Selective *N*-terminal Protein Bioconjugation Enabled by Conjugate Addition/ Ring Expansion (CARE) Cascades

Owen Rhys Hughes

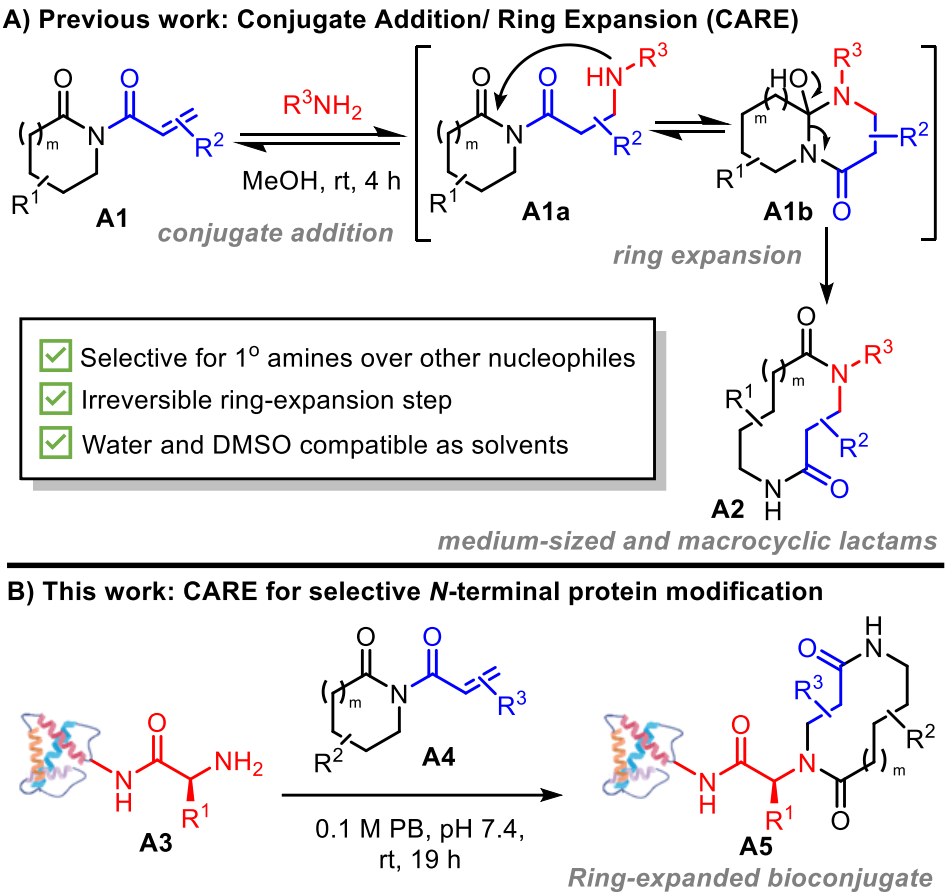
Doctor of Philosophy

University of York
Department of Chemistry
November 2025

Abstract

Protein bioconjugation has revolutionised modern day biomedical and biotechnological research by enabling proteins to be engineered with properties that greatly surpass the limitations of the 20 canonical amino acids. Site-selective approaches are vital for preserving protein function and producing homogenous bioconjugates, both of which facilitate regulatory approval and improve reproducibility. In this context, protein *N*-termini have emerged as attractive sites for chemical modification owing to their distinct structure and accessibility. Despite recent advances, *N*-terminal selective bioconjugation remains limited by the need to balance chemo- and regioselectivity with conjugate stability and reagent accessibility. Even the most advanced reagents are not universally compatible with all proteins, highlighting the need for further innovation in this field to unlock its full potential.

Herein, this thesis describes the development of a novel, site-selective *N*-terminal protein bioconjugation strategy using conjugate addition/ ring expansion (CARE) cascade reactions. Building on the small molecule ring expansion strategy reported by the Unsworth group, we conceptualised, modified, and optimised an approach for *N*-terminal bioconjugation *via* ring expansion with cyclic acryloyl imides. The advantages of *N*-terminal CARE bioconjugation include high selectivity for *N*-termini over lysine ε-amino groups, the formation of stable macrocyclic bioconjugates *via* an irreversible ring expansion, and the potential for further derivatisation through pendant bioorthogonal tags on CARE-assembled bioconjugates. We prove the utility of this approach by modifying a series of medicinally relevant proteins, including nanobodies used for prostate cancer cell and tissue imaging, and human chemokines implicated in the human immune response.


A) Previous work: Conjugate Addition/ Ring Expansion (CARE)
R3NH2
MeOH, rt, 4 h
conjugate addition
A1
A1a
A1b
ring expansion
Selective for 1° amines over other nucleophiles
Irreversible ring-expansion step
Water and DMSO compatible as solvents
A2
medium-sized and macrocyclic lactams
B) This work: CARE for selective N-terminal protein modification
A3
A4
0.1 M PB, pH 7.4,
rt, 19 h
A5
Ring-expanded bioconjugate

Author's Declaration

I, Owen Rhys Hughes, declare that this Thesis is a presentation of original work. I am the sole author of this Thesis. The work presented in this Thesis was carried out at the University of York between October 2022 and August 2025.

This work has not previously been presented for an award at this, or any other, University. All sources are appropriately acknowledged in the main text or in the references listed in Chapter 8. The work submitted for this Thesis is my own, except where it forms part of jointly authored publications and pre-print articles, in which case the contributions of others are clearly noted in the footnotes.

Some of the work described in this Thesis is reported in a pre-print article, currently under review:

O. R. Hughes, A. Tufail, E. Hutton, J. Nabarro, R. Howarth, N. Yates, A. Whitwood, C. Robson, N. Signoret, M. A. Fascione, C. D. Spicer, W. P. Unsworth, *ChemRxiv*, 2025, doi:10.26434/chemrxiv-2025-06817.

Acknowledgements

Firstly, I would like to thank my supervisors Will Unsworth, Chris Spicer and Martin Fascione for giving me the opportunity to undertake this PhD within their groups. Throughout my time in York, their advice, support, and guidance have been invaluable and truly inspiring. I was often reminded by others of how difficult it must be having three supervisors, but I genuinely could not have asked for a better trio! It has been a pleasure working with them, and I cannot thank them enough for everything they have done for me.

I would also like to thank all the technical staff within the department, particularly Karl for his patience with my never-ending mass spec samples, Heather and Matt for their support in NMR, and Adrian for his assistance with X-ray crystallography. A special shout out goes to Lee Duff, our lab technician, who must be the most patient man in the world with the never-ending list of jobs we kept sending his way. I would also like to extend a special thank you to Ed Bergström for answering my many questions about the chromatography facilities and for saving many of my precious samples whenever the instruments encountered issues.

Thank you to all of the WPU group for brightening up some difficult days in the lab and for keeping me constantly entertained during long days at the fume hood, particularly with our impromptu games of Among Us, or with Ben K's and my favourite game of 'Do you know this 2010s pop culture moment?' — even if you never knew any of them… My proudest moments throughout my PhD include topping the tennis ranking table, being the best imposter, and always having the cleanest fume hood. I'd like to give special thanks to Will O, who has put up with me since day one. You have always been an amazing friend and have been part of all my favourite memories in York. Our many drunken nights together and trips to Old Trafford have been real highlights and memories I will always remember — especially the 4–3 FA Cup win against Liverpool! A huge thanks also goes to my desk neighbours, Jack W, Selin, and Lachlan, for putting up with my huffing and puffing, and for tolerating me slamming my head on the desk after a bad result!

I would also like to thank both the Fascione and Spicer groups for supporting me as I navigated the unfamiliar world of chemical biology. Their knowledge and patience have been incredible, and I am extremely grateful for all their help.

To Corinna, you are truly a one-of-a-kind person. Since meeting you on the very first day of undergraduate Chemistry in Liverpool, I have always considered you to be one of the most special people in my life. I am eternally grateful for all the advice and support you have given

me throughout this PhD, and I will always be grateful that you are my biggest chemistry cheerleader! I can't wait to keep making a lifetime of memories with you and getting ourselves into trouble with whoever crosses us when we're up to mischief! I don't have the words to begin thanking you for everything you have done for me, but I hope you know just how much it means.

To Mads, Charl, Hannah, and Katie T — my longest and best-est friends. I could not have got through this PhD without each and every one of you. You have always been there for me to lean on during tough times, and for that, I am truly grateful. You all have a way of cheering me up like no one else, and I count myself as the luckiest person in the world to have such incredible friends. Here's to a lifetime of fun together, always.

To Rob, my incredible partner throughout all of this: you have brightened up my life since coming into it just 18 months ago, and you have been the most encouraging and supportive person I could have ever wished for.

Finally, I would like to thank my amazing family. Life has not always been easy or kind to us, but we have always found a way to get through it with a smile. To my incredible mum and dad, I love you both so much. Your constant love, support and encouragement have been a lifesaver during some difficult times, and I cannot begin to thank you enough for all that you have done. You have sacrificed so much to keep Kieran, Niamh, and I happy and healthy, and I can only hope to one day be as selfless as you both have been for us. You have given me the best start in life, and I owe all of this to you. This Thesis is dedicated to you both — for without you, it would not have been at all possible.

Abbreviations

2-EBA	2-Ethynylbenzaldehydes
2-PCA	2-Pyridinecarboxaldehyde
AA	Amino acid
ABPP	Activity-based protein profiling
Ac	Acetyl
AcOH	Acetic acid
ADC	Antibody drug conjugate
AF	Alexa Fluor™
AIBN	Azobisisobutyronitrile
AOP	Area of peak
Ar	Aryl
Arg	Arginine
BCArg	*Bacillus caldovelox* arginase
BPC	Base peak chromatogram
BPH	Benign prostatic hyperplasia
Bs	Brosyl
BSA	Bovine serum albumin
Bu	Butyl
CARE	Conjugate addition/ ring expansion
CBT	2-Cyanobenzothiazole
COSY	Correlation spectroscopy
CPO	Cyclopropenone
CRE	Cyclisation ring expansion
CuAAC	Copper catalysed azide-alkyne cycloaddition
Cys	Cysteine
DAPI	4'-6-Diamidino-2-phenylindole
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCM	Dichloromethane
DEPT	Distortionless enhancement by polarisation transfer
DIPEA	*N*,*N*'-Diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide

DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
Eq	Equivalents
ESI	Electrospray ionisation
Et	Ethyl
FCC	Flash column chromatography
FDA	U.S. Food and Drug Administration
Fmoc	Fluorenylmethyloxycarbonyl
GRK	G-protein coupled receptor kinases
GSH	Glutathione
h	Hour(s)
HASP	Hydrophilic acylated surface protein
HCTU	2-(6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HEPES	*N*-(2-hydroxyethyl)piperazine-*N*'-(2-ethanesulfonic acid)
HIV	Human immunodeficiency virus
HMBC	Heteronuclear multiple bond correlation
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
HSAB	Hard and soft (Lewis) acids and bases
HSQC	Heteronuclear single quantum coherance
HWE	Horner-Wadsworth-Emmons
IEDDA	Inverse electron-demand Diels-Alder
IL-8	Interleukin-8
LC	Liquid chromatography
LC-MS	Liquid chromatography-mass spectrometry
LHMDS	Lithium bis(trimethylsilyl)amide
LNCaP	Lymph node carcinoma of the prostate
M.P.	Melting point
MALDI-ISD	Matrix-assisted laser desorption ionisation in-source decay
Me	Methyl
MetAP	Methionine aminopeptidases

min	Minutes
Mol	Molar
MS	Mass spectrometry
MWCO	Molecular weight cut-off
n-BuLi	n-Butyllithium
NHS	*N*-hydroxysuccinimide
NMR	Nuclear magnetic resonance
Ns	Nosyl
OPAL	Organocatalyst-mediated cross aldol ligation
PB	Phosphate buffer
PBS	Phosphate-buffered saline
PCa	Prostate cancer
PEG	Polyethyleneglycol
Ph	Phenyl
PSMA	Prostate-specific membrane antigen
RNase A	Ribonuclease A
rt	Room temperature
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Ser	Serine
SPAAC	Strain-promoted azide-alkyne cycloaddition
SPANC	Strain-promoted alkyne-nitrone cycloaddition
SPPS	Solid phase peptide synthesis
SuRE	Successive ring expansion
T3P	Propylphosphonic anhydride
TA4C	Triazolecarbaldehydes
TBS	*tert*-Butyldimethylsilyl
Tf	Triflyl
TFA	Trifluoroacetic acid
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
Thr	Threonine
TIPS	Triisopropylsilane

TLC	Thin layer chromatography
TMEDA	Tetramethylethylenediamine
TMS	Trimethylsilyl
Tris	2-Amino-2-(hydroxymethyl)propane-1,3-diol
Ts	Tosyl
TsOH	*p*-Toluenesulfonic acid
UV	Ultraviolet
XRD	X-ray diffraction

List of Tables

List of Figures

Chapter 1

Chapter 2

11226; [**5.03**+H]$^+$ found 11418, calculated 11407.* Nb In all experiments using CjX183-D and mutants, the observed mass was ~11 Da higher than the theoretical mass and thus was taken as the unmodified protein mass in all instances. The reasons for this observation are unknown but are consistent with previous reports.[63] Experimental details can be found in Chapter 7.5.2.1

Figure 5.02: (A) CjX183 R51K bioconjugation with CARE using imide **5.02**. (B) LC-MS data obtained from the crude reaction mixture. **MS** (ESI$^+$): [**5.04**+H]$^+$ found 11209, calculated 11198; [**5.05(s)**+H]$^+$ found 11390, calculated 11379; [**5.05(d)**+H]$^+$ found 11571, calculated 11560. Experimental details can be found in Chapter 7.5.2.1

Figure 5.03: CARE bioconjugation of model proteins with imide **5.02** at 37 °C and rt. Experimental details can be found in Chapter 7.5.2.1

Figure 5.04: Screening conditions for RNase A bioconjugation with imide **5.02**. Reactions were run at 50 μM of RNase A at pH 6.0, 6.5, 7.4 and 8.0 with 5, 10, 20 or 50 equivalents of imide **5.02** (stock adjusted accordingly in DMSO) and incubated at rt, 1000 rpm for 19 h before an aliquot was taken and analysed by LC-MS. Experimental details can be found in Chapter 7.5.2.2.

Figure 5.05: Screening conditions for cytochrome C bioconjugation with imide **5.02**. Reactions were run at 50 μM of cytochrome C at pH 6.0, 6.5, 7.4 and 8.0 with 5, 10, 20 or 50 equivalents of imide **5.02** (stock adjusted accordingly in DMSO) and incubated at rt, 1000 rpm for 19 h before an aliquot was taken and analysed by LC-MS. Experimental details can be found in Chapter 7.5.2.2.

Figure 5.06: Screening conditions for RNase A bioconjugation with **5.07**. Reactions were run at 50 μM of RNase A at pH 6.0, 6.5, 7.4 and 8.0 with 5, 10, 20 or 50 equivalents of imide **5.07** (stock adjusted accordingly in DMSO) and incubated at rt, 1000 rpm for 19 h before an aliquot was taken and analysed by LC-MS. Experimental details can be found in Chapter 7.5.2.2.

Figure 5.07: Screening conditions for JVZ-007 bioconjugation with imide **5.07**. Reactions were run at 50 μM of JVZ-007 at pH 6.0, 6.5, 7.4 and 8.0 with 5, 10, 20, 50, 100 or 200 equivalents of imide **5.07** (stock adjusted accordingly in DMSO) and incubated at rt, 1000 rpm for 19 h before an aliquot was taken and analysed by LC-MS. Experimental details can be found in Chapter 7.5.2.2.

Chapter 6

Chapter 7

List of Schemes

Chapter 1

Chapter 2

Chapter 3

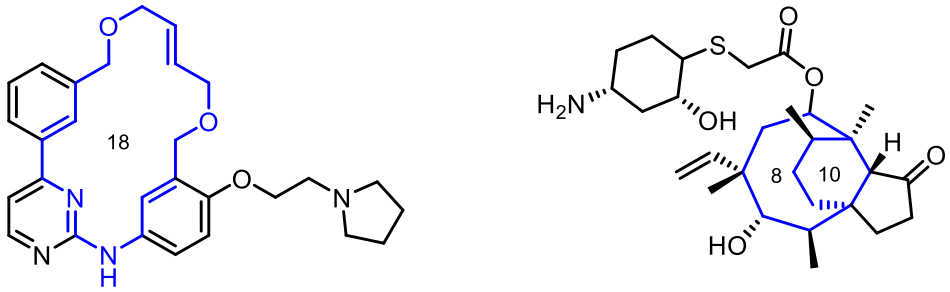
Table of Contents

Chapter 1: Introduction

The first half of the introduction Chapter of this Thesis (Sections 1.1–1.4) summarises the key structural features, chemistry, and methods to synthesise medium-sized rings and macrocycles. This includes discussion of ring-expansion methods — a major focus of this PhD project. The prevalence of medium-sized rings and macrocycles in nature and their emerging significance in therapeutic applications is also covered. Following this, Section 1.5 includes discussion on bioconjugation — focussing on methods that target protein *N*-termini — and concludes with an exploration of how our ring-expansion chemistry presents itself as a promising and versatile bioconjugation strategy.

1.1 Medium-sized rings and macrocycles

Medium-sized rings (8–11 members) and macrocycles (12+ members) are covalent cyclic molecules that are widely embedded in many naturally occurring bioactive compounds.[1] In contrast to 'normal'-sized rings (5–7 members), medium-sized rings can offer distinct structural rigidity and unique three-dimensional shapes that can enhance binding affinity to biological targets, and help to improve bioavailability and cell permeability.[2–5] Compared to medium-sized rings, macrocycles often exhibit higher conformational flexibility, allowing them to change their shapes. This feature, coupled to their relatively large surface area, helps improve their capability to interact with biological targets that have large, flat binding sites.[6,7] Cyclisation has also been shown to lower the entropic cost of receptor binding by removing some degrees of conformational freedom, while removing charged termini can result in molecules with greater cell permeability, compared to their corresponding linear analogues.[8,9] Both medium-sized rings and macrocycles have been explored in fields such as functional materials,[10] photovoltaics,[11] and polymer science,[12] while recently there has been a significant increase in research focussed on their application in therapeutics.[13–15] Despite their abundance in nature, the discovery of novel therapeutics containing medium-sized rings and macrocycles has proven challenging. Indeed, a recent review reports that, since the 20th century, only 60 macrocyclic drugs have been approved by the U.S. Food and Drug Administration (FDA) (**1.000** and **1.001**, Figure 1.01).[16]



1.000
Vonjo® (pacritinib)

1.001
Xenleta® (lefamulin)

Figure 1.01: The structures of recently FDA-approved macrocyclic and medium-sized ring therapeutics.

1.2 Synthesising medium-rings and macrocycles by macrocyclisation

Traditionally, these cyclic systems are assembled *via* end-to-end macrocyclisation, in which reactive groups at the ends of a linear precursor react intramolecularly through a cyclic transition state containing the same number of atoms as the resulting macrocycle. However, favouring the intramolecular macrocyclisation (**1.002** → **1.003**, Figure 1.01) over competing intermolecular reactions, such as dimerisation or oligomerisation (**1.002** → **1.004**, Figure 1.02) can be very challenging.[17,18] Macrocyclisation reactions are hindered by the thermodynamic barrier associated with the entropically unfavourable nature of cyclisation. This entropic penalty is comparatively less severe for medium-sized rings, but their synthesis remains equally (or more) challenging due to the destabilising transannular interactions often encountered in medium-sized rings.[19,20]

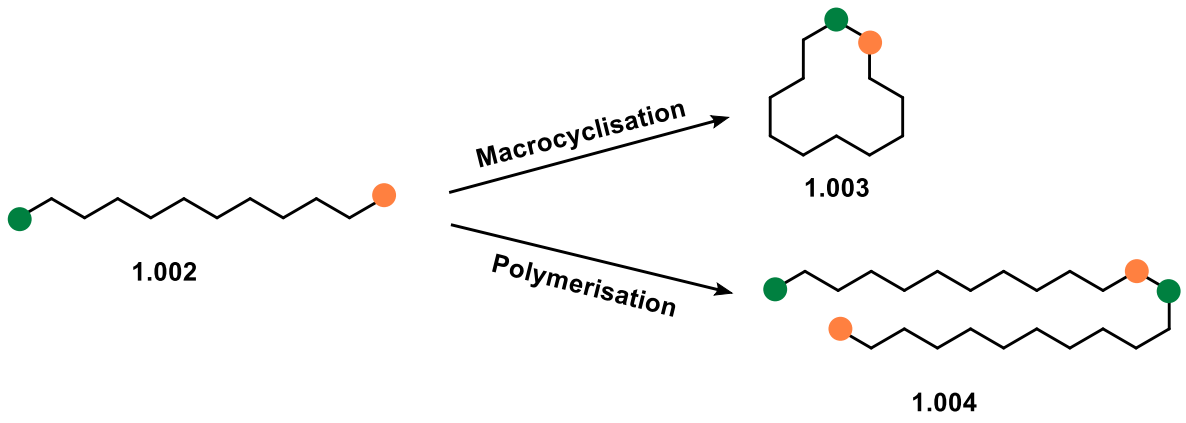
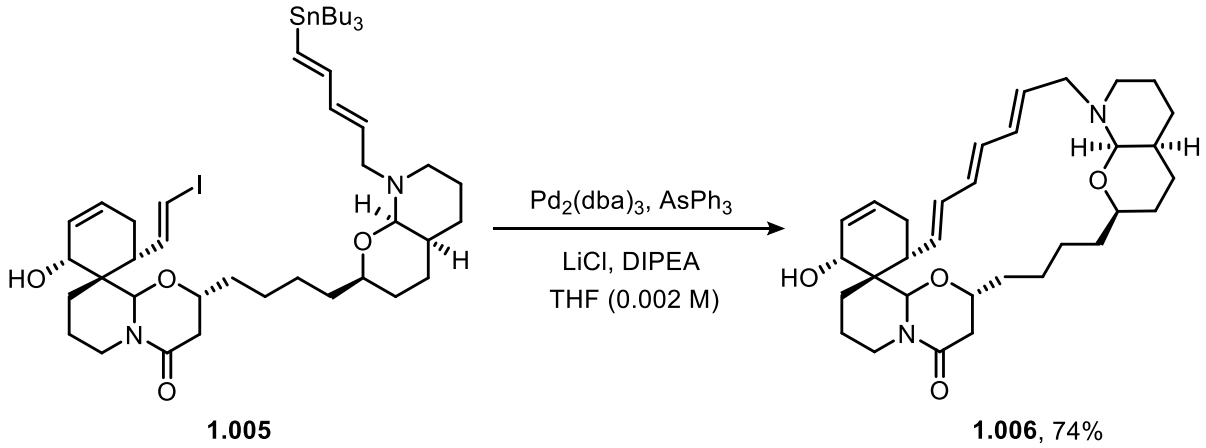


Figure 1.02: End-to-end macrocyclisation *vs*. competing dimerisation and oligomerisation reactions.

The success of macrocyclisation is highly dependent on the type of ring-closing reaction used. In some cases, careful optimisation of the reaction conditions can promote the intramolecular pathway sufficiently well that end-to-end cyclisation methods are a viable approach. For

example, in the total synthesis of the reported structure of ‘upenamide (**1.006**, Scheme 1.01) by Taylor *et al.*, a Stille cross-coupling reaction served as the key ring-closing step. Conducted under high-dilution conditions (~0.002 M), the reaction proceeded efficiently to afford the product in 74% yield, with minimal disruption from competing intermolecular coupling.[21]

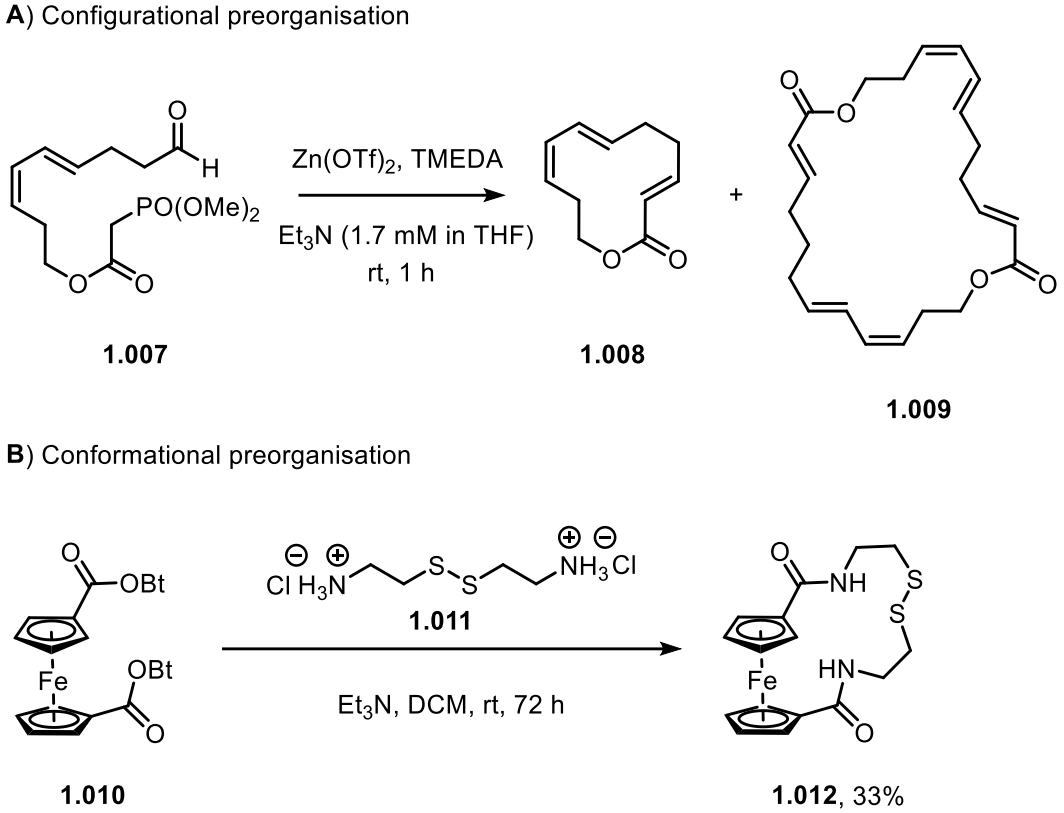


1.005 **1.006**, 74%

Scheme 1.01: Stille cross-coupling under high dilution conditions towards the synthesis of ‘upenamide.

However, high dilution conditions are impractical, particularly at large scale, and economically inefficient. Alternatively, pseudo-high-dilution techniques — including slow reagent addition,[22] the use of solid supports,[23,24] or biphasic solvent systems[25] — can be used to promote large ring cyclisation reactions under more practical conditions. Moreover, the careful design of linear precursors plays a crucial role in determining the outcome of a macrocyclisation reaction. Two common strategies used to promote intramolecular cyclisation are configurational and conformational preorganisation.[26] In configurational preorganisation, the structural features of the linear precursor are carefully chosen to help position the reactive groups closer together in space. An instructive example is shown in Scheme 1.02A, where a Horner-Wadsworth-Emmons (HWE) ring-closing reaction is employed for the synthesis of macrocycles **1.008** and **1.009**. The success of this transformation is critically dependent on the stereochemistry of diene precursor **1.007**, which helps to pre-organise the linear molecule so that the reactive aldehyde and phosphonate ester groups are brought into close proximity. The reaction conditions can be tuned to favour the formation of either the monomeric or dimeric macrocyclic products.[27] In conformational preorganisation, specific ligands or linkers are introduced to influence the molecule’s overall shape, effectively ‘folding’ it into a conformation whereby the reactive centres are brought into close proximity. For example, suitably disubstituted ferrocenes can be employed to assemble parallel chains in a pre-

organised folded conformation, thereby positioning the reactive chain-end groups in close proximity and favouring macrocyclisation (Scheme 1.01B).[28]



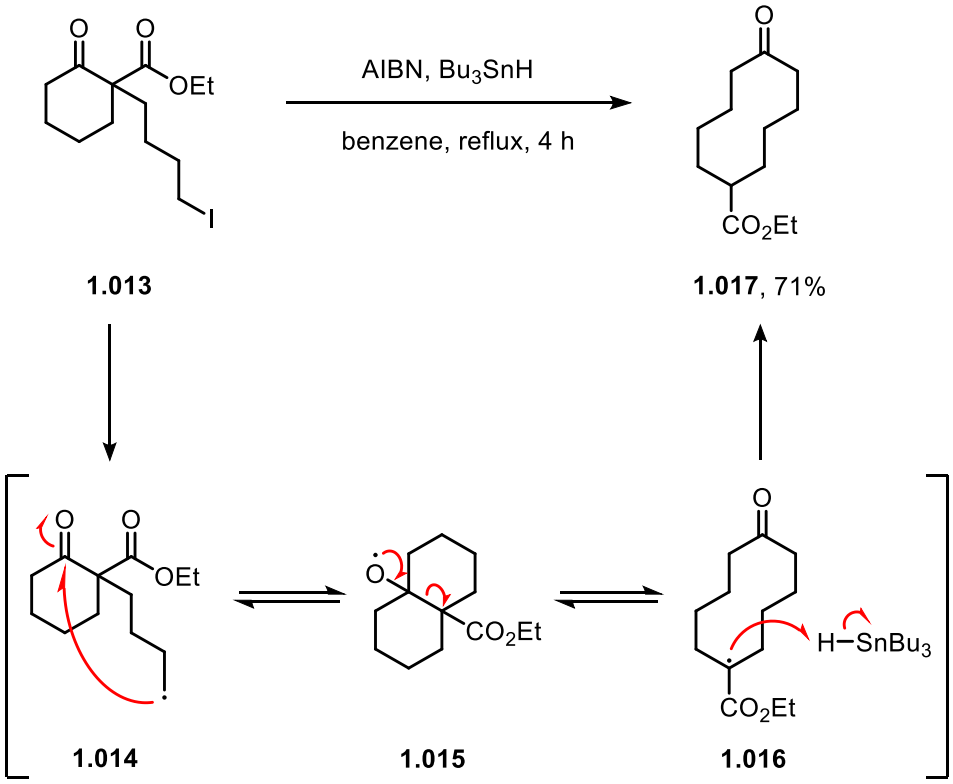
Scheme 1.02: Configurational and conformational preorganisation in macrocyclisation reactions.

1.3 Ring expansion reactions

Ring expansion strategies have distinct advantages over the macrocyclisation reactions described above, most pertinently that they can enable direct end-to-end macrocyclisation reactions to be completely avoided. Ring expansion can enable the synthesis of medium-sized rings or macrocycles from pre-assembled smaller rings *via* rearrangement reactions, with the rearrangement reactions often proceeding *via* 'normal'-sized ring transition states. By avoiding the formation of higher-energy medium- and macrocyclic transition states, ring expansion can be considerably more thermodynamically favourable. Ring expansion methods can enable more modular and efficient access to medium-sized rings and macrocycles that were previously difficult to obtain with traditional synthetic methods. Generally, ring expansion methods can be classified into three main classes: fragmentation, radical-mediated, and pericyclic approaches, each of which are discussed in the following sections.[29]

1.3.1 Radical-mediated ring expansion

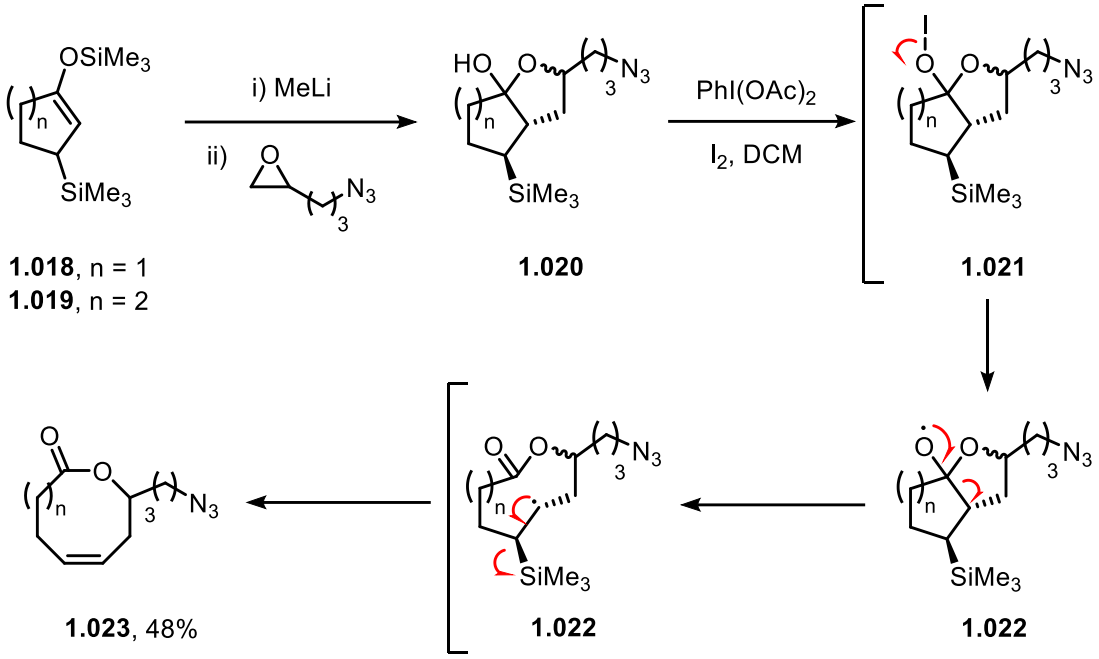
The Dowd-Beckwith carbon-centred radical ring expansion is arguably the most widely recognised radical-mediated ring expansion method (Scheme 1.03). The reaction begins with the formation of an alkyl radical **1.014** on a pendant carbon chain, which then undergoes an intramolecular addition to an electron-deficient ketone, generating an alkoxy radical (**1.015**). The subsequent β-scission of the unstable alkoxy radical is the key step, producing the ring-expanded intermediate **1.016** with a more stabilised alkyl radical. This radical is then quenched by hydrogen atom abstraction from tributyltin hydride, affording the desired product **1.017** in good yield.



Scheme 1.03: Dowd-Beckwith radical ring expansion.

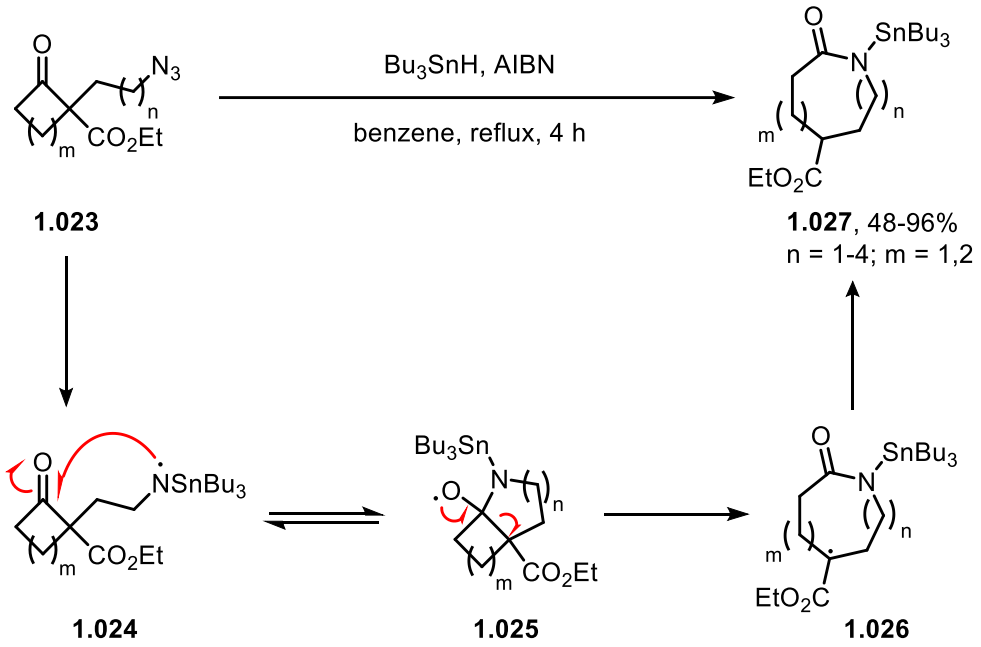
Radical-mediated ring expansions can also be initiated by heteroatom-centred radicals, such as *O*- and *N*-centred radicals. Posner *et al.* reported an oxidative radical fragmentation strategy for the synthesis of ring-expanded lactones (Scheme 1.04). The transformation begins with the lithiation of 5- and 6-membered silyl enol ethers (**1.018**/**1.019**), which are then reacted with epoxides containing masked *O*- or *N*-nucleophiles. The resulting hemiketal (**1.020**) undergoes the oxidative radical fragmentation *via* an alkoxy-radical intermediate (**1.022**), fragmenting in a similar way to the Dowd-Beckwith reaction. Only the *cis*-olefin product is obtained from this

reaction, which is likely due to the effective orbital overlap required for the elimination of a trimethylsilyl radical (**1.022 → 1.023**).[30, 31]



Scheme 1.04: Posner *et al.*'s oxidative radical fragmentation ring expansion.

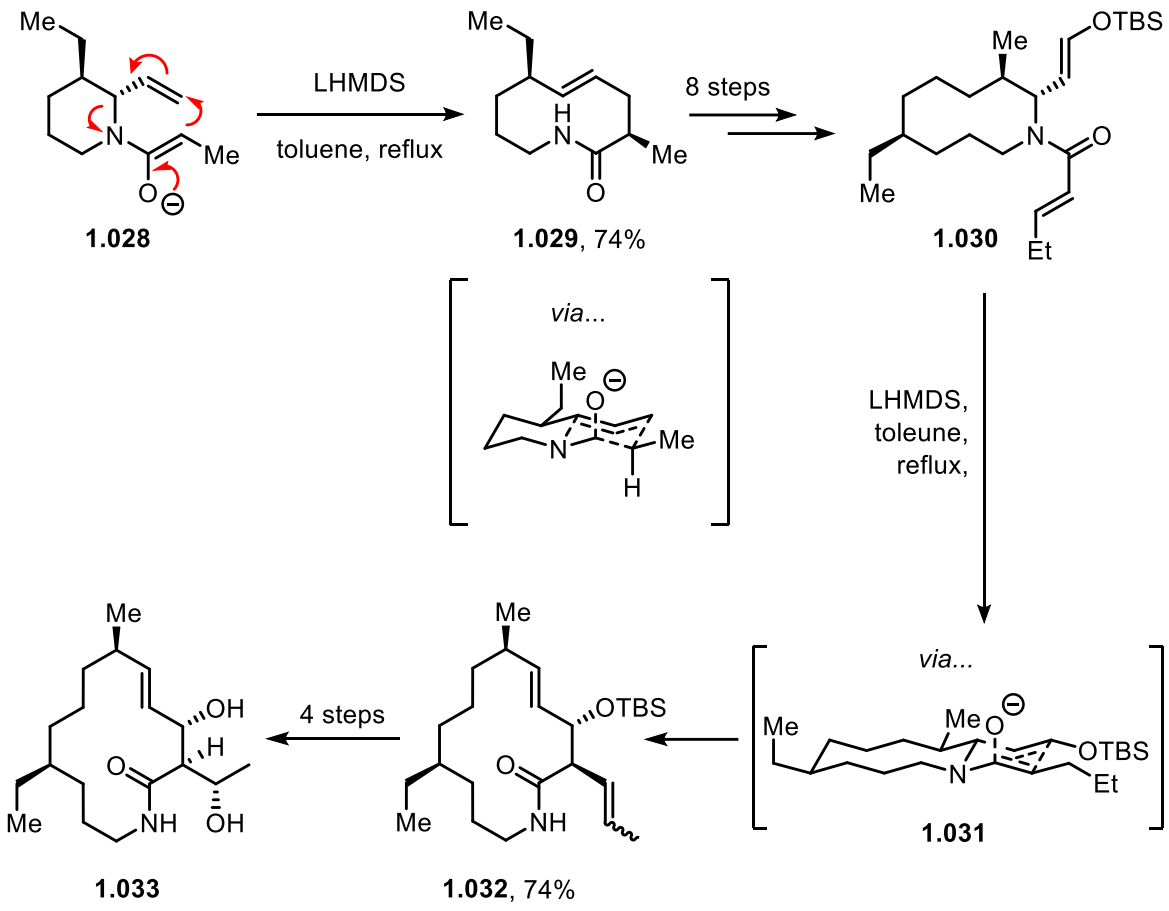
N-centred radicals can also be employed in a similar manner for the synthesis of lactams. Kim *et al.* pioneered the work in this area by developing a novel method for generating *N*-centred radicals in the presence of ketones (**1.023 → 1.024**, Scheme 1.05). Again, the reaction proceeds in a similar way to the Dowd-Beckwith reaction and is driven by the formation of an ester-stabilised radical (**1.026**).[32]



Scheme 1.05: Kim *et al.*'s *N*-centred radical ring expansion.

1.3.2 Pericyclic ring expansion

Pericyclic ring expansions are perhaps the most underutilised class of ring expansion reaction in synthesis. Their application is often limited by the need for a clear thermodynamic driving force to drive the equilibrium towards product formation. Such driving forces may arise from the relief of ring strain, the formation of stronger bonds, or ground-state destabilisation. The most commonly reported pericyclic ring expansions proceed *via* [3,3]-sigmatropic rearrangements. An example is shown in Scheme 1.06, where two aza-Claisen rearrangements were used to expand the 6-membered starting material (**1.028**) into the key 14-membered lactam intermediate (**1.032**) in the total synthesis of Fluvirucine A_2 (**1.033**).[33]



Scheme 1.06: Iterative aza-Claisen pericyclic ring expansions towards the synthesis of Fluvirucine A_2.

Other pericyclic ring expansions have been reported in the literature, most of which proceed *via* [3,3]-sigmatropic rearrangements, including 3-aza-Cope[34] and anionic oxy-Cope rearrangements.[35] However, other types of sigmatropic rearrangement have also been employed for ring expansion, although they are much less common. Examples include ring expansions using [1,4]- and [2,3]-sigmatropic rearrangements.[36,37]

1.3.3 Fragmentation ring expansions

Fragmentation reactions, the simplest and most thoroughly studied class of ring expansion, are the most relevant to this project. These reactions involve the cleavage of a bridging bond in a fused bicyclic system, resulting in a ring-expanded product. The bicyclic system may either be present in the starting material (**1.034**) or formed *in situ* through a side-chain insertion approach (**1.037**). A 'direct' fragmentation approach typically involves the elimination of a leaving group, meaning these ring expansion reactions are usually irreversible (Figure 1.03A). Alternatively, a side-chain insertion approach involves the reaction of a pendant nucleophilic group with an electrophile within the preassembled ring, typically forming new C-N, C-O, or

C-C bonds and relying on the formation of a more stable product to drive the equilibrium towards ring expansion (Figure 1.03B).

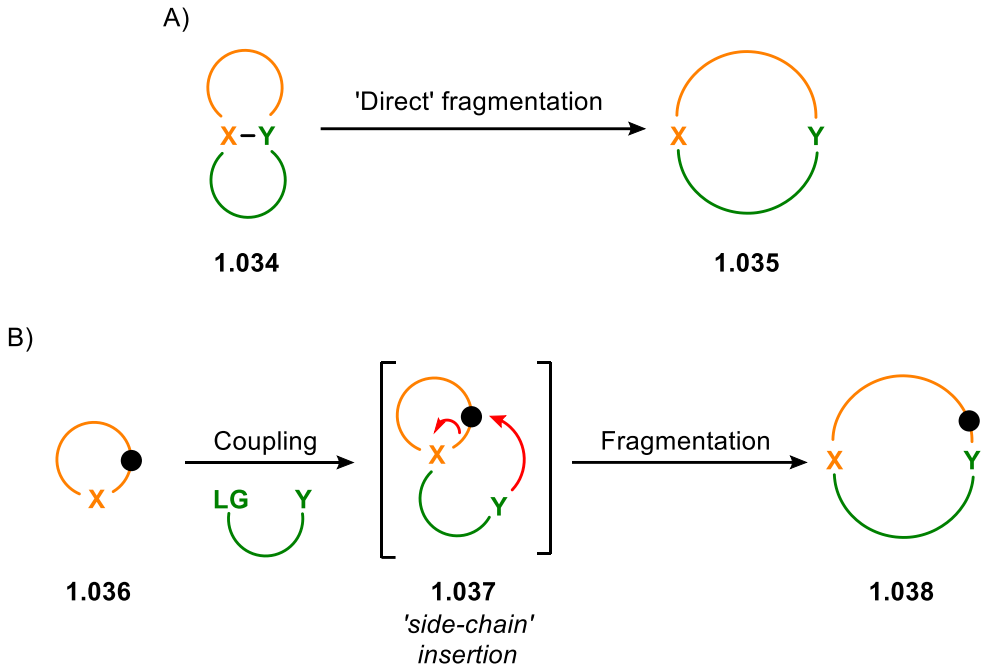
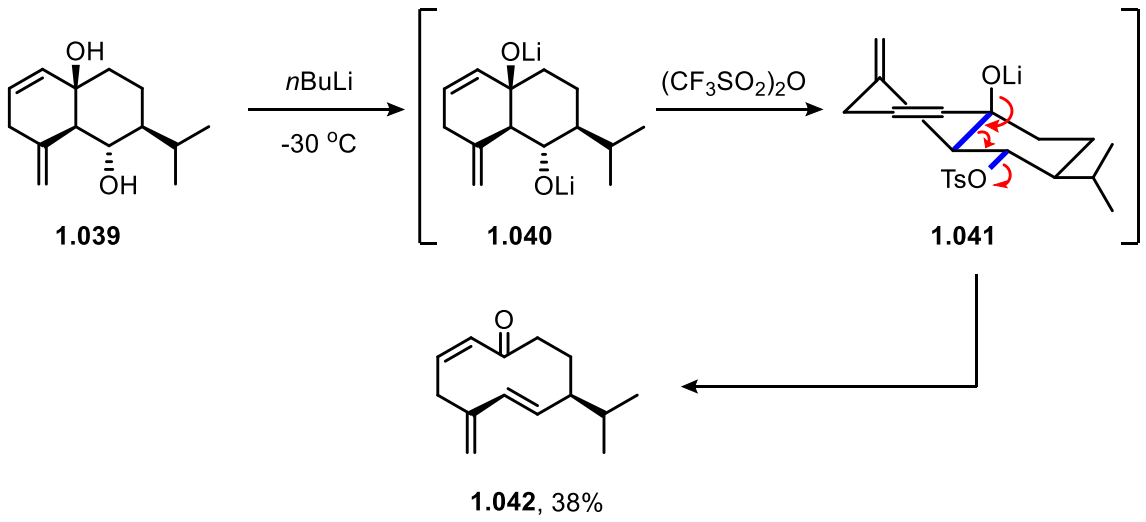


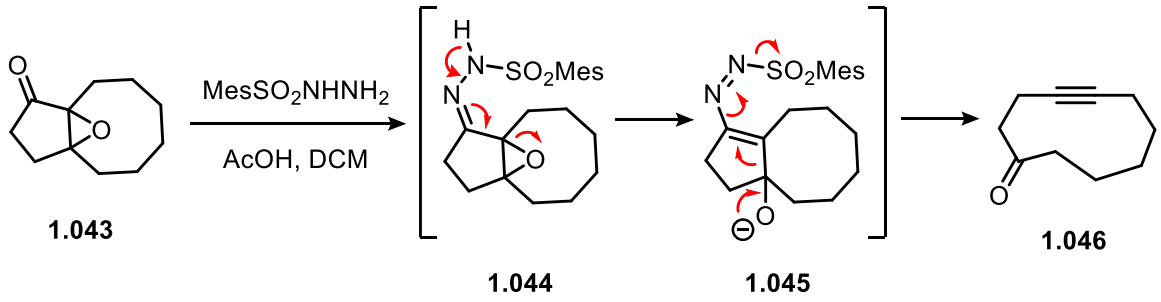
Figure 1.03: A general schematic of the types of fragmentation ring expansions. (A) 'Direct' fragmentation of a bicyclic system. (B) Side-chain insertion forming a transient bicycle followed by fragmentation.

A classic example of a 'direct' fragmentation is shown in Scheme 1.07: a Grob-fragmentation used in the synthesis of periplanone B, a sex excitant pheromone of *Periplanata americana*.[38,39] A typical Grob fragmentation involves the 'pushing' of electrons from an electron-donating group (such as an OH group in this case) toward a leaving group (such as OTs) located 4 atoms away, resulting in the cleavage of a C-C bond. The antiperiplanar arrangement of bridging bond and the tosyl group in intermediate **1.041** is imperative to the outcome of this reaction. The reaction begins with lithiation of both alcohols in **1.039**, which allows selective tosylation of the less sterically hindered alkoxide. The resulting intermediate **1.041** undergoes a Grob-fragmentation to produce the ring-expanded product **1.042** in 38% yield.



Scheme 1.07: A Grob-fragmentation used in the synthesis of periplanone B.

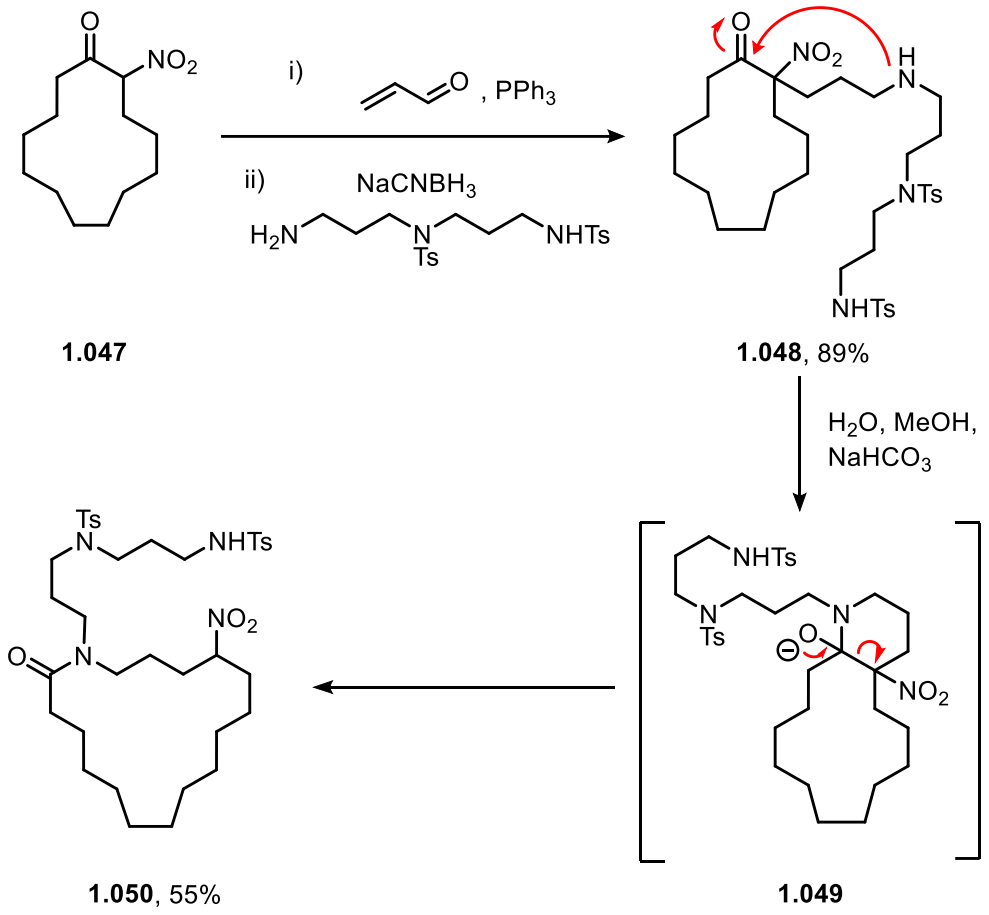
Eschenmoser fragmentations are another example of a 'direct' fragmentation reaction. They involve the reaction of α,β-epoxy ketones (**1.043**) with aryl sulfonyl hydrazines to form ring-expanded cycloalkynes (**1.046**). In the example shown in Scheme 1.08, the release of nitrogen gas provides a strong thermodynamic driving force for the formation of highly strained cycloalkynes, enabling access to structures that would almost certainly be inaccessible *via* end-to-end macrocyclisation methods.[40,41]



Scheme 1.08: An Eschenmoser fragmentation

As noted earlier, fragmentation ring expansions are not limited to bicyclic starting materials. It can also be applied to non-bicyclic systems, provided that the reaction proceeds through a bicyclic intermediate, such as in side-chain insertion. This work was pioneered by Hesse *et al.* with their work on C-N bond formation *via* side-chain insertion for the synthesis of macrocyclic spermidine alkaloids.[42] As exemplified in Scheme 1.09, a fragmentation ring-expansion is facilitated by the insertion of a pendant secondary amine into an internal carbonyl group. The 13-membered intermediate **1.047** first undergoes a Michael addition and reductive amination to give **1.048**, which subsequently undergoes side-chain insertion under mildly basic conditions

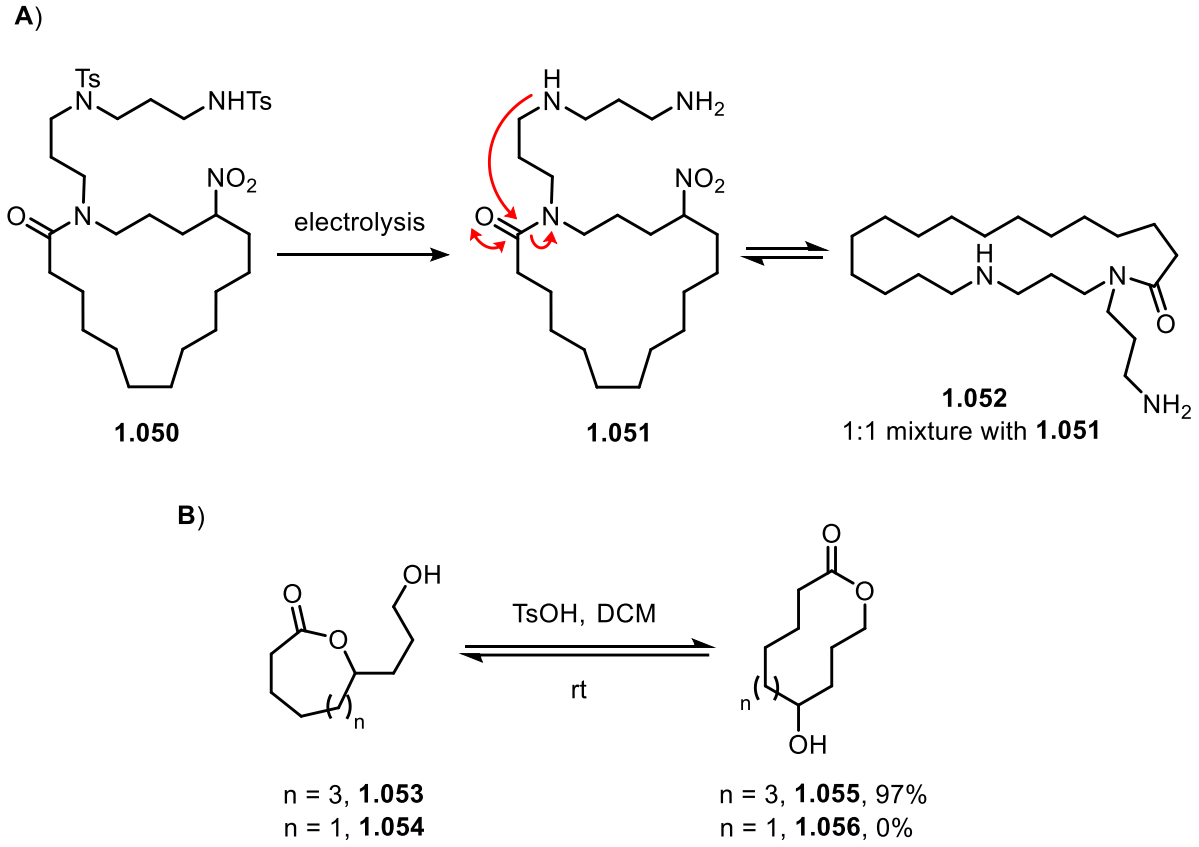
to afford the bicyclic intermediate **1.049**. The key C-C bond fragmentation is driven by the expulsion of a relatively good α-nitro carbanion leaving group, and the formation of the more thermodynamically stable lactam product, **1.050**.



Scheme 1.09: Fragmentation ring expansion enabled by side-chain insertion in the synthesis of a macrocyclic spermidine alkaloid derivative.

Subsequent treatment of **1.050** with electrolysis removes both pendant amine protecting groups, allowing successive ring expansion *via* a second side-chain insertion to form 21-membered macrocycle, **1.052**. However, this second ring expansion does not reach completion, resulting in a 1:1 mixture of **1.051** and **1.052** (Scheme 1.10A). The lack of thermodynamic drive between the two isomers, due to no overall change in functional group, is the likely cause of this inactivity. Reversible ring expansions are governed by the relative stabilities of the ring sizes before and after the expansion, and a clear driving force for rearrangement is needed for effective ring expansion. This concept has been investigated in a series of transesterification reactions by Corey *et al.*, in which the sole thermodynamic driving force is the relative stability of the different ring sizes (Scheme 1.10B).[43] A notable example is the efficient ring expansion of the 9-membered lactone (**1.053**) to the 12-membered lactone (**1.055**), with >95% conversion

in just 2 hours. In this case, the reaction is driven by the relief of transannular strain in the 9-membered ring, upon expanding to a 12-membered ring. In contrast, the 7-membered lactone (**1.054**) does not undergo ring expansion to the 10-membered lactone (**1.056**); 'normal'-sized rings are typically more stable than medium-sized rings and this reflected in this example, with **1.154** presumably being more stable than its isomer **1.156**.



Scheme 1.10: Fragmentation ring expansions in the absence of a thermodynamic driving force. (A) Successive ring expansion for the synthesis of a macrocyclic spermidine alkaloid. (B) Intramolecular transesterification demonstrating the relative stabilities of each ring size.

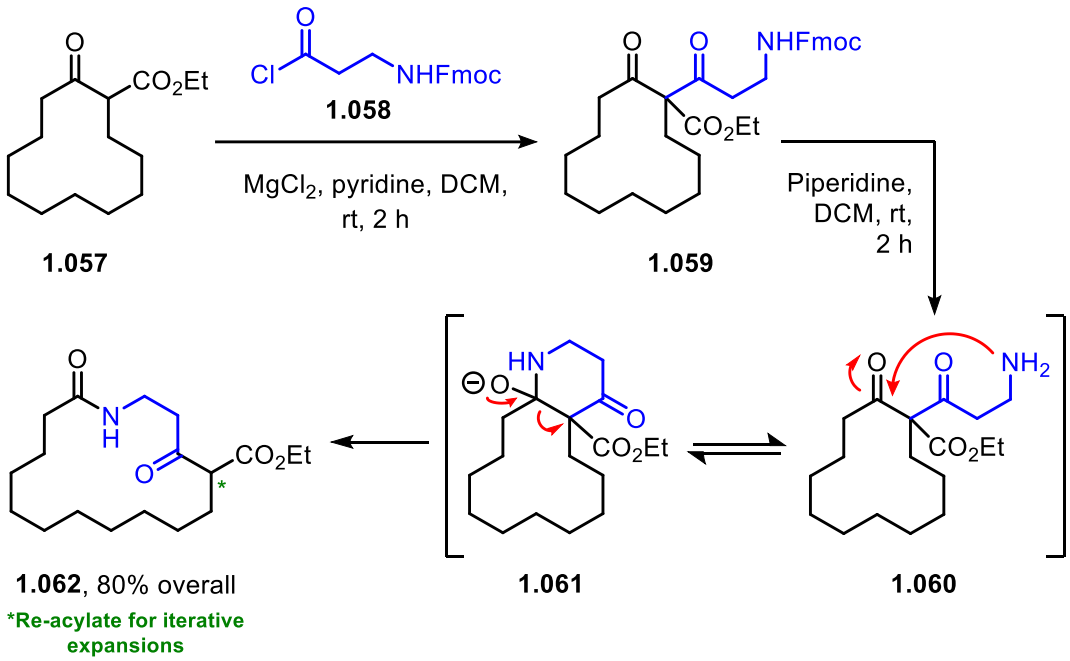
1.4 Unsworth group ring expansion methods

The work of Hesse *et al.* on side-chain insertion-based fragmentation ring expansions helped to inspire the ring expansion methods designed and developed within the Unsworth group. The group has developed 3 distinct ring expansion reaction approaches: successive ring expansion (SuRE), cyclisation ring expansion (CRE) cascade reactions, and conjugate addition/ ring expansion (CARE), discussed in turn below.

1.4.1 Successive ring expansion (SuRE)

SuRE reactions employ a side-chain insertion strategy and operate by first functionalising an existing cyclic system with a linear fragment containing a protected nucleophile (usually *N*- or

O-nucleophiles).[44–47] In the example shown below, β-keto ester **1.057** is acylated with acyl chloride **1.058** to give the product **1.059**. Upon treatment with piperidine, the Fmoc protecting group is cleaved, exposing the free amine, which subsequently attacks the internal carbonyl to form the key bicyclic intermediate **1.061**. Fragmentation of this intermediate, driven by the formation of a stable amide group, affords the final product **1.062**. This ring expansion method enables the synthesis of products with a broad range of ring sizes, either by using cyclic starting materials of various sizes or by performing iterative ring expansions on the macrocyclic products by re-acylating the α-position of the β-keto ester. Notably, the Unsworth group later adapted SuRE for use on lactam starting materials, which operate similarly, but *via* acylation of the nitrogen rather than the α-position of the β-keto ester.[46]

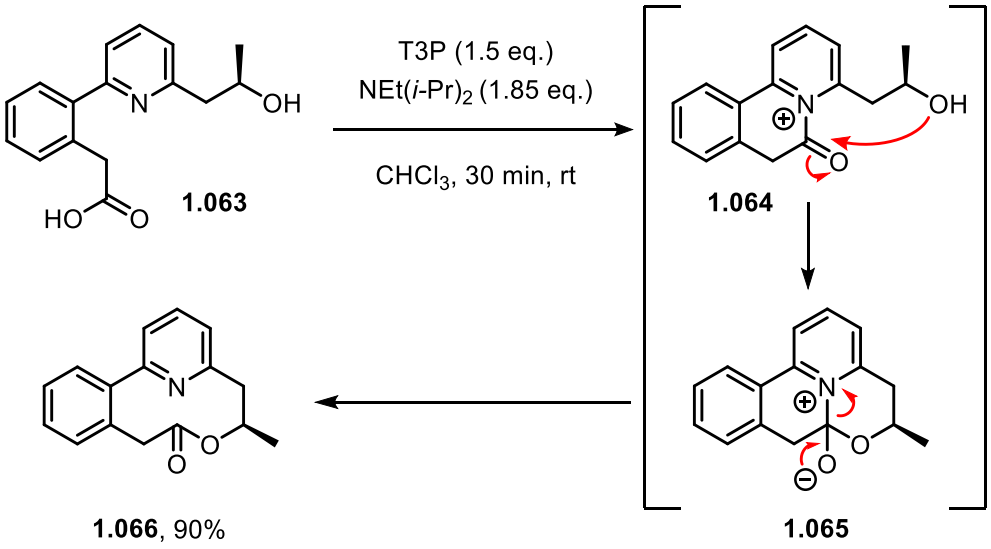


Scheme 1.11: Successive ring expansion (SuRE).

1.4.2 Cyclisation/ ring expansion (CRE)

The second ring-expansion approach developed within the Unsworth group is cyclisation/ ring expansion (CRE).[48,49] CRE does not require a preformed cyclic starting material; instead, both the cyclic and bicyclic intermediates are formed *in situ* prior to the key fragmentation. Because both the cyclic and bicyclic intermediates are formed through 'normal'-sized ring transition states, the overall transformation is designed to be more kinetically favourable than the alternative direct end-to-end macrocyclisation. An example CRE reaction is shown in Scheme 1.12 below. In this instance, the central pyridine acts as an internal nucleophilic catalyst, enabling the initial cyclisation with the activated carboxylic acid to form charged intermediate

1.064. The terminal nucleophile, in this case a hydroxy group, then undergoes side-chain insertion at the activated carbonyl to construct the bicyclic system intermediate (**1.065**) which immediately undergoes fragmentation. The ring-expanded lactone product (**1.066**) is obtained in a 90% yield. Crucially, in the absence of the internal pyridine catalyst, cyclisation does not occur, and none of the analogous 10-membered ring product is obtained. This control reaction helps to demonstrate that the CRE reaction proceeds *via* the cascade ring expansion mechanism shown, rather than *via* direct end-to-end cyclisation.



Scheme 1.12: Cyclisation/ ring expansion cascade reaction.

This ring expansion method has been extensively studied and has proven to be a highly versatile strategy for accessing complex, functionalised medium-sized rings and macrocycles. Within the classic CRE scaffold (**1.067**), there are 3 sites of potential variation (**Z**, **E**, and **Nu**; Figure 1.04). The first is the internal nucleophilic catalyst (Z), which can be substituted with various functional groups, including tertiary amines, piperidines, sulfides, and selenoethers. Depending on the method of activation, the electrophilic component (E) can include a range of groups such as activated carboxylic acids, carbamoyl chlorides, and thiocarbamoyl chlorides. Lastly, the nucleophilic component can be derived from alcohols, phenols, amines, thiols, and others. With such a broad scope for variation, the Unsworth group has successfully adapted the CRE cascade reaction to synthesise and wide range of functionalised medium-sized rings and macrocycles (as shown in Figure 1.04).

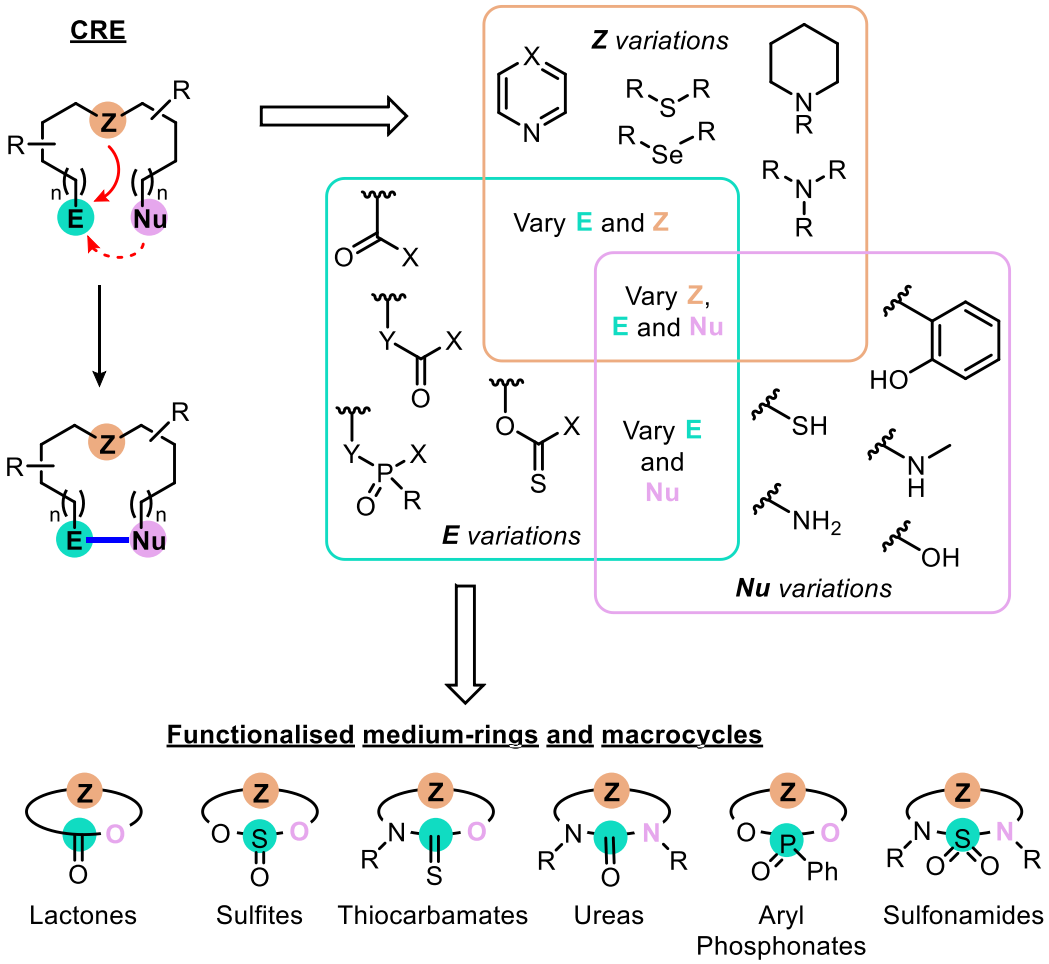
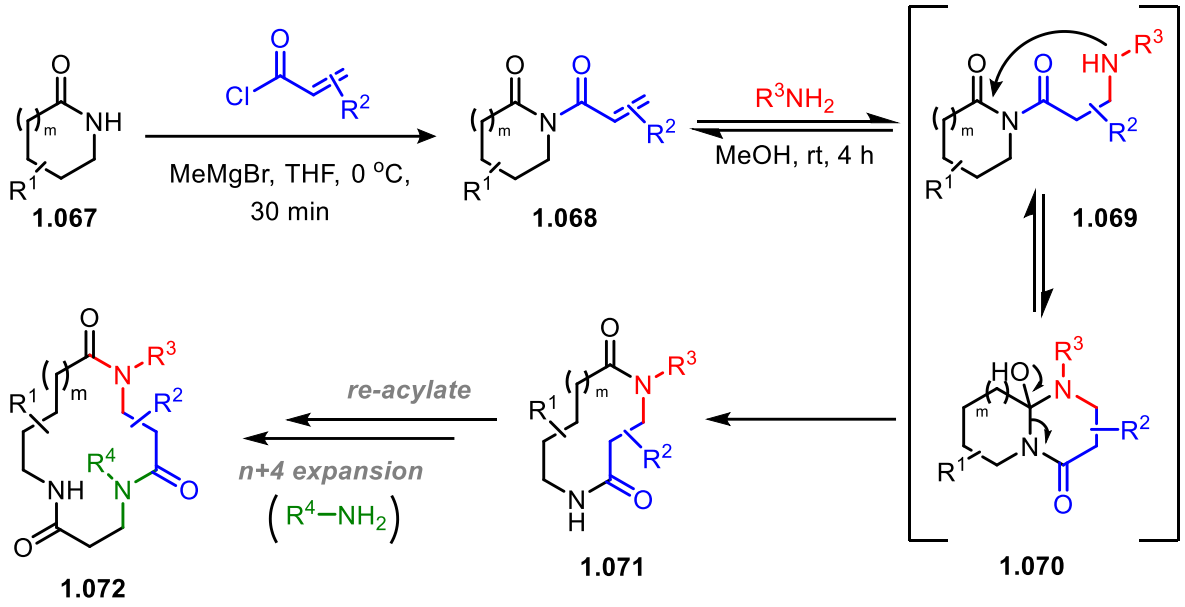


Figure 1.04: Sites of variation within a typical CRE scaffold.

1.4.3 Conjugate addition/ ring expansion (CARE)

The final class of ring expansion reaction developed within the Unsworth group is the conjugate addition/ ring expansion (CARE) cascade approach (Scheme 1.13).[50–53] CARE takes inspiration from the aforementioned SuRE chemistry but eliminates the need for protecting groups. As with SuRE, CARE begins with *N*-acylation of a lactam (*e.g.* **1.067**), but in contrast to SuRE, this is done using acryloyl chloride to afford acryloyl imide intermediate **1.068**. This key acryloyl imide motif is central to the CARE chemistry as it enables a subsequent Michael addition following reaction with a primary amine in methanol, affording intermediate **1.069**. The newly installed secondary amine then attacks the imide's internal carbonyl to form bicyclic intermediate **1.070**, which fragments into the ring-expanded lactam **1.071**. Iterative ring expansions in CARE are well-tolerated and are carried out by re-acylating the lactam product and reacting the resulting acryloyl imide with a second primary amine (**1.071** → **1.072**).



Scheme 1.13: A conjugate addition/ ring expansion (CARE) cascade reaction.

Extensive optimisation of the CARE chemistry was carried out to identify the best conditions to maximise the isolated yields of medium-sized ring and macrocyclic products. Although methanol was identified as the optimal solvent, all 18 solvents tested in the solvent screening showed at least some conversion into the 10-membered lactam product **1.071**. Importantly for the work described in this Thesis, water and DMSO — two solvents that are better tolerated by biological systems than most — proved to be effective solvents for the CARE chemistry. Furthermore, the CARE chemistry proved to be highly tolerant of structural modifications. Variations in the lactam ring size, substitutions on the lactam and Michael acceptor (R^1 and R^2), and a wide range of functionalised primary amines were all well accommodated. Perhaps most importantly, CARE exhibited high selectivity for primary amine nucleophiles, even in the presence of potentially competing functional groups such as alcohols, phenols, and anilines when these were appended to the primary amine.[50] Given the method's robustness toward multifunctional substrates and its compatibility with biologically compatible solvents, it was proposed that future applications of CARE could be directed toward the modification of biological systems.

1.5 Protein bioconjugation

Protein bioconjugation is a term used to describe the formation of a covalent linkage between two molecules, at least one of which is a protein.[54,55] In recent years, protein bioconjugation has garnered considerable attention in modern biomedical and biotechnological research, as chemical modifications can generate proteins with altered, or novel, structure and function.[56] Techniques for non-specific bioconjugation — in which reactive groups attach to biomolecules at multiple, uncontrolled sites— such as fluorescent dye labelling[57] or the preparation of first-generation antibody drug conjugates (ADCs)[58], are well established and commonly used. However, the limitations of these approaches soon became apparent: the lack of control results in the conjugation of multiple payloads to nucleophilic sites on a single protein leading to problems with conjugate stability (*i.e.* risk of unfolding), heterogeneity, and reproducibility. The need for precise protein chemistries became evident, prompting extensive research over the past three decades into developing methodologies that address these shortcomings. Over the years, researchers have developed strategies that rely on the selective targeting of specific amino acid residues, with lysine and cysteine standing out as the most effective and widely exploited of them all.[59] Nevertheless, these methods continue to face challenges in site-selectivity, since many proteins contain multiple identical copies of these residues. While unnatural amino acids provide a powerful route to highly specific bioconjugation, they rely on their efficient incorporation into biomolecules, which can limit their practicality.[60] Consequently, increasing attention has turned to the protein termini (*N*- and *C*-termini) as unique and promising sites for precise protein modification. In particular, the *N*-terminus has emerged as a promising target for site-selective conjugation, since its (typically) solvent-exposed α-amino group represents an accessible and well-defined chemical handle.[61] Importantly, chemical modifications at the *N*-terminus tend to be minimally disruptive to the protein's native structure and function since they are usually positioned well away from active- or binding-sites. Furthermore, the adjacent amide bond lowers the pK_a of the α-ammonium group (6–8) compared with the ε-ammonium groups of lysine residues (~10), providing a basis for selectivity.[62] Broadly, methods for *N*-terminal bioconjugation fall into two categories (Figure 1.05): (1) selective strategies, which require careful pH and stoichiometric control to achieve site-selectivity based on differences in pK_a, and (2) specific strategies, which typically rely on the involvement of a nearby sidechain to direct the modification. Both approaches have proven valuable across a range of biological applications, yet selecting the most appropriate approach requires careful consideration. Generally, an effective bioconjugation strategy must satisfy several key criteria: the reagents should be stable, the modification should proceed

under biocompatible conditions (*e.g.* mild pH and temperature), and the modified protein should retain the structure and function of its native form. However, a recent study from the Spicer group demonstrates that even the most sophisticated *N*-terminal bioconjugation strategies are not universally compatible with all proteins, concluding that there is no "one-method-fits-all" approach to protein bioconjugation.[63] Thus, there remains a need for further innovation in this field to expand the *N*-terminal bioconjugation toolkit.

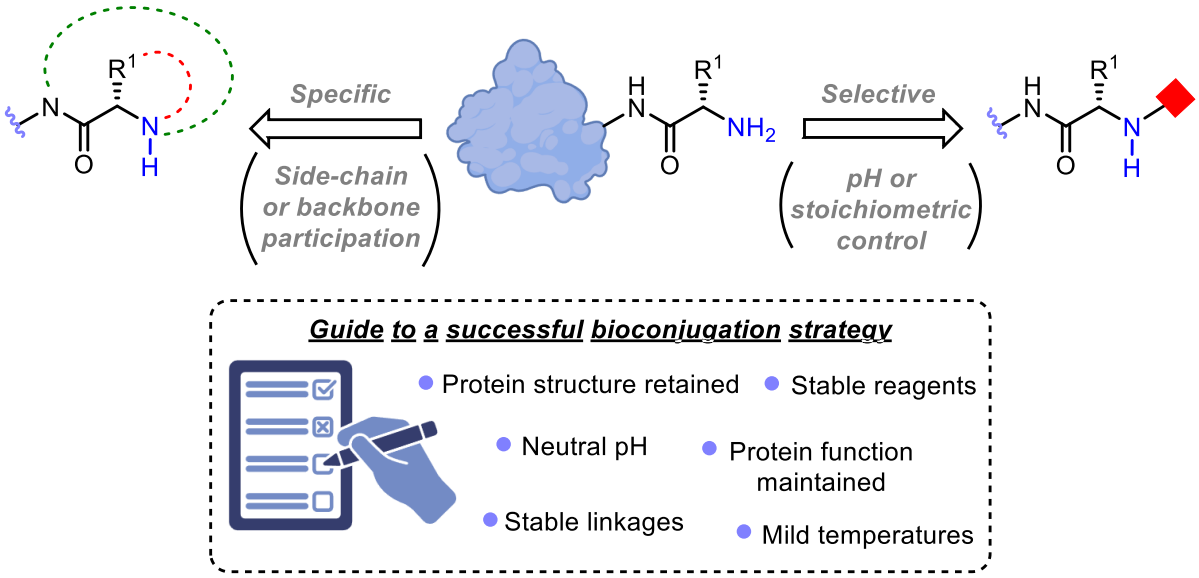


Figure 1.05: Specific *vs.* selective *N*-terminal protein bioconjugation.

1.5.1 Specific *N*-terminal protein bioconjugation

As discussed previously, relying solely on pH and stoichiometric control cannot always guarantee full selectivity and may lead to off target labelling in proteins that contain multiple reactive sites. The most sophisticated *N*-terminal protein bioconjugation strategies can overcome these limitations by exploiting the participation of the *N*-terminal residue's side chain or the protein backbone to ensure specific modification of the α-amino group.

1.5.1.1 N-terminal residue side chain participation

Cysteine is the most commonly exploited residue in these approaches. When located at the *N*-terminus, a cysteine residue presents a 1,2-aminothiol motif that can be readily targeted by chemistries specific for this functional group. Unfortunately, cysteine residues are rarely found at the *N*-terminus of naturally occurring proteins and are therefore typically introduced through protein engineering, which extends the experimental timeline.[64,65] When at the *N*-terminus, cysteine residues can undergo unwanted thiazolidine formation with pyruvic acid, an abundant metabolite of *E. coli*, which removes the 1,2-aminothiol motif for modification.[65] However,

with careful control of the reaction conditions, this secondary modification can be minimised. Moreover, introducing an unnatural cysteine at the *N*-terminus can lead to protein misfolding due to the formation of non-native disulfide bonds with other free thiols present in the protein.[66] Nonetheless, several chemistries have been reported to target the 1,2-aminothiol motif in proteins, some of which are illustrated in Figure 1.06.

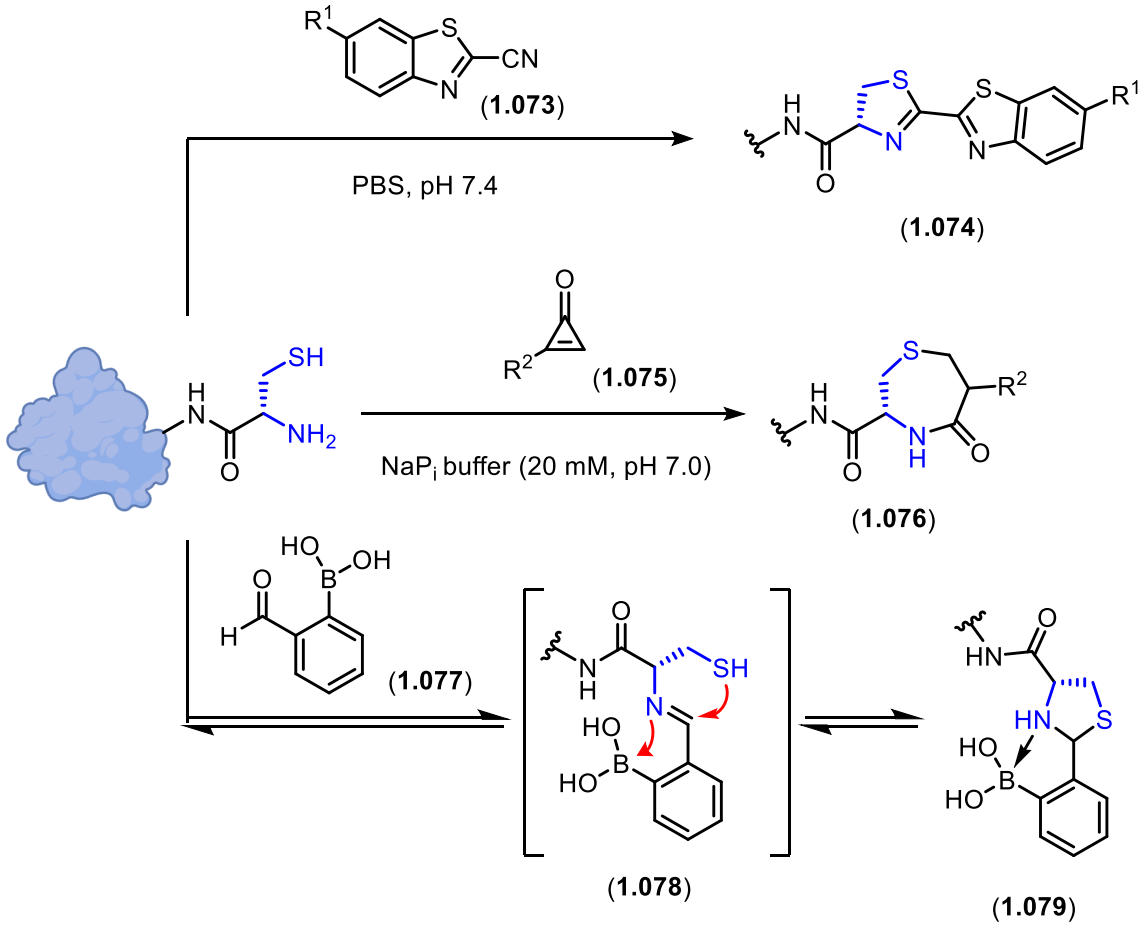


Figure 1.06: Specific *N*-terminal cysteine protein bioconjugation strategies.

Since their discovery, 2-cyanobenzothiazole reagents (CBTs) (**1.073**) have proven highly effective for the rapid and selective modification of *N*-terminal cysteine residues.[67,68] This transformation is a water-compatible condensation reaction that is selective for 1,2- and 1,3-aminothiol substrates over the free thiols of internal cysteine residues. Investigations into the kinetics of this condensation showed that its second-order rate constant (9.19 M^{-1} s^{-1}) is substantially greater than that of the copper catalysed azide-alkyne cycloaddition (CuAAC), a widely used bioconjugation method.[69] Furthermore, the resulting bioconjugates (**1.074**) are stable and have been applied in a range of *in vitro* and in *vivo* biological applications, including imaging with fluorescently tagged CBT probes.[70]

The selective reaction of *N*-terminal cysteine residues with aldehydes to form thiazolidines has been well studied but is often overlooked due to the relatively harsh reaction conditions needed (pH 4–5, high reactant conditions) and slow kinetics (incubation times >2 days).[71] Notably, a later study demonstrated that the condensation between 1,2-aminothiols and aliphatic aldehydes is in fact highly specific and rapid under physiological conditions, contradicting the earlier work.[72] However, a solution to these challenges was presented by Gao *et al.*, who reported the use of *ortho*-boronic acid substituted benzaldehydes (**1.078**) for the rapid and selective modification of 1,2-aminothiols in peptides and proteins.[73] The reagents are among the fastest reported in bioorthogonal chemistry, with rate constants up to 1 x 10^3 M^{-1} s^{-1} under physiologically compatible reaction conditions. The boronic acid moiety not only activates the iminoboronate for thiazolidine formation but also improves the stability of the resulting bioconjugate by forming a stable thiazolidino-boronate complex (**1.079**). Under physiological conditions, the thiazolidino-boronate complex is highly stable; however, even mild acidification induces dissociation. The authors proposed leveraging this property for the design of drug-protein conjugates capable of the release of small molecules within endosomes (pH <7).[74]

More recently, Bernardes *et al.* reported the use of monosubstituted cyclopropenone (CPO) (**1.075**) reagents for the modification of *N*-terminal cysteine residues under biocompatible conditions.[75] This strategy enables the construction of dual-functionalised bioconjugates by first modifying the *N*-terminus with a CPO, followed by selective modification of an internal cysteine residue using a conventional cysteine-targeting reagent. This is possible due to the selectivity of CPOs for the 1,2-aminothiol motif, allowing them to discriminate against free thiols present elsewhere in the same protein, in other proteins, or in nucleophilic agents such as DTT. The ring-expanded 1,4-thiazepan-5-one bioconjugates (**1.076**) are stable under physiological conditions and can be functionalised with bioorthogonal handles for further derivatisation.

Serine and threonine, like cysteine, present a 1,2-dinucleophile motif at the *N*-terminus, which can be readily exploited for site-specific modification. Perhaps more appealing for these residues, though, is the ability to introduce a reactive carbonyl *via* a rapid, one-pot α-oxo-aldehyde/ketone formation using oxidation (**1.080** → **1.081**; Figure 1.07A).[76] Biomimetic transamination of non-serine or threonine residues is reported for the installation of α-oxo-aldehydes at *N*-termini, but these reactions can be sluggish and low-yielding.[77,78] Owing to the higher reactivity of aldehydes compared to ketones, *N*-terminal serine residues are more

commonly subjected to transamination than threonine residues and can be selectively targeted using a variety of chemoselective reactions, as illustrated in Figure 1.07B. Van Delft's strain-promoted alkyne-nitrone cycloaddition (SPANC) relies on the site-selective installation of nitrones into biomolecules *via* treatment of α-oxo-aldehydes with *p*-methoxybenzenethiol, *N*-methylhydroxylamine, and *p*-anisidine.[79] The formation of nitrones from α-oxo-aldehydes is thought to proceed *via* a similar mechanism to oxime formation from aldehydes and hydroxylamines, with *p*-anisidine acting as a nucleophilic catalyst to enhance the rate of ligation.[80] Treatment of nitrone **1.083** with ring-strained 4-dibenzylcyclooctynol (**1.084**) gave a stable isooxazoline bioconjugate (**1.085**) in high yield. Strategic functionalisation of the nitrone (by replacing *N*-methyl) afforded derivatives with markedly improved reaction kinetics, especially when bearing electron-withdrawing groups such as esters and amides. This approach was successfully applied to the modification of interleukin-8 (IL-8), a chemokine that has traditionally been modified by non-specific labelling techniques with limited control.[81]

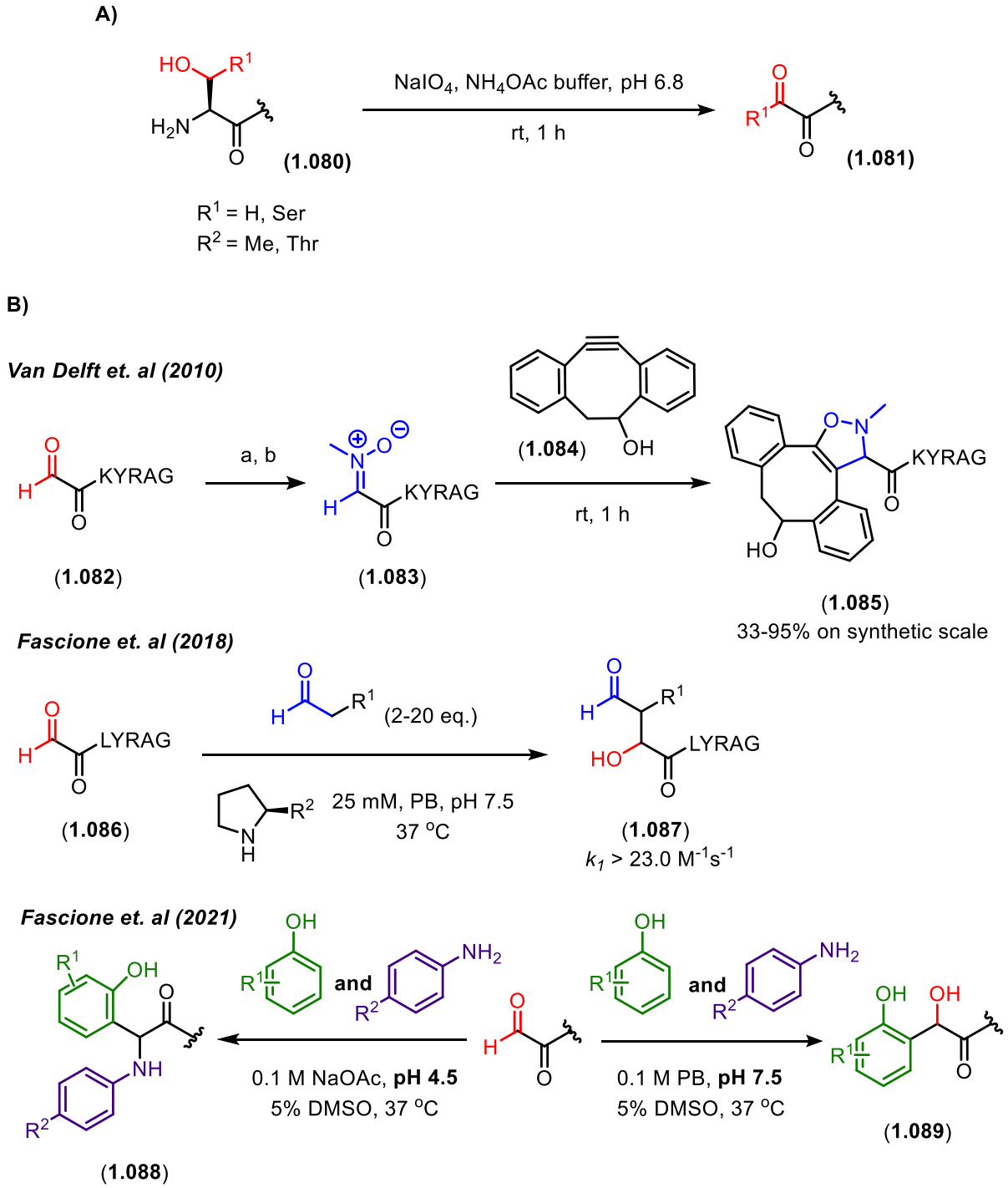


Figure 1.07: (A) The oxidation of *N*-terminal serine and threonine residues to α-oxo-aldehydes using periodate. (B) Chemoselective reactions for the ligations of α-oxo-aldehydes to functionalised small molecules. (a): $NaIO_4$, NH_4OAc buffer, pH 6.8, rt, 1 h; (b): *p*-$MeOC_6H_4SH$, rt, 1 h; then: *p*-$MeOC_6H_4NH$, MeHNOH-HCl, rt, 20 min.

The water compatible L-proline-catalysed cross aldol reaction of aldehydes was of great inspiration to the Fascione group's organocatalyst-mediated cross aldol ligation (OPAL) chemistry.[82,83] OPAL is a rapid bioconjugation strategy which utilises commercially available secondary amine organocatalysts and small aldehyde probes, in small excess, to form stable C-

C bonds at neutral pH at both *N*-terminal and internal aldehyde sites.[84] Notably, both the choice of secondary amine organocatalyst and the functionalisation of the small aldehyde donor are critical for achieving optimal ligation rates; under optimised reaction conditions, rate constants as high as 24 M^{-1} s^{-1} have been reported. Importantly, under neutral aqueous conditions, the OPAL predominantly forms the β-hydroxy aldehyde product (**1.087**), which is significantly less reactive towards further cross-aldol reactions than in its enal form, thereby ensuring control of over-modification and the formation of heterogenous bioconjugates. Since the aldehyde handle in retained in the product, further functionalisation of the product is possible (*e.g.* by oxime ligation). This was exploited by the authors to generate bi-functionalised bioconjugates of hydrophilic acylated surface proteins (HASPs) from *Leishmania donovani* through a tandem OPAL/ oxime ligation reaction.

The Fascione group further demonstrated the utility of α-oxo-aldehydes by developing two novel bioconjugation strategies that can show divergent reactivity with careful pH control.[85] As expected, a reaction between an α-oxo-aldehyde and aniline results in the formation of an imine *via* a Mannich-type reaction at pH 4.5. In the presence of phenol, the imine can undergo a Betti-type reaction to afford the bi-functionalised bioconjugate **1.088**.[86] Unusually, at neutral pH, the α-oxo-aldehyde is simply converted to the aldol condensation product with the phenol, giving the monofunctionalised bioconjugate **1.089** as the only product. Thus, the α-oxo-aldehyde is displaying pH divergent reactivity where the desired bioconjugate can be favoured with judicious choice of reaction conditions. Though both bioconjugates were unstable, it is suggested that the nature of the phenol component may directly influence bioconjugate stability. If this could be improved, such constructs could potentially be used for in the oxidative release of small molecules from proteins.

Whilst 1,2-dinucleophiles at the *N*-terminus (Cys, Ser, Thr) are attractive targets for specific *N*-terminus modification, new reagents that target other residues have broadened the scope of this approach. Tam *et al.* first proposed a modified Pictet-Spengler reaction as a strategy for the site-specific modification of *N*-terminal tryptophan residues in peptides (Figure 1.08).[87] The reaction begins with an intermolecular condensation between a β-arylenylamine (**1.090**) and an aldehyde, forming the protonated imine intermediate **1.091a**. This intermediate subsequently undergoes a spontaneous 5-*endo-trig* cyclisation between the iminium ion and an aromatic C-nucleophile, which, after a slow migration step (**1.091b → 1.092**) and rearomatisation, affords the tetrahydroisoquinoline product **1.092**. In their study, unprotected peptide *C*-terminal aldehydes (**1.093**) were ligated with *N*-terminal tryptophan-containing peptides (**1.094**) through a Pictet-Spengler reaction, affording a stable and irreversible tetrahydro-β-carboline linkage (**1.095**). Despite its promise, the method suffered from slow kinetics (~1 x 10^{-4} M^{-1} s^{-1}),[88] required high reagent concentrations (> 50 mM), and the reaction conditions were completely incompatible with those required with protein modification, with glacial acetic acid being the only solvent of choice.

However, this work inspired other researchers in the field to develop improved methods based on reactions of this kind.[89,90] Notably, Bertozzi *et al.* reported the "Pictet-Spengler ligation", an improved approach with rate constants 4–5 orders of magnitude greater than those previously reported.[91] They developed a modified *N*-alkoxytryptamine probe (**1.097**), in which the replacement of the typical aliphatic amine of tryptamine with an aminooxy group, resulted in faster iminium formation. Furthermore, by positioning the aminooxy moiety at the 2-position of the indole, they freed up the indole's naturally more nucleophilic 3-position for cyclisation. Upon ligation of glyoxyl-modified proteins (**1.096**) with **1.097**, the oxocarboline bioconjugates (**1.098**) exhibited significantly enhanced hydrolytic stability compared to model oximes, with 90% remaining intact after 2 days and the sole degradation product being the result of hydrolysis.

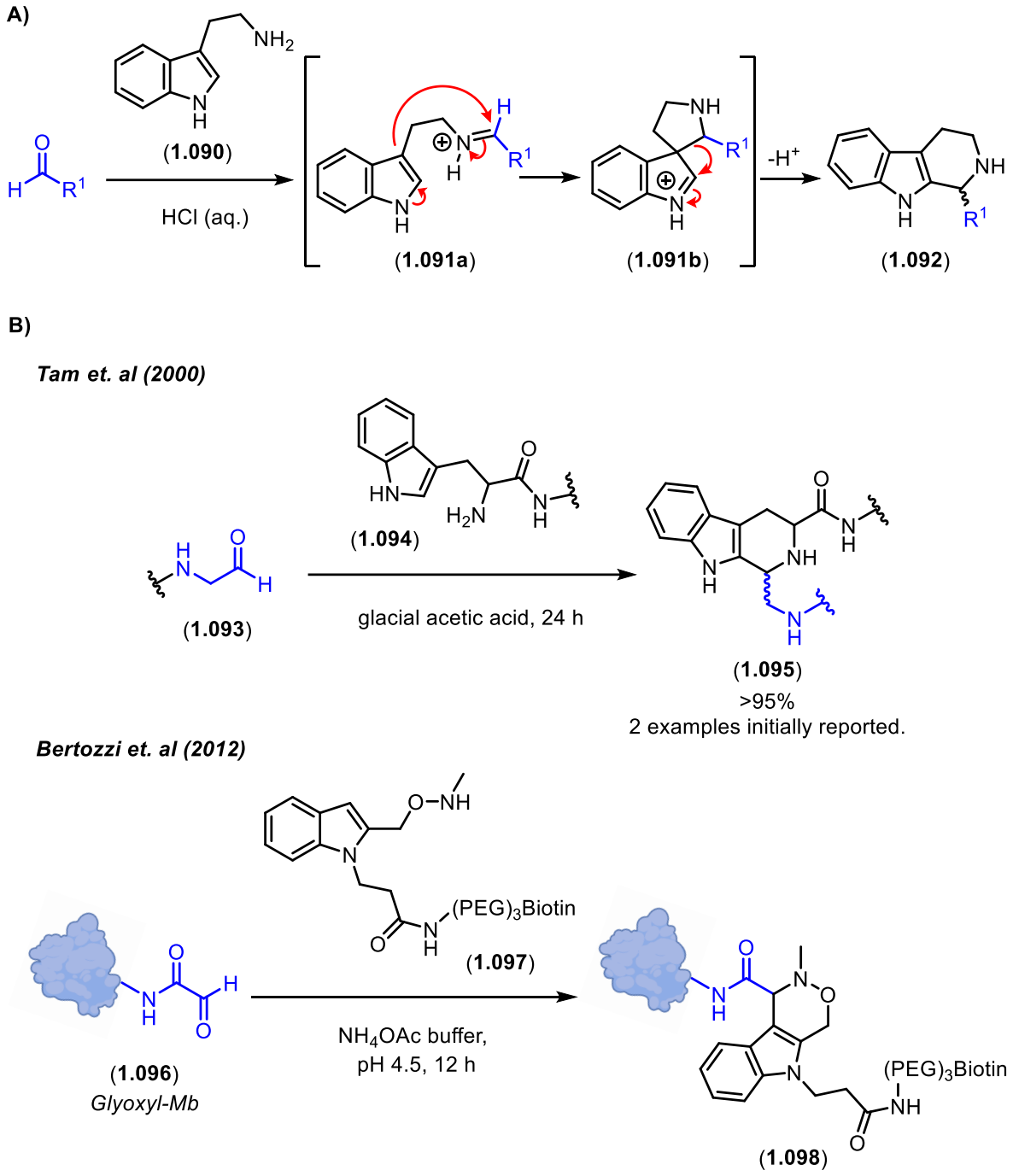
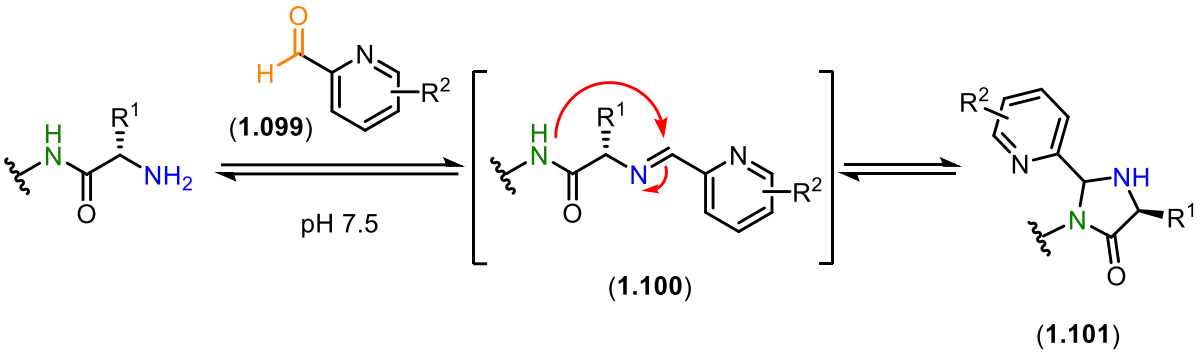


Figure 1.08: The Pictet-Spengler reaction as a bioorthogonal reaction for bioconjugation. (A) The Pictet-Spengler reaction. (B) Modified Pictet-Spengler approaches for the ligation of biomolecules.

1.5.1.2 Polypeptide backbone participation

The gold-standard reagents for site-specific *N*-terminal modification are those that exploit the polypeptide backbone to achieve selectivity. Although examples of this class of reagents are limited, 2-pyridinecarboxaldehydes (2-PCAs) (**1.099**) are by far the most well-known. First reported by Francis *et al.* in 2015, these reagents exhibit α-amino specificity by exploiting the neighbouring backbone amide, whose nitrogen is essential for cyclisation to the

imidazolidinone product (**1.100** → **1.101**; Scheme 1.14).[92] Whilst lysine ε-amino groups can react with 2-PCAs to form imine intermediates, these species are hydrolytically unstable and, without a neighbouring amide to facilitate cyclisation, do not lead to stable modification. This approach proceeds smoothly in a variety of biological buffers at neutral pH, making it broadly compatible with many proteins. The imidazolidinone products (**1.101**) are somewhat stable across a broad pH range (3–11) and are susceptible to hydrolysis at 37 °C.



Scheme 1.14: 2-Pyridinecarboxaldehydes (2-PCAs) as reagents for the specific *N*-terminal modification of peptides and proteins.

Furthermore, these conjugations can be relatively slow and partially reversible. To address these limitations, the Spicer group investigated the effect of functionalisation of 2-PCA reagents and found that 3-methoxy-2-pyridinecarboxaldehydes accelerate the rate of conjugation and produce more stable bioconjugates.[93] Nonetheless, all 20 naturally occurring amino acids are compatible with this chemistry; screening *N*-terminal peptide analogues found an average conversion of >85%, underscoring the importance of PCAs in chemical biology laboratories for both peptide and protein modifications. Continued investigation into improved 2-PCA reagents will reveal the full potential of *N*-terminal specific labelling.

1.5.2 Selective *N*-terminal protein bioconjugation

Methods that are more generally applicable to any *N*-termini cannot rely on side-chain participation and therefore must exploit regiochemical differences between α-amino groups and other competing nucleophiles to ensure selective labelling. Since lysine ε-amino groups pose the greatest selectivity challenge to *N*-terminal modification, most methods target the difference in pK_a between the α- and ε- amino groups to achieve selectivity (α-amino group pK_a: ~6–8; ε-amino group pK_a: ~10).[62] Given the abundance of examples in the literature, only a select few will be discussed in detail. For a more comprehensive overview, there are reviews covering a wide range of approaches.[55,94]

Phenyl ketene reagents (**1.102**; Figure 1.09) have been reported by Wong *et al.* as selective *N*-terminal protein modification reagents under mild conditions (pH 6.3).[95] These reagents have been successfully used to modify the *N*-termini of a variety of medicinally relevant proteins, such as insulin, lysozyme, RNase A, and BCArg, with high levels of selectivity. Although a large excess of acylating agent can be used without compromising selectivity, most *N*-terminal residues still exhibit poor conversions, with less than 50% conversion observed for 17 of the *N*-terminal residues tested. Moreover, certain functionalised ketene probes are both difficult to prepare and highly unstable, making them unsuitable for *N*-terminal protein modifications and less reliable than other well-established methods.

A preferred alternative, also proposed by Wong *et al.* is the use of 2-ethynylbenzaldehydes (2-EBAs) (**1.104**) for the selective *N*-terminal modification of proteins.[96] The reaction, carried out in slightly acidic phosphate buffer, begins with condensation of the α-amino group and the arylaldehyde to form a 2-alkynylarylaldimine intermediate (**1.105**). A subsequent 6-*endo-dig* cyclisation between the imine and alkyne produces a stable isoquinolinium bioconjugate (**1.105** → **1.106**). All 20 of the naturally occurring amino acid residues proved compatible in a screening of XSKFR peptides, with selectivity greater than 99% for the *N*-terminus in some cases. A series of functionalised 2-EBA probes were synthesised, all exhibiting high *N*-terminal selectivity and conversion. Alkyne- and fluorescein-functionalised 2-EBA probes were used to selectively modify the *N*-termini of the same panel of medicinally relevant proteins, with conversions between 10–84%. However, in the Spicer group's study mentioned earlier, a functionalised 2-EBA probe caused some degree of over-modification in every case, and all the resulting bioconjugates exhibited partial instability over the course of a week.[63]

The *N*-terminal selectivity of sulfonamide-linked alkyne boronic acids (**1.107**) is assisted by a copper-mediated amine arylation *via* a copper-chelation complex (**1.108**). These reagents, reported by Ball *et al.*, contain *o*-electron-withdrawing groups that promote N-H arylation at the α-amino group, with no observable off-target modification at lysine ε-amino groups.[97] This approach was only demonstrated on peptides, likely due to its reliance on copper(II) catalysis — which can pose cytotoxicity concerns — and the high levels of organic solvents (20–30% MeCN or DMSO) required to solubilise the reagents.

The first example of protein phthalimidation was reported by Rai *et al.*, who developed a catalyst-free methodology that proceeded efficiently in aqueous buffer under biocompatible conditions.[98] The site-selectivity between α- and ε-amino groups can be governed by precise

pH control and the relative accessibility of each site. The reaction begins with the nucleophilic addition of the α-amino group to the *N*-hydroxyphthalimide (**1.110**), forming a latent amphoteric intermediate (**1.111**). Ring closure of the intermediate is the rate-determining and irreversible step and is responsible for the overall chemoselectivity of the reaction (**1.111** → **1.112**). Notably, the resulting phthalimidoamine bioconjugate possess an electrophilic imide which can be targeted by a secondary nucleophile, enabling further late-stage functionalisation. The authors demonstrated the successful application of this chemistry to the modification of RNase A and employed functionalised *N*-hydroxyphthalimide derivatives bearing fluorescent probes and bioorthogonal handles for further derivatisation.

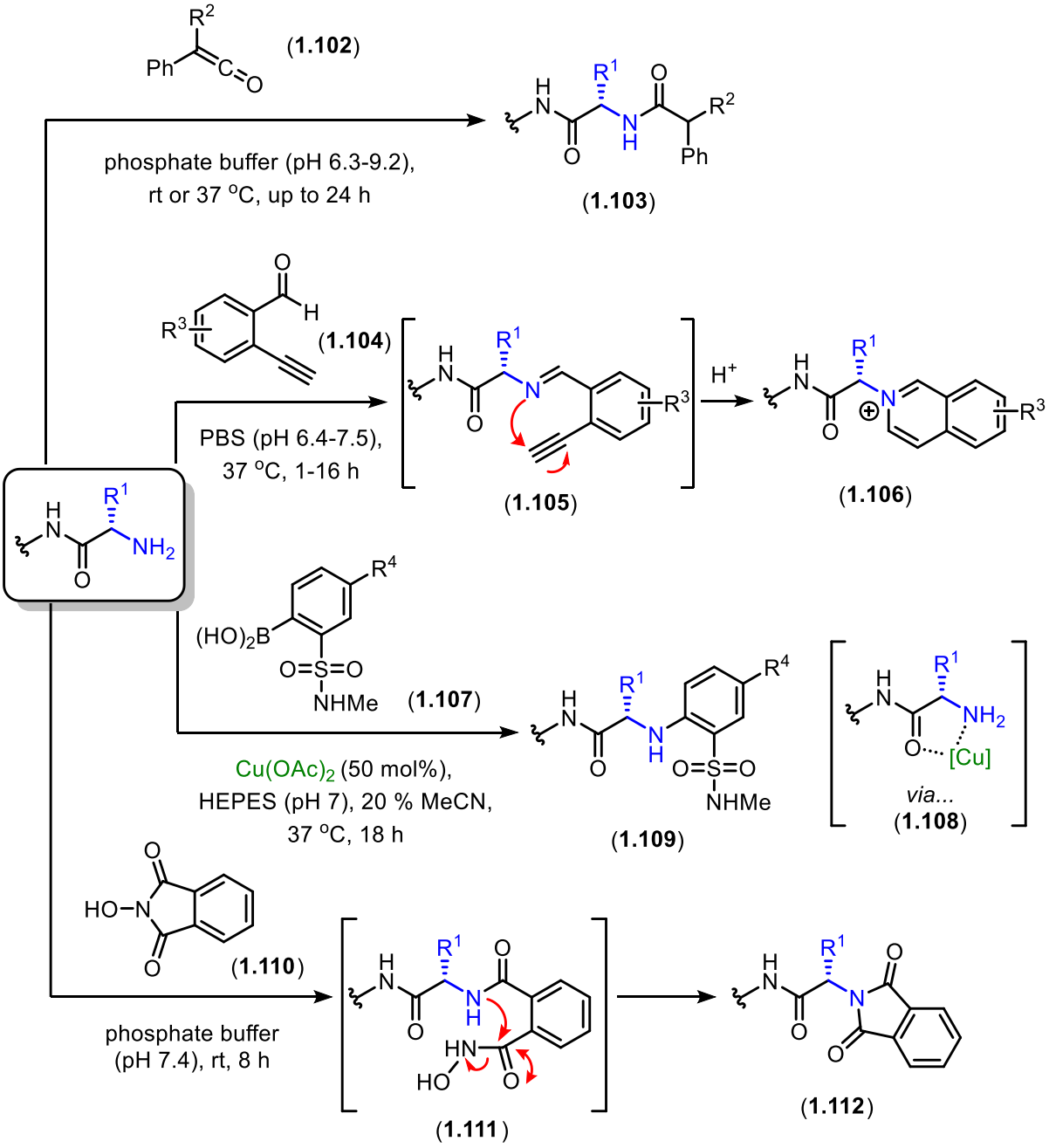


Figure 1.09: Selective *N*-terminal bioconjugation strategies.

1.5.2.1 Aza-Michael bioconjugation strategies

Among the various bioconjugation strategies previously discussed, those based on aza-Michael addition are of particular relevance to this work. Reports of such methods are limited, and those that exist have largely focussed on the modification of lysine ε-amino groups. Aza-Michael additions are inherently reversible, meaning that bioconjugation strategies based on this chemistry often struggle to form stable products and therefore rely on carefully designed reagents to ensure stable, irreversible modification. This includes the work of Jiang *et al.* who reported the use of *N*-phenylvinylsulfonamides (**1.114**; Figure 1.10A) as Michael acceptors for

the selective modification of lysine ε-amino groups.[99] Whilst they were able to demonstrate specificity for K29 in the B chain of insulin (Figure 1.10B), this is not entirely surprising, as this residue represents the only other nucleophilic amine available for modification aside from the *N*-termini of the A- and B-chains. Moreover, all cysteine residues in insulin are involved in disulfide bridges, which the authors acknowledge would pose a significant chemoselectivity issues if this were not the case. Lastly, the reaction conditions required for this chemistry are too harsh for most proteins, as it requires strongly basic conditions (pH 9) and a high concentration of organic solvent (2:1; DBU buffer: MeCN).

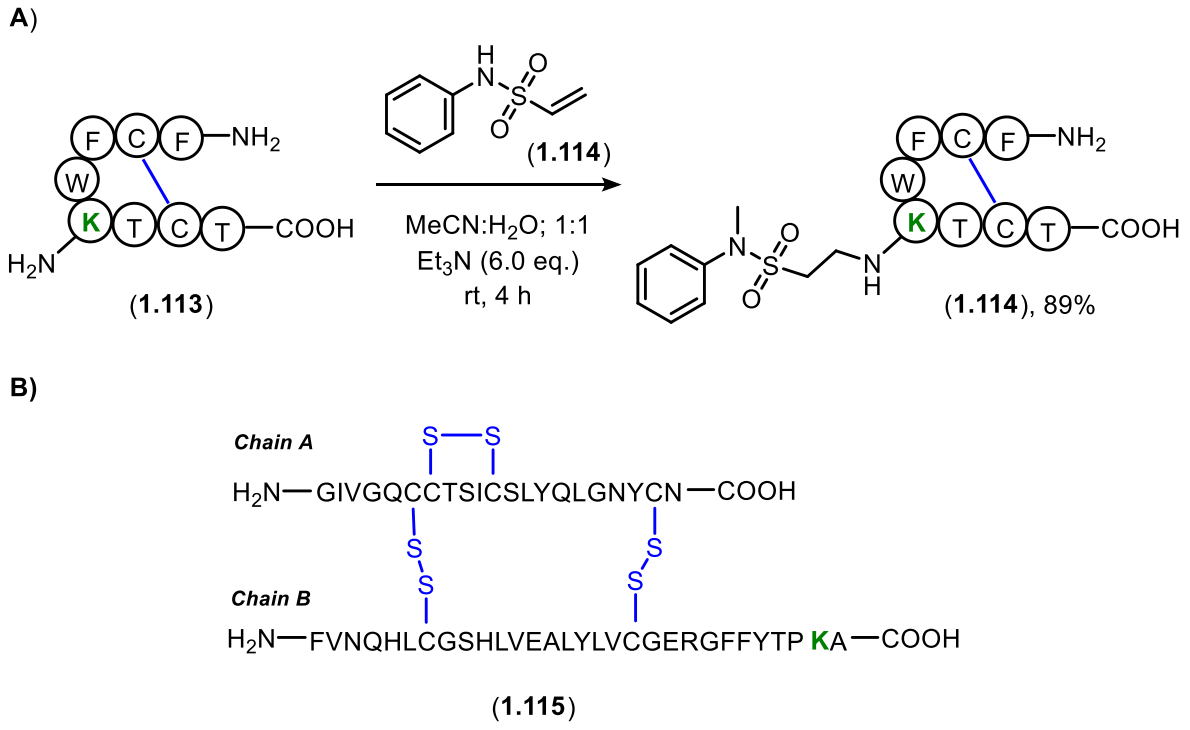
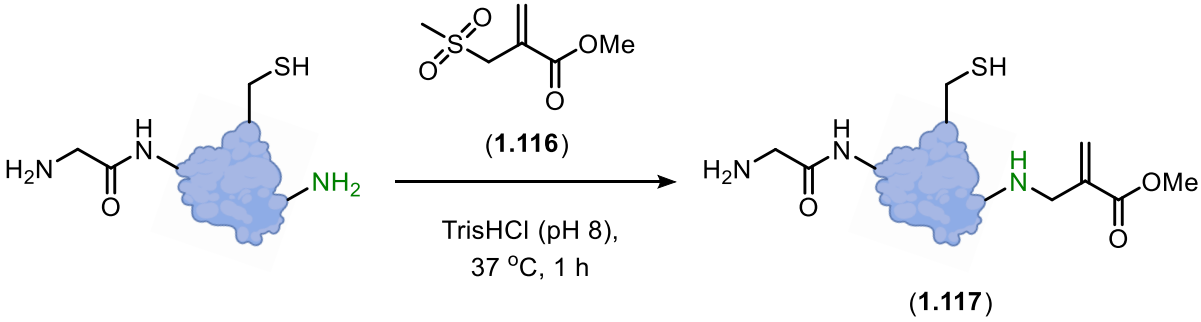


Figure 1.10: *N*-phenylvinylsulfonamides as probes for the modification of ε-amino groups *via* aza-Michael addition.

The work of Bernardes *et al.* took a different approach, designing Michael acceptor reagents that worked *via* an addition-elimination mechanism.[100] Their sulfonyl acrylate reagents (**1.116**; Scheme 1.15) exhibited preferential selectivity for ε-amino groups, even in the presence of other nucleophiles —including free thiols—and could be finely tuned to account for the local microenvironment of the target residues to improve site-selectivity. This selectivity is hypothesised to arise from transient hydrogen-bonding interactions between the sulfonyl acrylate and the lysine residue, which helps direct the reagent to the target site. Following the aza-Michael addition of the ε-amino group to the acrylate, elimination of methanesulfinic acid drives the reaction towards the formation of the acrylate-linked bioconjugate (**1.117**). A

particularly attractive aspect of this approach is that the resulting bioconjugates can be further functionalised through the acrylate handle introduced by the elimination.



Scheme 1.15: Sulfonyl acrylates as probes for the modification of ε-amino groups *via* aza-Michael addition/elimination.

Although a large number of reagents are available for site-selective bioconjugation, even the most advanced methods remain unsuitable for certain biomolecular targets. The ongoing challenge of balancing chemo- and regioselectivity, conjugate stability, and synthetic accessibility underscores the need for continued innovation in this area.

1.6 Project aims

As detailed in section 1.4.3, the conjugate addition/ring expansion (CARE) chemistry, developed by the Unsworth group, enables the well-controlled assembly of medium-sized rings and macrocycles from simple acryloyl imides and primary amines. In the initial exploration of primary amine substrates, those derived from the canonical amino acids glycine and alanine proved to be compatible with the reaction. With this in mind, and considering CARE's selectivity for primary amine nucleophiles, its tolerance for water as a solvent, and the irreversible nature of the ring expansion step, it was proposed that this approach could be adapted as a novel strategy for *N*-terminal selective protein bioconjugation (Figure 1.11). Given the amine-specific nature of the chemistry, achieving site-selectivity would rely on judicious control of the reaction conditions to preferentially target α-amino groups over ε-amino groups of lysine residues.

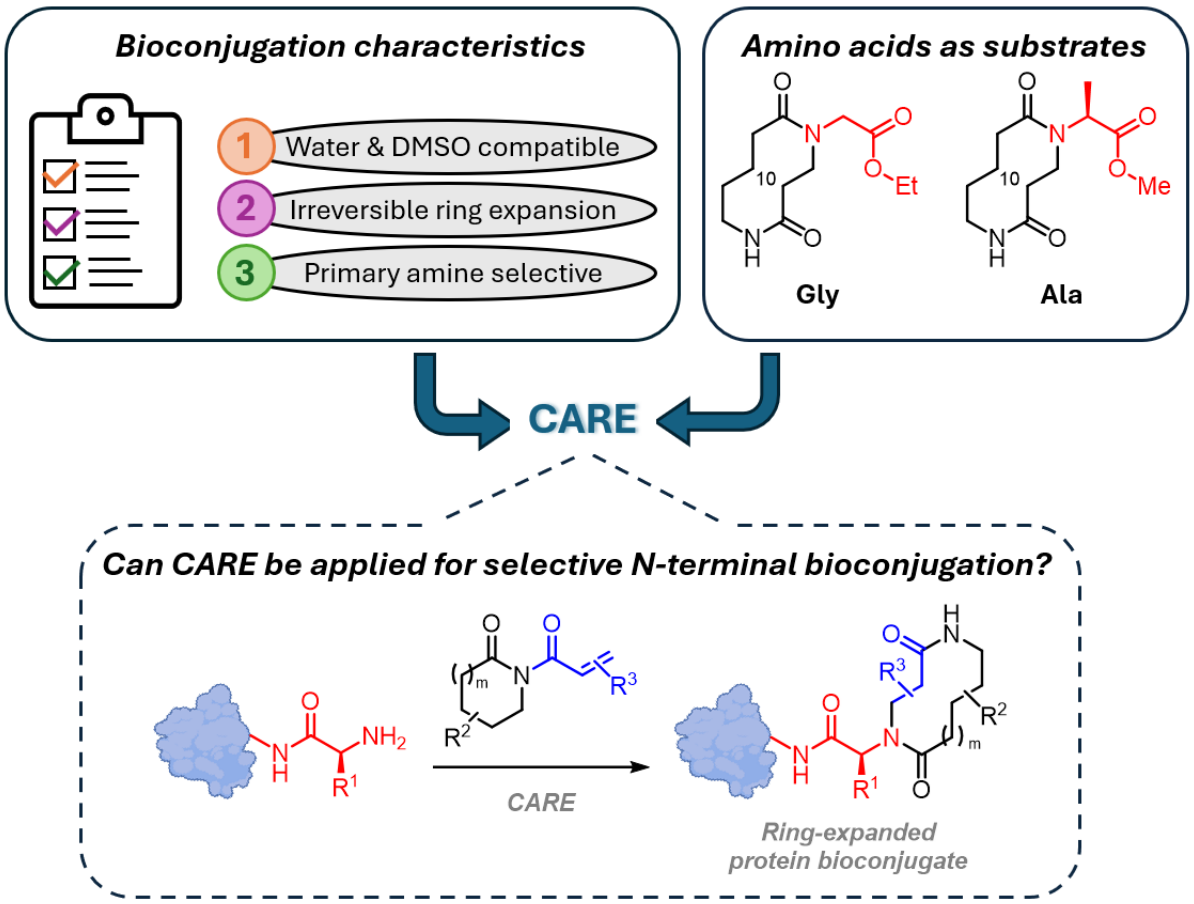


Figure 1.11: Conceptual summary of the rationale forming the foundation of this PhD project.

The initial aim of the project was to investigate CARE's compatibility with naturally occurring amino acid monomers to ensure that the reaction was not adversely affected by previously untested functional groups, such as guanidines and imidazoles. This work involved synthesising amino acid-derived medium-sized rings from canonical amino acid derivatives and investigating how steric bulk around the α-carbon affects the reaction outcome. The structure of these rings were confirmed by X-ray crystallography. Details of this work can be found in Chapter 2.

The next aim of the project was to optimise CARE for the modification of small peptides under biologically compatible reaction conditions. This involved investigating the effect of key parameters including reagent concentration, reaction pH and temperature, and the potential use of organocatalysts. A series of *N*-terminal residue mutants were synthesised and tested under the optimised reaction conditions to confirm that the reaction was broadly applicable across a range of *N*-terminal amino acids. The reaction kinetics of each *N*-terminal mutant was evaluated to identify any factors that might influence the reaction. Lastly, the reaction's

selectivity for α-amino *vs.* ε -amino groups and the stability of the peptide bioconjugates were evaluated. This work is discussed in Chapter 3.

To expand the applicability of our approach, the next objective of the project was to design novel acryloyl imide probes bearing bioorthogonal tags, enabling further reagent diversification. We developed synthetic routes to novel azide- and alkyne-tagged acryloyl imide probes using readily available starting materials and optimised their reactions for the *N*-terminal modification of model peptides. Details of this work are found in Chapter 4.

The final aim of the project was to optimise CARE for the site-selective *N*-terminal modification of proteins. This work involved the successful modification of a range of medicinally relevant proteins and the validation of α-amino group selectivity *via N*-terminal sequencing. Moreover, we demonstrated the long-term stability of our ring-expanded protein bioconjugates. Lastly, we showed the utility of our approach through the successful modification of proteins involved in cancer cell and tissue imaging, as well as human chemokine receptor activation. This work is described in Chapter 5.

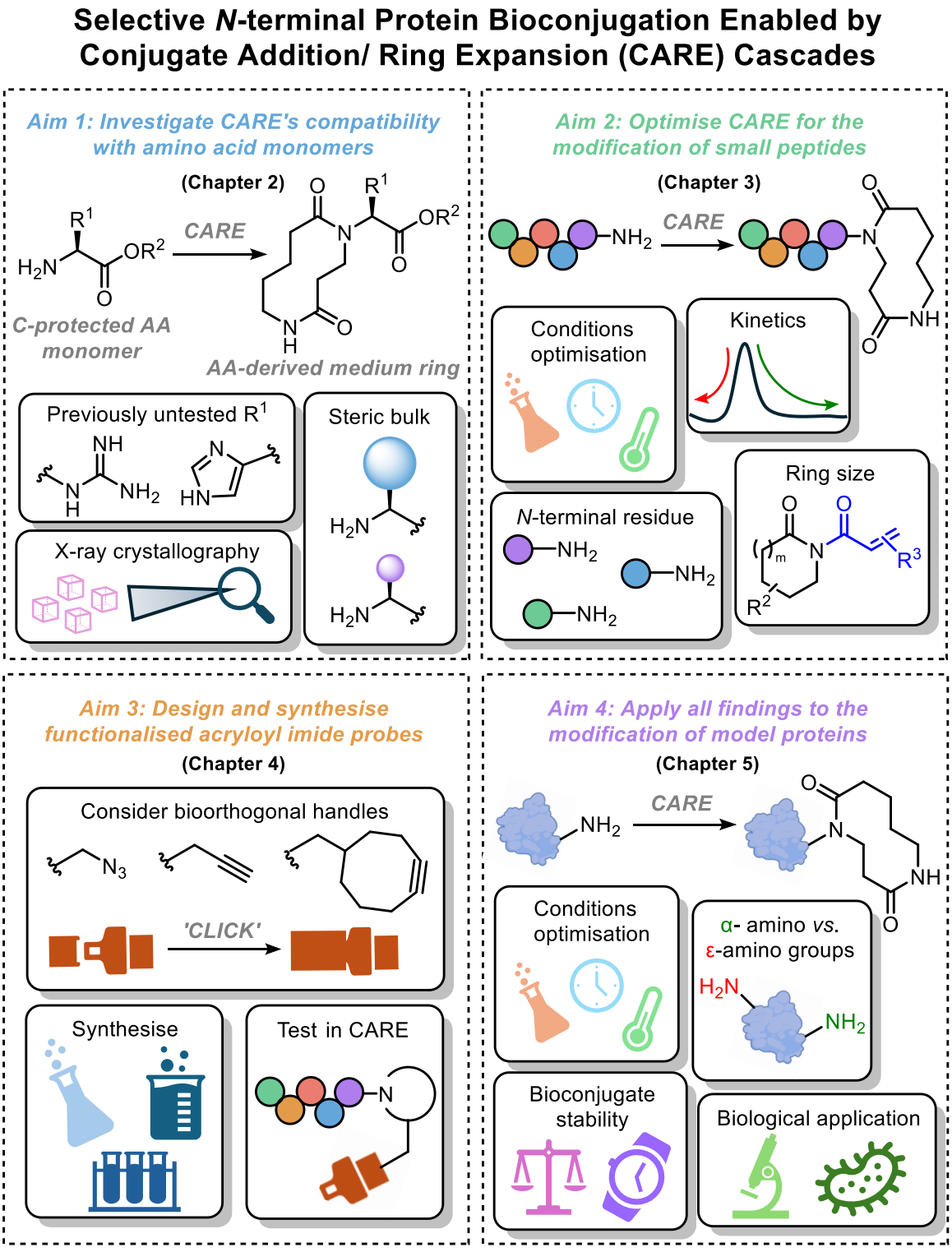
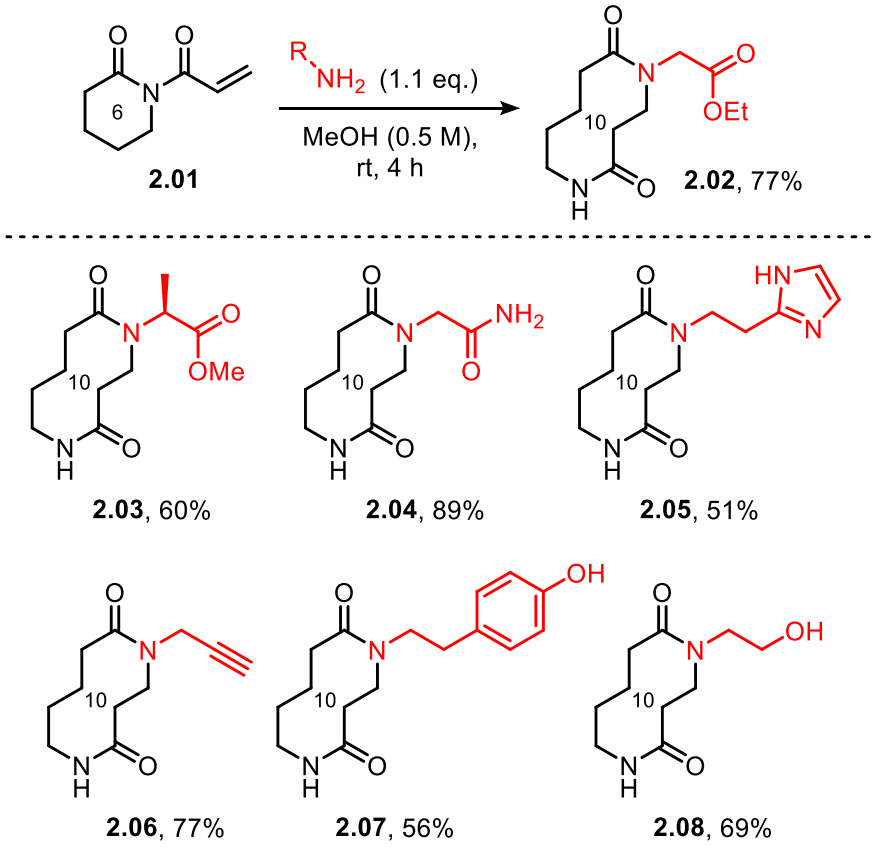


Figure 1.12: Conceptual summary of the 4 main aims of this PhD project.

Chapter 2: Expanding the scope of amino acids used in CARE

2.1 Previous work

As discussed previously in Chapter 1, primary amines derived from the canonical amino acids, glycine and alanine, have been explored as substrates for conjugate addition/ring expansion cascade (CARE) reactions and successfully incorporated into medium-sized ring bis lactams (**2.02** and **2.03,** Scheme 2.1). The robustness of CARE is also exemplified by its compatibility with complex amines containing an array of biologically-relevant functional groups including amides, aza-heterocycles, alkynes and unprotected phenols and alcohols (**2.04**–**2.08**, Scheme 2.1).[50]

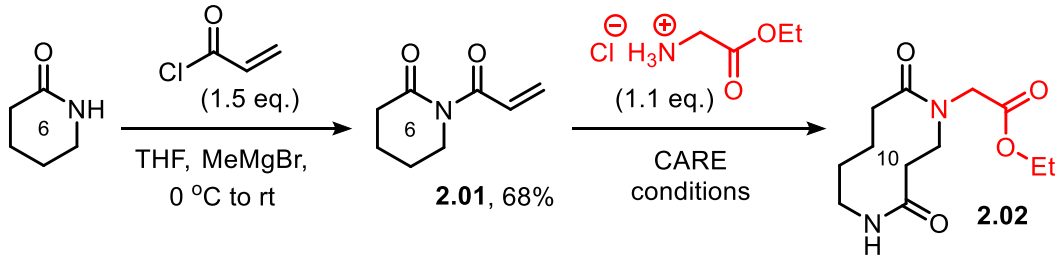


Scheme 2.1: Medium-sized ring bis lactams synthesised *via* CARE between acryloyl imide **2.01** and primary amines containing biologically-relevant functional groups.

Whilst the wide demonstrated scope of CARE with respect to variation of the amine provided reasonable confidence that it would be compatible with other proteinogenic amino acids, this chapter details the efforts undertaken to validate that compatibility. Specifically, the synthesis of 15 10-membered ring bis lactams derived from α-amino acids is described, along with the structural characterisation of these compounds by X-ray crystallography.

2.2 Optimising the existing approach

6-Membered acryloyl imide **2.01** was synthesised using the Unsworth group's optimised *N*-acylation conditions for δ-valerolactam with acryloyl chloride, using MeMgBr in THF at 0 °C, affording the desired product in a respectable 68% yield. With the key electrophilic imide now in hand, we first repeated its CARE reaction with glycine ethyl ester hydrochloride in methanol. To enable the transformation, the amino acid ester hydrochloride salt was first converted into its free base using triethylamine, since the nucleophilicity of the amine is significantly reduced in its protonated form. Carrying out the reaction with 2 equivalents of triethylamine afforded the desired ring-expanded product **2.02** in 76% yield (Entry 1, Table 2.1), which improved to 90% when 4 equivalents of base were used (Entry 2, Table 2.1). As expected, repeating the reaction in the absence of triethylamine resulted in no conversion of the starting materials (Entry 3, Table 2.1).



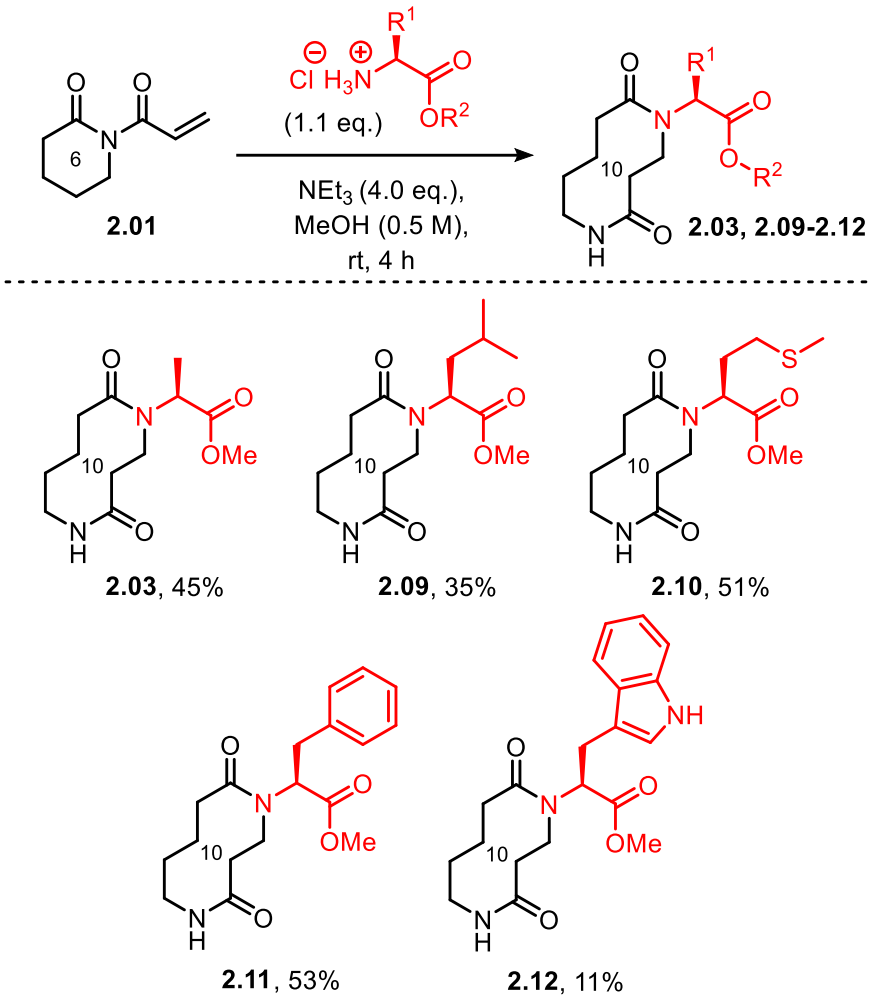
Scheme 2.2: CARE of acryloyl imide **2.01** with glycine ethyl ester hydrochloride.

Table 2.1: Conditions for the CARE reaction of acryloyl imide 2.01 with glycine ethyl ester hydrochloride

Entry	CARE Conditions	Isolated Yield of **2.02** (%)
1	NEt_3 (2.0 eq.), MeOH (0.5 M), rt, 4 h	76
2	NEt_3 (4.0 eq.), MeOH (0.5 M), rt, 4 h	90
3	MeOH (0.5 M), rt, 4 h.	0

2.3 CARE reactions of the hydrophobic amino acids

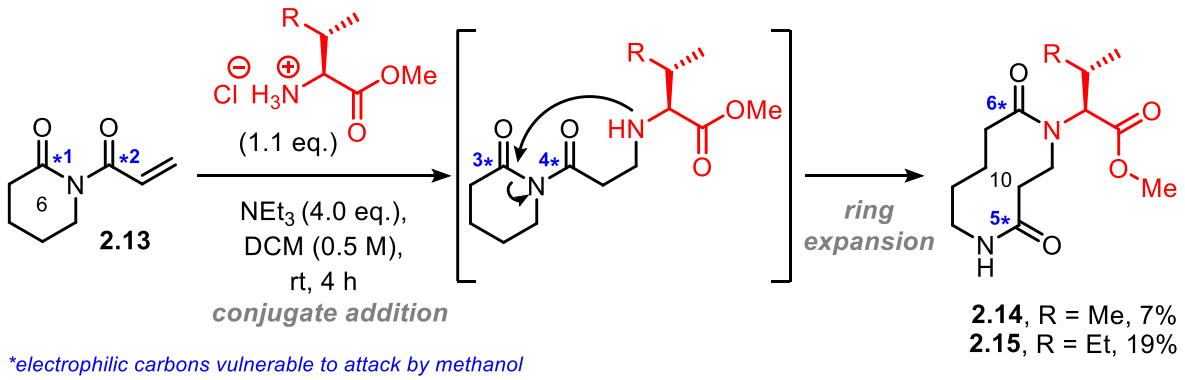
To confirm whether the new reaction conditions would work with substrates beyond amino acid derivatives other than glycine, alanine methyl ester hydrochloride was next tested, and pleasingly, ring-expanded product **2.03** was isolated in 45% yield. Given their simpler chemical functionality, our initial efforts were focussed on testing the compatibility of the hydrophobic amino acids. Ring expanded products **2.09–2.12** were successfully synthesised and isolated from acryloyl imide **2.01** and the corresponding amino acid ester hydrochloride salt using the optimised conditions previously described (Scheme 2.2).



Scheme 2.3: Medium-sized ring products isolated from the CARE reactions of acryloyl imide **2.01** with select hydrophobic amino acids.

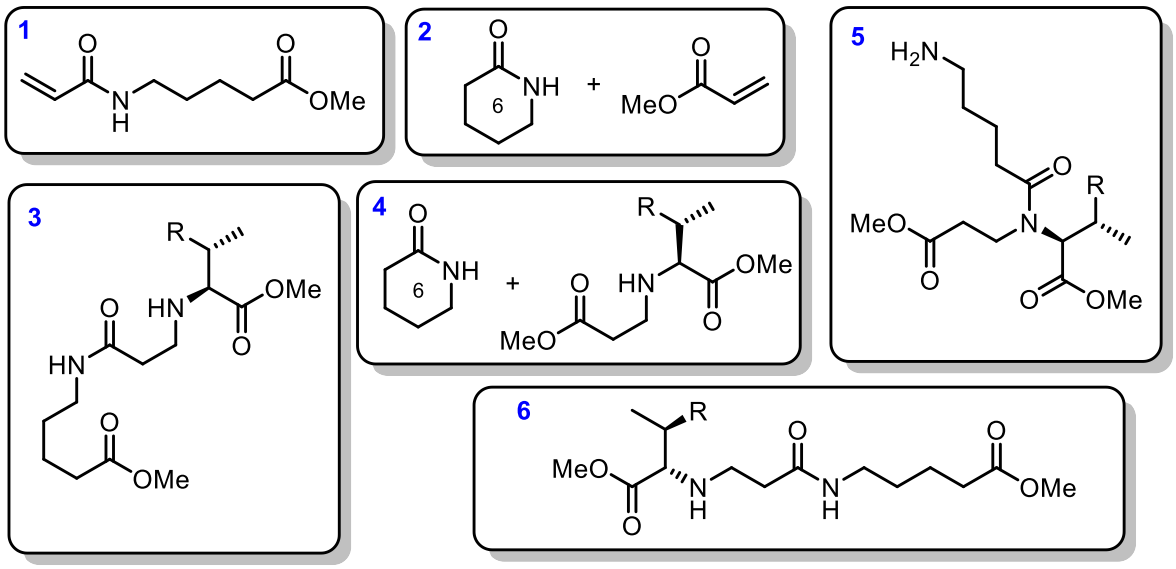
The CARE reactions of valine and isoleucine methyl ester hydrochlorides under these conditions led to the formation of a complex mixture of ring-expanded (**2.14** and **2.15**) and ring-opened products, which could not be separated by FCC. Similar side reactions have been observed previously in CARE reactions, and can be attributed to the use of bulky primary amines, with steric bulk hindering both the conjugate addition and ring-expansion steps.[50] This reduced reactivity promotes methanol-mediated ring-opening and deacylation of the acryloyl imide starting material (Scheme 2.3B). To remove the possibility of side product arising from reaction with the methanol solvent, these reactions were instead done in DCM to afford the target ring-expanded products **2.14** and **2.15** in 7% and 19% yields respectively (Scheme 2.3A).

A) CARE reactions of valine and isoleucine methyl ester hydrochloride



*electrophilic carbons vulnerable to attack by methanol

B) Possible methanol-mediated ring-opened and deacylated products



Scheme 2.4: (**A**) CARE reactions of valine and isoleucine methyl ester hydrochloride with acryloyl imide **2.01** and (**B**) the possible methanol-mediated ring-opened and deacylated byproducts.

Unsurprisingly, even when synthesised in DCM, compounds **2.14** and **2.15** were still some of the lowest yielding and this can be explained by the increased steric bulk around the amino acid's α carbon. This is exemplified by the clear improvement in yield from **2.14** to **2.09** when replacing an *iso*-propyl group with an *iso*-butyl group; this observation suggests that moving the bulky tertiary carbon further from the α carbon reduces its deleterious effect on the reaction outcome.

Our medium-sized ring bis lactam products contain two amide groups within their cyclic structure, each of which can adopt either a *cis* or *trans* conformation. As a result, these compounds can exist in four possible rotameric forms: *cis-cis*, *cis-trans*, *trans-cis*, *trans-trans*. These distinct rotameric forms can be distinguished by NMR; however, their presence often

complicates the analysis due to significant peak broadening. In most cases, the ^{13}C NMR spectrum is better resolved than the ^{1}H NMR spectrum and is therefore used as the primary diagnostic tool to characterise these compounds. 2D-NMR spectroscopic data (DEPT, COSY, HSQC and HMBC) can then be used to help elucidate the complex ^{1}H NMR spectrum. Due to the complexity of the spectra and the structural similarities between isomeric intermediates and the ring expanded product, it can be challenging to confirm whether the isolated material is the desired ring-expanded product or just the initial conjugate addition intermediate. For additional confirmation that the isolated compounds were indeed the ring-expanded products, a crystal of **2.12** was grown and submitted for XRD. As shown in Figure 2.1, the crystal structure of **2.12** confirmed the structure of the ring-expanded lactam, which had previously been suggested by NMR analysis. This result gives us confidence that the other isolated compounds are also the desired ring-expanded products, rather than intermediates from the conjugate addition.

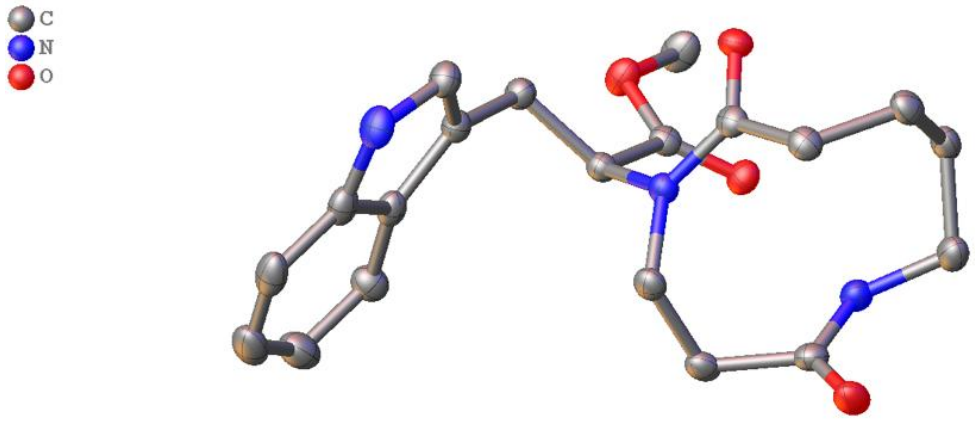


Figure 2.1: X-ray crystal structure of compound **2.12**.

2.4 CARE reactions of the polar uncharged amino acids

Next, we turned our attention to investigating more complex polar uncharged amino acids. The target ring-expanded products **2.16**–**2.19** (Figure 2.2) were synthesised and purified using the modified CARE conditions previously described. Tyrosine, serine and threonine all contain suitably nucleophilic alcohol groups to compete with the aza-Michael addition but in all cases, no products arising from competing oxy-Michael reactions were isolated.

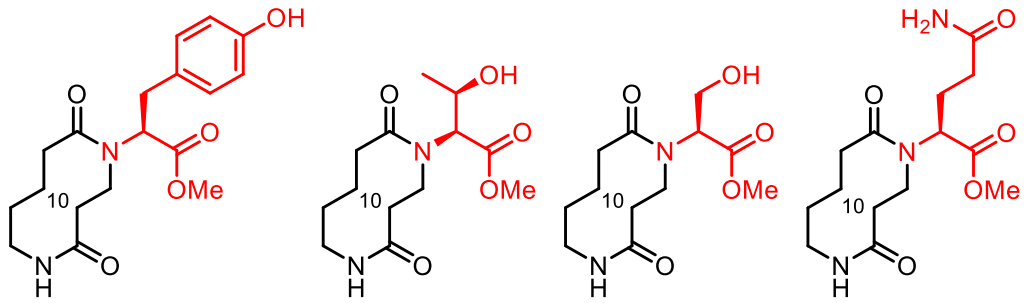


Figure 2.2: Ring-expanded products **2.16**–**2.19** from the polar uncharged amino acids.

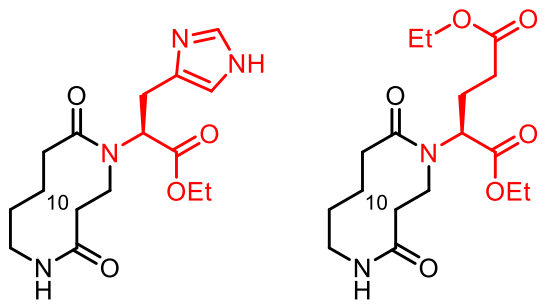
Once again, we observed the effect of increased steric bulk around the α carbon between **2.17** and **2.18** with the additional methyl group clearly reducing the isolated yield significantly. A reaction using an asparagine-derived amine hydrochloride was not attempted due to limited availability of this starting material. However, given its close structural similarity to glutamine methyl ester hydrochloride, we are confident that it would have given a similarly successful outcome.

2.5 CARE reactions of the polar charged amino acids

Lastly, we moved to trying these reactions with the charged amino acids. This group contains an array of chemically diverse side chains, including functional groups such as guanidines and imidazoles, which have yet to be explored for compatibility with the CARE chemistry. Importantly, this group contains lysine whose side chain contains an ε-amino group that presents a significant chemoselectivity issue. *N*-terminal selectivity can be achieved through careful pH control, taking advantage of the slight pK_a difference between the α- and ε-amino groups. This difference has been successfully exploited in numerous *N*-terminal modification chemistries including NHS esters[94,101–103] and 2-ethynylbenzaldehydes.[104] However, to avoid complications arising from this selectivity, lysine was excluded from this initial small-molecule investigation. Site-selectivity in lysine-containing substrates will instead be addressed when our chemistry is tried on peptides later in the project, where such differentiation becomes more relevant and experimentally tractable with careful pH control.

Compound **2.20** was synthesised from **2.01** and histidine ethyl ester hydrochloride and isolated in moderate yield, with no evidence of competing reactivity from the imidazole group in histidine's side chain. The only glutamic acid derivative available to us was L-glutamic acid diethyl ester hydrochloride, in which both carboxylic acid groups are protected as ethyl esters. Whilst this compound does not fully mimic the native form of glutamic acid residues *in vivo*,

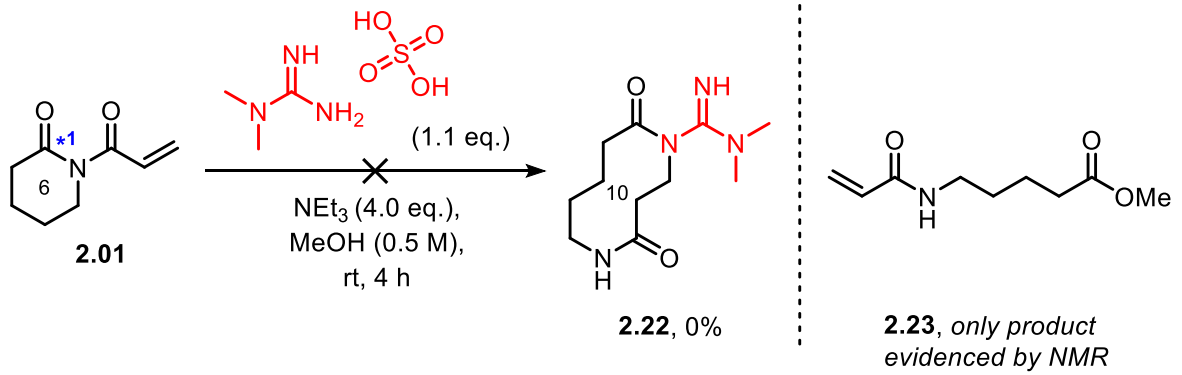
its compatibility with the CARE chemistry was still evaluated and medium-sized ring bis lactam **2.21** was isolated in 37% yield.



2.20, 27% **2.21**, 37%

Figure 2.3: Ring-expanded products **2.20** and **2.21** from the polar charged amino acids.

The other charged amino acids were not easily available in their ester-protected forms and therefore were not included in our small molecule screening. However, given the structural similarity between guanidines — as found in the side chain of arginine residues — and primary amines, we deemed it important to investigate whether these functional groups could participate as substrates in our reaction. Consequently, an analogous reaction to those carried out with the protected amino acids was carried out with *N*, *N*'-dimethylguanidine sulphate to determine if arginine residues could introduce another chemoselectivity issue to our approach (Scheme 2.4). The only product isolated from this reaction was the methanol ring-opened species of the acryloyl imide starting material (**2.23**), as previously discussed. This outcome suggests that there was no reaction between the *N*, *N*'-dimethylguanidine sulphate and acryloyl imide and thus, we concluded that arginine residues were unlikely to cause any significant chemoselectivity issues in subsequent work.

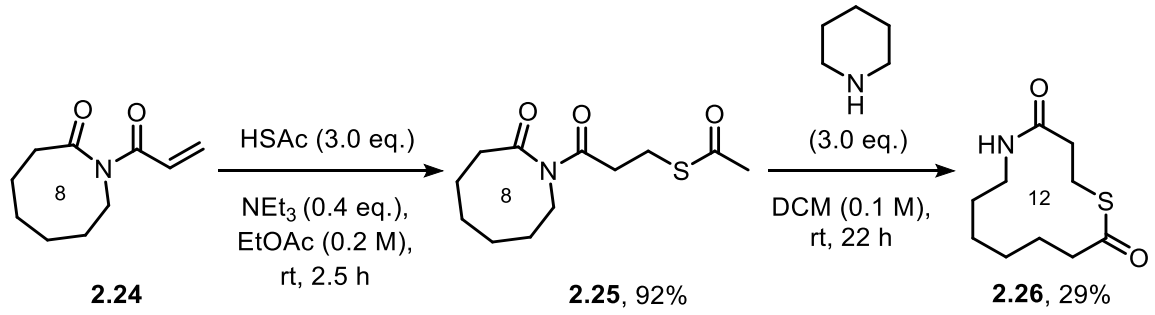


Scheme 2.5: Attempted CARE reaction of acryloyl imide **2.01** with *N*, *N*'-dimethylguanidine sulphate. Asterisk [*1] denotes site of methanol ring-opening to form **2.23**.

2.6 Sulphur conjugate addition of cysteine residues

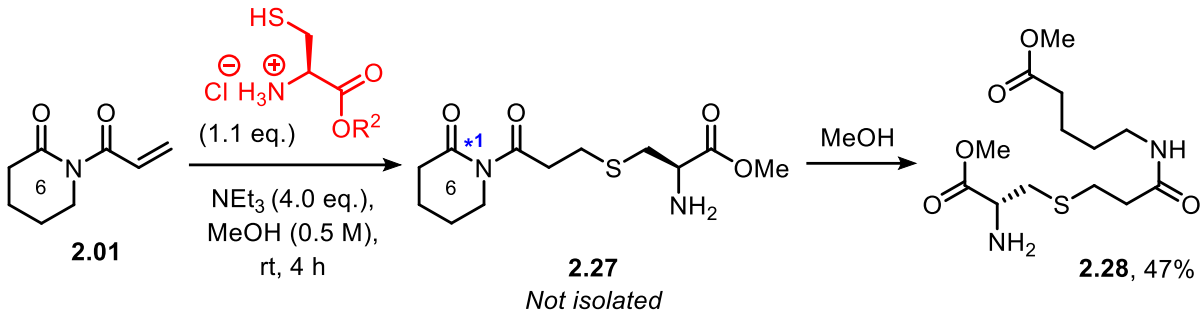
The conjugate addition of sulphur-containing nucleophiles into Michael acceptors has been well-studied and is commonly employed in the synthesis of many biologically active compounds.[105] Thiols are particularly effective at these reactions — thanks to a combination of advantageous electronic and structural effects — meaning they often outcompete other nucleophiles present in the reaction mixture. Thiolates are soft nucleophiles and typically react well with the soft electrophilic ß-carbons of Michael acceptors, as is consistent with the hard and soft (Lewis) acids and bases (HSAB) theory. Furthermore, sulphur-containing nucleophiles often better stabilise the reaction's transition state compared to nitrogen- and oxygen-containing nucleophiles due to the better overlap of its larger 3p orbital with the π* antibonding orbital of the conjugated system.

Previous work in the Unsworth group has shown that acryloyl imides are susceptible to conjugate addition from nucleophilic sulphur-containing reagents such as di- and mono-thioacetic acid (**2.24** → **2.25**, Scheme 2.5). Subsequent thioacetate cleavage triggers the typical ring expansion to afford thiolactones with the general structure **2.26**.[106]



Scheme 2.6: Previously optimised conditions for the conjugate addition of thioacetic acid into acryloyl imide **2.24** followed by thioacetate cleavage and ring expansion to afford thiolactone **2.25**.

With this in mind, we anticipated that the free thiol group of cysteine residues could interfere with our reactions. To confirm this, a usual CARE reaction was carried out using cysteine methyl ester hydrochloride under our modified conditions (Scheme 2.6). This resulted in the formation of a complex mixture of products that could not be fully separated by FCC; however, byproduct **2.28** was isolated from the reaction mixture in 47% yield. Its formation is thought to be a result of thia-Michael addition into **2.01** to form thioether **2.27** followed by methanol-mediated ring-opening to form **2.28**.



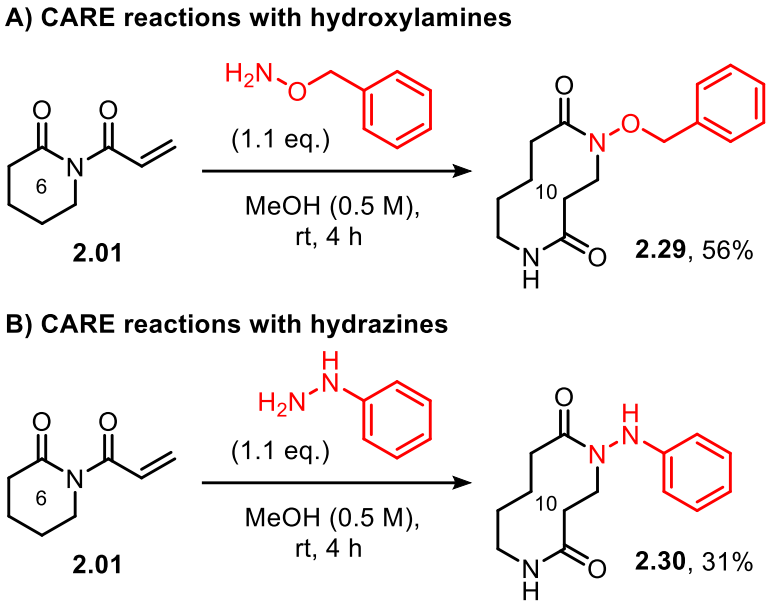
Scheme 2.7: Thia-Michael addition byproduct isolated from a reaction between acryloyl chloride **2.01** and cysteine methyl ester hydrochloride. Asterisk [*1] denotes site of methanol ring-opening to form **2.28**.

The isolated yield of **2.28** suggests that thia-Michael addition is competing with the intended aza-Michael addition in this instance. Therefore, it is reasonable to assume that the other byproducts of this reaction were likely to be a mixture of aza- and thia-Michael-derived products.

2.7 CARE reactions of alternative amine derivatives

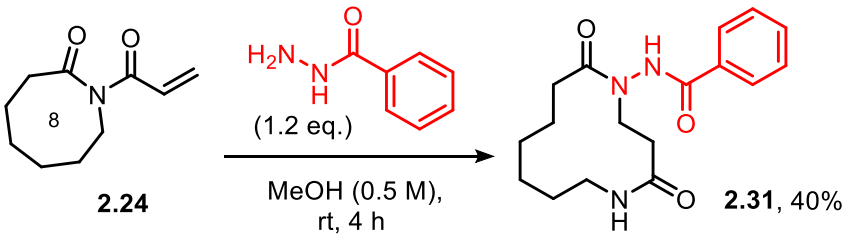
Since CARE reactions are selective for primary amine nucleophiles, there has been little exploration of other potentially compatible substrates. Hydrazines, hydrazides and hydroxylamines were of particular interest due to their structural similarity to primary amines as well as their presence in both therapeutics — such as antituberculars,[107] antidepressants[108] and chemotherapies[109] — and important biosynthetic pathways.[110] Most notably, the formation of oximes and hydrazones through the reactions of hydroxylamines and hydrazines with carbonyl compounds is an essential and widely used strategy for the facile linkage of biomolecules to various probes.[111]

We began by reacting our standard acryloyl imide **2.01** with *O*-benzylhydroxylamine under the standard CARE conditions (Scheme 2.7A). Pleasingly, the desired ring expanded product **2.29** was isolated in 56% yield. An analogous reaction was then successfully carried out using phenylhydrazine (Scheme 2.7B), with the target medium-sized ring **2.30** being isolated in 31% yield. Whilst these isolated yields are less than that obtained for similarly structured primary amines (83% with benzylamine), it may be of interest to employ these functionalities in future work to expand on the potential applications of this ring-expansion chemistry.



Scheme 2.8: CARE reaction of *O*-benzylhydroxylamine (**A**) and phenylhydrazine (**B**) with acryloyl imide **2.01**.

We next turned our attention to investigating hydrazides as substrates in the CARE reaction. Given the electron-withdrawing nature of the adjacent carbonyl group, electron density is delocalised away from the terminal nitrogen, making hydrazides less nucleophilic than primary amines. Therefore, it was unclear whether they would be sufficiently reactive towards the acryloyl imide to initiate the reaction. However, a reaction between acryloyl imide **2.31** and benzhydrazide afforded the expected product in 40% yield (Scheme 2.8).



Scheme 2.9: CARE reaction of benzhydrazide with acryloyl imide **2.24**.

2.8 Chapter summary

To summarise, derivatives of 15 of the canonical amino acids were investigated as amine substrates in the CARE methodology to form a series of functionalised medium-sized ring bis lactams. Whilst the isolated yields for these reactions were varied (7–89%), no competing reactions with the amino acid side chains was observed in all but one case — the notable exception being cysteine methyl ester hydrochloride. We postulate that the variability in yield is due to the reduced nucleophilicity of amino acids bearing large or branched side chains, due

to steric crowding. This aligns well with the trend reported from the findings of the published CARE primary amine scope, where bulky amines are less effective.[50] Furthermore, we have confirmed that other biologically-important amine derivatives — such as hydrazines, hydrazides and hydroxylamines — are all compatible with this chemistry. Whilst these functional groups will not be explored further as a part of this research, they stand as promising avenues for investigation within the group. The findings from this study laid the groundwork for the rest of this project, where we next optimised the CARE chemistry for the modification of model peptides.

Chapter 3: Optimising the existing CARE approach for the modification of model peptides

3.1 Rethinking CARE: a strategy for site-selective bioconjugation

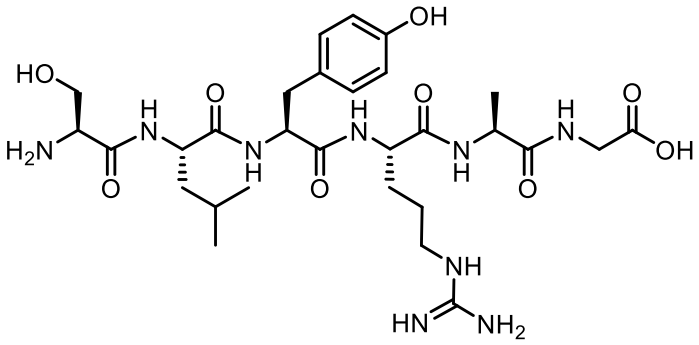
Having successfully confirmed the compatibility of over half of the canonical amino acids as CARE substrates, the next step was to optimise the chemistry for the modification of model peptides, an important step in developing this novel protein bioconjugation strategy. Although CARE was initially designed and optimised for the synthesis of medium-sized rings and macrocycles, it also meets several key criteria for bioconjugation application. More specifically, it has proven compatibility in both water and DMSO—solvents commonly employed in protein modification—and exhibits selectivity for primary amines, as discussed and exemplified in Chapter 2.[50] Furthermore, the irreversible ring-expansion step of CARE addresses one of the major limitations of many established bioconjugation methods by driving the equilibrium towards complete conversion of the conjugate addition intermediate to ring-expanded product, whilst preventing any reversion to the starting material. Importantly, this feature should eliminate the need for more forcing conditions to drive conversion, thereby decreasing the levels of off-target modifications at other residues and the formation of heterogeneous bioconjugates.[112,113]

This Chapter describes the synthesis of model peptides and their use as substrates in a CARE cascade reaction for the assembly of medium-sized rings and macrocycles at their *N*-termini. Specifically, we established a set of optimal reaction conditions, conducted a detailed investigation into the reaction kinetics, and evaluated the stability of the resulting conjugates. Furthermore, we examined the ε-amino group of lysine residues as a potential competing reaction site. The development of our approach—from conceptualisation to completion—is described herein.

3.2 SLYRAG: a model for method development

Peptides are commonly used as protein surrogates in the development of novel bioconjugation methods as their significantly lower molecular weight often makes them easier to handle, modify, and characterise compared to fully sized proteins. Moreover, peptides are much simpler to synthesise and customise *via* solid phase peptide synthesis (SPPS) and, therefore, different sequence variations can be more efficiently screened before trying on more complex macromolecular structures. We chose to use well established 6-mer SLYRAG **3.01** as a testbed for our initial optimisation studies due to its favourable solubility and stability, as well as its extensive prior use within our groups.[84,114–116] In addition, its internal tyrosine residue provides

a UV-active chromophore to aid detection by HPLC-MS. The peptide was assembled from preloaded Fmoc-Gly-2-Chlorotrityl resin *via* Fmoc-SPPS before cleavage in a trifluoroacetic acid (TFA), triisopropylsilane (TIPS) and water cocktail and purification by reversed phase automated column chromatography.



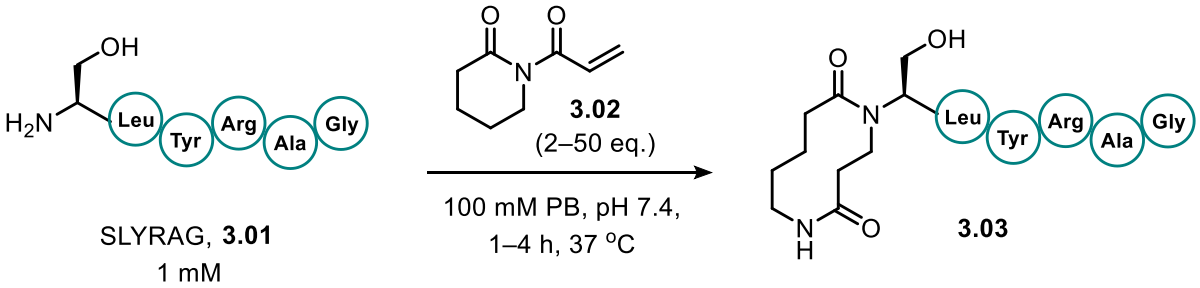
SLYRAG, **3.01**
$[M+H]^+$ = 666.36 Da

Figure 3.01: The structure of 6-mer SLYRAG **3.01**.

3.3 Optimising the CARE reaction on SLYRAG

3.3.1 Establishing optimal reaction conditions

We began by reacting SLYRAG (**3.01)** with 2 molar equivalents of 6-membered acryloyl imide, **3.02** in pH 7.4 phosphate buffer at 37 °C. After 4 hours, only 2% conversion to *N*-terminally modified SLYRAG derivative, **3.03** was observed by LC-MS (Entry 1, Table 3.01). We responded to this poor result by increasing the number of equivalents of **3.02** up to 50 molar equivalents, which, pleasingly, led to ≥95% conversion to **3.03** (Entry 6, Table 3.01). Mindful that large excesses of reagent can lead to off-target modifications on proteins, we aimed to reduce the number of equivalents used whilst still achieving complete modification within a reasonable reaction time. Table 3.01 details the full set of conditions screened in the optimisation of CARE with acryloyl imide **3.02** on SLYRAG (**3.01**).



Scheme 3.01: CARE of acryloyl imide **3.02** with SLYRAG **3.01** under varying conditions.

Table 3.01: Effect of imide **3.02** equivalents and reaction time on the conversion of SLYRAG (**3.01**) to *N*-terminally modified derivative **3.03**.

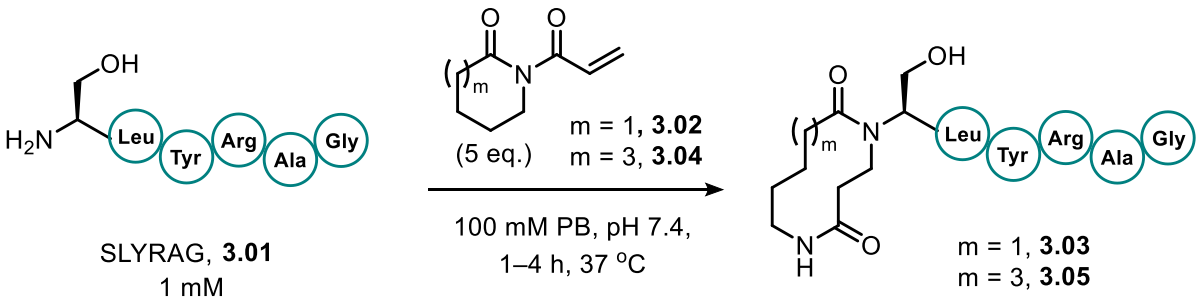
Entry	Equivalents of **3.02**	Time (h)	Conversion to **3.03** (%) [a]
1	2	4	2
2	5	1	73
3	5	2	93
4	5	3	≥95
5	10	4	≥95
6	50	4	≥95

[a] Conversion was determined by LC-MS through integration of the UV-280 nm absorbance trace. For further details, see Chapter 7.3.3.1.

Adequate or complete conversion could be achieved when using as little as 5 molar equivalents of **3.02** *e.g.* 73% after 1 h, 93% after 2 h and complete conversion by 3 h (Entries 2–4, Table 3.01). Pleased with these results, we turned our attention to investigating the effect of ring size on the transformation.

3.3.2 Investigating the effect of ring size

It has long been known that the ring size of the starting material plays a crucial role in the outcome of ring-expansion reactions—especially when the expansion takes the ring into a new size category, most notably when expanding from a "normal"-sized ring (5–7 members) into a medium-sized ring (8–11 members), or from a medium-sized ring into a macrocycle (12+ members).[117] Previous work within the Unsworth group has proven that "normal-to-medium" ring-expansions are generally the most difficult due to the "uphill" thermodynamic expense arising from the destabilising transannular interactions and ring strain of medium-sized ring systems.[118] Our typical "6-to-10" expansion (as shown in Scheme 3.01) falls within this category and relies on the formation of a more stable lactam to drive the equilibrium towards ring-expansion. In contrast, an "8-to-12" expansion benefits not only from this functional group exchange as a thermodynamic driving force, but also from an additional contribution associated with the release of the medium-sized ring's transannular ring strain upon becoming macrocyclic. Therefore, medium-sized ring acryloyl imide starting materials often work better than analogous normal-sized ring acryloyl imide starting materials. Considering this, we synthesised 8-membered acryloyl imide **3.04** in a moderate 52% yield and conducted CARE reactions with SLYRAG (**3.01**) using both **3.04** and the 6-membered imide **3.02**, to determine how significantly the starting material's ring size affects the outcome of these bioconjugations (Scheme 3.02).



Scheme 3.02: CARE of acryloyl imides **3.02** and **3.04** with SLYRAG.

Table 3.02: Effect of ring size (**3.02** vs. **3.04**) on the conversion of SLYRAG **3.01** to *N*-terminally modified SLYRAG derivatives, **3.03** and **3.05**.

Entry	Imide	Product	Time (h)	Conversion (%) [a]
1	**3.02**	**3.03**	1	73
2	**3.02**	**3.03**	2	93
3	**3.02**	**3.03**	3	≥95
4	**3.02**	**3.03**	4	≥95
5	**3.04**	**3.05**	1	55
6	**3.04**	**3.05**	2	≥95
7	**3.04**	**3.05**	3	≥95
8	**3.04**	**3.05**	4	≥95

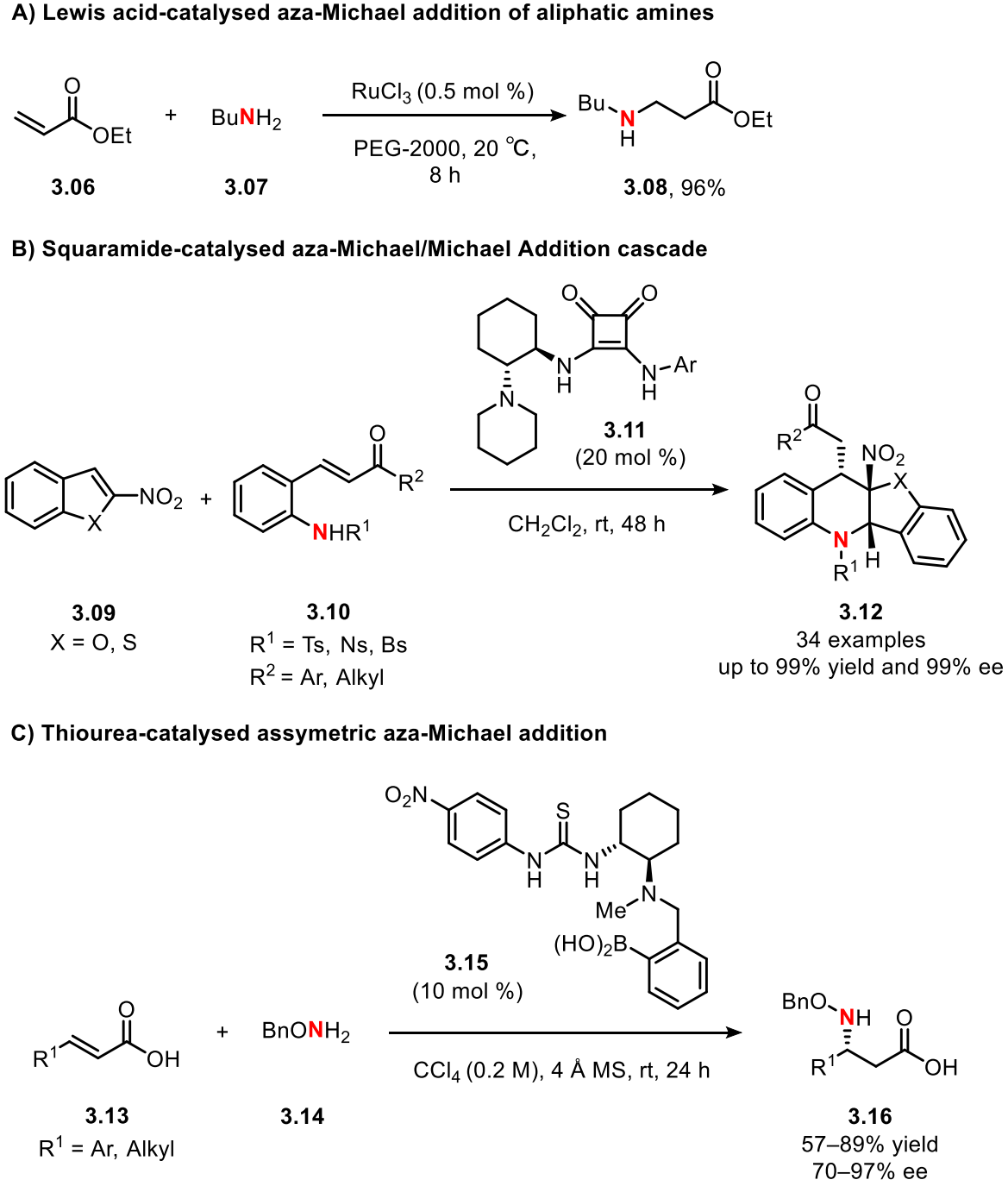
[a] Conversion was determined by LC-MS through integration of the UV-280 nm absorbance trace. For further details, see Chapter 7.3.3.1.

Table 3.02 indicates that whilst ring size does not significantly impact the final outcome of these reactions. The initial conjugate addition step proceeds marginally more slowly with **3.04** than with **3.02** (Entries 1 & 5), but both imides worked well. These findings demonstrate that the ability to assemble relatively complex, functionalised medium-sized rings and macrocycles at SLYRAG's *N*-terminus is possible using different acryloyl imide probes.

3.3.3 Exploring the use of organocatalysts

The aza-Michael reaction has become increasingly fundamental to the synthesis of carbon-nitrogen bonds—such as the β-amino carbonyl motif found in numerous pharmaceuticals and natural products—under relatively mild conditions.[119,120] Considerable effort has been devoted to improving the efficiency of these reactions by using organocatalysts to increase the electrophilicity of the Michael acceptor towards nucleophilic attack. Transition metal salts and hydrogen bond catalysts, such as squaramides (**3.11**) and thioureas (**3.15**), have proven

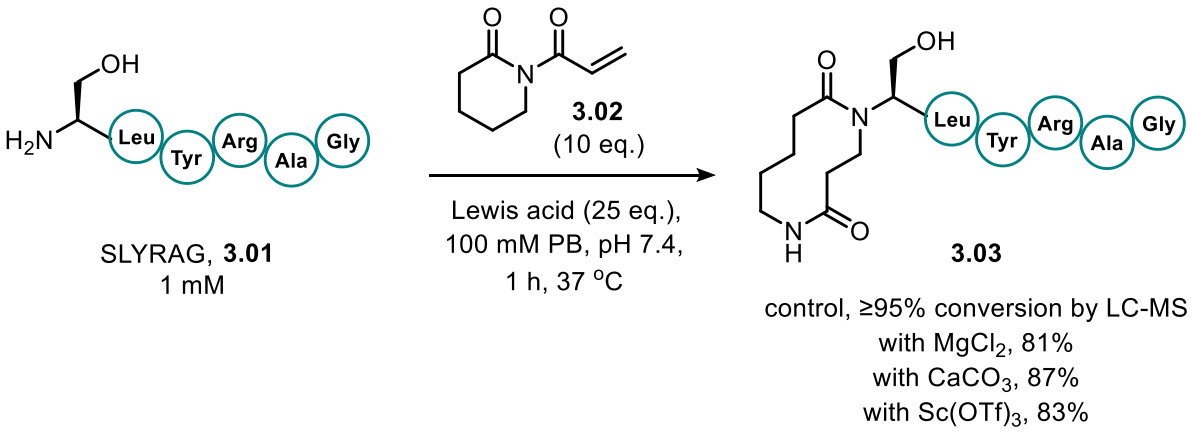
particularly effective in this regard, especially for aza-Michael additions into β-substituted Michael acceptors (Scheme 3.03).[121–123]



Scheme 3.03: Organocatalysed aza-Michael addition of amines into α, β-unsaturated carbonyl compounds with Lewis acids (**A**), squaramides (**B**) and thioureas (**C**).

With this in mind, we sought to investigate whether the addition of an appropriate organocatalyst into the reaction mixture could enhance the rate of the initial aza-Michael addition step. To explore this, we selected a range of Lewis acids and a thiourea catalyst already

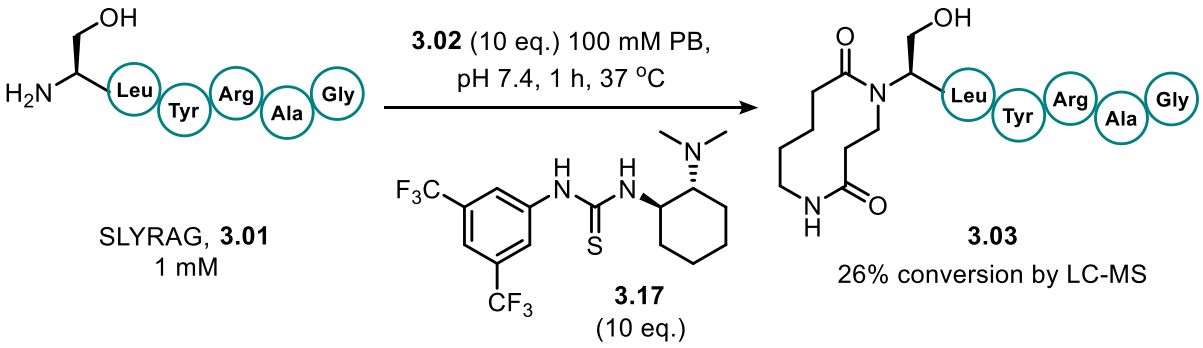
available within our inventory. Additionally, we identified a squaramide catalyst of interest that required a routine synthesis from readily available starting materials. SLYRAG (**3.01**) was first reacted with acryloyl imide **3.02** under the previously established conditions, with the addition of 25 molar equivalents of a Lewis acid−$MgCl_2$, $CaCO_3$, or $Sc(OTf)_3$ (Scheme 3.04).



Scheme 3.04: Investigating Lewis acids as organocatalysts for the CARE of acryloyl imide **3.02** with SLYRAG (**3.01**). Samples were incubated under the specified conditions before conversion was determined by LC-MS (details in Chapter 7.3.3.2).

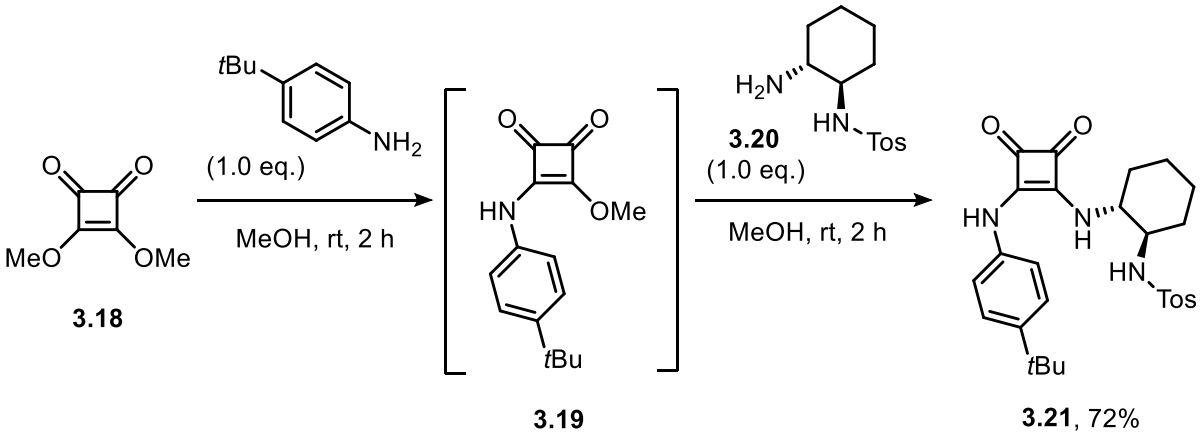
As shown in Scheme 3.04, the addition of a Lewis acid to the reaction mixture had no beneficial effect on the outcome of the reaction and instead led to a modest decrease in conversion to **3.03**. However, due to the very limited scope of Lewis acids tried in this study, we can only tentatively conclude that their use has little effect on the reaction. A broader investigation would be required before confirming this finding with confidence.

The reaction was then repeated with the addition of 25 molar equivalents of thiourea **3.04** (Scheme 3.05). In this case, there was a clear detrimental effect on the reaction, with only 26% conversion of the starting material **3.01** to the *N*-terminally modified derivative, **3.03**. No additional peaks were observed in the LC-MS trace of the crude reaction mixture, suggesting that the thiourea is negatively impacting the CARE reaction, but not by causing the formation of an unwanted byproduct.



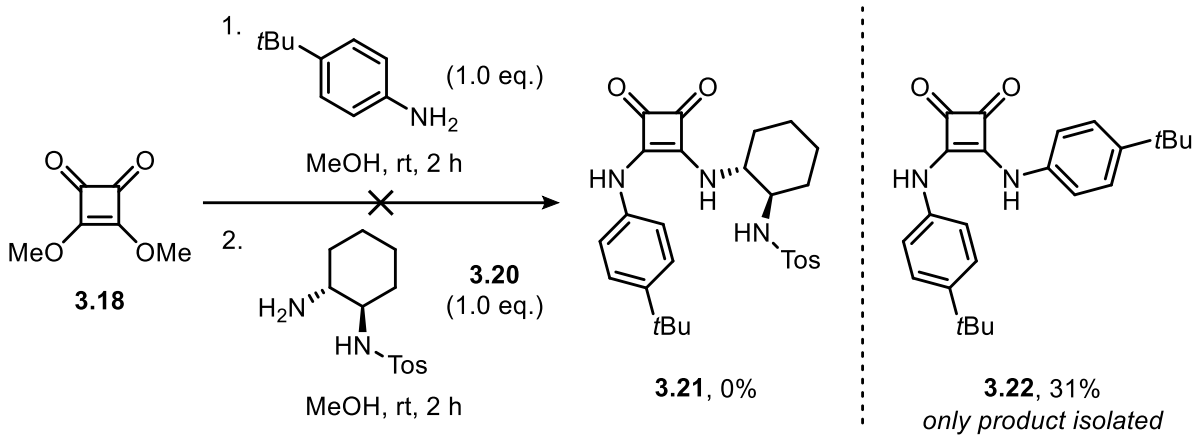
Scheme 3.05: Investigating a thiourea as an organocatalyst for the CARE of acryloyl imide **3.02** with SLYRAG **3.01**. Samples were incubated under the specified conditions before conversion was determined by LC-MS (details in Chapter 7.3.3.2).

Squaramide **3.21** and some structural derivatives have been reported as efficient and enantioselective organocatalysts for the Henry (nitroaldol) reaction, allowing the synthesis of chiral β-amino alcohols. Functionalised squaramides can contain both hydrogen-bond donors and hydrogen-bond acceptors that create a complex network of non-covalent interactions with the reagents and substrate *in situ*, significantly improving the reactivity and stereoselectivity of the transformation. A derivative of squaramide **3.21** was identified as the most effective organocatalyst in their screen.[124]



Scheme 3.06: Literature procedure for the one-pot synthesis of squaramide **3.21**.

Scheme 3.06 shows the reported one-pot synthesis of **3.21**, a faster and more efficient alternative to the traditional two-step synthesis of squaramides.[125] However, attempts to repeat this procedure were unsuccessful, affording only the symmetrical squaramide **3.22** in 31% yield, resulting from the double substitution of 4-*tert*-butyl aniline into **3.18** (Scheme 3.07).

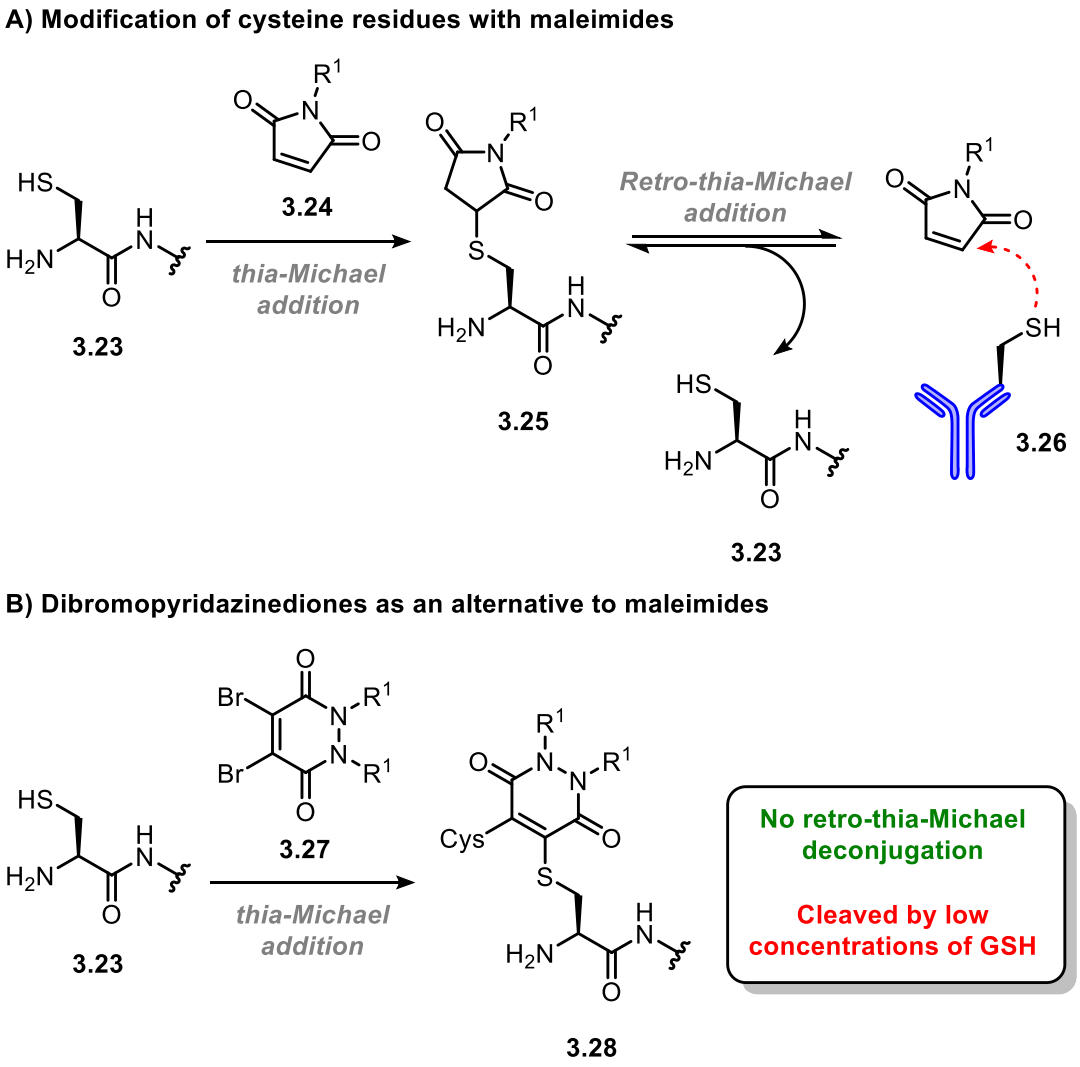


Scheme 3.07: Unsuccessful attempts at repeating the synthesis of squaramide **3.21**.

Given the discouraging results obtained with the other organocatalysts, the investigation of additives to enhance the aza-Michael reaction was concluded at this stage. Since the CARE reaction was already successful in the absence of a catalyst, further optimisation in this area was deemed a lower priority to other aspects of the project.

3.4 Determining the stability of CARE conjugates

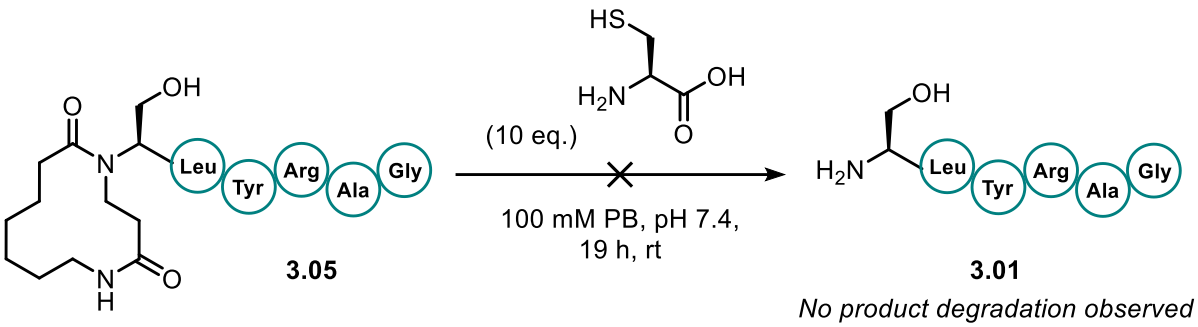
Whilst significant progress has been made in developing novel methods for *N*-terminal bioconjugation, many widely used approaches still suffer from conjugate instability, which limits their broader applications in chemical biology.[63] Chemical bioconjugation can introduce structural and environmental changes that affect stability; however, sometimes instability can arise from an inherent vulnerability of the chemistry itself. Maleimides (**3.24**) have been extensively studied as reagents for rapid and selective cysteine modification in numerous applications (**3.23** → **3.25**, Scheme 3.08A)[126], but it is commonly reported that deconjugation *via* retro-thia-Michael addition can occur (**3.25** → **3.26**, Scheme 3.08A).[127] Although methods to obviate this issue have been reported, the limited research available makes it difficult to substantively conclude their effectiveness, and each approach presents its own set of advantages and limitations. For example, dibromopyridazinediones (**3.27**) have been identified as potential alternatives to maleimides for rapid, selective cysteine bioconjugation. The presence of a leaving group on the vinylic bond in the pyridazinedione scaffold means that a retro-thia-Michael stable unsaturated thio-succinimide linkage is formed upon cysteine conjugation (**3.23**→**3.28**, Scheme 3.08B). The resulting conjugates are also more hydrolytically stable than those obtained from maleimides and can undergo di-addition, making them particularly useful for disulfides. However, they are still susceptible to thiol-exchange and GSH-mediated cleavage *in vivo*.[128]



Scheme 3.08: (A) Thia-Michael and retro-thia-Michael addition of maleimides at cysteine residues. (B) Retro-thia-Michael-stable dibromopyridazinediones as alternative reagents for cysteine bioconjugation.

It is not only cysteine bioconjugates that suffer from instability. In fact, a recent study from the Spicer group found that several *N*-terminal modification chemistries afforded protein conjugates that were unstable under a range of different conditions (*i.e.* various pHs and temperatures). Moreover, the study also found that the stability was not only determined by the chemistry used, but it also depended on the protein being modified. This suggests that there is a strong protein-dependence of conjugate stability—which, perhaps, is not surprising given the same well-known protein-dependence of conjugation efficiency. Interestingly, they also reported that there was little difference in the decrease of protein conjugation between the addition of 1000 or 2000 molar equivalents of a competing α-amino group, indicating that the instability is often due to a dissociative method of cleavage rather than by direct substitution with competing *N*-termini.[63]

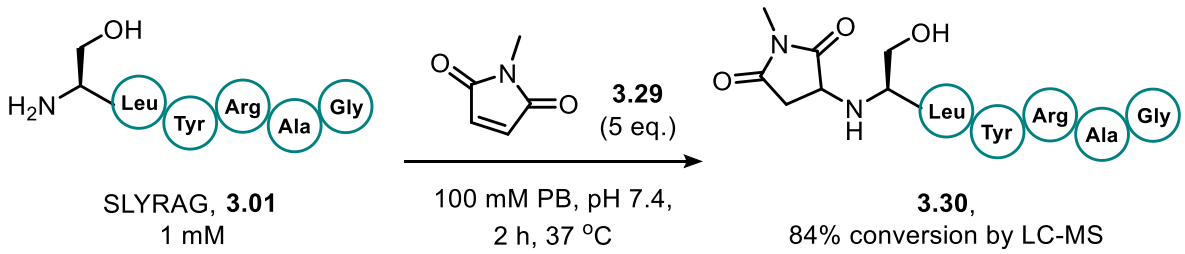
In a similar fashion, we tested the conjugate stability of *N*-terminally modified SLYRAG derivative, **3.05**, by incubating a purified sample in the presence of 10 molar equivalents of cysteine at room temperature for 19 hours (Scheme 3.09). If the ring-expansion was unexpectedly reversible under these conditions, we would expect to observe product degradation over time, as the cysteine thiol would trap the reformed Michael acceptor (**3.05** → **3.01**, Scheme 3.09). Notably, no such degradation was observed; instead, **3.05** was fully retained, demonstrating that CARE conjugates are stable and irreversible under these conditions.



Scheme 3.09: Investigating the stability of peptide conjugate **3.05** over time. Samples were incubated under the specified conditions before level of conjugation was determined by LC-MS (details in Chapter 7.3.3.3).

3.5 Benchmarking CARE against conventional Michael acceptors

To emphasise the significance of this stability, we compared our acryloyl imides with a well-established Michael acceptor used for protein bioconjugation. SLYRAG **3.01** was reacted with 5 equivalents of *N*-methylmaleimide (**3.29**), resulting in 84% conversion to the *N*-terminally modified derivative **3.30** after 2 hours—which is noticeably lower than the conversion achieved with our CARE approach under the same conditions (Scheme 3.10).

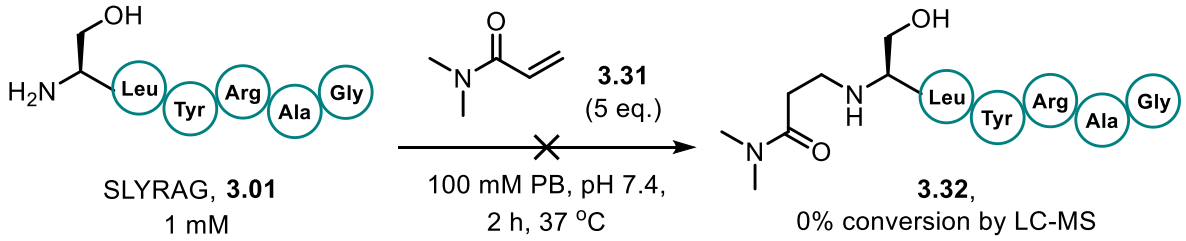


Scheme 3.10: *N*-methylmaleimide as a comparative Michael acceptor for the modification of SLYRAG **3.01** (details in Chapter 7.3.3.4).

An equivalent conjugate stability assay was performed on **3.30** by incubation with 10 molar equivalents of cysteine under the conditions previously described (REF Scheme 3.09). After 19 hours, a 24% reduction in the amount of **3.30** was observed, along with a corresponding

increase in unmodified SLYRAG **3.01**, indicating reversibility of the aza-Michael step and/or instability of the resulting bioconjugate. This finding testifies to the importance of the irreversible ring-expansion step in the CARE reaction, which clearly provides enhanced stability to the bioconjugate.

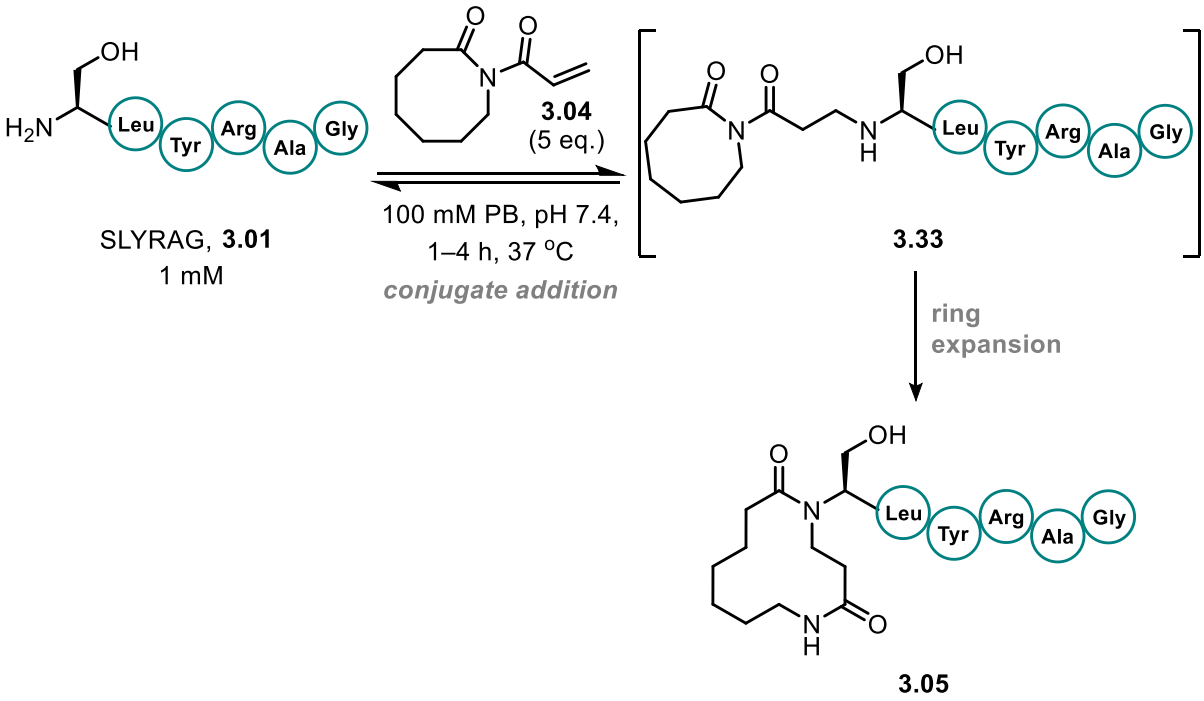
Moreover, the reaction of SLYRAG **3.01** with *N*, *N*-dimethylacrylamide (**3.31**) failed to show any conversion, and the *N*-terminally modified SLYRAG derivative **3.32** was not observed by LC-MS, demonstrating that the increased reactivity of acryloyl imides relative to acrylamide electrophiles is pertinent to the reaction outcome (Scheme 3.11).



Scheme 3.11: *N*, *N*-dimethylacrylamide as a comparative Michael acceptor for the modification of SLYRAG **3.01** (details in Chapter 7.3.3.4).

3.6 Kinetic analysis of the CARE reaction on SLYRAG

Upon closer examination of the LC-MS data for the reaction between SLYRAG **3.01** and imide **3.04**, we identified an additional peak in the chromatogram with the same mass as the expected product. This peak was observed only at earlier time points and gradually decreased in intensity as the reaction progressed, disappearing completely by the time the reaction was complete. We attributed this peak to the intermediate aza-Michael product **3.33**, which subsequently undergoes ring-expansion to form the product **3.05** (Scheme 3.12).

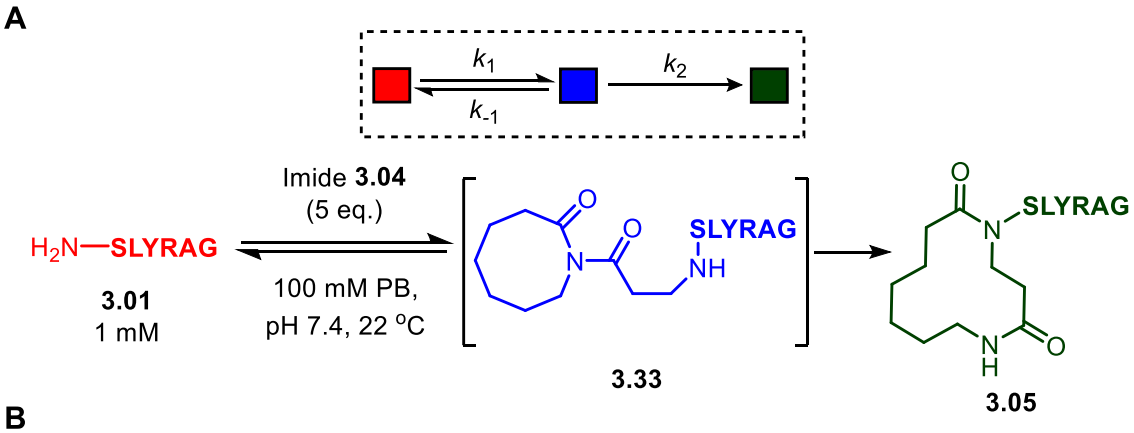


Scheme 3.12: CARE of imide **3.04** with SLYRAG **3.01**.

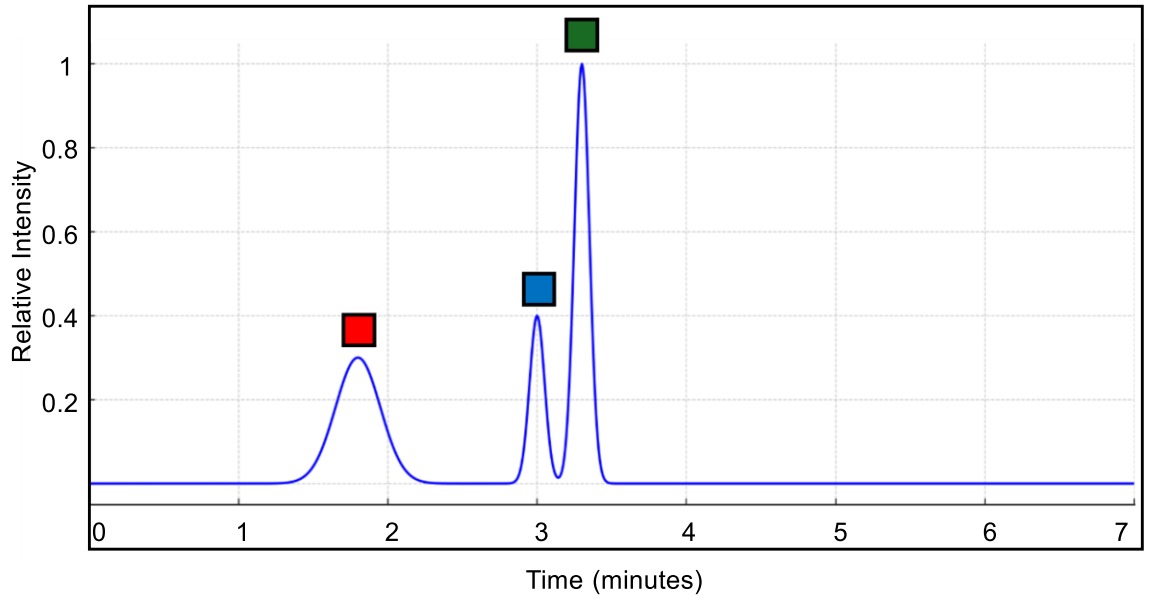
We realised that a molecular species concentration could be calculated by LC-MS using the area of the peak (AOP) of the UV-280 nm absorbance trace and the respective initial concentration of each species (Equation 3.01). At this wavelength, the absorbance of tyrosine's aromatic ring can be monitored, and since it remains unaffected by our CARE modifications, it provides a stable and reliable signal for tracking the concentration of each peptide. Therefore, we carried out a CARE reaction between imide **3.04** and SLYRAG (**3.01)** in an LC-MS vial, recording measurements at 15-minute intervals to monitor changes in the concentration of each species over time (Figure 3.02A & B). Using the concentrations of the starting material **3.01**, the aza-Michael intermediate **3.33**, and the ring-expanded product **3.05**, we employed COPASI—a biochemical modelling and simulation application—to fit our experimentally-calculated concentration changes to a two-step kinetic computational model. From this, we were able to extract rate constants for both the conjugate addition (k_1) and ring-expansion (k_2) steps (Figure 3.02C). Importantly, these rate constants illustrate the importance of the irreversible ring-expansion step in this transformation. In the absence of this step, a reaction governed solely by the equilibrium between aza-Michael and retro-aza-Michael addition would be expected to reach only ~16% conversion despite the reasonably high equilibrium constant.

$$\frac{AOP\ (\mathbf{3.05})}{AOP\ (\mathbf{3.01}) + AOP\ (\mathbf{3.33}) + AOP\ (\mathbf{3.05})} \times 0.001 = Concentration\ of\ \mathbf{3.05}\ (M) \qquad \textit{Equation 3.01}$$

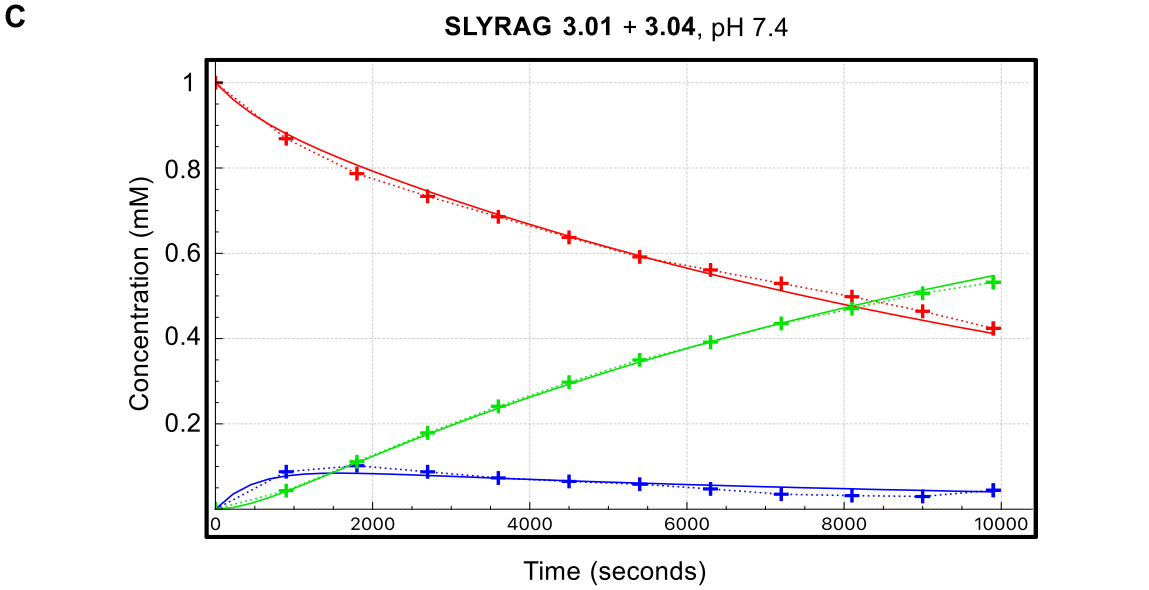
A



B



C



Entry	pH	Conversion after 3 h (%)	k_1 ($M^{-1}s^{-1}$)	k_{-1} (s^{-1})	k_2 (s^{-1})
1	7.4	53	0.040 ± 0.005	$9.7 \times 10^{-4} \pm 2.6 \times 10^{-4}$	$9.0 \times 10^{-4} \pm 2.0 \times 10^{-4}$

Figure 3.02: Kinetic analysis of CARE cascade with acryloyl imide **3.04** and SLYRAG (**3.01)**. (A) Reaction scheme (details in Chapter 7.3.3.5); (B) Illustrative LC-MS chromatogram used to monitor reaction progress; (C) A plot of molecular species concentration against time under these conditions.

3.6.1. Confirming the effect of pH on reaction kinetics

To investigate the effect of pH on the reaction kinetics, we conducted a series of identical LC-MS experiments but in phosphate buffers across a range of biologically compatible pHs. The results shown in Table 3.03 demonstrate the decrease in rate of aza-Michael reactions and an increased rate of retro-aza-Michael reactions at lower pHs (*e.g.* pH 6.5), consistent with the expected increase in amine protonation at acidic pHs.

Table 3.03: Effect of pH on the reaction kinetics between SLYRAG **3.01** and acryloyl imide **3.04**.

Entry	pH	Conversion after 3 h (%) [a]	k_1 ($M^{-1}s^{-1}$)	k_{-1} (s^{-1})	k_2 (s^{-1})
1	6.5	28	0.027±0.02	0.001±0.001	$6.6x10^{-4}$±$0.7x10^{-4}$
2	7.4	53	0.040±0.005	$9.70x10^{-4}$±$2.6x10^{-4}$	$9.0x10^{-4}$±2.0 $x10^{-4}$
3	8.0	58	0.023±0.003	$1.20x10^{-4}$±$1.0x10^{-4}$	0.001±$5.0x10^{-4}$

[a] Conversion was determined by LC-MS through integration of the UV-280 nm absorbance trace. For further details, see Chapter 7.3.3.5.

Whilst the results indicate a slight preference for reactions at pH 8.0, all further experiments were performed at pH 7.4 to maximise *N*-terminal selectivity, particularly in anticipation of competing ε-amino groups in more complex biomolecules.

3.6.2. Exploring the influence of the *N*-terminal residue on CARE

3.6.2.1. Synthesising *N*-terminal SLYRAG mutants

We next focussed on determining the effect of the *N*-terminal residue on the reaction kinetics. Until this point, all peptidyl experiments had been conducted on SLYRAG, which has an *N*-terminal serine residue. To assess the compatibility of other residues, we synthesised a few SLYRAG variants, each with different *N*-terminal amino acid residues. To do this, we synthesised a batch of LYRAG (**3.34**) *via* SPPS using preloaded Fmoc-Gly-2-chlorotrityl resin. The LYRAG-loaded resin was then divided into separate portions, and each *N*-terminal residue was manually coupled using standard peptide coupling conditions with 2-(6-chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HCTU) and *N, N*-diisopropylethylamine (DIPEA) in DMF (Figure 3.03B). Peptides were cleaved from the resin using the standard TFA/TIPS/H_2O cleavage cocktail then precipitated in cold diethyl ether, resuspended in Milli-Q water, and lyophilised. The identity of each *N*-terminal mutant

was confirmed by LC-MS. Where purity was sufficient, the crude peptides were used directly in subsequent kinetics experiments, whilst in cases of significant contamination, purification was performed using reversed-phase automated column chromatography. The peptides were stored as powders at −20°C.

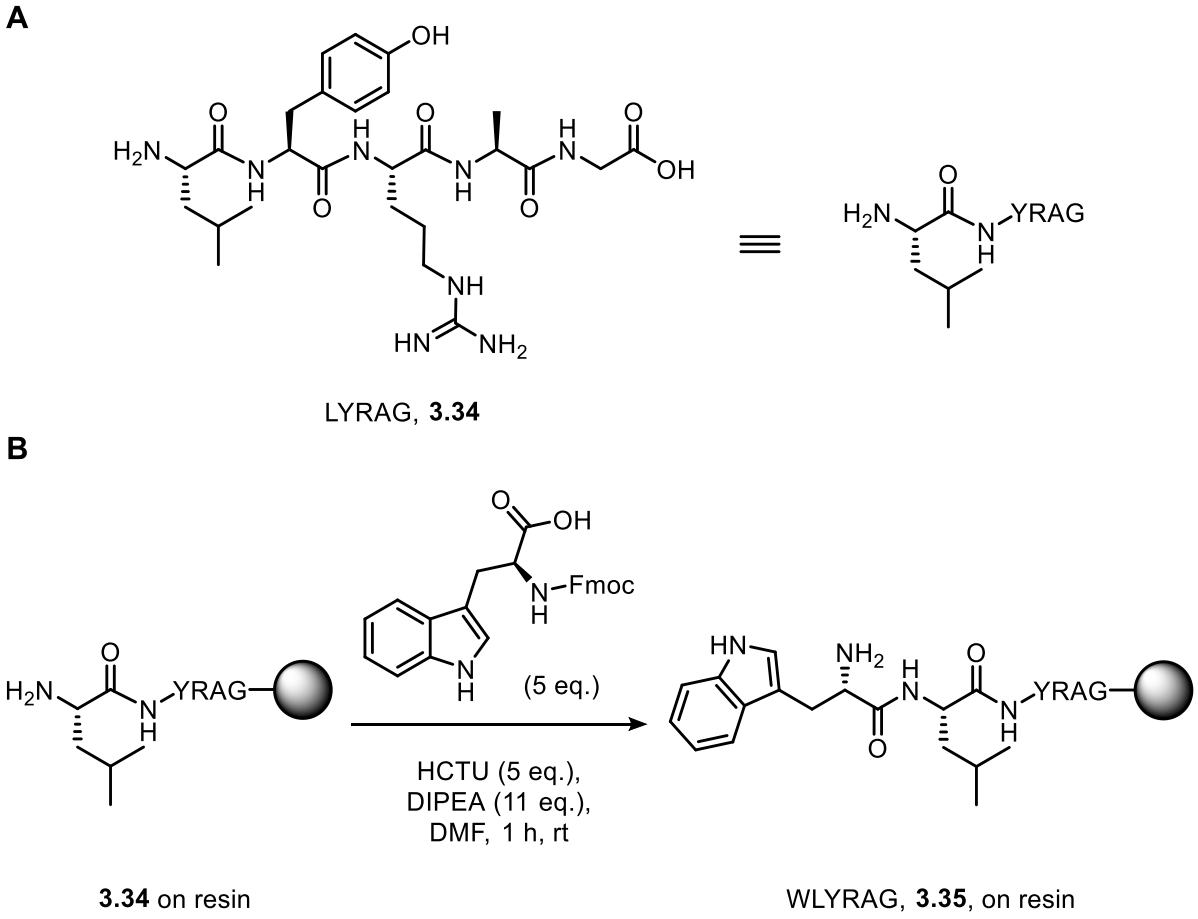


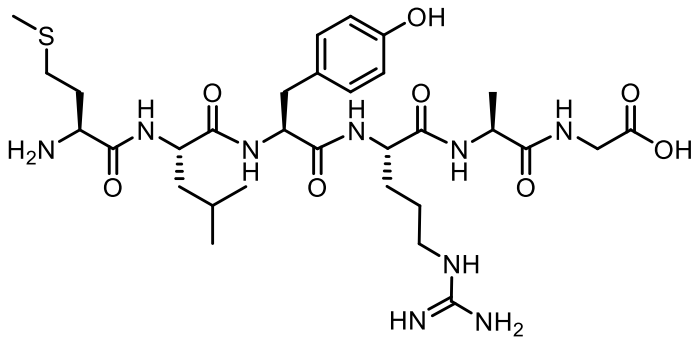
Figure 3.03: (A) Structure of LYRAG peptide. (B) Coupling of *N*-terminal residue to LYRAG using standard peptide coupling conditions. WLYRAG shown as an example. Details in Chapter 7.3.1.

3.6.2.2 WLYRAG as a CARE substrate

We began this study by applying our chemistry to WLYRAG (**3.35**), an *N*-terminal LYRAG derivative containing tryptophan at the *N*-terminus, using the same reaction conditions as those optimised for SLYRAG (**3.04** (5 eq.), 100 mM phosphate buffer, pH 7.4, 22 °C, 3 h). In this case, we observed a small reduced rate for the conjugate addition step (k_1) in WLYRAG compared to SLYRAG ($k_1 = 0.040$ M^{-1} s^{-1} for SLYRAG, *vs.* 0.024 M^{-1} s^{-1} for WLYRAG). This observation agrees well with the lower α-amine pK_a of serine relative to tryptophan at the *N*-terminus, which would be expected to increase the rate of conjugate addition.[62] Moreover, the rate of ring expansion (k_2) in WLYRAG was also lower than that for SLYRAG ($k_2 = 9.0 \times 10^{-4}$ s^{-1} for SLYRAG, *vs.* 1.0×10^{-4} s^{-1} for WLYRAG). We attribute this decrease to the increase in

the size of the *N*-terminal side-chain, an effect we have previously identified as detrimental in our small molecule work discussed in Chapter 1. See Figure 3.07 for the concentration *vs.* time plot and tabulated rate constants.

3.6.2.3 MLYRAG as a CARE substrate

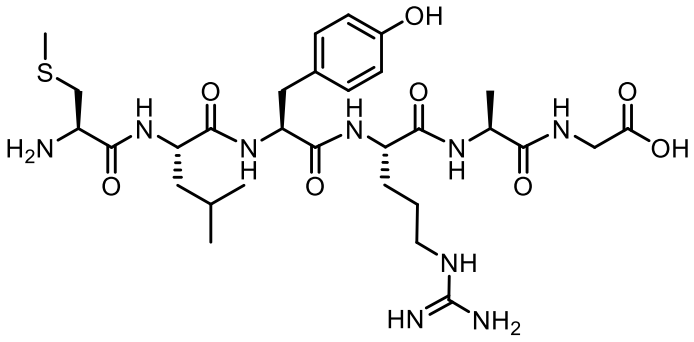


MLYRAG, **3.36**

Figure 3.04: Structure of MLYRAG peptide.

Next, we synthesised MLYRAG (**3.36**) which has an *N*-terminal methionine residue. This derivative was of particular interest to us, as the AUG start codon in eukaryotic translation encodes methionine, meaning that after translation, all proteins naturally possess an *N*-terminal methionine residue. However, these are subsequently removed during post-translational processing by methionine aminopeptidases (MetAP), depending on the size of the sidechain on the adjacent residue — generally, if the radius of gyration is less than 1.29 Å, the *N*-terminal methionine is cleaved.[129,130] Interestingly, having an N-terminal methionine residue led to a substantial increase in the rate of conjugate addition (k_1), which again aligns well with the lower α-amine pK_a of methionine compared to serine (k_1= 0.113 M^{-1} s^{-1} for MLYRAG *vs.* 0.040 M^{-1} s^{-1} for SLYRAG). In fact, methionine has one of the lowest α-amine pK_a values of all the amino acids, which likely accounts for the significant increase in k_1.[62] Furthermore, we observed a similar k_2 dependency on the size of the *N*-terminal side-chain, with MLYRAG exhibiting a slightly decreased rate of ring-expansion compared to SLYRAG (k_2= 3.0 x10^{-4} s^{-1} for MLYRAG *vs.* 9.0 x10^{-4} s^{-1} for SLYRAG). See Figure 3.07 for the concentration *vs.* time plot and tabulated rate constants.

3.6.2.4 C(Me)LYRAG as a CARE substrate

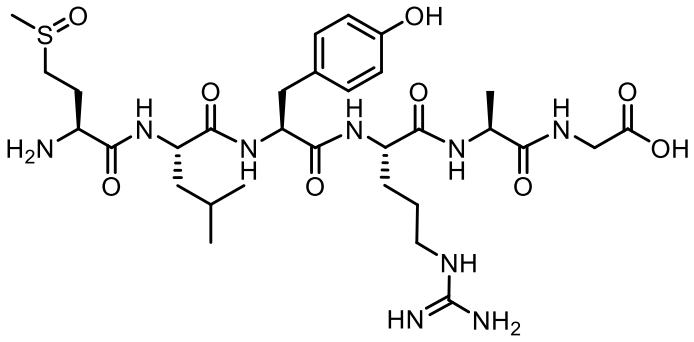


C(Me)LYRAG, **3.37**

Figure 3.05: Structure of C(Me)LYRAG peptide.

To further investigate the sensitivity of the ring-expansion rate to the size of the *N*-terminal side chain, we synthesised C(Me)LYRAG (**3.37**), which has one fewer methylene (CH_2) in its chain side compared to MLYRAG. As with MLYRAG, C(Me)LYRAG exhibited an increased k_1 relative to SLYRAG; however, the magnitude of this increase was significantly lower than that observed for MLYRAG (k_1= 0.051 M^{-1} s^{-1} for C(Me)LYRAG *vs.* 0.113 M^{-1} s^{-1} for MLYRAG). Whilst there is no reported data on the α-amino group pK_a of *N*-terminal methylated cysteine residues in peptides or proteins, the pK_a of the α-amino group in the free amino acid is reported as 9.15, which is comparable to that of methionine (9.2).[62] This suggests that either an additional factor is affecting k_1 in this case, or that the α-amino group's pK_a of methylated cysteine within a peptide is not as low as that of methionine. Nevertheless, C(Me)LYRAG does exhibit an increased rate of ring-expansion (k_2) compared to MLYRAG, further reinforcing that the size of the *N*-terminal side chain is a key factor influencing k_2 (k_2= 7.2 x10^{-4} s^{-1} for C(Me)LYRAG *vs.* 3.0 x10^{-4} s^{-1} for MLYRAG). Intriguingly, the CARE reaction with C(Me)LYRAG reached a significantly greater level of completion than with any other SLYRAG derivatives, with ~85% conversion to ring-expanded product compared to <60% for all other substrates. See Figure 3.07 for the concentration *vs.* time plot and tabulated rate constants.

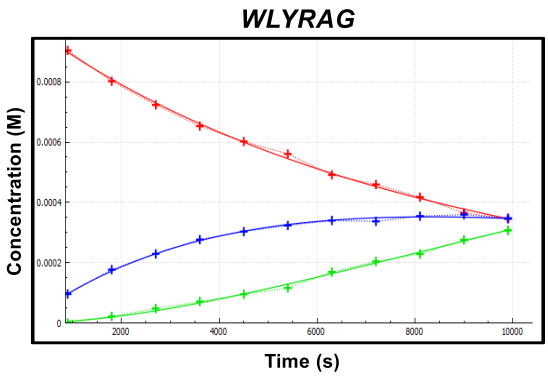
3.6.2.5 M(=O)LYRAG as a CARE substrate

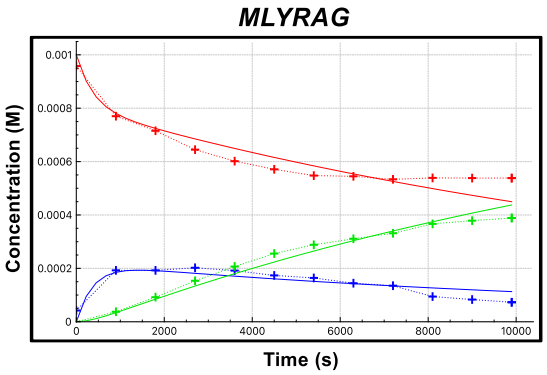


M(=O)LYRAG, **3.38**

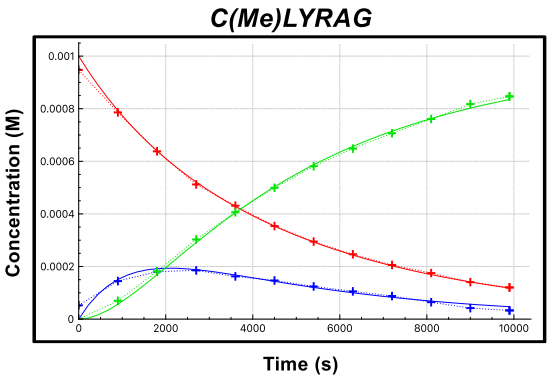
Figure 3.06: Structure of M(=O)LYRAG peptide.

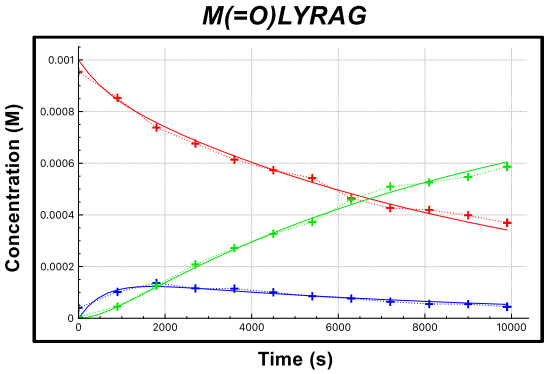
Lastly, we synthesised M(=O)LYRAG (**3.38**) to investigate the effect of a different sulphur valency -as introduced by the sulfoxide group -in contrast to the sulphide-containing side-chains of MLYRAG and C(Me)LYRAG. Additionally, we were interested in investigating a non-canonical amino acid with a side-chain functionality that had not yet been tested for compatibility with the CARE chemistry. Our standard LC-MS kinetics experiment found that the k_1 value for M(=O)LYRAG (0.051 M^{-1} s^{-1}) was identical to that of C(Me)LYRAG, which is unsurprising given the similar α-amino group pK_a values of the corresponding free amino acids (9.11 for M(=O)LYRAG *vs.* 9.15 for C(Me)LYRAG). Likewise, we see a similar rate of ring-expansion in M(=O)LYRAG as in C(Me)LYRAG (k_2= 7.0 x10^{-4} s^{-1} for M(=O)LYRAG *vs.* 7.2 x10^{-4} s^{-1} for C(Me)LYRAG). See Figure 3.07 for the concentration *vs.* time plot and tabulated rate constants.





k_1	0.024 ± 0.005 $M^{-1}s^{-1}$	0.113 ± 0.03 $M^{-1}s^{-1}$
k_{-1}	8.0 x10^{-6} ± 0.8 x10^{-6} s^{-1}	1.3 x10^{-3} ± 0.5 x10^{-3} s^{-1}
k_2	1.0 x10^{-4} ± 0.02 x10^{-4} s^{-1}	3.0 x10^{-4} ± 0.1 x10^{-4} s^{-1}





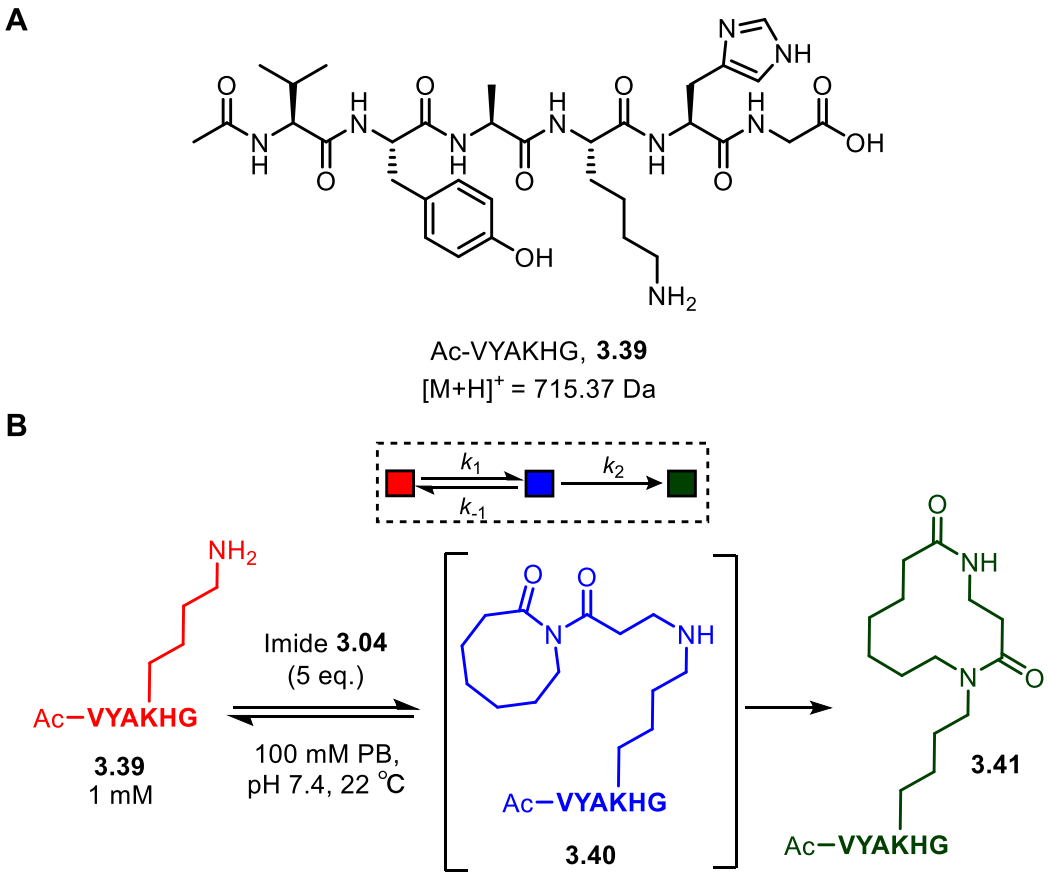
k_1	0.051 ± 0.003 $M^{-1}s^{-1}$	0.051 ± 0.006 $M^{-1}s^{-1}$
k_{-1}	2.0 x10^{-5} ± 4.3 x10^{-5} s^{-1}	8.0 x10^{-4} ± 1.8 x10^{-4} s^{-1}
k_2	7.2 x10^{-4} ± 0.1 x10^{-4} s^{-1}	7.0 x10^{-4} ± 0.2 x10^{-4} s^{-1}

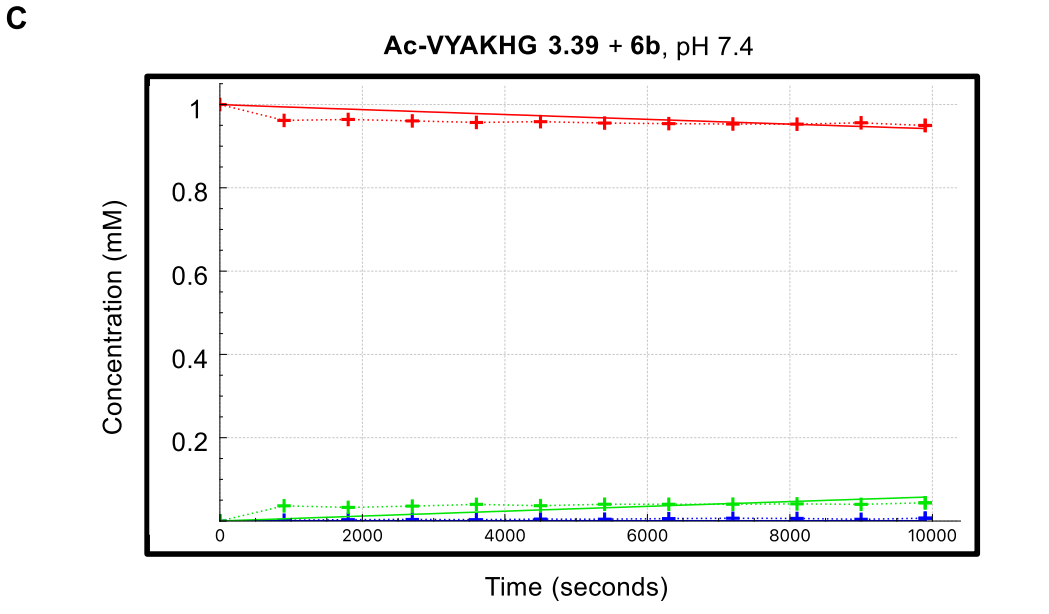
Figure 3.07: Plots of concentration against time for the *N*-terminal modification of XLYRAG derivatives with acryloyl imide **3.04**. Fits are based on the two-step kinetic model describe in Chapter 3.6. Experimental details can be found in Chapter 7.3.3.5.

3.7 Kinetic analysis of CARE at ε-amino groups

Having successfully demonstrated that α-amino groups in several peptidyl substrates are compatible with our CARE reactions, we next turned our attention to exploring site-selectivity. Specifically, we decided to investigate the kinetics of a CARE reaction at an ε-amino group to determine how the introduction of a competing reactive site might affect our *N*-terminal selectivity. To do this, we synthesised Ac-VYAKHG (**3.39**), a 6-mer containing an internal lysine residue and bearing an *N*-terminal acetate cap and subjected it to the same LC-MS

analysis under the same reaction conditions as those used for the *N*-terminal SLYRAG derivatives (Figure 3.08B).





Entry	pH	Conversion after 3 h (%)	k_1 ($M^{-1}s^{-1}$)	k_{-1} (s^{-1})	k_2 (s^{-1})
1	7.4	<5	$6.3 \times 10^{-3} \pm 283 \times 10^{-3}$ $M^{-1}s^{-1}$	0.05 ± 2.75 s^{-1}	0.01 ± 0.05 s^{-1}

Figure 3.08: Kinetic analysis of CARE cascade with acryloyl imide **3.04** and Ac-VYAKHG (**3.39)**. (A) Structure of Ac-VYAKHG peptide; (B) Reaction Scheme (details found in Chapter 7.3.3.5); (C) A plot of molecular species concentration against time under these conditions.

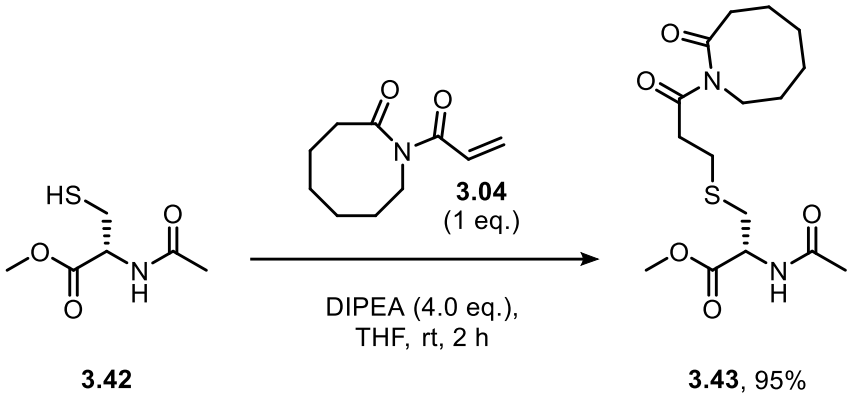
As shown in Figure 3.08C, negligible conversion of starting material to either the aza-Michael intermediate, **3.40** or the ring-expanded product, **3.41** was observed after 3 hours. Correspondingly, the rate of aza-Michael addition (k_1) was significantly reduced than those calculated for reactions at α-amino groups, whilst the rate of retro-aza-Michael addition (k_{-1}) was greatly increased. As a result, there was a greater than 2-order of magnitude increase in the equilibrium constant (K_1) for the conjugate addition step relative to the modification of SLYRAG's α-amino group (0.12 M^{-1} *vs.* 41 M^{-1} for SLYRAG). This sluggishness is likely due to increased protonation of ε-amino groups, resulting from their higher pK_a, which aligns well with the previously observed reduction in *N*-terminal labelling of SLYRAG at slightly acidic pH.

With that being said, the results also showed a noticeably increased rate of ring-expansion at the lysine's ε-amino group, which could be attributed to two possible factors: (1) that reduced steric hindrance around ε-amino groups may be favouring ring-expansion, or (2) that low conversion is leading to a high error in k_2, making the rate constants unreliable. Nonetheless, the results of this study indicate a clear preference for *N*-terminal modification over lysine modification under mild reaction conditions. We anticipate that under more forcing conditions — such as increasing the equivalents of imide or increased reaction pH — the selectivity would diminish, resulting in off-target modifications at lysine residues.

3.8 Comparative reactivity of acryloyl imides with cysteine thiols and *N*-termini

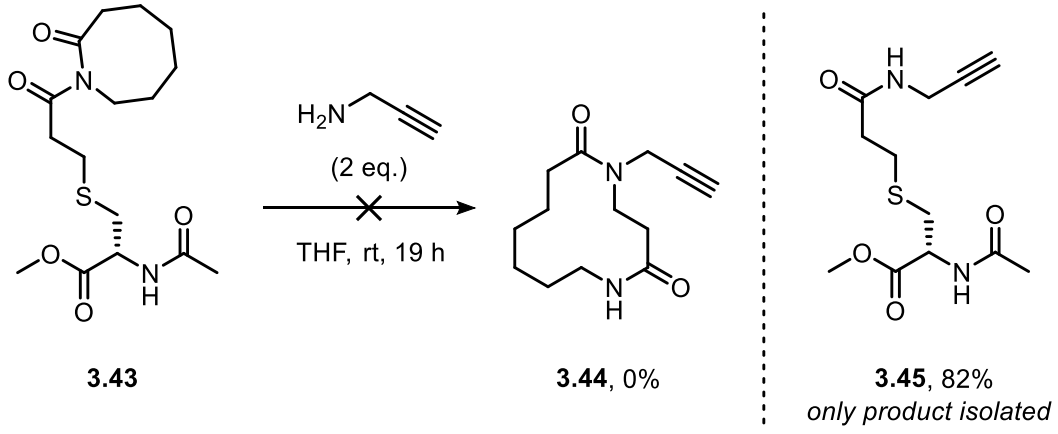
As previously discussed in Chapter 2.5, the conjugate addition of sulphur-containing nucleophiles presents a significant challenge for our CARE chemistry by competing with α-amino groups for the acryloyl imide. Fortunately, our CARE chemistry exhibits selectivity for primary amine nucleophiles, as only these can trigger the subsequent ring-expansion following conjugate addition. Although organic thiols may outcompete α-amino groups in the initial conjugate addition step, they lack the ability to induce the irreversible ring-expansion that ultimately drives the equilibrium towards product formation, as previously discussed. The reversibility of thia-Michael addition is well documented in the literature and is influenced by several factors, including the structure of the Michael acceptor, steric hindrance, solvent, pH and temperature.[131] Our expectation was that if thia-Michael addition to our acryloyl imides was reversible under the reaction conditions, the regenerated Michael acceptor would remain

available for reaction with α-amino groups. If this were the case, the presence of competing cysteine thiols would not significantly impact the efficiency of our CARE chemistry. As an initial test of thiol reactivity and reversibility, *N*-acetyl cysteine methyl ester (**3.42**) was reacted with the 8-membered acryloyl imide (**3.04**) and *N*,*N*'-diisopropylethylamine (DIPEA) in THF (Scheme 3.13). The desired thioether product (**3.43**) was isolated by FCC and obtained in 95% yield.



Scheme 3.13: Thia-Michael addition of *N*-acetyl cysteine methyl ester (**3.42**) with acryloyl imide **3.04**.

Next, to investigate the reversibility of the thia-Michael addition, we subjected the isolated thioether product (**3.43**) to a standard CARE reaction under our established small molecule conditions. If the formation of the thioether were reversible, CARE would be expected to occur as usual, resulting in the formation of the characteristic 12-membered bis-lactam product **3.44** (Scheme 3.14). However, none of the desired bis-lactam product **3.44** was isolated. Instead, the sole product obtained resulted from nucleophilic attack by propargylamine at the external carbonyl of the imide and ejection of the eight-membered lactam, forming **3.45** in 82% yield.



Scheme 3.14: Investigating the reversibility of thia-Michael addition on acryloyl imides by attempting a CARE reaction with thioether **3.43**.

However, this result only demonstrates that the thia-Michael addition is irreversible under these reaction conditions and may not accurately reflect the reactivity observed when modifying peptides or proteins. Compound **3.43** was later employed as a reagent to modify RNase A (see Chapter 5), but no modification was detected. This outcome suggests that the thia-Michael addition is also irreversible under these conditions.

3.9 Chapter summary

In summary, we have successfully optimised the CARE chemistry for the selective *N*-terminal modification of several peptide substrates. These simple and efficient reactions proceed with complete conversion in under 3 hours and are compatible with a range of ring sizes using our easy-to-prepare acryloyl imide reagents. Kinetic analysis revealed that the reaction rates are primarily governed by the reaction pH, the pK_a of the α-amino group of the *N*-terminal residue, and the size of its side chain. Furthermore, we have shown that our CARE bioconjugates exhibit superior stability compared to those formed from traditional Michael acceptors such as maleimides and acrylamides. Lastly, we have demonstrated that the CARE approach shows strong *N*-terminal selectivity over competing ε-amino groups of lysine residues; under mild conditions, this is exemplified by the greater than two-fold increase in the equilibrium constant for reactions at the *N*-terminus. With these results in hand, we next focussed on synthesising a more versatile acryloyl imide reagent — one that could be diversified pre- or post-bioconjugation — to enhance the utility of our approach before investigating our approach on model proteins, with these results discussed in the following Chapter 4.0.

Chapter 4: Synthesising a versatile acryloyl imide probe for reagent diversification

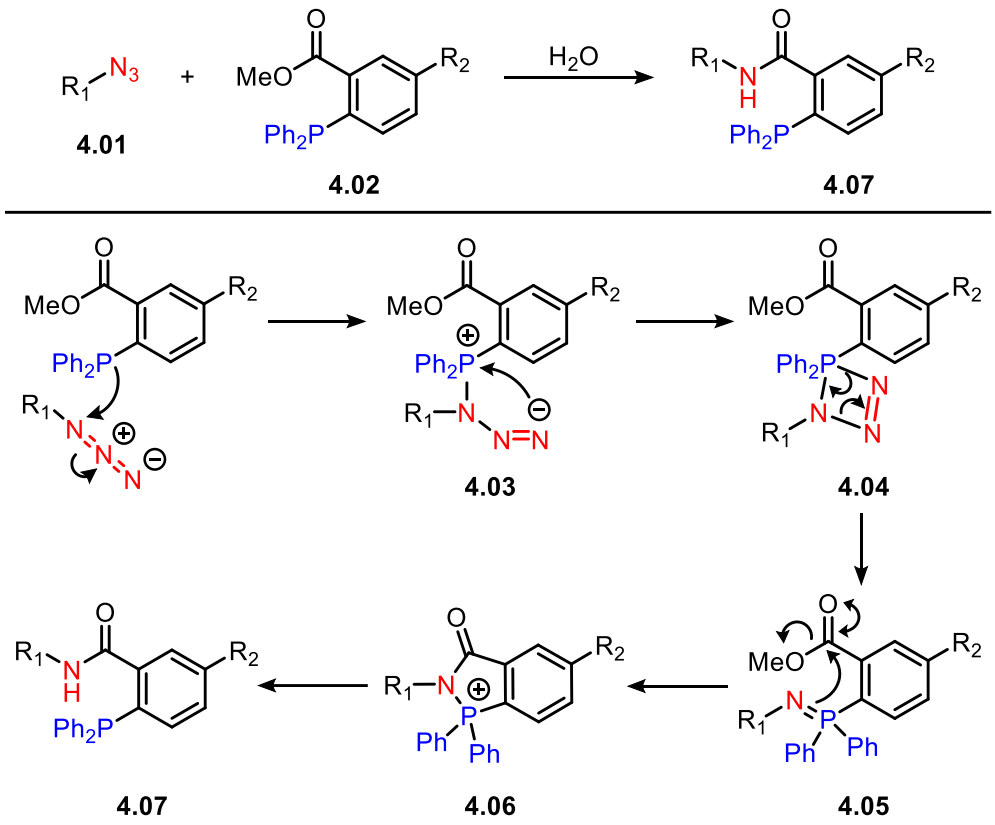
4.1 Exploring bioorthogonal chemistries for probe functionalisation

Building upon the success of our CARE chemistry for peptide bioconjugation, we next aimed to design a new acryloyl imide probe that would both expand upon these findings and further demonstrate the utility of our approach. To do this, we recognised the need to incorporate a bioorthogonal handle into the probe, allowing for post-CARE functionalisation and greater reagent diversification. Bioorthogonal handles incorporated into biomolecules enable rapid and selective reactions with complementary functional groups, without interfering with the biomolecules structure or its native biochemical processes. Importantly, the bioorthogonal handle must not contain any functional groups naturally present in biological systems, as this increases the selectivity of the reaction and reduces the likelihood of off-target modifications. Bioorthogonal chemistry has garnered significant attention in recent years, with numerous studies demonstrating its utility in a myriad of applications including imaging, activity-based protein profiling (ABPP), and drug delivery[132–136]. Well-known examples of bioorthogonal chemistries include the modified Staudinger ligation, tetrazine ligation, and, perhaps most notably, the copper-catalysed azide-alkyne cycloaddition (CuAAC) reaction.

This Chapter will begin with a consideration of each of these bioorthogonal chemistries, followed by a description of the synthetic work undertaken towards the development of a versatile acryloyl imide probe for CARE bioconjugate diversification. The application of these novel probes in our peptide model is then explored, including the identification of optimal reaction conditions and a detailed investigation of its reaction kinetics in comparison to the corresponding non-substituted probes.

4.1.1 The Staudinger Ligation

The Staudinger ligation (Scheme 4.01) was the first bioorthogonal reactions to be developed and was modified from the classic Staudinger reaction, in which azides react with phosphines to form amines *via* an iminophosphorane intermediate.[137,138] Bertozzi *et al.,* modified the classical Staudinger reaction by incorporating an electrophilic methyl ester into the phosphine (**4.02**), which acted as an intramolecular trap for the reactive iminophosphorane intermediate (**4.05**). This addition effectively prevents iminophosphorane hydrolysis and instead results in the formation of a stable amide bond (**4.07**).[139]

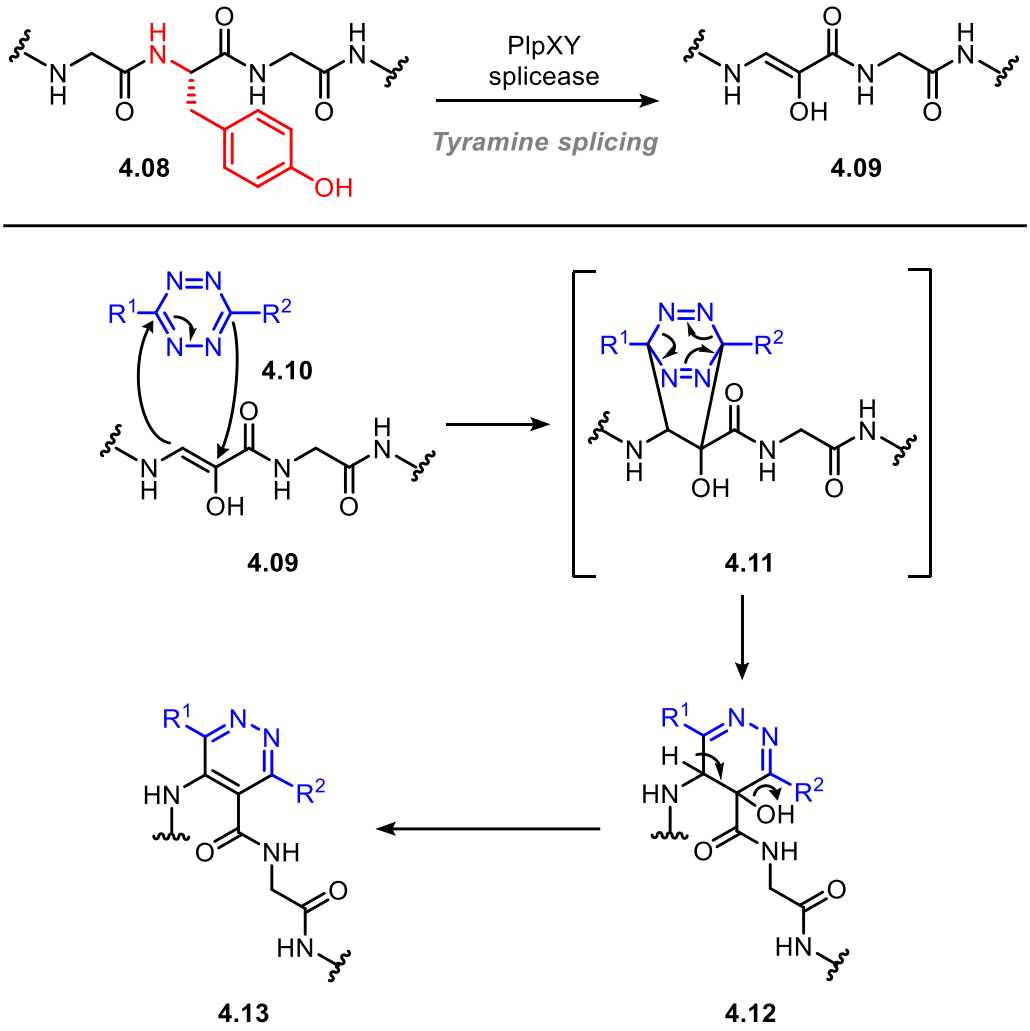


Scheme 4.01: The Staudinger ligation between azides and ester-containing phosphoranes.

Whilst the Staudinger ligation is highly selective for reactions between azide- and phosphine-containing molecules, it often proceeds slowly, meaning that in recent years it has largely been supplanted by other, faster ligation reactions. Notably, some substrates can exhibit reduced reactivity toward the intramolecular trapping of the iminophosphorane intermediate. This can lead to the accumulation of a primary amine byproduct, arising from the competing Staudinger reduction pathway. Such byproducts could interfere with the amine-selective CARE chemistry and reduce its selectivity for the *N*-terminus. On this basis, we concluded that the Staudinger ligation is not a suitable bioorthogonal reaction for CARE bioconjugate diversification.

4.1.2 Tetrazine ligation

Another useful bioorthogonal ligation involves the inverse electron-demand Diels-Alder (IEDDA) cycloaddition between tetrazines and highly reactive, strained dienophiles (Scheme 4.02). This transformation is among the fastest known bioorthogonal reactions, exhibiting rate constants up to 10^6 $M^{-1}s^{-1}$ with *trans*-cyclooctene as the dienophile. It proceeds with high conversions and specificity, and has been used across a range of applications, including PET imaging using pre-radiolabelled tetrazines.[140–142]



Scheme 4.02: PlpXY splicease-mediated tyramine splicing followed by tetrazine ligation using endogenous β-amino acid dienophiles.

Traditionally, a major limitation of this approach was the incorporation of reactive dienophiles into biomolecules. As shown in Scheme 4.02, this can be overcome by the *in-situ* generation of a reactive dienophile within the peptide backbone *via* PlpXY splicease-mediated tyramine excision from tyrosine (**4.08** → **4.09**). The newly installed IEDDA-reactive site (in **4.09**) can undergo an inverse [4+2] cycloaddition with tetrazine **4.10**, forming a dihydropyridazine intermediate (**4.11**). This intermediate then eliminates N_2 (**4.11** → **4.12**), followed by the loss of H_2O, to afford the pyridazine product (**4.12** → **4.13**).[143]

Despite the favourable kinetics and selectivity of this approach, tetrazines can exhibit poor stability in aqueous environments. Whilst the most reactive tetrazines towards IEDDA cycloaddition are generally the least stable, disubstituted tetrazines can offer improved stability at the expense of reduced — though still very fast — reactivity. Additionally, the synthesis of strained alkenes and highly reactive tetrazines can be very challenging, and these compounds

are often hydrophobic, further complicating their handling and use in biological environments. Furthermore, reactive tetrazines are susceptible to off-target reactions with thiols, making them unsuitable in the presence of free cysteine residues in biomolecules. Taking these limitations into consideration, we concluded that tetrazine ligation was not the easiest bioorthogonal reaction to employ for our investigation.

4.1.3 Copper-catalysed azide-alkyne cycloaddition (CuAAC)

The copper-catalysed azide-alkyne cycloaddition (CuAAC) stands as the archetypical "click" reaction, owing to its broad applicability, solvent compatibility and substrate tolerance.[144–46] The CuAAC proceeds with exceptional chemoselectivity and often high yields, even when coupling large, sterically hindered molecules. These features make the reaction highly versatile, particularly in applications involving polymeric or biological macromolecules.[147,148]

Whilst the mechanistic details of the CuAAC are still debated, the catalytic cycles depicted in Figure 4.01 are widely accepted. Jin *et al.,* confirmed that bis(copper) complexes are the catalytic species in the kinetically favoured pathway, whilst mono(copper) complexes, which are also present in solution, catalyse a less kinetically favourable route.[149] The catalytic cycle begins with the coordination of a Cu(I) ion to the terminal alkyne (**4.17**), followed by abstraction of the terminal proton to generate a σ-mono(copper) acetylide species (**4.18**). A second Cu(I) ion can then coordinate to **4.18** to form a bis(copper) acetylide complex (**4.24**), with a second π-bound copper atom at the carbon-carbon triple bond of the alkyne. The bis(copper) acetylide complex **4.24** can reversibly coordinate to an organic azide (**4.19**), positioning the reactants for cycloaddition. This begins with a nucleophilic attack by the β-carbon of the acetylide on the terminal nitrogen of the azide, leading to the formation of intermediate **4.25**. Ring closing of intermediate **4.25** forms triazolide **4.26**, which undergoes rapid protonation — either from the solvent or excess alkyne (**4.17**) — to release the 1,4-disubstituted 1,2,3-triazole product (**4.23**) and the regenerated Cu(I) catalyst.[150] The minor, mono(copper) pathway is also capable of catalysing the transformation, but extensive examination of the reaction has found that the σ-mono(copper) acetylide species (**4.18**) is much less stable and reactive with the organic azide.

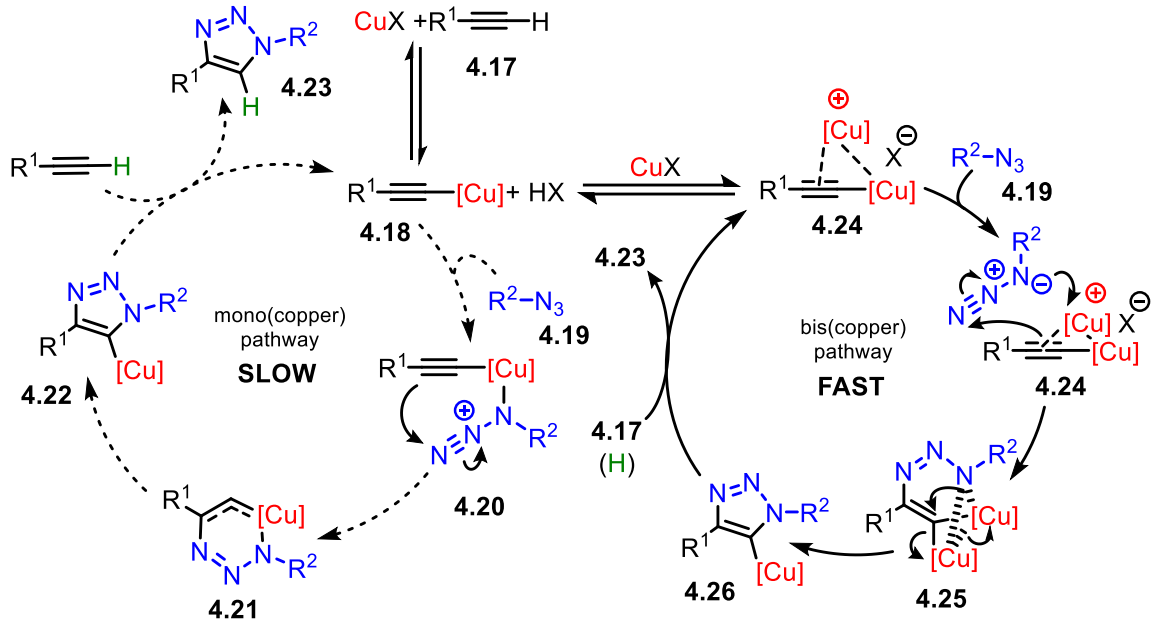


Figure 4.01: Mono- and bis-copper mechanistic pathways for the copper-catalysed azide-alkyne cycloaddition reaction.

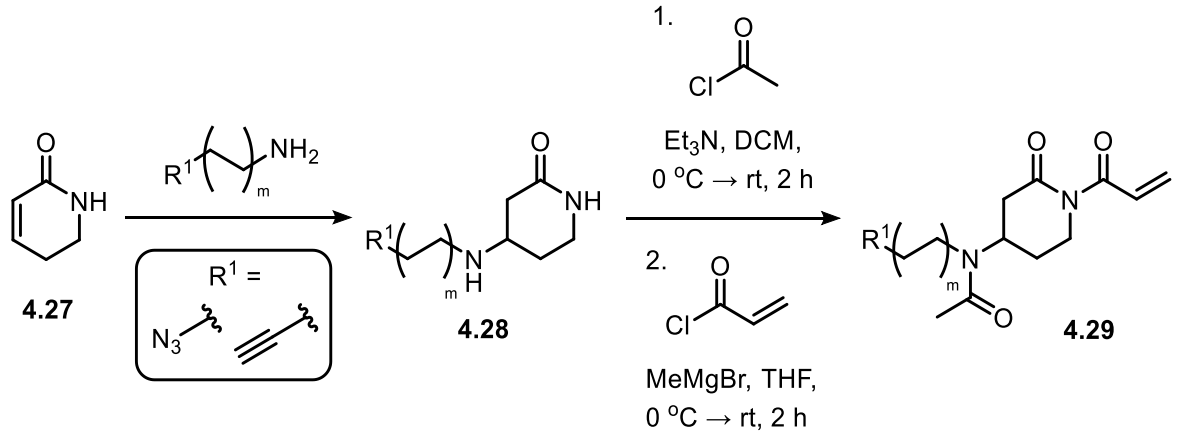
The use of CuAAC in biological systems can be challenging due to the cytotoxicity associated with Cu(II) ions, which are often introduced as a precursor to the active Cu(I) catalyst and can cause oxidative damage to cells. Reducing agents such as sodium ascorbate are often added to the reaction mixture to mitigate these oxidative effects; however, the byproducts of these reductions can themselves contribute to cellular damage.[151] Fortunately, Bertozzi *et al.*, circumvented this issue by removing the need for copper catalysts in this transformation. Strain-promoted azide-alkyne cycloaddition (SPAAC) uses cyclic alkynes containing highly strained *sp*-hybridised alkynyl carbons within their cyclic structure. The relief of ring strain upon reaction with azides significantly accelerates the rate of reaction, enabling copper-free cycloaddition with kinetics comparable to those with CuAAC.[152] However, it is important to note that SPAAC also has several limitations, including the poor water solubility of cycloalkynes, their challenging synthesis, and background reactivity with thiol-containing biomolecules.[151]

The stability of organic azides and alkynes, combined with the relative ease of incorporating them into small molecules, made this bioorthogonal chemistry very attractive for our project. Thus, this work begins with the synthesis of azide- and alkyne-functionalised acryloyl imide probes for application in our CARE chemistry. We then demonstrate the utility of these novel probes in our peptide model and compare their relative reaction rates to those of our original, unsubstituted probes.

4.2 Installing a clickable tag *via* Michael-addition

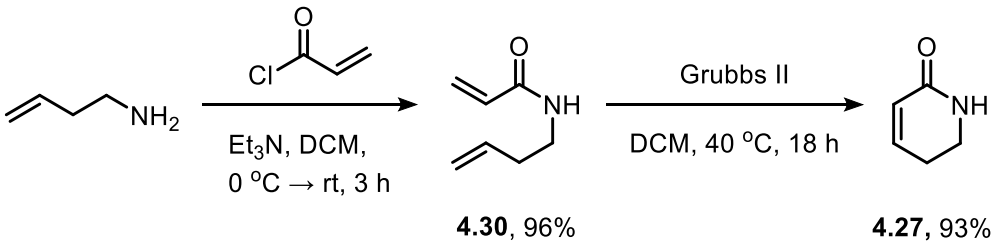
4.2.1 Synthesis of an α,β-unsaturated lactam intermediate

α,β-Unsaturated lactam **4.27** was identified as a versatile scaffold for the construction of tagged acryloyl imide probes *via* 1,4-conjugate addition with azide or alkyne-containing linkers (**4.27** → **4.28**). After protecting the resulting secondary amine, the key acryloyl imide motif required for the CARE chemistry could then be assembled through *N*-acylation of the lactam using the Unsworth group's well-established conditions (**4.28** → **4.29**).



Scheme 4.04: Planned assembly of azide- or alkyne-tagged acryloyl imide probes (**4.29**) from lactam **4.27**.

Several literature procedures were available for the synthesis of intermediate **4.27**, with the route summarised below in Scheme 4.05 chosen.[153] The synthesis began with the coupling of homoallylamine with acryloyl chloride to afford acrylamide **4.30** in 96% yield. Acrylamide **4.30** was then converted into the key intermediate **4.27** *via* a Grubbs (II)-catalysed ring-closing metathesis in an excellent 93% yield.

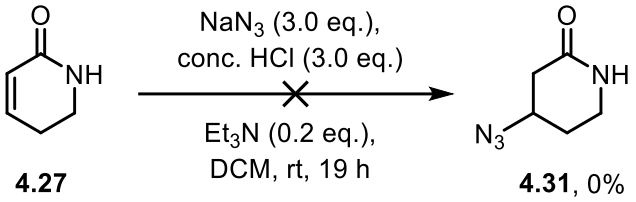


Scheme 4.05: Synthesis of key intermediate **4.27** from readily available starting materials.

4.2.2 1,4-Conjugate addition of azide ion into an α,β-unsaturated lactam

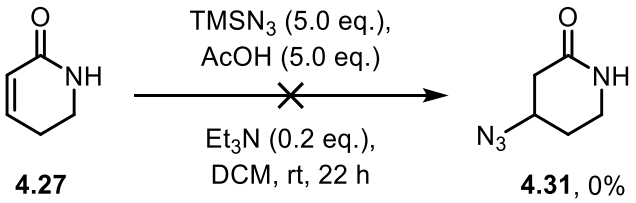
With lactam **4.27** in hand, we began exploring 1,4-conjugate addition reactions to introduce clickable tags into the structure. Initially, for the sake of simplicity, we investigated the direct 1,4-addition of azide ions into the α,β-unsaturated lactam **4.27**. Although there was

considerable precedent in the literature for reactions of this kind, most examples involved α,β-unsaturated ketones or esters rather than lactams.[154,155] Nonetheless, we attempted an acid-catalysed Michael addition of sodium azide into **4.27** in attempts to isolate β-azido lactam **4.31**. Unfortunately, the reaction was unsuccessful, with no observable conversion of the starting material to the desired product nor any byproducts (Scheme 4.06). It is important to note that in DCM, sodium azide can generate diazidomethane, a highly explosive compound, whilst in HCl it can produce HN_3, another toxic and potentially explosive species.



Scheme 4.06: Unsuccessful acid-catalysed 1,4-addition of azide ion to **4.27**.

Whilst the result was disappointing, it was not unexpected given the lack of precedent for sodium azide reacting with non-activated alkenes and its sparing solubility in most organic solvents. With this in mind, we carried out the reaction using trimethylsilyl azide—a safer and more soluble azide source—in an effort to isolate β-azido lactam **4.31**. However, this reaction was also unsuccessful, with no traces of the desired product observed in the crude ^{1}H NMR spectrum (Scheme 4.07).



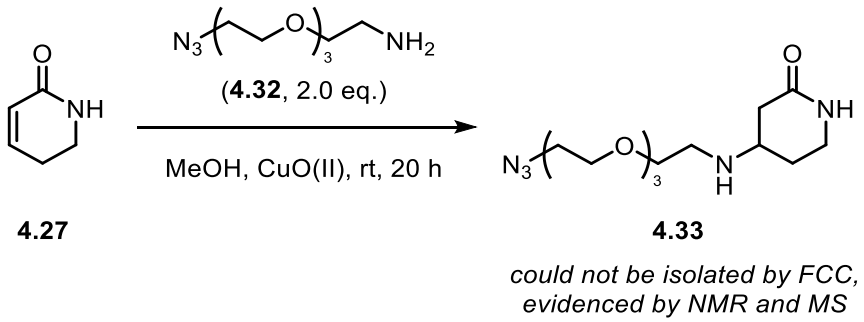
Scheme 4.07: Unsuccessful 1,4-addition of trimethylsilyl azide to **4.27**.

Since neither of the direct 1,4-addition approaches were successful in isolating the desired β-azido lactam, we instead focussed our efforts on alternative strategies for introducing 'clickable' tags onto the lactam.

4.2.3 Aza-Michael addition of azide-functionalised amines

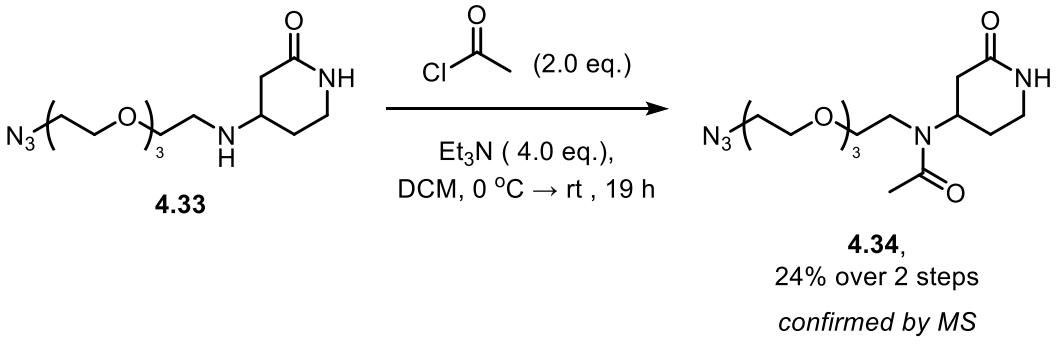
Given the reliability of aza-Michael additions, we thought that this approach would be a more dependable alternative to the methods previously described. Within our chemical inventory was azido-PEG-amine **4.32**, a potentially suitable amine substrate for this type of aza-Michael addition. Moreover, the incorporation of a long PEG linker ensures that the resulting product

adheres to the "rule of six", a guideline for the safe handling of explosive compounds by requiring at least six carbon atoms per energetic functional group (such as azides).[156] Whilst no aza-Michael reactions with this specific substrate have been reported in the literature, there is evidence that nanocrystalline copper (II) oxide can efficiently catalyse the conjugate addition of long aliphatic amines into α,β-unsaturated carbonyl compounds.[157] Based on this precedent, a reaction was therefore tried using the reported conditions (Scheme 4.08).



Scheme 4.08: Aza-Michael addition of azido-PEG-amine **4.32** into α,β-unsaturated lactam **4.27**.

The formation of the azide-tagged β-amino lactam 4.33 was supported by both the crude ^{1}H NMR and MS spectra. Specifically, we observed the disappearance of the vinyl proton signals present in **4.27**, along with the appearance of a molecular ion peak consistent with the expected mass of **4.33**. However, the compound could not be isolated by FCC due to coelution with excess unreacted starting material. Therefore, the crude reaction mixture was carried forward into a secondary amine-trapping step *via* acylation with acetyl chloride (Scheme 4.09). This step was necessary to block the free amine, thereby preventing chemoselectivity issues during the subsequent *N*-acylation of the lactam.



Scheme 4.09: Secondary amine acylation of **4.33** with acetyl chloride.

A product from the reaction was successfully isolated by FCC; however, its identity could not be confidently confirmed as the desired product, **4.34**. The ^{1}H NMR spectrum suffered from severe signal broadening, and the ^{13}C NMR spectrum from multiple overlapping signals,

suggesting the presence of rotamers or possible impurities. This made it extremely difficult to confirm the structure of the material with any certainty. Mass spectrometry data included a molecular ion peak consistent with the expected mass of **4.34**, but this alone was insufficient to conclusively confirm the successful synthesis of the target compound.

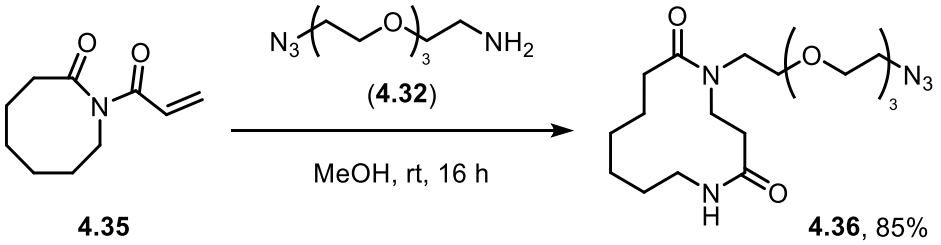
Given the low overall yield across two steps, the need for intermediate amine protection, and the high cost of the starting materials, this route was deemed unsuitable for the scalable synthesis of an azide- or alkyne-tagged acryloyl imide probe. Consequently, we returned to the drawing board and devised a new synthetic route towards a novel target compound.

4.3 Installing a clickable tag *via* CARE

4.3.1 Synthesis of an azide-tagged acryloyl imide

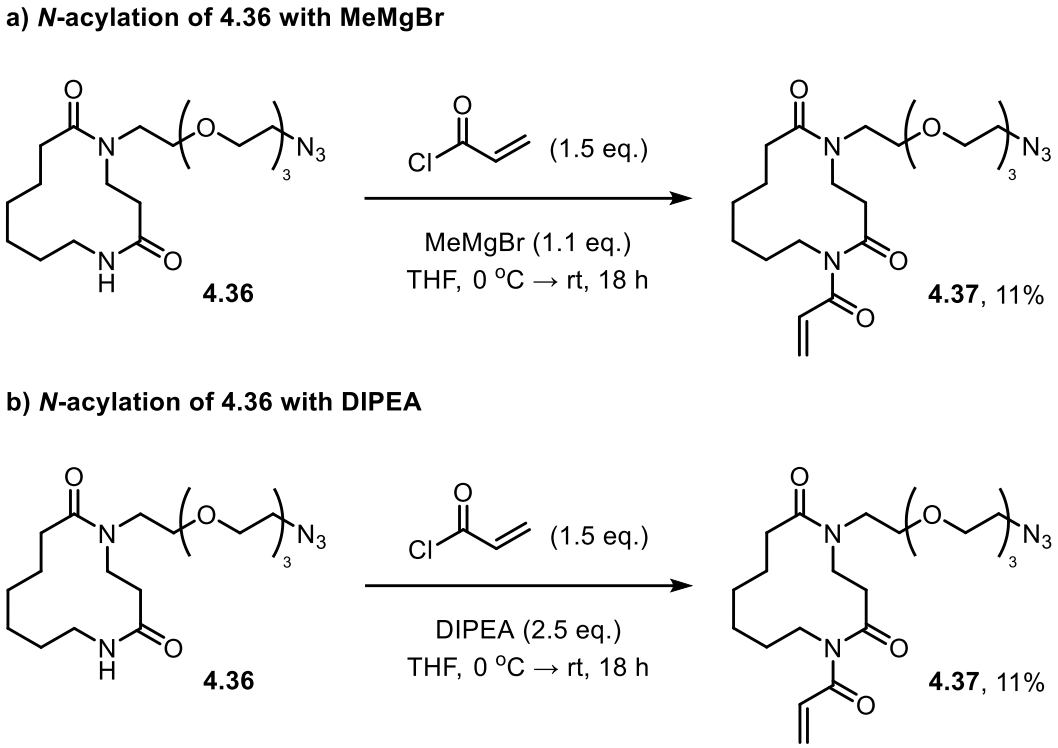
We recognised that our small molecule CARE chemistry could be used as a means of introducing a clickable tag directly into the probe's structure. Previous work within the Unsworth group has demonstrated that iterative CARE reactions are possible by re-acylating the medium-sized ring product from the initial expansion and subsequently reacting it with a second primary amine substrate.[50] This approach offers a modular route towards the incorporation of azide- or alkyne-functionalised amines directly into the probe's framework, without the need for complex, multi-step syntheses.

Following this reasoning, a standard CARE reaction was carried out between acryloyl imide **4.35** and azido-PEG-amine **4.32** under the usual conditions, affording the ring-expanded lactam **4.36** in a respectable 85% yield (Scheme 4.10).



Scheme 4.10: CARE reaction of acryloyl imide **4.35** with **4.32**.

Two different sets of conditions were tried for the *N*-acylation of **4.36** with acryloyl chloride: one using MeMgBr as the base, and the other using *N*,*N*-diisopropylethylamine (DIPEA). Coincidentally, both approaches afforded the desired acryloyl imide product **4.37** in 11% yield (Scheme 4.11).

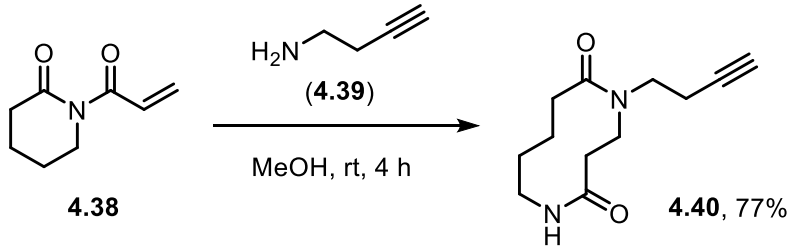


Scheme 4.11: *N*-acylation of ring-expanded lactam **4.36** with acryloyl chloride under two different conditions: (a) using MeMgBr and (b) using DIPEA as a base.

Despite the poor overall yield of just 9%, further optimisation of this pathway was not required, as a high yielding synthesis was not important to our intended application at this stage of the project. Notably, in protein modification, only small amount of reagent is required, meaning that limited material is sufficient for numerous modifications. Importantly, the successful isolation of **4.37** enabled us to investigate an azide-tagged acryloyl imide as a substrate in a CARE reaction using our optimised peptide model.

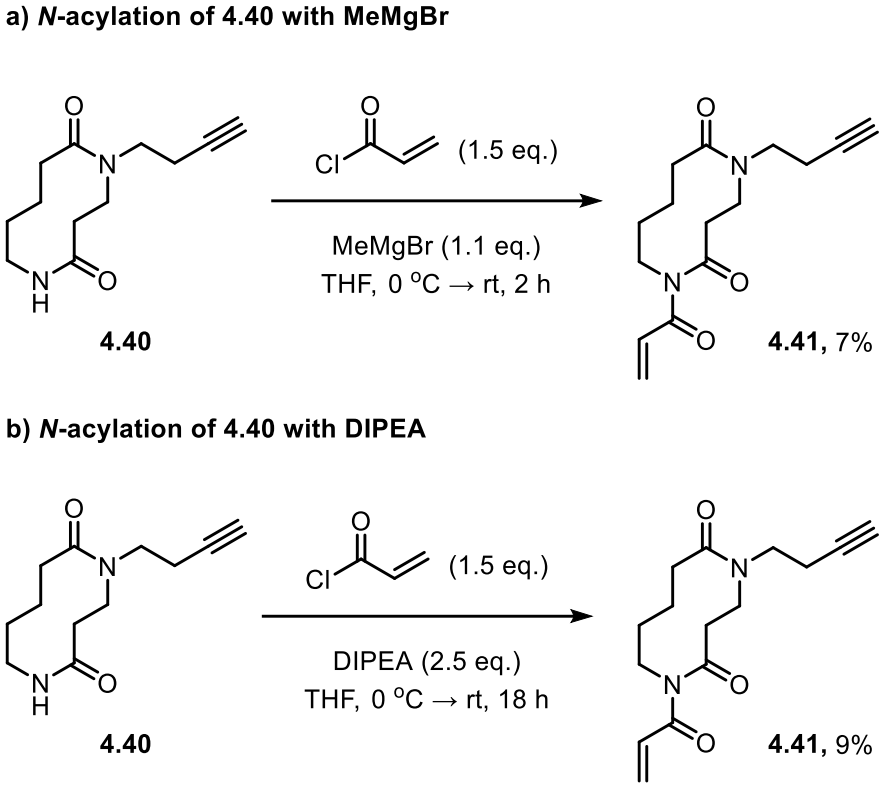
4.3.2 Synthesis of an alkyne-tagged acryloyl imide

We also set out to introduce an alkyne tag into our scaffold *via* a CARE reaction using an alkyne-functionalised primary amine. We began by reacting acryloyl imide **4.38** with **4.39** under the usual CARE conditions, affording the desired ring-expanded product **4.40** in 77% yield (Scheme 4.12).



Scheme 4.12: CARE reaction of acryloyl imide **4.38** with **4.39**.

However, the following *N*-acylation step proved extremely challenging due to the poor solubility of **4.40** in THF, meaning that both sets of reaction conditions afforded the *N*-acylated product **4.41** in <10% yield (Scheme 4.13). This explanation was supported by the observation of starting material signals in the crude ^{1}H NMR spectra, indicating that the low isolated yields were due to incomplete conversion rather than off-target acylation at the terminal alkyne.



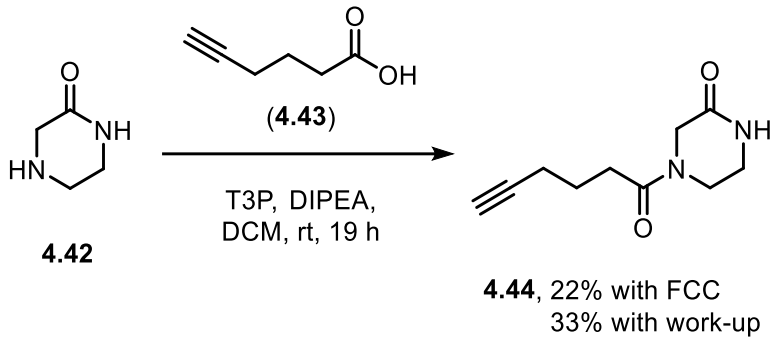
Scheme 4.13: *N*-acylation of ring-expanded lactam **4.40** with acryloyl chloride under two different conditions: (a) using MeMgBr and (b) using DIPEA as a base.

With imide **4.41** in hand, attention turned to testing it in peptide studies. However, when preparing stock solutions of **4.41**, we found it to be only sparingly soluble in DMSO. Given both its poor solubility and low isolated yields noted above, we instead turned our attention to alternative synthetic routes towards an alkyne-tagged acryloyl imide probe.

4.4 Installing a clickable tag *via* amide coupling

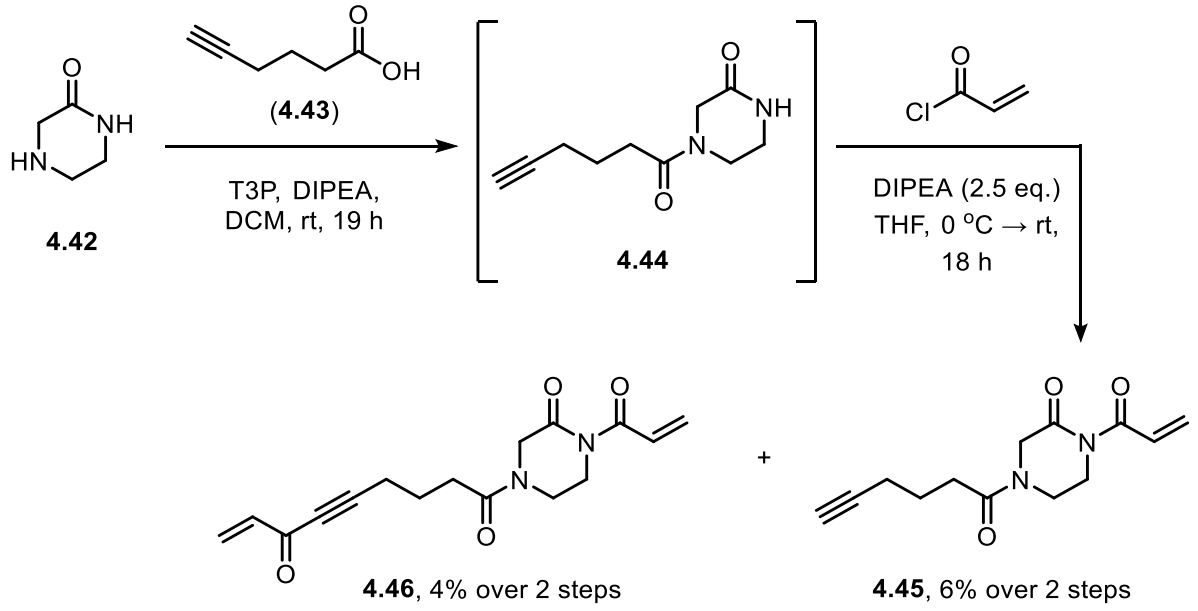
4.4.1 Synthesis of an alkyne-tagged acryloyl imide

Finally, we investigated amide coupling reactions as a potential route to accessing an alkyne-tagged acryloyl imide. This began with the coupling of piperazin-2-one (**4.42**) and 5-hexynoic acid (**4.43**) using propanephosphonic acid anhydride (T3P) and DIPEA, affording the alkyne-tagged lactam product, **4.44** in 22% yield (Scheme 4.14).



Scheme 4.14: Synthesis of alkyne-tagged lactam **4.44**.

Significant crude mass loss was observed during the purification of **4.44** by FCC, which accounts for the low isolated yield. Analysis of the crude reaction mixture by ^{1}H NMR after work-up confirmed the presence of exclusively the desired product. Notably, avoiding FCC improved the yield to 33%. This allowed us to carry the material directly into an *N*-acylation reaction to install the key acryloyl imide motif. The *N*-acylation of lactam **4.44** using MeMgBr resulted in the formation of multiple products, as observed by TLC, which were inseparable by FCC. Previous work within the Unsworth group has demonstrated that these strongly basic conditions can be incompatible with certain substrates by causing degradation of the starting material—this is likely what we are observing in this case. Fortunately, replacing MeMgBr with DIPEA led to the formation of the desired acryloyl imide product **4.45** in a combined yield of 6% over two steps (Scheme 4.15).



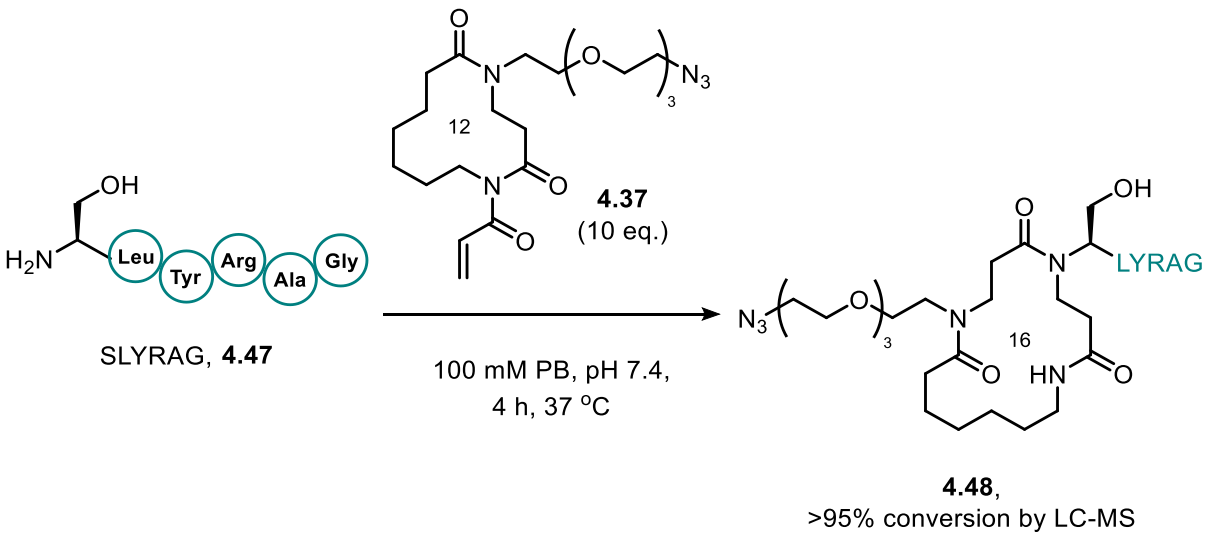
Scheme 4.15: Complete synthesis of alkyne-tagged acryloyl imide **4.45**.

As shown in Scheme 4.15, byproduct **4.46** was also isolated from the reaction mixture, arising from double acylation at both the lactam and the terminal alkyne. Whilst the combined isolated yield of **4.45** was very low, this was not a significant concern, as only a small amount of material was required for investigating its reactivity with SLYRAG.

4.5 Modification of SLYRAG using tagged acryloyl imides and CARE

4.5.1 Using an azide-tagged acryloyl imide

We began by reacting SLYRAG **4.47** with 10 molar equivalents of our azide-tagged acryloyl imide **4.37** under the previously optimised conditions described in Chapter 3.3.1 and outlined in Scheme 4.16. Aliquots of the reaction mixture were taken every hour over the 4-hour reaction time and analysed immediately by LC-MS to track the reaction's progress. The data showed the gradual consumption of **4.47** and the appearance of a strong signal with a molecular ion peak consistent with the expected mass of **4.48**. Unfortunately, both the UV and BPC chromatograms of pure **4.37** were poorly resolved, with the major product peak being accompanied by several smaller peaks throughout the spectrum. These additional signals may correspond to co-eluting contaminants or could suggest an instability of the product, leading to degradation. This made it difficult to carry out any kind of kinetic analysis since significant signal overlap prevented the accurate isolation and integration of individual peaks of interest.



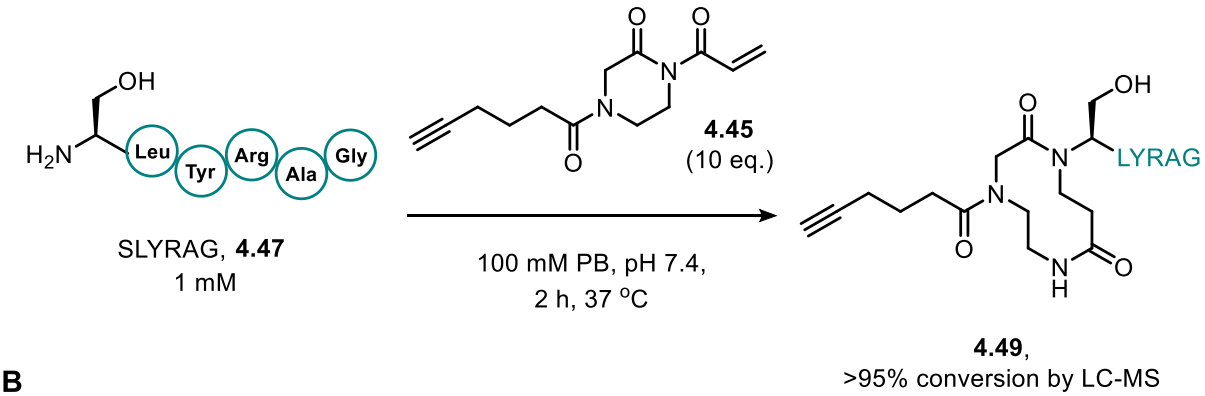
Scheme 4.16: CARE of azide-tagged acryloyl imide **4.37** with SLYRAG **4.47**.

Nonetheless, we were pleased to observe that our novel azide-tagged acryloyl imide **4.37** successfully modified SLYRAG in a manner comparable to the previously tested non-substituted acryloyl imides. This comparison reflects overall reactivity rather than reaction kinetics, due to challenges in isolating clean peaks for kinetic analysis.

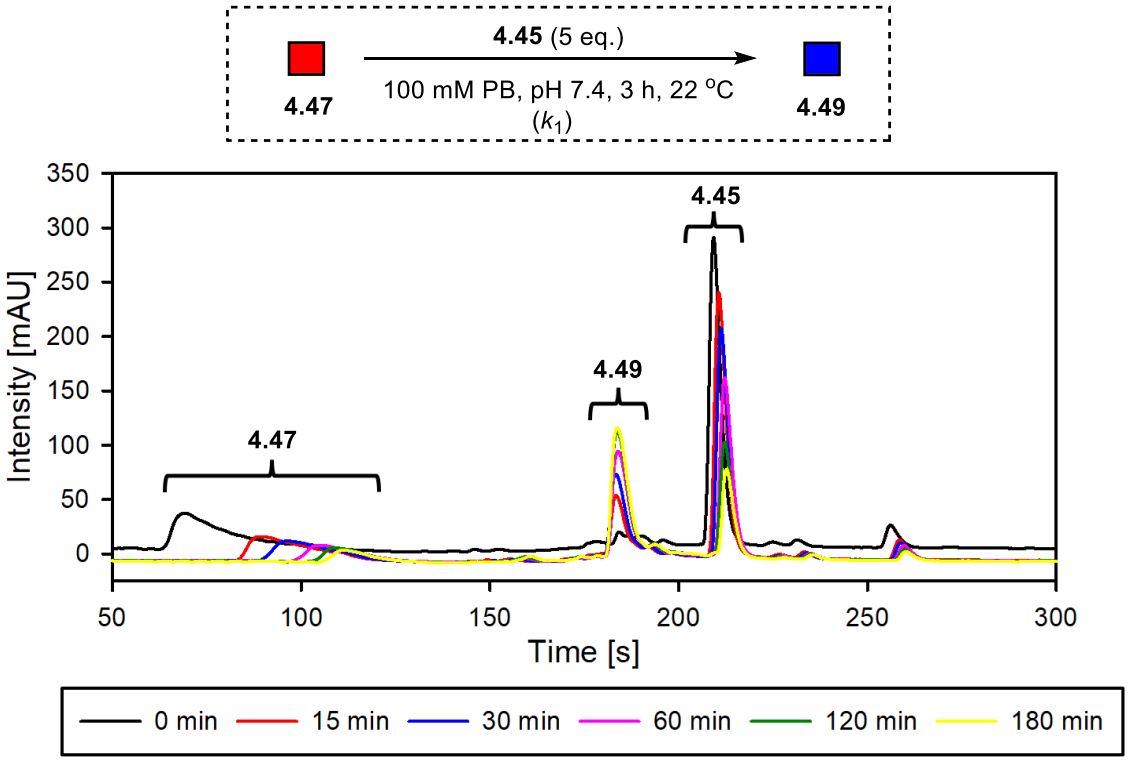
4.5.2 Using an alkyne-tagged acryloyl imide

Similarly, SLYRAG **4.47** was reacted with 10 molar equivalents of the alkyne-tagged acryloyl imide **4.45** under the same conditions used for **4.37**. In this case, the LC-MS chromatogram of the crude reaction mixture was significantly cleaner, with well-resolved peaks. The reaction reached >95% conversion to **4.49** within just 2 hours, exemplifying superior reactivity compared to the other acryloyl imides previously tested (Figure 4.02A). Interestingly, at all time points, LC-MS analysis found only a single species with the correct molecular ion peak for the desired product. Presumably, we observed the direct consumption of the starting material accompanied by the immediate formation of the final ring-expanded product, **4.49** (Figure 4.02B). This suggests that the ring-expansion step proceeds rapidly, preventing the accumulation of the aza-Michael intermediate, which therefore remains undetectable by LC-MS. As a result, we used a simplified one-step COPASI kinetic model to compute an experimentally derived rate constant (k_{eff}) that accounts for both the conjugate addition and ring-expansion steps (Figure 4.02C).

A



B



C

SLYRAG 4.47 + **4.45**, pH 7.4, 22 °C

Entry	pH	Conversion after 3 h (%)	k_{eff} ($M^{-1}s^{-1}$)
1	7.4	80	0.0525

Figure 4.02: (A) CARE of alkyne-tagged acryloyl imide **4.45** with SLYRAG **4.47** (experimental details can be found in Chapter 7.3.3.1); (B) Overlay plot of 280 nm UV traces showing the conversion of SLYRAG **4.47** to ring-expanded product **4.49** over time (experimental details can be found in Chapter 7.3.3.5); (C) Experimentally derived rate constant (k_1) from COPASI computational model for the reaction.

Given these positive results, we also prepared 0.1 M stock solutions of **4.45** in DMSO for protein bioconjugation and stored them at −20 °C. The successful preparation and application

of azide- and alkyne-tagged acryloyl imides marked a key step in the practical applications of our approach in biological systems.

4.6 Chapter summary

To summarise, this Chapter has outlined the various approaches explored for the synthesis of tagged acryloyl imide reagents for use in our CARE chemistry. We have developed synthetic routes to novel azide- and alkyne-tagged acryloyl imide reagents using simple chemical reactions and readily available starting materials. Although the overall yields for these reactions were low, no efforts were made to improve or optimise their syntheses and thus, future work could focus on improving the efficiency of these transformations. We demonstrated the utility of these probes for the selective *N*-terminal modification of peptides using our well-established SLYRAG model system, observing full conversion within a few hours in both cases. These studies showed that the alkyne-tagged probe (**4.45**) reacted significantly faster than the previously tested imide reagents (**3.02** and **3.04**, Chapter 3), with a notably enhanced rate of ring-expansion. Having optimised our CARE chemistry on model peptides in Chapter 3, and synthesising two biologically relevant probes in this Chapter, we were now prepared to tackle the most challenging aspect of this project: optimising CARE for the selective *N*-terminal modification of proteins.

Chapter 5: Optimising CARE for the modification of proteins

5.1 Protein bioconjugation: considerations for a 'CAREful' approach

So far, we have demonstrated that the CARE method can be used to selectively assemble medium-sized and macrocyclic rings on both individual amino acid units and model peptides under mild reaction conditions. We have investigated the kinetics of CARE reactions at both α- and ε-amino groups and have found a marked preference for reactivity at the α-amino group at neutral pH. However, we anticipate that achieving selectivity on proteins will be significantly more challenging, given that there is only a single α-amino group in a protein and often numerous competing ε-amino groups. Moreover, some proteins contain hyper-reactive lysine residues that exhibit significantly increased nucleophilicity over typical lysines at neutral pH.[158,159] This increased reactivity typically arises from local environmental factors — such as solvent exposure, interactions with nearby residues and structural constraints — that contribute towards the lowering of the lysine's pK_a and increasing nucleophilicity.[160] In some cases, the presence of hyper-reactive lysines has been used as an advantage; for example, bioconjugation strategies have been reported focussing on targeting the most reactive lysine residues to achieve site-selectivity in proteins containing multiple lysine residues.[161] In contrast, our groups have previously developed novel *N*-terminal protein bioconjugation strategies and established a panel of model proteins amenable to *N*-terminal modification.[63,84,93]

This Chapter describes the use of our acryloyl imide probes as novel reagents for selective *N*-terminal protein bioconjugation under mild conditions. We began by optimising a set of reaction conditions on the beforementioned model proteins — RNase A, myoglobin, CjX183-D and cytochrome C — to establish a strong foundation for *N*-terminal selective modification. Building on these results, we employed our novel azide- and alkyne-tagged probes developed in Chapter 4 to assemble *N*-terminally modified bioconjugates bearing a 'clickable' tag for further derivatisation. Conjugate stability was evaluated, and site-selectivity was confirmed through *N*-terminal sequencing by Matrix-Assisted Laser Desorption Ionisation In-Source Decay (MALDI-ISD). Finally, we demonstrate the versatility of our bioconjugates in several applications, including the imaging of prostate cancer tissue slices and the modulation of cell surface chemokine receptors.

5.2 Optimising reaction conditions for the modification of model proteins

5.2.1 Modifications with an untagged acryloyl imide probe

After establishing optimal reaction conditions for peptide modification, we first tested these conditions on CjX183-D (PDB entry 7B21), a small c-type cytochrome from *Cellvibrio japonicus*.[162] Notably, CjX183-D does not contain any lysine residues and, therefore, the *N*-terminus is the only amino group available for CARE modification. CjX183-D (**5.01**) was incubated with 5 equivalents of acryloyl imide **5.02** at pH 7.4 and 37 °C for 19 hours (Figure 5.01A). LC-MS analysis of the reaction mixture revealed 63% conversion into a modified CjX183-D species, with a deconvoluted mass consistent with the expected mass of the *N*-terminally modified product (**5.03**, Figure 5.01B).

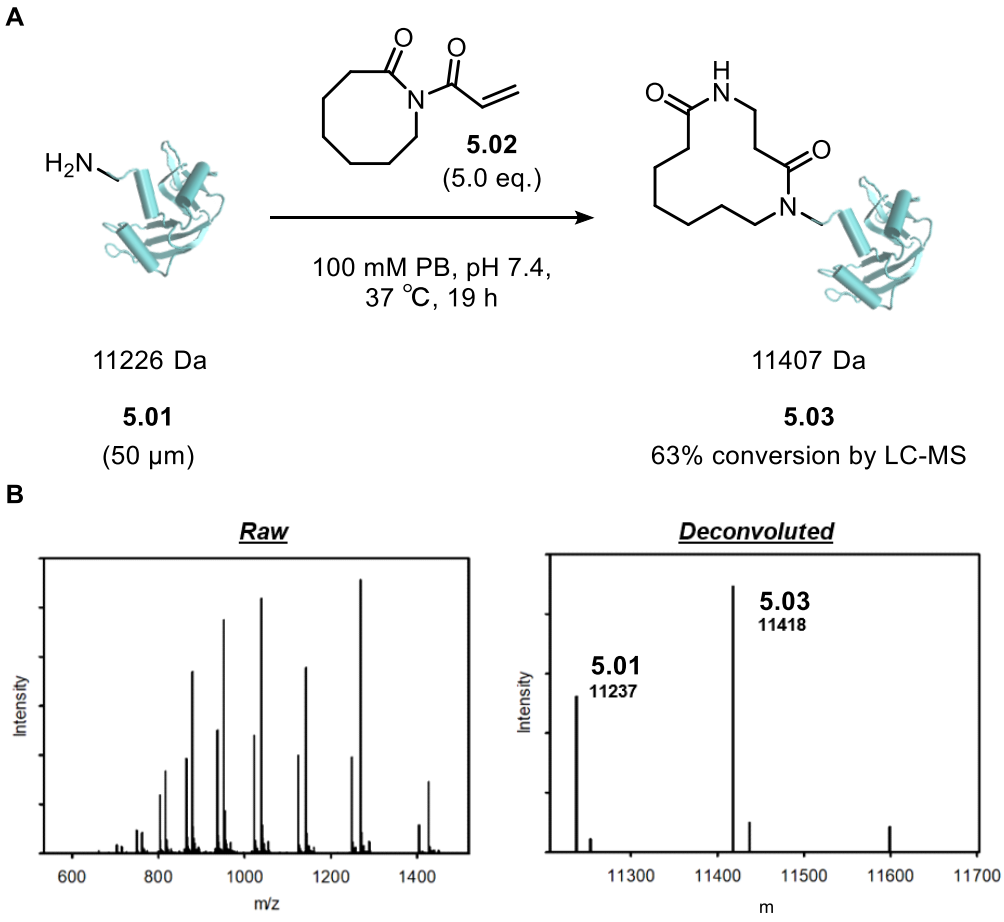


Figure 5.01: (A) CjX183-D bioconjugation with CARE using imide **5.02**. (B) LC-MS data obtained from the crude reaction mixture. **MS** (ESI$^+$): [**5.01**+H]$^+$ found 11237, calculated 11226; [**5.03**+H]$^+$ found 11418, calculated 11407.* Nb In all experiments using CjX183-D and mutants, the observed mass was ~11 Da higher than the theoretical mass and thus was taken as the unmodified protein mass in all instances. The reasons for this observation are unknown but are consistent with previous reports.[63] Experimental details can be found in Chapter 7.5.2.1

We were pleased to observe no doubly- or triply modified CjX183-D products by LC-MS. This result was not wholly surprising, considering the absence of lysine groups in CjX183-D, but nevertheless it was reassuring to find no evidence of modification at any of the four cysteine residues — two of which form disulfide bridges, whilst the other two are covalently bound to the heme unit. Next, we sought to increase the number of potential competing sites by using CjX183 R51K, a point-mutated derivative of **5.01** containing a single lysine residue. This allowed us to assess the *N*-terminal selectivity of our transformation *via* a direct comparison to the wild-type protein previously tested. CjX183 R51K (**5.04**) was subjected to the same conditions as those previously described, resulting in a comparable 59% conversion to modified products (**5.05**, Figure 5.02A). Of the 59% converted material, 91% was singularly modified (s) and 9% was doubly modified (d). In this instance, we attribute the formation of the doubly modified product to competing reactions at the newly introduced lysine residue, given the comparable conversion with **5.01** which lacks any internal lysine residues (Figure 5.02b).

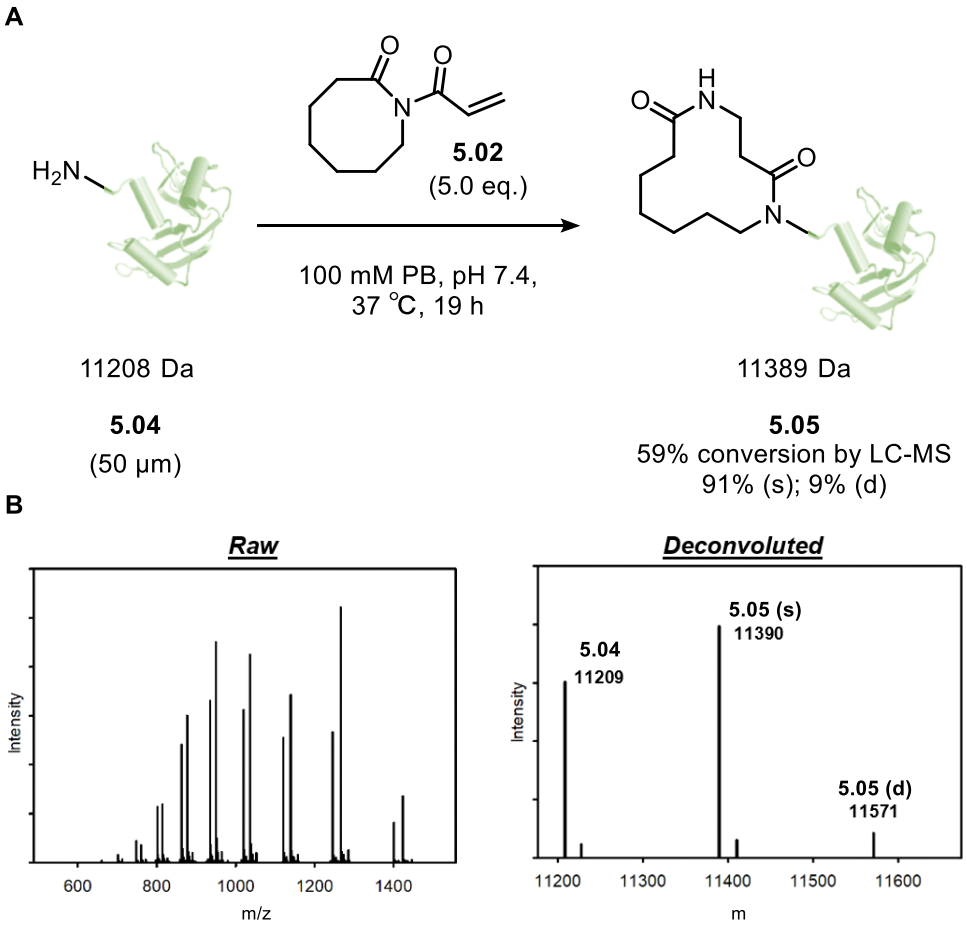


Figure 5.02: (A) CjX183 R51K bioconjugation with CARE using imide **5.02**. (B) LC-MS data obtained from the crude reaction mixture. **MS** (ESI^+): [**5.04**+H]$^+$ found 11209, calculated 11198; [**5.05(s)**+H]$^+$ found 11390, calculated 11379; [**5.05(d)**+H]$^+$ found 11571, calculated 11560. Experimental details can be found in Chapter 7.5.2.1

Encouraged by the clear evidence of protein modification with our acryloyl imide probes, we next applied the CARE chemistry to a wider range of proteins to further assess its versatility. Consequently, we applied the same reaction conditions to modify RNase A (PDB entry 1KF5), myoglobin (PDB entry 1MBN) and cytochrome C (PDB entry 2B4Z) using acryloyl imide **5.02**. As expected, a modest degree of double and triple modification was observed in all cases, likely due to competing reactions at lysine residues. However, we found that performing the reactions at room temperature, rather than 37 °C, improved the degree of single modification, albeit as a trade-off with reduced conversion (Figure 5.03). Nonetheless, given that selectivity remained our primary concern, all subsequent protein conjugation reactions were carried out at room temperature.

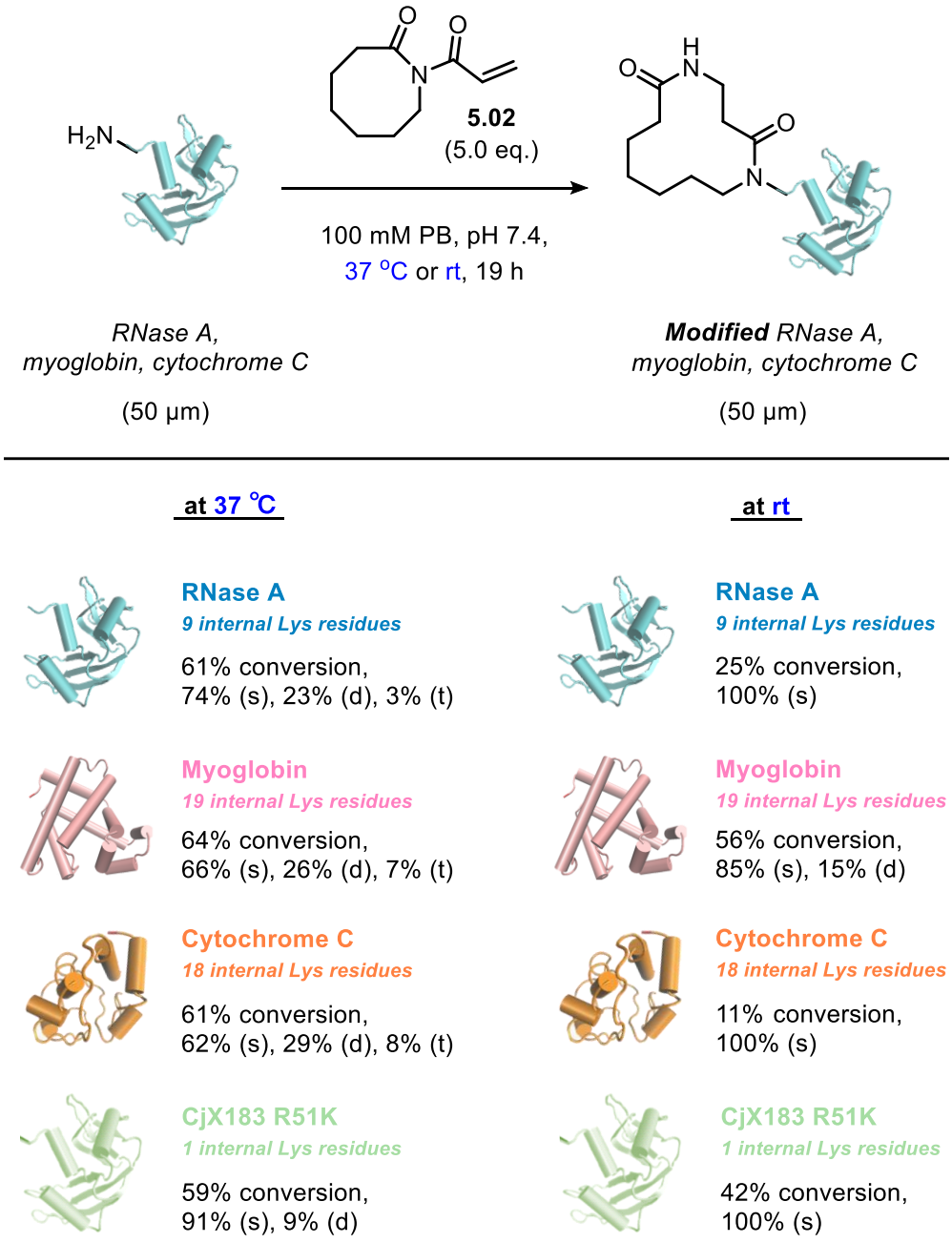


Figure 5.03: CARE bioconjugation of model proteins with imide **5.02** at 37 °C and rt. Experimental details can be found in Chapter 7.5.2.1

Based on these results, we further investigated the modification of RNase A by varying both the equivalents of **5.02** and the reaction pH. In total, we tested 5, 10, 20 and 50 equivalents of **5.02** across four pH values: 6.0, 6.5, 7.4, and 8.0. The data points can be grouped into four distinct quadrants based on a combination of selectivity and conversion, with thresholds set at 50% — crudely, values ≤50% were considered poor, whilst values ≥51% were considered good (Figure 5.04). As expected, increasing both the equivalents of **5.02** and the reaction pH led to greater conversion but reduced selectivity, consistent with the enhanced nucleophilicity of

lysine residues at higher pH due to their ε-amino group pK_a (~10.5). Notably, conversions of ~50% with 100% single modification on RNase A were achieved at pH 7.4 using 10 equivalents of **5.02** which is a significant improvement over the initial conditions tested.

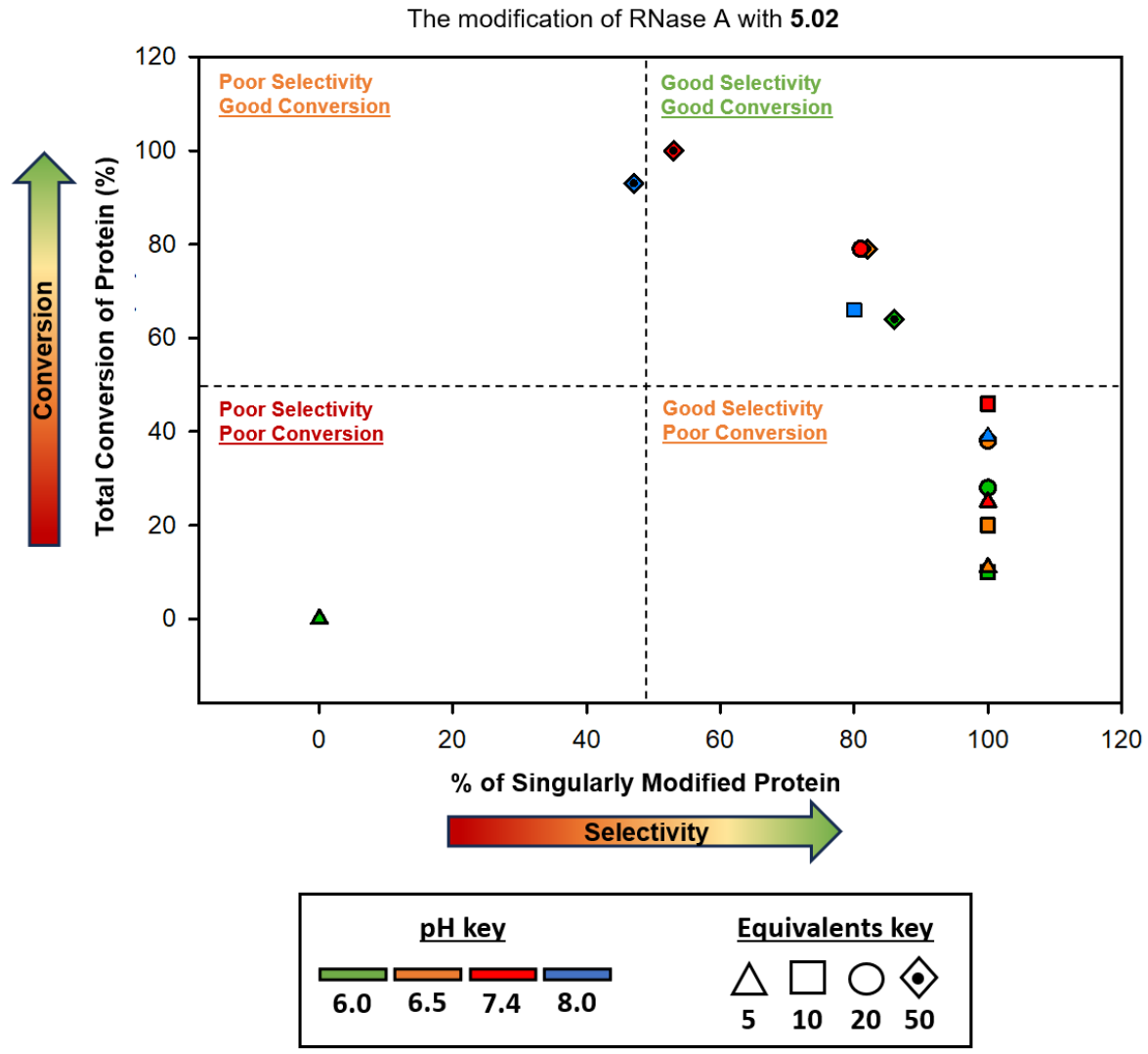


Figure 5.04: Screening conditions for RNase A bioconjugation with imide **5.02**. Reactions were run at 50 µM of RNase A at pH 6.0, 6.5, 7.4 and 8.0 with 5, 10, 20 or 50 equivalents of imide **5.02** (stock adjusted accordingly in DMSO) and incubated at rt, 1000 rpm for 19 h before an aliquot was taken and analysed by LC-MS. Experimental details can be found in Chapter 7.5.2.2 and tabulated data in Table 7.1.

A similar set of reactions were performed on cytochrome C and showed the same trend observed with RNase A — namely, that increasing both the pH and equivalents of **5.02** increased conversion but reduced selectivity (Figure 5.05). In this case, the best result achieved was ~40% conversion with 100% single modification on cytochrome C, obtained at pH 7.4 using 20 equivalents of **5.02**. This suggests that cytochrome C requires more forcing conditions

for effective modification and that good selectivity and good conversion are more difficult to achieve.

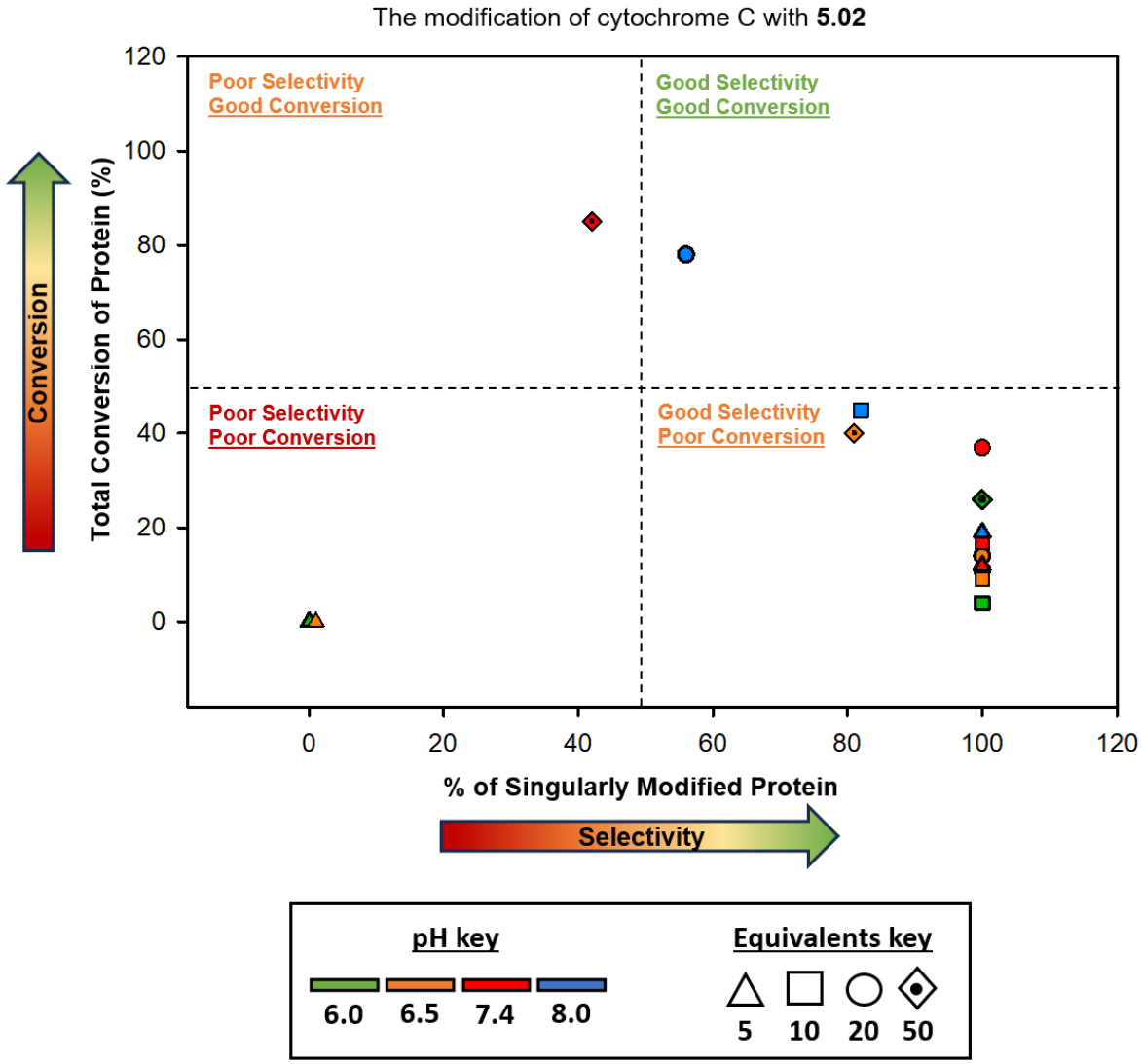
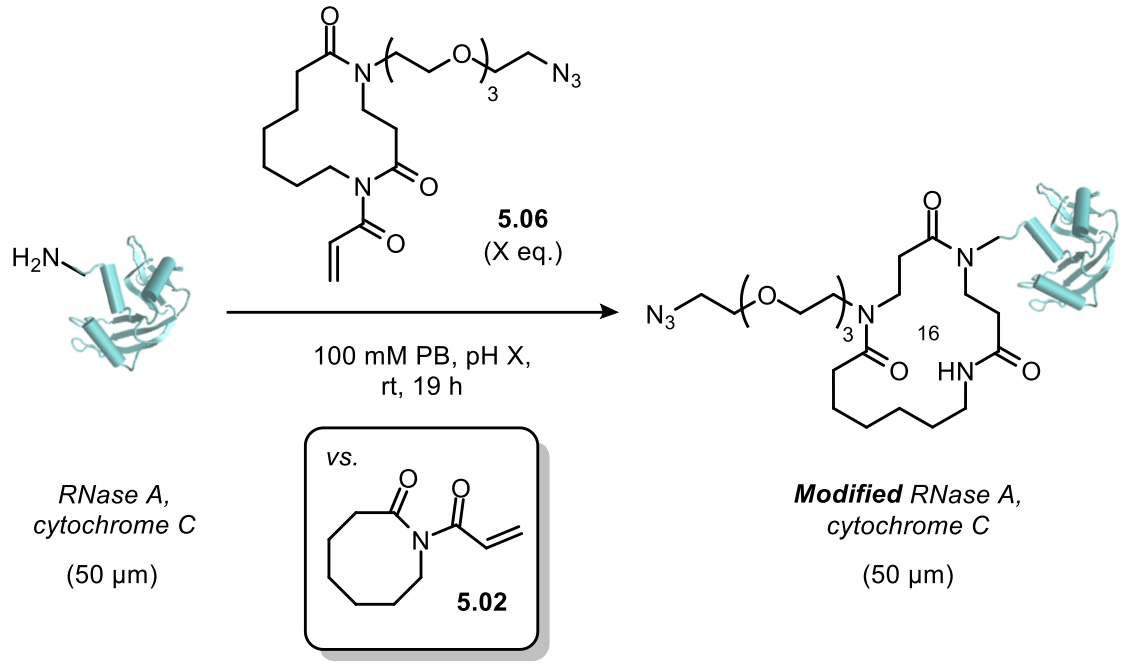


Figure 5.05: Screening conditions for cytochrome C bioconjugation with imide **5.02**. Reactions were run at 50 µM of cytochrome C at pH 6.0, 6.5, 7.4 and 8.0 with 5, 10, 20 or 50 equivalents of imide **5.02** (stock adjusted accordingly in DMSO) and incubated at rt, 1000 rpm for 19 h before an aliquot was taken and analysed by LC-MS. Experimental details can be found in Chapter 7.5.2.2.

5.2.2 Modifications with an azide-tagged acryloyl imide probe

Next, we turned our attention to demonstrating the utility of our approach by optimising the modification of RNase A using our azide- and alkyne-tagged acryloyl imide probes. We began by selecting the most efficient conditions previously identified for the modification of RNase A and cytochrome C and using these as a starting point for their modification with our azide-tagged acryloyl imide **5.06** (Scheme 5.01).



Scheme 5.01: CARE of RNase A and cytochrome C with imide **5.06** under different reaction conditions.

Table 5.01: Comparison of protein modification using imide **5.02** and **5.06** under various reaction conditions.

Protein	pH	Eq. of imide	Conversion with **5.02**(%)	Selectivity with **5.02**(%)	Conversion with **5.06**(%)	Selectivity with **5.06** (%)
RNase A	6.5	20	38	100 (s)	11	100 (s)
RNase A	7.4	10	46	100 (s)	10	100 (s)
RNase A	8.0	10	66	80 (s); 20 (d)	13	100 (s)
Cytochrome C	7.4	10	38	100 (s)	0	0
Cytochrome C	8.0	5	20	100 (s)	0	0

[a] Conversion was determined by LC-MS through the ratio of the deconvoluted peak's intensity for each species. For further details, see Chapter 7.5.2.2.

The data in Table 5.01 shows that, in all cases, greatly reduced or no modification was observed with **5.06** across the conditions tested. However, we were mindful of the finding described in Chapter 4.5.1, where the LC-MS trace of **5.06** was poorly resolved and accompanied by several smaller peaks throughout the spectrum. Assuming that these signals are from contaminants, it is likely that the effective concentration of **5.06** was lower than calculated, which may explain the reduced conversion. Nonetheless, no double or triple modifications were detected, suggesting that selectivity is conserved even with these larger, more functionally complex probes. Consequently, we repeated these experiments using increasing equivalents of **5.06** in an effort to drive the reaction towards completion. However, all repeated reactions were unsuccessful, with no conversion observed under any conditions. Upon further investigation,

we concluded that the stock solutions of **5.06** had degraded, accounting for the complete loss of reactivity. Whilst this issue could have been revisited, and potentially resolved, we instead shifted our focus to the alkyne-tagged acryloyl imide **5.07** in hopes of achieving a better outcome.

5.2.3 With an alkyne-tagged acryloyl imide probe

We evaluated a range of conditions for the modification of RNase A using 5, 10, 20 or 50 equivalents of acryloyl imide **5.07**, at reaction pH values of 6.0, 6.5, 7.4 and 8.0. These reactions were repeated using the same procedure as described with imide **5.02** (section 5.2.1), allowing us to make a direct comparison of the conversion and selectivity between the two probes. Interestingly, modification with **5.07** (Figure 5.06) resulted in a higher degree of single modification compared to **5.02** (Figure 5.04), as evidenced by the overall rightward shift of the data points. The conjugation of RNase A with imide **5.07** represents our first successful installation of an alkyne-functionalised, medium-sized ring *bis*-lactam selectively assembled at a protein's *N*-terminus, wherein the alkyne group provides a handle for subsequent post-CARE modification *via* CuAAC.

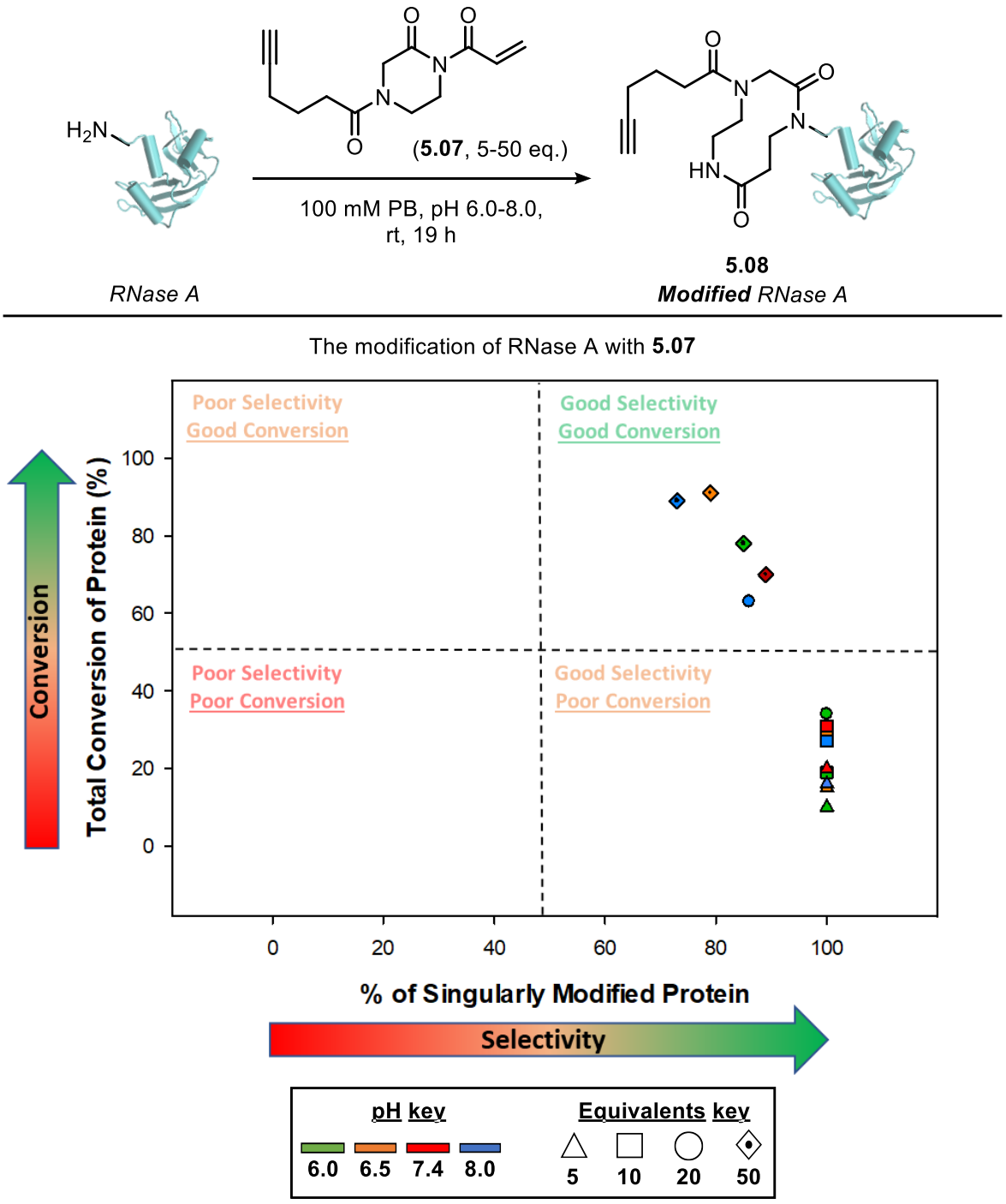


Figure 5.06: Screening conditions for RNase A bioconjugation with **5.07**. Reactions were run at 50 µM of RNase A at pH 6.0, 6.5, 7.4 and 8.0 with 5, 10, 20 or 50 equivalents of imide **5.07** (stock adjusted accordingly in DMSO) and incubated at rt, 1000 rpm for 19 h before an aliquot was taken and analysed by LC-MS. Experimental details can be found in Chapter 7.5.2.2 and tabulated data in Table 7.2.

5.3 JVZ-007 as a substrate for CARE modification

To pursue a more compelling downstream application, imide **5.07** was next used for the modification of nanobody JVZ-007.[163] This nanobody is capable of targeting surface prostate-specific membrane antigen (PSMA) on lymph node carcinoma of the prostate (LNCaP) cells

by binding with high affinity (K_d = 27.4 nM) to its extracellular domain.[163] PSMA is a surface protein which shows an increase in expression linked with prostate tumour grades and poor patient survival,[164] and a promising marker already clinically targeted by the ProstaScint® scan diagnostic technology. Consequently, JVZ-007 bioconjugates have been employed in various applications related to the imaging and treatment of PSMA-overexpressing prostate cancers.[165,166] We once again began by carrying out a comprehensive screening of reaction conditions to identify the most efficient for modification of JVZ-007 (Figure 5.07).

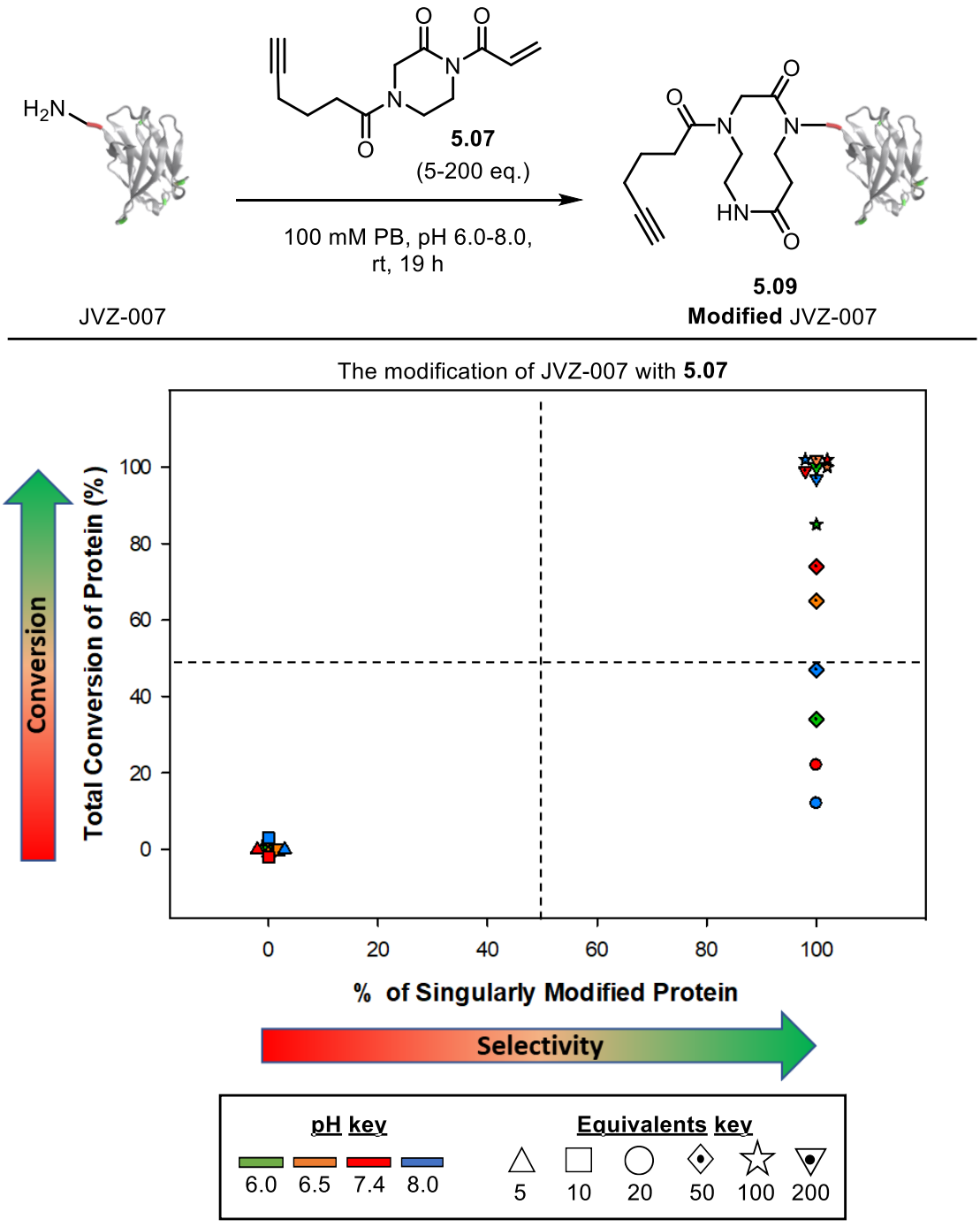


Figure 5.07: Screening conditions for JVZ-007 bioconjugation with imide **5.07**. Reactions were run at 50 µM of JVZ-007 at pH 6.0, 6.5, 7.4 and 8.0 with 5, 10, 20, 50, 100 or 200 equivalents of imide **5.07** (stock adjusted accordingly in DMSO) and incubated at rt, 1000 rpm for 19 h before an aliquot was taken and analysed by LC-MS. Experimental details can be found in Chapter 7.5.2.2 and tabulated data in Table 7.3.

Figure 5.07 shows that several of the conditions resulted in excellent conversion and selectivity with imide **5.07**. However, it is clear that this protein required more forcing conditions to initiate modification, as milder conditions (<20 equivalents of **5.07**) did not result in any

detectable modification. That said, this was not a significant concern, as even at 200 equivalents of **5.07**, less than 5% double modification was observed.

5.3.1 Confirming *N*-terminal site-selectivity of our approach

Having successfully established optimal conditions for the complete modification of JVZ-007, the next priority was to confirm the site of modification and ensure that no hyper-reactive lysine residues were being inadvertently modified. To do this, we collaborated with proteomics specialist Adam Dowle at the University's Bioscience Technology Facility, who performed intact MALDI-MS analysis and verification of *N*-terminal modification *via* Arg-*C* digestion on purified bioconjugates. Intact MALDI-MS analysis found <10% double modification of our JVZ-007 bioconjugates (*m/z* 12,618; Figure 5.08), with a clear preference for single modification (*m/z* 12,371; Figure 5.08).

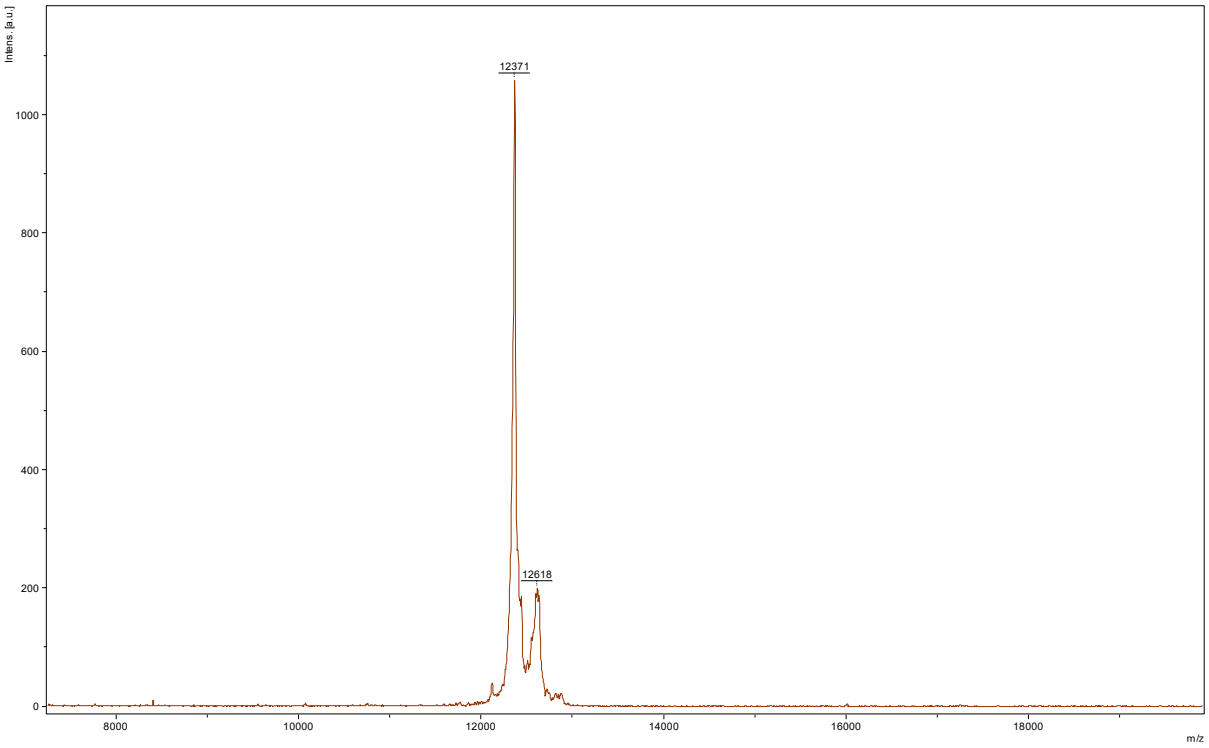


Figure 5.08: Intact protein M_W by MALDI of JVZ-007 conjugated with **5.07** found two *m/z* values at 12,371 Da [**5.09**(s)+H]$^+$ and 12,618 Da [**5.09**(d)+H]$^+$ in >10:1 ratio. Experimental details can be found in Chapter 7.5.2.3.

Furthermore, MALDI-ISD sequencing provided strong evidence that the modification occurred at the *N*-terminal residue, as shown in Figure 5.09. The MALDI-ISD spectrum can be read from left to right, with the most intense fragments deriving from residues closest to the *N*-terminus. We observe a deviation from the expected sequence mass from the first assignable sequence from the *N*-terminus (at *m/z* 1089). Based on succeeding sequence annotation, this signal is expected to be that of *N*-terminal sequence SEVQLVE — the theoretical non-modified

mass of this fragment would be 802.71 Da, meaning we are observing a mass addition of 286.24 Da [**5.09**(s)+K]$^+$ in the first 7 amino acids. This is corroborated by all subsequent fragment ions emanating from the *N*-terminus up until Arg-27, where we begin to see a mass addition of just 248 Da. This observation suggests that all ions in the *N*-terminal region preceding Arg-27 are potassiated ([**5.09**(s)+K]$^+$), and at the basic Arg-27 residue, this tendency switches to the usual protonated ([**5.09**(s)+H]$^+$) form. Nonetheless, the most plausible interpretation of these complex annotations is a 248 Da mass addition localised within the first seven amino acids — consistent with a single modification by **5.07**. Given that the only primary amine present within the SEVQLVE sequence is the α-amino group, and that no other nucleophilic amino groups are available to compete, it is highly likely that the modification had occurred at the intended site.

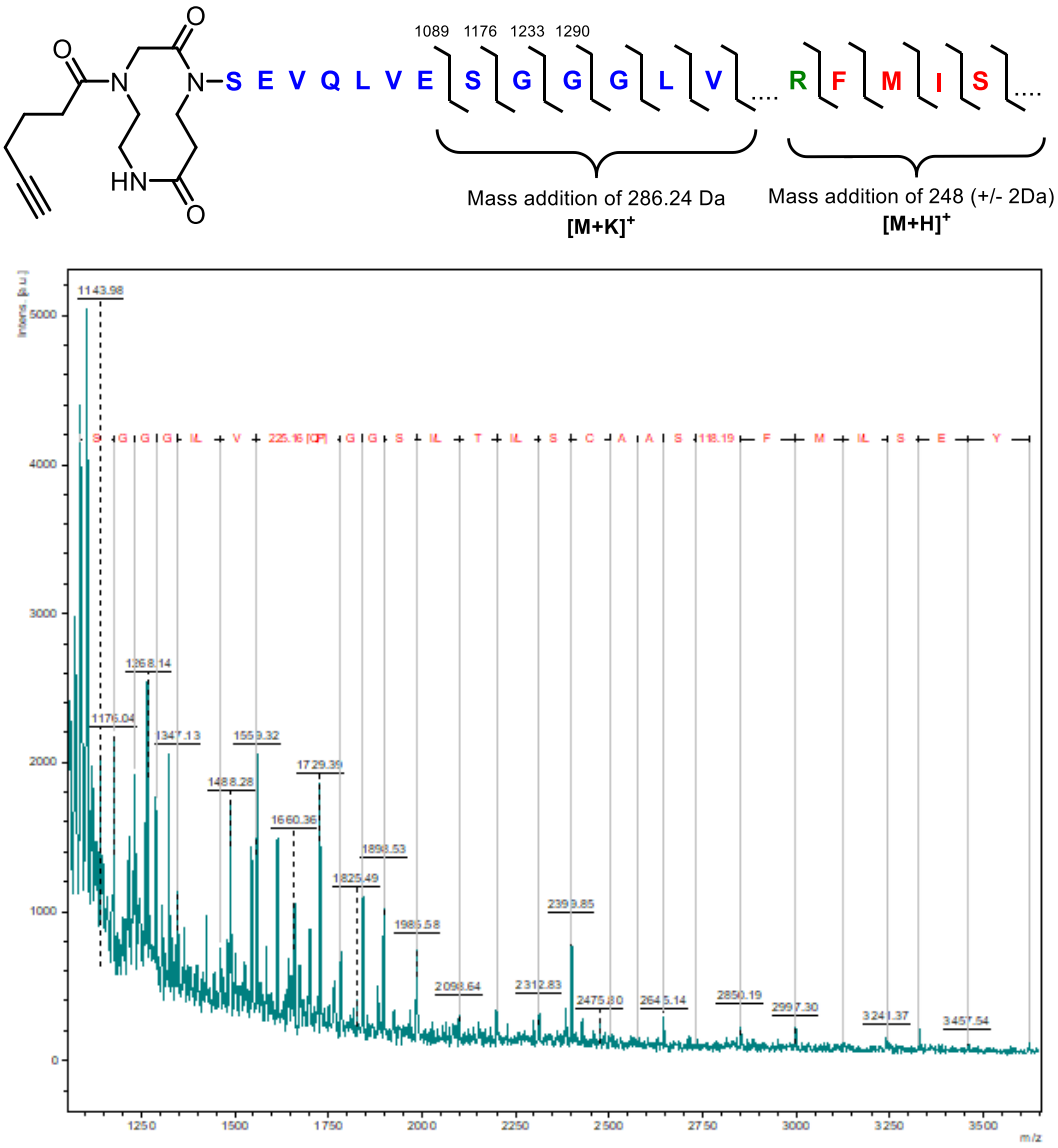


Figure 5.09: MALDI-ISD N-terminal sequencing spectrum of JVZ-007 + **5.07** conjugates found the *N*-terminal SEVQLVE sequence consistent with a mass addition of 284 Da [**5.09**(s)+K]$^+$ as would be expected for our modification at the *N*-terminus. Experimental details can be found in Chapter 7.5.2.4.

5.3.2 Assessing CARE bioconjugate stability

We next turned our attention to evaluating the stability of our bioconjugates, as several reviews of *N*-terminal bioconjugation highlight bioconjugate instability as a major limitation.[63,94] We anticipated that the introduction of an irreversible ring expansion step would mitigate degradation caused by competing reverse reactions. To assess conjugate stability, the modified JVZ-007 conjugates (**5.09**) were dialysed overnight at 37 °C. Following incubation, no evidence of degradation to the unmodified protein was observed, and the singularly modified JVZ-007 product was fully retained. Previous work within the Spicer group found that under the same conditions, bioconjugates deriving from a range of 2-pyridinecarboxyaldehydes (2-

PCAs), 2-ethynylbenzaldehydes (2-EBAs) and triazolecarbaldehydes (TA4Cs) reagents, commonly used for *N*-terminal modification, degraded by 55–85% after 48 hours.[63]

5.3.3 Post-CARE 'click' functionalisation

With optimal conditions for the site-selective modification of JVZ-007 established, we next explored post-CARE functionalisation — specifically, *via* the CuAAC 'click' reaction. Having successfully installed an alkyne-functionalised medium-sized *bis*-lactam at the *N*-terminus of JVZ-007, we were equipped to perform a CuAAC 'click' coupling with Alexa Fluor™ 647 azide, a far-red fluorescent dye widely used in flow cytometry and imaging.[167,168] Samples of **5.09** were first purified by multiple rounds of dialysis into 100 mM phosphate-buffered saline (PBS) at pH 7.4 to remove any residual **5.07** remaining in the sample. The conjugates were then treated with an excess of Alexa Fluor™ 647 azide in the presence of sodium ascorbate and $CuSO_4$, and the reaction mixture was gently agitated while protected from light. After 45 minutes, the reaction was stopped, and the mixture was subsequently purified by sequential rounds of dialysis into Milli-Q water. To confirm the successful conjugation of AF-647 to **5.09**, the protein was analysed by tricine SDS-PAGE, which showed a distinct pink fluorescent band at the expected molecular weight for **5.10** (*m/z* ~ 13,412 Da).

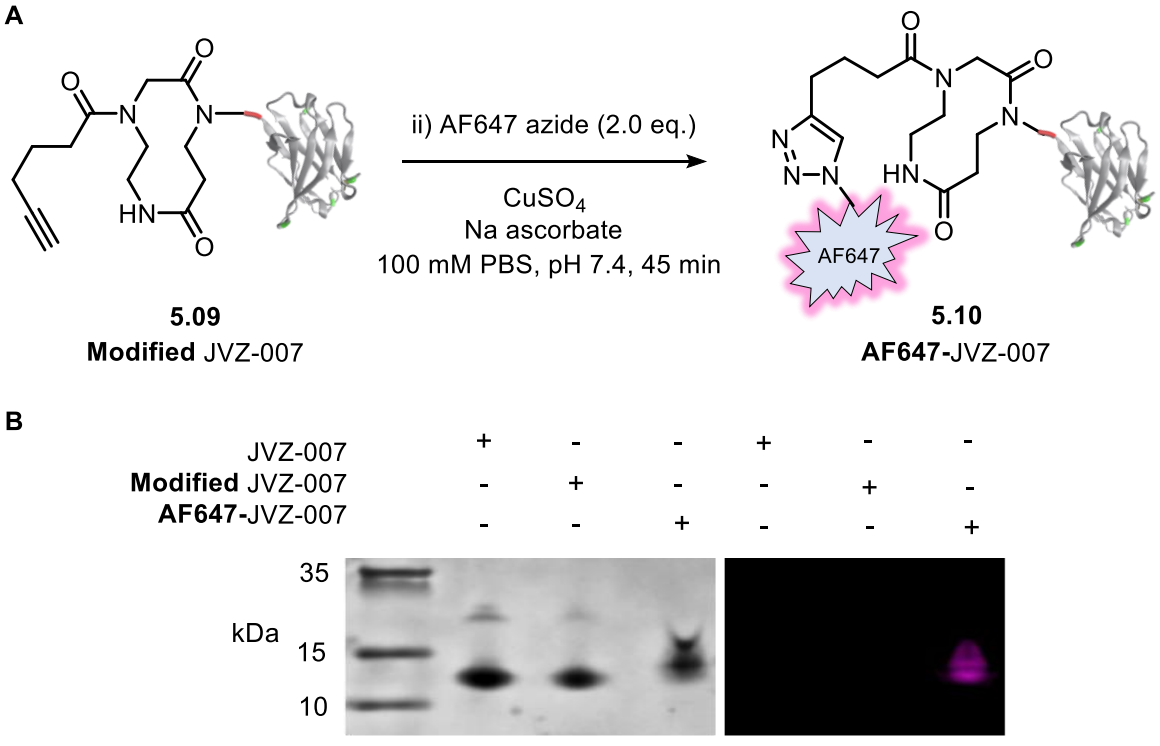


Figure 5.10: Fluorescent tagging of **5.09** with Alexa Fluor™ 647. (A) Reaction Scheme. (B) Analysis of 5.10 by TSDS-PAGE and visualisation by Coomassie staining (left) and fluorescent imagine (right). Experimental details can be found in Chapters 7.5.2.6 and 7.5.2.7.

We postulate that the second, higher-positioned spot corresponds to the <10% of doubly modified JVZ-007 conjugates, therefore potentially bearing two AF-647 fluorophores.

5.3.4 Immunofluorescent staining of prostate cancer tissue slices

As previously discussed, JVZ-007 selectively binds to the PSMA which is overexpressed on the surface of prostate cancer cells.[165] Our fluorescently labelled JVZ-007 conjugates (**5.10**) were used to visualise the PSMA on the surface of LNCaP cells using fluorescence microscopy (Figure 5.11).[a]

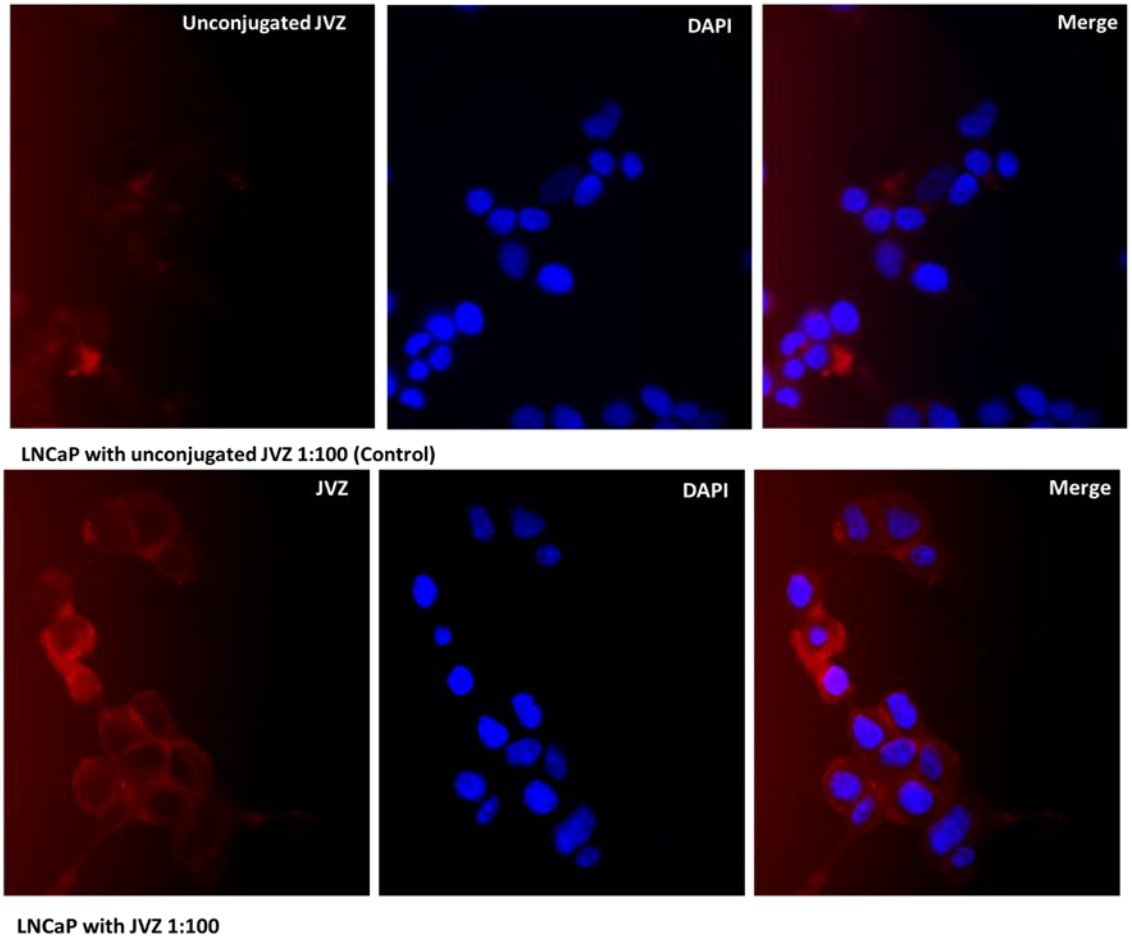


Figure 5.11: Fluorescent microscopy images of PSMA on LNCaP prostate cancer cells visualised using fluorescently labelled JVZ-007 (red) and DAPI (blue) (**5.10**, bottom row**)**. Experimental details can be found in Chapter 7.5.3.1

As anticipated for binding to cell surface PSMA, the fluorescence microscopy images show a membrane localised fluorescent signal when using the labelled JVZ-007 conjugate (**5.10**), in comparison to the DAPI nuclear stain. Encouraged by these promising preliminary images,

[a] Immunofluorescence experiments were kindly carried out by Esme Hutton.

subsequent immunofluorescent staining was performed on prostate cancer (PCa) tissue slices and compared with tissue slices from patients with benign prostatic hyperplasia (BPH), which does not overexpress PSMA at its surface. Figure 5.12 shows the immunofluorescent staining of these tissue samples and confirms that the labelled JVZ-007 conjugates (**5.10**) retained their activity and binding properties following both *N*-terminal CARE modification and CuAAC. This is evidenced by strong membranous staining in the PCa tissues, again consistent with PSMA overexpression at the cell surface. In contrast, the BPH samples exhibited markedly reduced staining, as expected given the lower levels of PSMA expression on their surface. Notably, there was a complete lack of staining in a control breast cancer sample confirming the preservation of JVZ-007's selectivity for PSMA.

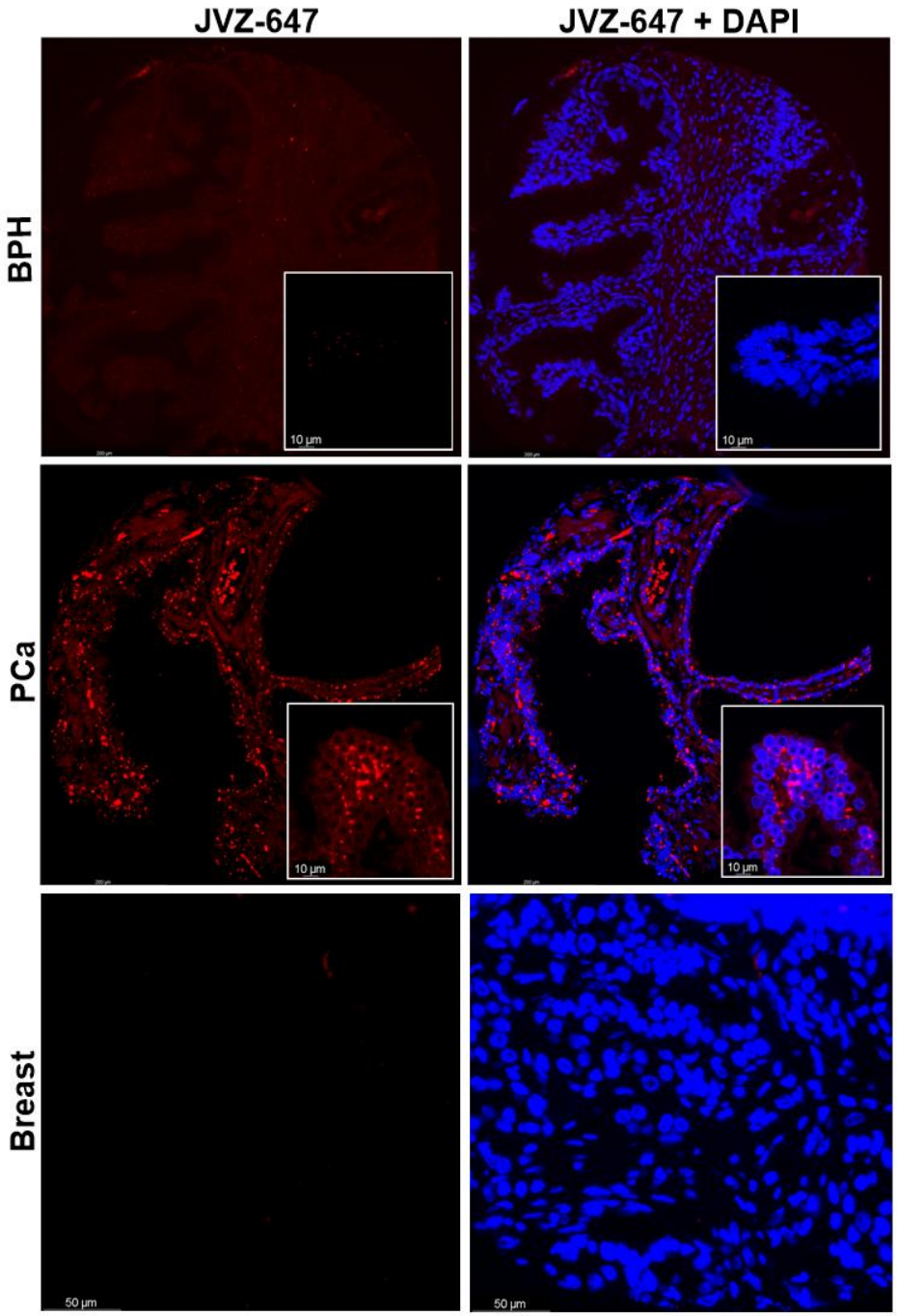


Figure 5.12: Immunofluorescent staining of BPH, PCa and breast cancer tissue using the fluorescent nanobody AF647-JVZ-007 (**5.10**) (left, red) and DAPI nuclear counterstain (right, blue). Experimental details can be found in Chapter 7.5.3.1.

5.4 CCL5 as a substrate for CARE modification

We sought a protein target that would allow us to explore the versatility of the CARE approach in a different context, instead focussing on the assembly of functionalised macrocycles of diverse size and structure. In this context, we identified human CCL5 chemokine (also known as RANTES) as our next target for modification. CCL5 (PDB entry 5COY) is a small (~ 9 kDa)

pro-inflammatory chemokine that attracts immune cells towards sites of inflammation through binding to G-protein coupled chemokine receptors on the surface of immune cells, CCR5 (Figure 5.13).[169] This ligand-receptor interaction represents a promising therapeutic target in a range of pathological conditions, including HIV, where CCR5 functions as the principal co-receptor for viral entry into host cells.[170-173] Interestingly, the *N*-terminal region of CCL5 plays a critical role in this event by mediating a high-affinity binding interaction deep within the extracellular binding pocket of CCR5.[174] As a result, antagonists that disrupt this interaction are of significant interest in the development of anti-HIV therapies. We hypothesised that assembling derivatised macrocycles of varying sizes and structure at the *N*-terminus of CCL5 could modulate the downstream signalling events — such as directed cell migration or chemotaxis — associated with CCR5 activation. The rationale behind this idea is informed by previous work on 5P12-RANTES, a modified CCL5 analogue with an altered *N*-terminus that disrupts dimerisation, rendering it a potent inhibitor of HIV infection.[170]

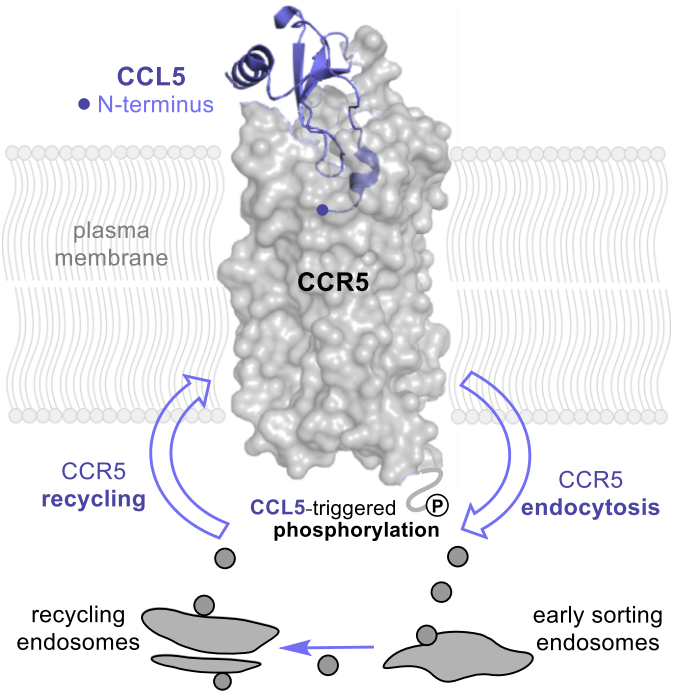


Figure 5.13: Wild type CCL5 *N*-terminal insertion and binding into the extracellular CCR5 cavity inducing downstream signalling events.

5.4.1 Establishing optimal conditions for modification

In the first instance, CCL5 was subjected to an *N*-terminal CARE modification using the underivatised six-membered acryloyl imide **5.11**, to identify the optimal conditions for efficient and site-selective modification. Several sets of conditions were tested by varying the

equivalents of **5.11** and the reaction pH, with full conversion and complete selectivity achieved using 75 equivalents of **5.11** at pH 6.0 (Figure 5.14).

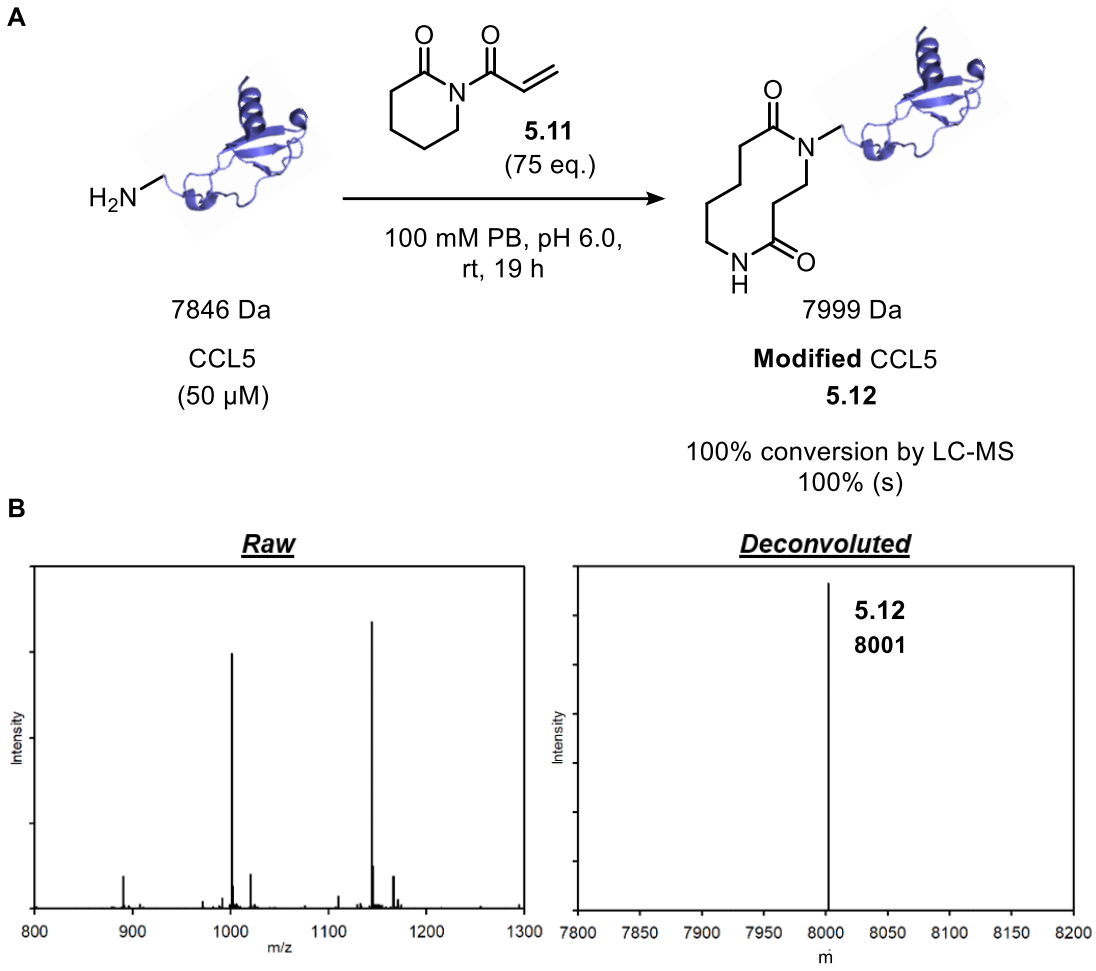
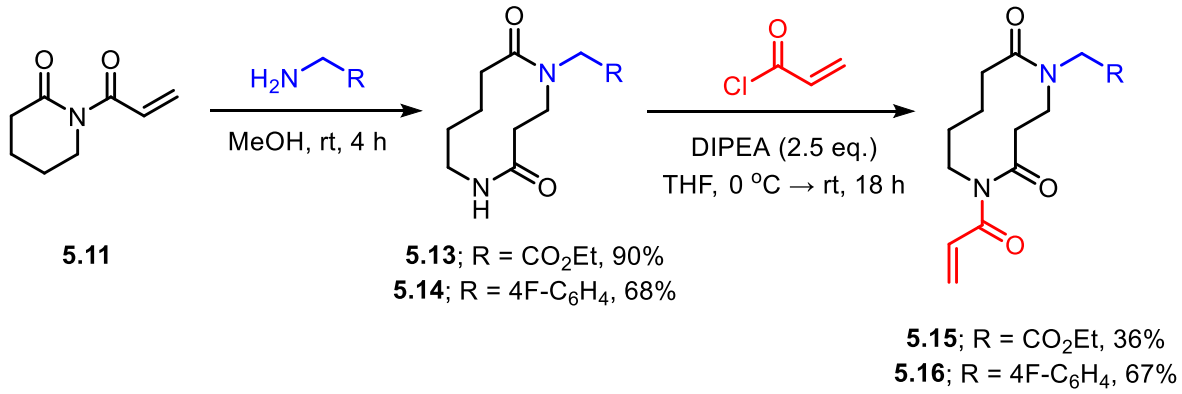


Figure 5.14: (A) CCL5 bioconjugation with CARE using **5.11**. (B) LC-MS data obtained from the crude reaction mixture. **MS** (ESI$^+$): [**5.012**+H]$^+$ found 8002, calculated 7999. Experimental details can be found in Chapter 7.5.2.2.

5.4.2 Synthesising structurally and functionally diverse bioconjugates

A key advantage of the CARE chemistry is its tolerance of iterative reactions, which allows successive ring expansions on the same system and, in some cases, enabling access to macrocycles with >20 atoms.[50] With this in mind, we sought to investigate the tolerance of our CARE bioconjugation strategy using larger acryloyl imide probes, with the aim of assembling more structurally diverse macrocycles at the *N*-terminus of CCL5. To make these probes, we chose to carry out 'off-protein' CARE reactions with **5.11** under previously established conditions. The resulting 10-membered *bis*-lactam products were subsequently re-acylated

with acryloyl chloride to regenerate the key acryloyl imide motif required for CARE reactivity (*e.g.* **5.13** → **5.15**, Scheme 5.02).



Scheme 5.02: 'Off-protein' CARE reactions towards the synthesis of structurally diverse acryloyl imides.

Acryloyl imides **5.11**, **5.15** and **5.16**, probes with varying size and functionalisation, were then employed for the CARE modification of the *N*-terminus of CCL5. In the cases of **5.15** and **5.16**, this represented a second, successive 'on-protein' ring expansion, resulting in the formation of a 14-membered macrocycle at the targeted site of reactivity. Acryloyl imide **5.15** was used to optimise the reaction conditions for these larger ring systems (Entries 2–9, Table 5.02), and the most effective conditions were subsequently applied to the modification of CCL5 with **5.16** (Entry 11, Table 5.02). The most effective conditions are summarised in Figure 5.15 along with the structure of each of the resulting modified CCL5 products (**5.12**, **5.17** and **5.18**).

Table 5.02: Comparison of CCL5 modification with imide **5.11**, **5.15** and **5.16** under various conditions.

Entry	Imide	pH	Eq. of imide	Conversion (%)	Selectivity (%)	Product
1	**5.11**	6.0	75	100	100 (s)	**5.12**
2	**5.15**	6.0	10	77	100 (s)	**5.17**
3	**5.15**	6.0	20	69	100 (s)	**5.17**
4	**5.15**	6.0	50	100	100 (s)	**5.17**
5	**5.15**	6.5	10	79	90–95 (s); 5–10 (d)	**5.17**
6	**5.15**	6.5	20	100	90–95 (s); 5–10 (d)	**5.17**
7	**5.15**	6.5	50	100	80–85 (s); 15–20 (d)	**5.17**
8	**5.15**	7.4	10	82	90–95 (s); 5–10 (d)	**5.17**
9	**5.15**	7.4	20	100	85–90 (s); 10–15 (d)	**5.17**
10	**5.15**	7.4	50	100	50 (s); 50 (d)	**5.17**
11	**5.16**	6.0	50	100	60 (s); 40 (d)	**5.18**

[a] Conversion was determined by LC-MS through the ratio of the deconvoluted peak's intensity for each species. For further details, see Chapter 7.5.2.2.

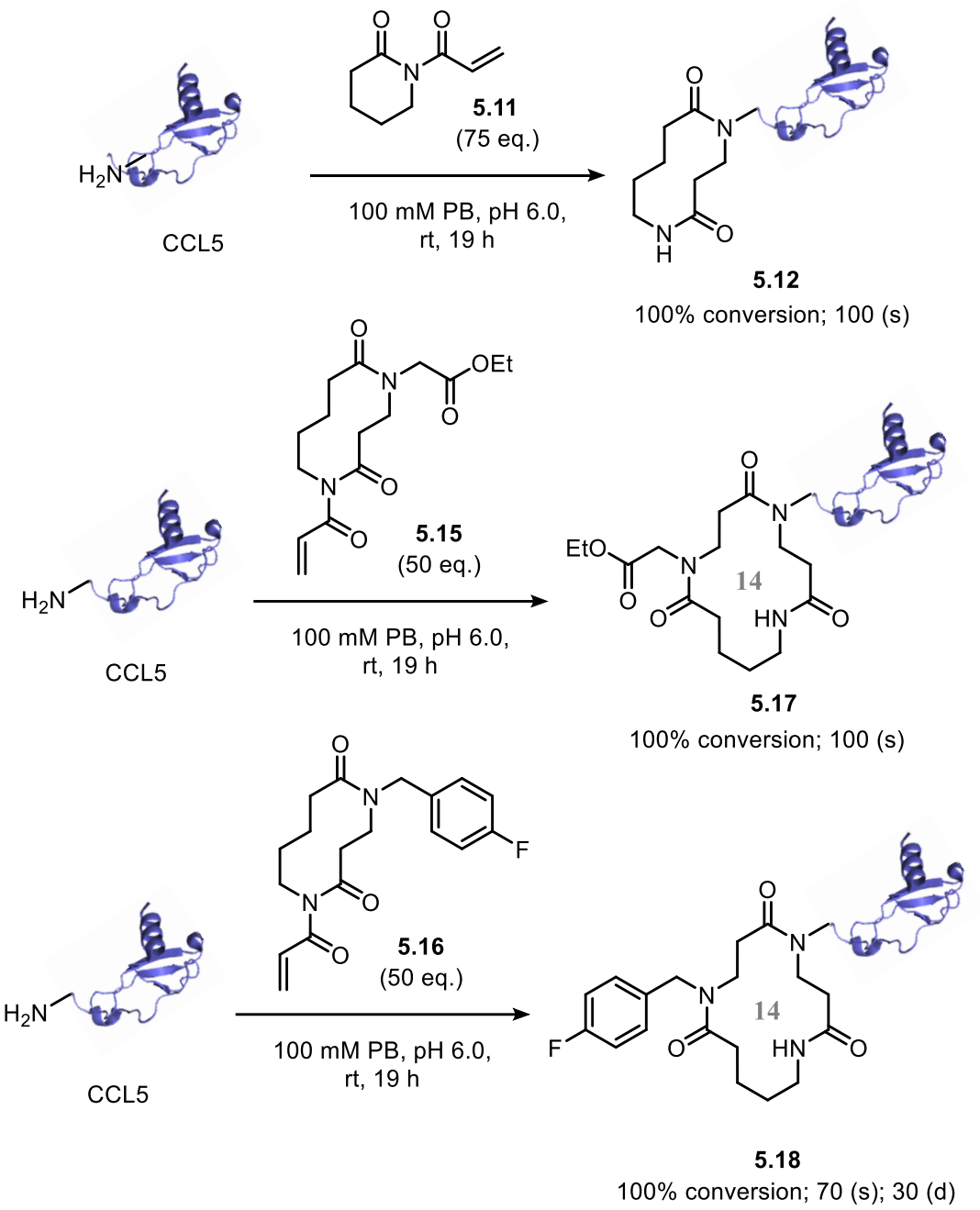


Figure 5.15: Optimised conditions for the *N*-terminal CARE reactions of CCL5 to afford 10-membered ring conjugate **5.12**, and 14-membered ring conjugates **5.17** and **5.18**.

In all cases, we were able to optimise the reaction conditions to enable complete conversion, and to access functionally and structurally diverse CCL5 bioconjugates for further investigation. The samples were dialysed to remove excess acryloyl imide reagent and to store it in a buffer in which the chemokine is most stable — first into 50 mM sodium acetate (pH 4.5), then into 1% acetic acid, and finally into 0.1% TFA — before being lyophilised and stored at −80 °C. A sample was submitted for MALDI-ISD *N*-terminal sequencing analysis; however, the results were inconclusive due to a complete absence of signal of the spectrum. Whilst this

means we cannot conclusively confirm that the modification is located at the *N*-terminus of CCL5, the similarity to our previous results with JVZ-007, which showed a comparable modification pattern, gives us confidence that the same site is being modified. Nevertheless, these samples were passed on to another member of the Fascione group, Afzaal Tufail, who carried out extensive investigations into the effects of *N*-terminal macrocycle assembly on the modulation of cell surface chemokine receptor downstream signalling.

5.4.3 The downstream activation effects of CCL5-CARE bioconjugates

The CCL5 bioconjugates **5.12**, **5.17** and **5.18** were used in a comparative exploration of CCL5-CCR5 binding and receptor recycling, relative to the wild-type chemokine. This began with an investigation into the agnostic properties of each bioconjugate, to determine whether the assembly of a relatively sterically bulky macrocycle at the ligand-receptor interaction site disrupts binding or activation of the complex. Pleasingly, each of the CCL5 bioconjugates exhibited agonistic properties comparable to wild-type CCL5, suggesting that the CARE modifications do not disrupt binding within the CCR5 ligand pocket. However, notable differences were observed between wild-type CCL5 and the CCL5-CARE bioconjugates in the downstream signalling events triggered upon ligand-receptor binding. In all instances, the modified CCL5 conjugates exhibited reduced receptor internalisation (Figure 5.16A), recycling (Figure 5.16B), and phosphorylation (Figure 5.16C), with the 14-membered aryl fluoride derivative, **5.18**, showing the most pronounced effect in all cases. Perhaps most notable is the complete loss of Ser349 phosphorylation which is typically mediated by G-protein coupled receptor kinases (GRKs) prior to agonist-induced endocytosis.[167] In this respect, **5.18** behaves similarly to 5P12-RANTES, which, as previously discussed, does not induce CCR5 phosphorylation. That being said, unlike 5P12-RANTES, **5.18** still activates the ligand-receptor complex towards endocytosis meaning the installation of macrocycles at the *N*-terminus of CCL5 could provide a novel approach to dissecting these complex chemokine ligand-receptor interactions. Broadly, these results suggest that CCR5-CARE bioconjugates could be used to remove CCR5 from human cell surfaces without inducing the detrimental downstream signalling events.[175,[b]]

[b] All experiments in this section (5.4.3) were carried out by Afzaal Tufail.

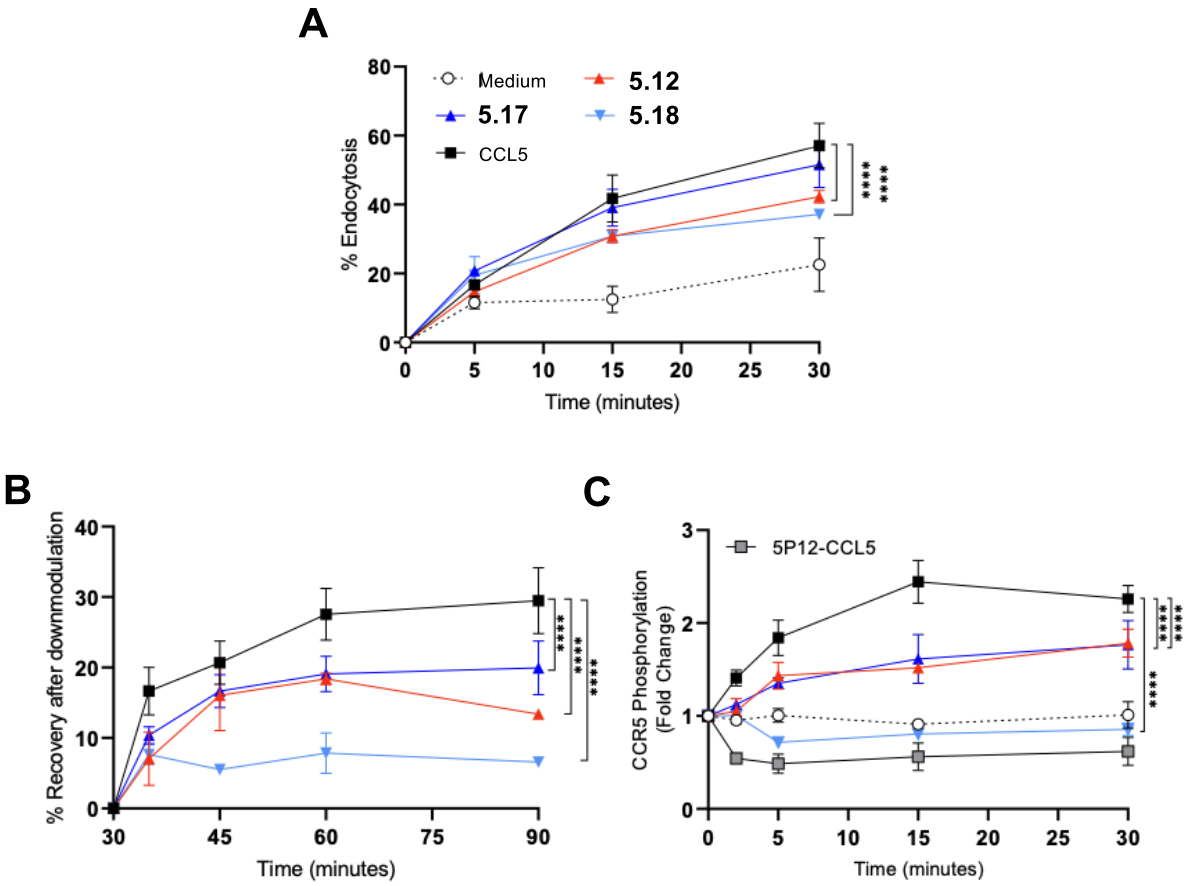


Figure 5.16: Kinetics of CCR5 endocytosis (A), recycling (B), and phosphorylation (C) upon treatment with CCL5 derivative. **** $P \leq 0.0001$ two-way ANOVA with Dunnett secondary test. Experimental details can be found in Chapter 7.5.3.2.

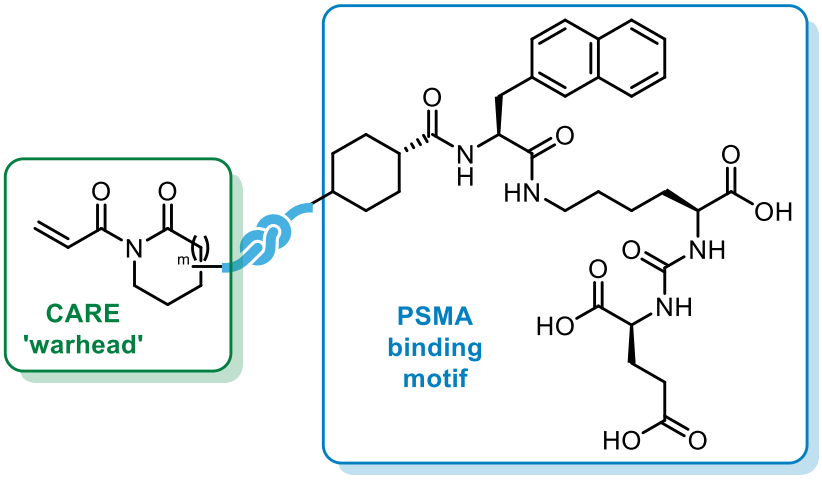
5.5 Chapter summary

In summary, this Chapter details the successful application of our novel *N*-terminal bioconjugation strategy, using CARE, for the modification of a range of medicinally relevant proteins. We have demonstrated how easy-to-prepare acryloyl imide reagents can be used to modify protein *N*-termini with high selectivity over competing ε-amino groups and other peptidyl nucleophiles. We confirmed the selectivity of our approach using MALDI-ISD *N*-terminal sequencing, which identified the site of modification, and demonstrates the superior stability of our bioconjugates — addressing one of the major challenges associated with other strategies currently available. That said, our findings underscore that there is no 'one-reaction-fits-all' approach to protein bioconjugation; rather, achieving full efficiency with this chemistry requires careful optimisation of the reaction conditions for each protein of interest. Lastly, we have demonstrated the utility of this approach through the successful modification of proteins involved in cancer cell and tissue imaging, as well as human chemokine receptor activation.

Chapter 6: Future work and outlook

The work presented in this Thesis has demonstrated the use of cyclic acryloyl imide reagents for *N*-terminal selective protein bioconjugation through conjugate addition/ ring expansion (CARE) cascade reactions. Whilst our work offers a novel, versatile tool for selective *N*-terminal protein bioconjugation, this strategy also reveals considerable scope for further development, with some clear opportunities to improve and expand its utility even further.

A priority for future work is the development of new, more functionally diverse acryloyl imide reagents. Whilst the alkyne-tagged acryloyl imide probe developed in this Thesis exhibited robust selectivity and stability, strategically varying its structure and functionalisation would provide a deeper understanding of the chemistry's tolerance towards diversification in future applications. This includes exploring a wider range of ring sizes, particularly the formation of larger macrocycles, and functionalising the reagents with biologically relevant groups such as affinity tags or cell-targeting motifs, including those that mimic the antigen-binding sequences of specific nanobodies. One promising example is the incorporation of a PSMA-targeting glutamate-urea peptide scaffold, such as that previously reported by the Fascione group.[176] Functionalising our acryloyl imide probes with this anti-PSMA motif (**6.01**) would enable the targeted delivery of CARE-modified sialidases to PSMA-overexpressing PCa cells. In principle, the delivered sialidases could cleave sialic acid residues associated with PSMA glycosylation on the cell surface, thereby potentially altering how these cancer cells are recognised by the human immune system (Figure 6.01).



6.01

Figure 6.01: A conceptual proposal of the structure of an anti-PSMA motif-containing acryloyl imide probe for *N*-terminal CARE modification.

Furthermore, a more comprehensive exploration of substrate scope would provide important insights into the generality and limitations of this approach to *N*-terminal bioconjugation. This includes investigating larger proteins with a broader range of *N*-terminal residues and diverse *N*-terminal environments, particularly with respect to solvent exposure and accessibility. In addition, a more comprehensive investigation into the effect of free cysteine residues on the CARE chemistry would provide important context for its wider application in biotechnological research. Although our studies in Chapter 3 indicated that free cysteine thiols can interfere with the reaction, we did not observe any noticeable off-target cysteine modification in the proteins tested. This may reflect specific structural features of the proteins already tested, and a more extensive substrate scope would help determine how disruptive free thiols are likely to be on *N*-terminal selective CARE modification. Moreover, identifying more protein targets whose function depends on key protein-protein interactions mediated at their *N*-terminus would be extremely valuable. As shown in this work with CCL5/CCR5, *N*-terminal macrocycle assembly can modulate key protein-proteins interactions and offers a unique approach to probing their mechanisms. An example of particular interest is the interaction between CXCL12 and its chemokine receptor CXCR4, which is widely regarded as a key regulator of cell migration and metastasis in several highly invasive cancers.[177,178] This interaction involves a deep insertion of the CXCL12 *N*-terminus into the binding pocket of CXCR4, resulting in a complex downstream signalling cascade. Investigating such system would allows us to further assess whether macrocycle assembly at the *N*-terminus can disrupt these interactions.

In summary, the findings from this Thesis provide a strong foundation for further exploration of conjugate addition/ ring expansion (CARE)-based bioconjugation strategies. Having been thoroughly optimised, its kinetics characterised, and applied to a range of medicinally relevant proteins, this chemistry already serves as a robust platform for *N*-terminal protein bioconjugation. Nevertheless, it has the potential to become even more versatile and widely applicable with further development.

Chapter 7: Experimental procedures and characterisation

7.1 General considerations

7.1.1 Reagent and solvent suppliers

Unless otherwise stated, all chemical reagents were purchased and used directly from commercial sources. Anhydrous CH_2Cl_2 and THF were taken from an Innovative Technology Inc. Puresolv® solvent purification system and used without additional drying. Anhydrous MeOH was purchased from a commercial source and used without additional drying. Water used in reactions and protein manipulations was deionized. RNase A from bovine pancreas (powder, 50 units/mg protein), myoglobin from equine skeletal muscle (powder, 95–100%) and cytochrome C from bovine heart (powder, 95–100%) were purchased from Sigma Aldrich.

7.1.2 Spectroscopic and spectrometric instruments and analysis

^{1}H NMR and ^{13}C NMR spectra were recorded on a JEOL ECX400 or JEOL ECS400 spectrometer (operating at 400 MHz and 100 MHz), a Bruker Avance III 300 NMR spectrometer (operating at 300 MHz, 75 MHz, and 282 MHz), a Bruker Avance I 500 MHz spectrometer (operating at 500 MHz, 125 MHz, and 470 MHz), a Bruker Avance III HD 500 NMR spectrometer (operating at 500 MHz, 125 MHz, and 470 MHz), a Bruker Avance III HD 600 NMR spectrometer (operating at 600 MHz, 151 MHz, and 565 MHz), or a Bruker Avance Neo 700 NMR spectrometer (operating at 700 MHz, 176 MHz, and 659 MHz). All spectral data was recorded at 295 K unless otherwise stated. All chemical shifts are quoted on the δ scale in ppm using residual solvent as the internal standard (^{1}H NMR: $CDCl_3$ = 7.26; CD_3OD = 3.31; D_2O = 4.69; DMSO-d_6 = 2.50 and ^{13}C NMR: $CDCl_3$ = 77.16, CD_3OD = 49.00, DMSO-d_6 – 39.52). Coupling constants (*J*) are reported in Hertz (Hz) to the nearest 0.5 Hz. The multiplicity abbreviations used are: br s broad singlet, s singlet, d doublet, br d broad doublet, t triplet, br t broad triplet, q quartet, p pentet, dd doublet of doublets, ddd doublet of doublets of doublets, dddd doublet of doublet of doublet of doublets, dt doublet of triplets, ddt doublet of triplets, td triplet of doublets, m multiplet. Signal assignment was achieved by analysis of 2D NMR techniques (DEPT, COSY, HMBC, HSQC) where required. Infrared (IR) spectra were recorded on a PerkinElmer UATR 2 spectrometer as a thin film dispersed from either CH_2Cl_2 or $CDCl_3$. Mass spectra (high-resolution) were obtained by the University of York Mass Spectrometry Service, using Electrospray Ionisation (ESI) on a Bruker Daltonics, Micro-tof spectrometer. Nominal and exact *m/z* values are reported in Daltons (Da). Specific rotations

($[\alpha]_D$) were recorded on a Bellingham + Stanley ADP450 polarimeter as a solution in $CDCl_3$. Melting points were determined using Gallenkemp apparatus.

7.1.3 Chromatographic instruments and materials

Thin layer chromatography was carried out on Merck silica gel $60F_{254}$ pre-coated aluminium foil sheets and were visualised using UV light (254 nm) and stained with basic aqueous potassium permanganate. In most cases, flash column chromatography was carried out using slurry packed Fluka silica gel (SiO_2), 35 – 70 μM, 60 Å, under a light positive pressure, eluting with the specified solvent system. In selected cases, products were purified using automated column chromatography; this was done using a Teledyne ISCO NextGen 300+ automated flash column chromatography unit equipped with UV–Vis (200–800 nm) and evaporative light scattering (ELS) detectors. Crude materials were loaded onto pre-packed RediSep Rf Gold columns (SiO_2: 40–60 mesh) or RediSep Gold® C18 reversed phase columns (SiO_2: 400–632 mesh) either by direct liquid injection or dry loading from adsorbed Celite.

Liquid chromatography-mass spectrometry (LC-MS) was performed on a HCTultra ETD II ion trap spectrometer, coupled to an Ultimate300 HPLC using an Accucore™ C18 column (150 x 2.1 mm, 2.6 μm particle size) or an Accucore™ 150-C4 column (100 x 2.1 mm, 2.6 μm particle size). Water (solvent A) and acetonitrile (solvent B), both containing 0.1% formic acid, were used as the mobile phase at a flow rate of 0.3 mL min^{-1}. LC traces were measured via UV absorption at 220, 270 and 280 nm. Peptide samples were eluted with a linear gradient 5–90% (increasing solvent B) over 13 minutes. Protein samples were eluted with a linear gradient 10–90% (increasing solvent B) over 15 minutes. Samples were analysed using the Bruker Data Analysis 4.4 software. Spectra were charge deconvoluted using ESI Compass 1.3. Both the raw ion series MS data and the deconvoluted spectra are shown for each sample. Expected masses were calculated relative to reported values of 13681 Da (RNase A), 16951 Da (myoglobin), 12327 Da (Cytochrome C), and 12124 (JVZ-007). Observed masses of CjX183-D WT and R51K mutant were typically ca. 10 Da higher than theoretical masses (11226 Da and 11198 Da respectively), so expected masses for CjX183-D were calculated from the smallest species present in each individual sample.[63]

LC-MS method A: LC-MS samples were prepared by aliquoting 5 μL of crude reaction mixture into 45 μL of 1:1 acetonitrile/water with 1 % (v/v) formic acid. The samples were eluted on an Accucore™ C18 2.6 μm column (50 × 2.1 mm) (Thermoscientific) with a linear

gradient 5-90 % (increasing solvent B, details found below) over 13 mins with a mobile phase flow rate of 0.3 mL/min at 30 ºC.

Solvent A: HPLC-grade water with 0.1 % (v/v) formic acid.

Solvent B: HPLC-grade acetonitrile with 0.1 % (v/v) formic acid.

Conversion from starting material to modified product (bioconjugation yield, %) was calculated using the peak area of the UV 280 nm chromatograph signals corresponding to each species. Since our starting material, intermediate and product all contain a tyrosine residue, their absorbance coefficient at 280 nm was equivalent. Relative absorbance at 280 nm was therefore used to calculate conversion using **Equation 1**:

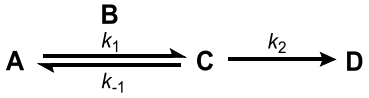
$$\frac{Area\ of\ peak\ (product)}{Area\ of\ peak\ (SM)+Area\ of\ peak\ (product)} \times 100 = Conversion\ (\%)$$

Equation 1

LC-MS method B: Kinetic experiments were conducted on a 100 µL scale with the samples incubated within the LC-MS instrumentation. Aliquots of the reaction mixture were injected every 15 mins and run as described in **LC-MS method A**. Each sample underwent a total of 12 injections over 3 h, and the reaction progress was monitored by calculating concentrations of each species using **Equation 2**, where AOP represents the peak area:

$$\frac{AOP\ (product)}{AOP\ (SM)+AOP\ (intermediate)+AOP\ (product)} \times 0.001 = Concentration\ (M) \qquad \textbf{\textit{Equation 2}}$$

The data from **Equation 2** was fitted in Copasi 4.34.251 to a model based on the following two-step reaction:



The aza-Michael addition (A+B → C) was defined as a reversible reaction (forward rate constant k_1, reverse rate constant k_{-1}), whilst the ring-expansion (C → D) was defined as an irreversible transformation (forward rate constant k_2). Rate constants were estimated using the evolutionary programming method built into the software, with 200 generations and a population size of 20. Parameters were restricted within the confines of: k_1 10^{-6} – 10^{7} $M^{-1}s^{-1}$; k_{-1} 10^{-8} – 10^{3} s^{-1}; k_2 10^{-8} – 10^{3} s^{-1}.

LC-MS method C: LC-MS samples were prepared by aliquoting 5 µL of crude reaction mixture into 45 µL of 1:1 acetonitrile/water with 1 % (v/v) formic acid. The samples were eluted on an Accucore™ 150-C4 2.6 µm column (100 × 2.1 mm) (Thermoscientific) with a linear gradient 10-90% (increasing solvent B) over 15 minutes with a mobile phase flow rate of 0.3 mL/min at 30 °C.

Solvent A: HPLC-grade water with 0.1% (v/v) formic acid.

Solvent B: HPLC-grade acetonitrile with 0.1% (v/v) formic acid.

Conversion from starting material to desired product (bioconjugation yield, %) was calculated by the MS peak intensity of each species after 'Protein Charge Deconvolution' using the ESI Compass 1.3 DataAnalysis V4.1 software (Bruker Daltonics). The deconvoluted *m/z* values were matched to their respective species and their relative intensities used to calculate conversion using Equation 3:

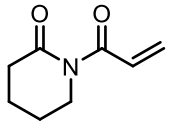
$$\frac{Intensity\ of\ peak\ (product)}{Int.of\ peak\ (SM)+Int.of\ peak\ (product)} \times 100 = Conversion\ (\%) \qquad \textbf{\textit{Equation 3}}$$

7.1.4 X-ray crystallography instruments and analysis

X-ray diffraction data were collected at 110 K on an Oxford Diffraction SuperNova diffractometer with Cu-K_α radiation (λ = 1.54184 Å) using an EOS CCD camera. The crystal was cooled with an Oxford Instruments Cryojet. Diffractometer control, data collection, initial unit cell determination, frame integration and unit-cell refinement were carried out with CrysAlisPro (Rigaku Ltd). Face-indexed absorption corrections were applied using spherical harmonics, implemented in SCALE3 ABSPACK scaling algorithm within CrysAliPro. OLEX2 (J. Appl. Cryst., 2009, 42, 339–341) was used for overall structure solution, refinement and preparation of computer graphics and publication data. Within OLEX2, the algorithms used for structure solution was ShelXT dual-space. (G. M. Sheldrick, Acta Cryst. 2015, A71, 3-8). Refinement by full-matric least-squares used the SHELXL algorithm (G. M. Sheldrick, Acta Cryst. 2015, A71, 3-8) within OLEX2.

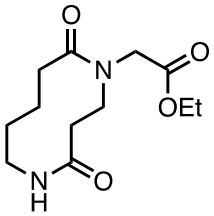
7.2 Experimental procedures and characterisation data for Chapter 2

Acryloyl-piperidin-2-one (2.01)



δ-Valerolactam (2.04 g, 20.6 mmol, 1.00 eq.) in dry THF (70.0 mL) was cooled to 0 ºC before the dropwise addition of MeMgBr (3.0 M in Et_2O, 7.40 mL, 22.6 mmol, 1.10 eq.) over 30 mins using a syringe pump. After the addition was complete, the reaction mixture was stirred at 0 ºC for 15 mins. At this point, acryloyl chloride (2.50 mL, 30.7 mmol, 1.50 eq.) was added as a single portion and the reaction was stirred for 1 h at 0 ºC. The reaction was then quenched with sat. aq. NH_4Cl (60.0 mL), the aqueous mixture extracted with Et_2O (100 mL) and the organic extracts washed with sat. $NaHCO_3$ (2 x 60 mL). The organic phase was dried over $MgSO_4$ and concentrated *in vacuo*. The crude material was purified by FCC (SiO_2, 4:1; hexane: ethyl acetate) to afford the *title compound* as a viscous, colourless oil (2.13 g, 68%); R_f = 0.31 (4:1; hexane: ethyl acetate); ν_{max}/cm^{-1} (thin film) 2950, 1677, 1542, 1384, 1289, 1207, 1154, 795, 568; δ_H (400 MHz, $CDCl_3$) 7.23 (dd, *J* = 16.9, 9.5 Hz, 1H, COC**H**=CH_2), 6.58 (dd, *J* = 16.9, 1.7 Hz, 1H, COCH=C**H**H'), 5.94 (dd, *J* = 9.5, 1.7 Hz, 1H, COCH=CH**H'**), 4.05 – 3.93 (m, 2H, C**H**$_2$NCO), 2.88 – 2.77 (m, 2H, C**H**$_2$CON), 2.18 – 2.04 (m, 4H, 2 x C**H**$_2$); δ_C (101 MHz, $CDCl_3$) 173.8 (**C**O), 169.6 (**C**O), 132.0 (**C**HCH$_2$), 127.9 (**C**H**C**H$_2$), 44.6 (**C**H$_2$NCO), 34.9 (**C**H$_2$CON), 22.7 (**C**H$_2$), 20.7 (**C**H$_2$); HRMS (ESI$^+$): calcd. for $C_8H_{11}NNaO_2$, 176.0682. Found [M+Na]$^+$, 176.0683. The data obtained match those previously reported.[50,179]

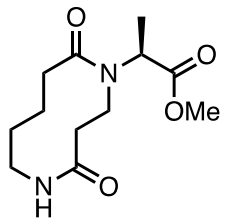
Ethyl 2-(4,10-dioxo-1,5-diazecan-1-yl) acetate (2.02)



To a flask charged with glycine ethyl ester hydrochloride (206 mg, 1.48 mmol, 1.10 eq.) in triethylamine (0.740 mL, 5.24 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (204 mg, 1.34 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred at room temperature for 4 h before the solvent was removed *in vacuo* and the crude material purified by FCC (SiO_2, 19:1; ethyl acetate: methanol → 9:1; ethyl acetate: methanol). The *title compound* was isolated as a white powder (306 mg, 90%). M.P. = 117 – 120 ºC; R_f = 0.51

(100% MeOH); ν_{max}/cm^{-1} (thin film) 3316, 2931, 1721, 1640, 1557, 1477, 1208, 1153, 1022, 861, 716, 566, 523; δ_H (400 MHz, $CDCl_3$) 7.63 – 7.56 (m, 1H, N**H**), 4.74 (d, *J* = 17.3 Hz, 1H, C**H**H'COO), 4.30 – 4.16 (m, 2H, C**H₂**CH_3), 4.17 – 4.02 (m, 1H, C**H**H'), 3.88 – 3.74 (m, 1H, C**H**H'NH), 3.32 (d, *J* = 17.3 Hz, 1H, CH**H'**COO), 3.24 (dt, *J* = 15.6, 3.7 Hz, 1H, C**H**), 2.91 (dd, *J* = 13.8, 4.9 Hz, 1H, CH**H'**NH), 2.69 – 2.55 (m, 1H, C**H**H'), 2.41 (td, *J* = 12.8, 3.2 Hz, 1H, C**H**H'), 2.27 – 2.17 (m, 1H, C**H**H'), 2.14 – 1.98 (m, 2H, 2 x C**H**H'), 1.68 – 1.57 (m, 2H, 2 x C**H**H'), 1.53 – 1.38 (m, 1H, C**H**H'), 1.29 (t, *J* = 7.3 Hz, 3H, CH_2C**H₃**); δ_C (101 MHz, $CDCl_3$) 174.5 (**C**O), 172.3 (**C**O), 170.8 (**C**O), 62.2 (**C**H_2CH_3), 51.5 (**C**H_2COO), 48.9 (**C**H_2), 39.0 (**C**H_2NH), 37.5 (**C**H_2), 27.9 (**C**H_2), 25.2 (**C**H_2), 24.1 (**C**H_2), 14.1 (CH_2**C**H_3); HRMS (ESI^+): calc. for $C_{12}H_{20}N_2NaO_4$, 279.1315. Found $[M+Na]^+$, 279.1313. The data obtained match those previously reported.[50]

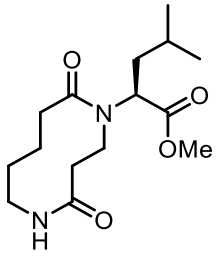
Ethyl 2-(4,10-dioxo-1,5-diazecan-1-yl) propanoate (2.03)



To a flask charged with *L*-alanine methyl ester hydrochloride (101 mg, 0.720 mmol, 1.10 eq.) in triethylamine (0.380 mL, 2.72 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (104 mg, 0.680 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred for 15 h before the solvent was removed *in vacuo* to afford the crude material as a white powder. This was purified by FCC using a gradient eluent system (SiO_2, ethyl acetate → 9:1; ethyl acetate: methanol) to afford the *title compound* as a pale orange oil (70.9 mg, 41%). In $CDCl_3$ the compound exists as a 25:11 (A:B) mixture of rotamers. R_f = 0.22 (9:1; ethyl acetate: methanol); $[\alpha]_D^{20}$ = –11.06 (9.7 mg/mL in CH_2Cl_2); ν_{max}/cm^{-1} (thin film) 3322, 2934, 1717, 1667, 1633, 1552, 1431, 1348, 1214, 1256, 1211, 1156, 1102, 1060, 763, 734, 697. ^{1}H NMR signals for the major rotamer (A) can be found at: δ_H (400 MHz, $CDCl_3$) 7.39 – 7.32 (m, 1H, N**H**), 4.01 – 3.82 (m, 1H, NC**H**H'CH_2), 3.74 (s, 3H, $CH_3CHCOOC$**H₃**), 3.47 (q, *J* = 6.9 Hz, 1H, CH_3C**H**$COOCH_3$), 3.22 (dt, *J* = 15.5, 3.6 Hz, 1H, NCH**H'**CH_2), 2.89 – 2.80 (m, 2H, NHC**H₂**), 2.73 – 1.90 (m, 6H, NHCHH'C**H₂** & 2 x C**H₂** [overlapping with minor rotamer signals]), 1.51– 1.48 (m, 2H, C**H₂** [overlapping with minor rotamer signals]), 1.46 (d, *J* = 6.9 Hz, 3H, C**H₃**$CHCOOCH_3$). Diagnostic ^{1}H NMR signals for the minor rotamer (B) can be found at: δ 7.49 – 7.41 (m, 1H, N**H**), 4.88 (q, *J* = 7.9 Hz, 1H, CH_3C**H**$COOCH_3$), 3.71 (s, 3H,

CHCOOC**H**$_3$), 3.42 – 3.39 (m, 2H, NC**H**$_2$CH$_2$), 2.73 – 1.90 (m, 6H, NHCHH'C**H**$_2$ & 2 x C**H**$_2$ [overlapping with major rotamer signals]), 1.51– 1.48 (m, 2H, C**H**$_2$ [overlapping with major rotamer signals]), 1.40 (d, *J* = 7.9 Hz, 3H, C**H**$_3$CHCOOCH$_3$) – not all ^{1}H NMR signals for the minor rotamer (B) could be found. ^{13}C NMR signals for the major rotamer can be found at: δ_C (101 MHz, $CDCl_3$) 173.4 (**C**O), 173.2 (**C**O), 170.7 (**C**O), 58.7 (**C**HCOO), 52.8 (**C**H$_3$OCO), 47.4 (N**C**H$_2$), 39.0 (NH**C**H$_2$), 36.9 (**C**H$_2$), 27.9 (**C**H$_2$), 24.8 (**C**H$_2$), 23.8 (**C**H$_2$), 14.3 (**C**H$_3$CH). Diagnostic ^{13}C NMR signals for the minor rotamer can be found at: 176.2 (**C**O), 174.8 (**C**O), 170.9 (**C**O), 53.2 (**C**H$_3$OCO), 40.5 (**C**H$_2$), 38.9 (**C**H$_2$), 38.1 (**C**H$_2$), 28.5 (**C**H$_2$), 25.0 (**C**H$_2$), 23.9 (**C**H$_2$), 14.8 (**C**H$_3$CH); HRMS (ESI$^+$): calc. for $C_{12}H_{20}N_2NaO_4$, 279.1315. Found $[M+Na]^+$, 279.1316. The data obtained match those previously reported.[50]

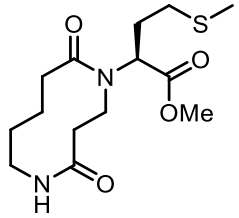
Methyl 2-(4,10-dioxo-1,5-diazecan-1-yl)-4-methylpentanoate (2.09)



To a flask charged with *L*-leucine methyl ester hydrochloride (131 mg, 0.720 mmol, 1.10 eq.) in triethylamine (0.380 mL, 2.76 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (106 mg, 0.690 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred for 17 h before being concentrated *in vacuo* to afford the crude material as a white solid. The crude material was purified by FCC using a gradient eluent system (2:1; hexane: ethyl acetate → 1:1; hexane: ethyl acetate) to afford the *title compound* as a white solid (72 mg, 35%). In $CDCl_3$ the compound exists as a ~20:7:4 (A: B: C) mixture of rotamers. M.P. = 132 – 136 ºC; $R_{f=}$ 0.12 (100% ethyl acetate); $[\alpha]_D^{20}$ = –15.79 (10.0 mg/mL in CH_2Cl_2); ν_{max}/cm^{-1} (thin film) 3328, 2954, 2243, 1715, 1665, 1633, 1554, 1445, 1428, 1317, 1257, 1156, 912, 727, 643. ^{1}H NMR signals for the major rotamer (A) can be found at: δ_H (400 MHz, $CDCl_3$) 7.54 – 7.47 (m, 1H, N**H**), 4.11 – 3.97 (m, 1H, C**H**H'), 3.92 – 3.75 (m, 4H, C**H**H'NH & COOC**H**$_3$ [overlapping with minor rotamer signal]), 3.45 (br dd, *J* = 8.9, 5.0 Hz, 1H, C**H**COO), 3.21 (dt, *J* = 15.5, 3.6 Hz, 1H, C**H**H'), 2.94 – 2.83 (m, 1H, CH**H'**NH), 2.61 – 2.36 (m, 2H, 2 x C**H**H' [overlapping with unidentifiable minor rotamer signals]), 2.22 (dd, *J* = 12.4, 3.9 Hz, 1H, C**H**H'), 2.12 – 1.86 (m, 3H, C**H**$_2$CH(CH$_3$)$_2$ & C**H**H' [overlapping with unidentifiable minor rotamer signals]), 1.79 – 1.34 (m, 4H, C**H**$_2$, CH$_2$C**H**(CH$_3$)$_2$ & C**H**H' [overlapping with unidentifiable minor rotamer

signals]), 0.97 – 0.87 (m, 6H, 2 x C**H**$_3$ [overlapping with minor rotamer signal]). Diagnostic ^{1}H NMR signals for the minor rotamers can be found at: δ 7.42 – 7.34 (m, 1H, N**H** (B)), 5.15 (dd, J = 10.6, 5.5 Hz, 1H, C**H**COO (B)), 3.92 – 3.75 (m, 2H, C**H**H'NH (B) [overlapping with major rotamer signal]), 3.74 (s, 3H, COOC**H**$_3$ (B)), 3.68 (s, 2H, COOC**H**$_3$ (C)), 3.32 (dt, J = 16.2, 3.6 Hz, 1H, CH**H'**NH (B)), 2.75 – 2.65 (m, 1H, C**H**H' (B)), 2.18 – 2.13 (m, 1H, C**H**H' (B)), 0.97 – 0.87 (m, 6H, 2 x C**H**$_3$ (B & C) [overlapping with major rotamer signal]) - not all ^{1}H NMR signals for the minor rotamers could be found or identified likely due to overlap with major rotamer signals. ^{13}C NMR signals for the major rotamer (A) can be found at: δ_C (101 MHz, $CDCl_3$) 173.8 (**C**O), 173.5 (**C**OOCH$_3$), 170.9 (**C**O), 61.9 (**C**HCOO), 53.0 (COO**C**H$_3$), 48.8 (**C**H$_2$), 39.1 (**C**H$_2$NH), 38.4 (**C**H$_2$CH(CH$_3$)$_2$), 37.0 (**C**H$_2$), 28.2 (**C**H$_2$), 25.0 (**C**H(CH$_3$)$_2$), 24.8 (**C**H$_2$), 23.7 (**C**H$_2$), 23.4 (**C**H$_3$), 22.3 (**C**H$_3$). Diagnostic ^{13}C NMR signals for the minor rotamers can be found at δ 176.0 (**C**O (B)), 175.4 (**C**O (B)), 171.2 (**C**O (B)), 55.1 (**C**HCOO (B)), 52.9 (COO**C**H$_3$ (B)), 52.0 (COO**C**H$_3$ (C)), 40.4 (**C**HH'NH (B)), 38.9 (**C**H$_2$ (B)), 37.8 (**C**H$_2$ (B)), 28.7 (**C**H$_2$ (B)), 25.2 (**C**H$_2$ (B)), 25.2 (C**H**(CH$_3$)$_2$ (B)), 24.0 (**C**H$_2$ (B)), 23.1 (**C**H$_3$ (B)), 23.0 (**C**H$_3$ (C)), 21.9 (**C**H$_3$ (C)), 21.4 (**C**H$_3$ (B)) – not all ^{13}C NMR signals for the minor rotamers could be found; HRMS (ESI$^+$): Calcd. for $C_{15}H_{26}N_2NaO_4$, 321.1785; Found $[M+Na]^+$, 321.1789.

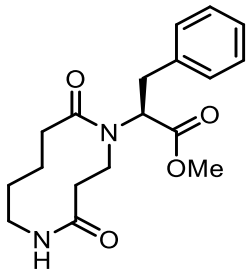
Methyl 2-(4,10-dioxo-1,5-diazecan-1-yl)-4-(methylthio)butanoate (2.10)



To a flask charged with *L*-methionine methyl ester hydrochloride (144 mg, 0.72 mmol, 1.10 eq.) in triethylamine (0.360 mL, 2.60 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (100 mg, 0.650 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred at room temperature for 16 h before being concentrated *in vacuo* to afford the crude material as an orange oil. The crude material was purified by FCC using a gradient eluent system (2:1; hexane: ethyl acetate → 1:1; hexane: ethyl acetate → ethyl acetate → 9:1; ethyl acetate: methanol) to afford the *title compound* as a yellow oil (105 mg, 51%). In $CDCl_3$ the compound exists in a ~10:2:0.5 (A:B:C) mixture of rotamers. R_f = 0.64 (MeOH); $[\alpha]_D^{20}$ = –37.03 (10.0 mg/mL in CH_2Cl_2); ν_{max}/cm^{-1} (thin film) 3329, 2921, 2242, 1720, 1634, 1429, 1251, 908, 724, 645. ^{1}H NMR signals for the major rotamer (A) can be found at: δ_H (400 MHz, $CDCl_3$) 7.48 –

7.39 (m, 1H, N**H**), 3.99 – 3.88 (m, 1H, CONC**H**H'CHH'), 3.83 – 3.65 (m, 5H, C**H**COOCH$_3$, CHCOOC**H$_3$** & NHC**H**H'), 3.37 (br dt, *J* = 15.5, 3.7 Hz, 1H, CONCH**H'**CHH'), 2.87 – 2.78 (m, 1H, NHCH**H'**), 2.71 – 2.16 (m, 5H, CONCHH'C**H**H' & 4 x C**H**H'), 2.20 – 2.08 (m, 1H, CONCHH'CH**H'**), 2.06 – 1.91 (m, 6H, SC**H$_3$**, C**H$_2$** & C**H**H'), 1.61 – 1.30 (m, 3H, C**H$_2$** & CH**H'**). Diagnostic ^{1}H NMR signals for the minor rotamers can be found at: δ 7.24 – 7.17 (m, 1H, N**H** (B)), 6.56 (br t, *J* = 5.9 Hz, 1H, N**H** (C)), 3.64 (s, 3H, COOC**H$_3$** (C)), 3.56 (s, 3H, COOC**H$_3$** (B)) – not all ^{1}H NMR signals for the minor rotamers could be found, likely due to overlap with major rotamer signals. ^{13}C NMR signals for the major rotamer can be found at: δ_C (101 MHz, $CDCl_3$) 173.8 (**C**O), 173.2 (**C**O), 170.7 (**C**O), 61.0 (**C**HCOO), 52.9 (COO**C**H$_3$), 48.5 (CON**C**HH'CHH'), 38.8 (NH**C**HH'), 36.9 (CONCHH'**C**HH'), 30.8 (**C**H$_2$), 28.0 (**C**H$_2$), 27.6 (**CH$_2$**), 24.6 (**C**H$_2$), 23.5 (**C**H$_2$), 15.1 (S**C**H$_3$). Diagnostic ^{13}C NMR signals for the minor rotamers can be found at: 175.2 (**C**O), 175.0 (**C**O), 174.8 (**C**O), 172.4 (**C**O), 172.0 (**C**O), 170.9 (**C**O), 59.7, 55.9, 51.9 (COO**C**H$_3$ (C)), 51.5 (COO**C**H$_3$ (B)), 46.9, 44.0, 40.5, 38.6 , 37.6, 36.0, 35.0, 33.8, 33.4, 32.3, 30.4, 28.9, 25.0, 23.8, 22.1, 15.5, 15.2 – not all ^{13}C NMR signals could be found or assigned from rotamers B and C due to overlap with major rotamer signals; HRMS (ESI$^+$): Calcd. for $C_{14}H_{24}N_2NaO_4S$, 339.1349; Found $[M+Na]^+$, 339.1350.

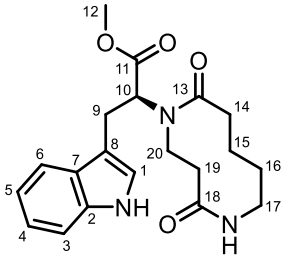
Methyl 2-(4,10-dioxo-1,5-diazecan-1-yl)-3-phenylpropanoate (2.11)



To a flask charged with *L*-phenylalanine methyl ester hydrochloride (157 mg, 0.720 mmol, 1.10 eq.) in triethylamine (0.360 mL, 2.60 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (102 mg, 0.65 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred for 4 h before being concentrated *in vacuo* and the crude material being purified by FCC (ethyl acetate) to afford the *title compound* as a white solid (114 mg, 53%). In $CDCl_3$ the compound exists in a ~20:3 (A: B) mixture of rotamers. R_f = 0.16 (ethyl acetate); $[\alpha]_D^{20}$ = – 119.34 (10.0 mg/mL in CH_2Cl_2); ν_{max}/cm^{-1} (thin film) 3328, 2932, 1717, 1668, 1553, 1454, 1319, 1257, 1043, 1019, 754, 702, 634. ^{1}H NMR signals for the major rotamer (A) can be found at: δ_H (400 MHz, $CDCl_3$) 7.36 – 7.01 (m, 5H, 5 x ArC**H**), 3.84 (s, 3H, CHCOOC**H$_3$**), 3.83 – 3.72 (m, 1H, CONC**H**H'), 3.62 – 3.35 (m, 3H, C**H**COOCH$_3$, CHC**H**H'ArC & C**H**H'), 3.30

(dd, *J* = 13.6, 4.3 Hz, 1H, CHCH**H'**ArC), 2.93 – 2.77 (m, 1H, CONCH**H'**), 2.52 – 2.37 (m, 1H, C**H**H'), 2.35 – 2.20 (m, 2H, 2 x C**H**H'), 2.13 – 1.93 (m, 3H, 3 x C**H**H'), 1.65 – 1.37 (m, 3H, C**H₂** & C**H**H'). Diagnostic ^{1}H NMR signals for the minor rotamers can be found at: δ 7.97 – 7.89 (m, 1H, N**H** (B)), 3.68 (s, 3H, COOC**H₃** (B)), 3.15 – 2.92 (m, 2H, C**H₂** (B)) – not all ^{1}H NMR signals for the minor rotamers could be found likely due to overlap with major rotamer signals. ^{13}C NMR signals for the major rotamer (A) can be found at: δ_C (101 MHz, $CDCl_3$) 173.9 (**C**O), 172.7 (**C**O), 170.7 (**C**O), 138.0 (Ar**C**), 129.3 (2 x Ar**C**H), 128.8 (2 x Ar**C**H), 126.9 (Ar**C**H), 65.3 (**C**HCOOCH$_3$), 53.1 (CHCOO**C**H$_3$), 48.6 (**C**H$_2$), 39.0 (**C**H$_2$), 36.7 (**C**H$_2$), 34.6 (CH**C**H$_2$ArC), 28.1 (**C**H$_2$), 24.9 (**C**H$_2$), 23.8 (**C**H$_2$). Diagnostic ^{13}C NMR signals for the minor rotamer were not found; HRMS (ESI$^+$): Calcd. for $C_{18}H_{24}N_2NaO_4$, 355.1628; Found $[M+Na]^+$, 355.1635.

Methyl 2-(4,10-dioxo-1,5-diazecan-1-yl)-3-(1*H*-indol-3-yl)propanoate (2.12)



To a flask charged with tryptophan methyl ester hydrochloride (185 mg, 0.73 mmol, 1.10 eq.) in triethylamine (0.380 mL, 2.76 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (106 mg, 0.690 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred for 17 h before the solvent was removed *in vacuo* to afford the crude material as a white solid. The crude material was purified by FCC (19:1; ethyl acetate: methanol) but it was not possible to isolate pure fractions of the *title compound* due to coelution with a reaction by-product. These fractions were further purified through recrystallisation in $CDCl_3$ (28 mg, 11%). R_f = 0.38 (9:1; ethyl acetate: methanol); $[\alpha]_D^{20}$ –117.96 (10.0 mg/mL in $CDCl_3$); ν_{max}/cm^{-1} (thin film) 3309, 3181, 2944, 1719, 1627, 1434, 1250, 1029, 744, 724, 645; δ_H (400 MHz, $CDCl_3$) 8.10 (s, 1H, ArN**H**), 7.51 (d, *J* = 8.1 Hz, 1H, ArC**H**), 7.37 (d, *J* = 8.1 Hz, 1H, ArC**H**), 7.26 – 7.16 (m, 1H, ArC**H**), 7.13 – 7.08 (m, 1H, ArC**H**), 6.99 (br d, *J* = 2.3 Hz, 1H, **H**-1), 3.90 (s, 3H, **H**-12), 3.87 – 3.60 (m, 4H, **H**-10 & 3 x C**H**H'), 3.56 – 3.42 (m, 2H, **H**-9 & C**H**H'), 2.91 (d, *J* = 13.9 Hz, 1H, C**H**H'), 2.51 – 2.35 (m, 2H, **H**-9' & C**H**H'), 2.27 (td, *J* = 12.7, 3.1 Hz, 1H, C**H**H'), 2.15 – 2.06 (m, 2H, 2 x C**H**H'), 1.99 – 1.90 (m, 1H, C**H**H'), 1.67 – 1.45 (m, 2H, 2 x C**H**H'); δ_C (101 MHz, $CDCl_3$) 173.8 (**C**O), 173.0 (**C**O), 170.9 (**C**O), 136.3 (Ar**C**), 127.3 (Ar**C**), 122.9 (**C**-1), 122.5 (Ar**C**H), 119.9 (Ar**C**H), 118.3 (Ar**C**H), 112.2 (Ar**C**), 111.6 (Ar**C**H), 64.4 (**C**-10), 53.1

(**C**-12), 48.6 (**C**-9), 39.1 (**C**H$_2$), 36.8 (**C**H$_2$), 28.3 (**C**H$_2$), 25.0 (**C**H$_2$), 24.3 (**C**H$_2$), 23.9 (**C**H$_2$); HRMS (ESI$^+$): Calcd. for $C_{20}H_{26}N_3NaO_4$, 394.1737; Found [M+Na]$^+$, 394.1746. For X-ray crystallographic data, see CCDC 2391160.

X-ray crystallographic data for 2.12:[c]

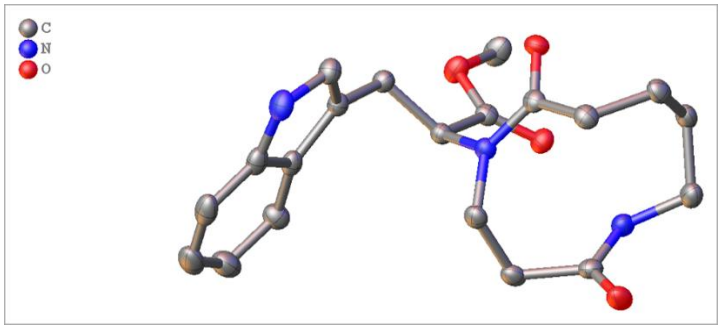


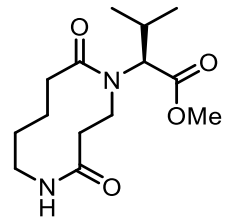
Table 1 Crystal data and structure refinement for wpu23007.

Identification code	wpu23007
Empirical formula	$C_{20}H_{25}N_3O_4$
Formula weight	371.43
Temperature/K	110.00(10)
Crystal system	orthorhombic
Space group	$P2_12_12_1$
a/Å	9.14295(10)
b/Å	10.25505(13)
c/Å	20.8149(3)
α/°	90
β/°	90
γ/°	90
Volume/Å^3	1951.63(4)
Z	4
ρ_{calc}g/cm^3	1.264
μ/mm^{-1}	0.727
F(000)	792.0
Crystal size/mm^3	0.218 × 0.155 × 0.043

[c]X-ray crystallography was carried out by Dr Adrian Whitwood.

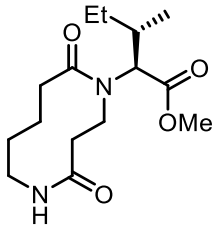
Radiation	Cu Kα (λ = 1.54184)
2Θ range for data collection/°	8.496 to 134.128
Index ranges	$-8 \leq h \leq 10$, $-11 \leq k \leq 12$, $-24 \leq l \leq 24$
Reflections collected	9377
Independent reflections	3479 [R_{int} = 0.0348, R_{sigma} = 0.0369]
Data/restraints/parameters	3479/0/345
Goodness-of-fit on F^2	1.029
Final R indexes [I>=2σ (I)]	R_1 = 0.0265, wR_2 = 0.0640
Final R indexes [all data]	R_1 = 0.0277, wR_2 = 0.0645
Largest diff. peak/hole / e $Å^{-3}$	0.17/-0.14
Flack parameter	0.11(9)

Methyl 2-(4,10-dioxo-1,5-diazecan-1-yl)-3-methylbutanoate (2.14)



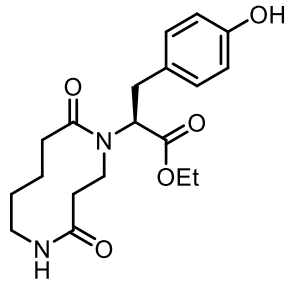
To a flask charged with *L*-valine methyl ester hydrochloride (128 mg, 0.720 mmol, 1.10 eq.) in triethylamine (0.370 mL, 2.68 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (102 mg, 0.670 mmol, 1.00 eq.) as a solution in DCM (1.50 mL). 2–3 Drops of MeOH were added to the flask to aid the dissolution of the starting material. The solution was stirred for 23 h before the solvent was removed *in vacuo* to afford the crude material as a white powder. The crude material was then purified by FCC (SiO_2, 1:1; hexane: ethyl acetate → 1:2; hexane: ethyl acetate) to afford the *title compound* as a fine white powder (12.5 mg, 7%). R_f = 0.31 (100% ethyl acetate); $[\alpha]_D^{20}$ = –5.36 (10.0 mg/mL in CH_2Cl_2); ν_{max}/cm^{-1} (thin film) 2956, 1687, 1733, 1571, 1447, 1291, 1262, 1197, 1174, 1153, 1024, 912, 728, 646, 576; δ_H (400 MHz, $CDCl_3$) 3.81– 3.63 (m, 5H, COOC**H₃** & C**H₂**), 3.16 – 3.05 (m, 2H, C**H₂**), 3.01 (d, *J* = 6.1 Hz, 1H, C**H**COO), 2.95 – 2.87 (m, 1H, C**H**H'), 2.80 – 2.68 (m, 1H, C**H**H'), 2.59 – 2.49 (m, 2H, C**H₂**), 1.95 – 1.87 (m, 1H, C**H**$(CH_3)_2$), 1.87 – 1.75 (m, 4H, 2 x C**H₂**), 0.92 (t, *J* = 6.6 Hz, 6H, CH(C**H₃)₂**); δ_C (101 MHz, $CDCl_3$) 175.9 (**C**O), 175.5 (**C**O), 173.5 (**C**O), 67.8 (**C**HCOO), 51.6 (COO**C**H_3), 44.3 (**C**H_2), 44.0 (**C**H_2), 40.3 (**C**H_2), 35.0 (**C**H_2), 31.6 (**C**H$(CH_3)_2$), 22.5 (**C**H_2), 20.4 (**C**H_2), 19.2 (**C**H_3), 18.9 (**C**H_3); HRMS (ESI^+): calc. for $C_{14}H_{24}N_2NaO_4$, 307.1628. Found $[M+Na]^+$, 307.1635.

Ethyl 2-(4,10-dioxo-1,5-diazecan-1-yl)-3-methylpentanoate (2.15)



To a flask charged with methyl *L*-isoleucinate hydrochloride (130 mg, 0.720 mmol, 1.10 eq.) in triethylamine (0.360 mL, 2.60 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (100 mg, 0.650 mmol, 1.00 eq.) as a solution in DCM (1.50 mL). The solution was stirred overnight for 23 h before being concentrated *in vacuo* to afford the crude material as a white paste. The crude material was purified by FCC using a gradient eluent system (2:1; hexane: ethyl acetate → 1:1; hexane: ethyl acetate → 2:1; ethyl acetate: hexane → 100% ethyl acetate) to give the *title compound* as a colourless oil (37.5 mg, 19%). R_f = 0.20 (2:1; ethyl acetate: hexane); $[\alpha]_D^{20}$ = 0.70 (10.0 mg/mL in CH_2Cl_2); ν_{max}/cm^{-1} (thin film) 2959, 2876, 1734, 1711, 1667, 1438, 1197, 1174, 730, 659; δ_H (400 MHz, $CDCl_3$) 3.73 – 3.68 (m, 5H, C**H₂** & C**H₃**OCO), 3.16 – 3.00 (m, 3H, C**H**COO & C**H₂**), 2.98 – 2.85 (m, 1H, C**H**H'), 2.78 – 2.67 (m, 1H, C**H**H'), 2.60 – 2.50 (m, 2H, C**H₂**), 1.89 – 1.76 (m, 4H, 2 x C**H₂**), 1.73 – 1.59 (m, 1H, C**H**CH_3), 1.56 – 1.44 (m, 1H, C**H**H'CH_3), 1.27 – 1.08 (m, 1H, CH**H'**CH_3), 0.97 – 0.84 (m, 6H, 2 x C**H₃**); δ_C (101 MHz, $CDCl_3$) 175.9 (**C**O), 175.6 (**C**O), 173.5 (**C**O), 66.6 (**C**HCOO), 51.5 (COO**C**H_3), 44.3 (**C**H_2), 44.0 (**C**H_2), 40.5 (**C**H_2), 38.3 (**C**HCH_3), 35.0 (**C**H_2), 25.8 (**C**H_2CH_3), 22.5 (**C**H_2), 20.4 (**C**H_2), 15.6 (**C**H_3), 11.6 (**C**H_3); HRMS (ESI^+): calc. for $C_{15}H_{26}N_2NaO_4$, 321.1785. Found $[M+Na]^+$, 321.1784.

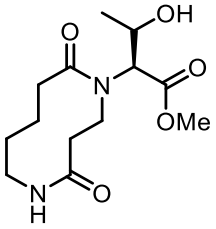
Ethyl 2-(4,10-dioxo-1,5-diazecan-1-yl)-3-(4-hydroxyphenyl)propanoate (2.16)



To a flask charged with *L*-tyrosine ethyl ester hydrochloride (357 mg, 1.45 mmol, 1.10 eq.) in triethylamine (0.750 mL, 5.40 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (206 mg, 1.35 mmol, 1.00 eq.) as a solution in MeOH (3.00 mL). The solution was stirred for 15 h overnight before the crude material was purified by FCC (20:1; ethyl acetate: methanol) to

afford the *title compound* as a pale orange oil (60.8 mg, 26%). R_f = 0.34 (15:1; ethyl acetate: methanol); ν_{max}/cm^{-1} (thin film) 3301, 2936, 2248, 1710, 1632, 1515, 1431, 1322, 1250, 1028, 907, 725, 645, 540; δ_H (400 MHz, $CDCl_3$) 7.65 – 7.58 (m, 1H, N**H**), 6.91 (d, *J* = 8.2 Hz, 2H, 2 x Ar**H**), 6.76 (d, *J* = 8.2 Hz, 2H, 2 x Ar**H**), 4.38 – 4.22 (m, 2H, COOC**H₂**CH_3), 3.88 – 3.75 (m, 1H, C**H**H'NH), 3.56 – 3.50 (m, 1H, C**H**COOEt), 3.50 – 3.44 (m, 1H, C**H**H'), 3.42 – 3.29 (m, 1H, C**H**H'ArC), 3.21 (dd, *J* = 13.8, 4.1 Hz, 1H, CH**H'**ArC), 3.03 – 2.78 (m, 1H, CH**H'**NH), 2.51 – 2.27 (m, 3H, 3 x C**H**H'), 2.15 – 1.98 (m, 3H, 3 x C**H**H'), 1.67 – 1.45 (m, 3H, 3 x C**H**H'), 1.35 (t, *J* = 7.1 Hz, 3H, $COOCH_2C$**H₃**); δ_C (101 MHz, $CDCl_3$) 174.0 (**C**O), 172.4 (**C**O), 171.6 (**C**O), 156.0 (Ar**C**OH), 130.2 (2 x Ar**C**H**)**, 128.6 (Ar**C**), 115.7 (2 x Ar**C**H**)**, 65.7 (**C**HCOOEt), 62.5 ($\mathbf{C}H_2CH_3$), 48.6 ($\mathbf{C}H_2$), 39.3 (**C**HH'NH), 36.4 ($\mathbf{C}H_2$), 33.6 (**C**HH'ArC), 28.1 ($\mathbf{C}H_2$), 24.7 ($\mathbf{C}H_2$), 23.8 ($\mathbf{C}H_2$), 14.1 ($CH_2\mathbf{C}H_3$); HRMS (ESI^+): Calcd. for $C_{19}H_{26}N_2NaO_5$, 385.1734; Found $[M+Na]^+$, 385.1745.

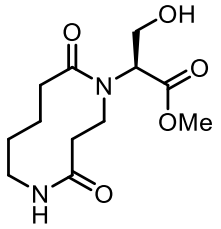
Methyl 2-(4,10-dioxo-1,5-diazecan-1-yl)-2-hydroxybutanoate (2.17)



To a flask charged with *L*-threonine methyl ester hydrochloride (128 mg, 0.75 mmol, 1.10 eq.) in triethylamine (0.380 mL, 2.72 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (104 mg, 0.680 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred for 20 h before the solvent was removed *in vacuo* to afford the crude material as an orange oil. The crude material was purified by FCC using a gradient eluent system (2:1; ethyl acetate: hexane → 100% ethyl acetate → 9:1; ethyl acetate: methanol) to afford the *title compound* as a pale orange oil (36 mg, 19%). In $CDCl_3$ the compound exists as a ~9:1:1 (A: B: C) mixture of rotamers. R_f = 0.15 (9:1; ethyl acetate: methanol); $[\alpha]_D^{20}$ = –15.79 (10.0 mg/mL in CH_2Cl_2); ν_{max}/cm^{-1} (thin film) 3294, 2939, 1741, 1625, 1555, 1438, 1212, 1163, 728, 506. ^1H NMR signals for the major rotamer (A) can be found at: δ_H (400 MHz, $CDCl_3$) 7.49 – 7.42 (m, 1H, N**H**), 6.33 (s, 1H, O**H**), 4.53 – 4.43 (m, 1H, CH_3C**H**OH), 4.17 – 4.05 (m, 1H, CONC**H**H'CHH'), 3.93 – 3.78 (m, 4H, COOC**H₃** & NHC**H**H'), 3.80 – 3.62 (m, 1H, C**H**COOMe [overlapping with minor rotamer signals]), 3.34 – 3.18 (m, 1H, CONCH**H'**CHH' [overlapping with minor rotamer signals]), 3.01 – 2.82 (m, 1H, NHCH**H'**), 2.74 – 2.57 (m, 1H, C**H**H'), 2.45 – 2.37 (m, 1H, CONCHH'C**H**H'), 2.28 – 1.97 (m, 3H, CONCHH'CH**H'** & 2 x

C**H**H'), 1.68 – 1.41 (m, 3H, C**H$_2$** & C**H**H'), 1.25 (d, *J* = 6.4 Hz, 3H, C**H$_3$**CHOH). Diagnostic ^{1}H NMR signals for the minor rotamers can be found at: δ 4.61 – 4.54 (m, 1H), 3.80 – 3.62 (m, 1H, COOC**H$_3$** (B) & COOC**H$_3$** (C) [overlapping with major rotamer signal]), 3.34 – 3.18 (m, 4H [overlapping with major rotamer signal]), 2.39 – 2.28 (m, 2H), 1.86 – 1.69 (m, 6H), 1.37 (d, *J* = 6.1 Hz, 3H, C**H$_3$**CHOH (B)), 1.20 (d, *J* = 6.1 Hz, 3H, C**H$_3$**CHOH (C)) - not all ^{1}H NMR signals for the minor rotamers were found due to overlap with major rotamer and impurity signals. ^{13}C NMR signals for the major rotamer (A) can be found at: δ_C (101 MHz, $CDCl_3$) 176.5 (**C**O), 172.5 (**C**O), 171.5 (**C**O), 69.0 (**C**HCOOMe), 66.7 (**C**HOH), 53.2 (COO**C**H$_3$), 49.5 (CON**C**HH'CHH'), 39.1 (NH**C**HH'), 36.8 (CONCHH'**C**HH'), 28.5 (**C**H$_2$), 24.9 (**C**H$_2$), 23.8 (**C**H$_2$), 21.2 (**C**H$_3$CHOH). Diagnostic ^{13}C NMR signals for the minor rotamers can be found at δ 170.6, 68.1, 67.8, 65.2, 52.6 (COO**C**H$_3$ (B)), 52.2, 51.7 (COOCH$_3$ (C)), 42.4, 44.6, 38.9, 37.3, 36.5, 33.5, 31.6, 29.0, 28.6, 25.5, 24.4, 22.4, 22.2, 20.9, 19.9 (**C**H$_3$CHOH) – not all ^{13}C NMR signals for the minor rotamers could be found or identified due to overlap with major rotamer; HRMS (ESI$^+$): Calcd. for $C_{13}H_{22}KN_2O_5$, 325.1160; Found $[M+K]^+$, 325.1158.

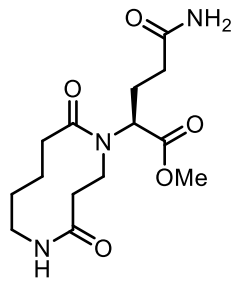
Methyl 2-(4,10-dioxo-1,5-diazecan-1-yl)-3-hydroxypropanoate (2.18)



To a flask charged with *L*-serine methyl ester hydrochloride (113 mg, 0.726 mmol, 1.09 eq.) in triethylamine (0.370 mL, 2.64 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (101 mg, 0.660 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred for 17 h before the solvent was removed *in vacuo* to afford the crude material as a white powder. The crude material was purified by FCC using a gradient eluent system (2:1; hexane: ethyl acetate → 100% ethyl acetate → 20:1; ethyl acetate: methanol → 9:1; ethyl acetate: methanol) to afford the *title compound* as a colourless oil (83.4 mg, 47%). In $CDCl_3$ the compound exists as a 14:7:1 (A: B: C) mixture of rotamers. R_f = 0.33 (100% ethyl acetate); $[\alpha]_D^{20}$ = –6.63 (10.0 mg/mL in CH_2Cl_2); ν_{max}/cm^{-1} (thin film) 3315, 2935, 1731, 1632, 1564, 1433, 1209, 1036, 918, 726, 645. ^{1}H NMR signals for the major rotamer (A) can be found at: δ_H (400 MHz, $CDCl_3$) 7.31 – 7.22 (m, 1H, N**H**), 6.19 (s, 1H, O**H**), 4.15 (dd, *J* = 11.6, 5.1 Hz, 1H, C**H**H'OH), 4.09 – 3.94 (m, 2H, CH**H'**OH & C**H**H'), 3.87 – 3.59 (m, 5H, COOC**H$_3$**, C**H**COO & C**H**H'NH [overlapping with minor rotamer signals]), 3.44 – 3.34 (m, 1H, C**H**H'), 2.91 – 2.82 (m, 1H,

CH**H'**NH [overlapping with unidentified minor rotamer signals]), 2.67 – 2.48 (m, 1H, C**H**H' [overlapping with minor rotamer signals]), 2.47 – 2.35 (m, 1H, C**H**H'), 2.23 – 2.16 (m, 1H, C**H**H' [overlapping with minor rotamer signals]), 2.08 – 1.96 (m, 2H, 2 x C**H**H' [overlapping with minor rotamer signals]), 1.66 – 1.50 (m, 2H, 2 x C**H**H' [overlapping with minor rotamer signals]), 1.47 – 1.35 (m, 1H, C**H**H'). Diagnostic ^{1}H NMR signals for the minor rotamers can be found at: δ 8.30 – 8.24 (m, 1H, N**H** (C)), 7.52 – 7.45 (m, 1H, N**H** (B)), 4.37 – 4.31 (m, 2H, C**H**H'OH (B & C)), 4.08 – 4.02 (m, 1H, CH**H'**OH (B)), 3.87 – 3.59 (m, 6H, C**H$_3$**OCO (B & C) [overlapping with major rotamer signals]), 3.55 – 3.47 (m, 2H, 2 x C**H**H' (C)), 3.31 – 3.15 (m, 5H), 2.67 – 2.48 (m, 2H, C**H$_2$** (C) [overlapping with major rotamer signals]), 2.29 (t, *J* = 6.4 Hz, 2H, C**H$_2$** (B)), 1.81 – 1.65 (m, 4H, 2 x C**H$_2$** (B)) – not all ^{1}H NMR signals for the minor rotamers were found due to overlap with major rotamer and impurity signals. ^{13}C NMR signals for the major rotamer (A) can be found at: δ_C (101 MHz, $CDCl_3$) 174.8 (**C**O), 171.6 (**C**O), 171.0 (**C**O), 64.5 (**C**HCOO), 60.7 (**C**H$_2$OH), 52.8 (**C**H$_3$OCO), 48.5 (**C**H$_2$), 39.2 (**C**H$_2$NH), 37.0 (**C**H$_2$), 28.3 (**C**H$_2$), 25.0 (**C**H$_2$), 23.9 (**C**H$_2$). Diagnostic ^{13}C NMR signals for the minor rotamers can be found at: δ 175.1 (**C**O), 172.7 (**C**O), 171.7 (**C**O), 171.3 (**C**O), 64.3 (**C**HCOO), 63.0 (**C**H$_2$OH (C)), 62.9, 59.2 (**C**H$_2$OH (B)), 56.0, 52.5 (**C**H$_3$OCO (B/C)), 52.2 (**C**H$_3$OCO (B/C)), 45.7 (**C**H$_2$ (C)), 39.0 (**C**H$_2$NH), 42.4, 37.6 (**C**H$_2$), 33.6, 31.5 (**C**H$_2$ (B)), 28.8, 25.7, 24.5, 22.3 (**C**H$_2$ (B)), 20.9 (**C**H$_2$ (B)) – not all ^{13}C NMR signals for the minor rotamers could be found or identified due to overlap with major rotamer and impurity signals; HRMS (ESI$^+$): Calcd. for $C_{12}H_{20}N_2NaO_5$, 295.1264; Found $[M+Na]^+$, 295.1268.

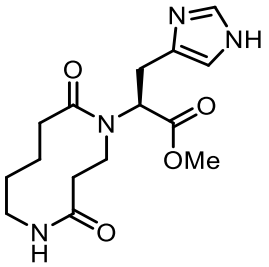
Methyl 5-amino-2-(4,10-dioxo-1,5-diazecan-1-yl)-5-oxopentanoate (2.19)



To a flask charged with glutamine methyl ester hydrochloride (141 mg, 0.72 mmol, 1.10 eq.) in triethylamine (0.370 mL, 2.68 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (102 mg, 0.670 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred for 18 h before the solvent was removed *in vacuo* to afford the crude material as a white powder. This was purified by FCC using a gradient eluent system (9:1; ethyl acetate: methanol → 6:1; ethyl acetate: methanol → 3:1; ethyl acetate: methanol) to afford the *title compound* as a

colourless oil (64 mg, 31%). In $CDCl_3$ the compound exists as a 10:1 (A: B) mixture of rotamers. R_f = 0.21 (3:1; ethyl acetate: methanol); $[\alpha]_D^{20}$ –18.08 (10.0 mg/mL in CH_2Cl_2); ν_{max}/cm^{-1} (thin film) 3320, 3206, 2934, 1716, 1629, 1557, 1432, 1257, 1210, 1175, 1041, 920, 728, 646, 582. ^1H NMR signals for the major rotamer (A) can be found at: δ_H (400 MHz, $CDCl_3$) 7.49 – 7.42 (m, 1H, N**H**), 5.99 (s, 1H, CON**H**H'), 5.91 (s, 1H, CONH**H'**), 4.01 – 3.91 (m, 1H, C**H**H'), 3.86 – 3.67 (m, 5H, C**H**COO, COOC**H₃** & C**H**H'NH), 3.39 (dt, J = 15.5, 3.6 Hz, 1H, C**H**H'), 2.96 – 2.82 (m, 1H, CH**H'**NH), 2.63 – 2.12 (m, 7H, 3 x C**H₂** & C**H**H' [overlapping with unidentifiable minor rotamer signals]), 2.12 – 1.91 (m, 2H, 2 x C**H**H'), 1.66 – 1.34 (m, 2H, CH_2 & C**H**H'). Diagnostic ^1H NMR signals for the minor rotamers can be found at: δ 7.28 – 7.20 (m, 1H, N**H** (B)), 3.62 (s, 3H, COOC**H₃** (B)) - not all ^1H NMR signals for the minor rotamer (B) were found due to overlap with major rotamer signals and signal weakness. ^{13}C NMR signals for the major rotamer (A) can be found at: δ_C (101 MHz, $CDCl_3$) 175.2 (**C**O), 174.1 (**C**O), 173.2 (**C**O), 171.1 (**C**O), 62.0 (**C**HCOO), 53.0 (COO**C**H_3), 48.5 (**C**H_2), 39.0 (**C**HH'NH), 37.0 (**C**H_2), 31.2 (**C**H_2), 28.1 (**C**H_2), 24.8 (**C**H_2), 24.6 (**C**H_2), 23.7 (**C**H_2). Diagnostic ^{13}C NMR signals for the minor rotamer were not found. HRMS (ESI^+): Calcd. for $C_{14}H_{23}N_3NaO_5$, 336.1530; Found $[M+Na]^+$, 336.1527.

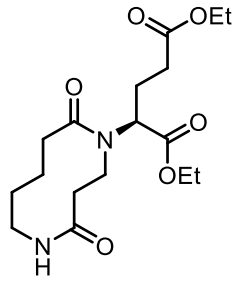
Methyl 2-(4,10-dioxo-1,5-diazecan-1-yl)-3-(1*H*-imidazol-4-yl) propanoate (2.20)



To a flask charged with *L*-histidine methyl ester dihydrochloride (174 mg, 0.72 mmol, 1.10 eq.) in triethylamine (0.360 mL, 2.60 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (99 mg, 0.650 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred for 18 h, before the addition of acetone (2.00 mL), which led to the precipitation of contaminating quaternary ammonium salts, which were removed from the reaction mixture by filtration. The filtrate was concentrated *in vacuo* to give the crude material as an orange oil. This was purified by FCC (6:1; ethyl acetate: methanol) to afford the *title compound* as a white powder (57.1 mg, 27%). M.P. = 47 – 52 ºC; R_f = 0.18 (ethyl acetate: methanol); $[\alpha]_D^{20}$ = –5.12 (10.0 mg/mL in $CDCl_3$); ν_{max}/cm^{-1} (thin film) 3319, 2933, 1717, 1633, 1563, 1433, 1263, 729, 634; δ_H (400 MHz, $CDCl_3$) 7.65 – 7.54 (m, 2H, Ar**H** & N**H**), 6.77 (s, 1H, Ar**H**), 4.21 – 4.11

(m, 1H, C**H**COOCH$_3$), 3.89 – 3.71 (m, 4H, CHCOOC**H$_3$** & C**H**H'NH), 3.76 – 3.53 (m, 1H, C**H**H'), 3.48 – 3.29 (m, 2H, CHC**H$_2$**), 3.00 – 2.88 (m, 2H, CH**H'**NH & C**H**H'), 2.52 – 2.38 (m, 2H, 2 x C**H**H'), 2.15 – 1.92 (m, 3H, 3 x C**H**H'), 1.64 – 1.37 (m, 3H, C**H$_2$** & C**H**H'); δ_C (101 MHz, $CDCl_3$) 173.9 (**C**O), 173.0 (**C**O), 171.3 (**C**O), 136.2 (Ar**C**H), 135.0 (Ar**C**), 115.3 (Ar**C**H), 62.8 (**C**HCOO), 53.1 (COO**C**H$_3$), 48.0 (**C**H$_2$), 39.1 (**C**H$_2$NH), 37.0 (**C**H$_2$), 28.1 (**C**H$_2$), 27.0 (CH**C**H$_2$), 25.0 (**C**H$_2$), 24.0 (**C**H$_2$); HRMS (ESI$^+$): Calcd. for $C_{15}H_{22}N_4NaO_4$, 345.1533; Found [M+Na]$^+$, 345.1531.

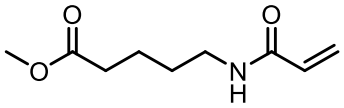
Diethyl 2-(4,10-dioxo-1,5-diazecan-1-yl)pentanedioate (2.21)



To a flask charged with glutamic acid diethyl ester hydrochloride (173 mg, 0.72 mmol, 1.10 eq.) in triethylamine (0.360 mL, 2.60 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (100 mg, 0.650 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred for 18 h before the solvent was removed *in vacuo* to afford the crude material as a white solid. The crude material was purified by FCC using a gradient eluent system (1:1; hexane: ethyl acetate → ethyl acetate) to afford the *title compound* as a colourless oil (85 mg, 37%) which in $CDCl_3$ exists as a complex mixture of rotamers, with the most distinguishable in a 12:1 (A: B) ratio with the major rotamer's signals. R_f = 0.12 (100% ethyl acetate); $[\alpha]_D^{20}$ –18.08 (10.0 mg/mL in $CDCl_3$); ν_{max}/cm^{-1} (thin film) 3323, 2981, 2934, 2240, 1717, 1668, 1639, 1446, 1254, 1181, 1023, 919, 727. ^{1}H NMR signals for the major rotamer (A) can be found at: δ_H (400 MHz, $CDCl_3$) 7.51 – 7.44 (m, 1H, N**H**), 4.28 – 4.13 (m, 2H, COOC**H$_2$**CH$_3$ [overlapping with minor rotamer signals]), 4.15 – 3.98 (m, 2H, COOC**H$_2$**CH$_3$ [overlapping with minor rotamer signals]), 3.99 – 3.90 (m, 1H, C**H**H'), 3.82 – 3.68 (m, 1H, C**H**H'NH), 3.63 – 3.56 (m, 1H, C**H**COOEt), 3.23 (dt, *J* = 15.4, 3.7 Hz, 1H, C**H**H'), 2.88 – 2.79 (m, 1H, CH**H'**NH), 2.56 – 2.21 (m, 6H, C**H**H'CHH'COOEt, CH**H'**CHH'COOEt, CHH'C**H**H'COOEt, CHH'CH**H'**COOEt & 2 x C**H**H'), 2.15 (br dd, *J* = 12.2, 3.8 Hz, 1H, C**H**H'), 2.05 – 1.89 (m, 2x C**H**H'), 1.60 – 1.34 (m, 3H, C**H$_2$** & C**H**H'), 1.26 (t, *J* = 7.1 Hz, 3H, C**H$_3$**), 1.18 (t, *J* = 7.1 Hz, 3H, C**H$_3$** [overlapping

with minor rotamer signals]). Diagnostic ^{1}H NMR signals for the minor rotamer (B) can be found at: δ 7.30 – 7.23 (m, 1H, N**H** (B)), 4.28 – 4.13 (m, 1H, COOC**H**H'CH_3 (B) & COOCH**H'**CH_3 (B) [overlapping with major rotamer signals]), 4.12 – 3.99 (m, 2H, COOC**H₂**CH_3 (B) [overlapping with major rotamer signals]), 3.44 – 3.34 (m, 1H, C**H**COOEt (B)), 1.81 – 1.68 (m, 1H, C**H**H' (B)), 1.24 – 1.12 (m, 6H, 2 x C**H₃** (B) [overlapping with major rotamer signals]) – due to the presence of a complex mixture of rotamers, not all the ^{1}H NMR signals for the distinguishable minor rotamer (B) or other minor rotamers could be found. ^{13}C NMR signals for the major rotamer (A) can be found at: δ_C (101 MHz, $CDCl_3$) 173.8 (**C**O), 173.3 (**C**O), 172.5 (**C**O), 170.8 (**C**O), 62.2 (COO**C**H_2CH_3), 62.1 (C**H**COOEt), 60.6 (COO**C**H_2CH_3), 48.5 (**C**H_2), 38.9 (**C**HH'NH), 36.9 (CHH'**C**HH'COOEt), 30.4 (**C**H_2), 27.9 (**C**H_2), 24.7 (**C**H_2), 24.3 (**C**HH'CHH'COOEt), 23.6 (**C**H_2), 14.2 (**C**H_3), 14.0 (**C**H_3). Diagnostic ^{13}C NMR signals for the minor rotamer (B) can be found at δ 60.9 (COO**C**H_2CH_3), 60.4 (COO**C**H_2CH_3), 53.7 (**C**HCOOEt), 29.7 (**C**H_2), 14.1 (**C**H_3), 14.0 (**C**H_3). The remaining ^{13}C NMR signals for all minor rotamers can be found at: 177.7, 176.0, 175.6, 175.3, 173.1, 173.1, 172.6, 172.0, 171.0, 61.5, 60.8, 56.4, 55.4, 55.2, 53.7, 52.5, 52.0, 40.5, 38.8, 37.7, 31.2, 30.6, 30.5, 29.7, 29.1, 28.5, 25.1, 24.8, 23.9, 23.3; HRMS (ESI^+): Calcd. for $C_{17}H_{28}N_2NaO_6$, 379.1840; Found $[M+Na]^+$, 379.1837.

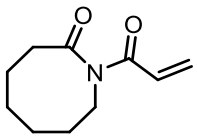
Methyl 5-acrylamidopentanoate (2.23)



To a flask charged with dimethylguanidine sulphate (197 mg, 0.72 mmol, 1.10 eq.) in triethylamine (1.00 mL, 7.20 mmol, 10.0 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (101 mg, 0.65 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred at room temperature for 4 h before the solvent was removed *in vacuo* and the crude material purified by FCC (SiO_2, 1:1; ethyl acetate: hexane → 2:1; ethyl acetate: hexane → 3:1; ethyl acetate: hexane). The *title compound* was isolated as a reaction byproduct as a yellow oil (98.5 mg, 82%); R_f = 0.52 (9:1; ethyl acetate: methanol); δ_H (400 MHz, $CDCl_3$) 6.26 (dd, *J* = 17.0, 1.5 Hz, 1H, NCOC**H**CHH'), 6.08 (dd, *J* = 17.0, 10.2 Hz, 1H, NCOCHC**H**H'), 5.88 (br s, 1H, N**H**), 5.62 (dd, *J* = 10.2, 1.5 Hz, 1H, , NCOCHCH**H'**), 3.66 (s, 3H, C**H₃**), 3.33 (m, 2H, NHC**H₂**), 2.34 (t, *J* = 7.1 Hz, 2H, C**H₂**CO_2Me), 1.74 – 1.51 (m, 4H, 2 x C**H₂**); δ_C (101 MHz, $CDCl_3$) 174.1 (**C**O), 165.7 (**C**O), 131.0 (NCO**C**HCHH'), 126.4 (NCOCH**C**HH'), 51.7 (C**H₃**), 39.2

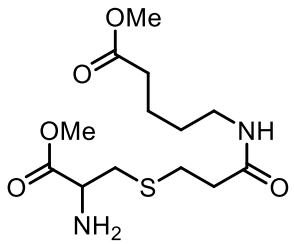
(NH**C**H_2), 33.6 (**C**H_2COMe), 29.0 (**C**H_2), 22.2 (**C**H_2); HRMS (ESI^+): Calcd. for $C_9H_{15}NNaO_3$, 208.0944; Found $[M+Na]^+$, 208.0945. The data obtained match those previously reported.[50]

1-Acryloylazocan-2-one (2.24)



1-Aza-2-cyclooctanone (500 mg, 3.93 mmol, 1.00 eq.) in dry THF (20.0 mL) was cooled to 0 ºC before the dropwise addition of MeMgBr (3.0 M in Et_2O, 1.43 mL, 4.32 mmol, 1.10 eq.) over 30 mins using a syringe pump. After complete addition, the reaction mixture was stirred at 0 ºC for 15 mins. At this point, acryloyl chloride (0.48 mL, 5.90 mmol, 1.50 eq.) was added as a single portion and the reaction was stirred for 1 h at 0 ºC. The reaction was then quenched with sat. aq. NH_4Cl (20.0 mL), the aqueous mixture extracted with Et_2O (30.0 mL) and the organic extracts washed with sat. $NaHCO_3$ (2 x 30.0 mL). The organic phase was dried over $MgSO_4$ and concentrated *in vacuo* to give the crude material as a yellow oil. Purification by FCC (SiO_2, 4:1; hexane: ethyl acetate) gave the *title compound* as low-melting point, colourless crystals (366 mg, 52%); M.P. = 29 – 33 ºC; R_f = 0.15 (4:1; hexane: ethyl acetate); δ_H (400 MHz, $CDCl_3$) 6.88 (dd, *J* = 16.6, 10.5 Hz, 1H, COC**H**=CH_2), 6.31 (dd, *J* = 16.6, 1.7 Hz, 1H, COCH=C**H**H'), 5.68 (dd, *J* = 10.5, 1.7 Hz, 1H, COCHCH**H'**), 3.96 – 3.89 (m, 2H, C**H₂**NCO), 1.95 – 1.84 (m, 2H, C**H₂**CO), 1.77 (p, *J* = 6.2 Hz, 2H, C**H₂**), 1.80 – 1.65 (m, 2H, C**H₂**), 1.66 – 1.53 (m, 2H, C**H₂**), 1.51 – 1.43 (m, 2H, C**H₂**); δ_C (101 MHz, $CDCl_3$) 178.9 (**C**O), 169.4 (**C**O), 132.0 (CO**C**HCH_2), 127.7 (COCH**C**H_2), 43.8 (**C**H_2NCO), 36.7 (C**H₂**CO), 30.0 (**C**H_2), 29.3 (**C**H_2), 26.3 (**C**H_2), 24.0 (**C**H_2). The data obtained match those previously reported.[179]

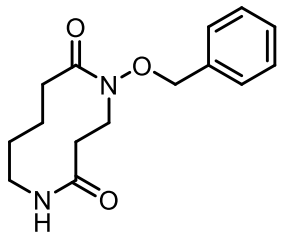
Methyl 5-(3-((2-amino-3-methoxy-3-oxopropyl)thio)propanamido)pentanoate (2.28)



To a flask charged with cysteine methyl ester hydrochloride (124 mg, 0.76 mmol, 1.10 eq.) in triethylamine (0.400 mL, 2.76 mmol, 4.00 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (107 mg, 0.690 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred for 18 h before the solvent was removed *in vacuo* to afford the crude material as a colourless oil. The

crude material was purified by FCC using a gradient eluent system (100% ethyl acetate → 9:1; ethyl acetate: methanol). The *title compound* was isolated as one of many reaction byproducts as a colourless oil (105 mg, 47%). R_f = 0.11 (9:1; ethyl acetate: methanol); ν_{max}/cm^{-1} (thin film) 3281, 2949, 2871, 1732, 1637, 1551, 1496, 1437, 1353, 1199, 1168, 1110, 660, 499; δ_H (400 MHz, $CDCl_3$) 6.83 (s, 2H, N**H₂**), 6.60 – 6.43 (m, 1H, N**H**), 3.67 (s, 3H, COOC**H₃**), 3.64 – 3.56 (m, 4H, $CH_2CHCOOCH_3$ & $CH_2CHCOOCH_3$), 3.26 – 3.20 (m, 2H, C**H₂**), 3.17 (q, *J* = 6.5 Hz, 2H, C**H₂**), 2.95 – 2.80 (m, 1H, C**H**H'CHCOOCH₃), 2.81 – 2.67 (m, 3H, CH**H'**CHCOOCH₃ & C**H₂**), 2.38 (t, *J* = 7.2 Hz, 2H, C**H₂**), 2.30 – 2.22 (m, 1H, C**H₂**), 1.62 – 1.53 (m, 1H, C**H**), 1.53 – 1.39 (m, 1H, C**H**); δ_C (101 MHz, $CDCl_3$) δ 174.4 (**C**O), 174.0 (**C**O), 171.1 (**C**O), 54.1 (CH_2**C**HCOOCH₃), 52.3 (COO**C**H₃), 51.5 (CH_2CHCOO**C**H₃), 42.2 (**C**H₂), 39.0 (**C**H₂), 37.2 (**C**H₂CHCOOCH₃), 36.6 (**C**H₂), 31.4 (**C**H₂), 28.4 (**C**H₂), 22.1 (**C**H₂); HRMS (ESI^+): Calcd. for $C_{13}H_{25}N_2O_5S$, 321.1479; Found $[M+Na]^+$, 321.1482.

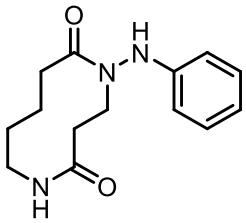
5-(Benzyloxy)-1,5-diazecane-2,6-dione (2.29)



To a flask charged with benzylhydroxylamine (0.08 mL, 0.72 mmol, 1.10 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (102 mg, 0.670 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred for 17 h before the solvent was removed *in vacuo* to afford the crude material as a yellow oil. The crude material was purified by FCC using a gradient eluent system (2:1; hexane: ethyl acetate → 1:1; hexane: ethyl acetate → 100% ethyl acetate → 20:1; ethyl acetate: methanol → 6:1; ethyl acetate: methanol) to afford the *title compound* as a pale-yellow oil (100 mg, 56%). R_f = 0.35 (6:1; ethyl acetate: methanol); ν_{max}/cm^{-1} (thin film) 3300, 2930, 1649, 1535, 1441, 1176, 736, 699, 509; δ_H (400 MHz, $CDCl_3$) 7.41 – 7.19 (m, 5H, 5 x Ar**H**), 5.54 – 5.49 (m, 1H, N**H**), 4.80 (s, 2H, OC**H₂**), 4.64 – 4.49 (m, 1H, C**H**), 3.52 – 3.35 (m, 1H, C**H**H'NH), 3.19 – 3.08 (m, 1H, C**H**), 2.96 – 2.78 (m, 2H, CH**H'**NH & C**H**), 2.54 – 2.45 (m, 1H, C**H**), 2.17 – 2.08 (m, 1H, C**H**), 2.05 – 1.93 (m, 1H, C**H**), 1.91 – 1.82 (m, 1H, C**H**), 1.82 – 1.67 (m, 2H, 2 x C**H**), 1.43 – 1.27 (m, 1H, C**H**); δ_C (101 MHz, $CDCl_3$) 179.5 (**C**O), 170.9 (**C**O), 135.2 (Ar**C**CH_2), 130.5 (2 x Ar**C**), 130.0 (Ar**C**), 129.6 (2 x Ar**C**), 78.2 (**C**H_2), 46.1

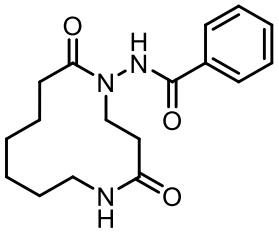
(**C**H_2), 40.7 (**C**H_2NH), 35.2 (**C**H_2), 32.3 (**C**H_2), 28.5 (**C**H_2), 24.2 (**C**H_2); HRMS (ESI$^+$): Calcd. for $C_{15}H_{21}N_2O_3$, 277.1552; Found [M+H]$^+$, 277.1547.

5-(Phenylamino)-1,5-diazecane-2,6-dione (2.30)



To a flask charged with phenylhydrazine (0.07 mL, 0.72 mmol, 1.10 eq.) was added 1-acryloyl-piperidin-2-one **2.01** (100 mg, 0.650 mmol, 1.00 eq.) as a solution in MeOH (1.50 mL). The solution was stirred for 18 h before the solvent was removed *in vacuo* to afford the crude material as an orange solid. The crude material was purified by FCC using a gradient eluent system (1:1; hexane: ethyl acetate → 100% ethyl acetate → 20:1; ethyl acetate: methanol) to afford the *title compound* as a waxy peach solid (57.8 mg, 31%). R_f = 0.16 (100% ethyl acetate); ν_{max}/cm^{-1} (thin film) 3286, 2934, 1628, 1601, 1495, 1443, 1155, 1059, 829, 751, 728, 693, 501; δ_H (400 MHz, $CDCl_3$) 7.29 – 7.07 (m, 2H, 2 x Ar**H**), 6.90 (t, *J* = 7.8 Hz, 1H, Ar**H** (*p*)), 6.73 – 6.64 (m, 2H, 2 x Ar**H**), 6.35 (s, 1H, ArCN**H**), 6.16 – 6.08 (m, 1H, N**H**), 4.39 (ddd, *J* = 13.9, 9.3, 2.2 Hz, 1H, C**H**), 3.95 – 3.80 (m, 1H, C**H**H'NH), 3.24 (dt, *J* = 13.9, 9.3 Hz, 1H, C**H**), 3.08 – 2.94 (m, 1H, C**H**), 2.92 – 2.80 (m, 1H, C**H**), 2.79 – 2.66 (m, 1H, CH**H'**NH), 2.52 – 2.41 (m, 1H, C**H**), 2.06 – 1.75 (m, 4H, 4 x C**H**), 1.71 – 1.58 (m, 1H, C**H**); δ_C (101 MHz, $CDCl_3$) 179.7 (**C**O), 172.2 (**C**O), 145.5 (Ar**C**NH), 129.4 (2 x Ar**C**H), 121.3 (Ar**C**H (*p*)), 113.6 (2 x Ar**C**H), 44.7 (**C**H_2), 40.3 (**C**H_2NH), 34.7 (**C**H_2), 34.3 (**C**H_2), 26.9 (**C**H_2), 24.6 (**C**H_2); HRMS (ESI$^+$): Calcd. for $C_{14}H_{19}N_2NaO_2$, 284.1375; Found [M+Na]$^+$, 284.1369.

***N*-(4,12-dioxo-1,5-diazacyclododecan-1-yl)benzamide (2.31)**



To a flask charged with benzhydrazide (90.2 mg, 0.66 mmol, 1.20 eq.) was added 1-acryloylazocan-2-one **2.24** (100 mg, 0.55 mmol, 1.00 eq.) as a solution in MeOH (2.00 mL). The solution was stirred for 4 h before the solvent was removed *in vacuo* to afford the crude

material as a yellow powder. The crude material was purified by FCC using a gradient eluent system (4:1; ethyl acetate: hexane → 9:1; ethyl acetate: methanol) to afford the *title compound* as a fine white powder (69.4 mg, 40%) which in DMSO-d_6 existed as an 18:2:1 (A:B:C) mixture of rotamers. M.P = 215 – 220 °C. R_f = 0.05 (4:1; ethyl acetate: hexane); ν_{max}/cm^{-1} (thin film) 3291, 2932, 1663, 1611, 1542, 1444, 1408, 1286, 1255, 1217, 1172, 1059, 902, 804, 693. ^{1}H NMR signals for the major rotamer (A) can be found at: δ_H (400 MHz, DMSO-d_6) 9.75 (s, 1H, NN**H**), 7.05 – 6.50 (m, 5H, 5 x Ar**H**), 3.88 (t, *J* = 13.2 Hz, 1H, NC**H**H'CHH'CO), 2.75 – 2.54 (m, 1H, C**H**), 2.01 – 1.93 (m, 1H, NCH**H'**CHH'CO), 1.92 – 1.63 (m, 2H, 2 x C**H**) , 1.67 – 1.47 (m, 2H, NCHH'C**H**H'CO & C**H**), 1.19 – 1.11 (m, 1H, NCHH'CH**H'**CO), 1.04 – 0.80 (m, 2H, 2 x C**H**), 0.84 – 0.63 (m, 1H, C**H**), 0.67 – 0.18 (m, 6H, 6 x C**H**); Diagnostic ^{1}H NMR signals for the minor rotamers can be found at: (s, 1H, NN**H**), 9.17 (s, 1H, NN**H**), 6.28 (s, 1H, CON**H**CH_2), 2.96 – 2.88 (m, 2H, NC**H$_2$**CH_2CO), 2.32 – 2.20 (m, 2H, CONHC**H$_2$**), 1.44 – 1.32 (m, 2H, C**H$_2$**), 0.84 – 0.63 (m, 2H, C**H$_2$**). ^{13}C NMR signals for the major rotamer (A) can be found at: δ_C (101 MHz, DMSO-d_6) 174.3 (**C**O), 170.0 (**C**O), 166.3 (**C**O), 132.4 (Ar**C** & Ar**C**H), 128.8 (Ar**C**H), 127.6 (Ar**C**H), 43.4 (N**C**HH'CHH'CO), 36.8 (**C**HH'), 34.5 (NCHH'**C**HH'CO), 28.9 (**C**HH'), 25.6 (**C**HH'), 24.7 (**C**HH'), 22.3 (2 x **C**HH'); Diagnostic ^{13}C NMR signals for the minor rotamers can be found at: δ 178.1 (**C**O), 175.7 (**C**O), 174.8 (**C**O), 171.7 (**C**O), 170.6 (**C**O), 166.5 (**C**O), 132.2 (Ar**C**H), 128.6 (Ar**C**H), 128.4 (Ar**C**H), 128.3 (Ar**C**H), 127.6 (Ar**C**H), 127.5 (Ar**C**H), 127.4 (Ar**C**H), 48.3, 42.7 (NC**H$_2$**CH_2CO), 40.7 (CONH**C**H_2), 36.3, 35.5, 32.0 (**C**H_2), 31.7 (**C**H_2), 31.0 (**C**H_2), 29.2 (**C**H_2), 28.7 (**C**H_2), 27.7 (**C**H_2), 25.8 (**C**H_2), 25.4 (**C**H_2), 25.3 (**C**H_2), 24.2 (**C**H_2), 23.6 (**C**H_2), 23.1 (**C**H_2), 23.0 (**C**H_2); HRMS (ESI$^+$): Calcd. for $C_{17}H_{24}N_3NaO_3$, 340.1632; Found $[M+Na]^+$, 340.1634.

7.3 Experimental procedures and characterisation data for Chapter 3

7.3.1 General approach to peptide synthesis

Solid-phase peptide synthesis (SPPS) was performed on a CEM Liberty Lite Automated Microwave Peptide Synthesiser, according to the manufacturer's standard protocols. Briefly, Fmoc-protected amino acids (5.5 equiv. (11 equiv. of Fmoc-arginine), 0.2 M in DMF) were coupled in the presence of excess *N*, *N*'-diisopropylcarbodiimide (DIC, 125 equiv, 1.0 M in DMF) and Oxyma Pure (100 equiv, 1.0 M in DMF), as coupling agent and base respectively, under microwave irradiation at a temperature of 25 ºC for 5 seconds, 78 ºC for 30 seconds, 88 ºC for 20 seconds, and 90 ºC for 60 seconds. Fmoc deprotection was performed using 20% piperidine in DMF at 25 ºC for 5 minutes. Syntheses were performed on a 0.1 mmol scale, using Rink Amide MBHA resin (*C*-terminal amide, 0.5 mmol/g loading, 1% DVB, 100-200 mesh, Fluorochem) or Wang resin pre-loaded with the *C*-terminal amino acid (*C*-terminal acid, glycine), unless otherwise stated. Where specified, *N*-terminal acetyl capping was performed in DMF (5 mL) with acetyl chloride (2 equiv.) and DIPEA (4 equiv.) at room temperature for 2 h. Prior to cleavage, the resin was washed sequentially with DCM (3 × 15 mL) and methanol (3 × 15 mL). Peptides were cleaved from the resin in 20 mL of cleavage cocktail (90% TFA, 5% H_2O, 3% TIPS, 2% DTT for Cys-containing sequences) for 4 hrs (18 hrs for Arg-containing sequences). After filtration, the resin was washed extensively with DCM (3 × 50 mL) and the filtrate concentrated *in vacuo* to ~2 mL volume. The residue was dropped into ice cold diethyl ether (~50 mL), and the resultant precipitate collected by centrifugation (3000 rpm, 5 min), resuspended in diethyl ether (~50 mL), and centrifuged again. The residual solid was allowed to air dry for 10 min, then dissolved in deionized water (10 mL) and dried by lyophilisation. Peptides were typically pure by LC-MS analysis and used directly. In cases where the LC-MS analysis showed the sample to contain impurities, the peptide was purified by automated reverse-phase column chromatography using a Teledyne ISCO NextGen 300+ automated flash column chromatography unit on a RediSep Gold® C18 reversed phase column (SiO_2: 400–632 mesh).

7.3.2 Characterisation data of unmodified peptides

SLYRAG (3.01)

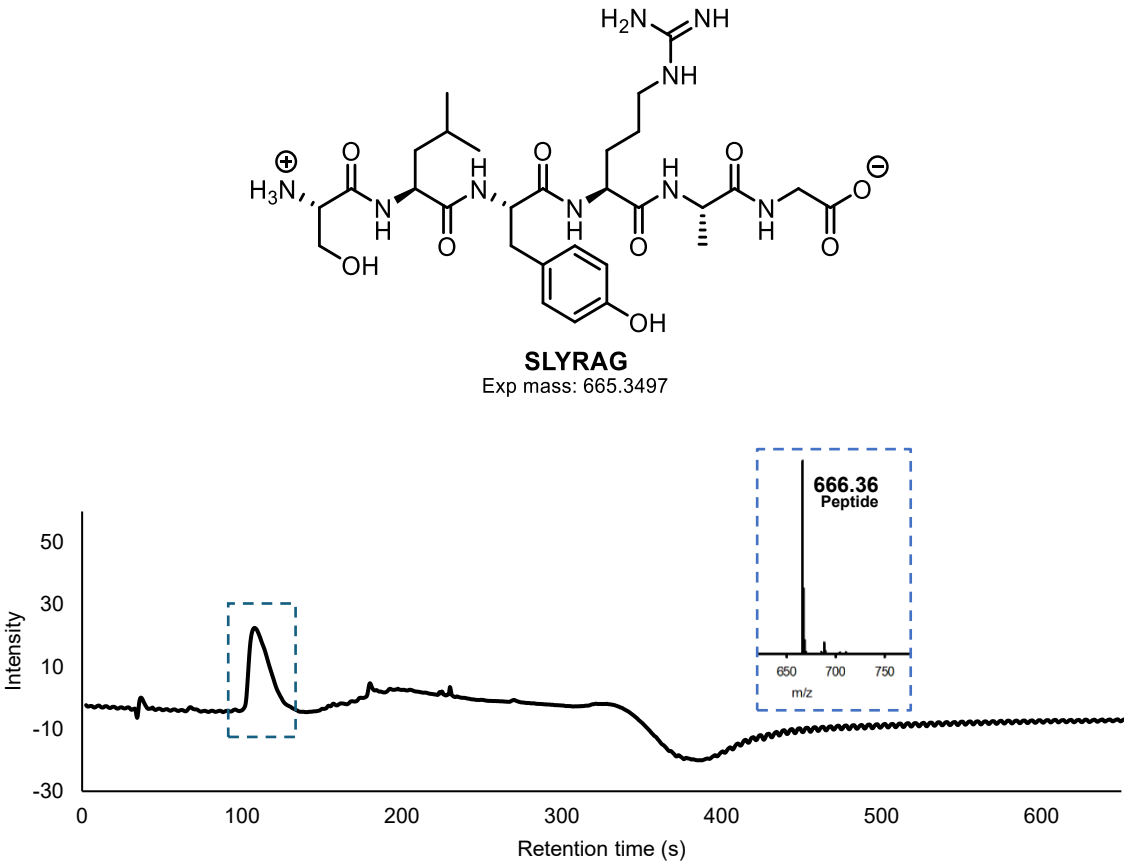
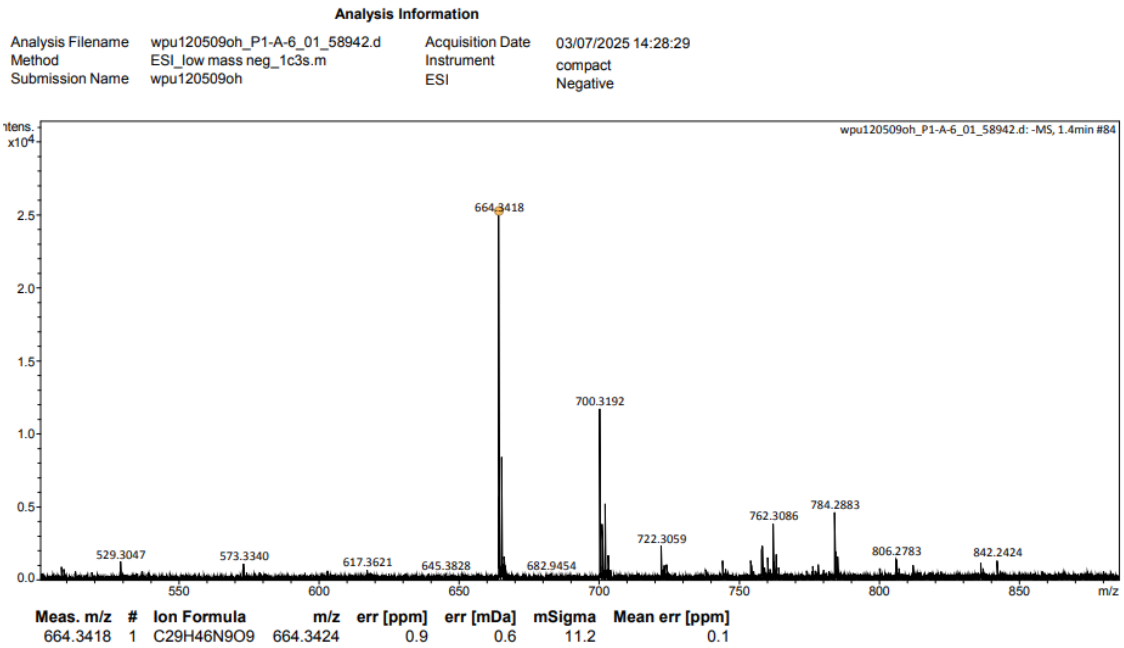


Figure 7.01: Structure and 280 nm chromatogram of hexapeptide, SLYRAG made using automated solid phase peptide synthesis.



Meas. m/z	#	Ion Formula	m/z	err [ppm]	err [mDa]	mSigma	Mean err [ppm]
664.3418	1	C29H46N9O9	664.3424	0.9	0.6	11.2	0.1

Figure 7.02: HRMS spectrum of hexapeptide SLYRAG.

WLYRAG (3.35)

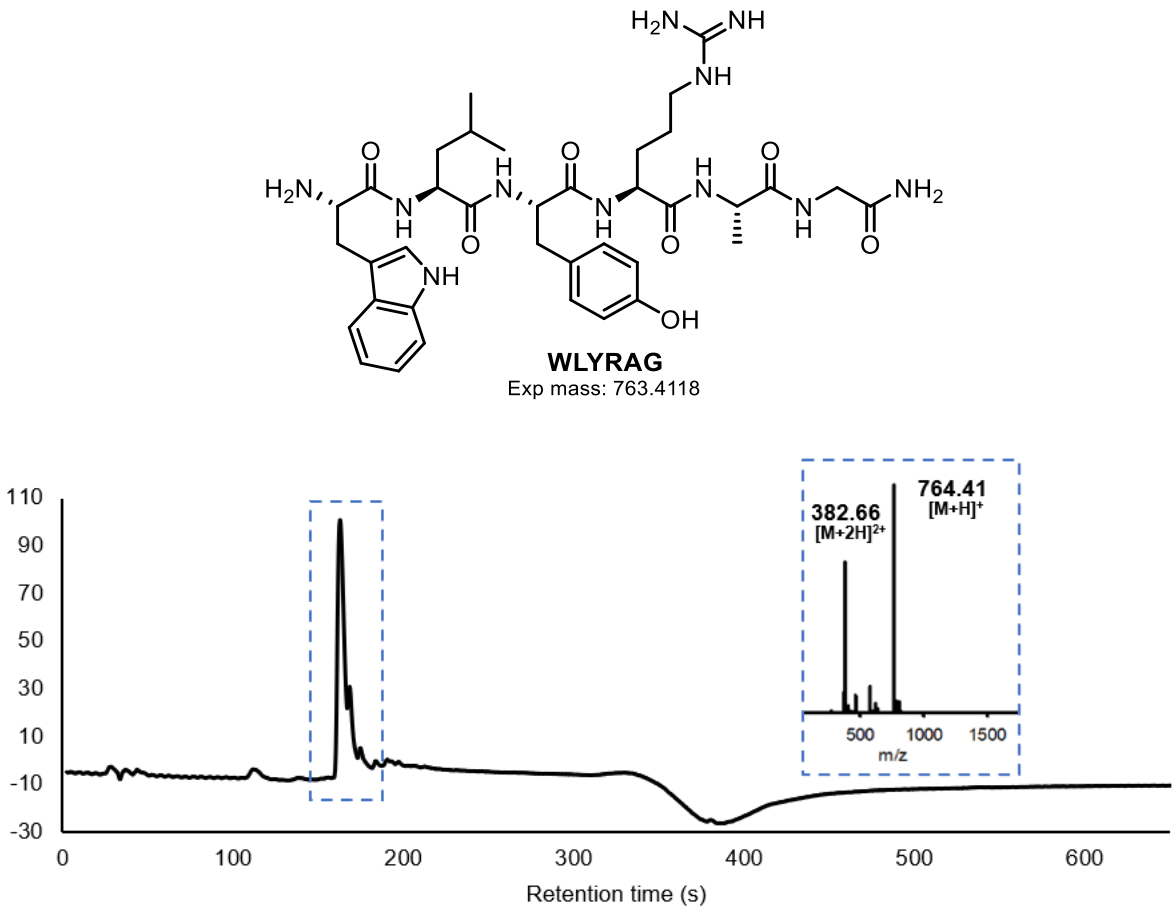


Figure 7.03: Structure and 280 nm chromatogram of hexapeptide, WLYRAG made using automated solid phase peptide synthesis.

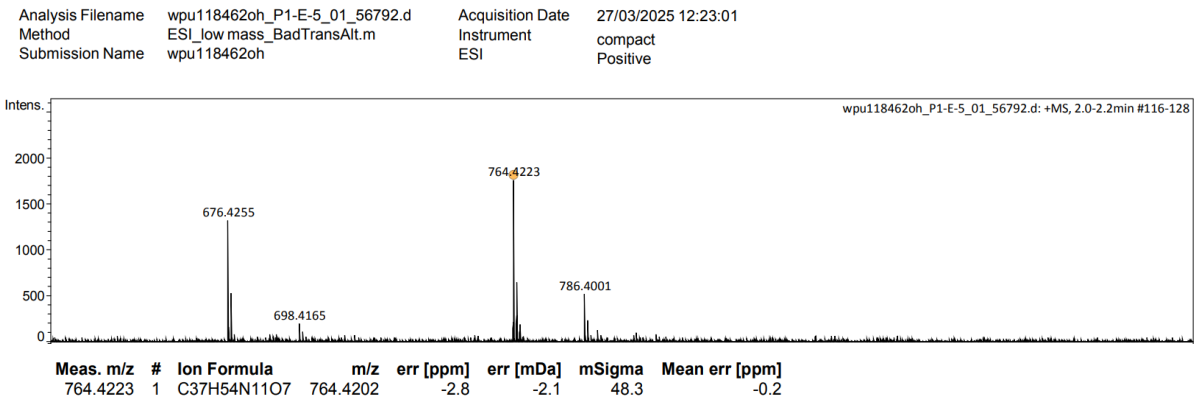


Figure 7.04: HRMS spectrum of hexapeptide WLYRAG.

MLYRAG (3.36)

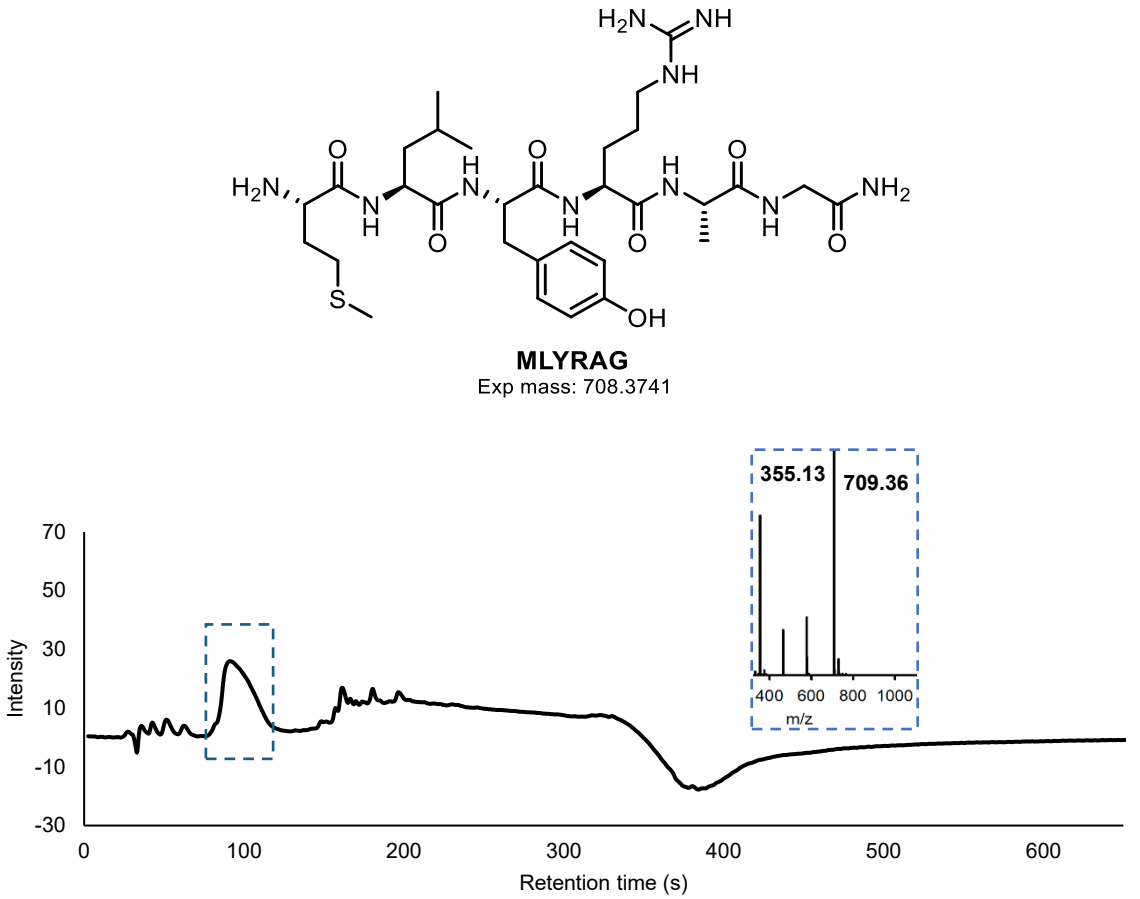
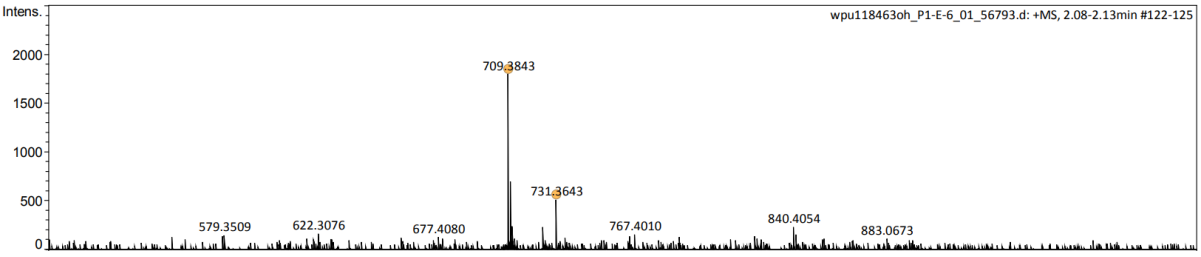


Figure 7.05: Structure and 280 nm chromatogram of hexapeptide, MLYRAG made using automated solid phase peptide synthesis.

Analysis Filename	wpu118463oh_P1-E-6_01_56793.d	Acquisition Date	27/03/2025 12:27:03
Method	ESI_low mass_BadTransAlt.m	Instrument	compact
Submission Name	wpu118463oh	ESI	Positive



Meas. m/z	#	Ion Formula	m/z	err [ppm]	err [mDa]	mSigma	Mean err [ppm]
709.3843	1	C31H53N10O7S	709.3814	-4.2	-3.0	17.1	-3.2
731.3643	1	C31H52N10NaO7S	731.3633	-1.3	-1.0	n.a.	-0.6

Figure 7.06: HRMS spectrum of hexapeptide MLYRAG.

C(Me)LYRAG (3.37)

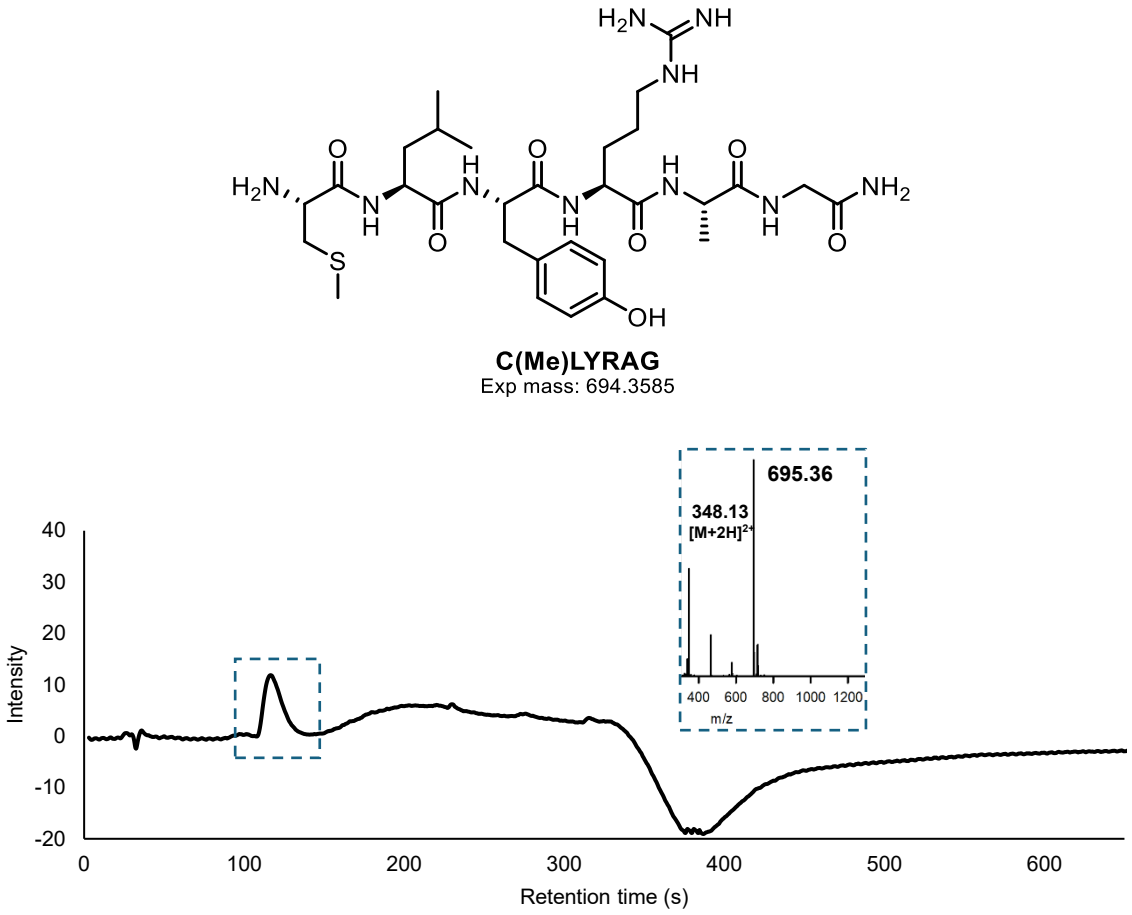
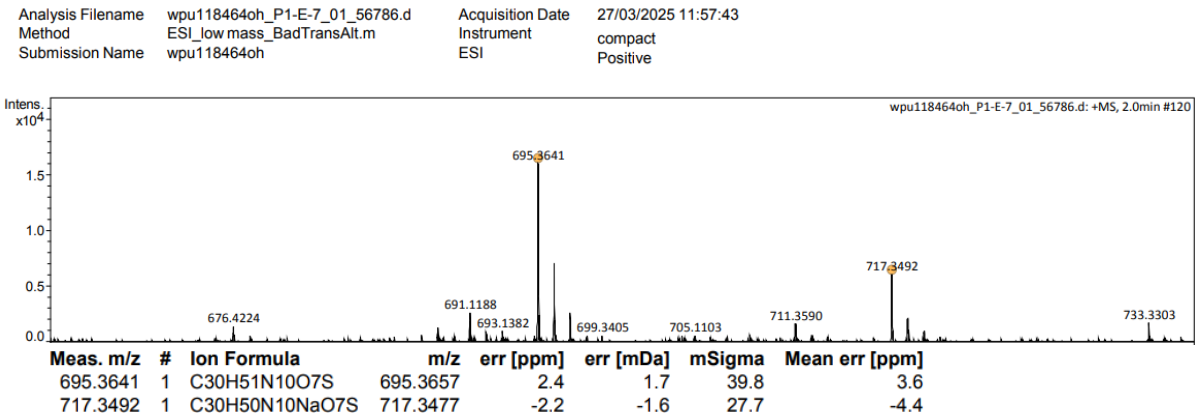


Figure 7.07: Structure and 280 nm chromatogram of hexapeptide, C(Me)LYRAG made using automated solid phase peptide synthesis.



Meas. m/z	#	Ion Formula	m/z	err [ppm]	err [mDa]	mSigma	Mean err [ppm]
695.3641	1	C30H51N10O7S	695.3657	2.4	1.7	39.8	3.6
717.3492	1	C30H50N10NaO7S	717.3477	-2.2	-1.6	27.7	-4.4

Figure 7.08: HRMS spectrum of hexapeptide C(Me)LYRAG.

M(=O)LYRAG (3.38)

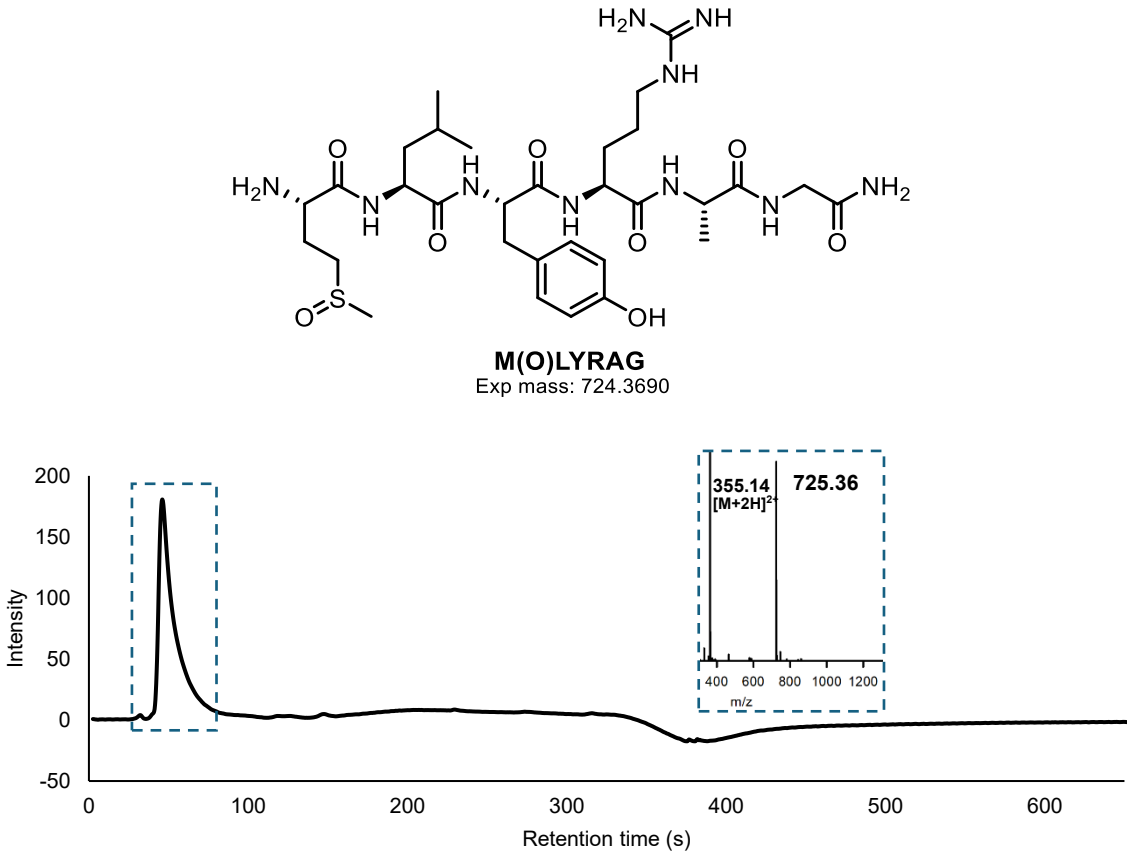
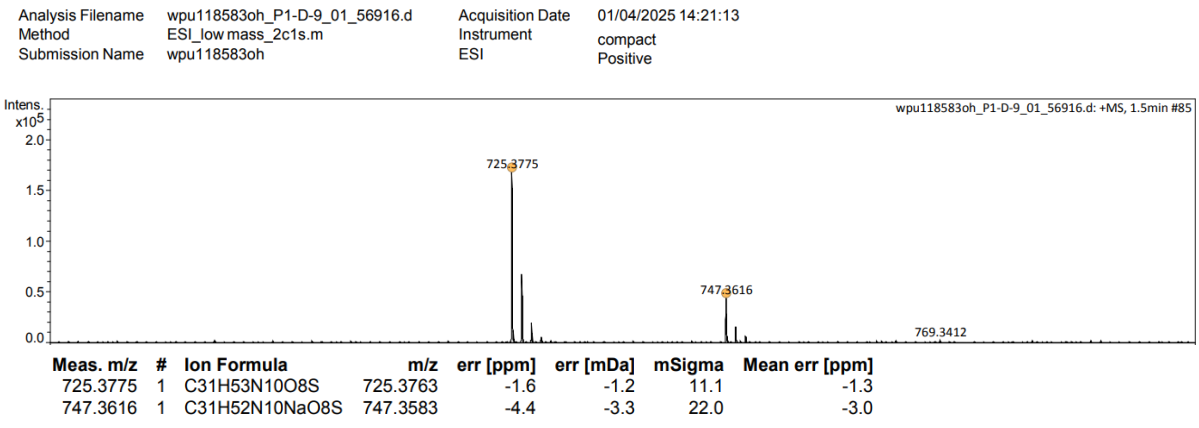


Figure 7.09: Structure and 280 nm chromatogram of hexapeptide, M(=O)LYRAG made using automated solid phase peptide synthesis.



Meas. m/z	#	Ion Formula	m/z	err [ppm]	err [mDa]	mSigma	Mean err [ppm]
725.3775	1	C31H53N10O8S	725.3763	-1.6	-1.2	11.1	-1.3
747.3616	1	C31H52N10NaO8S	747.3583	-4.4	-3.3	22.0	-3.0

Figure 7.10: HRMS of hexapeptide M(=O)LYRAG.

Ac-VYAKHG (3.39)

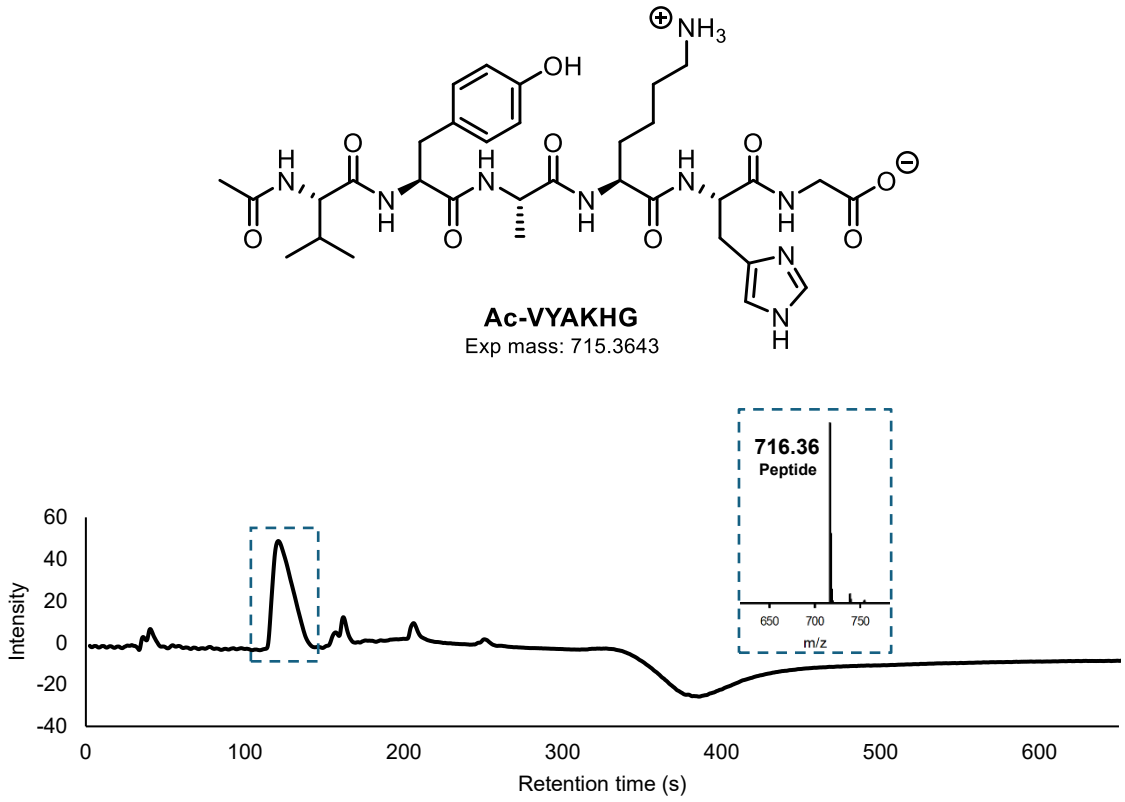
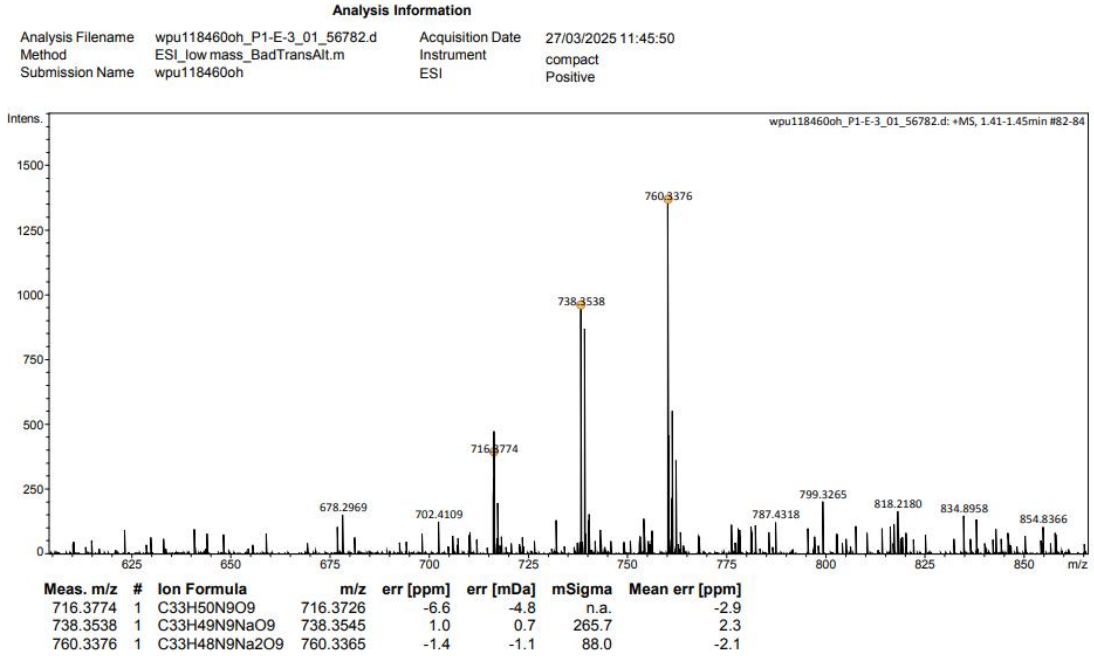


Figure 7.11: Structure and 280 nm chromatogram of hexapeptide, Ac-VYAKHG made using automated solid phase peptide synthesis.



Meas. m/z	#	Ion Formula	m/z	err [ppm]	err [mDa]	mSigma	Mean err [ppm]
716.3774	1	C33H50N9O9	716.3726	-6.6	-4.8	n.a.	-2.9
738.3538	1	C33H49N9NaO9	738.3545	1.0	0.7	265.7	2.3
760.3376	1	C33H48N9Na2O9	760.3365	-1.4	-1.1	88.0	-2.1

Figure 7.12: HRMS of hexapeptide Ac-VYAKHG

7.3.3 General procedures and materials for peptide modifications

7.3.3.1 CARE of acryloyl imides with SLYRAG

To a solution of SLYRAG **3.01** (10 µL, 5 mM stock in HPLC-grade water, 1 mM final concentration, 1.00 eq.), sodium phosphate buffer (10 µL, 0.5 M stock, 0.1 M, pH 7.4) and HPLC-grade water (25 µL) was added reagent (**3.02**, **3.04**, **4.37** or **4.45**) (5 µL; 20 mM, 50 mM, 100 mM or 500 mM stock in DMSO; 2, 5, 10 or 50 mM; 2, 5, 10 or 50 eq.). The reaction was incubated at 37 ºC with agitation (1000 rpm) and aliquots were taken every hour for up to 4 hours and analysed by LC-MS in accordance with the procedure outlined in **LC-MS method A** (section 7.1.3).

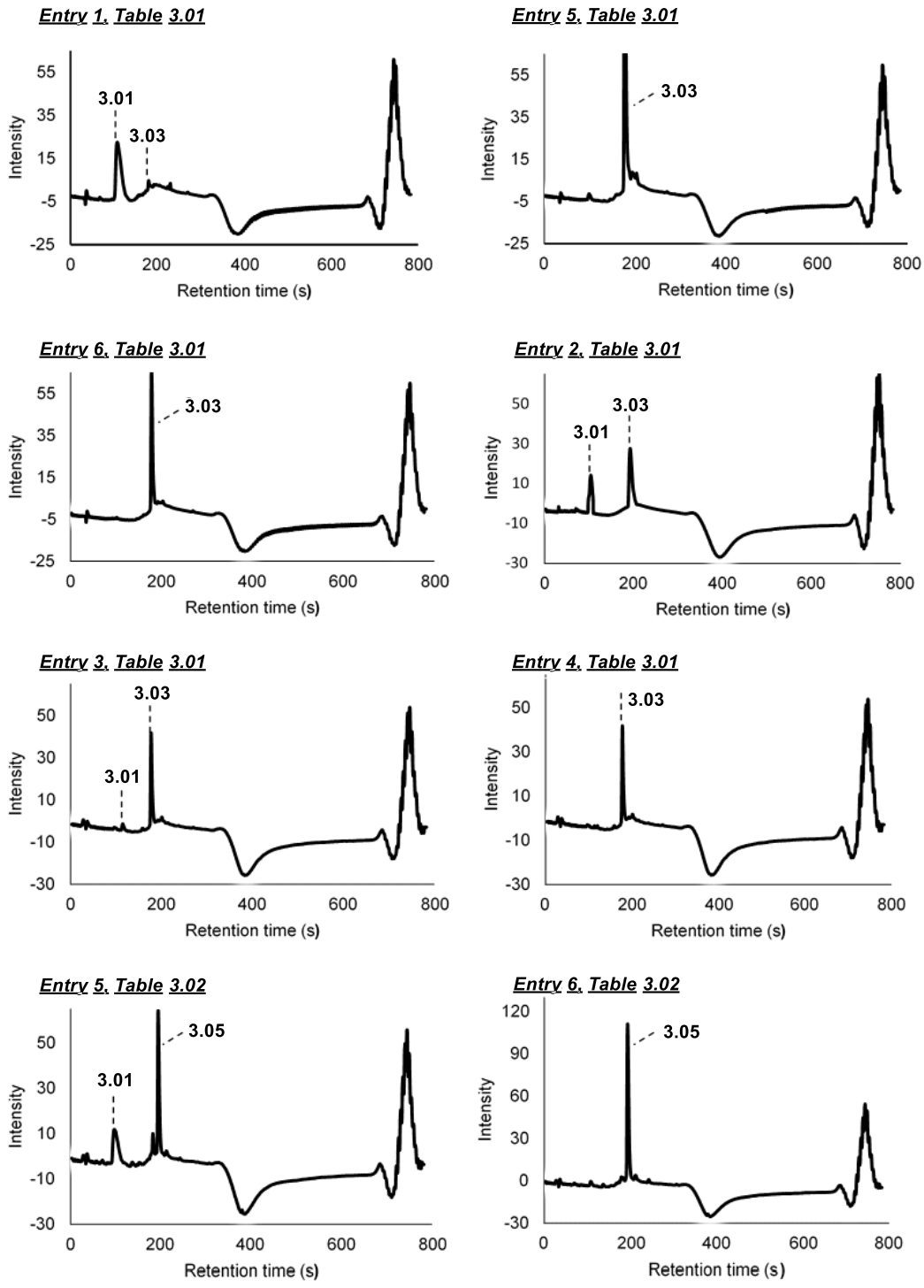


Figure 7.13: 280 nm UV chromatograms from LC-MS analyses of the crude reaction mixtures of SLYRAG **3.01** with acryloyl imides **3.02** and **3.04** under varying conditions (Tables 3.01 and 3.02, Chapter 3).

7.3.3.2 CARE of acryloyl imides with SLYRAG using an organocatalyst

To a solution of SLYRAG **3.01** (10 µL, 5 mM stock in HPLC-grade water, 1 mM final concentration, 1.00 eq.), sodium phosphate buffer (10 µL, 0.5 M stock, 0.1 M, pH 7.4) and HPLC-grade water (10 µL) was added reagent (**3.02**) (5 µL; 100 mM stock in DMSO; 10 mM;

10 eq.) and an organocatalyst ($MgCl_2$, $CaCO_3$, $Sc(OTf)_3$ or thiourea **3.17**) (12.5 µL; 100 mM stock in HPLC-grade water; 25 mM; 25 eq.). The reaction was incubated at 37 ºC with agitation (1000 rpm) for 1 hour before the crude reaction mixture was analysed by LC-MS in accordance with the procedure outlined in **LC-MS method A** (section 7.1.3).

7.3.3.3 Conjugate stability testing of CARE-SLYRAG bioconjugates

A crude sample of CARE-SLYRAG bioconjugate **3.05** was first purified by passing it through an LC-18 reverse phase cartridge, eluting with 30% MeCN in HPLC-grade water. The fractions containing modified peptide **3.05** were collected and lyophilised. The sample was resuspended in HPLC-grade water before the addition of cysteine (100 µL,100 mM stock in HPLC-grade water; 10 mM, 10 eq.). The mixture was incubated at room temperature with agitation (1000 rpm) whilst aliquots were taken every 15 mins and analysed by LC-MS in accordance with the procedure outlined in **LC-MS method A** (section 7.1.3).

7.3.3.4 *N*-terminal modification of SLYRAG using maleimides and acrylamides

To a solution of SLYRAG **3.01** (20 µL, 5 mM stock in HPLC-grade water, 1 mM final concentration, 1.00 eq.), sodium phosphate buffer (20 µL, 0.5 M stock, 0.1 M, pH 7.4) and HPLC-grade water (50 µL) was added reagent (*N*-methylmaleimide or *N*,*N*-dimethylacrylamide) (10 µL; 100 mM stock in DMSO; 10 mM; 10 eq.). The reactions were incubated at 37 ºC with agitation (1000 rpm) for 4 h before the crude reaction mixtures were purified on an LC-18 reverse phase cartridge, eluting with 30 % acetonitrile in HPLC-grade water. The fractions containing modified peptide (**3.30**, **3.32**) were flash frozen and lyophilised. The stability of the conjugates was investigated by resuspending the samples in HPLC-water before the addition of cysteine (100 µL; 100 mM stock in HPLC-grade water, 10 mM, 10 eq.). The mixtures were then incubated at room temperature with agitation (1000 rpm) whilst aliquots were taken every 15 mins and analysed by LC-MS in accordance with the procedure outlined in **LC-MS method A** (section 7.1.3).

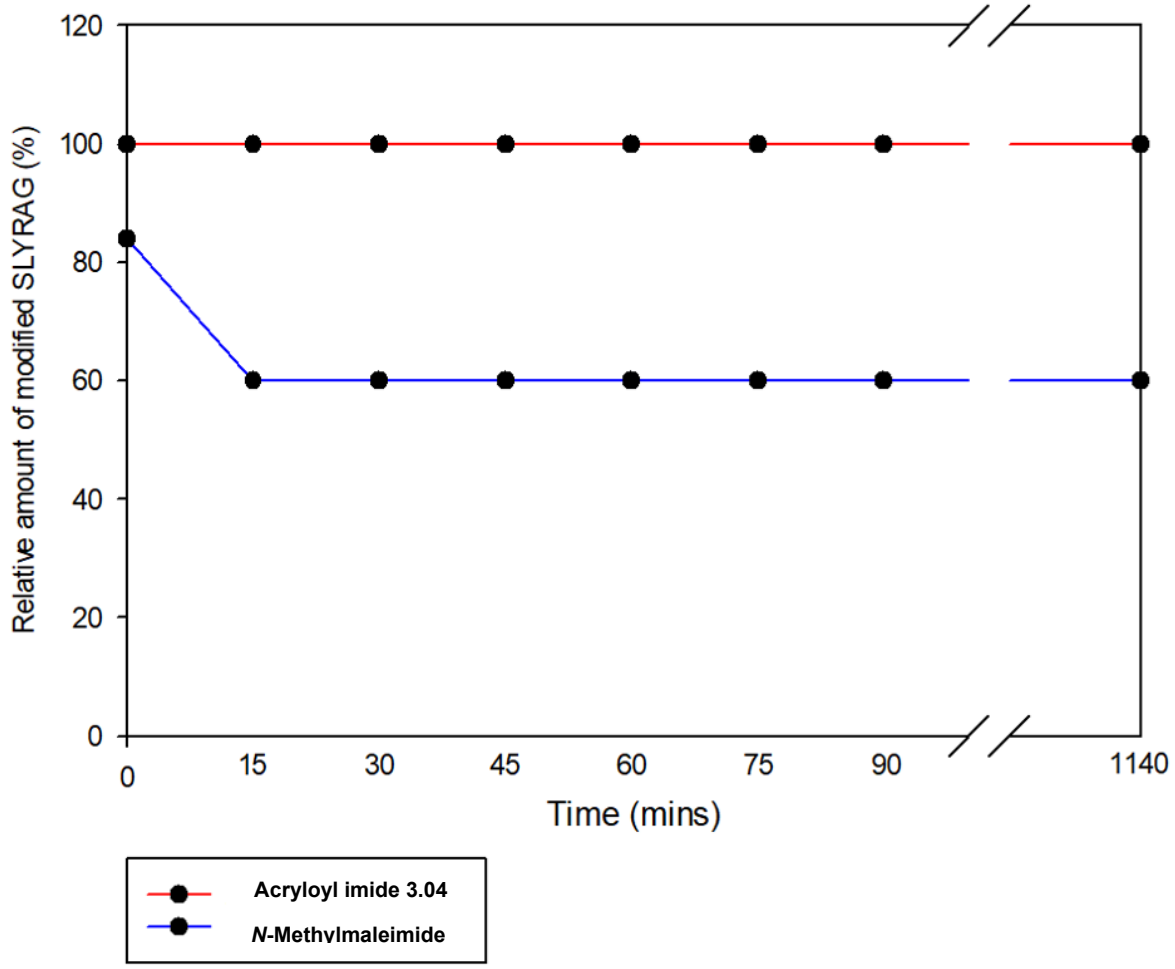


Figure 7.14: Stability of SLYRAG **3.01** modified with **3.04** and *N*-methylmaleimide over time when incubated with 10 equivalents of cysteine.

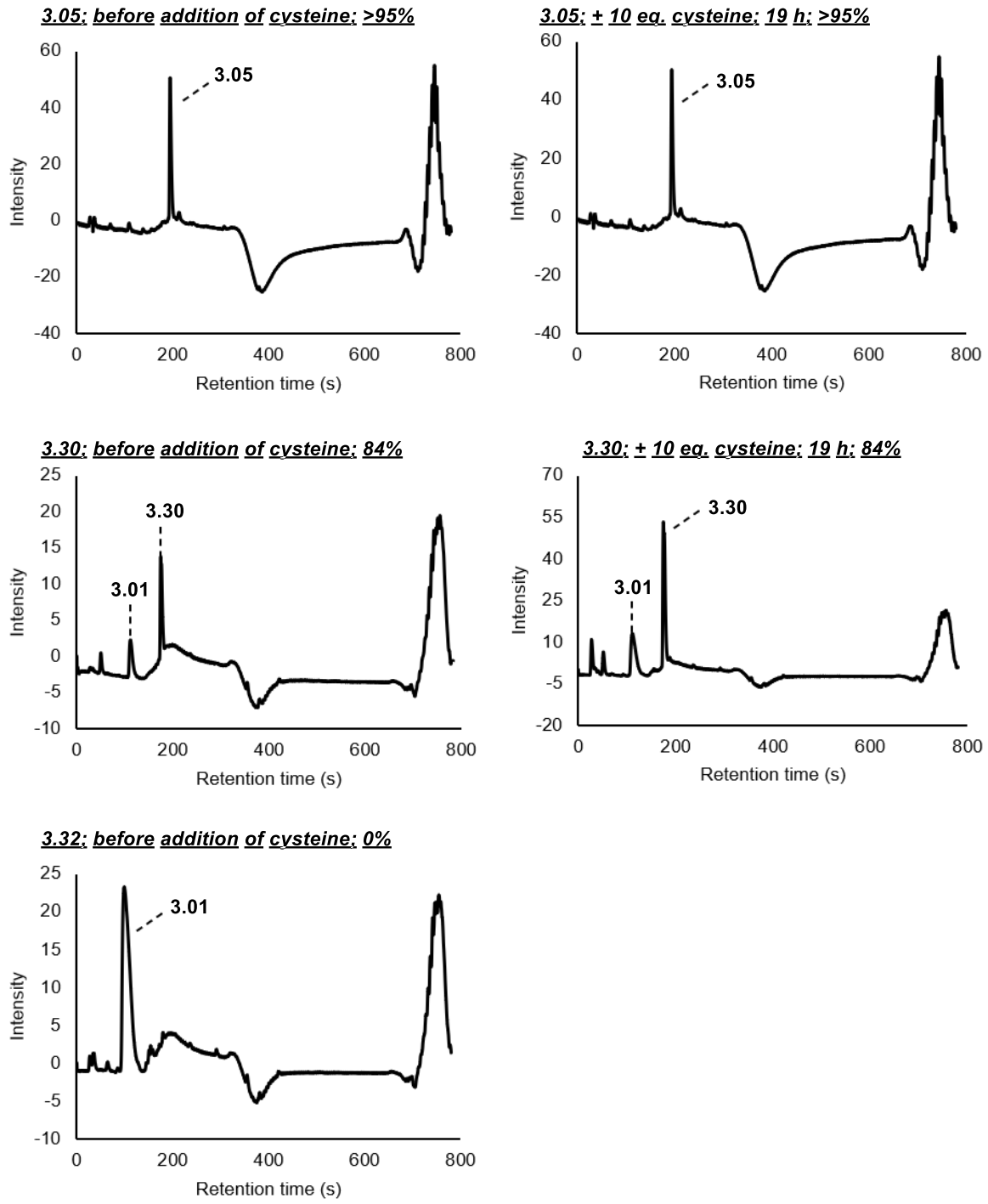
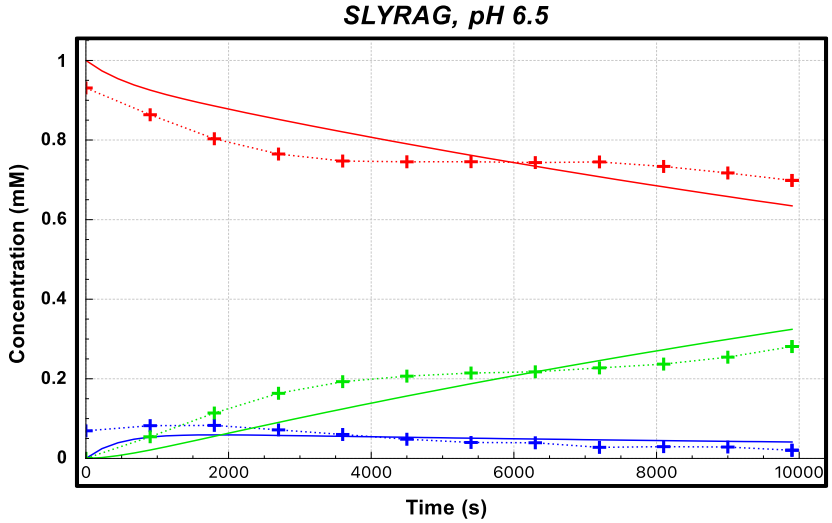


Figure 7.15: 280 nm UV chromatograms from LC-MS analyses of the purified conjugates of SLYRAG **3.01** with **3.04**, N-methylmaleimide (**3.29**) and *N*,*N*-dimethylacrylamide before and after incubation with 10 eq. of cysteine for 19 h.

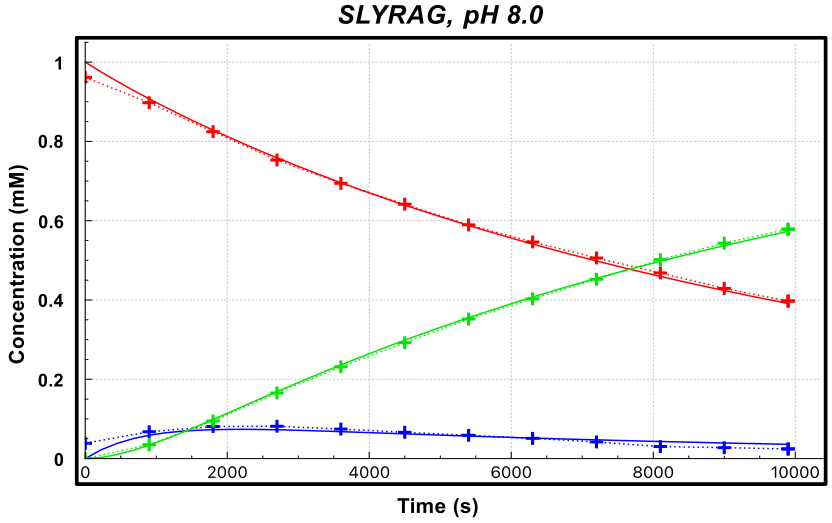
7.3.3.5 Kinetic analysis of the CARE reaction on 6-mer XLYRAG peptides

To a solution of peptide (20 µL, 5 mM stock in HPLC-grade water, 1 mM final concentration, 1.00 eq.), sodium phosphate buffer (20 µL, 0.5 M stock, 0.1 M, pH 7.4) and HPLC-grade water (50 µL) in an LC-MS vial was added reagent (**3.02**, **3.04**, **4.45**) (10 µL, 50 mM in DMSO, 5

mM, 5.00 eq.). The reaction was incubated at 22 °C with no agitation. Samples were analysed every 15 minutes for 3 hours by LC-MS with the procedure outlined in **LC-MS method B** (section 7.1.3).



Entry	pH	Conversion after 3 h (%)	k_1 ($M^{-1}s^{-1}$)	k_{-1} (s^{-1})	k_2 (s^{-1})
1	6.5	28	0.027 ± 0.02	0.001 ± 0.001	6.60 ± 0.7 x10^{-4}

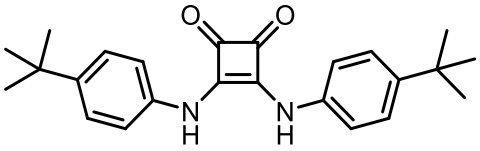


Entry	pH	Conversion after 3 h (%)	k_1 ($M^{-1}s^{-1}$)	k_{-1} (s^{-1})	k_2 (s^{-1})
3	8.0	58	0.023 ± 0.003	1.20 ± 1 x10^{-4}	0.001 ± 5 x10^{-4}

Figure 7.16: Plots of concentration against time for the *N*-terminal modification of SLYRAG with **3.04** at different pHs. Reactions were run as described above (section 7.3.3.5). Fits are based on two-step kinetic model as described in section 7.1.3 **LC-MS method B**. (Table 3.03, Chapter 3).

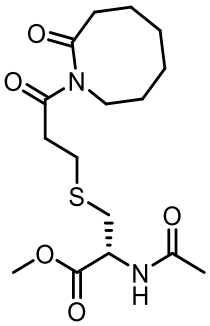
7.3.4 Small molecule synthetic procedure and characterisation data

3,4-Bis((4-(tert-butyl)phenyl)amino)cyclobut-3-ene-1,2-dione (3.22)



To a solution of dimethyl squarate **3.18** (28.4 mg, 0.20 mmol, 1.00 eq.) in anhydrous MeOH (0.50 mL) was added a solution of 4-*tert* butyl aniline (0.03 mL, 0.20 mmol, 1.00 eq.) in anhydrous MeOH (1.00 mL). The reaction was stirred at room temperature for 2 h, by which time an off-white precipitate had formed. *N*-((1R,2R)-2-aminocyclohexyl)-4-methylbenzenesulfonamide (**3.20**) (53.6 mg, 0.20 mmol, 1.00 eq.) was then added as a solution in anhydrous MeOH (1.00 mL) and the mixture stirred at room temperature for a further 2 h. After this time, the resulting off-white precipitate was collected by vacuum filtration and washed with a minimum volume of cold MeOH. The *title* compound was isolated as a reaction byproduct as a peach coloured waxy solid (23.0 mg, 31%). R_f = 0.27 (1:1; ethyl acetate: hexane); δ_H (400 MHz, $CDCl_3$) 7.24 – 7.17 (m, 4H, 4 x Ar**H**), 6.71 – 6.62 (m, 4H, 4 x Ar**H**), 1.30 (s, 18H, 2 x C(C**H**$_3$)$_3$); HRMS (ESI$^+$): Calcd. for $C_{24}H_{28}N_2NaO_2$, 399.2043; Found $[M+Na]^+$, 399.2053.

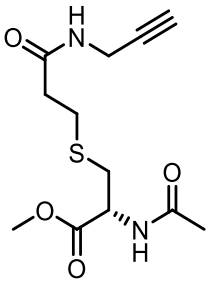
Methyl N-acetyl-S-(3-oxo-3-(2-oxoazocan-1-yl)propyl)-L-cysteinate (3.43)



To a flask charged with *N*-acetyl-*L*-cysteine methyl ester (**3.42**) (294 mg, 1.66 mmol, 3.00 eq.) and 1-acryloylazocan-2-one (**3.04)** (100 mg, 0.55 mmol, 1.00 eq.) in THF (3.00 mL) was added DIPEA (0.38 mL, 2.20 mmol, 4.00 eq.). The reaction mixture was stirred at room temperature for 2 h before being concentrated *in vacuo* and the crude material purified by FCC (SiO_2, 4:1; ethyl acetate: hexane) to afford the *title compound* as a waxy white solid (187 mg, 95 %). R_f = 0.15 (4:1; ethyl acetate: hexanes); $[\alpha]_D^{20}$ 22.58 (10.0 mg/mL in $CDCl_3$); ν_{max}/cm^{-1} (thin film) 3307, 2929, 2860, 1744, 1674, 1533, 1438, 1372, 1244, 1196, 1175, 1125, 1092, 999, 594; δ_H

(400 MHz, $CDCl_3$) 6.60 (d, *J* = 7.8 Hz, 1H, N**H**$COCH_3$), 4.83 (dt, *J* = 7.8, 4.8 Hz, 1H, CH_2C**H**$COOCH_3$), 3.94 – 3.86 (m, 2H, CONC**H₂**$CH_2CH_2CH_2CH_2CH_2$), 3.74 (s, 3H, $CH_2CHCOOC$**H₃**), 3.12 (t, *J* = 7.0 Hz, 2H, COC**H₂**CH_2S), 3.07 – 2.92 (m, 2H, C**H₂**$CHCOOCH_3$), 2.81 (t, *J* = 7.0 Hz, 2H, $COCH_2C$**H₂**S), 2.68 – 2.60 (m, 2H, $CONCH_2CH_2CH_2CH_2CH_2C$**H₂**), 2.04 (s, 3H, $NHCOC$**H₃**), 1.91 – 1.81 (m, 2H, $CONCH_2CH_2CH_2CH_2C$**H₂**CH_2), 1.74 – 1.64 (m, 2H, $CONCH_2C$**H₂**$CH_2CH_2CH_2CH_2$), 1.63 – 1.51 (m, 2H, $CONCH_2CH_2CH_2C$**H₂**CH_2CH_2), 1.50 – 1.38 (m, 2H, $CONCH_2CH_2C$**H₂**$CH_2CH_2CH_2$); δ_C (101 MHz, $CDCl_3$) 178.4 (**C**$ONCH_2CH_2CH_2CH_2CH_2CH_2$), 174.8 (**C**$OCH_2CH_2S$), 171.4 ($CH_2CH$**C**$OOCH_3$), 170.1 (NH**C**$OCH_3$), 77.4, 52.7 ($CH_2CHCOO$**C**$H_3$), 52.2 ($CH_2$**C**$HCOOCH_3$), 43.6 (CON**C**$H_2CH_2CH_2CH_2CH_2CH_2$), 39.8 (CO**C**$H_2CH_2S$), 37.2 ($CONCH_2CH_2CH_2CH_2CH_2$**C**$H_2$), 34.7 (**C**$H_2CHCOOCH_3$), 29.5 ($CONCH_2$**C**$H_2CH_2CH_2CH_2CH_2$), 29.2 ($CONCH_2CH_2CH_2CH_2$**C**$H_2CH_2$), 28.2 ($COCH_2$**C**$H_2S$), 26.3 ($CONCH_2CH_2CH_2$**C**$H_2CH_2CH_2$), 24.0 ($CONCH_2CH_2$**C**$H_2CH_2CH_2CH_2$), 23.2 (NHCO**C**$H_3$); HRMS ($ESI^+$): Calcd. for $C_{16}H_{26}N_2NaO_5S$, 381.1455; Found $[M+Na]^+$, 381.1455.

Methyl N-acetyl-S-(3-oxo-3-(prop-2-yn-1-ylamino)propyl)cysteinate (3.45)

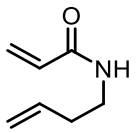


Methyl *N*-acetyl-S-(3-oxo-3-(2-oxoazocan-1-yl)propyl)-*L*-cysteinate (**3.43**) (100 mg, 0.28 mmol, 1.00 eq.) was dissolved in dry THF (0.80 mL) before the complete addition of propargylamine (0.04 mL, 0.56 mmol, 2.00 eq.). The reaction was stirred at room temperature for 19 h before the solvent was removed *in vacuo* and the crude material purified by FCC (SiO_2, 100 % ethyl acetate). The *title compound* was obtained as a reaction byproduct as a viscous orange oil (65.5 mg, 82 %). R_f = 0.17 (100 % ethyl acetate); δ_H (400 MHz, $CDCl_3$) 6.85 (br d, *J* = 7.9 Hz, 1H, N**H**$COCH_3$), 6.59 (br t, *J* = 5.5 Hz, 1H, SCH_2CH_2CON**H**CH_2CCH), 4.81 (dt, *J* = 7.9, 5.3 Hz, 1H, CH_2C**H**$COOCH_3$), 4.05 – 3.99 (m, 2H, SCH_2CH_2CONHC**H₂**CCH), 3.75 (s, 3H, $CH_2CHCOOC$**H₃**), 3.04 – 2.90 (m, 2H, C**H₂**$CHCOOCH_3$), 2.83 (t, *J* = 7.0 Hz, 2H, SCH_2C**H₂**$CONHCH_2CCH$), 2.56 – 2.39 (m, 2H, SC**H₂**$CH_2CONHCH_2CCH$), 2.23 (t, *J* = 2.5 Hz, 1H, $SCH_2CH_2CONHCH_2CC$**H**), 2.07 (s, 3H, $NHCOC$**H₃**); δ_C (101 MHz, $CDCl_3$) 171.3

($CH_2CH\mathbf{C}OOCH_3$), 171.0 ($SCH_2CH_2\mathbf{C}ONHCH_2CCH$), 170.8 ($NH\mathbf{C}OCH_3$), 79.5 ($SCH_2CH_2CONHCH_2C\mathbf{C}H$), 71.7 ($SCH_2CH_2CONHCH_2\mathbf{C}CH$), 52.9 ($CH_2CHCOO\mathbf{C}H_3$), 52.5 ($CH_2C\mathbf{H}COOCH_3$), 36.2 ($S\mathbf{C}H_2CH_2CONHCH_2CCH$), 34.5 ($\mathbf{C}H_2CHCOOCH_3$), 29.3 ($SCH_2CH_2CONH\mathbf{C}H_2CCH$), 28.4 ($SCH_2\mathbf{C}H_2CONHCH_2CCH$), 23.1 ($NHCO\mathbf{C}H_3$); HRMS ($ESI^+$): Calcd. for $C_{12}H_{18}N_2NaO_4S$, 309.0879; Found $[M+Na]^+$, 309.0876 (1.1 ppm error).

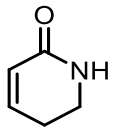
7.4 Experimental procedures and characterisation data for Chapter 4

_N_-(but-3-en-1-yl)acrylamide (4.30)



A solution of homoallylamine (1.29 mL, 14.06 mmol, 1.00 eq.) in anhydrous DCM (30.0 mL) was cooled to 0 °C before the dropwise addition of triethylamine (2.92 mL, 21.09 mmol, 1.50 eq.). The mixture was stirred at this temperature for 10 mins before the dropwise addition of acryloyl chloride (1.37 mL, 16.87 mmol, 1.20 eq.). The reaction mixture was slowly allowed to reach room temperature whilst stirring for 3 h before being quenched with water (20.0 mL). The organic portion was collected and the aqueous portion extracted with DCM (3 x 20.0 mL). The combined organic portions were washed with sat. aq. NH_4Cl (3 x 20 mL) before being dried over $MgSO_4$ and concentrated *in vacuo*. The *title compound* did not require FCC and was isolated as a yellow oil (1.68 g, 96%); δ_H (400 MHz, $CDCl_3$) 6.25 (dd, *J* = 17.1, 1.5 Hz, 1H, C**H**H'CHCONH), 6.05 (dd, *J* = 17.1, 10.4 Hz, 1H, CHH'C**H**CONH), 5.84 – 5.68 (m, 1H, CHH'C**H**CH_2CH_2), 5.61 (dd, *J* = 10.4, 1.5 Hz, 1H, CH**H'**CHCONH), 5.14 – 5.03 (m, 2H, C**HH'**$CHCH_2CH_2$), 3.40 (app q, *J* = 6.5 Hz, 2H, CHH'$CHCH_2C$**H_2**), 2.34 – 2.23 (m, 2H, CHH'CHC**H_2**CH_2). The data obtained match those previously reported.[153]

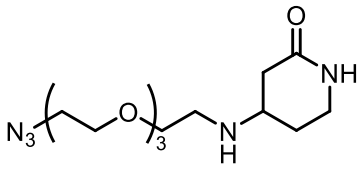
5,6-Dihydropyridin-2(1H)-one (4.27)



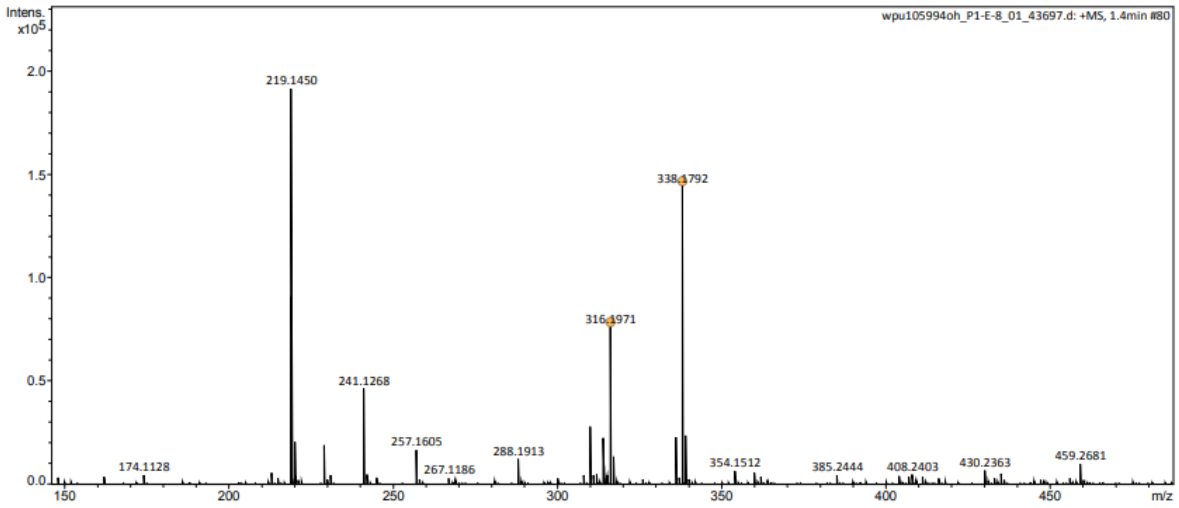
A solution of *N*-(but-3-en-1-yl)acrylamide (**4.30**) (300 mg, 2.40 mmol, 1.00 eq.) in anhydrous DCM (400 mL) was purged with argon before the addition of Grubbs 2nd generation catalyst (102 mg, 0.12 mmol, 5 mol%). The reaction mixture was heated to reflux for 18 h before being filtered through Celite and concentrated *in vacuo* to give the crude material as white solid. The crude material was purified by FCC (SiO_2, 100% ethyl acetate) to give the *title compound* as a waxy white solid (110 mg, 93%); δ_H (400 MHz, $CDCl_3$) 6.94 (br s, 1H, N**H**), 6.58 (dt, *J* = 10.0, 4.2 Hz, 1H, CH_2CH_2C**H**CHCONH), 5.83 (app dq, *J* = 10.0, 1.9 Hz, 1H, CH_2CH_2CHC**H**CONH), 3.36 (td, *J* = 7.1, 2.6 Hz, 2H, C**H_2**$CH_2CHCHCONH$), 2.35 – 2.22 (m, 2H, CH_2C**H_2**CHCHCONH); δ_C (101 MHz, $CDCl_3$) 166.7 (**C**O), 141.5

(CH_2CH_2**C**HCHCONH), 124.7 (CH_2CH_2CH**C**HCONH), 39.4 (**C**$H_2CH_2CHCHCONH$), 23.7 (CH_2**C**$H_2CHCHCONH$). The data obtained match those previously reported.[180]

4-((2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)amino)piperidin-2-one (4.33)



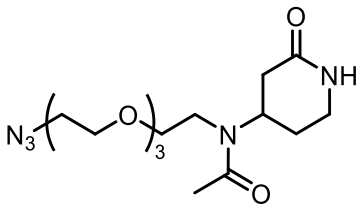
5,6-Dihydropyridin-2(1H)-one (**4.27**) (68.0 mg, 0.70 mg, 1.20 eq.), 1-amino-11-azido-3,6,9-trioxaundecane (**4.32**) (0.12 mL, 0.58 mmol, 1.00 eq.), and CuO(II) (5.00 mg, 10 mol%) were added to a flask and dissolved in anhydrous MeOH (5.00 mL). The reaction mixture was heated to reflux and stirred for 20 h before being concentrated *in vacuo* to give the crude material as a dark orange oil. The *title compound* could not be isolated by FCC but was evidenced by crude MS analysis. The crude material was therefore carried forward to the synthesis of **4.34**. HRMS (ESI^+): Calcd. for $C_{13}H_{25}N_5NaO_4$, 338.1799; Found $[M+Na]^+$, 338.1792.



Meas. m/z	#	Ion Formula	m/z	err [ppm]	err [mDa]	mSigma	Mean err [ppm]
316.1971	1	C13H26N5O4	316.1979	2.5	0.8	6.7	2.3
338.1792	1	C13H25N5NaO4	338.1799	1.9	0.6	1.6	1.7

Figure 7.17: HRMS of the crude reaction mixture containing compound **4.33**.

N-(2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)-N-(2-oxopiperidin-4-yl)acetamide (4.34)



The crude reaction mixture containing 4-((2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)amino)piperidin-2-one (**4.33**) was redissolved in anhydrous DCM (5.00 mL) and cooled to 0 °C before the dropwise addition of triethylamine (0.39 mL, 2.80 mmol, 4.00 eq.). The mixture was stirred at this temperature for 10 mins before the dropwise addition of acetyl chloride (0.075 mL, 1.05 mmol, 1.50 eq.). The reaction mixture was slowly allowed to reach room temperature whilst stirring overnight before the solvent was removed *in vacuo* and the crude material purified by FCC (8:1; ethyl acetate: methanol) to give the *title compound* as a yellow oil (59.8 mg, 24% over 2 steps). The ^{1}H NMR suffered from severe signal broadening and the ^{13}C NMR spectrum from multiple overlapping signals, suggesting the presence of rotamers or possible impurities. This made it difficult to confirm the structure with certainty. Mass spectrometry data included a molecular ion peak consistent with the expected mass of **4.34**. HRMS (ESI^{+}): Calcd. for $C_{15}H_{27}N_5NaO_5$, 380.1904; Found [M+Na] $^{+}$, 380.1905.

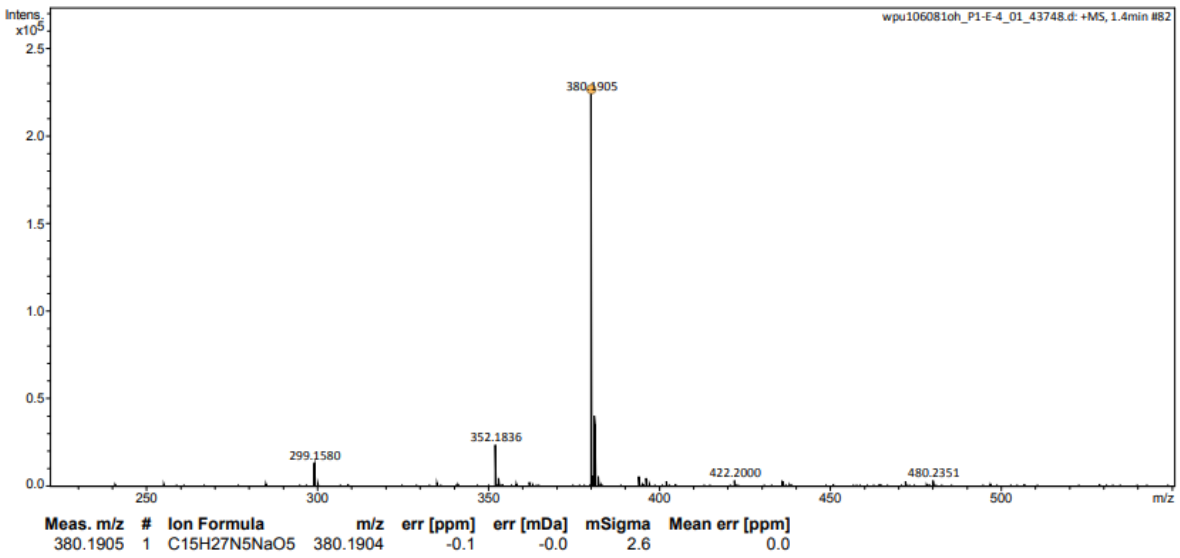


Figure 7.18: HRMS of purified fractions containing **4.34**.

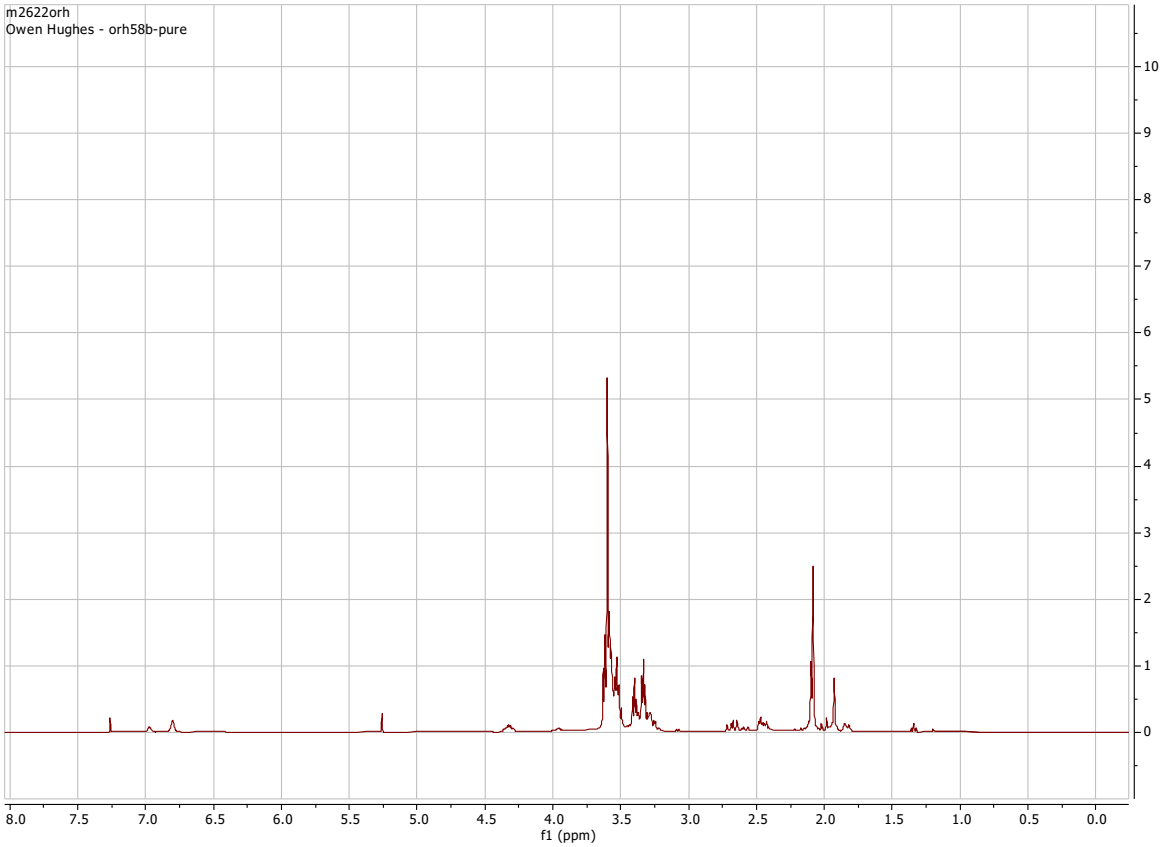


Figure 7.19: ^{1}H NMR of purified fraction containing **4.34**.

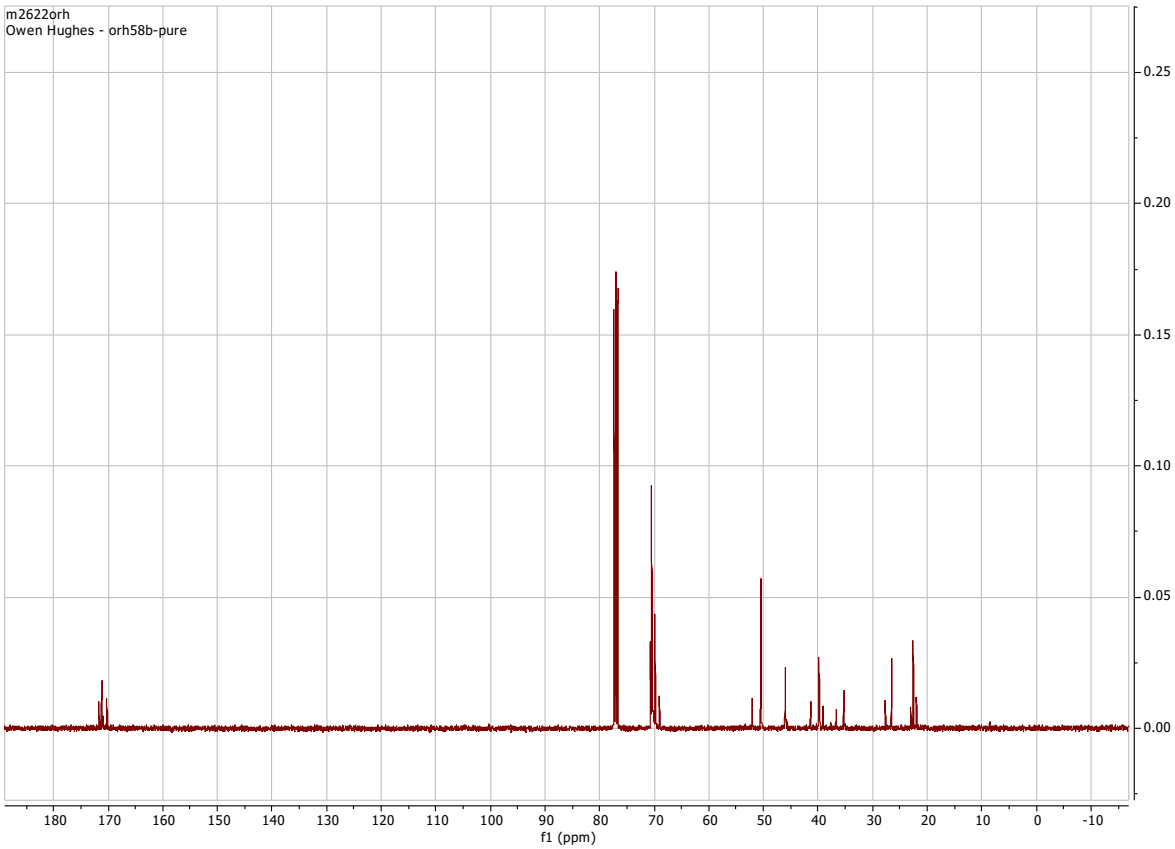
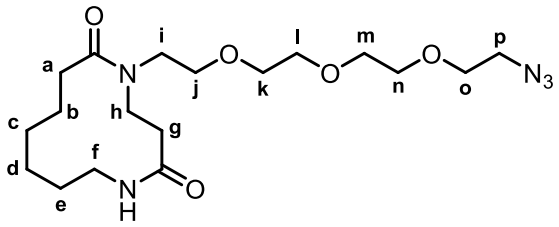


Figure 7.20: ^{13}C NMR of purified fraction containing **4.34**.

5-(2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)-1,5-Diazacyclododecane-2,6-dione (4.36)

The ^{1}H spectrum for this compound was extremely complex, with multiple overlapping signals. The compound's identity is best confirmed using its ^{13}C NMR spectrum which is better resolved.



To a flask charged with 1-Acryloylazocan-2-one (**4.35**) (500 mg, 2.76 mmol, 1.00 eq.) in MeOH (7.00 mL) was added 1-amino-11-azido-3,6,9-trioxaundecane (**4.32**) (0.660 mL, 3.31 mmol, 1.20 eq.). The solution was stirred for 18 h before the solvent was removed *in vacuo* to afford the crude material as an orange oil. The crude material was purified by FCC (9:1; ethyl acetate: methanol) to afford the *title compound* as a golden oil (933 mg, 85%). In $CDCl_3$ the compound exists as a 5:2 (A:B) mixture of rotamers as determined by NMR. R_f = 0.15 (9:1; ethyl acetate: methanol); ν_{max}/cm^{-1} (thin film) 3313, 2923, 2864, 2102, 1635, 1550, 1449, 1349, 1289, 1103, 931, 642, 574; δ_H (400 MHz, $CDCl_3$) 6.97 – 6.89 (m, 1H, N**H** (A)), 6.56 – 6.48 (m, 0H, N**H** (B)), 4.59 (ddd, *J* = 14.7, 10.1, 4.9 Hz, 2H, C**H₂** (B)), 3.78 – 3.52 (m, 16H, $\mathbf{H}_{j\text{-}o}$ (A)), 3.52 – 3.43 (m, 1H, C**H₂** (B)), 3.43 – 3.28 (m, 4H, 2 x C**H₂** (A)), 3.02 – 2.87 (m, 1H, C**H** (A)), 2.72 (dt, *J* = 15.6, 4.9 Hz, 1H, C**H** (A)), 2.62 – 2.48 (m, 1H, C**H** (A)), 2.47 – 2.38 (m, 1H, C**H** (A)), 2.31 – 2.10 (m, 2H, C**H₂** (B)), 1.89 – 1.76 (m, 2H, C**H₂** (B)), 1.67 – 1.53 (m, 2H, C**H₂** (A)), 1.52 – 1.31 (m, 4H, 2 x C**H₂** (A)), 1.23 – 1.00 (m, 2H, C**H₂** (A)); δ_C (101 MHz, $CDCl_3$) 175.8 (**C**O (B)), 175.6 (**C**O (A)), 171.5 (**C**ON (A)), 170.3 (**C**ON (B)), 70.9 ($\mathbf{C}_{j\text{-}o}$ (A)), 70.9 ($\mathbf{C}_{j\text{-}o}$ (B)), 70.7 ($\mathbf{C}_{j\text{-}o}$ (B)), 70.7 ($\mathbf{C}_{j\text{-}o}$ (B)), 70.7 ($\mathbf{C}_{j\text{-}o}$ (A)), 70.6 ($\mathbf{C}_{j\text{-}o}$ (A)), 70.5 ($\mathbf{C}_{j\text{-}o}$ (A)), 70.1 ($\mathbf{C}_{j\text{-}o}$ (B)), 70.0 ($\mathbf{C}_{j\text{-}o}$ (A)), 69.2 ($\mathbf{C}_{j\text{-}o}$ (B)), 68.9 ($\mathbf{C}_{j\text{-}o}$ (A)), 50.6 ($\mathbf{C}_{p}$ (A)), 50.2 (**C**H_2 (A)), 48.5 (**C**H_2 (A)), 48.3 (**C**H_2 (B)), 41.8 (**C**H_2 (B)), 41.3 (**C**H_2 (B)), 38.8 ($\mathbf{C}_{f}$ (B)), 37.2 (**C**H_2 (A)), 35.9 ((**C**H_2 (A))), 35.5 (**C**H_2 (B)), 32.4 (**C**H_2 (B)), 32.0 (**C**H_2 (B)), 31.8 ((**C**H_2 (A))), 28.1 (**C**H_2 (B)), 27.4 (**C**H_2 (B)), 26.9 (**C**H_2 (A)), 26.8 (**C**H_2 (A)), 25.8 (**C**H_2 (B)), 25.6 (**C**H_2 (B)), 24.5 (**C**H_2 (B)), 24.1 (**C**H_2 (A)), 23.5 (**C**H_2 (B)), 22.0 (**C**H_2 (A)), 21.8 (**C**H_2 (B)). HRMS (ESI^+): Calcd. For $C_{18}H_{33}N_5NaO_5$, 422.2374; Found $[M+Na]^+$, 422.2373.

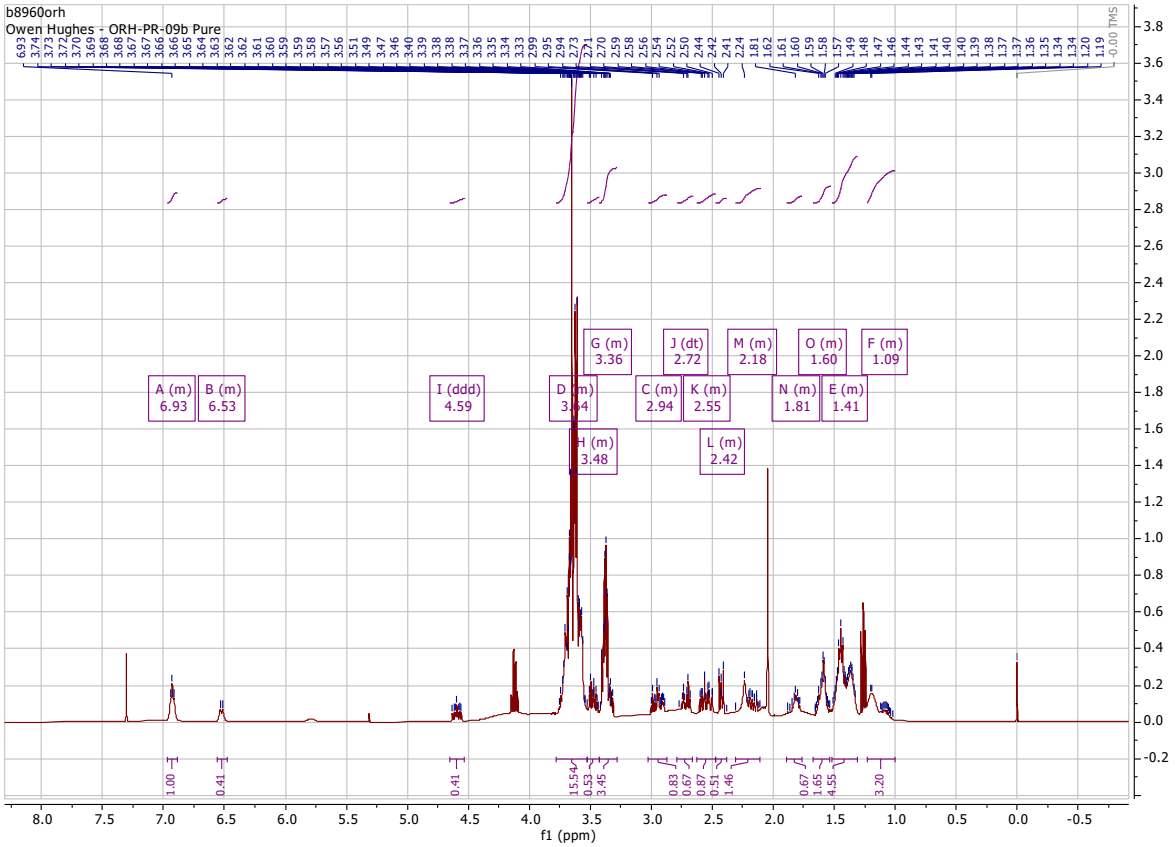


Figure 7.21: Figure 6.19: ^{1}H NMR of purified fractions containing **4.34**.

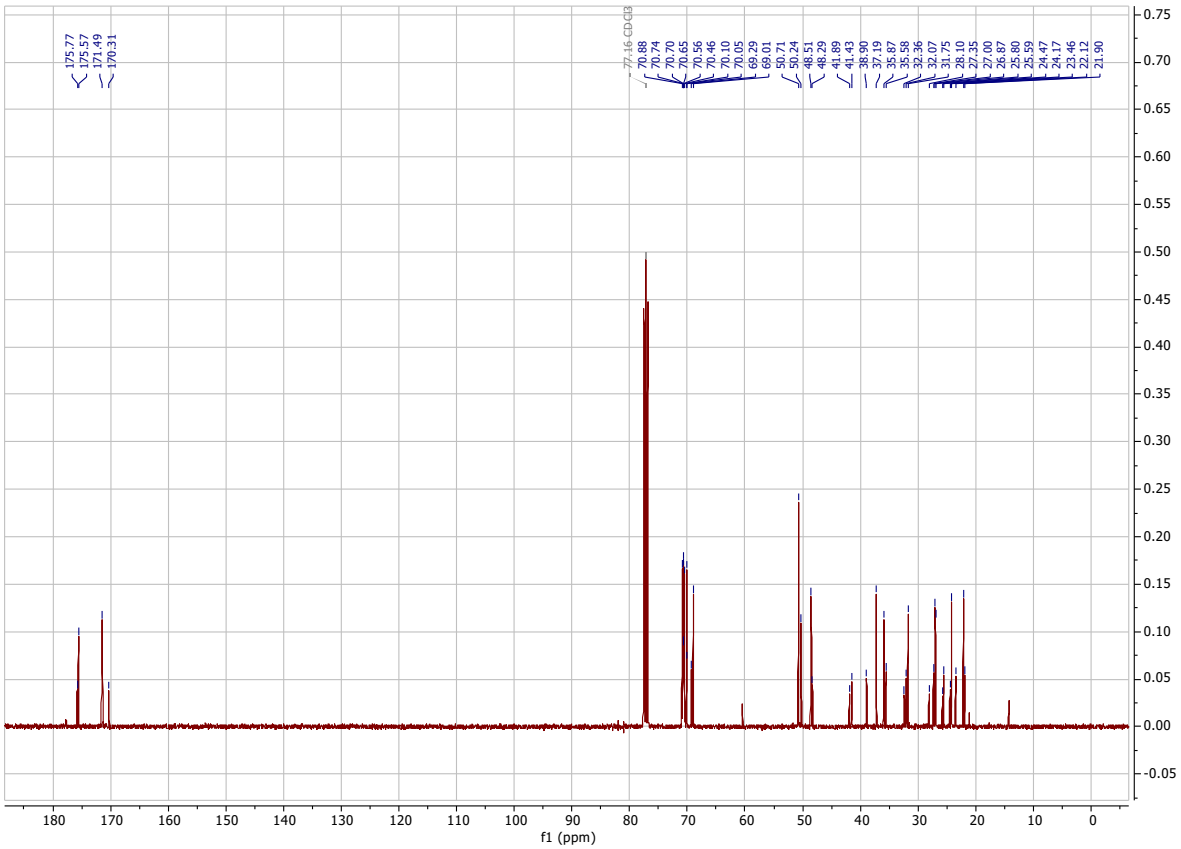
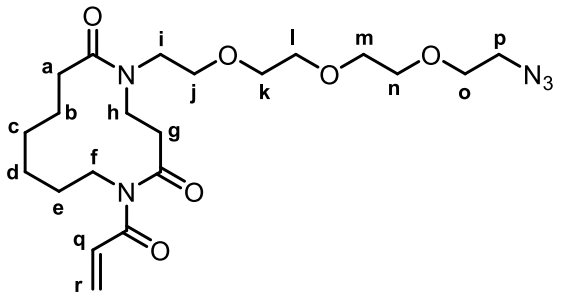


Figure 7.22: ^{13}C NMR of purified fractions containing **4.34**.

1-acryloyl-5-(2-(2-(2-(2-azidoethoxy)ethoxy)ethoxy)ethyl)-1,5-diazacyclododecane-2,6-dione (4.37)



To a cooled solution of 5-(2-(2-(2-(2-azidoethoxy) ethoxy) ethoxy) ethyl)-1,5-diazacyclododecane-2,6-dione (**4.36**) (100 mg, 0.25 mmol, 1.00 eq.) in THF (1.50 mL) was added MeMgBr (3.0 M in Et_2O, 0.09 mL, 0.275 mmol, 1.10 eq.) dropwise. The reaction mixture was stirred at 0 ºC for 30 minutes before the addition of acryloyl chloride (0.03 mL, 0.375 mmol, 1.50 eq.) as a single portion. The solution was then allowed to warm to room temperature and stirred for 2 hours. The reaction mixture was concentrated *in vacuo*, re-dissolved in DCM (5.00 mL), and the solid filtered off; the filtrate was purified by FCC (19:1; ethyl acetate: methanol) to yield the *title compound* as a colourless oil (12.0 mg, 11%). In $CDCl_3$ the compound exists as a 2:1 (A:B) mixture of rotamers as determined by NMR; R_f = 0.41 (9:1; ethyl acetate: methanol); ν_{max}/cm^{-1} (thin film) 3415, 2924, 2864, 2104, 1683, 1636, 1454, 1409, 1360, 1282, 1207, 1144, 1120, 796; δ_H (400 MHz, $CDCl_3$) 6.63 (dd, *J* = 16.8, 10.3 Hz, 1H, **H$_q$** (A)), 6.62 (dd, *J* = 16.8, 10.2 Hz, 1H, **H$_q$** (B)), 6.48 (dd, *J* = 16.7, 1.9 Hz, 1H, **H$_r$** (B)), 6.42 (dd, *J* = 16.9, 1.8 Hz, 1H, **H$_r$** (A)), 5.86 (dd, *J* = 10.2, 1.9 Hz, 1H, **H$_r$**, (B)), 5.80 (dd, *J* = 10.3, 1.8 Hz, 1H, **H$_r$**, (A)), 4.51 – 4.42 (m, 1H, **H$_j$** (A)), 4.17 – 4.06 (m, 1H, C**H** (A)), 4.04 – 3.84 (m, 2H, 2 x C**H** (A)), 3.83 – 3.69 (m, 1H, C**H** (A)), 3.69 – 3.51 (m, 12H, 6 x C**H$_2$** (A)), 3.50 – 3.40 (m, 1H, C**H** (A)), 3.39 (t, *J* = 5.1 Hz, 2H, **H$_p$** (A)), 3.14 – 2.99 (m, 1H, **H$_j$**, (A)), 2.81 – 2.69 (m, 1H, **H$_i$**, (A)), 2.54 (ddd, *J* = 14.8, 9.6, 2.4 Hz, 1H, **H$_a$** (A)), 2.35 – 2.31 (m, 2H, C**H$_2$** (B)), 2.10 (ddd, *J* = 14.8, 8.6, 2.7 Hz, 1H, **H$_a$**, (A)), 1.97 – 1.93 (m, 2H, **H$_b$** (A), C**H** (A)), 1.86 – 1.76 (m, 1H, C**H** (B)), 1.73 – 1.65 (m, 1H, C**H** (B)), 1.64 – 1.57 (m, 1H, **H$_b$**, (A)), 1.51 – 1.44 (m, 1H, C**H** (A)), 1.42 – 1.34 (m, 2H, C**H$_2$** (A)), 1.30 – 1.20 (m, 2H, C**H$_2$** (B)), 1.21 – 1.00 (m, 1H, C**H** (A)), 0.91 – 0.74 (m, 1H, C**H** (A)). The 13C NMR spectrum was weak, presumably as the sample was too dilute. Diagnostic ^{13}C NMR signals for the major rotamer (A) can be found at: δ_C (101 MHz, $CDCl_3$) 176.7 (**C**O), 175.1 (**C**O), 175.0 (**C**O), 130.3 (**C$_q$**), 130.0 (**C$_r$**), 70.8 (**C**H$_2$), 70.7 (**C**H$_2$), 70.7 (**C**H$_2$), 70.6 (**C**H$_2$), 70.0 (**C**H$_2$), 69.2 (**C**H$_2$), 50.7 (**C$_p$**),

43.9 (**C**H_2), 42.6 (**C**$_j$), 37.7 (**C**H_2), 37.0 (**C**$_i$), 32.7 (**C**$_a$), 27.7 (**C**$_b$), 27.0 (**C**H_2), 24.1 (**C**H_2), 22.6 (**C**H_2). HRMS (ESI$^+$): Calcd. For $C_{21}H_{35}N_5NaO_6$, 476.2480; Found $[M+Na]^+$, 476.2479.

Alternative method – acylation with DIPEA: A stirring solution of 5-(2-(2-(2-(2-azidoethoxy) ethoxy) ethoxy) ethyl)-1,5-diazacyclododecane-2,6-dione (**4.36**) (100 mg, 0.25 mmol, 1.00 eq.) and DIPEA (0.110 mL, 0.630 mmol, 2.50 eq.) in DCM (1.00 mL) was cooled to 0 ºC. A 0 ºC cooled solution of acryloyl chloride (0.030 mL, 0.375 mmol, 1.50 eq.) in DCM (1.00 mL) was added dropwise to the reaction mixture. The solution was stirred under an argon atmosphere for 2 h at 0 ºC before being warmed to rt and stirred for a further 17 h. The reaction mixture was then quenched with sat. aq. NH_4Cl (5.00 mL), extracted with Et_2O (2 x 5.00 mL), and the organic extracts washed with sat. aq. $NaHCO_3$ (2 x 5.00 mL). The combined organic mixture was dried over $MgSO_4$ and concentrated *in vacuo* to give the crude material as a yellow oil. Purification by FCC (SiO_2, 19:1; ethyl acetate: methanol) gave the *title compound* as a colourless oil (11.7 mg, 11%).

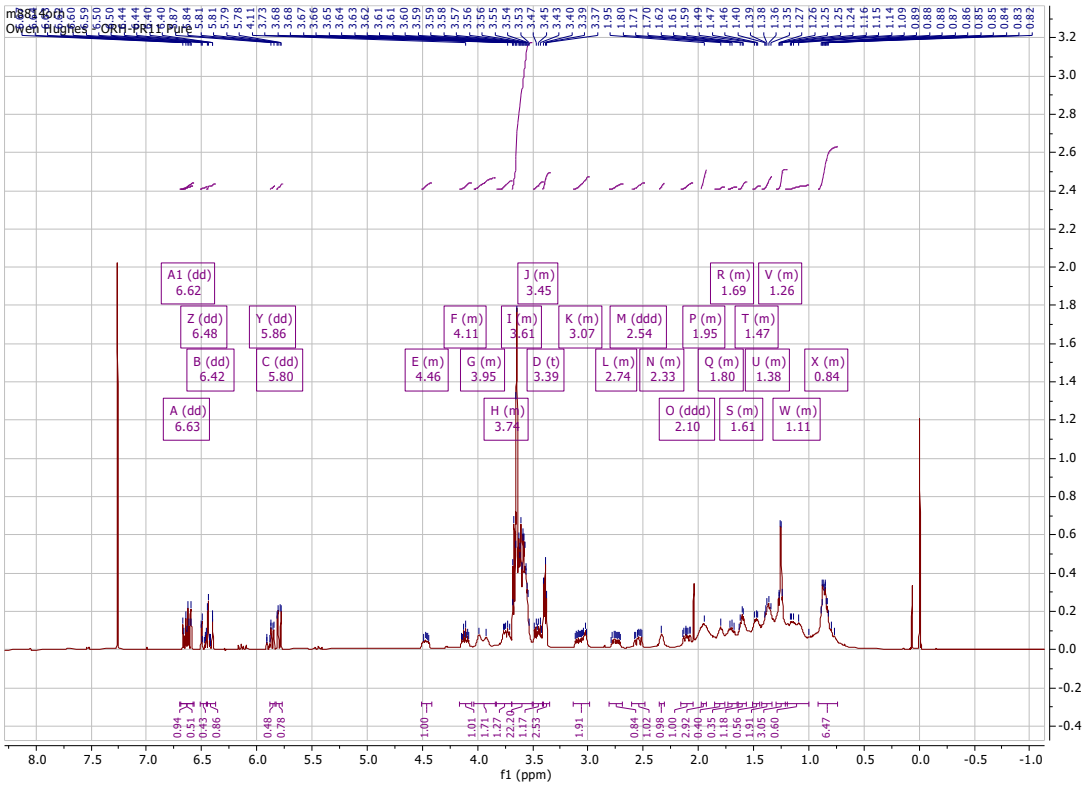


Figure 7.23: ^{1}H NMR spectrum of purified fractions containing **4.37**.

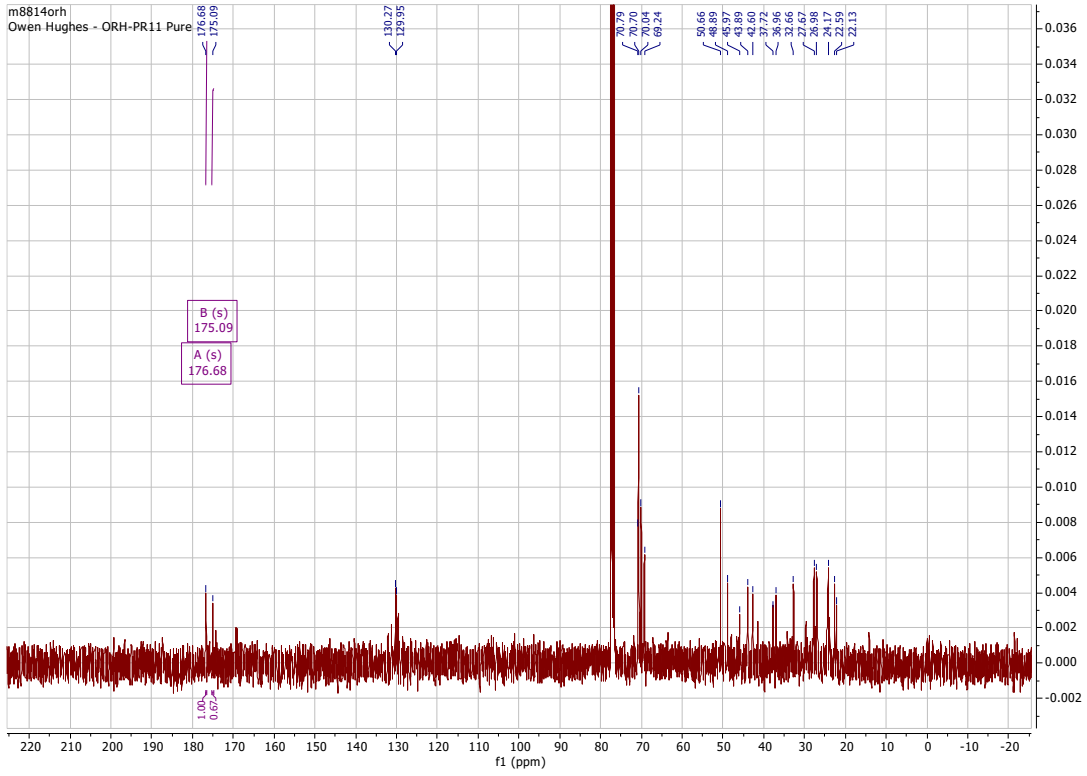
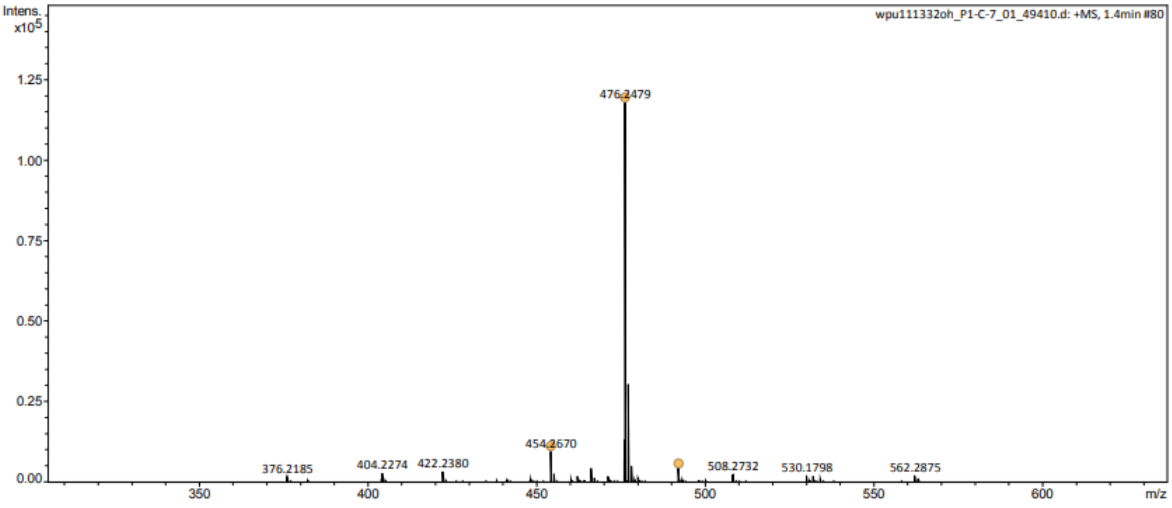


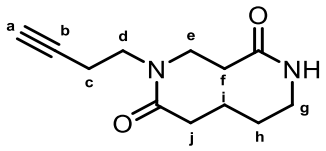
Figure 7.24: ^{13}C NMR spectrum of purified fractions containing **4.37**.



Meas. m/z	#	Ion Formula	m/z	err [ppm]	err [mDa]	mSigma	Mean err [ppm]
454.2670	1	C21H36N5O6	454.2660	-2.2	-1.0	12.6	-0.7
476.2479	1	C21H35N5NaO6	476.2480	0.1	0.0	3.5	-0.0
492.2245	1	C21H35KN5O6	492.2219	-5.2	-2.6	20.0	-4.2

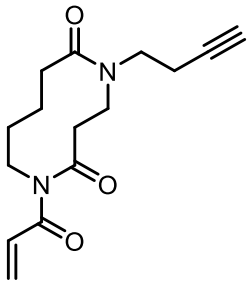
Figure 7.25: HRMS of purified fractions containing **4.34**.

5-(But-3-yn-1-yl)-1,5-diazecane-2,6-dione (4.40)



To a flask charged with 1-acryloyl-piperidin-2-one **2.01** (200 mg, 1.60 mmol, 1.00 eq.) in MeOH (3.00 mL) was added 1-amino-3-butyne (0.16 mL, 1.92 mmol, 1.20 eq.). The reaction mixture was stirred for 18 h before being concentrated *in vacuo* and purified by FCC (19:1; ethyl acetate: methanol). The *title compound* was obtained as a waxy white solid (272 mg, 77%). R_f = 0.18 (9:1; ethyl acetate: methanol); δ_H (400 MHz, $CDCl_3$) 5.76 (s, 1H, N**H**), 4.02 (t, *J* = 14.1 Hz, 1H, $\mathbf{H_e}$), 3.83 (q, *J* = 11.7 Hz, 1H, $\mathbf{H_g}$), 3.71 – 3.60 (m, 1H, $\mathbf{H_d}$), 3.51 – 3.41 (m, 1H, $\mathbf{H_{d'}}$), 3.41 – 3.29 (m, 1H, $\mathbf{H_{e'}}$), 2.96 – 2.89 (m, 1H, $\mathbf{H_{g'}}$), 2.86 – 2.75 (m, 1H, $\mathbf{H_c}$), 2.67 – 2.55 (m, 1H, C**H**H'), 2.54 – 2.41 (m, 2H, $\mathbf{H_{c'}}$ & $\mathbf{H_f}$), 2.38 – 2.28 (m, 1H, $\mathbf{H_{f'}}$), 2.24 – 2.03 (m, 2H, $\mathbf{H_a}$ & CH**H'**), 1.75 – 1.54 (m, 3H, $\mathbf{H_h}$ & $\mathbf{CH_2}$), 1.50 – 1.40 (m, 1H, $\mathbf{H_{h'}}$); δ_C (101 MHz, $CDCl_3$) 174.6 (**C**O), 171.1 (**C**O), 83.1 ($\mathbf{C_b}$), 70.7 ($\mathbf{C_a}$), 47.5 ($\mathbf{C_e}$), 47.5 ($\mathbf{C_d}$), 39.3 ($\mathbf{C_g}$), 38.5 ($\mathbf{C_f}$), 28.8 (**C**H_2), 26.3 ($\mathbf{C_h}$), 24.2 (**C**H_2), 18.1 ($\mathbf{C_c}$). HRMS (ESI^+): Calcd. For $C_{12}H_{19}N_2O_2$, 223.1368; Found $[M+Na]^+$, 223.1365.

1-Acryloyl-5-(but-3-yn-1-yl)-1,5-diazecane-2,6-dione (4.41)



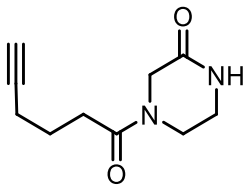
To a cooled solution of 5-(but-3-yn-1-yl)-1,5-diazecane-2,6-dione (**4.40**) (50.0 mg, 0.23 mmol, 1.00 eq.) in THF (1.00 mL) was added MeMgBr (3.0 M in Et_2O, 0.08 mL, 0.25 mmol, 1.10 eq.) dropwise. The reaction mixture was stirred at 0 ºC for 30 minutes before the addition of acryloyl chloride (0.027 mL, 0.34 mmol, 1.50 eq.) as a single portion. The solution was then allowed to warm to room temperature and stirred for 2 hours. The reaction was then quenched with sat. aq. NH_4Cl (2.00 mL), the aqueous mixture extracted with Et_2O (2.00 mL) and the organic extracts washed with sat. aq. $NaHCO_3$ (2 x 2.00 mL). The organic phase was dried over $MgSO_4$ and concentrated *in vacuo*. The crude material was purified by FCC (SiO_2, 1:1; hexane: ethyl acetate) to afford the *title compound* as a colourless film (4.45 mg, 7%). R_f =

0.08 (1:1; hexane: ethyl acetate); δ_H (400 MHz, $CDCl_3$) 6.67 (dd, *J* = 16.7, 10.2 Hz, 1H, NCOC**H**CHH'), 6.50 (dd, *J* = 16.7, 1.1 Hz, 1H, NCOCHC**H**H'), 5.92 (dd, *J* = 10.2, 1.1 Hz, 1H, NCOCHCH**H'**), 4.03 – 3.65 (m, 4H, 2 x C**H₂**), 3.62 – 3.16 (m, 4H, 2 x C**H₂**), 2.50 (td, *J* = 6.8, 2.6 Hz, 2H, C**H₂**), 2.46 – 2.26 (m, 2H, C**H₂**), 1.96 (t, *J* = 2.8 Hz, 1H, CH_2CH_2CC**H**), 1.94 – 1.84 (m, 2H, C**H₂**), 1.70 – 1.60 (m, 2H, C**H₂**); δ_C (101 MHz, $CDCl_3$) 172.6 (**C**O), 168.8 (**C**O), 132.1 (NCO**C**HCHH'), 129.3 (NCOCH**C**HH'), 82.0 (CH_2CH_2**C**CH), 69.6 (CH_2CH_2**C**CH), 46.4 (**C**H_2), 45.7 (2 x **C**H_2), 45.0 (**C**H_2), 38.8 (**C**H_2), 29.2 (**C**H_2), 24.1 (**C**H_2), 17.4 (**C**H_2). HRMS (ESI^+): Calcd. For $C_{15}H_{20}N_2NaO_3$, 299.1366; Found $[M+Na]^+$, 299.1366.

Note: A **C**O signal was missing from the ^{13}C NMR spectrum, presumably as the sample was too dilute and only sparingly soluble in $CDCl_3$. This made both the ^1H and ^{13}C NMR spectra poorly resolved, and difficult to assign with confidence.

Alternative method – acylation with DIPEA: A stirring solution of 5-(but-3-yn-1-yl)-1,5-diazecane-2,6-dione (**4.40**) (50.0 mg, 0.23 mmol, 1.00 eq.) and DIPEA (0.10 mL, 0.580 mmol, 2.50 eq.) in THF (1.00 mL) was cooled to 0 ºC. A 0 ºC cooled solution of acryloyl chloride (0.027 mL, 0.34 mmol, 1.50 eq.) in THF (0.25 mL) was added dropwise to the reaction mixture. The solution was stirred under an argon atmosphere for 2 h at 0 ºC before being warmed to rt and stirred for a further 17 h. The reaction mixture was then quenched with sat. aq. NH_4Cl (2.00 mL), extracted with Et_2O (2 x 5.00 mL), and the organic extracts washed with sat. aq. $NaHCO_3$ (2 x 5.00 mL). The combined organic mixture was dried over $MgSO_4$ and concentrated *in vacuo* to give the crude material as a yellow oil. Purification by FCC (SiO_2, 19:1; ethyl acetate: methanol) gave the *title compound* as a colourless film (5.70 mg, 9%).

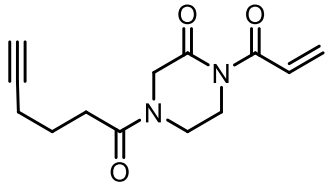
4-(Hex-5-ynoyl)piperazin-2-one (4.44)



A solution of 5-hexynoic acid (**4.43**) (0.12 mL, 1.09 mmol, 1.10 eq.), propylphosphonic anhydride (T3P, 0.95 mL, 1.49 mmol, 1.50 eq.) and *N*,*N'*-diisopropylethylamine (0.52 mL, 3.00 mmol, 3.00 eq.) in anhydrous DCM (18.0 mL) was stirred at rt for 10 minutes before the addition of piperazin-2-one (**4.42**) (100 mg, 0.990 mmol, 1.00 eq.). The reaction mixture was stirred under these conditions for 18 h before being diluted with DCM (20.0 mL) and washed with sat. aq. $NaHCO_3$ (2 x 20.0 mL) and brine (2 x 20.0 mL). The organic extracts were dried

over $MgSO_4$ and concentrated *in vacuo* to give the crude product which was purified by FCC (SiO_2, 19:1; ethyl acetate: methanol) to afford the *title compound* as a colourless oil (41.9 mg, 22%). In solution in $CDCl_3$, the compound exists as a mixture of two rotamers in an approximately 5:4 (A:B) ratio; R_f = 0.17 (19:1; ethyl acetate: methanol); ν_{max}/cm^{-1} (thin film) 3250, 2923, 2114, 1671, 1633, 1492, 1435, 1338, 1237, 1110, 982, 653, 536, 482; δ_H (400 MHz, $CDCl_3$) 7.59 (s, 1H, N**H** (B)), 7.30 (s, 1H, N**H** (A)), 4.21 (s, 2H, C**H₂**CONH (B)), 4.13 (s, 2H, C**H₂**CONH (A)), 3.79 (t, *J* = 5.5 Hz, 2H, CONHCH₂C**H₂** (A)), 3.68 (t, *J* = 5.3 Hz, 2H, CONHCH₂C**H₂** (B)), 3.44 – 3.39 (m, 2H, CONHC**H₂**CH₂ (B)), 3.38 – 3.29 (m, 2H, CONHC**H₂**CH₂ (A)), 2.53 – 2.40 (m, 4H, NCOC**H₂** (A,B [overlapping]), 2.32 – 2.21 (m, 4H, CH₂C**H₂**CCH (A,B [overlapping]), 1.99 – 1.95 (m, 2H, CH₂CH₂CC**H** (A,B [overlapping]), 1.91 – 1.77 (m, 4H, C**H₂**CH₂CCH (A,B [overlapping]); δ_C (101 MHz, $CDCl_3$) 171.0 (**C**O (A)), 170.8 (**C**O (B)), 168.6 (**C**O (B)), 167.3 (**C**O (A)), 83.6 (CH₂CH₂**C**CH (B)), 83.5 (CH₂CH₂**C**CH (A)), 69.5 (CH₂CH₂C**C**H (A)), 69.4 (CH₂CH₂C**C**H (B)), 48.7 (**C**H₂CONH (B)), 46.0 (**C**H₂CONH (A)), 42.3 (CONHCH₂**C**H₂ (B)), 41.2 (CONH**C**H₂CH₂ (B)), 40.8 (CONH**C**H₂CH₂ (A)), 38.4 (CONHCH₂**C**H₂ (A), 31.6 (NCO**C**H₂ (A)), 31.5 (NCO**C**H₂ (B)), 23.5 (**C**H₂CH₂CCH (B)), 23.4 (**C**H₂CH₂CCH (A)), 17.9 (CH₂**C**H₂CCH (A,B [overlapping])); HRMS (ESI^+): calc. for $C_{10}H_{15}N_2NaO_2$, 217.0947. Found $[M+Na]^+$, 217.0948.

1-Acryloyl-4-(hex-5-ynoyl)piperazine-2-one (4.45)



A solution of 4-(hex-5-ynol)piperazin-2-one (**4.44**) (126 mg, 0.510 mmol, 1.00 eq.) in THF (2.50 mL) was cooled to 0 ºC before the addition of *N*, *N'*-diisopropylethylamine (0.220 mL, 1.27 mmol, 2.50 eq.). The mixture was stirred at this temperature for 10 minutes before the dropwise addition of acryloyl chloride (0.060 mL, 0.770 mmol, 1.50 eq.) in THF (0.50 mL). Upon complete addition, the reaction mixture was warmed to room temperature and stirring continued, monitored every hour by TLC until complete conversion of starting material was observed (~4 hours). The reaction mixture was then quenched with sat. aq. NH_4Cl (10.0 mL) and extracted with Et_2O (2 x 10.0 mL). The organic extracts were then washed with sat. aq. $NaHCO_3$ (2 x 10.0 mL) and brine (2 x 10.0 mL), dried over $MgSO_4$ and concentrated *in vacuo* to give the crude product. This was purified by FCC (SiO_2, 2:1; hexane: ethyl acetate → 1:1;

hexane: ethyl acetate) to afford the *title compound* as a golden oil (21.3 mg, 17%). In solution in $CDCl_3$, the compound exists as a mixture of two rotamers in an approximately 4:3 (A:B) ratio; R_f = 0.19 (1:1; hexane: ethyl acetate); ν_{max}/cm^{-1} (thin film) 3280, 2933, 2115, 1686, 1647, 1434, 1407, 1390, 1366, 1307, 1191, 1111, 1028, 978, 965, 796, 641; δ_H (400 MHz, $CDCl_3$) 7.23 – 7.09 (m, 2H, COC**H**=CHH' (A, B [overlapping])), 6.50 – 6.40 (m, 2H, COCH=C**H**H' (A, B [overlapping])), 5.84 (d, *J* = 10.4 Hz, 1H, COCH=CH**H'** (A)), 5.83 (d, *J* = 10.4 Hz, 1H, COCHCH**H'** (B)), 4.40 (s, 2H, CONC**H₂**CO (A)), 4.32 (s, 2H, CONC**H₂**CO (B)), 4.04 – 3.96 (m, 2H, CONC**H₂**CH_2 (A)), 3.94 – 3.87 (m, 2H, CONC**H₂**CH_2 (B)), 3.84 – 3.77 (m, 2H, $CONCH_2C$**H₂** (B)), 3.77 – 3.71 (m, 2H, $CONCH_2C$**H₂** (A)), 2.54 – 2.43 (m, 4H, NCOC**H₂** (A, B [overlapping])), 2.30 (td, *J* = 6.7, 2.7 Hz, 4H, CH_2C**H₂**CCH (A, B [overlapping])), 2.04 – 1.94 (m, 2H, CH_2CH_2CC**H** (A, B [overlapping])), 1.88 (app p, *J* = 6.7 Hz, 4H, C**H₂**CH_2CCH (A, B [overlapping])); δ_C (101 MHz, $CDCl_3$) 171.1 (**C**O), 170.8 (**C**O), 168.5 (**C**O), 167.8 (**C**O), 167.7 (**C**O), 167.6 (**C**O), 130.9 (CO**C**HCHH' (A, B [overlapping])), 130.8 (COCH**C**HH' (A, B [overlapping])), 83.6 (CH_2CH_2**C**CH (A)), 83.5 (CH_2CH_2**C**CH (B)), 69.6 (CH_2CH_2C**C**H (B)), 69.5 (CH_2CH_2C**C**H (A)), 50.3 (CON**C**H_2CO (B)), 47.6 (CON**C**H_2CO (A)), 42.9 ($CONCH_2$**C**H_2 (A)), 41.9 (CON**C**H_2CH_2 (A, B [overlapping])), 39.8 ($CONCH_2$**C**H_2 (B)), 31.7 (NCO**C**H_2 (B)), 31.5 (NCO**C**H_2 (A)), 23.3 (**C**H_2CH_2CCH (A, B [overlapping]), 17.9 (CH_2**C**H_2CCH (A, B [overlapping]). HRMS (ESI^+): calc. for $C_{13}H_{16}N_2NaO_3$, 271.1053. Found $[M+Na]^+$, 271.1053.

7.5 Experimental procedures and characterisation data for Chapter 5

7.5.1 Protein expression and sequences

Protein expression:

Expression of CjX183-D,[63] CjX183-D R51K,[63] JVZ-007,[181] and CCL5,[182] was undertaken as previously reported by our groups.

Protein sequences:

CjX183-D:

GYLVGDATRGANLWNTQTCVACHGVDGERNASGTPALTPLNPNRDLYRHSRDTQD
RALRD FISMWMPQGNEGSCTGQCAADIEAFIRTWHHHHHH

Theoretical mass (+ heme): 11226 Da

* Nb. In all experiments using CjX183-D and mutants, the observed mass was ~11 Da higher than the theoretical mass and thus was taken as the unmodified protein mass in all instances. The reasons for this observation are unknown but are consistent with previous reports.[63]

CjX183-D R51K:

GYLVGDATRGANLWNTQTCVACHGVDGERNASGTPALTPLNPNRDLYRHSKDTQD
RALRD FISMWMPQGNEGSCTGQCAADIEAFIRTWHHHHHH

Theoretical mass (+ heme): 11198 Da

* Nb. In all experiments using CjX183-D and mutants, the observed mass was ~11 Da higher than the theoretical mass and thus was taken as the unmodified protein mass in all instances. The reasons for this observation are unknown but are consistent with previous reports.[63]

RNase A:[183]

KETAAAKFERQHMDSSTSAASSSNYCNQMMKSRNLTKDRCKPVNTFVHESLADVQ
AVCSQKNVACKNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHIIVACE
GNPYVPVHFDASV

Theoretical mass: 13681 Da

Myoglobin:[184]

GLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMK
ASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEAIIHVL
HSRHPGDFGADAQGAMNKALELFRKDIAAKYKELGYQG

Theoretical mass:[185] 16951 Da

Cytochrome C:[186]

GDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFTYTDANKNK
GITWKEETLMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKKATNE

Theoretical mass: 12327 Da

JVZ-007:

SEVQLVESGGGLVQPGGSLTLSCAASRFMISEYSMHWVRQAPGKGLEWVSTINPAG
TTDYAESVKGRFTISRDNAKNTLYLQMNSLKPEDTAVYYCDGYGYRGQGTQVTVSS

Theoretical mass: 12121 Da

CCL5:

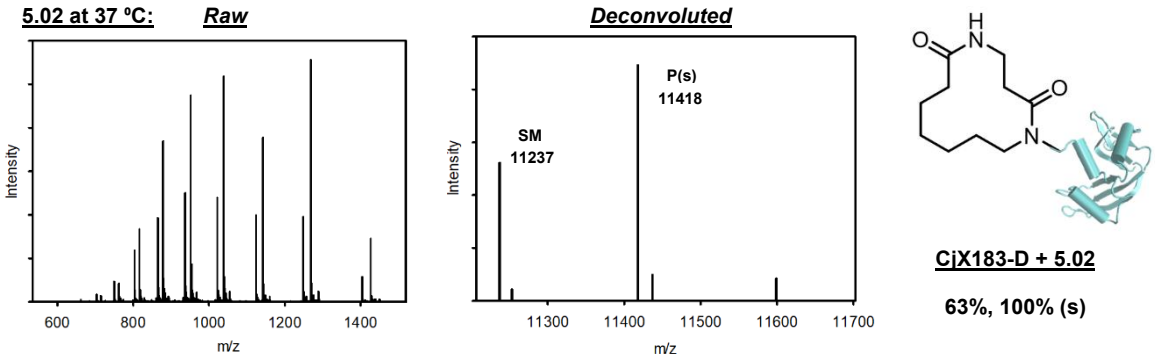
SPYSSDTTPCCFAYIARPLPRAHIKEYFYTSGKCSNPAVVFVTRKNRQVCANPEKKW
VREYINSLEMS

Theoretical mass: 7847 Da

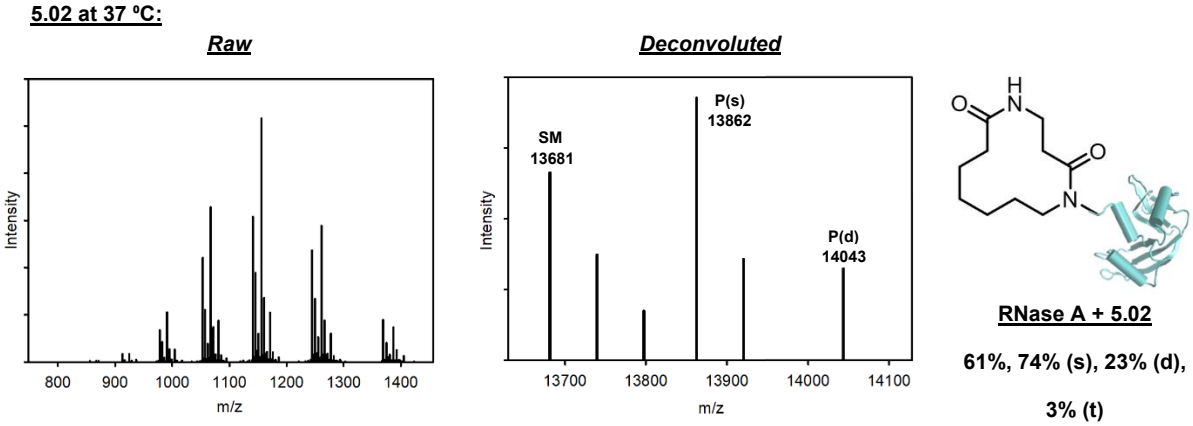
7.5.2 General procedures and materials for protein modification

7.5.2.1 Conversion experiments

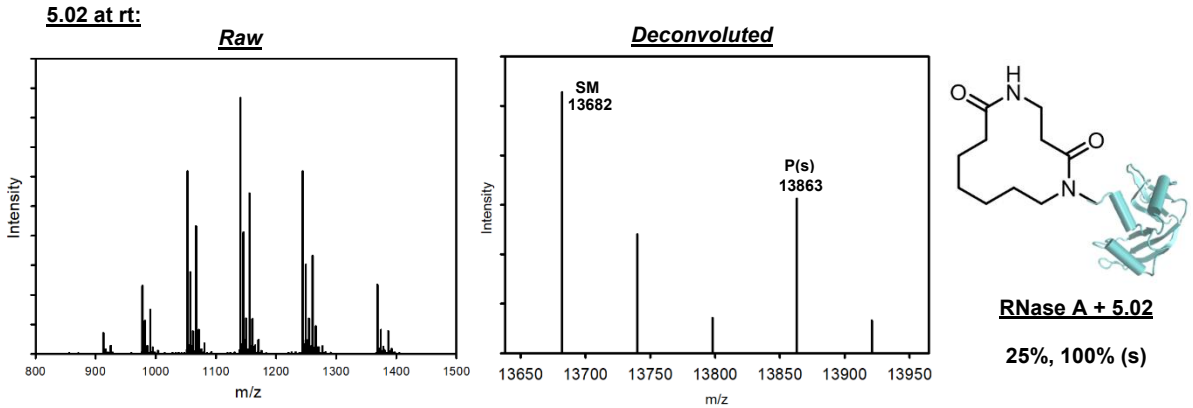
For the modification of CjX183-D, RNase A, myoglobin and cytochrome C: A stock solution of **5.02** (2.00 µL, 2.5 mM in DMSO, 5 nmol, 5 eq.) was added to a mixture containing an aliquot of a stock solution of protein (6.67 µL, 150 µM in 0.1 M, pH 7.4 sodium phosphate buffer, 1 nmol, 50 µM final concentration, 1 eq.), sodium phosphate buffer (4 µL, 0.5 M , pH 7.4) and HPLC-grade water (7.33 µL). The mixture was incubated at rt or 37 °C with agitation (1000 rpm) for 19 h before analysis by LC-MS in accordance with the procedure outlined in **LC-MS method C** (section 7.1.3).



MS (ESI$^+$): [**SM**+H]$^+$ found 11237, calculated 11226; [SM+H_2O+H]$^+$ found 11253, calculated 11244; [**P(s)**+H]$^+$ found 11418, calculated 11407; [**P(s)**+H_2O+H]$^+$ found 11437, calculated 11425.



MS (ESI$^+$): [**SM**+H]$^+$ found 13681, calculated 13681; [**SM**+58]$^+$ found 13739; [**SM**+116]$^+$ found 13798; [**P(s)**+H]$^+$ found 13862, calculated 13862; [**P(s)**+58]$^+$ found 13920; [**P(d)**+H]$^+$ found 14043, calculated 14043.



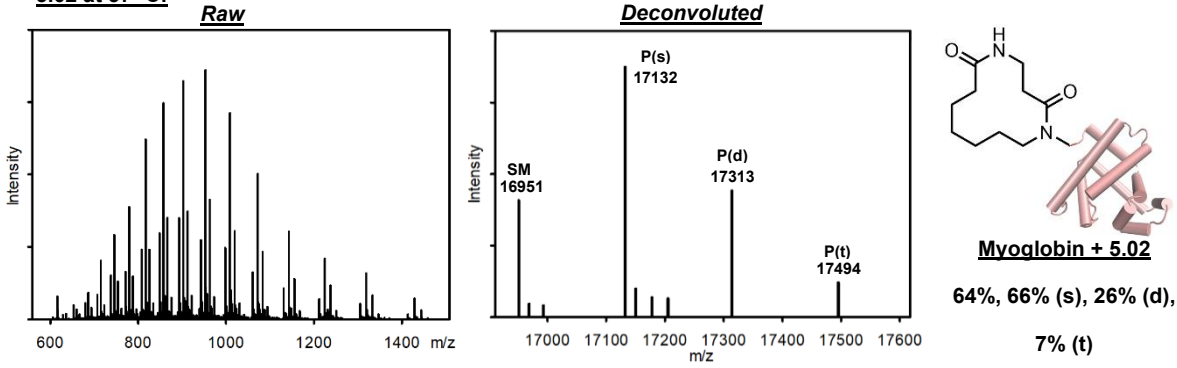
MS (ESI$^+$): [**SM**+H]$^+$ found 13682, calculated 13681; [**SM**+58]$^+$ found 13739; [**SM**+116]$^+$ found 13978; [**P(s)**+H]$^+$ found 13863, calculated 13862; [**P(s)**+58]$^+$ found 13920.

Table 7.1: Tabulated data for Figure 5.04 – Screening conditions for RNase A bioconjugation with imide **5.02.**

Entry	pH	Equivalents of **5.02**	Selectivity (%)	Conversion (%)	Symbol
1	6.0	5	0	0	
2	6.0	10	100	10	
3	6.0	20	100	28	
4	6.0	50	86	64	
5	6.5	5	100	11	
6	6.5	10	100	20	
7	6.5	20	100	38	
8	6.5	50	82	79	
9	7.4	5	100	25	
10	7.4	10	100	46	
11	7.4	20	81	79	
12	7.4	50	53	100	
13	8.0	5	100	39	
14	8.0	10	80	66	
15	8.0	50	47	93	

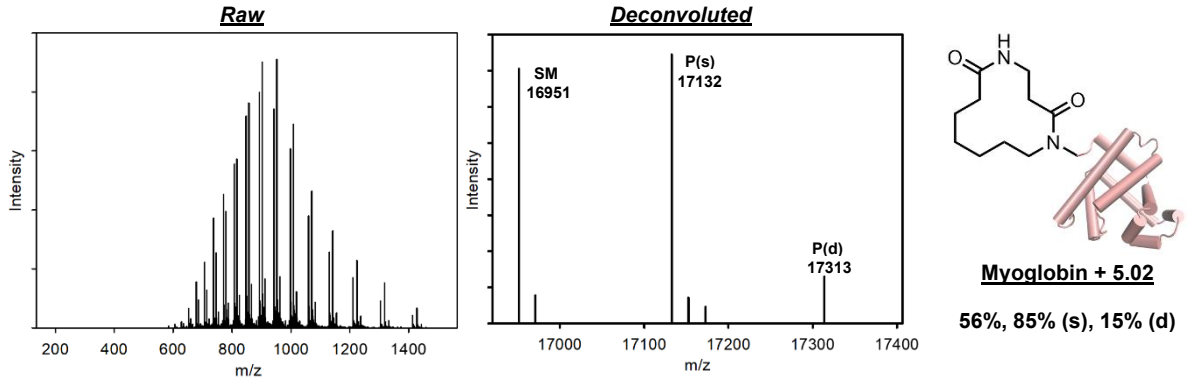
[a] Reactions were run at 50 µM RNase A at room temperature, 1000 rpm for 19 h before an aliquot was taken and analysed by LC-MS.

5.02 at 37 ºC:



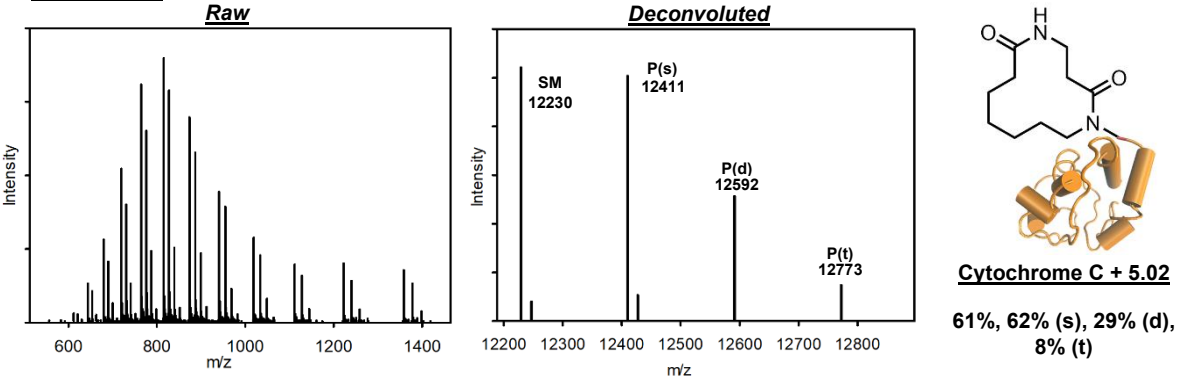
MS (ESI$^+$): [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16969, calculated 16969; [**SM**+MeCN+H]$^+$ found 16993, calculated 16993; [**P(s)**+H]$^+$ found 17132, calculated 17132; [**P(s)**+H_2O+H]$^+$ found 17150, calculated 17150; [**P(s)**+MeCN+H]$^+$ found 17177, calculated 17177; [**P(s)**+DMSO+H]$^+$ found 17205, calculated 17205; [**P(d)**+H]$^+$ found 17313, calculated 17313; [**P(t)**+H]$^+$ found 17494, calculated 17494.

5.02 at rt:

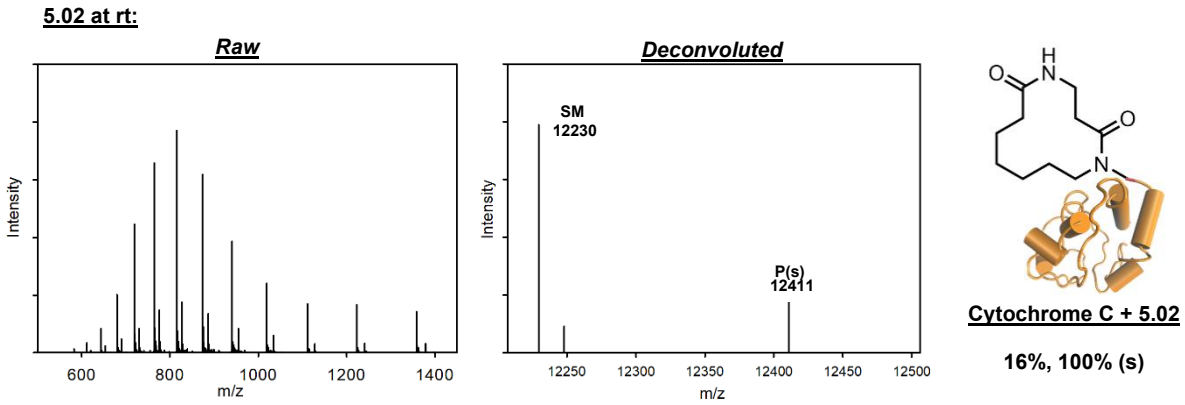


MS (ESI$^+$): [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16970, calculated 16969; [**P(s)**+H]$^+$ found 17132, calculated 17132; [**P(s)**+H_2O+H]$^+$ found 17152, calculated 17150; [**P(s)**+MeCN+H]$^+$ found 17173, calculated 17174; [**P(d)**+H]$^+$ found 17313, calculated 17313.

5.02 at 37 ºC:

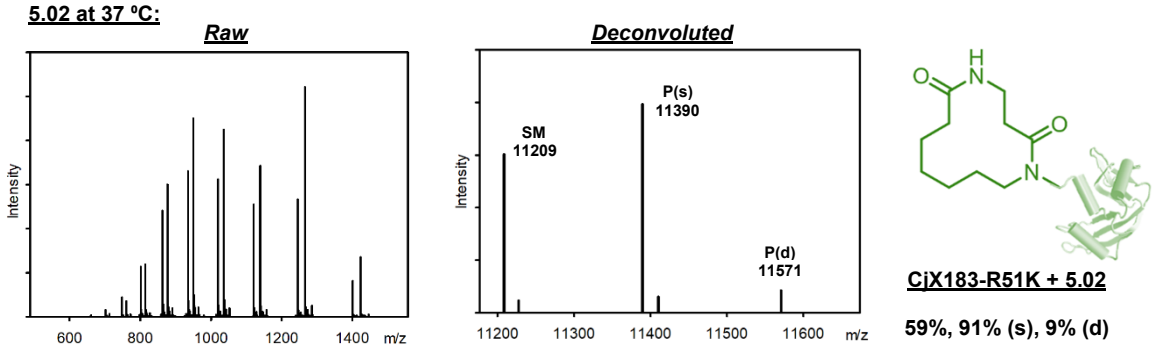


MS (ESI$^+$): [**SM**+H]$^+$ found 12230, calculated 12327; [**SM**+H_2O+H]$^+$ found 12247, calculated 12345; [**P(s)**+H]$^+$ found 12411, calculated 12508; [**P(s)**+H_2O+H]$^+$ found 12428, calculated 12526; [**P(d)**+H]$^+$ found 12592, calculated 12689; [**P(t)**+H]$^+$ found 12773, calculated 12870.

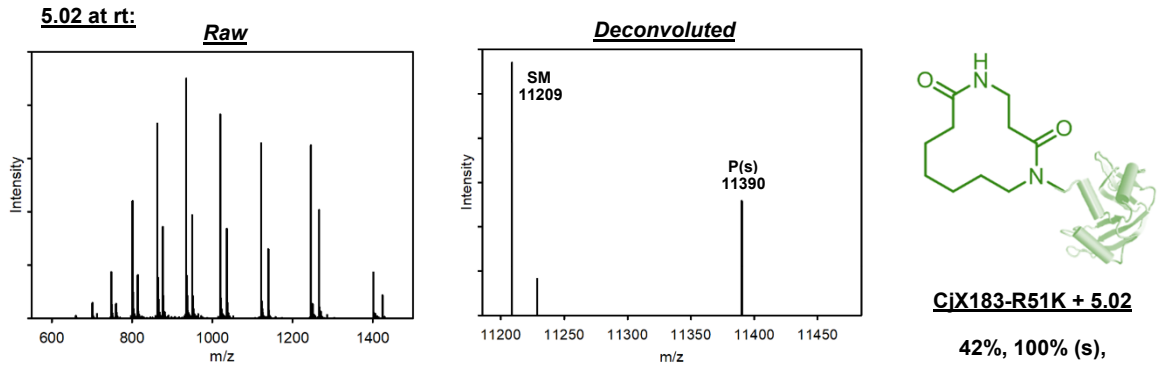


MS (ESI$^+$): [**SM**+H]$^+$ found 12230, calculated 12327; [**SM**+H_2O+H]$^+$ found 12247, calculated 12345; [**P(s)**+H]$^+$ found 12411, calculated 12508.

For the modification of CjX183-D R51K: A stock solution of **5.02** (2.00 µL, 2.5 mM in DMSO, 5 nmol, 5 eq.) was added to a mixture of a stock solution of CjX183-D R51K (9.17 µL, 109 µM in 0.1 M, pH 7.4 sodium phosphate buffer, 1 nmol, 50 µM final concentration, 1 eq.), sodium phosphate buffer (4 µL, 0.5 M stock, 0.1 M, pH 7.4) and HPLC-grade water (4.83 µL). The mixture was incubated at rt or 37 ºC with agitation (1000 rpm) for 19 h before analysis by LC-MS in accordance with the procedure outlined in **LC-MS method C** (section 7.1.3).



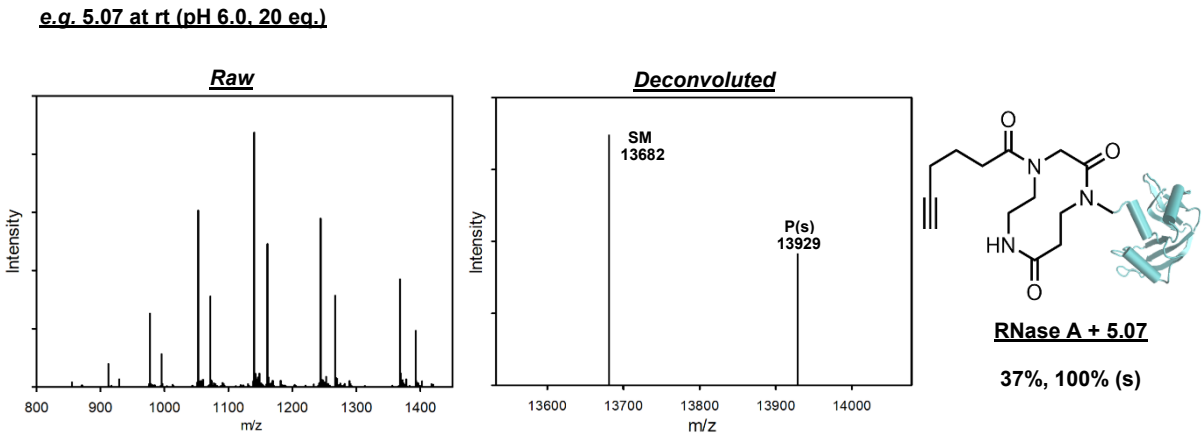
MS (ESI$^+$): [**SM**+H]$^+$ found 11209, calculated 11198; [**SM**+H_2O+H]$^+$ found 11228, calculated 11216; [**P(s)**+H]$^+$ found 11390, calculated 11379; [**P(s)**+H_2O+H]$^+$ found 11410, calculated 11397; [**P(d)**+H]$^+$ found 11571, calculated 11560.



MS (ESI$^+$): [**SM**+H]$^+$ found 11209, calculated 11198; [**SM**+H_2O+H]$^+$ found 11228, calculated 11216; [**P(s)**+H]$^+$ found 11390, calculated 11379.

7.5.2.2 Conditions screening experiments

For the modification of RNase A and cytochrome C: A stock solution of **5.02**, **5.06** or **5.07** (2.00 µL, 2.5, 5, 10 or 25 mM stock in DMSO; 5, 10, 20 or 50 nmol; 5, 10, 20 or 50 eq.) was added to a mixture containing an aliquot of a stock solution of protein (6.67 µL, 150 µM in 0.1 M, pH 7.4 sodium phosphate buffer, 1 nmol, 50 µM final concentration, 1 eq.), sodium phosphate buffer (4 µL, 0.5 M stock, 0.1 M, pH 6.0, 6.5, 7.4 or 8.0) and HPLC-grade water (7.33 µL). The mixture was incubated at rt with agitation (1000 rpm) for 19 h before analysis by LC-MS in accordance with the procedure outlined in **LC-MS method C** (section 7.1.3).



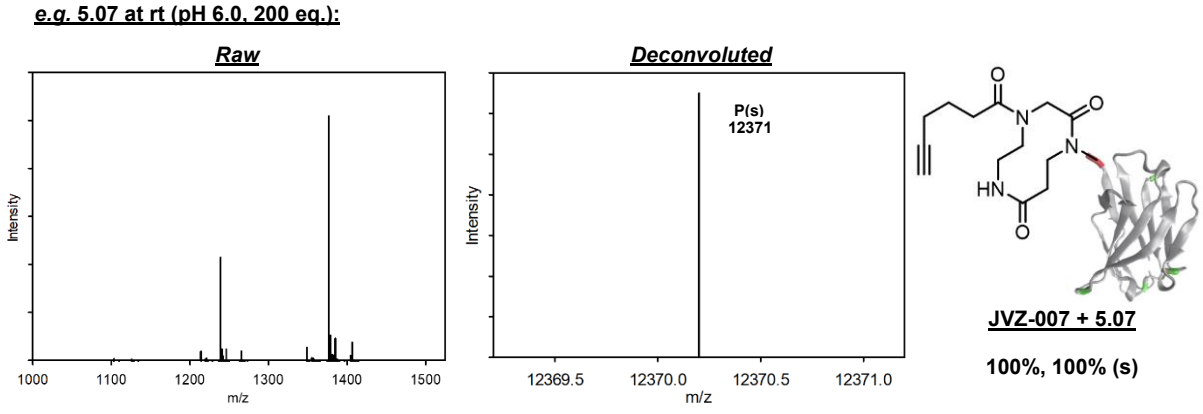
MS (ESI$^+$): [**SM**+H]$^+$ found 13682, calculated 13681; [**P(s)**+H]$^+$ found 13929, calculated 13929.

Table 7.2: Tabulated data for Figure 5.06 – Screening conditions for RNase A bioconjugation with imide **5.07.**

Entry	pH	Equivalents of **5.07**	Selectivity (%)	Conversion (%)	Symbol
1	6.0	5	100	10	
2	6.0	10	100	19	
3	6.0	20	100	34	
4	6.0	50	85	78	
5	6.5	5	100	15	
6	6.5	10	100	30	
7	6.5	50	79	91	
8	7.4	5	100	20	
9	7.4	10	100	31	
10	7.4	20	89	70	
11	7.4	50	78	100	
12	8.0	5	100	16	
13	8.0	10	100	27	
14	8.0	20	86	63	
15	8.0	50	73	89	

[a] Reactions were run at 50 µM RNase A at room temperature, 1000 rpm for 19 h before an aliquot was taken and analysed by LC-MS.

For the modification of JVZ-007: A stock solution of **5.07** (2.00 µL; 2.5, 5, 10, 25, 50, 100 mM stock in DMSO; 5, 10, 20, 50, 100 or 200 nmol; 5, 10, 20, 50, 100 or 200 eq.) was added to a mixture containing an aliquot of a stock solution of JVZ-007 (5.92 µL, 168.9 µM in 0.1 M, pH 7.4 sodium phosphate buffer, 1 nmol, 50 µM final concentration, 1 eq.), sodium phosphate buffer (4 µL, 0.5 M stock, 0.1 M, pH 6.0, 6.5, 7.4 or 8.0) and HPLC-grade water (8.08 µL). The mixture was incubated at rt with agitation (1000 rpm) for 19 h before analysis by LC-MS in accordance with the procedure outlined in **LC-MS method C** (section 7.1.3).



MS (ESI$^+$): [**P(s)**+H]$^+$ found 12371, calculated 12372.

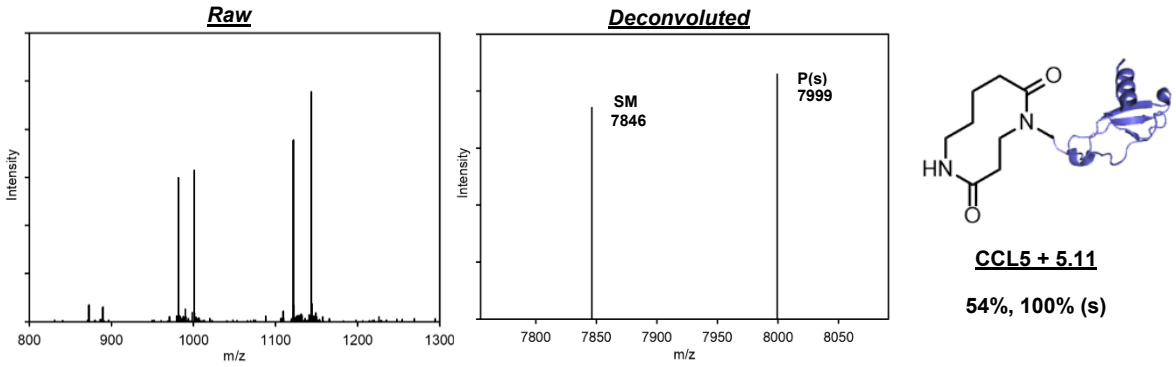
Table 7.3: Tabulated data for Figure 5.07 – Screening conditions for JVZ-007 bioconjugation with imide **5.07.**

Entry	pH	Equivalents of **5.07**	Selectivity (%)	Conversion (%)	Symbol
1	6.0	5	0	0	
2	6.0	10	0	0	
3	6.0	20	0	0	
4	6.0	50	100	34	
5	6.0	100	100	85	
6	6.0	200	100	100	
7	6.5	5	0	0	
8	6.5	10	0	0	
9	6.5	20	0	0	
10	6.5	50	100	65	
11	6.5	100	100	100	
12	6.5	200	100	100	
13	7.4	5	0	0	
14	7.4	10	0	0	
15	7.4	20	100	22	
16	7.4	50	100	74	
17	7.4	100	100	100	
18	7.4	200	100	100	
19	8.0	5	0	0	
20	8.0	10	0	0	
21	8.0	20	100	12	
22	8.0	50	100	47	
23	8.0	100	100	100	
24	8.0	200	100	100	

[a] Reactions were run at 50 μM JVZ-007 at room temperature, 1000 rpm for 19 h before an aliquot was taken and analysed by LC-MS.

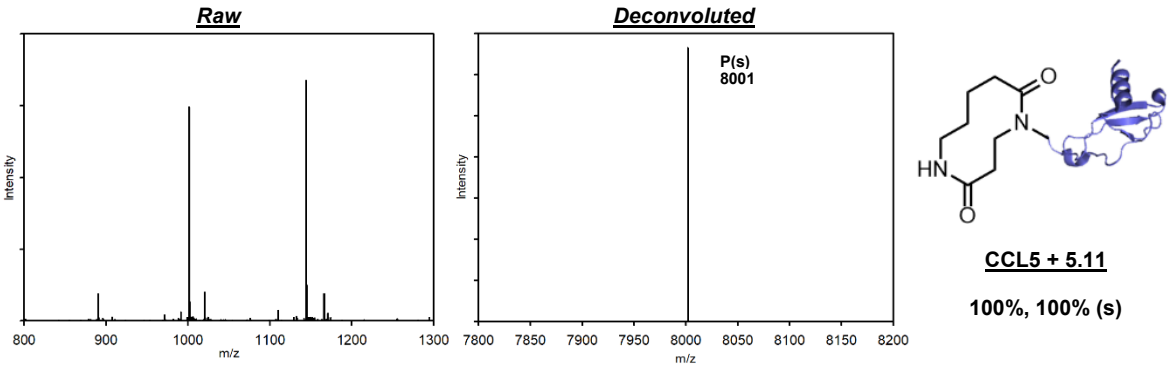
For the modification of CCL5: A stock solution of **5.11** (2.00 µL, 5, 10, 25 or 37.5 mM stock in DMSO; 10, 20, 50 or 75 nmol; 10, 20, 50 or 75 eq.) was added to a stock solution of CCL5 (18.0 µL, 55.6 µM in 0.1 M, pH 6.0, 6.5, 7.4 or 8.0 sodium phosphate buffer, 1 nmol, 50 µM final concentration, 1 eq.). The mixture was incubated at rt with agitation (1000 rpm) for 19 h before analysis by LC-MS in accordance with the procedure outlined in **LC-MS method C** (section 7.1.3).

5.11 at rt (pH 6.0, 50 eq.):



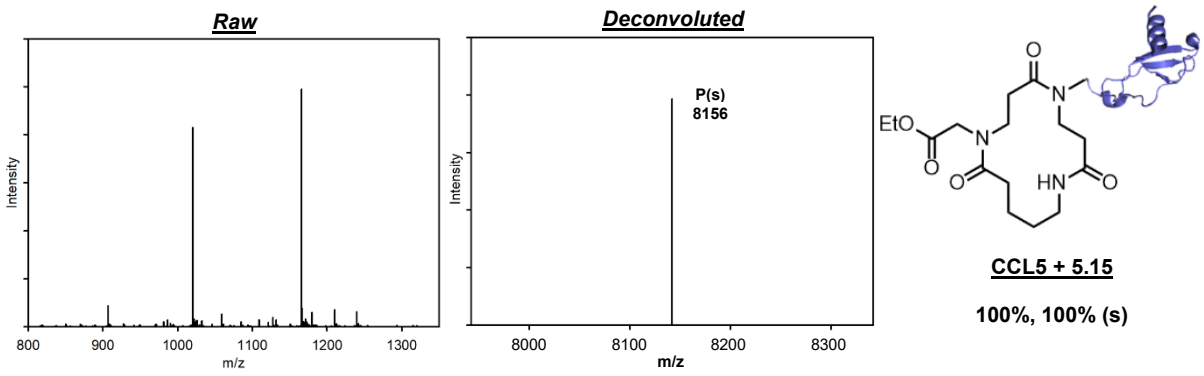
MS (ESI$^+$): [**SM**+H]$^+$ found 7846, calculated 7846; [**P(s)**+H]$^+$ found 7999, calculated 7999.

5.11 at rt (pH 6.5, 75 eq.):



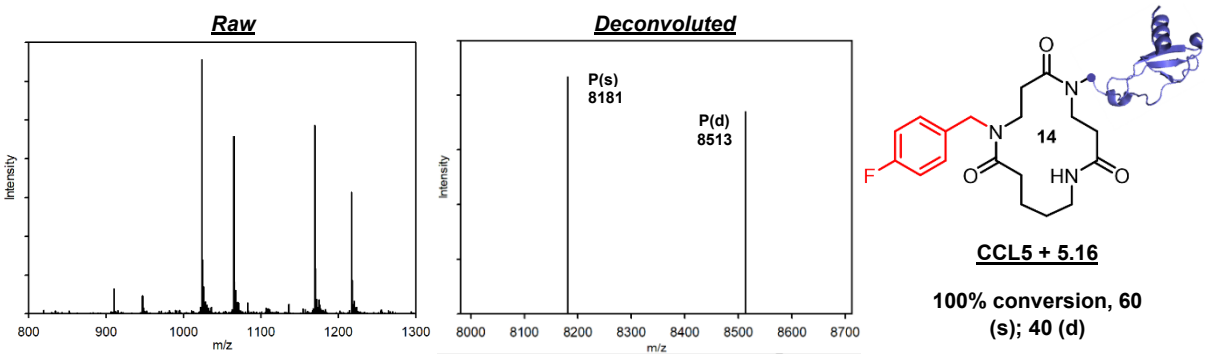
MS (ESI$^+$): [**P(s)**+H]$^+$ found 8002, calculated 7999.

5.15 at rt (pH 6.0, 50 eq.):



MS (ESI$^+$): [**P(s)**+H]$^+$ found 8156, calculated 8156.

5.16 at rt (pH 6.0, 50 eq.):



MS (ESI$^+$): [**P(s)**+H]$^+$ found 8181, calculated 8178; [P(d)+H]$^+$ found 8513, calculated 8510.

7.5.2.3 Protein sequencing by linear MALDI-MS

A 1 mL aliquot of 100 mM protein solution in aqueous 10 mM ammonium acetate was applied to a ground steel MALDI target plate, followed immediately by an equal volume of a freshly-prepared 10 mg/mL solution of sinapinic acid (Sigma, 85429) in 50% aqueous (v:v) acetonitrile containing 0.1% , trifluoroacetic acid (v:v).

Positive-ion MALDI mass spectra were acquired using a Bruker ultrfleXtreme in linear mode, equipped with a Nd:YAG smart beam laser. Bruker FlexControl (version 3.4) was used to operate acquisition. Laser power was manually adjusted for each sample. MS spectra were acquired over a range of 5,000-20,000 *m/z*. Spectra were smoothed using SavitzkyGolay, 5 *m/z*, 3 cycles, before centroid peak picking to average masses with a signal to nose threshold of 3. Bruker flexAnalysis software (version 3.4) was used to perform spectral processing and peak list generation.

7.5.2.4 Terminal verification by MALDI-ISD

A 1 mL aliquot of 100 mM protein solution in aqueous 10 mM ammonium acetate was applied to a ground steel MALDI target plate, followed immediately by an equal volume of a freshly-prepared saturated solution of 1,5-diaminonaphthalene matrix (Sigma, 56451) in 50% aqueous (v:v) acetonitrile containing 0.1% , trifluoroacetic acid (v:v).

Positive-ion MALDI-ISD mass spectra were acquired using a Bruker ultrafleXtreme in reflectron mode, equipped with a Nd:YAG smart beam laser. Bruker FlexControl (version 3.4) was used to operate acquisition. MS spectra were acquired over a range of 800-4,000 *m/z*. Monoisotopic peak detection used a SNAP averagine algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) with a minimum S/N of 3. Laser power was manually adjusted for each sample. Bruker flexAnalysis software (version 3.4) was used to perform spectral processing, peak list generation and annotation. In source decay, c-ion series were manually annotated onto the obtained spectrum for N-terminal identification, allowing for a mass error of 0.5 Da.

7.5.2.5 Determination of conjugate stability

JVZ-007 was modified with reagent **5.07** (100 equiv.) in accordance with procedure previously described (section 7.5.5.2), on a 50 µM scale. The reaction outcome was determined using **LC-MS method C** (section 7.1.3) after 19 h before the crude reaction mixture was placed directly into a 3.5 kDa MWCO dialysis membrane and incubated in HPLC-grade water (45.0 mL) at 37 °C with agitation (800 rpm) for 24 h. After this time, the sample was analysed using the same LC-MS method to determine the stability of the conjugates.

7.5.2.6 Copper-catalysed azide-alkyne cycloaddition (CuAAC) of modified JVZ-007 samples

Crude reaction mixtures of JVZ-007 conjugates modified with **5.07** were dialysed into phosphate-buffered saline (100 mM, pH 7.4) (SnakeSkin™ Dialysis Tubing, 3.5K MWCO). To a solution of **5.09** (50 µL, 40 µM in 0.1 M, pH 7.4 phosphate-buffered saline, 2 nmol, 1 eq.) was added a stock solution of AZ647-N_3 (0.8 µL, 5 mM stock in DMSO, 4 nmol, 2 eq.), a stock solution of $CuSO_4$ (2 µL, 200 mM stock in HPLC-water, 400 nmol, 200 eq.) and a stock solution of (+)-sodium L-ascorbate (8.3 µL, 15 mg/mL stock in HPLC-water, 628 nmol, ~315 eq.). The reaction vessel was sealed and protected from light and agitated for 45 mins. After this time the sample was dialysed into phosphate-buffered saline (100 mM, pH 7.4) (SnakeSkin™ Dialysis Tubing, 3.5K MWCO) to remove excess reagents before being visualised by tricine SDS-PAGE.

7.5.2.7 Tricine sodium dodecyl sulphate polyacrylamide gel electrophoresis (TSDS-PAGE)

TSDS-PAGE analysis was performed on a 6% (w/v) stacking polyacrylamide gel atop a 15% (w/v) resolving polyacrylamide gel using the procedures and recipes found in the literature.[187] To 15 µL of sample (50 µM) was added 5 µL of loading buffer (2% SDS, 2 mM 2-mercaptoethanol, 4% glycerol, 40 mM Tris-HCl and 0.01% bromophenol blue) before being heated at 40 ºC for 1 h. 5 µL of protein ladder (PageRuler Plus Prestained Protein Ladder (Thermoscientific)) was added to the first well and 15 µL of boiled sample added to successive wells. The gel was ran at 60 mA, 100 V for 45 mins before increasing to 100 mA, 150 V for around 3.5 h, or until the gel front was approaching the bottom of the gel. The gels were then carefully removed from the PAGE apparatus and covered in fixing solution (50% MeOH, 10% AcOH, 100 mM ammonium acetate) and rocked for 1 h. The gels were then stained in 0.1% Coomassie Brilliant Blue R-250 (50% MeOH, 10% AcOH) for 24 h before destaining in 50% MeOH, 10% AcOH. Images were captured on a Syngene G:BOX Chemi XRQ equipped with Synoptics 4.0 MP camera, with GeneSys software.

7.5.3 Experimental procedures for functional testing of modified proteins

7.5.3.1 Immunofluorescence staining and microscopy

Tissue microarrays: Tissue microarray (TMA) slides were de-waxed, and antigen retrieval was carried out by pressure cooking in Tris-EDTA (pH 9.0). Slides were blocked using 1% BSA in TBS for 1 h. JVZ nanobody was added to the slides at a final concentration of 0.2 µM in TBS and incubated for 1 h at RT. Slides were mounted using mounting medium containing DAPI (Abcam, #ab104139) and imaged using a Leica fluorescence microscope. Images were processed using LASX software.

Prostate cancer cell-line: LNCaP cells were grown on coverslips. Cells were washed 2x with cold PBS, then blocked with 1% human serum in PBS (HS/PBS) for 30 min at 4°C. The blocking solution was removed and the JVZ conjugate was added at a 1:100 dilution in HS/PBS. Coverslips were incubated for 3 h at 4 °C. Cells were washed 3 x 5 min with PBS then fixed with 3% formaldehyde in PBS for 15 minutes at room temperature. After 3 x PBS washes, coverslips were mounted using Mowiol containing DAPI at a 1:5000 dilution. Coverslips were imaged using a Zeiss Axioplan fluorescence microscope and processed using image J. Images were taken at an exposure time of 3s.

7.5.3.2 CCL5 and macrocycle derivatives activation and down-modulation of CCR5

Endocytosis assay: CHO-CCR5 cells detached in PBS-10mM EDTA were resuspended at 2 × 106 cells/mL in Binding Medium (BM: 1xRPMI w/o phenol red plus 1% BSA pH 7.0) pre-labelled in suspension with the anti-CCR5 specific monoclonal antibody MC-5. This is known to bind the *N*-terminus of the receptor without affecting chemokine binding[182] and used at 5µg/mL in BM before treating cells with 100 nM of chemokine tools (in-house produced wild-type CCL5[182] or the indicated macrocycle derivative) for up to 30 mins at 37°C, keeping one control of untreated cells (medium). Following washes in fresh BM, cells were fixed overnight with a 1% formaldehyde solution in PBS before detecting MC-5-bound CCR5 remaining at the surface using an Alexa Fluor 633 anti-mouse IgG fluorescent secondary antibody (Invitrogen) in FACS Buffer (PBS, 1% FCS with 0.05% sodium azide). Cell-associated Mean Fluorescence Intensity was measured with a CytoFLEX S flow cytometer (Beckman Coulter) and results are expressed as percentage of signal from cells prior to treatment.[182]

Recycling assay: CHO-CCR5 cells were detached, resuspended in BM and treated with chemokine tools as above to trigger downmodulation. After 30 mins, cells were washed and resuspended in fresh BM supplemented with 800 nM of the CCR5 antagonist TAK-779 for up to 60 mins with further incubation at 37°C, whilst keeping on ice a sample for reference of downmodulation. TAK-779 is used as a displacer of ligands remaining bound to CCR5, avoiding further cycles of receptor internalisation.[188] Collected cells kept on ice were then fixed in 1% formaldehyde before labelling for 1 h at RT with Alexa Fluor 488-coupled MC-5 at 5µg/ml in FACS Buffer. The sample analysis was done with a CytoFLEX S, as above. Results are expressed as percentage of fluorescence signal recovered compared to cells taken at the end of the downmodulation step.

Phosphorylation assay: CHO-CCR5 cells were treated with chemokine tools as above for up to 30 mins. Aliquots were removed at the indicated time and fixed in equal volume of 6% formaldehyde solution in PBS for 20 min. Cells were washed 1x with FACs buffer and kept in ice-cold methanol (−80 °C) until staining was carried out. Cells were removed from the freezer and were washed 3x with FACs buffer and incubated with the Anti-CCR5 E11/19-APC (IgG1) specific for CCR5 phosphorylated Ser349 (purchased from Biolegends as recommended by the manufacturer) for 1 h at 4 °C.[182] Cells were then washed 3 x in FACs buffer and resuspended in 200 µL FACs buffer and analysed using a CytoFLEX S. Data were expressed as a

mean ± SEM of the stated number of experiments and were analysed with GraphPad Prism v9.4.1. The indicated statistical tests were used, where appropriate.

References

1. D. J. Newman, G. M. Cragg, in *Macrocycles in Drug Discovery*, ed. J. Levin, The Royal Society of Chemistry, 2014, ch.1, pp. 1.

2. K. R. Romines, K. D. Watenpaugh, P. K. Tomich, W. Jeffrey Howe, J. K. Morris, K. D. Lovasz, A. M. Mulichak, B. C. Finzel, J. C. Lynn, M. M. Homg, F. J. Schwende, M. J. Ruwart, G. L. Zipp, K. T. Chong, L. A. Dolak, L. N. Toth, G. M. Howard, B. D. Rush, K. F. Wilkinson, P. L. Possert, R. J. Dalga, R. R. Hinshaw, *J. Med. Chem.*, 1995, **38**, 1884.

3. D. F. Veber, S. R. Johnson, H.Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, *J. Med. Chem.*, 2002, **45**, 2615.

4. K. M. George, M.C. Frantz, K. Bravo-Altamirano, C. R. LaValle, M. Tandon, S. Leimgruber, E. R. Sharlow, J. S. Lazo, Q.J. Wang, P. Wipf, *Pharmaceutics*, 2011, **3**, 186.

5. R. A. Bauer, T. A. Wenderski, D. S. Tan, *Nat. Chem. Biol*., 2013, **9**, 21.

6. F. Giordanetto, J. Kihlberg, *J. Med. Chem.*, 2014, **57**, 278.

7. E. Marsault, M. L. Peterson, *J. Med. Chem.*, 2011, **54**, 1961.

8. Y. U. Kwon, T. Kodadek, *Chem. Biol.*, 2007, **14**, 671.

9. T. Rezai, B. Yu, G. L. Millhauser, M. P. Jacobson, R. S. Lokey, *J. Am. Chem. Soc.*, 2006, **128**, 2510.

10. Y. Zhou, K. Jie, R. Zhao, F. Huang, *Adv. Mater.*, 2020, **32**, 1904824.

11. C. Zhong, Y. Yan, Q. Peng, Z. Zhang, T. Wang, X. Chen, J. Wang, Y. Wei, T. Yang, L. Xie, *Nanomater.*, 2023, **13**, 1750.

12. K. Velmurugan, M. Mohan, B. Li, K. Wang, M. Zuo, X.Y. Hu, *Mater. Adv.*, 2020, **1**, 2646.

13. D. G. Jimenez, V. Poongavanam, J. Kihlberg, *J. Med. Chem.,* 2023, **66**, 5377.

14. T. Kim, E. Baek, J. Kim, *Pharmaceuticals*, 2025, **18**, 617.

15. E. M. Driggers, S. P. Hale, J. Lee, N. K. Terrett, *Nat. Rev. Drug Discovery*, 2008, **7**, 608.

16. Y. Du, A. Semghouli, Q. Wang, H. Mei, L. Kiss, D. Baecker, V. A. Soloshonok, J. Han, *Arch. Pharm.*, 2025, **358**, e2400890.

17. G. Illuminati, L. Mandolini, *Acc. Chem. Res.*, 1981, **14**, 95.

18. J. C. Collins, K. James, *Med. Chem. Commun.*, 2012, **3**, 1489.

19. X.N. Sun, A. Liu, K. Xu, Z. Zheng, K. Xu, M. Dong, B. Ding, J. Li, Z.Y. Zhang, C. Li, *Aggregate*, 2024, **5**, 607.

20. A. C. Cope, M. M. Martin, M. A. McKervey, *Q. Rev. Chem. Soc.*, 1966, **20**, 119.

21. W. P. Unsworth, K. A. Gallagher, M. Jean, J. P. Schmidt, L. J. Diorazio, R. J. K. Taylor, *Org. Lett.,* 2013, **15**, 262.

22. M. Malesevic, U. Strijowski, D. Bachle, N. Sewald, *J. Biotechnol.*, 2004, **112**, 73.

23. C. Rosenbaum, H. Waldmann, *Tetrahedron Lett.*, 2001, **42**, 5677.

24. Q. Fan, T. Wang, J. Dai, J. Kuttner, G. Hilt, J. M. Gottfried, J. Zhu, *ACS Nano*, 2017, **11**, 5070.

25. E. Godin, A.C. Bedard, M. Raymond, S. K. Collins, *J. Org. Chem.*, 2017, **82**, 7576.

26. V. Marti-Centelles, M. D. Pandey, M. I. Burguete, S. V. Luis, *Chem. Rev.*, 2015, **115**, 8736.

27. B. J. Larsen, Z. Sun, P. Nagorny, *Org. Lett.*, 2013, **15**, 2998.

28. S. Chowdhury, G. Schatte, H. B. Kraatz, *Dalton Trans.*, 2004, **11**, 1726.

29. J. R. Donald, W. P. Unsworth, *Chem. Eur. J.*, 2017, **23**, 8780.

30. G. H. Posner, M. A. Hatcher, W. A. Maio, *Org. Lett.*, 2005, **7**, 4301.

31. G. H. Posner, M. A. Hatcher, W. A. Maio, *Tetrahedron*, 2016, **72**, 6025.

32. S. Kim, G. H. Joe, J. Y. Do, *J. Am. Chem. Soc.*, 1993, **115**, 3328.

33. Y.S. Lee, J.W. Jung, S.H. Kim, J.K. Jung, S.M. Paek, N.J. Kim, D.J. Chang, J. Lee, Y.G. Suh, *Org. Lett.*, 2010, **12**, 2040.

34. M. H. Weston, K. Nakajima, M. Parvez, T. G. Back, *J. Org. Chem.*, 2008, **73**, 4630.

35. M. L. Steigerwald, W. A. Goddard, D. A. Evans, *J. Am. Chem. Soc.*, 1979, **101**, 1994.

36. J. C. Orejarena Pacheco, T. Opatz, *J. Org. Chem.*, 2014, **79**, 5182.

37. D. Lednicer, C. R. Hauser, *J. Am. Chem. Soc*., 1957, **79**, 4449.

38. S. G. Cauwberghs, P. J. De Clercq, *Tetrahedron Lett.*, 1988, **29**, 6501.

39. C. A. Grob, W. Baumann, *Helv. Chim. Acta.*, 1955, **38**, 594.

40. A. Eschenmoser, D. Felix, G. Ohloff, *Helv. Chim. Acta.*, 1967, **50**, 708.

41. D. M. Gordon, S. J. Danishefsky, G. K. Schulte, *J. Org. Chem.*, 1992, **57**, 7052.

42. R. Walchli, A. Guggisberg, M. Hesse, *Tetrahedron Lett.*, 1984, **25**, 2205.

43. E. J. Corey, D. J. Brunelle, K. C. Nicolaou, *J. Am. Chem. Soc.*, 1977, **99**, 7359.

44. C. Kitsiou, J. J. Hindes, P. I'Anson, P. Jackson, T. C. Wilson, E. K. Daly, H. R. Felstead, P. Hearnshaw, W. P. Unsworth, *Angew. Chem. Ind. Ed.*, 2015, **54**, 15794.

45. L. G. Baud, M. A. Manning, H. L. Arkless, T. C. Stephens, W. P. Unsworth, *Chem. Eur. J.*, 2017, **23**, 2225.

46. T. C. Stephens, M. Lodi, A. Steer, Y. Lin, M. Gill, W. P. Unsworth, *Chem. Eur. J.*, 2017, **23**, 13314.

47. T. C. Stephens, A. Lawer, T. French, W. P. Unsworth, Chem. Eur. J., 2018, **24**, 13947.

48. A. Lawer, A. Rossi-Ashton, T. C. Stephens, B. J. Challis, R. G. Epton, J. M. Lynam, W. P. Unsworth, *Angew. Chem. Int. Ed.*, 2019, **58**, 13942.

49. I. Zalessky, J. M. Wooton, J. K. F. Tam, D. E. Spurling, W. C. Glover-Humphreys, J. R. Donald, W. E. Orukotan, L. C. Duff, B. J. Knapper, A. C. Whitwood, T. F. N. Tanner, A. H. Miah, J. M. Lynam, W. P. Unsworth, *J. Am. Chem. Soc.*, 2024, **146**, 5702.

50. K. Y. Palate, Z. Yang, A. C. Whitwood, W. P. Unsworth, *RSC Chem. Biol.*, 2022, **3**, 334.

51. Z. Yang, M. Arnoux, D. Hazelard, O. R. Hughes, J. Nabarro, A. C. Whitwood, M. A. Fascione, C. D. Spicer, P. Compain, W. P. Unsworth, *Org. Biomol. Chem.*, 2024, **22**, 2985.

52. S. Yilmaz, H. Zhou, C. Ozsan, A. Frei, P. O'Brien, W. P. Unsworth, *Org. Biomol. Chem.*, 2025, **23**, 8053.

53. W. E. Orukotan, B. J. Knapper, D. Dimitrova, W. P. Unsworth, *Chem. Eur. J.*, 2025, **0**, e02887.

54. N. Stephanopoulos, M. B. Francis, *Nat. Chem. Biol.*, 2011, **7**, 876.

55. P. Chauhan, V. Ragendu, M. Kumar, R. Molla, S. D. Mishra, S. Basa, V. Rai, *Chem. Soc. Rev.*, 2024, **53**, 380.

56. C. D. Spicer, B. G. Davis, *Nat. Commun.*, 2014, **5**, 4740.

57. P. Theisen, C. McCullum, K. Upadhya, K. Jacobson, H. Vu, A. Andrus, *Tetrahedron Lett.*, 1992, **33**, 5033.

58. J. M. Sasso, R. Tenchov, R. Bird, K. A. Iyer, K. Ralhan, Y. Rodriguez, Q. A. Zhou, *Bioconjugate Chem.*, 2023, **34**, 1951.

59. For a review of lysine bioconjugation, see: M. Haque, N. Forte, J. R. Baker, *Chem Commun.*, 2021, **57**, 10689–10702. For a review of cysteine bioconjugation, see: Y. Liao, X. Jiang, *Explor. Drug Sci.*, 2024, **2**, 540.

60. C. H. Kim, J. Y. Axup, P. G. Schultz, *Curr. Opin. Chem. Biol.*, 2013, **17**, 412.

61. E. Jacob, R. Unger, *Bioinformatics*, 2007, **23**, 225.

62. T. J. Sereda, C. T. Mant, A. M. Quinn, R. S. Hodges, *J. Chromatogr.*, 1993, **646**, 17.

63. L. J. Barber, N. D. J. Yates, M. A. Fascione, A. Parkin, G. R. Hemsworth, P. G. Genever, C. D. Spicer, *RSC Chem. Biol.*, 2023, **4,** 56.

64. X. Zou, Z. Liu, F. Song, W. Zhou, J. Hang, C. Feng, T. Yuan, J. Dong, W. Shi, F. Tang, W. Huang, *Commun. Biol.,* 2025, **8**, 1165.

65. I. E. Gentle, D. P. De Souza, M. Baca, *Bioconjug. Chem.*, 2004, **15**, 658.

66. J. Fu, J. Gao, Z. Liang, D. Yang, *Molecules*, 2020, **26**, 171.

67. H. Ren, F. Xiao, K. Zhan, Y. P. Kim, H. Xie, J. Rao, *Angew. Chem. Int. Ed.*, 2009, **48**, 9658.

68. G. Liang, H. Ren, J. Rao, *Nat. Chem*., 2010, **2**, 54.

69. J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 16793.

70. J. Jeon, B. Shen, L. Xiong, Z. Miao, K. H. Lee, J. Rao, F. T. Chin, *Bioconjug. Chem.,* 2012, **23**, 1902.

71. For an example protocol, see: G. J. L. Bernardes, M. Steiner, I. Hartmann, D. Neri, G. Casi, *Nat. Protoc.*, 2013, **8**, 2079. For the original work, see: G. Casi, N. Huguenin-Dezot, K. Zuberbuhler, J. Schueuermann, D. Neri, *J. Am. Chem. Soc.*, 2012, **134**, 5887.

72. D. Bermejo-Velasco, G. N. Nawale, O. P. Oommen, J. Hilborn, O. P. Varghese, *Chem. Commun.*, 2018, **54**, 12507.

73. A. Bandyopadhyay, S. Cambray, J. Gao, *Chem. Sci.*, 2016, **7**, 4589.

74. X. Dong, R. K. Brahma, C. Fang, S. Q. Yao, *Chem Sci.*, 2022, **13**, 4239.

75. A. Istrate, M. B. Geeson, C. D. Navo, B. D. Sousa, M. C. Marques, R. J. Taylor, T. Journeaux, S. R. Oehler, M. R. Mortensen, M. J. Deery, A. D. Bond, F. Corzana, G. Jimenez-Oses, G. J. L. Bernardes, *J. Am. Chem. Soc.*, 2022, **144**, 10396.

76. K. F. Geoghegan, J. G. Stroh, *Bioconjug. Chem.*, 1992, **3**, 138.

77. L. S. Witus, C. Netirojjanakul, K. S. Palla, E. M. Muehl, C.H. Weng, A. T. Iavarone, M. B. Francis, *J. Am. Chem. Soc.*, 2013, **135**, 17223.

78. R. J. Spears, M. A. Fascione, *Org. Biomol. Chem.*, 2016, **14**, 7622.

79. X. Ning, R. P. Temming, J. Dommerholt, J. Guo, D. B. Ania, M. F. Debets, M. A. Wolfert, G. J. Boons, F. L. Van Delft, *Angew. Chem. Int. Ed.*, 2010, **122**, 3129.

80. A. Dirksen, T. M. Hackeng, P. E. Dawson, *Angew. Chem. Int. Ed.*, 2006, **118**, 7743.

81. R. V. Hay, R. S. Skinner, O. C. Newman, S. L. Kunkel, L. R. Lyle, B. Shapiro, M. D. Gross, *Nucl. Med. Commun.*, 1997, **18**, 367.

82. M. J. Gaunt, C. C. C. Johansson, A. McNally, N. T. Vo, *Drug Discovery Today*, 2007, **12**, 8.

83. A. B. Northrup, D. W. C. MacMillan, *J. Am. Chem. Soc.*, 2002, **124**, 6798.

84. R. J. Spears, R. L. Brabham, D. Budhadev, T. Keenan, S. McKenna, J. Walton, J. A. Brannigan, A. M. Brzozowski, A. J. Wilkinson, M. Plevin, M. A. Fascione, *Chem. Sci.*, 2018, **9**, 5585.

85. T. Keenan, R. J. Spears, S. Akkad, C. S. Mahon, N. E. Hatton, J. Walton, A. Noble, N. D. Yates, C. G. Baumann, A. Parkin, N. Signoret, M. A. Fascione, *ACS Chem. Biol.*, 2021, **16**, 2387.

86. J. Phillips, E. Barrall, *J. Org. Chem.*, 1956, **21**, 692.

87. X. Li, L. Zhang, S. E. Hall, J. P. Tam, *Tetrahedron Lett.*, 2000, **41**, 4069.

88. J. J. Maresh, L.A. Giddings, A. Friedrich, E. A. Loris, S. Panjikar, B. L. Trout, J. Stockigt, B. Peters, S. E. O'Connor, *J. Am. Chem. Soc.*, 2008, **130**, 710.

89. T. Sasaki, K. Kodama, H. Suzuki, S. Fukuzawa, K. Tachibana, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 4550.

90. S. Pomplun, M. Y. H. Mohamed, T. Oelschlaegel, C. Wellner, F. Bergmann, *Angew. Chem. Int. Ed.*, 2019, **58**, 3542.

91. P. Agarwal, J. van der Weijden, E. M. Sletten, D. Rabuka, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA*, 2012, **110**, 46.

92. J. I MacDonald, H. K. Munch, T. Moore, M. B. Francis, *Nat. Chem. Biol.*, 2015, **11**, 326.

93. L. J. Barber, K. S. Stankevich, C. D. Spicer, *J. Am. Chem. Soc. Au*, 2025, **5**, 1983.

94. A. Tantipanjaporn, M.K. Wong, *Molecules*, 2023, **28**, 1083.

95. A. O.Y. Chan, C.M. Ho, H.C. Chong, Y.C. Leung, J.S. Huang, M.K. Wong, C. M. Che, *J. Am. Chem. Soc.*, 2012, **134**, 2589.

96. J.R. Deng, N. C.H. Lai, K. K.Y. Kung, B. Yang, S.F. Chung, A. S.L. Leung, M.C. Choi, Y.C. Leung, M.K. Wong, *Commun. Chem.*, 2020, **3**, 67.

97. M. K. Miller, H. Wang, K. Hanaya, O. Zhang, A. Berlaga, Z. T. Ball, *Chem Sci.*, 2020, **11**, 10501.

98. R. Singudas, S. R. Adusumalli, P. N. Joshi, V. Rai, *Chem Commun.*, 2015, **51**, 473.

99. H. Chen, R. Huang, Z. Li, W. Zhu, J. Chen, Y. Zhan, B. Jiang, *Org. Biomol. Chem.*, 2017, **15**, 7339.

100. M. J. Matos, B. L. Oliveira, N. Martinez-Saez, A. Guerreiro, P. M. S. D. Cal, J. Bertoldo, M. Maneiro, E. Perkins, J. Howard, M. J. Deery, J. M. Chalker, F. Corzana, G. Jimenez-Oses, G. J. L. Bernardes, *J. Am. Chem. Soc.*, 2018, **140**, 4004.

101. G. W. Anderson, J. E. Zimmerman, F. M. Callahan, *J. Am. Chem. Soc.,* 1963, **86**, 1839.

102. I. Sélo, L. Négroni, C. Créminon, J. Grassi, J. M. Wal, *J. Immunol. Methods*, 1996, **199**, 127.

103. H. Jiang, G. D. D'Agostino, P. A. Cole, D. R. Dempsey, *Methods Enzymol.,* 2020, **639**, 333.

104. J. R. Deng, N. C. H. Lai, K. K. Y. Kung, B. Yang, S. F. Chung, *Commun. Chem.,* 2020, **3**, 67.

105. P. Wadwha, A. Kharbanda, A. Sharma, *Asian J. Org. Chem.,* 2018, **7**, 634.

106. K. Y. Palate, PhD Thesis, University of York, 2022.

107. C. Vilchèze, W. R. Jacobs Jr, *J. Mol. Biol.,* 2019, **431**, 3450.

108. D. Matveychuk, E. M. Mackenzie, D. Kumpula, M-S. Song, A. Holt, S. Kar, K. G. Todd, P. L. Wood, G. B. Baker, *Cell Mol. Neurobiol.,* 2021, **42**, 225.

109. J-P. Armand, V. Ribrag, J-L. Harrousseau, L. Abrey, *Ther. Clin. Risk. Manag.,* 2007, **3**, 213.

110. D. J. Stuehr, N. S. Kwon, C. F. Nathan, O. W. Griffiths, P. L. Feldman, J. Wiseman, *J. Biol. Chem.,* 1991, **266**, 6259.

111. D. K. Kölmel, E. T. Kool, *Chem. Rev*., 2017, **117**, 10358.

112. M. W. Jones, G. Mantovani, C. A. Blindauer, S. M. Ryan, X. Wang, D. J. Brayden, D. M. Haddleton, *J. Am. Chem. Soc.*, 2012, **134**, 7406.

113. G. E. Means, R. E. Feeney, *Bioconjugate Chem.*, 1990, **1**, 2.

114. R. J. Spears, C. McMahon, M. Shamsabadi, C. Bahou, I. A. Thanasi, L. N. C. Rochet, N. Forte, F. Thoreau, J. R. Baker, V. Chudasama, *Chem. Commun.*, 2022, **58**, 645.

115. N. D. J. Yates, S. Akkad, A. Noble, T. Keenan, N. E. Hatton, N. Signoret, M. A. Fascione, *Green Chem.*, 2022, **24**, 8046.

116. A. Dirksen, T. M. Hackeng, P. E. Dawson, *Angew. Chem. Int. Ed.*, 2006, **45**, 7581.

117. A. K. Clarke, W. P. Unsworth, *Chem. Sci.*, 2020, **11**, 2876.

118. Z. Yang, I. Zalessky, R. G. Epton, A. C. Whitwood, J. M. Lynam, W. P. Unsworth, *Angew. Chem. Int. Ed.*, 2023, **135**, e202217178.

119. M. Sánchez-Roselló, M. Escolano, D. Gaviña, C. Del Pozo, *Chem. Rec.*, 2022, **22**, e202100161.

120. A. Y. Rulev, *Russ. Chem. Rev.*, 2011, **80**, 197.

121. For Lewis-acid-catalysed aza-Michael reactions see: H. Zhang, Y. Zhang, L. Liu, H. Xu, Y. Wang, *Synthesis*, 2005, **13**, 2129.

122. For Squaramide-catalysed aza-Michael reactions see: X-J. Zhou, J-Q. Zhao, X-M. Chen, J-R. Zhou, Y-P. Zhang, Y-Z. Chen, Z-M. Zhang, X-Y. Xu, W-C. Yuan, *J. Org. Chem.*, 2019, **84**, 4381.

123. For Thiourea-catalysed aza-Michael reactions see: N. Hayama, T. Azuma, Y. Kobayashi, Y. Takemoto, *Chem. Pharm. Bull.*, 2016, **64**, 704.

124. J. V. Alegre-Requena, E. Marqués-López, R. P. Herrera, *Adv. Synth. Catal.*, 2016, **358**, 1801.

125. J. V. A. Requena, PhD Thesis, University of Zaragoza, 2017.

126. J. M. Ravasco, H. Faustino, A. Trindale, P. M. Gois, *Chem. Eur. J.*, 2019, **25**, 43.

127. B-Q. Shen, K. Xu, L. Liu, H. Raab, S. Bhakta, M. Kenrick, K. L. Parsons-Reponte, J. Tien, S. F. Yu, E. Mai, *Nat. Biotechnol.*, 2012, **30**, 184.

128. V. Chudasama, M. E. B. Smith, F. F. Schumacher, D. Papaioannou, G. Waskman, J. R. Baker, S. Caddick, *Chem Commun.*, 2011, **47**, 8781.

129. F. Frottin, A. Martinez, P. Peynot, R. C. Holz, C. Giglione, T. Meinnel*, Mol. Cell. Proteomics*, 2006, **5**, 2336.

130. R. P. Moerschell, Y. Hosokawa, S. Tsunasawa, F. Sherman, *J. Biol. Chem.*, 1990, **265**, 19638.

131. D. Berne, V. Ladmiral, E. Leclerc, S. Caillol, *Polymers*, 2022, **14**, 4457.

132. R. E. Bird, S. A. Lemmel, X. Yu, Q. A. Zhou, *Bioconjugate Chem.*, 2021, **32**, 2457.

133. Y. Fu, X. Zhang, L. Wu, M. Wu, T. D. James, R. Zhang, *Chem. Soc. Rev.*, 2025, **54**, 201.

134. S. H. L. Verhelst, K. M. Bonger, L. I. Willems, *Molecules*, 2020, **25**, 5994.

135. S. Hapuarachchige, Y. Kato, D. Artemov, *Sci. Rep.*, 2016, **6**, 24298.

136. J. M. Fairhall, J. C. Camilli, B. H. Gibson, S. Hook, A. B. Gamble, *Bioorg. Med. Chem*., 2021, **46**, 116361.

137. Y. G. Gololobov, L. F. Kasukhin, *Tetrahedron*, 1992, **48**, 1353.

138. H. Staudinger, J. Meyer, *Helv. Chim. Acta.*, 1919, **2**, 635.

139. H. C. Chang, C. R. Bertozzi, *Acc. Chem. Res.*, 2001, **34**, 727.

140. M. L. Blackman, M. Royzen, J. M. Fox, *J. Am. Chem. Soc.*, 2008, **130**, 13518.

141. B. M. Zeglis, K. K. Sevak, T. Reiner, P. Mohindra, S. D. Carlin, P. Zanzonico, R. Weissleder, J. S. Lewis, *J. Nucl. Med.*, 2013, **54**, 1389.

142. S. Liu, M. Hassink, R. Selvaraj, L-P. Yap, R. Park, H. Wang, X. Chen, J. M. Fox, Z. Li, P. S. Conti, *Mol. Imaging.,* 2013, **12**, 121.

143. D. Richter, E. Lakis, J. Piel, *Nat. Chem.*, 2023, **15**, 1422.

144. V. V. Rostovstev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem. Int. Ed.,* 2002, **41**, 2596.

145. C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.*, 2002, **67**, 3057.

146. E. Haldón, M. C. Nicasio, P. J. Pérez, *Org. Biomol. Chem.,* 2015, **13**, 9528.

147. D. Rösner, T. Schneider, D. Schneider, M. Scheffner, A. Marx, *Nat. Protoc.,* 2015, **10**, 1594.

148. D. Döhler, P. Michael, W. H. Binder, *Acc. Chem. Res.*, 2017, **50**, 2610.

149. L. Jin, D. R. Tolentino, M. Melaimi, G. Bertrand, *Sci. Adv.,* 2015, **1**, e1500304.

150. B. T. Worrell, J. A. Malik, V. V. Fokin, *Science*, 2013, **340**, 457.

151. L. Li, Z. Zhang, *Molecules,* 2016, **21**, 1393.

152. N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.,* 2004, **126**, 15046.

153. S. R. Punganuru, H. R. Madala, S. N. Venugopal, R. Samala, C. Mikelis, K. S. Srivenugopal, *Eur. J. Med. Chem.*, 2016, **107**, 233.

154. S-G. Kim, T-H. Park, *Synth. Commun.,* 2007, **37**, 1027.

155. D. J. Guerin, T. E. Horstmann, S. J. Miller, *Org. Lett.*, 1999, **1**, 1107.

156. H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.*, 2001, **40**, 2004.

157. M. Lakshmi Kantam, S. Laha, J. Yadav, S. Jha, *Tetrahedron Lett.,* 2009, **50**, 4467.

158. M. J. Matos, B. L. Oliveira, N. Martinez-Sáez, A. Guerreiro, P. M. S. D. Cal, J. Bertoldo, M. Maneiro, E. Perkins, J. Howard, M. J. Deery, J.M. Chalker, F. Corzana, G. Jiménez-Osés, G. J. L. Bernardes, *J. Am. Chem. Soc*., 2018, **140**, 4004.

159. X. Chen, K. Muthoosamy, A. Pfisterer, B. Neumann, T. Weil, *Bioconjugate Chem.*, 2012, **23**, 500.

160. S. M. Hacker, K. M. Backus, M. R. Lazear, S. Forli, B. E. Correia, B. F. Cravatt, *Nat. Chem.*, 2017, **9**, 1181.

161. M. Haque, N. Forte, J. R. Baker, *Chem. Commun.*, 2021, **57**, 10689.

162. J. Branch, B. S. Rajagopal, A. Paradisi, N. Yates, P. J. Lindley, J. Smith, K. Hollingsworth, W. B. Turnbell, B. Henrissat, A. Parkin, A. Berry, G. R. Hemsworth, *Biochem. J.*, 2021, **478**, 2927.

163. K. L. S. Chatalic, J. Veldhoven-Zweistra, M. Bolkestein, S. Hoeben, G. A. Koning, O. C. Boerman, M. de Jong, W. M. van Weerden, *J. Nucl. Med.*, 2015, **56**, 1094.

164. M. S. Hofman, L. Emmett, *Eur. Urol.,* 2019, **76**, 479.

165. J. S. Ross, C. E. Sheehan, H. A. G. Fisher, R. P. Kaufman, K. Gray, I. Webb, G. S. Gray, R. Mosher, B. V. S. Kallakury, *Clin. Cancer Res.*, 2003, **9**, 6357.

166. J. Su, X. Liu, S. Guo, J. Zhang, X. Wei, X. Li, *J. Cancer Res. Clin. Oncol.*, 2023, **149**, 6703.

167. N. Panchuk-Voloshina, J. Bishop-Stewart, M. K. Bhalgat, P. J. Millard, F. Mao, W-Y. Leung, R. P. Haugland, *J. Histochem. Cytochem.*, 1999, **47**, 1179.

168. M. Bates, G. T. Dempsey, K. H. Chen, X. Zhuang, *ChemPhysChem*, 2012, **13**, 99.

169. A. Zlotnik, O. Yoshie, *Immunity*, 2012, **36**, 705.

170. M. Mack, B. Lucklow, PJ. Nelson, J. Cihak, G. Simmons, PR. Clapham, N. Signoret, M. Marsh, M. Stangassiner, F. Borlat, T. N. Wells, D. Schlöndorff, A. E. Proudfoot, *J. Exp. Med.*, 1998, **187**, 1215.

171. H. Gaertner, F. Cerini, J. M. Escola, O. Hartley, *PNAS*, 2008, **105**, 17706.

172. J. M. Escola, G. Kuenzi, H. Gaertner, M. Foti, O. Hartley, *J. Biol. Chem.*, 2010, **285**, 41772.

173. C. Pastore, G. R. Picchio, F. Galimi, R. Fish, O. Hartley, R. E. Offord, D. E. Mosier, *Antimicrob. Agents. Chemother.*, 2003, **47**, 509.

174. P. Isaikina, C. Tsai, N. Dietz, F. Pamula, A. Grahi, K. N. Goldie, R. Guixà-González, C. Branco, M. Paolini-Bertrand, N. Calo, F. Cerini, G. F. X. Schertler, O. Hartley, H. Stahlberg, T. Maier, X. Deupi, S. Grzesiek, *Sci. Adv.*, 2021, **7**, eabg8685.

175. L. Vangelista, M. Secchi, P. Lusso, *Vaccine*, 2009, **26**, 3008.

176. A. R. Noble, S. Akkad, N. D. J. Yates, J. M. Jeffries, N. Signoret, M. A. Fascione, *RSC Adv.*, 2024, **14**, 23796.

177. Y. Liu, A. Liu, X. Li, Q. Liao, W. Zhang, L. Zhu, R. D. Ye, *Cell Rep.*, 2024, **43**, 114578.

178. J. Penfield, L. Zhang, *Commun. Chem.*, 2024, **7**, 205.

179. W. E. Orukotan, K. Y. Palate, B. Pogrányi, P. Bobinski, R. G. Epton, L. Duff, A. C. Whitwood, G. Grogan, J. M. Lynam, W. P. Unsworth, *Chem. Eur. J.*, 2023, **29**, e202303270.

180. LC. Han, P. A. Stanley, P. J. Wood, P. Sharma, A. I. Kuruppu, T. D. Bradshaw, J. E. Moses, *Org. Biomol. Chem.*, 2016, **14**, 7585.

181. N. D. Yates, S. Akkad, A. R. Noble, T. Keenan, N. E. Hatton, N. Signoret, M. A. Fascione, *Green Chem.*., 2022, **24**, 8046.

182. A. Tufail, S. Akkad, A. R. Noble, M. A. Fascione, N. Signoret, *Sci. Rep.,* 2024, **14**, 24188.

183. R. T. Raines, *Chem. Rev.*, 1998, **98**, 1045.

184. M. Dautrevaux, Y. Boulanger, K. Han, G. Biserte, *Eur. J. Biochem.*, 1969, **11**, 267.

185. J. Zaia, R. S. Annan, K. Biemann, *Rapid Commun. Mass Spectrom.*, 1992, **6**, 32.

186. T. Nakashima, H. Higa, H. Matsubara, A. M. Benson, K. T. Yasunobu, *J. Biol. Chem.*, 1966, **241**, 1166.

187. H. Schagger, *Nat Protoc*., 2006, **1**, 16.

188. N. Signoret, T. Christoph, M. Oppermann, M. Marsh, *Traffic,* 2004, **5**, 529.