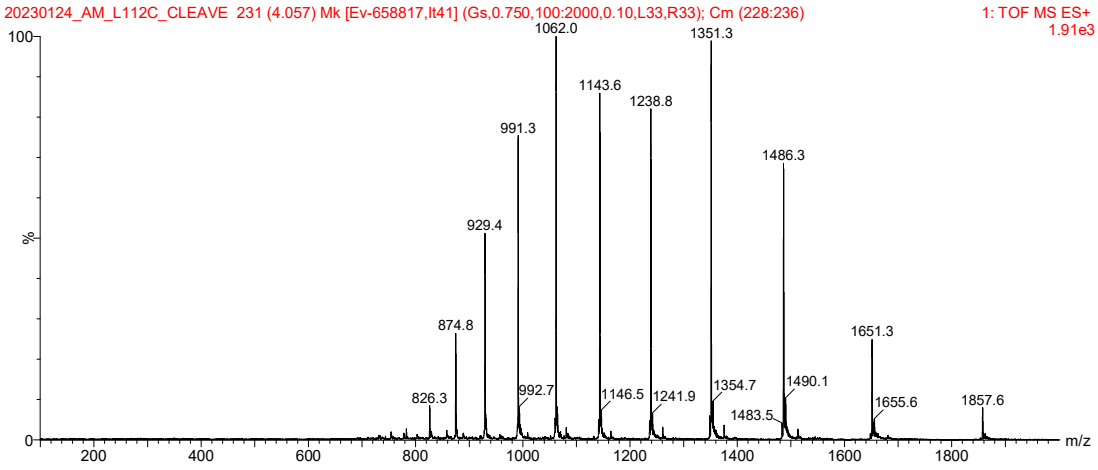
Expected mass: 14855.02 Mass Found 14853.6 (accounts for disulphide bond formation)

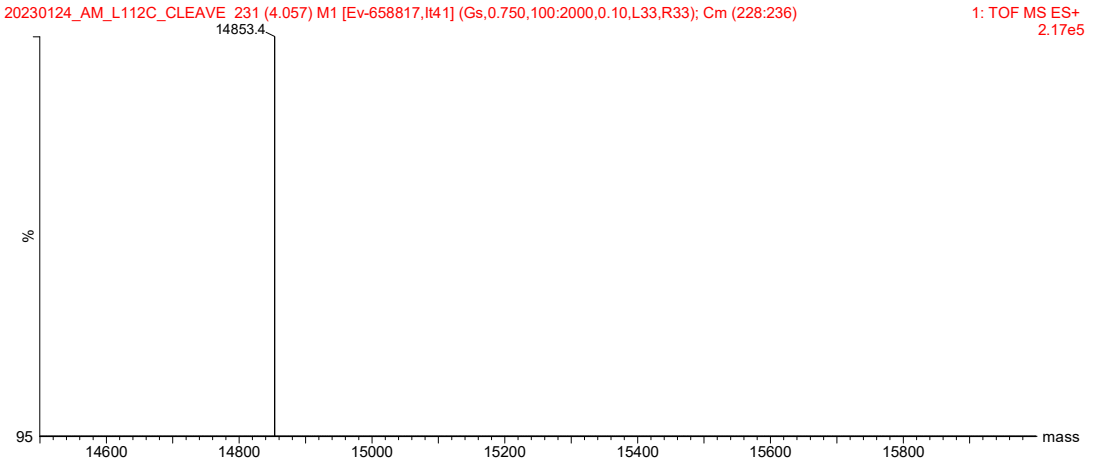




1.2.10 TTSCP L112C

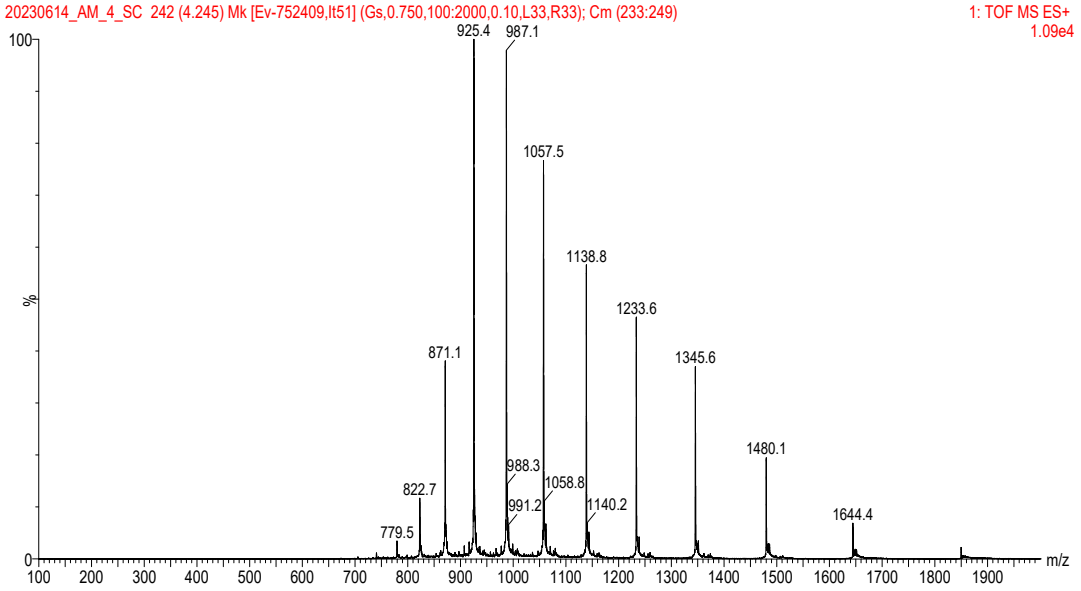
Expected mass: 14855.02 Mass Found 14853.4 (accounts for disulphide bond formation)

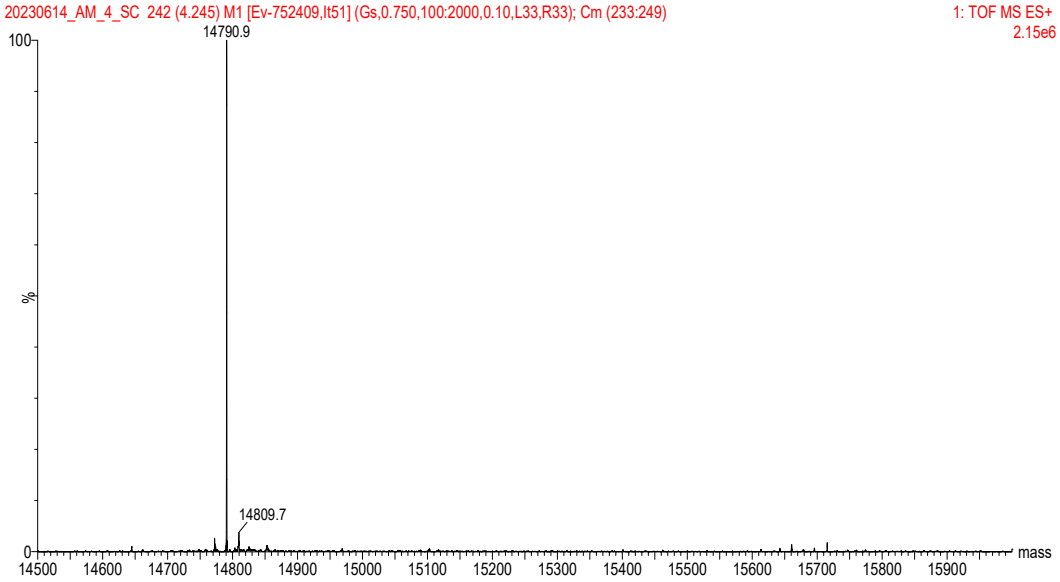




1.2.11 TTSCP ΔDSB L102C

Expected mass: 14790.90 Mass Found 14790.9

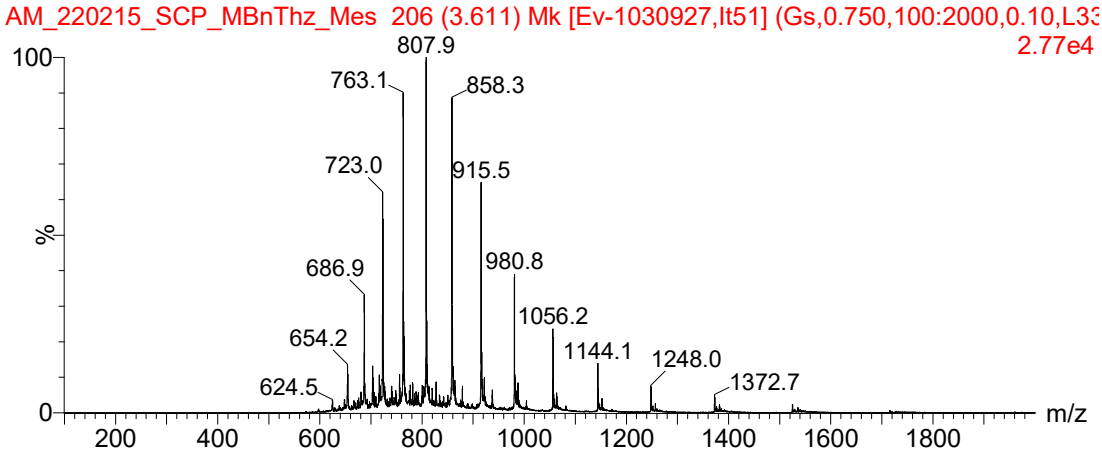


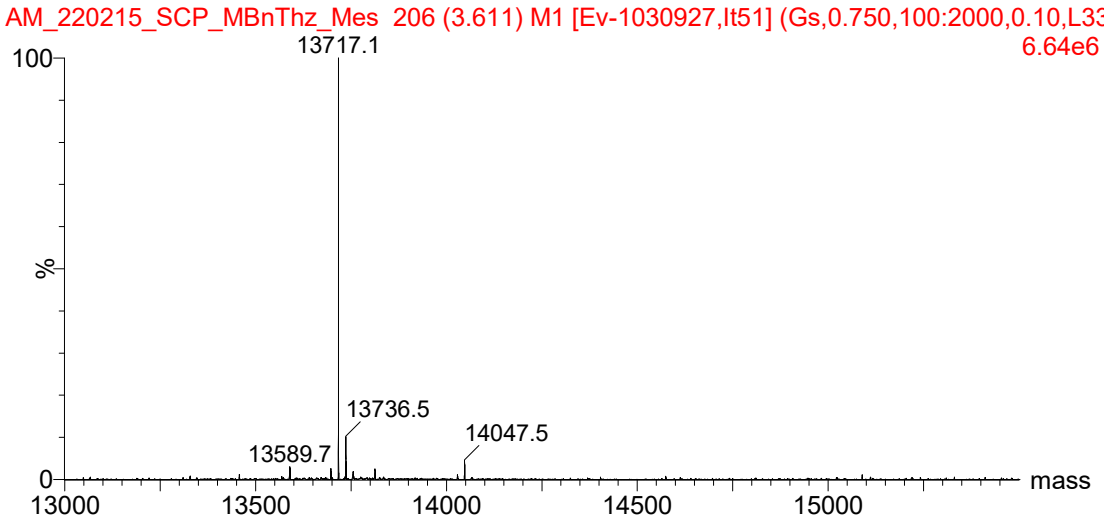


1.3 Protein LC-MS Functionalised with MBnTHz

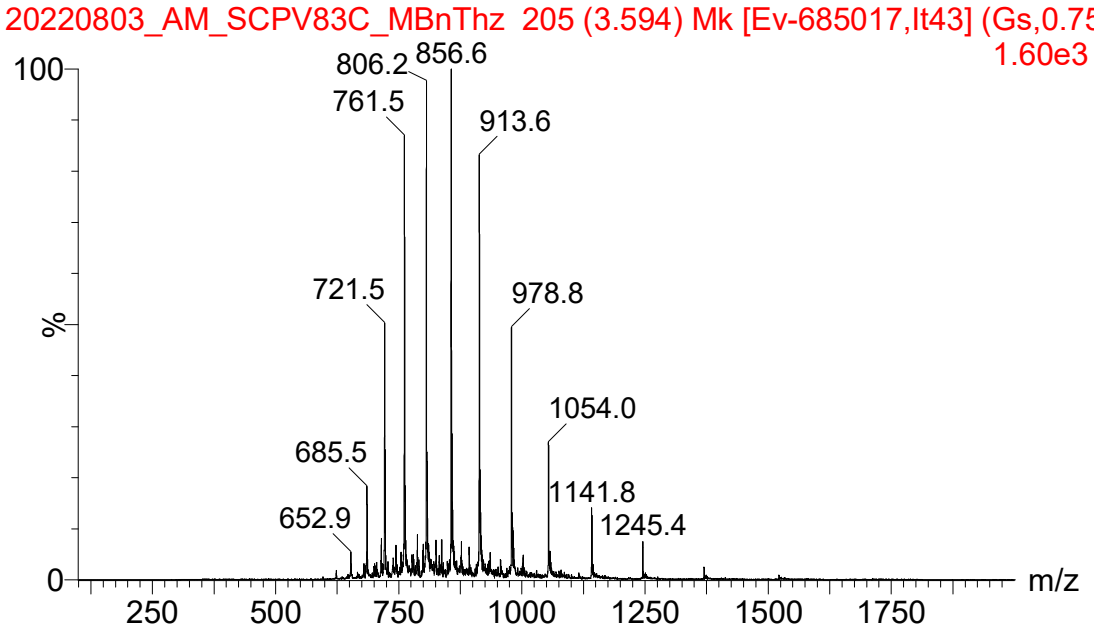
1.3.1 hSCP A100C MBnThz

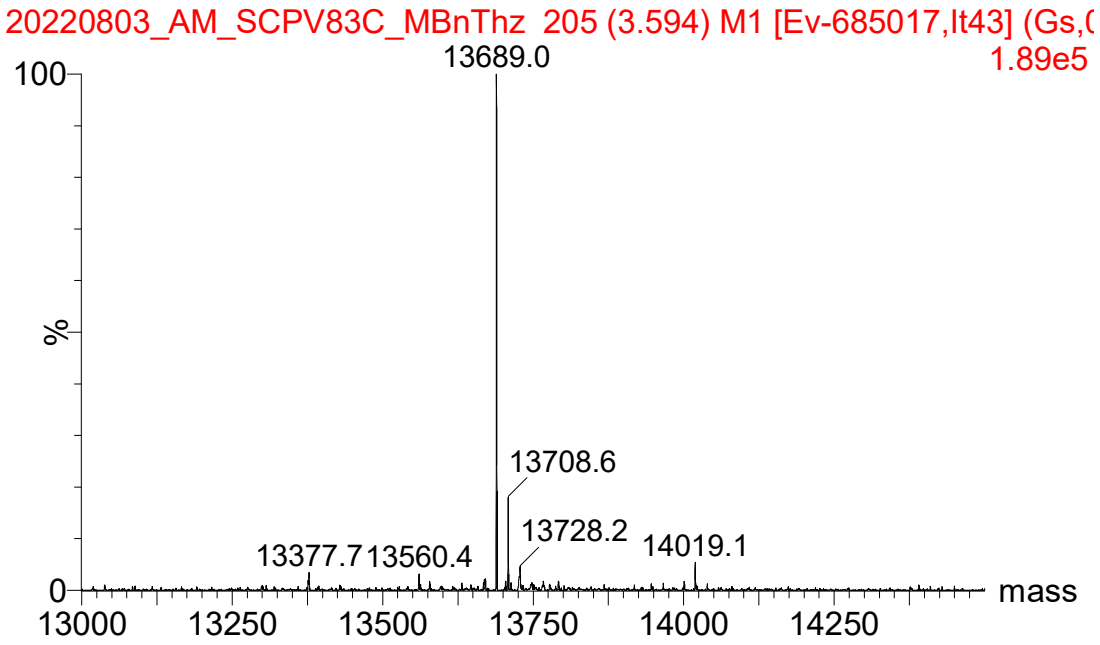
Expected mass: 13717.66 Mass Found 13717.1 Mono-functionalisation 100%



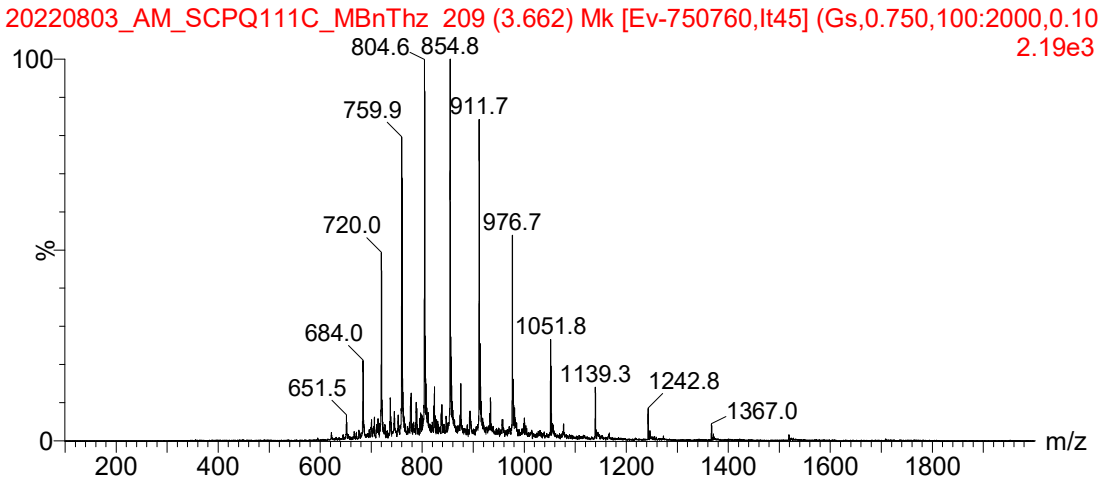


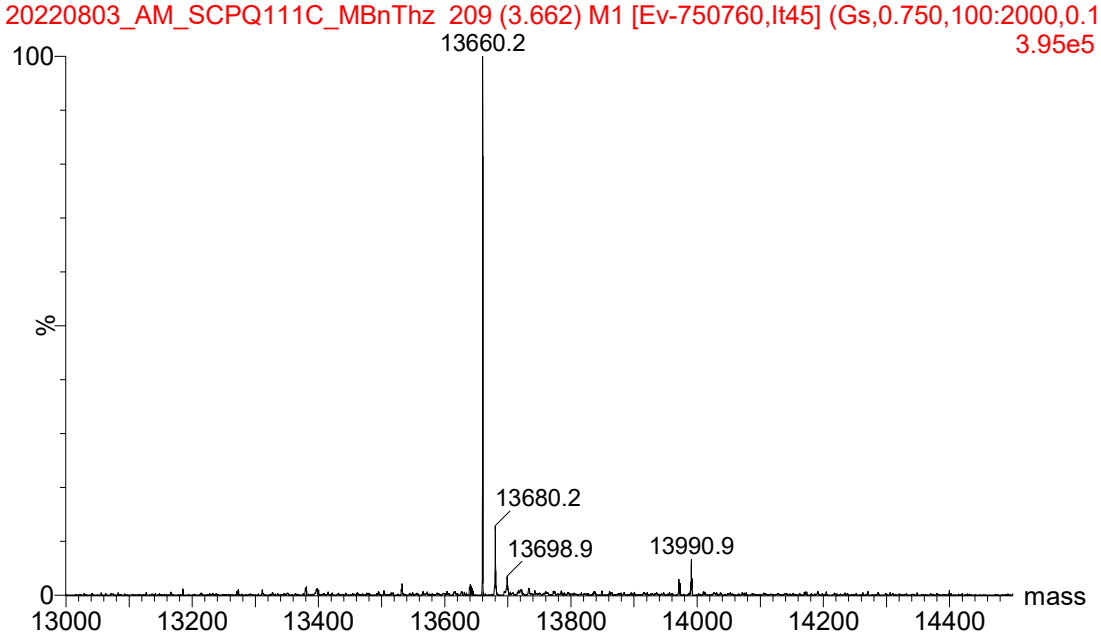
1.3.2 hSCP V83C MBnThz



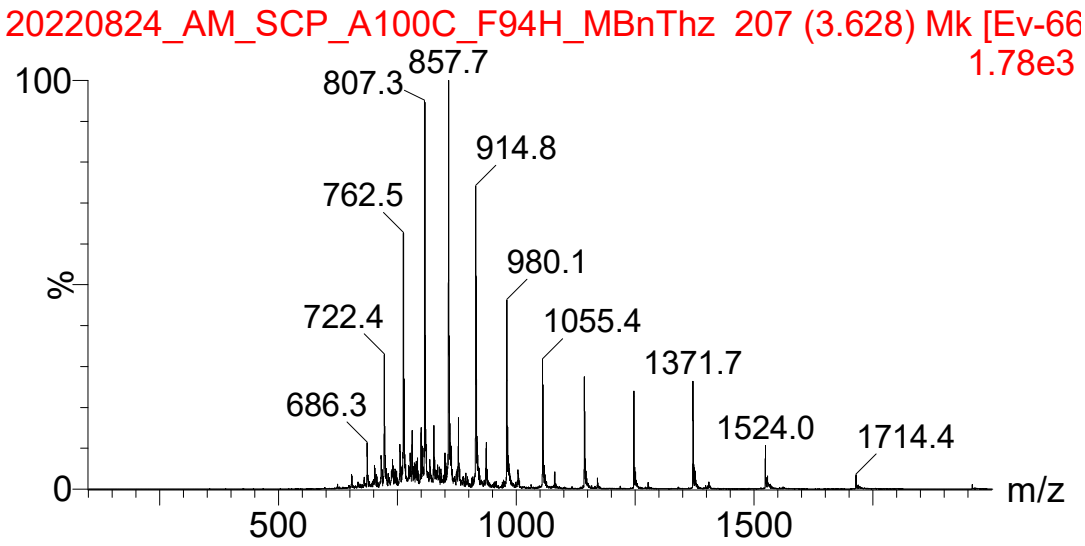


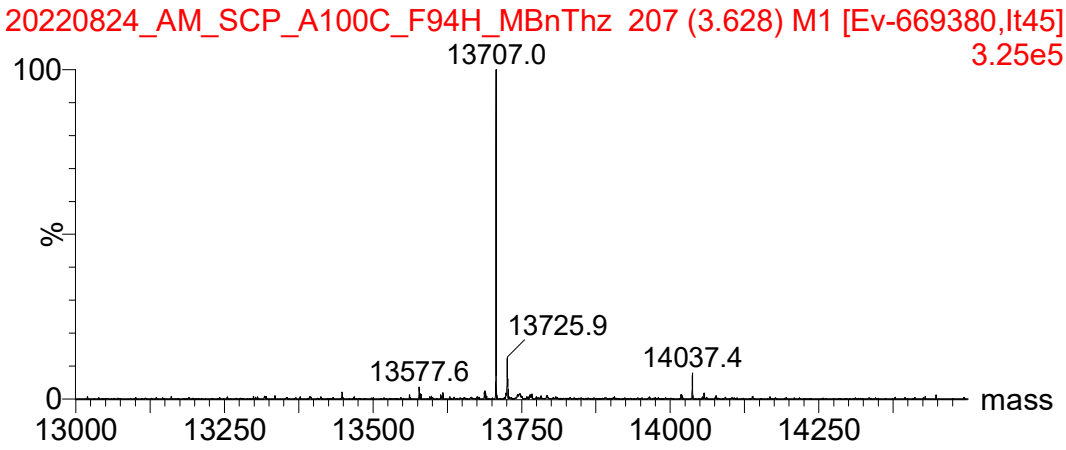
1.3.3 hSCP Q111C MBnThz



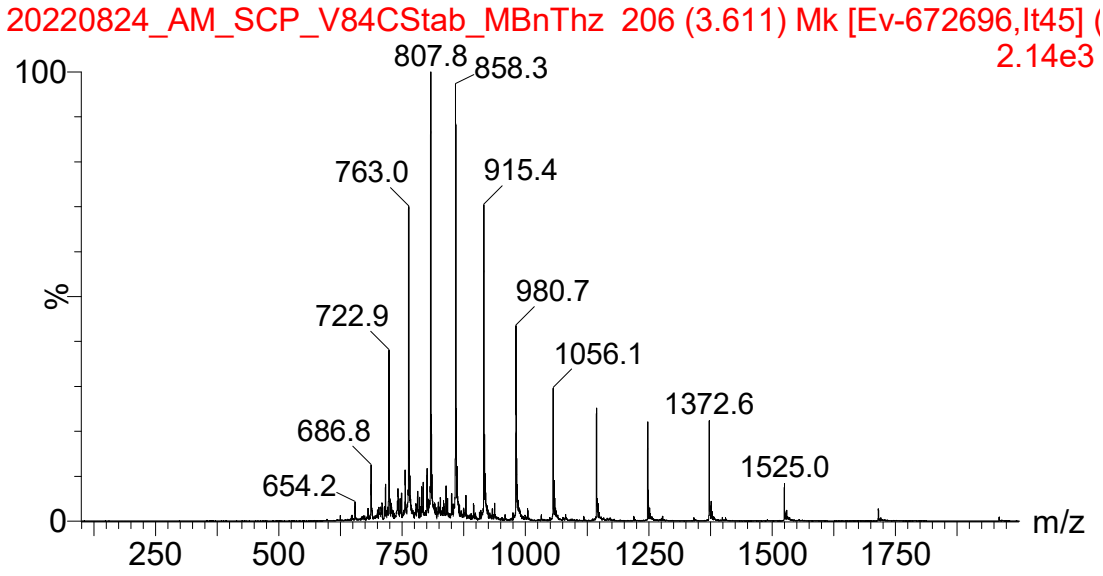


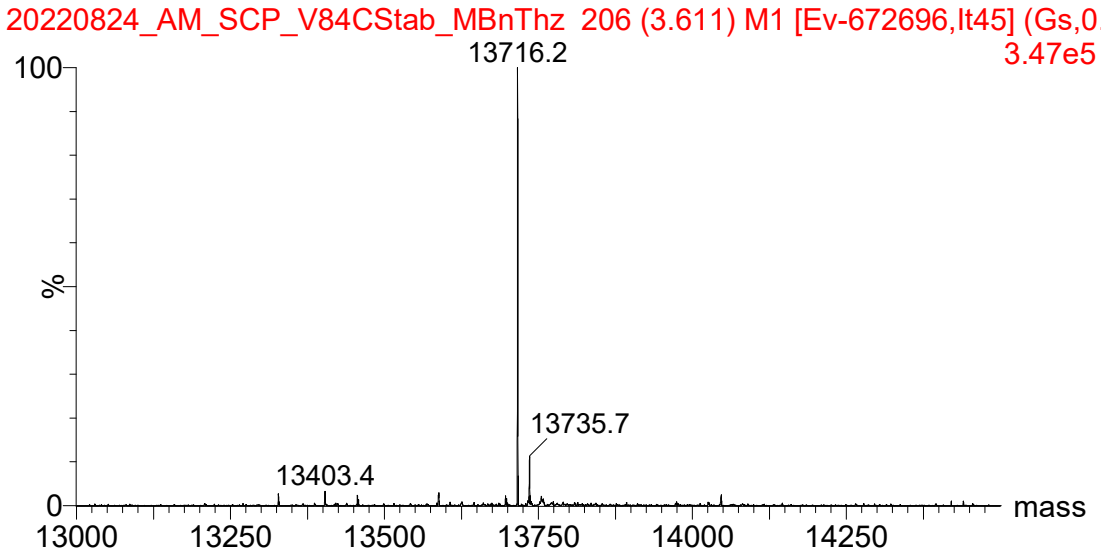
1.3.4 hSCP A100C F94H MBnThz





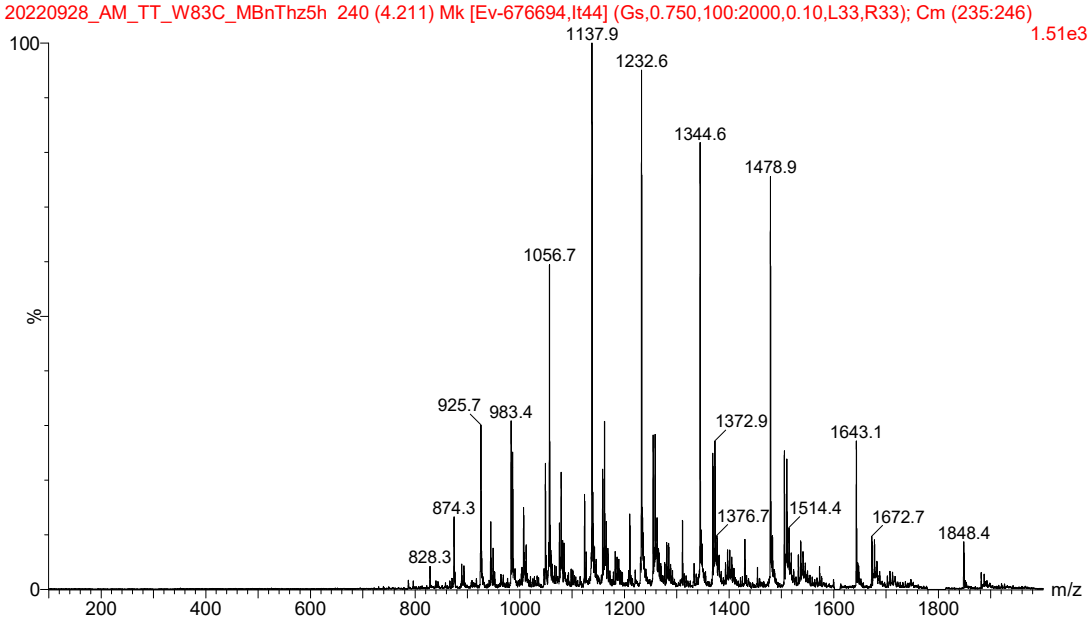
1.3.5 hSCP V83C Stable MBnThz

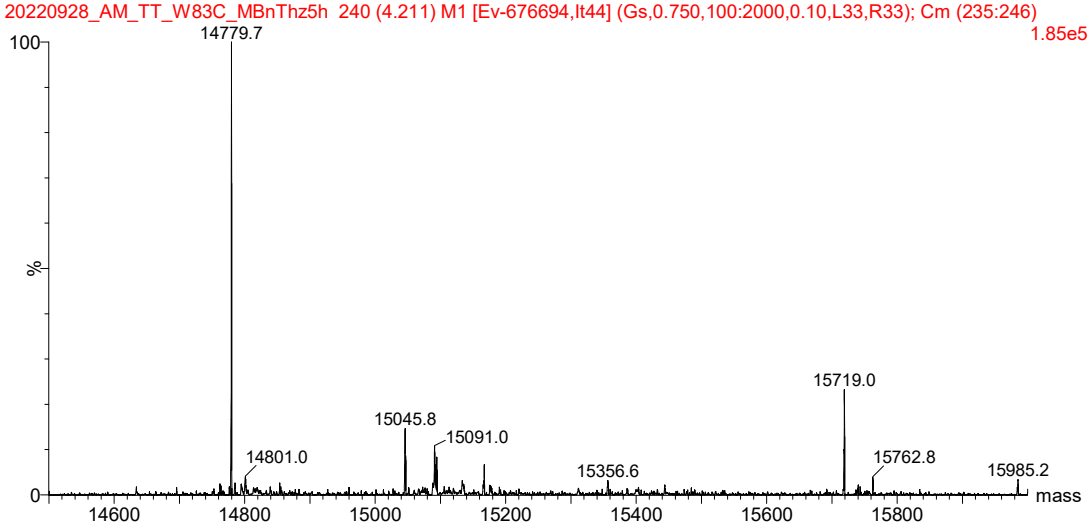




1.3.6 TTSCP W83C MBnThz

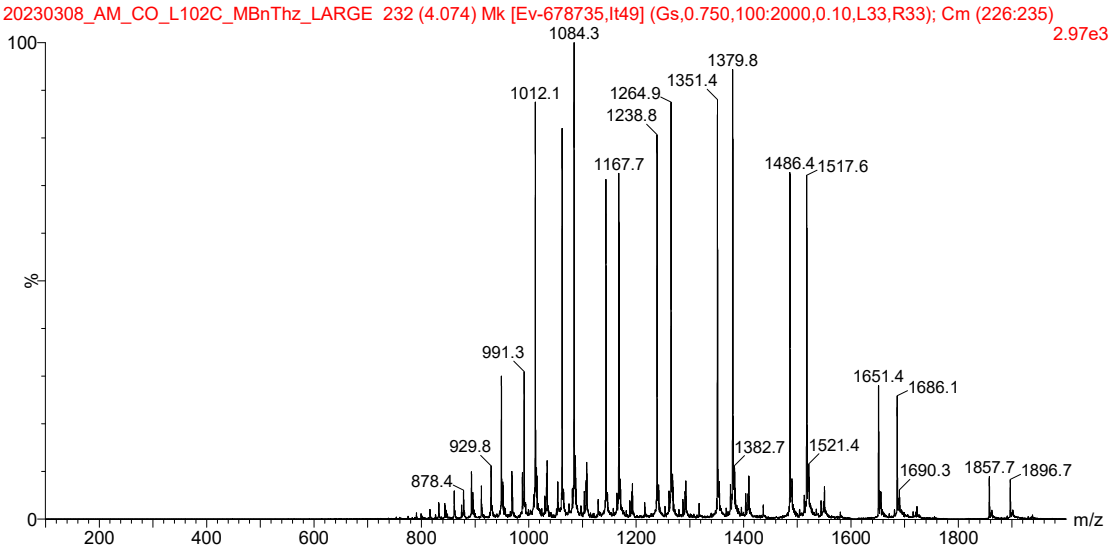
10 equiv 5 h RT 20 mM Mes pH6

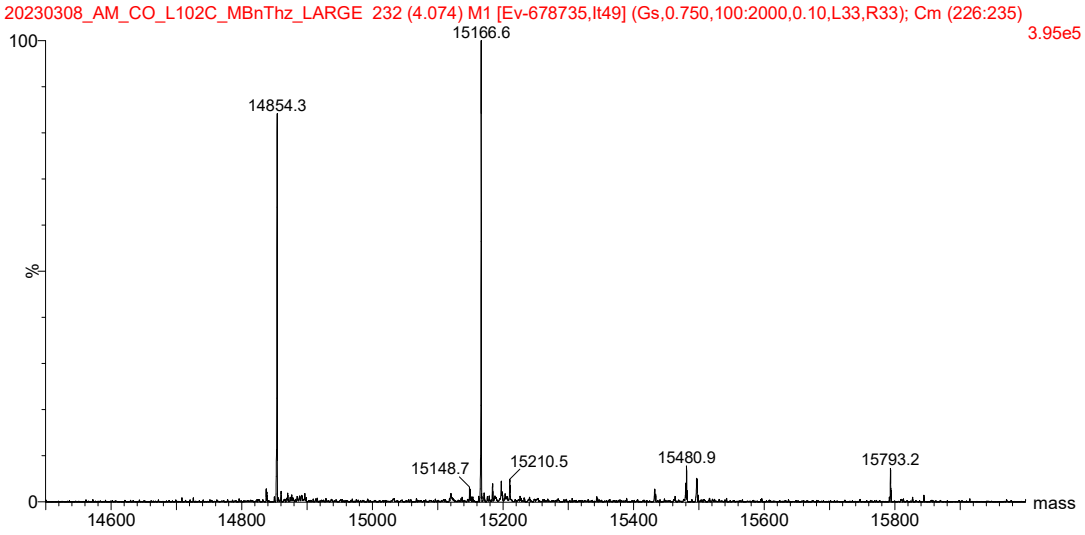

20220928_AM_TT_W83C_MBnThz5h 240 (4.211) Mk [Ev-676694,It44] (Gs,0.750,100:2000,0.10,L33,R33); Cm (235:246)
1.51e3
100
%
0
1137.9
1232.6
1344.6
1478.9
1056.7
925.7
983.4
1372.9
1643.1
874.3
1376.7
1514.4
1672.7
1848.4
828.3
200
400
600
800
1000
1200
1400
1600
1800
m/z


20220928_AM_TT_W83C_MBnThz5h 240 (4.211) M1 [Ev-676694,It44] (Gs,0.750,100:2000,0.10,L33,R33); Cm (235:246)
1.85e5
100
%
0
14779.7
15719.0
15045.8
15091.0
14801.0
15762.8
15356.6
15985.2
14600
14800
15000
15200
15400
15600
15800
mass

1.3.7 TTSCP L102C MBnThz

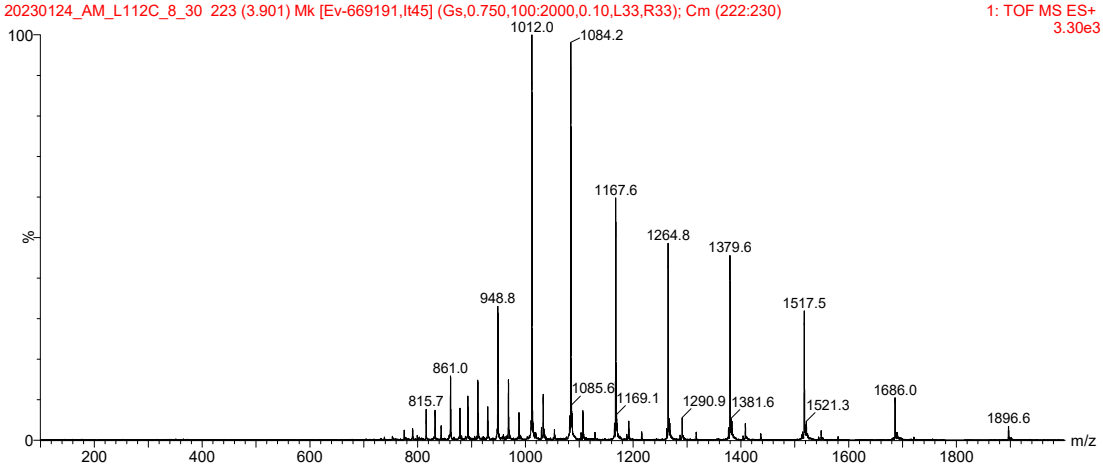
Expected mass 15168.02 Mass Found 15166.6 (accounts for disulphide bond still present) Mono-functionalisation Percentage: 50%

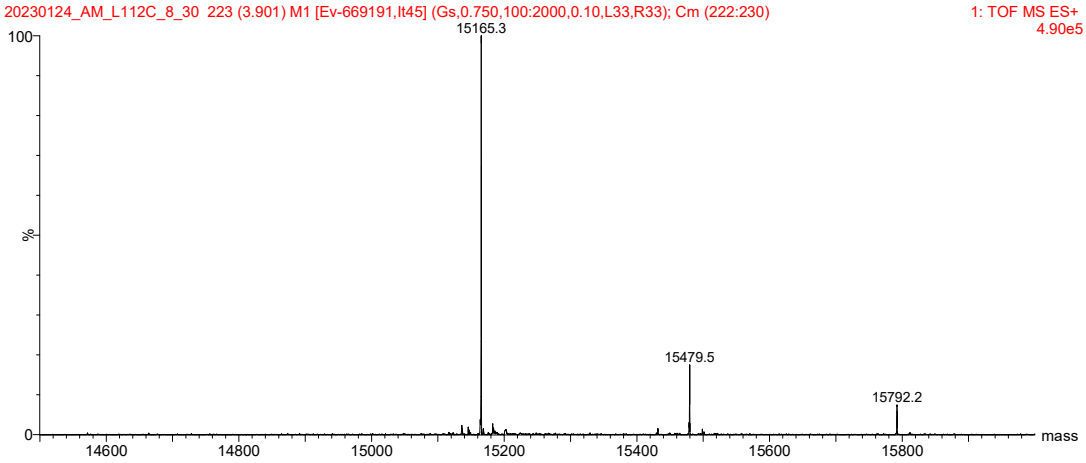




1.3.8 TTSCP L112C MBnThz

Expected mass 15168.02 Mass Found 15165.3 (accounts for disulphide bond still present) Mono-functionalisation Percentage: 85%

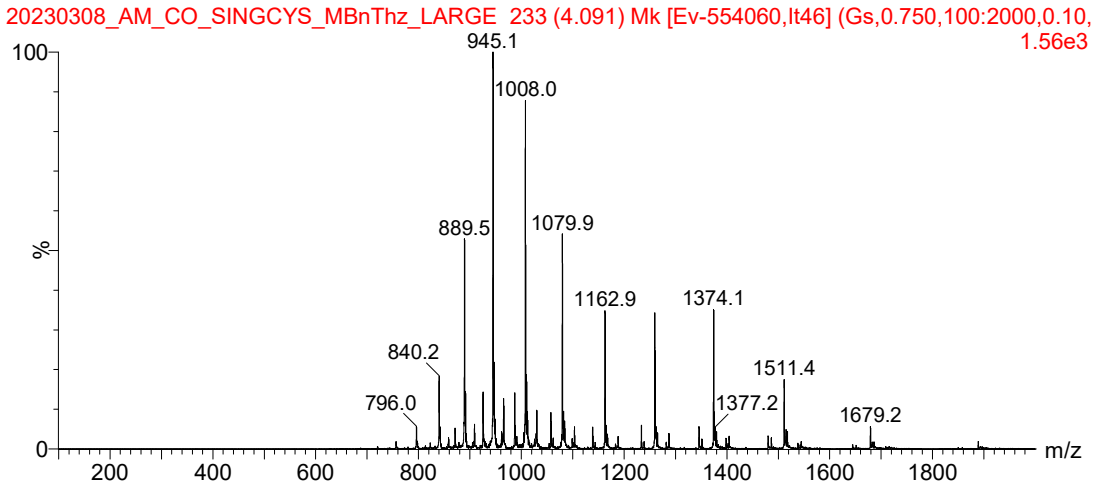


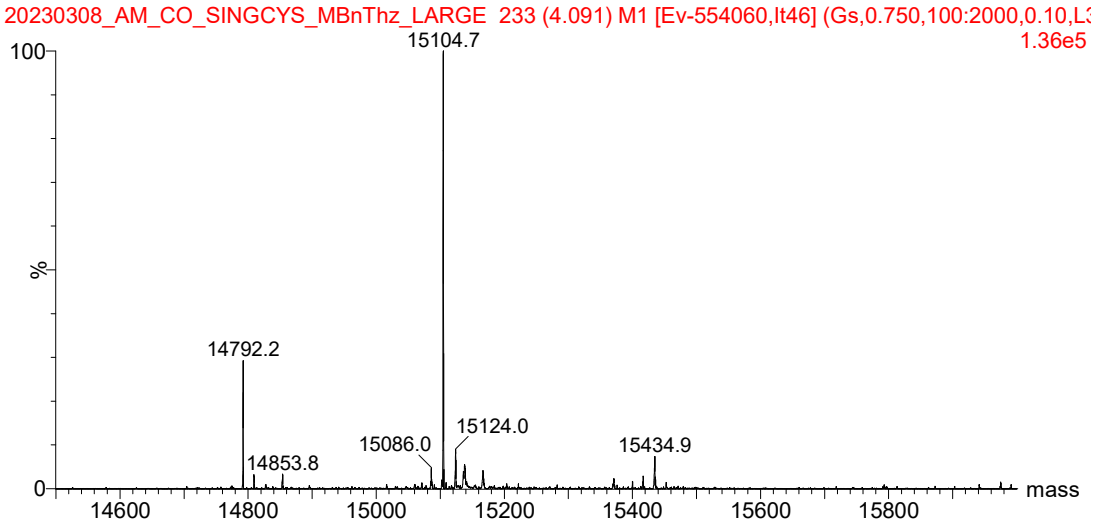


1.3.9 TTSCP ΔDSB L102C MBnThz

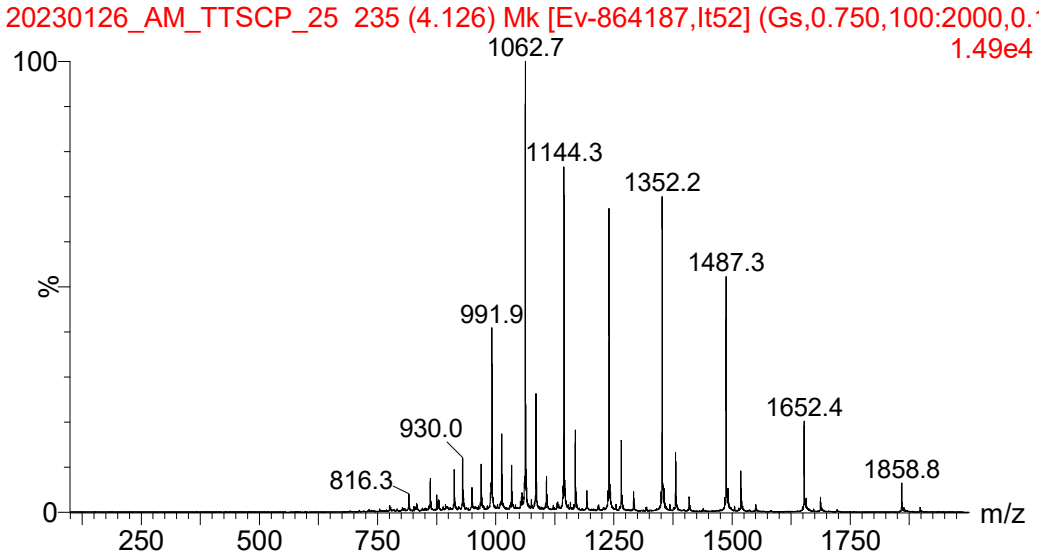
Expected mass 15103.90 Mass Found 15104.7 Mono-functionalisation

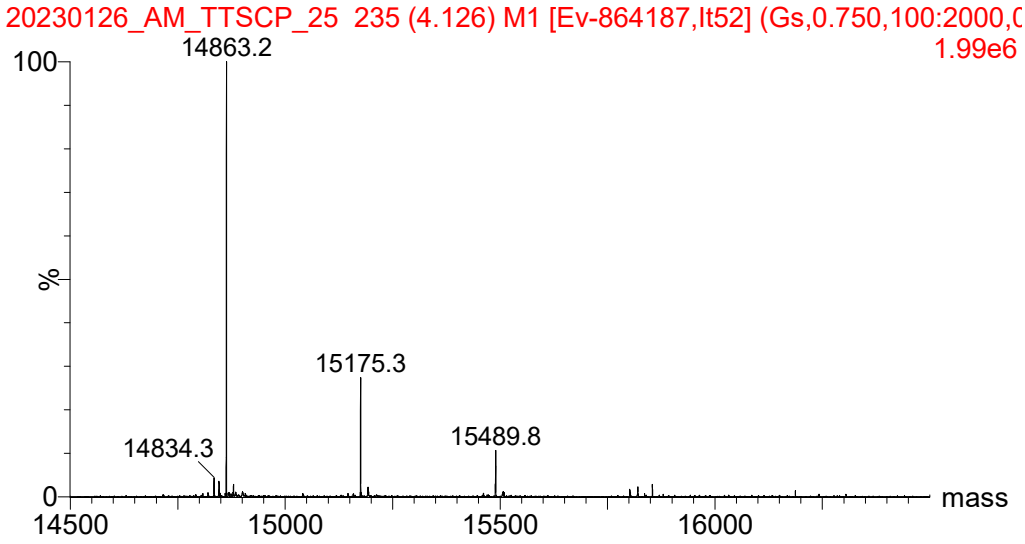
Percentage: 75%





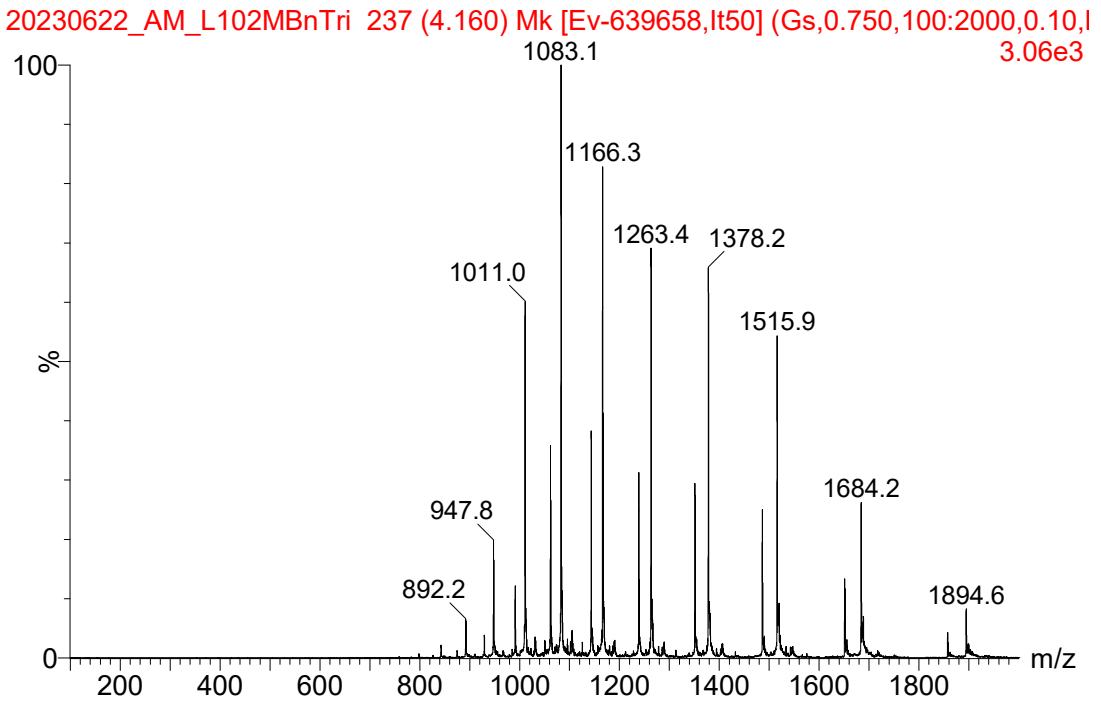
1.3.10 TTSCP MBnThz

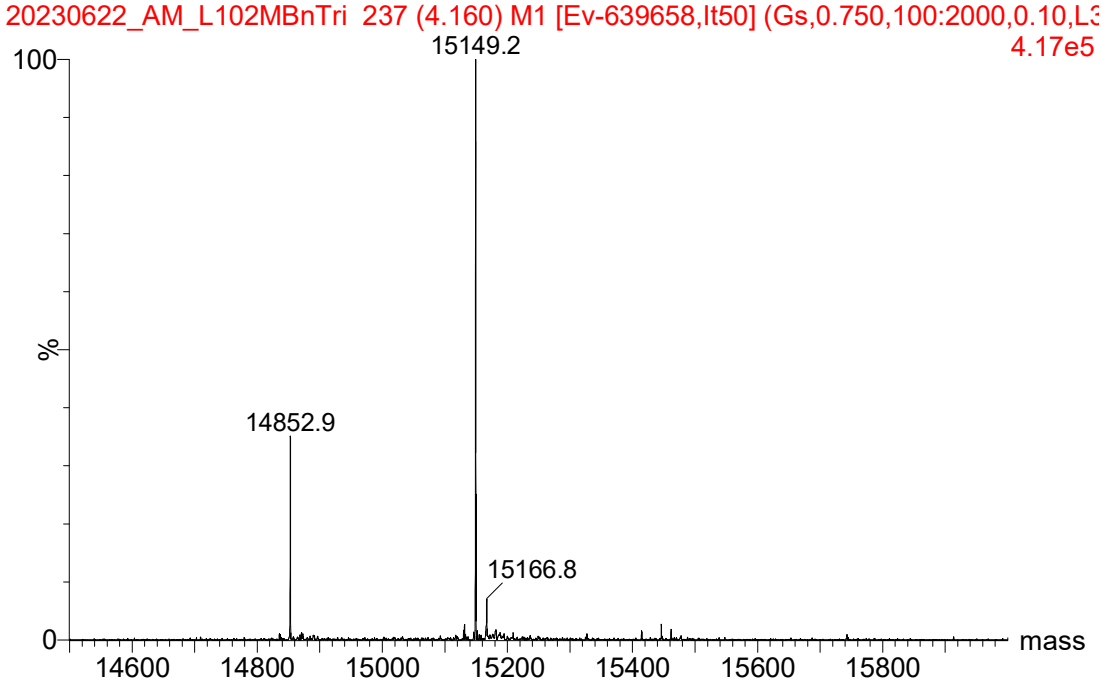




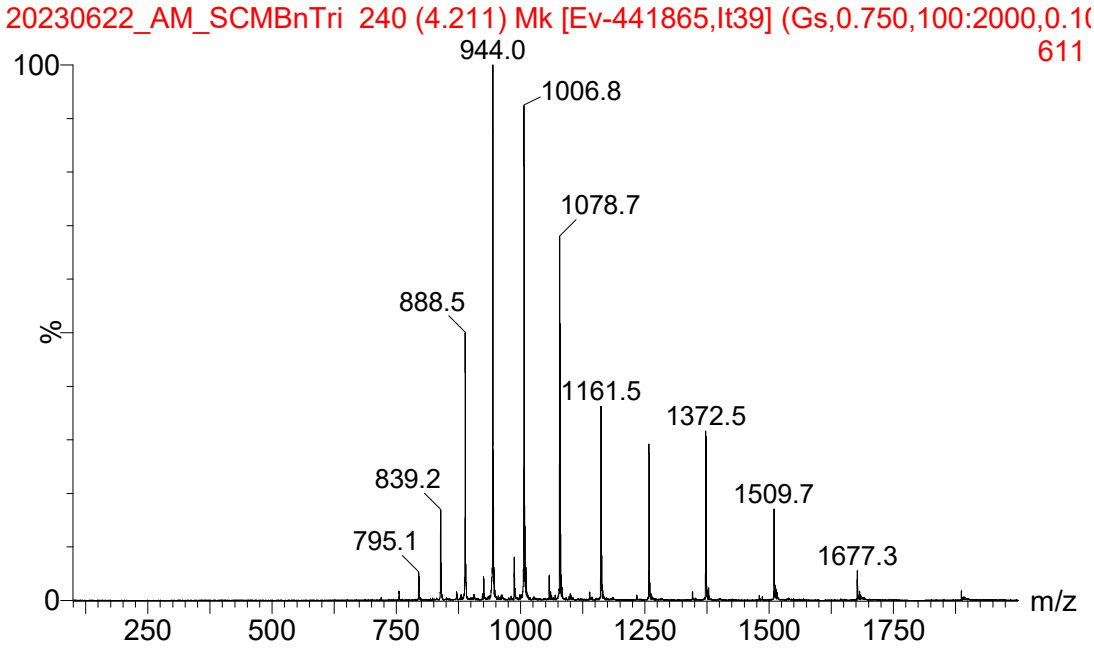
1.4 Protein LC-MS Functionalised with MBnTri

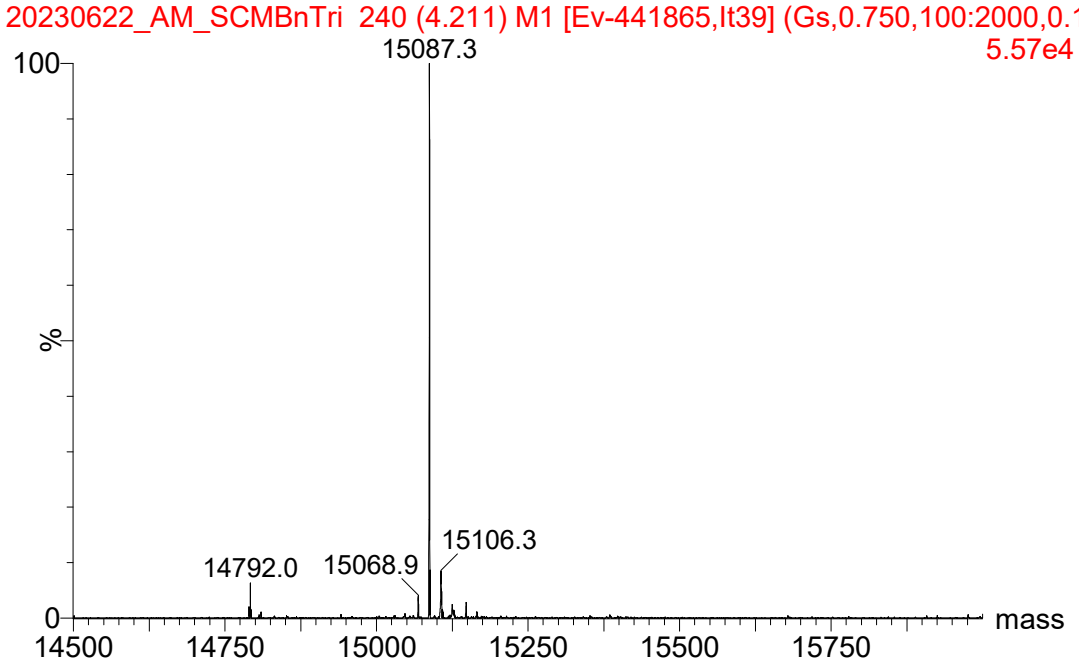
TTSCP L102C





TTSCP ΔDSB L102C





1.5 HPLC Calibration

HPLC method: Isocratic elution at 45% MeCN 0.1% TFA over 16 min, 30 °C

Standards were made up in 10% DMSO then diluted with an equal volume of MeCN (as reaction samples would be) and used to make calibration curve. Product elution time 12.02 min, Starting material elution time 12.61 min

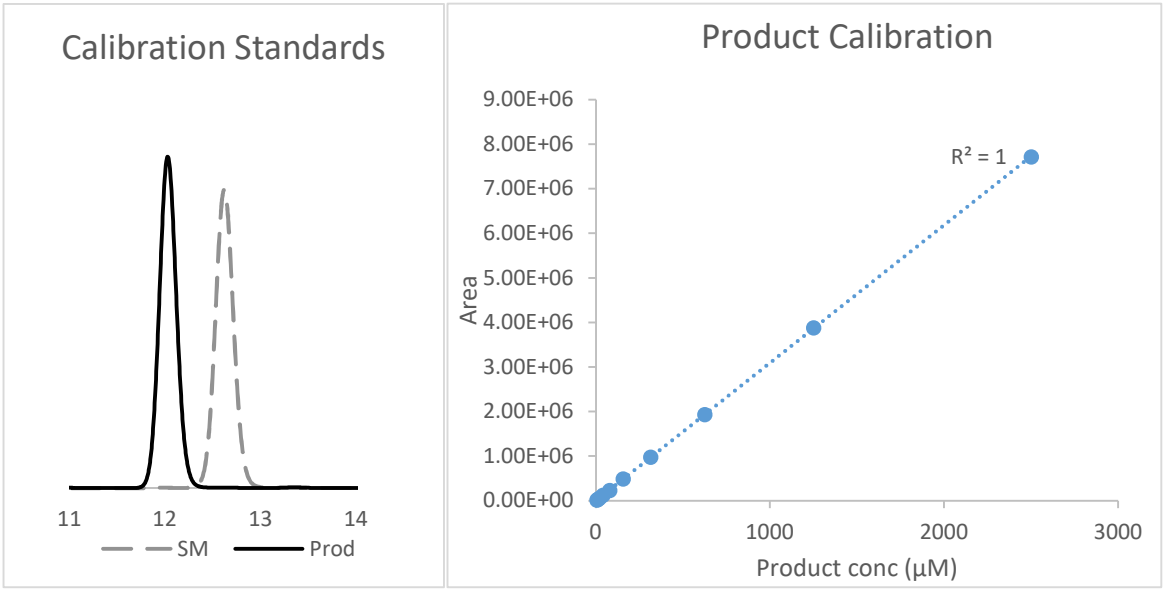
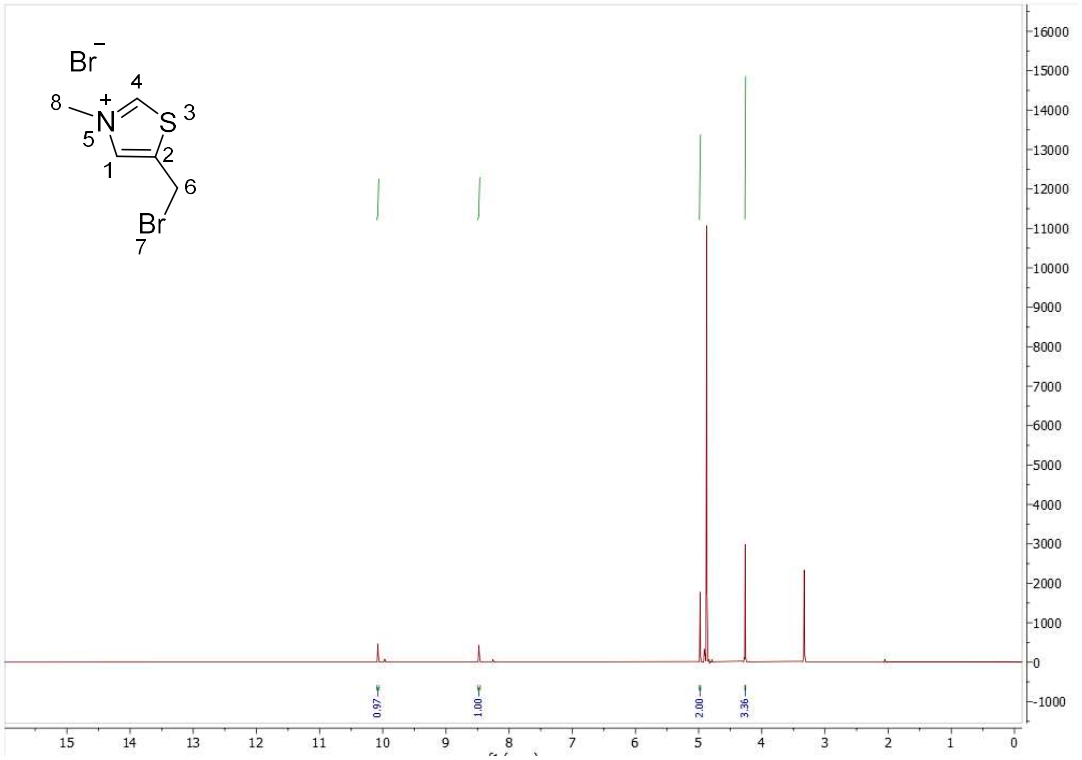


Figure 113 Calibration standards of 18 (intramolecular Stetter starting material, SM grey dashed line) and 19 (intramolecular Stetter product, black line) and the calibration curve plotted for 18.

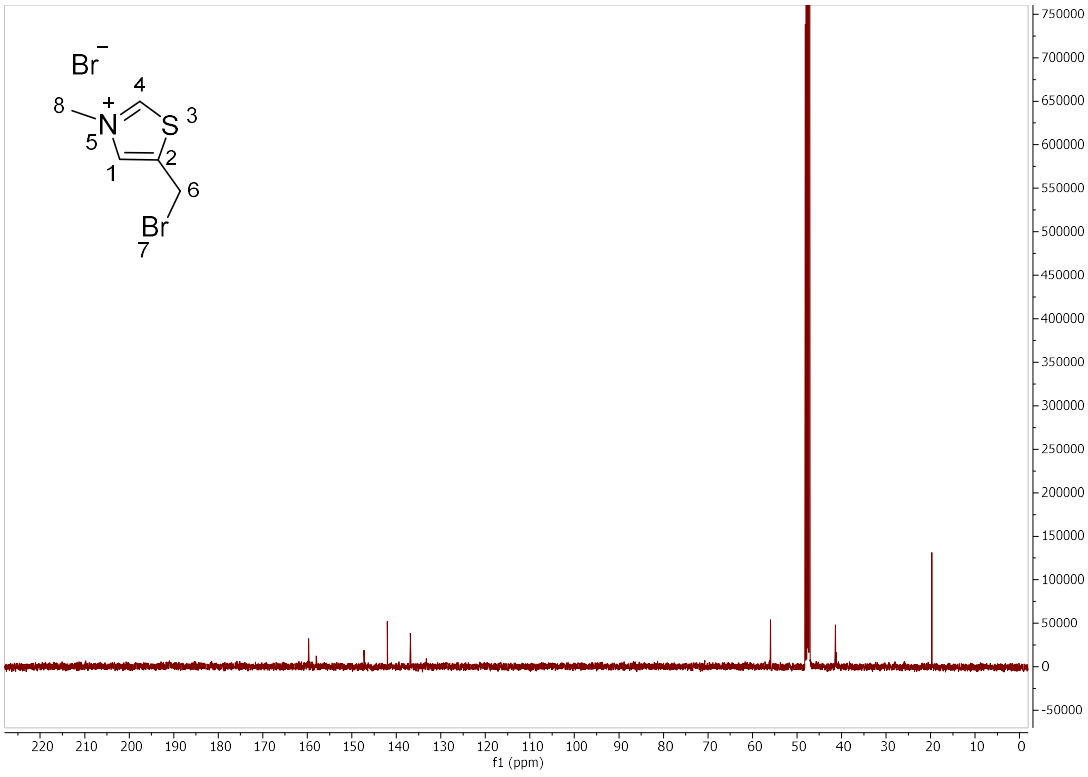
1.6 Characterisation Data (NMR, MS, IR)

1.6.1 MeThzBr Compound 29

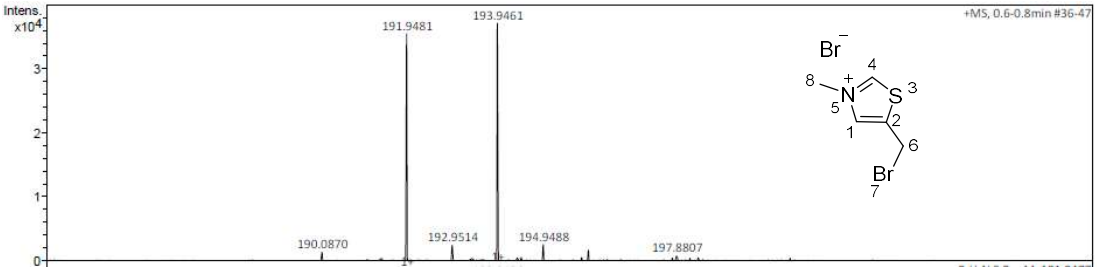
^{1}H NMR (500 MHz, MeOD)



^{13}C NMR (126 MHz, MeOD)

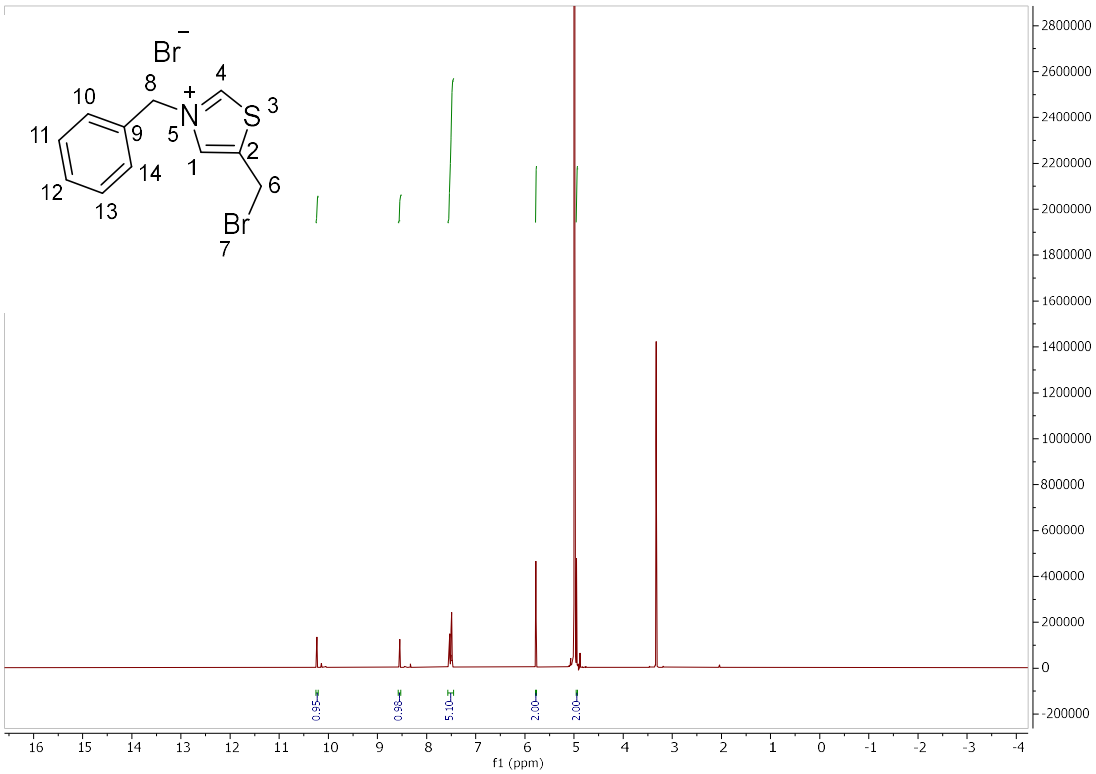


LCMS

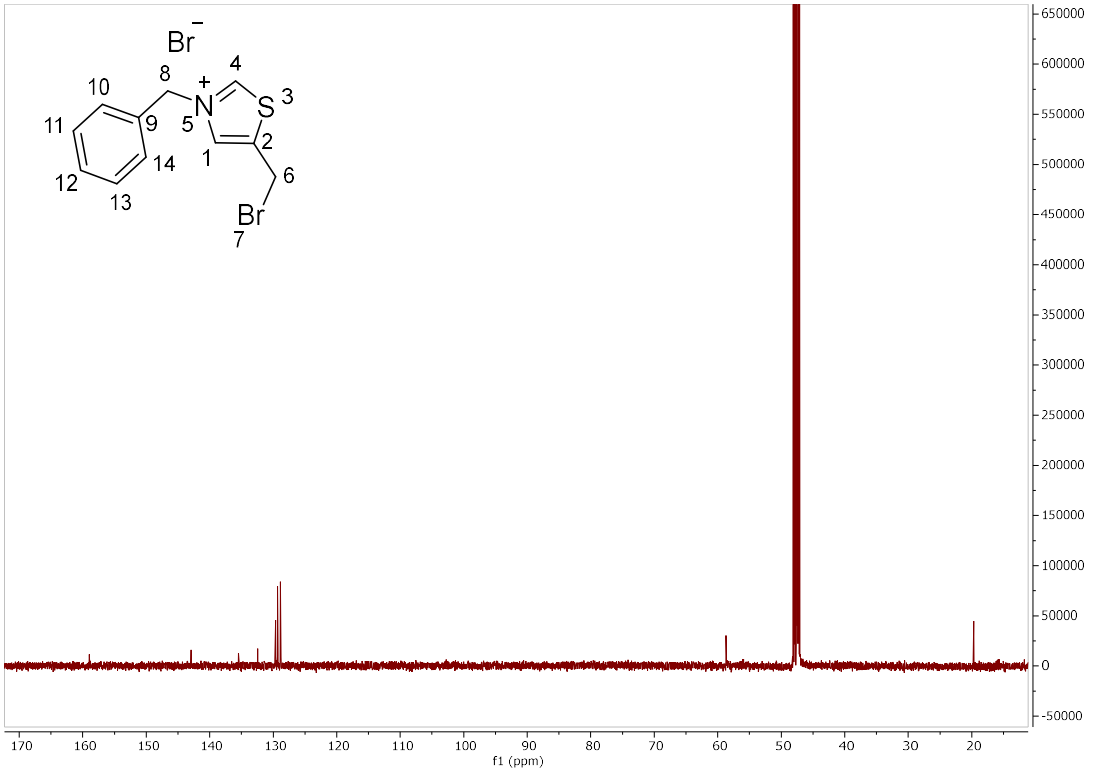

Intens.
x10^4
3
2
1
0
191.9481
193.9461
190.0870
192.9514
194.9488
197.8807
+MS, 0.6-0.8min #36-47
Br
8
N
4
3
S
5
1
2
6
Br
7

BnThzBr Compound 31

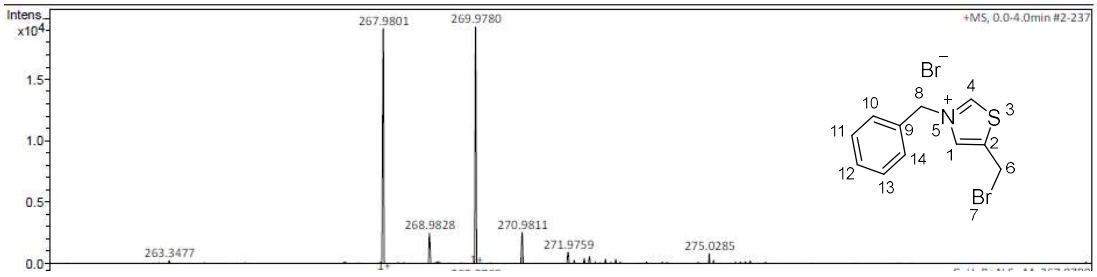
^{1}H NMR (500 MHz, MeOD)



^{13}C NMR (126 MHz, MeOD)

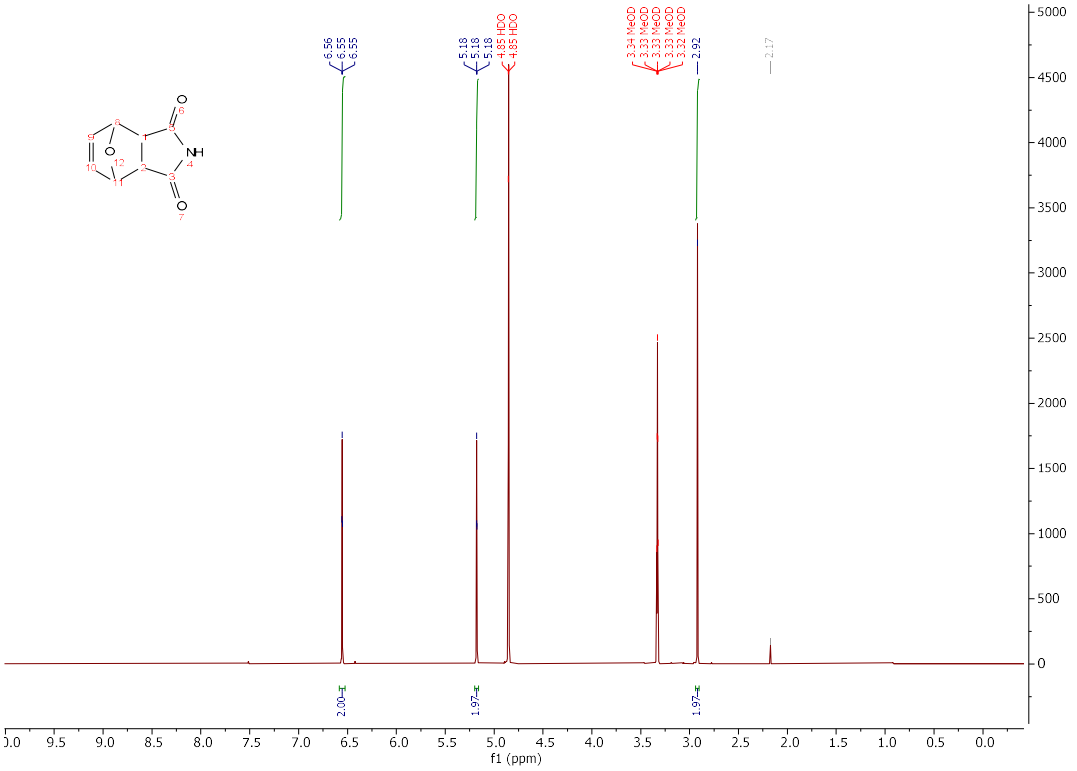


LCMS

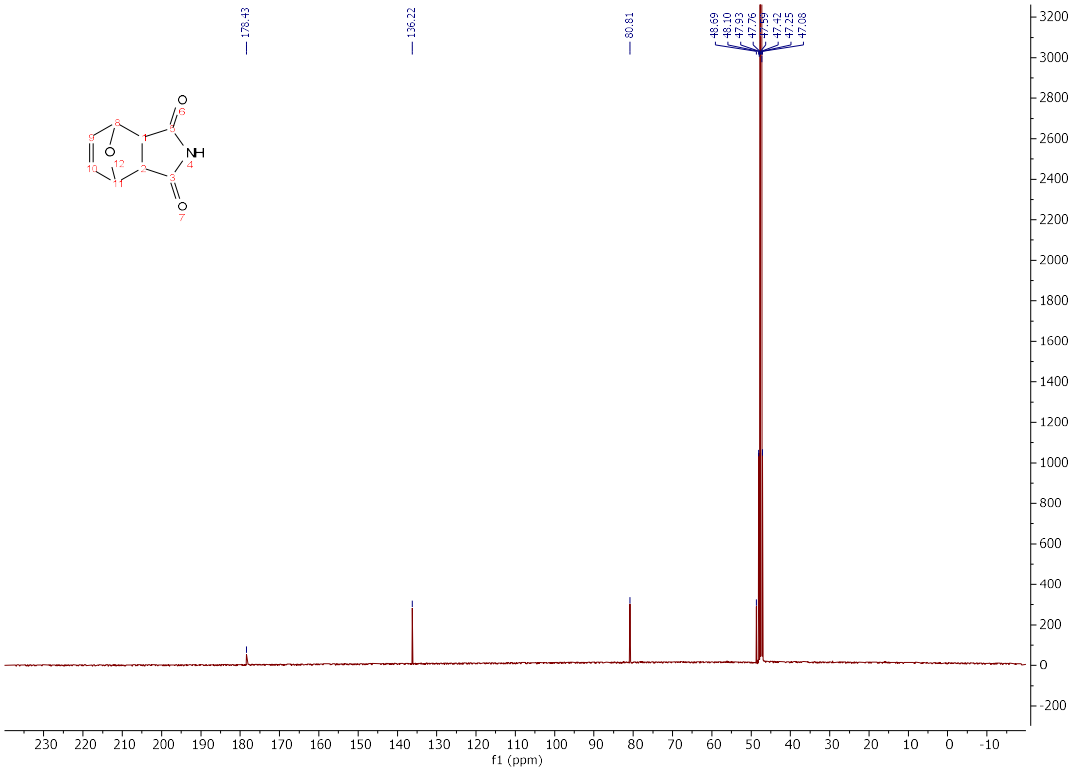


1.6.2 Compound 34

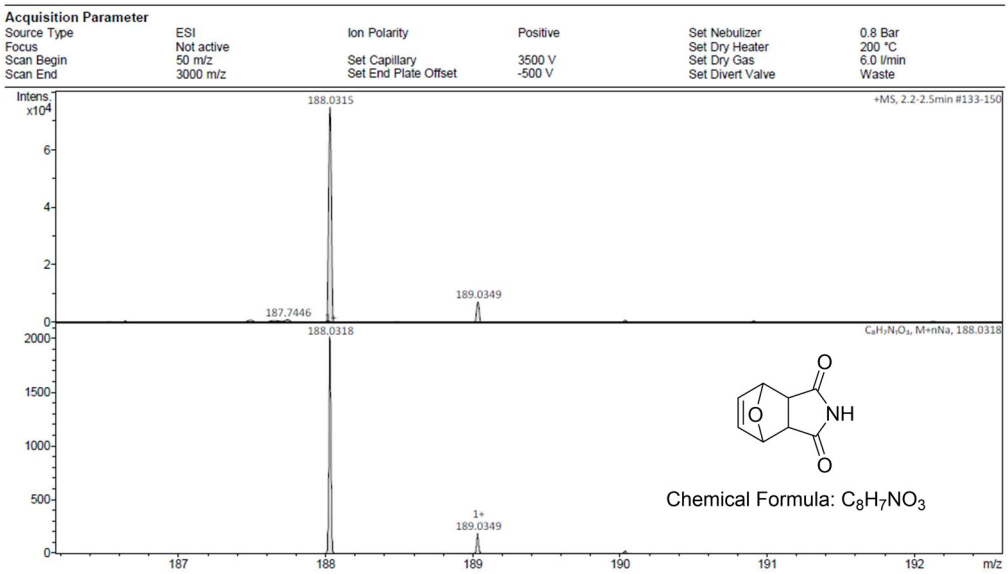
H NMR (500 MHz, MeOD)



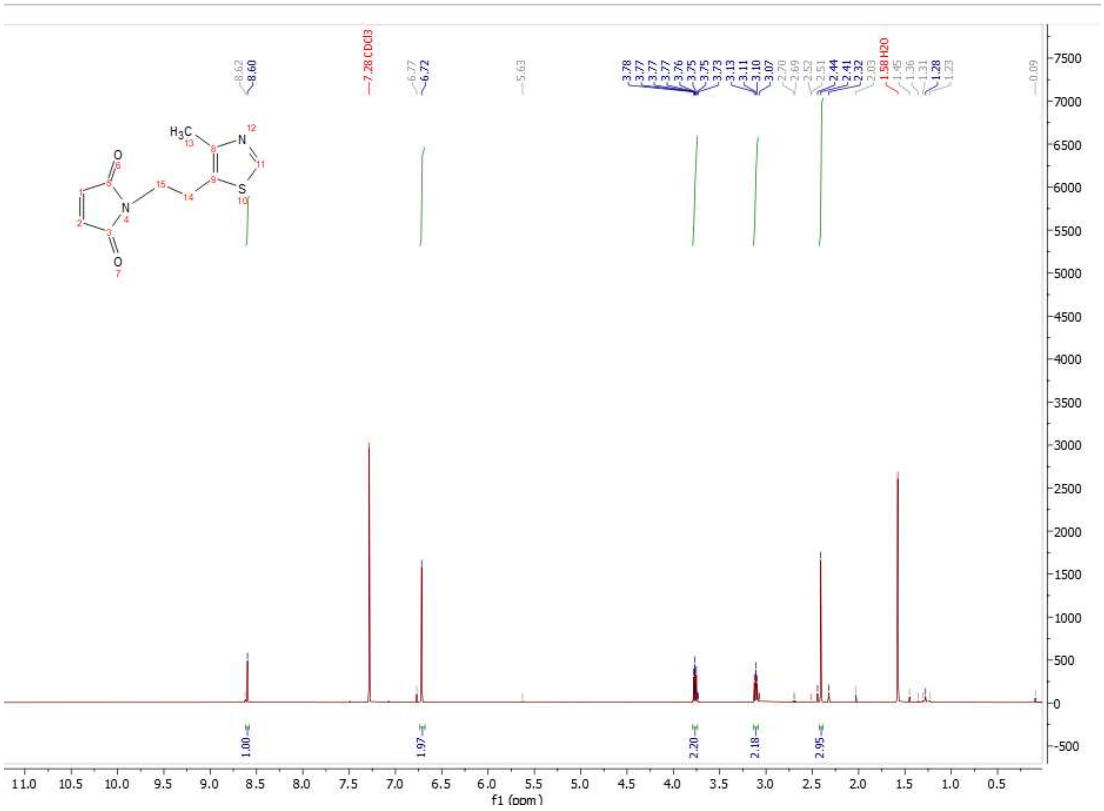
^{13}C NMR (126 MHz, MeOD)

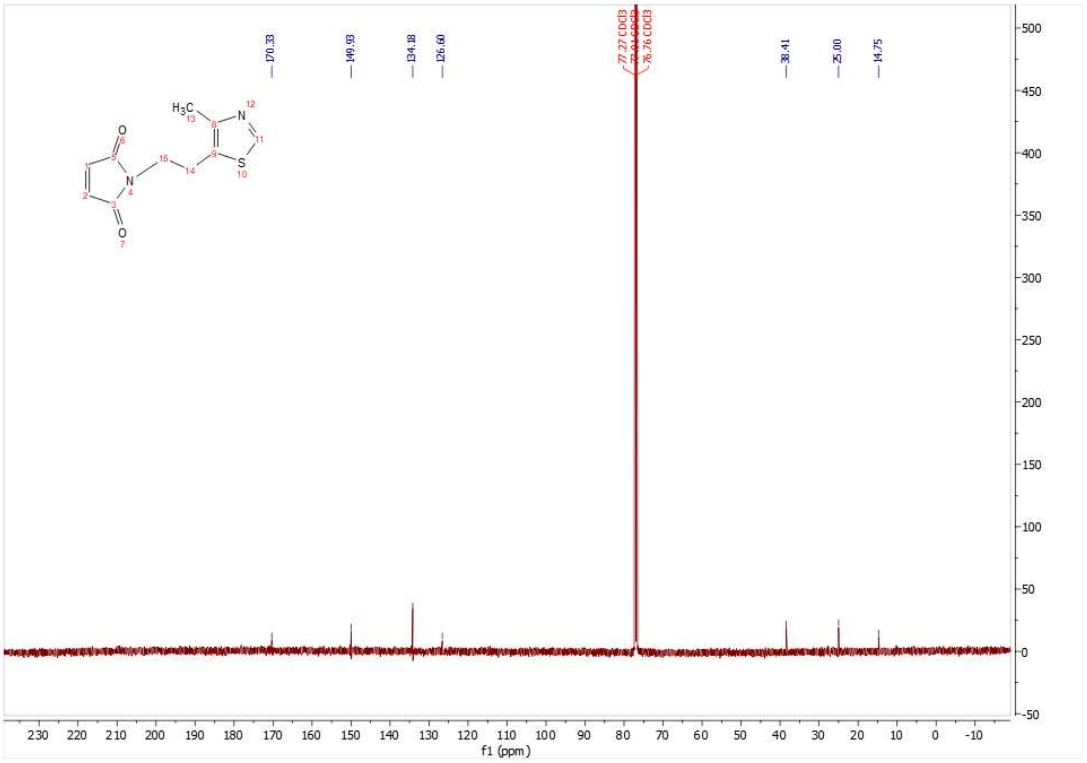


LCMS

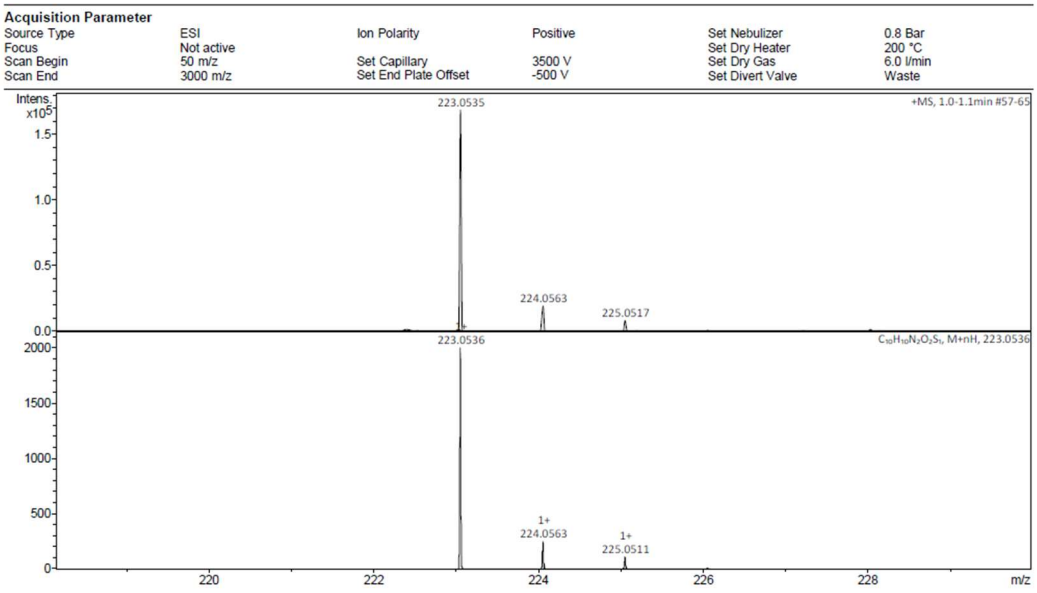


1.6.3 Compound 36

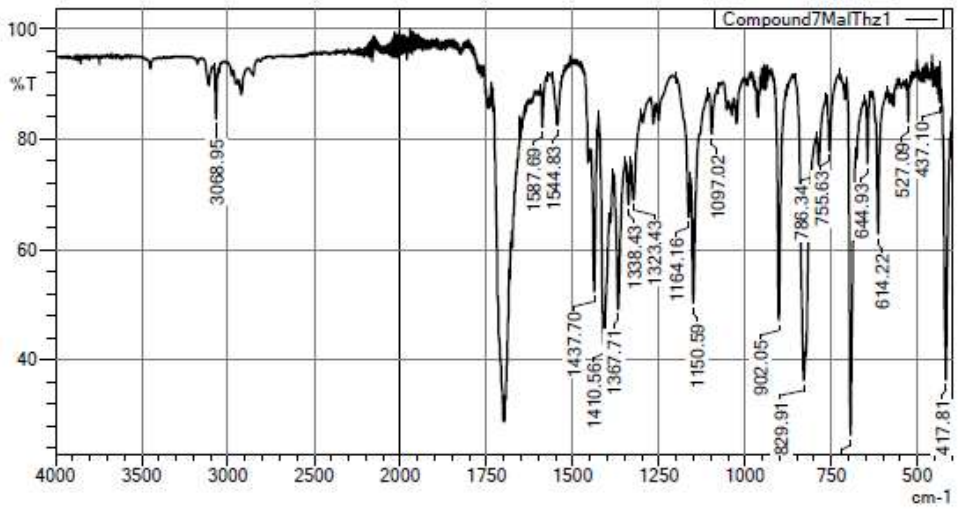




LCMS compound 36

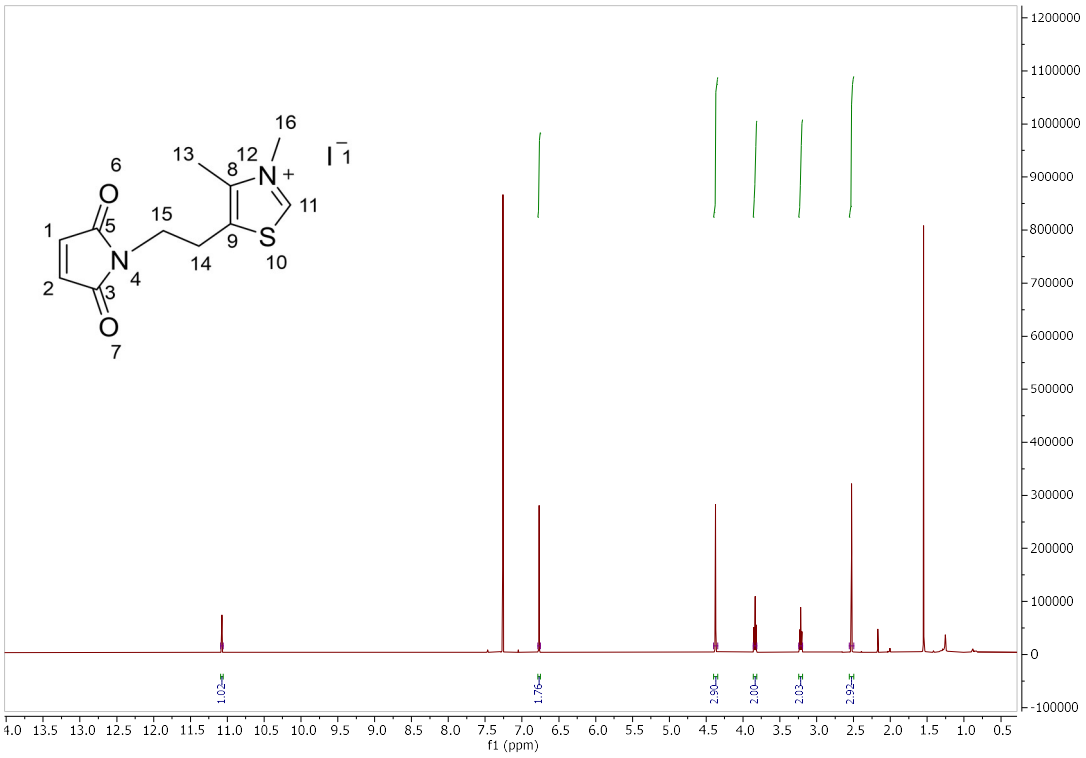


IR compound 36

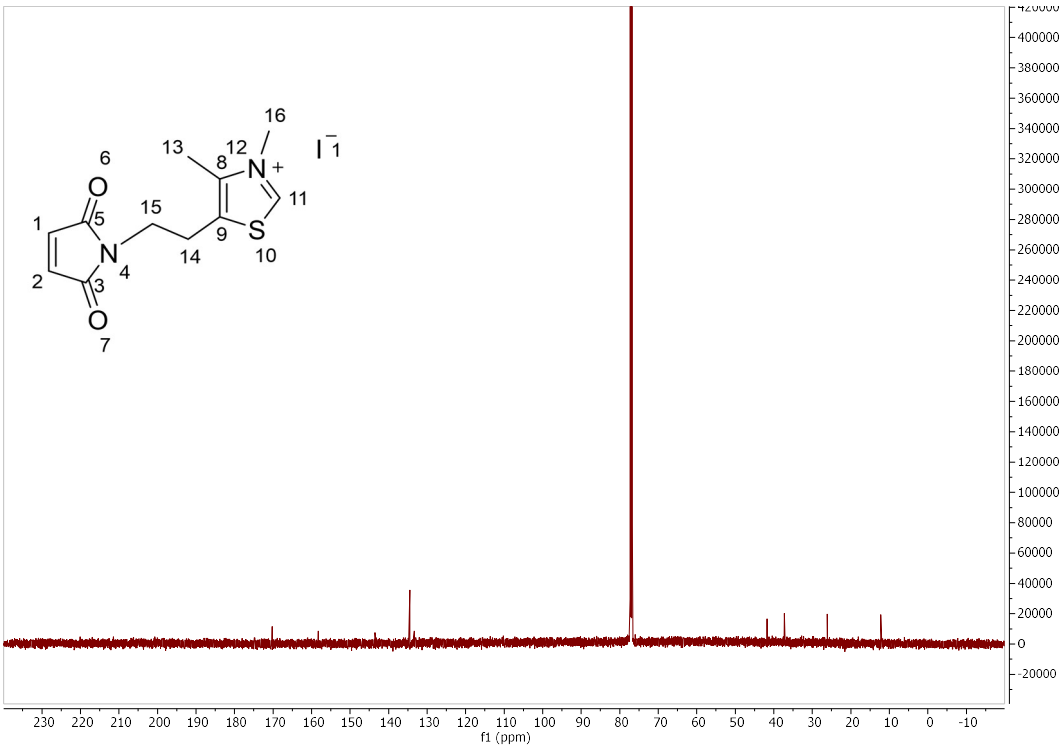


1.6.4 MMeThz compound 37

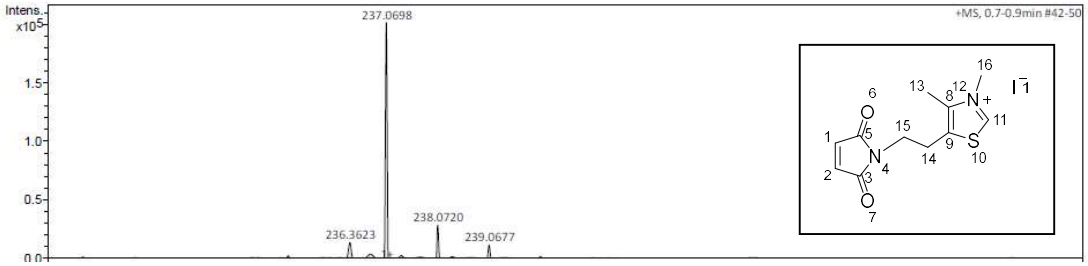
^{1}H NMR (500 MHz, Chloroform-*d*)



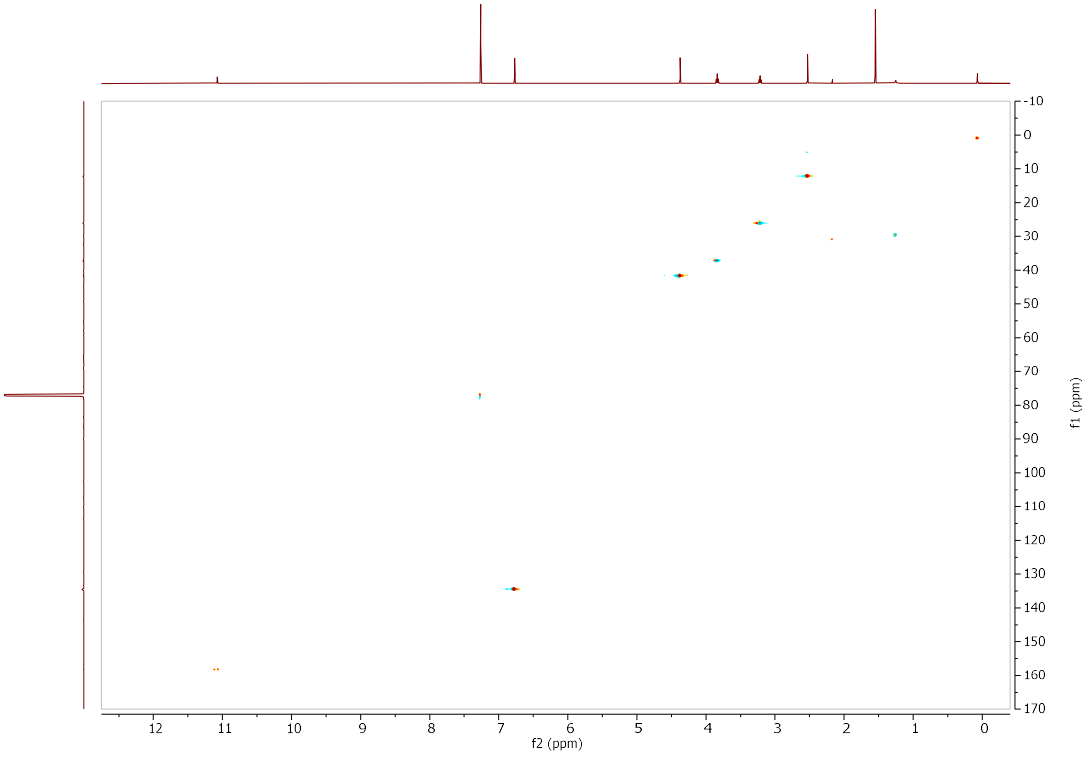
^{13}C NMR (126 MHz, $CDCl_3$)



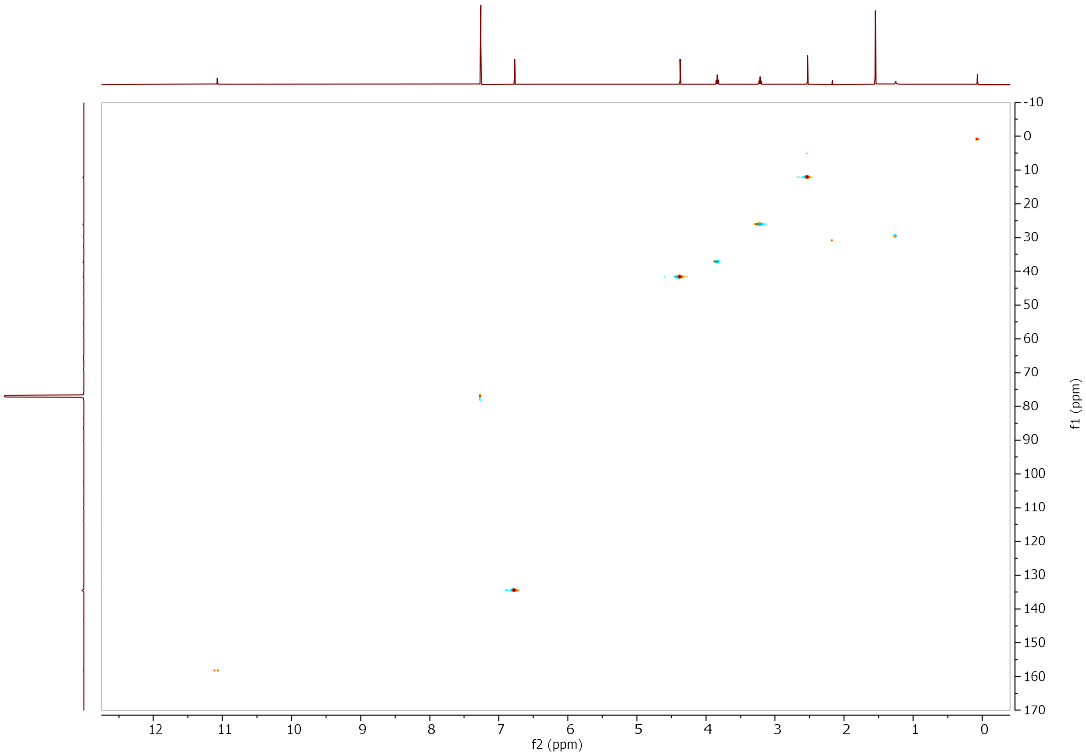
LCMS


Intens.
x10^5
1.5
1.0
0.5
0.0
237.0698
236.3623
238.0720
239.0677
+MS, 0.7-0.9min #42-50
I⁻

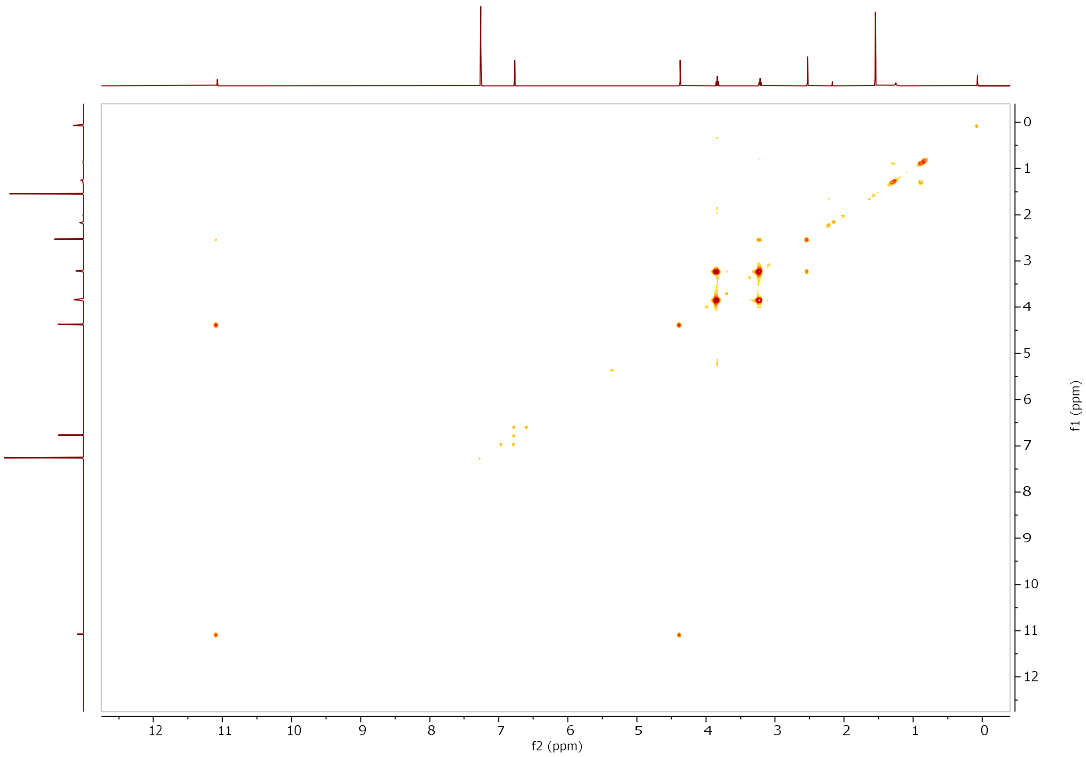
HSQC


f2 (ppm)
f1 (ppm)

HMBC

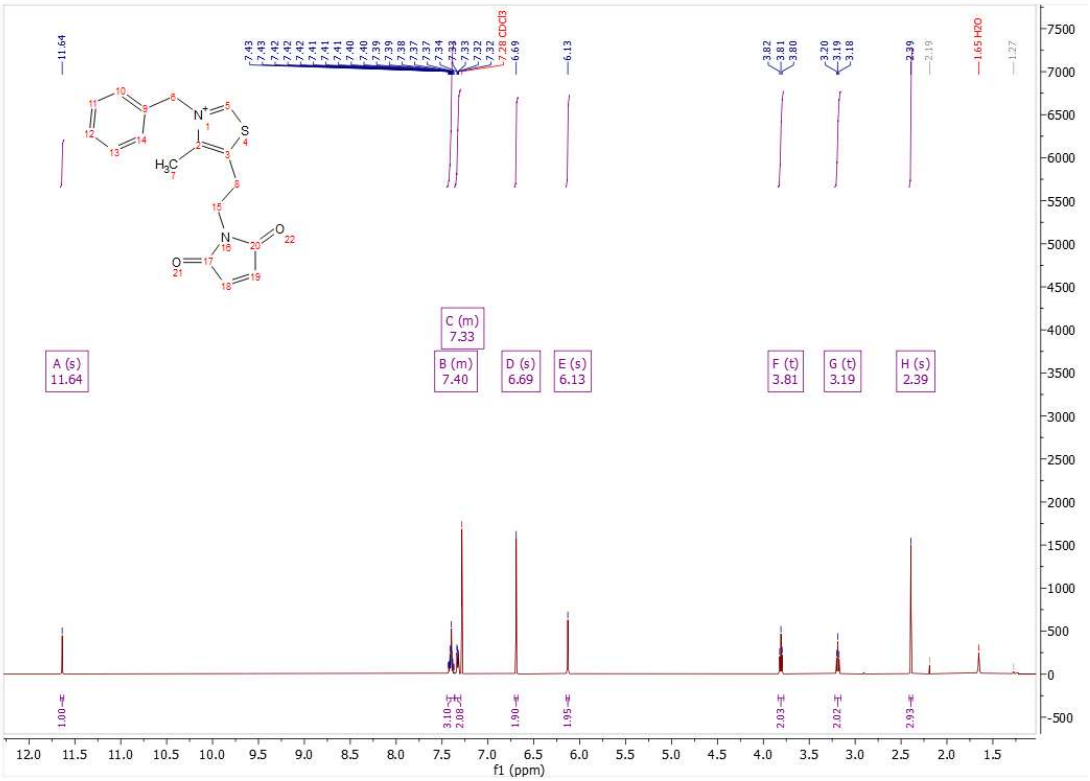


COSY

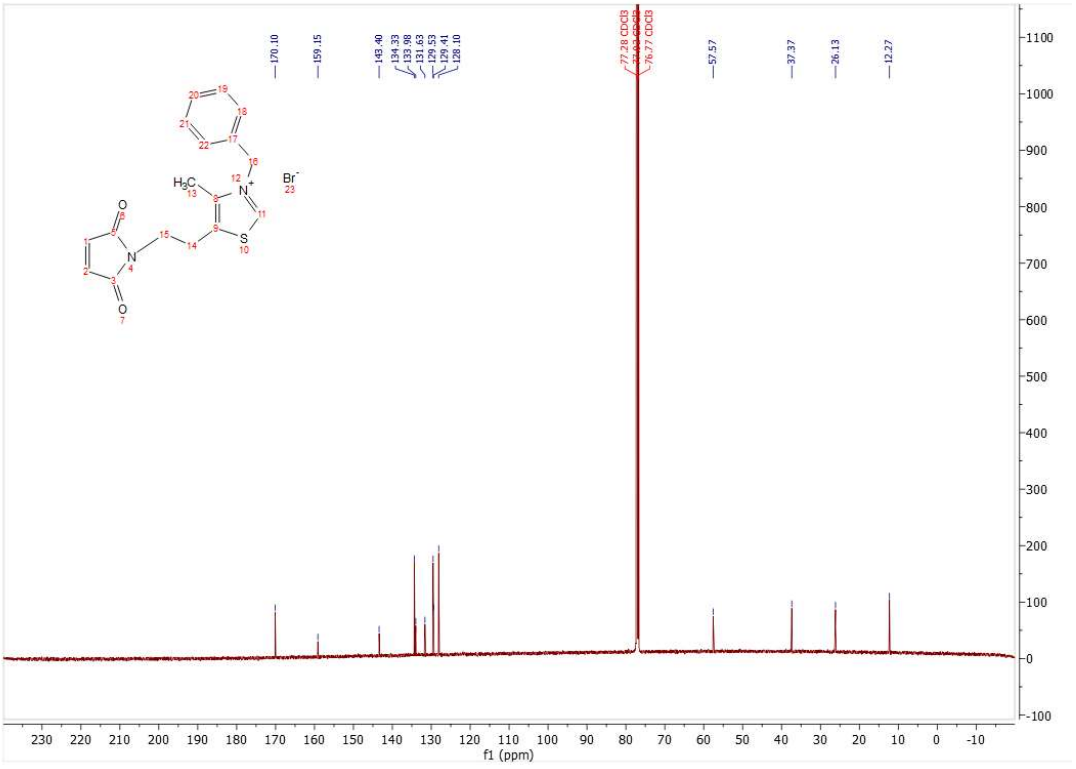


1.6.5 MBnThz Compound 38

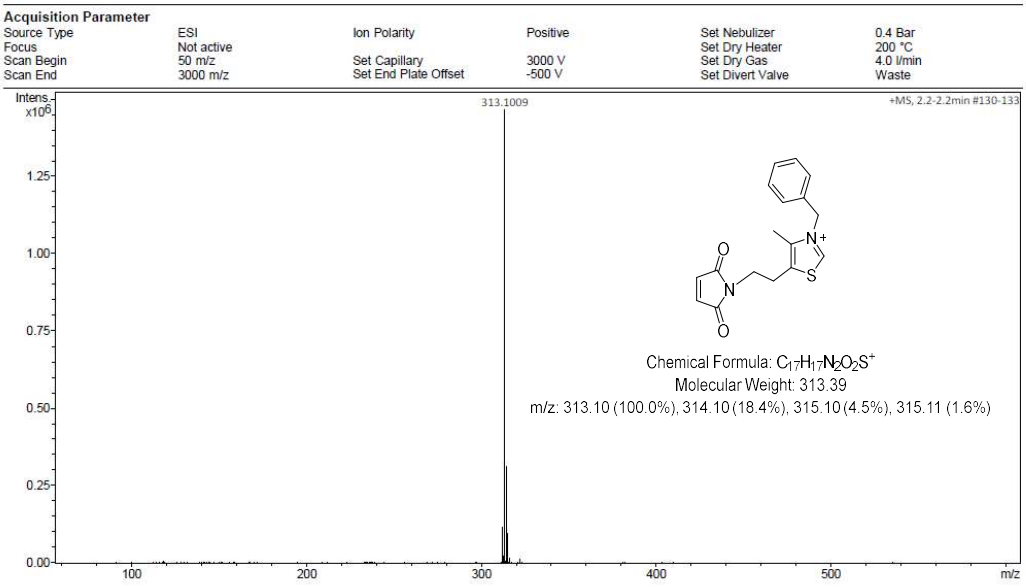
^{1}H NMR (500 MHz, Chloroform-*d*)



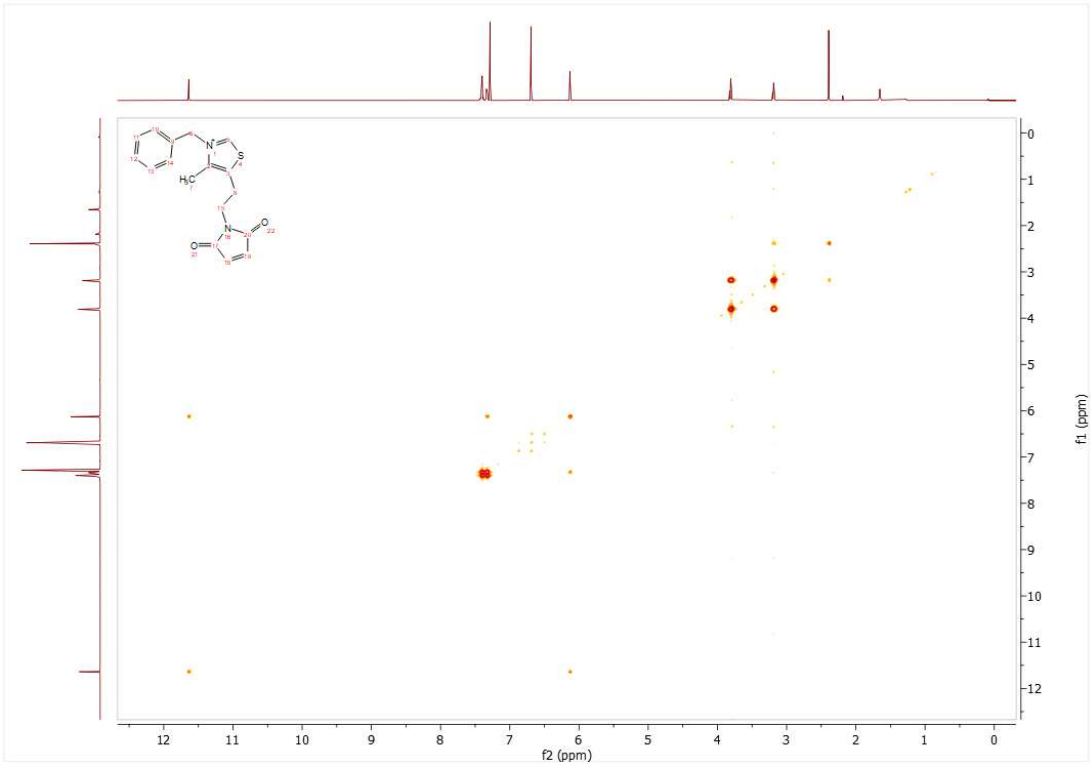
^{13}C NMR (126 MHz, $CDCl_3$)



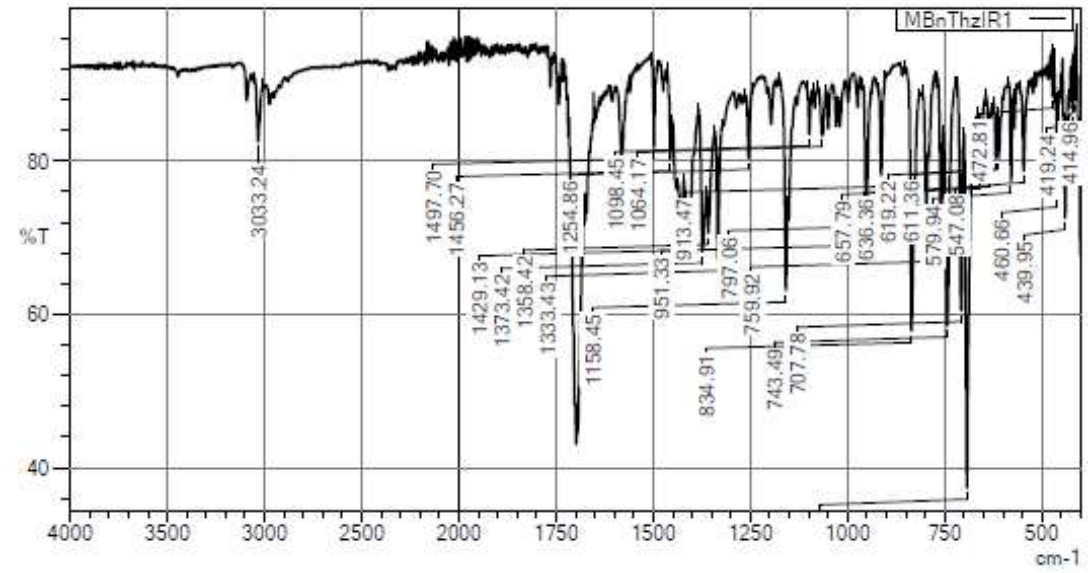
LCMS


Acquisition Parameter
Source Type ESI
Ion Polarity Positive
Set Nebulizer 0.4 Bar
Focus Not active
Set Dry Heater 200 °C
Scan Begin 50 m/z
Set Capillary 3000 V
Set Dry Gas 4.0 l/min
Scan End 3000 m/z
Set End Plate Offset -500 V
Set Divert Valve Waste
Intens.
x10^6
313.1009
+MS, 2.2-2.2min #130-133
Chemical Formula: C17H17N2O2S+
Molecular Weight: 313.39
m/z: 313.10 (100.0%), 314.10 (18.4%), 315.10 (4.5%), 315.11 (1.6%)
m/z

COSY

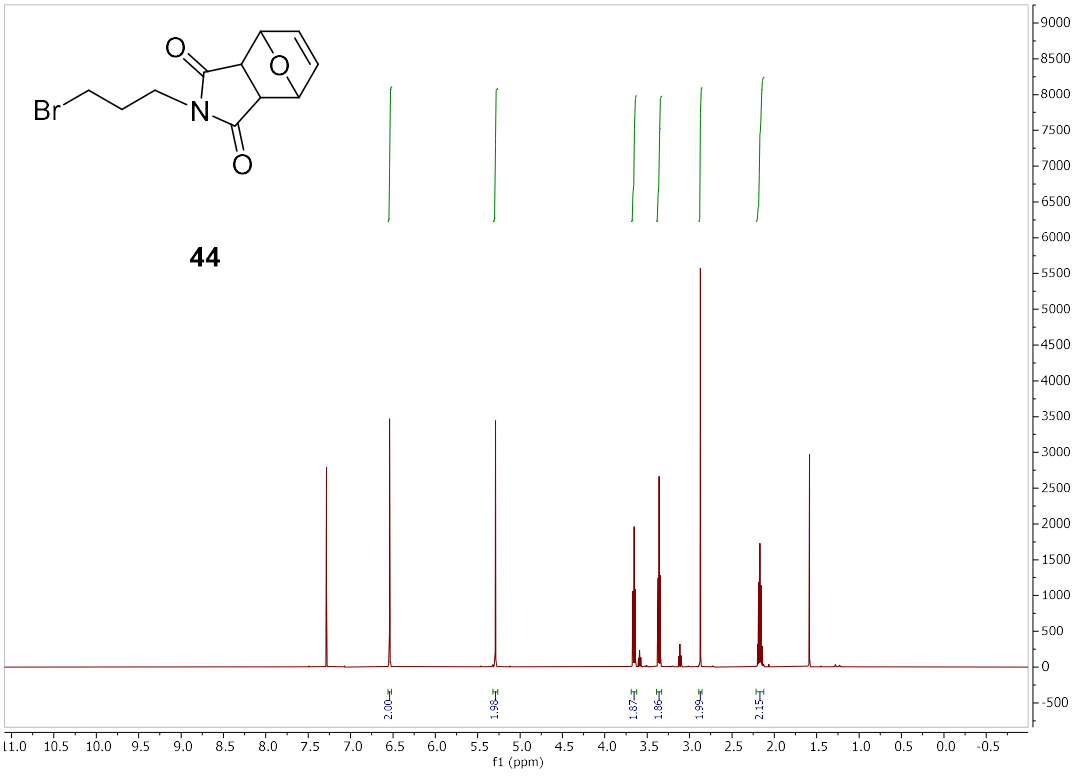

f2 (ppm)
f1 (ppm)

IR

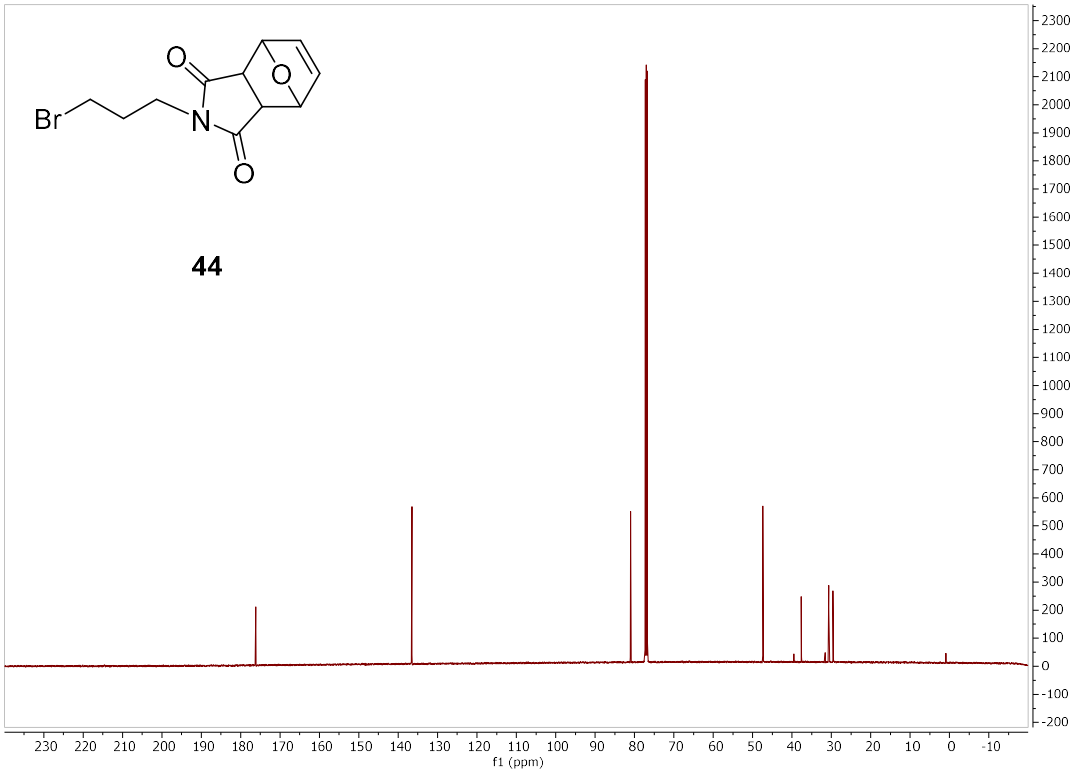


1.6.6 MBnTri compound 46

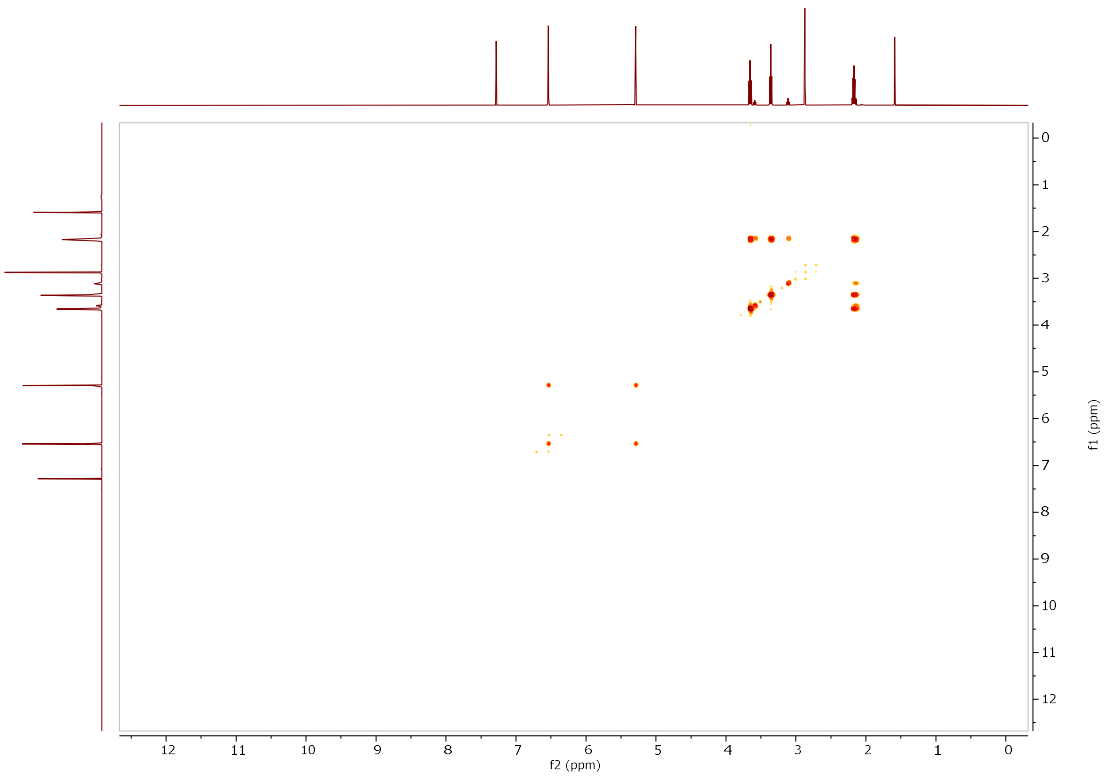
^{1}H NMR (500 MHz, Chloroform-*d*) Compound 44



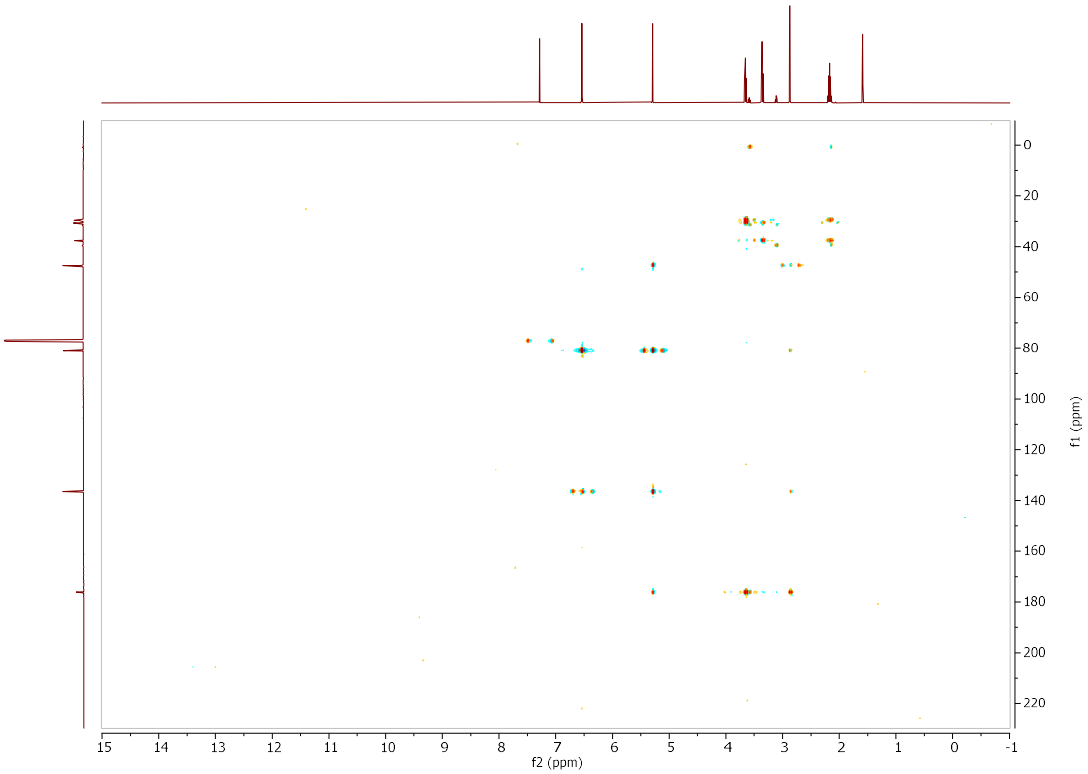
^{13}C NMR (126 MHz, CDCl3) Compound 44



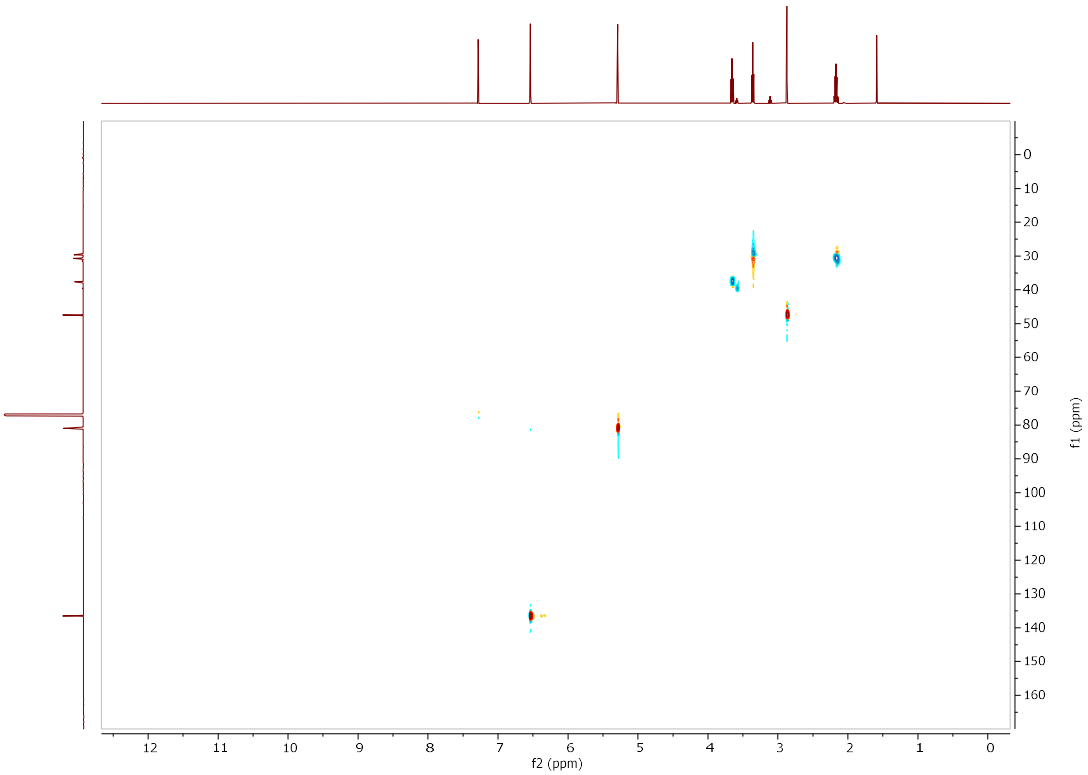
COSY Compound 44



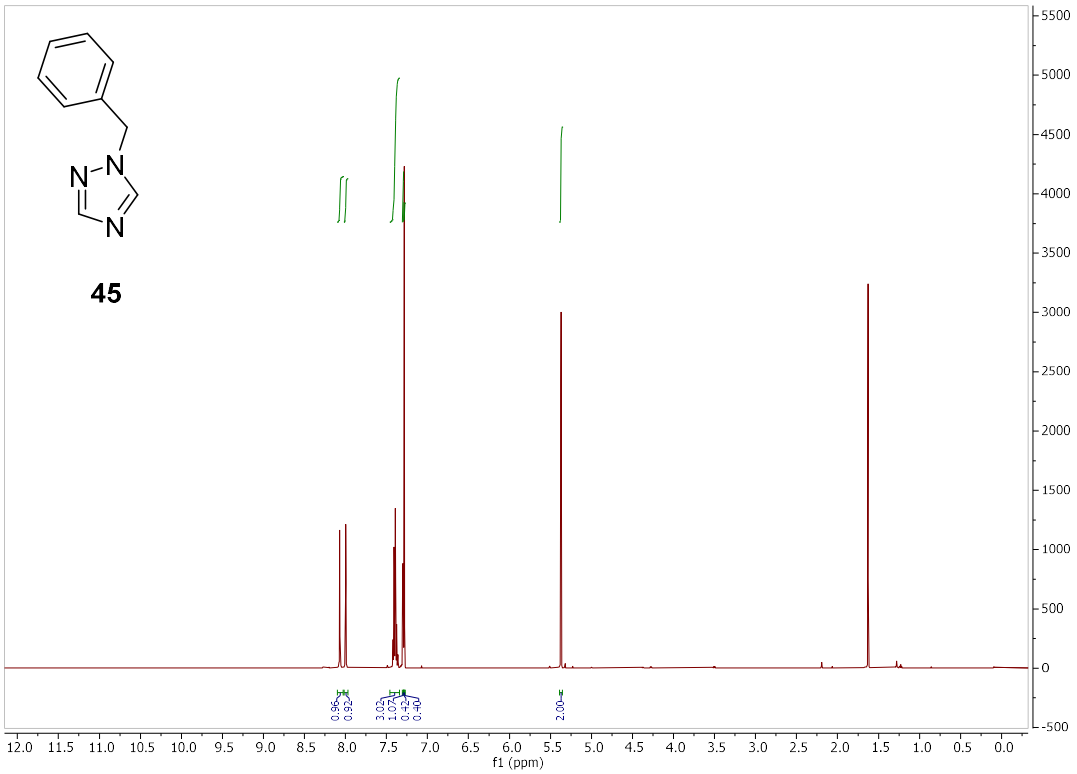
HMBC Compound 44



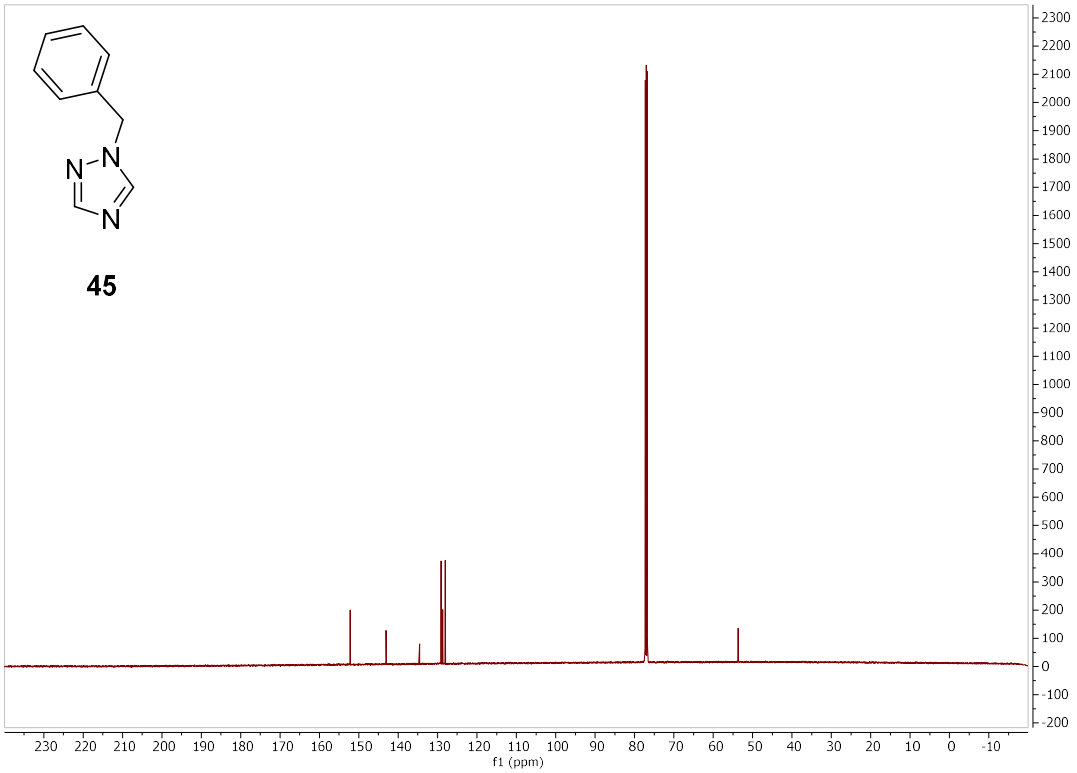
HSQC Compound 44



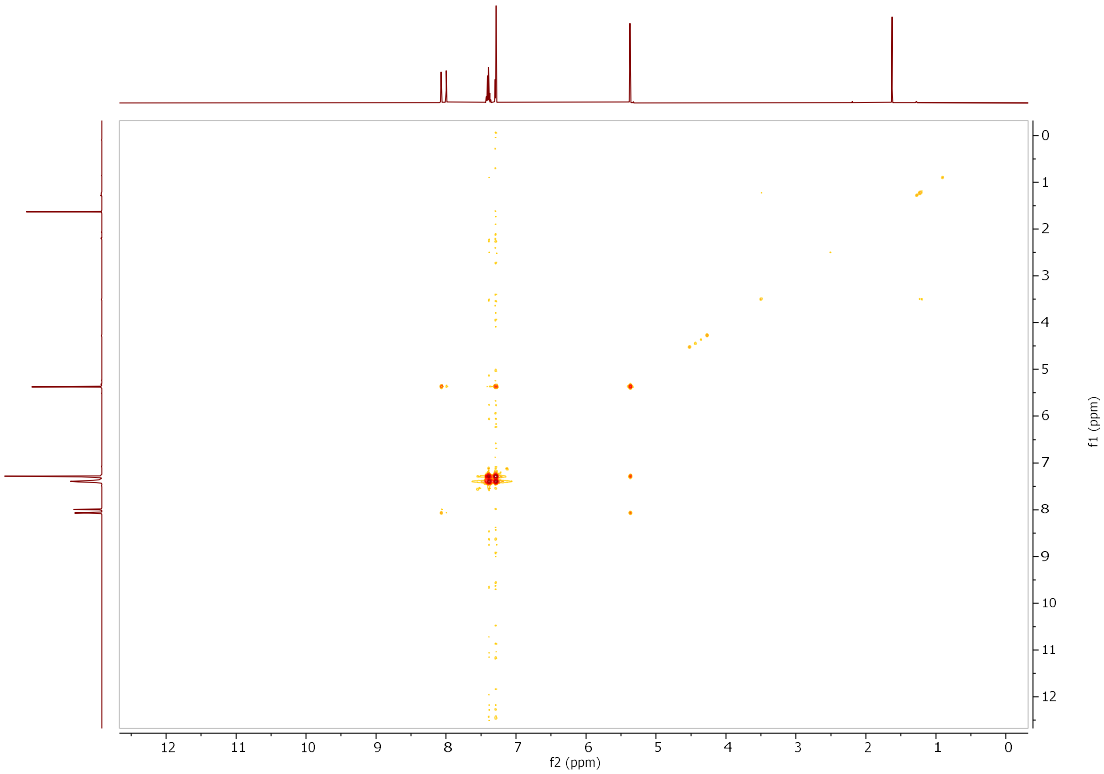
^{1}H NMR (500 MHz, Chloroform-d) Compound 45



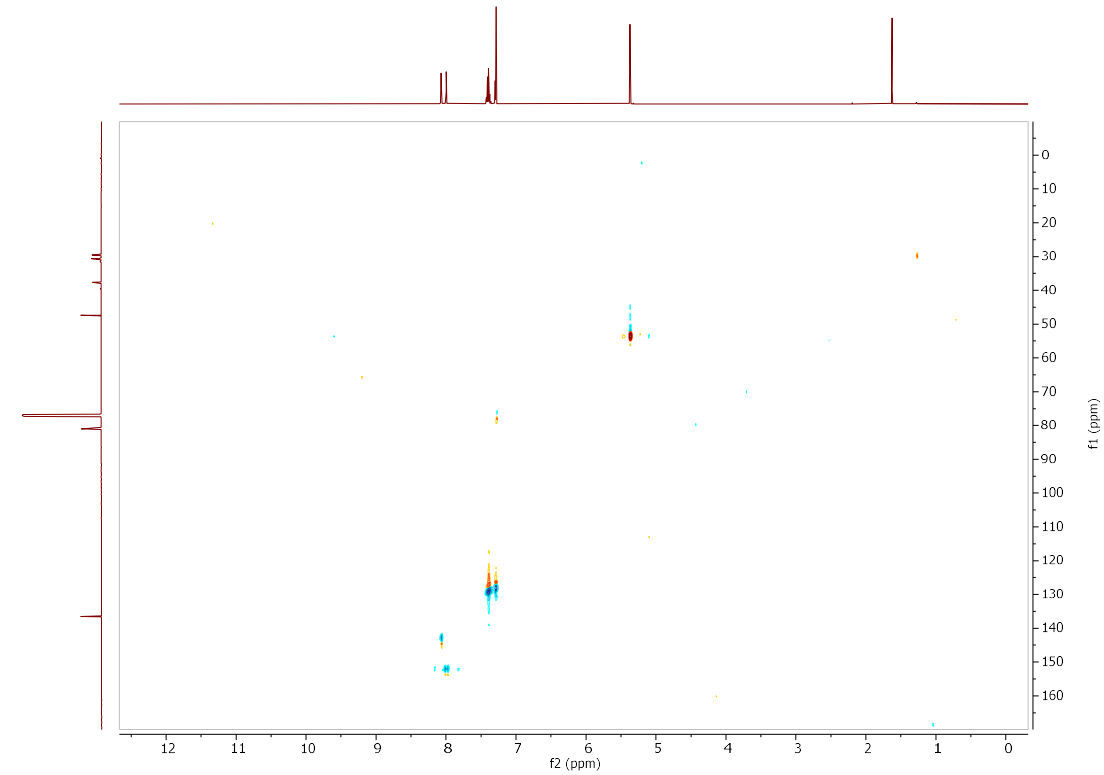
^{13}C NMR (126 MHz, CDCl3) Compound 45



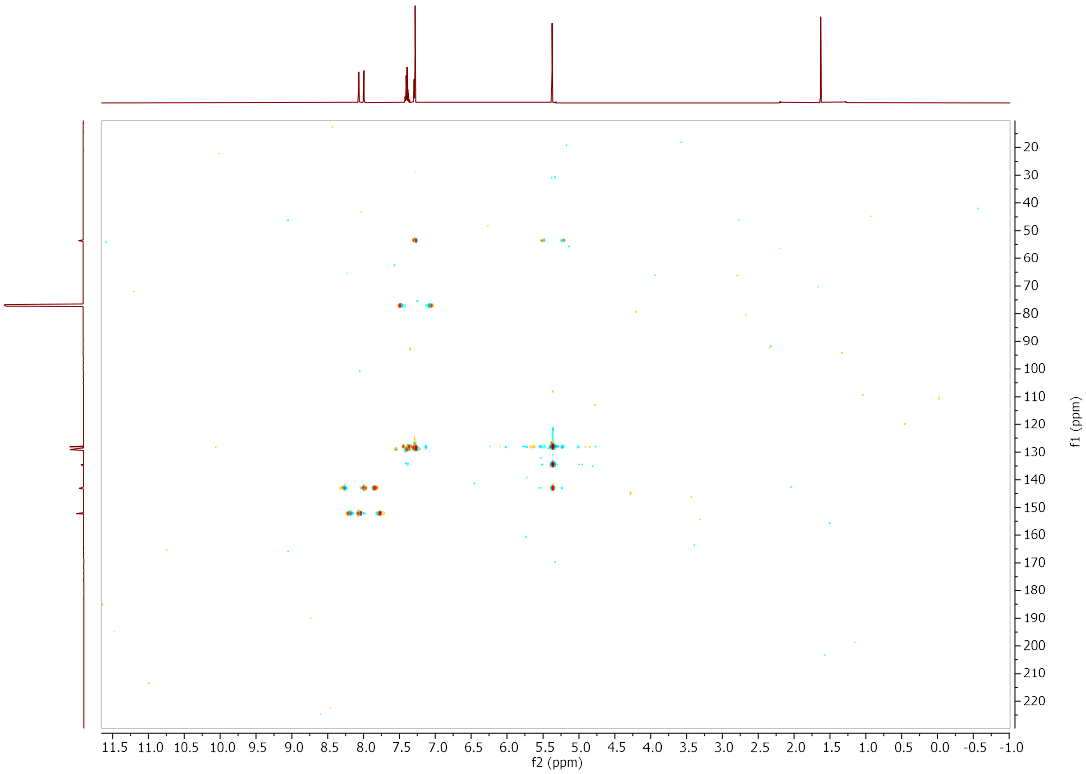
COSY Compound 45



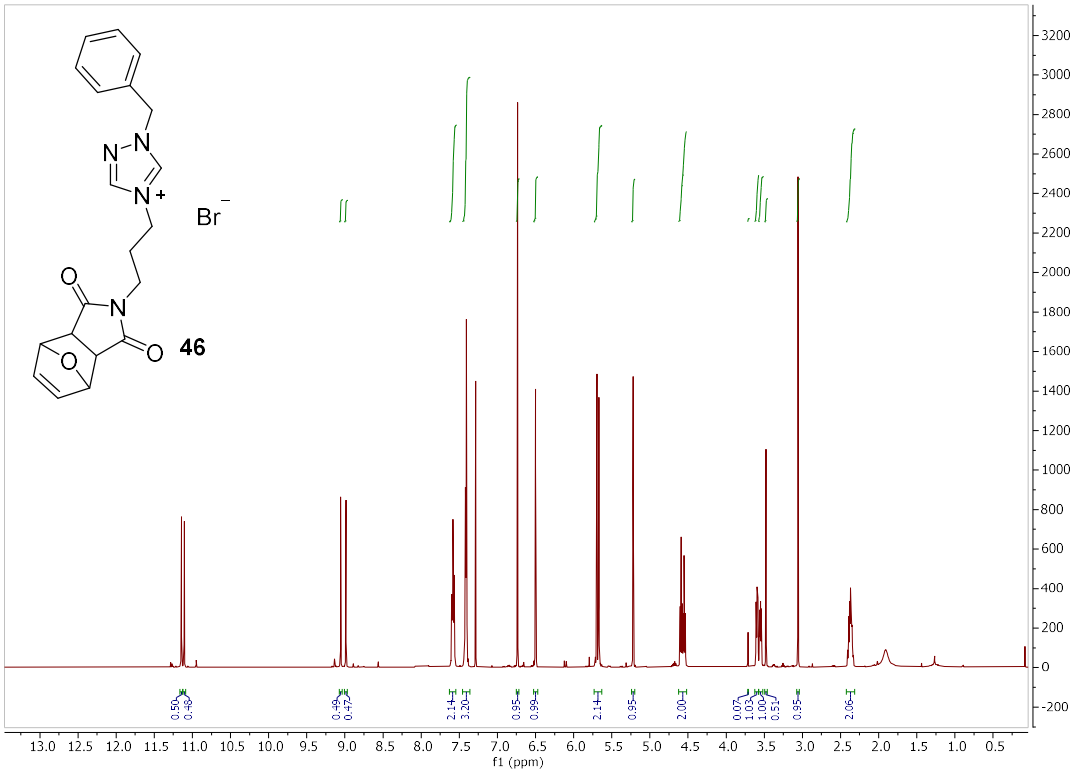
HSQC Compound 45



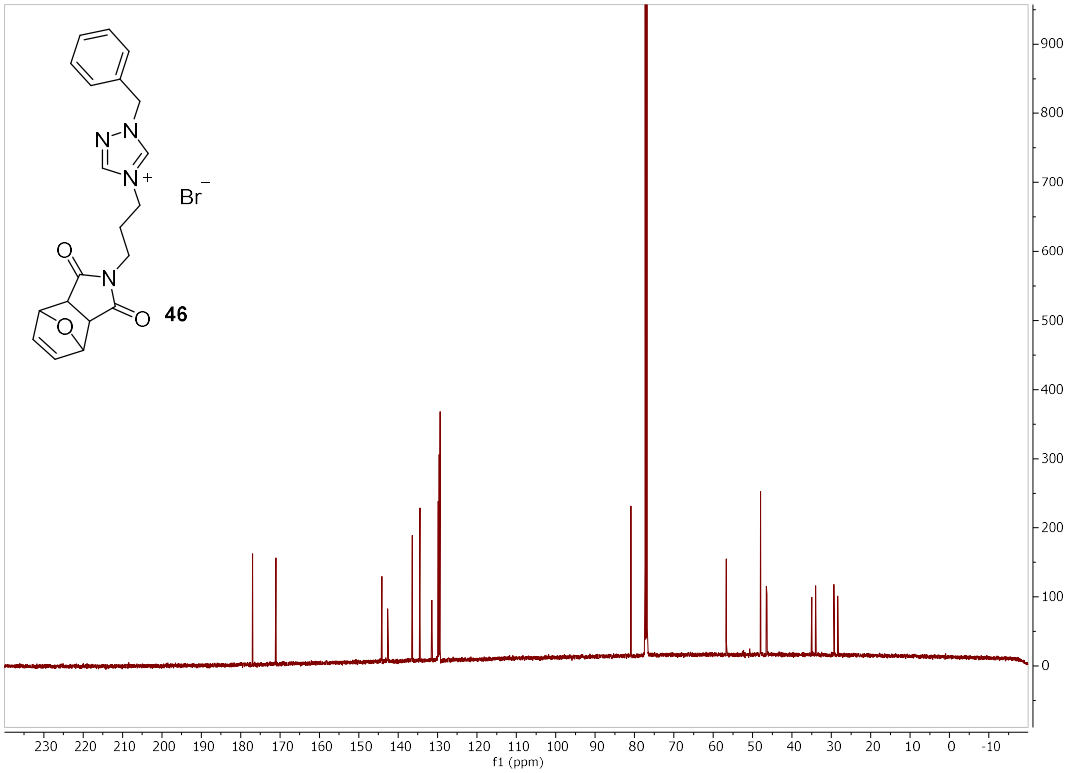
HMBC Compound 45



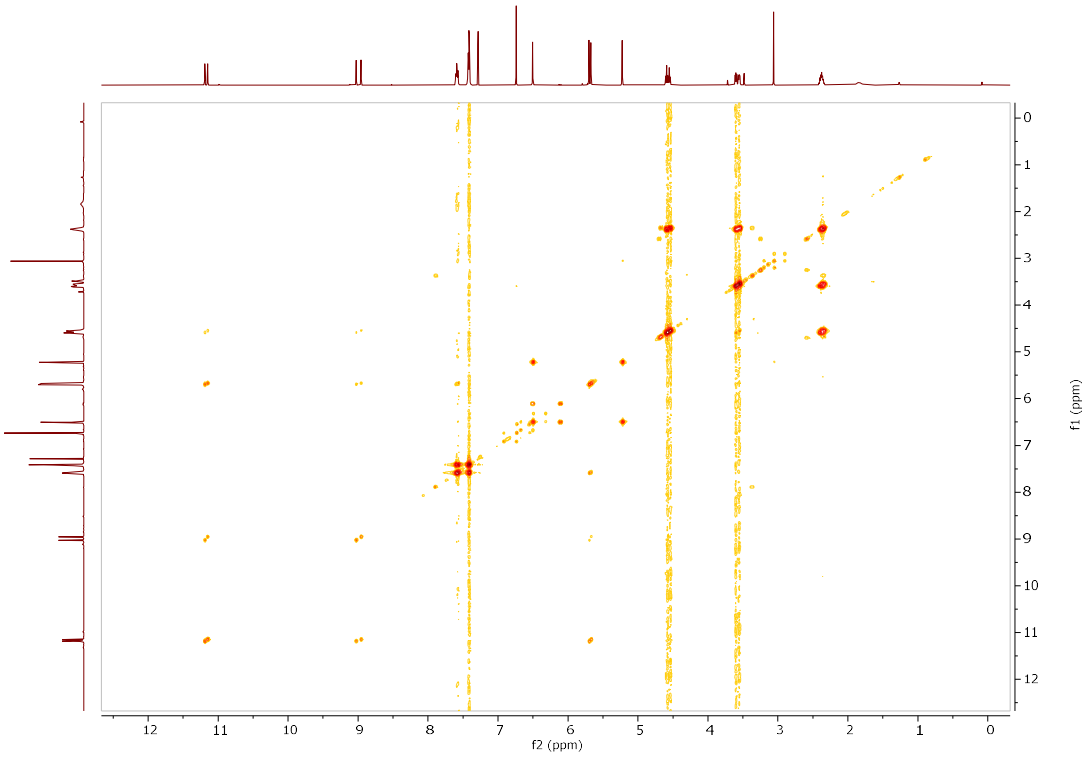
1H NMR (500 MHz, Chloroform-d) Compound 46



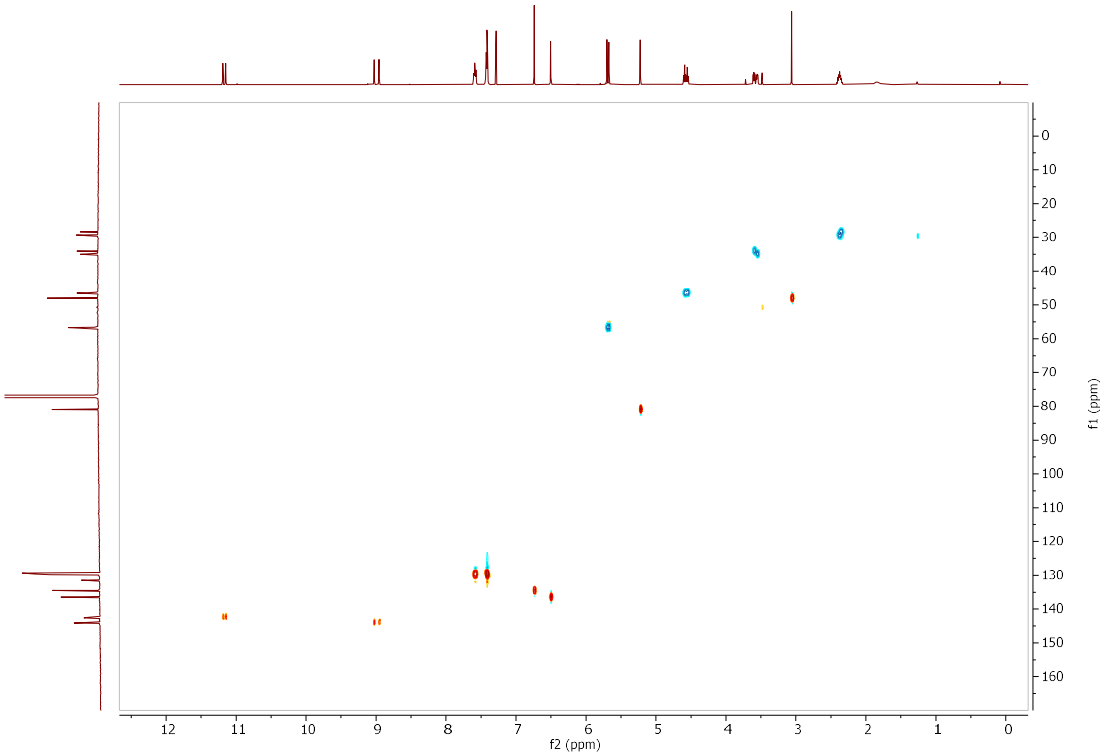
^{13}C NMR (126 MHz, CDCl3) Compound 46



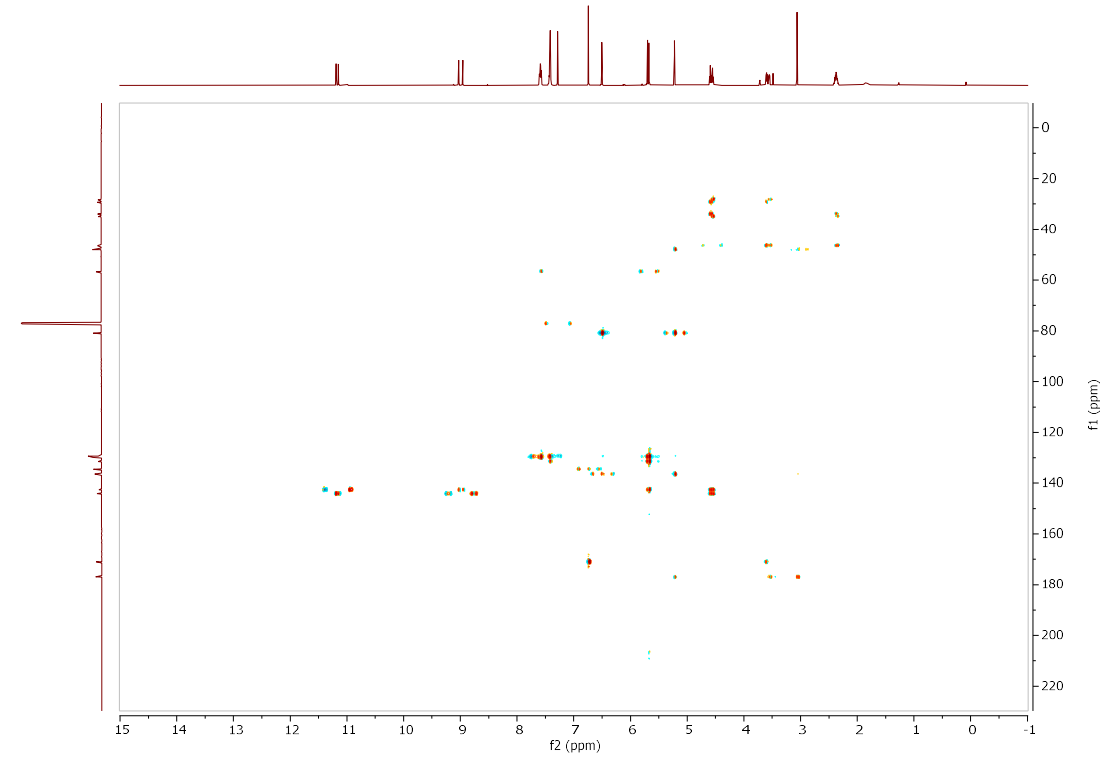
COSY Compound 46



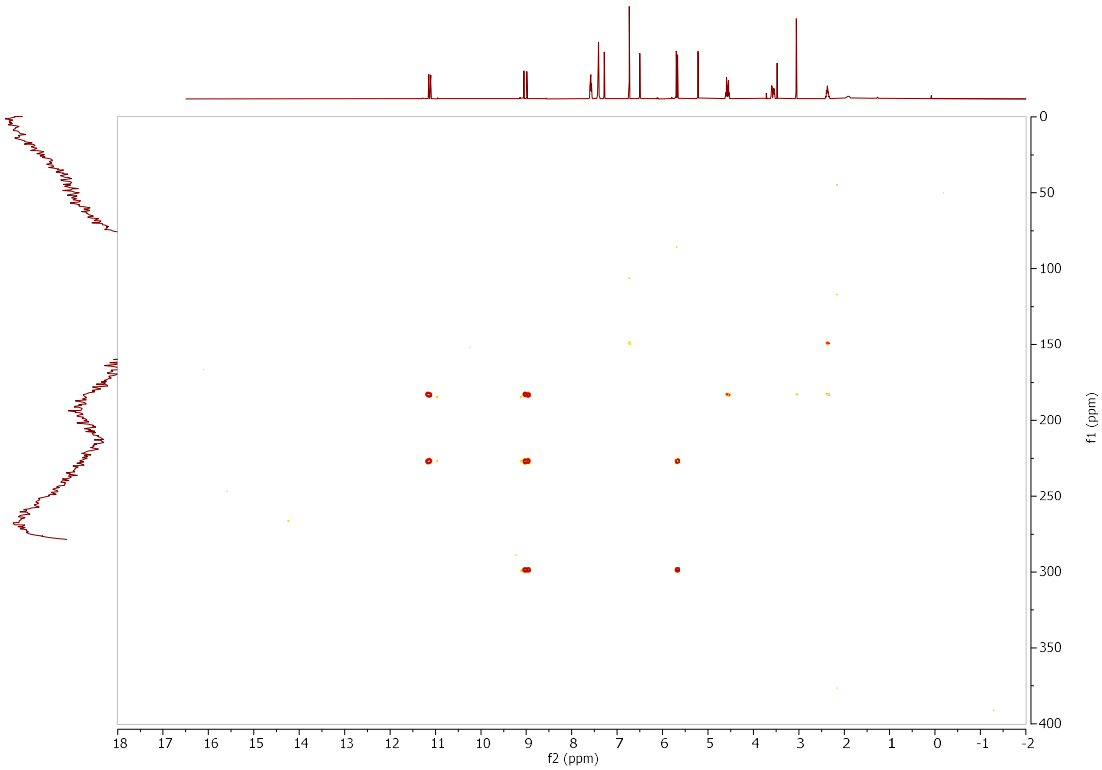
HSQC Compound 46



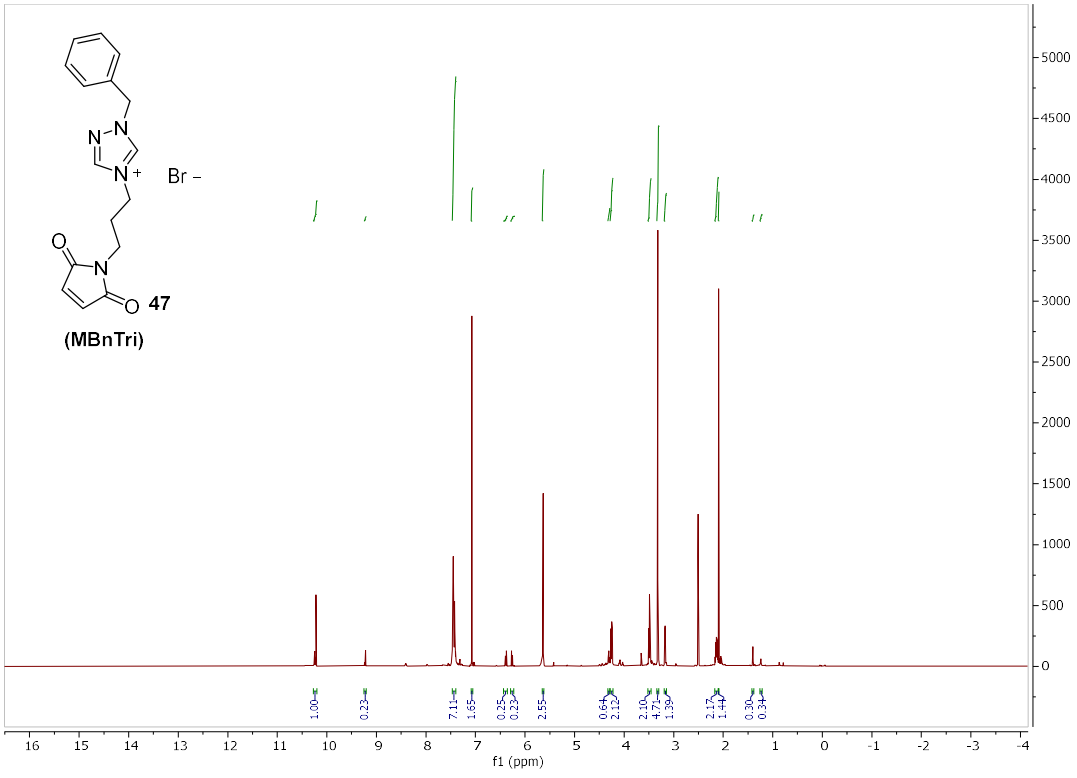
HMBC Compound 46



^{1}H-^{15}N HMBC Compound 46

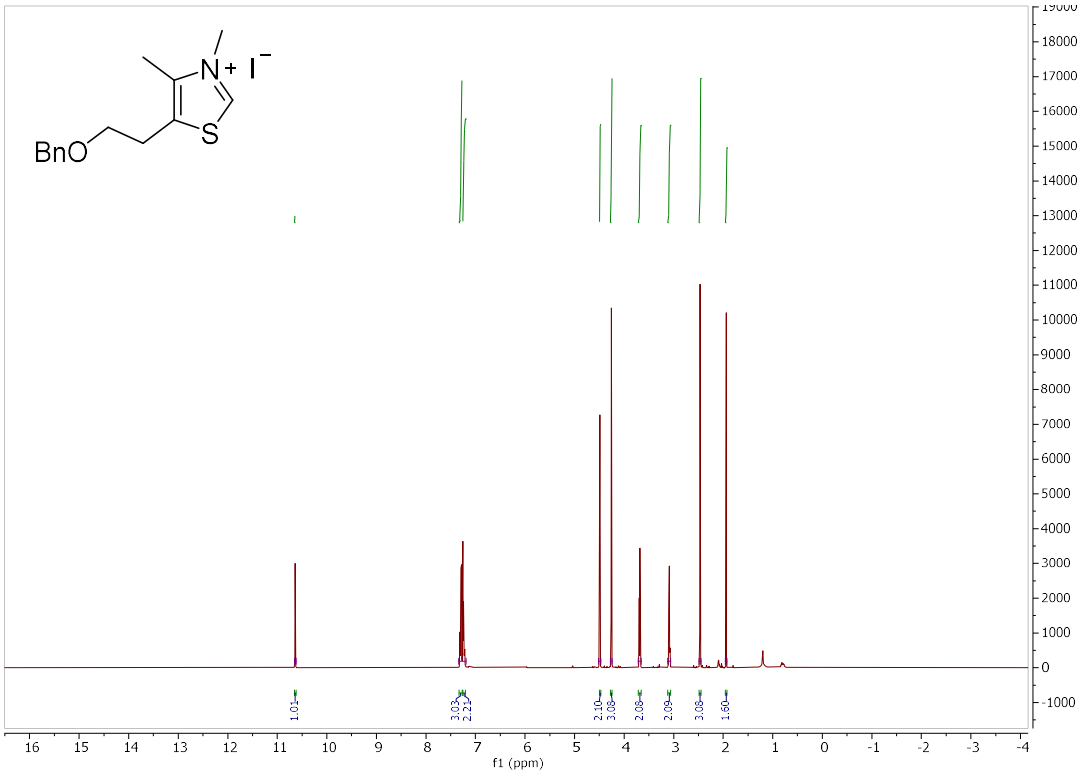


^{1}H NMR (500 MHz, DMSO-d_6) Compound 47



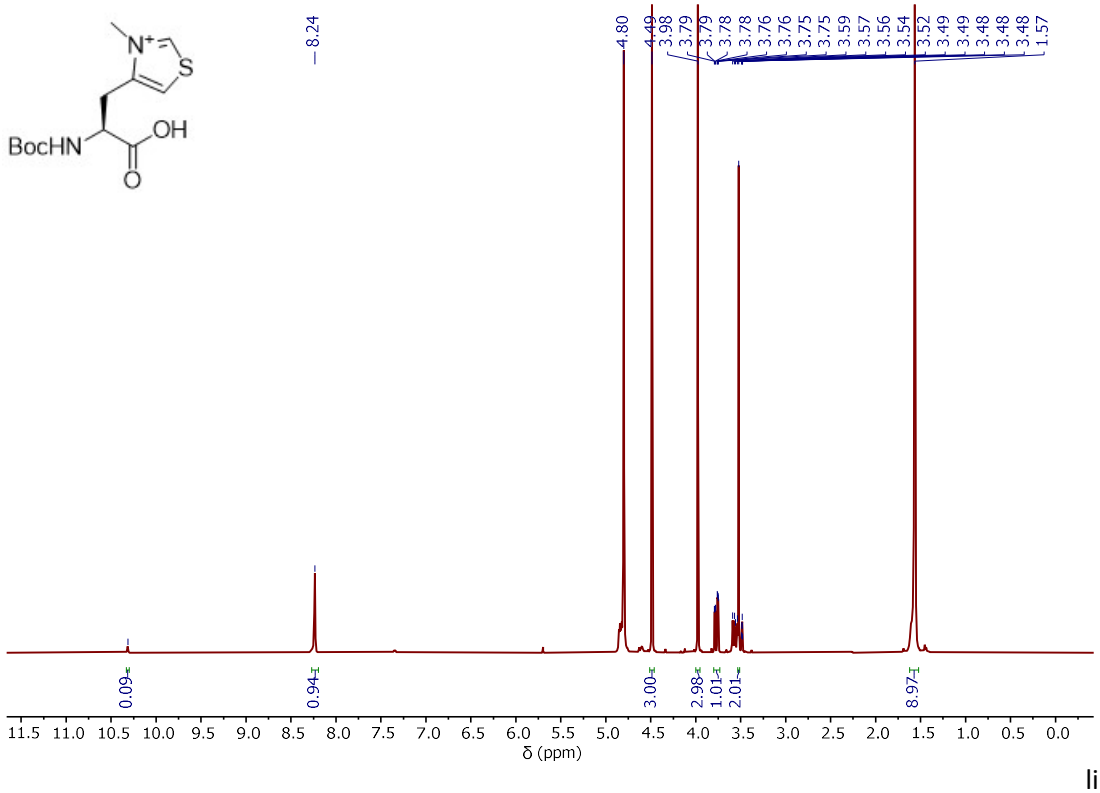
1.6.7 Compound 50

^{1}H NMR (500 MHz, Chloroform-*d*)



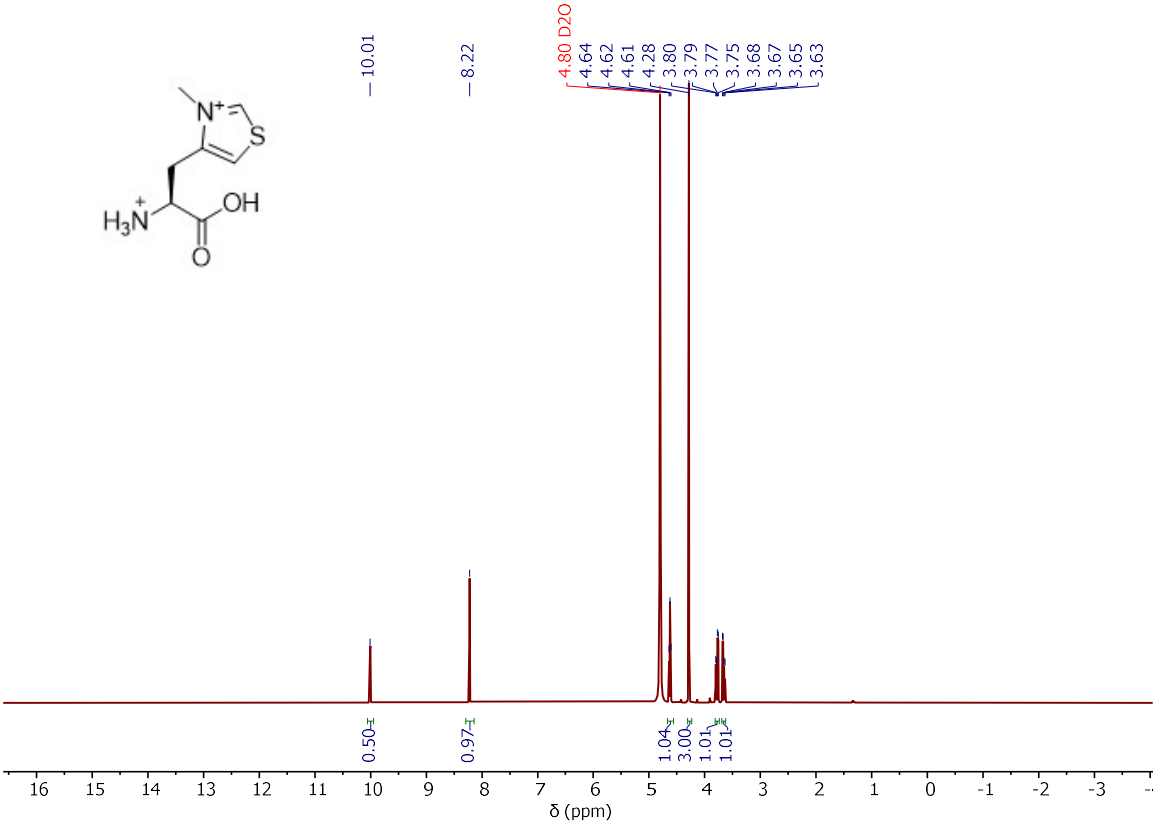
1.6.8 Me4ThzA

^{1}H NMR (500 MHz, MeOD)

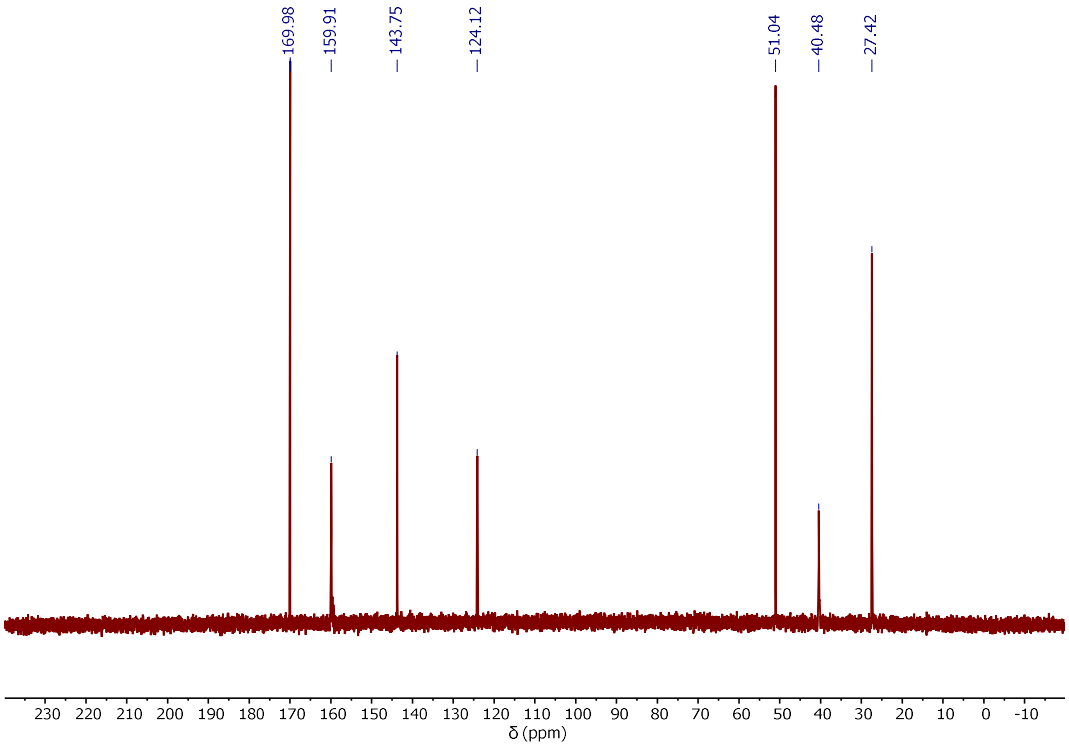


li

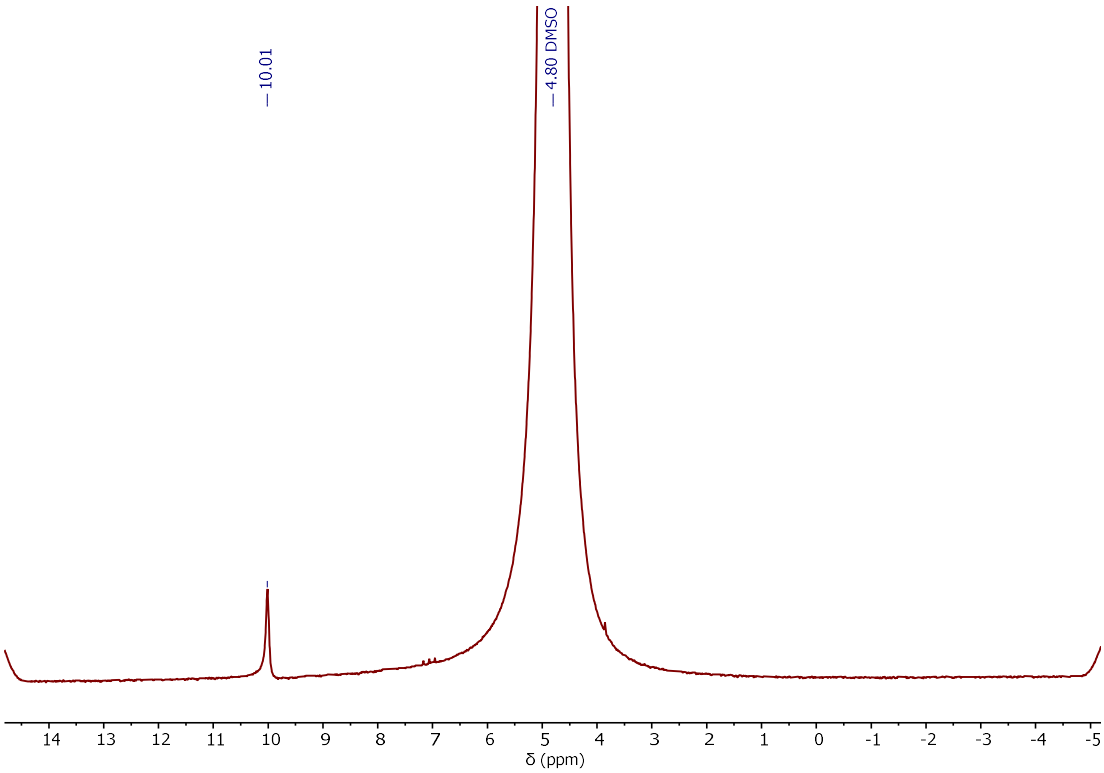
^{1}H NMR (500 MHz, D_2O)



^{13}C NMR (126 MHz, D_2O)

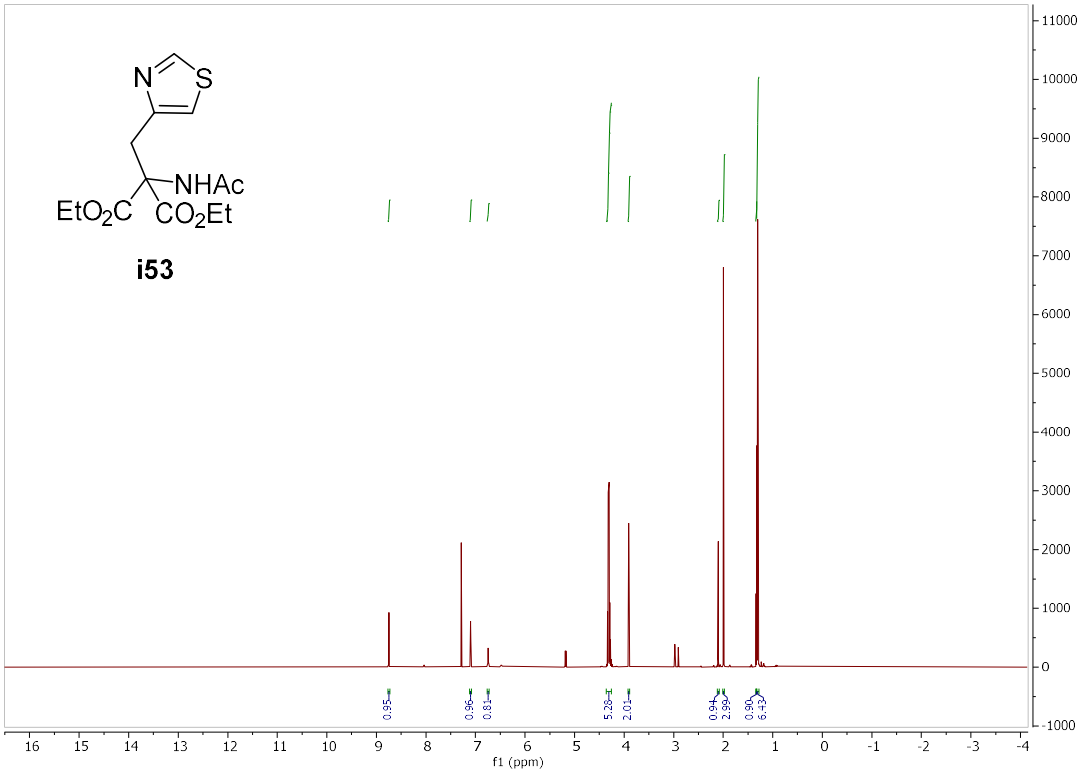


^{2}H NMR (500 MHz, D_2O)



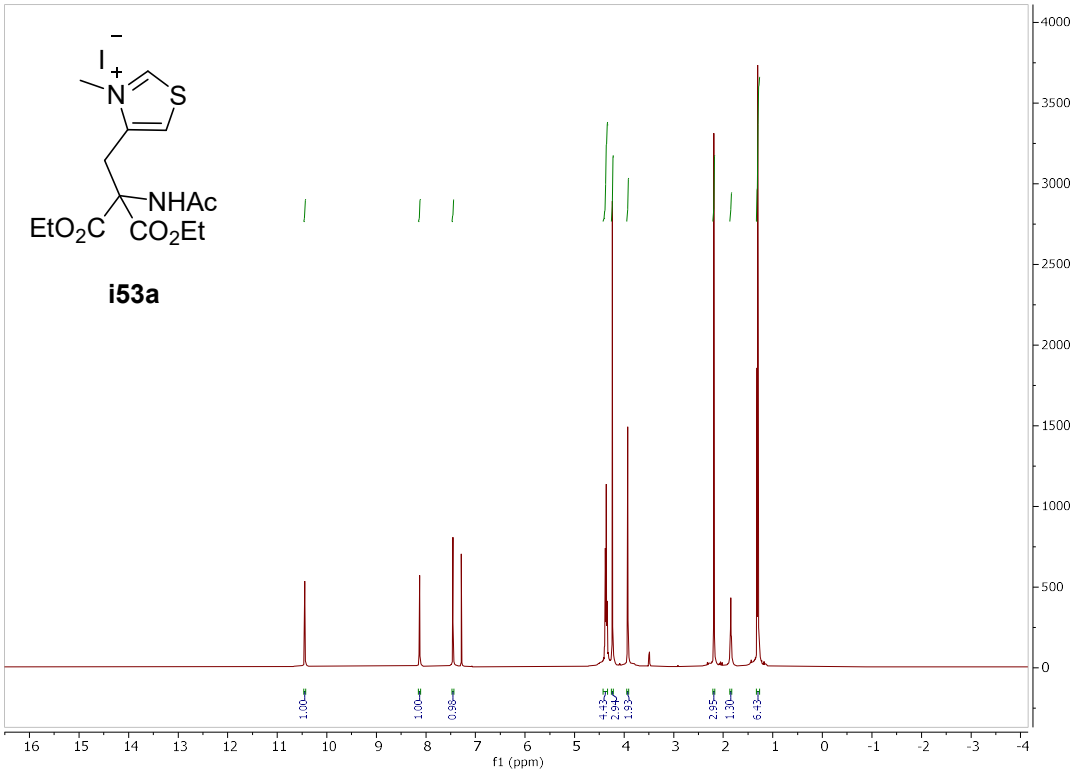
1.6.9 Racemic 4ThzA

Intermediate i53, ^{1}H NMR (500 MHz, $CDCl_3$)

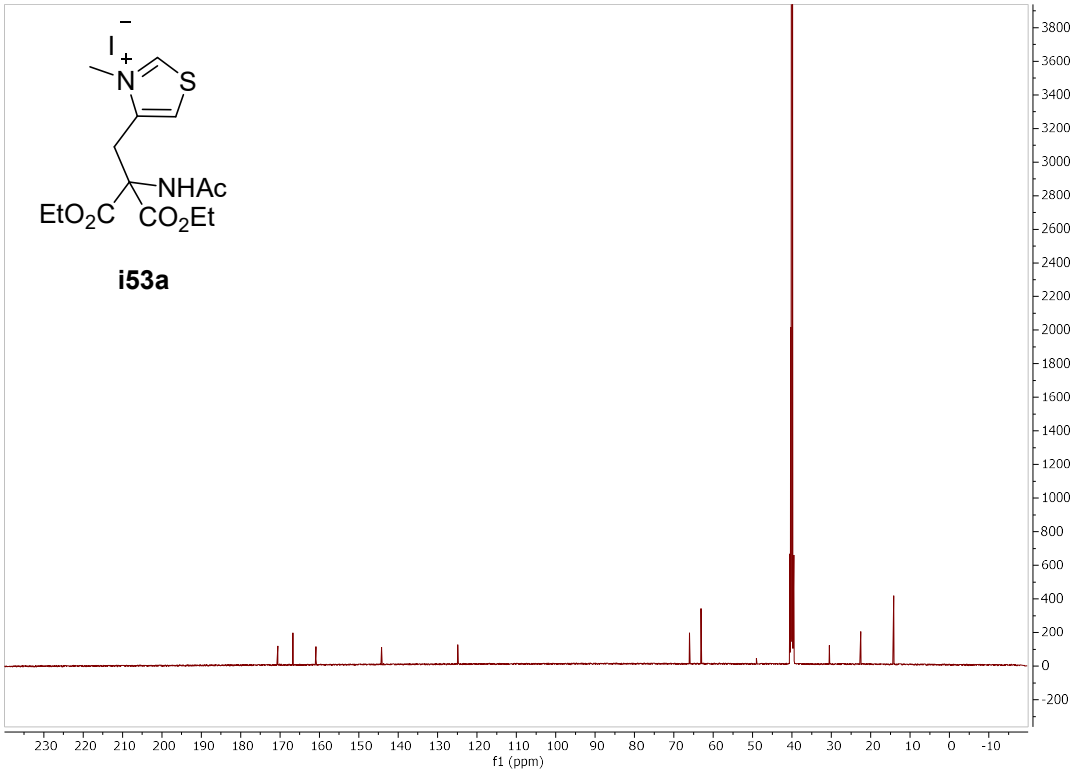


1.6.10 Racemic Me4ThzA

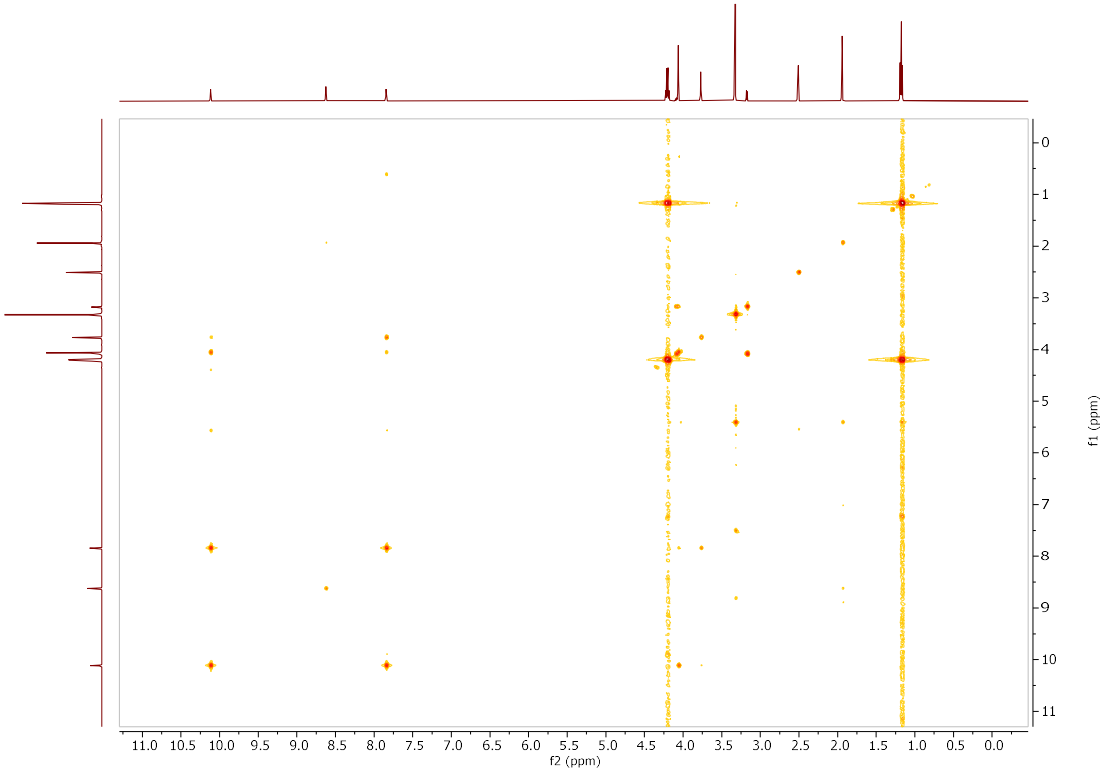
Intermediate i53a, ^{1}H NMR (500 MHz, $CDCl_3$)



^{13}C NMR (126 MHz, DMSO)

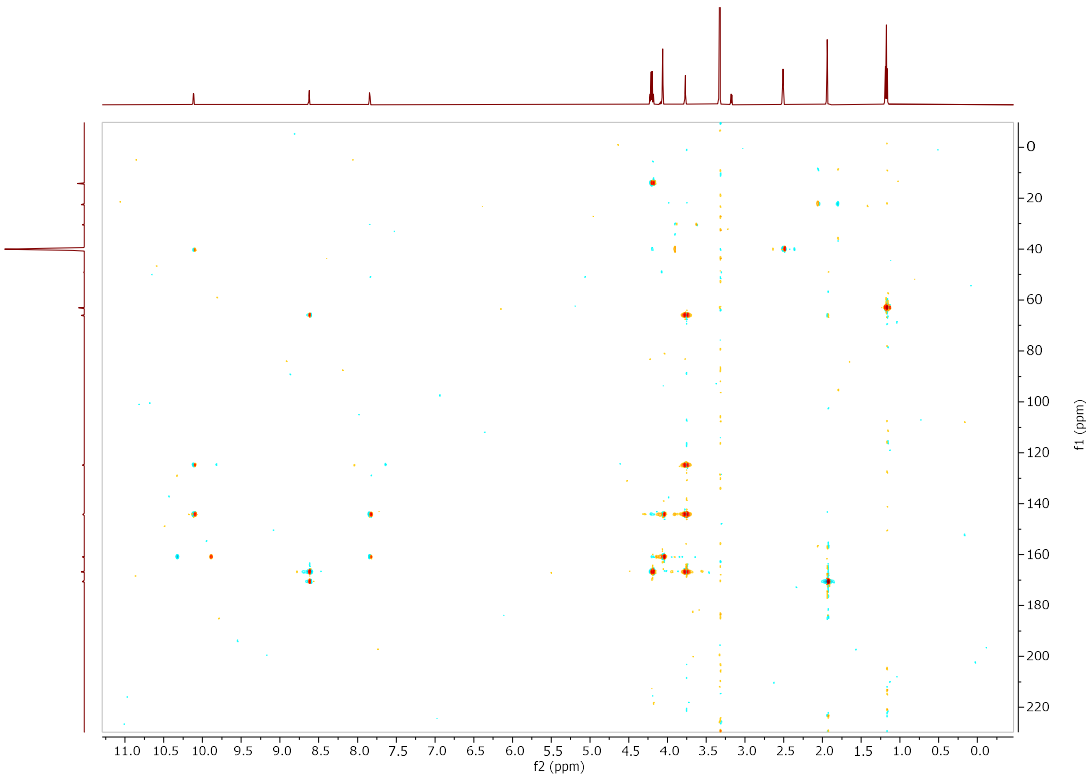


COSY i53a (DMSO)

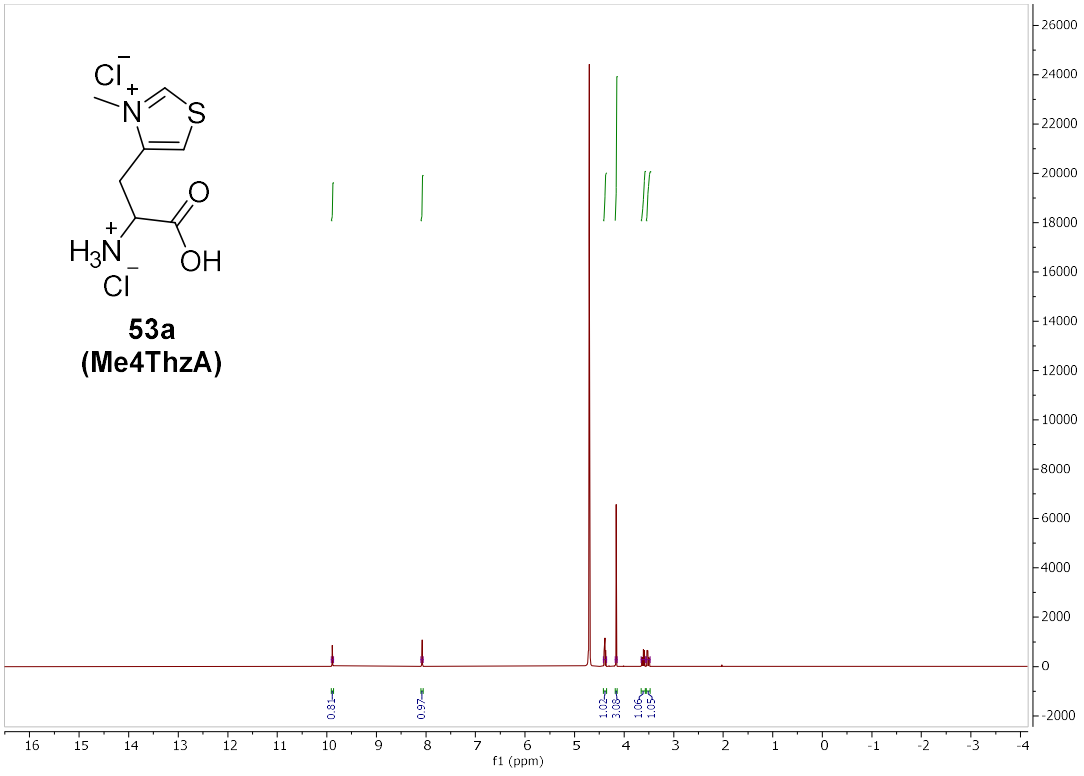


HSQC i53a (DMSO)

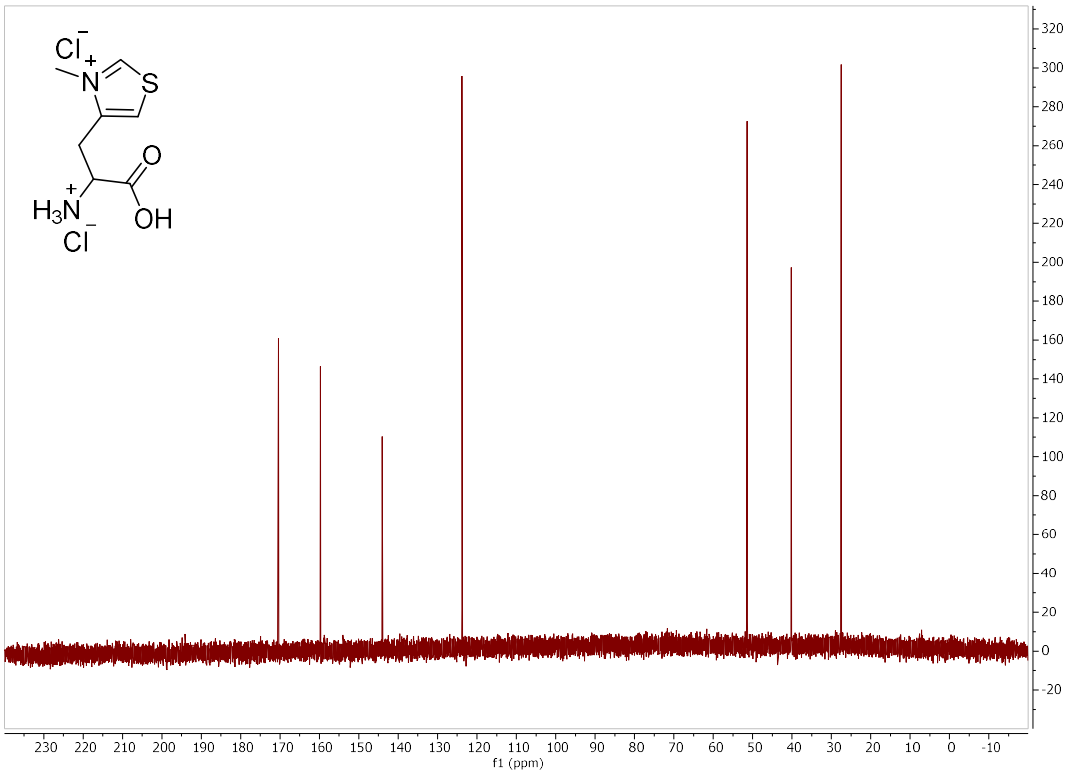
HMBC i53a (DMSO)



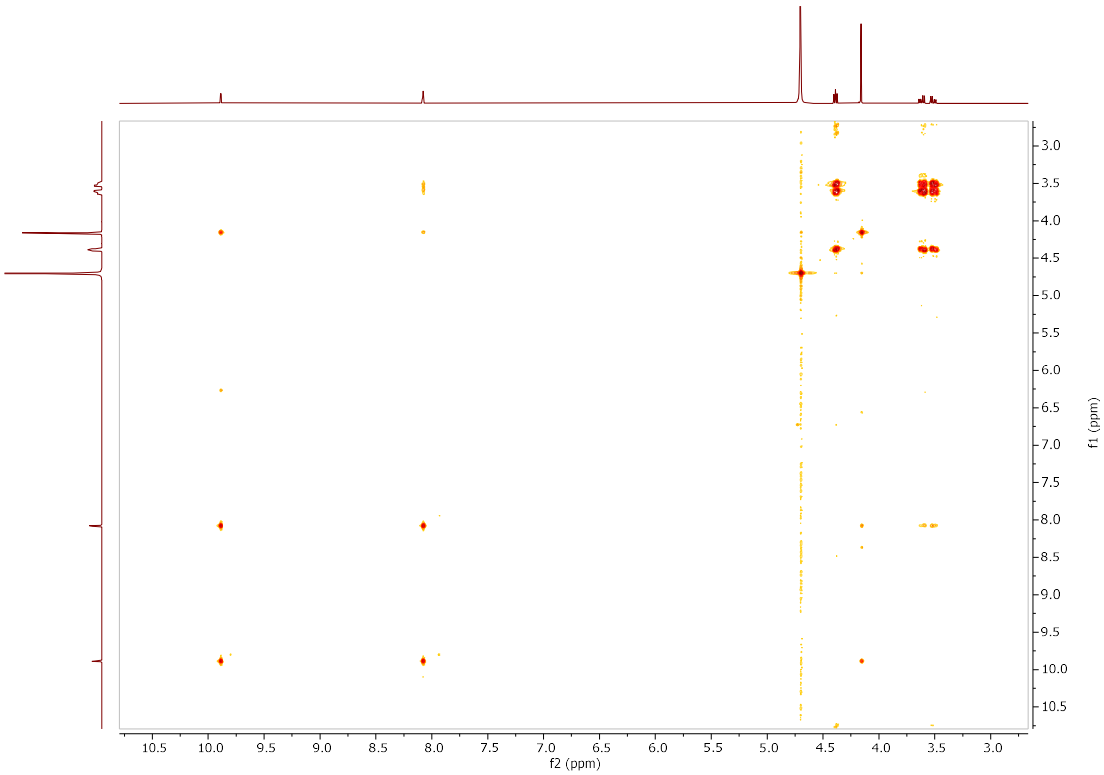
^{1}H NMR (500 MHz, D_2O)



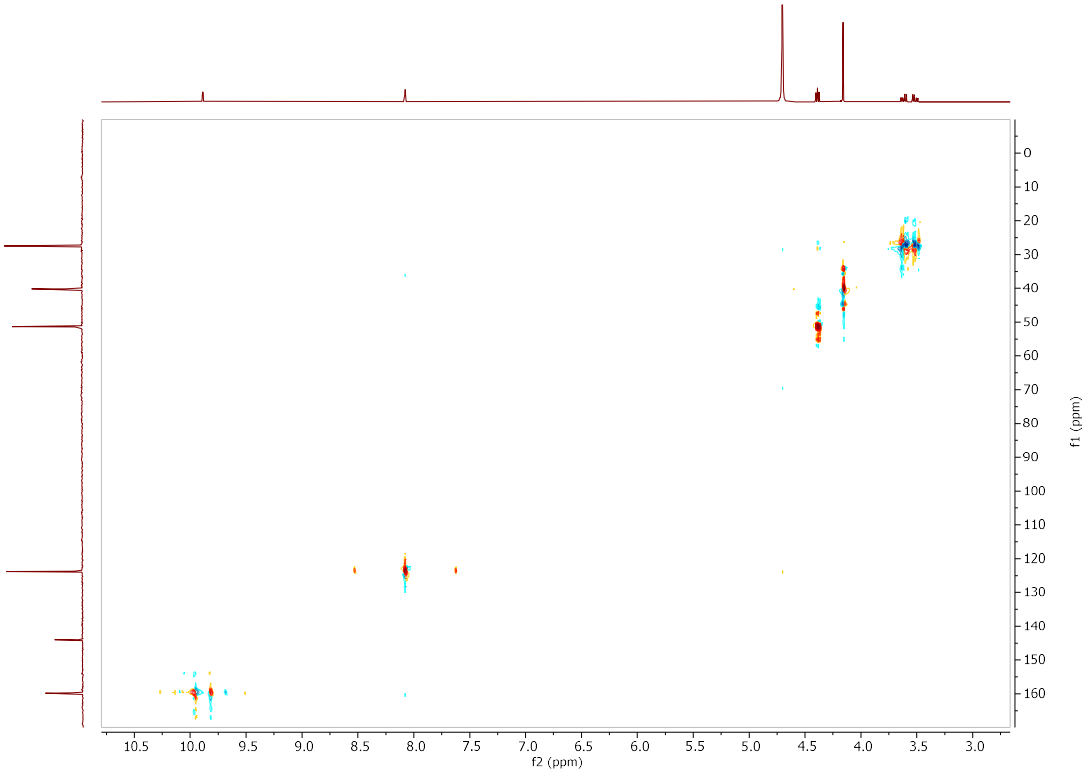
^{13}C NMR (126 MHz, D_2O)



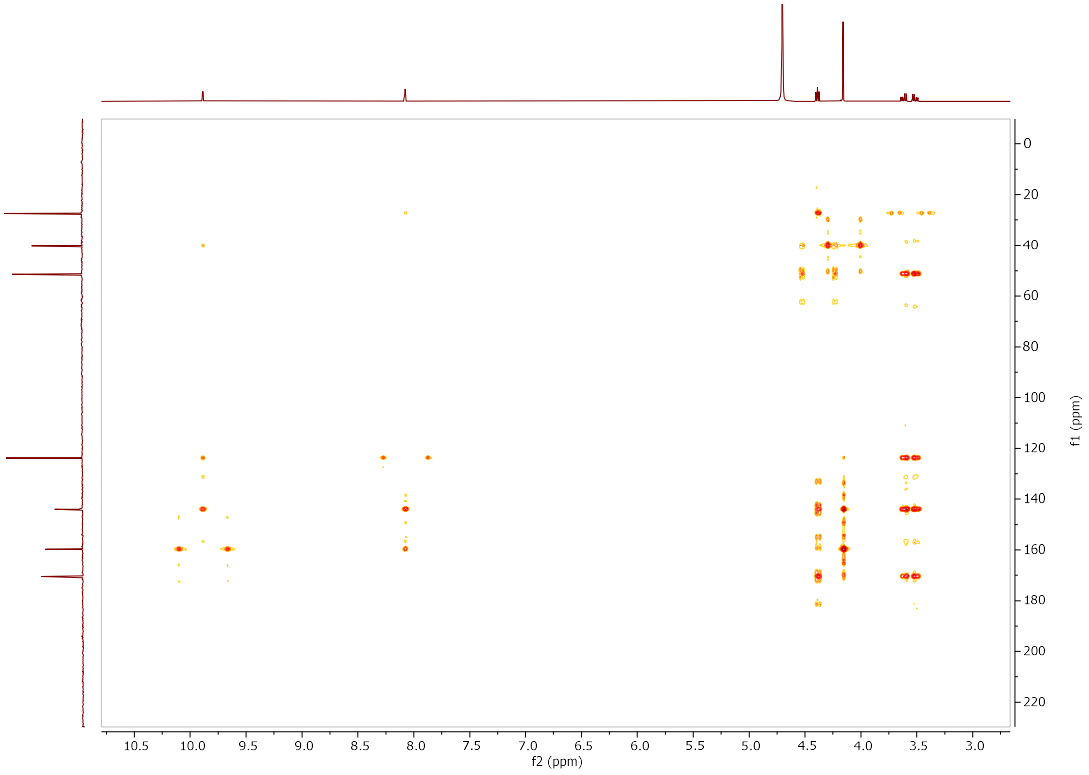
COSY 53a (D_2O)



HSQC 53a (D_2O)

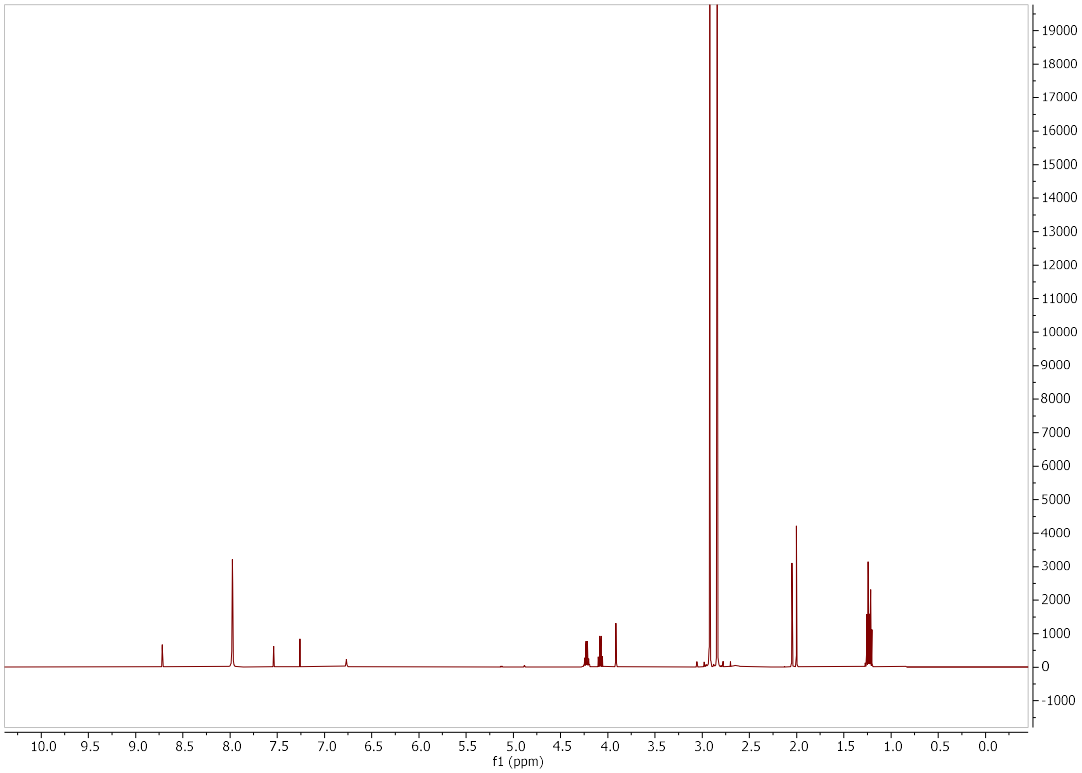


HMBC 53a (D_2O)

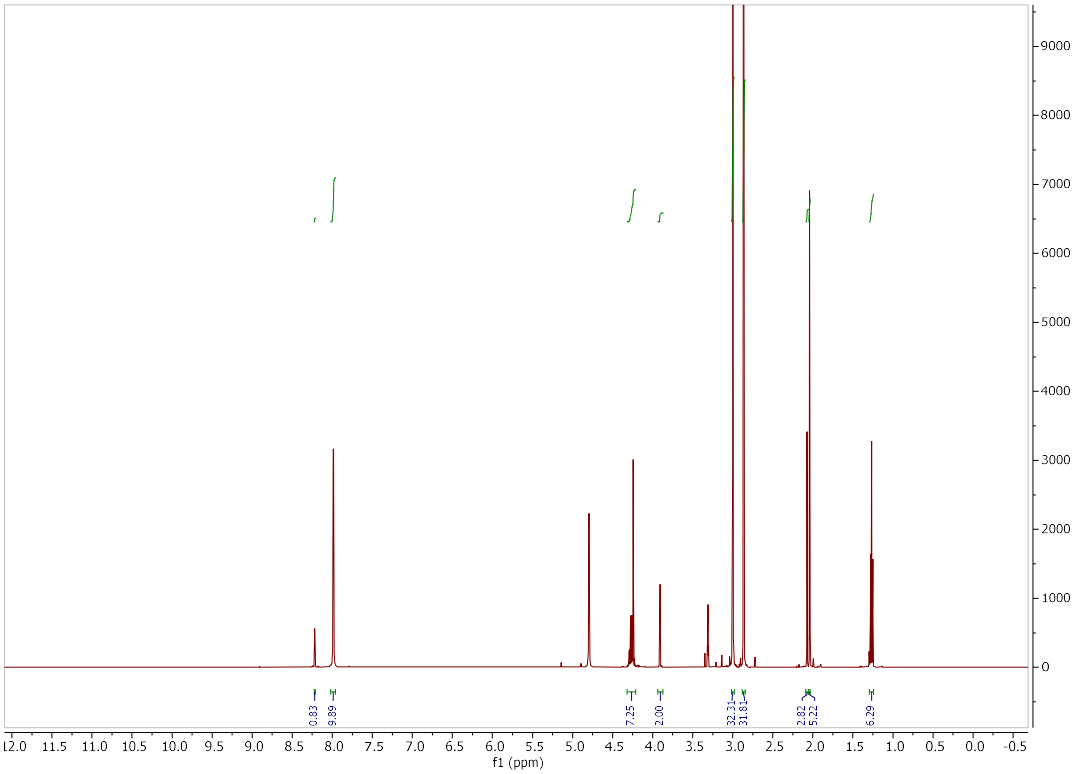


1.6.11 5ThzA

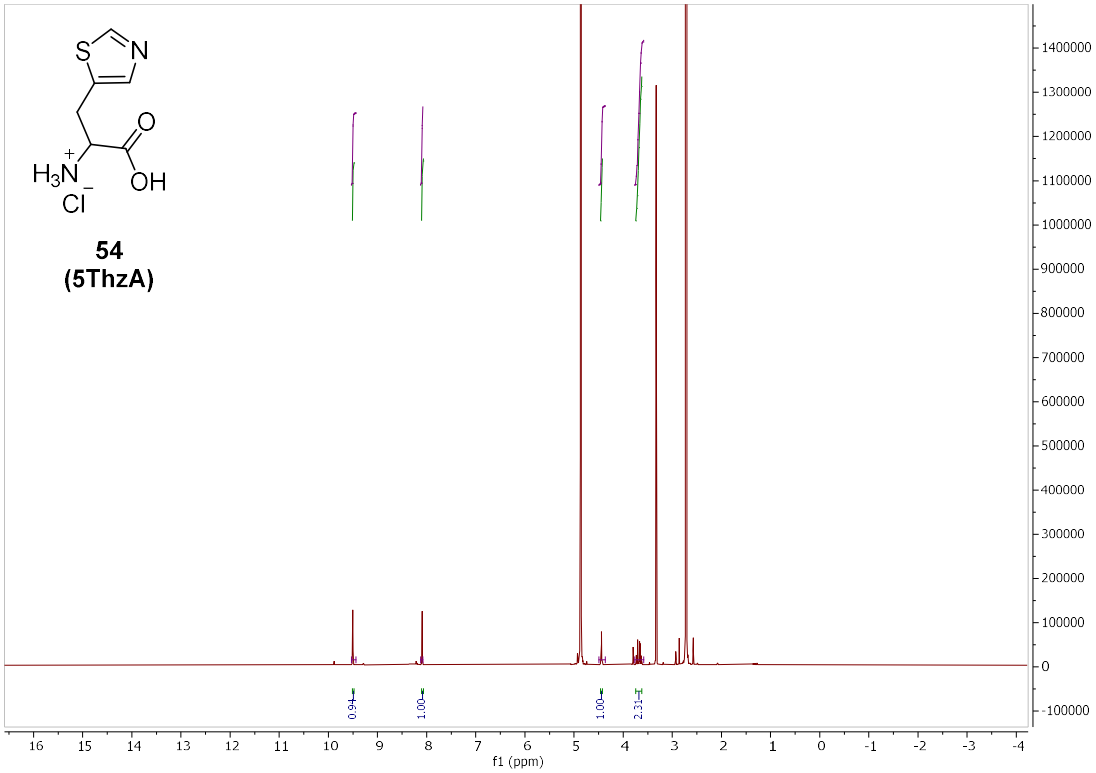
Intermediate **i54**



Intermediate **i55**

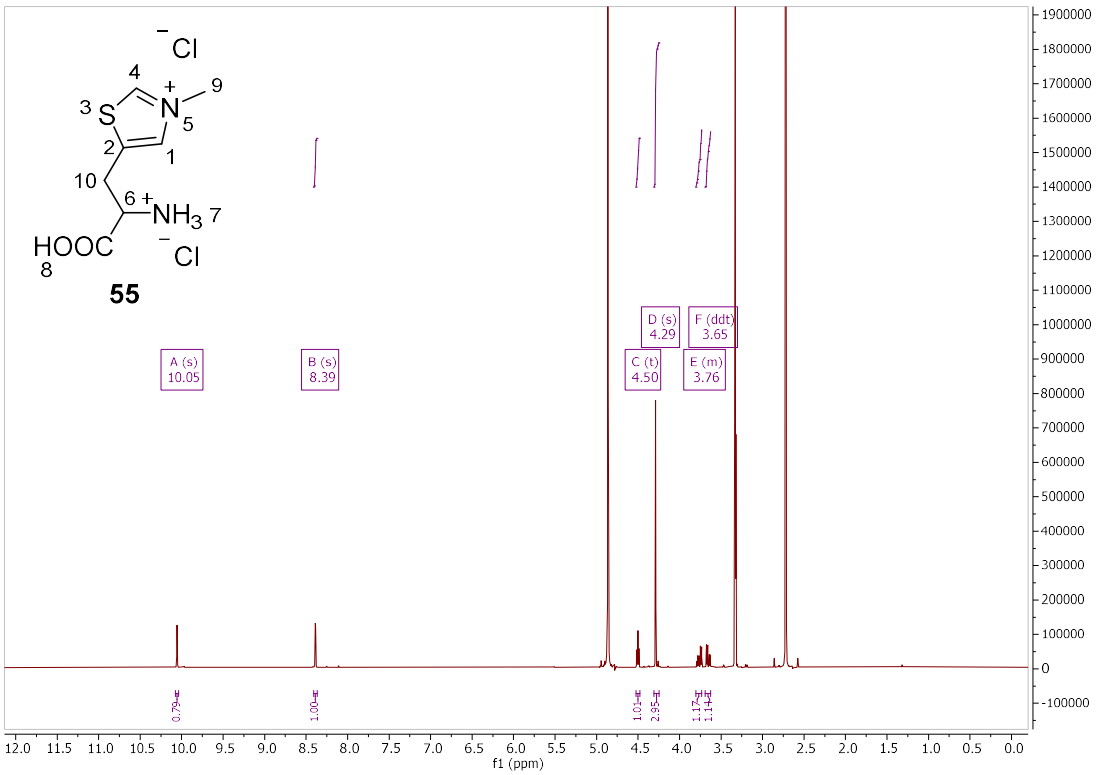


^{1}H NMR (500 MHz, Methanol-d_4)

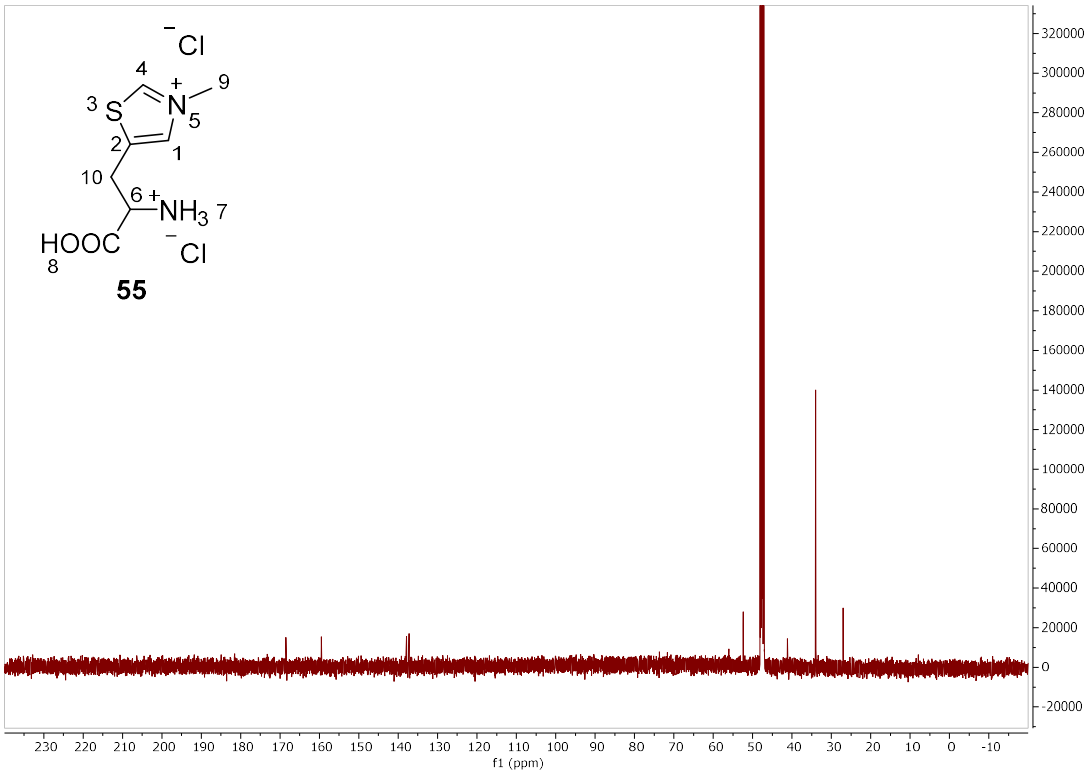


1.6.12 Me5ThzA

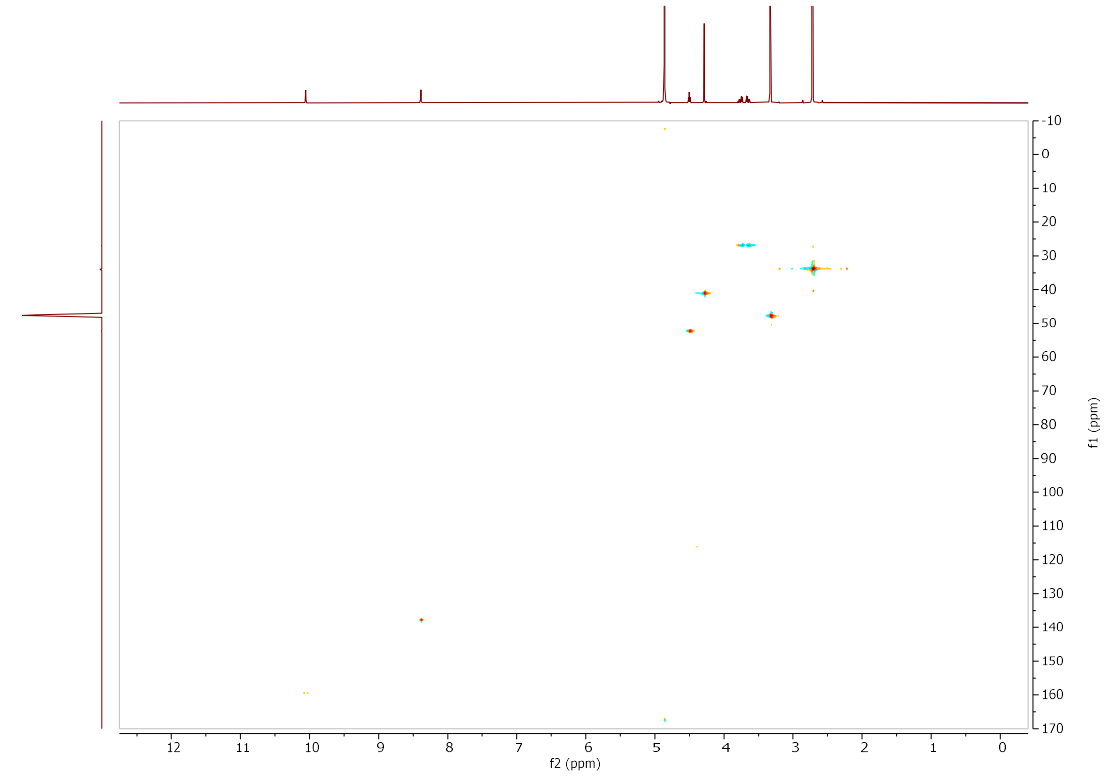
^{1}H NMR (500 MHz, Methanol-d_4)



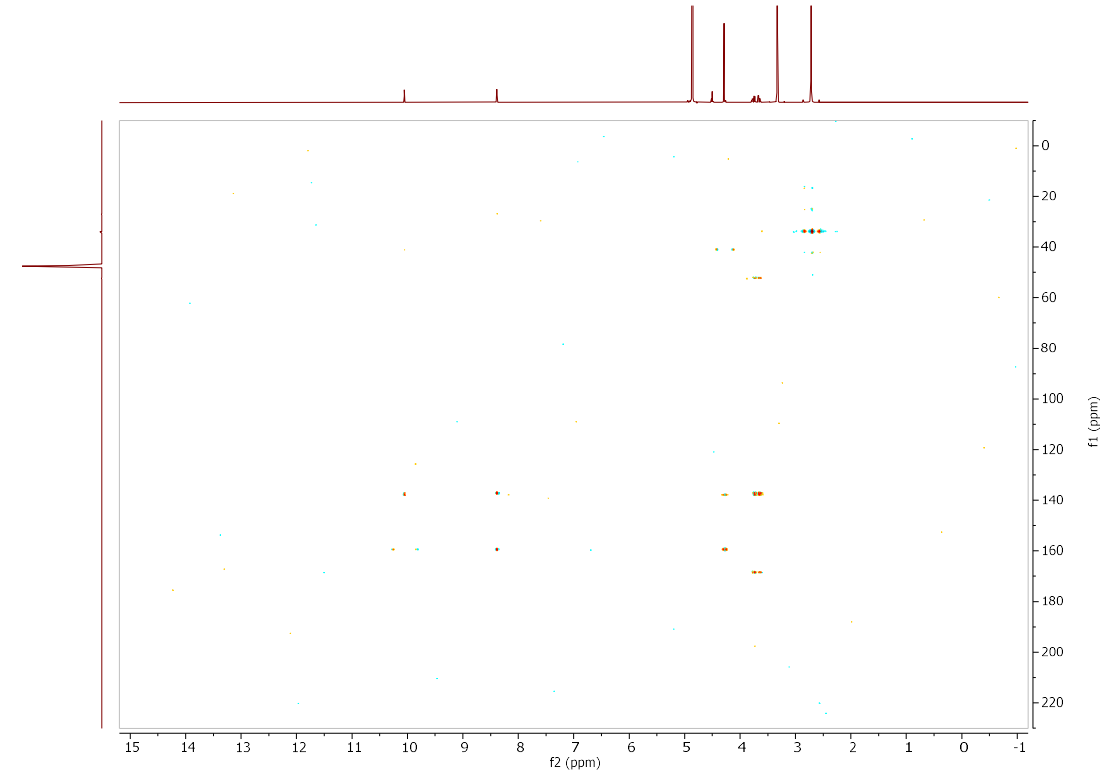
^{13}C NMR (126 MHz, MeOD)



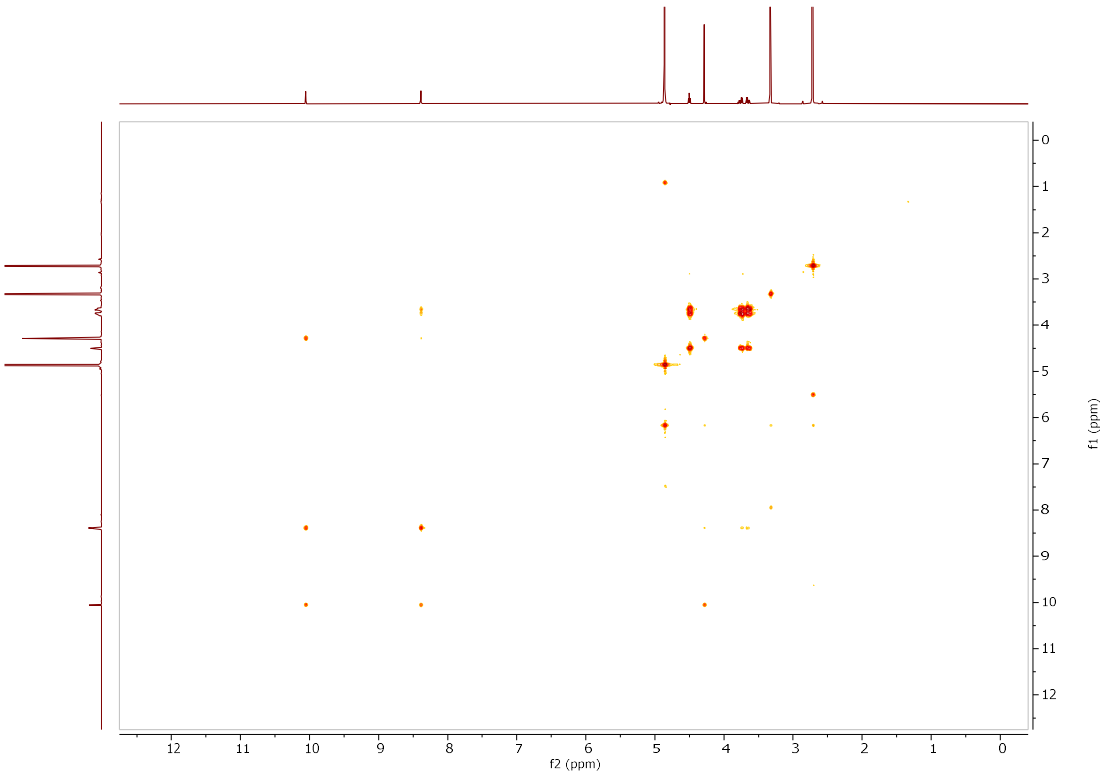
HSQC- compound 55



HMBC compound 55

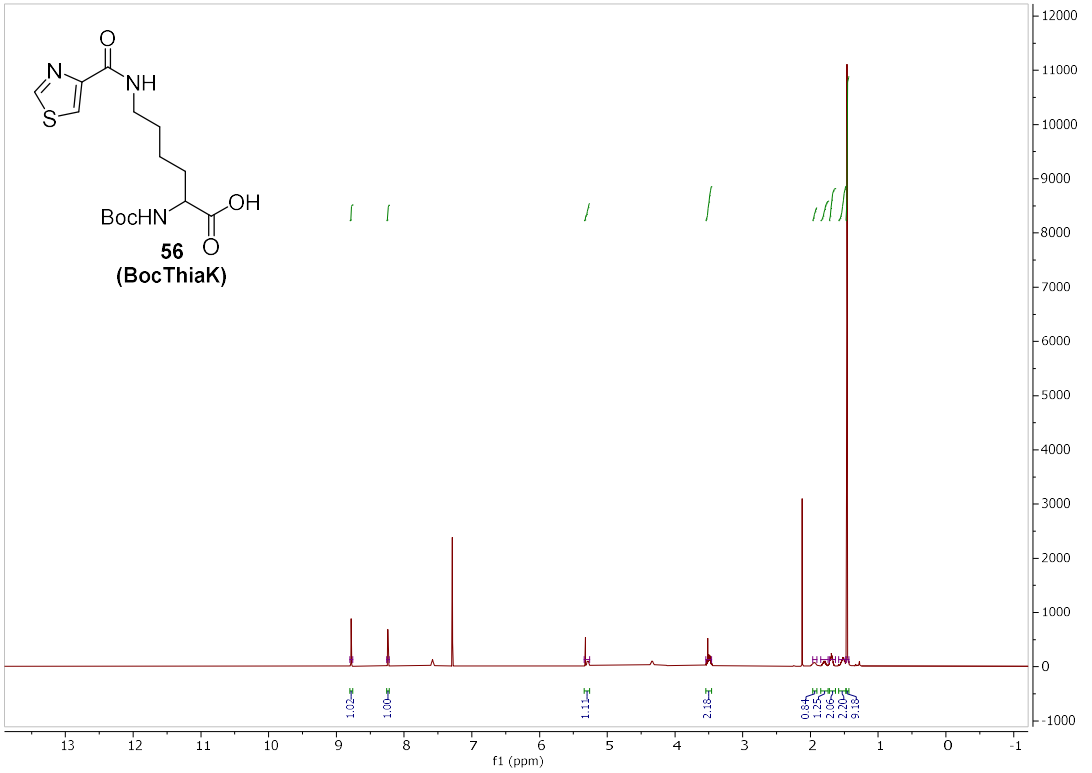


COSY compound 55

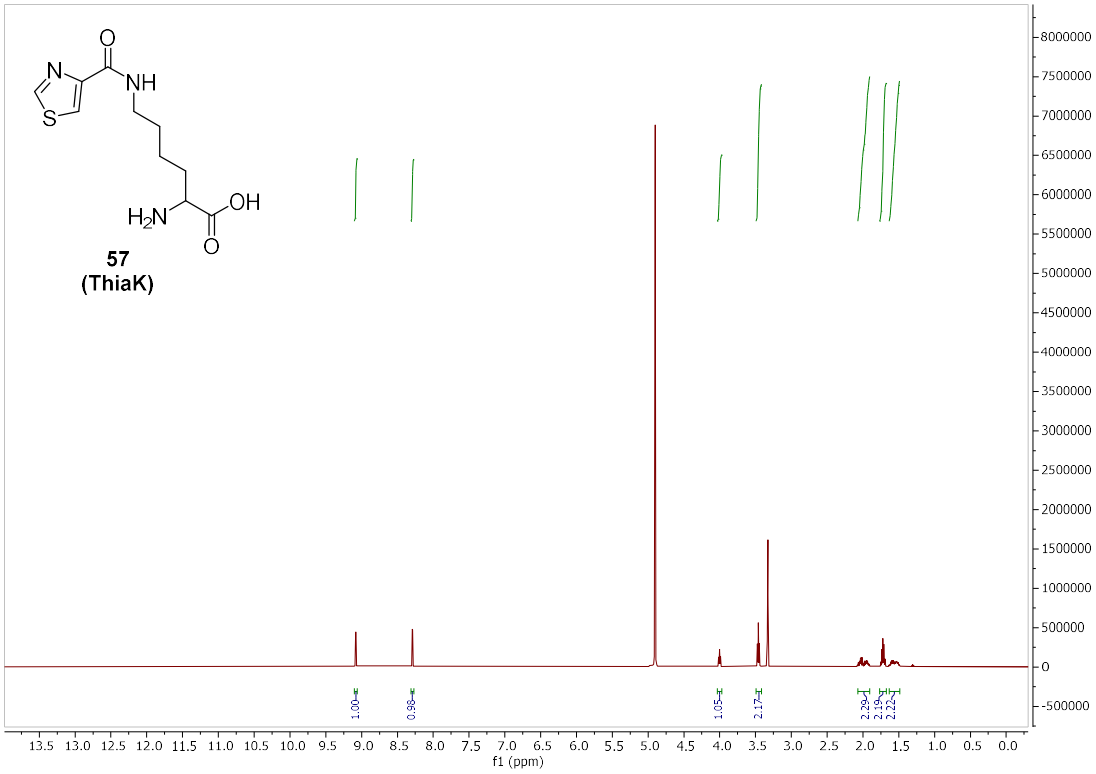


1.6.13 ThiaKOMe

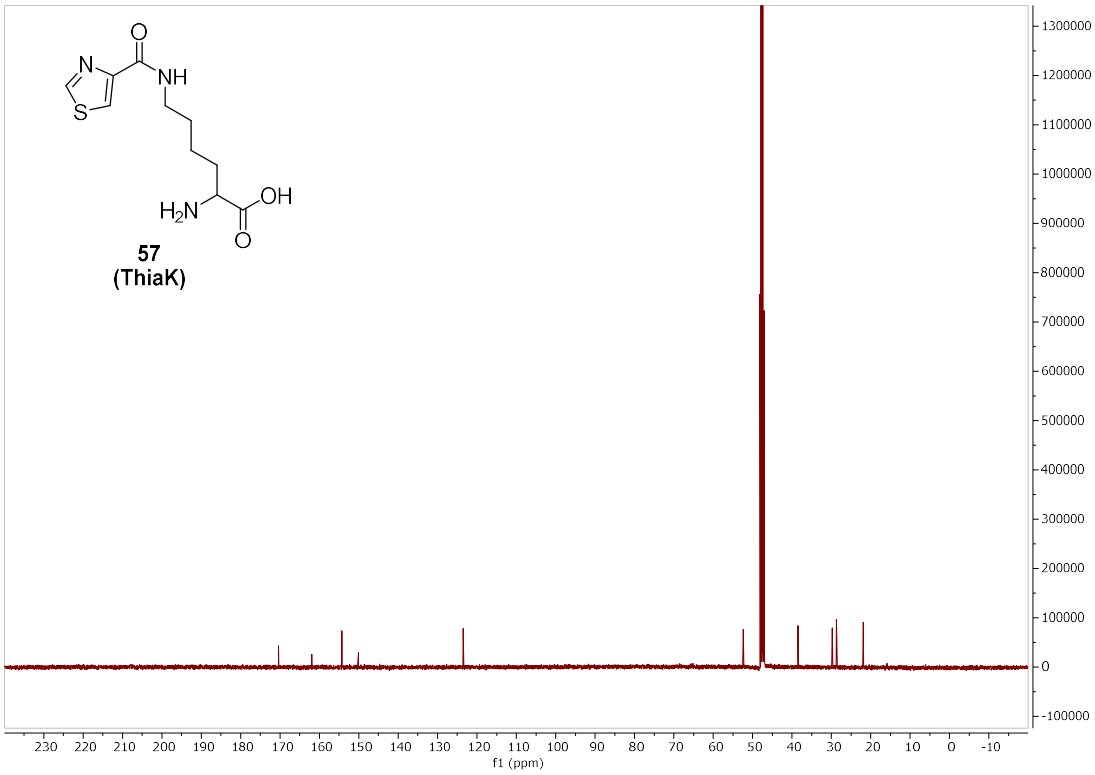
^{1}H NMR (500 MHz, Chloroform-*d*) Intermediate 56



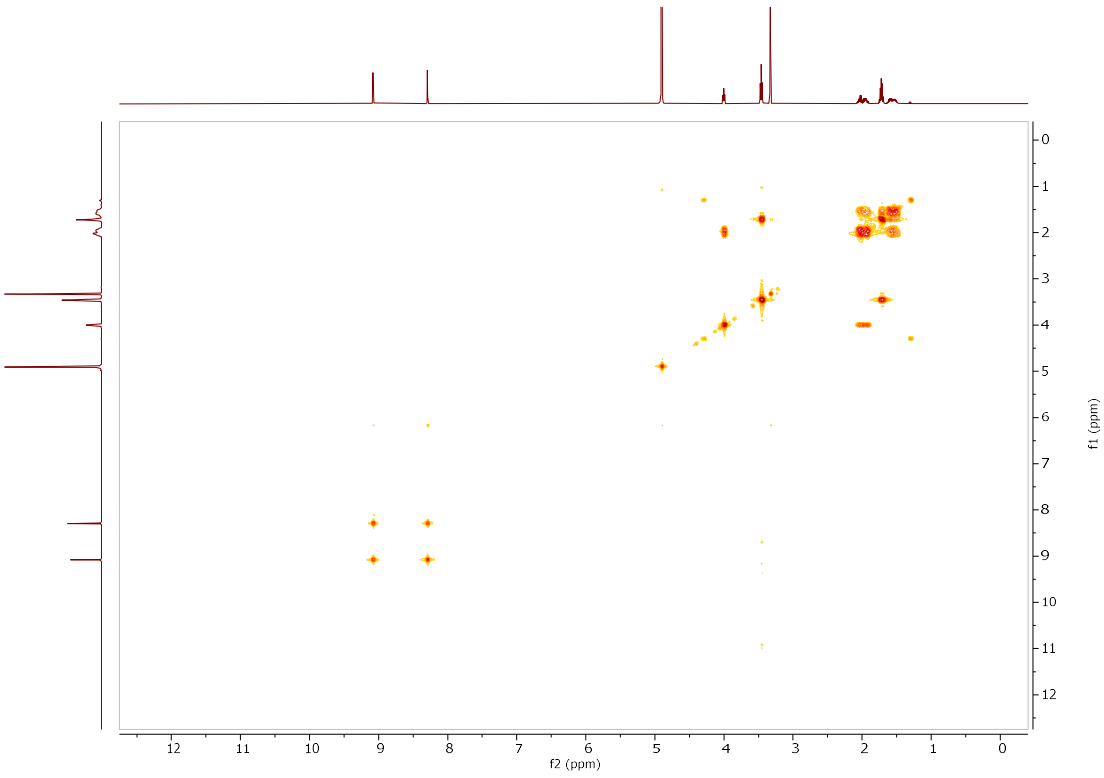
^{1}H NMR (500 MHz, Methanol-d_4) Compound 57



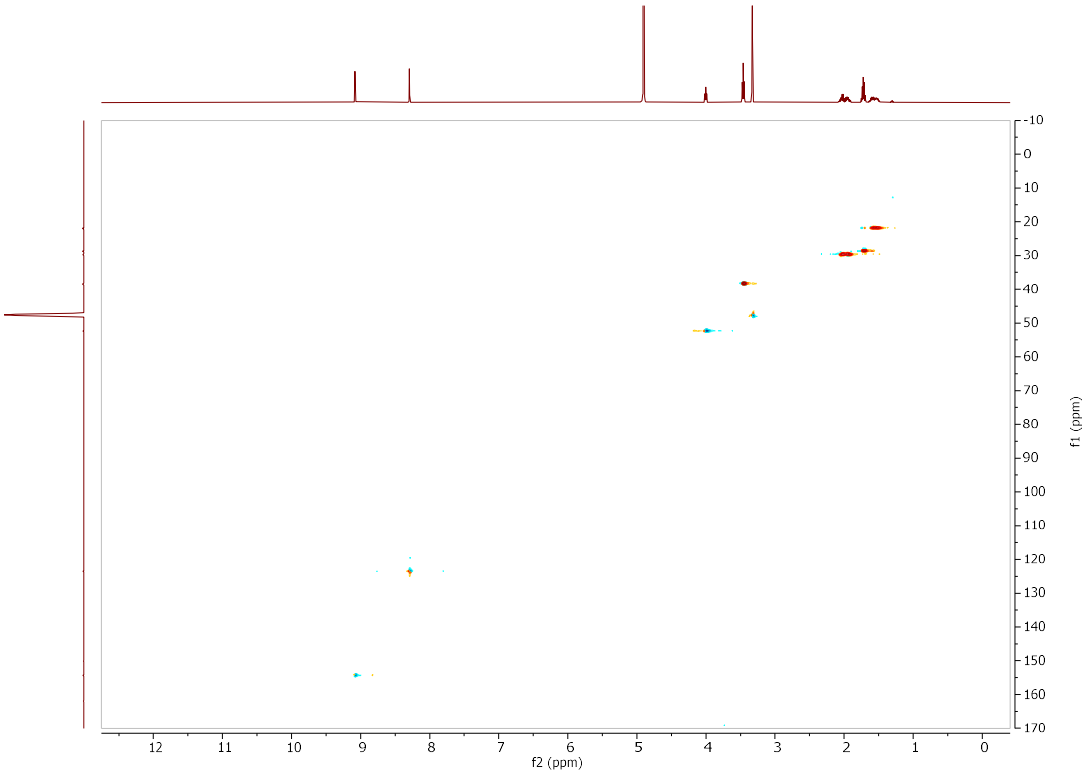
^{13}C NMR (126 MHz, MeOD) Compound 57



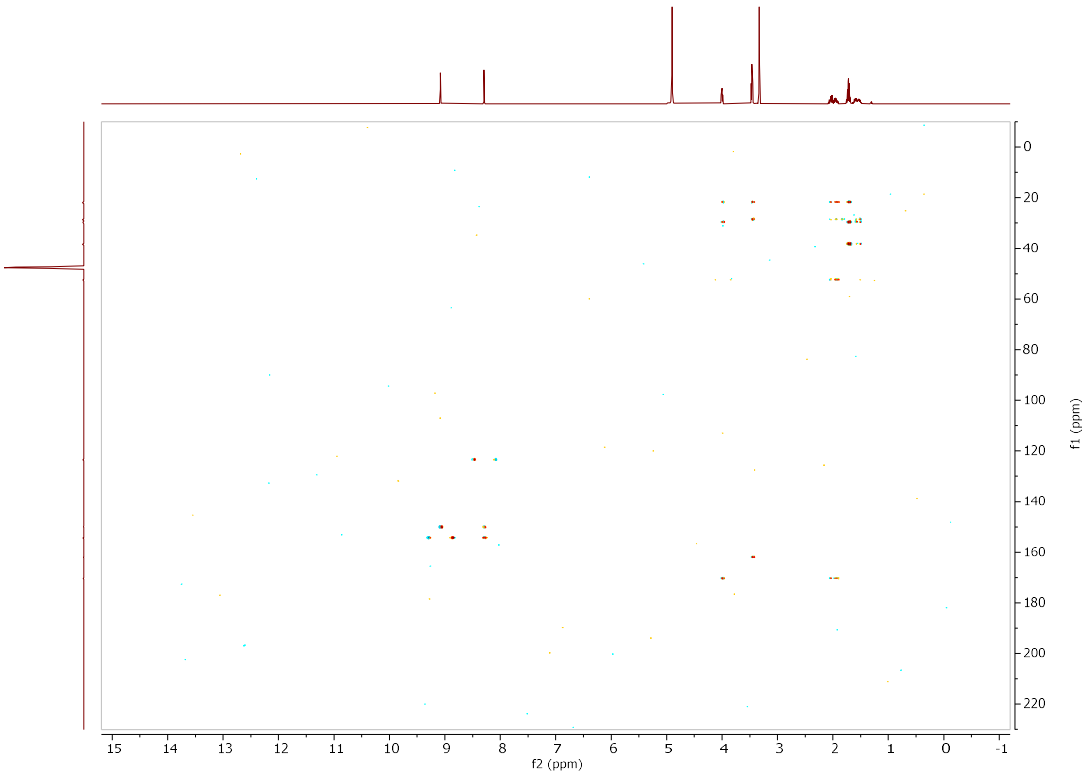
COSY Compound 57, MeOD



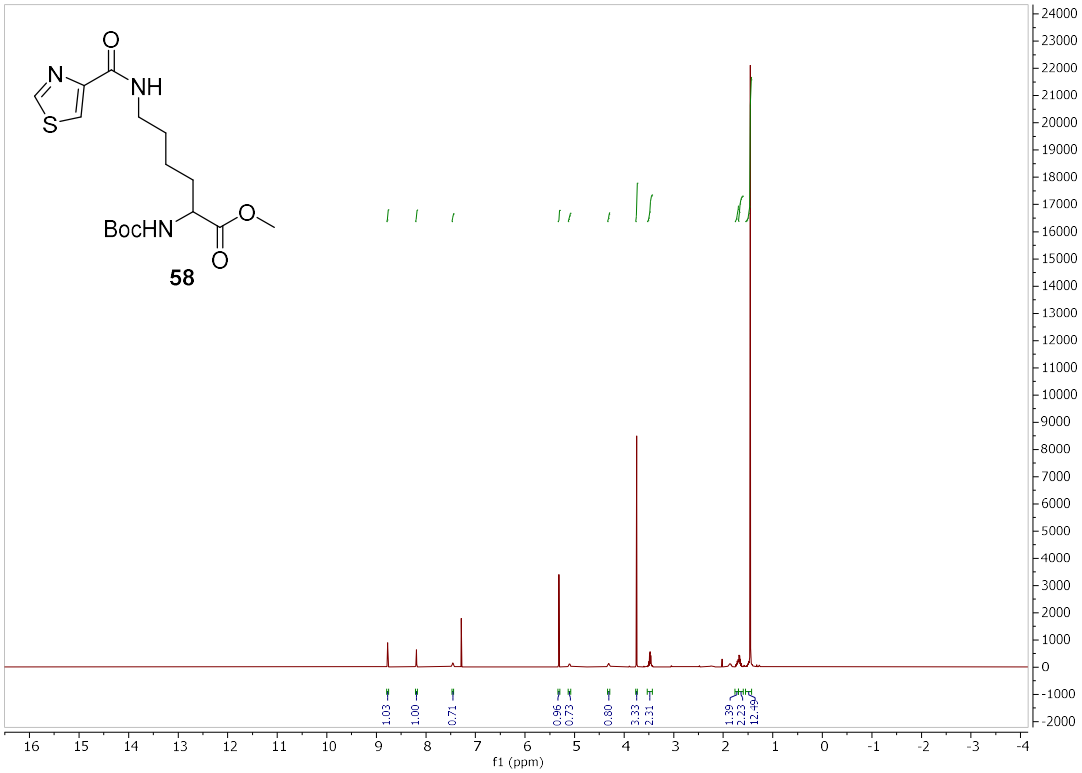
HSQC Compound 57, MeOD



HMBC Compound 57, MeOD

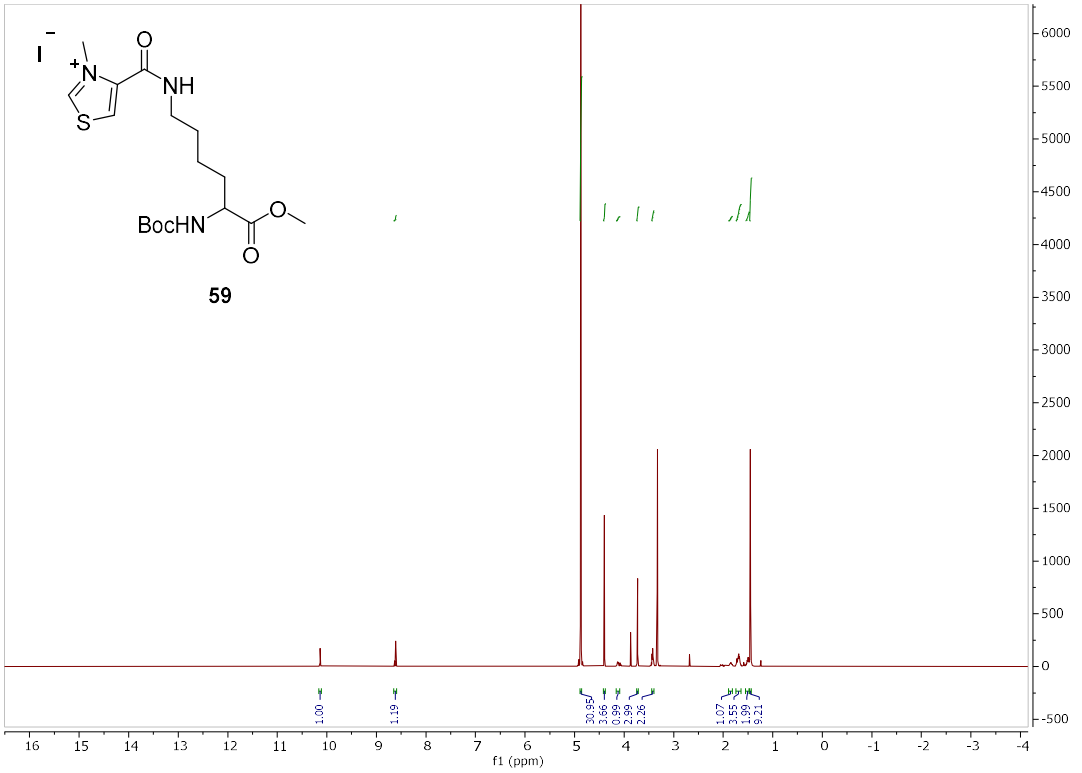


^{1}H NMR (500 MHz, Chloroform-*d*) Intermediate 58

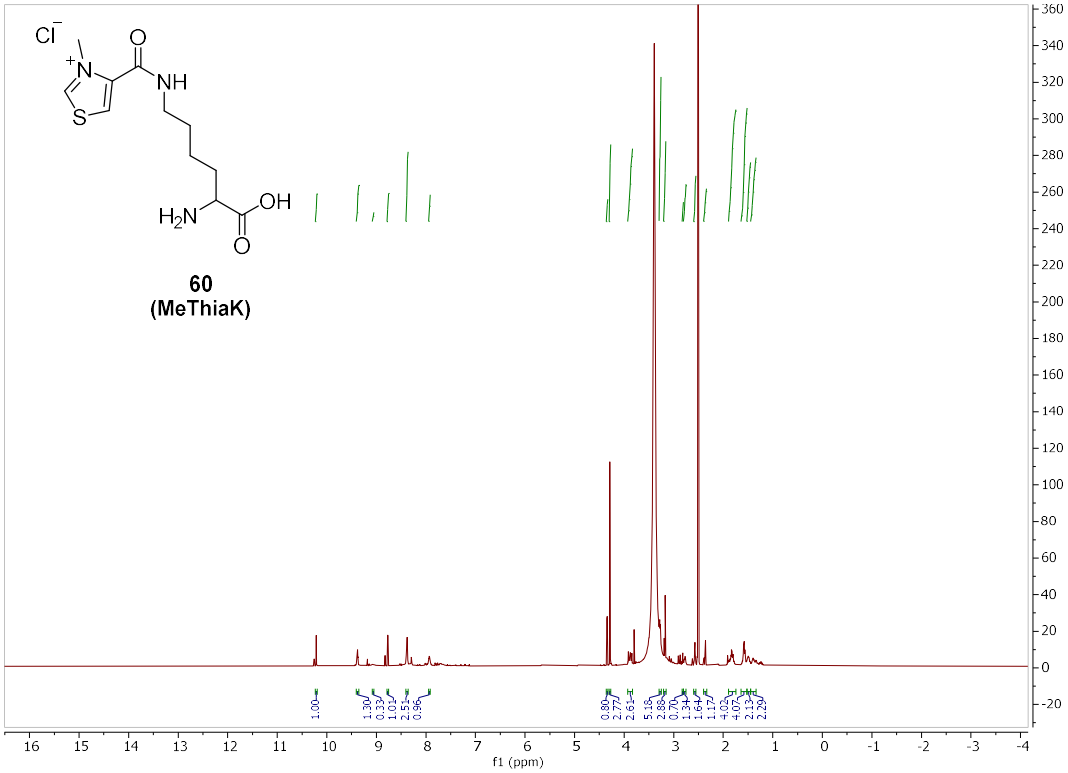


1.6.14 MeThiaKOMe

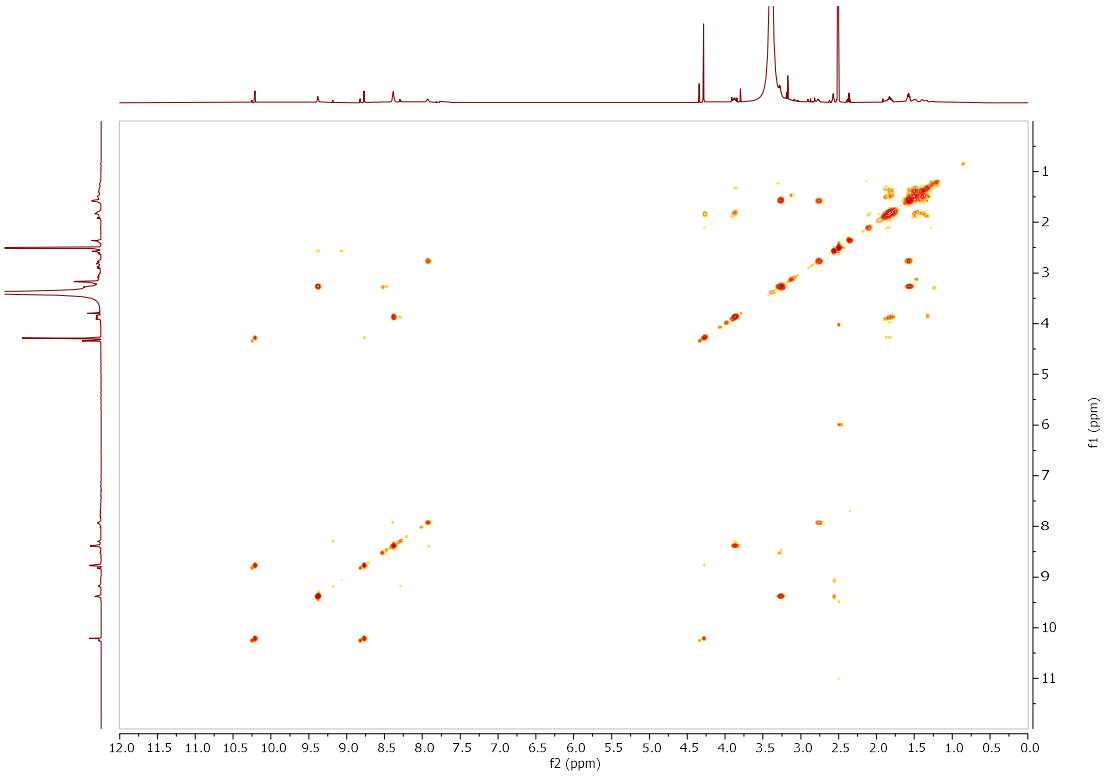
Intermediate 59



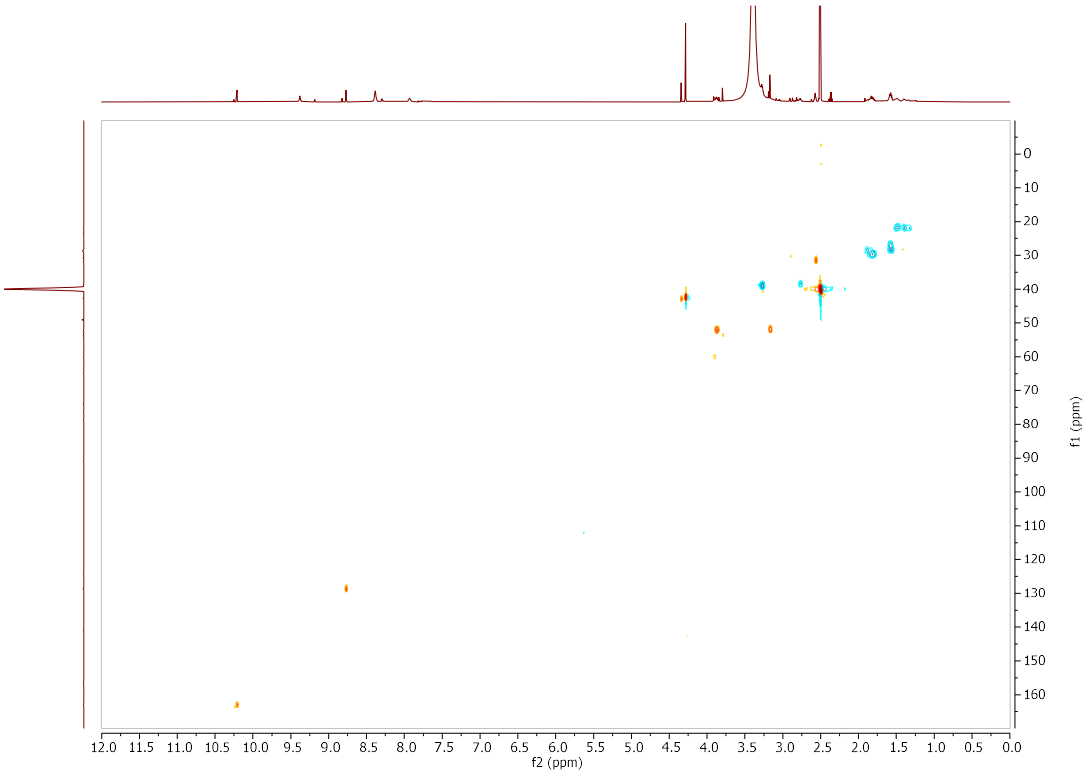
^{1}H NMR (500 MHz, DMSO-d_6) Compound 60



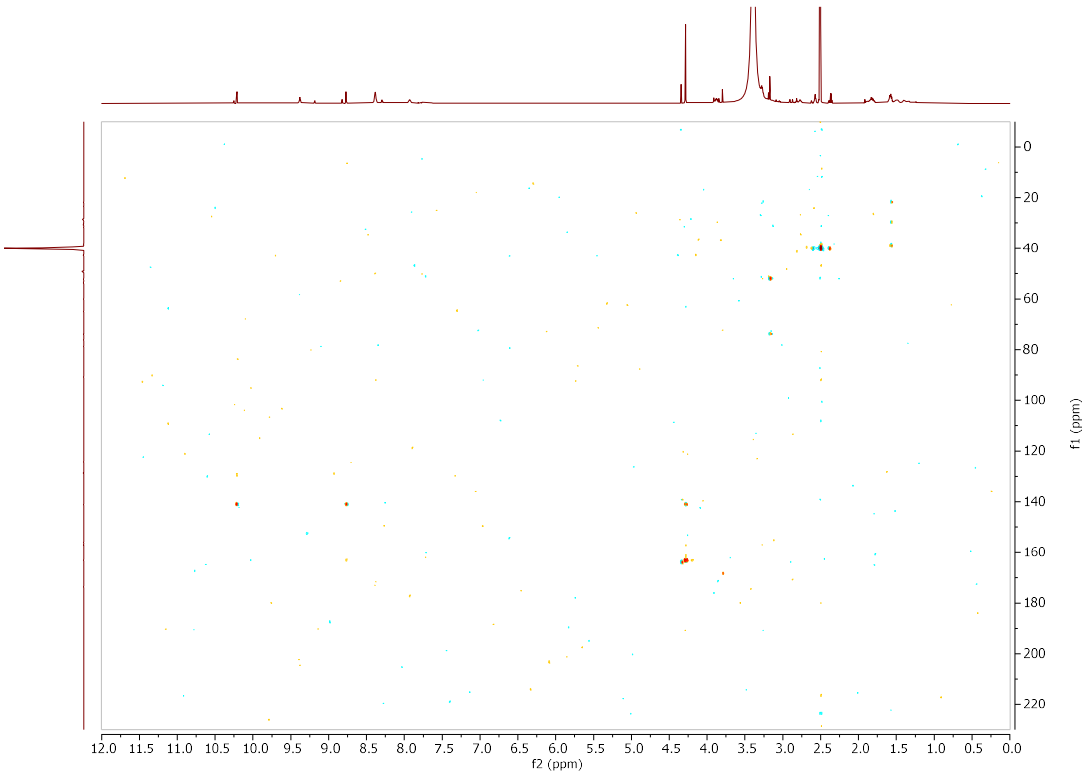
COSY compound 60



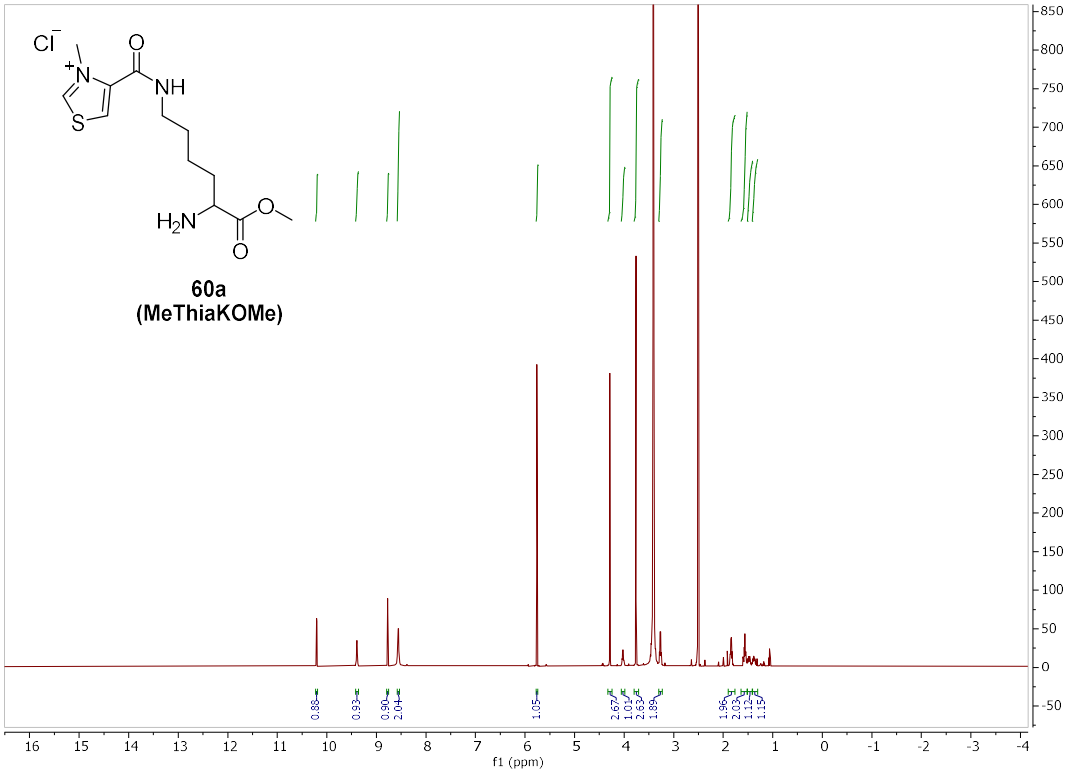
HSQC compound 60



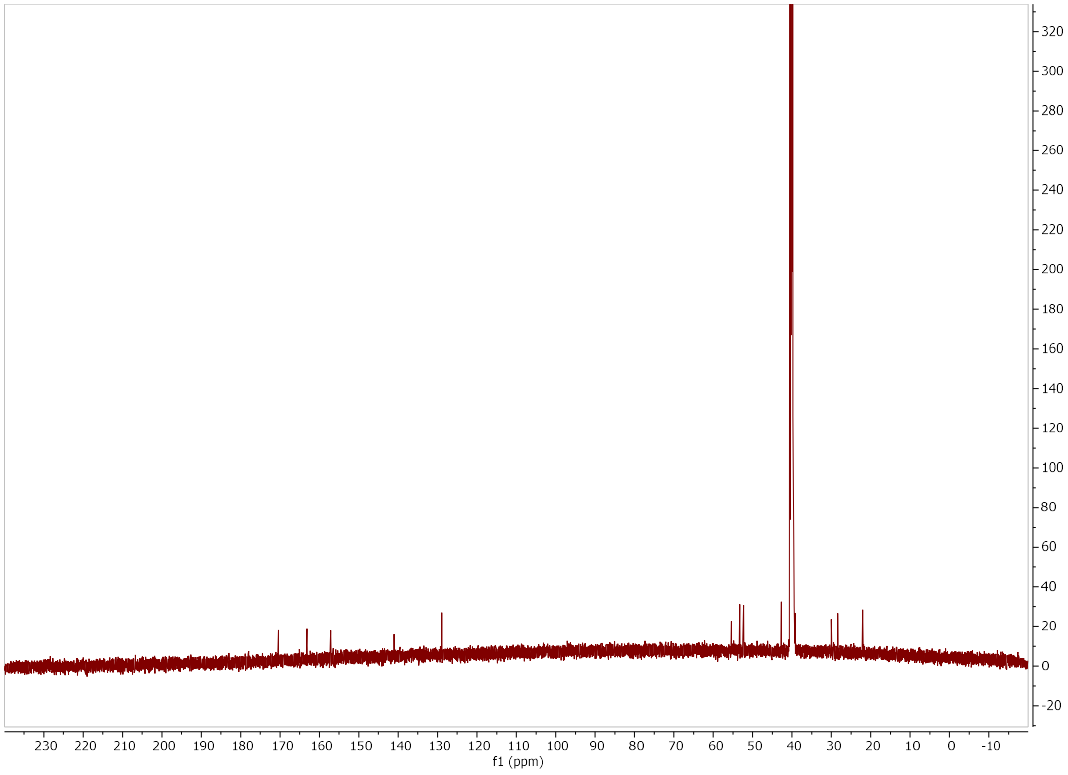
HMBC compound 60



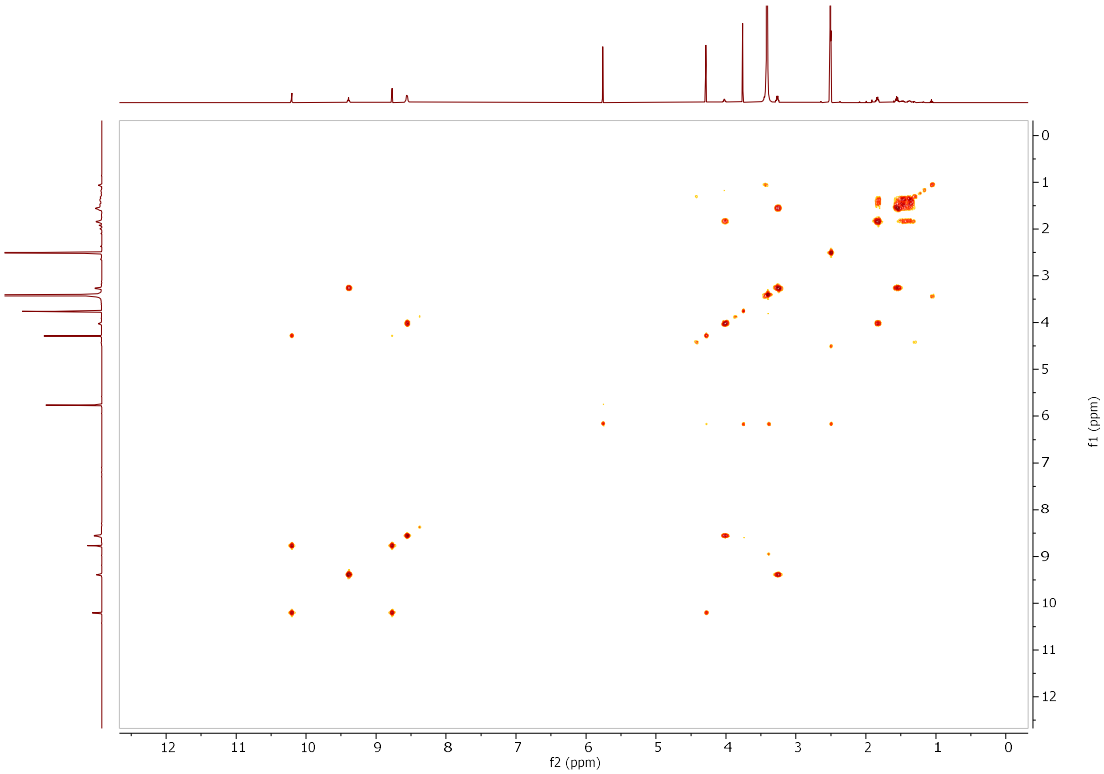
^{1}H NMR (500 MHz, DMSO-d_6) Compound 60a



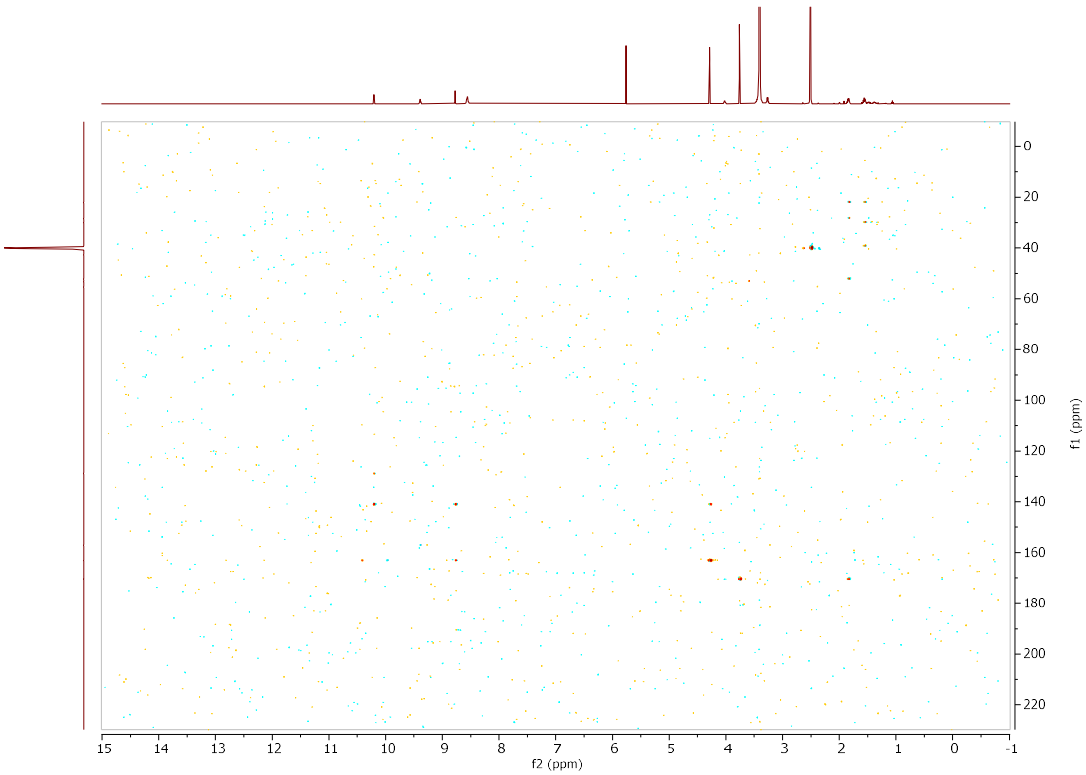
^{13}C NMR (500 MHz, DMSO-d_6) Compound 60a



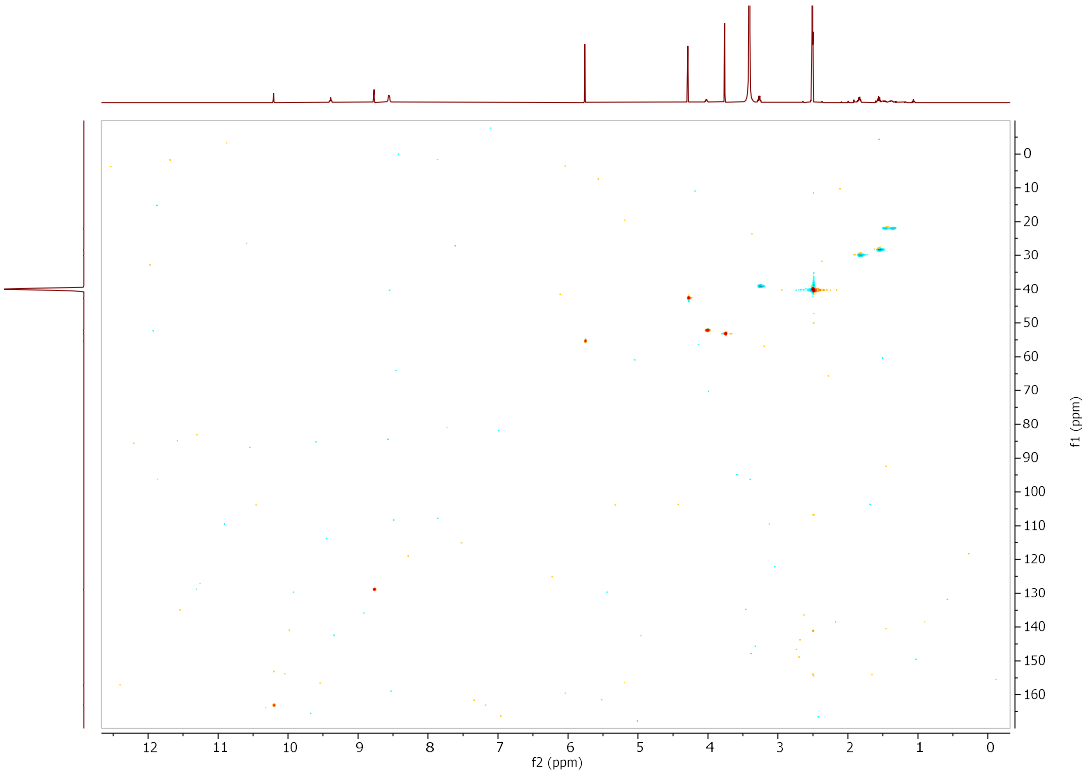
COSY Compound 60a



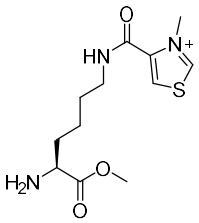
HMBC Compound 60a



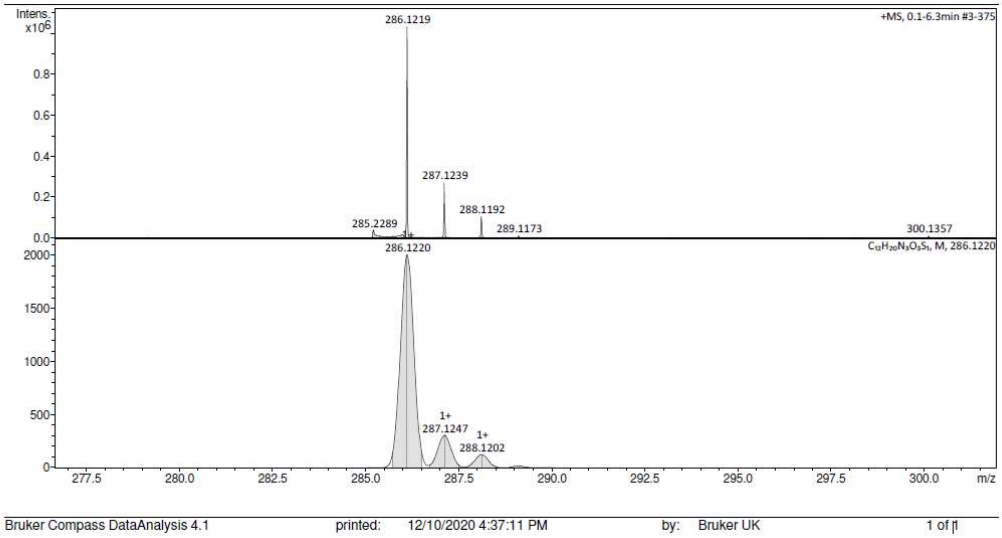
HSQC Compound 60a



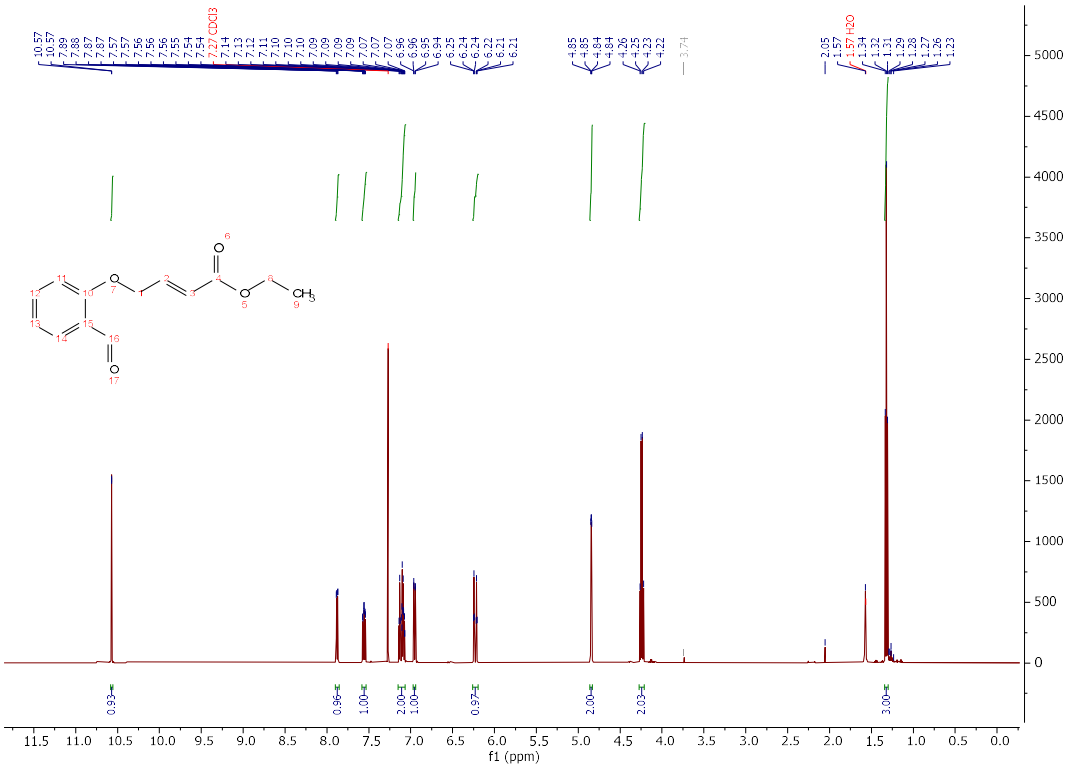
LCMS Compound 60a

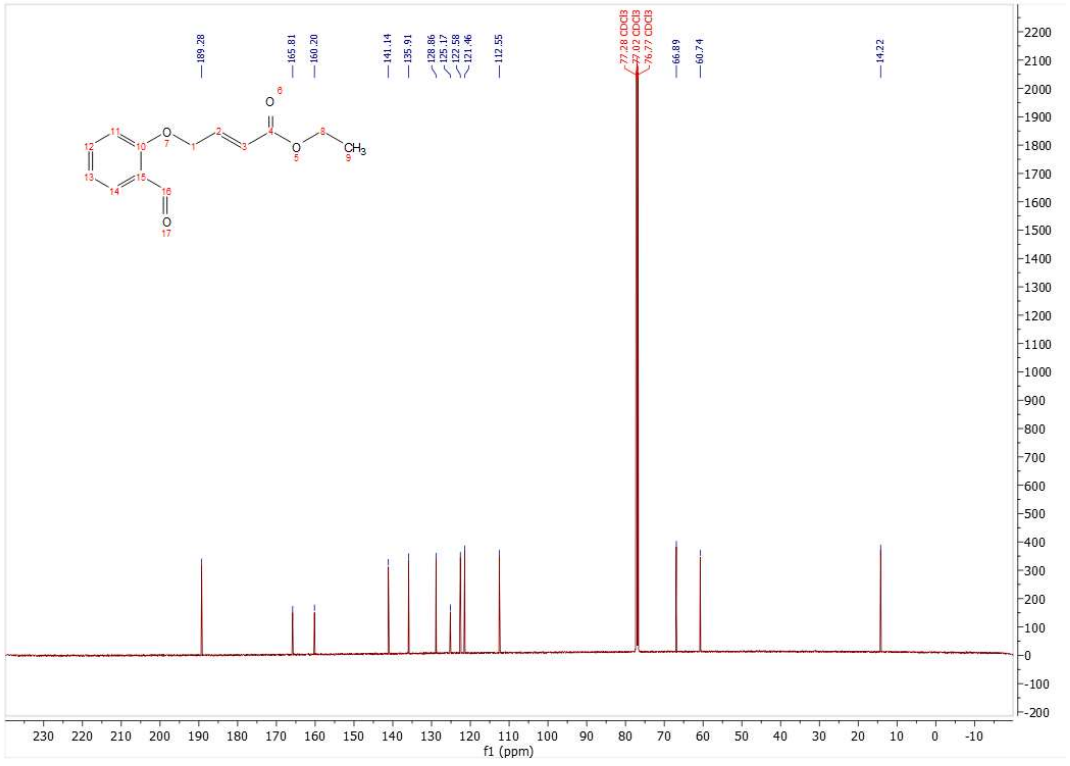


Chemical Formula: $C_{12}H_{20}N_3O_3S^+$
m/z: 286.12 (100.0%), 287.13 (13.0%), 288.12 (4.5%), 287.12 (1.1%)

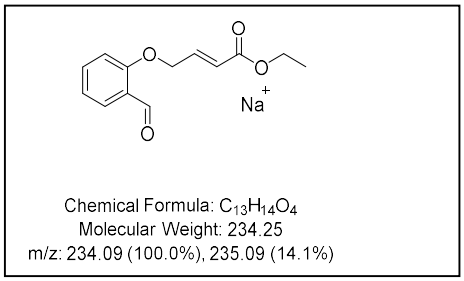


1.6.15 Starting material 18

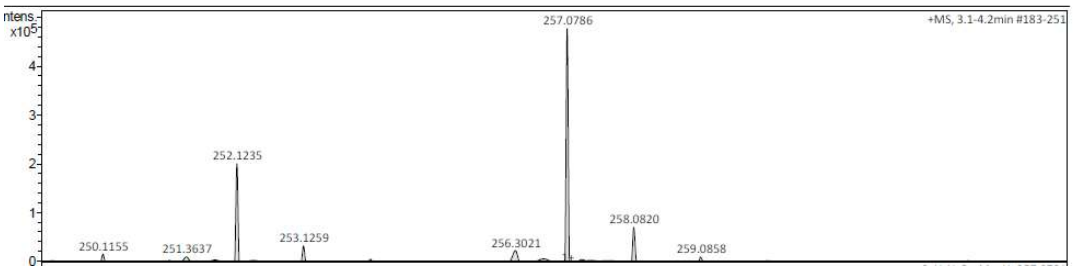




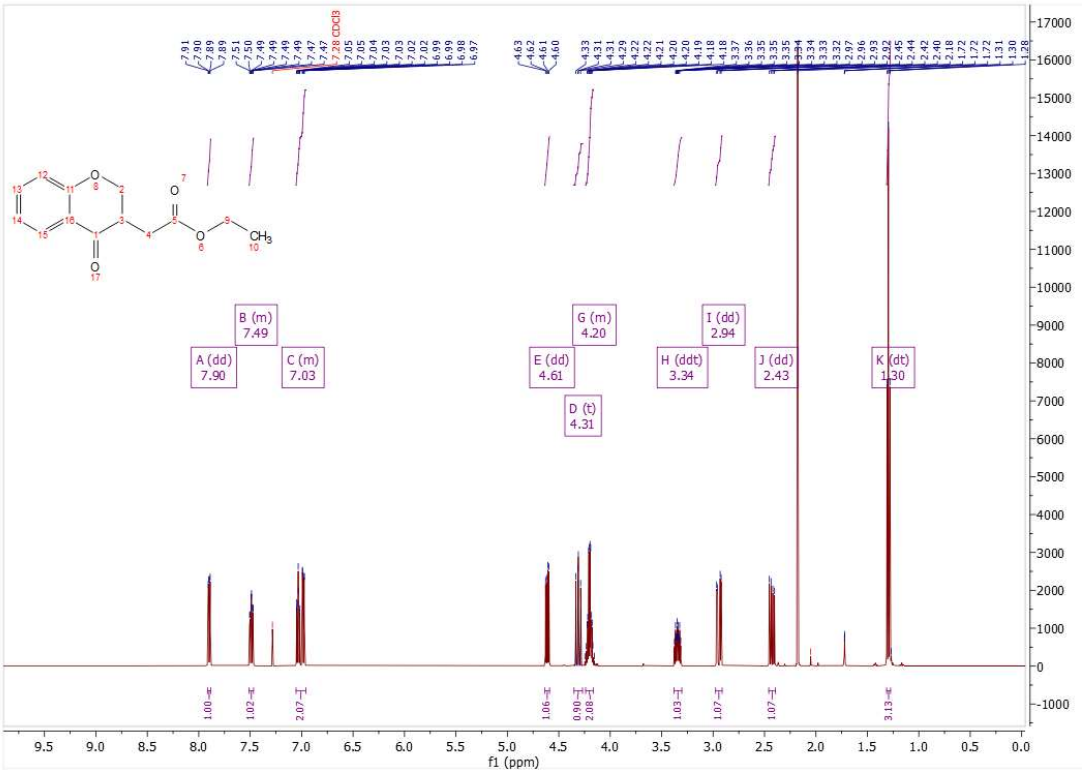
LCMS compound 18

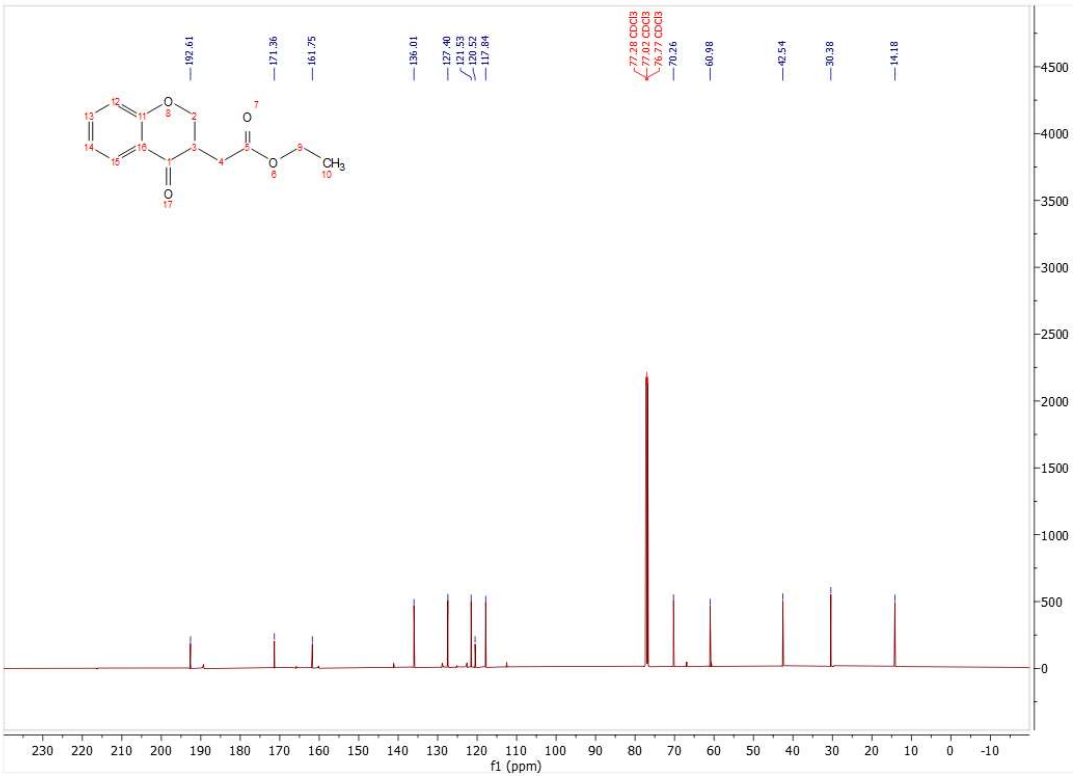


Chemical Formula: $C_{13}H_{14}O_4$
Molecular Weight: 234.25
m/z: 234.09 (100.0%), 235.09 (14.1%)



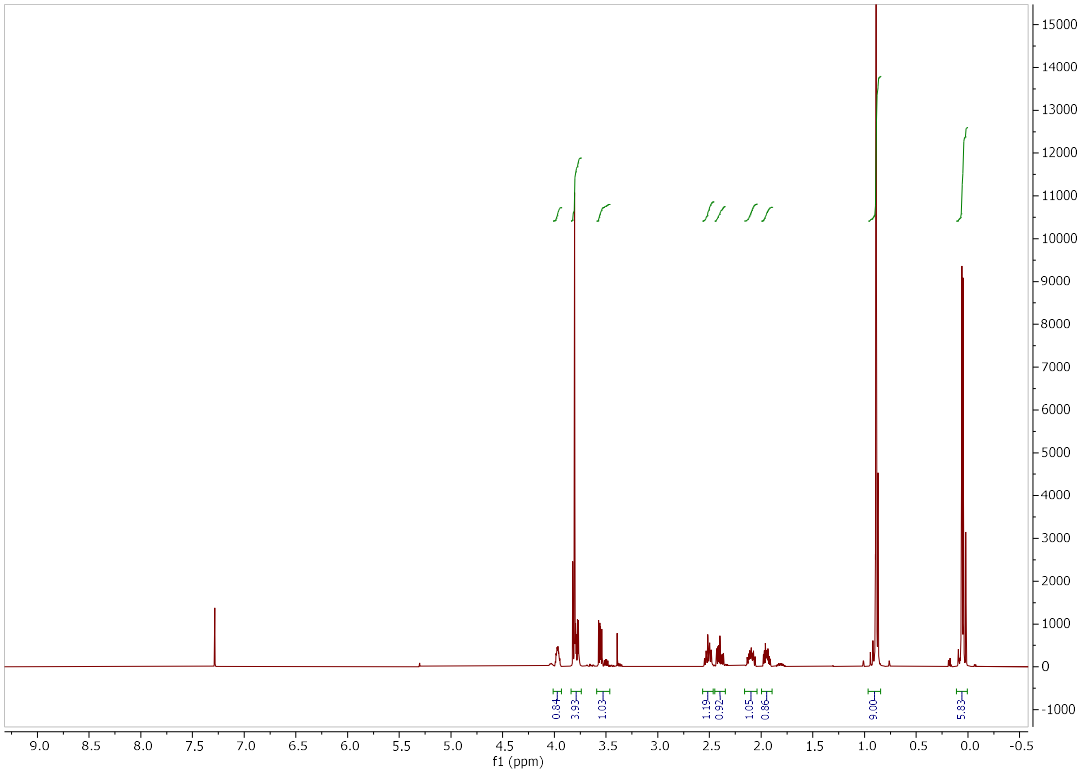
1.6.16 Product 19





1.6.17 Compound 42

^{1}H NMR (500 MHz, Chloroform-*d*) ,Intermediate 41



^{1}H NMR (500 MHz, Chloroform-*d*) Compound 42