Kinetics and Stability of N-Terminal Protein Modification Chemistries

Lydia Jane Barber

PhD

University of York
Department of Chemistry

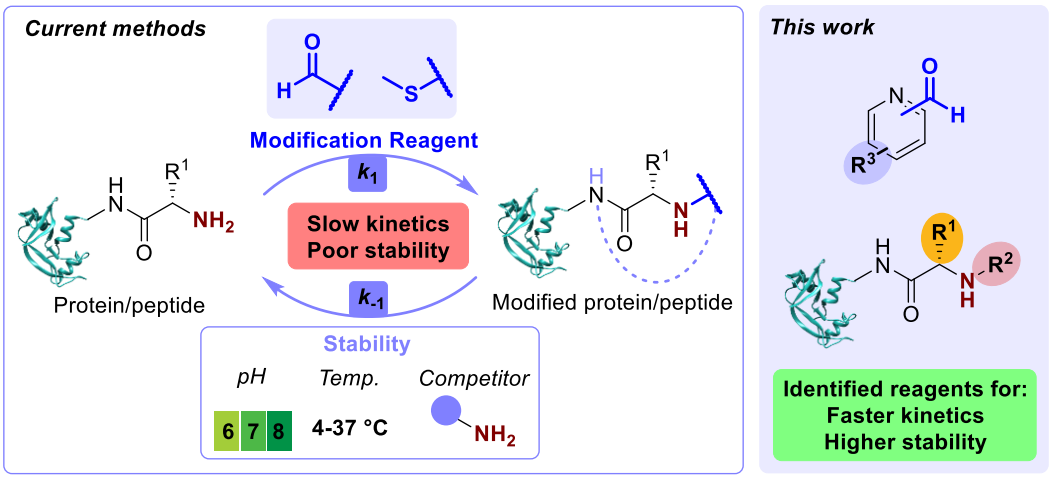
September 2024

Abstract

N-terminal targeting has emerged as a powerful means to functionalise proteins. However, selectivity is often poor or the conjugates formed suffer from instability. We have undertaken a detailed comparative study of the conversion, selectivity, and stability of leading N-terminal modification strategies to provide key insight into the formation and utility of the resultant protein-conjugates (Chapter 2). Critically, all N-terminal modification strategies were found to exhibit slow kinetics and some extent of reversibility, with reaction efficiency and selectivity found to be highly protein dependent. This work highlights the need for the screening of a toolbox of complementary N-terminal modification strategies to ensure optimal properties are achieved for a given target protein and application.

We next used our initial study as a platform to develop new modification reagents that overcome current limitations, through: *i)* changing the nature of the N-terminus or N-terminal residue (Chapter 3.1 and 3.2); *ii)* harnessing the reactivity of neighbouring residues through the development of proximity-driven chemistries (Chapter 3.3); and *iii)* influencing N-terminal labelling equilibria through reagent functionalisation (Chapter 4).

The functionalisation of N-terminal targeting protein modification reagents can influence the series of equilibria governing the overall modification process. For example, hydroxyl substituents can participate in three possible roles: *i)* intramolecular general acid catalysis; *ii)* hydrogen bonding catalysis; or *iii)* carbonyl activation *via* resonance donation. By varying both the nature and location of substituents, we isolated the different roles to identify derivatives with both faster peptide modification kinetics and higher conversion of model proteins. Our approach provides a valuable step forward in the goal to achieve a bioconjugation strategy applicable to a wide range of protein targets with high efficiency, selectivity, and stability. Finally, we applied our reagent screening approach to improve the biological activity of important and industrially relevant recombinant proteins through mimicking post-translational modifications (Chapter 5).



Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references, and appropriate credit has been given where others have conducted experiments which have contributed to this work. I acknowledge the publication “Selectivity and stability of N-terminal targeting protein modification chemistries” which arose from this work.

LYDIA JANE BARBER

26/09/2024

Acknowledgements

Firstly, I would like to thank Chris Spicer and Paul Genever, and the BBSRC White Rose DTP for providing the project. In particular, I would like to express my gratitude to Chris Spicer for being an incredible supervisor over the past 4 years. Thank you for your guidance and ideas, for always being available for support, for all of your helpful feedback, and for encouraging me to learn new skills. Having the opportunity to work in such a supportive and encouraging environment has been an amazing experience.

I would also like to thank past and present members of the Spicer group. I was lucky to cross paths with so many amazing scientists and friends in Chemical Biology and E214 labs – thank you all so much for your help and input, and for making it such a fun place to work. I would also like to say a huge thank you to our wonderful technicians: Julia, Roksana, and Iman, for all the work they do to keep the labs running smoothly.

Thank you to all the people who have performed collaborative work which has contributed to this thesis. Thank you to Laetitia Raynal for the synthesis of reagents for proximity-driven labelling, Ksenia Stankevich for DFT calculations and boronic acid synthesis, Julia Capar for synthesis of a dansyl lysine-containing peptide, Reuben Breetveld for submitting most of my NMR samples, and Nick Yates for CjX183-D protein expression. Thank you to technical staff at the University of York: Ed Bergstrom, Karl Heaton, Adam Dowle, and Chris Taylor for your expertise in LC-MS, MS, and tryptic digests. Thank you also to Heather Fish and Alex Heyam for submitting NMRs for all peptide experiments.

I would like to thank the R&D and bioassay teams at Qkine for providing relevant training and making me feel so welcome during my iCASE project. In particular, thank you to Helen Bell, Kerry Price, and Greg Hughes for contributions to the Sonic Hedgehog project, to Kalani for your support with accommodation, and to Catherine Elton and Marko Hyvönen for the generous funding contribution. Also, thank you to Dijana Matak Vinkovic for LC-MS training and support at the University of Cambridge.

Finally, I would like to thank my family for all your love, support, and cups of tea. Izzy and Ellie, thank you for all the adventures and for always being there.

Table of Contents

List of Abbreviations

Amino acids

3-letter code	*1-letter code*	
Ala	A	Alanine
Arg	R	Arginine
Asn	N	Asparagine
Asp	D	Aspartic acid
Cys	C	Cysteine
Glu	E	Glutamic acid
Gln	Q	Glutamine
Gly	G	Glycine
His	H	Histidine
Ile	I	Isoleucine
Leu	L	Leucine
Lys	K	Lysine
Met	M	Methionine
Phe	F	Phenylalanine
Pro	P	Proline
Ser	S	Serine
Thr	T	Threonine
Trp	W	Tryptophan
Tyr	Y	Tyrosine
Val	V	Valine

General

ADC	Antibody-Drug Conjugate
Ac	Acyl
AG	Affinity-guided
AGOX	Affinity-guided oxime
AHA	Azidohomoalanine
AI	Acyl imidazole
aq.	Aqueous
ArB(Epin)	Aryl boronic 1,1,2,2-tetraethylethylene glycol ester
ArB(pin)	Aryl pinacol boronate
ATP	Adenosine triphosphate
BA	Benzaldehdye
Boc	*tert*-Butyloxycarbonyl
B_2Pin_2	Bis(pinacolato)diboron

CHC		Citric acid, HEPES and CHES
CHES		N-cyclohexyl-2-aminoethanesulfonic acid
cLogP		Logarithm of the calculated partition coefficient
Clostripain LC		Clostripain light chain
CoA		Coenzyme A
CS		Calf serum
CuAAC	“Click”	Copper-catalysed azide-alkyne cycloaddition
Da		Daltons
dba		Dibenzylideneacetone
DCC		*N,N’*-dicyclohexylcarbodiimide
Di-Ala		Dialanine
DIC		*N,N’*-diisopropylcarbodiimide
DMA		Dimethylacetamide
DMAP		4-Dimethylaminopyridine
DMEM		Dulbecco’s Modified Eagle Medium
DMF		*N,N’*-dimethylformamide
DMSO		Dimethyl sulfoxide
dppf		1,1′-Ferrocenediyl-bis(diphenylphosphine)
DSF		Differential scanning fluorimetry
DTT		Dithiothreitol
2-EBA		2-Ethynylbenzaldehyde
EC_{50}		Half maximal effective concentration
EDAC/EDC		1-Ethyl-3-(3'-dimethylaminopropyl)carbodiimide hydrochloride
EDTA		Ethylenediaminetetraacetic acid
Equiv.		Equivalents
ESI		Electrospray ionisation
Et		Ethyl
FDC		First derivative of the curve
Fmoc		Fluorenylmethyoxycarbonyl
HASP		Hydrophilic acylated surface protein
HEPES		4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
Hhat		Hedgehog acetyltransferase
hiPSCs		Human induced pluripotent stem cells
His-tag		Six histidine-tag
HOBt		1-Hydroxybenzotriazole hydrate
HPG		Homopropargyglycine
HRMS		High resolution mass spectra
HSA		Human serum albumin
HSAB		Hard-soft acid-base

HSC	Haematopoietic stem cell
Hz	Hertz
IBX	2-Iodoxybenzoic acid
IME	2-Imino-2-methoxyethyl
IR	Infrared
IRIS	Immune reconstitution inflammatory syndrome
K_a	Acid dissociation constant
K_d	Dissociation constant
K_1	Equilibrium constant
k_1	Forwards rate constant
k_{-1}	Reverse rate constant
LC-MS	Liquid chromatography-mass spectrometry
LD	Ligand-directed
LG	Leaving group
M	Molarity M/L
MALDI	Matrix-assisted laser desorption/ionisation
MC	Melting curve value
m-CPBA	*Meta*-chloroperoxybenzoic acid
Me	Methyl
MES	2-(4-Morpholino)ethanesulfonic acid
MMT	L-malic acid, MES and Tris
MOPS	3-(N-morpholino)propanesulfonic acid
m.p.	Melting point
Ms	Mesyl
MWCO	Molecular weight cut-off
m/z	Mass to charge ratio
NASA	*N*-acyl-*N*-alkyl sulfonamide
NHS	*N*-hydroxysuccinimide
NTMT	N-terminal methyltransferase
Ox	Oxazoline
PBS	Phosphate buffered saline
2-PCA	2-Pyridinecarboxaldehyde
4-PCA	4-Pyridinecarboxaldehyde
PCR	Polymerase chain reaction
PDB	RCSB Protein Data Bank
PEG	Polyethylene glycol
Ph	Phenyl
Pi	Phosphate
pI	Isoelectric point

PLB	Passive lysis buffer
PLP	Pyridoxal-5'-phosphate
Ptch	Patched-1 receptor
PTM	Post-translational modification
PVA	Polyvinyl alcohol
rpm	Revolutions per minute
rt	Room temperature
SAH	*S*-adenosylhomocysteine
SAM	*S*-adenosyl-methionine
SCF	Stem-cell factor
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Shh	Sonic Hedgehog protein
SMO	Smoothened receptor
SPAAC	Strain-promoted azide-alkyne cycloaddition
SPPS	Solid-phase peptide synthesis
SrtA	Sortase A
SSG	Succinic acid, sodium phosphate and glycine
TA4C	Triazolecarbaldehyde
^{t}Bu	*tert*-Butyl
TCEP	Tris(2-carboxyethyl)phosphine
TEG	Triethylene glycol
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TIPS	Triisopropylsilane
TLC	Thin layer chromatography
T_m	Melting point
TMS	Trimethylsilyl
TPO	Thrombopoietin
TPPTS	Sodium triphenylphosphine trisulfonate
Ts	Tosyl
UCB	Umbilical cord blood
WT	Wild-type

NMR spectroscopy

app	Apparent
br	Broad
COSY	Correlation Spectroscopy
d	Doublet
HMBC	Heteronuclear Multiple Bond Correlation

HSQC	Heteronuclear Single Quantum Coherence
J	Coupling constant
m	Multiplet
NMR	Nuclear magnetic resonance
Pr	Product
s	Singlet
t	Triplet
q	Quartet

Protein/peptide modifications

A	Aldehyde reagent
CF	Correction factor
d	Double modification
Hy	Hydrate form
Im	Imine form
P	Modified protein
s	Single modification
se	Sextuple modification
SM	Unmodified starting material
sp	Septuple modification
t	Triple modification
q	Quadruple modification
qu	Quintuple modification

Chapter 1

Introduction

Chapter 1: Introduction

Note: Some text from this chapter is based on the following publication: Barber, L. J.; Yates, N. D. J.; Fascione, M. A.; Parkin, A.; Hemsworth, G. R.; Genever, P. G.; Spicer, C. D. *RSC Chem. Biol.* **4**, 56–64 (2023).

1.1 Challenges of Bioconjugation

A chemically-modified protein is a protein where the chemical structure has been altered *via* a chemical reaction. In nature, chemical modifications of proteins are typically enzyme-mediated and occur as co- or post-translational modifications (PTMs), to install moieties including phosphates, sugars, and fatty acids onto protein surfaces. PTMs such as these play an essential role in protein structure and function, expanding the potential molecular diversity of proteins by up to two orders of magnitude.[1] Alternatively, chemical modifications of proteins can be artificially introduced, expanding the structure and function of proteins beyond what can be achieved in nature. Artificially chemically-modified proteins are at the forefront of innovation in biomedicine and biotechnology, with site-selective modification chemistries driving applications in fields as diverse as therapeutics,[2–8] diagnostics,[9–16] catalysis,[17–21] and research applications[8].

A key feature underpinning these technologies is the control of the site of modification, essential to avoid a heterogeneous mixture of conjugates and potential loss of protein function.[22] *Precise* protein targeting is essential, primarily to preserve biological activity, i.e. keeping the active site and protein structure intact.[23] Other advantages of site-selectivity include *i)* enhanced consistency to ensure the uniformity of conjugates, reducing variability in research/clinical applications;[23,24] *ii)* increased accuracy for tracking and visualising imaging tools;[25] and *iii)* enabling complex bioconjugate design.[26] The growth in applications in this field has therefore been limited by the need to develop technologies for *fast*, *selective*, and *stable* bioconjugation.[27] Despite the extensive range of applications of bioconjugation, achieving site-selectivity during bioconjugation remains a complex challenge, requiring exquisite regio- and chemo-selectivity under restrictive conditions.[1]

1.1.1 Chemoselectivity

Protein surfaces are decorated with a wide range of functionalities arising from the side-chains of the 20 canonical amino acids constituting the polypeptide backbone (*Fig. 1.2*). The size, charge, hydrophobicity and reactivity of side-chains contribute to chemical and structural features essential for the structure and function of proteins, including *i)* interactions such as ionic bonds, hydrogen bonds, and disulfide bridges; and *ii)* the localisation of polar side-chains on the outside of protein surfaces to interact with water, with hydrophobic side-chains buried further inside the protein structure to minimise unfavourable interactions, and specific residues located within active sites to give function. This diverse range of functionalities include amines (Lys and N-terminus), amides (Asn, Gln, peptide backbone), carboxylic acids (Asp, Glu, C-terminus), hydroxyls (Ser, Thr), thiols (Cys) and thioethers (Met), in addition to aromatic groups such as phenols (Tyr), indoles

(Trp), and imidazoles (His). The development of bioconjugation strategies targeting only one chemical functionality is therefore critical to avoid a heterogeneous mixture of conjugates and potential loss of protein function (as outlined above).[22] For example, off-target protein modification can impact the efficacy and pharmacokinetics of therapeutics.[27] Differences in pK_a and hard-soft acid-base (HSAB) properties between side-chains of different amino acids can be exploited to achieve chemoselectivity, as will be discussed in Chapter 1.3;[27] however it can be particularly challenging to distinguish side-chains from the same reagent class, with subtle differences in reactivity and pK_a induced by the local environment of different residues (*Fig. 1.1*).[28]

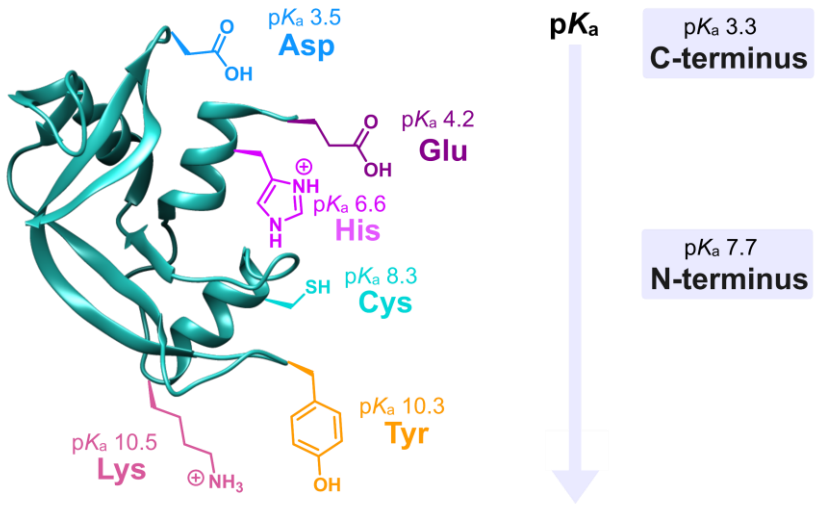


Figure 1.1. Average pK_a values for Asp, Glu, Cys, His, Lys and Tyr side-chains and protein termini reported by Grimsley *et al.*[28] Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[29].

1.1.2 Regioselectivity

As illustrated in *Fig. 1.2*, the availability of amino acid side chains on protein surfaces is influenced by both the overall frequency of the amino acid in the protein sequence (*Fig. 1.2a*),[30] and the level of hydrophobicity/protein folding interactions which influences side-chain surface exposure (*Fig. 1.2b*).[31] Bioconjugation reactions targeting side-chains with a high surface abundance (e.g. Lys) often have poor regioselectivity and lead to heterogeneous mixtures of conjugates, due to the difficulty in targeting a single side-chain in the presence of others. As outlined above, this loss in control of the modification site can potentially cause a loss in protein function through: *i)* disruption of the interactions outlined above essential for protein structure; and *ii)* altering the recognition site of the protein so that binding events critical for protein function cannot occur.[22] Reactions targeting naturally occurring amino acids, most commonly cysteine, are therefore reliant on the presence of a singly reactive residue on the protein surface, in an accessible and suitable position for modification that will not adversely affect protein activity.[32] However, it should be noted that each amino acid side-chain experiences a different local environment within a protein structure, with differences in solvent accessibility and pK_a between copies of the same amino acid throughout a protein.[27]

In unique cases this can allow one residue to be selectively targeted in proteins containing multiple other copies of the same amino acid.[33–36]

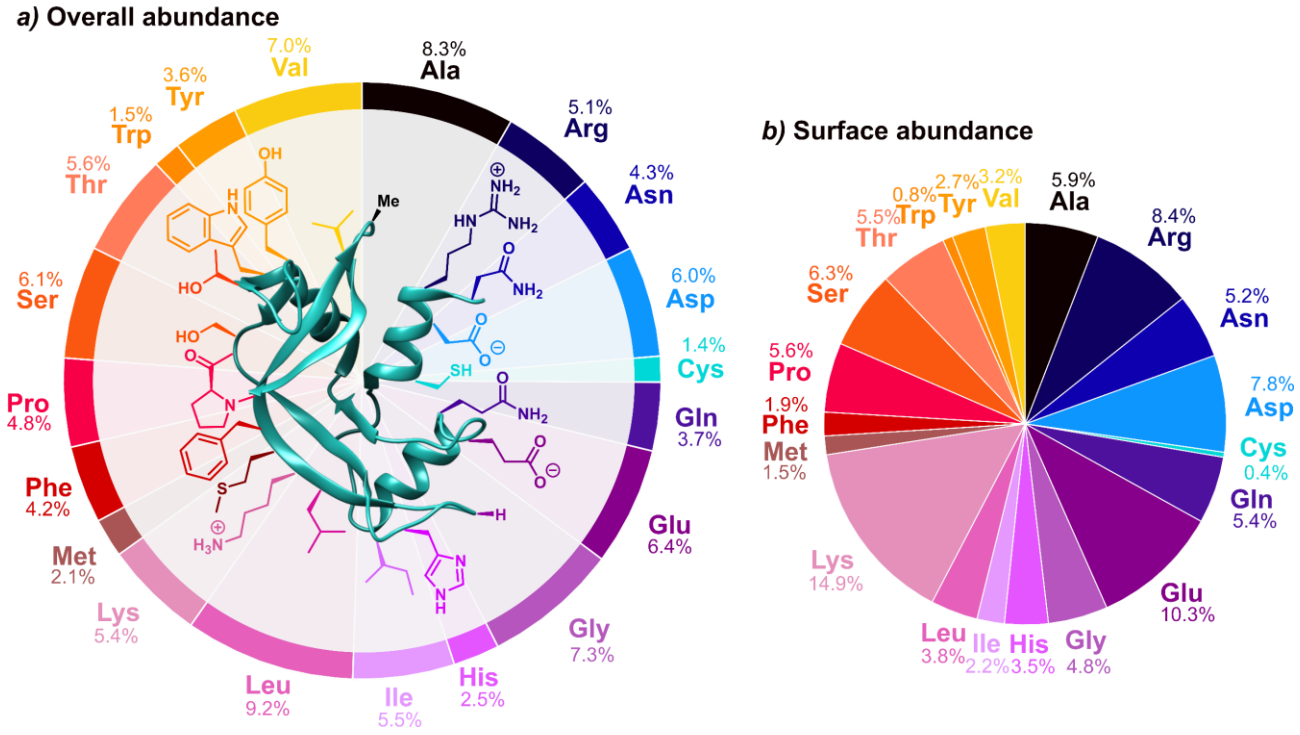


Figure 1.2. ***a)*** Overall frequencies of amino acids in globular proteins determined by Nacar *et al.* from a data-set of 4882 dissimilar protein chains, excluding extremophile proteins;[30] ***b)*** Composition of surface-accessible area reported by Conte *et al.* from a data-set of 23 oligomeric proteins.[31] Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[29].

1.1.3 Restrictive conditions

Critically, bioconjugation strategies must be biocompatible, and applied under aqueous conditions to ensure protein stability and solubility.[27] For most proteins, the use of organic cosolvents can negatively impact protein stability and function,[37] so the reagents applied should also be water-soluble. In cases where organic cosolvents are necessary for reagent solubility, the minimum levels necessary should be used to prevent protein denaturation.[27]

Reactions must be carried out at high dilution due to the propensity of many proteins to aggregate at high protein concentrations. This gives rise to several potential difficulties. Firstly, bioconjugation reactions are often slow, where reactions are typically carried out at concentrations several orders of magnitude more dilute (proteins are often used at nM-μM concentrations) than conventional chemical synthesis (mM-M). Secondly, the reactions require the stability of both reagents and conjugates in a large excess of water.[27]

Proteins have evolved to maintain their native structure and function at physiological salt levels. Additionally, protein water-solubility is determined by the amino acid composition and level of exposure of hydrophobic regions and can be influenced by varying parameters such as the pH of the surrounding environment, due to

the tendency of proteins to be the least soluble at their isoelectric point (pI), or by introducing solubilising additives such as detergents. Reactions must therefore be tolerant of buffers and solubilising additives, often required for protein stability or solubilisation, respectively.[22]

It is also essential that bioconjugation reactions proceed at near physiological pH/temperature to avoid protein denaturation, with near-complete conversion to avoid a heterogeneous mixture of modified and unmodified protein.[1] Conditions often required for reactivity in synthetic organic chemistry such as elevated pressure or temperature, or activation using acids/bases/various catalysts, are therefore not available in bioconjugation technologies.[27]

1.1.4 Properties required for application

Another important consideration for bioconjugation is the resulting conjugate properties required for the given application. For example, conjugates applied *in vivo* must exhibit plasma stability at 37 °C.[22] Antibody-Drug Conjugates (ADCs) are currently the most prominent example of therapeutic bioconjugates,[3] and are a class of pharmaceuticals for targeted therapy, most commonly cancer, which use monoclonal antibodies to deliver cytotoxic agents to tumour sites. ADCs comprise a tumour-targeting antibody chemically linked to a cytotoxic agent *via i)* a cleavable linker, which releases the cytotoxic agent once exposed to environmental differences between blood plasma and tumour cells; or *ii)* a non-cleavable linker, where the cytotoxic agent is released following enzymatic cleavage of the antibody component. These design features allow efficient targeting of tumour sites with low off-target toxicity. The safety of ADCs is influenced by the conjugate stability during circulation of the ADC through the blood: high conjugate stability ensures that the cytotoxic component is released only once the tumour site has been reached, thus minimising side-effects from damage to non-cancerous cells.[2] Whilst high conjugate stability is essential, site-selectivity is not always as necessary in the design of ADCs, which are often heterogeneous mixtures conjugated *via* multiple cysteine residues. However, in design of new ADCs, with >100 currently under development, more precise control of the conjugation site may be important to achieve high efficacy and to preserve protein structure and function.[2]

For other applications, different properties may be critical. For example, rapid labelling of proteolytic fragments with mass-sensitive probes is important during the analysis of complex mixtures by mass spectrometry, so a bioconjugation strategy with fast kinetics would be most suitable in this context.[38] Due to the diversity in the proteome and the variation in properties essential from one application to another, a toolkit of different strategies encompassing a plethora of different reactions and techniques, targeting different amino acids, has become available over the past 30 years.

1.2 Overview of Bioconjugation Strategies

Key covalent bioconjugation strategies include *i)* bioconjugation reaction *via* unnatural amino acid incorporation; *ii)* enzyme-mediated conjugation; or *iii)* chemical reactions targeting naturally occurring motifs. Whilst each method presents its own advantages and suitability for different case studies, in this thesis we have focused on chemical reactions targeting naturally occurring motifs: this is the most widely utilised and most accessible approach and is relevant to our work, so will be discussed in the most detail.

1.2.1 Bio-orthogonal reaction via unnatural amino acid incorporation

Firstly, bioconjugation can be achieved through the site-specific installation of unnatural amino acids bearing bio-orthogonal handles for modification. For example, azide or alkyne handles can be incorporated utilising unnatural amino acids such as azidohomoalanine (AHA) or homopropargyglycine (HPG);[39] site-specific bioconjugation can then be carried out *via* biorthogonal reactions such as traceless Staudinger ligation (*Fig. 1.3a*)[40,41] or copper-catalysed azide-alkyne cycloaddition (CuAAC, i.e., "Click" chemistry; *Fig. 1.3b*, *1.3d*).[42] Alternatively to CuAAC, strain-promoted azide-alkyne cycloaddition (SPAAC), developed by Bertozzi and co-workers, provides a powerful approach for biorthogonal reaction without the use of toxic metal catalysts (*Fig. 1.3c*).[43] In this way, unnatural amino acid incorporation allows site-selective modification through extension of the chemical functionality beyond that offered by canonical amino acids. Whilst the site-specific installation of unnatural amino acids is a powerful technology with extensive applications in biocatalysis,[39] this approach typically requires genetic code expansion, which may not always be suitable or accessible.[44]

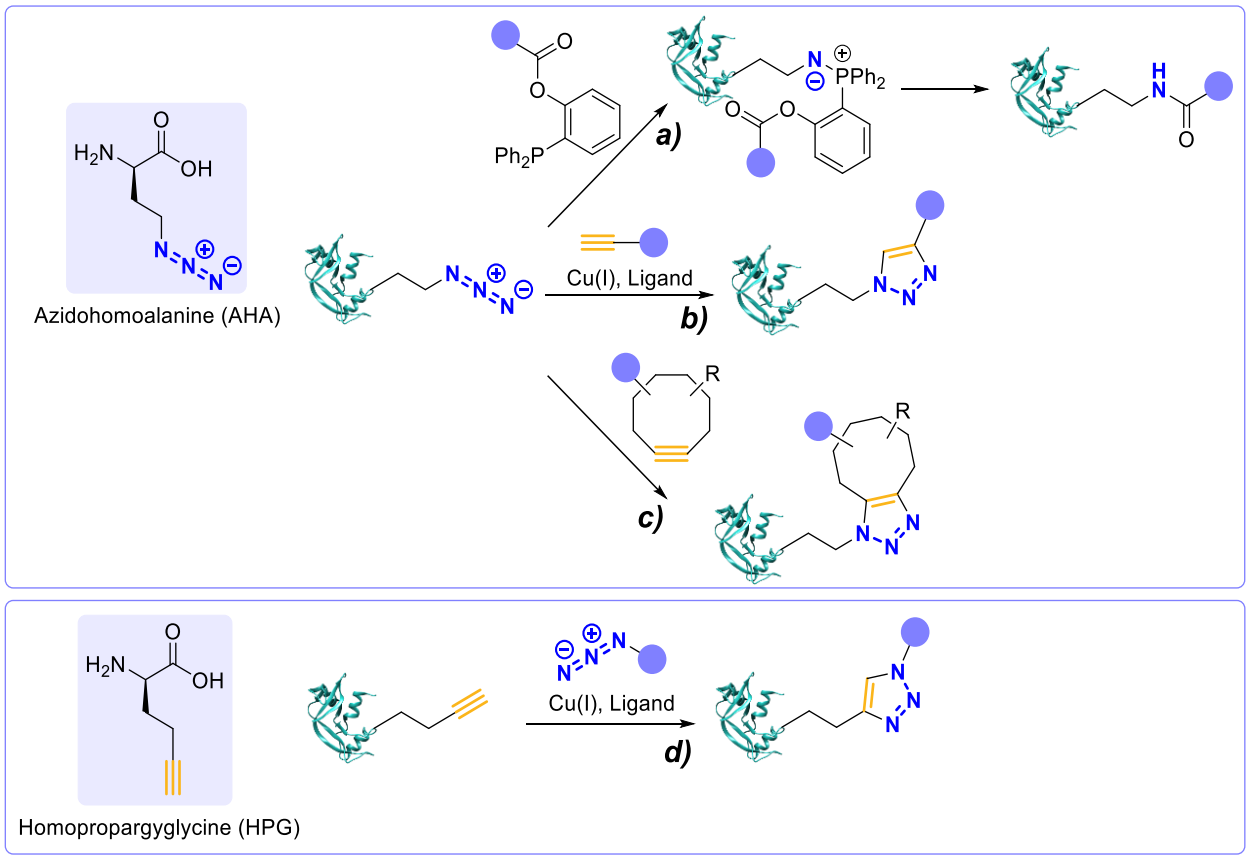


Figure 1.3. Bio-orthogonal reaction of unnatural amino acids (azidohomoalanine and homopropargyglycine) *via* traceless Staudinger ligation or click chemistry.[39–43] Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[29].

1.2.2 Enzyme-mediated conjugation

Alternatively, the inherent specificity of enzymes can be exploited for site-specific covalent bond formation between ligation units, and is most commonly carried out using Sortase A (SrtA).[3] SrtA is a calcium-assisted transpeptidase which anchors surface proteins to peptidoglycan cell walls in bacteria through recognition and ligation of "GGG" and "LPXTG" peptide sequences (where "X" denotes any amino acid). Firstly, SrtA cleaves the peptide bond between threonine (T) and glycine (G) in the "LPXTG" motif *via* nucleophilic attack of the thiol group of Cys184, followed by displacement of the cysteine residue with the N-terminal glycine of peptidoglycan.[45] This enzyme activity can be harnessed for bioconjugation *via* expression of the target protein with: *i)* a C-terminal sortase-tag (LPXTG), followed by ligation to a labelling peptide with an N-terminal glycine (*Fig. 1.4a*); or *ii) vice versa* (*Fig. 1.4b*).[46] Note that C-terminal protein labelling is limited by the requirement for engineering of the LPXTG tag into the protein and low accessibility of C-termini of cell-surface proteins. Contrastingly, N-terminal protein labelling involves minimal engineering to install an N-terminal glycine, and cell-surface protein N-termini are typically extracellular.[45]

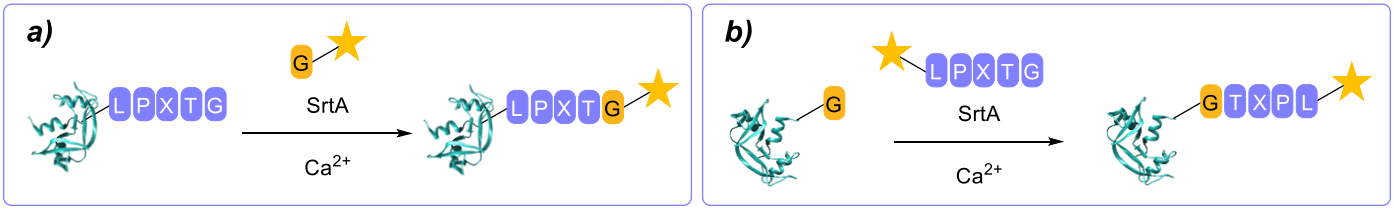


Figure 1.4. Enzyme-mediated conjugation using Sortase A (SrtA) of ***a)*** protein C-termini; and ***b)*** protein N-termini; "X" denotes any amino acid. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[29].

1.2.3 Chemical modification of naturally occurring motifs

Both the unnatural amino acid incorporation and enzyme-mediated conjugation strategies briefly outlined above typically require additional genetic/protein engineering steps to prepare the required protein substrate in addition to the covalent bond formation step. Instead of installing reactive/recognition motifs onto proteins, alternative strategies exploit the reactivity of the natural functionalities present on protein surfaces for chemical conjugation. These approaches fall into two categories: *i) chemoselective* approaches to target individual classes of amino acid residues, of which there may be multiple copies and heterogeneous labelling; and *ii) regioselective* approaches, which target a unique site of the protein. In this second class, proximity-driven strategies have emerged as powerful approaches to label unique sites on a diverse range of protein targets. Proximity-driven strategies will be introduced later in section 3.3, so will not be discussed further here. A summary of common *chemoselective* chemical modification strategies is provided below.

1.3 Chemoselective Chemical Modification Strategies

1.3.1 Traditional strategies

Traditional chemical conjugation strategies target naturally occurring amino acid residues on protein surfaces.[47] These methods classically harness the nucleophilicity of lysine and cysteine residues, as the most nucleophilic amino acids, *via* reaction with suitable electrophiles (*Fig. 1.5*).[8]

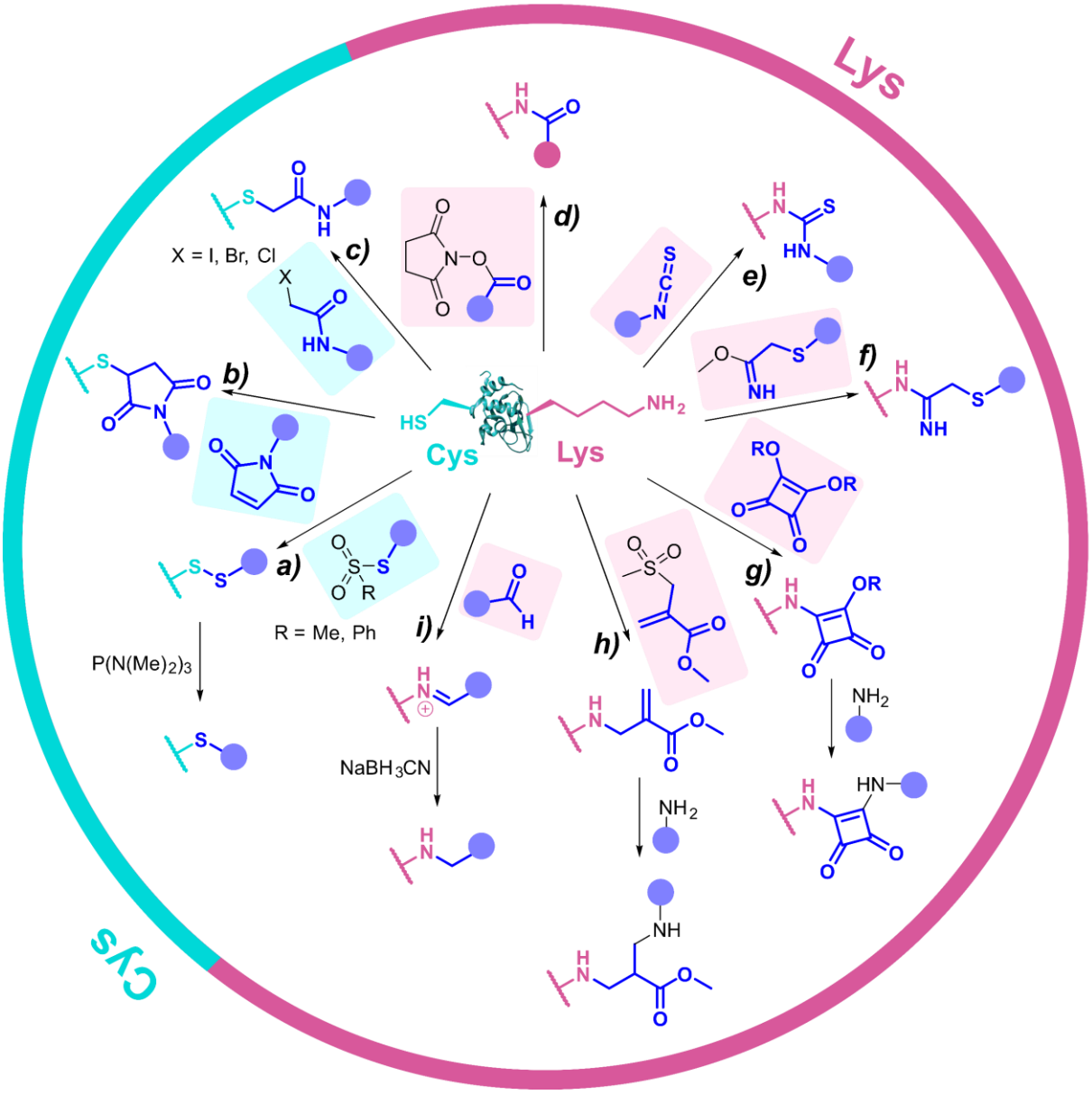


Figure 1.5. Traditional chemical conjugation strategies through reaction of lysine and cysteine residues with ***a)*** sulfones,[48,49] ***b)*** maleimides,[50,51] and ***c)*** halocarbonyls[52,53] for cysteine modification; and ***d)*** *N*-hydroxysuccinimide (NHS) esters,[54] ***e)*** isothiocyanates,[54] ***f)*** 2-imino-2-methoxyethyl (IME) reagents,[55] ***g)*** squaric acid esters,[56] ***h)*** sulfonyl acrylates,[33] and ***i)*** aldehydes[57] for lysine modification (Figure based on review by Naowarojina *et al.*).[39] Figure built using structural data obtained by Peretoltchine *et al.* (lysozyme, PDB 3RZ4; to be published).

1.3.1.1 Cysteine

Cysteine is the most nucleophilic amino acid residue of the 20 canonical amino acids at neutral pH and can be labelled preferentially using a combination of *i)* pH control, and *ii)* hard-soft acid-base (HSAB) theory.

i) pH control. pH control is the selection of a suitable pH at which off-target nucleophiles reside in their protonated form and are therefore less reactive than the target nucleophile. Chemoselective cysteine modification can be achieved under pH control due to the differences in pK_a between sulfhydryl (pK_a 8.3) and ε-ammonium groups (pK_a 10.5) of cysteine and lysine respectively.[22] Cysteine labelling is typically carried out at approx. pH 8;[58] below pH 9, ε-amino groups are protonated so react much slower than cysteine residues.[58] The difference in reactivity between thiol (RSH) and ammonium groups (RNH_3^+) can be further expanded by increasing the pH to a level at which cysteine resides in the thiolate form (RS^-), exploiting the 10-fold increase in nucleophilic reactivity of the thiolate form relative to the thiol form.[59] However, note that the pK_a of cysteine residues in proteins is highly variable depending on its local environment (e.g., the single highly reactive cysteine residue in human serum albumin has a pK_a of 6.7).[22] Attributes of the thiol microenvironment such as close proximity to cationic residues, an N-terminal α-ammonium group, or the π* orbital of a carboxylate residue, can lead to lower pK_a values.[60] These differences in pK_a result in varying levels of the fraction of thiolate form,[60] resulting in differences in reactivity between proteins.[22]

ii) HSAB theory. According to HSAB theory, reaction between "soft" nucleophiles and "soft" electrophiles (i.e., reagents with high polarisability), or "hard" nucleophiles with "hard" electrophiles (i.e., reagents with low polarisability) is kinetically and thermodynamically preferred. Lysine residues are "harder" nucleophiles than cysteine residues, so preferential labelling of lysine or cysteine can be achieved using "harder" or "softer" electrophiles, respectively.[1,22,61] Using the above principles, some common "soft" electrophiles for cysteine labelling include sulfones (*Fig. 1.5a*),[48,49] maleimides (*Fig. 1.5b*),[50,51] and halocarbonyls (*Fig. 1.5c*)[52,53]. Note that in practice, reactivity is not as clear-cut as predicted by HSAB theory and is used as an indication of reactivity only, often in combination with pH control: for example, lysine modification with maleimides can occur at higher pH.[62]

The low abundance of cysteine residues (as illustrated previously in *Fig. 1.2*) makes them suitable targets for site-selective labelling. However, cysteines are rarely present as a free sulfhydryl, commonly playing a critical role in the active site or forming disulfide bridges integral to the tertiary and quaternary structure of the protein.[1,22] Extra manipulation is often required to install/expose free cysteine residues *via* recombinant methods, mutagenesis, or reduction of disulfide bridges.[22] For example, disulfide bond reduction can be carried out using tris(2-carboxyethyl)phosphine (TCEP), followed by re-bridging *via* the introduction of a labelling motif using functionalities such as dibromomaleimides[63] or diiodomaleimides (*Fig. 1.6*).[64]

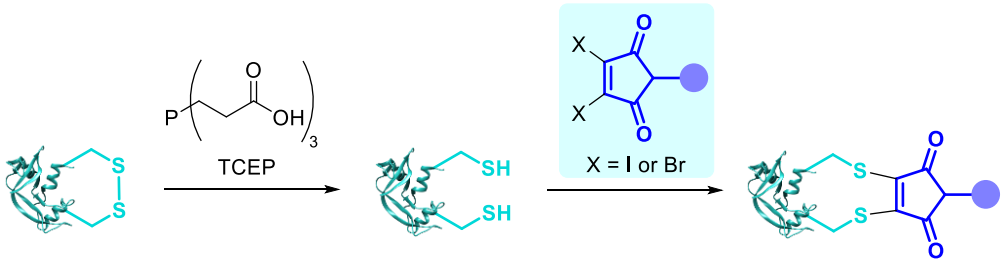


Figure 1.6. Example strategy for *de novo* re-bridging reduced disulfide bonds *via* functionalities such as dibromomaleimides[63] or diiodomaleimides (TCEP = tris(2-carboxyethyl)phosphine).[64] Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[29].

1.3.1.2 Lysine

Although the ε-amino groups of lysine residues are protonated at physiological pH (pK_a 10.5),[22] lysine residues are the second-most reactive nucleophiles out of the 20 canonical amino acids. Preferential targeting over cysteine residues is not always critical due to the rarity of free sulfhydryl groups, but can be achieved using: *i)* "Hard" electrophiles according to HSAB principles, including *N*-hydroxysuccinimide (NHS) esters (*Fig. 1.5d*),[54] isothiocyanates (*Fig. 1.5e*),[54] 2-imino-2-methoxyethyl (IME) reagents (*Fig. 1.5f*),[55] and squaric acid esters (*Fig. 1.5g*);[56] *ii)* acetylation, due to the low stability of intermediates for cysteine acetylation;[22] or *iii)* reductive alkylation *via* imine formation upon reaction with aldehydes in the presence of sodium cyanoborohydride (*Fig. 1.5i*),[57] where double bond formation between the sulfhydryl group and aldehydes cannot occur due to the divalent nature of sulfur.[22]

In contrast to cysteine residues, surface lysine residues are often accessible and highly abundant (as illustrated previously in *Fig. 1.2b*). Lysine targeting is therefore rarely site-selective due to the presence of multiple reactive sites, often resulting in heterogeneous mixtures of many products;[1] additionally, the ε-amino groups of lysine residues (pK_a 10.5) are also in competition with the α-amido amines of protein N-termini (pK_a 7.7).[22]

Despite these challenges, there are some notable examples where targeting of single highly reactive lysine residues has been achieved. The pK_a of lysine residues in proteins is variable depending on their local microenvironment. Matos *et al.* exploited this variation in pK_a and computationally designed sulfonyl acrylates to target a single lysine residue in various proteins; the residue with the lowest pK_a was kinetically favoured at slightly basic pH (*Fig. 1.5h*), providing a site-selective labelling strategy.[33] However, note the utility of this approach depends on the accessibility of the residue with the lowest pK_a,[39] with reagent design and synthesis required on a case-by-case basis. Other notable examples include the targeting of single highly reactive lysine residues in *i)* humanised catalytic antibody h38C2 with heteroaryl methylsulfonyl-functionalised small molecules (*Fig. 1.7a*);[34] *ii)* kappa antibodies with fluorophenyl esters (*Fig. 1.7b*);[35] and *iii)* anti-virulence bacterial target DHQ1 enzyme with hydroxylammonium compounds (*Fig. 1.7c*);[36] but these are all unique exceptions which cannot be universally applied.

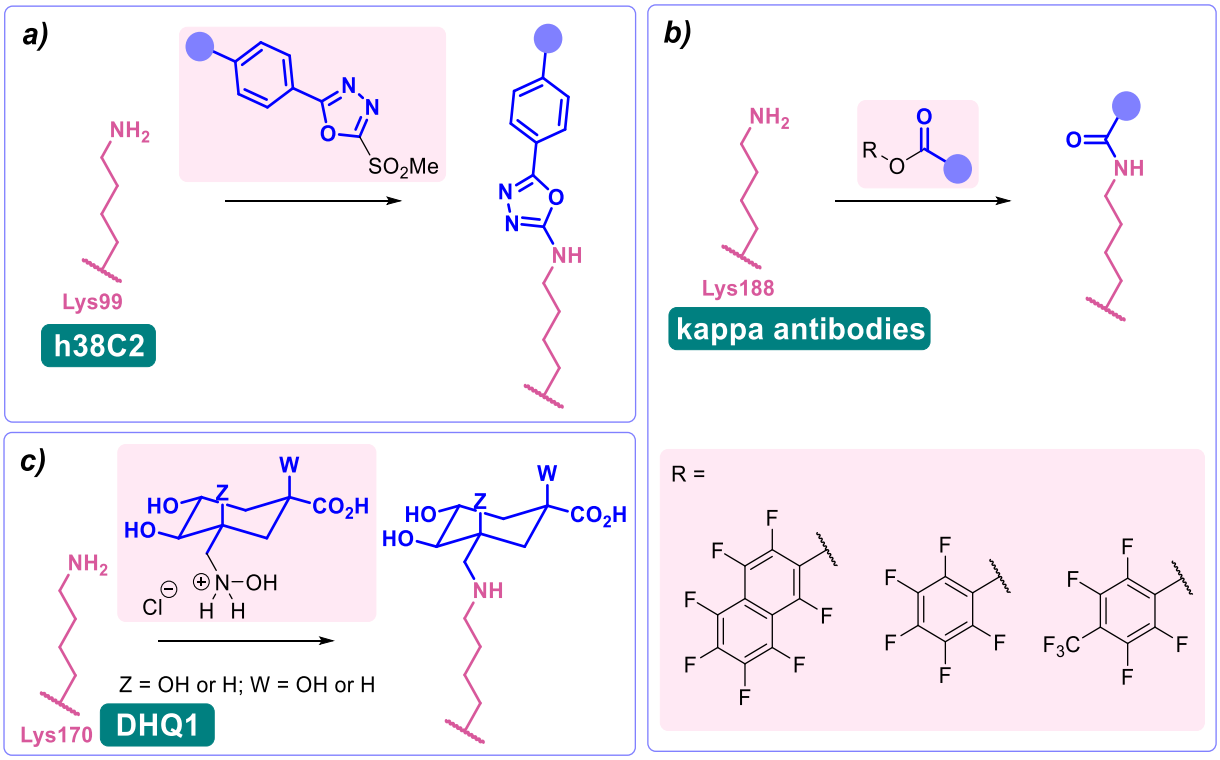


Figure 1.7. Methods targeting single highly reactive lysine residues in ***a)*** humanised catalytic antibody h38C2;[34] ***b)*** kappa antibodies;[35] and ***c)*** anti-virulence bacterial target DHQ1 enzyme.[36]

1.3.2 Hydrophobic residues with low surface abundance

As discussed above, the targeting of unique amino acid residues is key to site-selective labelling. In this context, hydrophobic aromatic residues such as tyrosine and tryptophan are suitable targets due to naturally low surface abundance, although *regioselective* labelling remains challenging in cases where multiple copies of the residues are present.

1.3.2.1 Tryptophan

Tryptophan is the rarest amino acid encountered on protein surfaces and is often buried deep within proteins to stabilise their structure through cation-π, π-π, and hydrophobic interactions. In some cases, tryptophan sites are solvent-exposed and can contribute to protein functions such as signal transduction, offering a site for chemical bioconjugation. However, chemical modification of solvent-accessible tryptophan residues is limited by its weak nucleophilicity at neutral pH.[65] Whilst some tryptophan-targeting reagents have been reported for fluoroalkylation,[66] alkynylation,[67,68] or C-H arylation,[69] application of these methods is hindered by poor chemoselectivity, high levels of required organic co-solvent, acidic conditions, high temperatures, or catalysis with toxic metals unsuitable for conjugate application *in vivo*. In recent developments, these challenges in chemoselectivity and biocompatibility have been overcome through several labelling strategies, including oxidative cyclisation (*Fig. 1.8a*)[65] and radical reactions (*Fig. 1.8b*)[70].

1.3.2.2 Tyrosine

Whilst slightly more surface-abundant than tryptophan residues, few tyrosine residues are typically present on protein surfaces;[71] where multiple copies of tyrosine residues are present, differences in accessibility can sometimes allow *regioselective* targeting.[72] The reactivity of tyrosine side-chains depends on the protonation state.[39] At high pH values close to the pK_a of the tyrosine phenol (pK_a 10.3),[28] tyrosine can be labelled using π-allylpalladium complexes (*Fig. 1.8c*),[73] with chemoselectivity over alcohols due to the higher pK_a values of Ser and Thr (pK_a ~13).[22] At lower pH values, tyrosine can react as a π-nucleophile,[22] with reaction at the aromatic ε-carbon *ortho* to the hydroxyl group *via* diazonium couplings (*Fig. 1.8d*),[74] ene-like click reactions (*Fig. 1.8e*),[75] or Mannich-like reactions (*Fig. 1.8f*)[76].

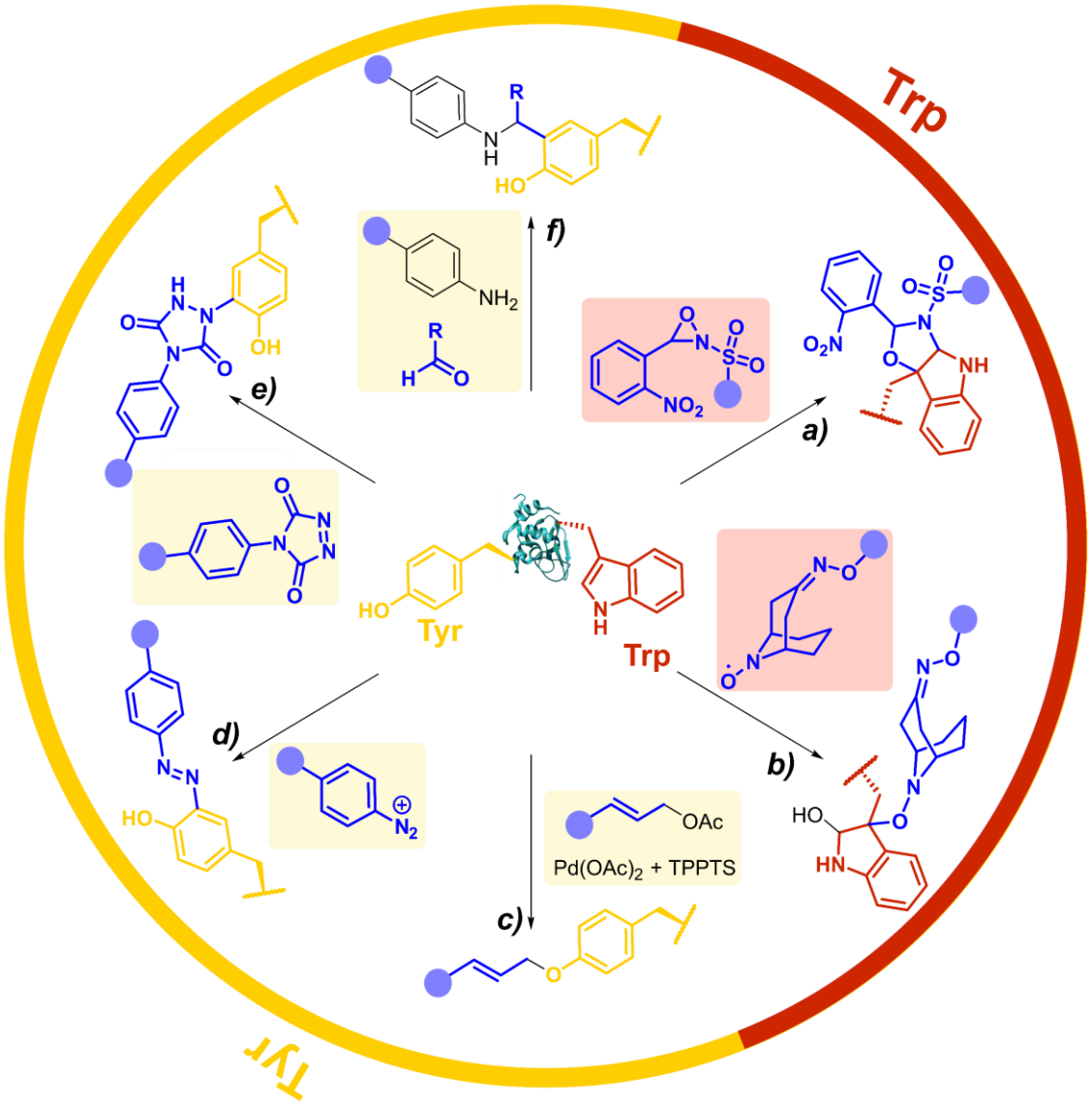


Figure 1.8. Chemical conjugation strategies for hydrophobic residues *via* ***a)*** oxidative cyclisation,[65] or ***b)*** radical reactions[70] for tryptophan modification; and ***c)*** π-allylpalladium complexes,[73] ***d)*** diazonium couplings,[74] ***e)*** ene-like click reactions,[75] or ***f)*** Mannich-like reactions[76] for tyrosine modification (TPPTS = sodium triphenylphosphine trisulfonate). Figure built using structural data obtained by Peretoltchine *et al.* (lysozyme, PDB 3RZ4; to be published).

1.3.3 N-terminal modification

Achieving site-selectivity using the labelling strategies outlined above relies on the presence of a uniquely reactive motif. As discussed previously, whilst this can be achieved by targeting residues with low abundance where only one copy is accessible, or by exploiting subtle differences in pK_a and accessibility between multiple copies of the same amino acid, these approaches are not universal, so applications are limited. In this context, N-terminal targeting has emerged as a powerful approach to label unique sites on a diverse range of protein targets.

The development of chemical reactions that target the N-termini of proteins is particularly attractive for site-selective modification; for single-chain proteins, the N-terminus is a unique chemical motif. Although up to 80-90% of the eukaryotic proteome is N-terminally acetylated,[77] in most bacterial and secreted proteins (e.g. antibodies) the N-terminus is free, and chemically and sterically accessible[78] (see Chapter 3 for further discussion about co- and post-translational N-terminal modifications). Moreover, N-termini are commonly positioned away from active or binding sites, providing a convenient handle to achieve labelling without a cost to bioactivity. In recent years, there has therefore been a surge in interest in the development of N-terminal targeting protein chemistries. Note that technologies for selective C-terminal labelling are much more limited due to poor accessibility of C-termini, with competition between C-termini and highly abundant carboxylic acid residues such as glutamic and aspartic acids.[78,79]

As the only α-amido amine present in native proteins, the N-terminus provides a chemically unique target for chemical modification: *i)* The lower pK_a of the N-terminal ammonium group (~6.0-8.0), relative to the ε-ammonium of lysine (~10.5), marks it as the most nucleophilic amine at physiological pH, allowing selective modification with electrophiles under careful pH control;[78] and *ii)* the α-amide itself can participate in conjugation, providing a unique chemical motif.[80] With the goal of identifying generalisable strategies for N-terminal protein modification, we chose to study only reactions that possess near-universal sequence compatibility in this thesis. Whilst powerful techniques have been developed which take place exclusively at N-terminal cysteines,[81,82] serines,[83] glycines,[84] or biomimetic sequences,[85] these are not as generally translatable so have not been included here (although N-terminal labelling by side chain participation will be discussed later in Chapter 3). Similarly, we have not discussed reactions which modify N-termini preferentially but which also readily modify lysine when > 1 equiv. reagent is used,[86] as well as for which organic co-solvents[87] or metal catalysts[88,89] are necessary.

In this context, the pioneering work of the Francis group on 2-pyridinecarboxaldehydes (2-PCAs, *Fig. 1.9a*) stands out as one of the earliest examples of a reagent that can target all protein N-terminal residues (other than with proline at the second position).[80] Since then, a number of general conjugation strategies have been reported. We classify these into two categories: *i) Selective* strategies, which target the N-terminus through precise control of conditions such as pH and stoichiometry, to minimise side reactions at lysine. These strategies typically rely on the pK_a of the N-terminus; and *ii) Specific* strategies which can *only* take place at the N-terminus, typically due to contributions from the α-amide.

The most classical *selective* reactions rely on benzaldehyde (BA) handles that can reversibly form imines at protein N-termini, and be trapped *via* reductive amination (*Fig. 1.9e*).[90] More recently, 2-ethynylbenzaldehydes (2-EBA) have been shown to similarly target N-termini *via* the formation of isoquinolinium salts following initial imine formation (*Fig. 1.9c*).[91] Selective modification has also been reported using oxazolines (Ox), *via* azolation (*Fig. 1.9d*).[38] However, selectivity is typically highly sensitive to reaction conditions and these strategies often exhibit poor reactivity, where high protein concentrations (> 0.7 mM) were reported for BA and Ox modifications, at which we may expect target proteins to aggregate.[38,90]

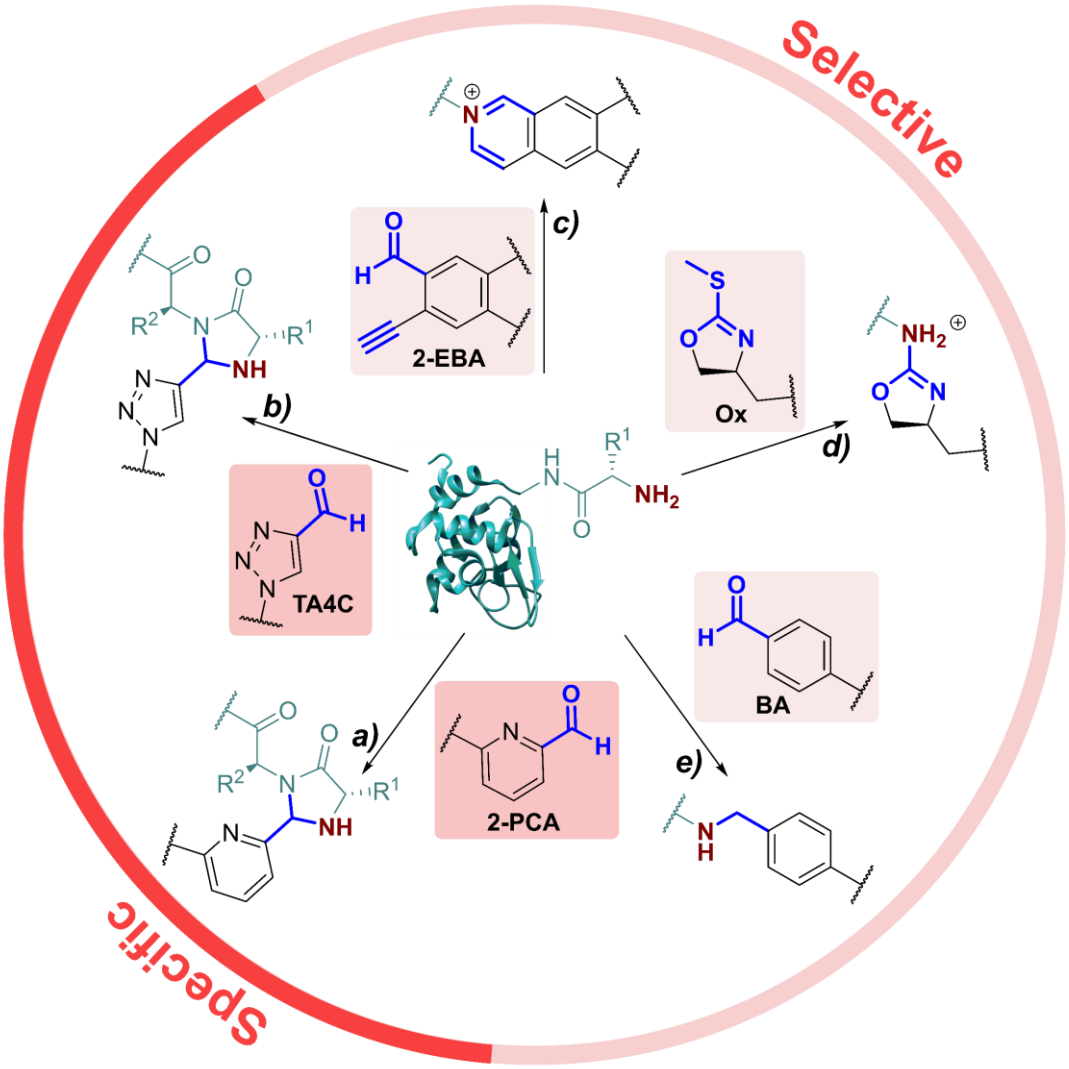


Figure 1.9. Previously reported universal strategies for targeting protein N-termini, which can be split into selective and specific methods.[38,80,90–92] Figure built using structural data obtained by Peretoltchine *et al.* (lysozyme, PDB 3RZ4; to be published).

In contrast, the 2-PCAs developed by the Francis group undergo cyclisation *via* nucleophilic attack of the α-amide at the intermediate imine formed *via* initial condensation at the N-terminus (*Fig. 1.10*).[80] The resulting imidazolidinone can only form at the N-terminus, making these reagents *specific*. Imine formation at the ε-amine of lysine is also possible, but hydrolysis back to the aldehyde is favoured in an aqueous environment.

A similar mode of reactivity has been reported for triazolecarbaldehyde (TA4C) reagents introduced recently by Onoda *et al.* (*Fig. 1.9b*).[92]

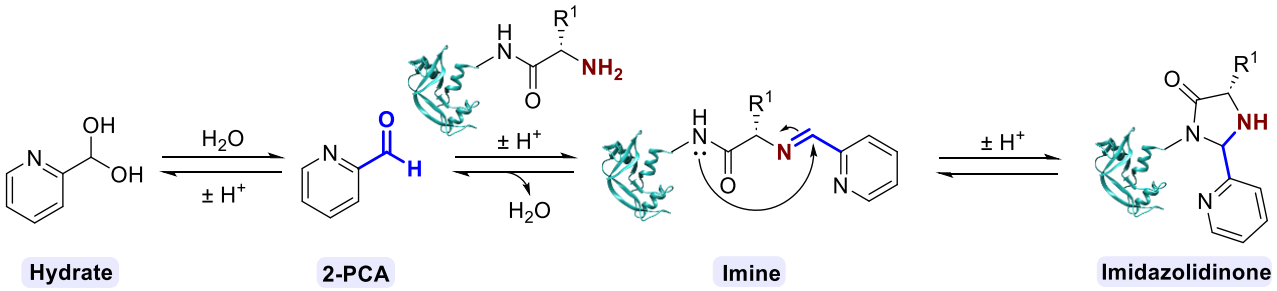


Figure 1.10. Reactions governing reactivity of PCAs with protein N-termini to form imidazolidinone conjugates. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[29].

The design of 2-PCA reagents for N-terminal labeling was inspired by a previous study by Francis and co-workers, in which they examined the suitability of aldehyde reagents for N-terminal transamination.[93] Pyridoxal-5'-phosphate (PLP) is an enzyme cofactor which mediates 4% of all enzymatic reactions.[94] PLP-dependent enzymes typically have a lysine residue in the active site; following initial condensation of the aldehyde moiety of PLP at the ε-ammonium of lysine, a glyoxyl imine is formed upon tautomerisation of the imine.[95] The next step is dictated by interactions of PLP with active site residues, with a wide range of reaction pathways including decarboxylation, racemization, transamination, elimination, and replacement reactions.[94] Inspired by nature, Francis and co-workers identified PLP as a suitable N-terminal transamination reagent compatible with N-terminal valine, glycine, lysine, and methionine residues.[93] In the absence of an enzyme (i.e., during a biomimetic bioconjugation reaction), imine formation is possible upon reaction of PLP with both protein N-termini and lysine residues. However, tautomerisation is only possible for imines generated at the N-terminus due to the lower pK_a of the α-proton of the imine formed at the N-terminus, relative to the α-proton of the imine formed at lysine residues. Subsequent hydrolysis of the glyoxyl imine *via* the transamination pathway yields an N-terminal aldehyde/ketone moiety as a handle for further reactions, such as with alkoxyamine or hydrazine groups (*Fig. 1.11*).[95] Note that this non-enzyme-mediated N-terminal labelling has not been observed *in vivo*, presumably due to low concentrations of free PLP in the cytoplasm.[93]

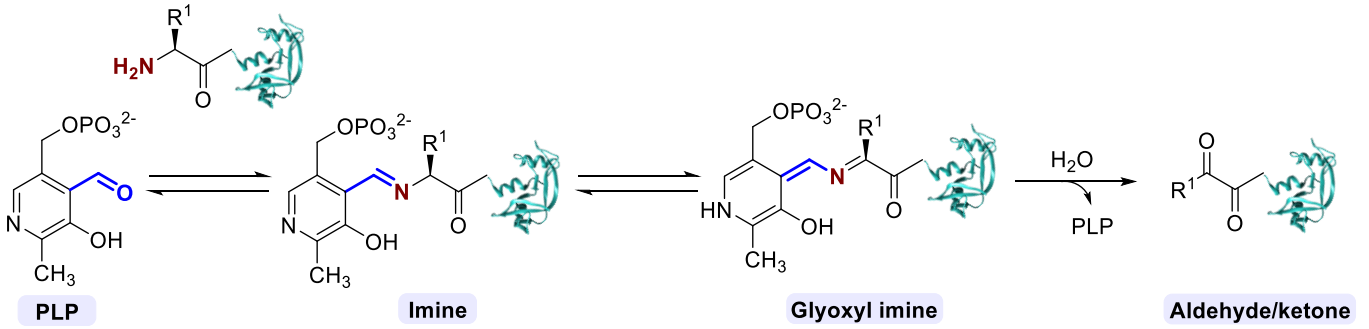


Figure 1.11. Pyridoxal-5'-phosphate (PLP)-catalysed transamination of protein N-termini. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[29].

Notably, the screening of aldehyde reagents in the above study indicated the ability of 2-PCAs to form condensation products with peptide N-termini, with no evidence of the transamination pathway;[93] further peptide studies confirmed high selectivity of 2-PCA reagents for imidazolidinone formation over the transamination pathway.[80] Francis and co-workers explored the reactivity and versatility of N-terminal PCA labelling through probing *i)* the nature of the PCA reagent itself, where unfunctionalised 2-PCA emerged as the most effective reagent in comparison to 4-PCAs, *ortho*-hydroxy functionalized PCAs, and benzaldehydes; and *ii)* the N-terminal amino acid, where conversion ranged from ~50-100% across a panel of peptides with the 20 canonical amino acids as the N-terminal residue. Notably, their study revealed a second addition to the N-terminus for N-terminal glycine, with cyclisation not possible for proteins containing proline in the second position (e.g., asparaginase or bradykinin) due to the absence of an adjacent amide NH group required for cyclisation.[80]

Following initial studies, Francis and co-workers designed 2-PCAs with *meta*-piperazine groups to ensure water-solubility, and to allow ease of functionalisation with a wide range of different functionalities, including polymers.[80] Possible applications of this chemical method have since been explored, for example in drug delivery, where bioconjugation of amphiphilic polymers to protein-based drugs can aid intracellular protein delivery (*Fig. 1.12a*);[96] and tissue engineering, where immobilisation of extracellular matrix proteins onto hydrogels can promote cell adhesion and spreading (*Fig. 1.12b*)[97]. However, we question the suitability of 2-PCA reagents for application of conjugates *in vivo*. Francis and co-workers reported a 20-30% decrease in modified protein upon incubation of imidazolidinone conjugates formed after reaction with 2-PCAs at 37 °C for 12 h.[80] As discussed previously, plasma stability under physiological conditions is critical for conjugates applied *in vivo*, for example where premature release of cytotoxic agents can lead to side-effects and reduced efficacy of ADCs,[2] which typically have a half-life of approx. 2-5 days.[98]

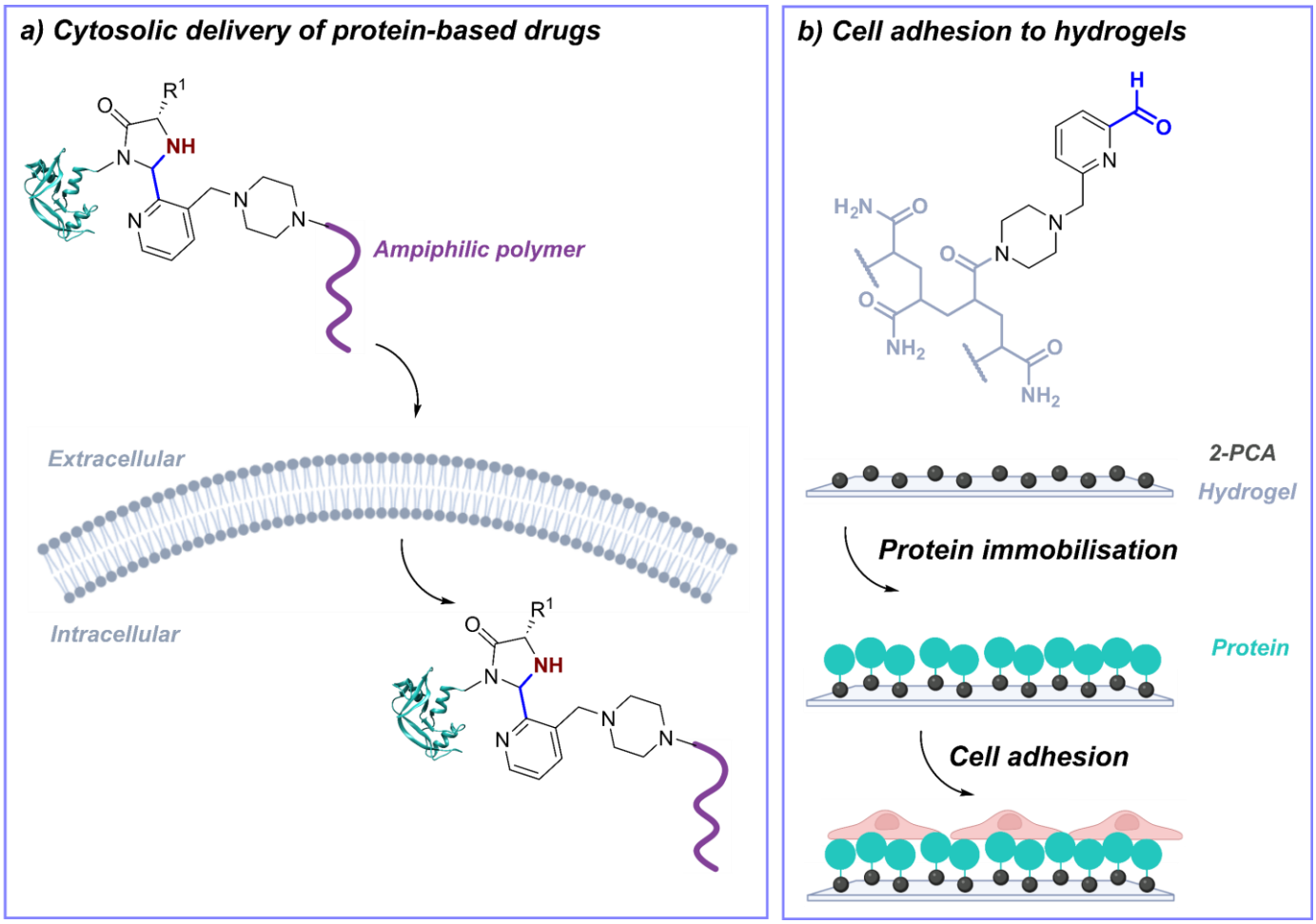


Figure 1.12. Potential applications of 2-PCA bioconjugates for ***a)*** drug delivery,[96] or ***b)*** tissue engineering.[97] Figure created with BioRender.com, using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[29].

1.4 Conclusion and Aims

Advances in recent years have driven the development of a diverse toolkit of bioconjugation strategies applicable to a range of fields. N-terminal labelling is a prominent example of a site-selective strategy applicable to a wide range of protein targets. Despite promising advances in technologies for N-terminal labelling, most reported strategies for N-terminal modification lack universal sequence compatibility or show poor reactivity, selectivity, or conjugate stability. Additionally, differences in target proteins and experimental set-ups make comparisons between different approaches challenging and the impact of these drawbacks on protein modification are therefore not always clear. We have therefore identified a need to elucidate more mechanistic and kinetic insight into the factors governing selectivity, reactivity, and stability of conjugates, in order to develop new modification strategies addressing current limitations in the field.

In this thesis, we initially set out to study the conversion, selectivity, kinetics, and stability of the most promising strategies for generalised N-terminal modification (Chapter 2). Upon identifying key limitations, we next attempted to improve kinetics and stability of current technologies through: *i)* changing the nature of the N-terminus or N-terminal residue (Chapter 3.1 and 3.2); *ii)* harnessing the reactivity of neighbouring residues through the development of proximity-driven chemistries (Chapter 3.3); and *iii)* influencing N-terminal labelling equilibria through reagent functionalisation (Chapter 4). We then applied various bioconjugation strategies to mimic post-translational modifications (PTMs) of an industrially relevant protein, in collaboration with our industrial partners at Qkine (Chapter 5). We hoped to enhance the bioactivity of one of their commercially available proteins, Sonic Hedgehog protein (Shh).

1.5 References

1. Spicer, C. D.; Davis, B. G. *Nat. Commun.* **5**, 4740 (2014).
2. Fu, Z.; Li, S.; Han, S.; Shi, C.; Zhang, Y. *Signal Transduction Targeted Ther.* **7**, 93 (2022).
3. Debon, A.; Siirola, E; Snajdrova, R. *JACS Au* **3**, 1267–1283 (2023).
4. Strosberg, J.; El-Haddad, G.; Wolin, E.; Hendifar, A.; Yao, J.; Chasen, B.; Mittra, E.; Kunz, P. L.; Kulke, M. H.; Jacene, H.; Bushnell, D.; O'Dorisio, T. M.; Baum, R. P.; Kulkarni, H. R.; Caplin, M.; Lebtahi, R.; Hobday, T.; Delpassand, E.; Cutsem, E. V.; Benson, A.; Srirajaskanthan, R.; Pavel, M.; Mora, J.; Berlin, J.; Grande, E.; Reed, N.; Seregni, E.; Öberg, K.; Sierra, M. L.; Santoro, P.; Thevenet, T.; Erion, J. L.; Ruszniewski, P.; Kwekkeboom, D.; Krenning, E. *N. Engl. J. Med.* **376**, 125–135 (2017).
5. Nakase, I.; Ueno, N.; Katayama, M.; Noguchi, K.; Takatani-Nakase, T.; Kobayashi, N. B.; Yoshida, T.; Fujii, I.; Futaki, S. *Chem. Commun.* **53**, 317–320 (2017).
6. Yarong, L.; Fang, Y.; Zhou, Y.; Zandi, E.; Lee, C.; Joo, K.; Wang, P. *Small* **9**, 333–486 (2013).
7. Dovgan, I.; Koniev, O.; Kolodych, S; Wagner, A. *Bioconjugate Chem.* **30**, 2483–2501 (2019).
8. Kalia, J.; Raines, R. *Curr. Org. Chem.* **14**, 138–147 (2010).
9. Billingsley, M. M.; Riley, R. S.; Day, E. S. *PLoS One* **12**, e0177592 (2017).
10. Liu, Q.-L.; Yan, X.-H.; Yin, X.-M.; Situ, B.; Zhou, H.-K.; Lin, L.; Li, B.; Gan, N.; Zheng, L. *Molecules* **18**, 12675–12686 (2013).
11. Ren, X.; Meng, X; Tang, F. *Sens. Actuators, B* **110**, 358–363 (2005).
12. Huang, Y.; Mao, K.; Zhang, B.; Zhao, Y. *Mater. Sci. Eng., C* **70**, 763–771 (2017).
13. Shevtsov, M. A.; Nikolaev, B. P.; Yakovleva, L. Y.; Marchenko, Y. Y.; Dobrodumov, A. V.; Mikhrina, A. L.; Martynova, M. G.; Bystrova, O. A.; Yakovenko, I. V.; Ischenko, A. M. *Int. J. Nanomed.* **9**, 273–287 (2014).
14. Tiefenauer, L. X.; Kuehne, G.; Andres, R. Y. *Bioconjugate Chem.* **4**, 347–352 (1993).
15. Parker, D. *Chem. Soc. Rev.* **19**, 271–291 (1990).
16. Parakh, S.; Lee, S. T.; Gan, H. K.; Scott, A. M. *Cancers* **14**, 1454 (2022).
17. Maghraby, Y. R.; El-Shabasy, R. M.; Ibrahim, A. H.; Azzazy, H. M. E.-S. *ACS Omega* **8**, 5184–5196 (2023).
18. Kobayashi, M.; Nagasawa, T.; Yamada, H. *Trends Biotechnol.* **10**, 402–408 (1992).
19. Basso, A.; Serban, S. *Mol. Catal.* **479**, 110607 (2019).
20. Ribeiro, B. D.; Castro, A. M. D.; Coelho, M. A. Z.; Freire, D. M. G. *Enzyme Res.* **2011**, 615803 (2011).
21. Fernández-Lorente, G.; Terreni, M.; Mateo, C.; Bastida, A.; Fernández-Lafuente, R.; Dalmases, P.; Huguet, J.; Guisán, J. M. *Enzyme Microb. Technol.* **28**, 389–396 (2001).
22. Fischer, N. H.; Oliveira, M. T.; Diness, F. *Biomater. Sci.* **11**, 719–748 (2023).
23. Sadiki, A.; Vaidya, S. R.; Abdollahi, M.; Bhardwaj, G.; Dolan, M. E.; Turna, H.; Arora, V.; Sanjeev, A.; Robinson, T. D.; Koid, A.; Amin, A.; Zhou, Z. S. *Antibody Ther.* **3**, 271–284 (2020).
24. Sarrett, S. M.; Rodriguez, C.; Rymarczyk, G.; Hosny, M. M.; Keinänen, O.; Delaney, S.; Thau, S.; Krantz, B. A.; Zeglis, B. M. *Bioconjugate Chem.* **33**, 1750–1760 (2022).

25. Adumeau, P.; Raavé, R.; Boswinkel, M.; Heskamp, S.; Wessels, H. J. C. T.; Gool, A. J.; Moreau, M.; Bernhard, C.; Costa, L. D.; Goncalves, V.; Denat, F. *Bioconjugate Chem.* **33**, 530–540 (2022).
26. Burrow, N. K.; Gourdie, R. K.; King, E. A.; Travis, C. R.; Goff, C. M.; Nimmo, Z. M.; Berg, J. M.; Boyt, E. L.; Young, D. D. *ChemBioChem* **24**, e202300565 (2023).
27. Reddy, N. C.; Kumar, M.; Molla, R.; Rai, V. *Org. Biomol. Chem.* **18**, 4669–4691 (2020).
28. Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. *Protein Sci.* **18**, 247–251 (2009).
29. Chatani, E.; Hayashi, R.; Moriyama, H.; Ueki, T. *Protein Sci.* **11**, 72–81 (2002).
30. Nacar, C. *Clin. Exp. Health Sci.* **13**, 261–266 (2023).
31. Conte, L. L.; Chothia, C.; Janin, J. *J. Mol. Biol.* **285**, 2177–2198 (1999).
32. De Rosa, L.; Di Stasi, R.; Romanelli, A.; D'Andrea, L. D. *Molecules* **26**, 3521 (2021).
33. Matos, M. J.; Oliveira, B. L.; Martínez-Sáez, N.; Guerreiro, A.; Cal, P. M. S. D.; Bertoldo, J.; Maneiro, M.; Perkins, E.; Howard, J.; Deery, M. J.; Chalker, J. M.; Corzana, F.; Jiménez-Osés, G.; Bernardes, G. J. L. *J. Am. Chem. Soc.* **140**, 4004–4017 (2018).
34. Hwang, D.; Tsuji, K.; Park, H.; Burke, T. R.; Rader, C. *Bioconjugate Chem.* **30**, 2889–2896 (2019).
35. Pham, G. H.; Ou, W.; Bursulaya, B.; DiDonato, M.; Herath, A.; Jin, Y.; Hao, X.; Loren, J.; Spraggon, G.; Brock, A.; Uno, T.; Geierstanger, B. H.; Cellitti, S. E. *ChemBioChem* **19**, 799–804 (2018).
36. Maneiro, M.; Lence, E.; Sanz-Gaitero, M.; Otero, J. M.; Raaij, M. J.; Thompson, P.; Hawkins, A. R.; González-Bello, C. *Org. Chem. Front.* **6**, 3127–3135 (2019).
37. Vegt, N. F. A.; Nayar, D. *J. Phys. Chem. B* **121**, 9986–9998 (2017).
38. Tang, K. C.; Raj, M. *Angew. Chem., Int. Ed.* **60**, 1797–1805 (2021).
39. Naowarojna, N.; Cheng, R.; Lopez, J.; Wong, C.; Qiao, L.; Liu, P. *Synth. Syst. Biotechnol.* **6**, 32–49 (2021).
40. Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. *Org. Lett.* **2**, 1939–1941 (2000).
41. Saxon, E.; Armstrong, J. I.; Bertozzi, C. R. *Org. Lett.* **2**, 2141–2143 (2000).
42. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **41**, 2423–2618 (2002).
43. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **126**, 15046–15047 (2004).
44. Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. *Chem. Sci.* **6**, 50–69 (2015).
45. Morgan, H. E.; Turnbull, W. B.; Webb, M. E. *Chem. Soc. Rev.* **51**, 4121–4145 (2022).
46. Warden-Rothman, R.; Caturegli, I.; Popik, V.; Tsourkas, A. *Anal. Chem.* **85**, 11090–11097 (2013).
47. Spicer, C. D.; Pashuck, E. T.; Stevens, M. M. *Chem. Rev.* **118**, 7702–7743 (2018).
48. Gamblin, D. P.; Garnier, P.; Ward, S. J.; Oldham, N. J.; Fairbanks, A. J.; Davis, B. G. *Org. Biomol. Chem.* **1**, 3642–3644 (2003).
49. Bernardes, G. J. L.; Grayson, E. J.; Thompson, S.; Chalker, J. M.; Errey, J. C.; Oualid, F. E.; Claridge, T. D. W.; Davis, B. G. *Angew. Chem., Int. Ed.* **47**, 2153–2317 (2008).
50. Moore, J. E.; Ward, W. H. *J. Am. Chem. Soc.* **78**, 2414–2418 (1956).
51. Tsao, T.-C.; Bailey, K. *Biochim. Biophys. Acta* **11**, 102–113 (1953).
52. Goddard, D. R.; Michaelis, L. *J. Biol. Chem.* **112**, 361–371 (1935).

53. Lundell, N.; Schreitmüller, T. *Anal. Biochem.* **266**, 31–47 (1999).
54. Brinkley, M. *Bioconjugate Chem.* **3**, 2–13 (1992).
55. Robinson, M. A.; Charlton, S. T.; Garnier, P.; Davis, B. G. *Proc. Natl. Acad. Sci.* **101**, 14527–14532 (2004).
56. Wurm, F.; Steinbach, T.; Klok, H.-A. *Chem. Commun.* **49**, 7815 (2013).
57. Gildersleeve, J. C.; Oyelaran, O.; Simpson, J. T.; Allred, B. *Bioconjugate Chem.* **19**, 1485–1490 (2008).
58. Baslé, E.; Joubert, N.; Pucheault, M. *Chem. Biol.* **17**, 213–227 (2010).
59. Bulaj, G.; Kortemme, T.; Goldenberg, D. P. *Biochemistry* **37**, 8965–8972 (1998).
60. Boll, L. B.; Raines, R. T. *ChemBioChem* **23**, e202200258 (2022).
61. LoPachin, R. M.; Geohagen, B. C.; Nordstroem, L. U. *Toxicology* **418**, 62–69 (2019).
62. Ravasco, J. M. J. M.; Faustino, H.; Trindade, A.; Gois, P. M. P. *Chem. - Eur. J.* **25**, 43–59 (2019).
63. Smith, M. E. B.; Schumacher, F. F.; Ryan, C. P.; Tedaldi, L. M.; Papaioannou, D.; Waksman, G.; Caddick, S.; Baker, J. R. *J. Am. Chem. Soc.* **132**, 1960–1965 (2010).
64. Forte, N.; Livanos, M.; Miranda, E.; Morais, M.; Yang, X.; Rajkumar, V. S.; Chester, K. A.; Chudasama, V.; Baker, J. R. *Bioconjugate Chem.* **29**, 486–492 (2018).
65. Xie, X.; Moon, P. J.; Crossley, S. W. M.; Bischoff, A. J.; Li, D. H. G.; Dao, N.; Gonzalez-Valero, A.; Reeves, A. G.; McKenna, J. M.; Elledge, S. K.; Wells, J. A.; Toste, F. D.; Chang, C. J. *Nature* **627**, 680–687 (2024).
66. Rahimidashaghoul, K.; Klimánková, I.; Hubálek, M.; Korecký, M.; Chvojka, M.; Pokorný, D.; Matoušek, V.; Fojtík, L.; Kavan, D.; Kukačka, Z.; Novák, P. *Chem. - Eur. J.* **25**, 15779–15785 (2019).
67. Hansen, M. B.; Hubalek, F.; Skrydstrup, T.; Hoeg-Jensen, T. *Chem. - Eur. J.* **22**, 1572–1576 (2016).
68. Ruan, Z.; Sauermann, N.; Manomi, E.; Ackermann, L. *Angew. Chem., Int. Ed.* **56**, 3172–3176 (2017).
69. Schischko, A.; Ren, H.; Kaplaneris, N.; Ackermann, L. *Angew. Chem., Int. Ed.* **56**, 1576–1580 (2017).
70. Seki, Y.; Ishiyama, T.; Sasaki, D.; Abe, J.; Sohma, Y.; Oisaki, K.; Kanai, M. *J. Am. Chem. Soc.* **138**, 10798–10801 (2016).
71. Moelbert, S.; Emberly, E.; Tang, C. *Protein Sci.* **13**, 567–841 (2004).
72. Sato, S.; Matsumura, M.; Kadonosono, T.; Abe, S.; Ueno, T.; Ueda, H.; Nakumara, H. *Bioconjugate Chem.* **31**, 1417–1424 (2020).
73. Chen, S.; Li, X.; Ma, H. *ChemBioChem* **10**, 1200–1207 (2009).
74. Jones, M. W.; Mantovani, G.; Blindauer, C. A.; Ryan, S. M.; Wang, X.; Brayden, D. J.; Haddleton, D. M. *J. Am. Chem. Soc.* **134**, 7406–7413 (2012).
75. Ban, H.; Gavrilyuk, J.; Barbas, C. F. *J. Am. Chem. Soc.* **132**, 1523–1525 (2010).
76. Joshi, N. S.; Whitaker, L. R.; Francis, M. B. *J. Am. Chem. Soc.* **126**, 15942–15943 (2004).
77. Polevoda, B.; Sherman, F. *J. Mol. Biol.* **325**, 595–622 (2003).
78. Rosen, C. B.; Francis, M. B. *Nat. Chem. Biol.* **13**, 697–705 (2017).
79. Bloom, S.; Liu, C.; Kölmel, D. K.; Qiao, J. X.; Zhang, Y.; Poss, M. A.; Ewing, W. R.; MacMillan, D. W. C. *Nat. Chem.* **10**, 205–211 (2018).
80. MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. *Nat. Chem. Biol.* **11**, 326–331 (2015).
81. Zhang, L.; Tam, J. P. *Anal. Biochem.* **233**, 87–93 (1996).
82. Bandyopadhyay, A.; Cambray, S.; Gao, J. *Chem. Sci.* **7**, 4589–4593 (2016).

83. Spears, R. J.; Brabham, R. L.; Budhadev, D.; Keenan, T.; McKenna, S.; Walton, J.; Brannigan, J. A.; Brzozowski, M.; Wilkinson, A. J.; Plevin, M.; Fascione, M. A. *Chem. Sci.* **9**, 2927–2933 (2018).
84. Purushottam, L.; Adusumalli, S. R.; Singh, U.; Unnikrishnan, V. B.; Rawale, D. G.; Gujrati, M.; Mishra, R. K.; Rai. V. *Nat. Commun.* **10**, 2539 (2019).
85. Witus, L. S.; Moore, T.; Thuronyi, B. W.; Esser-Kahn, A. P.; Scheck, R. A.; Lavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc.* **132**, 16812–16817 (2010).
86. Singudas, R.; Adusumalli, S. R.; Joshi, P. N.; Rai, V. *Chem. Commun.* **51**, 473–476 (2015).
87. Raj, M.; Wu, H.; Blosser, S. L.; Vittoria, M. A.; Arora, P. S. *J. Am. Chem. Soc.* 150512174012005 (2015).
88. Lin, Z.; Liu. B.; Lu, M.; Wang, Y.; Ren, X.; Liu, Z.; Luo, C.; Shi, W.; Zou, X.; Song, X.; Tang, F.; Huang, H.; Huang, W. *J. Am. Chem. Soc.* **146**, 23752–23763 (2024).
89. Hanaya, K.; Yamoto, K.; Taguchi, K.; Matsumoto, K.; Higashibayashi, S.; Sugai, T. *Chem. - Eur. J.* **28**, e202201677 (2022).
90. Chen, D.; Disotuar, M. M.; Xiong, X.; Wang, Y.; Chou, D. H.-C. *Chem. Sci.* **8**, 2717–2722 (2017).
91. Deng, J.-R.; Lai, N. C.-H.; Kung, K. K.-Y.; Yang, B.; Chung, S.-F.; Leung, A. S.-L.; Choi, M.-C.; Leung, Y.-C.; Wong, M.-K. *Commun. Chem.* **3**, 67 (2020).
92. Onoda, A.; Inoue, N.; Sumiyoshi, E.; Hayashi, T. *ChemBioChem* **21**, 1274–1278 (2020).
93. Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. *Angew. Chem., Int. Ed.* **45**, 5307–5311 (2006).
94. Liang, J.; Han, Q.; Tan, Y.; Ding, H.; Li, J. *Front. Mol. Biosci.* **6**, 4 (2019).
95. Tantipanjaporn, A.; Wong, M.-K. *Molecules* **28**, 1083 (2023).
96. Sangsuwan, R.; Tachachartvanich, P.; Francis, M. B. *J. Am. Chem. Soc.* **141**, 2376–2383 (2019).
97. Lee, J. P.; Kassianidou, E.; MacDonald, J. I.; Francis, M. B.; Kumar, S. *Biomaterials* **102**, 268–276 (2016).
98. Hinrichs, M. J. M.; Dixit, R. *AAPS J.* **17**, 1055–1064 (2015).

Chapter 2

Investigation of leading N-terminal targeting protein modification chemistries

Chapter 2: Investigation of leading N-terminal targeting protein modification chemistries

Note: This chapter is based on the following publication: Barber, L. J.; Yates, N. D. J.; Fascione, M. A.; Parkin, A.; Hemsworth, G. R.; Genever, P. G.; Spicer, C. D. *RSC Chem. Biol.* **4**, 56–64 (2023). Where experiments were performed by other researchers, this is clearly stated in the text.

2.1 Introduction

In this chapter, we carried out a detailed comparative study of the most promising strategies for generalised N-terminal chemical modification, with the aim of elucidating more mechanistic and kinetic insight into the factors governing selectivity, reactivity, and stability (*Fig. 2.1*). By doing so, we aimed to provide valuable insight into the appropriate N-terminal targeting chemistry for a given protein/application, and a platform to develop new modification strategies which address current limitations in the field.

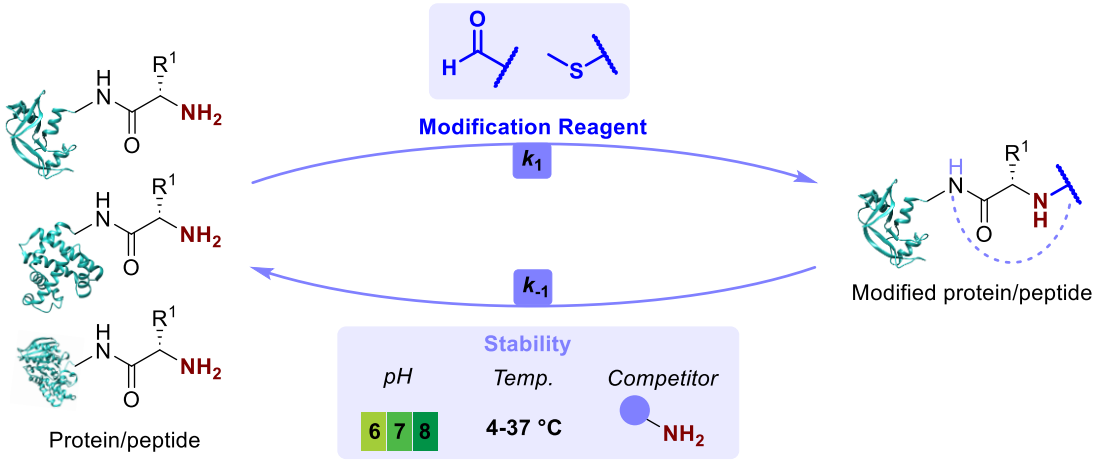


Figure 2.1. Overview of our comparative study of leading strategies for generalised N-terminal chemical modification of proteins.

2.2 Results and Discussion

2.2.1 Study design

Based on prior reports discussed in Chapter 1, we synthesised a panel of reagents, **2.1**-**2.7** which encompassed the key reagent classes for N-terminal protein labelling (*Fig. 2.2*). Reagents **2.1**-**2.3** are based on the 2-pyridinecarboxaldehydes (2-PCAs) reported by the Francis group,[1] with 6-amino-2-PCAs (**2.2**) having been reported to increase reactivity relative to 6-methylamino derivatives (**2.3**).[2] We also considered the synthetic accessibility of 6-methyloxy analogue **2.1**, which can be synthesised directly from the corresponding alcohol *via* an etherification reaction (*Fig. 2.3d*). Triazolecarbaldehyde (TA4C) **2.4**, 2-ethynylbenzaldehyde (2-EBA) **2.5**, oxazoline (Ox) **2.6**, and benzaldehyde (BA) **2.7** were similarly synthesised based on the most reactive structures reported in the relevant papers.[3–6] A triethylene glycol monomethyl

ether side chain was incorporated into each reagent to ensure water-solubility and eliminate any differences in reactivity that might arise from the side chain.

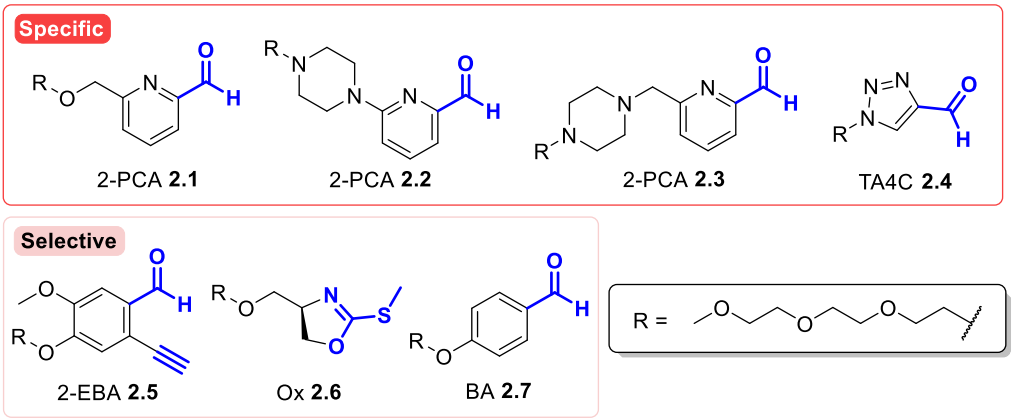


Figure 2.2. Structures of reagents **2.1**-**2.7**, for N-terminal protein modification.

2.2.2 Reagent synthesis

*i) 2-PCA **2.1***

To prepare 2-PCA **2.1**, we first tosylated triethylene glycol monomethyl ether following the procedure reported by Kohmoto *et al.* to afford compound **2.8** (85%).[7] We then used selenium dioxide to convert 2,6-pyridinedimethanol to mono-oxidised species **2.9** (64%) following the procedure reported by MacDonald *et al.* (*Fig. 2.3a*).[1] Next, we carried out the *O*-alkylation of 2-PCA **2.9** with tosylate **2.8** and NaH. Unexpectedly, the aldehyde group was reduced to an alcohol and subsequently *O*-alkylated, resulting in the formation of dialkylated species **2.11** (17%). We suspected the aldehyde reduction to be due to the presence of a hydride source (NaH) so the reaction was repeated using potassium *tert*-butoxide as an alternative base (*Fig. 2.3b*). Interestingly, aldehyde reduction to alcohol species **2.10** (10%) was observed even in the absence of hydride. To avoid aldehyde reduction and alkylation, we next attempted to prepare 2-PCA **2.1** *via* reversed reactivity; compound **2.9** was mesylated to compound **2.12** following the procedure reported by MacDonald *et al.*,[1] and then used directly in the reaction with triethylene glycol monomethyl ether (**2.13**) under basic conditions (*Fig. 2.3c*). Again, no product was observed, and mass spectrometry indicated the presence of reduced species **2.10**. We eventually overcame the aldehyde reduction issue *via* protection of the aldehyde as acetal **2.14** (72%), followed by *O*-alkylation to acetal **2.15** (79%), and then deprotection of the acetal to afford the target compound (61%, *Fig. 2.3d*).

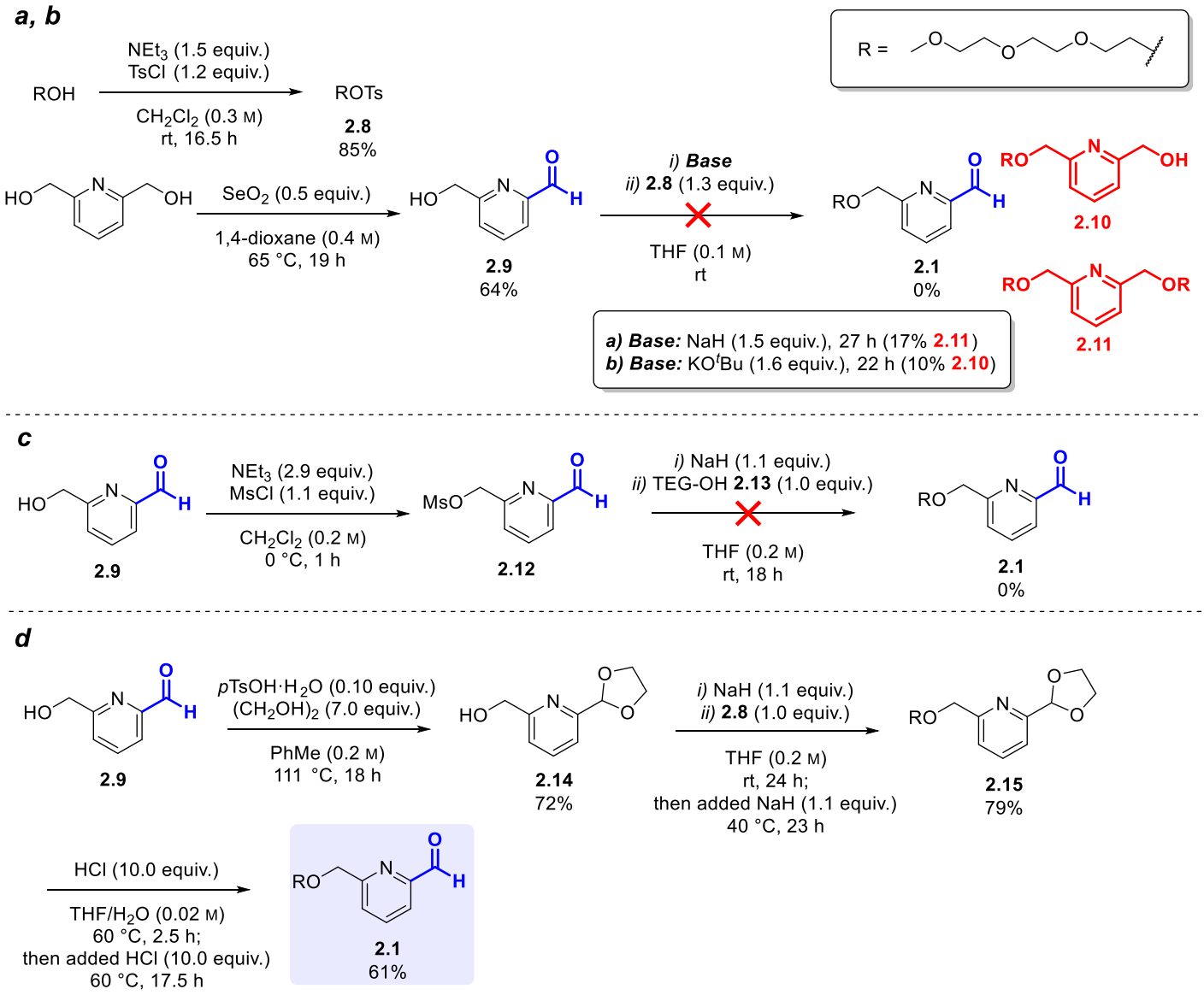


Figure 2.3. Synthetic strategies for the preparation of 2-PCA **2.1**.

*ii) 2-PCA **2.2***

To prepare 2-PCA **2.2**, we first carried out the *Boc*-protection of piperazine following a procedure reported by Bala *et al.* to afford piperazine **2.16** (79%, *Fig. 2.4*),[8] and 6-bromo-2-pyridinecarboxaldehyde was protected as acetal **2.17** following a procedure previously reported by Wang *et al.* (94%).[9] 2-PCA **2.19** was then prepared *via* a procedure adapted from Koo *et al.*,[2] where palladium catalysed cross-coupling of acetal **2.17** with piperazine **2.16** afforded acetal **2.18** (85%), followed by *Boc* and acetal deprotection to afford 2-PCA **2.19** (71%). Subsequent *N*-alkylation afforded compound **2.20** (52%). Conversion of the aldehyde moiety to the dimethyl acetal analogue unexpectedly proceeded readily during flash chromatography (10% MeOH:CH_2Cl_2), so a final deprotection step was required to obtain 2-PCA **2.2** as an aldehyde species (51%).

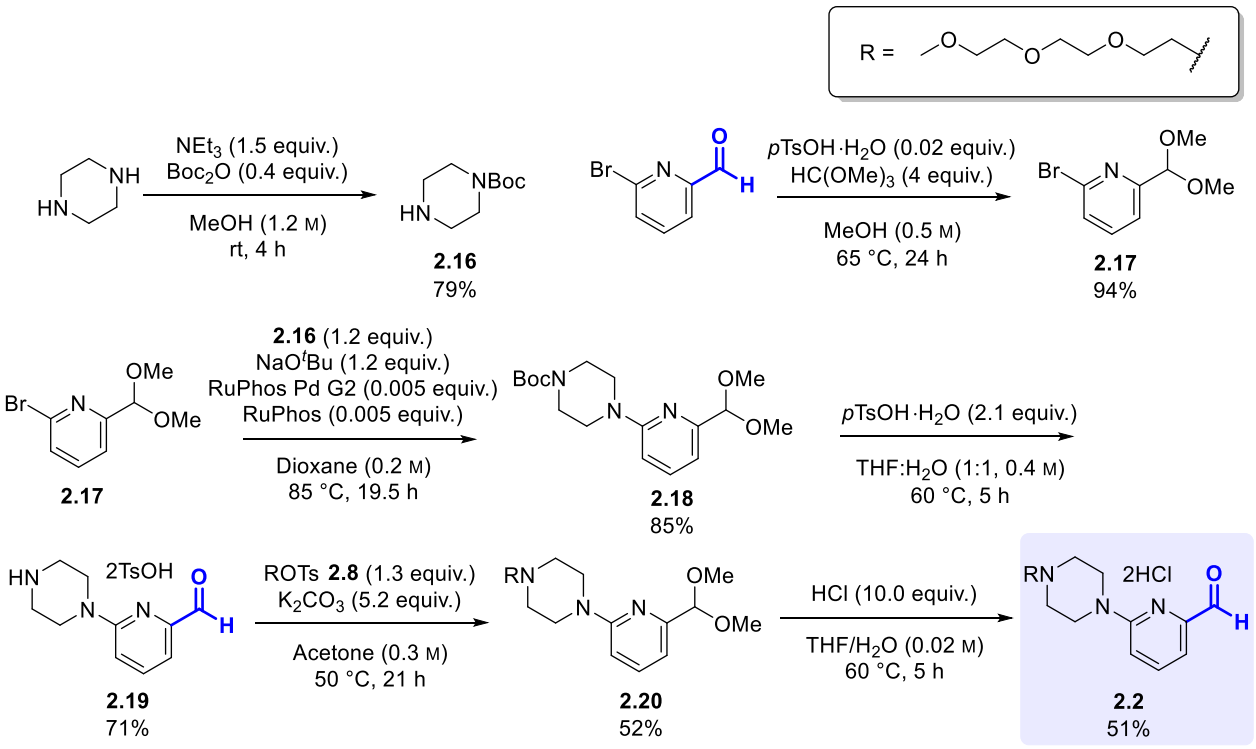


Figure 2.4. Synthesis of 2-PCA **2.2**.

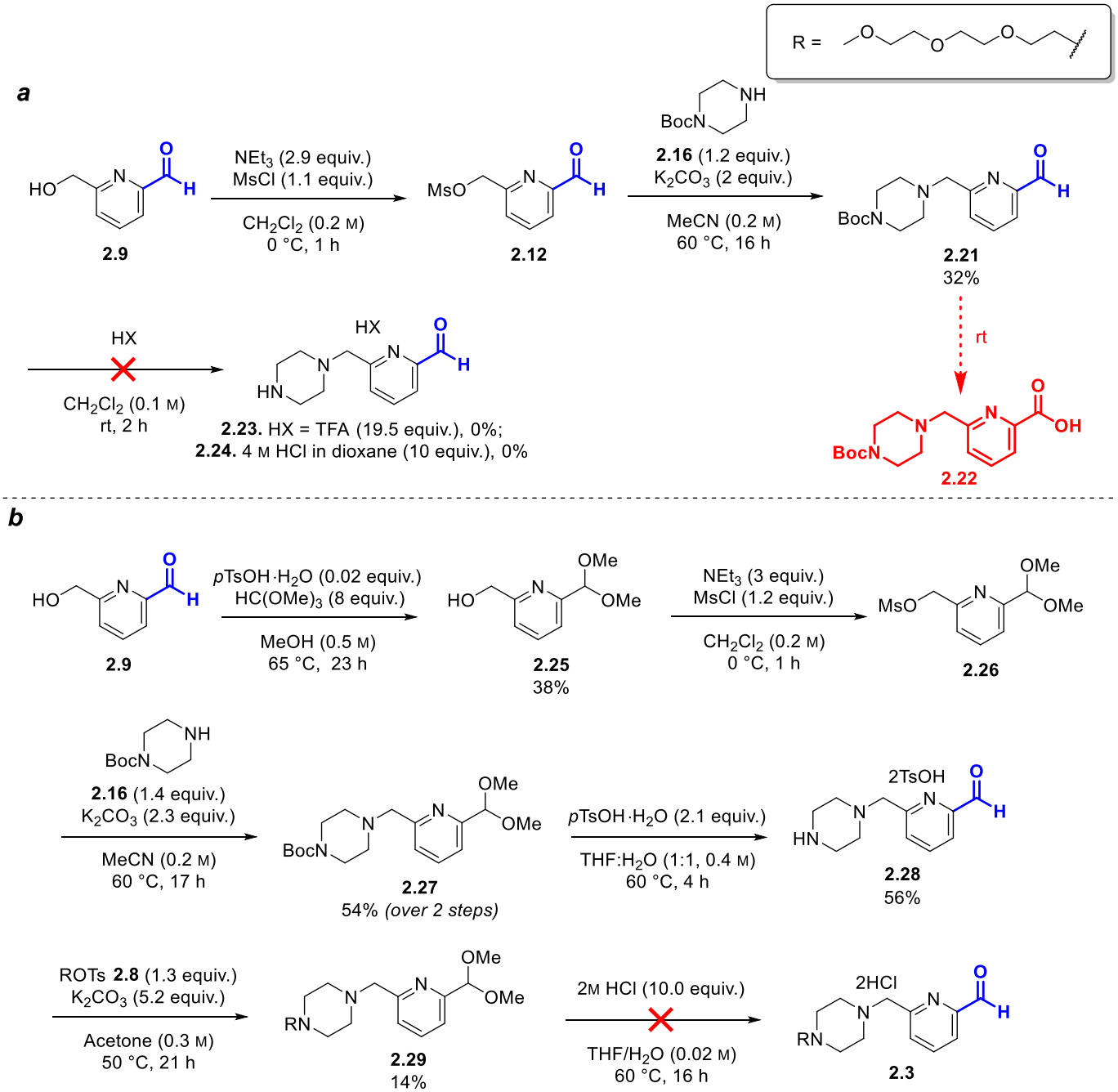
*iii) 2-PCA **2.3***

Initially, we hoped to prepare 2-PCA **2.3** *via* intermediate **2.21**, following the procedure reported by MacDonald *et al.* (*Fig. 2.5a*).[1] This involved the preparation of mesylate **2.12** from aldehyde **2.9** (as done previously in *Fig. 2.3c*), followed by an *N*-alkylation reaction with *Boc*-protected piperazine **2.16** (32%, reagent **2.16** prepared previously in *Fig. 2.4*). However, 2-PCA **2.21** was found to be prone to autooxidation to its corresponding carboxylic acid **2.22**. Autooxidation of aldehydes is a process where benzoylperoxy radicals are generated in the presence of atmospheric oxygen at ambient temperature, followed by carboxylic acid formation *via* Baeyer Villiger rearrangement. As reported by Sankar *et al.*, low concentrations of benzyl alcohol are known to inhibit autoxidation of benzaldehyde by quenching benzoylperoxy radicals involved in the oxidation process, *via* hydrogen atom extraction from the -CH_2 of the benzylic position.[10] Aldehyde **2.9** may have been less prone to autoxidation than aldehyde **2.21** due a similar radical interception by the benzyl alcohol group. From this point onwards, all aldehyde compounds were stored under argon in the freezer and no further autooxidation issues were encountered.

We next attempted the use of acetal protecting groups to prevent aldehyde autooxidation (*Fig. 2.5b*). Firstly, acetal **2.25** was prepared from 2-PCA **2.9** using conditions adapted from those used previously for the preparation of acetal **2.17** in *Fig. 2.4* (38%). Next, compound **2.25** was mesylated to compound **2.26** following the procedure reported by MacDonald *et al.*,[1] and then used directly in the *N*-alkylation reaction with *Boc*-protected piperazine **2.16** under basic conditions (54%). 2-PCA **2.28** (56%) was obtained upon *Boc* and acetal deprotection, followed by *N*-alkylation under basic conditions to afford acetal **2.29** (14%).

Unfortunately, this synthetic approach had limited success due to the degradation of compound **2.29** upon deprotection of the dimethyl acetal formed upon flash column chromatography (5% MeOH:CH_2Cl_2, *Fig. 2.5b*).

To avoid the degradation issues upon deprotection, we next decided to alkylate (and *Boc*-deprotect) the piperazine moiety prior to reaction with mesylate **2.12** (*Fig. 2.5c*). Firstly, *N*-alkylation of piperazine **2.16** upon reaction with tosylate **2.8** (prepared previously in *Fig. 2.3a*) afforded piperazine **2.30** (60%), followed by *Boc*-deprotection to afford piperazine **2.31**. Mesylate **2.12** was prepared from aldehyde **2.9** (as done previously above), followed by *N*-alkylation upon reaction with piperazine **2.31** to afford 2-PCA **2.3** (32%).



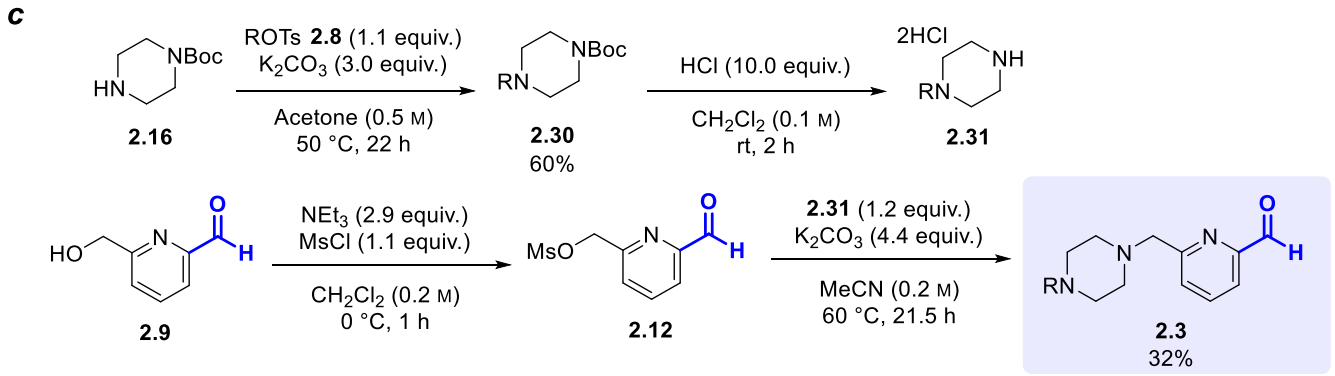


Figure 2.5. Synthetic strategies for the preparation of 2-PCA **2.3**.

*iv) TA4C **2.4***

TA4C **2.4** was synthesised *via* preparation of azide **2.32** upon reaction of tosylate **2.8** (prepared previously in *Fig. 2.3a*) with sodium azide, as reported by Bartl *et al.* (91%).[11] Azide **2.32** was next reacted with 3,3-diethoxyprop-1-yne in a copper-catalysed azide alkyne cycloaddition to afford TA4C **2.4** (61%, lit. procedure adapted from Onoda *et al.*,[3] *Fig. 2.6*). Tandem acetal cleavage was promoted by the protic solvent environment at an elevated temperature (70 °C).[12]

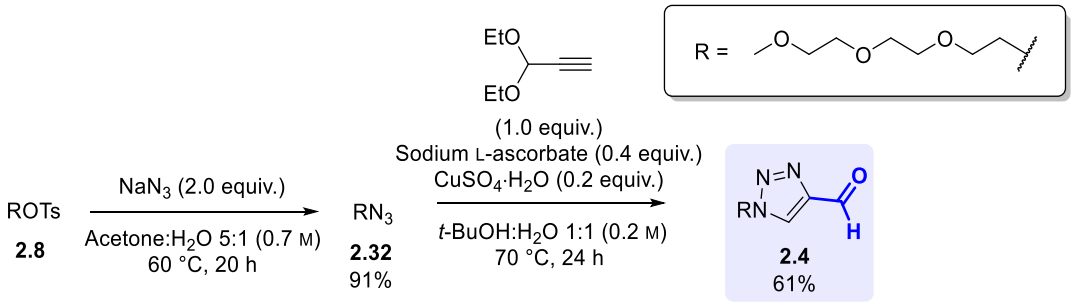


Figure 2.6. Synthesis of TA4C **2.4**.

*v) 2-EBA **2.5***

2-EBA **2.33** was prepared *via* Sonogashira coupling as reported by Deng *et al.* (70%),[4] followed by *O*-alkylation and TMS deprotection, to afford 2-EBA **2.5** (8%, *Fig. 2.7*). The conditions for *O*-alkylation were adapted from Göring *et al.*;[13] TLC indicated the survival of the TMS group under initial conditions, so we introduced an additional deprotection step *via* concentration and addition of methanol (lit. procedure adapted from Deng *et al.*).[4] Note that the low yield of the deprotection/alkylation step (8%) was due to difficulties in product isolation. Following the first purification, the product fraction collected split into 4 spots by TLC, suggesting instability and resulting in the requirement for a second purification and a significant decrease in yield.

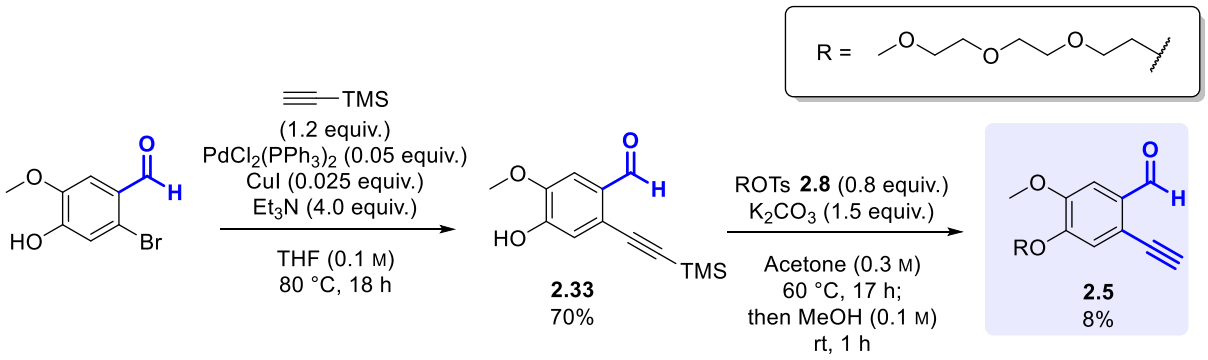


Figure 2.7. Synthesis of 2-EBA **2.5**.

*vi) Ox **2.6***

Ox **2.36** was prepared from L-serine methyl ester hydrochloride according to the procedure reported by Tang *et al* (*Fig. 2.8*).[6] This involved the preparation of oxazolidine-2-thione **2.34** from L-serine methyl ester hydrochloride, followed by *S*-methylation using methyl iodide to afford oxazoline **2.35** (53%). Compound **2.34** was found to degrade over 24 h at rt, so *S*-alkylation to Ox **2.35** was carried out immediately without characterisation to minimise degradation. The ester was then reduced using sodium borohydride to afford Ox **2.36** (57%), followed by *O*-alkylation under conditions previously applied in *Fig. 2.3a* (for the *O*-alkylation of reagent **2.9**) to afford Ox **2.6** (50%). Note that an additional portion of NaH, in combination with elevation of the temperature to 50 °C, was used to push the reaction further towards completion.

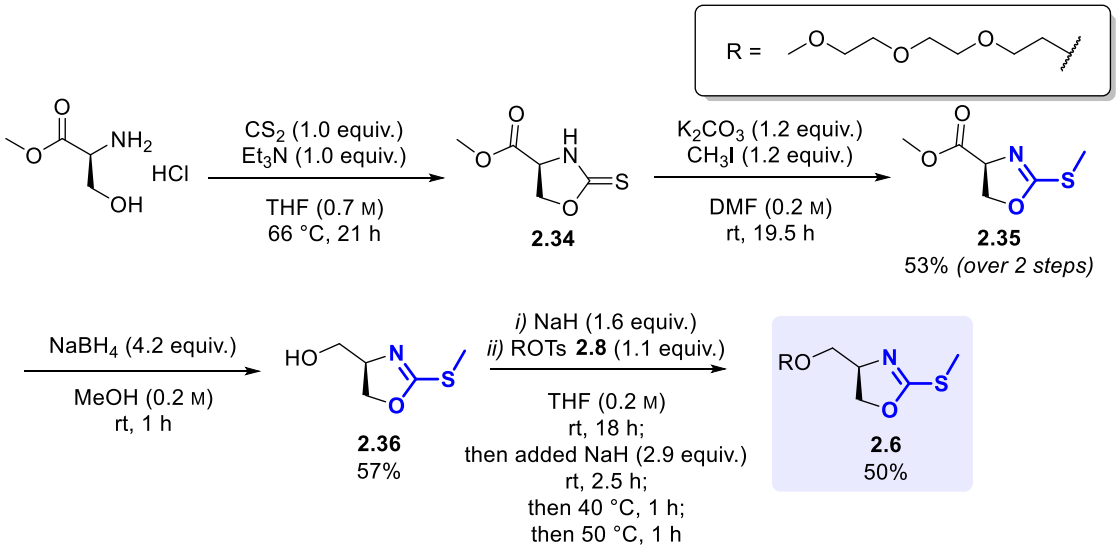


Figure 2.8. Synthesis of Ox **2.6**.

*vii) BA **2.7***

BA **2.7** was prepared *via* *O*-alkylation of benzaldehyde under basic conditions (71%, *Fig. 2.9*); during the reaction, autooxidation of BA **2.7** to its corresponding carboxylic acid was observed, supporting our earlier

observation that the storage of aldehyde-based protein modification reagents under argon in the freezer is critical to ensure stability.

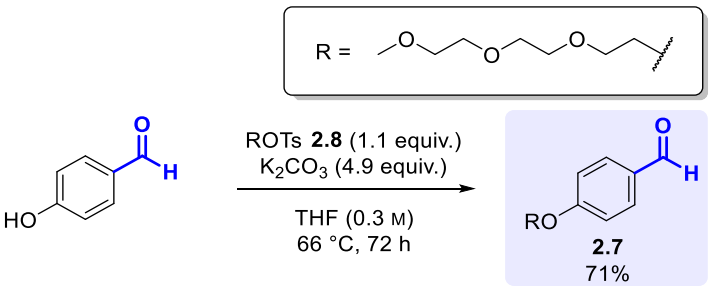


Figure 2.9. Synthesis of BA **2.7**.

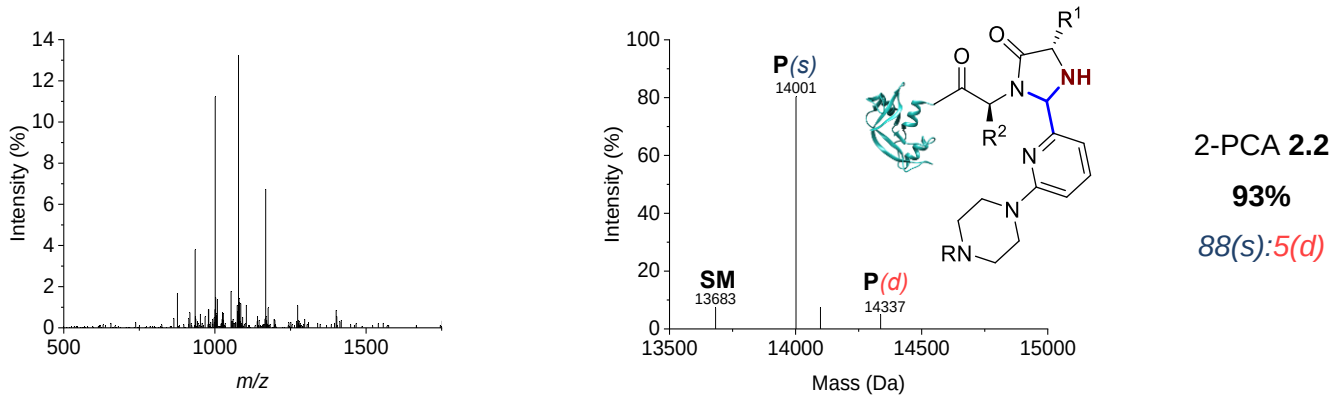
2.2.3 Protein modification: conversion and selectivity

With reagents **2.1**-**2.7** in hand, their reactivity was initially validated by undertaking modification of either RNase A or myoglobin under identical conditions to those previously reported in the literature for each reagent (conditions detailed in *Table 2.1*).[1–6]

					Conversion (%)	
Compound	Protein	Buffer	Temp. / °C	Time / h	***Observed***	***Previously reported***
2-PCA **2.1** (100 equiv.)	RNase A (50 µM)	Na phosphate (50 mM, pH 7.5)	37	23	**84**	**90**[1]
2-PCA **2.2** (100 equiv.)	RNase A (50 µM)	Na phosphate (50 mM, pH 7.5)	37	23	**93** *88(s):5(d)*	**90**[1]
2-PCA **2.3** (100 equiv.)	RNase A (50 µM)	Na phosphate (50 mM, pH 7.5)	37	23	**94** *86(s):8(d)*	**90**[1]
TA4C **2.4** (200 equiv.)	RNase A (50 µM)	K phosphate (10 mM, pH 7.5)	37	16	**86** *81(s):5(d)*	**80**[3]
2-EBA **2.5** (5 equiv.)	RNase A (100 µM)	PBS (50 mM, pH 7.5)	37	16	**15**	**Up to 44**[4]
Ox **2.6** (50 equiv.)	Myoglobin (3 mM)	Na phosphate (10 mM, pH 7.5)	25	16	**100** *20(s):31(d):31(t):14(q):4(qu)*	**70**[6]
BA **2.7** (2 equiv.)	RNase A (700 µM)	Citric acid (25 mM, pH 6.1), with 5 equiv. $NaBH_3CN$	25	23	**7**	**70 (approx.)**[5]

Table 2.1. Validation of literature protein modifications. Modification of an example protein with compounds **2.1**-**2.7** under respective reported conditions (s = single, d = double, t = triple, q = quadruple, qu = quintuple modification).

The reactions detailed above were then analysed prior to removal of the excess reagent *via* intact protein liquid chromatography-mass spectrometry (LC-MS). Raw mass spectra were deconvoluted using using ESI Compass 1.3 to combine the signals from all charge states; an example set of raw/deconvoluted spectra for the modification of RNase A with 2-PCA **2.2** is illustrated below in *Fig. 2.10* (see experimental section for the full dataset for reagents **2.1**-**2.7**). Conversions were then calculated from the relative intensities of modified/unmodified protein signals in the deconvoluted spectra. Our calculations assumed that the protein modifications evaluated had no influence on the ionisability of the protein; we expected this to be a reasonable assumption due to the small size of the moieties attached relative to the size of the model proteins, although the modifications may have also influenced the charge state of the N-terminus. We also assumed that single modifications (s) could be attributed to the N-terminus, due to N-terminal specificity/selectivity observed *via* tryptic digests in previous reports,[1] whilst we assumed double modifications (d) to be a combination of N-terminal modification, with an additional modification at a lysine residue. For reagents with poor selectivity (e.g., in the above experiment, multiple mass additions were observed for Ox **2.6**), signals corresponding to a single mass addition may represent a heterogenous mixture of conjugates, with single modifications in different locations. Note the signals observed in the deconvoluted mass spectra often displayed a range of adducts, most commonly including phosphate, water, acetonitrile or methanol; the relative intensity of each species was calculated as the sum of intensities of the respective adducts observed.



MS (ESI$^+$) [**SM**+H]$^+$ found 13683, calculated 13681;[14] [**P**(s)+H]$^+$ found 14001, calculated 14001; [**P**(s)+H_3PO_4+H]$^+$ found 14099, calculated 14098; [**P**(d)+H_2O+H]$^+$ found 14337, calculated 14338.

Figure 2.10. Example raw (left) and deconvoluted (right) mass spectra for the modification of RNase A with 2-PCA **2.2** under reported conditions outlined in Procedure **2A**. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[15].

Conversions observed for 2-PCAs **2.1**-**2.3** and TA4C **2.4** were comparable to the literature (the reported values for each reaction are illustrated in *Fig. 2.11*), however 2-EBA **2.5** and BA **2.7** gave lower conversions than reported, whilst Ox **2.6** gave a higher conversion than reported. Deng *et al.* previously found that 2-EBA conjugation efficiency and selectivity varied greatly with protein target, reagent structure, number of equivalents, and reaction pH, and so differences resulting from subtle changes to conditions are not unexpected;[4] the poor selectivity of Ox **2.6** was also presumably due to sensitivity of the reaction to

experimental conditions. For **2.7**, possible negative contributions from aldehyde oxidation or from the instability of the intermediate imine, known complications of the use of BAs for protein modification, may be the source of our discrepancies.

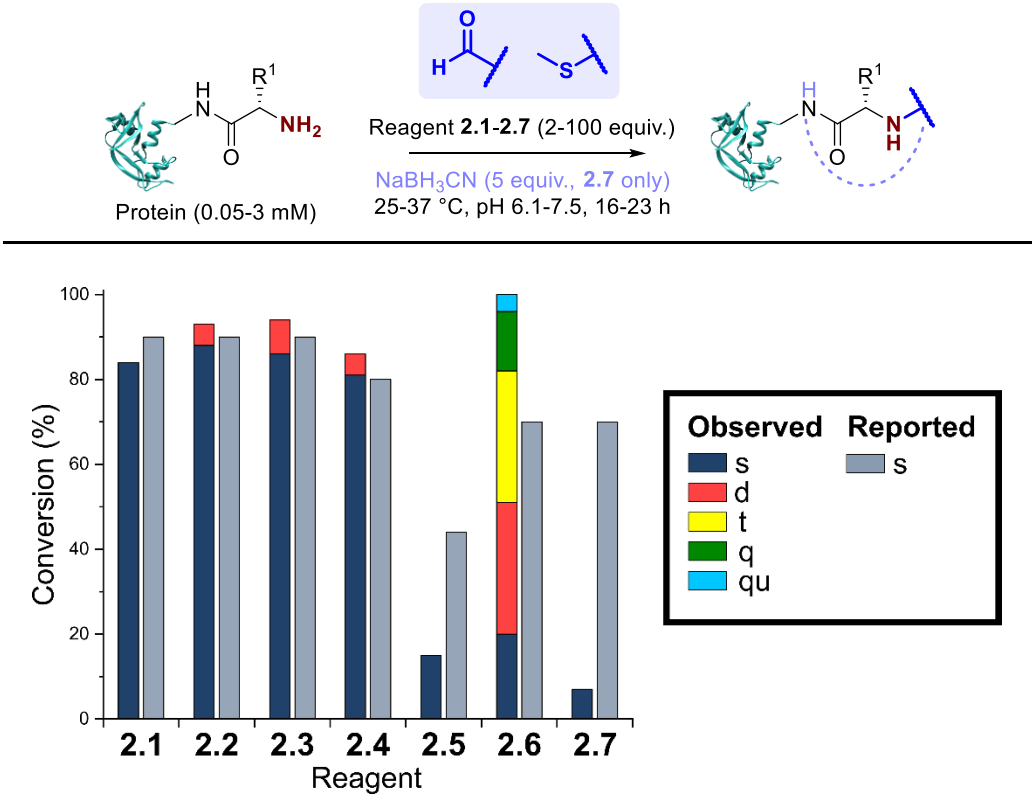


Figure 2.11. Validation of literature protein modifications. Modification of an example protein with compounds **2.1**-**2.7** under respective reported conditions and comparison to literature reports, as outlined in *Table S2.1* (s = single, d = double, t = triple, q = quadruple, qu = quintuple modification). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[15]. For 2-PCAs **2.1**-**2.3**, RNase A (50 μM) was modified with 100 equiv. reagent, in 50 mM pH 7.5 sodium phosphate buffer at 37 °C for 23 h (*procedure **2A***);[1] for TA4C **2.4**, RNase A (50 μM) was modified with 200 equiv. reagent, in 5% DMSO in 10 mM pH 7.5 potassium phosphate buffer at 37 °C for 16 h (*procedure **2B***);[3] for 2-EBA **2.5**, RNase A (100 μM) was modified with 5 equiv. reagent, in 10% DMSO in 50 mM pH 7.5 phosphate buffered saline at 37 °C for 16 h (*procedure **2C***);[4] for Ox **2.6**, myoglobin (3 mM) was modified with 50 equiv. reagent, in 10 mM pH 7.5 sodium phosphate buffer at 25 °C for 16 h (*procedure **2D***);[6] and for BA **2.7**, RNase A (700 μM) was modified with 2 equiv. reagent and 5 equiv. sodium cyanoborohydride, in 25 mM pH 6.1 citrate buffer at 25 °C for 23 h (*procedure **2E***).[5]

Next, we sought to identify suitable modification conditions broadly applicable to all of the reagents, to allow their conversion and selectivity to be directly compared and to maximise applicability to a wide range of protein substrates. We screened the modification of RNase A with reagents **2.1**-**2.7** under conditions outlined in reports for 2-PCA (*procedure **2A***),[1] 2-EBA (*procedure **2C***),[4] Ox (*procedure **2D***)[6] and BA (*procedure **2E***)[5] modification, with sodium cyanoborohydride required in all reactions involving **2.7** to reduce the intermediate imine (*Fig. 2.12*). Note that the reported procedures for Ox and BA modification were unsuitable for the following reasons: *i)* For Ox **2.6**, protein concentrations of 3 mM were used in the initial report by Tang *et al.*,[6] but such concentrations are not generalisable due to the propensity of many proteins to undergo aggregation; and *ii)* the reported reactivity of BA **2.7** is highest at pH 6.1 (although we were unable to reproduce this, with

highest reactivity in our hands observed at pH 7.5), but again this leads to a loss of generalisability.[16] In agreement with the above observations, our study highlighted the reported procedure for 2-PCA modification as the most suitable for use as a generalised procedure (*procedure* ***2A***: 50 μM protein, 100 equiv. reagent, pH 7.5, 37 °C, 23 h): this was the only set of conditions for which reactivity of RNase A with each of the reagents **2.1**-**2.7** was observed.

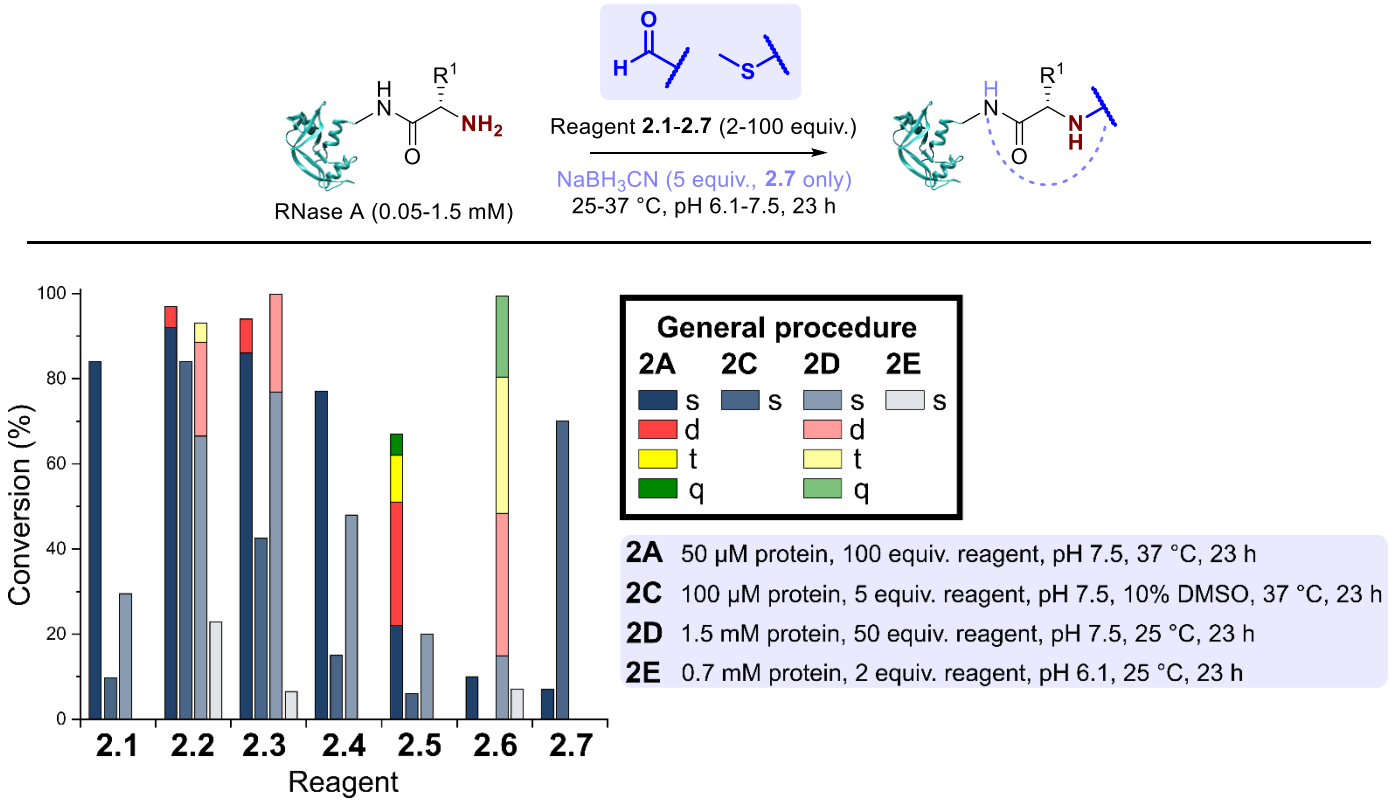
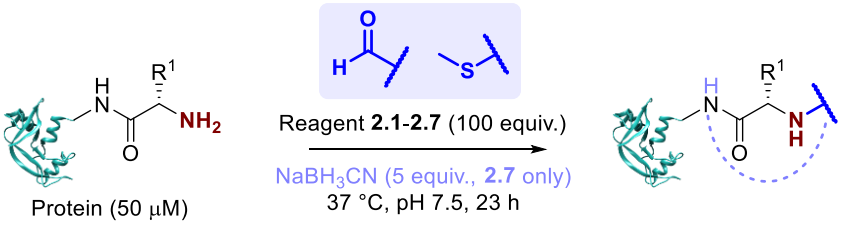


Figure 2.12. Screening of reagents under differing conditions (s = single, d = double, t = triple, q = quadruple modification). Modification of RNase A with compounds **2.1**-**2.7** under respective reported conditions: ***General procedure 2A*** (50 μM protein, 100 equiv. reagent, pH 7.5, 37 °C, 23 h);[1] ***General procedure 2C*** (100 μM protein, 5 equiv. reagent, pH 7.5, 10% DMSO, 37 °C, 23 h);[4] ***General procedure 2D*** (1.5 mM protein, 50 equiv. reagent, pH 7.5, 25 °C, 23 h);[6] ***General procedure 2E*** (0.7 mM protein, 2 equiv. reagent, pH 6.1, 25 °C, 23 h).[5] Modifications with reagent **2.7** were carried out in the presence of $NaBH_3CN$ (5 equiv.). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[15].

With protein modification validated and generalised conditions identified, we next made a direct comparison between the reactivity of **2.1**-**2.7**. Our representative reaction conditions were applied to a panel of model proteins (*Table 2.2*). RNase A (N-terminal Lys, 10 total Lys residues), equine myoglobin (N-terminal Gly, 20 Lys), and clostripain light chain (LC, N-terminal Asn, 14 Lys) were chosen due to their ease of MS detection. Our lack of protein purification prior to analysis during this study is notable, allowing us to observe varying levels of multi-site protein modification, due to either transient or stable adducts with other amino acids. Our results showed that all reagents exhibited some degree of off-target reactivity, highlighting the need for extensive protein purification to remove excess reagent, as discussed later.

Studying this panel of proteins revealed that each reagent had distinct protein labelling preferences, with no generalisable trends observed whereby a decrease in conjugation efficiency for one reagent was mirrored in a decrease for all others too. In general, 2-PCAs **2.1**-**2.3** showed the highest reactivity, with modifications of RNase A of 84-94% and of clostripain LC of 94-100%. Conjugation efficiency for myoglobin was reduced (41-57%) with little difference in reactivity between the three 2-PCA constructs being studied. Koo *et al.* previously reported an increase in reactivity when the piperazine was attached directly to the PCA ring *via* an Ar-NR_2 bond, as in reagent **2.2**, rather than the Ar-CH_2NR_2 seen in reagent **2.3**.[2] However, these differences were not observed here with **2.1**-**2.3** behaving similarly. The reactivity of **2.1** indicated that the piperazine motif did not affect conjugation efficiency, opening up alternative structures for functional 2-PCAs.

In general, TA4C **2.4** showed a similar trend across the different proteins but lower reactivity than the 2-PCA reagents, with conjugation to myoglobin (6%) notably low. In contrast, 2-EBA **2.5**, Ox **2.6**, and BA **2.7** all exhibited higher reactivity with myoglobin (100, 42, and 46% modification respectively) than for RNase A (74, 22, and 10% modification). Though this may indicate that labelling preferences are conserved when reagents are grouped as either N-terminal *specific* or N-terminal *selective*, a larger study would be required to confirm this.



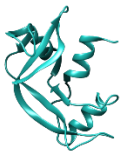
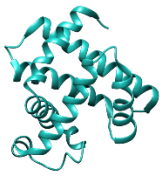
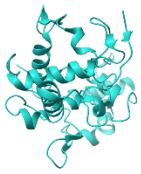
	RNase A (Lys)	Myoglobin (Gly)	Clostripain LC (Asn)	CjX183-D WT (Gly)
				
2-PCA **2.1**	**84%**	**57%** *55(s):2(d)*	**97%** *89(s):8(d)*	**100%** *49(s):39(d):12(t)*
2-PCA **2.2**	**93%** *88(s):5(d)*	**41%** *39(s):2(d)*	**94%**	**100%** *15(s):34(d):35(t):16(q)*
2-PCA **2.3**	**94%** *86(s):8(d)*	**47%**	**100%**	**95%** *40(s):43(d):10(t):3(q)*
TA4C **2.4**	**77%**	**6%**	**90%** *84(s):6(d)*	**69%** *55(s):11(d):3(t)*
2-EBA **2.5**	**74%** *29(s):29(d):11(t):5(q)*	**100%** *32(d):17(t):24(q):22(se):5(sp)*	**100%** *24(s):37(d):38(t)*	**97%**
Ox **2.6**	**22%**	**42%** *34(s):8(d)*	**37%** *34(s):3(d)*	**100%**
BA **2.7**	**10%**	**46%** *42(s):4(d)*	**37%** *29(s):7(d):1(t)*	**98%** *13(s):78(d):6(t)*

Table 2.2. Conversions for the modification of a panel of proteins with reagents **2.1-2.7**. Modifications with reagent **2.7** were carried out in the presence of $NaBH_3CN$ (5 equiv.). Selectivity is shown in *italics*, and entries colour coded based on the yield of singly modified protein (green = highest; red = lowest; s = single, d = double, t = triple, q = quadruple, qu = quintuple, se = sextuple, sp = septuple modification). The N-terminal residue of each protein is given in brackets. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3);[15] Maurus *et al.* (Myoglobin, PDB 1WLA);[17] González-Páez *et al.* (Clostripain LC, PDB 6MZO);[18] Grabarczyk *et al.* (CjX183-D, PDB 4V2K)[19].

During our studies, it became clear that the N-terminal selectivity of 2-EBA **2.5** was poor in our hands. At 100 equiv. reagent, high levels of double, triple, and even quadruple modification were observed. At 200 equiv. this off-target reactivity became so pronounced that protein signals could not be deconvoluted, presumably due to a mix of heterogeneously modified proteins bearing different numbers and sites of modification (*Fig. 2.13*).[20] Single site modification could be observed by using just 10 equiv. of **2.5**, but at the cost of conversion.

Moreover, given the poor selectivity observed at higher reagent loadings, it is plausible that for proteins bearing a single modification conjugation had taken place at another site than the N-terminus.

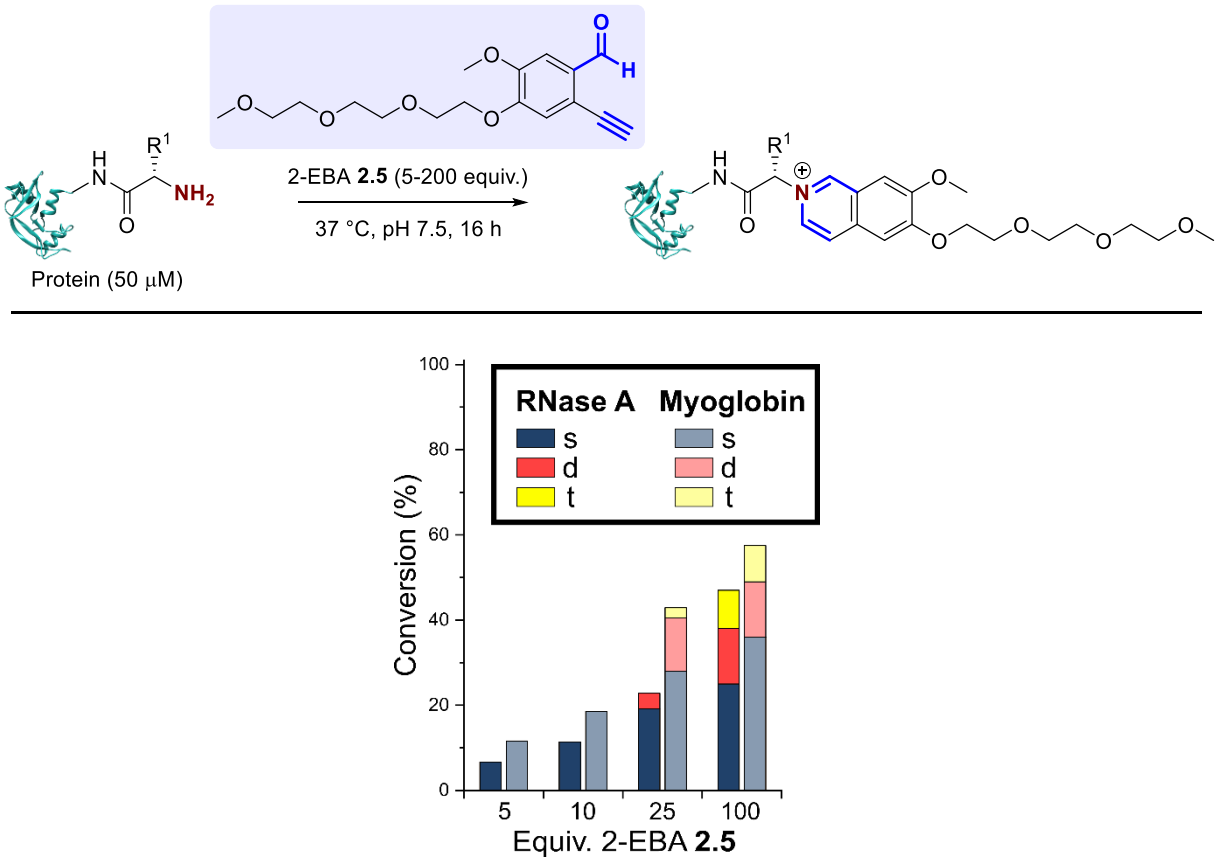


Figure 2.13. Optimisation of 2-EBA concentration. Modification of RNase A and myoglobin with 2-EBA **2.5** (5, 10, 25, 100 equiv.; s = single, d = double, t = triple modification). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[15].

Site-selectivity was generally lower for the N-terminal *selective* reagents, with **2.6** and **2.7** also exhibiting various levels of off-target modification. This is not surprising, given the subtle differences in basicity and nucleophilicity of the α- and ε-amines of the N-terminus and lysine side chains respectively, particularly given the importance of the surrounding environment in dictating protonation state and nucleophilicity. However, the lack of purification prior to analysis also allowed us to observe low levels of off-target reactivity for N-terminal specific reagents **2.1**-**2.4**. We hypothesised that these modifications were predominantly a result of transient modifications at lysine, which would be expected to hydrolyse over time upon removal of excess reagent.

To probe this hypothesis, we studied the modification of CjX183-D, a naturally occurring cytochrome from *Cellvibrio japonicus* which does not contain any lysine residues.[21] N-terminal selectivity was compared to the mutant CjX183-D R51K, in which a single lysine had been installed (*Fig. 2.14*): these proteins were expressed and purified by Dr. Nicholas Yates, as detailed in our publication.[22] This study revealed a number of interesting results. For 2-EBA **2.5** and Ox **2.6**, exclusive single-site modification was observed for CjX183-D WT, while significant levels of double modification were observed for CjX183-D R51K, strongly indicating that off-target

modification with these reagents takes place primarily at lysine. Selectivity was poor for BA **2.7**, but was worse for CjX183-D R51K, highlighting lysine as a possible, but not exclusive site for off-target modification.

To our surprise, 2-PCAs **2.1**-**2.3** and TA4C **2.4** had poor selectivity for the N-terminus of CjX183-D, with high levels of double, triple, and even quadruple modification being observed. This suggests that these reagents can undergo off-target modification at amino acids other than lysine, further highlighting the high dependence of modification on protein identity. Interestingly, no change in selectivity was observed for CjX183-D R51K, suggesting that the single lysine is not modified to an appreciable degree. It may be that the unique surface properties of CjX183-D create hyper-reactive residues that would not typically react with **2.1**-**2.4** in other proteins. The multi-site adducts observed had a mass of +18 Da relative to the mass that would result following reaction and dehydration of the 2-PCA/TA4C (as expected for imine formation), indicating a potentially new mode of reactivity, or perhaps hemiaminal formation. Unfortunately, as discussed next, our attempts to identify the site and nature of modification were unsuccessful, with low surface coverage after digestion complicated by the covalently conjugated haem group of the cytochrome.

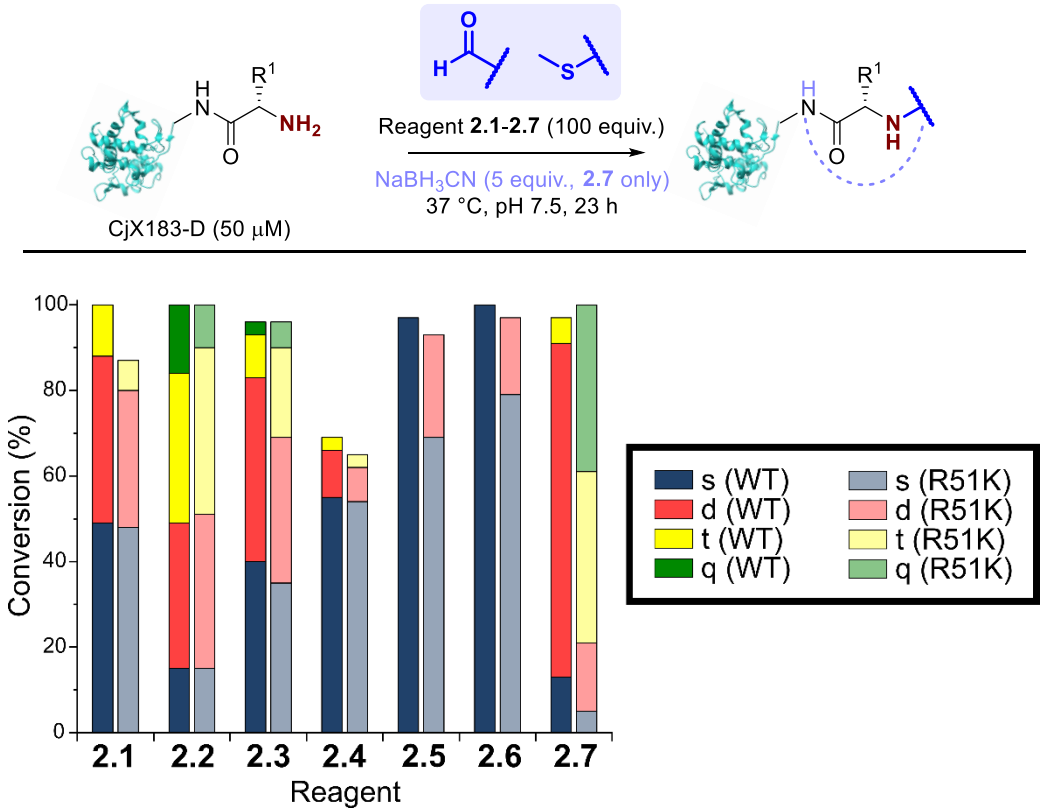


Figure 2.14. Conversion for the modification of CjX183-D WT (0 Lys) and CjX183-D R51K (1 Lys) prior to purification (s = single, d = double, t = triple, q = quadruple modification). Modifications with reagent **2.7** were carried out in the presence of $NaBH_3CN$ (5 equiv.). Figure built using structural data obtained by Grabarczyk *et al.* (CjX183-D, PDB 4V2K)[19].

2.2.4 CjX183-D off-target reactivity

Having observed unexpected off-target reactivity for CjX183-D, we next sought to determine the site and nature of the additional modifications. MALDI-MS and LC-MS/MS analysis of the CjX183-D WT modified with 2-PCA **2.3** and an unmodified control post-trypsin digestion localised the modification to the first two amino acids of the N-terminal peptide, supporting our previous assumption that the modification occurred preferentially at the N-terminus (*Fig. S2.8*; experimental work and analysis carried out by Dr. Adam Dowle and Dr. Chris Taylor). However, no additional modifications were located. Both the Cys-containing peptides were not detected (GANLWNTQTCVACHGVDGER and DFISMWMPQGNEGSCTGQCAADIEAFIR) due to haem covalent attachment, sequestering these peptides and impairing identification. Next, AspN and GluC digestion of the CjX183-D WT/2-PCA **2.3** conjugate was carried out to cover regions previously not covered by the original trypsin digestion and suggested Asp26 (DATRGANLWNTQTCVACHGVDGERNASGTPALTPLNPNR) as a possible site of double modification (*Fig. S2.9*; experimental work and analysis carried out by Dr. Adam Dowle and Dr. Chris Taylor). Whilst this mass addition cannot be unambiguously assigned to Asp26 due to the tendency of multiple modifications to concatenate within one peptide into one single modification and position, missing AspN cleavage at the Asp26 position suggested the chemical modification occurred at this location and thus prevented some of the cleavage.

We hypothesised that the additional modification of CjX183-D, tentatively observed at Asp26, may have occurred through enolate addition; note that our hypothesis was purely speculative, with no supporting literature evidence. To probe this hypothesis, we decided to directly compare the modification of CjX183-D WT with 2-PCA **2.37** and 4-PCA **2.38**. For 2-PCA **2.37**, the proximity of the pyridyl nitrogen atom with the β-hydrogen allows for deprotonation and subsequent dehydration (as shown in the dotted boxes, *Fig. 2.15*), whereas we expected the extended distance between the pyridyl nitrogen atom and the β-hydrogen for 4-PCA **2.38** to not facilitate the elimination stage of the reaction, thus preventing the secondary modification of the aspartic acid residue. Following elimination, we therefore expected the 2-PCA **2.37** conjugate to be a Michael acceptor and to react upon addition of β-mercaptoethanol (with double modification of both the N-terminus and Asp26 prior to β-mercaptoethanol addition), whereas the 4-PCA **2.38** conjugate was not expected to act as a Michael acceptor (with single N-terminal modification prior to β-mercaptoethanol addition).

Whilst 4% thiol addition was observed for 2-PCA **2.37** and not for 4-PCA **2.38**, 4% thiol addition was also observed for an unmodified control sample so could be attributed to off-target addition at another location of the protein (*Fig. 2.15a*). Surprisingly, up to triple modification was observed for 4-PCA **2.38**, with only single modification for 2-PCA **2.37**, in disagreement with our hypothesis. Due to the very low levels of thiol addition, the experiment was repeated with cysteine as an alternative thiol nucleophile, at higher concentration (100 equiv. instead of 7 equiv. β-mercaptoethanol), with incubation at 37 °C for a longer time (25 h instead of 5 h). We also included 2-PCA **2.3** in the experiment, as the previous AspN and GluC digestion (indicative of Asp26 modification) had been performed on a CjX183 / 2-PCA **2.3** conjugate (*Fig. 2.15b*). However, no thiol addition

was observed for any of the reagents: we have been unable to find any evidence to support the tentative conjugate digestion results. Note that a decrease in conversion was observed upon thiol addition due to competitive cleavage, as will be discussed later.

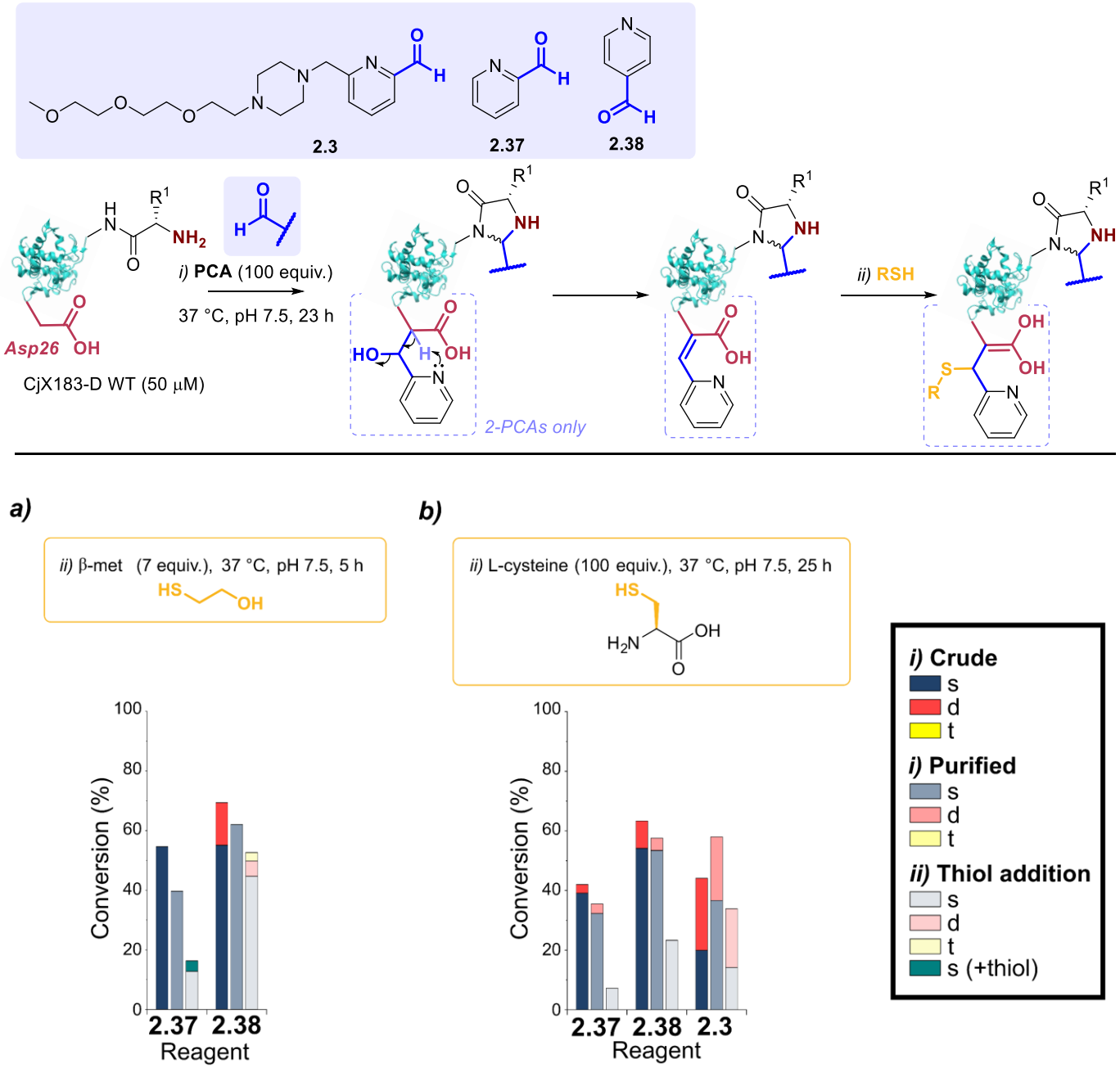


Figure 2.15. Dotted boxes show possible reactivity of 2-PCA **2.37** with the Asp26 residue of CjX183-D WT. Modification of CjX183-D WT with ***a)*** 2-PCA **2.37** and 4-PCA **2.38**, followed by addition of β-mercaptoethanol; ***b)*** PCAs **2.37**, **2.38**, and **2.3**, followed by addition of L-cysteine (s = single, d = double, t = triple modification). Figure built using structural data obtained by Grabarczyk *et al.* (CjX183-D, PDB 4V2K)[19].

2.2.5 N-terminal selectivity/specificity after purification

Having studied N-terminal modification prior to purification, we next performed an analogous study following removal of excess reagent *via* dialysis at 4 °C (*Fig. 2.16*). Though some loss of conjugation took place under these conditions, sufficient levels were maintained to allow qualitative comparisons. In this scenario, transient or unstable modifications, such as hydrolytically sensitive imines, would be cleaved. For RNase A and myoglobin, only single site modification was observed following the removal of excess **2.1**-**2.4**. For **2.5**, lower levels of off-target reactivity were found for RNase A conjugates after purification, indicating some level of transient/unstable modification prior to purification. However, significant levels of dual-site modification were still present for **2.5** and **2.6** after purification of myoglobin conjugates, presumably at lysine based on our previous experiments.

These results emphasise the benefits of N-terminal *specific* vs. N-terminal *selective* reagents for most proteins. However, high levels of CjX183-D dual modification were still observed for 2-PCAs **2.1**-**2.3** after purification (26-63%). This suggests that the hyper-reactive residues in CjX183-D that lead to this unexpected off-target reactivity produce conjugates of relatively high stability. It is plausible to consider that the singly modified protein that *is* observed may in fact be a mix of N-terminal and off-target conjugates, further reducing confidence in this modification as being targeted predominantly to the N-terminus. Variability in the stability of conjugates to purification between target proteins highlights the protein dependency of N-terminal labelling. For example, lower stability was observed for CjX183-D / TA4C **2.4** conjugates than for RNase A. As will be discussed later in Chapter 5, the modification efficiency and stability to purification of TA4C conjugates is highly influenced by nature of the N-terminal residue (*Fig. 5.38 and 5.45*).

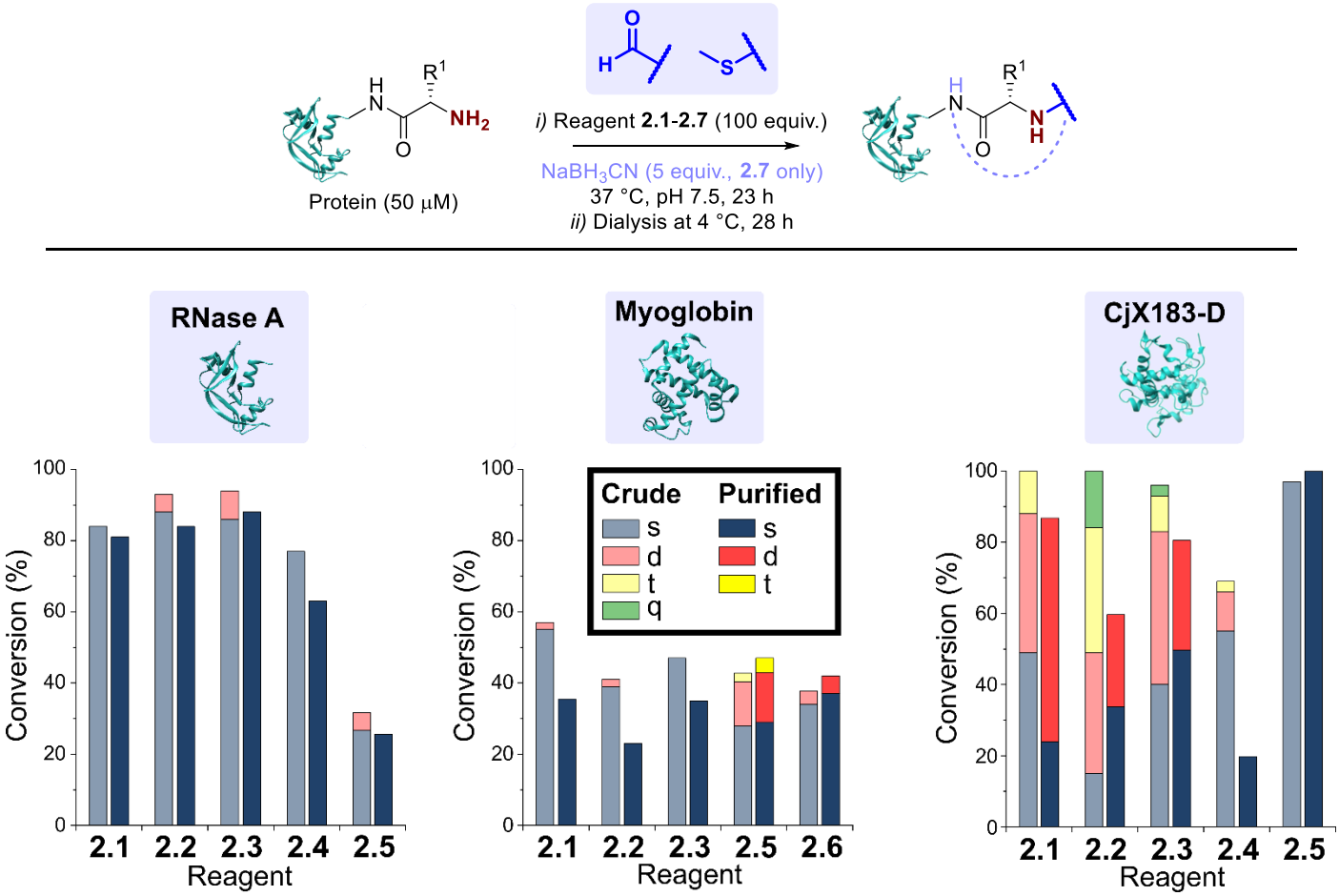


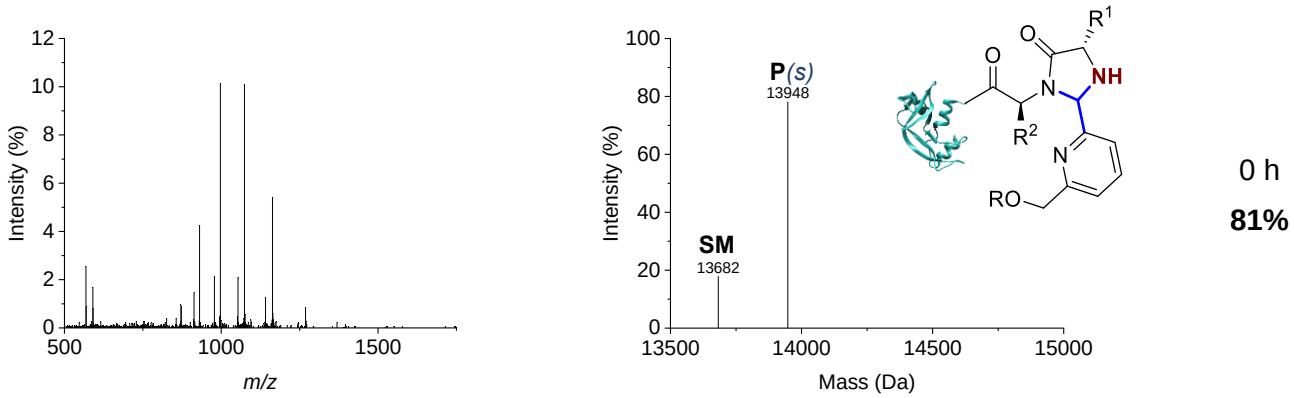
Figure 2.16. Conversions for the modification of RNase A, myoglobin, and CjX183-D before (crude) and after dialysis at 4 °C to remove excess reagent and unstable conjugation. Nb, 100 equiv. reagent **2.1**, **2.3**, **2.4**, **2.6**, and **2.7**; 67 equiv. **2.2** and 25 equiv. **2.5** were used to allow protein analysis and to prevent over-modification respectively (s = single, d = double, t = triple, q = quadruple modification). Modifications with reagent **2.7** were carried out in the presence of $NaBH_3CN$ (5 equiv.). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3);[15] Maurus *et al.* (Myoglobin, PDB 1WLA);[17] Grabarczyk *et al.* (CjX183-D, PDB 4V2K)[19].

2.2.6 Conjugate stability

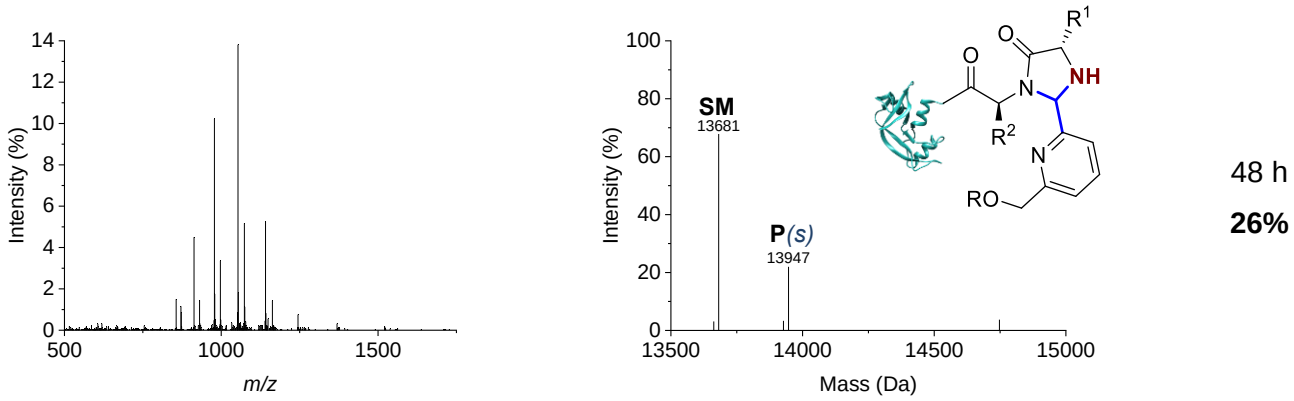
The stability of a modified protein to the conditions in which it is to be used is a critical, but often overlooked aspect of a bioconjugation strategy. MacDonald *et al.* previously reported that 2-PCA / RNase A conjugates underwent a 20-30% decrease in conjugation upon incubation in 50 mM sodium phosphate buffer at a range of pHs (pH 3-11) over 12 h at 37 °C,[1] while Deng *et al.* reported 2-EBA derivatives to possess superior stability, with no decomposition of 2-EBA / RNase A conjugates upon incubation in 50 mM phosphate buffered saline at a range of pHs (pH 3-11) over 12 h at 37 °C[4]. Similarly, Tang *et al.* found Ox-peptide conjugates to be stable upon incubation in 10 mM sodium phosphate buffer at a range of pHs (pH 3.5-10.5) at 37 °C for 48 h, though no studies on proteins were performed.[6] In contrast, the stabilities of TA4C and BA conjugations have not previously been investigated. Moreover, differences in target protein, study design, and conjugation analysis make it difficult to draw fair comparisons between the stability of the reagents used in these separate

reports. We therefore studied conjugate stability across a range of conditions that are representative of the expected applications of chemically modified proteins.

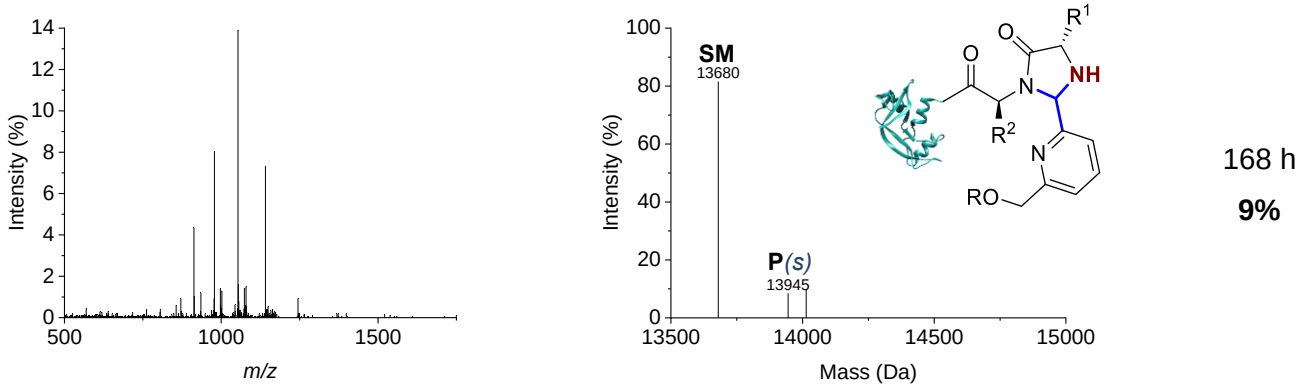
RNase A, myoglobin, and CjX183-D WT were modified with **2.1**-**2.7**, as described earlier, and purified by dialysis at 4 °C for 24 h to remove excess reagents. Purified conjugates were then subjected to a range of different conditions (pH 7 at 4 °C, 22 °C, and 37 °C; pH 6 and 8 at 22 °C) under dialysis conditions in sodium phosphate buffer, to remove any unconjugated reagents that were released over time and prevent reattachment. Conjugate stability was then monitored by LC-MS at different time points over 1 week; an example set of raw/deconvoluted spectra for the incubation of RNase A / 2-PCA **2.1** conjugates at 37 °C (pH 7) is illustrated below in *Fig. 2.17*. In some cases, e.g. the modification of myoglobin with TA4C **2.4**, the low initial conversion prior to purification prevented an accurate analysis being undertaken, and these samples were therefore omitted from the study.



MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681;[14] [**P**(s)+H]$^+$ found 13948, calculated 13946.



MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 13663, calculated 13663; [**SM**+H]$^+$ found 13681, calculated 13681;[14] [**P**(s)-H_2O+H]$^+$ found 13927, calculated 13928; [**P**(s)+H]$^+$ found 13947, calculated 13946.



MS (ESI$^+$) [**SM**+H]$^+$ found 13680, calculated 13681;[14] [**P**(s)+H]$^+$ found 13945, calculated 13946.

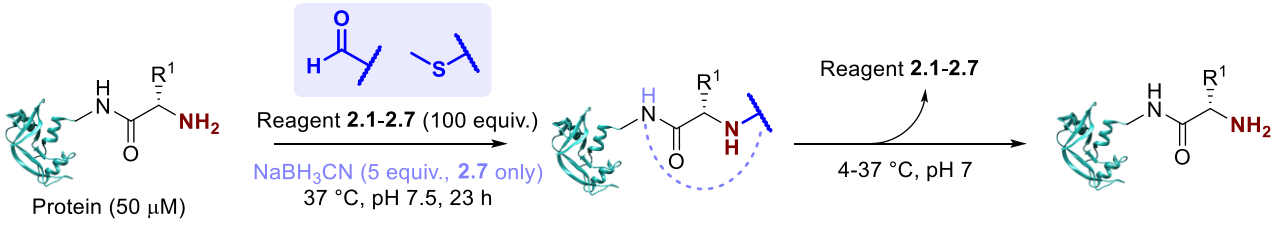
Figure 2.17. Example raw (left) and deconvoluted (right) mass spectra for the incubation of RNase A / 2-PCA **2.1** conjugates at 37 °C (pH 7) over time. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[15].

With our previous observations that conjugation efficiency for each reagent was strongly protein dependent, it was unsurprising to observe similar effects for the conjugate stability. At 4 °C, all RNase A conjugates were found to be reasonably stable over a 48 h time period (0-9% loss), with 2-PCAs **2.1**-**2.3** and 2-EBA **2.5** only losing ~2-12% conjugation over a week (*Fig. 2.18*). In contrast, modifications of myoglobin were generally found to be less stable at 4 °C, with 2-PCA **2.2** being notable for the ~30% loss in conjugation after 48 h. This instability was even more pronounced for CjX183-D WT, with major drops in conjugation for both 2-PCA **2.2** (44%) and TA4C **2.4** (59%).

At 22 °C, RNase A conjugates began to show instability (*Fig. 2.18*). For N-terminal *specific* reagents **2.1**-**2.4**, between 4-27% of conjugation was lost after 48 h, with 31-64% loss after a week. This instability was even more significant at 37 °C, with between 55-85% loss of conjugation for **2.1**-**2.4** after 48 h, and near complete cleavage after 7 days incubation. 2-EBA **2.5** adducts were more stable, but still exhibited a 29% loss after 7 days at 22 °C, or 56% at 37 °C. For myoglobin, differences in stability between 4 °C and 22 °C were minimal, but a more significant drop in conjugation upon incubation at 37 °C. Interestingly, among the 2-PCA reagents tested, myoglobin conjugates with **2.1** were considerably more stable (~16% loss after 1 week), than those formed with **2.2** (69% loss after just 48 h). Koo *et al.* previously reported that Ar-NR_2 based 2-PCAs, such as **2.2**, reacted more rapidly than the corresponding Ar-CH_2NR_2 analogues,[2] but these results suggest a more complex relationship between rate of conjugate formation and subsequent stability, which may differ on a protein by protein basis. A more complex analysis will be discussed later in Chapter 4. Across all of the proteins and conditions tested, we found that pH had little effect on conjugate stability within the range pH 6-8 tested (*Fig. 2.19*).

The stability of off-target modifications relative to the desired N-terminal conjugation has important implications for the preparation of homogenous constructs. Double modifications observed for 2-EBA **2.5** and Ox **2.6**, discussed above, were found to persist over extended incubation times, albeit with a small reduction over time, particularly for myoglobin. The unexpected dual-site modifications of CjX183-D with 2-PCAs **2.2**-

2.3 were also found to be surprisingly stable. Though some loss of dual conjugation was observed, over 1 week at 22 °C, and even 37 °C, significant levels persisted. This stability was less pronounced for 2-PCA **2.1**, with almost complete conversion to the singly modified protein at both 22 °C and 37 °C. However, given the results discussed above it is not clear if this modification is at the N-terminus, and it is plausible to consider a scenario in which all N-terminal modification is cleaved (given the instability of RNase A and myoglobin modifications) and the undesired off-target modification has been retained.


Protein (50 μM)
Reagent 2.1-2.7 (100 equiv.)
NaBH3CN (5 equiv., 2.7 only)
37 °C, pH 7.5, 23 h
Reagent 2.1-2.7
4-37 °C, pH 7

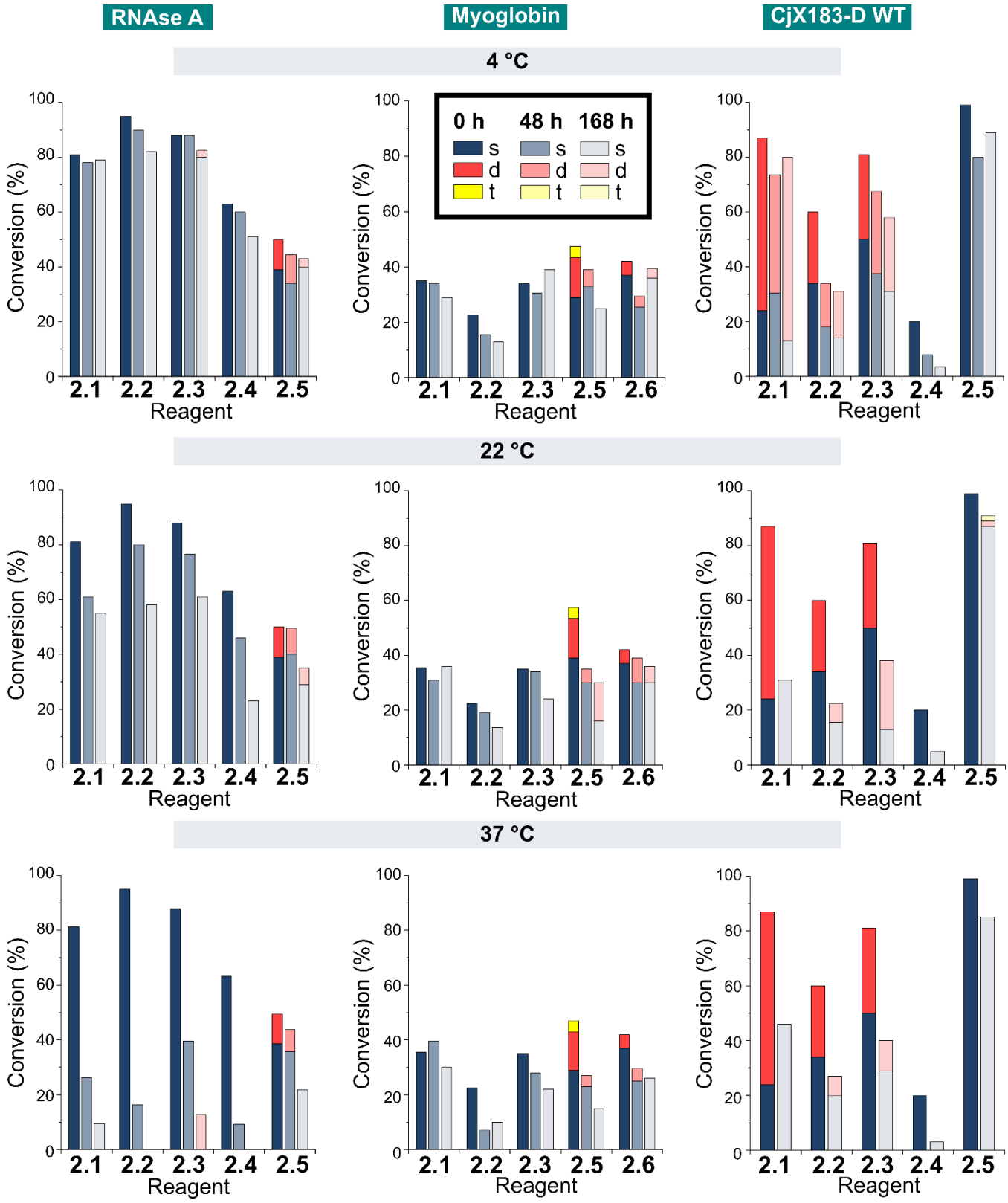

RNAse A
Myoglobin
CjX183-D WT
4 °C
22 °C
37 °C
0 h
48 h
168 h
s
d
t
Conversion (%)
Reagent
2.1
2.2
2.3
2.4
2.5
2.6

Figure 2.18. Stability of protein conjugates over time. After initial conjugation at 37 °C for 23 h, proteins were purified by dialysis at 4 °C. Samples were then incubated under the specified conditions at pH 7 and conjugation monitored over time by LC-MS (s = single, d = double, t = triple modification). Modifications with reagent **2.7** were carried out in the presence of $NaBH_3CN$ (5 equiv.). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[15].

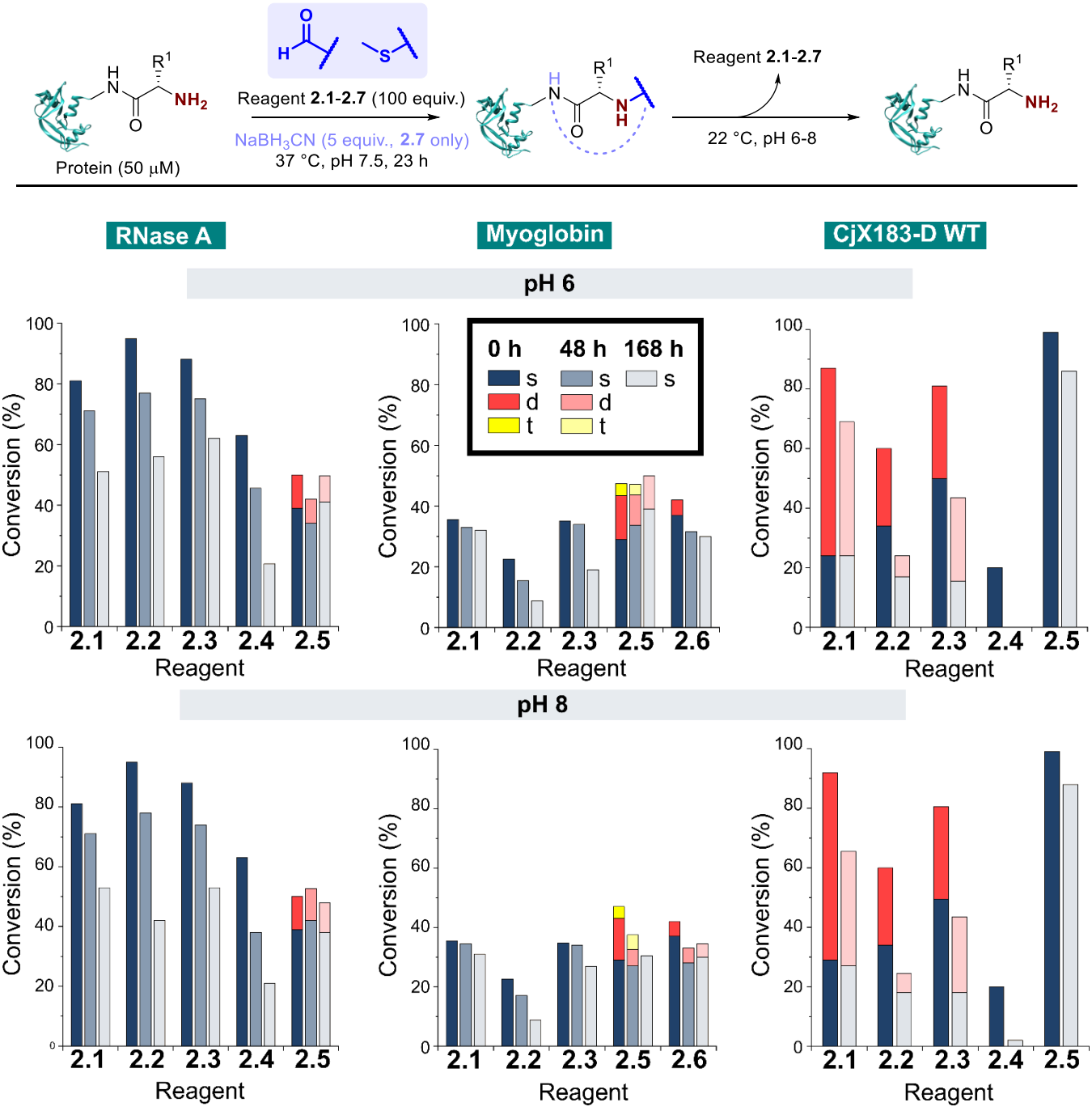


Figure 2.19. Stability of protein conjugates over time. After initial conjugation at 37 °C for 23 h, proteins were purified by dialysis at 4 °C. Samples were then incubated under the specified conditions at 22 °C and conjugation monitored over time by LC-MS (s = single, d = double, t = triple modification). Modifications with reagent **2.7** were carried out in the presence of $NaBH_3CN$ (5 equiv.). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[15].

2.2.7 Conjugate stability to competing N-termini

In many applications of modified proteins there are likely to be high concentrations of other unlabeled proteins and peptides (e.g. in blood, protein concentration is in the range 60-80 mg/mL). The stability of an N-terminal conjugate in the presence of an excess of potentially competing N-termini is therefore a critical factor. Koo *et al.* demonstrated that proteins could be stably immobilised on polymeric supports for up to 9 days *via* 2-PCA handles, but in this scenario the local 2-PCA concentration was high (up to 27 mM), driving the equilibrium towards polymer-protein conjugates, even if individual bonding events were dynamic.[2] When excess 2-PCA was reacted with hydroxylamine a shift in equilibrium was observed and the protein released from the polymer matrix, emphasising this point. We therefore sought to understand the effect an excess of competitive N-termini might have on the stability of protein conjugates formed with **2.1**-**2.7**.

To do this, RNase A was first modified with 100 equiv. of reagents **2.1**-**2.7** as described above. Without purification, the dipeptide Ala-Ala (DiAla) was added as a simple model competitor, possessing an α-amido amine at varying concentrations (*Fig. 2.20*, 1000, or 2000 equiv. w.r.t. protein; 10 or 20 equiv. w.r.t. reagent **2.1**-**2.7**). This large excess was expected to serve two roles, first reacting with excess reagent and nullifying its presence, and secondly scavenging any that was released as a result of dynamic binding at the N-terminus. Reactions were incubated at 37 °C for 24-72 h, and in all cases a large drop in protein conjugation was observed relative to the control (0 equiv. DiAla). Though a small decrease in stability was observed upon doubling the DiAla concentration, the effect was minimal implying a dissociative mechanism of cleavage, i.e. imidazolidinone ring opening and imine hydrolysis were the primary steps, followed by conjugation of DiAla with the free reagent. As seen during the stability studies described above, conjugates formed from 2-PCA **2.2** and TA4C **2.4** were particularly susceptible to cleavage, while conjugates formed with Ox **2.6** were less susceptible to competition. In contrast, when these experiments were repeated with CjX183-D, **2.6** was completely cleaved over 72 h, again demonstrating high protein dependence.

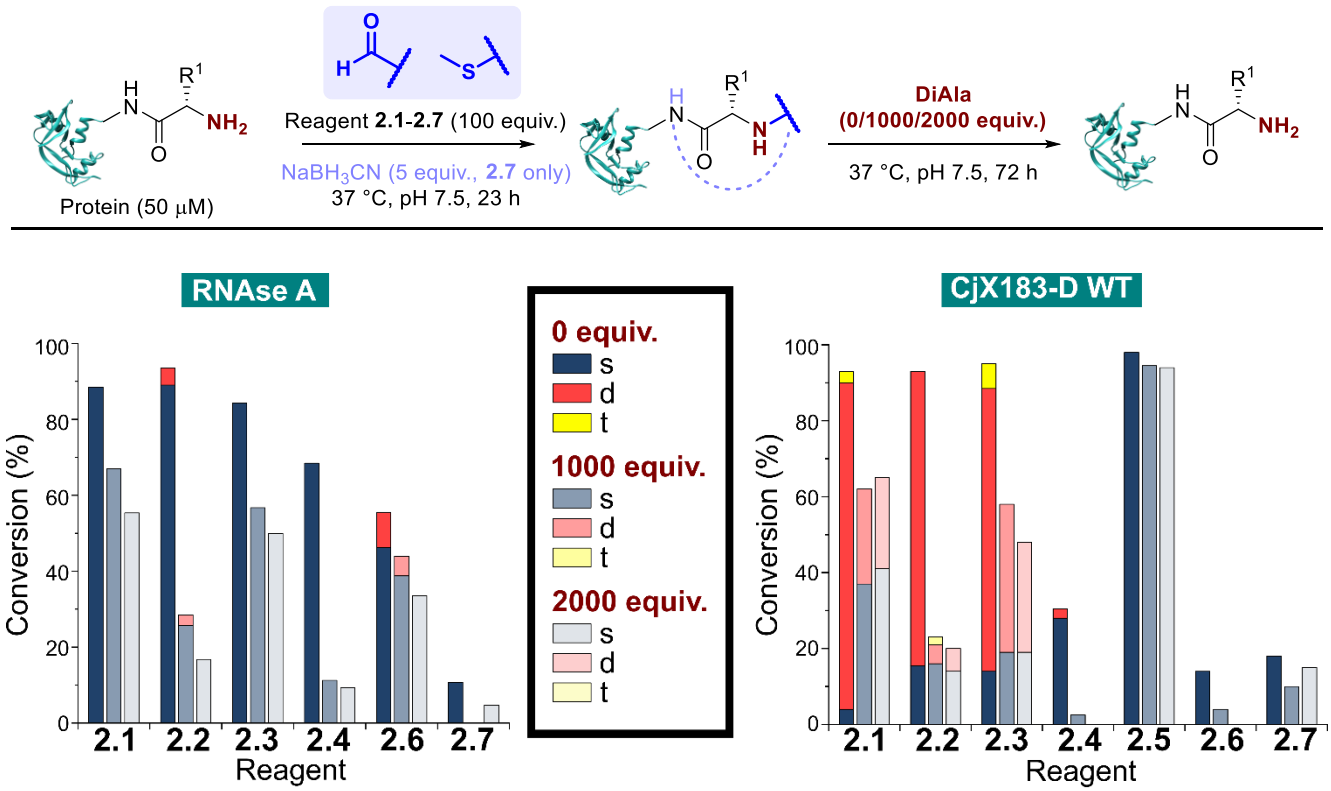


Figure 2.20. Stability of protein conjugates in the presence of DiAla as a competitive N-terminus mimic. After initial conjugation at 37 °C for 23 h, the stated number of equivalents of DiAla (w.r.t. protein) were added without purification, and the samples incubated at 37 °C for 72 h (s = single, d = double, t = triple modification). Modifications with reagent **2.7** were carried out in the presence of $NaBH_3CN$ (5 equiv.). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[15].

For 2-EBA **2.5**, the weak protein signal that resulted from the use of 100 equiv. of reagent (due to poor selectivity at higher reagent loadings, as demonstrated in *Fig. 2.13*) necessitated the experiment be performed with just 25 equiv. w.r.t. RNase A, followed by incubation with 250, or 500 equiv. of DiAla. Under these conditions minimal protein modification was observed for **2.6** and **2.7** and so comparisons to these reagents could not be made. However, relative to **2.1**-**2.4** conjugates formed with **2.5** were found to be the most stable and least sensitive to competitive cleavage (*Fig. 2.21*).

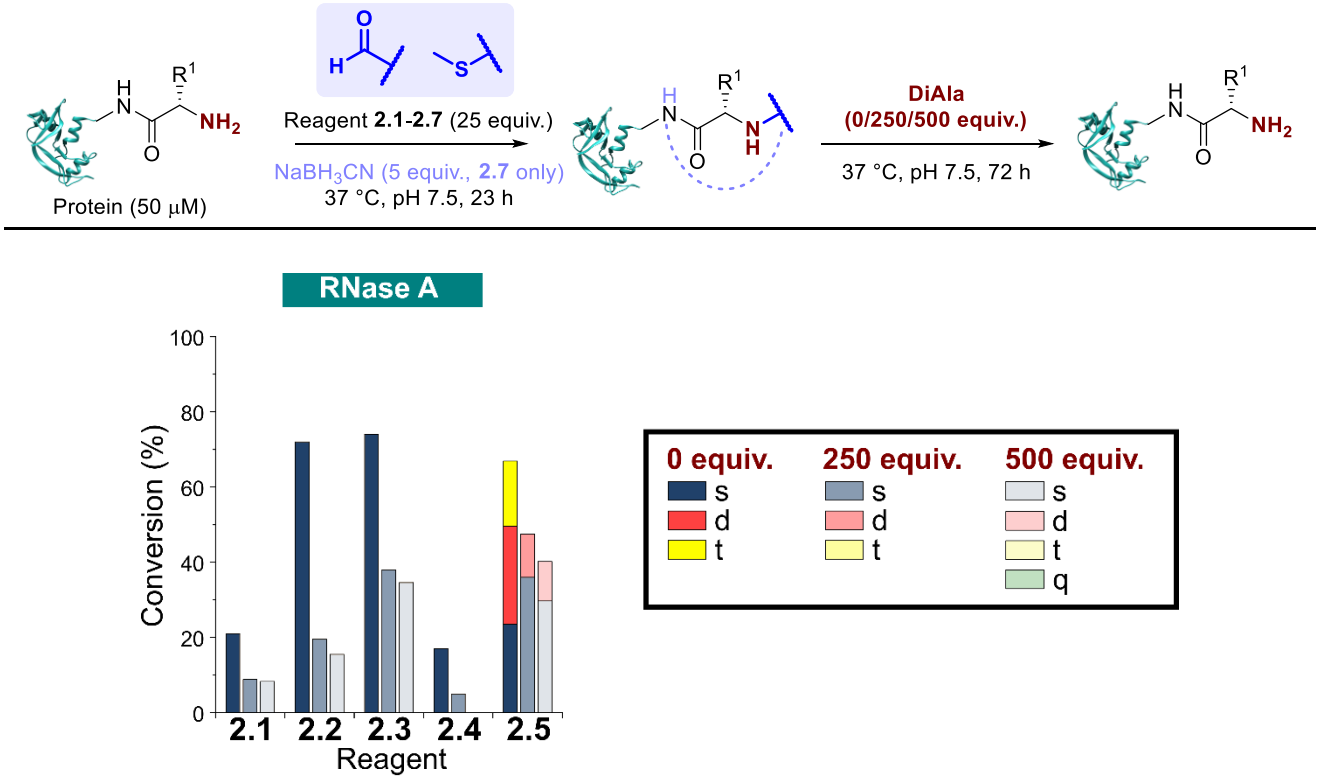


Figure 2.21. Stability of protein conjugates in the presence of DiAla as a competitive N-terminus mimic. After initial conjugation at 37 °C for 23 h, the stated number of equivalents of DiAla (w.r.t. protein) were added without purification, and the samples incubated at 37 °C for 72 h respectively (s = single, d = double, t = triple, q = quadruple modification). Modifications with reagent **2.7** were carried out in the presence of $NaBH_3CN$ (5 equiv.). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[15].

2.2.8 Kinetics of N-terminal modification

Having observed N-terminal conjugate instability for all of the reagents tested, we sought to gain better insight into the underlying conjugation chemistry. In preliminary studies, we attempted to follow conjugate formation over time by UV-vis spectroscopy as we hypothesised that we would observe a shift in absorbance due to expected changes in aromatic conjugation. Unfortunately, we suspected the minimal conversion at low concentrations, in combination with subtle changes in the conjugation of the molecules upon peptide modification, limited the viability of UV-vis spectroscopy as a method to determine kinetic parameters.

Due to poor reactivity at the concentrations required by UV-vis spectroscopy, we next decided to follow the conjugate formation for each of the targets using ^{1}H NMR, so that higher concentrations could be used. DiAla (50 mM) was therefore reacted with reagents **2.1**-**2.7** (50 mM) under second-order reaction conditions at 37 °C in pD 7.3 sodium phosphate buffer, and conjugate formation was followed over time by ^{1}H NMR spectroscopy (*Fig. 2.23a*). For aldehyde species **2.1**-**2.4** and **2.7**, both hydrate and aldehyde forms of the modification reagent were observed, with intermediate imine also observed in reaction mixtures for reagents **2.1**-**2.4**. Ratios of the different species present were calculated *via* the relative integrals of diagnostic ^{1}H NMR signals, for example as illustrated below in *Fig. 2.22* for TA4C **2.4** (see *Fig. S2.11* for representative spectra for all reagents).

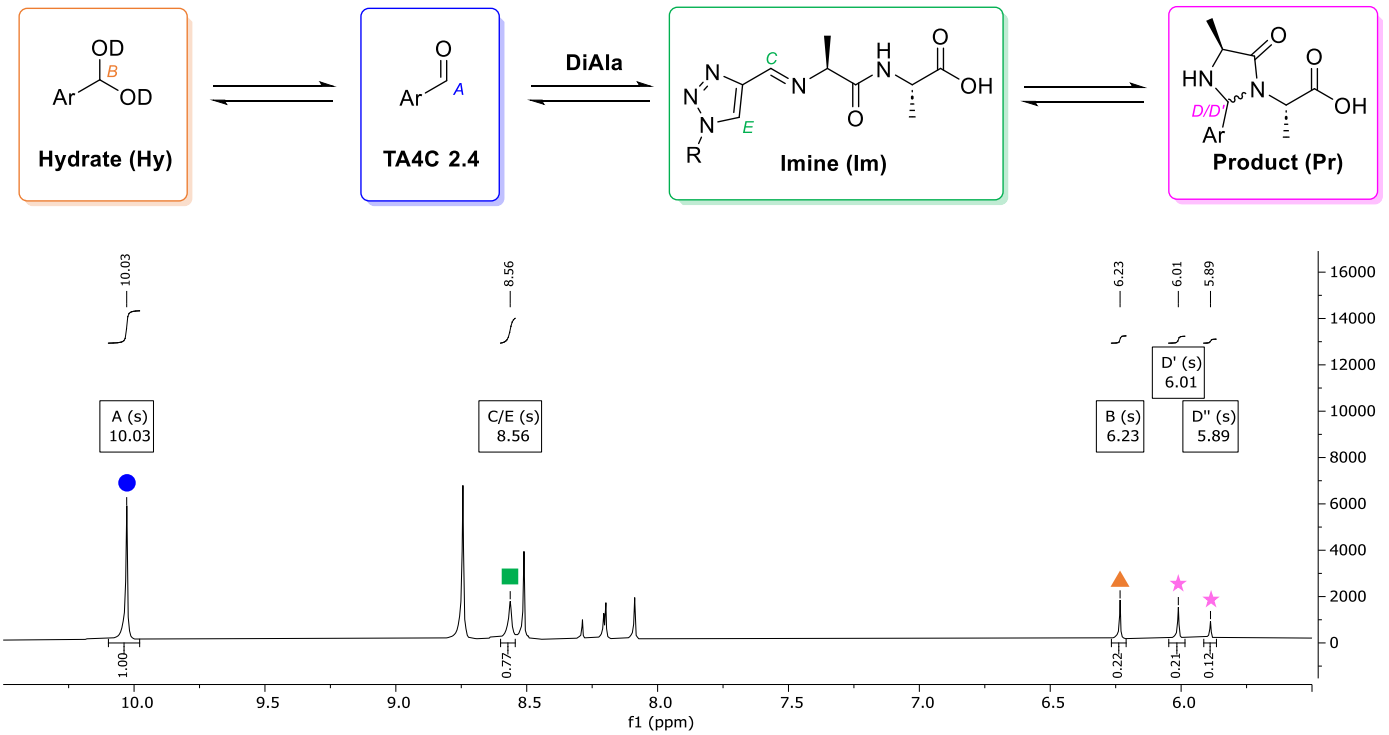
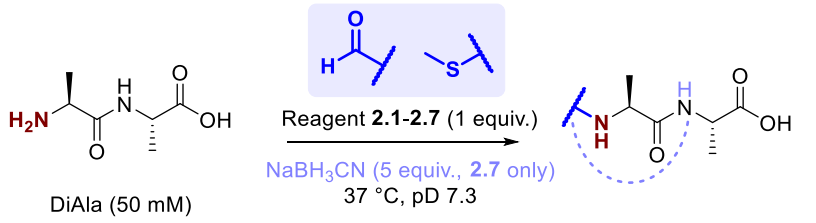


Figure 2.22. Representative ^{1}H NMR spectrum of kinetics experiment for TA4C **2.4** at 16 h time-point, showing the diagnostic signals used for kinetic modelling (**circle** = modification reagent, **star** = imidazolidinone, **triangle** = hydrate species, **circle** = aldehyde species, **square** = imine species).

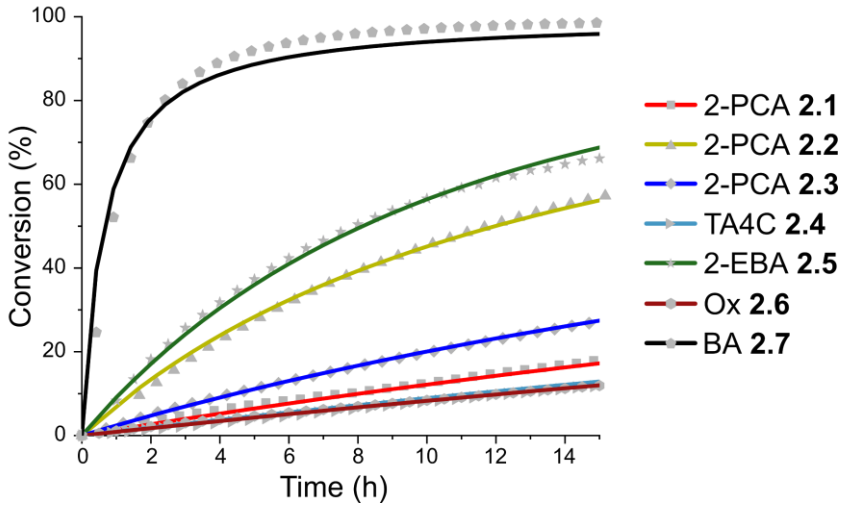
Data were then fit to a reversible or irreversible second-order kinetic model, as appropriate, allowing us to calculate forward and backward rate constants, k_1 and k_{-1} respectively, for conjugate formation, as well as the dissociation constant, K_d, for each reaction where appropriate (*Fig. 2.23b*).[23] Note that the kinetic model did not integrate hydrate/imine formation or imine reduction, and made the assumptions that cyclisation was the rate-limiting step of the reaction for reagents **2.1**-**2.4**, with cyclisation/imine reduction occurring immediately upon imine formation for reagents **2.5** and **2.7** respectively. A more complex and accurate analysis will be discussed later in Chapter 4. These studies revealed that reactions between the N-terminus of DiAla and reagents **2.1**-**2.6** were all fairly slow ($k_1 < 5 \times 10^{-4}$ M^{-1} s^{-1}) relative to many other bioconjugation strategies,[24] explaining the need for large excesses of reagent and long reaction times to drive significant conversions during protein modification. Under these conditions, the data fit an irreversible model for reagents **2.1**, **2.3**, and **2.6**, though this does not preclude reversibility at levels below our threshold for detection (K_d < 20 μM). Interestingly, our data conflicts with those of Tang *et al.* who reported a second order rate constant of ~1 × 10^{-2} M^{-1} s^{-1} for the reaction of an unmodified Ox reagent with the tripeptide GAF.[6] This may indicate that glycine possesses particularly high reactivity with Ox reagents, relative to bulkier α-substituted amino acids, but further studies are required to investigate this effect.

In contrast, for 2-PCA **2.2** (K_d ~ 600 μM), TA4C **2.4** (K_d ~ 4 mM), and 2-EBA **2.5** (K_d ~ 3 mM) significant levels of reversibility were observed. These results partly rationalise the protein conjugate stability data we observed, whereby conjugates were slowly cleaved following the removal of excess reagent, with **2.2** and

2.4 being the most sensitive in many cases (though not always). The high K_d recorded for 2-EBA **2.5** was more surprising, given protein modification with **2.5** seemed to lead to relatively stable conjugation (albeit with challenges associated with controlling the site of modification). This highlights the challenges faced in extrapolating results from small molecule models to challenging protein substrates, where effects from the local environment can make a major contribution, both to stability and instability.



a)



b)

Reagent	k_1 / M^{-1} s^{-1}	k_{-1} / s^{-1}	K_d / mM
2.1	$7.6 \times 10^{-5} \pm 0.1 \times 10^{-5}$	-	-
2.2	$4.2 \times 10^{-4} \pm 0.02 \times 10^{-4}$	$2.5 \times 10^{-7} \pm 0.3 \times 10^{-7}$	0.58
2.3	$1.4 \times 10^{-4} \pm 0.01 \times 10^{-4}$	-	-
2.4	$5.4 \times 10^{-5} \pm 0.1 \times 10^{-5}$	$2.3 \times 10^{-7} \pm 0.2 \times 10^{-7}$	4.27
2.5	$5.4 \times 10^{-4} \pm 0.04 \times 10^{-4}$	$1.6 \times 10^{-6} \pm 0.1 \times 10^{-6}$	2.91
2.6	$5.0 \times 10^{-5} \pm 0.1 \times 10^{-5}$	-	-
2.7	$5.1 \times 10^{-3} \pm 0.3 \times 10^{-3}$	-	-

Figure 2.23. ***a)*** Plots of conversion against time for the reactions of DiAla with reagents **2.1**-**2.7** under second order conditions at a concentration of 50 mM, as measured by ^{1}H NMR analysis. Fits are based on second order reversible or irreversible models. ***b)*** Tabulated forward (k_1) and backward (k_{-1}) rate constants, and dissociation constants (K_d) where relevant, for the reaction of reagents **2.1**-**2.7** with DiAla, as calculated by ^{1}H NMR.

2.3 Conclusions

We have carried out a detailed comparative study of the current leading strategies for N-terminal protein modification, probing the conversion, selectivity, stability, and kinetics of modification under standardised conditions. While protein modification chemistries are typically developed on small molecule or peptide model systems, our results emphasise the important influence protein structure plays in dictating conjugation efficiency – in no case was one reagent found to outperform the others, and large differences in protein-dependent conjugate behaviour were observed throughout the study. Predicting the most suitable reagent for efficient modification of a target protein is therefore challenging, with no 'one size fits all' reagent for N-terminal modification.

Multiple factors are at play, likely intertwined to act in both a synergistic and antagonistic fashion, including steric accessibility at the protein surface, interactions between reagents and the neighbouring environment, and the influence of localized differences in pH/pK_a. For aldehydes **2.1**-**2.5** and **2.7**, this variability is exacerbated by the nature of the reagents themselves, with differing propensities to equilibrate between aldehyde and hydrate, and the dynamic nature of imine chemistry potentially leading to different product structures being formed depending on the local environment. The influence of hydration and imine formation on imidazolidinone formation will be discussed further in Chapter 4. It is therefore critical that results from small molecule and peptide models are adequately validated on a range of protein substrates. Our results advocate for the screening of a panel of reagents to ensure optimal conjugation is realised. Moreover, by highlighting the limitations of current methods, particularly with regards to conjugate stability, this study supports the need for further investigation in the field, to provide new *fast*, *selective*, and *stable* N-terminal modification strategies that address these limitations. Efforts to develop such methods will be discussed in Chapters 3 and 4.

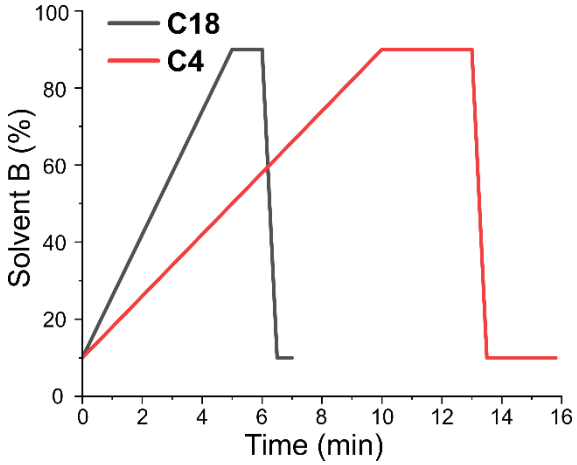
2.4 Experimental

2.4.1 General considerations

Proton nuclear magnetic resonance (^{1}H NMR) spectra were recorded on a Bruker AVII (300 MHz), Jeol ECX-400 (400 MHz), Bruker AVIIIHD (500 MHz) or Bruker AVIIIHD (600 MHz) spectrometer. Carbon nuclear magnetic resonance (^{13}C NMR) spectra were recorded on a Jeol ECX-400 (100 MHz) or Bruker AVII (75 MHz) spectrometer. NMR shifts were assigned using COSY, HSQC and HMBC spectra. 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 Hz with the following splitting abbreviations: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, app = apparent, br = broad. Melting points (m.p.) were recorded on a Gallenkamp melting point apparatus. Infrared (IR) spectra were recorded on a Perkin Elmer UATR Two FT-IR spectrometer or a Bruker Alpha II ATR spectrometer with Opus build 8.5.29. High resolution electrospray ionisation (ESI) mass spectra (HRMS) were recorded on a Bruker Compact TOF-MS or a Jeol AccuTOF GCx-plus spectrometer. Nominal and exact *m/z* values are reported in Daltons (Da).

Thin layer chromatography (TLC) was carried out using aluminium backed sheets coated with 60 F_{254} silica gel (Merck). Visualization of the silica plates was achieved using a UV lamp (λ_{max} = 254 nm), potassium permanganate (5% $KMnO_4$ in 1M NaOH with 5% potassium carbonate), or ninhydrin (1.5% ninhydrin, 3% AcOH in *n*-butanol). Flash column chromatography was carried out using Geduran Si 60 (40-63 μm) (Merck). Mobile phases are reported as % volume of more polar solvent in less polar solvent. Anhydrous solvents were dried over a PureSolv MD 7 Solvent Purification System. Deionized water was used for chemical reactions and for protein manipulations. All other solvents were used as supplied (Analytical or HPLC grade), without prior purification. Reagents were purchased from Sigma-Aldrich, VWR, or Fluorochem and used as supplied, unless otherwise indicated. Brine refers to a saturated solution of sodium chloride. Petrol refers to the fraction of petroleum ether boiling in the range 40-60 °C. Anhydrous magnesium sulfate ($MgSO_4$) was used as the drying agent after reaction workup unless otherwise stated.

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 × 2.1 mm, 2.6 μm particle size) or Accucore C4 (100 × 4.6 mm, 5 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. The gradient was programmed as shown below:

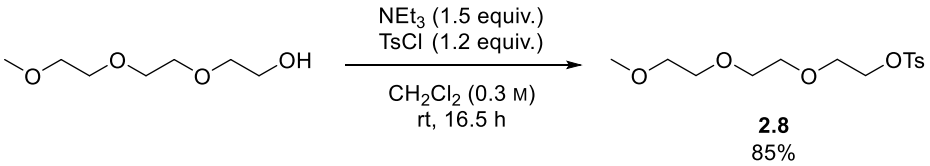


Spectra were analysed using the Bruker Data Analysis 4.4 software. Spectra were charge deconvoluted using ESI Compass 1.3 for RNase A (13000-16000 Da), myoglobin (16000-20000 Da), CjX183-D WT and R51K (11000-15000 Da). Clostripain light chain (LC) spectra were charge deconvoluted using MaxEnt (8000-20000 Da). Data are presented showing the raw ion series MS data on the left, and the deconvoluted spectra on the right. Expected masses were calculated relative to reported [**SM**+H]$^+$ values of 13681 Da (RNase A)[14], 16951 Da (myoglobin)[25] and 14941 Da (Clostripain LC)[1]. 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. **SM** = unmodified starting material, **P** = modified product, s = single modification, d = double modification, t = triple modification, q = quadruple modification, qu = quintuple modification, se = sextuple modification, sp = septuple modification.

2.4.2 Reagent synthesis

i) *2-PCA* ***2.1***

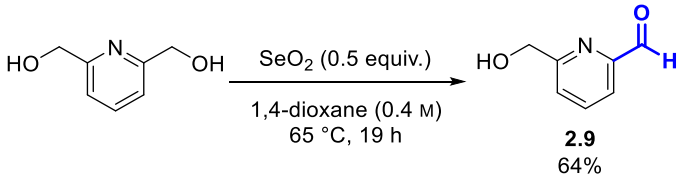
2-(2-(2-Methoxyethoxy)ethoxy)ethyl 4-methylbenzenesulfonate (2.8)



A solution of *p*-toluenesulfonyl chloride (1.88 g, 9.9 mmol, 1.2 equiv.) in anhydrous CH_2Cl_2 (5.8 mL) was added to a solution of triethylene glycol monomethyl ether (1.3 mL, 8.1 mmol, 1.0 equiv.) and triethylamine (1.7 mL, 12.2 mmol, 1.5 equiv.) in anhydrous CH_2Cl_2 (19.3 mL, 0.3 M) under nitrogen, and the reaction mixture was stirred at room temperature for 16.5 h. The reaction was then washed sequentially with saturated aqueous $NaHCO_3$ solution (10 mL) and brine (10 mL), dried over $MgSO_4$, filtered, and the solvent removed under reduced pressure. The resulting pale yellow oil was purified by flash column chromatography (66% EtOAc:Petrol, R_f 0.52 in 75% EtOAc:Petrol), and pure fractions were concentrated under reduced pressure

to afford the title compound (2.20 g, 6.9 mmol, 85%) as a pale yellow oil with spectroscopic data in accordance with the literature.[7] **^{1}H NMR** (300 MHz, $CDCl_3$) δ_H: 2.44 (3H, s, ArMe), 3.36 (3H, s, OMe), 3.49-3.71 (10H, m, -OC$\underline{H}_2$), 4.15 (2H, t, *J* = 4.9 Hz, -C$\underline{H}_2$OTs), 7.33 (2H, d, *J* = 8.3 Hz, Ar$\underline{H}$2), 7.79 (2H, d, *J* = 8.3 Hz, Ar$\underline{H}$3).

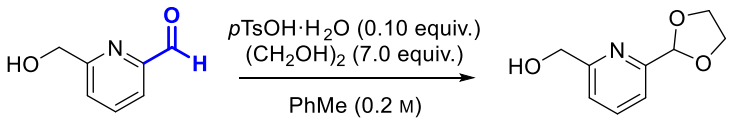
6-(Hydroxymethyl)-2-pyridinecarboxaldehyde (2.9)



2.9
64%

Selenium dioxide (0.59 g, 5.3 mmol, 0.5 equiv.) was added to a solution of 2,6-pyridinedimethanol (1.50 g, 10.8 mmol, 1.0 equiv.) in 1,4-dioxane (30 mL, 0.4 M) at room temperature. The resulting mixture was sonicated for 2 min and then stirred at 65 °C for 19 h. After cooling to room temperature, CH_2Cl_2 (50 mL) was added, the resulting mixture was filtered through Celite, and the filtrate was concentrated under reduced pressure to afford a yellow oil. The residue was purified by flash column chromatography (2.5% MeOH:CH_2Cl_2, R_f 0.20), and pure fractions were concentrated under reduced pressure to afford the title compound (0.95 g, 6.9 mmol, 64%) as a yellow oil with spectroscopic data in accordance with the literature.[1] **^{1}H NMR** (300 MHz, $CDCl_3$) δ_H: 3.59 (1H, br s, -O$\underline{H}$), 4.87 (2H, s, -C$\underline{H}_2$-), 7.47-7.57 (1H, m, Ar$\underline{H}$4), 7.83-7.93 (2H, m, Ar$\underline{H}$3, Ar$\underline{H}$5), 10.08 (1H, s, -C$\underline{H}$O).

(6-(1,3-Dioxolan-2-yl)pyridin-2-yl)methanol (2.14)

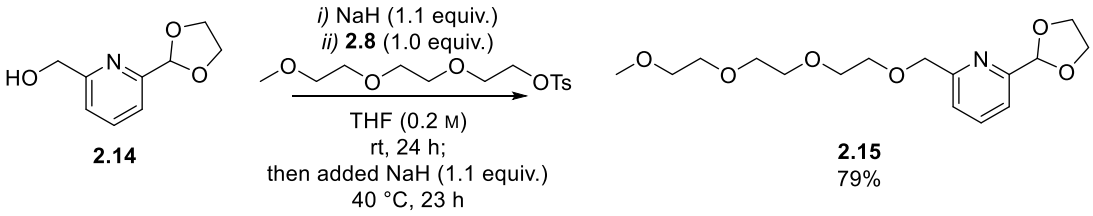


2.9 **2.14**
72%

A mixture of **2.9** (3.56 g, 26 mmol, 1.0 equiv.), *p*-toluenesulfonic acid monohydrate (0.49 g, 2.6 mmol, 0.10 equiv.), and ethylene glycol (10.1 mL, 181 mmol, 7.0 equiv.) in toluene (150 mL, 0.2 M) was refluxed under Dean-Stark conditions for 18 h. After cooling to room temperature, the reaction was concentrated under reduced pressure and the residue purified by flash column chromatography (0-2% MeOH:CH_2Cl_2, R_f 0.38 in 100% EtOAc). Pure fractions were concentrated under reduced pressure to afford the title compound (3.38 g, 18.7 mmol, 72%) as a yellow oil. **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H: 4.03-4.19 (5H, m, -OC$\underline{H}_2$ & -O$\underline{H}$), 4.76 (2H, s, ArC$\underline{H}_2$), 5.84 (1H, s, -C$\underline{H}$O$_2$R), 7.26 (1H, d, *J* = 7.8 Hz, Ar$\underline{H}$5), 7.44 (1H, d, *J* = 7.3 Hz, Ar$\underline{H}$3), 7.67 (1H, dd, *J* = 7.8, 7.3 Hz, Ar$\underline{H}$4); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C: 64.42 (Ar$\underline{C}$H$_2$), 65.80 (-OC$\underline{H}_2$), 103.68 (-$\underline{C}$HO$_2$R), 119.31 (Ar$\underline{C}$3), 121.09 (Ar$\underline{C}$5), 137.77 (Ar$\underline{C}$4), 156.19 (Ar$\underline{C}$2), 159.41 (Ar$\underline{C}$6); $\boldsymbol{\lambda_{max}}$ (ATR, cm^{-1}) 3350, 2890,

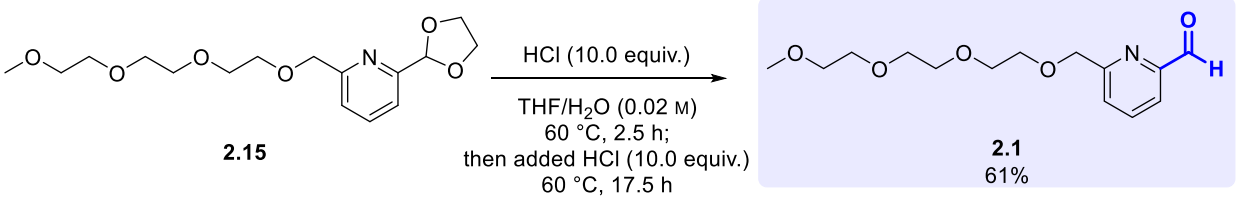
1598, 1455, 1368, 1212, 1104, 1027, 991, 945, 781, 649; **HRMS** (ESI$^+$) $C_9H_{11}NO_3$ [M+Na]$^+$ found 204.0632, calculated 204.0631.

2-(1,3-Dioxolan-2-yl)-6-(2,5,8,11-tetraoxadodecyl)pyridine (2.15)



Sodium hydride (60%, 12 mg, 330 µmol, 1.1 equiv.) was added slowly to a solution of **2.14** (54 mg, 300 µmol, 1.0 equiv.) in anhydrous THF (1.4 mL, 0.2 M), and after 10 min a solution of compound **2.8** (94 mg, 300 µmol, 1.0 equiv.) in anhydrous THF (0.5 mL) was added. The reaction mixture was then stirred at room temperature for 24 h under nitrogen. At this point TLC indicated that the reaction was incomplete, so another portion of sodium hydride (60%, 12 mg, 330 µmol, 1.1 equiv.) was added, and the reaction mixture was stirred at 40 °C for 23 h. After cooling to room temperature, the reaction was quenched through careful dropwise addition of water (5 mL), and the aqueous extracted with ethyl acetate (3 × 20 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The resulting yellow oil was purified by flash column chromatography (75-100% EtOAc:Petrol, then 0-2.5% MeOH:EtOAc, R_f 0.14 in 100% EtOAc). Pure fractions were concentrated under reduced pressure to afford the title compound (77 mg, 235 µmol, 79%) as a yellow oil. **^{1}H NMR** (300 MHz, $CDCl_3$) δ_H: 3.35 (3H, s, Me), 3.49-3.57 (2H, m, -C$\underline{H}_2$OMe), 3.59-3.76 (10H, m, -C$\underline{H}_2$C$\underline{H}_2$OC$\underline{H}_2$C$\underline{H}_2$OC$\underline{H}_2$CH$_2$OMe), 3.99-4.21 (4H, m, -C$\underline{H}_2$OCHAr), 4.71 (2H, s, -C$\underline{H}_2$Ar), 5.81 (1H, s, -C$\underline{H}$Ar), 7.42 (1H, dd, *J* = 7.7, 1.1 Hz, Ar$\underline{H}$3), 7.49 (1H, dd, *J* = 7.7, 1.1 Hz, Ar$\underline{H}$5), 7.74 (1H, dd, J_1 = J_2 = 7.7 Hz, Ar$\underline{H}$4); **^{13}C NMR** (75 MHz, $CDCl_3$) δ_C: 59.10 (Me), 65.63 (-$\underline{C}$H$_2$OCHAr), 70.38 (-$\underline{C}$H$_2$OCH$_2$Ar), 70.62 (-$\underline{C}$H$_2$-), 70.63 (-$\underline{C}$H$_2$-), 70.73 (-$\underline{C}$H$_2$-), 70.77 (-$\underline{C}$H$_2$-), 72.03 (-$\underline{C}$H$_2$OMe), 73.83 (-$\underline{C}$H$_2$Ar), 103.52 (-$\underline{C}$HAr), 119.18 (Ar$\underline{C}$3), 121.65 (Ar$\underline{C}$5), 137.68 (Ar$\underline{C}$4), 156.19 (Ar$\underline{C}$2), 158.58 (Ar$\underline{C}$6); **λ_{max}** (ATR, cm^{-1}) 2875, 1596, 1580, 1459, 1350, 1280, 1247, 1200, 1099, 1027, 988, 966, 945, 885, 851, 804, 751, 725, 651, 548; **HRMS** (ESI$^+$) $C_{16}H_{26}NO_6$ [M+H]$^+$ found 328.1756, calculated 328.1755.

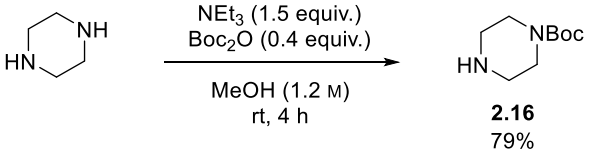
6-(2,5,8,11-Tetraoxadodecyl)picolinaldehyde (2.1)



HCl (2 M in water, 1 mL, 2.0 mmol, 10.0 equiv.) was added to a solution of compound **2.15** (65 mg, 200 µmol, 1.0 equiv.) in THF (12 mL, 0.02 M), and the reaction mixture was stirred at 60 °C for 2.5 h under nitrogen. At this time TLC indicated that the reaction was incomplete, so another portion of HCl (2 M in water, 1 mL, 2.0 mmol, 10.0 equiv.) was added and the reaction mixture was stirred at 60 °C for a further 17.5 h. After cooling to room temperature, the solvent was removed under reduced pressure and the residue obtained was purified by flash column chromatography (10% MeOH:CH_2Cl_2, R_f 0.24). ^{1}H NMR revealed the fractions collected were impure, so the resulting yellow residue was purified by flash column chromatography (75% EtOAc:Petrol). Pure fractions were concentrated under reduced pressure to afford the title compound (34 mg, 0.12 mmol, 61%) as a yellow oil. **^{1}H NMR** (400 MHz, DMSO-d_6) $\delta_{H:}$ 3.22 (3H, s, Me), 3.38-3.44 (2H, m, -C<u>H</u>$_2$OMe), 3.48-3.56 (6H, m, -C<u>H</u>$_2$C<u>H</u>$_2$OC<u>H</u>$_2$CH$_2$OMe), 3.58-3.64 (2H, m, -C<u>H</u>$_2$CH$_2$OCH$_2$Ar), 3.64-3.70 (2H, m, -C<u>H</u>$_2$OCH$_2$Ar), 4.70 (2H, s, -C<u>H</u>$_2$Ar), 7.75 (1H, d, *J* = 7.7 Hz, Ar<u>H</u>5), 7.85 (1H, d, *J* = 7.7 Hz, Ar<u>H</u>3), 8.06 (1H, dd, J_1 = J_2 = 7.7 Hz, Ar<u>H</u>4), 9.96 (1H, s, C<u>H</u>O); **^{13}C NMR** (101 MHz, DMSO-d_6) $\delta_{C:}$ 58.06 (Me), 69.62 (-<u>C</u>H$_2$CH$_2$-), 69.74 (-<u>C</u>H$_2$CH$_2$OCH$_2$Ar), 69.81 (-<u>C</u>H$_2$CH$_2$-), 69.89 (-<u>C</u>H$_2$CH$_2$- and -<u>C</u>H$_2$OCH$_2$Ar), 71.29 (-<u>C</u>H$_2$OMe), 72.68 (-<u>C</u>H$_2$Ar), 120.50 (Ar<u>C</u>3), 125.82 (Ar<u>C</u>5), 138.40 (Ar<u>C</u>4), 151.67 (Ar<u>C</u>2), 159.42 (Ar<u>C</u>6), 193.55 (<u>C</u>HO); **λ_{max}** (ATR, cm^{-1}) 2923, 2871, 1713, 1593, 1457, 1351, 1250, 1210, 1102, 992, 943, 851, 799, 782, 661, 619; **HRMS** (ESI$^+$) $C_{14}H_{22}NO_5$ $[M+H]^+$ found 284.1497, calculated 284.1492.

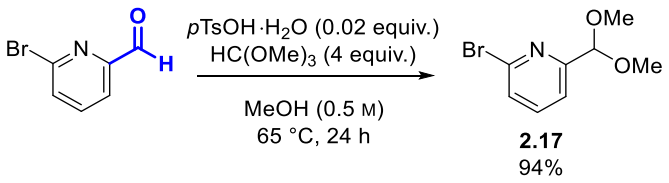
ii) *2-PCA* ***2.2***

***tert*-Butyl piperazine-1-carboxylate (2.16)**



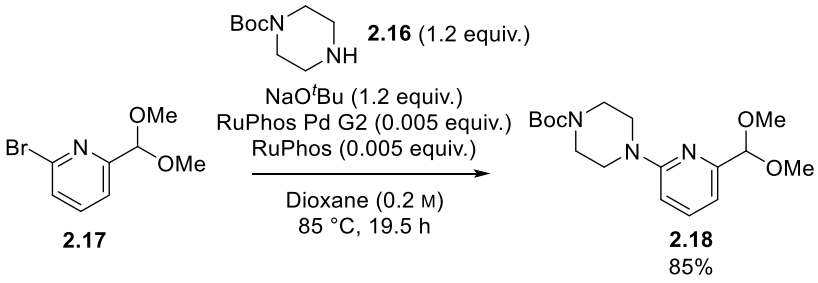
A solution of di-*tert*-butyl dicarbonate (1.14 g, 5.2 mmol, 0.4 equiv.) in methanol (5 mL) was added dropwise to a solution of triethylamine (2.8 mL, 20.1 mmol, 1.5 equiv.) and piperazine (1.16 g, 13.4 mmol, 1.0 equiv.) in methanol (12 mL, 1.2 M) over 5 min. The reaction mixture was stirred at room temperature for 4 h and then concentrated under reduced pressure. Ethyl acetate (20 mL) was added and resulting white precipitate was removed by vacuum filtration. The filtrate was washed with water (3 × 5 mL), dried over $MgSO_4$, filtered, and concentrated under reduced pressure to afford the title compound (0.79 g, 4.3 mmol, 79%) as a white solid with spectroscopic data in accordance with the literature.[26] **^{1}H NMR** (300 MHz, $CDCl_3$) $\delta_{H:}$ 1.46 (9H, s, Boc), 2.80 (4H, t, *J* = 5.3 Hz, -C<u>H</u>$_2$NBoc), 3.38 (4H, t, *J* = 5.3 Hz, -C<u>H</u>$_2$NH); **mp** 46-49 °C {Lit.[26] 46 °C}.

2-Bromo-6-(dimethoxymethyl)pyridine (2.17)



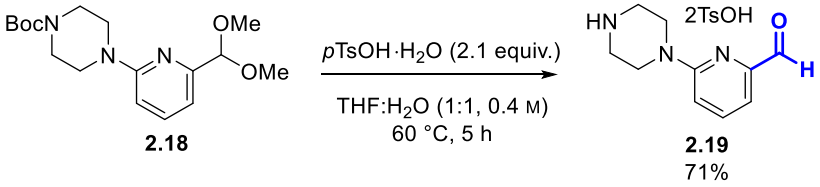
Trimethyl orthoformate (1.2 mL, 10.9 mmol, 4.0 equiv.) was added to a solution of *p*-toluenesulfonic acid monohydrate (10 mg, 0.1 mmol, 0.02 equiv.) and 6-bromo-2-pyridinecarbaldehyde (0.51 g, 2.7 mmol, 1.0 equiv.) in methanol (5 mL, 0.5 M), and the reaction mixture was heated under reflux for 24 h. After cooling to room temperature, the mixture was concentrated under reduced pressure. The residue obtained was partitioned between CH_2Cl_2 (25 mL) and saturated aqueous $NaHCO_3$ solution (20 mL). The layers were separated, and the aqueous layer was extracted with CH_2Cl_2 (5 × 25 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated under reduced pressure to afford the title compound (0.59 g, 2.5 mmol, 94%) as a brown oil with spectroscopic data in accordance with the literature.[9] **^{1}H NMR** (300 MHz, $CDCl_3$) δ_H: 3.41 (6H, s, -OMe), 5.30 (1H, s, CH(OMe)$_2$), 7.45 (1H, dd, *J* = 7.6, 1.2 Hz, ArH5), 7.52 (1H, dd, *J* = 7.6, 1.2 Hz, ArH3), 7.59 (1H, dd, $J_1 = J_2$ = 7.6 Hz, ArH4).

***tert*-Butyl 4-(6-(dimethoxymethyl)pyridine-2-yl)piperazine-1-carboxylate (2.18)**



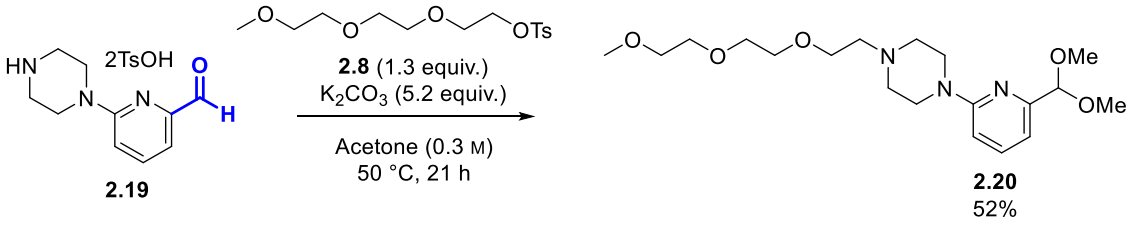
Anhydrous 1,4-dioxane (4 mL, 0.2 M) was added to a mixture of acetal **2.17** (0.40 g, 1.7 mmol, 1.0 equiv.), piperazine **2.16** (0.39 g, 2.1 mmol, 1.2 equiv.) and sodium *tert*-butoxide (0.20 g, 2.1 mmol, 1.2 equiv.) under nitrogen. A solution of RuPhos (4 mg, 0.01 mmol, 0.005 equiv.) and RuPhos Pd G2 (7 mg, 0.01 mmol, 0.005 equiv.) in anhydrous 1,4-dioxane (4 mL) was then added and the reaction mixture was stirred at 85 °C for 19.5 h. After cooling to room temperature, the mixture was partitioned between ethyl acetate (30 mL) and water (10 mL). The aqueous layer was extracted with ethyl acetate (3 × 30 mL) and the combined organics were dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The orange residue obtained was purified by flash column chromatography (5% NEt_3, 10% EtOAc:Petrol, R_f 0.31). Pure fractions were concentrated under reduced pressure to afford the title compound (0.50 g, 1.5 mmol, 85%) as a colourless oil with spectroscopic data in accordance with the literature.[2] **^{1}H NMR** (300 MHz, $CDCl_3$) δ_H: 1.48 (9H, s, Boc), 3.40 (6H, s, OMe), 3.53 (8H, app s, -CH$_2$CH$_2$NBoc), 5.17 (1H, s, -CH(OMe)$_2$), 6.59 (1H, d, *J* = 8.5, ArH3), 6.88 (1H, d, *J* = 7.3 Hz, ArH5), 7.52 (1H, dd, *J* = 8.5, 7.3 Hz, ArH4).

6-(Piperazin-1-yl)picolinaldehyde ditosylate salt (2.19)



Acetal **2.18** (55 mg, 163 µmol, 1.0 equiv.) and *p*-toluenesulfonic acid monohydrate (59 mg, 342 µmol, 2.1 equiv.) were dissolved in THF:H_2O (1:1, 0.9 mL, 0.2 M), and the resulting solution was stirred at 60 °C for 5 h. After cooling to room temperature, methanol (2 mL) was added, followed by diethyl ether (30 mL) causing precipitation. The orange residue was collected by filtration, and then dissolved in methanol (10 mL). Subsequent removal of the solvent under reduced pressure gave the title compound (62 mg, 115 µmol, 71%) as an orange residue with spectroscopic data in accordance with the literature.[2] **^{1}H NMR** (400 MHz, DMSO-d_6) δ_H: 2.28 (6H, s, Me), 3.22 (4H, t, *J* = 5.3 Hz, -C$\underline{H}_2$NR), 3.81 (4H, t, *J* = 5.3 Hz, -C$\underline{H}_2$NR), 7.13 (4H, d, *J* = 7.7 Hz, TsAr$\underline{H}$3), 7.25 (1H, d, *J* = 8.6 Hz, Ar$\underline{H}$3), 7.30 (1H, d, *J* = 7.2 Hz, Ar$\underline{H}$5), 7.50 (4H, d, *J* = 7.7 Hz, TsAr$\underline{H}$2), 7.83 (1H, dd, *J* = 8.6, 7.2 Hz, Ar$\underline{H}$4), 9.82 (1H, s, -C$\underline{H}$O).

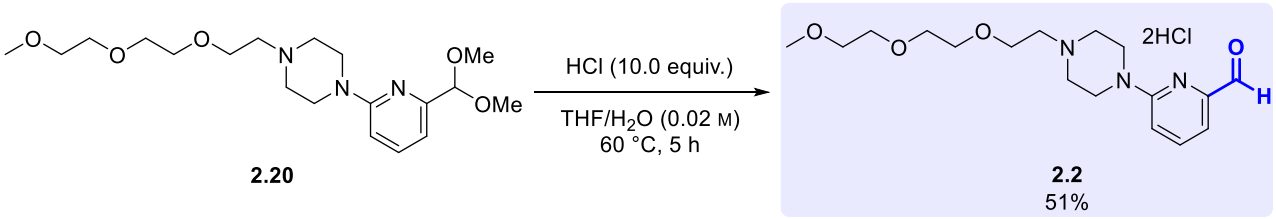
1-(6-(Dimethoxymethyl)pyridin-2-yl)-4-(2-(2-(2-methoxyethoxy)ethoxy)ethyl)piperazine (2.20)



A solution of compound **2.19** (79 mg, 248 µmol, 1.4 equiv.) in acetone (0.34 mL) was added to a mixture of K_2CO_3 (133 mg, 887 µmol, 5.3 equiv.) and compound **2.8** (98 mg, 183 µmol, 1.0 equiv.) in acetone (0.34 mL), and the reaction mixture was stirred at 50 °C for 21 h. The solvent evaporated over this time period and, after cooling to room temperature, the resulting residue was dissolved in CH_2Cl_2 (30 mL). The organics were washed sequentially with water (10 mL) and brine (10 mL), dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The brown residue obtained was purified by flash column chromatography (10% MeOH:CH_2Cl_2, R_f 0.30). Pure fractions were concentrated under reduced pressure to afford the title compound (32 mg, 84 µmol, 52%) as an orange oil, as a 3:97 mixture of aldehyde:dimethyl-acetal. *Spectroscopic data is provided for the major dimethyl-acetal product:* **^{1}H NMR** (300 MHz, $CDCl_3$) δ_H: 2.54-2.71 (6H, m, -C$\underline{H}_2$NC$\underline{H}_2$CH$_2$NAr), 3.37 (3H, s, -CH$_2$O$\underline{Me}$), 3.38 (6H, s, ArCH(O$\underline{Me}$)$_2$), 3.51-3.60 (6H, m, -C$\underline{H}_2$NAr & -C$\underline{H}_2$OMe), 3.61-3.69 (8H, m, -C$\underline{H}_2$OC$\underline{H}_2$C$\underline{H}_2$OC$\underline{H}_2$CH$_2$OMe), 5.15 (1H, s, ArC$\underline{H}$OMe), 6.56 (1H, d, *J* = 8.5 Hz, Ar$\underline{H}$3), 6.83 (1H, d, *J* = 7.3 Hz, Ar$\underline{H}$5), 7.48 (1H, dd, *J* = 8.5, 7.3 Hz, Ar$\underline{H}$4); **^{13}C NMR** (75 MHz, $CDCl_3$) δ_C: 45.12 (-$\underline{C}$H$_2$NAr), 53.50 (-$\underline{C}$H$_2$CH$_2$NAr), 53.87 (ArCHO$\underline{Me}$), 57.98 (-$\underline{C}$H$_2$NCH$_2$CH$_2$NAr), 59.15 (-CH$_2$O$\underline{Me}$), 69.01 (-$\underline{C}$H$_2$CH$_2$NCH$_2$CH$_2$NAr), 70.51 (-$\underline{C}$H$_2$-), 70.65 (-$\underline{C}$H$_2$-), 70.74 (-$\underline{C}$H$_2$-), 72.07 (-$\underline{C}$H$_2$OMe),

104.84 (ArCHOMe), 106.62 (ArC3), 110.29 (ArC5), 137.95 (ArC4), 155.47 (ArC6), 159.06 (ArC2); λ_{max} (ATR, cm^{-1}) 2827, 1798, 1762, 1593, 1451, 1398, 1334, 1252, 1109, 1052, 1004, 985, 962, 851, 795, 733, 612, 570, 512; **HRMS** (ESI^+) $C_{19}H_{34}N_3O_5$ $[M+H]^+$ found 384.2501, calculated 384.2493.

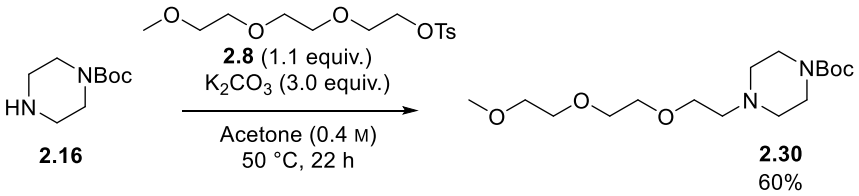
6-(4-(2-(2-(2-Methoxyethoxy)ethoxy)ethyl)piperazin-1-yl)picolinaldehyde dihydrochloride salt (2.2)



HCl (2 M in water, 422 µL, 840 µmol, 10.0 equiv.) was added to a solution of compound **2.20** (32 mg, 84 µmol, 1.0 equiv.) in THF (4.2 mL, 0.02 M) and the reaction mixture was stirred at 60 °C for 5 h. After cooling to room temperature, the mixture was concentrated under reduced pressure and the residue was purified by flash column chromatography (10% MeOH:CH_2Cl_2, R_f 0.30). Pure fractions were concentrated under reduced pressure to afford the title compound (14 mg, 0.04 mmol, 51%) as a yellow oil. **^{1}H NMR** (500 MHz, DMSO-d_6) δ_H: 3.08-3.19 (2H, m, -CH$_2$-), 3.24 (3H, s, Me), 3.29-3.36 (2H, m, -CH$_2$-), 3.37-3.65 (12H, m, -NCH$_2$CH$_2$NAr & -CH$_2$-), 3.84-3.90 (2H, m, -CH$_2$-), 4.43-4.49 (2H, m, -CH$_2$-), 7.28 (1H, *J* = 8.2 Hz, ArH3), 7.30 (1H, d, *J* = 7.2 Hz, ArH5), 7.84 (1H, dd, *J* = 8.2, 7.2 Hz, ArH4), 9.82 (1H, s, CHO). The compound was used without further analysis or purification and stored at -20 °C due to sensitivity to oxidation.

iii) 2-PCA ***2.3***

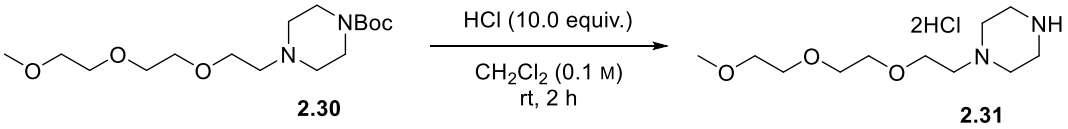
Tert-butyl 4-(2-(2-(2-methoxyethoxy)ethoxy)ethyl)piperazine-1-carboxylate (2.30)



Potassium carbonate (1.28 g, 9.3 mmol, 3.0 equiv.) was added to a solution of piperazine **2.16** (0.58 g, 3.1 mmol, 1.0 equiv.) in acetone (6.2 mL, 0.4 M) and the reaction mixture was stirred at room temperature for 10 min. A solution of compound **2.8** (1.08 g, 3.4 mmol, 1.1 equiv.) in acetone (1 mL) was then added and the reaction mixture was heated to 50 °C for 22 h. After cooling to room temperature, the reaction was concentrated under reduced pressure, and the residue dissolved in ethyl acetate (20 mL). The organics were washed with water (2 × 10 mL) and brine (10 mL), dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The residue was purified by flash column chromatography (5% MeOH:CH_2Cl_2, R_f 0.35 in 10% MeOH:CH_2Cl_2). Pure fractions were concentrated under reduced pressure to afford the title compound

(0.61 g, 1.8 mmol, 60%) as a yellow oil. **^{1}H NMR** (400 MHz, $CDCl_3$) $\delta_{H:}$ 1.44 (9H, s, Boc), 2.43 (4H, t, *J* = 5.1 Hz, -C$\underline{H_2}$CH$_2$NBoc), 2.58 (2H, t, *J* = 5.8 Hz, -C$\underline{H_2}$NCH$_2$CH$_2$NBoc), 3.37 (3H, s, OMe), 3.42 (4H, t, *J* = 5.1 Hz, -C$\underline{H_2}$NBoc), 3.51-3.55 (2H, m, -OC$\underline{H_2}$-), 3.58-3.65 (8H, m, -OC$\underline{H_2}$-); **^{13}C NMR** (101 MHz, $CDCl_3$) $\delta_{C:}$ 28.54 (C$\underline{Me_3}$), 43.20 (-$\underline{C}$H$_2$NBoc), 53.47 (-$\underline{C}$H$_2$CH$_2$NBoc), 57.93 (-$\underline{C}$H$_2$NCH$_2$CH$_2$NBoc), 59.17 (OMe), 68.92 (-CH$_2$-), 70.47 (-CH$_2$-), 70.65 (-CH$_2$-), 70.74 (-CH$_2$-), 72.05 (-CH$_2$-), 79.69 ($\underline{C}$Me$_3$), 154.87 -N$\underline{C}$O$_2$CMe$_3$); **λ_{max}** (ATR, cm^{-1}) 3574, 2975, 2868, 2816, 1693, 1456, 1418, 1365, 1291, 1244, 1169, 1108, 1024, 1004, 939, 866, 769, 576, 531, 461; **HRMS** (ESI$^+$) $C_{16}H_{33}N_2O_5$ $[M+H]^+$ found 333.2389, calculated 333.2384.

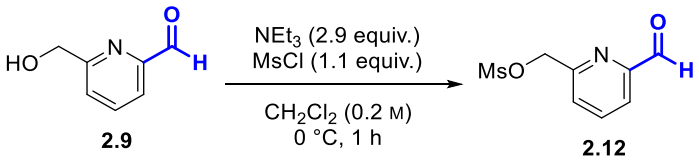
1-(2-(2-(2-Methoxyethoxy)ethoxy)ethyl)piperazine (2.31)



2.30 **2.31**

HCl (4 M in 1,4-dioxane, 3.8 mL, 15.2 mmol, 10.0 equiv.) was added to a solution of compound **2.30** (0.50 g, 1.5 mmol, 1.0 equiv.) in CH_2Cl_2 (13 mL, 0.1 M) and the reaction mixture was stirred at room temperature for 2 h. The reaction mixture was then concentrated under reduced pressure and azeotroped with CH_2Cl_2 (3 × 25 mL). The brown oil obtained (0.62 g) was used immediately in the subsequent reaction without characterisation.

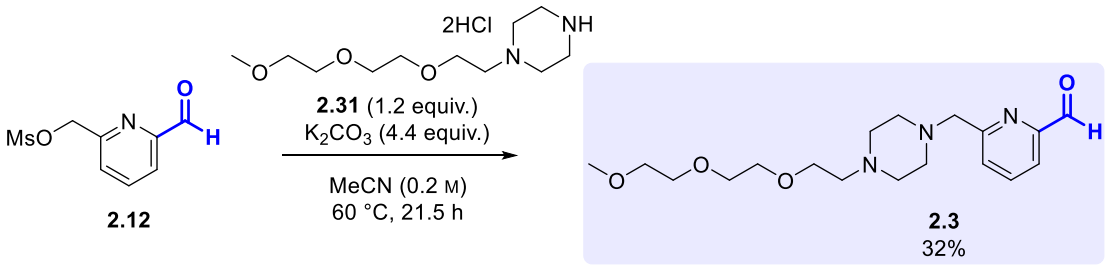
6-(Formylpyridin-2-yl)methyl methanesulfonate (2.12)



2.9 **2.12**

Methanesulfonyl chloride (0.06 mL, 803 µmol, 1.1 equiv.) was added dropwise to a solution of triethylamine (0.30 mL, 2.2 mmol, 2.9 equiv.) and compound **2.9** (0.10 g, 730 µmol, 1.0 equiv.) in anhydrous CH_2Cl_2 (4 mL, 0.2 M) under nitrogen at 0 °C, and stirred at 0 °C for 1 h. The reaction mixture was then quenched with saturated aqueous $NaHCO_3$ solution (10 mL) and the aqueous extracted with CH_2Cl_2 (3 × 30 mL). The combined organics were dried over Na_2SO_4 and concentrated under reduced pressure. The brown oil obtained (0.14 g) was used immediately in the subsequent reaction without characterisation.

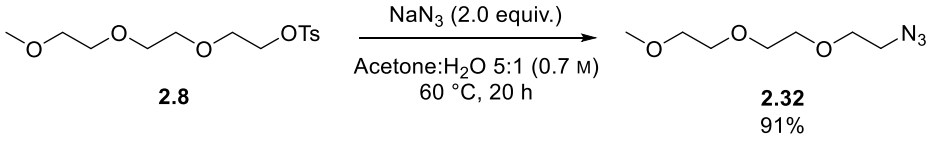
6-((4-(2-(2-(2-Methoxyethoxy)ethoxy)ethyl)piperazin-1-yl)methyl)picolinaldehyde (2.3)



A solution of compound **2.31** (0.27 g, 900 µmol, 1.2 equiv.) in acetonitrile (4 mL, 0.2 M) was added to compound **2.12** (0.16 g, 750 µmol, 1.0 equiv.), and K_2CO_3 (0.45 g, 3.3 mmol, 4.4 equiv.) was then added to the reaction mixture. The reaction was stirred at 60 °C for 21.5 h and, after cooling to room temperature, the solvent was removed under reduced pressure. CH_2Cl_2 (30 mL) and saturated aqueous $NaHCO_3$ solution (25 mL) were added to the orange residue obtained. The aqueous layer was separated and extracted with CH_2Cl_2 (3 × 25 mL), and the combined organics were then dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The orange residue obtained was purified by flash column chromatography (10% MeOH:CH_2Cl_2, R_f 0.16), and pure fractions were concentrated under reduced pressure to afford the title compound (84 mg, 239 mmol, 32%) as a yellow oil. **^{1}H NMR** (400 MHz, $CDCl_3$) $\delta_{H:}$ 2.51-2.64 (10H, m, -NC$\underline{H}_2$C$\underline{H}_2$N-, -OC$\underline{H}_2$-), 3.35 (3H, s, OMe), 3.50-3.54 (2H, m, -OC$\underline{H}_2$-), 3.58-3.65 (8H, m, -OC$\underline{H}_2$-), 3.75 (2H, s, -NC$\underline{H}_2$Ar), 7.63-7.69 (1H, m, Ar$\underline{H}$5), 7.79-7.84 (2H, m, Ar$\underline{H}$3,4), 10.05 (1H, s, C$\underline{H}$O); **^{13}C NMR** (101 MHz, $CDCl_3$) $\delta_{C:}$ 53.25 (-$\underline{C}$H$_2$NCH$_2$Ar), 53.62 (-$\underline{C}$H$_2$CH$_2$NCH$_2$Ar), 57.81 (-O$\underline{C}$H$_2$-), 59.14 (OMe), 64.12 (-$\underline{C}$H$_2$Ar), 68.97 (-O$\underline{C}$H$_2$-), 70.44 (-O$\underline{C}$H$_2$-), 70.61 (-O$\underline{C}$H$_2$-), 70.70 (-O$\underline{C}$H$_2$-), 72.02 (-$\underline{C}$H$_2$OMe), 120.28 (Ar$\underline{C}$3), 127.54 (Ar$\underline{C}$5), 137.43 (Ar$\underline{C}$4), 152.43 (Ar$\underline{C}$2), 159.85 (Ar$\underline{C}$6), 193.81 ($\underline{C}$HO); **λ_{max}** (ATR, cm^{-1}) 2874, 2812, 1712, 1591, 1456, 1352, 1303, 1255, 1200, 1105, 1013, 991, 936, 909, 833, 783, 734, 664, 619; **HRMS** (ESI$^+$) $C_{18}H_{30}N_3O_4$ $[M+H]^+$ found 352.2228, calculated 352.2231.

iv) *TA4C* ***2.4***

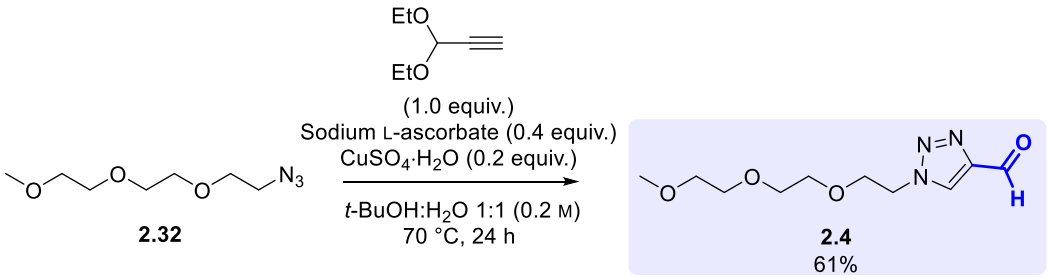
1-Azido-2-(2-(2-methoxyethoxy)ethoxy)ethane (2.32)



Sodium azide (0.54 g, 8.3 mmol, 2.0 equiv.) was added to a solution of compound **2.8** (1.33 g, 4.2 mmol, 1.0 equiv.) in acetone:water (5:1, 6 mL, 0.7 M), and the reaction mixture was stirred at 60 °C for 20 h. After cooling to room temperature, the acetone was removed under reduced pressure. The remaining solution was diluted with water (20 mL) and extracted with diethyl ether (4 × 20 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated under reduced pressure to afford the title compound as a pink oil (0.72 g,

3.8 mmol, 91%) with spectroscopic data in accordance with the literature.[11] **^{1}H NMR** (300 MHz, $CDCl_3$) δ_H: 3.31-3.40 (5H, m, OMe & -C$\underline{H}_2$N$_3$), 3.49-3.55 (2H, m, -C$\underline{H}_2$CH$_2$N$_3$), 3.59-3.69 (8H, m, -OC$\underline{H}_2$-).

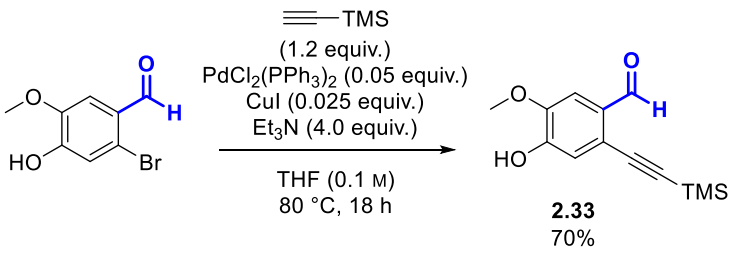
1-(2-(2-(2-Methoxyethoxy)ethoxy)ethyl)-1*H*-1,2,3-triazole-4-carbaldehyde (2.4)



Azide **2.32** (0.10 g, 530 µmol, 1.0 equiv.) was dissolved in a mixture of *t*-BuOH:water (1:1, 2.7 mL, 0.2 M), and the solution was degassed by N_2 bubbling for 25 min. $CuSO_4{\cdot}5H_2O$ (33 mg, 105 µmol, 0.2 equiv.), sodium L-ascorbate (43 mg, 210 µmol, 0.4 equiv.) and 3,3-diethoxyprop-1-yne (75 µL, 530 µmol, 1.0 equiv.) were then added under nitrogen and the reaction mixture was stirred at 70 °C for 24 h. After cooling to room temperature, the reaction was diluted with EtOAc (30 mL) and the organics washed with brine (10 mL). The aqueous layer was extracted with EtOAc (3 × 25 mL), and the combined organics were dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The residue obtained was purified by flash column chromatography (EtOAc, R_f 0.27), and pure fractions were concentrated under reduced pressure to afford the title compound (79 mg, 325 µmol, 61%) as a colourless oil. **^{1}H NMR** (300 MHz, $CDCl_3$) δ_H: 3.35 (3H, s, OMe), 3.48-3.69 (8H, m, MeOC$\underline{H}_2$C$\underline{H}_2$OC$\underline{H}_2$C$\underline{H}_2$-), 3.88 (2H, t, *J* = 4.9 Hz, ArCH$_2$C$\underline{H}_2$-), 4.60 (2H, t, *J* = 4.9 Hz, ArC$\underline{H}_2$-), 8.37 (1H, s, Ar$\underline{H}$5), 10.12 (1H, s, C$\underline{H}$O); **^{13}C NMR** (75 MHz, $CDCl_3$) δ_C: 50.53 (-$\underline{C}$H$_2$Ar), 58.96 (OMe), 68.83 (-$\underline{C}$H$_2$CH$_2$Ar), 70.42 (-$\underline{C}$H$_2$), 70.50 (-$\underline{C}$H$_2$), 70.53 (-$\underline{C}$H$_2$), 71.84 (-$\underline{C}$H$_2$), 126.89 (Ar$\underline{C}$5), 147.75 (Ar$\underline{C}$4), 185.00 ($\underline{C}$HO); **λ_{max}** (ATR, cm^{-1}) 3128, 2872, 1695, 1531, 1467, 1352, 1285, 1245, 1183, 1101, 1041, 930, 848, 780, 707, 665, 642, 523, 474; **HRMS** (ESI$^+$) $C_{10}H_{17}N_3NaO_4$ $[M+Na]^+$ found 266.1113, calculated 266.1111.

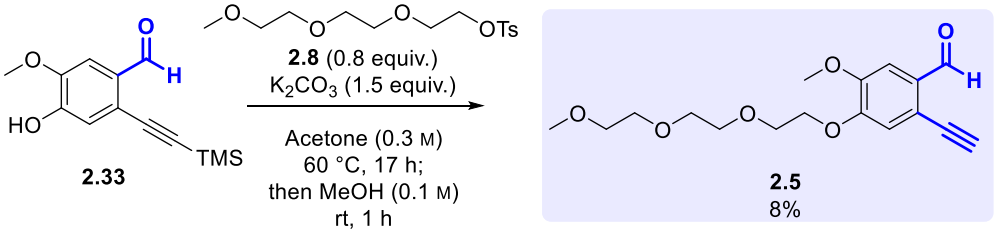
v) 2-EBA ***2.5***

4-hydroxy-5-methoxy-2-((trimethylsilyl)ethynyl)benzaldehyde (2.33)



A solution of triethylamine (2.8 mL, 20.1 mmol, 4.0 equiv.) and 2-bromo-4-hydroxy-5-methoxybenzaldehyde (1.16 g, 5.0 mmol, 1.0 equiv.) in anhydrous THF (50 mL, 0.1 M) was added to a mixture of bis(triphenylphosphine)palladium (II) dichloride (175 mg, 0.2 mmol, 0.05 equiv.) and copper (I) iodide (28 mg, 0.1 mmol, 0.025 equiv.) under nitrogen. Ethynyltrimethylsilane (0.83 mL, 6.0 mmol, 1.2 equiv.) was then added and the reaction mixture was stirred at 80 °C for 18 h. After cooling to room temperature, the mixture was filtered through Celite and the filtrate concentrated under reduced pressure. The residue obtained was purified by flash column chromatography (30% EtOAc:Petrol, R_f 0.46), and pure fractions were concentrated under reduced pressure to afford the title compound as a cream-coloured solid (0.88 g, 3.5 mmol, 70%). **^{1}H NMR** (400 MHz, $CDCl_3$) $\delta_{H:}$ 0.26 (9H, s, TMS), 3.95 (3H, s, OMe), 6.27 (1H, s, ArOH), 7.07 (1H, s, ArH3), 7.38 (1H, s, ArH6), 10.36 (1H, s, ArCHO); **^{13}C NMR** (101 MHz, $CDCl_3$) $\delta_{C:}$ -0.07 (TMS), 56.33 (OMe), 99.91 ($Me_3SiCCAr$), 100.90 ($Me_3SiCCAr$), 108.04 (ArC6), 118.68 (ArC3), 122.23 (ArC2), 130.37 (ArC1), 147.50 (ArC5), 150.91 (ArC4), 190.88 (ArCHO); **λ_{max}** (ATR, cm^{-1}) 3144, 2959, 2153, 1662, 1603, 1560, 1506, 1470, 1434, 1396, 1354, 1293, 1252, 1212, 1171, 1121, 1014, 901, 877, 842, 757, 734, 718, 591, 487; **mp** 129-132 °C; **HRMS** (ESI^+) $C_{13}H_{16}NaO_3Si$ $[M+Na]^+$ found 271.0769, calculated 271.0761.

2-Ethynyl-5-methoxy-4-(2-(2-(2-methoxyethoxy)ethoxy)ethoxy)benzaldehyde (2.5)

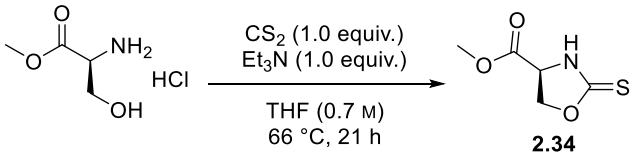


A solution of compound **2.8** (0.44 g, 1.4 mmol, 0.8 equiv.) in acetone (1 mL) was added to a mixture of K_2CO_3 (0.34 g, 2.5 mmol, 1.5 equiv.) and compound **2.33** (0.40 g, 1.6 mmol, 1.0 equiv.) in acetone (4 mL, 0.3 M), and the reaction mixture was stirred at 60 °C for 17 h. After cooling to room temperature, the reaction mixture was concentrated under reduced pressure and the residue obtained was dissolved in methanol (9.7 mL, 0.1 M) and stirred at room temperature for 1 h. The reaction mixture was then concentrated under reduced pressure, and saturated aqueous $NaHCO_3$ solution (20 mL) was added. The aqueous was extracted with CH_2Cl_2 (3 × 60 mL), and the combined organics were washed with saturated aqueous $NaHCO_3$ solution (50 mL), dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The resulting residue was purified by flash column chromatography (66% EtOAc:Petrol, R_f 0.21), and pure fractions were concentrated under reduced pressure to afford the title compound as an orange oil (37 mg, 0.11 mmol, 8%). **^{1}H NMR** (300 MHz, CD_3OD) $\delta_{H:}$ 3.34 (3H, s, $-CH_2OMe$), 3.50-3.54 (2H, m, $-OCH_2$), 3.61-3.63 (2H, m, $-OCH_2$), 3.63-3.67 (2H, m, $-OCH_2$), 3.68-3.74 (2H, m, $-OCH_2$), 3.84-3.91 (2H, m, $-OCH_2$), 3.89 (3H, s, ArOMe), 3.91 (1H, s, ArCCH), 4.21-4.28 (2H, m, $ArOCH_2$), 7.15 (1H, s, ArH3), 7.37 (1H, s, ArH6), 10.32 (1H, s, CHO); **^{13}C NMR** (75 MHz, CD_3OD) $\delta_{C:}$ 56.46 (ArOMe), 59.07 ($-CH_2OMe$), 70.05 ($-OCH_2$), 70.48 ($-OCH_2$), 71.38 ($-OCH_2$), 71.58 ($-OCH_2$), 71.85 ($-OCH_2$), 72.96 ($-OCH_2$), 79.91 (ArCCH), 84.80 (ArCCH), 109.45 (ArC6), 117.66 (ArC3), 121.61 (ArC2),

131.99 (Ar<u>C</u>1), 151.80 (Ar<u>C</u>5), 154.75 (Ar<u>C</u>4), 191.34 (Ar<u>C</u>HO); **λ$_{max}$** (ATR, cm^{-1}) 3251, 2874, 1682, 1590, 1566, 1507, 1457, 1397, 1347, 1276, 1219, 1096, 1054, 1013, 948, 874, 856, 759, 684, 645, 588, 555, 475; **HRMS** (ESI$^+$) $C_{17}H_{23}O_6$ $[M+H]^+$ found 323.1493, calculated 323.1489.

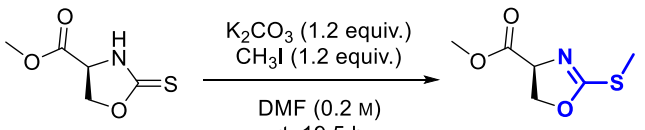
vi) *Ox* ***2.6***

Methyl (*S*)-2-thioxooxazolidine-4-carboxylate (2.34)



Triethylamine (0.9 mL, 6.5 mmol, 1.0 equiv.) was added to a solution of L-serine methyl ester hydrochloride (1.02 g, 6.5 mmol, 1.0 equiv.) in THF (9.2 mL, 0.7 M) at 0 °C. Carbon disulfide (0.39 mL, 6.5 mmol, 1.0 equiv.) was then added and the reaction mixture was warmed to 66 °C and stirred for 21 h [WARNING: Stench]. After cooling to room temperature, the reaction mixture was diluted with CH_2Cl_2 (50 mL) and the organics washed with saturated aqueous $NaHCO_3$ solution (2 × 20 mL), dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The resulting residue was purified by flash column chromatography (50% EtOAc:Petrol, R_f 0.50 in 66% EtOAc:Petrol), and pure fractions were concentrated under reduced pressure to afford a yellow oil (0.61 g), which was used immediately in the next step without characterisation.

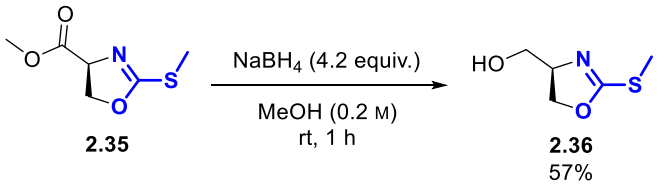
Methyl (*S*)-2-(methylthio)-4,5-dihydrooxazole-4-carboxylate (2.35)



53% *(over 2 steps)*

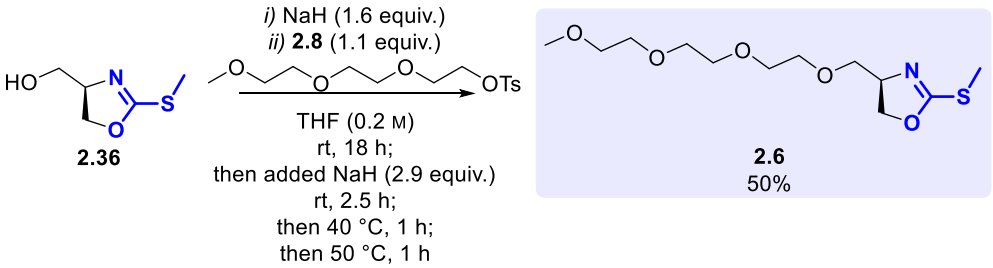
Methyl iodide (283 µL, 4.5 mmol, 1.2 equiv.) was added dropwise to a mixture of compound **2.34** (0.61 g, 3.8 mmol, 1.0 equiv.) and K_2CO_3 (0.63 g, 4.6 mmol, 1.2 equiv.) in DMF (20 mL, 0.2 M), and the reaction mixture was stirred at room temperature for 19.5 h. Brine (30 mL) was then added, and the aqueous was extracted with EtOAc (3 × 50 mL). The combined organics were washed with brine (3 × 50 mL), dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The residue was purified by flash column chromatography (33% EtOAc:Petrol, R_f 0.48 in 66% EtOAc:Petrol), and pure fractions were concentrated under reduced pressure to afford the title compound (0.35 g, 2.0 mmol, 53%) as a yellow oil, with spectroscopic data in accordance with the literature.[6] **^{1}H NMR** (300 MHz, $CDCl_3$) δ_H 2.49 (3H, s, SMe), 3.79 (3H, s, OMe), 4.47-4.66 (2H, m, -NCHC<u>H</u>$_2$), 4.71-4.83 (1H, m, -NC<u>H</u>).

(*R*)-(2-(methylthio)-4,5-dihydrooxazol-4-yl)methanol (2.36)



Sodium borohydride (0.18 g, 4.7 mmol, 4.2 equiv.) was added in small portions over 20 min to a solution of compound **2.35** (0.20 g, 1.1 mmol, 1.0 equiv.) in anhydrous methanol (5.5 mL, 0.2 M) at 0 °C under nitrogen. The reaction mixture was warmed to room temperature and stirred for 1 h, then quenched through dropwise addition of water (15 mL). The aqueous was extracted with ethyl acetate (3 × 45 mL) and the combined organics were dried over $MgSO_4$, filtered, and concentrated under reduced pressure to afford the target compound (94 mg, 0.6 mmol, 57%) as a pale yellow oil with spectroscopic data in accordance with the literature.[6] **^{1}H NMR** (300 MHz, $CDCl_3$) δ_H 2.47 (3H, s, SMe), 3.50-3.65 (1H, m, -C$\underline{H}_2$OH), 3.75-3.94 (1H, m, -C$\underline{H}_2$OH), 4.19-4.37 (2H, m, -NCHC$\underline{H}_2$OC-), 4.37-4.52 (1H, m, -NC$\underline{H}$).

(*R*)-4-(2,5,8,11-tetraoxadodecyl)-2-(methylthio)-4,5-dihydrooxazole (2.6)

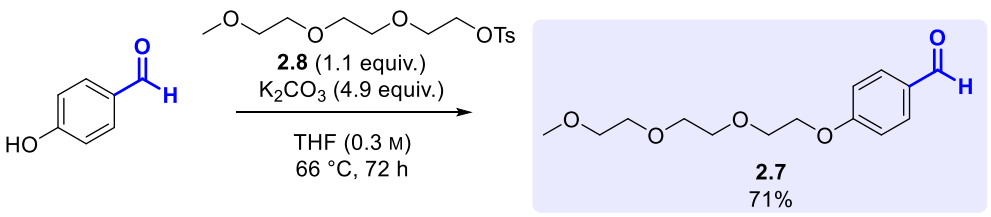


Sodium hydride (60% dispersion in mineral oil, 38 mg, 1.0 mmol, 1.6 equiv.) was added in small portions over 10 min to a solution of compound **2.36** (86 mg, 585 µmol, 1.0 equiv.) in anhydrous THF (3 mL, 0.2 M) under nitrogen. After 5 min, a solution of compound **2.8** (205 mg, 644 µmol, 1.1 equiv.) in THF (0.5 mL) was added and the reaction mixture was stirred at room temperature for 18 h. At this point TLC analysis demonstrated the reaction was incomplete, and so a further portion of sodium hydride (69 mg, 1.7 mmol, 2.9 equiv.) was added and the reaction mixture was stirred at room temperature for a further 2.5 h, then at 40 °C for 1 h, and 50 °C for 1 h. The reaction mixture was allowed to cool to room temperature, then quenched through dropwise addition of water (5 mL) and extracted with ethyl acetate (6 × 20 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The orange oil obtained was purified by flash column chromatography (33-100% EtOAc:Petrol, R_f 0.09 in 33% EtOAc:Petrol), and pure fractions were concentrated under reduced pressure to afford the title compound (85 mg, 0.29 mmol, 50%) as a yellow oil. **^{1}H NMR** (300 MHz, $CDCl_3$) δ_H 2.43 (3H, s, SMe), 3.35 (3H, s, OMe), 3.34-3.44 (1H, m, -C$\underline{H}_2$CHC$\underline{H}_2$OCSMe-), 3.49-3.55 (2H, m, -C$\underline{H}_2$-), 3.57-3.65 (10H, m, -C$\underline{H}_2$-), 3.65-3.72 (1H, m, -C$\underline{H}_2$CHN-), 4.19-4.35 (2H, m, -C$\underline{H}_2$OCSMe), 4.35-4.45 (1H, m, -NC$\underline{H}$); **^{13}C NMR** (75 MHz, $CDCl_3$) δ_C 14.51 (SMe), 59.09

(OMe), 66.23 (-N<u>C</u>H), 70.54 (-<u>C</u>H$_2$-), 70.60 (-<u>C</u>H$_2$-), 70.70 (2 × -<u>C</u>H$_2$-), 70.97 (-<u>C</u>H$_2$-), 72.01 (-<u>C</u>H$_2$OMe), 72.67 (-<u>C</u>H$_2$OCSMe), 73.22 (-<u>C</u>H$_2$CHN-), 167.62 (-<u>C</u>SMe); **λ$_{max}$** (ATR, cm^{-1}) 3433, 2872, 1759, 1670, 1604, 1455, 1349, 1329, 1287, 1250, 1199, 1100, 954, 927, 881, 851, 720, 648, 529; **HRMS** (ESI$^+$) $C_{12}H_{24}NO_5S$ $[M+H]^+$ found 294.1367, calculated 294.1370.

vii) *BA* ***2.7***

4-(2-(2-(2-Methoxyethoxy)ethoxy)ethoxy)benzaldehyde (2.7)



A solution of compound **2.8** (280 mg, 880 µmol, 1.1 equiv.) in THF (0.5 mL) was added to a mixture of K_2CO_3 (0.29 g, 2.1 mmol, 2.5 equiv.) and 4-hydroxybenzaldehyde (100 mg, 820 µmol, 1.0 equiv.) in THF (2.4 mL, 0.3 M), and the reaction stirred at 66 °C for 24 h. At this point TLC analysis demonstrated the reaction was incomplete, and so a further portion of K_2CO_3 (0.28 g, 2.0 mmol, 2.4 equiv.) was added and the reaction mixture was stirred at 66 °C for 48 h. The reaction mixture was allowed to cool to room temperature, then water (20 mL) was added and the aqueous was extracted with CH_2Cl_2 (3 × 20 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The brown oil obtained was purified by flash column chromatography (33-50% EtOAc:Petrol, R_f 0.18 in 33% EtOAc:Petrol), and pure fractions were concentrated under reduced pressure to afford the title compound (0.16 g, 0.6 mmol, 71%) as a pale-yellow oil with spectroscopic data in accordance with the literature.[27] **^{1}H NMR** (300 MHz, $CDCl_3$) δ_H 3.37 (3H, s, OMe), 3.50-3.57 (2H, m, -C<u>H</u>$_2$-), 3.60-3.70 (4H, m, -C<u>H</u>$_2$-), 3.71-3.78 (2H, m, -C<u>H</u>$_2$-), 3.83-3.93 (2H, t, *J* = 4.9 Hz, -ArOCH$_2$C<u>H</u>$_2$-), 4.21 (2H, t, *J* = 4.9 Hz, ArOC<u>H</u>$_2$), 7.02 (2H, d, *J* = 8.7 Hz, Ar<u>H</u>3), 7.81 (2H, d, *J* = 8.7 Hz, Ar<u>H</u>2), 9.87 (1H, s, C<u>H</u>O).

2.4.2 Protein expression

RNase A from bovine pancreas (powder, 50 units/mg protein) and myoglobin from equine skeletal muscle (powder, 95-100%) were purchased from Sigma Aldrich, and Clostripain histolyticum (Endoproteinase-Arg-C) was purchased from BioServ UK Limited. Expression and purification of CjX183-D WT and R51K was carried out by N. D. J. Yates, as detailed in our publication.[22]

Protein Sequences:

<u>CjX183-D</u>

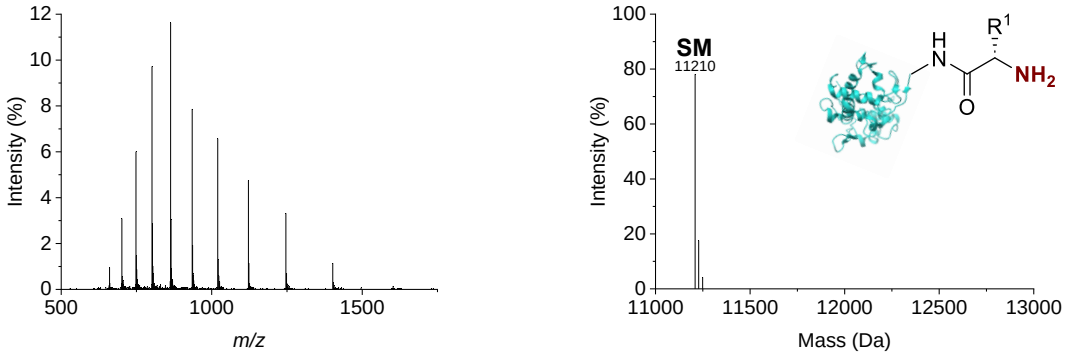
GYLVGDATRG ANLWNTQTCV ACHGVDGERN ASGTPALTPL NPNRDLYRHS RDTQDRALRD FISMWMPQGN EGSCTGQCAA DIEAFIRTWH HHHHH

Theoretical M_w (+heme): 11226 Da

CjX183-D R51K

GYLVGDATRG ANLWNTQTCV ACHGVDGERN ASGTPALTPL NPNRDLYRHS KDTQDRALRD FISMWMPQGN EGSCTGQCAA DIEAFIRTWH HHHHH

Theoretical M_w (+heme): 11198 Da



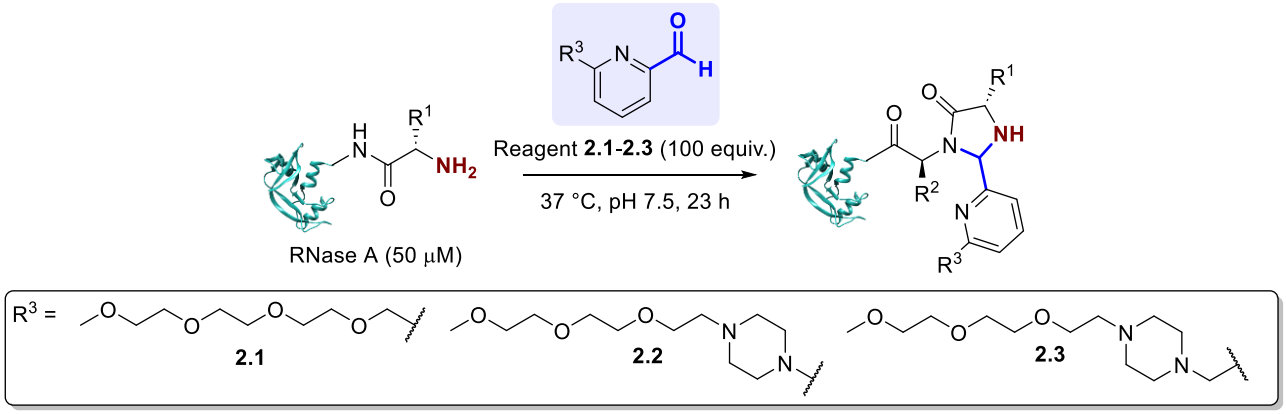
MS (ESI$^+$) [**SM**+H]$^+$ found 11210, calculated 11198.

Figure S2.1. Raw (left) and deconvoluted (right) mass spectra for CjX183-D R51K. Nb. In all experiments performed on CjX183-D, both wild type and mutants, the observed mass was ~ 12 Da higher than the calculated mass. Reasons for this difference from the theoretical mass are currently under investigation.

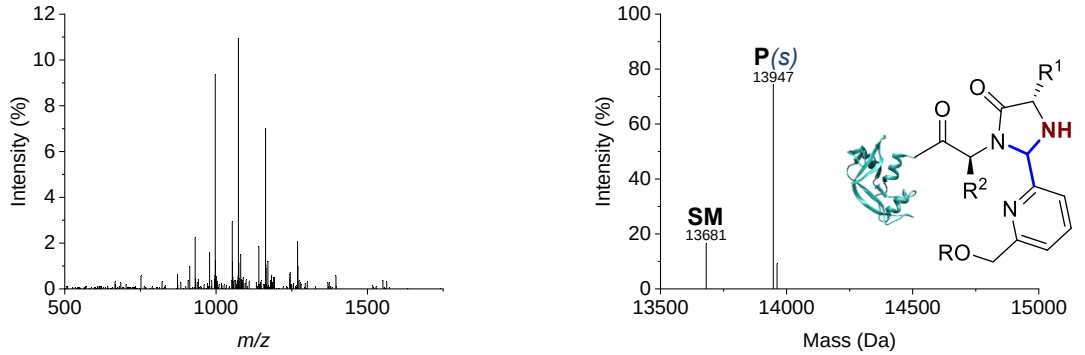
2.4.3 Protein modification

2.4.3.1 Validation of literature conditions and general procedures

2-PCAs 2.1-2.3



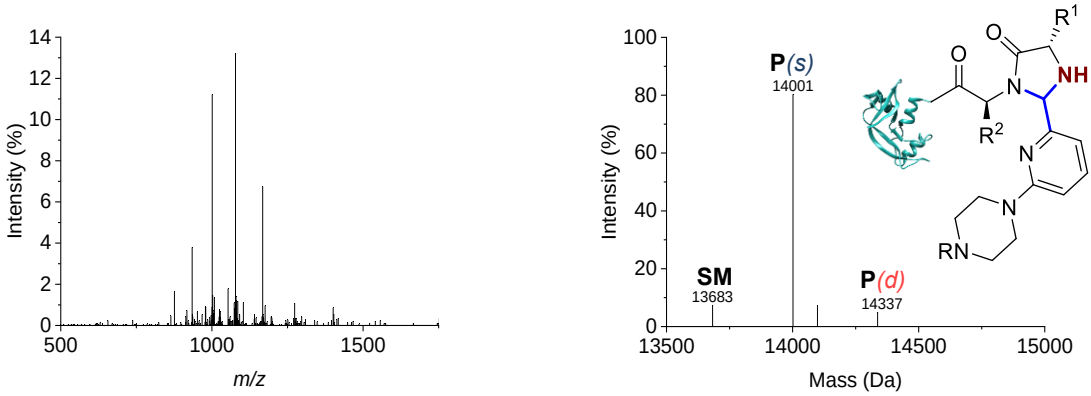
Procedure 2A: *Modification with 2-PCAs **2.1**-**2.3** under conditions reported by MacDonald et al.*[1] A stock solution of 2-PCA **2.1**, **2.2**, or **2.3** (50 µL, 10 mM, 500 nmol, 100 equiv., in 50 mM pH 7.5 sodium phosphate buffer) was added to a solution of RNase A (50 µL, 100 µM, 5 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer), and the mixture incubated at 37 °C for 23 h with agitation (1000 rpm). Conversion was determined by LC-MS analysis without purification.



2-PCA **2.1***

84%

MS (ESI$^+$) [**SM**+H]$^+$ found 13681, calculated 13681;[14] [**P**(s)+H]$^+$ found 13947, calculated 13946; [**P**(s)+H_2O+H]$^+$ found 13962, calculated 13964.

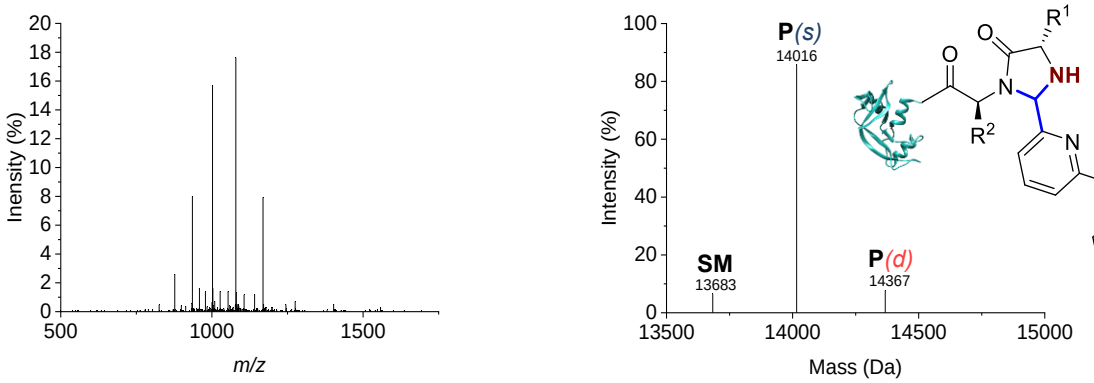


2-PCA **2.2***

93%

88(s):5(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 13683, calculated 13681;[14] [**P**(s)+H]$^+$ found 14001, calculated 14001; [**P**(s)+H_3PO_4+H]$^+$ found 14099, calculated 14098; [**P**(d)+H_2O+H]$^+$ found 14337, calculated 14338.



2-PCA **2.3***

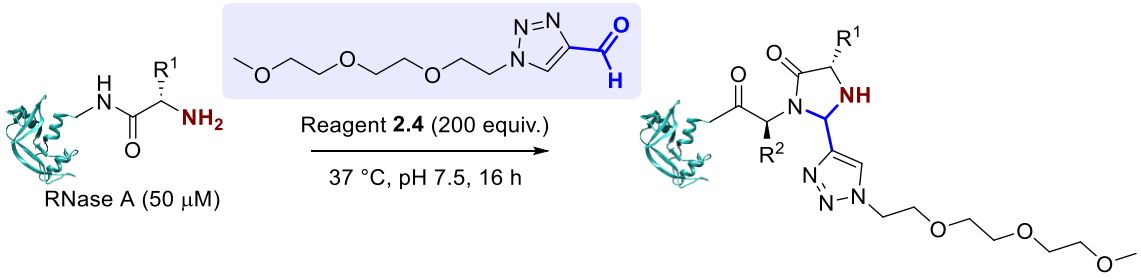
94%

86(s):8(d)

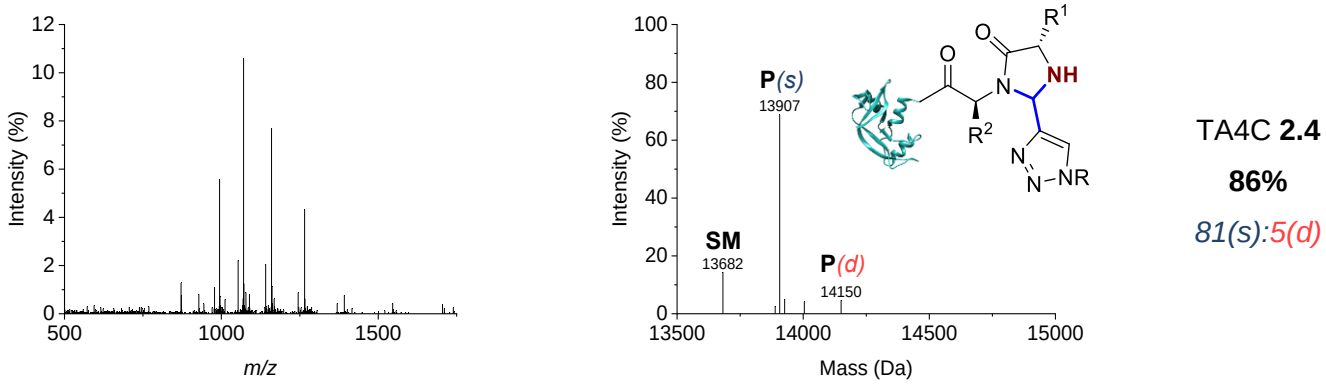
MS (ESI$^+$) [**SM**+H]$^+$ found 13683, calculated 13681;[14] [**P**(s)+H]$^+$ found 14016, calculated 14015; [**P**(d)+H_2O+H]$^+$ found 14367, calculated 14366.

Figure S2.2. Raw (left) and deconvoluted (right) mass spectra for the modification of RNase A with 2-PCAs **2.1**-**2.3** under reported conditions outlined in Procedure **2A**. **Following LC-MS of each protein sample, water was injected into the LC column to remove residual protein; for samples marked (*), conversions were determined from MS traces of residual protein due to co-elution with excess small molecule.*

TA4C 2.4



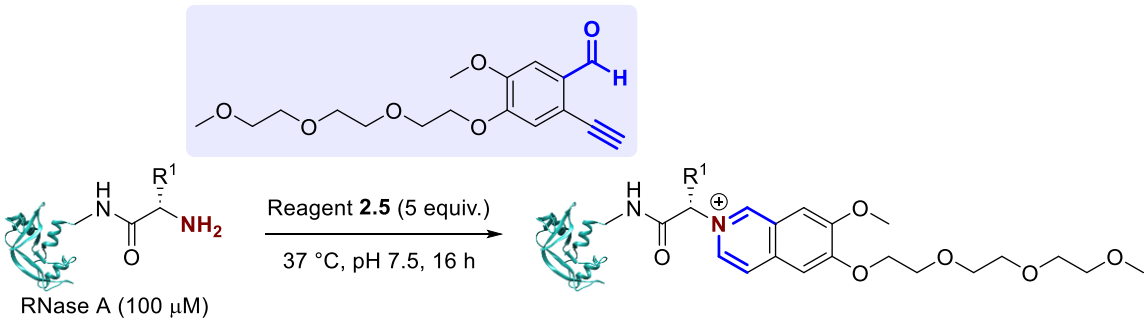
Procedure 2B: *Modification with TA4C **2.4** under conditions reported by Onoda et al.*[3] A stock solution of TA4C **2.4** (2.5 µL, 200 mM, 500 nmol, 166 equiv., in DMSO) was added to a solution of RNase A (47.5 µL, 52 µM, 3 nmol, 1 equiv., in 10 mM pH 7.5 potassium phosphate buffer), and the mixture incubated at 37 °C for 16 h with agitation (1000 rpm). Conversion was determined by LC-MS analysis without purification.



MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681;[14] [**P**(s)–H_2O+H]$^+$ found 13889, calculated 13888; [**P**(s)+H]$^+$ found 13907, calculated 13906; [**P**(s)+H_2O+H]$^+$ found 13926, calculated 13924; [**P**(s)+H_3PO_4+H]$^+$ found 14004, calculated 14004; [**P**(d)+H_2O+H]$^+$ found 14150, calculated 14150.

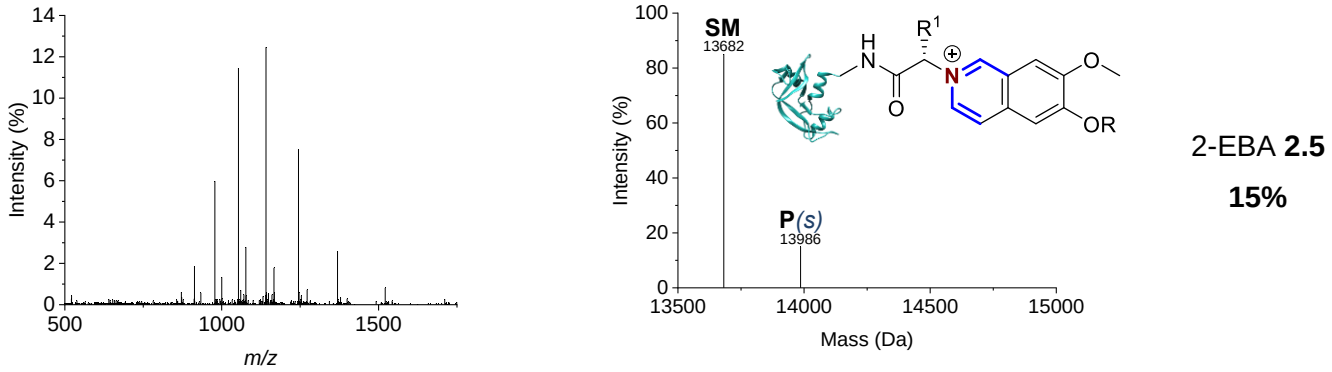
Figure S2.3. Raw (left) and deconvoluted (right) mass spectra for the modification of RNase A with TA4C **2.4** under reported conditions outlined in Procedure **2B**.

2-EBA 2.5



Procedure 2C: *Modification with 2-EBA **2.5** under conditions reported by Deng et al.*[4] A stock solution of 2-EBA **2.5** (15 µL, 5 mM, 75 nmol, 5 equiv., in DMSO) was added to a solution of RNase A (135 µL, 111 µM,

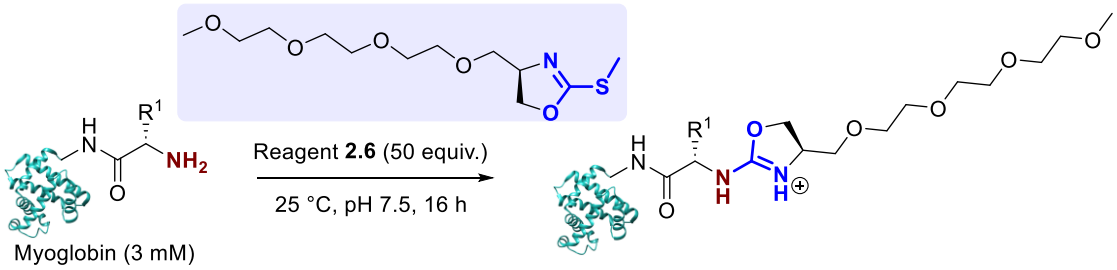
15 nmol, 1 equiv., in 50 mM pH 7.5 phosphate buffered saline), and the mixture incubated at 37 °C for 16 h with agitation (1000 rpm). Conversion was determined by LC-MS analysis without purification.



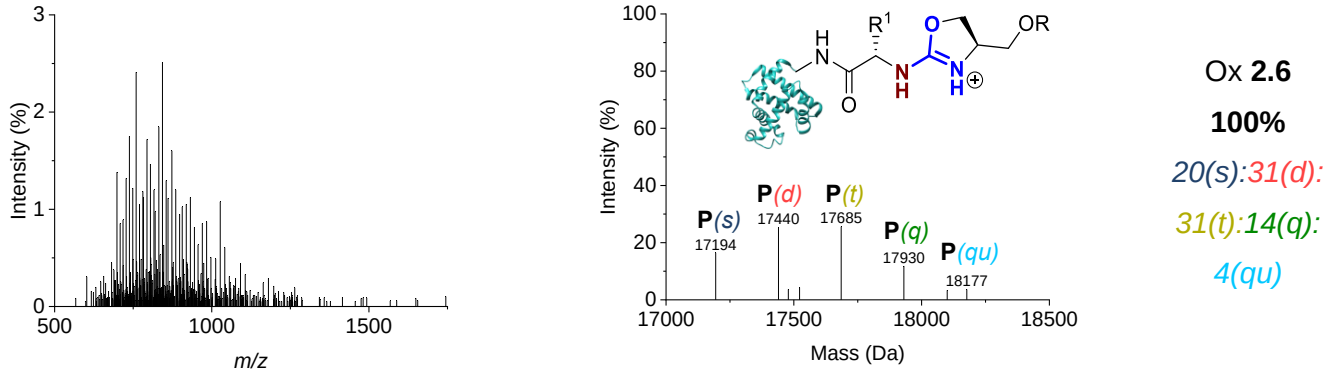
MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681;[14] [**P**(s)+H]$^+$ found 13986, calculated 13985.

Figure S2.4. Raw (left) and deconvoluted (right) mass spectra for the modification of RNase A with 2-EBA **2.5** under reported conditions outlined in Procedure **2C**.

Ox 2.6



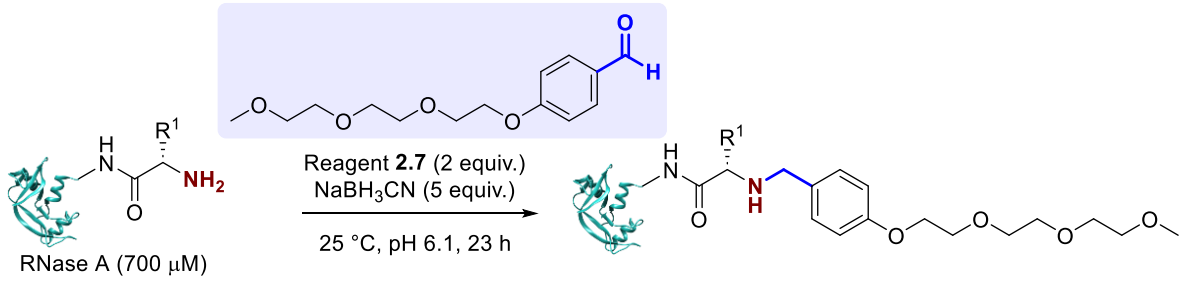
Procedure 2D: *Modification with Ox **2.6** under conditions reported by Tang et al.*[6] A stock solution of Ox **2.6** (15 µL, 300 mM, 4.5 µmol, 50 equiv., in 10 mM pH 7.5 sodium phosphate buffer) was added to a solution of myoglobin (15 µL, 6 mM, 90 nmol, 1 equiv., in 10 mM pH 7.5 sodium phosphate buffer), and the mixture incubated at 25 °C for 16 h with agitation (1000 rpm). Samples were diluted ×60 with water prior to analysis of conversion by LC-MS analysis without purification.



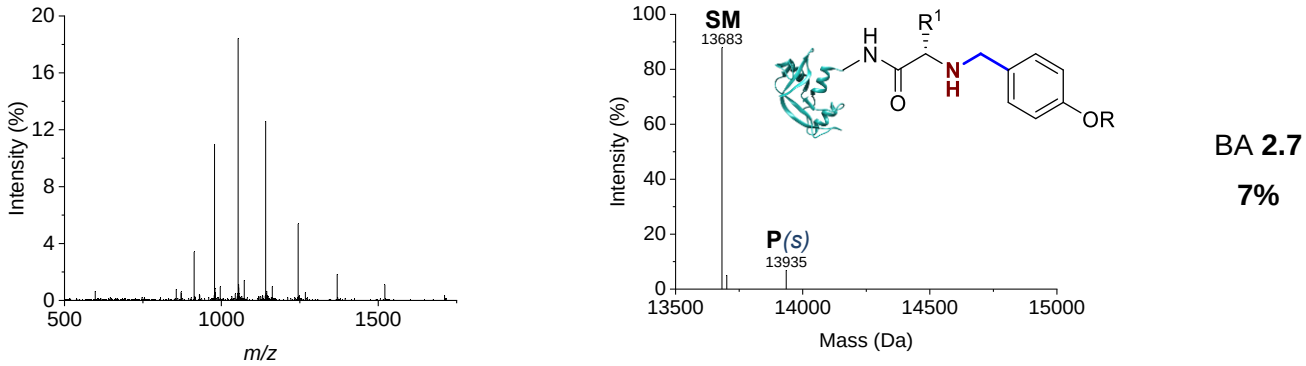
MS (ESI$^+$) [**P**(s)+H]$^+$ found 17194, calculated 17197; [**P**(d)+H]$^+$ found 17440, calculated 17442; [**P**(t)+H]$^+$ found 17685, calculated 17687; [**P**(q)+H]$^+$ found 17930, calculated 17932; [**P**(qu)+H]$^+$ found 18177, calculated 18177.

Figure S2.5. Raw (left) and deconvoluted (right) mass spectra for the modification of myoglobin with Ox **2.6** under reported conditions outlined in Procedure **2D**.

BA 2.7



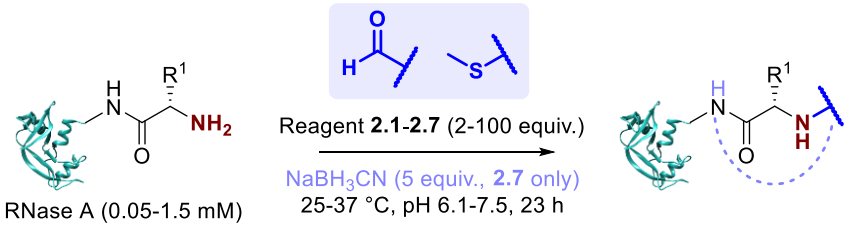
Procedure 2E: *Modification with BA **2.7** under conditions reported by Chen et al.*[5] Stock solutions of BA **2.7** (22.5 µL, 3 mM, 70 µmol, 2 equiv., in 25 mM pH 6.1 citrate buffer) and sodium cyanoborohydride (2.5 µL, 70 mM, 175 µmol, 5 equiv., in water) were sequentially added to a solution of RNase A (25 µL, 1.4 mM, 35 nmol, 1 equiv., in in 25 mM pH 6.1 citrate buffer), and the mixture incubated at 25 °C for 23 h with agitation (1000 rpm). Samples were diluted ×10 with 25 mM pH 6.1 citrate buffer prior to analysis of conversion by LC-MS analysis without purification.



MS (ESI$^+$) [**SM**+H]$^+$ found 13683, calculated 13681;[14] [**SM**+H_2O+H]$^+$ found 13701, calculated 13699; [**P**(s)+H]$^+$ found 13935, calculated 13933.

Figure S2.6. Raw (left) and deconvoluted (right) mass spectra for the modification of RNase A with BA **2.7** under reported conditions outlined in Procedure **2E**.

2.4.3.2 Screening of reagents under differing conditions



Each of the reagents **2.1**-**2.7** were used to modify RNase A under the conditions outlined in the **General Procedures 2A**, **2C**, **2D** and **2E** for 23 h.

For the modifications with Procedure 2A: reagents **2.1**-**2.7** were applied under the conditions outlined in **General Procedure 2A**. For compound **2.7**, 5 equiv. of sodium cyanoborohydride in 5 µL of water was also added, and compound **2.7** was added as a more concentrated solution (45 µL, 11 mM) to keep the total volume constant.

For the modifications with Procedure 2C: reagents **2.1**-**2.7** were applied under the conditions outlined in **General Procedure 2C**, on a 50 µL scale. For compound **2.7**, 5 equiv. of sodium cyanoborohydride in 2.5 µL of water was also added, and compound **2.7** was added as a more concentrated solution (22.5 µL, 1.1 mM) to keep the total volume constant.

For the modifications with Procedure 2D: reagents **2.1**-**2.7** were applied under the conditions outlined in **General Procedure 2D**, on a 20 µL scale, with 1.5 mM final protein concentration and 50 equiv. compound **2.1**-**2.7**. Samples were diluted ×10 with 10 mM pH 7.5 sodium phosphate buffer prior to analysis of conversion by LC-MS analysis without purification. For compound **2.7**, 5 equiv. of sodium cyanoborohydride in 1 µL of water was also added, and compound **2.7** was added as a more concentrated solution (9 µL, 167 mM) to keep the total volume constant.

For the modifications with Procedure 2E: reagents **2.1**-**2.7** were applied under the conditions outlined in **General Procedure 2E**. For compounds **2.1**-**2.6**, 2.5 µL 25 mM pH 6.1 citrate buffer was added instead of the sodium cyanoborohydride solution.

Compound	Conversion (%)			
	General Procedure **2A**	General Procedure **2C**	General Procedure **2D**	General Procedure **2E**
2-PCA **2.1**	**84**	**10**	**29**	**0**
2-PCA **2.2**	**92**	**84**	**93** *67(s):22(d):5(t)*	**23**
2-PCA **2.3**	**94** *86(s):8(d)*	**42**	**100** *77(s):23(d)*	**6**
TA4C **2.4**	**77**	**15**	**48**	**0**
2-EBA **2.5**	**67** *22(s):29(d):11(t):5(q)*	**6**	**20**	**0**
Ox **2.6**	**22**	**6**	**87** *54(s):33(d)*	**0**

BA **2.7**	**10**	**0**	**100** *15(s):33(d):32(t):19(q)*	**7**

Table S2.1. Screening of reagents under differing conditions. Modification of RNase A with compounds **2.1**-**2.7** under respective reported conditions outlined in the **General Procedures 2A**, **2C**, **2D** and **2E**.

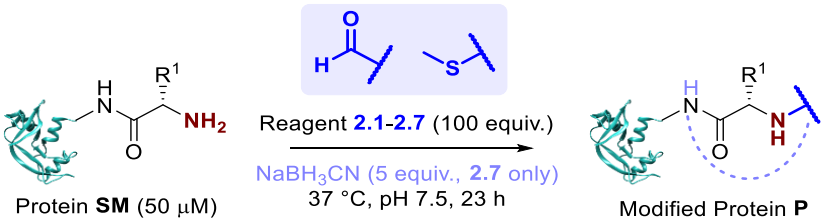
2.4.3.3 Optimisation of 2-EBA concentration

To minimise off-target reactivity with 2-EBA **2.5**, RNase A and myoglobin were modified according to **General Procedure 2A**, varying the equiv. of 2-EBA while maintaining all other conditions (5,10, 25,100, 200 equiv.).

Number of 2-EBA equiv.	**Conversion (%)**	
	RNase A	Myoglobin
5	**7**	**12**
10	**11**	**18**
25	**23** *19(s):4(d)*	**43** *28(s):13(d):2(t)*
100	**47** *25(s):13(d):9(t)*	**58** *36(s):13(d):9(t)*
200	*No protein detected*	

Table S2.2. Optimisation of 2-EBA concentration. Modification of RNase A and myoglobin with 2-EBA **2.5** (5, 10, 25, 100 equiv.).

2.4.3.4 Protein panel modification



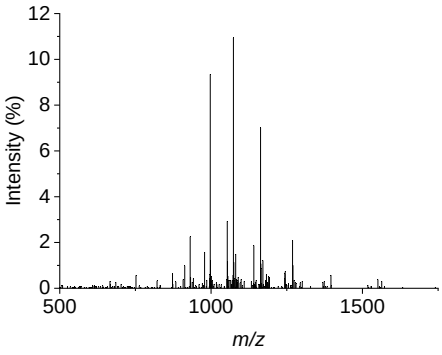
For the modification of RNase A and myoglobin: reagents **2.1**-**2.7** were applied under the conditions outlined in **General Procedure 2A**. For compound **2.7**, 5 equiv. of sodium cyanoborohydride in 5 µL of water was also added, and compound **2.7** was added as a more concentrated solution (45 µL, 11 mM) to keep the total volume constant.

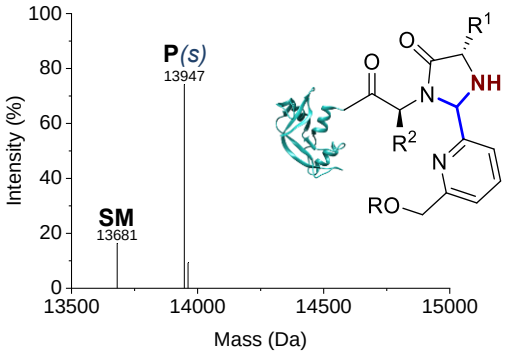
For the modification of Clostripain LC: As described above, on a 19.5 µL scale.

For the modification of CjX183-D WT: As described above, on a 22.9 µL scale.

For the modification of CjX183-D R51K: As described above, on a 30.6 µL scale.

RNase A

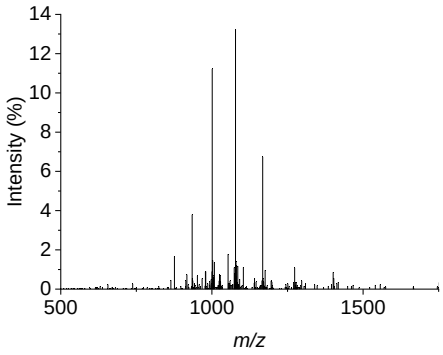


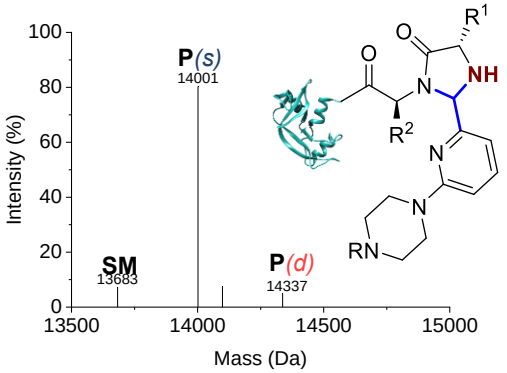


2-PCA **2.1***

84%

MS (ESI$^+$) [**SM**+H]$^+$ found 13681, calculated 13681;[14] [**P**(s)+H]$^+$ found 13947, calculated 13946; [**P**(s)+H_2O+H]$^+$ found 13962, calculated 13964.



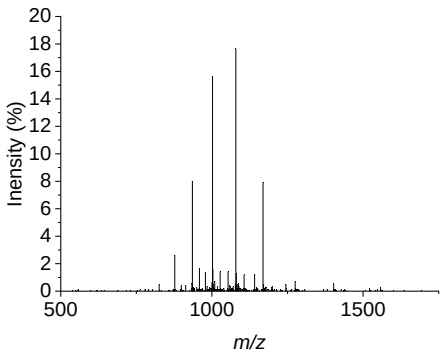


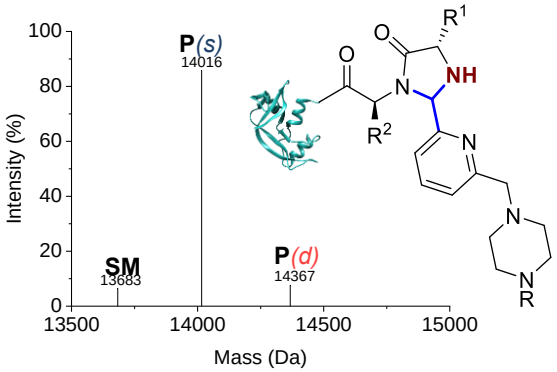
2-PCA **2.2***

93%

88(s):5(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 13683, calculated 13681;[14] [**P**(s)+H]$^+$ found 14001, calculated 14001; [**P**(s)+H_3PO_4+H]$^+$ found 14099, calculated 14098; [**P**(d)+H_2O+H]$^+$ found 14337, calculated 14338.



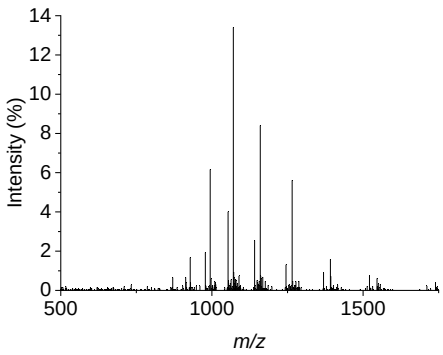


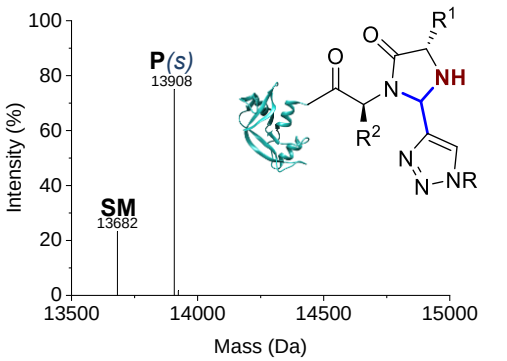
2-PCA **2.3***

94%

86(s):8(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 13683, calculated 13681;[14] [**P**(s)+H]$^+$ found 14016, calculated 14015; [**P**(d)+H_2O+H]$^+$ found 14367, calculated 14366.

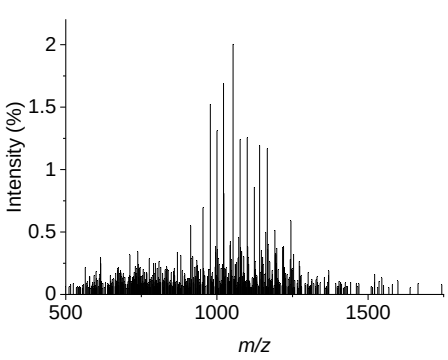


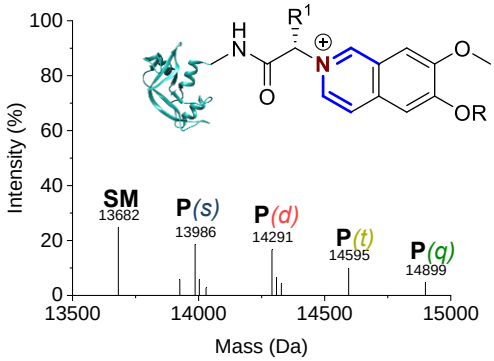


TA4C **2.4**

77%

MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681;[14] [**P**(s)+H]$^+$ found 13908, calculated 13906; [**P**(s)+H_2O+H]$^+$ found 13924, calculated 13924.





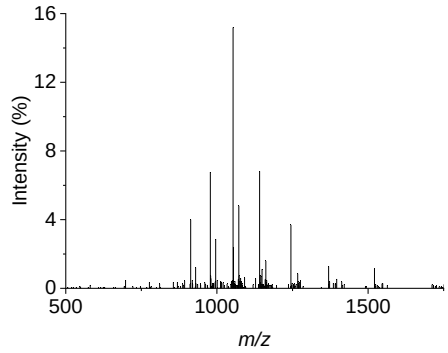
2-EBA **2.5**

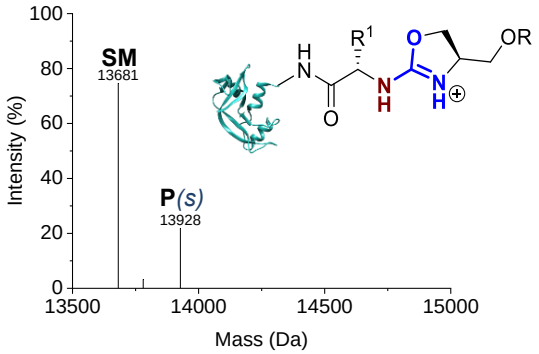
74%

29(s):29(d):

11(t):5(q)

MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681;[14] [**P**(s)+H]$^+$ found 13986, calculated 13985; [**P**(s)+H_2O+H]$^+$ found 14003, calculated 14003; [**P**(s)+MeCN+H]$^+$ found 14030, calculated 14026; [**P**(d)+H]$^+$ found 14291, calculated 14290; [**P**(d)+H_2O+H]$^+$ found 14309, calculated 14308; [**P**(d)+K]$^+$ found 14328, calculated 14328; [**P**(t)+H]$^+$ found 14595, calculated 14594; [**P**(q)+H]$^+$ found 14899, calculated 14898.

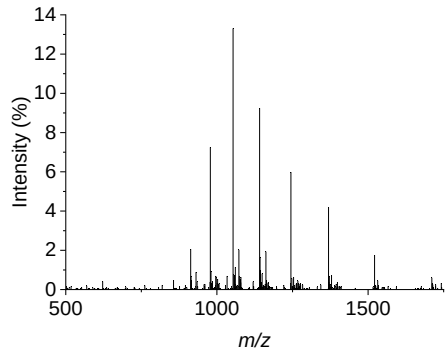


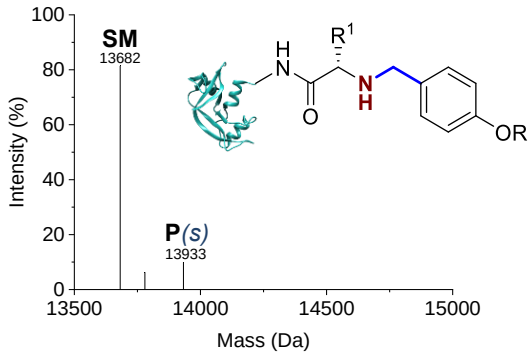


Ox **2.6**

22%

MS (ESI$^+$) [**SM**+H]$^+$ found 13681, calculated 13681;[14] [**SM**+H_3PO_4+H]$^+$ found 13780, calculated 13779; [**P**(s)+H]$^+$ found 13928, calculated 13926.



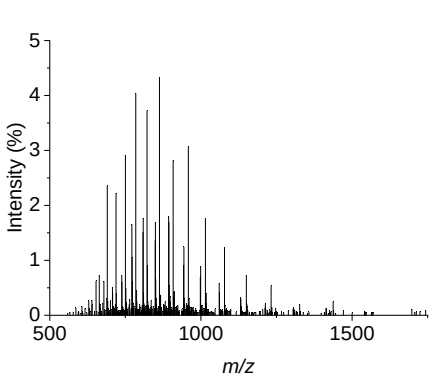


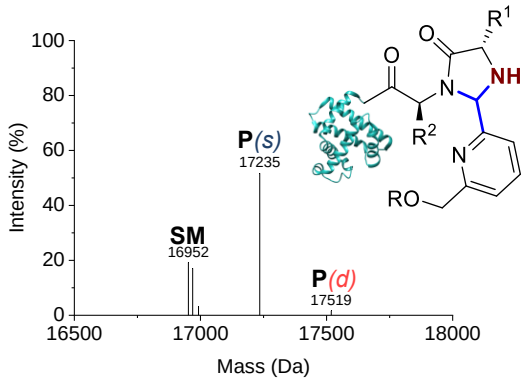
BA **2. 7**

10%

MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681;[14] [**SM**+H_3PO_4+H]$^+$ found 13780, calculated 13779; [**P**(s)+H]$^+$ found 13933, calculated 13933.

Myoglobin



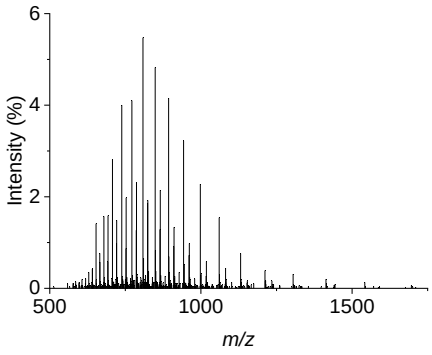


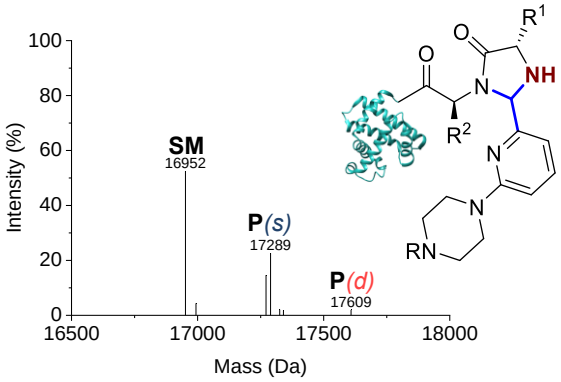
2-PCA **2.1**

57%

55(s):2(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951;[25] [**SM**+H_2O+H]$^+$ found 16969, calculated 16970; [**SM**+MeCN+H]$^+$ found 16992, calculated 16993; [**P**(s)+H_2O+H]$^+$ found 17235, calculated 17235; [**P**(d)+2H_2O+H]$^+$ found 17519, calculated 17518.



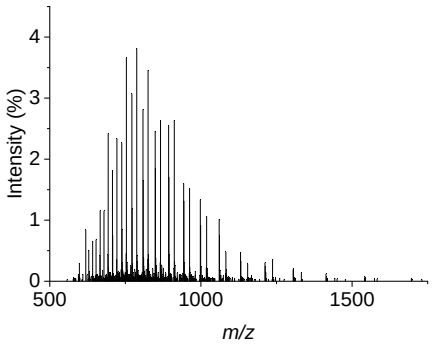


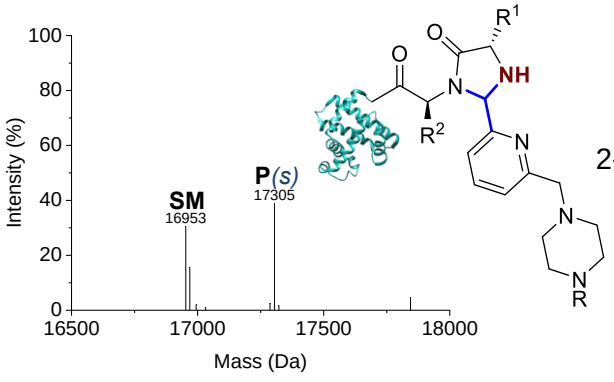
2-PCA **2.2**

41%

39(s):2(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951;[25] [**SM**+MeCN+H]$^+$ found 16994, calculated 16993; [**P**(s)+H]$^+$ found 17271, calculated 17271; [**P**(s)+H_2O+H]$^+$ found 17289, calculated 17289; [**P**(d)+H_2O+H]$^+$ found 17609, calculated 17608.

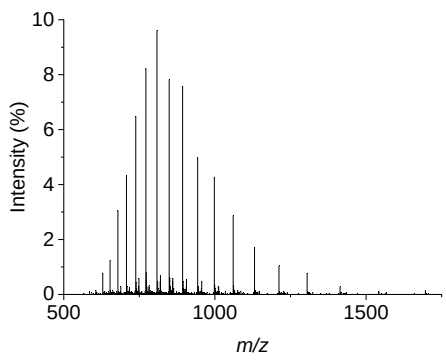


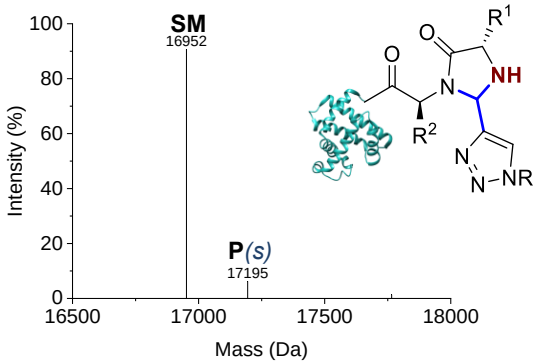


2-PCA **2.3**

47%

MS (ESI$^+$) [**SM**+H]$^+$ found 16953, calculated 16951;[25] [**SM**+H_2O+H]$^+$ found 16969, calculated 16970; [**SM**+MeCN+H]$^+$ found 16994, calculated 16993; [**SM**+DMSO+H]$^+$ found 17032, calculated 17030; [**P**(s)+H]$^+$ found 17287, calculated 17285; [**P**(s)+H_2O+H]$^+$ found 17305, calculated 17303; [**P**(s)+K]$^+$ found 17323, calculated 17324.

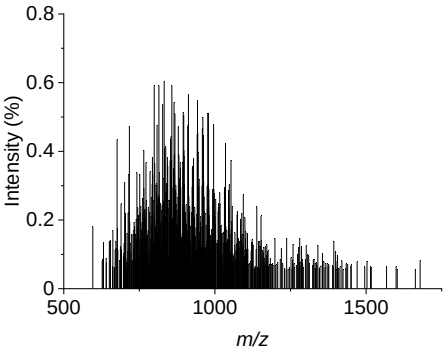


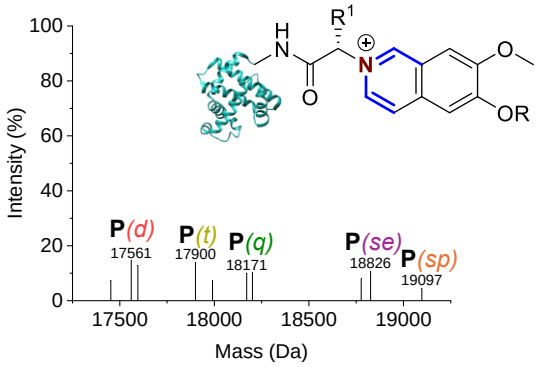


TA4C **2.4**

6%

MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951;[25] [**P**(s)+H_2O+H]$^+$ found 17195, calculated 17195.





2-EBA **2.5**

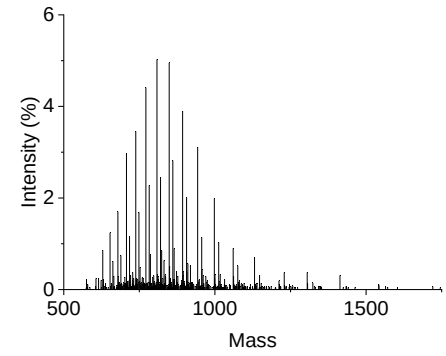
100%

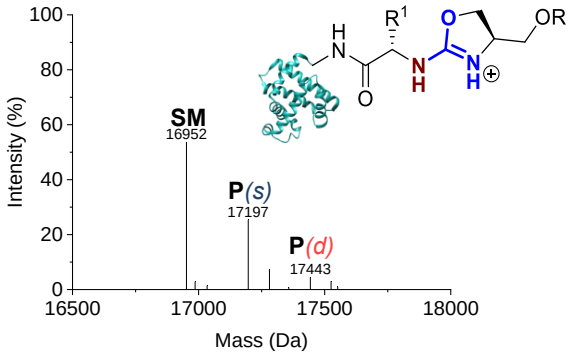
32(d):17(t):

24(q):22(se):

5(sp)

MS (ESI$^+$) [**P**(d)+H]$^+$ found 17561, calculated 17560; [**P**(d)+MeOH+H]$^+$ found 17596, calculated 17592; [**P**(t)+MeOH+H]$^+$ found 17900, calculated 17896; [**P**(q)+H]$^+$ found 18171, calculated 18168; [**P**(q)+MeOH+H]$^+$ found 18202, calculated 18200; [**P**(se)+H]$^+$ found 18776, calculated 18776; [**P**(se)+MeOH+H_2O+H]$^+$ found 18826, calculated 18826; [**P**(sp)+H_2O+H]$^+$ found 19097, calculated 19098.



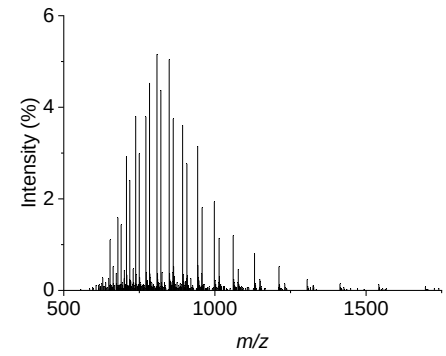


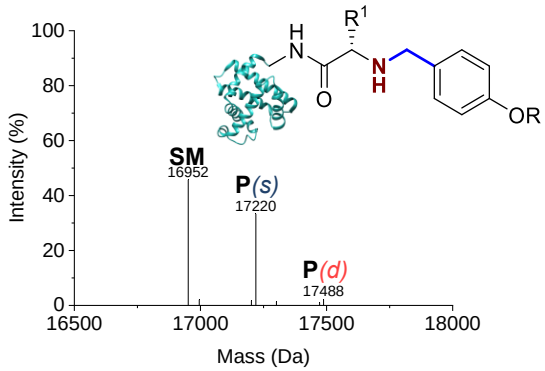
Ox **2.6**

42%

34(s):8(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951;[25] [**SM**+MeOH+H]$^+$ found 16986, calculated 16984; [**SM**+2MeCN+H]$^+$ found 17034, calculated 17034; [**P**(s)+H]$^+$ found 17197, calculated 17197; [**P**(s)+2MeCN+H]$^+$ found 17281, calculated 17279; [**P**(d)+H]$^+$ found 17443, calculated 17442; [**P**(d)+2MeCN+H]$^+$ found 17552, calculated 17524.





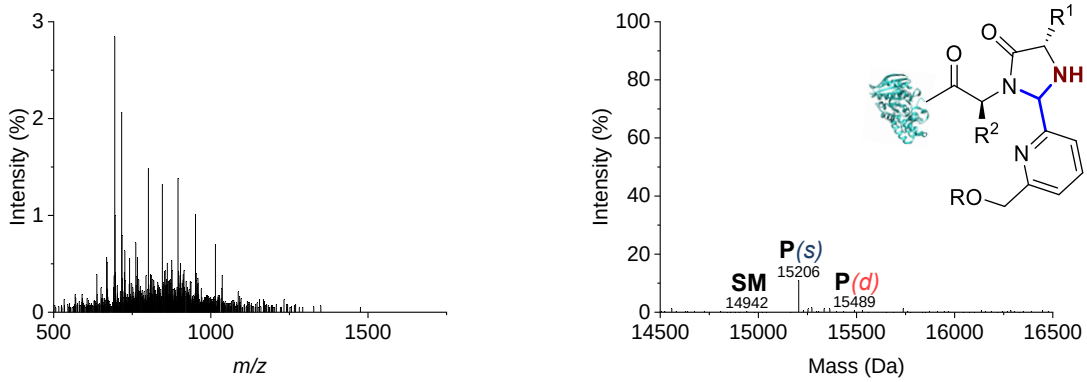
BA **2.7**

46%

42(s):4(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951;[25] [**SM**+MeCN+H]$^+$ found 16995, calculated 16993; [**P**(s)+H]$^+$ found 17202, calculated 17204; [**P**(s)+H_2O+H]$^+$ found 17220, calculated 17222; [**P**(s)+H_3PO_4+H]$^+$ found 17301, calculated 17302; [**P**(d)+H_2O+H]$^+$ found 17473, calculated 17474; [**P**(d)+MeOH+H]$^+$ found 17488, calculated 17488.

Clostripain LC

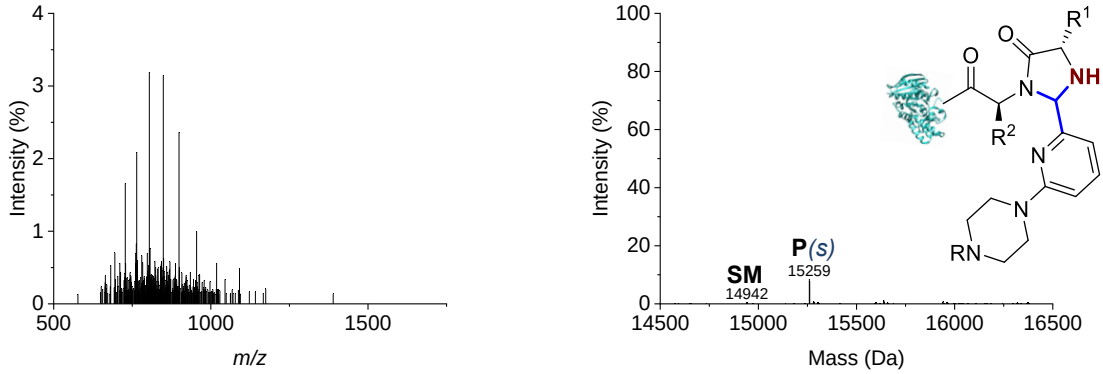


2-PCA **2.1**

97%

89(s):8(d)

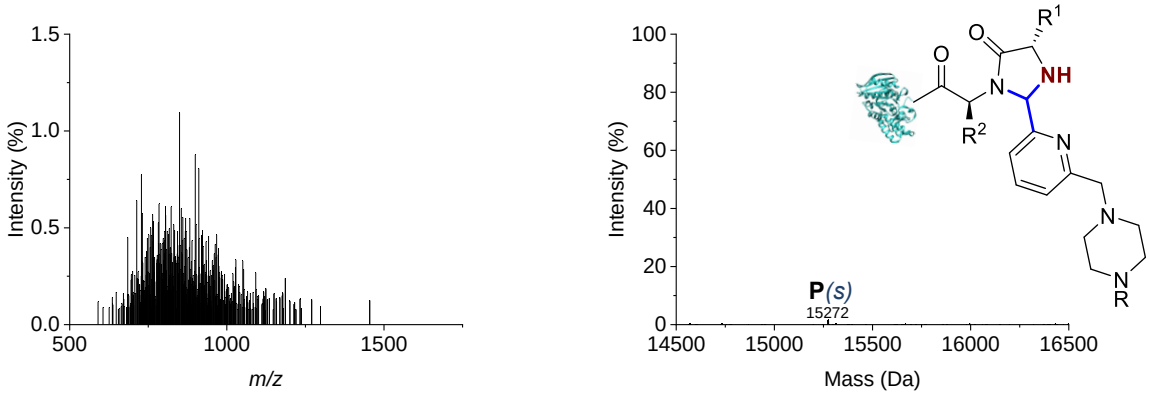
MS (ESI$^+$) [**SM**+H]$^+$ found 14942, calculated 14941;[1] [**SM**+MeCN+H]$^+$ found 14982, calculated 14982; [**P**(s)+H]$^+$ found 15206, calculated 15206; [**P**(s)+Na]$^+$ found 15229, calculated 15228; [**P**(s)+MeCN+H]$^+$ found 15253, calculated 15247; [**P**(d)+H_2O+H]$^+$ found 15489, calculated 15489; [**P**(d)+H_2O+MeCN+H]$^+$ found 15531, calculated 15530.



2-PCA **2.2**

94%

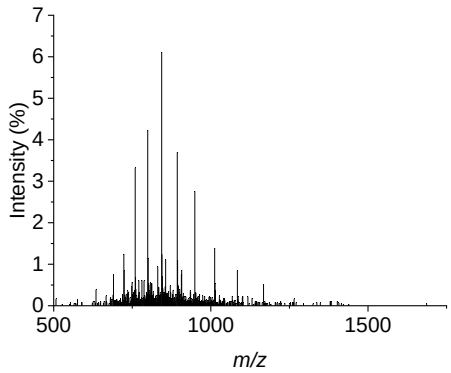
MS (ESI$^+$) [**SM**+H]$^+$ found 14942, calculated 14941;[1] [**P**(s)+H]$^+$ found 15259, calculated 15260; [**P**(s)+Na]$^+$ found 15282, calculated 15282; [**P**(s)+MeCN+H]$^+$ found 15300, calculated 15301.

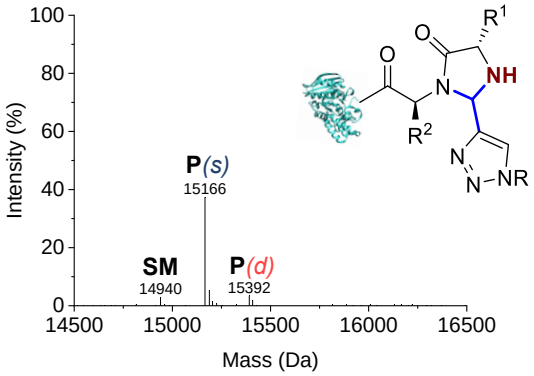


2-PCA **2.3**

100%

MS (ESI$^+$) [**P**(s)+H]$^+$ found 15272, calculated 15274.



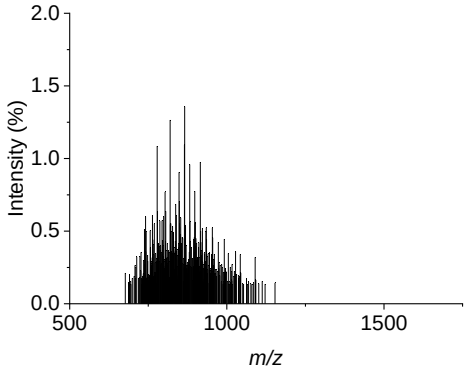


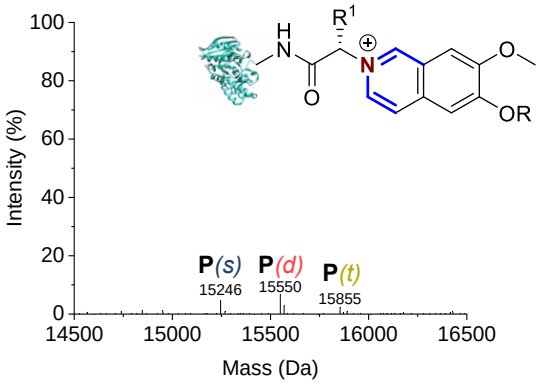
TA4C **2.4**

90%

84(s):6(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 14940, calculated 14941;[1] [**SM**+Na]$^+$ found 14962, calculated 14963; [**SM**+MeCN+H]$^+$ found 14983, calculated 14982; [**P**(s)+H]$^+$ found 15166, calculated 15166; [**P**(s)+Na]$^+$ found 15188, calculated 15189; [**P**(s)+MeCN+H]$^+$ found 15205, calculated 15207; [**P**(d)+H]$^+$ found 15392, calculated 15391; [**P**(d)+H_2O+H]$^+$ found 15409, calculated 15409; [**P**(d)+2H_2O+H]$^+$ found 15428, calculated 15427.





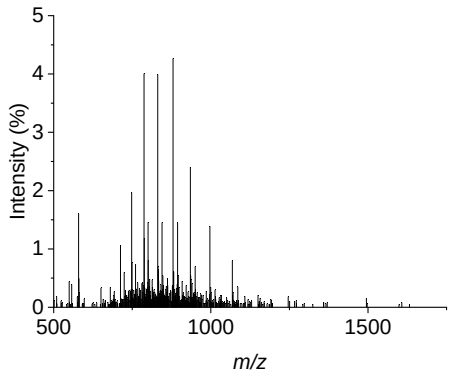
2-EBA **2.5***

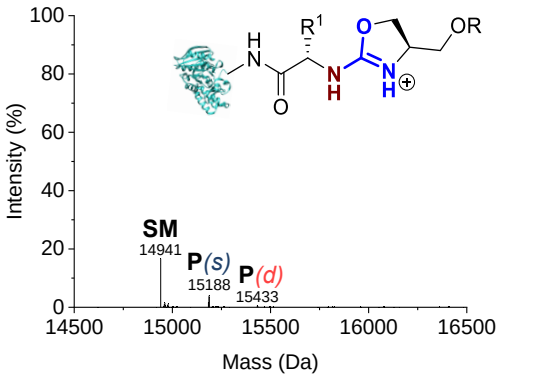
100%

24(s):37(d):

38(t)

MS (ESI$^+$) [**P**(s)+H]$^+$ found 15246, calculated 15245; [**P**(s)+Na]$^+$ found 15268, calculated 15267; [**P**(d)+H]$^+$ found 15550, calculated 15549; [**P**(d)+H_2O+H]$^+$ found 15569, calculated 15567; [**P**(d)+K]$^+$ found 15587, calculated 15587; [**P**(t)+H]+ found 15855, calculated 15853; [**P**(t)+H_2O+H]$^+$ found 15871, calculated 15871; [**P**(t)+2H_2O+H]$^+$ found 15890, calculated 15889.



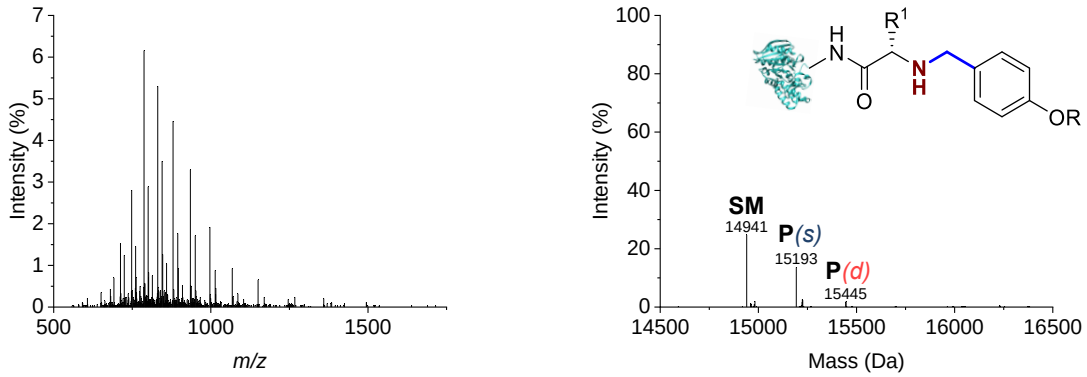


Ox **2. 6**

37%

34(s):3(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 14941, calculated 14941;[1] [**SM**+H_2O+H]$^+$ found 14959, calculated 14959; [**SM**+Na]$^+$ found 14962, calculated 14963; [**P**(s)+H]$^+$ found 15188, calculated 15187; [**P**(s)+H_2O+H]$^+$ found 15204, calculated 15205; [**P**(s)+Na]$^+$ found 15206, calculated 15209; [**P**(d)+H]$^+$ found 15433, calculated 15433.



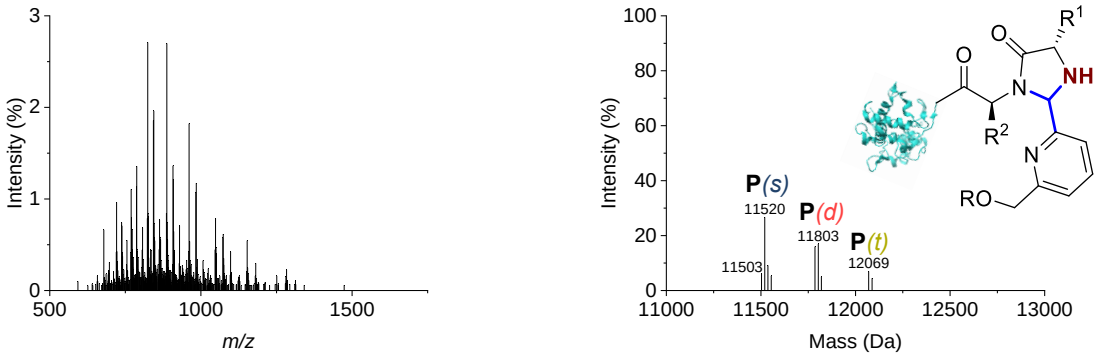
BA **2.7**

37%

29(s):7(d):1(t)

MS (ESI$^+$) [**SM**+H]$^+$ found 14941, calculated 14941;[1] [**SM**+H_2O+H]$^+$ found 14960, calculated 14959; [**SM**+Na]$^+$ found 14962, calculated 14963; [**SM**+K]$^+$ found 14979, calculated 14979; [**SM**+MeCN+H]$^+$ found 14983, calculated 14982; [**P**(s)+H]$^+$ found 15193, calculated 15192; [**P**(s)+H_2O+H]$^+$ found 15211, calculated 15211; [**P**(s)+Na]$^+$ found 15215, calculated 15215; [**P**(s)+K]$^+$ found 15232, calculated 15230; [**P**(d)+H]$^+$ found 15445, calculated 15443; [**P**(t)+H]$^+$ found 15696, calculated 15694.

CjX183-D WT

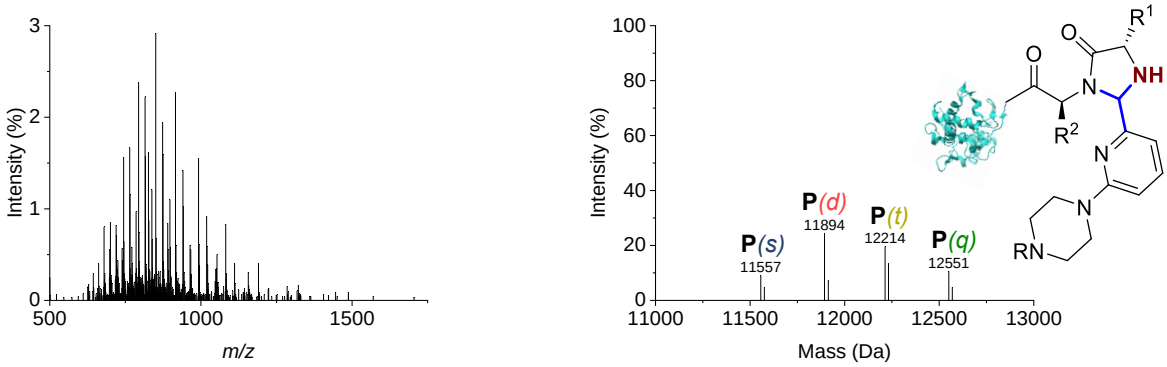


2-PCA **2.1***

100%

49(s):39(d):12(t)

MS (ESI$^+$) [**P**(s)+H]$^+$ found 11503, calculated 11491; [**P**(s)+H_2O+H]$^+$ found 11520, calculated 11521; [**P**(s)+2H_2O+H]$^+$ found 11537, calculated 11539; [**P**(s)+3H_2O+H]$^+$ found 11555, calculated 11557; [**P**(d)+H_2O+H]$^+$ found 11786, calculated 11786; [**P**(d)+2H_2O+H]$^+$ found 11803, calculated 11804; [**P**(d)+3H_2O+H]$^+$ found 11820, calculated 11822; [**P**(t)+2H_2O+H]$^+$ found 12069, calculated 12069; [**P**(t)+3H_2O+H]$^+$ found 12088, calculated 12087.

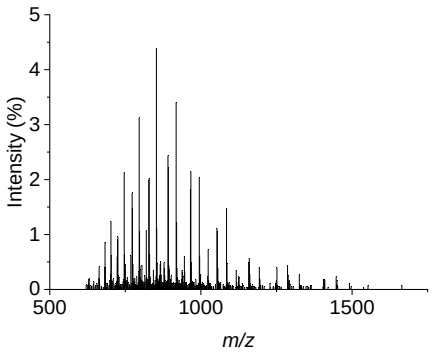


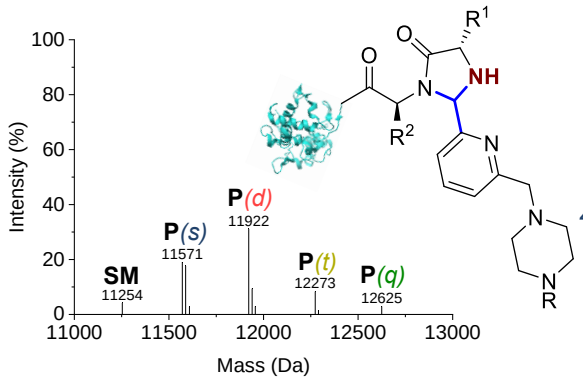
2-PCA **2.2**

100%

15(s):34(d):35(t):16(q)

MS (ESI$^+$) [**P**(s)+H]$^+$ found 11557, calculated 11545; [**P**(s)+H_2O+H]$^+$ found 11576, calculated 11575; [**P**(d)+H_2O+H]$^+$ found 11894, calculated 11894; [**P**(d)+K]$^+$ found 11914, calculated 11914; [**P**(t)+H_2O+H]$^+$ found 12214, calculated 12213; [**P**(t)+2H_2O+H]$^+$ found 12232, calculated 12231; [**P**(q)+2H_2O+H]$^+$ found 12551, calculated 12550; [**P**(q)+3H_2O+H]$^+$ found 12569, calculated 12568.



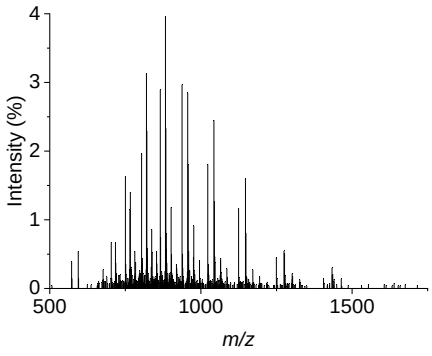


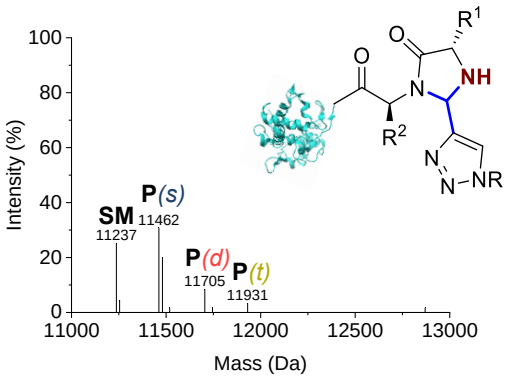
2-PCA **2.3**

96%

40(s):43(d):10(t):3(q)

MS (ESI$^+$) [**SM**+H_2O+H]$^+$ found 11254, calculated 11244; [**P**(s)+H]$^+$ found 11571, calculated 11569; [**P**(s)+H_2O+H]$^+$ found 11588, calculated 11587; [**P**(s)+K]$^+$ found 11608, calculated 11607; [**P**(d)+H_2O+H]$^+$ found 11922, calculated 11920; [**P**(d)+2H_2O+H]$^+$ found 11940, calculated 11938; [**P**(d)+3H_2O+H]$^+$ found 11956, calculated 11956; [**P**(t)+2H_2O+H]$^+$ found 12273, calculated 12271; [**P**(t)+3H_2O+H]$^+$ found 12291, calculated 12289; [**P**(q)+3H_2O+H]$^+$ found 12625, calculated 12622.



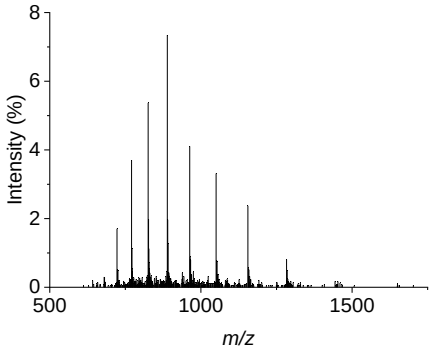


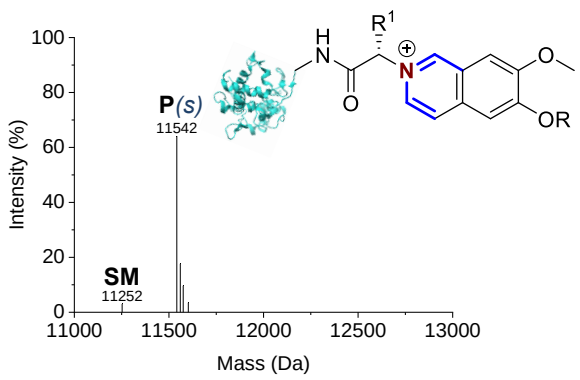
TA4C **2.4**

69%

55(s):11(d):3(t)

MS (ESI$^+$) [**SM**+H]$^+$ found 11237, calculated 11226; [**SM**+H_2O+H]$^+$ found 11254, calculated 11255; [**P**(s)+H]$^+$ found 11462, calculated 11462; [**P**(s)+H_2O+H]$^+$ found 11481, calculated 11480; [**P**(s)+H_2O+K]$^+$ found 11519, calculated 11518; [**P**(d)+H_2O+H]$^+$ found 11705, calculated 11705; [**P**(d)+H_2O+K]$^+$ found 11745, calculated 11743; [**P**(t)+H_2O+H]$^+$ found 11931, calculated 11930.

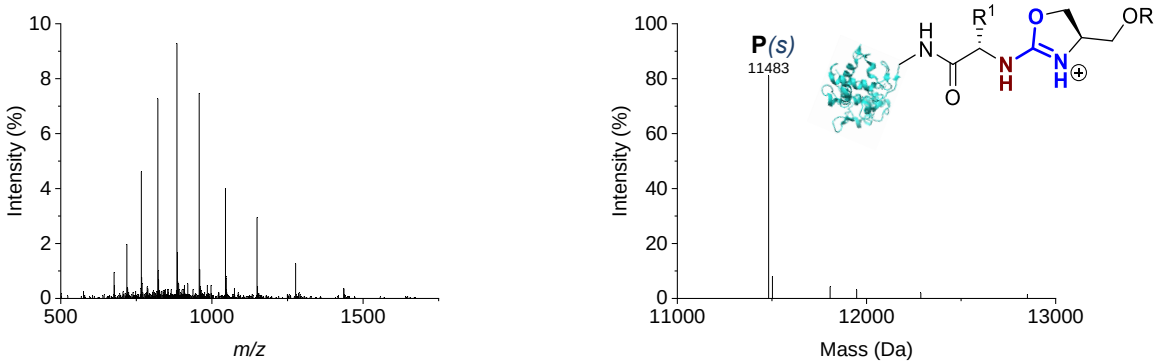




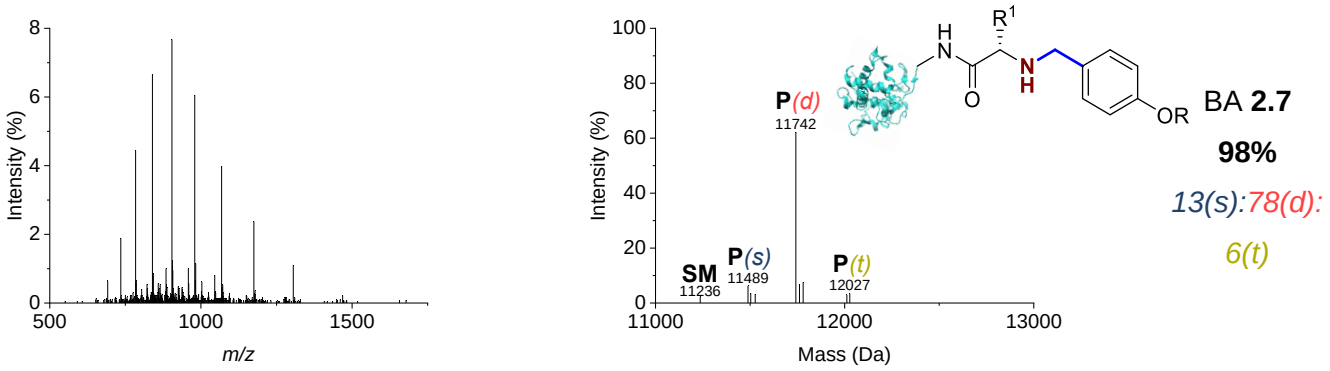
2-EBA **2.5**

97%

MS (ESI$^+$) [**SM**+H_2O+H]$^+$ found 11252, calculated 11244; [**P**(s)+H]$^+$ found 11542, calculated 11539; [**P**(s)+H_2O+H]$^+$ found 11560, calculated 11557; [**P**(s)+K]$^+$ found 11575, calculated 11577; [**P**(s)+MeCN+Na]$^+$ found 11603, calculated 11602.

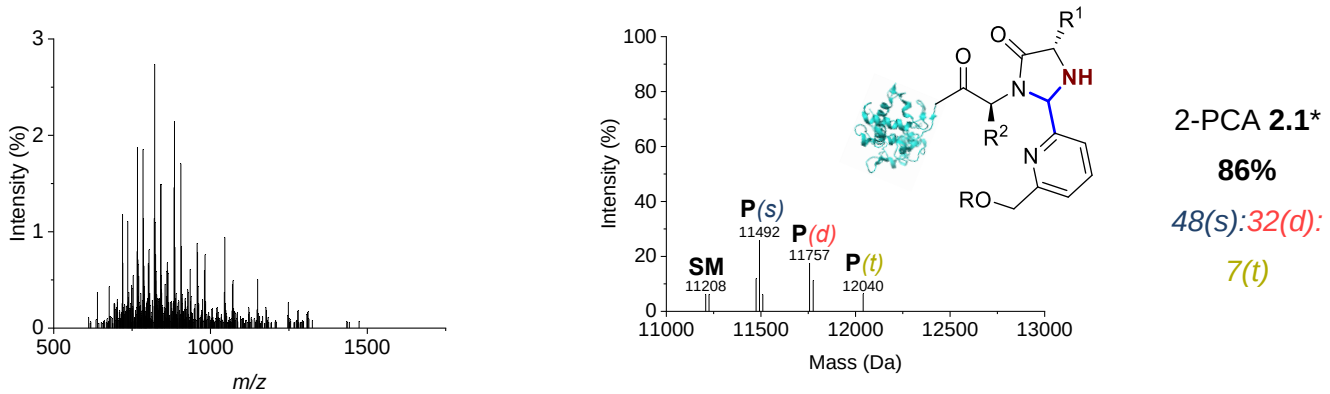


MS (ESI$^+$) [**P**(s)+H]$^+$ found 11483, calculated 11471; [**P**(s)+H_2O+H]$^+$ found 11503, calculated 11501.

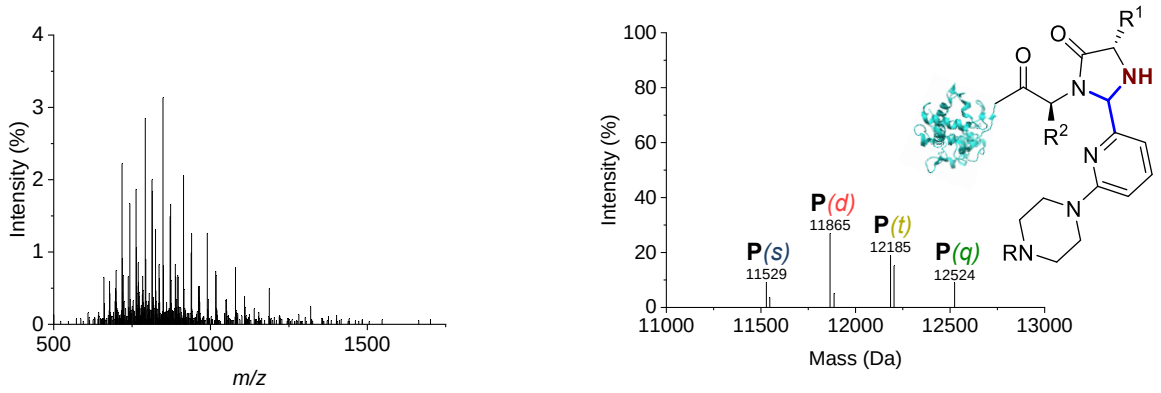


MS (ESI$^+$) [**SM**+H]$^+$ found 11236, calculated 11226; [**P**(s)+H]$^+$ found 11489, calculated 11488; [**P**(s)+H_2O+H]$^+$ found 11504, calculated 11506; [**P**(s)+K]$^+$ found 11527, calculated 11526; [**P**(d)+H]$^+$ found 11742, calculated 11741; [**P**(d)+H_2O+H]$^+$ found 11761, calculated 11759; [**P**(d)+K]$^+$ found 11781, calculated 11778; [**P**(t)+H_2O+H]$^+$ found 12012, calculated 12011; [**P**(t)+K]$^+$ found 12027, calculated 12031.

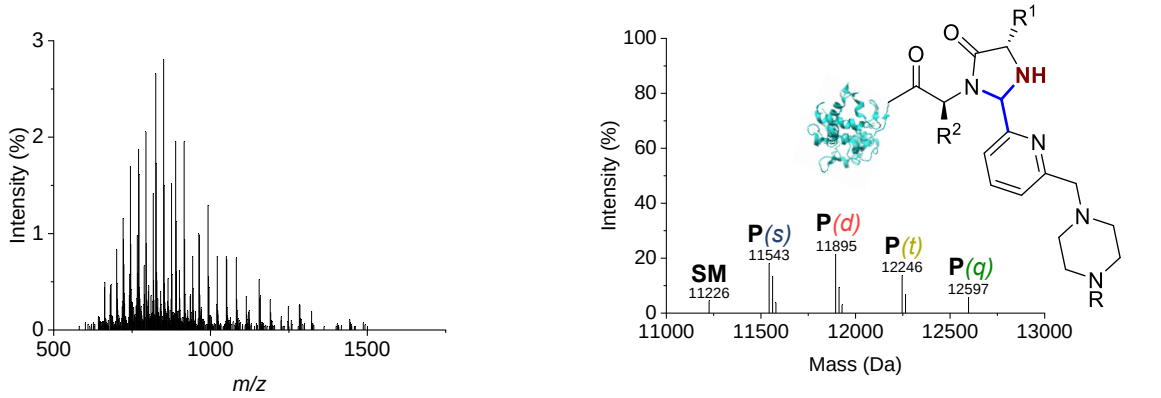
CjX183-D R51K



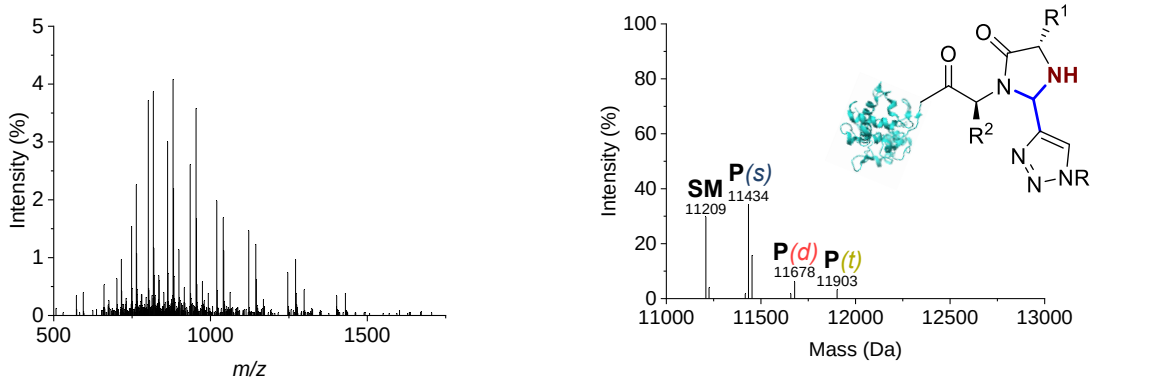
MS (ESI$^+$) [**SM**+H]$^+$ found 11208, calculated 11198; [**SM**+H_2O+H]$^+$ found 11225, calculated 11226; [**P**(s)+H]$^+$ found 11474, calculated 11473; [**P**(s)+H_2O+H]$^+$ found 11492, calculated 11491; [**P**(s)+2H_2O+H]$^+$ found 11510, calculated 11509; [**P**(d)+H_2O+H]$^+$ found 11757, calculated 11756; [**P**(d)+2H_2O+H]$^+$ found 11776, calculated 11774; [**P**(t)+2H_2O+H]$^+$ found 12040, calculated 12039.



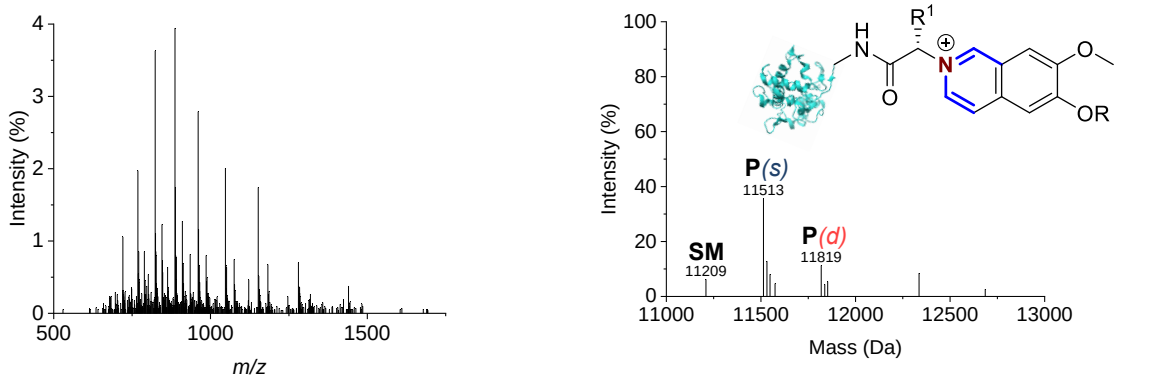
MS (ESI$^+$) [**P**(s)+H]$^+$ found 11529, calculated 11517; [**P**(s)+H_2O+H]$^+$ found 11547, calculated 11547; [**P**(d)+H_2O+H]$^+$ found 11865, calculated 11866; [**P**(d)+2H_2O+H]$^+$ found 11887, calculated 11884; [**P**(t)+H_2O+H]$^+$ found 12185, calculated 12185; [**P**(t)+2H_2O+H]$^+$ found 12203, calculated 12203; [**P**(q)+2H_2O+H]$^+$ found 12524, calculated 12522.



MS (ESI$^+$) [**SM**+H_2O+H]$^+$ found 11226, calculated 11216; [**P**(s)+H]$^+$ found 11543, calculated 11541; [**P**(s)+H_2O+H]$^+$ found 11562, calculated 11559; [**P**(s)+2H_2O+H]$^+$ found 11579, calculated 11577; [**P**(d)+H_2O+H]$^+$ found 11895, calculated 11892; [**P**(d)+2H_2O+H]$^+$ found 11913, calculated 11910; [**P**(d)+3H_2O+H]$^+$ found 11929, calculated 11928; [**P**(t)+2H_2O+H]$^+$ found 12246, calculated 12244; [**P**(t)+3H_2O+H]$^+$ found 12264, calculated 12262; [**P**(q)+3H_2O+H]$^+$ found 12597, calculated 12595.



MS (ESI$^+$) [**SM**+H]$^+$ found 11209, calculated 11198; [**SM**+H_2O+H]$^+$ found 11226, calculated 11227; [**P**(s)–H_2O+H]$^+$ found 11417, calculated 11416; [**P**(s)+H]$^+$ found 11434, calculated 11434; [**P**(s)+H_2O+H]$^+$ found 11453, calculated 11452; [**P**(d)+H]$^+$ found 11658, calculated 11659; [**P**(d)+H_2O+H]$^+$ found 11678, calculated 11677; [**P**(t)+H_2O+H]$^+$ found 11903, calculated 11902.

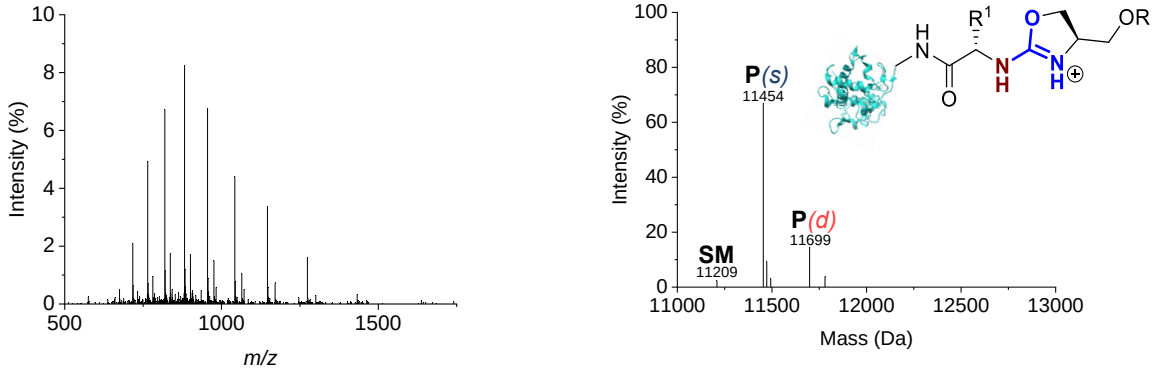


2-EBA **2.5**

92%

69(s):24(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 11209, calculated 11198; [**P**(s)+H]$^+$ found 11513, calculated 11513; [**P**(s)+H_2O+H]$^+$ found 11532, calculated 11531; [**P**(s)+2H_2O+H]$^+$ found 11548, calculated 11549; [**P**(s)+MeCN+Na]$^+$ found 11574, calculated 11576; [**P**(d)+H]$^+$ found 11819, calculated 11818; [**P**(d)+H_2O+H]$^+$ found 11838, calculated 11836; [**P**(d)+2H_2O+H]$^+$ found 11853, calculated 11854.

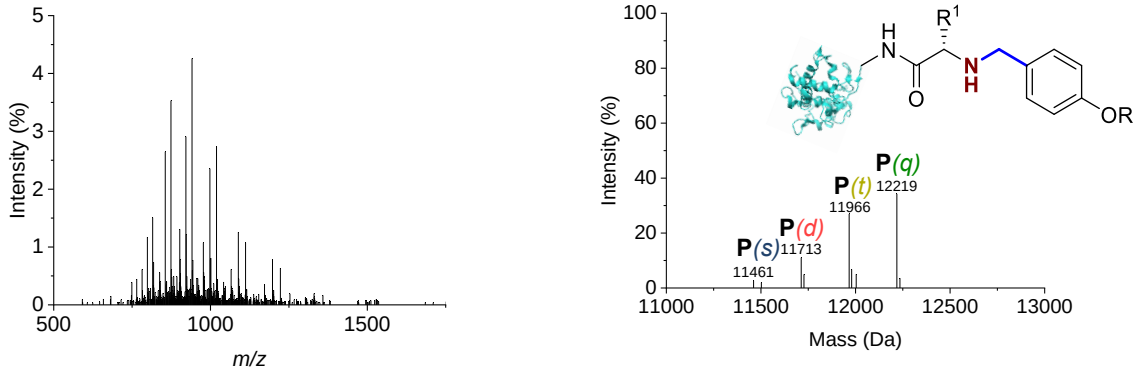


Ox **2.6**

98%

79(s):18(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 11209, calculated 11198; [**P**(s)+H]$^+$ found 11454, calculated 11455; [**P**(s)+H_2O+H]$^+$ found 11473, calculated 11473; [**P**(s)+K]$^+$ found 11494, calculated 11492; [**P**(d)+H]$^+$ found 11699, calculated 11700; [**P**(d)+2MeCN+H]$^+$ found 11780, calculated 11782.



BA **2.7**

100%

5(s):16(d):

40(t):39(q)

MS (ESI$^+$) [**P**(s)+H]$^+$ found 11461, calculated 11450; [**P**(s)+MeCN+H]$^+$ found 11502, calculated 11502; [**P**(d)+H]$^+$ found 11713, calculated 11713; [**P**(d)+H_2O+H]$^+$ found 11728, calculated 11731; [**P**(t)+H]$^+$ found 11966, calculated 11965; [**P**(t)+H_2O+H]$^+$ found 11980, calculated 11983; [**P**(t)+K]$^+$ found 12004, calculated 12003; [**P**(q)+H]$^+$ found 12219, calculated 12217; [**P**(q)+H_2O+H]$^+$ found 12235, calculated 12235.

Figure S2.7. Raw (left) and deconvoluted (right) mass spectra for RNase A, myoglobin, clostripain LC, CjX183-D WT and CjX183-D R51K conjugates. **Following LC-MS of each protein sample, water was injected into the LC column to remove residual protein; for samples marked (*), conversions were determined from MS traces of residual protein due to co-elution with excess small molecule.*

2.4.3.5 CjX183-D off-target reactivity

i) Trypsin digestion

CjX183-D WT was modified with compound **2.3** according to **General Procedure 2A**, on a 40 µL scale. A control sample was also prepared: a solution of CjX183-D WT (40 µL, 50 µM, 2 nmol, in 50 mM pH 7.5 sodium phosphate buffer) was incubated at 37 °C for 23 h with agitation (1000 rpm).

The protein was trypsin-digested in gel; digestion and analysis by MALDI-MS and LC-MS/MS was carried out by Dr. Adam Dowle and Dr. Chris Taylor. MALDI-MS was performed using a Bruker ultraflex-III and LC-MS/MS was performed in DDA mode using a Bruker maXis qTOF, with elution of peptides from a 50 cm microPillar column over 30 min runs.

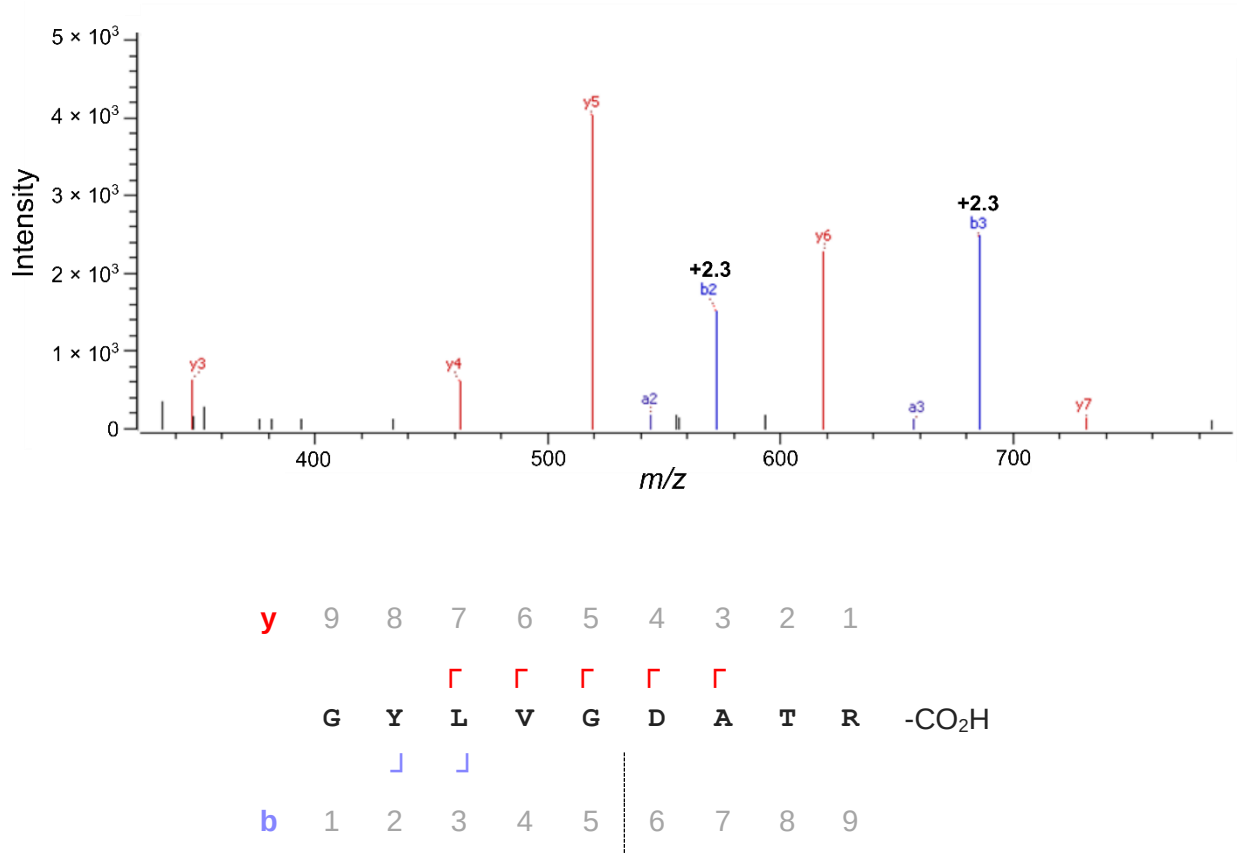


Figure S2.8. Tandem mass spectrum and sequence map from LC-MS of the singly charged positive ion at m/z 1302, due to the modification of the N-terminal peptide GYLVGDATR with compound **2.3** (+333 Da). The **b** and **a** (with losses of ammonia) ions are shown in blue; **y** ions shown in red; fragments modified with compound **2.3** are indicated as **+2.3**. Fragment ion masses indicated that the modification occurred within the first two N-terminal amino acids due to the modification of N-terminal ion **b2** and absence of modification of C-terminal ion **y7**.

ii) AspN and GluC digestion

CjX183-D WT was modified with compound **2.3** according to **General Procedure 2A**, on a 40 µL scale. A control sample was also prepared: a solution of CjX183-D WT (40 µL, 50 µM, 2 nmol, in 50 mM pH 7.5 sodium phosphate buffer) was incubated at 37 °C for 23 h with agitation (1000 rpm).

Protein digestion and analysis was carried out by Dr. Adam Dowle and Dr. Chris Taylor. The protein was in gel digested post reduction with DTE and alkylation with iodoacetamide. Samples were digested in parallel with AspN and GluC proteases. The individual digests were analysed by LC-MS/MS in DDA mode using an Orbitrap Fusion. Samples were eluted onto the MS over 1 h with elution from a 50 cm EN PepMap column driven by a Waters mClass UPLC. MS1 spectra were acquired in the Orbitrap mass analyser and MS2 spectra were acquired in the linear iontrap. Data analysis was carried out through Byonic, PEAKS and FragPipe, using both fixed and wildcard modifications.

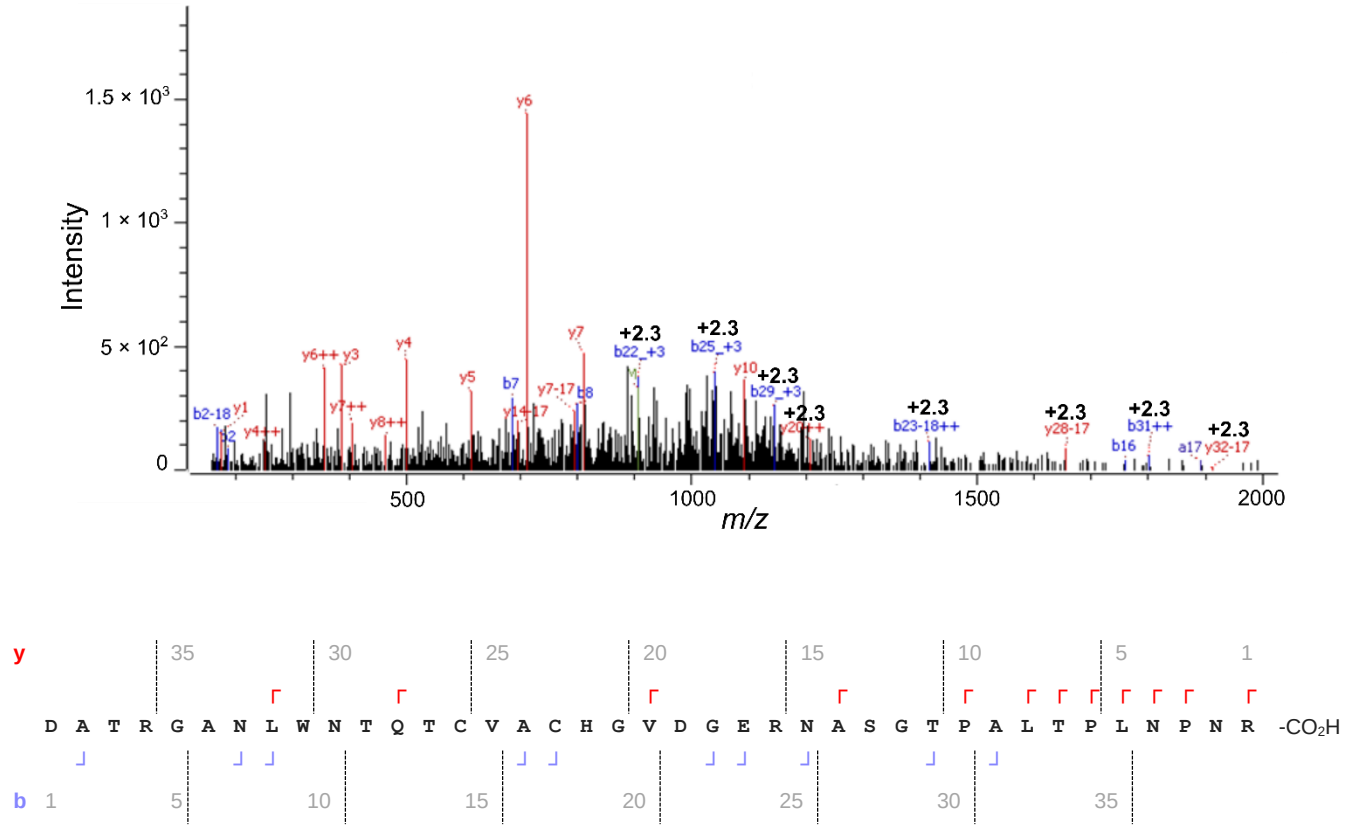


Figure S2.9. Tandem mass spectrum and sequence map from LC-MS of the singly charged positive ion at m/z 4525, due to the modification of the peptide DATRGANLWNTQTCVACHGVDGERNASGTPALTPLNPNR with compound **2.3** (+333 Da), assuming 100% alkylation of the Cys residues in the protein (+57 Da). The **b** and **a** (with losses of ammonia) ions are shown in blue; **y** ions shown in red; fragments modified with compound **2.3** are indicated as **+2.3**. Fragment ion masses indicated that the modification occurred within the ^{20}VDG$_{22}$ amino acids due to the modification of both N-terminal ion **b22** and C-terminal ion **y20**.

iii) Probing the nature of the additional modification

β-Mercaptoethanol nucleophile

CjX183-D WT was modified with compounds **2.37** and **2.38** according to **General Procedure 2A**. Protein conjugates were purified by dialysis to remove excess reagent (4 °C, 3.5 kDa MWCO; 1 × water, 2 h; 1 × water, 4 h; 1 × water, 20 h; 1 × water, 5 h; 1 × 50 mM Na phosphate buffer, pH 7.5, 16 h). β-mercaptoethanol (1 µL, 15 nmol, 7 equiv.) was added to an aliquot of the protein conjugate solutions (45 µL, 50 µM, 2.25 nmol,

1 equiv., in 50 mM pH 7.5 sodium phosphate buffer), and the mixture incubated at 37 °C for 5 h with agitation (1000 rpm). A control sample was prepared: β-mercaptoethanol (1 μL, 15 nmol, 7 equiv.) was added to a solution of CjX183-D WT (45 μL, 50 μM, 2.25 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer), and the mixture incubated at 37 °C for 5 h with agitation (1000 rpm). Conversion was determined by LC-MS of the reaction mixtures before and after dialysis, and 5 h after thiol addition.

Compound	PCA modification conversion (%)		
	Crude	Purified	After thiol addition
None (control)	-	-	**0** *4(unmodified+thiol)*
2.37	**55**	**40**	**16** *13(s):4(s+thiol)*
2.38	**69** *55(s):14(d)*	**62**	**53** *45(s):5(d):3(t)*

Table S2.3. Reactivity of CjX183 conjugates with β-mercaptoethanol. Modification of CjX183-D WT with compounds **2.37** and **2.38** under conditions outlined in **General Procedure 2A**, and conversions observed upon addition of β-mercaptoethanol as a nucleophile for Michael addition (7 equiv.).

L-Cysteine nucleophile

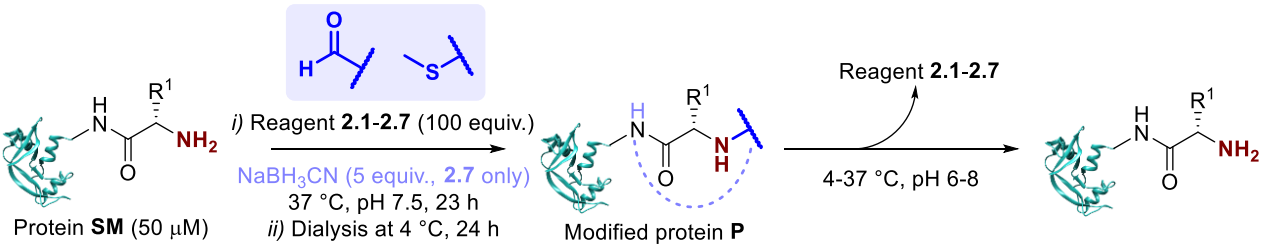
CjX183-D WT was modified with compounds **2.3**, **2.37** and **2.38** according to **General Procedure 2A**. Protein conjugates were purified by dialysis to remove excess reagent (4 °C, 3.5 kDa MWCO; 1 × water, 18 h; 3 × water, 2 h; 1 × water, 18 h; 1 × 50 mM Na phosphate buffer, pH 7.5, 5 h). A solution of L-cysteine (20 μL, 10 mM, 0.2 μmol, 100 equiv.) was added to an aliquot of the protein conjugate solutions (40 μL, 50 μM, 2 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer), and the mixture incubated at 37 °C for 25 h with agitation (1000 rpm). A control sample was prepared: L-cysteine (20 μL, 10 mM, 0.2 μmol, 100 equiv.) was added to a solution of CjX183-D WT (40 μL, 50 μM, 2 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer), and the mixture incubated at 37 °C for 25 h with agitation (1000 rpm). Conversion was determined by LC-MS of the reaction mixtures before and after dialysis, and 25 h after thiol addition.

Compound	PCA modification conversion (%)		
	Crude	Purified	After thiol addition
None (control)	-	-	**0**
2.3	**44** *20(s):24(d)*	**58** *37(s):21(d)*	**34** *14(s):20(d)*
2.37	**42** *39(s):3(d)*	**36** *32(s):3(d)*	**7**
2.38	**63** *54(s):9(d)*	**58** *53(s):4(d)*	**23**

Table S2.4. Reactivity of CjX183 conjugates with L-cysteine. Modification of CjX183-D WT with compounds **2.3**, **2.37** and **2.38** under conditions outlined in **General Procedure 2A**, and conversions observed upon addition of L-cysteine as a nucleophile for Michael addition (100 equiv.).

2.4.3.6 Conjugate stability

i) pH and temperature stability



RNase A, myoglobin, and CjX183-D WT were modified with compounds **2.1**-**2.7** according to **General Procedure 2A**, on a 150 μL scale. For compound **2.7**, 5 equiv. of sodium cyanoborohydride in 7.5 μL of water was also added, and compound **2.7** was added as a more concentrated solution (67.5 μL, 11 mM) to keep the total volume constant. For modifications of RNase A and myoglobin, 25 equiv. 2-EBA **2.5** was used. For the modification of myoglobin, 67 equiv. 2-PCA **2.2** were used. Protein conjugates **P** were purified by dialysis to remove excess reagent (4 °C, 3.5 kDa MWCO; 1 × 100 mM PBS, pH 7.5, 4 h; 2 × water, 4 h; 1 × water, 16 h).

Following purification, aliquots of protein (20 μL) were diluted with water (80 μL), placed inside a 3.5 kDa MWCO dialysis membrane, and incubated in sodium phosphate buffer (40 mL, 50 mM) at either pH 6, 7, or 8, and at either 4°C, 22 °C, or 37 °C. Samples were taken from the dialysis membranes at *t* = 0, 48, and 168 h, and analysed via LC-MS.

Temp. /°C	pH	Protein	Compound	Conversion (%)		
				0 h	48 h	168 h
4	7	RNase A	2-PCA **2.1**	**81**	**79**	**79**
			2-PCA **2.2**	**84**	**90**	**82**
			2-PCA **2.3**	**88**	**88**	**82** *80(s):3(d)*
			TA4C **2.4**	**63**	**59**	**51**
			2-EBA **2.5**	**49** *39(s):11(d)*	**45** *34(s):11(d)*	**43** *40(s):3(d)*
			Ox **2.6**	**0**	-	-
			BA **2.7**	**4**	**0**	-
4	7	Myoglobin	2-PCA **2.1**	**35**	**35**	**29**
			2-PCA **2.2**	**23**	**15**	**13**
			2-PCA **2.3**	**35**	**31**	**29**
			TA4C **2.4**	**0**	-	-
			2-EBA **2.5**	**48** *29(s):14(d):4(t)*	**38** *33(s):6(d)*	**25**
			Ox **2.6**	**42** *37(s):5(d)*	**30** *26(s):4(d)*	**40** *36(s):4(d)*
			BA **2.7**	**0**	-	-
4	7	CjX183-D WT	2-PCA **2.1**	**87**	**74**	**80**

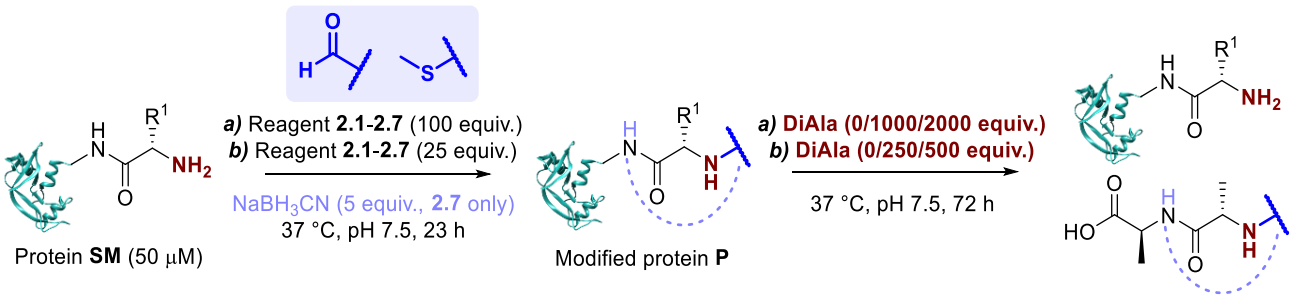
				24(s):63(d)	*31(s):43(d)*	*13(s):68(d)*
			2-PCA **2.2**	**60**	**34**	**31**
				34(s):26(d)	*18(s):16(d)*	*14(s):17(d)*
			2-PCA **2.3**	**81**	**67**	**58**
				50(s):31(d)	*37(s):30(d)*	*31(s):27(d)*
			TA4C **2.4**	**20**	**8**	**4**
			2-EBA **2.5**	**100**	**81**	**89**
			Ox **2.6**	**2**	-	-
			BA **2.7**	**0**	-	-
22	7	RNase A	2-PCA **2.1**	**81**	**61**	**56**
			2-PCA **2.2**	**84**	**81**	**58**
			2-PCA **2.3**	**88**	**77**	**61**
			TA4C **2.4**	**63**	**46**	**23**
			2-EBA **2.5**	**49**	**49**	**35**
				39(s):11(d)	*40(s):10(d)*	*29(s):6(d)*
			Ox **2.6**	**0**	-	-
			BA **2.7**	**4**	**0**	-
22	7	Myoglobin	2-PCA **2.1**	**35**	**31**	**36**
			2-PCA **2.2**	**23**	**19**	**14**
			2-PCA **2.3**	**35**	**34**	**24**
			TA4C **2.4**	**0**	-	-
			2-EBA **2.5**	**48**	**35**	**30**
				29(s):14(d):4(t)	*30(s):5(d)*	*16(s):14(d)*
			Ox **2.6**	**42**	**39**	**36**
				37(s):5(d)	*30(s):9(d)*	*30(s):6(d)*
			BA **2.7**	**0**	-	-
22	7	CjX183-D WT	2-PCA **2.1**	**87**	-	**70**
				24(s):63(d)		*19(s):51(d)*
			2-PCA **2.2**	**60**	-	**23**
				34(s):26(d)		*16(s):7(d)*
			2-PCA **2.3**	**81**	-	**62**
				50(s):31(d)		*13(s):25(d)*
			TA4C **2.4**	**20**	-	**5**
			2-EBA **2.5**	**100**	-	*87(s):3(d):2(t)*
			Ox **2.6**	**2**	-	-
			BA **2.7**	**0**	-	-
37	7	RNase A	2-PCA **2.1**	**81**	**26**	**9**
			2-PCA **2.2**	**84**	**16**	**0**
			2-PCA **2.3**	**88**	**40**	**13**
						0(s):13(d)
			TA4C **2.4**	**63**	**9**	**0**
			2-EBA **2.5**	**49**	**44**	**22**
				39(s):11(d)	*36(s):8(d)*	
			Ox **2.6**	**0**	-	-
			BA **2.7**	**4**	**0**	-
37	7	Myoglobin	2-PCA **2.1**	**35**	**30**	**30**

			2-PCA **2.2**	**23**	**7**	**10**
			2-PCA **2.3**	**35**	**28**	**23**
			TA4C **2.4**	**0**	-	-
			2-EBA **2.5**	**48** *29(s):14(d):4(t)*	**27** *23(s):4(d)*	**15**
			Ox **2.6**	**42** *37(s):5(d)*	**30** *25(s):5(d)*	**26**
			BA **2.7**	**0**	-	-
37	7	CjX183-D WT	2-PCA **2.1**	**87** *24(s):63(d)*	-	**46**
			2-PCA **2.2**	**60** *34(s):26(d)*	-	**20**
			2-PCA **2.3**	**81** *50(s):31(d)*	-	**41** *29(s):11(d)*
			TA4C **2.4**	**20**	-	**4**
			2-EBA **2.5**	**100**	-	**85**
			Ox **2.6**	**2**	-	-
			BA **2.7**	**0**	-	-
22	6	RNase A	2-PCA **2.1**	**81**	**71**	**51**
			2-PCA **2.2**	**95**	**77**	**56**
			2-PCA **2.3**	**88**	**75**	**62**
			TA4C **2.4**	**63**	**46**	**21**
			2-EBA **2.5**	**49** *39(s):11(d)*	**42** *34(s):8(d)*	**50** *41(s):9(d)*
			Ox **2.6**	**0**	-	-
			BA **2.7**	**4**	**0**	-
22	6	Myoglobin	2-PCA **2.1**	**35**	**33**	**32**
			2-PCA **2.2**	**23**	**15**	**9**
			2-PCA **2.3**	**35**	**34**	**19**
			TA4C **2.4**	**0**	-	-
			2-EBA **2.5**	**48** *29(s):14(d):4(t)*	**47** *34(s):10(d):4(t)*	**50** *39(s):11(d)*
			Ox **2.6**	**42** *37(s):5(d)*	**31**	**30**
			BA **2.7**	**0**	-	-
22	6	CjX183-D WT	2-PCA **2.1**	**87** *24(s):63(d)*	-	**69** *24(s):45(d)*
			2-PCA **2.2**	**60** *34(s):26(d)*	-	**24** *17(s):7(d)*
			2-PCA **2.3**	**81** *50(s):31(d)*	-	**44** *15(s):28(d)*
			TA4C **2.4**	**20**	-	**0**
			2-EBA **2.5**	**100**	-	**86**
			Ox **2.6**	**2**	-	-
			BA **2.7**	**0**	-	-
22	8	RNase A	2-PCA **2.1**	**81**	**71**	**53**

			2-PCA **2.2**	**95**	**78**	**42**
			2-PCA **2.3**	**88**	**74**	**53**
			TA4C **2.4**	**63**	**38**	**21**
			2-EBA **2.5**	**49** *39(s):11(d)*	**53** *42(s):11(d)*	**48** *38(s):10(d)*
			Ox **2.6**	**0**	-	-
			BA **2.7**	**4**	**0**	-
22	8	Myoglobin	2-PCA **2.1**	**35**	**34**	**31**
			2-PCA **2.2**	**23**	**17**	**9**
			2-PCA **2.3**	**35**	**34**	**27**
			TA4C **2.4**	**0**	-	-
			2-EBA **2.5**	**48** *29(s):14(d):4(t)*	**38** *27(s):5(d):5(t)*	**30**
			Ox **2.6**	**42** *37(s):5(d)*	**33** *28(s):5(d)*	**35** *30(s):5(d)*
			BA **2.7**	**0**	-	-
22	8	CjX183-D WT	2-PCA **2.1**	**87** *24(s):63(d)*	-	**65** *27(s):39(d)*
			2-PCA **2.2**	**60** *34(s):26(d)*	-	**25** *18(s):6(d)*
			2-PCA **2.3**	**81** *50(s):31(d)*	-	**44** *18(s):26(d)*
			TA4C **2.4**	**20**	-	**2**
			2-EBA **2.5**	**100**	-	**88**
			Ox **2.6**	**2**	-	-
			BA **2.7**	**0**	-	-

Table S2.5. Stability of protein conjugates to pH and temperature over time. Modification of RNase A, myoglobin, and CjX183-D WT with compounds **2.1**-**2.7** under conditions outlined in **General Procedure 2A**, and conversions observed over time upon exposure to a range of conditions (pH 6-8, 4-37 °C).

ii) Stability to competitive dipeptide



a) For the modification of RNase A with 100 equiv. reagent: RNase A was modified with compounds **2.1**-**2.7** according to **General Procedure 2A**, on a 200 µL scale. For compound **2.7**, 5 equiv. of sodium cyanoborohydride in 10 µL of water was also added, and compound **2.7** was added as a more concentrated

solution (90 µL, 11 mM) to keep the total volume constant. The reaction was then split into 3 aliquots (65 µL). To these, were added a different concentration of L-alanyl-L-alanine (DiAla; 26 µL; 0, 125, or 250 mM; 0, 1000, or 2000 equiv.) in sodium phosphate buffer (50 mM pH 7.5). The reactions were incubated at 37 °C for 72 h with agitation (1000 rpm). Aliqouts (10 µL) were taken from each mixture at *t* = 24 and 72 h, diluted with water (30 µL), and analysed by LC-MS.

b) For the modification of RNase A with 25 equiv. reagent: Reactions were run as described above, but using 25 equiv. of reagents **2.1**-**2.7** and 0, 250 or 500 equiv. DiAla. LC-MS analysis was carried out at *t* = 72 h.

a) For the modification of CjX183-D WT with 100 equiv. reagent: CjX183-D WT was modified with compounds **2.1**-**2.7** according to **General Procedure 2A**, on a 150 µL scale. For compound **2.7**, 5 equiv. of sodium cyanoborohydride in 7.5 µL of water was also added, and compound **2.7** was added as a more concentrated solution (67.5 µL, 11 mM) to keep the total volume constant. The reaction was then split into 3 aliquots (40 µL). To these, were added a different concentration of L-alanyl-L-alanine (DiAla; 16 µL; 0, 125, or 250 mM; 0, 1000, or 2000 equiv.) in sodium phosphate buffer (50 mM pH 7.5). The reactions were incubated at 37 °C for 72 h with agitation (1000 rpm). Aliqouts (10 µL) were taken from each mixture at *t* = 24 and 72 h, diluted with water (30 µL), and analysed by LC-MS.

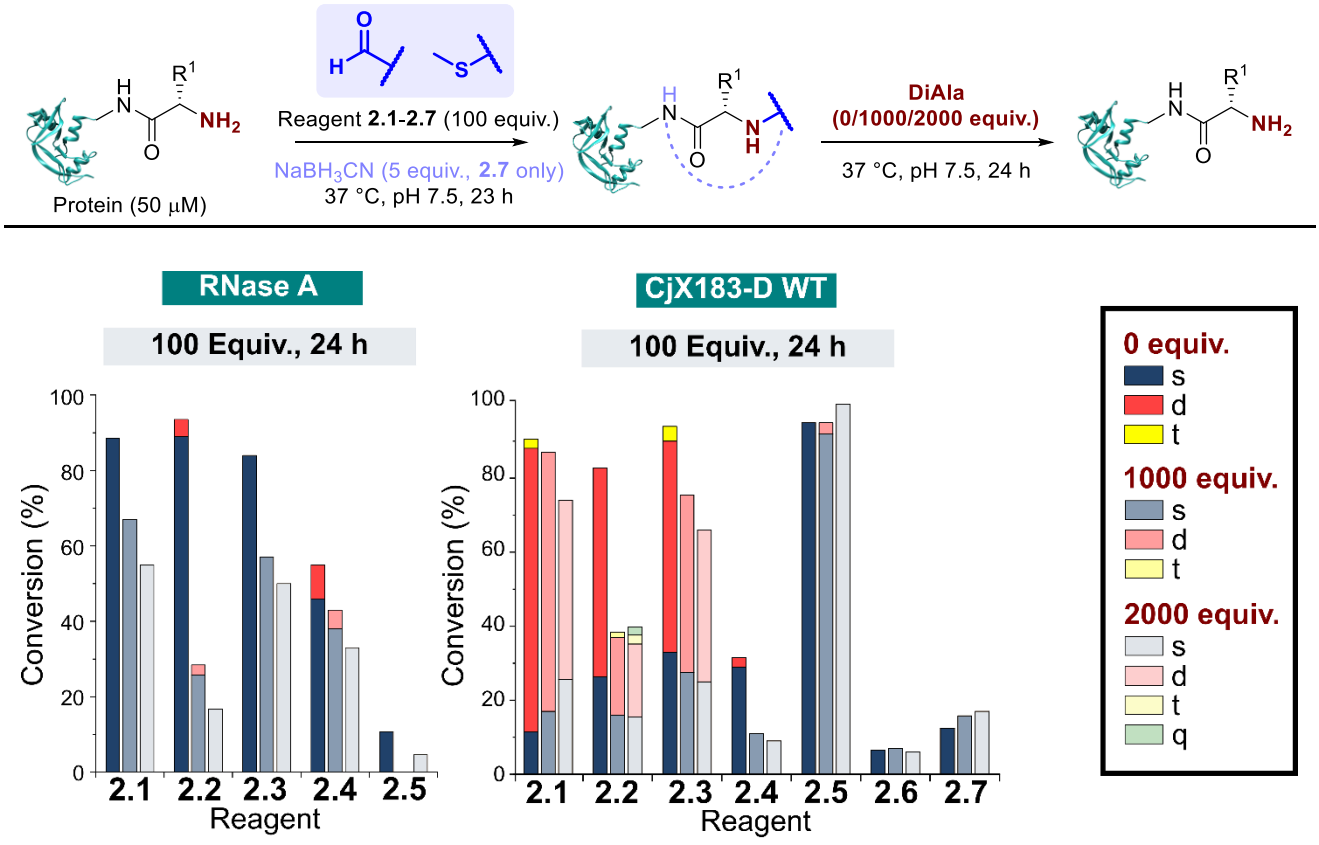


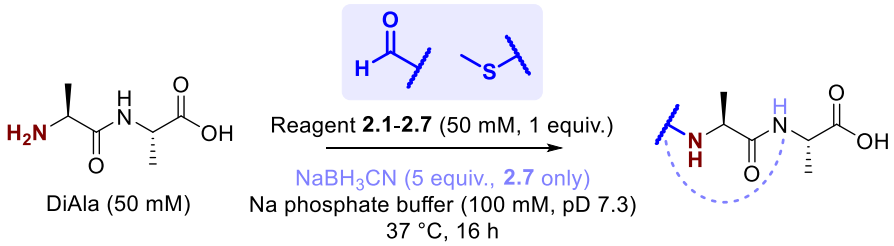
Figure S2.10. Stability of protein conjugates to competitor over time. Modification of RNase A and CjX183-D WT with compounds **2.1**-**2.7** under conditions outlined in **General Procedure 2A**, and conversions observed over time upon addition of DiAla as a competitor (0-2000 equiv.).

Protein	No. equiv. protein modifier	Compound	Time /h	Conversion (%)		
				0 equiv. DiAla	1000 equiv. DiAla	2000 equiv. DiAla
RNase A	100	2-PCA **2.1**	24	**88**	**78**	**73**
			72	**88**	**67**	**55**
		2-PCA **2.2**	24	**91**	**65**	**59**
			72	**94** *89(s):5(d)*	**28** *26(s):3(d)*	**17**
		2-PCA **2.3**	24	**81**	**77**	**69**
			72	**84**	**57**	**50**
		TA4C **2.4**	24	**67**	**40**	**35**
			72	**68**	**11**	**9**
		Ox **2.6**	24	**52** *39(s):13(d)*	**41** *36(s):5(d)*	**38** *34(s):4(d)*
			72	**56** *46(s):9(d)*	**44** *39(s):5(d)*	**34**
		BA **2.7**	24	**8**	**5**	**5**
			72	**11**	**0**	**5**
RNase A	25	2-PCA **2.1**	72	**21**	**9** *(250 equiv. DiAla)*	**8** *(500 equiv. DiAla)*
		2-PCA **2.2**	72	**72**	**20** *(250 equiv. DiAla)*	**15** *(500 equiv. DiAla)*
		2-PCA **2.3**	72	**74**	**38** *(250 equiv. DiAla)*	**35** *(500 equiv. DiAla)*
		TA4C **2.4**	72	**17**	**5** *(250 equiv. DiAla)*	**0** *(500 equiv. DiAla)*
		2-EBA **2.5**	72	**67** *24(s):26(d):17(t)*	**48** *36(s):12(d)* *(250 equiv. DiAla)*	**40** *30(s):10(d)* *(500 equiv. DiAla)*
		Ox **2.6**	72	**0**	**0**	**0**
		BA **2.7**	72	**0**	**0**	**0**
CjX183-D WT	100	2-PCA **2.1**	24	**91** *11(s):77(d):3(t)*	**88** *17(s):70(d)*	**74** *26(s):49(d)*
			72	**94** *4(s):87(d):3(t)*	**62** *37(s):25(d)*	**65** *41(s):24(d)*
		2-PCA **2.2**	24	**83** *26(s):56(d)*	**39** *16(s):21(d):1(t)*	**40** *15(s):20(d):2(t):2(q)*
			72	**93** *16(s):78(d)*	**23** *16(s):5(d):2(t)*	**20** *14(s):6(d)*
		2-PCA **2.3**	24	**93** *33(s):57(d):4(t)*	**76** *28(s):48(d)*	**66** *25(s):41(d)*
			72	**95**	**57**	**48**

		14(s):75(d):6(t)	*19(s):38(d)*	*19(s):29(d)*
TA4C **2.4**	24	**32** *29(s):3(d)*	**11**	**9**
	72	**30** *28(s):2(d)*	**3**	**0**
2-EBA **2.5**	24	**95**	**96** *92(s):3(d)*	**100**
	72	**98**	**95**	**94**
Ox **2.6**	24	**7**	**7**	**6**
	72	**14**	**4**	**0**
BA **2.7**	24	**12**	**16**	**17**
	72	**18**	**10**	**15**

Table S2.6. Stability of protein conjugates to competitor over time. Modification of RNase A and CjX183-D WT with compounds **2.1**-**2.7** under conditions outlined in **General Procedure 2A**, and conversions observed over time upon addition of DiAla as a competitor (0-2000 equiv.).

2.4.3.7 ^{1}H NMR Kinetic studies



A solution of DiAla (150 µL, 100 mM, 15 µmol, 1 equiv.) was added to solutions of reagents **2.1**-**2.7** (150 µL, 100 mM, 15 µmol, 1 equiv.), both in deuterated sodium phosphate buffer (100 mM, pD 7.3; *note that this value was calculated by addition of a +0.4 correction factor to measured pH value of 6.9, as described by Covington et al.*[28]). Due to discrepancies during the weighing out of small masses of **2.1**-**2.7**, exact concentrations were determined by ^{1}H NMR spectroscopy (2-PCA **2.1**, 45 mM; 2-PCA **2.2,** 39 mM; 2-PCA **2.3,** 46 mM; TA4C **2.4,** 33 mM; 2-EBA **2.5,** 17 mM; Ox **2.6**, 36 mM, BA **2.7**, 50 mM) in deuterated Na phosphate buffer (100 mM, pD 7.3). For compound **2.7**, the experiment was carried out on a 600 µL scale: 5 equiv. of sodium cyanoborohydride in 60 µL of deuterated Na phosphate buffer (100 mM, pD 7.3) was also added, and compound **2.7** was added as a more concentrated solution (240 µL, 125 mM) to keep the total volume constant. The reactions were incubated at 37 °C for 16 h and conversion was followed by ^{1}H NMR spectroscopy at 30 min intervals.

For the modifications with reagents 2.1-2.4. At each timepoint ($t = x\ h$), the concentration of aldehyde [**A**], hydrate [**Hy**], imine [**Im**], and sum of imidazolidinone product [**Pr**] species were calculated *via* the relative integral ratios of diagnostic ^{1}H NMR signals outlined in *Fig. S2.11*. Concentration values were normalised using a correction factor (**CF**) in *Equation 2.1* to remove background noise from the imidazolidinone signals, based on the assumption that the initial rate was linear over the first 3 time-points, and that $[\mathbf{Pr}]_{t=0\ h} = 0$.

Correction factors were calculated as the y-intercept of the linear regression line of the plot of the sum of integrals of product diastereomers over time: see the experimental section in Chapter 3 and 4 for further illustration of normalisation plots.

$$\textbf{Imidazolidinone Conversion} = 100 \times \frac{\int \mathbf{Pr} - \mathbf{CF}}{\int \mathbf{A} + \int \mathbf{Hy} + \int \mathbf{Im} + \int \mathbf{Pr} - \mathbf{CF}} \qquad \textit{(Equation 2.1)}$$

For the modifications with reagents 2.5-2.6. At each timepoint ($t = x\ h$), the concentration of reagent **2.5**/**2.6** [**SM**] and product [**Pr**] species were calculated *via* the relative integral ratios of diagnostic ^{1}H NMR signals outlined in *Fig. S2.11*. Note that for 2-EBA **2.5**, no hydrate or imine were observed. Concentration values were normalised as outlined above, using a correction factor (**CF**) in *Equation 2.2*.

$$\textbf{Conversion} = 100 \times \frac{\int \mathbf{Pr} - \mathbf{CF}}{\int \mathbf{SM} + \int \mathbf{Pr} - \mathbf{CF}} \qquad \textit{(Equation 2.2)}$$

For the modification with reagent 2.7. At each timepoint ($t = x\ h$), the concentration of aldehyde [**A**], hydrate [**Hy**], and product [**Pr**] species were calculated *via* the relative integral ratios of diagnostic ^{1}H NMR signals outlined in *Fig. S2.11*. Concentration values were normalised as outlined above, using a correction factor (**CF**) in *Equation 2.3*.

$$\textbf{Conversion} = 100 \times \frac{\int \mathbf{Pr} - \mathbf{CF}}{\int \mathbf{A} + \int \mathbf{Hy} + \int \mathbf{Pr} - \mathbf{CF}} \qquad \textit{(Equation 2.3)}$$

For all reagents, data were fit to a second order irreversible or reversible kinetic model in Copasi 4.34.251. A simplified model was used;

A + B = C;

For the modifications with reagents 2.1-2.4. **A** represented the overall concentration of the DiAla [**DiAla**] and imine [**Im**] species, **B** represented the overall concentration of aldehyde [**A**], hydrate [**Hy**], and imine [**Im**] forms of reagent **2.1**-**2.4**, and **C** represented the overall concentration of imidazolidinone conjugate [**Pr**]. The model did not integrate hydrate and imine formation, and relied on the assumption that imidazolidinone formation was the rate-limiting step of the reaction.

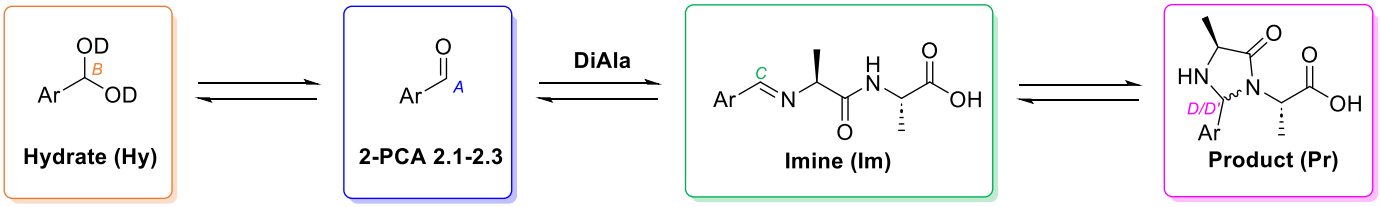
For the modification with reagent 2.5. **A** represented the concentration of DiAla [**DiAla**], **B** represented the concentration of aldehyde [**SM**], and **C** represented the overall concentration of isoquinolinium conjugate [**Pr**]. The model did not integrate hydrate and imine formation, and relied on the assumption that conversion of intermediate imine to isoquinolinium conjugates proceeded faster than the time-scale of ^{1}H NMR (as no imine was observed).

For the modification with reagent 2.6. **A** represented the overall concentration of the DiAla [**DiAla**], **B** represented the concentration of reagent **2.6** [**SM**], and **C** represented the overall concentration of oxazoline conjugate [**Pr**].

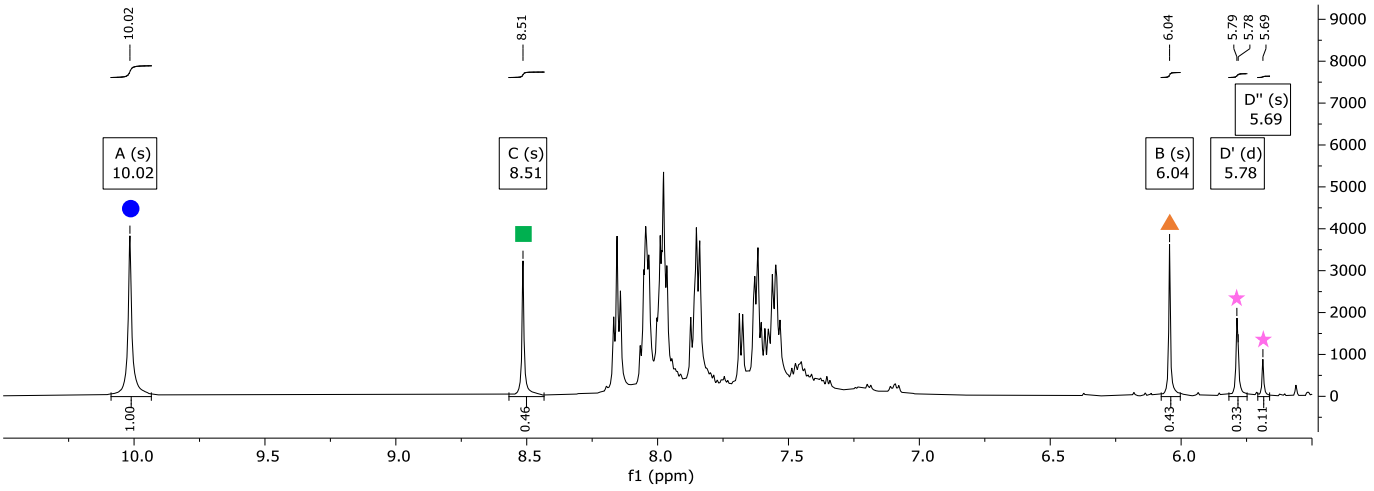
For the modifications with reagents 2.7. **A** represented the overall concentration of the DiAla [**DiAla**], **B** represented the overall concentration of aldehyde [**A**] and hydrate [**Hy**] forms of reagent **2.7**, and **C** represented the overall concentration of reduced conjugate [**Pr**]. The model did not integrate hydrate and imine formation, and relied on the assumption that reduction of the intermediate imine proceeded faster than the time-scale of ^{1}H NMR (as no imine was observed).

k_1 and k_{-1} 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^{-11}-10^{5} M^{-1} s^{-1}; k_{-1} 10^{-11}-10^{5} M^{-1} s^{-1}.

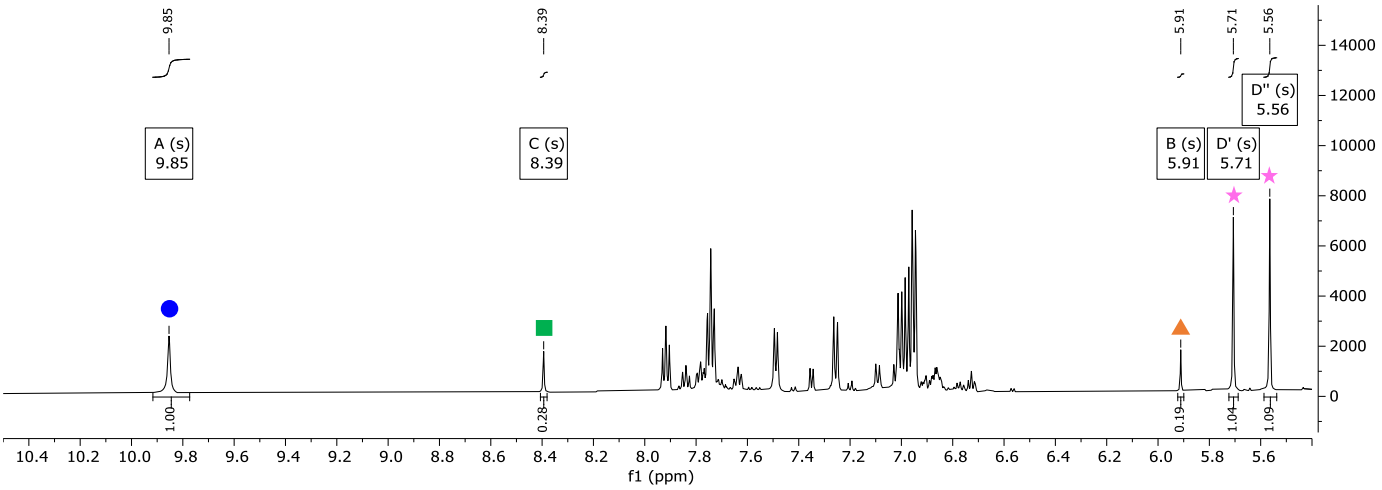
2-PCAs 2.1-2.3



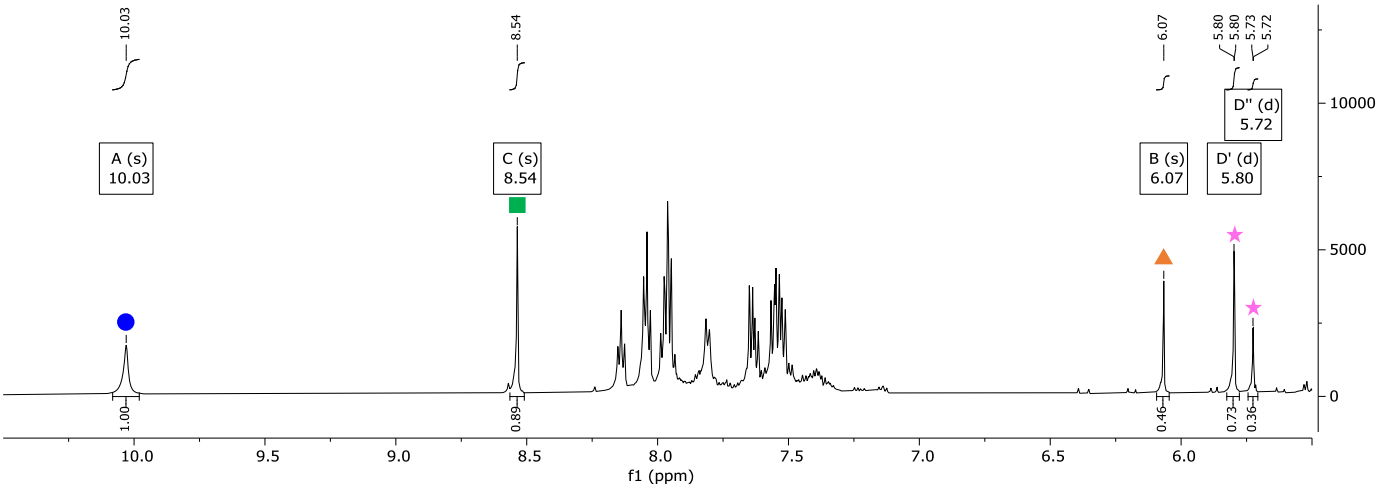
i) *2-PCA* ***2.1*** (t = 16 h)



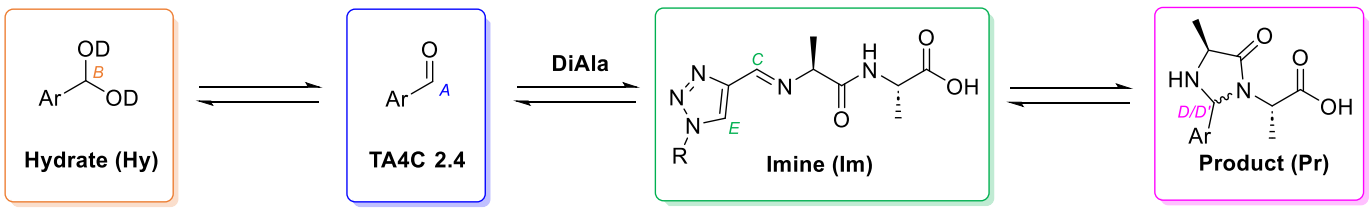
ii) *2-PCA* ***2.2*** (t = 16 h)

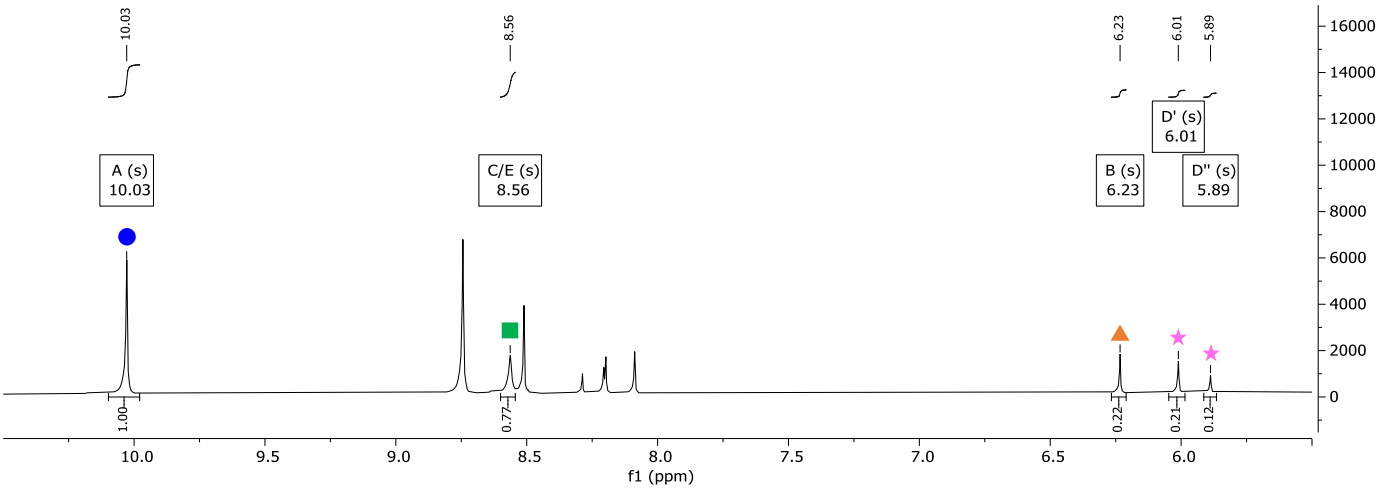


iii) *2-PCA* ***2.3*** (t = 16 h)

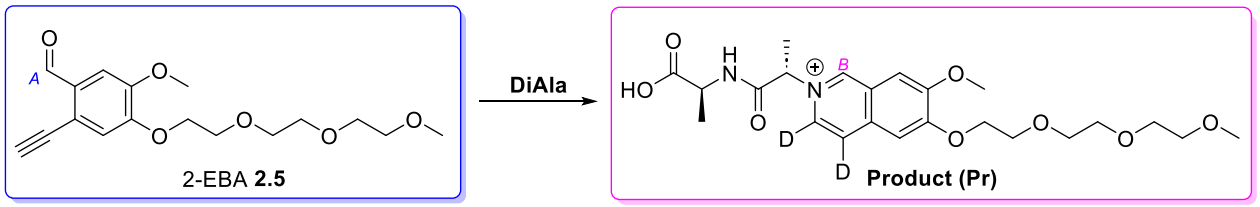


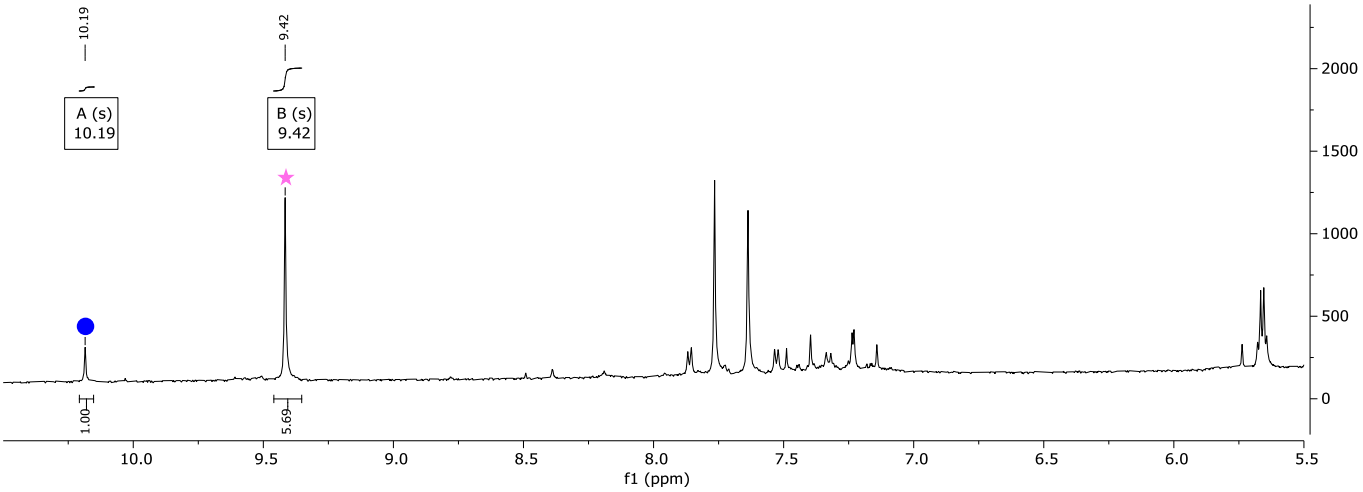
iv) *TA4C* ***2.4*** (t = 16 h)



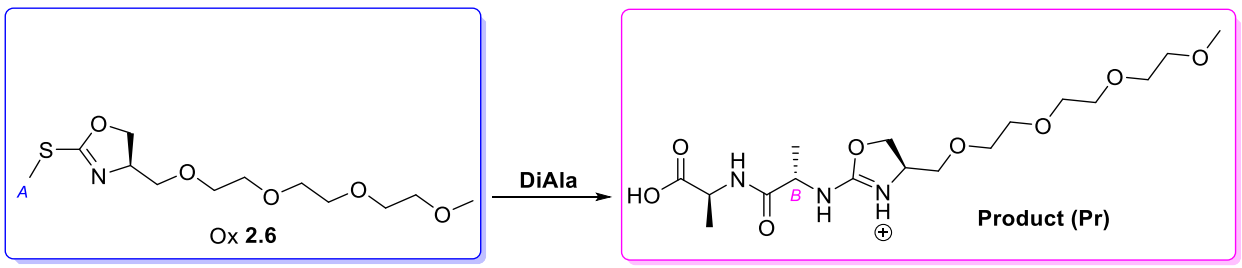


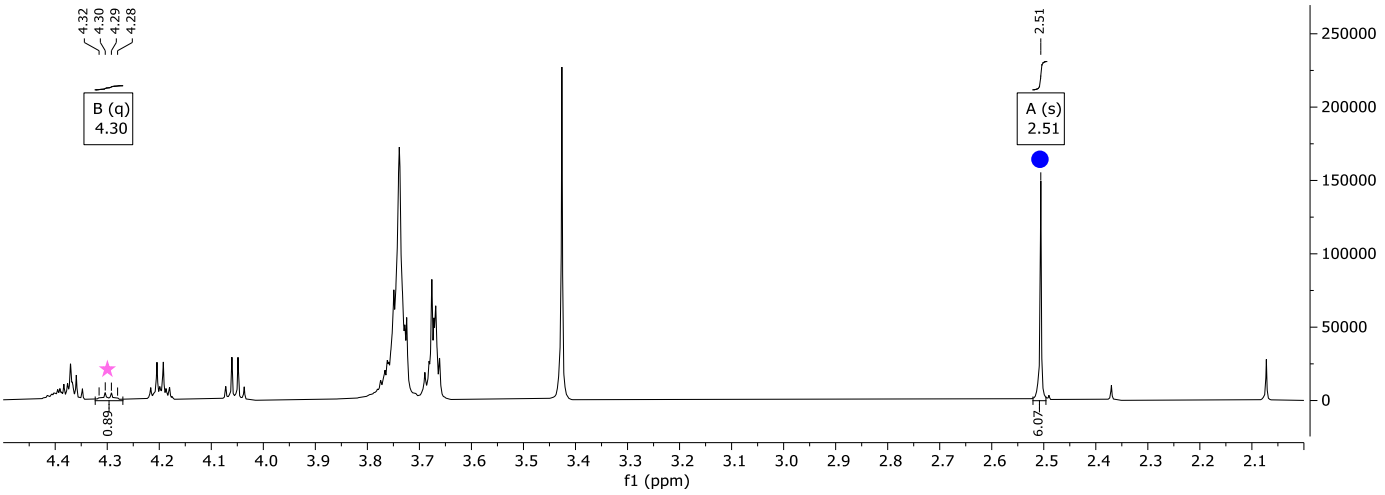
v) *2-EBA* ***2.5*** (t = 16 h)





vi) *Ox* ***2.6*** (t = 16 h)





vii) *BA* ***2.7*** (t = 1 h)

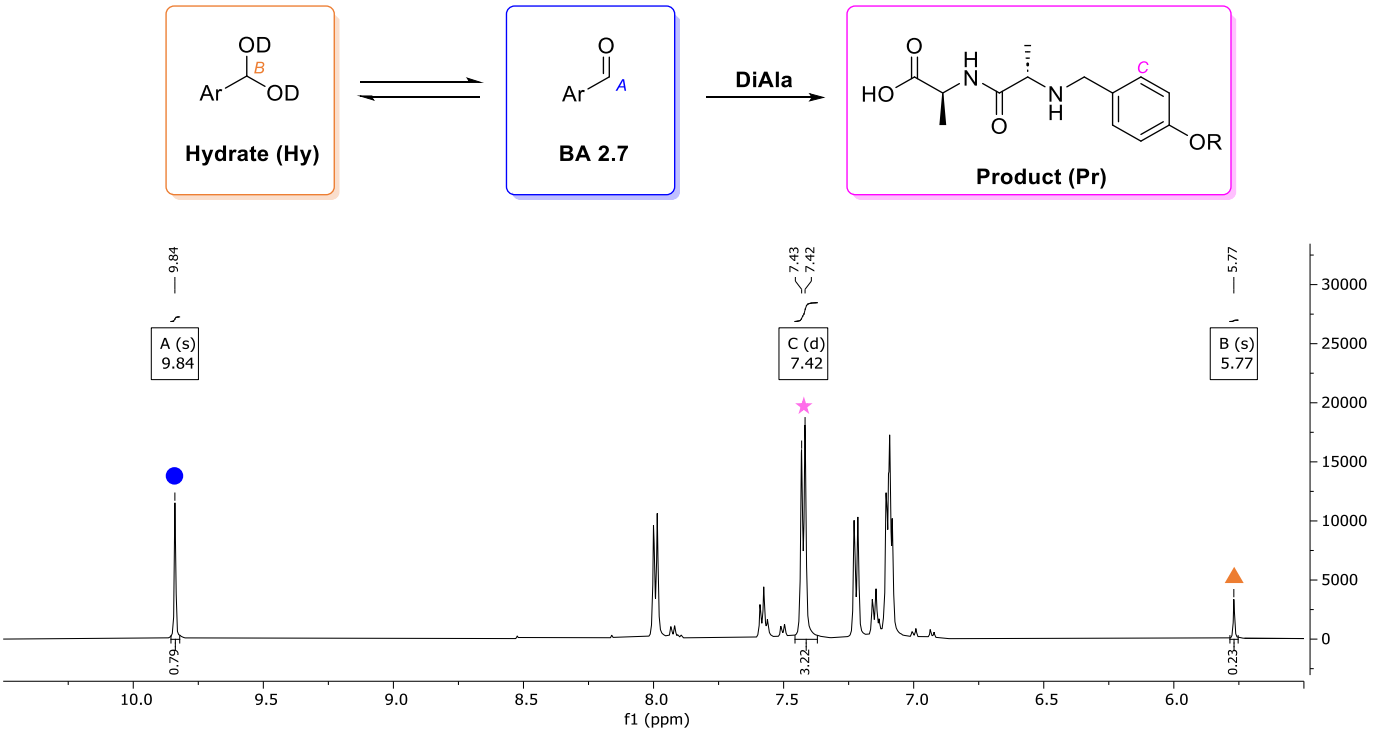


Figure S2.11. Representative ^{1}H NMR spectra of kinetics experiments showing the diagnostic signals used for kinetic modelling (**circle** = modification reagent, **star** = conjugate; for aldehyde reagents: **triangle** = hydrate species, **circle** = aldehyde species, **square** = imine species).

2.5 References

1. Macdonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. *Nat. Chem. Biol.* **11**, 326–331 (2015).
2. Koo, B.; Dolan, N. S.; Wucherer, K.; Munch, H. K.; Francis, M. B. *Biomacromolecules* **20**, 3933–3939 (2019).
3. Onoda, A.; Inoue, N.; Sumiyoshi, E.; Hayashi, T. *ChemBioChem* **21**, 1274–1278 (2020).
4. Deng, J.-R.; Lai, N. C.-H.; Kung, K. K.-Y.; Yang, B.; Chung, S.-F.; Leung, A. S.-L.; Choi, M.-C.; Leung, Y.-C.; Wong, M.-K. *Commun. Chem.* **3**, 67 (2020).
5. Chen, D.; Disotuar, M. M.; Xiong, X.; Wang, Y.; Chou, D. H.-C. *Chem. Sci.* **8**, 2717–2722 (2017).
6. Tang, K. C.; Raj, M. *Angew. Chem., Int. Ed.* **60**, 1797–1805 (2021).
7. Kohmoto, S.; Mori, E.; Kishikawa, K. *J. Am. Chem. Soc.* **129**, 13364–13365 (2007).
8. Bala, V.; Jangir, S.; Mandalapu, D.; Gupta, S.; Chhonker, Y. S.; Lal, N.; Kushwaha, B.; Chandasana, H.; Krishna, S.; Rawat, K.; Maikhuri, J. P.; Bhatta, R. S.; Siddiqi, M. I.; Tripathi, R.; Gupta, G.; Sharma, V. L. *Bioorg. Med. Chem. Lett.* **25**, 881–886 (2015).
9. Wang, F.; Tan, X.; Lv, H.; Zhang, X. *Chem. - Asian J.* **11**, 2103–2106 (2016).
10. Sankar, M.; Nowicka, E.; Carter, E.; Murphy, D. M.; Knight, D. W.; Bethell, D.; Hutchings, G. J. *Nat. Commun.* **5**, 3332 (2014).
11. Bartl, J.; Reinke, L.; Koch, M.; Kubik, S. *Chem. Commun.* **56**, 10457–10460 (2020).
12. Fletcher, J. T.; Christensen, J. A.; Villa, E. M. *Tetrahedron Lett.* **58**, 4450–4454 (2017).
13. Göring, S.; Taymans, J. M.; Baekelandt, V.; Schmidt, B. *Bioorg. Med. Chem. Lett.* **24**, 4630–4637 (2014).
14. Wang, R.; Yan, F.; Qiu, D.; Jeong, J.-S.; Jin, Q.; Kim, T.-Y.; Chen, L. *Bioconjugate Chem.* **23**, 705–713 (2012).
15. Chatani, E.; Hayashi, R.; Moriyama, H.; Ueki, T. *Protein Sci.* **11**, 72–81 (2002).
16. Raj, M.; Wu, H.; Blosser, S. L.; Vittoria, M. A.; Arora, P. S. *J. Am. Chem. Soc.* **137**, 6932–6940 (2015).
17. Maurus, R.; Overall, C. M.; Bogumil, R.; Luo, Y.; Mauk, A. G.; Smith, M.; Brayer, G. D. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1341**, 1–13 (1997).
18. González-Páez, G. E.; Roncase, E. J.; Wolan, D. W. *Acta Crystallogr., Sect. D: Struct. Biol.* **75**, 325–332 (2019).
19. Grabarczyk, D. B.; Chappell, P. E.; Eisel, B.; Johnson, S.; Lea, S. M.; Berks, B. C. *J. Biol. Chem.* **290**, 9209–9221 (2015).
20. Spicer, C. D.; Davis, B. G. *Chem. Commun.* **47**, 1698–1700 (2011).
21. Branch, J.; Rajagopal, B. S.; Paradisi, A.; Yates, N.; Lindley, P. J.; Smith, J.; Hollingsworth, K.; Turnbull, W. B.; Henrissat, B.; Parkin, A.; Berry, A.; Hemsworth, G. R. *Biochem. J.* **478**, 2927–2944 (2021).
22. Barber, L. J.; Yates, N. D. J.; Fascione, M. A.; Parkin, A.; Hemsworth, G. R.; Genever, P. G.; Spicer, C. D. *RSC Chem. Biol.* **4**, 56–64 (2023).
23. Dirksen, A.; Dirksen, S.; Hackeng, T. M.; Dawson, P. E. *J. Am. Chem. Soc.* **128**, 15602–15603 (2006).
24. Spicer, C. D.; Davis, B. G. *Nat. Commun.* **5**, 4740 (2014).
25. Zaia, J.; Annan, R. S.; Biemann, K. *Rapid Commun. Mass Spectrom.* **6**, 32–36 (1992).

26. Denißen, M.; Hannen, R.; Itskalov, D.; Biesen, L.; Nirmalananthan-Budau, N.; Hoffman, K.; Reiss, G. J.; Resch-Genger, U.; Müller, T. J. J. *Chem. Commun.* **56**, 7407–7410 (2020).
27. Erbas-Cakmak, S.; Akkaya, E. U. *Angew. Chem., Int. Ed.* **52**, 11364–11368 (2013).
28. Covington, A. K.; Paabo, M.; Robinson, R. A.; Bates, R. G. *Anal. Chem.* **40**, 700–706 (1968).

Chapter 3

Strategies to improve N-terminal targeting

Chapter 3: Strategies to improve N-terminal targeting

Note: The experimental work of section 3.3.3 was conducted in collaboration with Dr. Laetitia Raynal, a former Spicer group member. Reagents designed and synthesised by Dr. Laetitia Raynal are indicated in relevant text and figures.

3.1 Side chain participation

3.1.1 Introduction

Cyclised conjugates can be formed by initial reaction at the N-terminus with an aldehyde reagent, followed by nucleophilic attack of the N-terminal side chain at the intermediate imine. This powerful method for site-selective modification has been typically employed for the modification of N-terminal cysteine (*Fig. 3.1a*),[1] tryptophan (*Fig. 3.1b*),[2] serine, and threonine residues (*Fig. 3.1c*).[3,4] However, it should be noted that the key limitation of this approach is the loss of generalisability, requiring the presence, or more commonly genetic instalment, of a specific N-terminal residue with the desired reactivity.

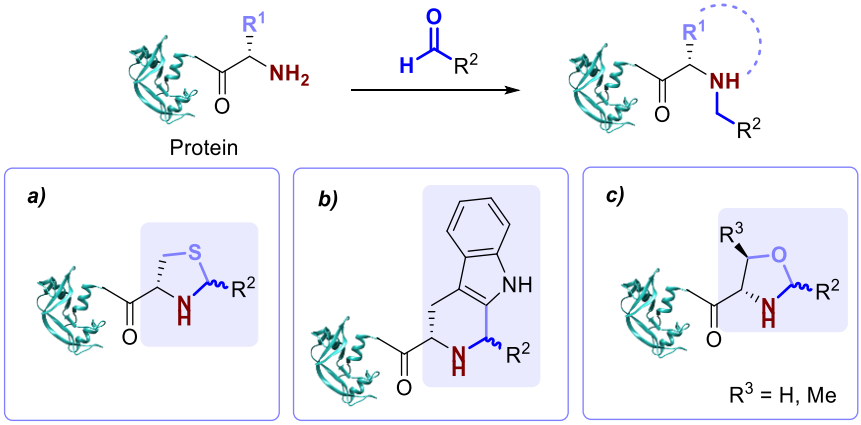


Figure 3.1. Previously reported strategies for targeting protein N-termini, involving condensation at the N-terminus *via* reaction with an aldehyde, followed by cyclisation with ***a)*** cysteine;[1] ***b)*** tryptophan;[2] or ***c)*** serine/threonine residues.[3,4] Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

In this context, we decided to probe the kinetics and stability of 2-pyridinecarboxaldehyde (2-PCA) modifications of peptides with an N-terminal asparagine (Asn) residue. As introduced in Chapter 2, 2-PCAs developed by the Francis group initially undergo a condensation reaction with the N-terminal amine to form an imine intermediate. Cyclisation occurs through nucleophilic attack of the α-amide nitrogen, resulting in imidazolidinone formation.[6] For proteins/peptides with an N-terminal Asn residue, we hypothesised that an alternative cyclisation route may be accessible *via* nucleophilic attack of the primary β-amide of the Asn residue. This would form a 6-membered hexahydropyrimidine rather than 5-membered imidazolidine; we expected the hexahydropyrimidine to be more stable due to reduced ring strain, with faster kinetics for primary amide *vs.* secondary amide attack and increased nucleophilicity (*Fig. 3.2*). Baldwin's rules support our hypothesis, whereby *6-endo-trig* reactions (i.e., hexahydropyrimidine formation) are more favoured than *5-*

endo-trig reactions (i.e., imidazolidinone formation) due to more effective orbital overlap for the 6-membered ring formation.[7]

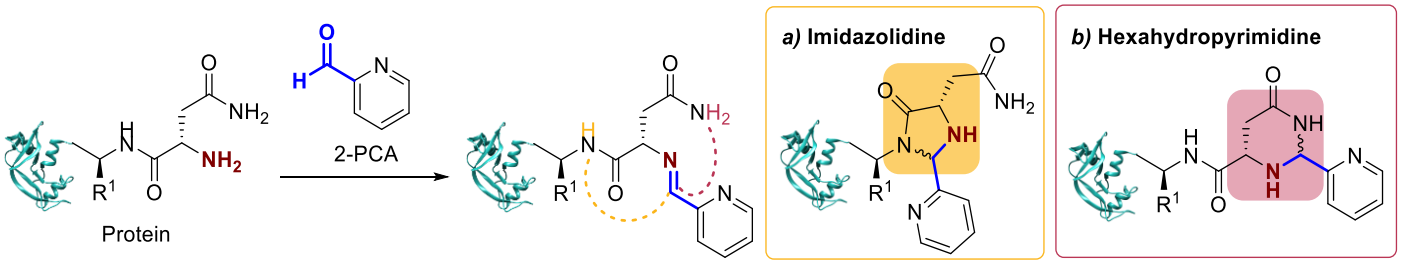
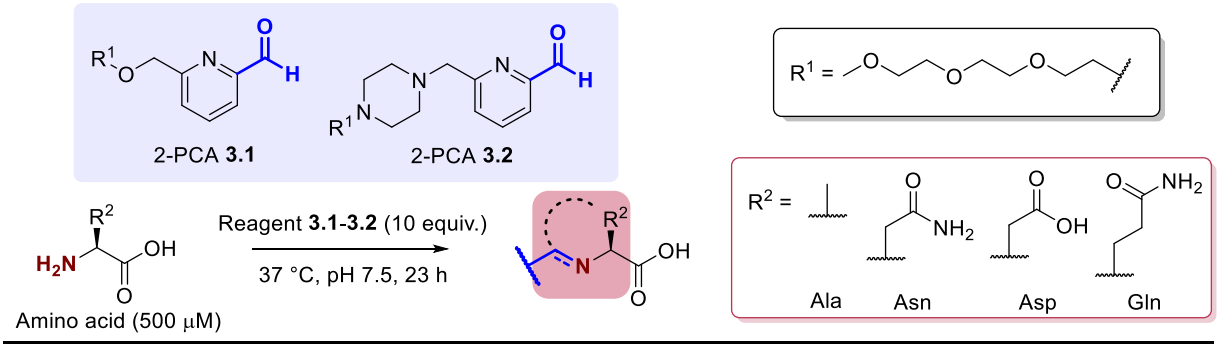


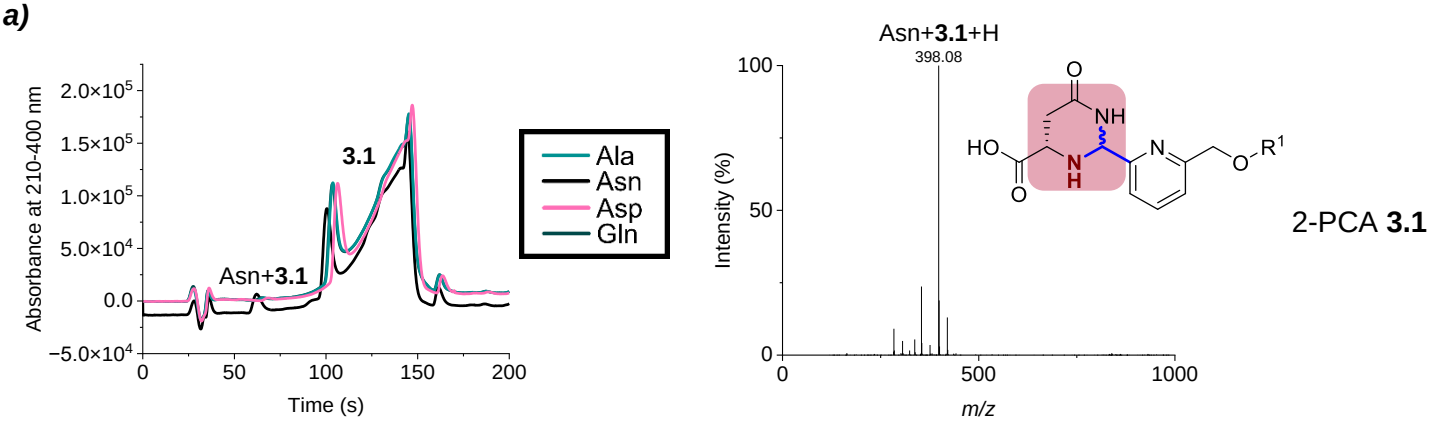
Figure 3.2. Side chain participation of the N-terminal Asparagine (Asn) residue in the cyclisation of PCA-based conjugates, to form ***a)*** an imidazolidinone *via* nucleophilic attack of the α-amide; or ***b)*** a hexahydropyrimidine *via* nucleophilic attack of the β-amide. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

3.1.2 Initial proof of concept

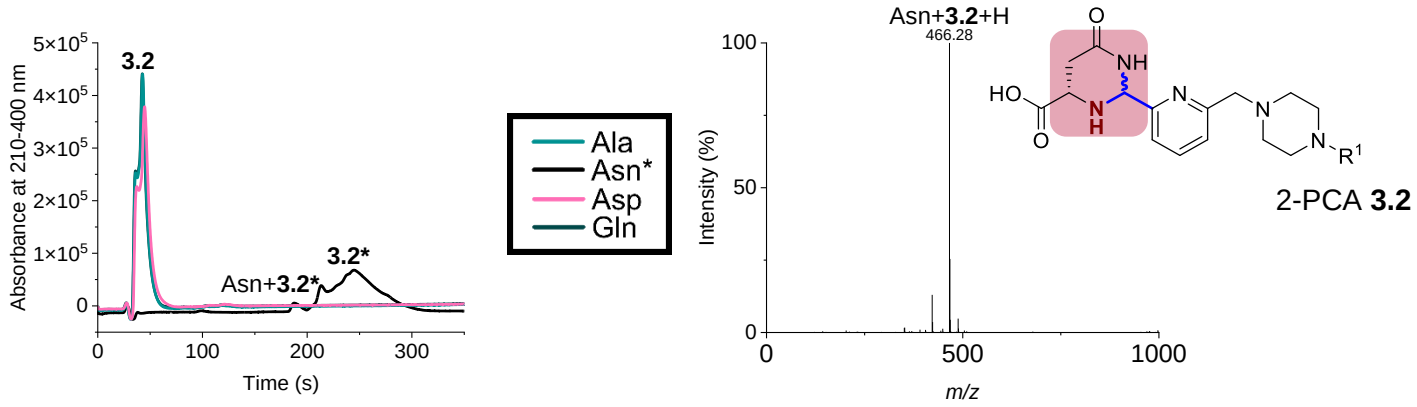
The reactivity of the Asn side chain was initially validated by undertaking modification of single amino acids NH_2-Asn-OH, aspartic acid (Asp), glutamine (Gln) and alanine (Ala) with reagents **3.1**-**3.2** (*Fig. 3.3*). Single amino acids were used rather than peptides due to the absence of the adjacent α-amide required for imidazolidinone formation, allowing any stable mass addition to be attributed to cyclisation with the β-amide of the Asn side chain. Reactions were analysed *via* LC-MS, and qualitative conversions were estimated based on the assumption that the absorption of reagents **3.1**-**3.2** at 210-400 nm was unaffected by conjugation. Note this was a poor assumption due to expected changes in aromatic conjugation, and used only as a comparative measurement due to the minimal changes in UV absorbance during the attempted kinetics experiments detailed in Chapter 2. NH_2-Asp-OH, NH_2-Gln-OH and NH_2-Ala-OH were also included in the study as examples of amino acids for which we expected transient imine formation but no cyclisation. As expected, NH_2-Asn-OH was the only amino acid for which modification was observed, indicative of a modification of higher stability than imine formation alone.



a)



MS (ESI$^+$) [Asn+**3.1**+H]$^+$ found 398.08, calculated 398.19.



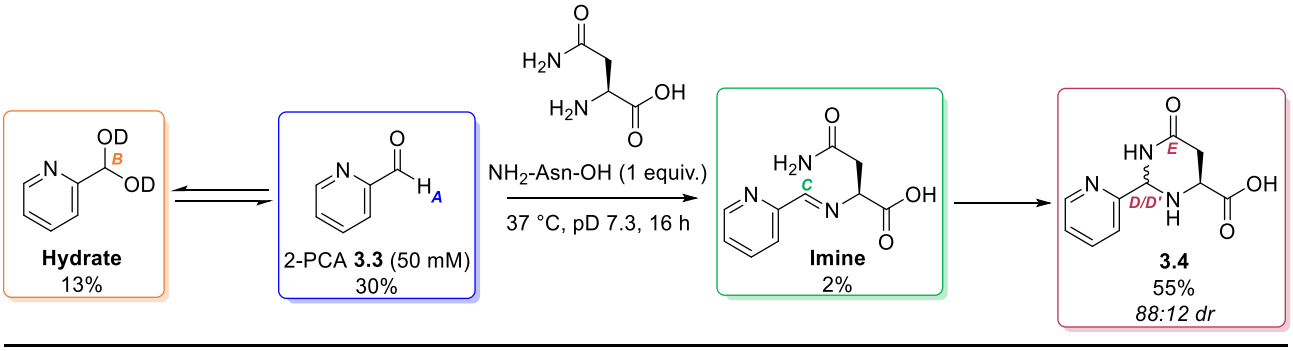
MS (ESI$^+$) [Asn+**3.2**+H]$^+$ found 466.28, calculated 466.27.

b)

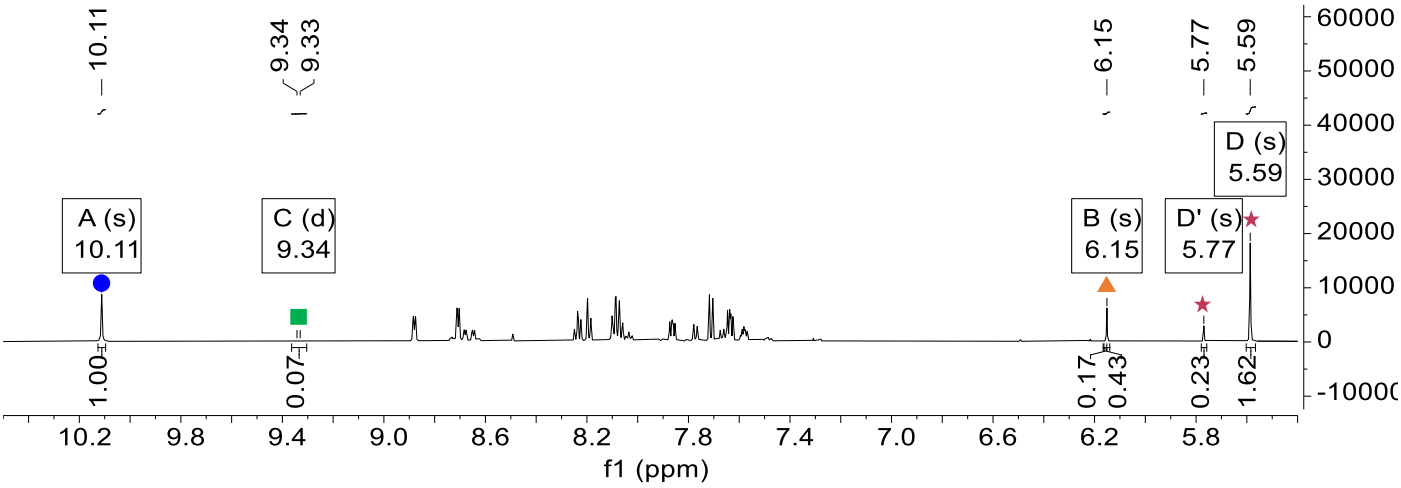
Reagent	Relative conversion (%)			
	Asn	Asp	Gln	Ala
2-PCA **3.1**	**25**	**0**	**0**	**0**
2-PCA **3.2**	**36**	**0**	**0**	**0**

Figure 3.3. Initial proof of concept for Asn reactivity. Modification of amino acids NH_2-Asn-OH, NH_2-Asp-OH, NH_2-Gln-OH and NH_2-Ala-OH with compounds **3.1**-**3.2**. ***a)*** UV-vis 210-400 nm trace and conjugate mass spectra for the modification of amino acids (Ala, Asn, Asp and Gln) with 2-PCAs **3.1**-**3.2**. **The Asn/PCA **3.2** reaction mixture was analysed with a slower column gradient than all other samples due to co-elution between the conjugate and excess reagent **3.2**.* ***b)*** Conversions reported are qualitative, calculated from relative integrals of the modified amino acid and excess unmodified PCA in the 210-400 nm UV absorbance trace.

To support our preliminary LC-MS study, NH_2-Asn-OH (50 mM) was reacted with 2-PCA **3.3** (50 mM) at 37 °C in deuterated pD 7.3 sodium phosphate buffer, and the reaction mixture was characterised by ^{1}H and ^{13}C NMR spectroscopy (*Fig. 3.4*). The formation of diastereomers (88:12 dr, *Fig. 4a*), and HMBC coupling between the $\underline{H}_{D'}$ ^{1}H NMR signal and $\underline{C}_{E}$ ^{13}C NMR signal (*Fig. 4b*), both provided further support for the formation of hexahydropyrimidine **3.4**.



a)



b)

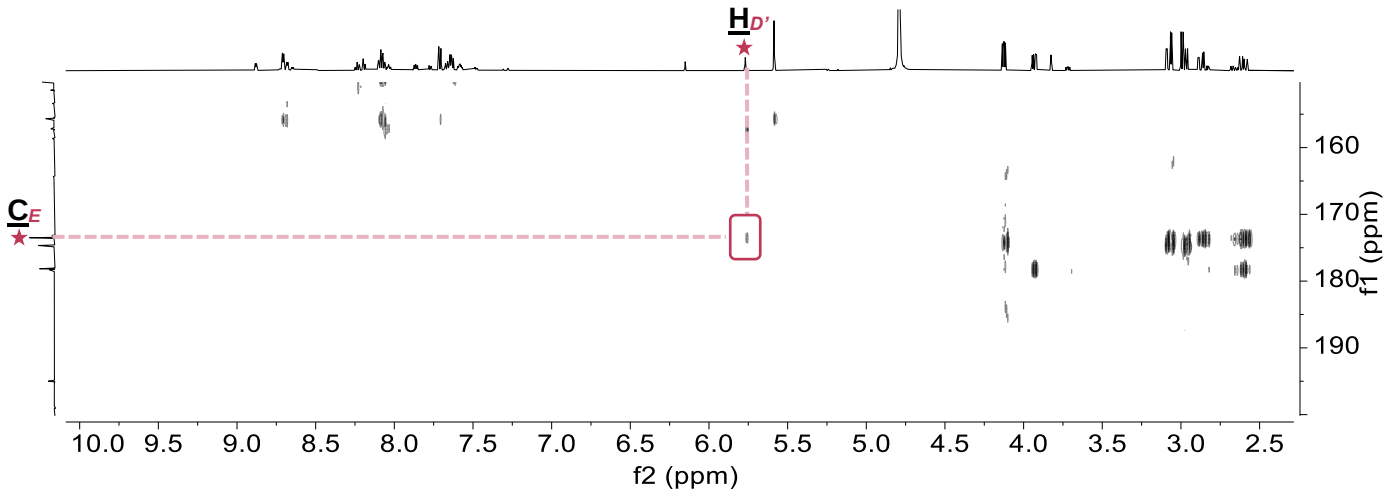


Figure 3.4. NMR spectra upon reaction of Asn with 2-PCA **3.3**, showing ***a)*** the diagnostic ^{1}H NMR signals used to calculate conversions and the diastereomeric ratio (dr); and ***b)*** HMBC coupling indicative of cyclisation to hexahydropyrimidine **3.4** (**triangle** = hydrate species, **circle** = aldehyde species, **square** = imine species, **star** = hexahydropyrimidine species).

3.1.3 Kinetics and stability of hexahydropyrimidines and imidazolidinones

Having demonstrated the ability of H_2N-Asn-OH to form hexahydropyrimidines, we next sought to compare the kinetics and stability of hexahydropyrimidine formation to imidazolidinone formation. We synthesised dipeptide Asn-Ala **3.5**, capable of nucleophilic attack of both the primary β-amide of the Asn residue, and the α-amide. Boc-Asn-OH was coupled with *tert*-butyl L-alaninate hydrochloride to afford dipeptide **3.6** in 31% yield, followed by Boc-deprotection to afford dipeptide **3.5** in 89% yield (*Fig. 3.5*). Note that the proceeding conjugation NMR experiments were conducted based on the assumption that the concentration of the buffer (100 mM) was sufficient to neutralise the TFA salt (50 mM), as supported by measurement of neutral pH of the reaction mixture at the final timepoint.

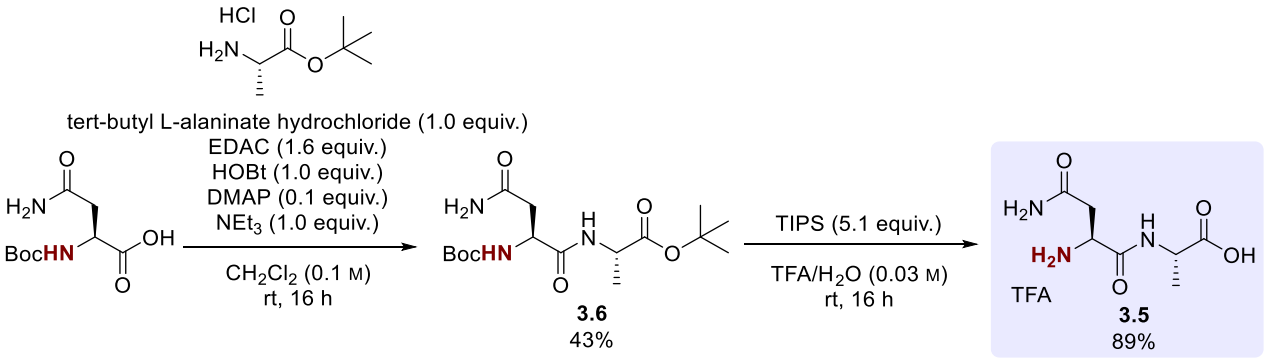


Figure 3.5. Synthesis of dipeptide **3.5**.

To determine the imidazolidinone to hexahydropyrimidine ratio upon conjugation, Asn-Ala **3.5** (50 mM) was reacted with reagent **3.3** (50 mM) under second-order reaction conditions at 37 °C in deuterated pD 7.3 sodium phosphate buffer, and conjugate formation was followed over time by ^{1}H NMR spectroscopy. Diagnostic product signals of imidazolidinone **3.7** and hexahydropyrimidine **3.8** were assigned by comparison to the relative diastereomeric ratios of analogous conjugates of DiAla (imidazolidinone **3.9**) and H_2N-Asn-OH (hexahydropyrimidine **3.4**), capable of only imidazolidinone or hexahydropyrimidine formation respectively (*Fig. 3.6*).

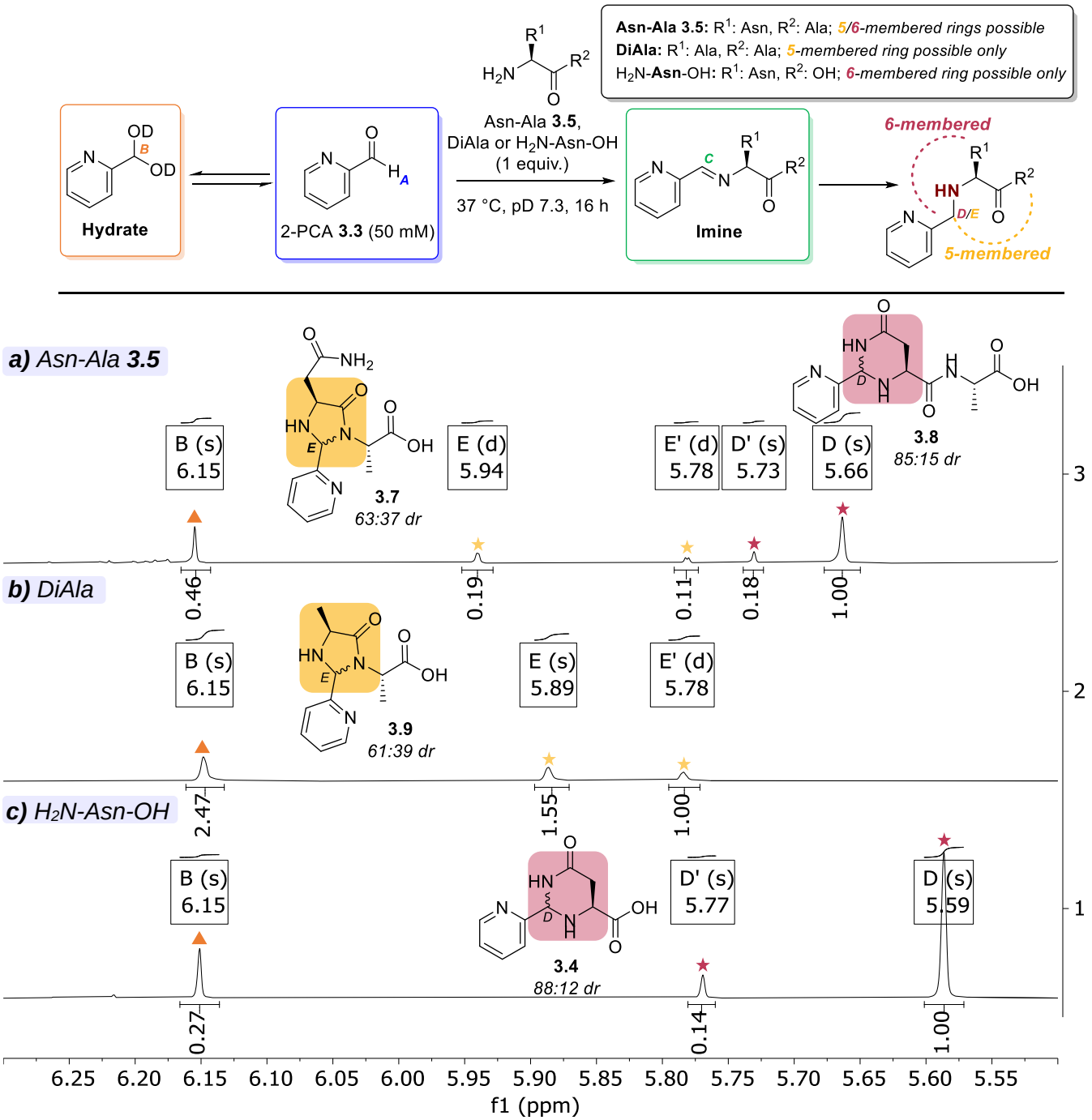
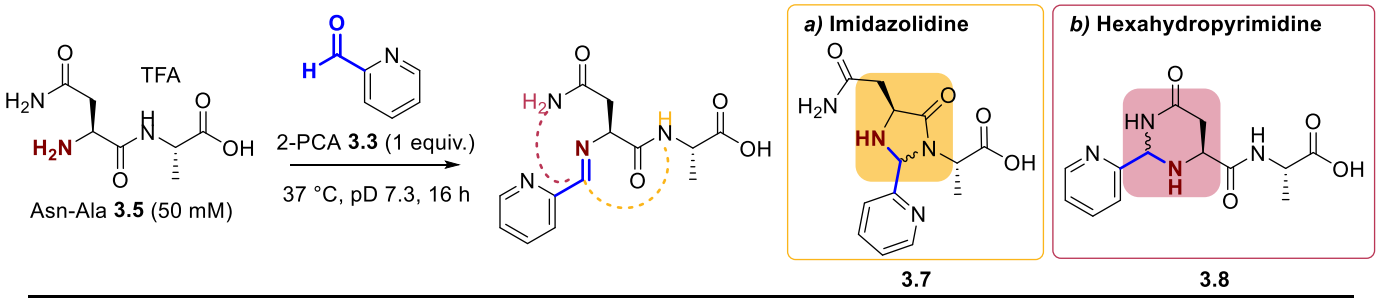
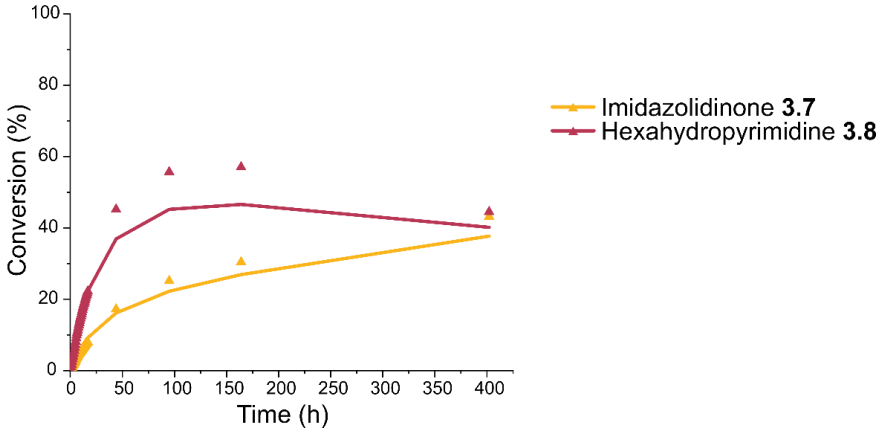


Figure 3.6. ^{1}H NMR spectra upon reaction of Asn-Ala **3.5**, DiAla and H_2N-Asn-OH with 2-PCA **3.3**, showing the diagnostic ^{1}H NMR signals used to calculate the diastereomeric ratio (dr; **triangle** = hydrate species, **star** = imidazolidinone species, **star** = hexahydropyrimidine species).

With diagnostic signals identified, data were then fit to a reversible second-order kinetic model encompassing both reactions, allowing us to calculate forward and backward rate constants, k_1 and k_{-1} respectively, as well as the dissociation constant, K_d, for each reaction (*Fig. 3.7*). Note that the kinetic model did not integrate hydrate and imine formation, and made the assumption that cyclisation was the rate-limiting step of the reaction: more detailed analysis will be discussed later in Chapter 4. The 6-membered ring (**3.8**) was found

to be the kinetic product, with formation favoured initially over the formation of the 5-membered ring (**3.7**) (k_1: 3.99 × 10^{-1} M^{-1} h^{-1} and 1.60 × 10^{-1} M^{-1} h^{-1} for hexahydropyrimidine and imidazolidinone formation respectively). Under these conditions, the data fit an irreversible model for imidazolidinone **3.7** formation, though this does not preclude reversibility at levels below our threshold for detection (k_{-1} < 10^{-6} h^{-1}); poor fitting of the data to the model also leads to a lack of certainty within this observation. In contrast, for hexahydropyrimidine **3.8** formation (K_d ~ 8 mM), significant levels of reversibility were observed. Over an extended time-period of 2 weeks, the 6-membered ring was seen to gradually convert to the more thermodynamically stable imidazolidinone product. In both cases, though the data did not closely fit the model, it was supportive of the hexahydropyrimidine being the kinetically favoured product, with the imidazolidine being the thermodynamic end-product given sufficient time.





Cyclisation route	k_1 / M^{-1} h^{-1}	k_{-1} / h^{-1}	**K_d / mM**
a) Imidazolidinone **3.7**	1.60 × 10^{-1} ± 0.05 × 10^{-1}	*	-
b) Hexahydropyrimidine **3.8**	3.99 × 10^{-1} ± 0.12 × 10^{-1}	3.20 × 10^{-3} ± 0.51 × 10^{-3}	8.0

Figure 3.7. Plots of conversion against time for the reaction of Asn-Ala **3.5** with reagent **3.3** to form imidazolidinone (**3.7**) or hexahydropyrimidine (**3.8**) conjugates under second order conditions at a concentration of 50 mM, as measured by ^{1}H NMR analysis. Fits are based on a second order reversible model encompassing both imidazolidinone and hexahydropyrimidine formation. **Fitted* k_{-1} *<*10^{-6} h^{-1}.

Our observations were supported by DFT calculations carried out by Dr. Ksenia Stankevich, a PDRA in the Spicer group. Computations predicted *i)* the 6-membered ring was the *kinetic* product, with a Gibbs free

energy change (ΔG) for the lowest-energy transition state 8.2 kcal/mol lower than for the 5-membered ring; and *ii)* the 5-membered ring was the *thermodynamic* product, with ΔG for the lowest-energy cyclised conjugate 4.7 kcal/mol lower than for the 6-membered ring.

3.1.4 Application to protein modification

Our previous work highlighted the challenges faced when translating results from small molecule and peptide models to protein substrates (Chapter 2.3). With this in mind, we next planned to validate our results *via* comparison of the kinetics and stability of an N-terminally modified protein with an analogue bearing a terminal Asn residue. However, as will be discussed further in Chapter 5, attempts to prepare an N-terminal Asn mutant of the protein sonic hedgehog (Shh C24N) by Dr. Helen Bell at our industrial collaborators Qkine were unsuccessful due to incomplete cleavage of the N-terminal methionine residue.

Translation of almost all proteins in *E. coli* is initiated at an AUG codon with the installation of *N*-formylmethionine (*Fig. 3.8a*). This N-terminal residue becomes exposed to enzymes early during translation before protein folding, upon exit of the first residues of the newly synthesised polypeptide chain from the ribosomal exit tunnel.[8] In prokaryotes, the *N*-formyl group is cleaved by a deformylase enzyme (*Fig. 3.8b*), followed by the removal of the N-terminal methionine residue by a methionine aminopeptidase (*Fig. 3.8c*) to expose the penultimate amino acid. Penultimate amino acids with large sidechains with a radius of gyration >1.29 Å (e.g. Asn) can block the methionine aminopeptidase activity through steric hindrance:[9] it is common for no cleavage to occur for penultimate Asn residues.[10] The extremely low abundance of naturally occurring exposed N-terminal Asn residues therefore limits the applicability of an approach exploiting Asn N-termini. Both genetic instalment of Asn, and the use of an engineered *E. coli* methionine aminopeptidase,[11] or an alternative protease cleavage strategy, would be required to expose Asn as the N-terminal residue. Similar issues limited the applicability of Pictet-Spengler reactions for targeting N-terminal tryptophans, developed by Li *et al.*, to proteolytically cleaved protein fragments or synthetic peptides.[2]

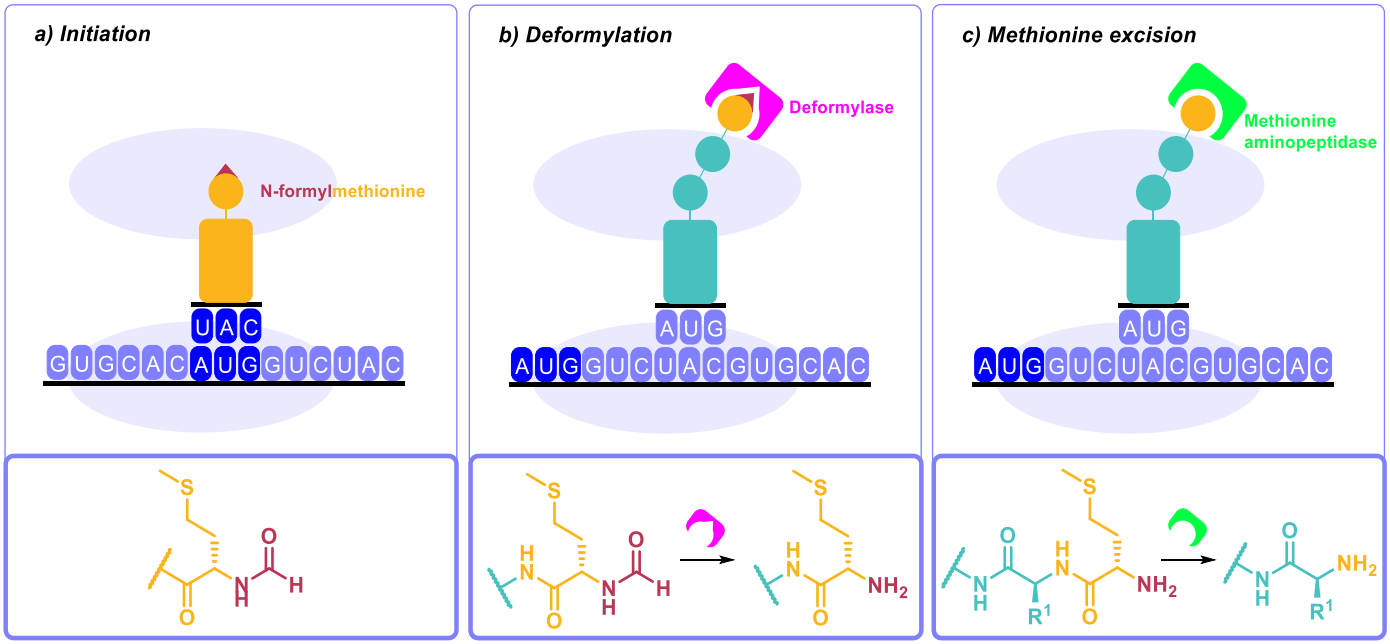


Figure 3.8. N-formylmethionine in initiation of protein translation (***a***), deformylation (***b***), and methionine excision (***c***) to expose the penultimate residue at the N-terminus.

3.1.5 Conclusion

We have demonstrated that peptide modification with 2-PCAs can be accelerated by side chain participation of an N-terminal Asn residue, at the expense of conjugate stability. For applications in which long-term conjugate stability is not required, this could be a useful strategy to ensure high levels of conversion, albeit with additional genetic engineering steps required to incorporate and expose a N-terminal Asn residue. Due to difficulties in preparing mutant proteins with an exposed N-terminal Asn, this has not been further explored.

3.2 N-methylated termini

3.2.1 Introduction

Incomplete Met cleavage and the resulting inaccessibility of proteins with an N-terminal asparagine residue prompted us to consider the impact of co-/post-translational modifications (PTMs) on the reactivity of protein N-termini in the eukaryotic proteome. Whilst most bacterial and secreted proteins have a free N-terminus,[3] a large proportion of the eukaryotic proteome is co/post-translationally modified at the N-terminus through acetylation/propionylation *(Fig. 3.9a*), myrisotylation/palmitoylation (*Fig. 3.9b*), ubiquitylation, or mono-/di-/tri-methylation (*Fig. 3.9c*) by processing enzymes.[9] For example, during methylation, the transfer of a methyl group from *S*-adenosyl-methionine (SAM) is catalysed by protein N-terminal methyltransferases (NTMTs) to the α-amine of protein N-termini. During the methyl transfer, SAM is converted into *S*-adenosylhomocysteine (SAH).[9,12,13]

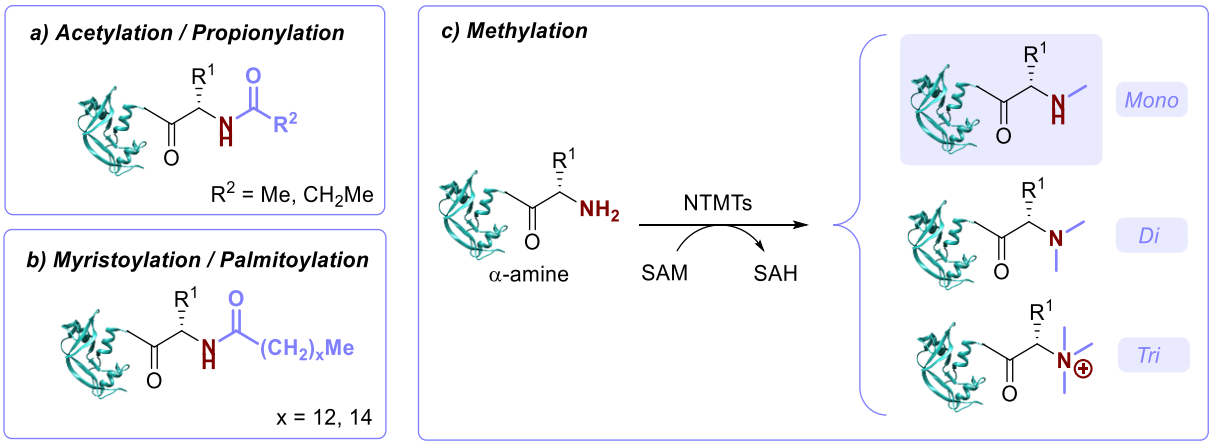


Figure 3.9. Common N-terminal co-/post-translational modifications of eukaryotic proteins. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

N-terminal PTMs play a key role in protein regulation and signalling:[9] for example, N-terminal methylation is important for the production of antibiotics, mobilisation of protein complexes, and protein-protein recognition,[12,13] and the role of N-terminal palmitoylation in Sonic Hedgehog signalling will be discussed later in Chapter 5. Synthetic addition of N-methyl groups to peptide N-termini is also of widespread interest in the development of therapeutics due to the influence on pharmacokinetic properties and applications in drug delivery.[12] Note that lysine di-methylation is also a key PTM, and chemical modification strategies targeting dimethyl amines of lysine residues have been developed recently by Raj *et al.* They exploited the tendency of tertiary amines to form iminium ions under oxidative conditions, followed by hydrolysis to aldehydes which could then be coupled with amines to form ***a)*** oxime, ***b)*** hydrazone, or ***c)*** thiazolidine conjugates (*Fig. 3.10*).[14] However, despite the high prevalence and utility of N-terminally methylated peptides and proteins, to our knowledge no methods for the chemical modification of α-methylated protein N-termini have been developed: in doing so, we hoped to expand the scope of proteins suitable for N-terminal targeting chemistries.

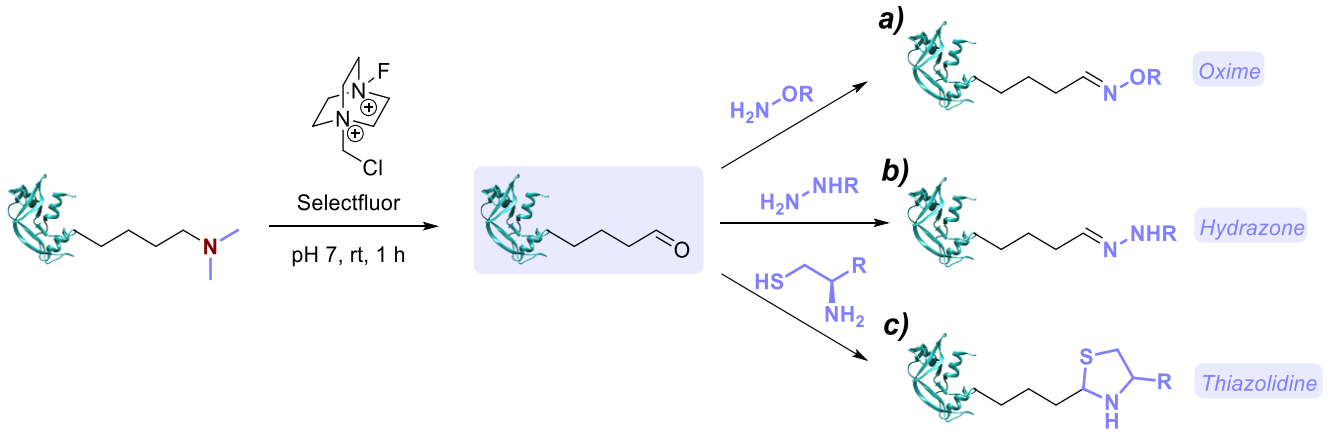


Figure 3.10. Reported strategy for targeting dimethyl lysines, involving aldehyde formation *via* oxidation and hydrolysis, followed by coupling with amines to form ***a)*** oxime, ***b)*** hydrazone, or ***c)*** thiazolidine conjugates.[14] Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

We expected the N-terminal amide of the PTMs outlined in *Fig. 3.9a* and *3.9b* to have reduced nucleophilicity in comparison to the unmodified N-terminal α-amine due to resonance stabilisation provided by the carbonyl group, whilst the impact of N-terminal methylation on the reactivity of the α-amine would be influenced by the degree of methylation (*Fig. 3.9c*). For example, tri-methylation results in a permanent positive charge on the N-terminal α-amine, and thus no nucleophilicity.[9] Contrastingly, we expected mono-methylation to result in increased electron density at the N-terminal α-amine (as compared to primary α-amines), increasing the nucleophilicity of the N-terminal amine and thus increasing the kinetics of chemical modifications with small molecule reagents. Indeed, Chen *et al.* observed increased nucleophilic reactivity of protein secondary N-alkyl α-amines *vs.* primary α-amines.[15] Not only would the application of N-terminal modification strategies outlined in Chapter 2 to N-terminally methylated proteins provide a way to selectively modify proteins with mono-methylated N-termini, enhanced nucleophilicity of the α-amine would also allow incubation of target proteins at lower temperatures/for shorter times for chemical modification, and thus enable the modification of low stability proteins.

3.2.2 Kinetics of primary vs secondary amine modification

Firstly, we decided to compare the kinetics of N-terminal peptide modifications between a primary and secondary N-terminal α-amine. We synthesised N-terminally methylated DiAla **3.10** as a simple example of a mono-methylated peptide. Boc-*N*-Me-Ala-OH was coupled with *tert*-butyl L-alaninate hydrochloride to afford dipeptide **3.11** in 48% yield. Upon isolation of dipeptide **3.11**, the N-terminal *Boc* group was unexpectedly cleaved. No by-products due to coupling at the N-terminus of N-Me-Ala-OH were observed, indicating that *Boc*-deprotection may have occurred during acidic workup, rather than during the reaction (*Fig. 3.11*). Deprotection of the C-terminal *tert*-butyl group of dipeptide **3.11** was next attempted with: *i)* TFA (10 equiv.), however the protecting group was not removed after 1.5 h at rt (monitored by ^{1}H NMR). The deprotection

was re-attempted with: *ii)* TFA (30 equiv.) in CH_2Cl_2 at rt for 142 h; or *iii)* HCl (10 equiv.) in THF at 60 °C for 4 h to afford compound **3.10** as TFA/HCl salts respectively (*Fig 3.10*). Note that whilst azeotroping with CH_2Cl_2 was insufficient for full HCl removal, the proceeding conjugation NMR experiments were conducted based on the assumption that the concentration of the buffer (100 mM) was sufficient to neutralise the excess HCl, as supported by measurement of neutral pH of the reaction mixture at the final timepoint.

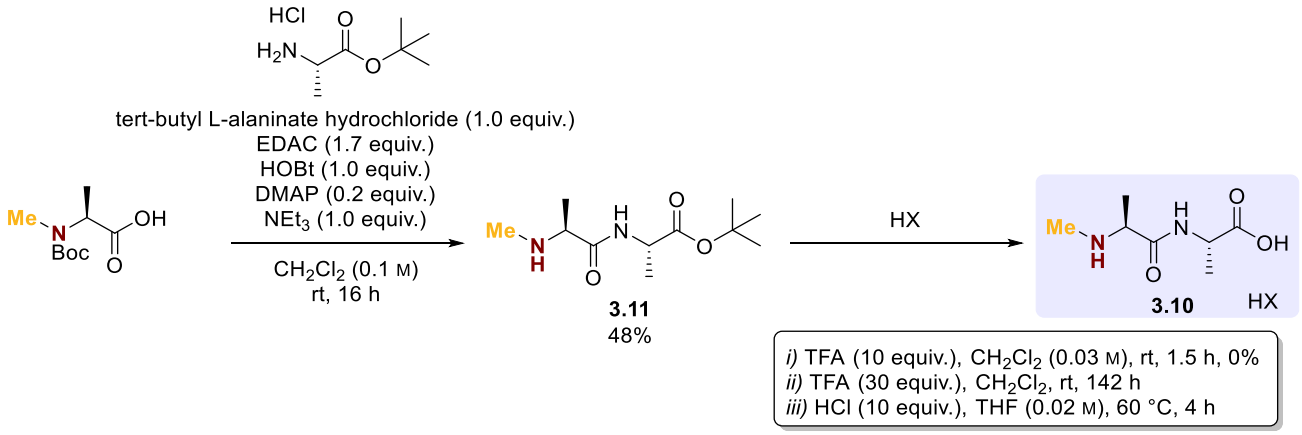
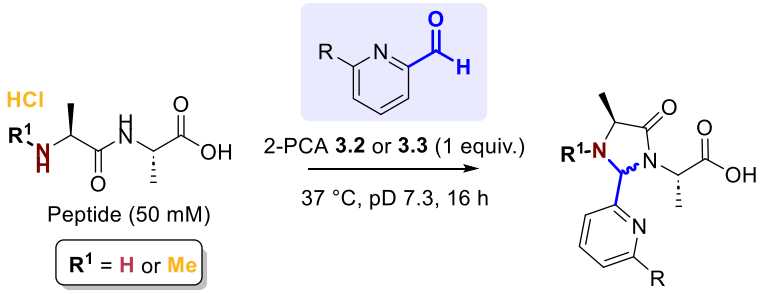
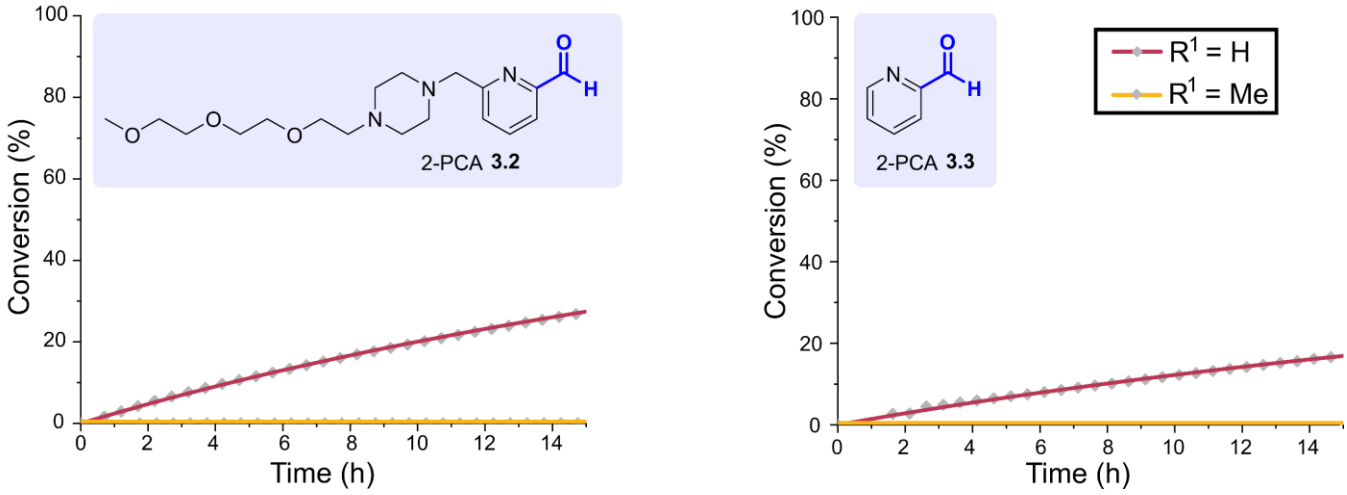


Figure 3.11. Synthesis of dipeptide **3.10**.

To determine the impact of N-terminal mono-methylation on the reactivity of the α-amine, DiAla and dipeptide **3.10** (50 mM) were reacted with 2-PCAs **3.2** and **3.3** (1 equiv.) under second-order reaction conditions at 37 °C in deuterated pD 7.3 sodium phosphate buffer, and conjugate formation was followed over time by ^{1}H NMR spectroscopy. Data were then fit to a reversible or irreversible second-order kinetic model, as appropriate, allowing us to calculate forward and backward rate constants, k_1 and k_{-1} respectively, as well as the dissociation constant, K_d, for each reaction where appropriate (*Fig. 3.12*). These studies revealed that reactions between the mono-methylated N-terminus of DiAla **3.10** and reagents **3.2** and **3.3** were extremely slow ($k_1 < 5 \times 10^{-3}$ M^{-1} h^{-1}) relative to the analogous DiAla primary α-amine ($k_1 > 2 \times 10^{-1}$ M^{-1} h^{-1}), with no imine formation observed; we suspected poor reactivity of the N-terminally methylated peptide to be due to a combination of increased basicity and steric hindrance. Under these conditions, the data fit an irreversible model for reaction of **3.10** with reagent **3.3**, though this does not preclude reversibility at levels below our threshold for detection ($k_{-1} < 10^{-6}$ h^{-1}). The reactivity of mono-methylated N-termini was much poorer than expected; we next set out to validate these results using a larger peptide.





2-PCA	R^1	k_1 / M^{-1} h^{-1}	k_{-1} / h^{-1}	**K_d / mM**
3.2	H	$5.5 \times 10^{-1} \pm 0.02 \times 10^{-1}$	$1.3 \times 10^{-3} \pm 0.54 \times 10^{-3}$	2.4
	Me	$4.1 \times 10^{-3} \pm 0.35 \times 10^{-3}$	*	-
3.3	H	$2.9 \times 10^{-1} \pm 0.03 \times 10^{-1}$	$9.3 \times 10^{-3} \pm 1.4 \times 10^{-3}$	31.7
	Me	-	-	-

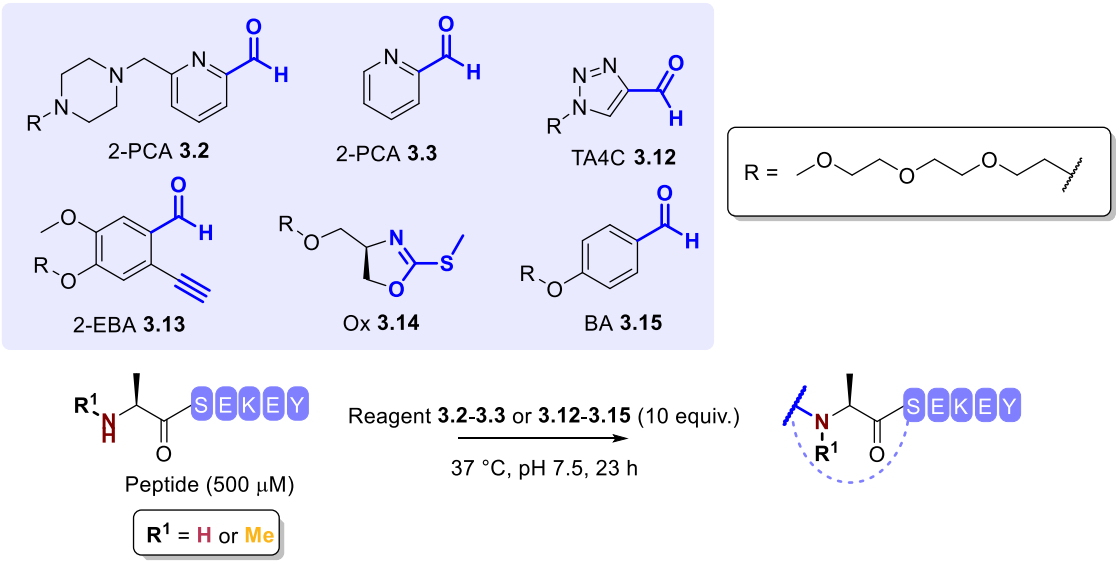
Figure 3.12. Plots of conversion against time for the reaction of DiAla and N-methylated DiAla **3.10** with reagents **3.2** or **3.3** under second order conditions at a concentration of 50 mM, as measured by ^{1}H NMR analysis. Fits are based on a second order reversible or irreversible model, as appropriate. **Fitted $k_{-1} < 10^{-6}$ h^{-1}.*

3.2.3 Conversion of primary vs secondary amine modification

In the previous ^{1}H NMR study, precipitation was observed during the modification of **3.10** with reagent **3.2**. We suspected the precipitation to be due to increased hydrophobicity from the N-terminal methyl group. Poor conversions of **3.10** may therefore have been due to loss of conjugates from solution: ^{1}H NMR analysis of the reaction mixture upon re-dissolving the precipitate revealed 24% conversion, although these higher conversion levels may have also been a result of extended reaction time (1 month at 37 °C). We next decided to follow modifications by LC-MS analysis to allow us to work at a lower peptide concentration (500 µM), at which we hoped less precipitation would occur. We prepared ASEKEY and its N$^{\alpha}$-methylated analogue by solid phase peptide synthesis. We included the tyrosine residue for detection in the UV 280 nm region so that the relative conversions could be estimated by comparison of the integrals in the 280 nm absorbance trace by LC-MS. However, this was a poor estimation due to absorbance of both the peptide and modification

reagent at 280 nm; in future studies detailed in Chapter 4, this issue was avoided *via* modification of peptides bearing unique chromophores. This study was therefore very approximate and was used only for an indication of the comparative reactivity of ASEKEY/N$^{\alpha}$-Me-ASEKEY for each reagent, based upon the assumption that N-methylation had no influence on absorbance. For some reagents (e.g., 2-EBA **3.13**), co-elution of signals with impurities from the modification reagent provided further ambiguity.

To probe a range of different structural moieties which we hoped would have different degrees of reactivity and steric interactions with the N-terminal methyl group, we screened the modification of ASEKEY and its N-methylated analogue with a panel of reagents readily available to us from previous studies in Chapter 2 (*Fig. 3.13*). It became clear that upon modification of mono-methylated peptides, there is a subtle balance between the increased nucleophilicity of the primary amine and increased steric clashes, with no general trend across all reagents. Conversion with TA4C **3.12** was increased for the methylated peptide, perhaps due to increased nucleophilicity of the N-terminal amine, whereas conversion with 2-PCA **3.3** and 2-EBA **3.13** was decreased for the methylated peptide, perhaps due to steric clashes. Note that cyclisation to an isoquinolinium with 2-EBA **3.13** was not possible for an N-terminal secondary amine, potentially also contributing to the decrease in conversion observed; the species observed may have been the uncyclised iminium ion intermediate. Reactivity of N$^{\alpha}$-Me-ASEKEY with 2-PCA **3.3** was considerably higher than seen previously in ^{1}H NMR experiments, though still notably lower than the non-methylated ASEKEY analogue, highlighting the need for multiple model systems due to influence of multiple factors on reactivity, e.g. the adjacent amino acid.



	Relative conversion (%)	
Reagent	R^1 = H	R^1 = Me
3.2	64	-
3.3	89	47
3.12	43	60
3.13	73	38
3.14	0	0
3.15	39	38

Figure 3.13. Modification of peptides ASEKEY and N^{α}-Me-ASEKEY with compounds **3.2**-**3.3** and **3.12**-**3.15**. Conversions reported are qualitative, calculated from relative integrals of the unmodified/modified peptide in the 280 nm UV absorbance trace.

3.2.4 Conclusion

As demonstrated by the ^{1}H NMR and LC-MS kinetics/conversion experiments, our aim to significantly increase the kinetics of chemical modifications by using N-terminally methylated peptides was unsuccessful. The reactivity of methylated *vs.* free N-termini of peptides with small molecule reagents varied between reagents, and illustrated a potential compromise between nucleophilicity and steric clashes. Whilst we decided not to pursue this approach any further, we have been able to demonstrate that modification of N-terminally methylated peptides is possible, so the scope of N-terminal chemical modification strategies could potentially be increased to include proteins which have been post-translationally modified by NTMTs.

3.3 Proximity-driven chemistries

3.3.1 Introduction

Having initially attempted to develop N-terminal modification strategies through changing the nature of the N-terminus or N-terminal residue, we next attempted to harness the reactivity of neighbouring residues through the development of proximity-driven chemistries. Proximity-driven chemistries rely on protein-ligand interactions bringing weakly reactive reagents into close contact with nucleophilic residues present on the protein surface, creating a pseudo-intramolecular environment that promotes otherwise unfavourable reactions, and thus providing site-selectivity (*Fig. 3.14*). These chemistries fall into two classifications: *i)* ligand-directed (LD) chemistries, i.e., reaction between an electrophile linked to the ligand and the protein residue (*Fig. 3.14a*); and *ii)* affinity-guided (AG) catalysis, i.e., reaction between an electrophile and the protein residue, promoted by a catalyst linked to the ligand (*Fig. 3.14b*).[16] A key aspect of these approaches is that ligand-binding is reversible, and therefore the end result is modification at the nucleophilic protein residue only.

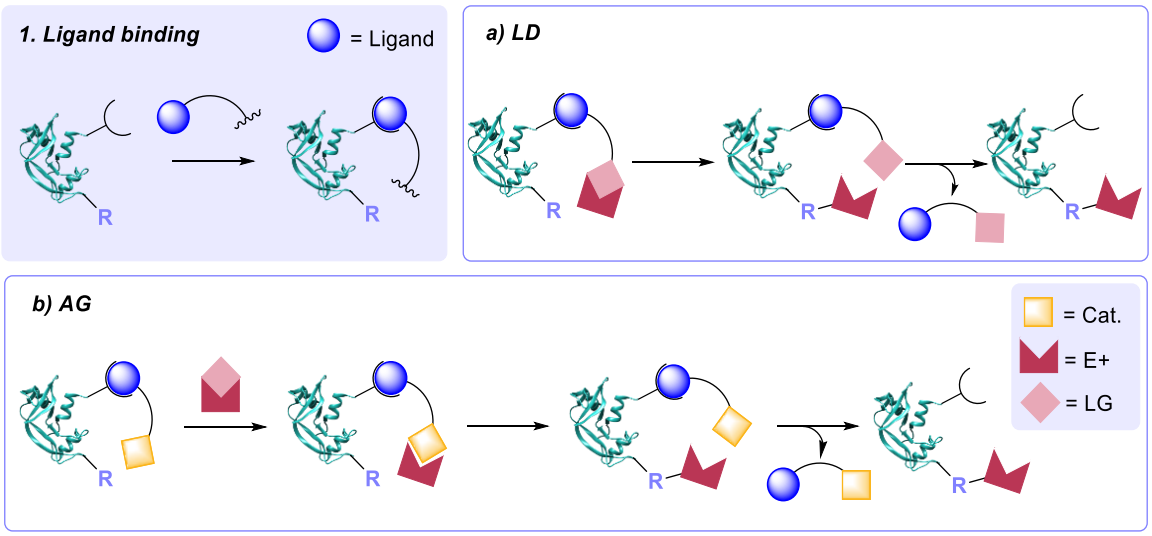


Figure 3.14. ***a)*** Ligand-directed (LD) and ***b)*** affinity-guided (AG) catalysis approaches to proximity-driven protein modification (E+ = electrophile, Cat. = catalyst, LG = leaving group). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

Examples of reactive functionalities incorporated in LD strategies include acyl imidazoles (AI) for Lys, Ser and Tyr modification (*Fig. 3.15a*), and *N*-acyl-*N*-alkyl sulfonamides (NASA) for irreversible Lys modification (*Fig. 3.15b*).[17] Lys modification can also be achieved using affinity-guided oxime (AGOX) chemistry, developed by Hamachi and co-workers, through the use of a ligand-modified pyridinium oxime catalyst and NASA acyl donor (*Fig. 3.15d*).[18] Diazirines have also been extensively employed as photo-crosslinkers in proximity labelling; the diazirine is promoted to an excited state upon irradiation with 350 nm light, followed by loss of N_2 and the formation of a carbene intermediate, which can cross-link *via* insertion into nearby C-H or X-H bonds following ligand binding (*Fig. 3.15c*).[19]

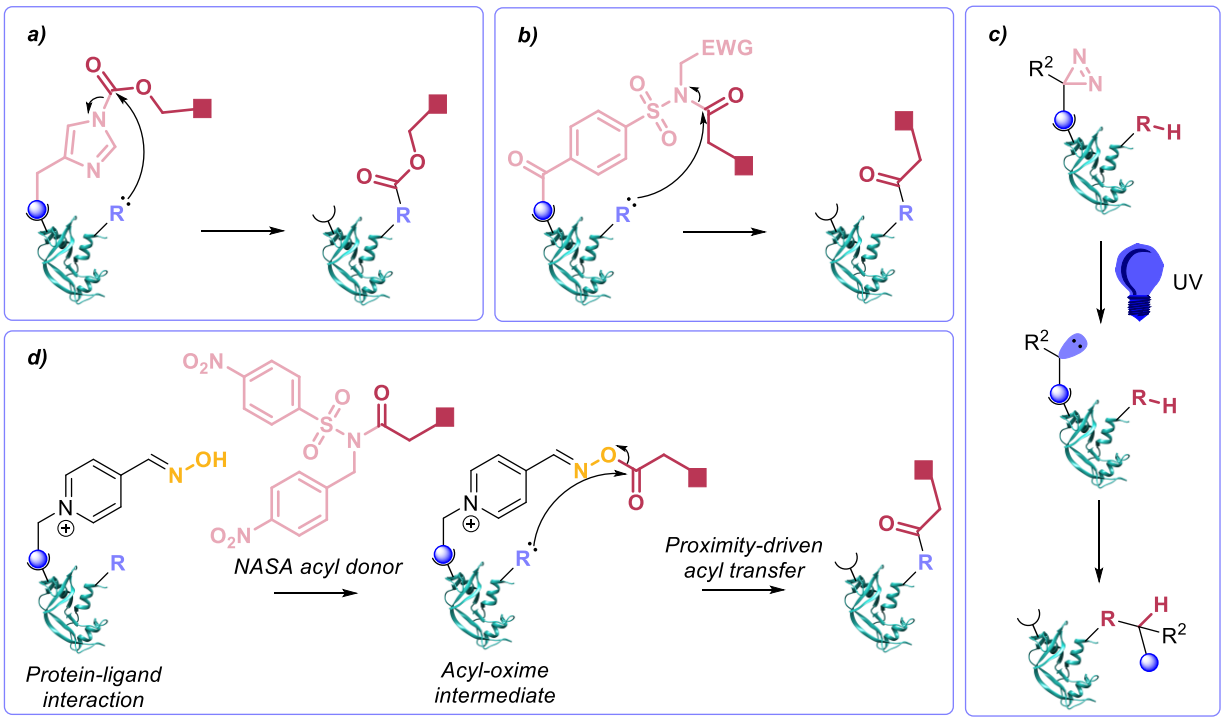


Figure 3.15. Examples of ligand-directed chemistries using ***a)*** acyl imidazoles; ***b)*** NASA; or ***c)*** diazirines; and ***d)*** AGOX chemistry through use of a protein ligand-modified pyridinium oxime catalyst and NASA acyl donor. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

Traditionally, in both classes of proximity-driven chemistries, a strong covalent bond is formed between the residue and electrophile, whilst the protein-ligand interaction is temporary and non-covalent, allowing post-labelling dissociation of the ligand.[16] More recently, "Linchpin" protein modification strategies have been demonstrated, in which a protein modification reagent is designed with a highly reactive and chemo-selective linchpin group for reversible covalent protein modification, linked by a spacer to a non-chemo-selective reagent of low reactivity for irreversible proximity modification (*Fig. 3.16*).[20–23]

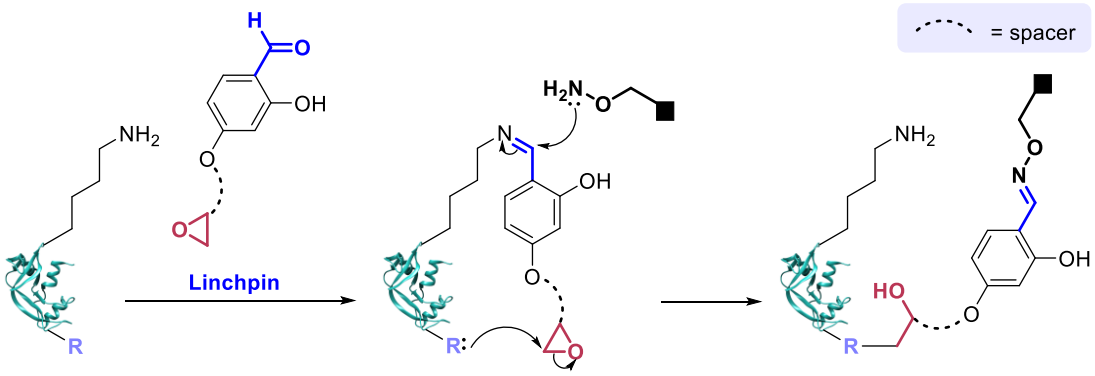


Figure 3.16. Linchpin approach exploiting dynamic covalent chemistry of imines.[21] Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

We proposed the use of a 2-PCA functional group in lieu of the conventional non-covalent ligation approach for both LD and AG modifications, effectively aiming to exploit a dynamic covalent 'linchpin' to achieve reversible binding. We hoped to use the linchpin approach to initially guide the protein modification to the N-terminus using the 2-PCA functionality, followed by irreversible reaction with a diazirine or oxime catalytic group linked to the 2-PCA reagent (*Fig. 3.17*). In Chapter 2, we have previously demonstrated that the 2-PCA modification can then be cleaved by addition of a competitor (e.g., DiAla) or dialysis at elevated temperatures.

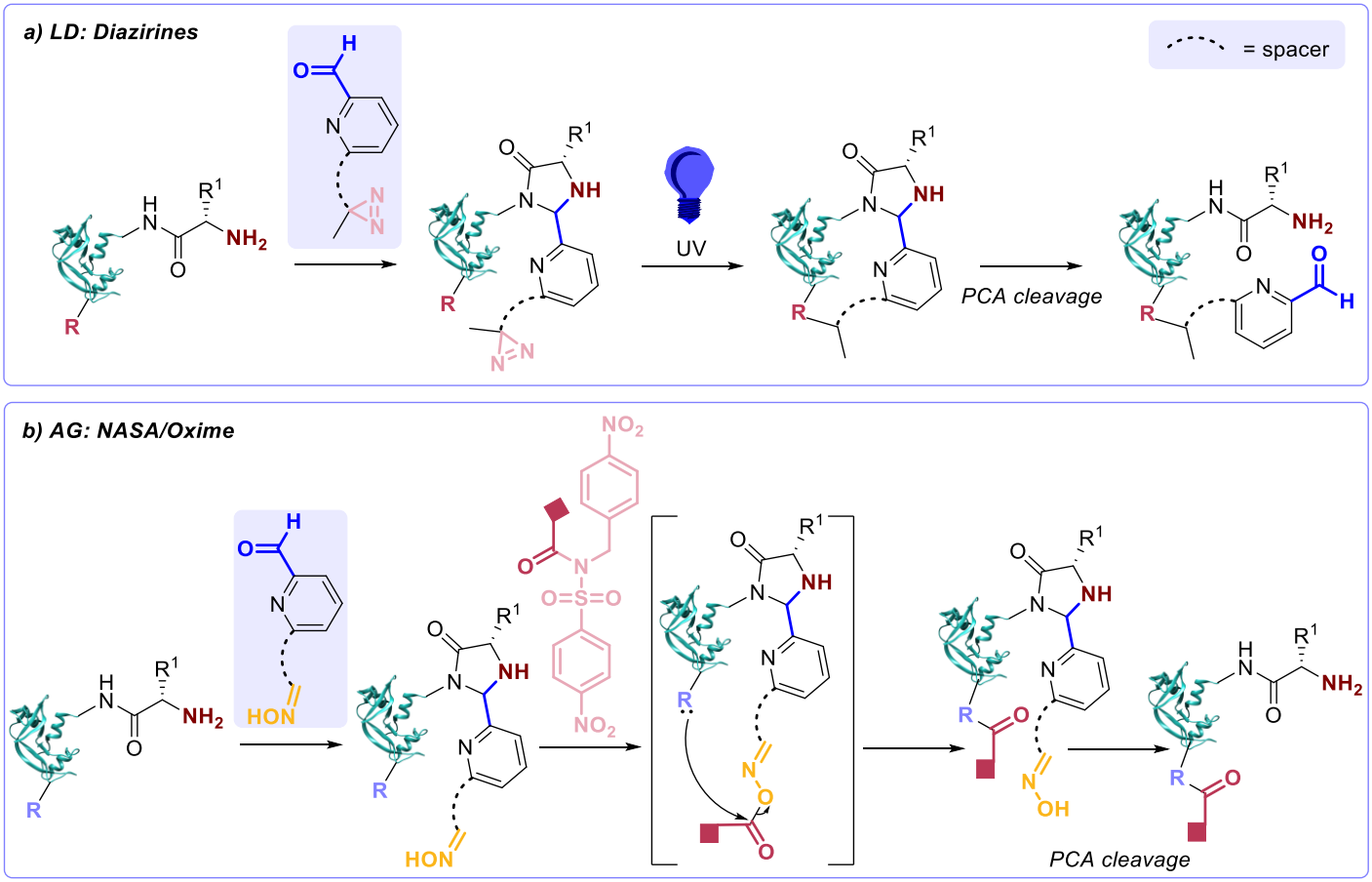


Figure 3.17. Proposed strategies for proximity-driven chemistries using 2-PCA as a linchpin for ***a)*** LD and ***b)*** AG modifications. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

3.3.2 Ligand-directed strategy: Diazirines

3.3.2.1 Initial proof of concept

Firstly, we attempted the LD strategy using a 2-PCA linchpin, with a diazirine functionality (2-PCA **3.16**) for irreversible reaction with nearby electrophilic groups on the surface of the protein upon irradiation with UV light. NHS ester **3.18** was prepared via EDC coupling between *N*-hydroxysuccinimide and carboxylic acid **3.17** (**3.17** prepared by Dr. Laetitia Raynal) in 45% yield, followed by reaction with 2-PCA **3.19** (see Chapter 2 for synthesis) to afford 2-PCA **3.16** in 57% yield (*Fig. 3.18*).

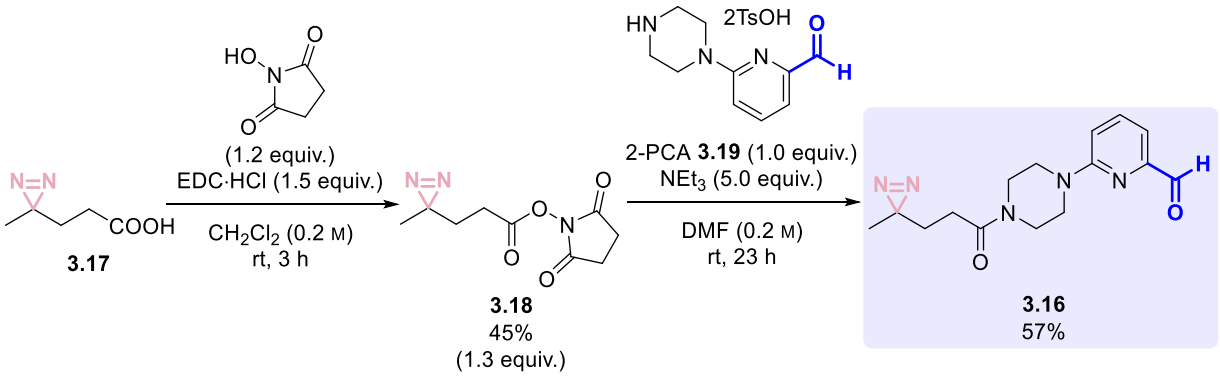


Figure 3.18. Synthesis of diazirine-functionalised 2-PCA **3.16**.

Model proteins myoglobin and RNase A were first N-terminally modified with 2-PCA **3.16**, and the conversion was determined prior to purification, and following removal of excess reagent *via* dialysis at 4 °C. Conversion was then determined for both crude and purified samples after UV irradiation for 15 min (*Fig. 3.19*). Unfortunately, no RNase A was observed by MS, perhaps due to signal suppression or protein precipitation. LC-MS revealed very low levels of myoglobin conversion to conjugate **3.20** (<10%), with no diazirine/residue reactivity upon UV irradiation to conjugate **3.21**. Observation of the formation of conjugate **3.22** indicated that the carbene formed under UV irradiation was perhaps too reactive or that there were no suitable electrophiles in close enough proximity, resulting in reaction of the carbene with water rather than protein residues. The high reactivity of carbenes with surrounding water molecules is what provides our approach with site-selectivity: the short solution half-life ($t_{1/2}$ < 1 ns) of carbenes limits their diffusion radius, ensuring that only carbenes in close proximity to the protein will react with the protein.[19] Reaction with surrounding water molecules therefore suggests the absence of suitable groups on the protein nearby and that the flexibility/length of the 2-PCA linker was unsuitable for this target protein. Whilst improved conversions and diazirine/residue reactivity could potentially be achieved by the screening of a wider panel of proteins and a range of 2-PCAs with varying linker lengths, we decided not to pursue this route any further due to the side-reactivity of the diazirine.

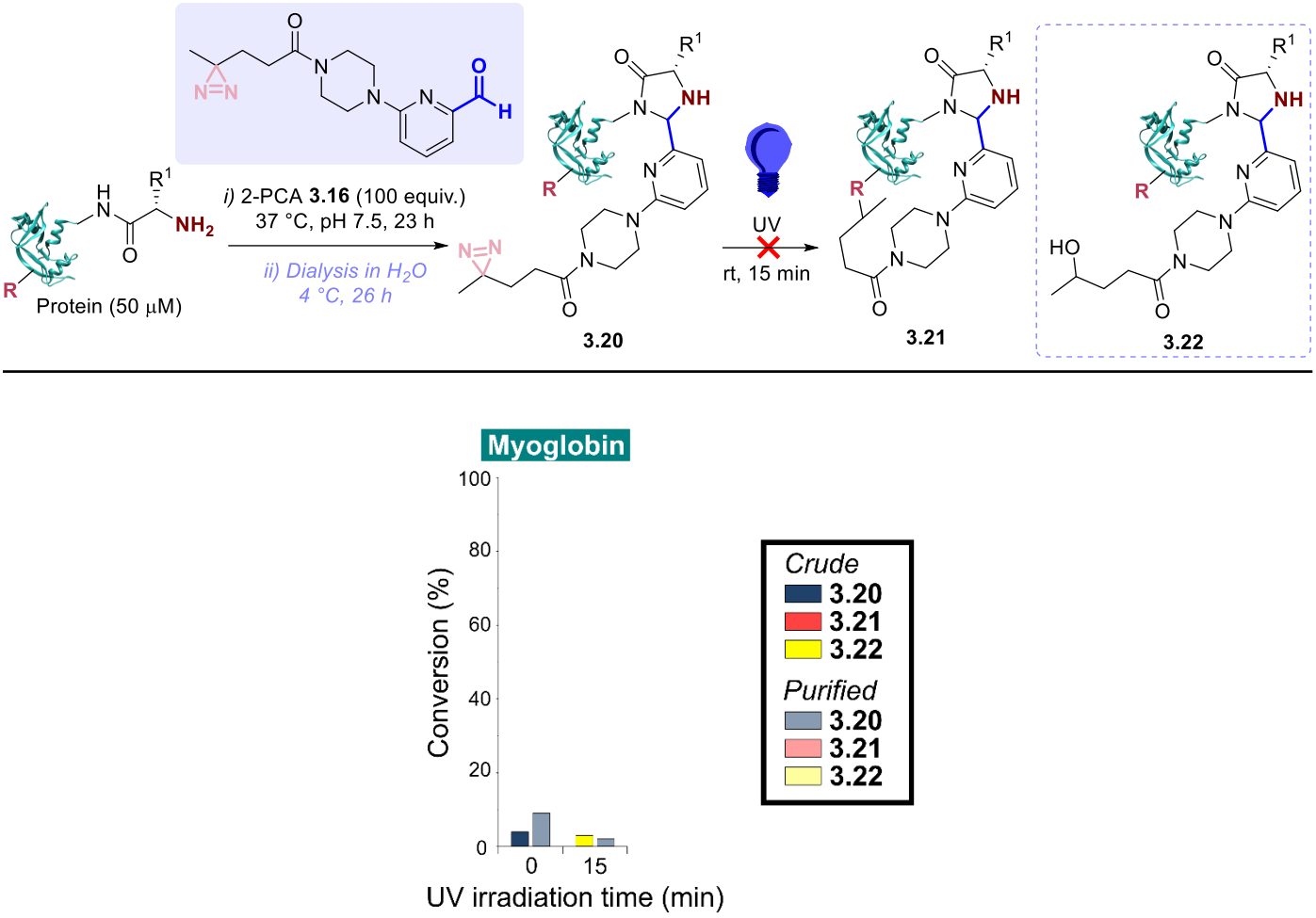


Figure 3.19. Ligand-directed 2-PCA/diazirine chemistry for the modification of proteins. Myoglobin and RNase A (50 μM) were modified with reagent **3.16** (100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 23 h, and the conversion was determined by LC-MS before and after purification by dialysis, and after UV irradiation for 15 min. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

3.3.3 Affinity-guided Strategy: NASA/Oximes

3.3.3.1 Initial proof of concept

We next turned our focus to NASA/oxime reagents: our aim was to modify proteins using 2-PCA linchpins, followed by addition of NASA reagents for oxime-catalysed acyl transfer to nucleophilic protein residues in close proximity to the N-terminus. Dr. Laetitia Raynal designed and prepared 2-PCA/oxime reagents **3.23** and **3.24**, with varying linker lengths to allow the optimal distance between the N-terminus and the nucleophilic protein residue be probed (*Fig. 3.20*).

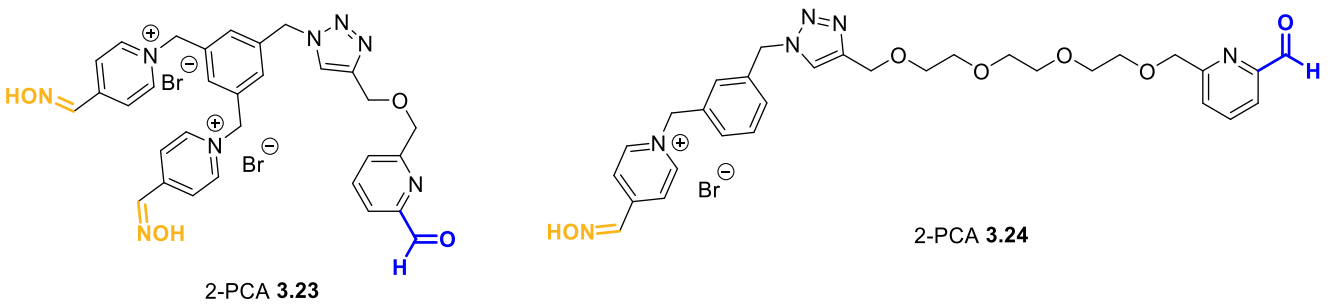


Figure 3.20. 2-PCA linchpins with oxime functionalities, designed and synthesised by Dr. Laetitia Raynal.

Unfortunately, 99-100% protein precipitation of both myoglobin and RNase A was induced by the presence of 2-PCA **3.23**, as determined by SDS-PAGE analysis, indicating the reagent was unsuitable for protein modification due to its negative influence on protein solubility (*Fig. 3.21a*). Protein precipitation also occurred upon modification of both RNase A and myoglobin with 2-PCA **3.24**, with 84% and 7% retention of soluble protein respectively (*Fig. 3.21b*). We therefore proceeded with reagent **3.24** only, due to reduced protein precipitation in comparison to reagent **3.23**. Optimisation of reaction conditions to reduce the level of protein precipitation with this reagent will be discussed later.

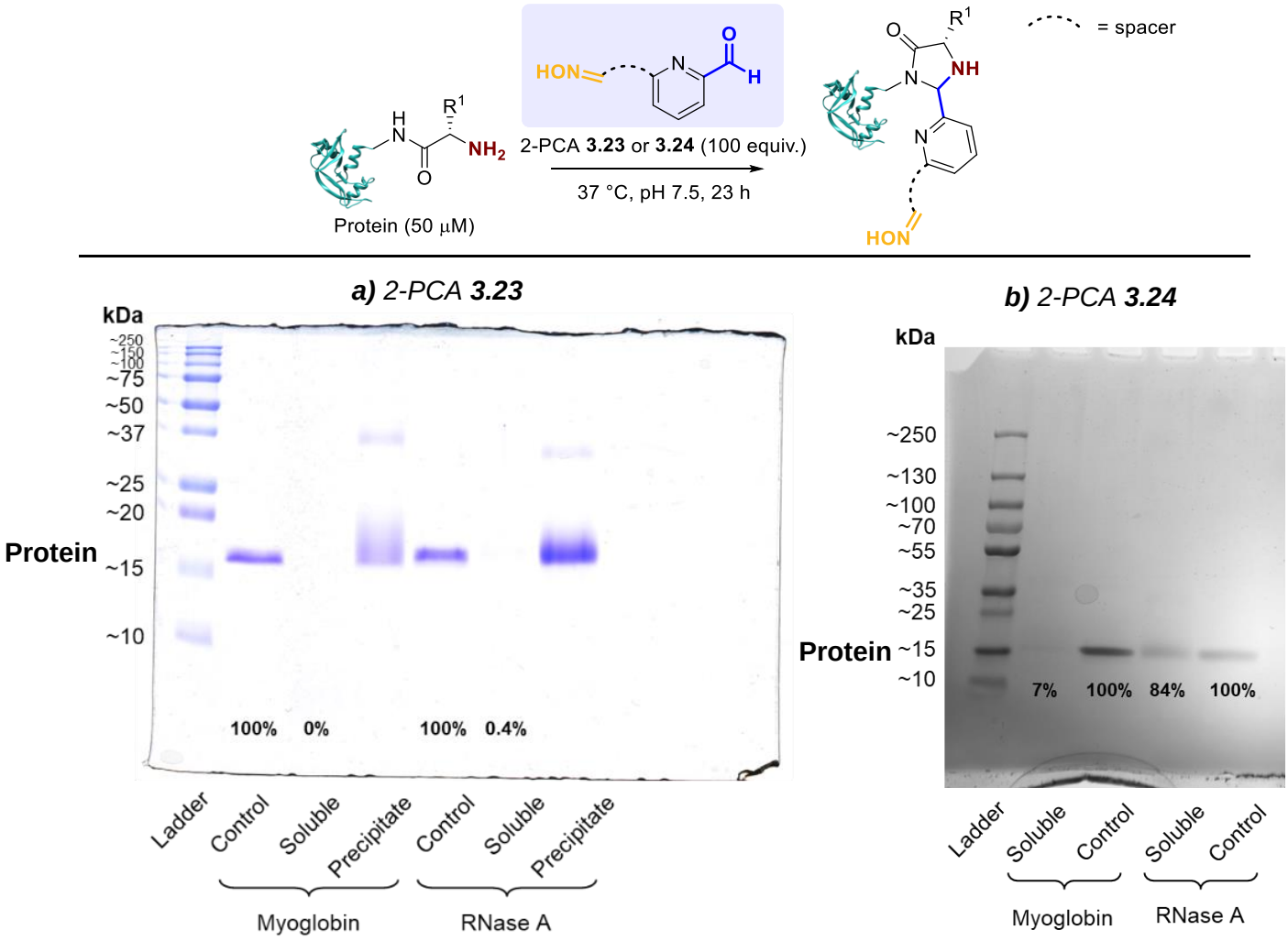
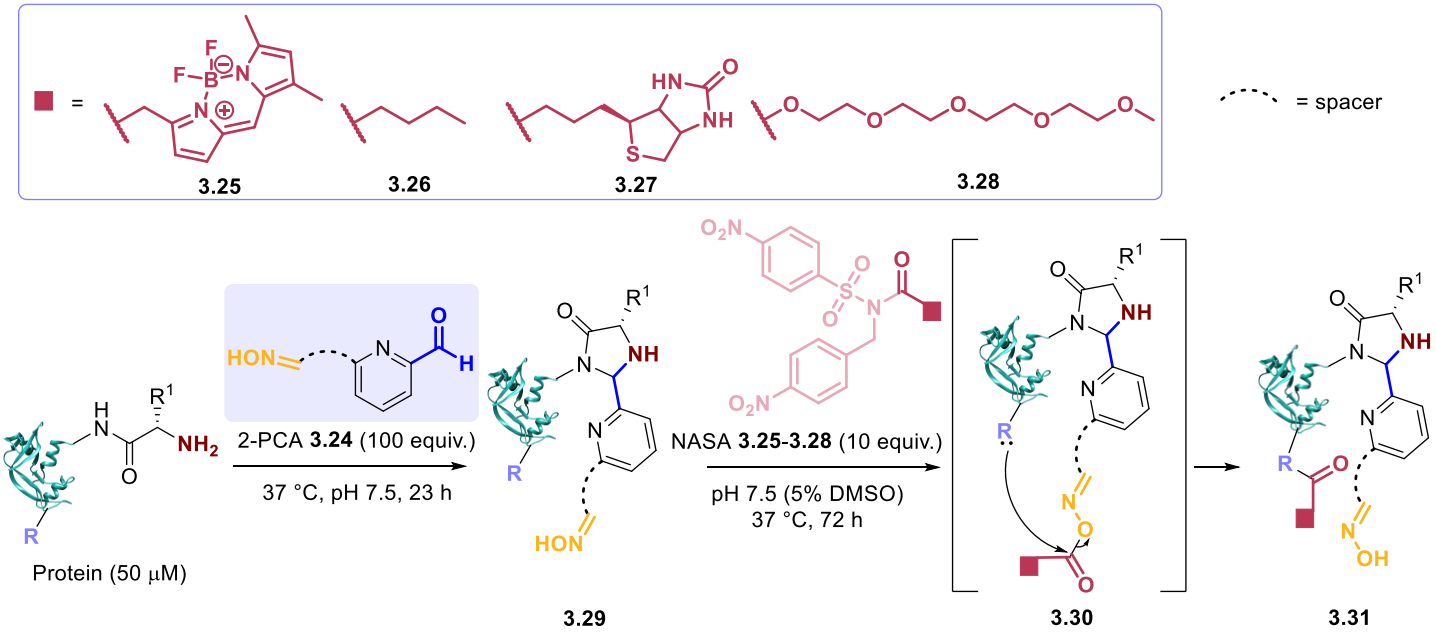


Figure 3.21. SDS-PAGE of myoglobin and RNase A modified with ***a)*** 2-PCA **3.23**; and ***b)*** 2-PCA **3.24**, in comparison to myoglobin/RNase A controls treated under identical conditions in the absence of 2-PCA reagent *(note the difference in appearance in gels was due to the use of different imaging instruments, with gel **a)** imaged in the Biology department, and gel **b)** imaged in the Chemistry department).* The streaky band seen by SDS-PAGE for both precipitates in gel ***a)*** may have been due to protein precipitation and subsequent re-homogenisation, resulting in protein degradation. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

A panel of model proteins (myoglobin, RNase A, insulin and CjX183-D WT) was modified with 2-PCA **3.24**, followed by incubation with NASA reagents **3.25**-**3.28** (synthesised by Dr. Laetitia Raynal) at 37 °C for 72 h. Conversion was followed by LC-MS over time (*Fig. 3.22*). Unfortunately, RNase A, insulin, and CjX183-D WT were not observed by MS due to signal suppression upon co-elution with impurities from 2-PCA **3.24**. For myoglobin, up to 15% acyl transfer was observed with NASA **3.26** and **3.27**, however no acyl addition was observed after 72 h. This perhaps suggests that there were no suitable nucleophilic groups available on the protein nearby and that the additions observed at 3 h and 24 h were transient intermediate **3.30**, rather than **3.31**, although we expected the intermediate to be very unstable and not observed. Alternatively, it could be that the modification took place at a serine residue to form a hydrolysable ester, which was then cleaved. Up to 2% acyl transfer was observed under control conditions in the absence of oxime catalysis, suggesting some degree of the acyl transfer observed may have been non-proximity driven (*Fig. S3.13*): again, the NASA may have added to the protein at a serine residue to form a hydrolysable ester. Note that the quality of the MS data was poor: calculated conversions were approximate, with strong background signals resulting in ambiguity. The low levels of soluble myoglobin observed previously in Fig. *3.21b* perhaps accounts for the poor-quality MS spectra observed.



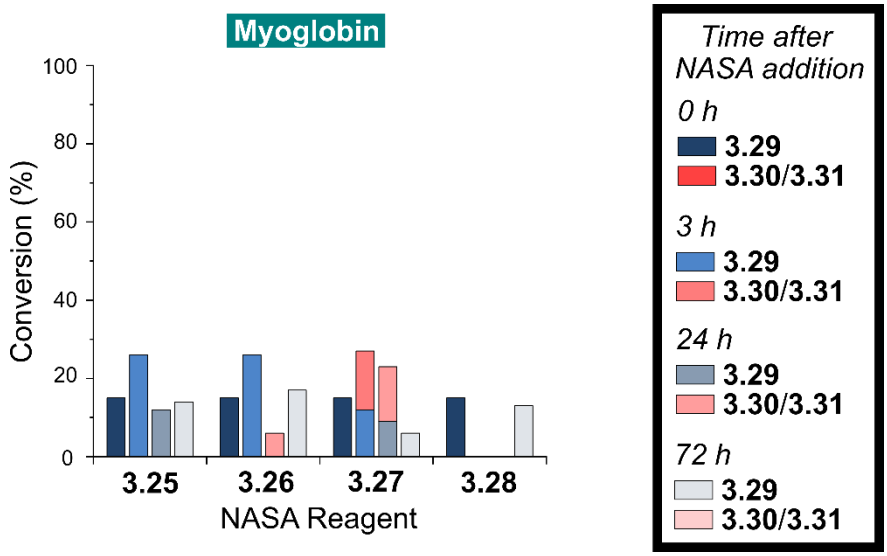


Figure 3.22. AGOX chemistry through use of 2-PCA pyridinium oxime catalyst **3.24** and NASA acyl donors. Myoglobin, RNase A, insulin and CjX183-D WT (50 μM) were modified with reagent **3.24** (100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 23 h, followed by the addition of NASA acyl donors **3.25**-**3.28** (10 equiv.) and incubation at 37 °C for 72 h. Conversion was followed by LC-MS. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

3.3.3.2 Attempts to improve the data quality

We next sought to explore the effect of different variables on the solubility of myoglobin and RNase A: parameters varied included 2-PCA concentration, temperature, and buffer concentration (*Fig. 3.23*). We also included an additional purification step *via* dialysis to remove excess 2-PCA **3.24** prior to the addition of NASA, to prevent co-elution with RNase A and the resulting MS signal suppression. However, dialysis for 25 h was ineffective for complete 2-PCA removal. Despite previous conversions of up to 30% in *Fig. 3.22*, minimal (<5%) conversion of myoglobin was observed under all conditions, so myoglobin was excluded from

further studies. A reduction in 2-PCA concentration from 5 mM to 2.5 mM had no influence on RNase A conversion, although with a decrease in conversion below 2.5 mM. Reduced conversion of RNase A was also observed upon lowering the reaction temperature from 37 °C to 25 °C, and reduced selectivity for single modification of RNase A was observed upon increasing the sodium phosphate buffer concentration from 50 mM to 100 mM.

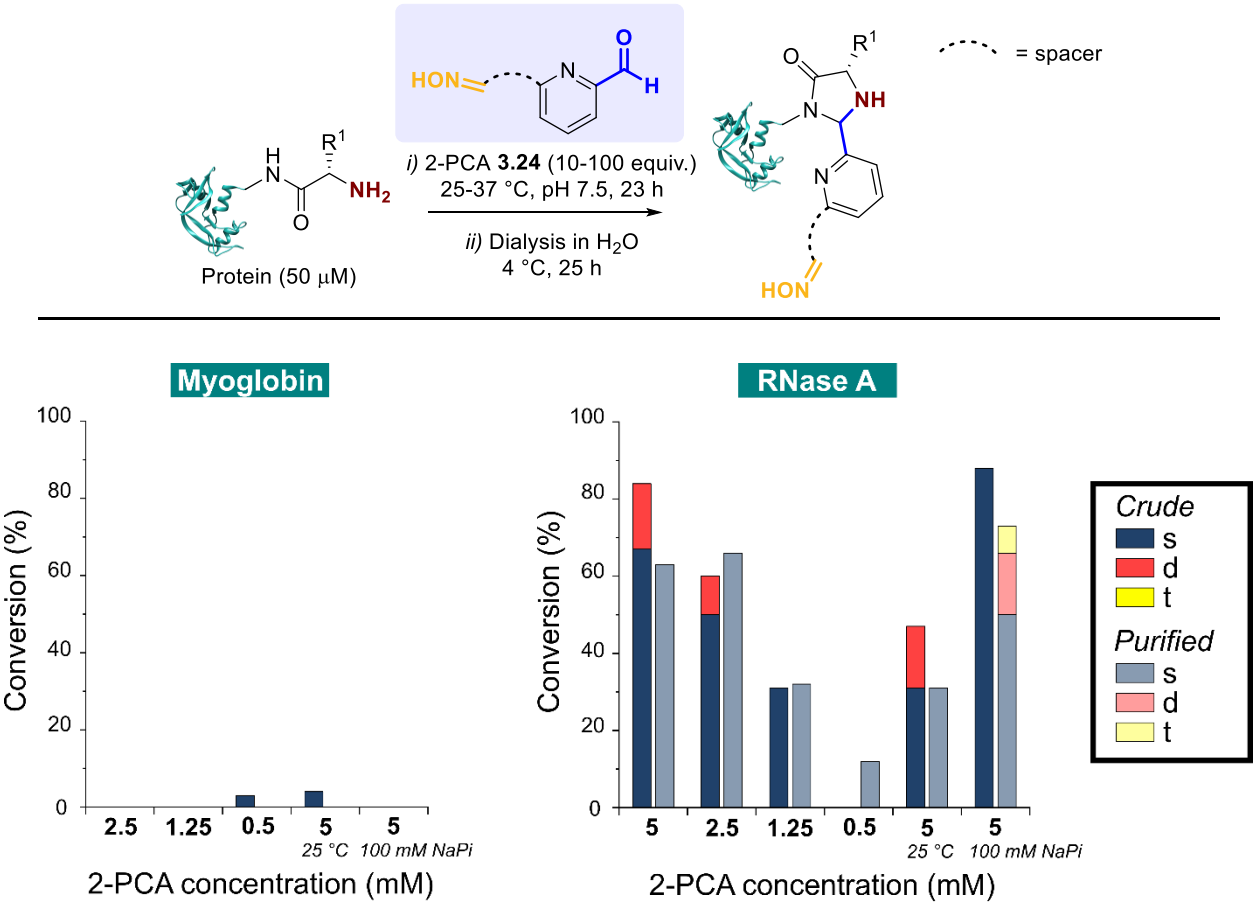


Figure 3.23. Precipitation experiment for the modification of myoglobin and RNase A with 2-PCA **3.24**. Myoglobin and RNase A (50 μM) were modified with reagent **3.24** (10-100 equiv.) in Na phosphate buffer (50-100 mM, pH 7.5) at 25-37 °C for 23 h, and the reaction mixture was analysed by LC-MS before and after purification by dialysis at 4 °C (s = single, d = double, t = triple modification; assume reaction was carried out at 37 °C in 50 mM sodium phosphate buffer unless otherwise indicated). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

We ran SDS-PAGE gels of the soluble fractions of the reactions to compare directly to a myoglobin/RNase A control prepared identically in the absence of the modification reagent (*Fig. 5.24*). SDS-PAGE revealed that protein precipitation occurred under most reaction conditions. For myoglobin, the level of precipitation increased with increasing 2-PCA concentration (from 0-99% precipitation at 1.25-5 mM 2-PCA), confirming that protein solubility was strongly influenced by the 2-PCA concentration. RNase A samples had generally higher solubility than myoglobin samples: reaction parameters therefore had a less pronounced influence on solubility, with no precipitation at 25 °C.

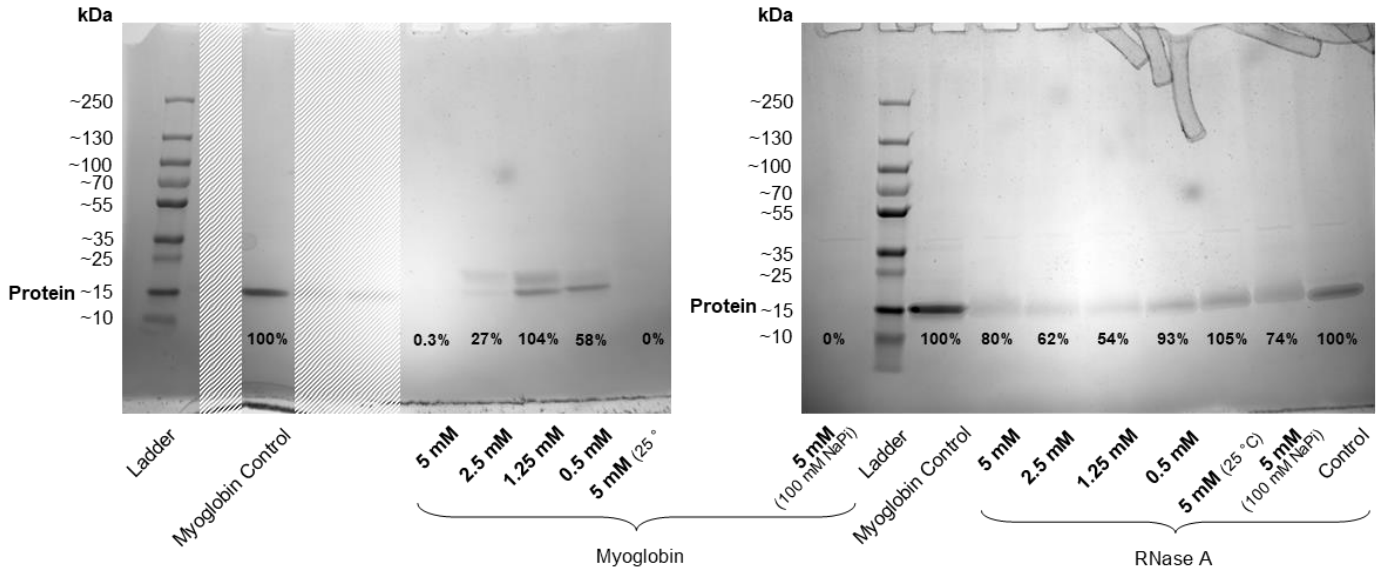


Figure 3.24. SDS-PAGE of the soluble fractions of reaction mixtures of myoglobin and RNase A modified with 2-PCA **3.24** under a range of conditions, in comparison to myoglobin/RNase A controls prepared identically in the absence of the modification reagent.

We next modified RNase A with 2-PCA **3.24** (2.5 mM) under reaction conditions selected as a compromise between conversion and solubility. Increased dialysis time (48 h) with a larger volume of water and additional water changes was also used to help to remove excess 2-PCA post reaction. However, the dialysis was not fully effective, resulting in co-elution and protein signal suppression; all results were therefore of poor quality. Note that for PCAs studied previously in Chapter 2, no problems were encountered during dialysis; these purification issues were unique to 2-PCA **3.24**. "Purified" conjugates were incubated with NASA reagents **3.26**, **3.27** and **3.32** (synthesised by Dr. Laetitia Raynal) at 37 °C for 18 h, and conversion was followed by LC-MS over time (*Fig. 5.25*). Unfortunately, N-terminal conversions were significantly lower (approx. 30%) than seen previously under similar conditions in *Fig. 5.23*, and no acyl transfer was observed. Whilst we suspected the acyl transfer step required optimisation of the levels of NASA acyl donor, our main focus moving forwards was in obtaining pure conjugates (**3.29**) prior to NASA addition.

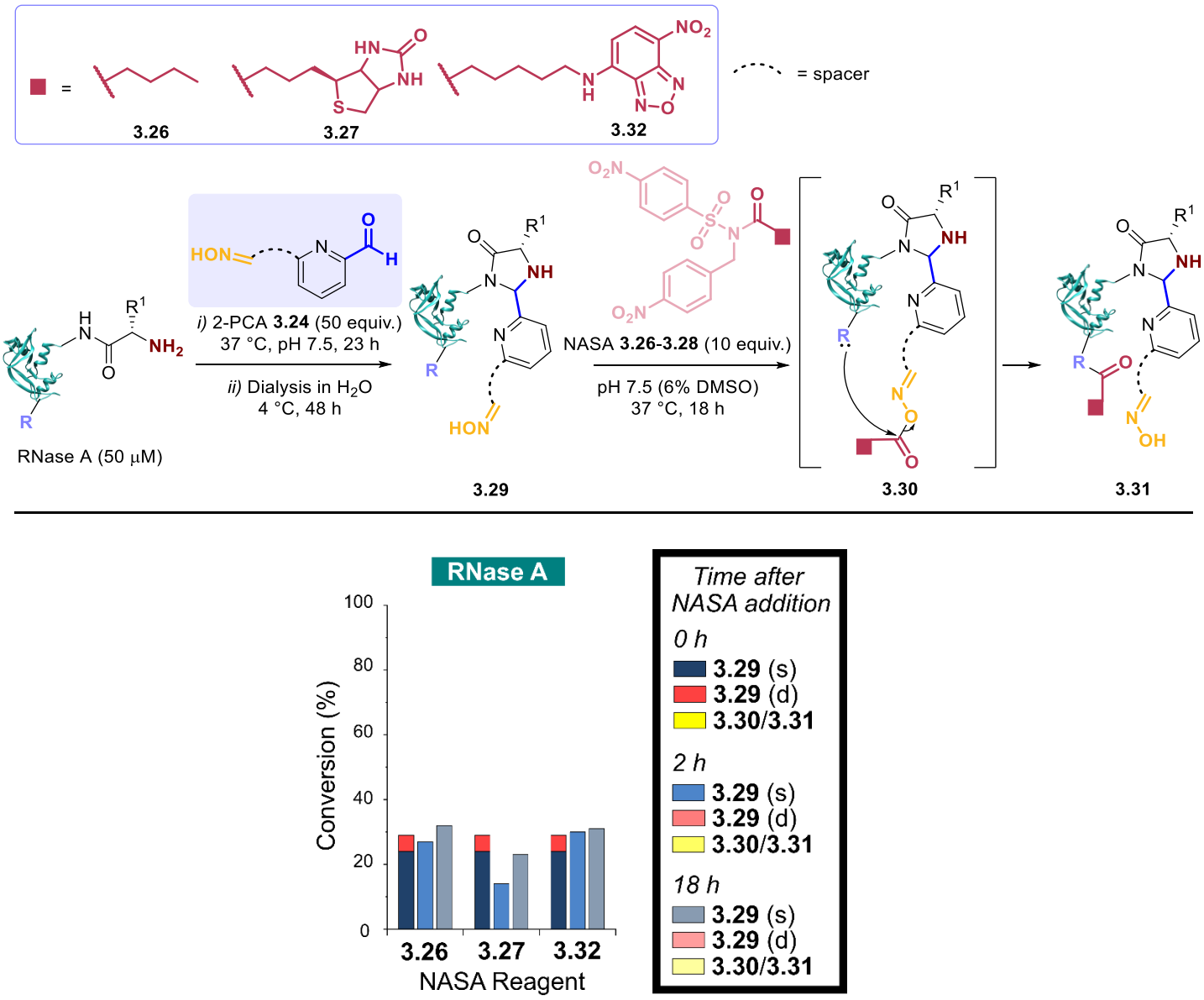


Figure 3.25. AGOX chemistry through use of 2-PCA pyridinium oxime catalyst **3.24** and NASA acyl donors, under conditions optimised for protein solubility. RNase A (50 μM) was modified with reagent **3.24** (50 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 23 h and purified, followed by the addition of NASA acyl donors **3.26**, **3.27** and **3.32** (10 equiv.) and incubation at 37 °C for 18 h. Conversion was followed by LC-MS. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

We next attempted to optimise the dialysis conditions to allow complete removal of excess 2-PCA to avoid the issues associated with LC co-elution and protein signal suppression. Three aqueous glycine scavenger solutions of varying concentrations (5 mM, 500 mM and 5 M) were compared for the purification of RNase A/2-PCA **3.24** conjugates. We hoped for glycine to remove excess 2-PCA *via* transient imine formation (*Fig. 3.26*). We chose an amino acid scavenger as whilst we expected dipeptide scavengers to be more effective than amino acids due to the potential for imidazolidinone formation, this would be at the cost of conversion due to competitive cleavage of the N-terminal modifications explored previously in Chapter 2. Unfortunately,

whilst increasing levels of glycine reduced the levels of excess 2-PCA, complete removal of 2-PCA was not achieved and the protein mass signals were still significantly suppressed, not allowing deconvolution. Due to difficulties in purification and no NASA reactivity observed, we decided that RNase A was a poor choice of protein for the 2-PCA/oxime/NASA approach with 2-PCA **3.24**, so no further studies were carried out.

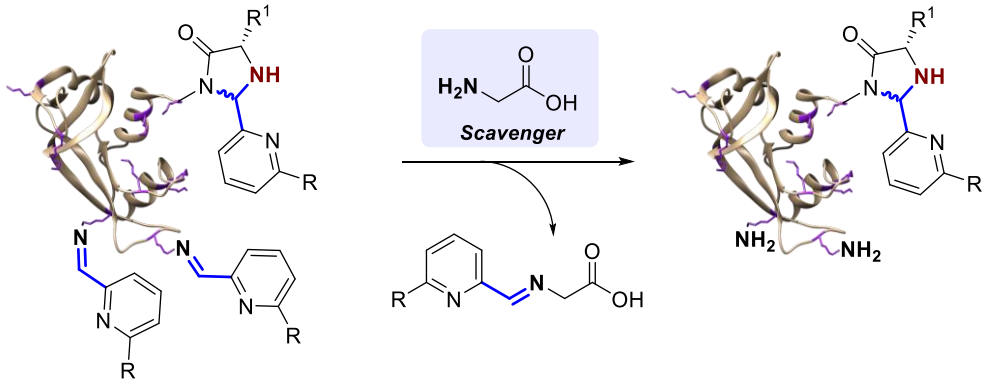


Figure 3.26. Glycine scavenger for 2-PCA removal by dialysis. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

3.3.4 Conclusion

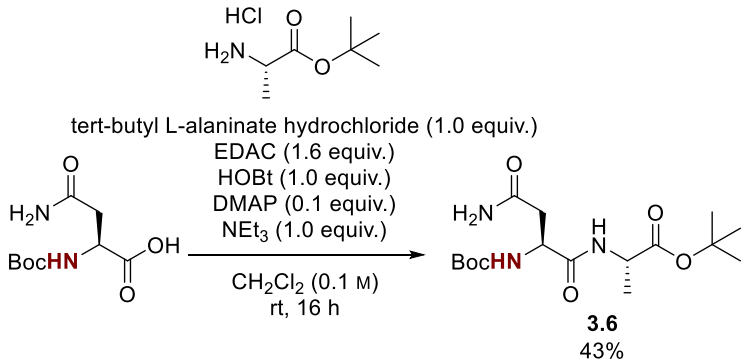
Despite initial promising results, using 2-PCA/oxime reagents for the AGOX strategy has presented many issues with conjugate purification, 2-PCA reactivity, protein solubility and NASA reactivity, so we decided not to pursue this avenue any further. In future work, more extensive 2-PCA reagent design could be explored to include the incorporation of solubilising groups such as sulfonic acids. The use of solubilising additives, such as those discussed later in Chapter 5, could also be attempted to reduce protein precipitation. In Chapter 5, we found size exclusion chromatography to be effective for the removal of modification reagents; this purification strategy could also be more effective than dialysis for removal of 2-PCA **3.24**.

3.4 Experimental

3.4.1 Reagent synthesis

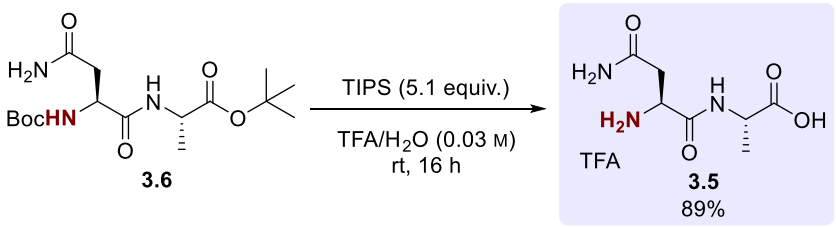
a) *Asn-Ala* ***3.5***

tert-Butyl (tert-butoxycarbonyl)-L-asparaginyl-L-alaninate (3.6)



1-Ethyl-3-(3'-dimethylaminopropyl)carbodiimide hydrochloride (1.32 g, 6.9 mmol, 1.6 equiv.), 1-hydroxybenzotriazole hydrate (0.58 g, 4.3 mmol, 1.0 equiv.), and 4-dimethylaminopyridine (55 mg, 0.4 mmol, 0.1 equiv.) were added to a solution of Boc-L-asparagine (1.00 g, 4.3 mmol, 1.0 equiv.) and *tert*-butyl L-alaninate hydrochloride (0.78 g, 4.3 mmol, 1.0 equiv.) in dichloromethane (45 mL, 0.1 M) at room temperature. The reaction mixture was cooled to 5 °C, and triethylamine (600 µL, 4.3 mmol, 1.0 equiv.) was added. The reaction mixture was stirred at 5 °C for 15 min, and then warmed to room temperature and stirred for 16 h. The reaction mixture was then quenched with saturated aqueous NH_4Cl solution (20 mL) and the aqueous extracted with CH_2Cl_2 (3 × 15 mL). The combined organics were washed with brine (2 × 15 mL) and with water (2 × 15 mL), dried over Na_2SO_4 and concentrated under reduced pressure. The residue obtained was purified by flash column chromatography (0-15% MeOH:CH_2Cl_2, R_f 0.56 in 10% MeOH:CH_2Cl_2) to afford the title compound (0.67 g, 1.9 mmol, 43%) as an off-white solid. **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H 1.35 (3H, d, *J* = 7.2 Hz, Me), 1.45 (18H, app s, Boc & ^{t}Bu), 2.58 (1H, dd, *J* = 15.6, 6.0 Hz, -C$\underline{H}_2$CONH$_2$), 2.91 (1H, dd, *J* = 15.6, 4.0 Hz, -C$\underline{H}_2$CONH$_2$), 4.31-4.42 (1H, m, C$\underline{H}$MeCOO^{t}Bu), 4.45-4.49 (1H, m, -C$\underline{H}$CH$_2$CONH$_2$), 5.62 (1H, br app s, N$\underline{H}$ or 1×N$\underline{H}_2$), 6.10 (1H, d, *J* = 8.1 Hz, N$\underline{H}$ or 1×N$\underline{H}_2$), 6.17 (1H, br app s, N$\underline{H}$ or 1×N$\underline{H}_2$), 7.39 (1H, d, *J* = 7.2 Hz, N$\underline{H}$ or 1×N$\underline{H}_2$); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C 18.27 (Me), 27.98 (-NHCOOC(C$\underline{H}_3$)$_3$ or -CHMeCOOC(C$\underline{H}_3$)$_3$), 28.33 (-NHCOOC(C$\underline{H}_3$)$_3$ or -CHMeCOOC(C$\underline{H}_3$)$_3$), 37.09 (-C$\underline{H}_2$CONH$_2$), 49.00 (-C$\underline{H}$MeCOO^{t}Bu), 50.99 (-C$\underline{H}$CH$_2$CONH$_2$), 80.33 (-NHCOOC(CH$_3$)$_3$ or -CHMeCOOC(CH$_3$)$_3$), 81.94 (-NHCOOC(CH$_3$)$_3$ or -CHMeCOOC(CH$_3$)$_3$), 155.75 (-NHCOO^{t}Bu), 170.74 (-CONHCHMe- or -CONH$_2$), 171.67 (-CHMeCOO^{t}Bu), 173.53 (-CONHCHMe- or -CONH$_2$); **λ_{max}** (ATR, cm^{-1}) 3403, 3346, 2985, 1725, 1689, 1660, 1519, 1440, 1369, 1323, 1291, 1241, 1167, 1045, 1023, 908, 853, 806, 779, 630, 531, 517, 493, 465; **HRMS** (ESI$^+$) $C_{16}H_{29}N_3NaO_6$ $[M+Na]^+$ found 382.1950, calculated 382.1949.

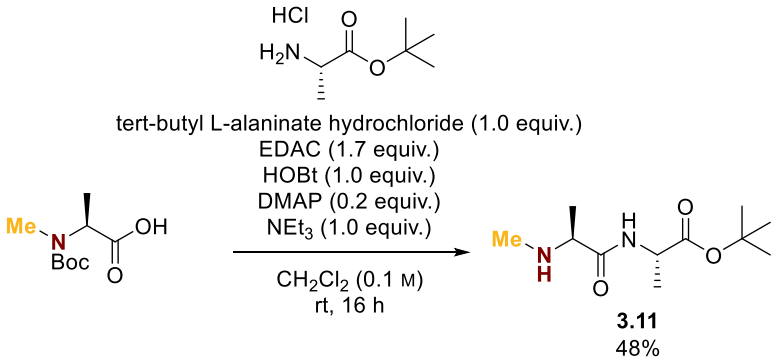
L-asparaginyl-L-alanine trifluoroacetic acid salt (3.5)



Peptide **3.6** (0.21 g, 0.6 mmol, 1.0 equiv.) was added to a solution of triisopropylsilane (0.60 mL, 2.9 mmol, 5.1 equiv.) in H_2O (1 mL). Trifluoroacetic acid (20 mL) was added dropwise and the reaction mixture was stirred at room temperature for 16 h. The reaction mixture was concentrated *in vacuo* to ~2 mL volume. The residue was dropped into ice cold diethyl ether (~ 120 mL), and the resultant precipitate collected by centrifugation (3000 rpm, 6 min). The residual solid was allowed to air dry for 10 min, then dissolved in deionized water (5 mL) and dried by lyophilisation to afford the title compound (0.16 g, 0.51 mmol, 89%) as a yellow residue. **^{1}H NMR** (400 MHz, D_2O) δ_H 1.43 (3H, d, *J* = 7.3 Hz, Me), 2.90 (1H, dd, *J* = 16.9, 8.0 Hz, -C$\underline{H}_2$CONH$_2$), 3.00 (1H, dd, *J* = 16.9, 4.9 Hz, -C$\underline{H}_2$CONH$_2$), 4.35 (1H, dd, *J* = 8.0, 4.9 Hz, -C$\underline{H}$CH$_2$CONH$_2$), 4.41 (1H, q, *J* = 7.3 Hz, -C$\underline{H}$MeCOOH); **^{13}C NMR** (101 MHz, D_2O) δ_C 16.01 (Me), 34.86 (-$\underline{C}$H$_2$CONH$_2$), 48.96 (-$\underline{C}$HMeCOOH), 49.74 (-$\underline{C}$HCH$_2$CONH$_2$), 168.45 (-$\underline{C}$ONHCHMeCOOH), 172.85 (-$\underline{C}$ONH$_2$), 176.02 (-$\underline{C}$OOH); **λ_{max}** (ATR, cm^{-1}) 3075, 1655, 1560, 1458, 1426, 1340, 1186, 1132, 990, 940, 838, 799, 722, 570, 517; **HRMS** (ESI$^+$) $C_7H_{13}N_3NaO_4$ $[M+Na]^+$ found 226.0792, calculated 226.0798.

b) *N$^{\alpha}$-Me-DiAla* ***3.10***

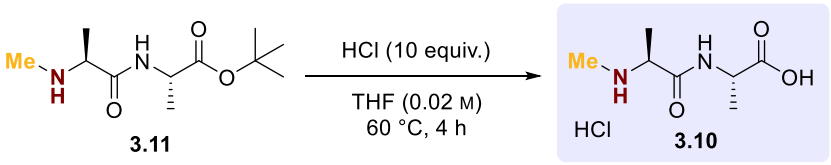
***tert*-Butyl methyl-L-alanyl-L-alaninate (3.11)**



1-Ethyl-3-(3'-dimethylaminopropyl)carbodiimide hydrochloride (318 mg, 1.7 mmol, 1.7 equiv.), 1-hydroxybenzotriazole hydrate (137 mg, 1.0 mmol, 1.0 equiv.), and 4-dimethylaminopyridine (23 mg, 0.2 mmol, 0.2 equiv.) were added to a solution of Boc-L-alanine (204 mg, 1.0 mmol, 1.0 equiv.) and *tert*-butyl L-alaninate hydrochloride (189 mg, 1.0 mmol, 1.0 equiv.) in anhydrous dichloromethane (10 mL, 0.1 M) at room temperature. The reaction mixture was cooled to 5 °C, and triethylamine (140 µL, 1.0 mmol, 1.0 equiv.) was added. The reaction mixture was stirred at 5 °C for 30 min, and then warmed to room temperature and stirred

for 16 h. The reaction mixture was then quenched with saturated aqueous NH_4Cl solution (20 mL) and the aqueous extracted with CH_2Cl_2 (3 × 15 mL). The combined organics were washed with brine (2 × 15 mL) and with water (2 × 15 mL), dried over Na_2SO_4 and concentrated under reduced pressure. The residue obtained was purified by flash column chromatography (0-5% MeOH:CH_2Cl_2, R_f 0.32 in 10% MeOH:CH_2Cl_2) to afford the title compound (111 mg, 0.5 mmol, 48%) as a yellow liquid. **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H 1.28 (3H, d, *J* = 7.0 Hz, -CHMeNHMe), 1.36 (3H, d, *J* = 7.2 Hz, -CHMeCOO^{t}Bu), 1.46 (9H, s, ^{t}Bu), 2.42 (3H, s, NHMe), 3.05 (1H, q, *J* = 7.0 Hz, -CHMeNHMe), 4.46 (1H, dq, *J* = 8.4, 7.2 Hz, -CHMeCOO^{t}Bu), 7.59 (1H, br d, *J* = 8.4 Hz, -NHCHMeCOO^{t}Bu); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C 18.78 (-CHMeCOO^{t}Bu), 19.74 (-CHMeNHMe), 28.10 (-COOC(CH_3)$_3$), 35.32 (-NHMe), 48.09 (-CHMeCOO^{t}Bu), 60.41 (-CHMeNHMe), 81.84 (-COOC(CH_3)$_3$), 172.36 (-COO^{t}Bu), 174.56 (-COCHMeNHMe); **λ_{max}** (ATR, cm^{-1}) 3317, 2978, 1733, 1656, 1518, 1453, 1368, 1315, 1225, 1149, 1053, 847, 744, 472; **HRMS** (ESI$^+$) $C_{11}H_{23}N_2O_3$ $[M+H]^+$ found 231.1704, calculated 231.1703.

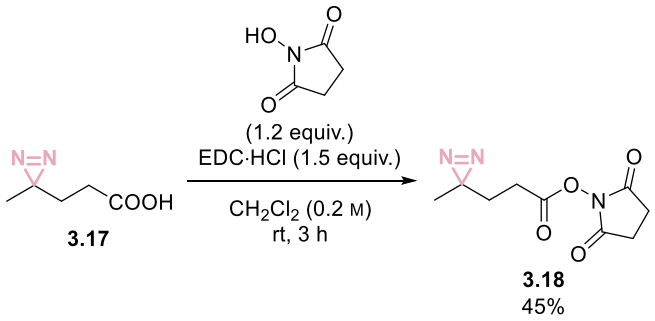
Methyl-L-alanyl-L-alanine hydrochloride (3.10)



Hydrochloric acid (2 M, 367 µL, 0.7 mmol, 10.0 equiv.) was added to a solution of peptide **3.11** (17 mg, 0.07 mmol, 1.0 equiv.) in THF (3.7 mL, 0.02 M) at room temperature. The reaction mixture was stirred at 60 °C for 4 h. After allowing to cool to room temperature, the reaction mixture was concentrated under reduced pressure and azeotroped with dichloromethane (3 × 5 mL) to afford the title compound as a off-white solid. The product was contaminated with residual HCl which could not be removed, but was effectively buffered in subsequent experiments. **^{1}H NMR** (400 MHz, D_2O) δ_H 1.43 (3H, d, *J* = 7.4 Hz, -CHMeCOOH), 1.53 (3H, d, *J* = 7.1 Hz, -CHMeNHMe), 2.67 (3H, s, NHMe), 3.94 (1H, q, *J* = 7.1 Hz, -CHMeNHMe), 4.40 (1H, q, *J* = 7.4 Hz, -CHMeCOOH); **^{13}C NMR** (101 MHz, D_2O) δ_C 15.28 (-CHMeNHMe), 15.88 (-CHMeCOOH), 31.03 (-NHMe), 48.70 (-CHMeCOOH), 57.00 (-CHMeNHMe), 169.75 (-COCHMeNHMe), 175.98 (-COOH); **HRMS** (ESI$^+$) $C_7H_{15}N_2O_3$ $[M+H]^+$ found 175.1077, calculated 175.1077.

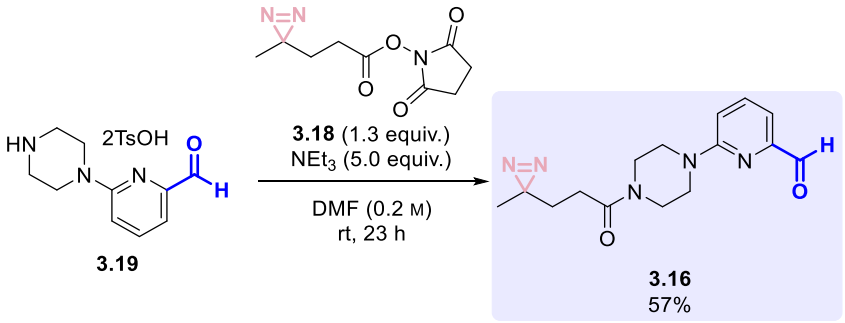
c) *Diazirine-functionalised 2-PCA* ***3.16***

2,5-Dioxopyrrolidin-1-yl 3-(3-methyl-3*H*-diazirin-3-yl)propanoate (3.18)



N-hydroxysuccinimide (0.22 g, 1.9 mmol, 1.2 equiv.) and EDC·HCl (0.45 g, 2.3 mmol, 1.5 equiv.) were added sequentially to a solution of compound **3.17** (0.21 g, 1.6 mmol, 1.0 equiv.) in dichloromethane (8 mL, 0.2 M), and the reaction mixture was stirred at room temperature for 3 h. The reaction mixture was diluted with CH_2Cl_2 (20 mL) and the organics washed with 1M HCl (5 mL), saturated aqueous $NaHCO_3$ solution (5 mL) and brine (5 mL). The organic layer was dried over $MgSO_4$, filtered, and concentrated under reduced pressure to afford the title compound (0.16 g, 0.72 mmol, 45%) as an orange liquid. **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H 1.07 (3H, s, Me), 1.80 (2H, t, *J* = 7.9 Hz, -C$\underline{H}_2$CN$_2$Me), 2.51 (2H, t, *J* = 7.9 Hz, -C$\underline{H}_2$CH$_2$CN$_2$Me), 2.83 (4H, s, -NCOC$\underline{H}_2$-); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C 19.51 (Me), 24.77 (-$\underline{C}$N$_2$Me), 25.60 (-NCO$\underline{C}$H$_2$-), 25.75 (-$\underline{C}$H$_2$CH$_2$CN$_2$Me), 29.52 (-$\underline{C}$H$_2$CN$_2$Me), 167.64 (-$\underline{C}$OCH$_2$CH$_2$CN$_2$Me), 169.01 (-N$\underline{C}$OCH$_2$-); **λ_{max}** (ATR, cm^{-1}) 1816, 1781, 1732, 1427, 1415, 1382, 1360, 1300, 1199, 1064, 1046, 993, 891, 809, 747, 642, 601, 579, 556, 454; **mp** 61-63 °C; **HRMS** (ESI$^+$) $C_9H_{11}N_3NaO_4$ $[M+Na]^+$ found 248.0640, calculated 248.0642.

6-(4-(3-(3-methyl-3*H*-diazirin-3-yl)propanoyl)piperazin-1-yl)picolinaldehyde (3.16)



Compound **3.18** (70.0 mg, 0.2 mmol, 1.3 equiv.) and triethylamine (0.11 mL, 0.8 mmol, 5.0 equiv.) were added sequentially to a solution of compound **3.19** (85.0 mg, 0.2 mmol, 1.0 equiv.) in DMF (0.8 mL, 0.2 M). The reaction mixture was stirred at room temperature for 23 h, then concentrated under reduced pressure. The liquid obtained was diluted with CH_2Cl_2 (30 mL), and the organics washed with water (10 mL) and brine (10 mL), dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The resulting brown liquid was purified by flash column chromatography (50% EtOAc in Petrol, R_f 0.22) and the title compound (27 mg, 0.11 mmol, 57%) was obtained as a yellow liquid. **^{1}H NMR** (300 MHz, $CDCl_3$) δ_H 1.07 (3H, s, Me), 1.75-1.93 (2H, m, -C$\underline{H}_2$CN$_2$Me), 2.10-2.23 (2H, m, -C$\underline{H}_2$CH$_2$CN$_2$Me), 3.50-3.64 (4H, m, -C$\underline{H}_2$NR), 3.68-3.88 (4H, m, -C$\underline{H}_2$NR), 6.87 (1H, d, *J* = 8.5 Hz, Ar$\underline{H}$3), 7.33 (1H, d, *J* = 7.2 Hz, Ar$\underline{H}$5), 7.67 (1H, dd, *J* = 8.5, 7.2 Hz, Ar$\underline{H}$4),

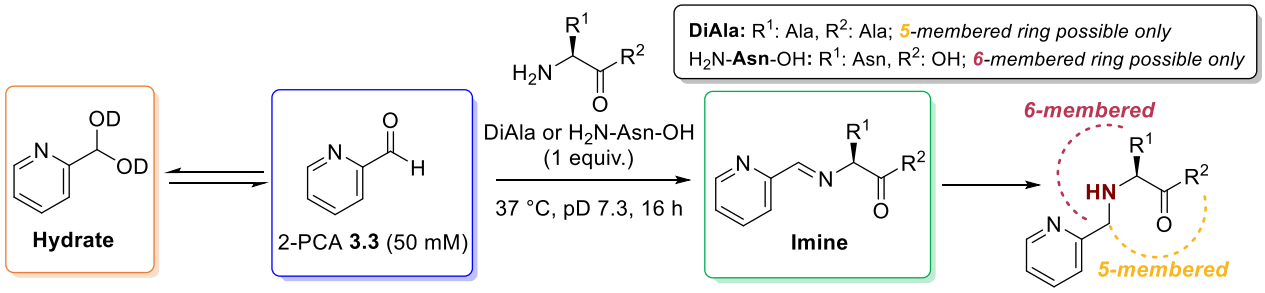
9.91 (1H, s, -CHO); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C 20.24 (Me), 25.66 (-CN_2Me), 27.29 (-$CH_2CH_2CN_2$Me), 29.65 (-CH_2CN_2Me), 41.33 (-CH_2NR), 45.16 (-CH_2NR), 111.40 (ArC3), 112.19 (ArC5), 138.60 (ArC4), 193.94 (CHO) *(note: missing quaternary signals due to high sample dilution)*; **λ_{max}** (ATR, cm^{-1}) 2923, 1709, 1640, 1593, 1440, 1385, 1228, 1023, 984, 794, 747; **HRMS** (ESI^+) $C_{15}H_{19}N_5NaO_2$ $[M+Na]^+$ found 324.1431, calculated 324.1431.

3.4.2 Side chain participation

3.4.2.1 Amino acid modification

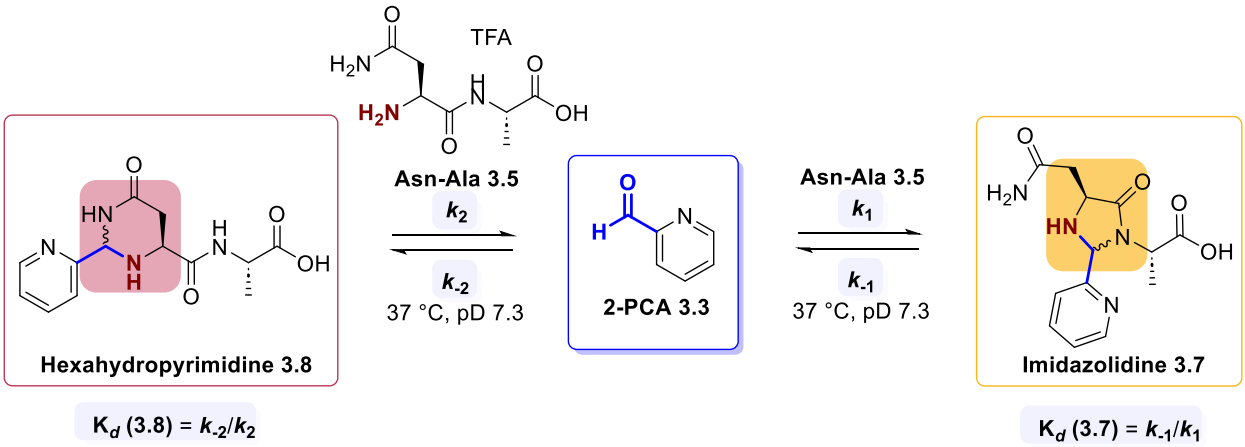
A solution of NH_2-Ala-OH, NH_2-Asn-OH, NH_2-Asp-OH or NH_2-Gln-OH (20 µL, 1 mM, 20 nmol, 1 equiv.) was added to solutions of reagents **3.1**-**3.2** (20 µL, 10 mM, 200 nmol, 10 equiv.), both in sodium phosphate buffer (50 mM, pH 7.5). The reactions were incubated at 37 °C for 23 h and qualitative conversion was determined by the relative integrals of the modified amino acid and excess unmodified PCA in the 210-400 nm UV trace from LC-MS analysis, without purification.

3.4.2.2 Identification of diagnostic ^{1}H NMR signals



A solution of DiAla or H_2N-Asn-OH (150 µL, 100 mM, 15 µmol, 1 equiv.) was added to a solution of reagent **3.3** (150 µL, 100 mM, 15 µmol, 1 equiv.), both in deuterated sodium phosphate buffer (100 mM, pD 7.3). The reaction was incubated at 37 °C for 16 h and the conjugate diastereomeric ratio was determined by ^{1}H NMR spectroscopy. The H_2N-Asn-OH reaction mixture was characterised by ^{1}H and ^{13}C NMR spectroscopy.

3.4.2.3 Kinetic studies



A solution of Asn-Ala **3.5** (150 µL, 100 mM, 15 µmol, 1 equiv.) was added to a solution of reagent **3.3** (150 µL, 100 mM, 15 µmol, 1 equiv.), both in deuterated sodium phosphate buffer (100 mM, pD 7.3). The reaction was incubated at 37 °C for 402 h and conversion was followed by ^{1}H NMR spectroscopy at 30 min intervals for 17 h, and at 44 h, 95 h, 164 h, and 402 h time-points.

At each timepoint ($t = x\ h$), the concentration of aldehyde [**PCA**], hydrate [**Hy**], imine [**Im**], sum of imidazolidinone [**3.7**] and sum of hexahydropyrimidine [**3.8**] species were calculated *via* the relative integral ratios of diagnostic ^{1}H NMR signals outlined in *Fig. S3.3*. Concentration values were normalised using a correction factor ($\mathbf{CF_1}$ and $\mathbf{CF_2}$) in *Equation 3.1* and *Equation 3.2* to remove background noise from the imidazolidinone and hexahydropyrimidine signals respectively, based on the assumption that the initial rate was linear over the first 3 time-points collected, and that $[\mathbf{3.7}]_{t=0\ h} = 0$ and $[\mathbf{3.8}]_{t=0\ h} = 0$. Correction factors were calculated as the y-intercept of the linear regression line of the plot of the sum of integrals of product diastereomers over time (for the first 3 time-points collected), as demonstrated in *Fig. S3.1*.

Imidazolidinone Conversion *(Equation 3.1)*

$$= \mathbf{100} \times \frac{\int \mathbf{3.7} - \mathbf{CF_1}}{\int \mathbf{PCA} + \int \mathbf{Hy} + \int \mathbf{Im} + \int \mathbf{3.7} + \int \mathbf{3.8} - \mathbf{CF_1} - \mathbf{CF_2}}$$

Hexahydropyrimidine Conversion *(Equation 3.2)*

$$= \mathbf{100} \times \frac{\int \mathbf{3.8} - \mathbf{CF_2}}{\int \mathbf{PCA} + \int \mathbf{Hy} + \int \mathbf{Im} + \int \mathbf{3.7} + \int \mathbf{3.8} - \mathbf{CF_1} - \mathbf{CF_2}}$$

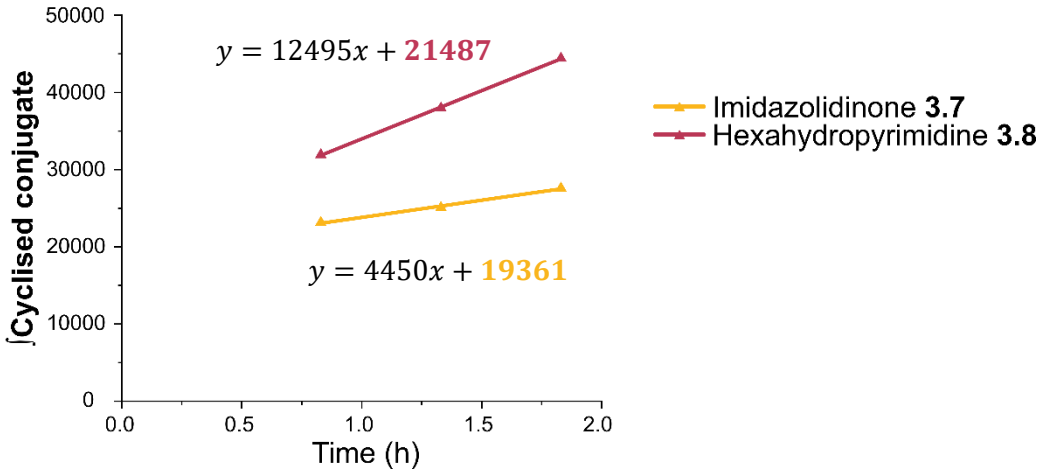


Figure S3.1. Calculation of correction factors for imidazolidinone (CF_1) and hexahydropyrimidine (CF_2) formation. Fit is a linear regression, based on the initial rates model.

Data were fit to a second order irreversible or reversible kinetic model in Copasi 4.34.251. A simplified model was used;

A + B = C; and

A + B = D;

in which **A** represented the overall concentration of the DiAla [**DiAla**] and imine [**Im**] species, **B** represented the overall concentration of aldehyde [**PCA**], hydrate [**Hy**], and imine [**Im**] forms of reagent **3.3**, **C** represented the overall concentration of imidazolidinone [**3.7**] conjugate, and **D** represented the overall concentration of hexahydropyrimidine [**3.8**] conjugate. The model did not integrate hydrate and imine formation, and relied on the assumption that imidazolidinone/hexahydropyrimidine formation was the rate-limiting step of the reaction. k_1, k_{-1}, k_2, and k_{-2} 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}-10^6 M^{-1} h^{-1}; k_{-1} 10^{-6}-10^6 M^{-1} h^{-1}.

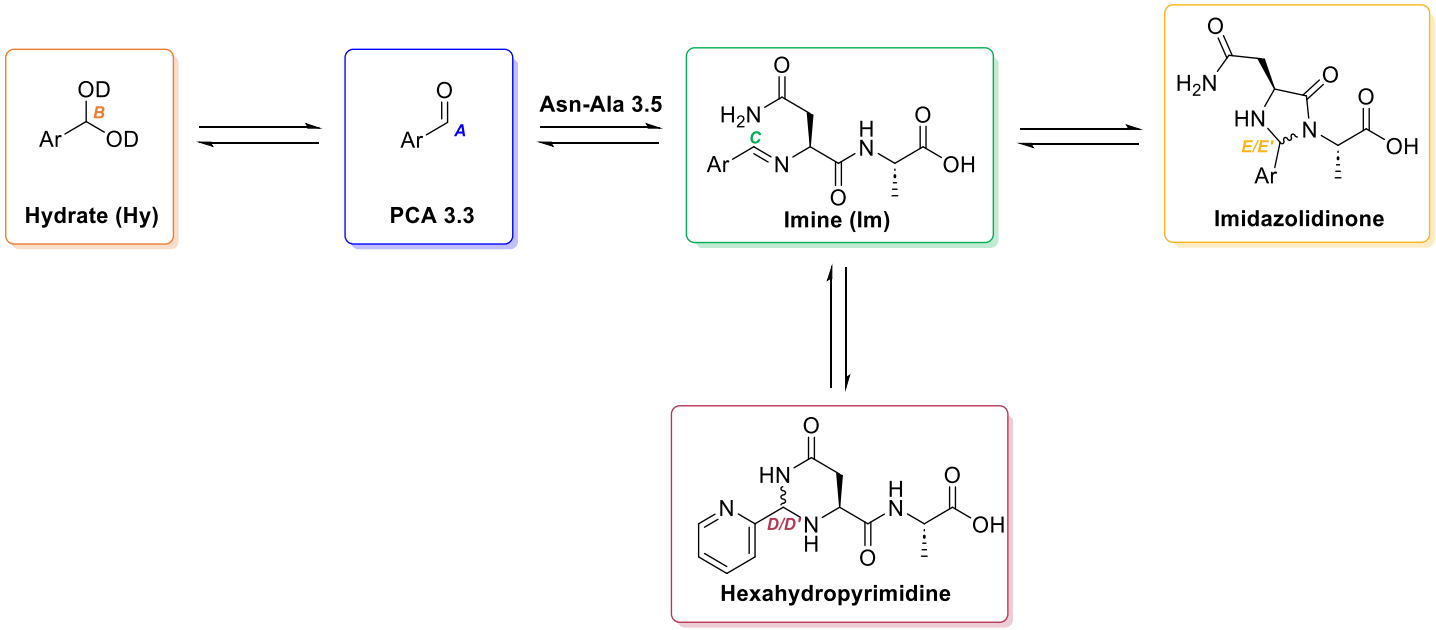


Figure S3.2. Overview of the diagnostic protons used to calculate imidazolidinone and hexahydropyrimidine rate constants.

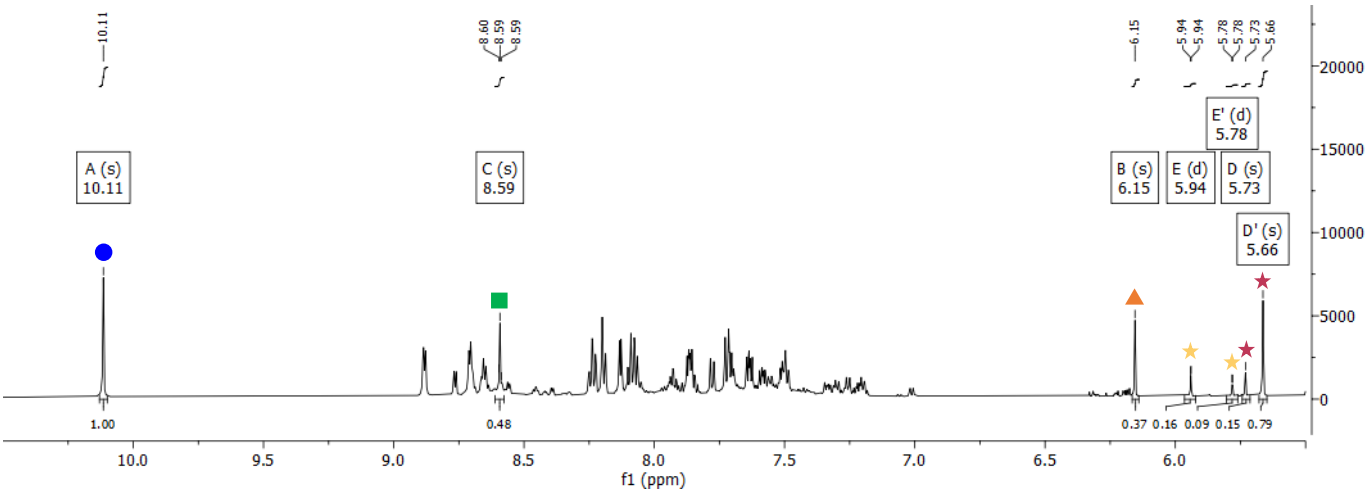


Figure S3.3. Representative ^{1}H NMR spectrum of kinetic experiment at t = 17 h, showing the diagnostic signals used to calculate imidazolidinone and hexahydropyrimidine rate constants (**triangle** = hydrate species, **circle** = aldehyde species, **square** = imine species, **star** = imidazolidinone species, **star** = hexahydropyrimidine species).

3.4.3 N-methylated termini

3.4.3.1 Kinetic studies

A solution of DiAla or N-methylated DiAla **3.10** (150 µL, 100 mM, 15 µmol, 1 equiv.) was added to a solution of reagent **3.2** or **3.3** (150 µL, 100 mM, 15 µmol, 1 equiv.), both in deuterated sodium phosphate buffer (100 mM, pD 7.3). The reaction was incubated at 37 °C for 17 h and conversion was followed by ^{1}H NMR spectroscopy at 30 min intervals for 17 h.

At each timepoint ($t = x\ h$), the concentration of aldehyde [**PCA**], hydrate [**Hy**], imine [**Im**], and sum of imidazolidinone product [**Pr**] species were calculated *via* the relative integral ratios of diagnostic ^{1}H NMR signals outlined in *Fig. S3.6*. Concentration values were normalised using a correction factor (**CF**) in *Equation 3.3* to remove background noise from the imidazolidinone signals, based on the assumption that the initial rate was linear over the time-points indicated in *Fig. S3.4*, and that $[\mathbf{Pr}]_{t=0\ h} = 0$. Correction factors were calculated as the y-intercept of the linear regression line of the plot of the sum of integrals of product diastereomers over time, as demonstrated in *Fig. S3.4*.

$$\textbf{Imidazolidinone Conversion} = 100 \times \frac{\int \mathbf{Pr} - \mathbf{CF}}{\int \mathbf{PCA} + \int \mathbf{Hy} + \int \mathbf{Im} + \int \mathbf{Pr} - \mathbf{CF}} \qquad \textit{(Equation 3.3)}$$

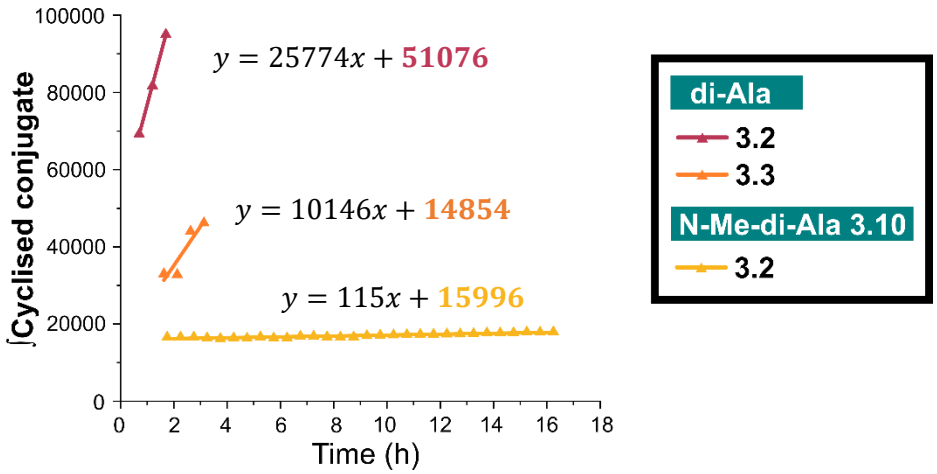


Figure S3.4. Calculation of correction factor (**CF**) for imidazolidinone formation. Fit is a linear regression, based on the initial rates model. The number of time-points included was selected for each series individually in attempt to negate influences from outlier points.

Data were fit to a second order irreversible or reversible kinetic model in Copasi 4.34.251. A simplified model was used;

A + B = C;

in which **A** represented the overall concentration of the DiAla [**DiAla**] and imine [**Im**] species, **B** represented the overall concentration of aldehyde [**PCA**], hydrate [**Hy**], and imine [**Im**] forms of reagent **3.2**/**3.3**, and **C** represented the overall concentration of imidazolidinone [**Pr**] conjugate. The model did not integrate hydrate and imine formation, and relied on the assumption that imidazolidinone formation was the rate-limiting step of the reaction. k_1 and k_{-1} 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}-10^6 M^{-1} h^{-1}; k_{-1} 10^{-6}-10^6 M^{-1} h^{-1}.

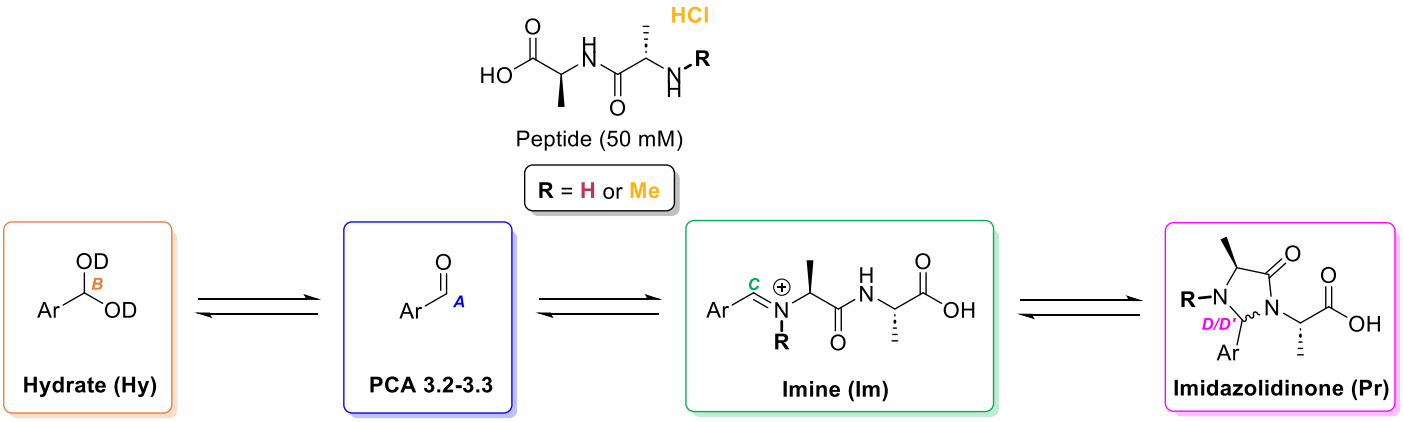
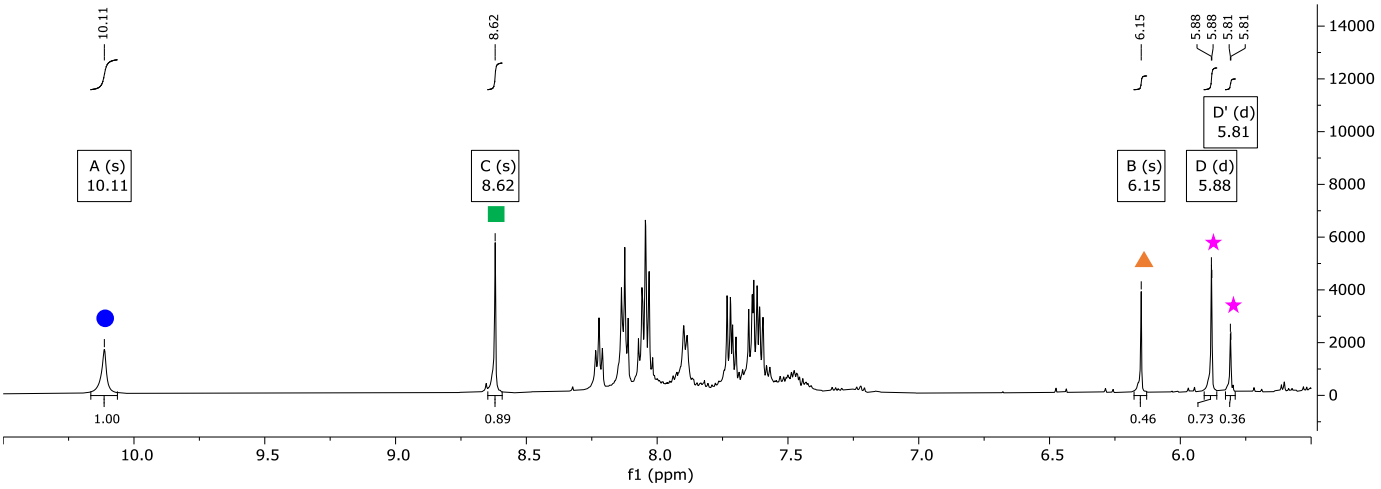
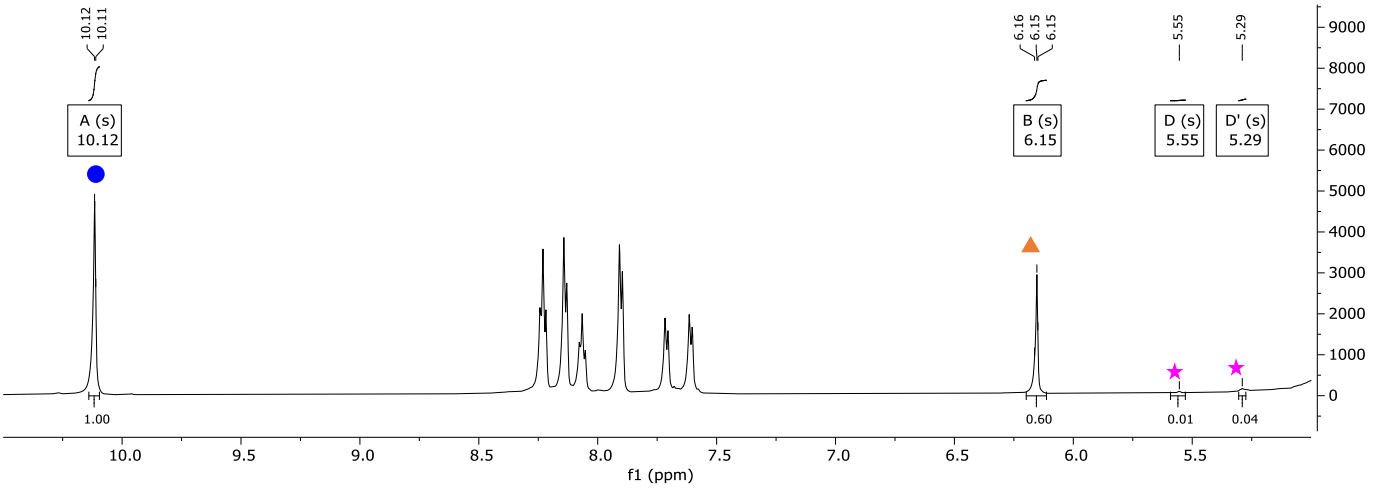


Figure S3.5. Overview of the diagnostic protons used to calculate imidazolidinone rate constants.

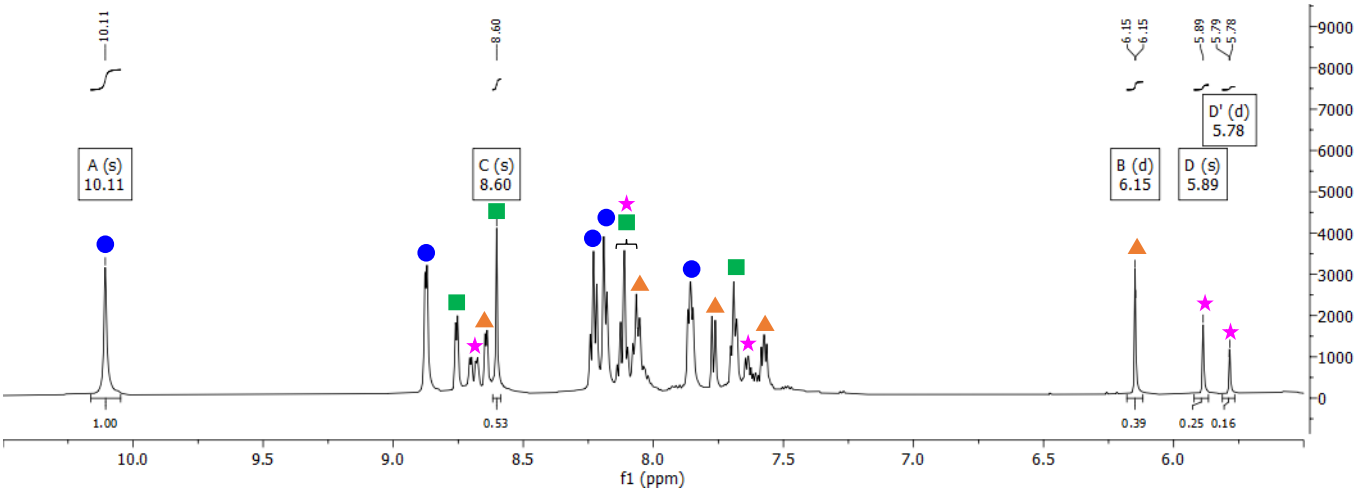
a) PCA **3.2**, R = **H** *(t = 16 h)*



b) PCA **3.2**, R = **Me** *(t = 16 h)*



c) PCA **3.3**, R = **H** *(t = 17 h)*



d) PCA **3.3**, R = **Me** *(t = 17 h)*

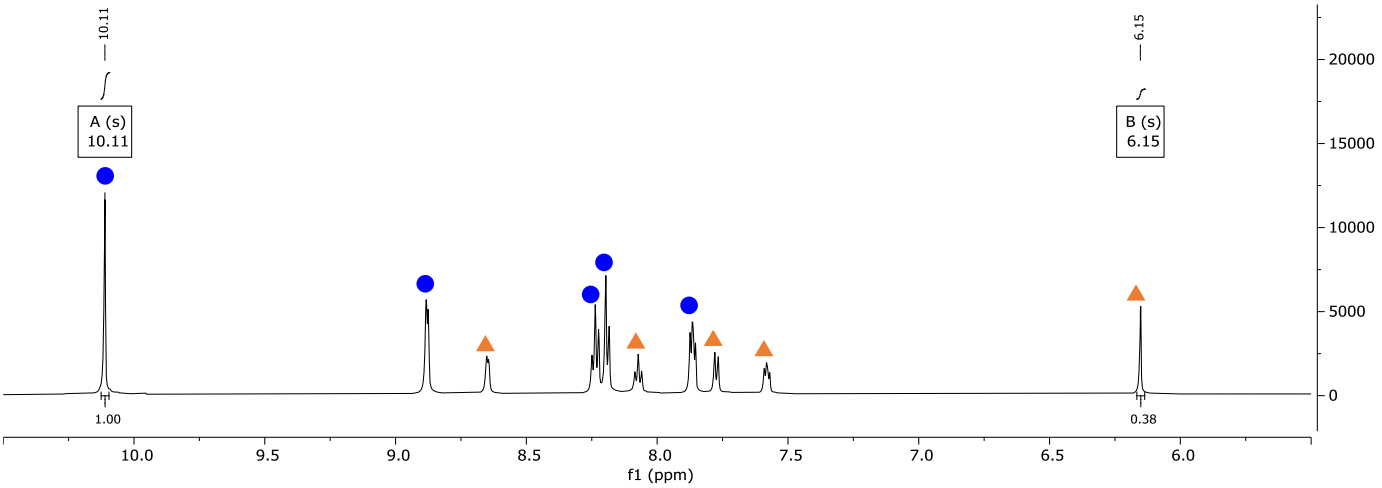
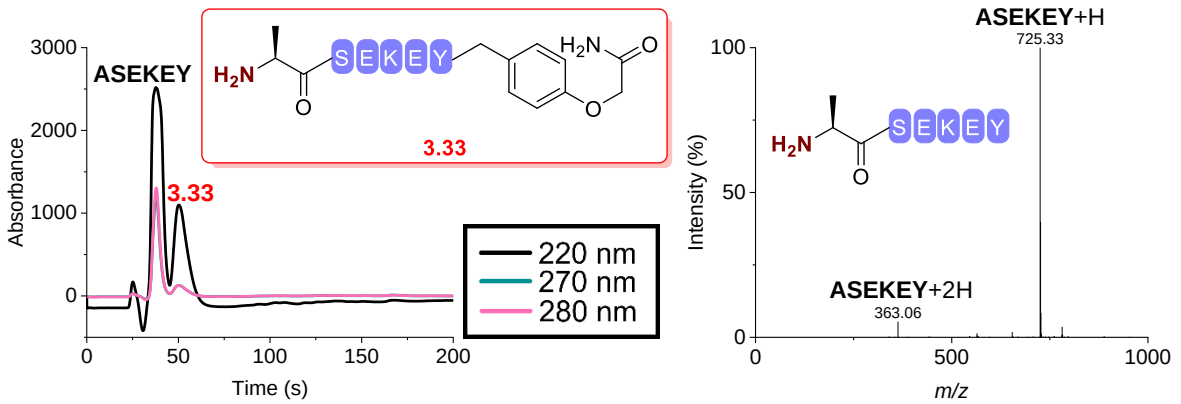


Figure S3.6. Representative ^{1}H NMR spectrum of kinetic experiment, showing the diagnostic signals used to calculate imidazolidinone rate constants (**triangle** = hydrate species, **circle** = aldehyde species, **square** = imine species, **star** = imidazolidinone species).

3.4.3.2 Peptide synthesis

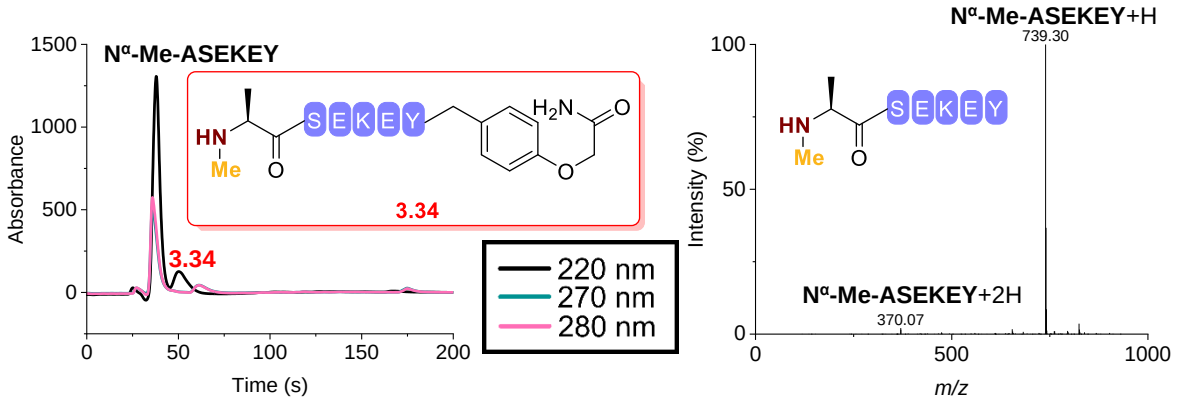
Solid-phase peptide synthesis (SPPS) was performed on a CEM Liberty Lite Automated Microwave Peptide Synthesiser, according to the manufacturer's standard protocols. Briefly, Fmoc-protected amino acids (5 equiv., 0.2 M in DMF) were coupled in the presence of *N,N'*-diisopropylcarbodiimide (DIC, 5 equiv., 1.0 M in DMF) and Oxyma Pure (5 equiv., 1.0 M in DMF), as coupling agent and base respectively, under microwave irradiation at a temperature of 90 °C for 2 minutes. Fmoc deprotection was performed using 20% piperidine in DMF at 90 °C for 60 seconds. Syntheses were performed on a 0.1 mmol scale, using Rink Amide MBHA resin (*C*-terminal amide, 0.5 mmol/g, 1% DVB, 100-200 mesh, Fluorochem). Prior to cleavage, the resin was washed sequentially with CH_2Cl_2 (3 x 15 mL) and methanol (3 x 15 mL). Peptides were cleaved from the resin in 20 mL of cleavage cocktail (90% TFA, 5% H_2O, 3% TIPS) for 4 hrs. After filtration, the resin was

washed extensively with CH_2Cl_2 (3 x 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, 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.



ASEKEY

MS (ESI$^+$) [**ASEKEY**+2H]$^{2+}$ found 363.06, calculated 363.37; [**ASEKEY**+H]$^+$ found 725.33, calculated 725.74.



Nα-Me-ASEKEY *(24 mg, 0.03 mmol, 24%)*

MS (ESI$^+$) [**Nα-Me-ASEKEY**+2H]$^{2+}$ found 370.07, calculated 370.18; [**Nα-Me-ASEKEY**+H]$^+$ found 739.30, calculated 739.36.

Figure S3.7. UV-vis 220, 270 and 280 nm traces, and mass spectra of ASEKEY and Nα-Me-ASEKEY. Note the impurities observed at approx. 50 s were due to "mis-cleavage" of the Rink Amide MBHA resin to afford the *C*-terminal-alkylated peptides (*m/z* +163 Da, **3.33** and **3.34**).[24,25]

3.4.3.3 Peptide modification

A solution of ASEKEY or Nα-Me-ASEKEY (15 µL, 1 mM, 15 nmol, 1 equiv.) was added to solutions of reagents **3.2**-**3.3** or **3.12**-**3.15** (15 µL, 10 mM, 150 nmol, 10 equiv.), both in sodium phosphate buffer (50 mM, pH 7.5). For compound **3.15**, 5 equiv. of sodium cyanoborohydride in 1.5 µL of water was also added, and compound **3.15** was added as a more concentrated solution (13.5 µL, 11 mM) to keep the total volume constant. The reactions were incubated at 37 °C for 23 h and qualitative conversion was determined by the relative integrals of the unmodified/modified peptide in the 280 nm UV trace from LC-MS analysis, without purification.

Reaction mixtures of reagents **3.2** and **3.3** were analysed with a slower column gradient than all other samples due to co-elution between unmodified peptide, modified peptide, and reagent **3.2**/**3.3** (*Fig. S3.8*).

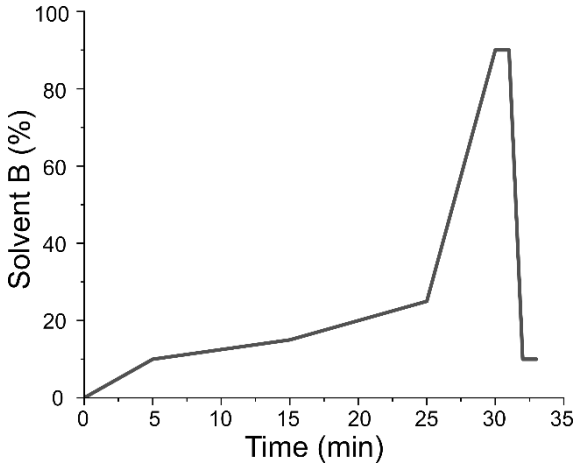
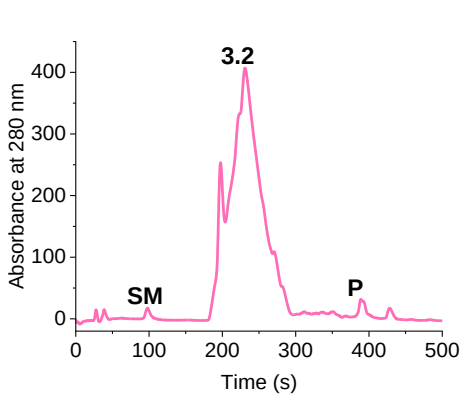
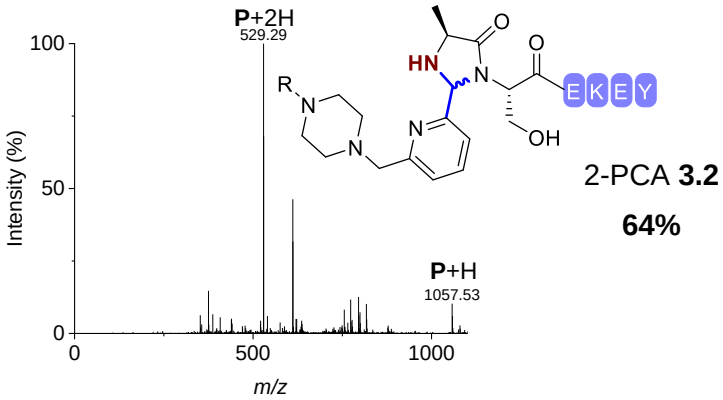


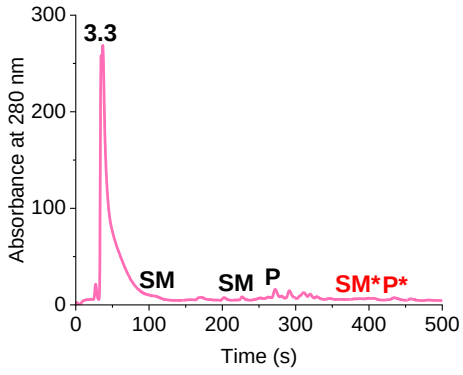
Figure S3.8. LC-MS gradient for the analysis of reaction mixtures of reagents **3.2** and **3.3**.

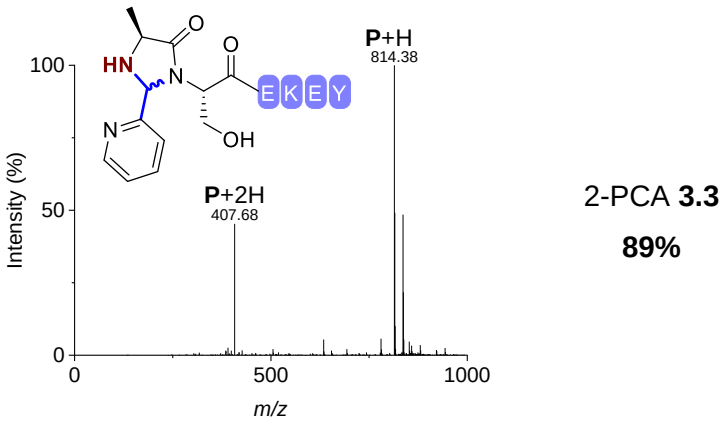
ASEKEY



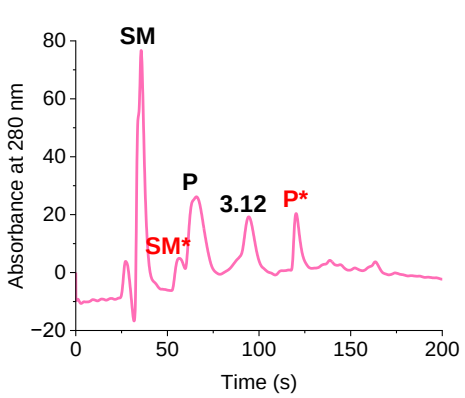


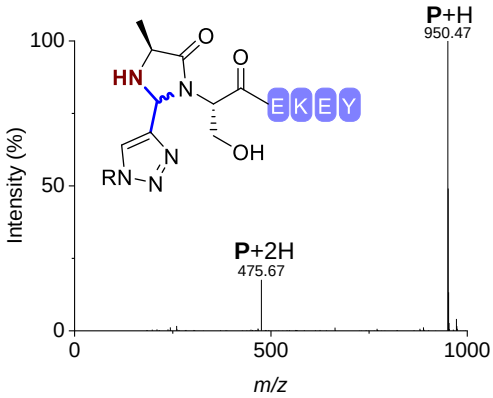
MS (ESI$^+$) [**P**+2H]$^{2+}$ found 529.29, calculated 529.78; [**P**+H]$^+$ found 1057.53, calculated 1058.55.





MS (ESI$^+$) [**P**+2H]$^{2+}$ found 407.68, calculated 407.69; [**P**+H]$^+$ found 814.38, calculated 814.38.

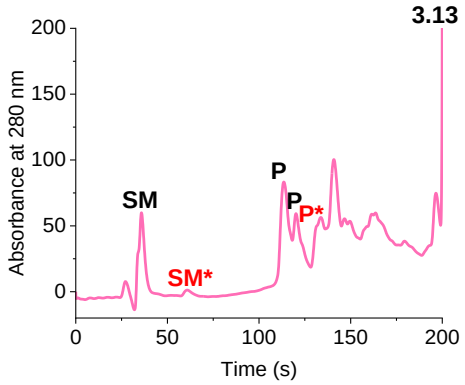


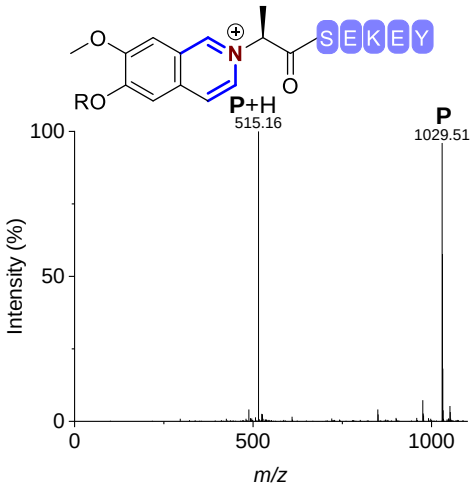


TA4C **3.12**

43%

MS (ESI$^+$) [**P**+2H]$^{2+}$ found 475.67, calculated 475.73; [**P**+H]$^+$ found 950.47, calculated 950.46.

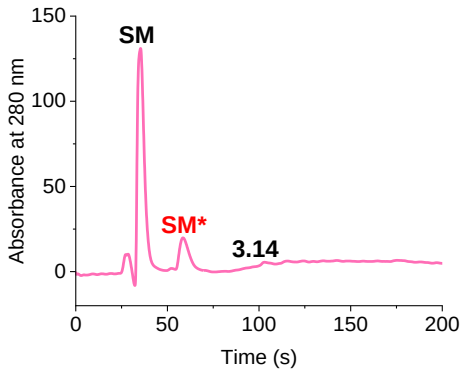




2-EBA **3.13**

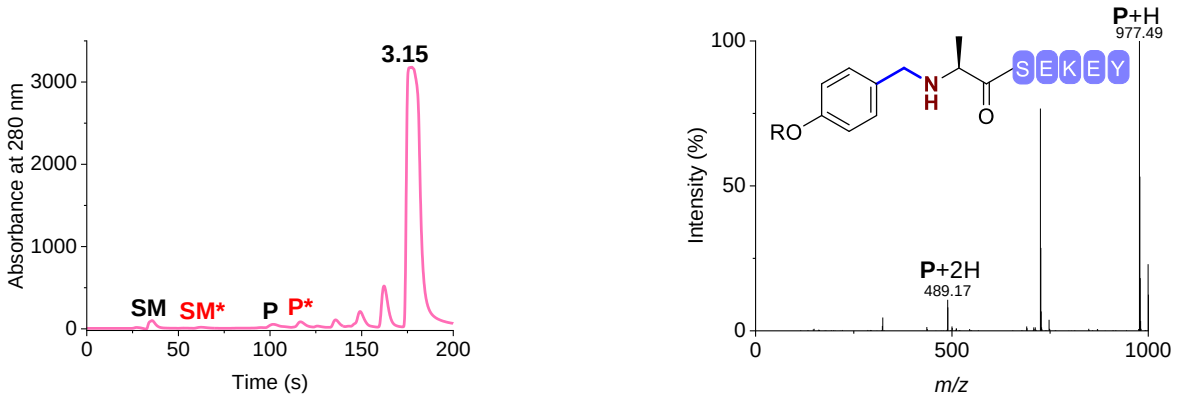
73%

MS (ESI$^+$) [**P**+H]$^{2+}$ found 515.16, calculated 515.24; [**P**]$^+$ found 1029.51, calculated 1029.48.



Ox **3.14**

0%



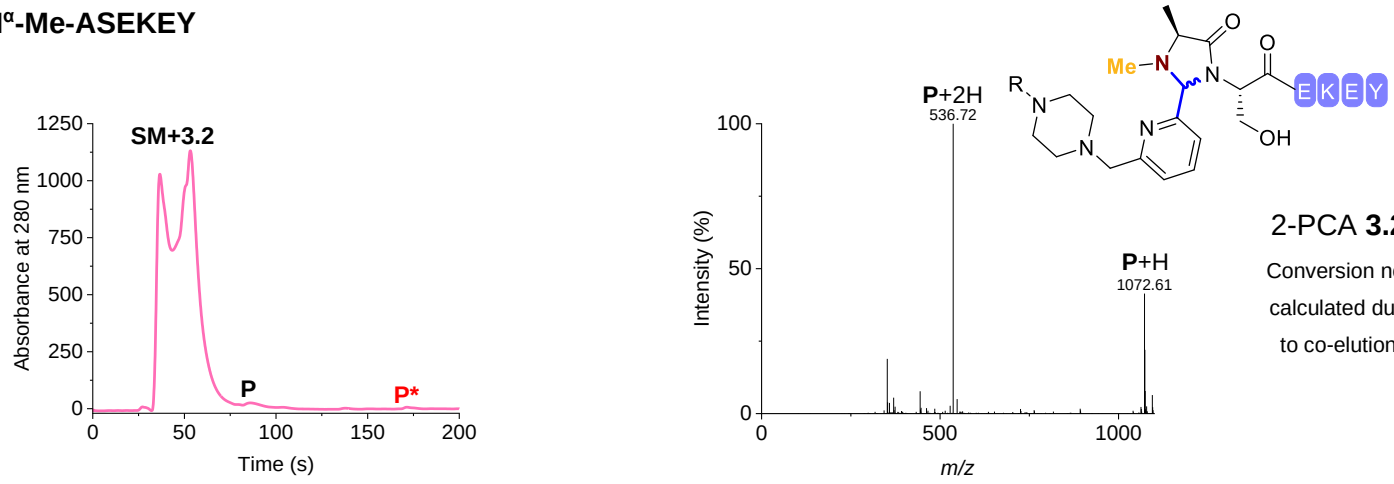
BA **3.15**

39%

MS (ESI$^+$) [**P**+2H]$^{2+}$ found 489.17, calculated 489.25; [**P**+H]$^+$ found 977.49, calculated 977.49.

Figure S3.9. UV-vis 280 nm trace and conjugate mass spectra for the modification of ASEKEY with reagents **3.2-3.3** and **3.12**-**3.15** (**SM** = unmodified ASEKEY, **P** = modified ASEKEY, * = "mis-cleavage" of the Rink Amide MBHA resin. Conversion values reported do not include mis-cleavage side products).

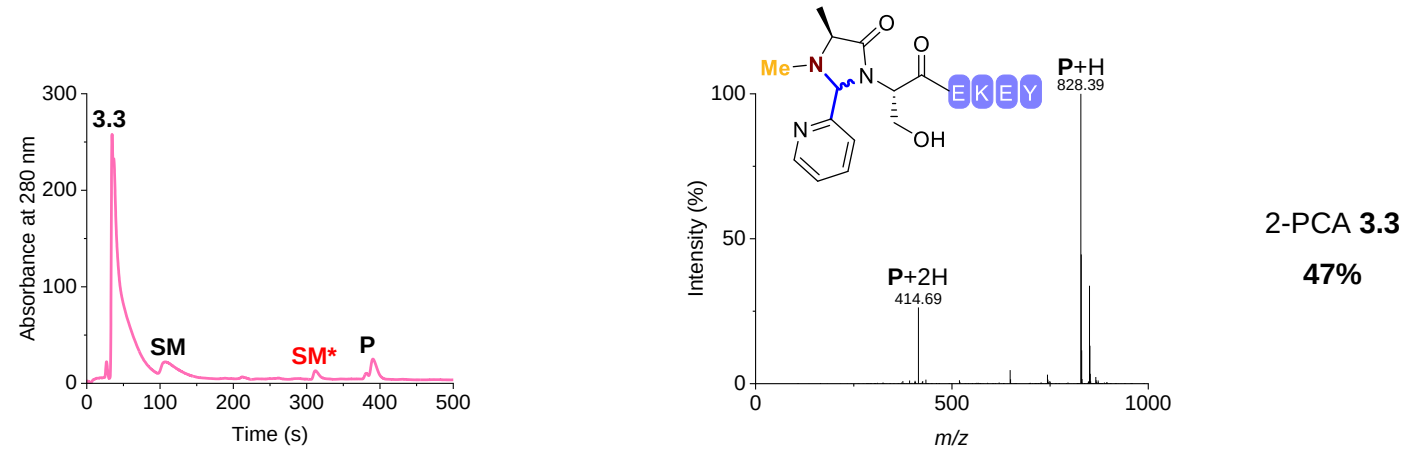
N$^{\alpha}$-Me-ASEKEY



2-PCA **3.2**

Conversion not calculated due to co-elution

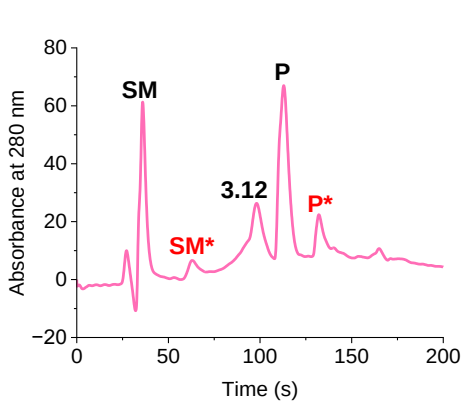
MS (ESI$^+$) [**P**+2H]$^{2+}$ found 536.72, calculated 536.79; [**P**+H]$^+$ found 1072.61, calculated 1072.57.

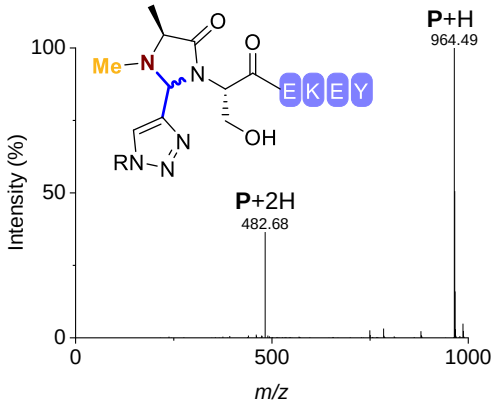


2-PCA **3.3**

47%

MS (ESI$^+$) [**P**+2H]$^{2+}$ found 414.69, calculated 414.70; [**P**+H]$^+$ found 828.39, calculated 828.39.

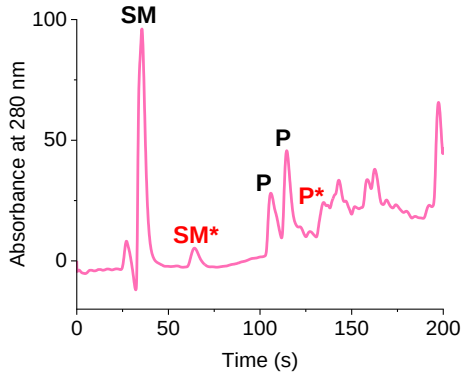


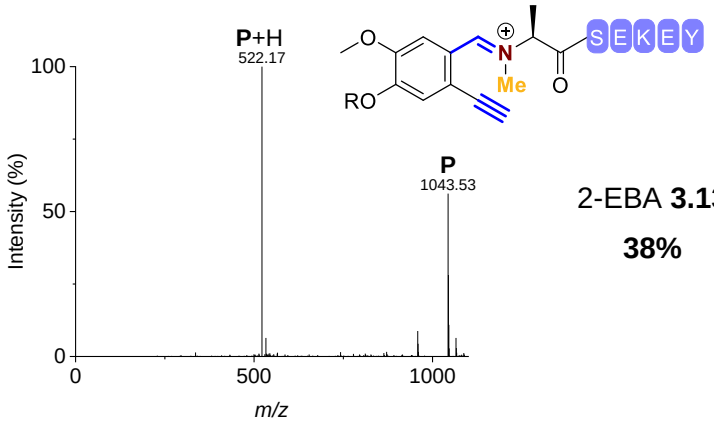


TA4C **3.12**

60%

MS (ESI$^+$) [**P**+2H]$^{2+}$ found 482.68, calculated 482.74; [**P**+H]$^+$ found 964.49, calculated 964.48.

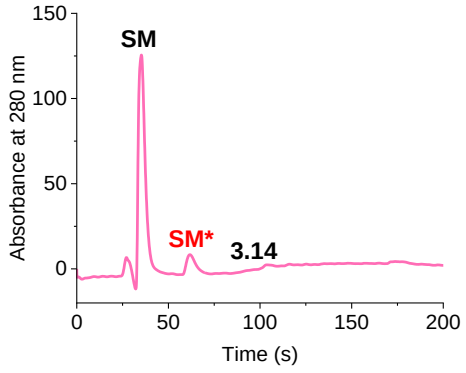




2-EBA **3.13**

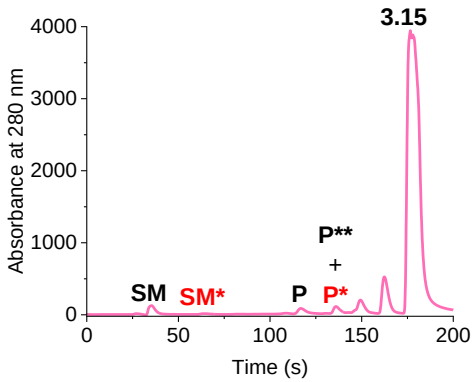
38%

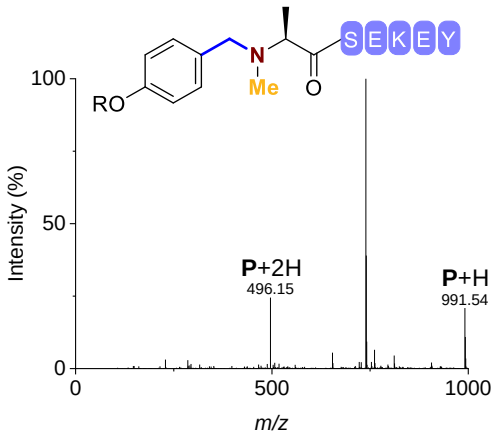
MS (ESI$^+$) [**P**+H]$^{2+}$ found 522.17, calculated 522.25; [**P**]$^+$ found 1043.53, calculated 1043.49.



Ox **3.14**

0%





BA **3.15**

38%

MS (ESI$^+$) [**P**+2H]$^{2+}$ found 496.15, calculated 496.25; [**P**+H]$^+$ found 991.54, calculated 991.50.

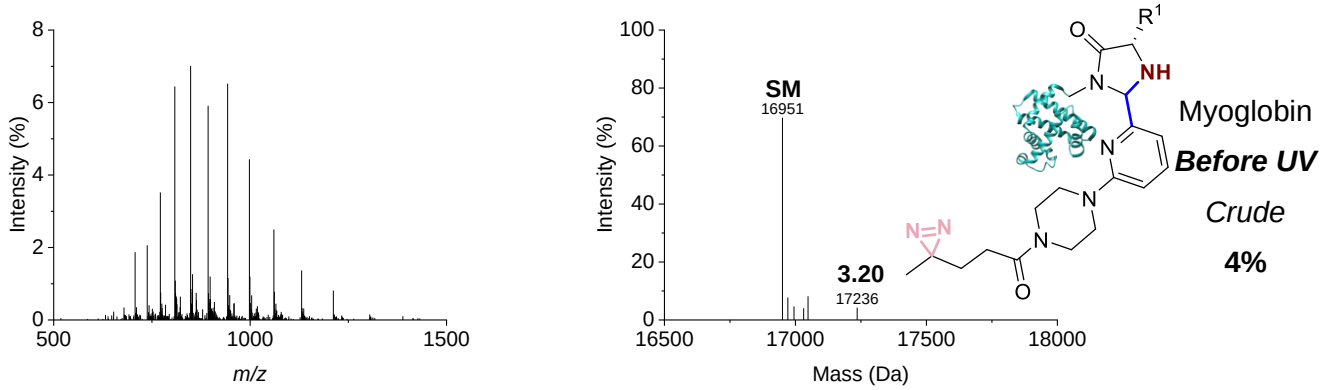
Figure S3.10. UV-vis 280 nm trace and conjugate mass spectra for the modification of N$^\alpha$-Me-ASEKEY with reagents **3.2-3.3** and **3.12-3.15** (**SM** = unmodified N$^\alpha$-Me-ASEKEY, **P** = modified N$^\alpha$-Me-ASEKEY, * = "mis-cleavage" of the Rink Amide MBHA resin; ** = imine (not reduced). Conversion values reported do not include mis-cleavage side products).

3.4.4 Proximity-driven chemistries

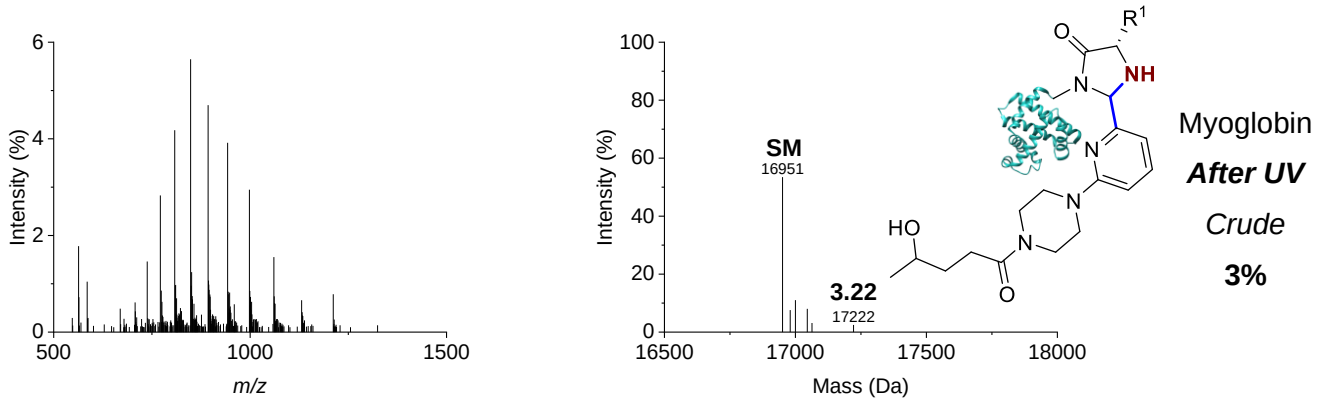
3.4.4.1 Ligand-directed strategy: Diazirines

Procedure 3A: *Modification with reagents under conditions adapted from MacDonald et al.*[6] A stock solution of the modification reagent (60 µL, 10 mM, 600 nmol, 100 equiv., in 50 mM pH 7.5 sodium phosphate buffer) was added to a solution of protein (60 µL, 100 µM, 6 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer), and the mixture incubated at 37 °C for 23 h with agitation (1000 rpm).

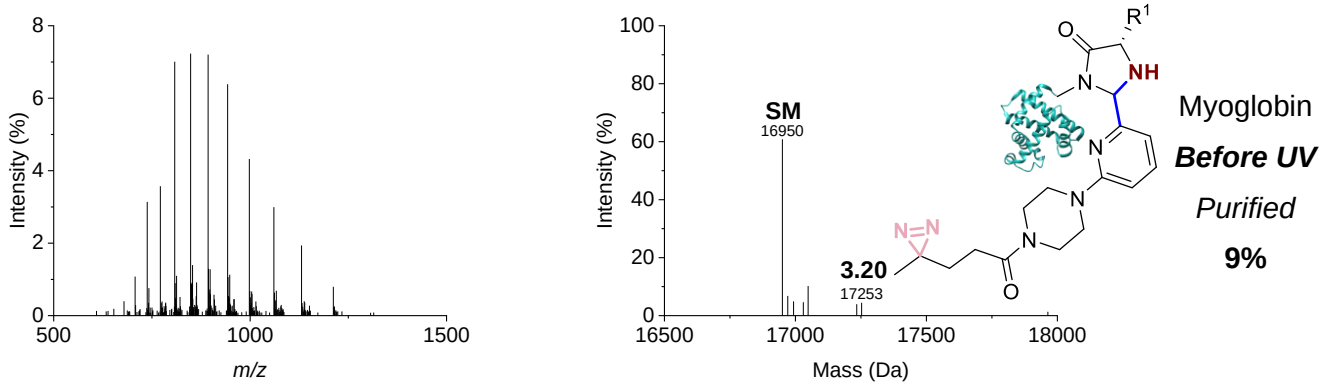
Reagent **3.16** was used to modify myoglobin and RNase A under the conditions outlined in **General Procedure 3A**, and conversion was determined by LC-MS analysis of the reaction mixture. Aliquots of the protein conjugates (70 µL) were purified by dialysis to remove excess reagent (4 °C, 3.5 kDa MWCO; 1 × water, 6 h; 1 × water, 16 h; 1 × water, 4 h), and conversion of the purified conjugates was determined by LC-MS analysis. Crude and purified protein conjugates (30 µL) were exposed to UV light at rt for 15 min, and conversion was determined by LC-MS analysis.



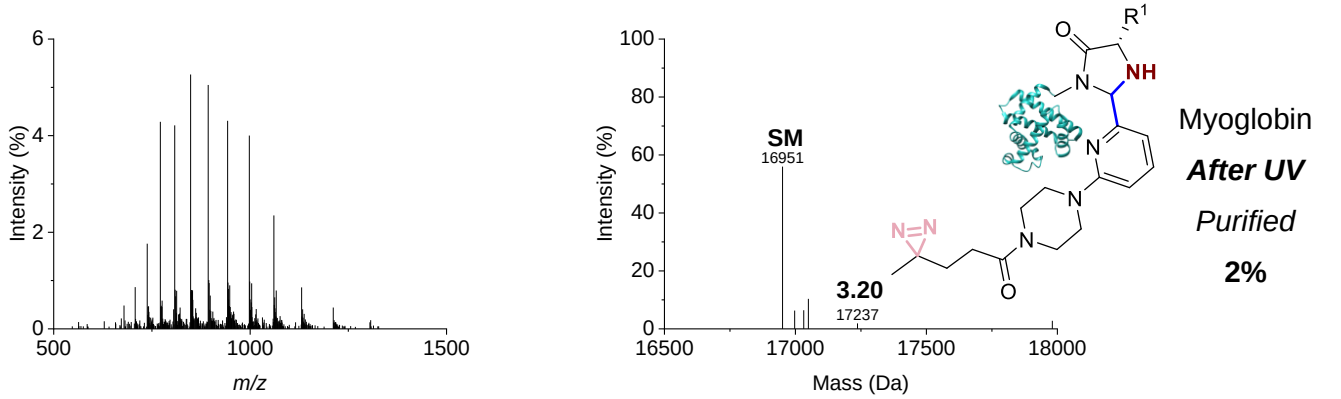
MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951;[26] [**SM**+H_2O+H]$^+$ found 16971, calculated 16970; [**SM**+MeCN+H]$^+$ found 16995, calculated 16993; [**SM**+DMSO+H]$^+$ found 17031, calculated 17029; [**SM**(s)+H_3PO_4+H]$^+$ found 17048, calculated 17049; [**3.20**+H]$^+$ found 17236, calculated 17234.



MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951;[26] [**SM**+MeOH+H]$^+$ found 16980, calculated 16983; [**SM**+MeCN+H]$^+$ found 17000, calculated 16993; [**SM**(s)+H_3PO_4+H]$^+$ found 17045, calculated 17049; [**3.22**+H]$^+$ found 17222, calculated 17224.



MS (ESI$^+$) [**SM**+H]$^+$ found 16950, calculated 16951;[26] [**SM**+H_2O+H]$^+$ found 16971, calculated 16970; [**SM**+MeCN+H]$^+$ found 16993, calculated 16993; [**SM**+DMSO+H]$^+$ found 17030, calculated 17029; [**SM**(s)+H_3PO_4+H]$^+$ found 17048, calculated 17049; [**3.20**+H]$^+$ found 17234, calculated 17234; [**3.20**+H_2O+H]$^+$ found 17253, calculated 17252.



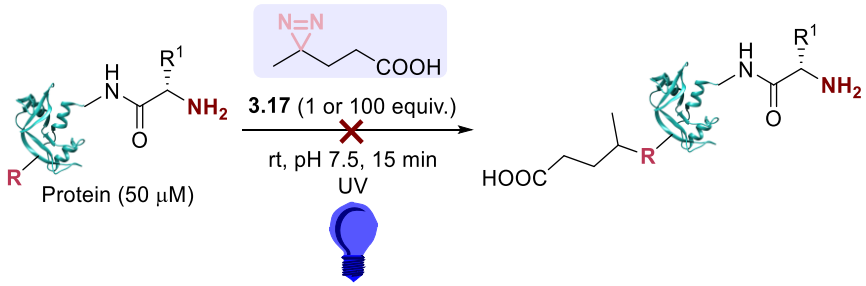
MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951;[26] [**SM**+MeCN+H]$^+$ found 16997, calculated 16993; [**SM**+DMSO+H]$^+$ found 17032, calculated 17029; [**SM**(s)+H_3PO_4+H]$^+$ found 17049, calculated 17049; [**3.20**+H]$^+$ found 17237, calculated 17234.

Figure S3.11. Raw (left) and deconvoluted (right) mass spectra for myoglobin conjugates. **SM** = unmodified protein.

Purification	**Conversion (%)**	
	Before UV irradiation	*After UV irradiation*
Crude	**4** (conjugate **3.20**)	**3** (conjugate **3.22**)
Purified	**9** (conjugate **3.20**)	**2** (conjugate **3.20**)

Table S3.1. Ligand-directed 2-PCA/diazirine chemistry for the modification of proteins. Myoglobin was modified with reagent **3.16** under conditions outlined in General Procedure **3A**, and the conversion was determined by LC-MS before and after purification by dialysis, and after UV irradiation for 15 min.

Control to verify if diazirine reactivity with the protein could only occur by proximity labelling. A stock solution of reagent **3.17** (15 µL, 10 mM or 100 µM, 150 or 1.5 nmol, 100 or 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer) was added to a solution of myoglobin or RNase A (15 µL, 100 µM, 1.5 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer), and the mixture exposed to UV light at rt for 15 min, and conversion was determined by LC-MS analysis. No non-specific reactivity between the proteins and the diazirine moiety was observed.



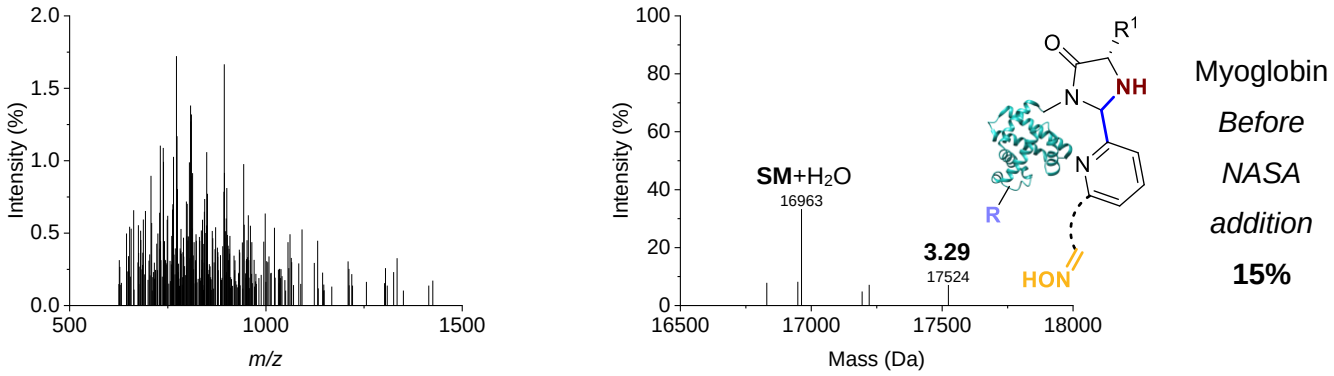
3.4.4.2 Initial SDS-PAGE for AGOX strategy

Reagent **3.23** was used to modify myoglobin and RNase A under the conditions outlined in **General Procedure 3A**, on a 20 µL scale. Control samples were prepared as detailed above, with 50 mM pH 7.5 sodium phosphate buffer (10 µL) in lieu of the solution of modification reagent. The reaction mixture was centrifuged at 1000 rpm for 10 s, and a 5.9 µL sample (myoglobin) or 3.7 µL sample (RNase A) was taken for SDS-PAGE (soluble). Pellets were reconstituted in PBS; control, soluble, and precipitate samples were analysed by 15% SDS-PAGE gels.

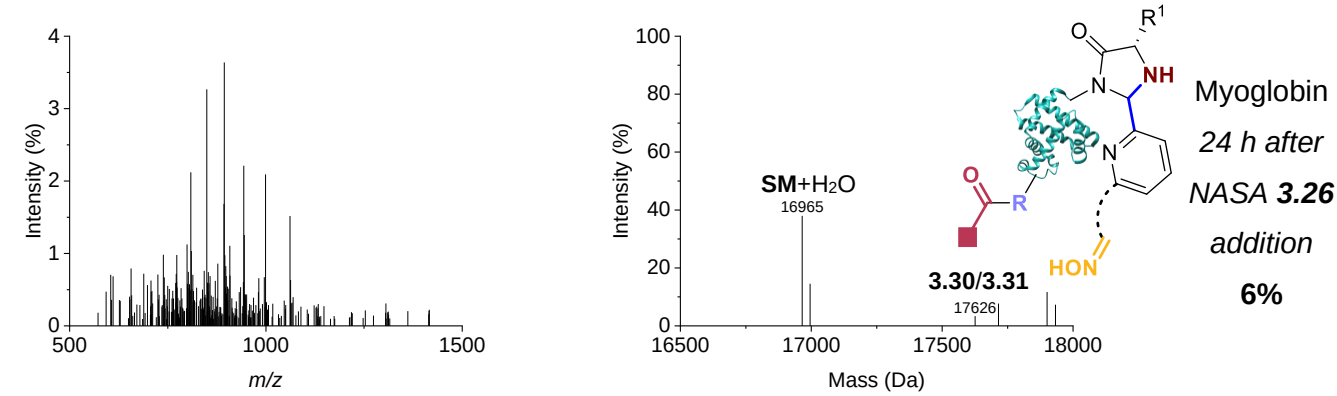
Reagent **3.24** was used to modify myoglobin and RNase A under the conditions outlined in **General Procedure 3A**, on a 75 µL scale. Control samples were prepared as detailed above, with 50 mM pH 7.5 sodium phosphate buffer (37.5 µL) in lieu of the solution of modification reagent. A 5.9 µL sample (myoglobin) or 7.3 µL sample (RNase A) was taken for SDS-PAGE (soluble). Control and soluble samples were analysed by 15% SDS-PAGE gels.

*3.4.4.3 Initial screening for AGOX strategy with 2-PCA **3.24***

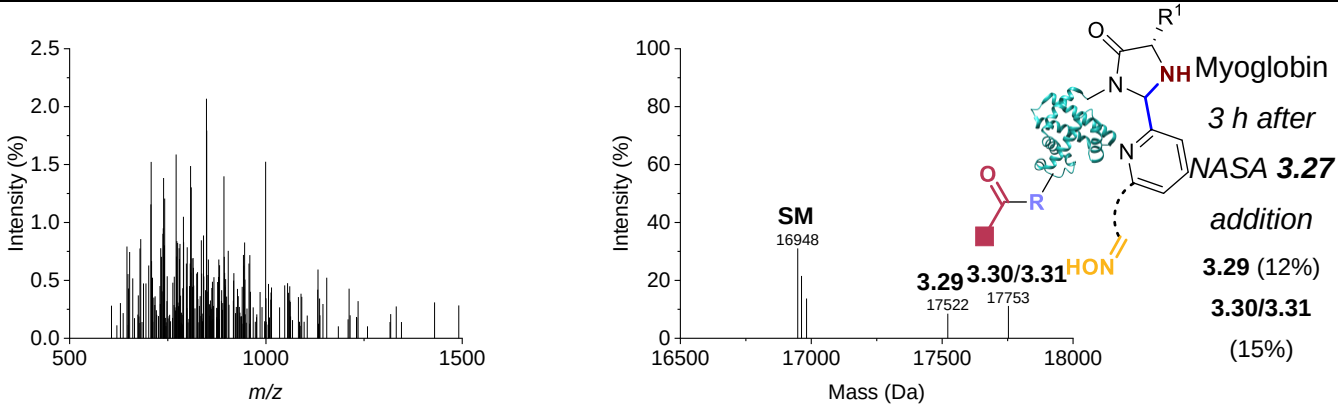
Reagent **3.24** was used to modify myoglobin, RNase A, insulin, and CjX183-D WT under the conditions outlined in **General Procedure 3A**, on an 80 µL scale. Conversion was determined by LC-MS analysis of the reaction mixture. A stock solution of NASA reagent **3.25**-**3.28** (17.5 µL, 500 µM, 8.75 nmol, 10 equiv., in 10% DMSO in 50 mM pH 7.5 sodium phosphate buffer) was added to aliquots of the above reaction mixture (17.5 µL), and the mixture incubated at 37 °C for 72 h with agitation (1000 rpm). Conversion was determined by LC-MS analysis of the reaction mixture 3 h, 24 h, and 72 h after NASA reagent addition.



MS (ESI$^+$) [**SM**+H]$^+$ found 16949, calculated 16951;[26] [**SM**+H_2O+H]$^+$ found 16963, calculated 16970; [**3.29**+H_2O]$^+$ found 17524, calculated 17525.



MS (ESI$^+$) [**SM**+H_2O+H]$^+$ found 16965, calculated 16970; [**SM**+MeCN+H]$^+$ found 16996, calculated 16993; [**3.30**/**3.31**+H_2O]$^+$ found 17626, calculated 17623.



MS (ESI$^+$) [**SM**+H]$^+$ found 16948, calculated 16951;[26] [**SM**+H_2O+H]$^+$ found 16963, calculated 16970; [**SM**+MeOH+H]$^+$ found 16982, calculated 16983; [**3.29**+H_2O]$^+$ found 17522, calculated 17525; [**3.30**/**3.31**+H_2O]$^+$ found 17753, calculated 17751.

Figure S3.12. Examples of raw (left) and deconvoluted (right) mass spectra for myoglobin conjugates. **SM** = unmodified protein.

NASA Reagent	Conversion to conjugate 3.29 (%)				Conversion to conjugate 3.30/3.31 (%)			
	Before	*3 h*	*24 h*	*72 h*	*Before*	*3 h*	*24 h*	*72 h*
3.25	**15**	**26**	**12**	**14**	**-**	**0**	**0**	**0**
3.26		**26**	**0**	**17**		**0**	**6**	**0**
3.27		**12**	**9**	**6**		**15**	**14**	**0**
3.28		**0**	**0**	**13**		**0**	**0**	**0**

*Table S3.2. AGOX chemistry through use of 2-PCA pyridinium oxime catalyst **3.24** and NASA acyl donors.* Myoglobin was modified with reagent **3.24** under conditions outlined in General Procedure **3A**, followed by the addition of NASA acyl donors **3.25**-**3.28** (10 equiv.) and incubation at 37 °C for 72 h. Conversion was followed by LC-MS.

Control samples were prepared as detailed above, with modification reagent **3.2**.

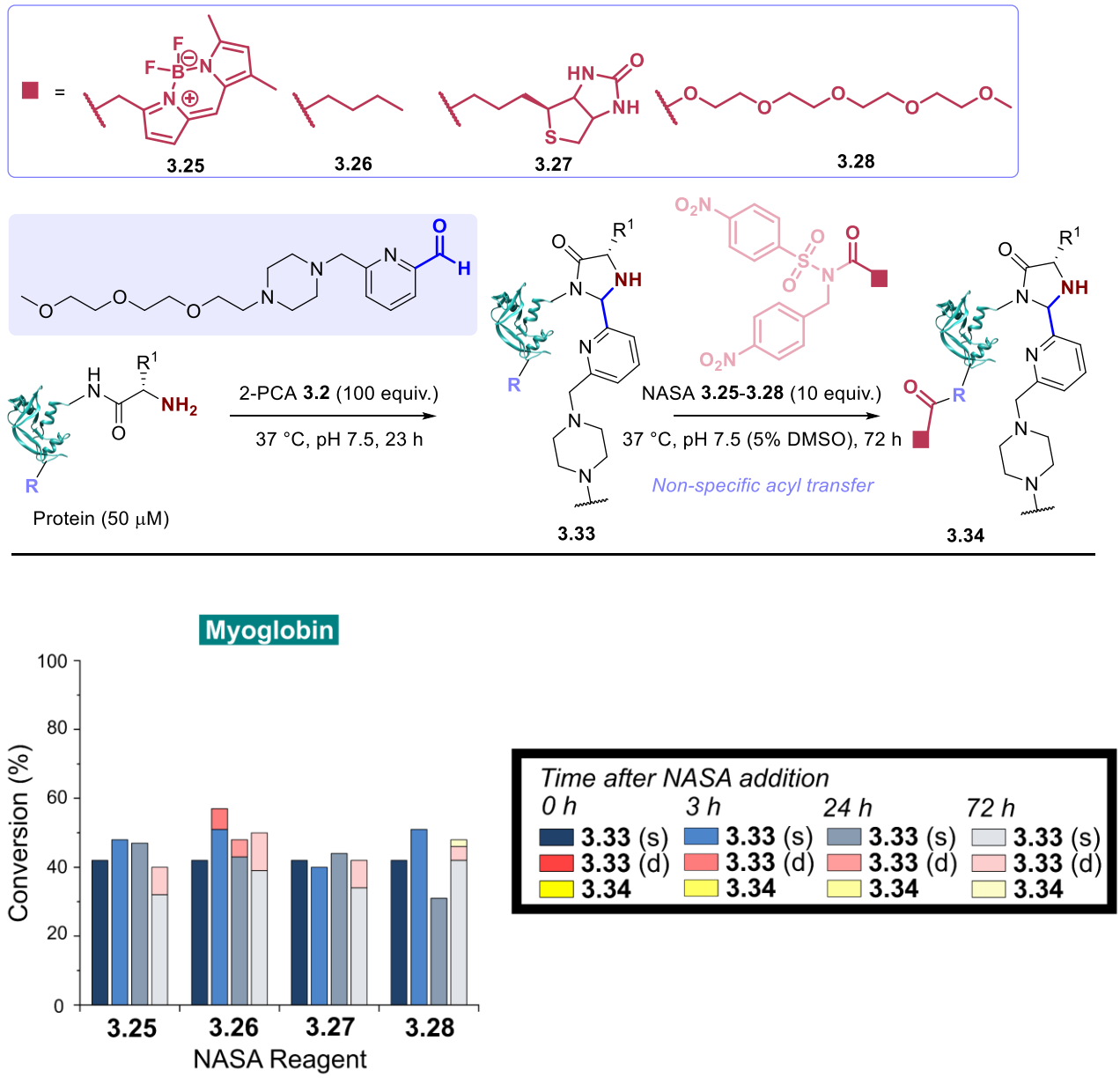
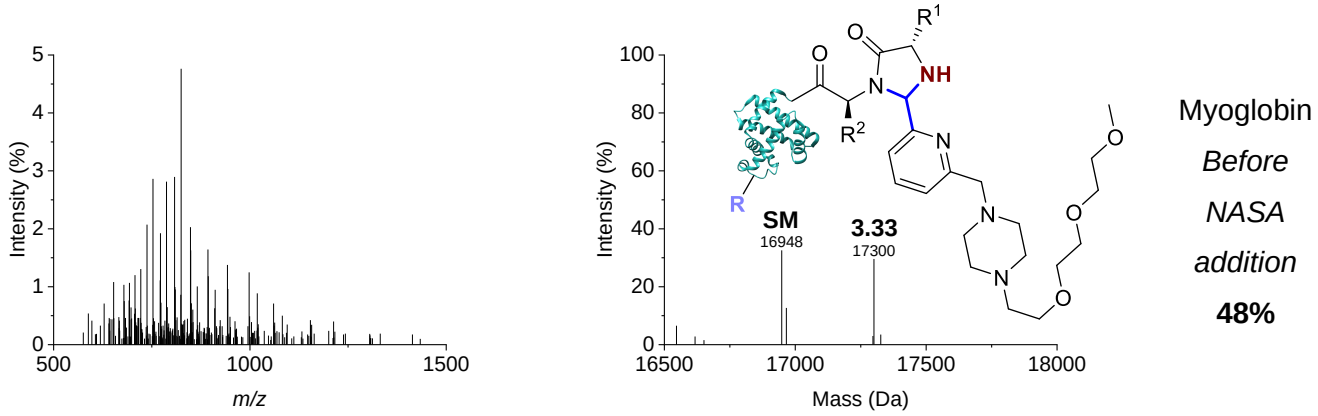
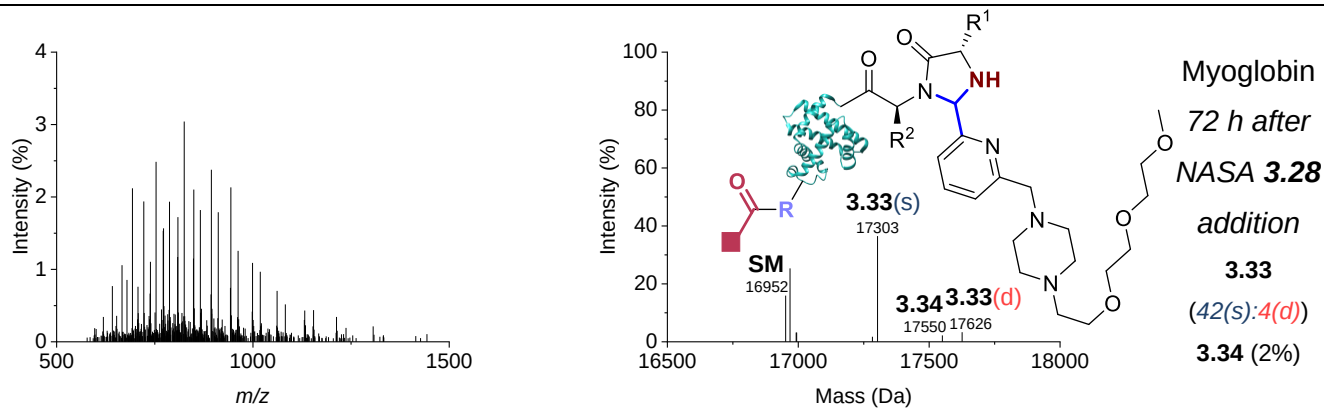


Figure S3.13. AGOX chemistry control, through use of 2-PCA **3.2** and NASA acyl donors. Myoglobin, RNase A, insulin and CjX183-D WT were modified with reagent **3.2** under conditions outlined in General Procedure **3A**, followed by the addition of NASA acyl

donors **3.25**-**3.28** (10 equiv.) and incubation at 37 °C for 72 h. Conversion was followed by LC-MS. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].



MS (ESI$^+$) [**SM**+H]$^+$ found 16948, calculated 16951;[26] [**SM**+H_2O+H]$^+$ found 16966, calculated 16970; [**3.33**+H_2O+H]$^+$ found 17297, calculated 17302; [**3.33**+H_2O+H]$^+$ found 17300, calculated 17302; [**3.33**+MeCN+H]$^+$ found 17326, calculated 17325.



MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951;[26] [**SM**+H_2O+H]$^+$ found 16968, calculated 16970; [**SM**+MeCN+H]$^+$ found 16992, calculated 16993; [**SM**+MeCN+H]$^+$ found 16994, calculated 16993; [**3.33**(s)+H]$^+$ found 17283, calculated 17284; [**3.33**(s)+H_2O+H]$^+$ found 17303, calculated 17302; [**3.34**+H_2O+H]$^+$ found 17550, calculated 17552; [**3.33**(d)+H_2O+H]$^+$ found 17626, calculated 17635.

Figure S3.14. Examples of raw (left) and deconvoluted (right) mass spectra for myoglobin conjugates. **SM** = unmodified protein.

NASA Reagent	Conversion to conjugate 3.33 (%)				Conversion to conjugate 3.34 (%)			
	Before	*3 h*	*24 h*	*72 h*	*Before*	*3 h*	*24 h*	*72 h*
3.25	**42**	**48**	**47**	**40** *32(s):8(d)*	-	**0**	**0**	**0**
3.26		**57** *51(s):6(d)*	**48** *51(s):6(d)*	**50** *39(s):11(d)*		**0**	**0**	**0**
3.27		**40**	**44**	**42** *34(s):8(d)*		**0**	**0**	**0**
3.28		**51**	**31**	**46** *42(s):4(d)*		**0**	**0**	**2**

*Table S3.3. Control for AGOX chemistry through use of 2-PCA **3.2** and NASA acyl donors.* Myoglobin was modified with reagent **3.2** under conditions outlined in General Procedure **3A**, followed by the addition of NASA acyl donors **3.25**-**3.28** (10 equiv.) and incubation at 37 °C for 72 h. Conversion was followed by LC-MS.

*3.4.4.4 Solubility studies for AGOX strategy with 2-PCA **3.24***

To explore the influence of 2-PCA concentration on conversion and solubility, RNase A and myoglobin were modified with reagent **3.24** according to **General Procedure 3A** on an 80 µL scale, varying the equiv. of 2-PCA while maintaining all other conditions (10, 25, 50, 100 equiv.). Control samples were prepared as detailed above, with 50 mM pH 7.5 sodium phosphate buffer (40 µL) in lieu of the solution of modification reagent. Reagent **3.24** was also applied under the conditions outlined in **General Procedure 3A** on an 80 µL scale, with increased buffer concentration (100 mM) or reduced temperature (25 °C). Conversion was determined by LC-MS analysis of the reaction mixture. A 5.9 µL sample (myoglobin) or 7.3 µL sample (RNase A) was taken for SDS-PAGE (soluble). Control and soluble samples were analysed by 15% SDS-PAGE gels. Aliquots of the protein conjugates (50 µL) were purified by dialysis to remove excess reagent (4 °C, 3.5 kDa MWCO; 1 × water, 18 h; 1 × water, 2 h; 1 × water, 5 h), and conversion of the purified conjugates was determined by LC-MS analysis.

2-PCA concentration (mM)	Temperature (°C)	Buffer concentration (mM)	**Conversion (%)**			
			Myoglobin		RNase A	
			Crude	*Purified*	*Crude*	*Purified*
0.5	37	50	**3**	**0**	**0**	**12**
1.25	37	50	**0**	**0**	**31**	**32**
2.5	37	50	**0**	**0**	**60** *50(s):10(d)*	**66**
5	37	50	**-**	**-**	**84** *67(s):17(d)*	**63**
5	25	50	**4**	**0**	**47** *31(s):16(d)*	**31**
5	37	100	**0**	**0**	**88**	**74** *50(s):16(d):7(t)*

*Table S3.4. Variation of 2-PCA concentration, temperature, and buffer concentration for the modification of myoglobin and RNase A with 2-PCA **3.24**.* Modification of myoglobin and RNase A with compound **3.24** under conditions adapted from General Procedure **3A**. s = single modification; d = double modification (*missing data due to poor quality mass spectra*).

3.4.4.5 Improved protocol for AGOX Strategy for RNase A modification

Reagent **3.24** was used to modify RNase A under the conditions outlined in **General Procedure 3A**, on a 150 µL scale with reduced equiv. of 2-PCA (50 equiv.). The reaction mixture was purified by dialysis to remove excess reagent (4 °C, 3.5 kDa MWCO; 1 × water, 3 h; 1 × water, 16 h; 1 × water, 5 h; 1 × water, 3 h; 1 ×

water, 17 h; 2 × water, 2 h), and conversion of the purified conjugate was determined by LC-MS analysis. A stock solution of NASA reagent **3.26**, **3.27** or **3.32** (2 µL, 7.5 mM, 15 nmol, 10 equiv., in DMSO) was added to aliquots of the above reaction mixture (30 µL), and the mixture incubated at 37 °C for 18 h with agitation (1000 rpm). Conversion was determined by LC-MS analysis of the reaction mixture 2 h and 18 h after NASA reagent addition.

NASA Reagent	**Conversion to conjugate 3.29 (%)**			**Conversion to conjugate 3.30/3.31 (%)**		
	Before	*2 h*	*18 h*	*Before*	*2 h*	*18 h*
3.26	**30** *24(s):5(d)*	**27**	**32**	**-**	**0**	**0**
3.27		**14**	**23**		**0**	**0**
3.32		**30**	**31**		**0**	**0**

*Table S3.5. AGOX chemistry through use of 2-PCA pyridinium oxime catalyst **3.24** and NASA acyl donors, under conditions optimised for protein solubility.* RNase A was modified with reagent **3.24** under conditions adapted from General Procedure **3A** and purified, followed by the addition of NASA acyl donors **3.26**, **3.27** or **3.32** (10 equiv.) and incubation at 37 °C for 18 h. Conversion was followed by LC-MS.

3.4.4.6 Dialysis optimisation

Reagent **3.24** was used to modify RNase A under the conditions outlined in **General Procedure 3A**, on a 200 µL scale. Aliquots of the reaction mixture (67 µL) were purified by dialysis to remove excess reagent (4 °C, 3.5 kDa MWCO; 1 × water, 4 h; 1 × water, 17 h; 1 × 5, 50, or 500 mM glycine in water, 6 h; 1 × water, 17 h; 1 × water, 5 h; 1 × water, 4 h; 1 × water, 15 h), and the purified conjugate was analysed by LC-MS.

3.5 References

1. Zhang, L.; Tam, J. P. *Anal. Biochem.* **233**, 87–93 (1996).
2. Li, X.; Zhang, L.; Hall, S. E.; Tam, J. P. *Tetrahedron Lett.* **41**, 4069–4073 (2000).
3. Rosen, C. B.; Francis, M. B. *Nat. Chem. Biol.* **13**, 697–705 (2017).
4. Liu, C. F.; Tam, J. P. *Proc. Natl. Acad. Sci.* **91**, 6584–6588 (1994).
5. Chatani, E.; Hayashi, R.; Moriyama, H.; Ueki, T. *Protein Sci.* **11**, 72–81 (2002).
6. MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. *Nat. Chem. Biol.* **11**, 326–331 (2015).
7. Baldwin, J. E. *J. Chem. Soc. D* 734–736 (1976).
8. Giglione, C.; Boularot, A.; Meinnel, T. *Cell. Mol. Life Sci.* **61**, 1455–1474 (2004).
9. Varland, S.; Osberg, C.; Arnesen, T. *Proteomics* **15**, 2385–2401 (2015).
10. Sherman, F.; Stewart, J. W.; Tsunasawa, S. *BioEssays* **3**, 27–31 (1985).
11. Liao, Y.-D.; Jeng, J.-C.; Wang, C.-F.; Wang, S.-C.; Chang, S.-T. *Protein Sci.* **13**, 1802–1810 (2004).
12. Chatterjee, J.; Rechenmacher, F.; Kessler, H. *Angew. Chem., Int. Ed.* **52**, 254–269 (2013).
13. Huang, R. *ChemBioChem* **20**, 976–984 (2019).
14. Emenike, B.; Czabala, P.; Farhi, J.; Swaminathan, J.; Anslyn, E. V.; Spangle, J.; Raj, M. *J. Am. Chem. Soc.* **146**, 10621–10631 (2024).
15. Chen, D.; Disotuar, M. M.; Xiong, X.; Wang, Y.; Chou, D. H.-C. *Chem. Sci.* **8**, 2717–2722 (2017).
16. Shiraiwa, K.; Cheng, R.; Nonaka, H.; Tamura, T.; Hamachi, I. *Cell Chem. Biol.* **27**, 970–985 (2020).
17. Sakamoto, S.; Hamachi, I. *Anal. Sci.* **35**, 5–27 (2019).
18. Tamura, T.; Song, Z.; Amaike, K.; Lee, S.; Yin, S.; Kiyonaka, S.; Hamachi, I. *J. Am. Chem. Soc.* **139**, 14181–14191 (2017).
19. Seath, C. P.; Trowbridge, A. D.; Muir, T. W.; MacMillan, D. W. C. *Chem. Soc. Rev.* **50**, 2911–2926 (2021).
20. Sahu, T.; Chilamari, M.; Rai, V. *Chem. Commun.* **58**, 1768–1771 (2022).
21. Adusumalli, S. R.; Rawale, D. G.; Thakur, K.; Purushottam, L.; Reddy, N. C.; Kalra, N.; Shukla, S.; Rai, V. *Angew. Chem., Int. Ed.* **59**, 10332–10336 (2020).
22. Adusumalli, S. R.; Rawale, D. G.; Singh, U.; Tripathi, P.; Paul, R.; Kalra, N.; Mishra, R. K.; Shukla, S.; Rai, V. *J. Am. Chem. Soc.* **140**, 15114–15123 (2018).
23. Rawale, D. G.; Thakur, K.; Sreekumar, P.; Sajeev, T. K.; Ramesh, A.; Adusumalli, S. R.; Mishra, R. K.; Rai, V. *Chem. Sci.* **12**, 6732–6736 (2021).
24. Rothbart, S. B.; Krajewski, K.; Strahl, B. D.; Fuchs, S. M., in *Methods in Enzymology*, ed. Wu, C.; Allis, C. D., Elsevier, vol. 512, 2012, Chapter 6, 107–135.
25. Stathopoulos, P.; Papas, S.; Tsikaris, V. *J Pept. Sci.* **12**, 227–232 (2006).
26. Zaia, J.; Annan, R. S.; Biemann, K. *Rapid Commun. Mass Spectrom.* **6**, 32–36 (1992).

Chapter 4

The effect of PCA functionalisation on reactivity and N-terminal targeting

Chapter 4: The effect of PCA functionalisation on reactivity and N-terminal targeting

4.1 Introduction

In Chapter 3, we discussed attempts to improve the kinetics and thermodynamic favourability of N-terminal targeting strategies through changing the nature of the protein (*i.e.* the N-terminal amino acid, Chapter 3.1), N-terminal post-translational modification (Chapter 3.2), or through harnessing the reactivity of nearby residues (Chapter 3.3). With limited success in each of these approaches, in addition to a loss of generalisability, we next turned our attention to the nature of the modification reagent.

The functionalisation of amine targeting reagents with moieties capable of intramolecular stabilisation has previously been employed to improve conjugate stability *via* hydrogen bonding or N → B coordination. For example, in the development of reversible covalent kinase inhibitors, Yao and co-workers modified lysines (*ε*-NH_2) with salicylaldehydes and *ortho*-borono aldehydes, with high potency and selectivity due to intramolecular imine stabilisation *via* hydrogen bonding (*Fig. 4.1a*) or N → B coordination (*Fig. 4.1b*) respectively.[1,2] Previously, *ortho*-borono aldehyde stabilisation has also been effectively incorporated by Bandyopadhyay *et al.* to increase reactivity and prevent hydrolysis of thiazolidine conjugates formed at an N-terminal cysteine (*Fig. 4.1c*).[3]

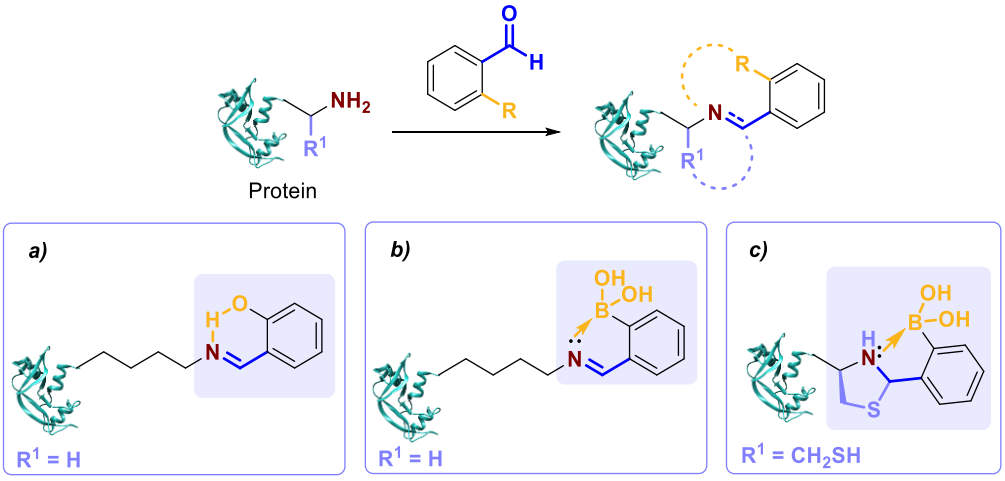


Figure 4.1. Previously reported strategies for intramolecular conjugate stabilisation, *via* ***a)*** hydrogen bonding;[4] or N → B coordination for imines[4] ***(b)*** or thiazolidines[3] ***(c)***. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

Notably, the intramolecular stabilisation strategies outlined above were applied to phenyl (rather than pyridyl) rings; and to our knowledge, neither of these stabilisation approaches have yet been applied to imidazolidinone conjugates. Whilst MacDonald *et al.* reported that the introduction of a hydroxy group *ortho* to the aldehyde had no effect on protein modifications with 2-PCAs (84% conversion observed for both

pyridine-2-carboxaldehyde (2-PCA) and 3-hydroxypyridine-2-carboxaldehyde), but resulted in an increase in conversion for 4-PCAs (from 21% conversion to 81% upon introduction of the *ortho*-hydroxy group), the reaction kinetics and resulting conjugate stability were not explored.[6] We proposed that intramolecular hydrogen bonding / N → B coordination, introduced *via*: *i)* an *ortho*-hydroxy/methoxy group (*Fig. 4.2a*); *ii)* pyridinium *N*-oxide (*Fig. 4.2b*); or *iii)* an *ortho*-boronic acid group (*Fig. 4.2c*) may result in stabilisation of imidazolidinone conjugates formed upon N-terminal modification with PCAs, and reduce their susceptibility to ring-opening and hydrolysis.

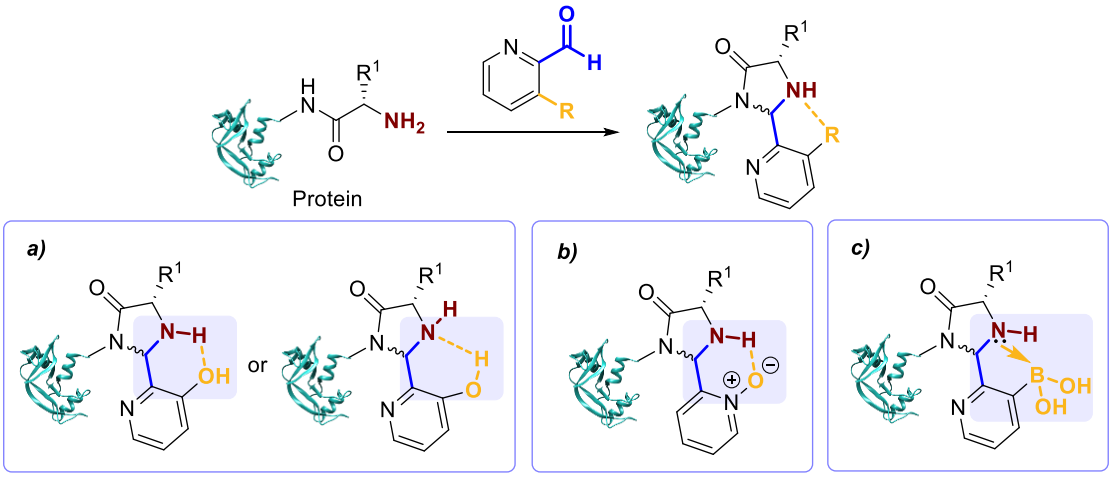


Figure 4.2. Our proposed strategy for intramolecular PCA conjugate stabilisation, *via* hydrogen bonding of hydroxy ***(a)*** or *N*-oxide ***(b)***; or ***c)*** N → B coordination. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

4.2 Results and Discussion

4.2.1 Study design

The reactivity of PCAs with protein N-termini to form imidazolidinone conjugates is governed by a series of equilibria between hydrate, aldehyde (PCA), imine and imidazolidinone forms (*Fig. 4.3*). The effects of PCA functionalisation on the kinetics and stability of protein conjugation were therefore difficult to predict. Previous work by Barman *et al*. explored the effect of 2-PCA and 4-PCA functionalisation on hydration and acetal formation. Even in their simple model system, the equilibria were found to be influenced by a range of multiple factors. They outlined three possible roles of hydroxy groups: *i)* intramolecular general acid-base catalysis; *ii)* hydrogen bonding catalysis; and *iii)* carbonyl deactivation *via* resonance donation.[7] Inspired by their work, we expected the impact of PCA functionalisation to be further influenced upon extending the model system from hydrate formation to imine formation and cyclisation.

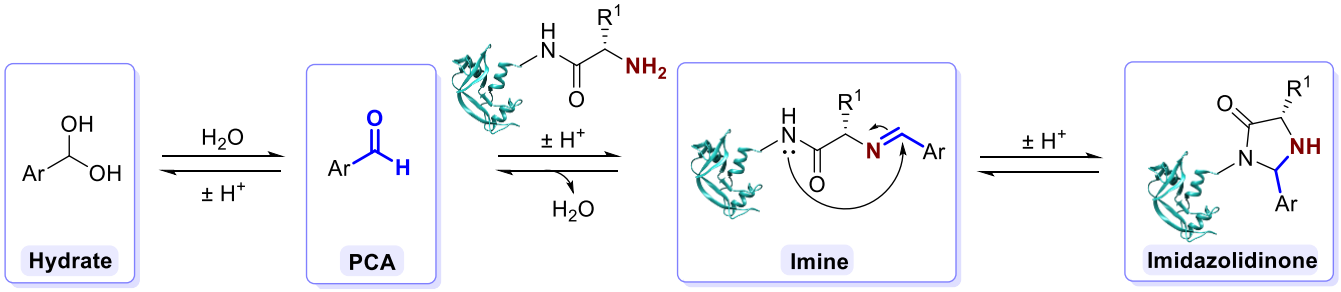


Figure 4.3. Reactions governing reactivity of PCAs with protein N-termini to form imidazolidinone conjugates.

We selected a range of PCAs **4.1**-**4.14** to study, with varying substituents and positions to allow the different roles of substituents to be evaluated in isolation (*Fig. 4.4*). We chose to study three different classes of PCA reagent: *i)* 2-PCAs (**4.1**-**4.6** and **4.13**), *ii)* 4-PCAs (**4.7**-**4.10**), and *iii)* 2-PCA salts (**4.11**-**4.12** and **4.14**). Comparison of 2-PCAs to their 4-PCA analogues allowed the governing role of the pyridine nitrogen atom to be distinguished for each step of N-terminal labelling. In previous reports, differences between 2-PCA and 4-PCA reactivity have been attributed to various factors, including electronic effects of the pyridine, the potential of the nitrogen to act as a general base/hydrogen bond acceptor (2-PCAs only), and ability of the nitrogen to form hydrogen bonds with solvent molecules, resulting in steric hindrance.[6,7] Methylated pyridiniums **4.11** and **4.14** were expected to negate any general base or hydrogen bonding properties of the nitrogen, in addition to inductively activating the aldehyde to nucleophilic attack. Note that in an aqueous environment, **4.6** was expected to be found predominantly in its 2-pyridone tautomer form,[8] thus reducing any contributions from the nitrogen as a general base, and reversing the hydrogen bonding potential of the nitrogen from a hydrogen bond acceptor to donor. However, note that the influence of the reaction equilibria on tautomerism of **4.6** was unknown.

The substitution pattern of hydroxy-functionalised PCAs was varied to distinguish the governing role proposed by Barman *et al.* of the hydroxy for each step of N-terminal labelling,[7] including *i)* general acid

catalysts; *ii)* hydrogen bond acceptors or donors (*ortho*-hydroxy **4.2** and **4.8** only); and *iii)* mesomeric donors to deactivate the carbonyl (*ortho*-/*para*-hydroxy **4.2**, **4.4**, and **4.8**). Methoxy-functionalised PCA analogues were capable of serving as hydrogen bond acceptors only (*ortho*-methoxy **4.3** and **4.9** only), with slightly weaker resonance contributions than their corresponding hydroxy-analogues (Hammet sigma values: –0.38 and –0.27 for *ortho*-hydroxy **4.2** and *ortho*-methoxy **4.3** respectively), and some added steric bulk.[7] Notably, hydrogen bond acceptors had the potential to stabilise hydrate and imidazolidinone forms only, whereas hydrogen bond donors had the potential to stabilise hydrate, aldehyde, imine and imidazolidinone forms. Fluoro-substitution (**4.10**) was included to evaluate the impact of inductive electron withdrawal on carbonyl activation; pyridinium *N*-oxide **4.12** was expected to be similarly activating, whilst also having hydrogen bond acceptor potential, and boronic acid **4.13** could potentially facilitate N → B coordination.

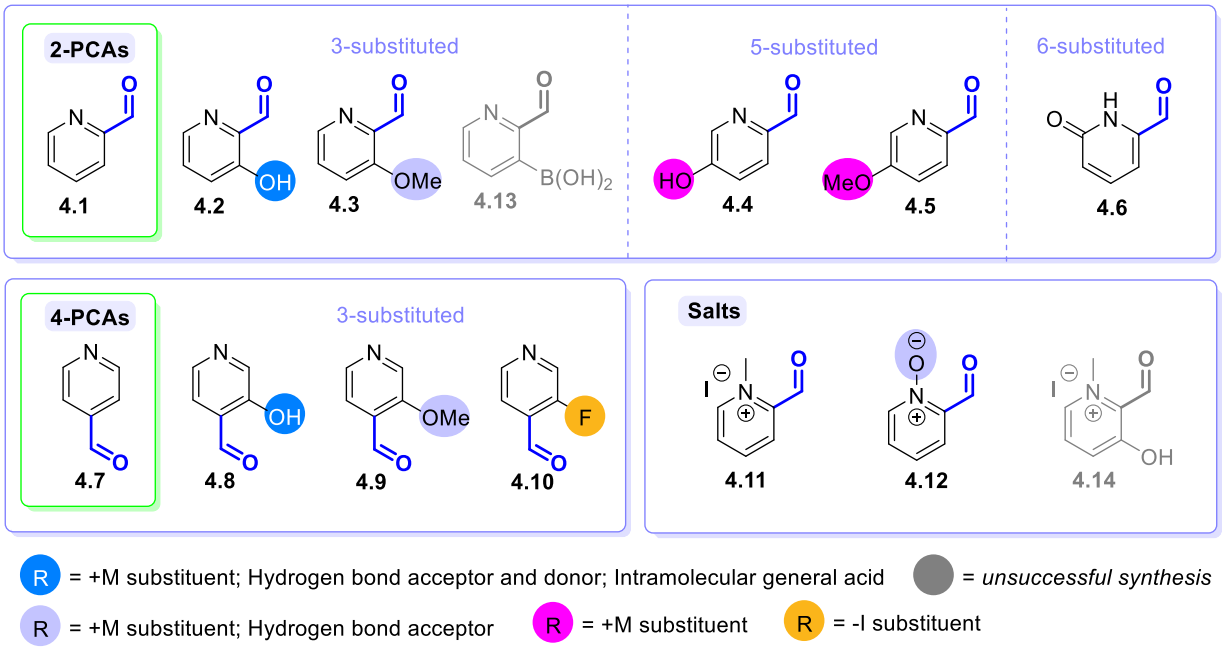


Figure 4.4. PCAs **4.1**-**4.14** used in this study.

4.2.2 Reagent synthesis

i) Pyridinium *N*-oxide **4.12**

To prepare 2-PCA **4.12**, we first carried out *N*-oxidation of 2-(hydroxymethyl)pyridine to pyridinium *N*-oxide **4.15** (77%), followed by alcohol oxidation to afford the target compound (47%, *Fig. 4.5*).

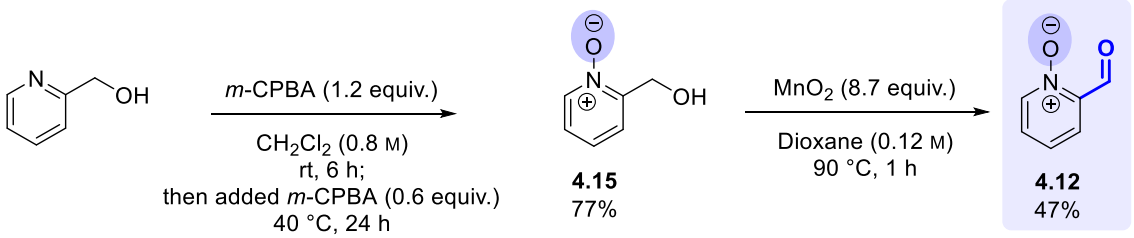


Figure 4.5. Synthesis of 2-PCA **4.12**.

ii) Methylated pyridines **4.11** and **4.14**

To prepare 2-PCAs **4.11** and **4.14**, we attempted the *N*-methylation of 2-PCAs **4.1** and **4.2**, to **4.11** (4%, *Fig. 4.6a*) and **4.14** (1%, *Fig. 4.6b*) respectively. Whilst a sufficient quantity of 2-PCA **4.11** was obtained for the study, the methylations were very poor-yielding in comparison to reported synthesis of 2-PCA **4.14** (lit.[7] 69% after refluxing in benzene for 24 h), due to poor reactivity in the alternative solvent used (CH_2Cl_2). We attempted the reaction of **4.2** in alternative solvents (e.g., toluene), with additional portions of methyl iodide and with variation in temperature to reduce evaporation in methyl iodide. Unfortunately, no improvements in yield were observed. Due to difficulties in synthesis, low purity, and poor modification in preliminary protein experiments under standard protein modification conditions (*Fig. 4.7*), 2-PCA **4.14** was discounted from the study.

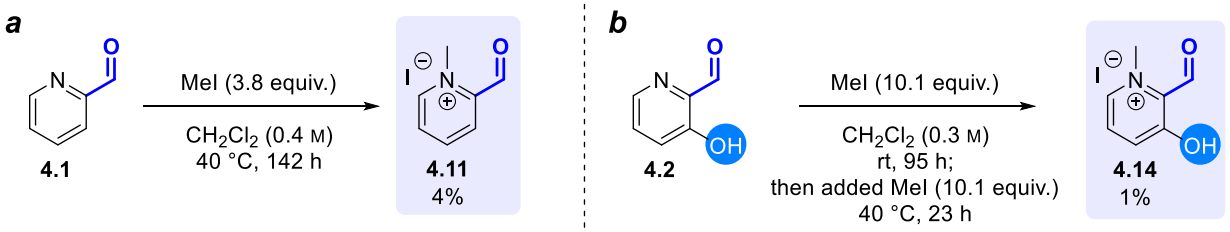


Figure 4.6. Synthesis of 2-PCAs **4.11** and **4.14**.

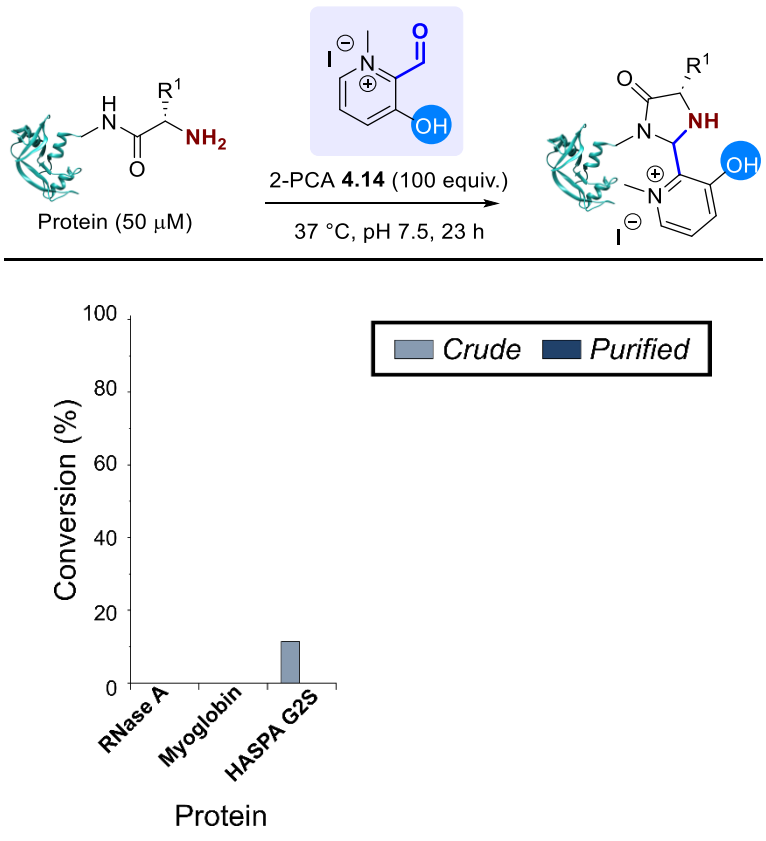


Figure 4.7. Preliminary study to validate the protein labelling reactivity of reagent **4.14**. RNase A, myoglobin and HASPA G2S (50 µM) were modified with reagent **4.14** (100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 23 h, and the conversion was determined by LC-MS before and after purification by dialysis. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

iii) Boronic acid **4.13**

To prepare boronic acid **4.13**, we first attempted to synthesise boronic ester **4.16** from 3-bromopyridine-2-carboxaldehyde with a range of different Pd catalysts and phosphine ligand sources *via* a Miyaura borylation (*Fig. 4.8a-c*). Whilst consumption of 3-bromopyridine-2-carboxaldehyde was observed in each reaction attempt, no conclusive evidence for the formation of ester **4.16** was obtained, with product isolation unsuccessful; issues were encountered with the difficulties in removal of B_2Pin_2 by flash column chromatography. We attempted to take the crude material forwards for the deprotection to boronic acid **4.13**; again, no target compound was identified by ^{1}H NMR spectroscopy or mass spectrometry. Moving forwards, alternative synthetic strategies using aryl boronic 1,1,2,2-tetraethylethylene glycol esters [ArB(Epin)s], developed by Oka *et al.* could provide easier purification than aryl pinacol boronates [ArB(pin)], for which significant loss in yield is commonly experienced upon purification by silica gel column chromatography.[9]

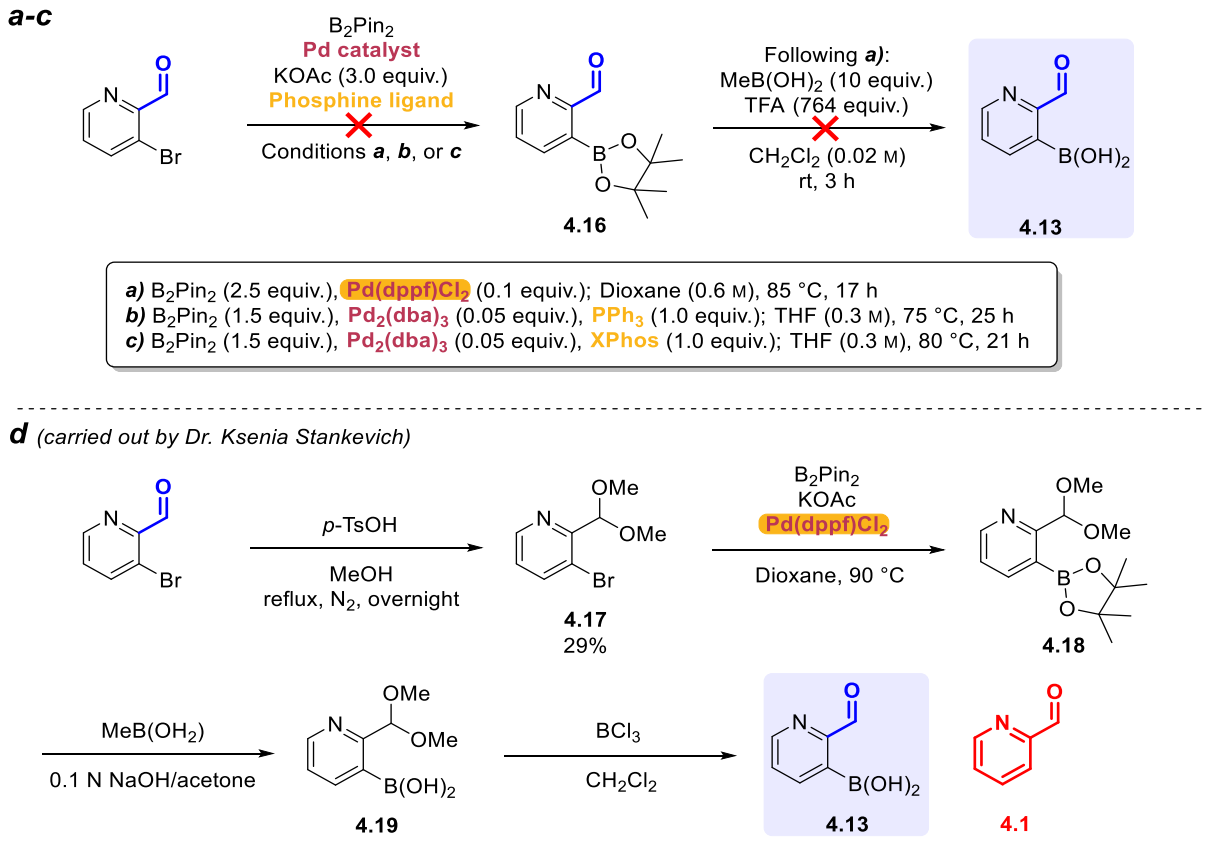


Figure 4.8. Synthesis of *ortho*-borono-pyridaldehyde **4.13** (note: synthesis outlined in part ***d)*** was carried out by Dr. Ksenia Stankevich).

2-PCA **4.13** was synthesised by Dr. Ksenia Stankevich, a PDRA in the Spicer group, as outlined in *Fig. 4.8d*. Firstly, the aldehyde was protected as dimethyl acetal **4.17**, followed by Miyaura borylation to boronic ester **4.18**, and boronic ester deprotection to boronic acid **4.19**. Dimethyl acetal was deprotected using BCl_3 following a protocol developed by Ganley *et al.*,[10] due to the tendency of boronic acids to undergo deborylation under both acidic and basic conditions.[11] Even with this precaution in place, 2-PCA **4.13** was obtained in 50% purity, as a mixture of the target compound and protodeborylated product **4.1**. A preliminary study was carried out to validate the protein labelling reactivity of reagent **4.13**: unfortunately, no modification of RNase A, myoglobin, or HASPA G2S was observed under standard protein modification conditions (*Fig. 4.9*). Due to initial difficulties in synthesis and poor reactivity with model proteins, further attempts to prepare 2-PCA **4.13** were not continued.

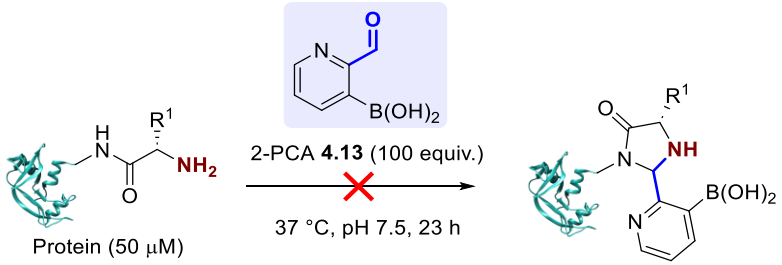


Figure 4.9. Preliminary study to validate the protein labelling reactivity of reagent **4.13**. RNase A, myoglobin and HASPA G2S (50 μM) were modified with reagent **4.13** (100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 23 h, and the conversion was determined by LC-MS before and after purification by dialysis. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

4.2.3 Hydrate formation

With PCAs **4.1**-**4.12** in hand, we first set out to determine the degree of hydration of each reagent. Note that whilst we applied conditions relevant to those used for protein labelling, Barman *et al.* studied hydrate formation in pure water at 25 °C.[7] PCAs **4.1**-**4.12** (50 mM) were incubated at 37 °C (100 mM deuterated Na phosphate buffer, pD 7.3) for > 5 min; based on reported kinetic studies, we expected this to exceed the time-scale for equilibration.[7] Under the conditions outlined above, we expected the pyridine to be deprotonated (pK_a of pyridinium ion ~5.25)[12] and for the hydroxyl groups to be partially deprotonated (e.g., pK_a of *ortho*-hydroxy of PCA **4.2** is 6.8),[7] depending on the substitution pattern; see below for further details. ^{1}H NMR spectra were recorded at 37 °C to determine the equilibrium constant between aldehyde and hydrate forms (*Fig. 4.10*); ratios were calculated *via* the relative integrals of diagnostic ^{1}H NMR signals outlined in *Fig. S4.3a*.

PCA functionalisation and substitution pattern had a large influence on the degree of hydration, with between 1% to >99% hydrate formation observed. In general, higher levels of hydration (7-19% higher) were observed for 4-PCAs in comparison to equivalent 2-PCAs; this is in agreement with the aldehyde/hydrate distributions observed by Barman *et al.*[7] Upon functionalisation, the levels of hydration were mostly governed by

electronic/resonance effects, with reagents with increased aldehyde electrophilicity exhibiting the highest degree of hydration (76-100% hydrate for fluorinated **4.10**, *N*-methylated **4.11**, and *N*-oxide **4.12**). The positive mesomeric effect of *ortho*/*para* hydroxyls resulted in reduced electrophilicity of the aldehyde, and a 27-37% reduction in hydration in comparison to unsubstituted parent PCAs **4.1** and **4.7**. The hydrogen bonding donor/acceptor potential of *ortho*-hydroxy PCA **4.2** had minimal influence on hydration when compared to *para*-hydroxy PCA **4.4**, which was not capable of serving as a hydrogen bonding donor/acceptor.

The degree of hydration of hydroxy-PCA derivatives discussed above suggested resonance donation from the oxygen dominated over any favourable effects from hydrogen bonding/acid catalysis. Interestingly, *ortho*-methoxy substitution resulted in a slight (4-7%) reduction in hydration in comparison to unsubstituted parent PCAs **4.1** and **4.7**, whilst *para*-methoxy **4.5** exhibited a much larger reduction in the degree of hydration (25% reduction). Whilst aldehyde deactivation was expected to be less pronounced for methoxy-derivatives in comparison to their equivalent hydroxy-derivatives due to slightly poorer resonance donor properties of the methoxy group, the differing extent of this effect between *ortho*- and *para*-substitution indicated the interplay of a combination of factors. The nature of these factors is difficult to distinguish, and could potentially be due to a combination of *i)* a steric role of the *ortho*-methoxy group; and *ii)* stabilisation of the hydrate form only through the hydrogen-bonding acceptor potential of *ortho*-methoxy derivatives (in contrast to *ortho*-hydroxy derivatives, for which stabilisation of the aldehyde form was also possible by serving as a hydrogen-bonding donor). Note that Barman *et al.* found that the *ortho*-methoxy group resulted in minimal deactivation of the aldehyde towards hydration,[7] consistent with our results. They suggested that hydroxy anions from the partial deprotonation of hydroxy-functionalised PCAs may have amplified the deactivation of the aldehyde for hydroxy-PCAs. To probe this hypothesis, Dr. Chris Spicer measured the pK_a of the hydroxyl groups of **4.2** (6.5), **4.4** (6.1), and **4.8** (6.9), indicating that the hydroxyl groups were predominantly in their anion form under reaction conditions, as predicted. Note that pyridine **4.6** showed similar hydration levels to unsubstituted PCA **4.1**.

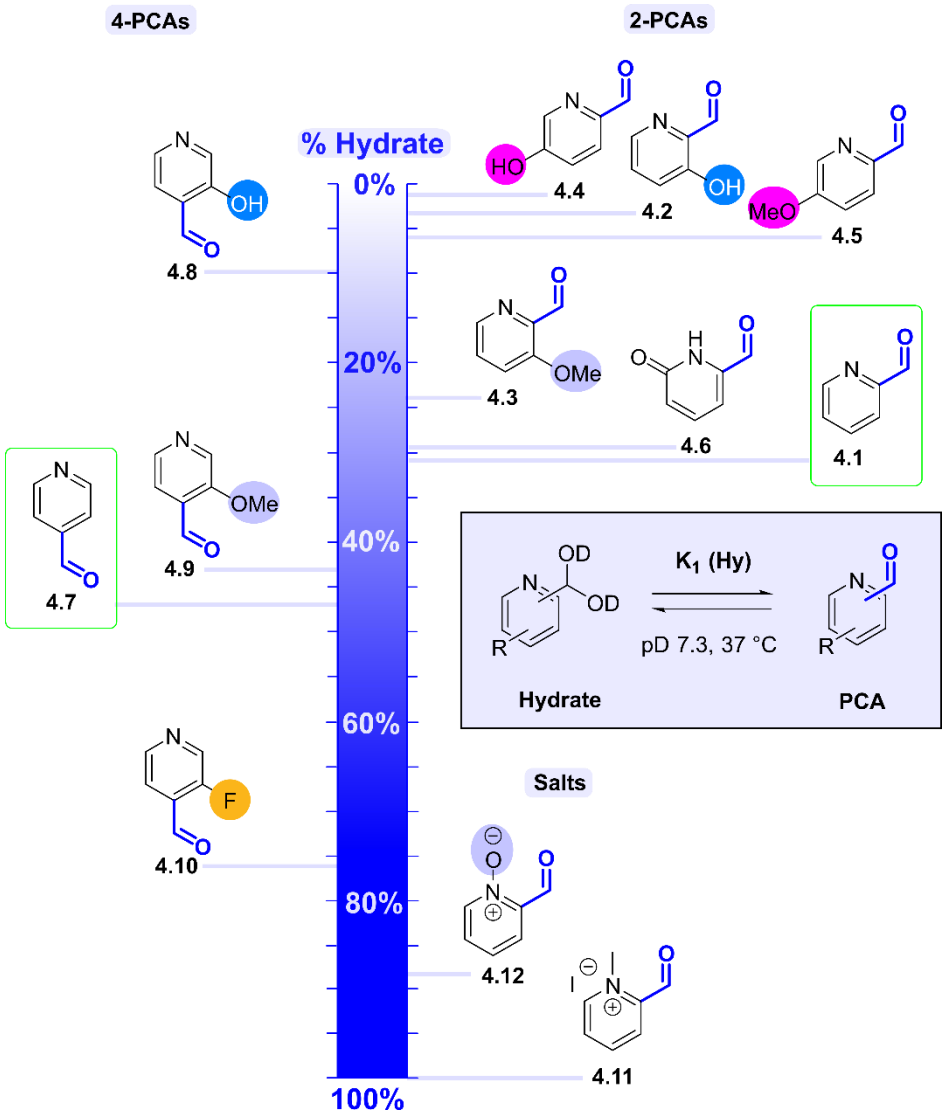


Figure 4.10. Percentage hydrate formation at 37 °C in pD 7.3 buffer for PCAs **4.1**-**4.12**.

4.2.4 Imine formation

Having studied the degree of hydration under conditions relevant to protein modification, we next extended our model to also include imine formation, as a key intermediate labelling step. As discussed previously, N-terminal selectivity can be partly attributed to the lower pK_a of the N-terminal ammonium group (~6.0-8.0) relative to the ε-ammoniums of lysine residues (~10.5), due to the electron-withdrawing effects of the adjacent amide.[13] We hoped to mimic this reduced pK_a using alanine dimethylamide **4.20** to model N-terminal reactivity; the adjacent nitrogen atom was blocked with two methyl groups to prevent cyclisation to an imidazolidinone, allowing the imine and hydrate formation to be studied in isolation. We expected the nature of the adjacent amide (tertiary *vs.* secondary) to have some effect on the pK_a of the N-terminal α-amine, but for this effect to be small in comparison to the corresponding carboxylic acid of L-alanine.

To prepare alanine dimethylamide **4.20**, we first carried out *N*-methylation of Boc-L-alanine to **4.21** (57%), followed by Boc-deprotection to **4.22** and ion exchange to afford the target compound (87%, *Fig. 4.11*).

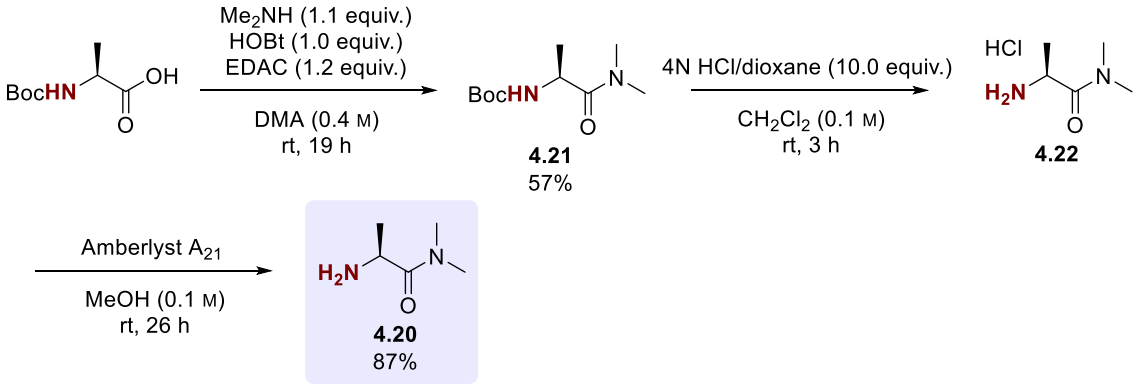


Figure 4.11. Synthesis of alanine dimethylamide **4.20**.

PCAs **4.1**-**4.12** (25 mM) were incubated with alanine dimethylamide **4.20** (25 mM) at 37 °C (100 mM deuterated Na phosphate buffer, pD 7.3) and allowed to equilibrate over 24-48 h. Note that imine formation first proceeds through a hemi-aminal intermediate, followed by dehydration to form an iminium ion. We expected hemi-aminal formation to be in competition with hydrate formation, and to be promoted for reagents with increased aldehyde electrophilicity, as seen previously for hydrate formation. On the other hand, the resulting iminium species would be stabilised better by electron-donating substituents. Under the conditions outlined above, we expected the degree of protonation of the imines formed to be highly variable due to large differences in the reported pK_a of iminium ions (e.g., pK_a ~5 for unsubstituted 4-PCA **4.7**; pK_a ~9 for *ortho*-hydroxy 4-PCA **4.8**), a feature which arises from the stabilisation of iminium ions for *ortho*-hydroxy-functionalised PCAs through intramolecular hydrogen bonding and resonance donation.[14] The reactivity of imines/iminium ion species towards hydrolysis or imidazolidinone formation is therefore likely to vary, depending on the substitution pattern. ^{1}H NMR spectra were recorded at 37 °C to determine the relative concentrations of aldehyde, hydrate and imine forms (*Fig. 4.12*); ratios were calculated *via* the relative integrals of diagnostic ^{1}H NMR signals outlined in *Fig. S4.3b*. Note that hemi-aminal intermediates were not observed. We identified two measures of imine formation, through comparison of: *i)* overall percentage imine formation (*Fig. 4.12a*); and *ii)* the equilibrium constant (K_2) which discounted the effects of hydrate formation (*Fig. 4.12b*).

i) Percentage imine formation

Firstly, we evaluated the percentage imine formation to provide an indication of the availability of the imine intermediate for the subsequent cyclisation step. Note that high levels of imine formation may also be indicative of a higher tendency for lysine side-chain modification and reduced N-terminal selectivity, as discussed later. Whilst we found hydration to be mostly governed by electronic/resonance effects, roles of PCA functionalisation on imine formation were more difficult to distinguish due to the balance between factors promoting hemi-aminal formation or subsequent dehydration. 4-PCA functionalisation had no influence on

the degree of imine formation (11-12%), with the majority of 2-PCAs also resulting in similar imine levels (8-15%). Notably, reduced imine formation (0-5%) was observed for *para*-substituted 2-PCAs **4.4**-**4.5**, due to decreased aldehyde electrophilicity from mesomeric donation, and less favourable hemi-aminal formation. Interestingly, whilst similar resonance effects were expected for *ortho*-substituted 2-PCAs **4.2**-**4.3**, increased levels of imine formation (8-22%) were observed in comparison to equivalent *para*-substituted 2-PCAs. This effect was most notable for 2-PCA **4.2** (22%) and could indicate the role of the *ortho*-hydroxy group in intramolecular hydrogen bonding, providing imine/iminium ion stabilisation. For *ortho*-methoxy **4.3**, the influence of intramolecular hydrogen bonding was expected to be less notable due to its potential to serve as a hydrogen bonding acceptor, but not donor, providing iminium ion stabilisation only. High levels of imine formation were observed for N-oxide **4.12** (15%); this may have been due to both aldehyde activation (i.e., promoting hemi-aminal formation), and iminium ion stabilisation *via* hydrogen bonding.

The degree of imine formation was influenced by competing hydration: for example, a high degree of hydration (100%) was observed for 2-PCA **4.11**, thus depleting aldehyde levels available for imine formation (0%). Contrastingly, low levels of hydration of 2-PCA **4.2** may have contributed to the high degree of imine formation *via* increased aldehyde availability. We therefore chose to carry out a second analysis of the equilibrium constant (K_2) to negate the influence of hydration.

ii) Equilibrium constant for imine formation (K_2)

No PCA **4.11** was available in aldehyde form due to high levels of hydration, so this derivative was discounted from the analysis. As no imine formation was observed for PCA **4.4**, K_2 could not be calculated but was assumed to approach 0 M^{-1}. No aldehyde was observed for PCA **4.12** due to high levels of imine formation, so K_2 could not be calculated; instead, the observed equilibrium constant between hydrate and imine forms was calculated (K_{obs2}), but this cannot be directly compared to the other PCAs. For the remaining PCAs, the most notable difference in comparison to our analysis in part *i)* emerged for electron-deficient PCA analogues with high levels of hydration (e.g., fluorinated **4.10**; K_2 = 26.02 M^{-1}). Upon discounting hydration, it became clear that increased aldehyde electrophilicity promoted imine formation, presumably due to favourable amine addition. As discussed previously, higher levels of hydrate formation were observed for 4-PCAs in comparison to equivalent 2-PCAs (4-14% higher). It was therefore unsurprising that upon discounting the effects of hydration, clearer differences emerged between 4-PCAs, whilst the trends within the 2-PCA class were mostly retained. Interestingly, whilst increased imine formation was observed upon *ortho*-hydroxy substitution of 2-PCA **4.2** in comparison to the unsubstituted parent PCA **4.1** (8.37 M^{-1} increase in K_2), the opposite effect was observed for the equivalent 4-PCAs (4.71 M^{-1} decrease in K_2). A moderate increase in K_2 was observed for pyridone **4.6**, in comparison to PCA **4.1**. Earlier, we saw comparable behaviour of **4.6** and **4.1** in hydrate formation; similarly, we expected the reagents to show similar reactivity in the hemi-aminal formation step. This may indicate that the pyridone plays a role in promoting the dehydration step.

Unexpectedly, the ratio of aldehyde/hydrate/imine species of PCA **4.2** to alanine dimethylamide **4.20** decreased over time, with formation of a new unidentified species in addition to a colour change from a colourless to a bright blue solution. Full characterisation of the reaction mixture at 250 mM initial concentration was carried out after incubation at 37 °C for 4 days: only alanine dimethylamide **4.20** was observed, with additional very broad peaks in the aromatic region (δ_H ~6.5-8.5 ppm) and at approx. δ_H 1.5-3 ppm. With a pK_a of 6.8,[7] the *ortho*-hydroxy of PCA **4.2** is partially deprotonated at pH 7.5. We speculated that the broad peaks may have been due to polymerisation of the 2-PCA *via* nucleophilic activity of its deprotonated form, enabling attack at a neighbouring imine; this step is reversible, and could indicate further steps then taking place. The loss of defined 2-PCA ^{1}H NMR signals and appearance of broad signals was not observed for aldehyde/hydrate species of PCA **4.2** in isolation in *Fig. 4.10*: we suspected the polymerisation was catalysed by alanine dimethylamide **4.20**. Whilst concerning for the modification of **4.20**, we suspected the side-reaction may be less significant for peptides with an N-terminal adjacent secondary amide, for which intramolecular imidazolidinone formation could trap the intermediate imine species in preference to polymer formation, although further kinetic studies would be required to confirm this.

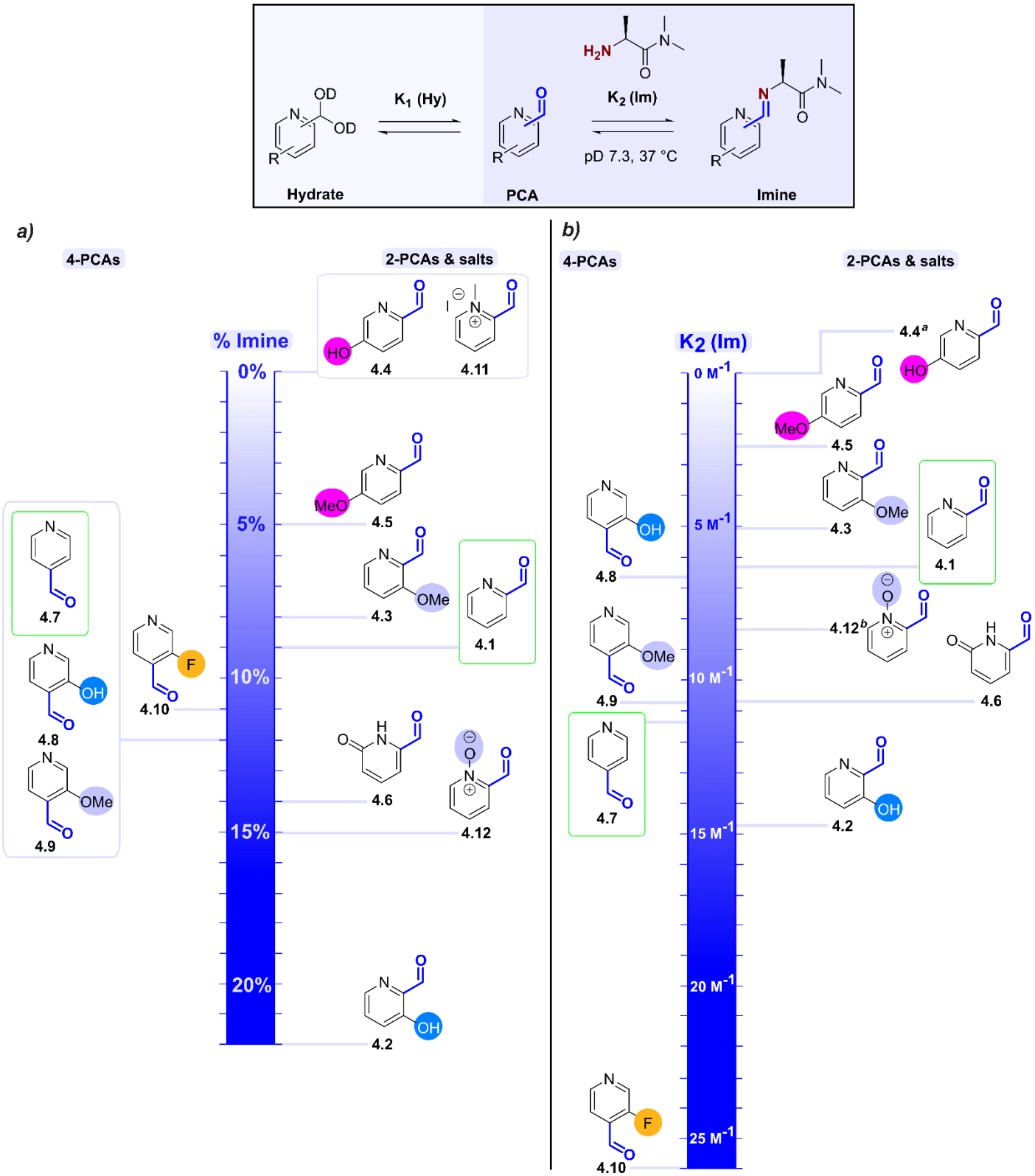


Figure 4.12. Reaction of PCAs **4.1**-**4.12** (25 mM) with alanine dimethylamide **4.20** (25 mM) at 37 °C in pD 7.3 buffer; ***a)*** Percentage imine formation; ***b)*** Calculated K_2 (Im) values for imine formation; [a]No imine was observed for **4.4**, and so K_2 (Im) could not be calculated but approaches 0 M^{-1}; [b]No aldehyde was observed for **4.12** and so K_{obs2} for hydrate-imine equilibrium is used.

4.2.5 Imidazolidinone formation

Having studied the degree of hydration and imine formation under conditions relevant to protein modification, we were now in a position to study the final labelling step: cyclisation to an imidazolidinone. PCAs **4.1**-**4.12** (50 mM) were incubated in a 1:1 ratio with DiAla (50 mM) at 37 °C (100 mM deuterated Na phosphate buffer, pD 7.3) for 16 h. ^{1}H NMR spectra were recorded at 30 min time-points at 37 °C to determine the relative ratio of aldehyde, hydrate, imine and the two imidazolidinone diastereomer forms. Ratios were calculated *via* the relative integrals of diagnostic ^{1}H NMR signals outlined in *Fig. S4.3c*.

We identified two measures of imidazolidinone formation, through comparison of: *i)* overall percentage imidazolidinone formation (*Fig. 4.13*); and *ii)* the rate constants (k_3 and k_{-3}) for imidazolidinone formation (*Fig. 4.14*). Note that for *ortho*-hydroxy PCAs **4.2** and **4.8**, and *N*-oxide **4.12**, the formation of additional species was observed, resulting in increased complexity and ambiguity in the ^{1}H NMR analysis. We hypothesised that hydrogen bonding may reorient the pyridine ring, resulting in the presence of multiple isomers; in the case of PCA **4.2**, increased ^{1}H NMR complexity may have also been a result of the PCA side-reactivity identified previously for imine formation. This led us to a third measure of imidazolidinone formation *iii)*, where the modification of a peptide was followed by LC-MS analysis.

i) Percentage imidazolidinone formation

Firstly, we compared the percentage imidazolidinone formation after 16 h incubation with each reagent. Note that the selected time-point was a snapshot of the reaction progress; for some analogues, equilibrium had not yet been reached. Reaction profiles can be seen in *Table S4.3* in the experimental section. The percentage imidazolidinone formation reflected the complex balance between hydrate, imine, and imidazolidinone species; notably, imidazolidinone formation was observed even for PCA analogues where imine availability was <1% (**4.4** & **4.11**). In general, 2-PCAs reached higher conversions over 16 h than equivalent 4-PCAs (8-43% higher). This observation could suggest a potential role of the pyridine nitrogen as a hydrogen bond acceptor/general base to accelerate nucleophilic attack of the α-amide.

The levels of imidazolidinone formation demonstrated a balance between electronic/resonance and hydrogen bonding effects. Low levels of imidazolidinone formation were observed for electron-deficient PCAs (fluorinated **4.10** and *N*-methylated **4.11**). Conversely, some increase in imidazolidinone formation in comparison to parent PCA **4.1** was observed for *para*-hydroxy and methoxy-functionalisation, where we expected resonance donation but no hydrogen bonding capability. This increase in imidazolidinone formation was amplified for *ortho*-hydroxy and methoxy-functionalisation; *ortho*-methoxy groups to a lesser extent. This suggested a potential contribution to imidazolidinone formation from hydrogen bonding. *N*-oxide **4.12** represented an electron-deficient analogue with hydrogen bonding potential, and reflected a compromise between the above effects, with a moderate increase in imidazolidinone formation in comparison to **4.1**.

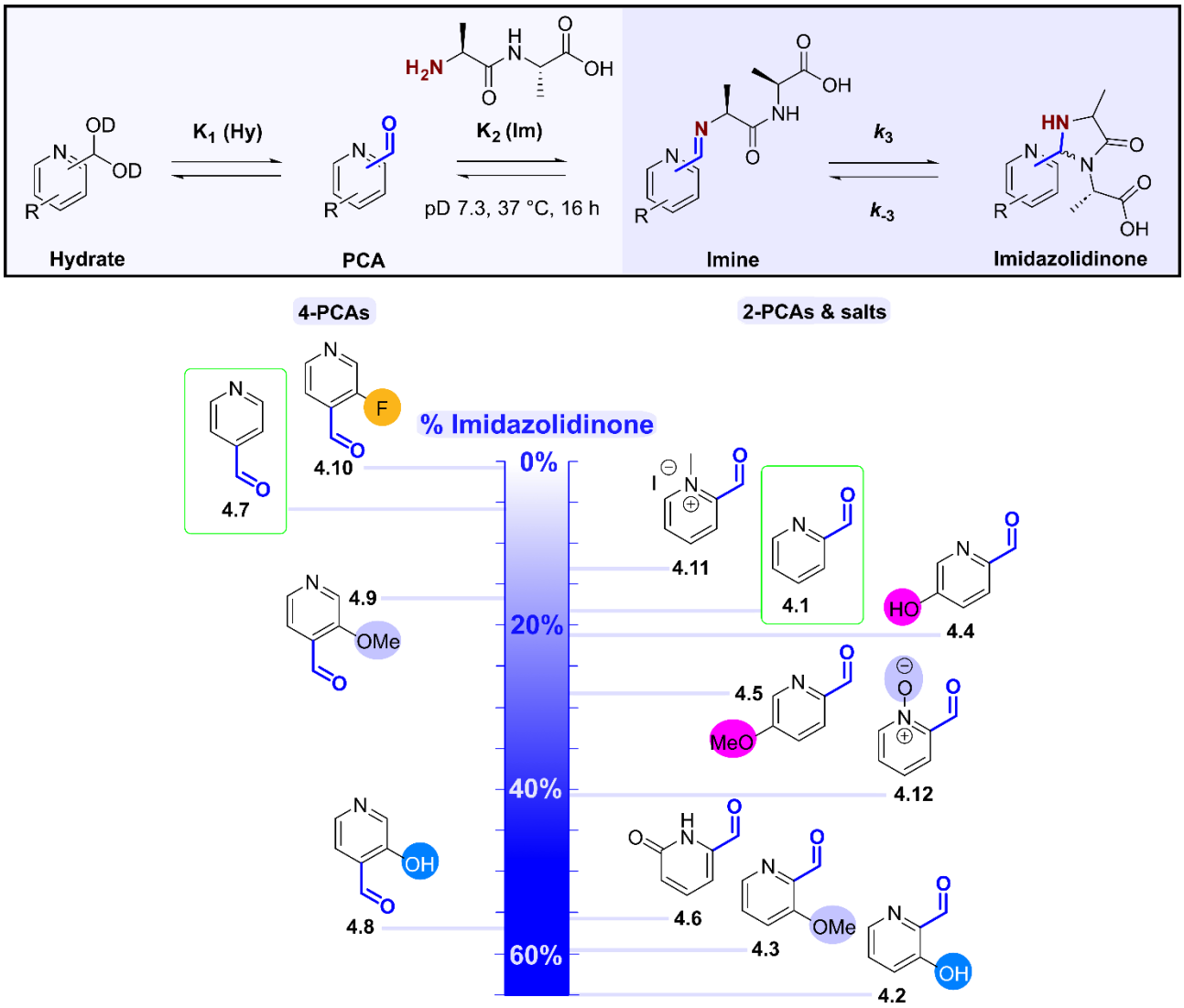


Figure 4.13. Percentage imidazolidinone formation after reaction of PCAs **4.1**-**4.12** (50 mM) with DiAla (50 mM) at 37 °C in pD 7.3 buffer for 16 h.

ii) Rate constants for imidazolidinone formation (k_3 and k_{-3})

During previous work discussed in Chapter 2, kinetic analysis revealed imidazolidinone formation to be the rate-limiting step of the labelling process, as modification of DiAla (50 mM) with alternative 2-PCA analogues (50 mM) at 37 °C was relatively slow (k_3 = 0.7-4.2 × 10^{-4} M^{-1} s^{-1}) in comparison to hydration/imine formation discussed above, where equilibrium was reached prior to initial time-point measurements. In our second analysis, we therefore fitted the data to a kinetic model integrating hydration, imine and imidazolidinone formation under equilibrium, under the assumption that imidazolidinone formation was orders of magnitude slower than either imine or aldehyde formation. This enabled us to evaluate reaction kinetics of reversible imidazolidinone formation through calculation of forwards and reverse rate constants (k_3 and k_{-3}), and the dissociation constant K_d to provide an indication of conjugate stability.

Note that for analogues **4.4** and **4.11**, alternative models were used due to the lack of a measured K_2 (Imine) that could be integrated into the model, so comparisons in k_3 and k_{-3} could not be made with the other reagents. Rate constants (k_{obs}) were calculated directly from aldehyde to imidazolidinone for PCA **4.4** (k_{obs4}, second-order; k_{obs-4}, first-order), and directly from hydrate to imidazolidinone for PCA **4.11** (k_{obs5}, second-order; k_{obs-5}, first-order), based on steady-state approximations detailed in the experimental section.

Trends were less clear here than in part *i)*, with differing effects of ring substitution between 2- and 4-PCAs, and *ortho*-hydroxy/methoxy substitution. In general, hydroxy/methoxy substitution accelerated imidazolidinone formation in comparison to parent unsubstituted PCAs, ranging from a minimal increase up to an order of magnitude. Whilst electron-donating groups were expected to decrease the electrophilicity of the imine, we expected electron-donating groups to provide stabilisation of iminium species (with a higher iminium ion pK_a), resulting in a higher level of protonated species and favoured nucleophilic attack for hydroxy/methoxy-substituted PCAs.[14] Contrastingly, very slow rates of cyclisation were observed for electron-deficient analogue **4.10**, with a slow overall increase in the percentage of imidazolidinone for **4.10** and N-methylated **4.11**, due to a lower level of imine protonation and reduced susceptibility to nucleophilic attack.

For the reverse reaction, *meta*- and *para*-hydroxy/methoxy analogues also accelerated ring opening, resulting in little difference in the overall dissociation constant. Contrastingly, *ortho*-hydroxy/methoxy substitution, in addition to *N*-oxide **4.12**, accelerated the forwards reaction but slowed down the reverse reaction (with *ortho*-hydroxy **4.2** as a notable exception, and with only slight enhancement of the forwards reaction for *ortho*-methoxy **4.9**); these analogues all had potential hydrogen bonding capability. Within this category, no value of k_{-3} could be fitted for *ortho*-methoxy PCAs **4.3** and **4.9**, or *N*-oxide **4.12**, indicating higher stability for these analogues and suitability to protein modification discussed later. Note this does not preclude reversibility at levels below our threshold for detection ($k_{-1} < 10^{-3}$ h^{-1}); a lack of certainty within this observation may also arise from the fitting of the data to the model. This difference between reversibility of *ortho*-hydroxy *vs.* methoxy/*N*-oxide analogues could potentially be as whilst imidazolidinone stabilisation was possible for all *ortho*-analogues through hydrogen bonding donor/acceptor potential, imine stabilisation was possible through the hydrogen bonding donor potential of hydroxy-analogues only (although hydrogen bonding acceptor potential of methoxy-analogues could provide iminium ion stabilisation).

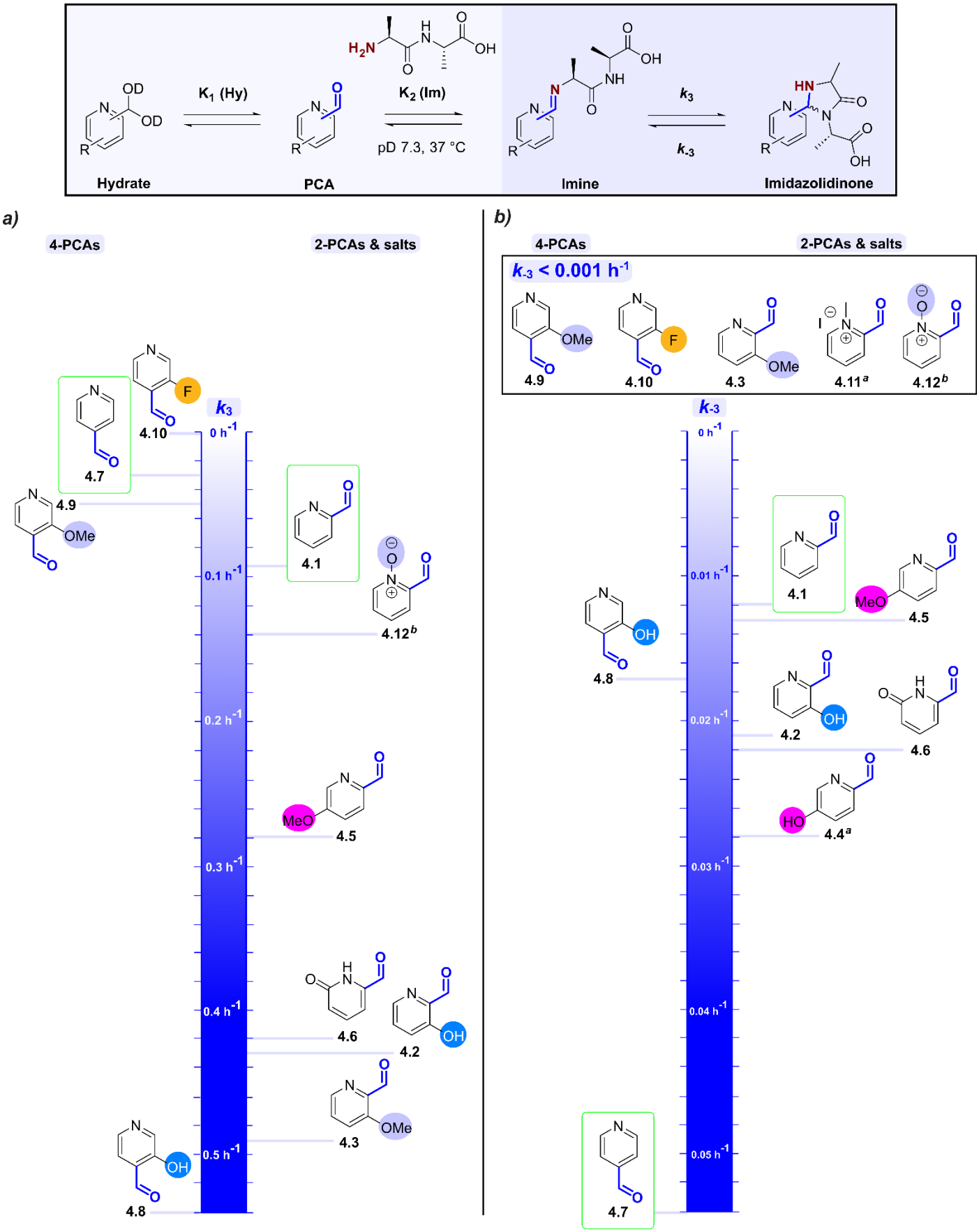


Figure 4.14. Reaction of PCAs **4.1**-**4.12** (50 mM) with DiAla (50 mM) at 37 °C in pD 7.3 buffer. ***a)*** Calculated forward rate constants, k_3. Nb. For PCAs **4.4** and **4.11** no imine was shown in previous experiments and so no k_3 could be calculated. Steady-state approximations could be applied to give the second-order observed rate constants k_{obs4} and k_{obs5} as 0.44 M^{-1} h^{-1} and 0.18 M^{-1} h^{-1}

respectively, as described in the experimental; ***b)*** Calculated reverse rate constants, k_{-3}; [a]For **4.4** and **4.11** reverse reaction rates, $k_{obs\text{-}4}$ and $k_{obs\text{-}5}$ respectively, were based on calculations encompassing the steady state approximation detailed above and in the experimental; [b]For **4.12** values of both k_3 and k_{-3} were derived based on K_{obs2} as described in *Fig. 4.12* and the experimental.

iii) LC-MS Kinetics for PCA analogues with complex spectra

Next, we monitored reaction kinetics *via* LC-MS as a final measure of imidazolidinone formation, to account for ambiguity in the above NMR kinetics experiments for PCA analogues where more than two diastereomers were formed. In Chapter 3, estimations of peptide conversion by LC-MS analysis were poor due to absorbance of both the peptide and modification reagent at the wavelength analysed. To avoid this issue, we prepared a short DiAla-terminated peptide bearing a unique dansyl chromophore (hereby referred to as AAK(dansyl); prepared by Julia Capar, a summer student in the group). As shown in *Fig. 4.15*, an absorption maximum was observed at 325 nm (ε = 3544 mol^{-1} dm^3 cm^{-1}) for AAK(dansyl), whereas PCAs **4.1** and **4.12** exhibited minimal absorbance at 325 nm (ε = 54 mol^{-1} dm^3 cm^{-1} and 81 mol^{-1} dm^3 cm^{-1} respectively). Signals in the 325 nm chromatogram could therefore be solely attributed to the dansyl group, allowing quantitative comparison. This relied on the assumption that the molar extinction coefficient of the dansyl group at 325 nm was not influenced upon modification of the N-terminus. Unfortunately, PCA **4.8** was found to absorb at 325 nm (1871 mol^{-1} dm^3 cm^{-1}), so this derivative was discounted from the analysis.

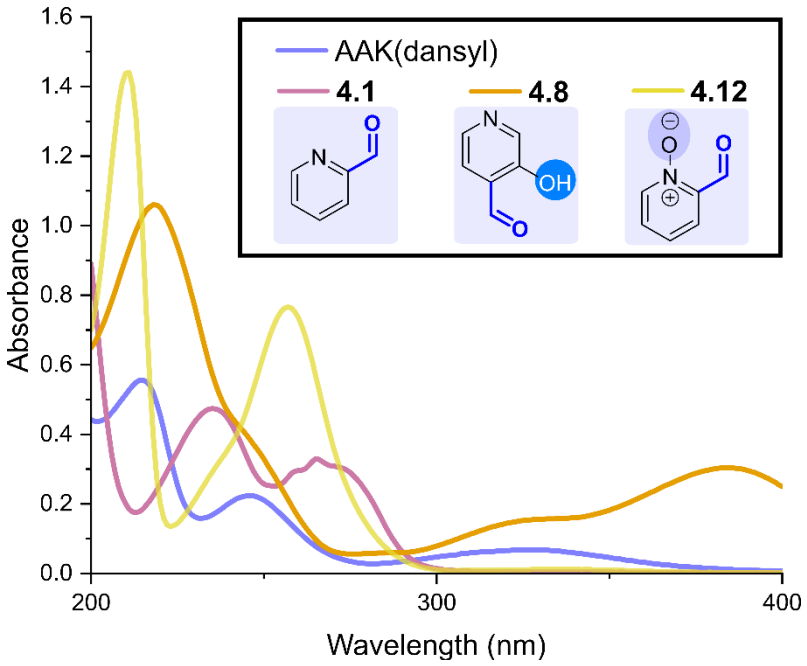
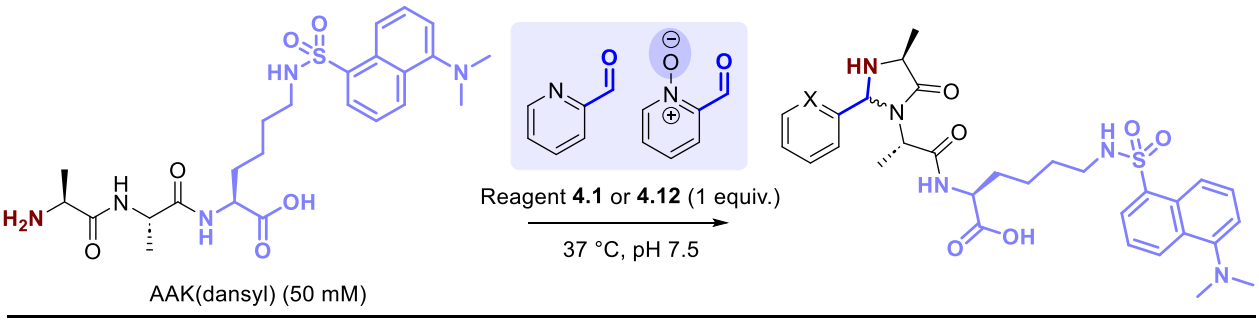


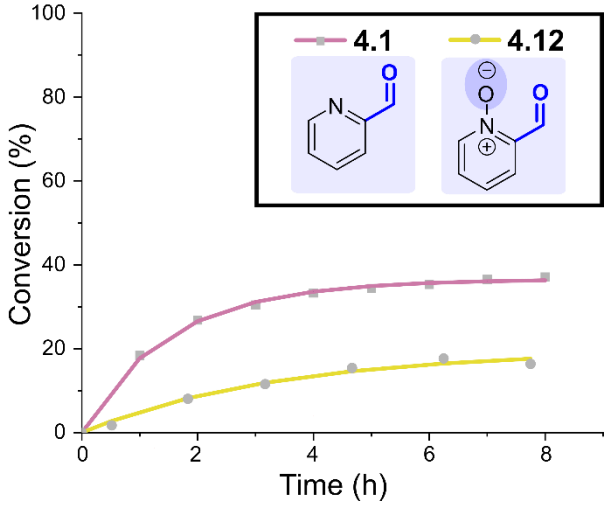
Figure 4.15. UV-vis spectra of AAK(dansyl) (19 μM), 2-PCA **4.1** (93 μM), 4-PCA **4.8** (81 μM) and 2-PCA **4.12** (81 μM) in 50 mM Na phosphate buffer (pH 7.5).

PCAs **4.1** and **4.12** (50 mM) were incubated in a 1:1 ratio with AAK(dansyl) (50 mM) at 37 °C (100 mM deuterated Na phosphate buffer, pH 7.5) for 8 h. The reaction mixtures were analysed by LC-MS at 1-1.5 h

time-points; conversion was calculated from the relative intensities of unmodified/modified peptide signals in the 325 nm absorbance trace. Note that *m/z* values corresponding to hydrate species were not observed by LC-MS, and masses of conjugates were assumed to be due to imidazolidinone formation, rather than the intermediate imine. We made this assumption based on the transient nature of imine species: we expected imine hydrolysis to occur during LC-MS analysis. In contrast to the ^{1}H NMR analysis discussed above, here the resultant data were fitted to a simplified kinetic model which did not encompass hydration or imine formation.

Surprisingly, PCA **4.1** reacted more quickly and reached higher conversions after 8 h than **4.12**, in contrast to the NMR experiments detailed above. These differences in reactivities with different peptide models highlights the protein-dependent nature of N-terminal protein modification chemistries; the 100-fold dilution of samples prior to LC-MS analysis may have also been a contributing factor to these differences. As illustrated in *Fig. S4.5*, whilst the intensities of only two peaks were compared per chromatogram, additional peaks (masses corresponding to unknown species) were present in each of the reaction mixtures. Attempts to remove these additional species by reverse phase purification were unsuccessful, providing additional calculation error.





PCA	k_6 / M^{-1} h^{-1}	k_{-6} / h^{-1}	**K_d / mM**
4.1	5.04 ± 0.12	$2.77 \times 10^{-1} \pm 0.10 \times 10^{-1}$	55.1
4.12	1.17 ± 0.11	$1.91 \times 10^{-1} \pm 0.37 \times 10^{-1}$	163.5

Figure 4.16. Plots of conversion against time for the reaction of AAK(dansyl) with reagent **4.1** or **4.12** under second order conditions at a concentration of 50 mM, as measured by LC-MS analysis of 325 nm chromatogram. Fits are based on a second order reversible model.

4.2.6 Protein modification

Having gained an insight of the factors governing the equilibria between aldehyde, hydrate, imine and imidazolidinone forms on a molecular level, we next decided to probe the influence of PCA functionalisation on the *i)* conversion and selectivity; and *ii)* stability of N-terminal protein labelling. From our kinetic study, *ortho*-methoxy **4.3** and *N*-oxide **4.12** emerged as the most promising candidates due to accelerated imidazolidinone formation where we could not fit a reversible reaction ($k_{-3} < 10^{-3}$ h^{-1}). We were also interested in probing the labelling efficiency of *ortho*-methoxy **4.9**, fluorinated **4.10**, and methylated **4.11**, for which we also could not fit a reversible reaction, albeit with a lower rate of imidazolidinone formation. Previously in Chapter 2, our work indicated that the success of N-terminal protein modification strategies is highly protein dependent due to differences in the local environment such as steric interactions and pH.[15] For this reason, we decided to carry out protein experiments with multiple example target proteins, screening the full range of modification reagents **4.1**-**4.12**, rather than limiting our work to the most promising candidates.

We selected RNase A (N-terminal K) and equine myoglobin (N-terminal G), as model proteins for which we previously observed variation in labelling efficiency and stability depending on the modification reagent used. We also included hydrophilic acylated surface protein A (HASPA G2S, N-terminal S) as a potentially interesting target protein for N-terminal modification. HASPs are present in *Leishmania* parasites responsible for the neglected tropical disease Leishmaniasis, and their high immunogenicity is utilised in the development of vaccines. Post-translational N-terminal myristoyl-/palmitoylation influences the ability of the protein to associate with plasma membranes, which can be mimicked using chemical methods:[16] PCA labelling could provide a means to achieve this without the requirement for installation of an N-terminal cysteine, as currently used in literature attempts to modify recombinant HASPs. Our attempts to mimic N-terminal myristoyl-/palmitoylation of another protein will be discussed further in Chapter 5.

i) Conversion and selectivity of N-terminal protein labelling

The above proteins were modified with reagents **4.1**-**4.12**, and conversion was determined by LC-MS analysis of the crude reaction mixture, and following purification by dialysis at 4 °C (*Fig. 4.17*). For myoglobin, most derivatives resulted in low levels of labelling (~5-30%), albeit with high stability to purification. We were pleased to see that *N*-oxide **4.12**, one of our most promising labelling candidates from kinetic studies, showed considerably higher levels of conversion (66%) than any of the other reagents, although with some off-target reactivity.

Modification of RNase A highlighted the protein dependency of labelling efficiency, with no PCA derivative outperforming the others across all proteins. Despite being the clear choice of reagent for myoglobin labelling, *N*-oxide **4.12** showed reduced conversion in comparison to unsubstituted parent PCA **4.1**. Instead, *ortho*-methoxy PCA **4.3** gave the highest levels of conversion, and the conjugate was stable to purification. Although slow imidazolidinone formation resulted in poor conversion with fluorinated **4.10** and *N*-methylated **4.11**, *ortho*-methoxy PCA **4.9** also showed high conversion and stability to purification. 2-PCAs showed higher conversions than equivalent 4-PCAs, and all methoxy-analogues (*ortho*- and *para*-) showed higher conversion and stability than equivalent hydroxy-analogues. Whilst initial conversion was high for *ortho*-hydroxy PCAs **4.2** and **4.8**, the conjugates showed poor stability to dialysis. We attributed decreases in conversion upon purification to the removal of transient/unstable modification: for *ortho*-hydroxy PCAs this may have been a combination of reversibility (from high k_{-3} values) and off-target lysine modification (due to imine stabilisation through hydrogen bonding).

Similar trends in conversion and stability to purification emerged upon modification of HASPA G2S in comparison to RNase A; note that an incomplete dataset was obtained due to insufficient quantities of protein. However, differences in selectivity between reagents became considerably more apparent. Notably, the highest levels of double modification were observed for reagents for which we had previously seen high percentages of imine formation (e.g., *N*-oxide **4.12**) and high equilibrium constants for imine formation (e.g., fluorinated **4.10**). Concerningly, the level of double modification was not reduced upon purification by dialysis, potentially reflecting the increased stability of the off-target imine formation.

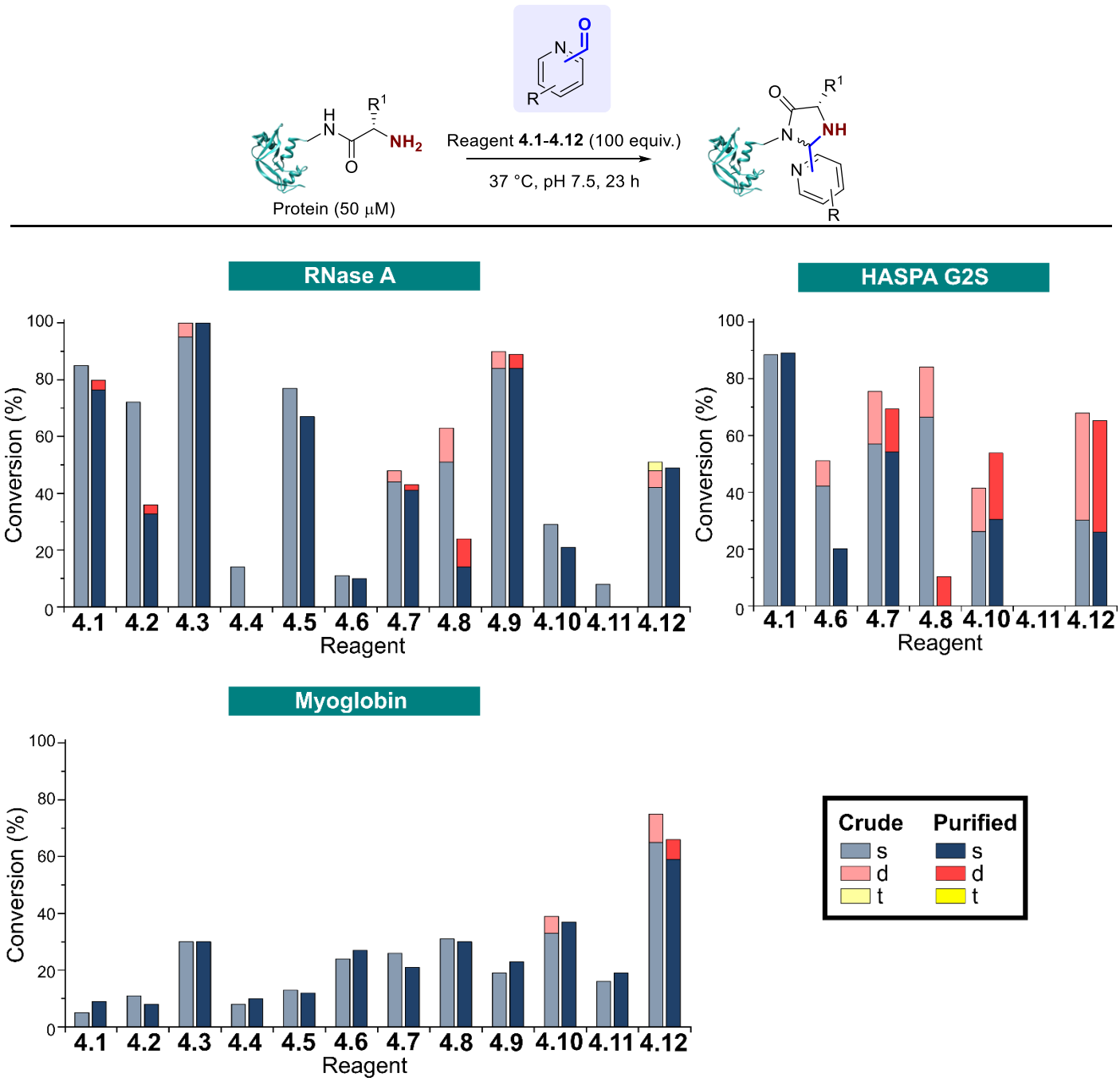


Figure 4.17. Screening protein labelling of RNase A, myoglobin, and HASPA G2S with reagents **4.1**-**4.12**. RNase A, myoglobin and HASPA G2S (50 μM) were modified with reagents **4.1**-**4.12** (100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 23 h, and the conversion was determined by LC-MS before and after purification by dialysis (s = single, d = double, t = triple modification). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

Throughout our previous work up to this point, we have expected transient interactions captured by analysis of crude samples to be removed upon purification, as reflected by low levels of multiple modifications generally observed in purified samples. However, without tryptic digestion of samples, we cannot confirm with full confidence that single modifications can be solely attributed to the N-terminus. We decided to compare selectivity for the N-terminus over lysine residues *via* modification of CjX183-D WT (0 lysine

residues) and CjX183-D R51K (1 lysine residue), as done previously in Chapter 2 (*Fig. 4.18*). Whilst we cannot locate the additional labelling sites (as attempted in Chapter 2), we can compare the extent of modification upon introduction of a lysine residue, as an indication of lysine reactivity. We were pleased to find that the introduction of the Lys residue in CjX183-D R51K resulted in very little change in the distribution of multiple modifications, indicating low reactivity of the lysine residue with PCAs. However, we cannot rule out lysine reactivity in other proteins due to differences in accessibility and reactivity of the side-chain in different protein environments.

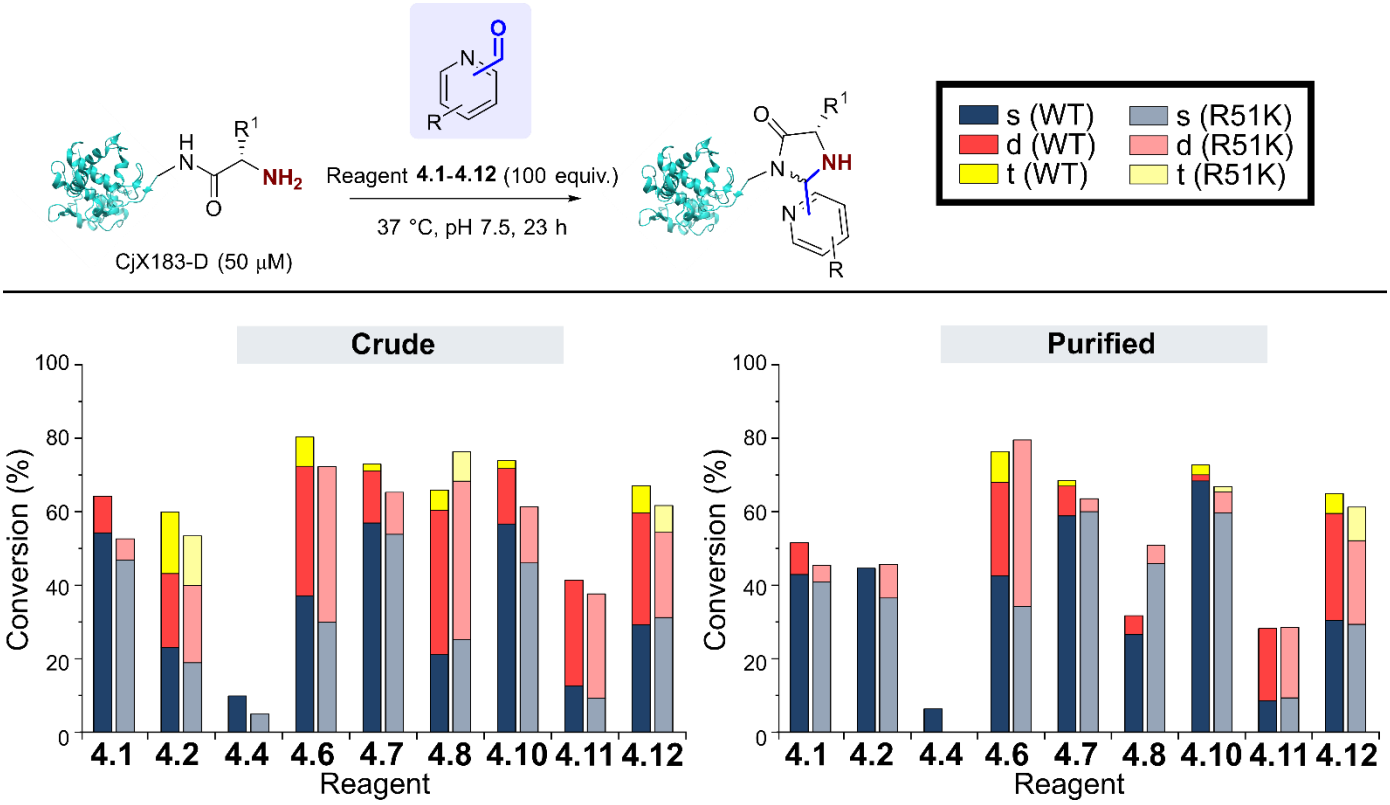


Figure 4.18. Screening protein labelling of CjX183-D WT (0 Lys) and CjX183-D R51K (1 Lys) with reagents **4.1**-**4.12** *(note: data was not collected for reagents **4.3**, **4.5** and **4.9** due to insufficient quantities of protein).* CjX183-D WT and CjX183-D R51K (50 μM) were modified with reagents **4.1**-**4.12** (100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 23 h, and the conversion was determined by LC-MS before and after purification by dialysis (s = single, d = double, t = triple modification). Figure built using structural data obtained by Grabarczyk *et al.* (CjX183-D, PDB 4V2K)[17].

ii) Stability of N-terminal protein labelling

We next evaluated conjugate stability, as detailed in Chapter 2. RNase A and myoglobin were modified with **4.1**-**4.12**, and purified by dialysis at 4 °C for 24 h to remove excess reagents. Purified conjugates were then subjected to a range of different conditions (pH 7 at 4 °C, 22 °C, and 37 °C; pH 6 and 8 at 22 °C) under dialysis conditions in sodium phosphate buffer, to remove any unconjugated reagents that were released over time and prevent reattachment. Conjugate stability was then monitored at different time points over 1 week (*Fig. 4.19-4.20*). In some cases, e.g. the modification of RNase A with **4.4** and **4.11**, the low initial conversion prior to purification prevented an accurate analysis being undertaken, and these samples were therefore omitted from the study.

Following conjugate purification, we had increased (but not certain) confidence that the single modifications could be attributed to the protein N-terminus. In general, all myoglobin conjugates showed high stability under all conditions. For RNase A, we unfortunately saw no improvement in conjugate stability for any of the derivatives explored in comparison to unsubstituted PCA **4.1**. Interestingly, despite the large drop in conversion of RNase A upon purification of *ortho*-hydroxy PCA **4.2** and **4.8** conjugates, the relative decrease in conversion of purified samples over time was not significantly worse than that of unsubstituted PCAs **4.1** and **4.7**. Very little further decrease in conversion was observed for conjugates at 4 °C, suggesting that once the transient modifications were removed by the initial dialysis at 4 °C, the remaining conjugates were of higher stability and had reduced susceptibility to hydrolysis.

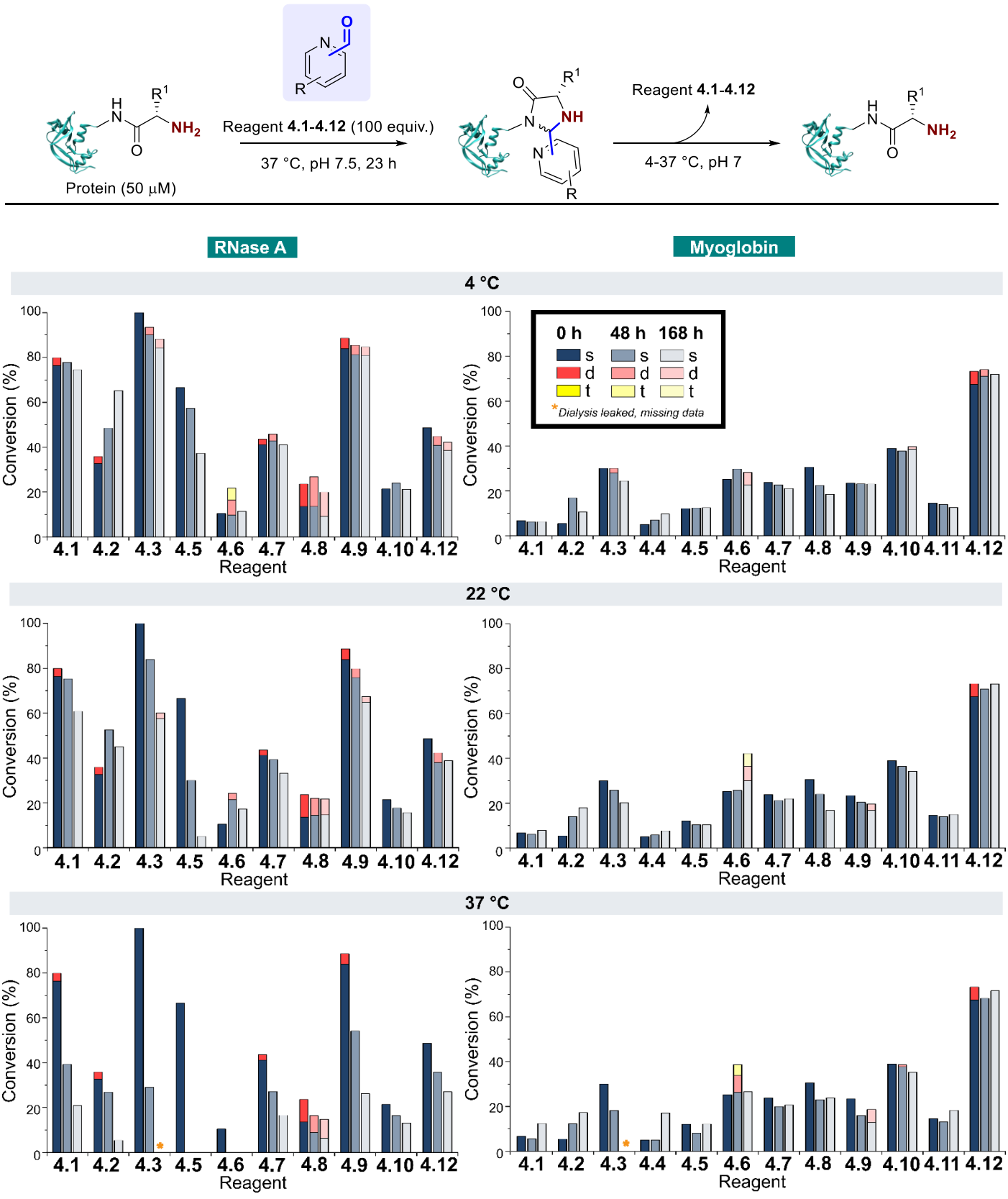


Figure 4.19. Stability of protein conjugates over time. After initial conjugation at 37 °C for 23 h, proteins were purified by dialysis at 4 °C. Samples were then incubated under the specified conditions at pH 7 and conjugation monitored over time by LC-MS (s = single, d = double, t = triple modification). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

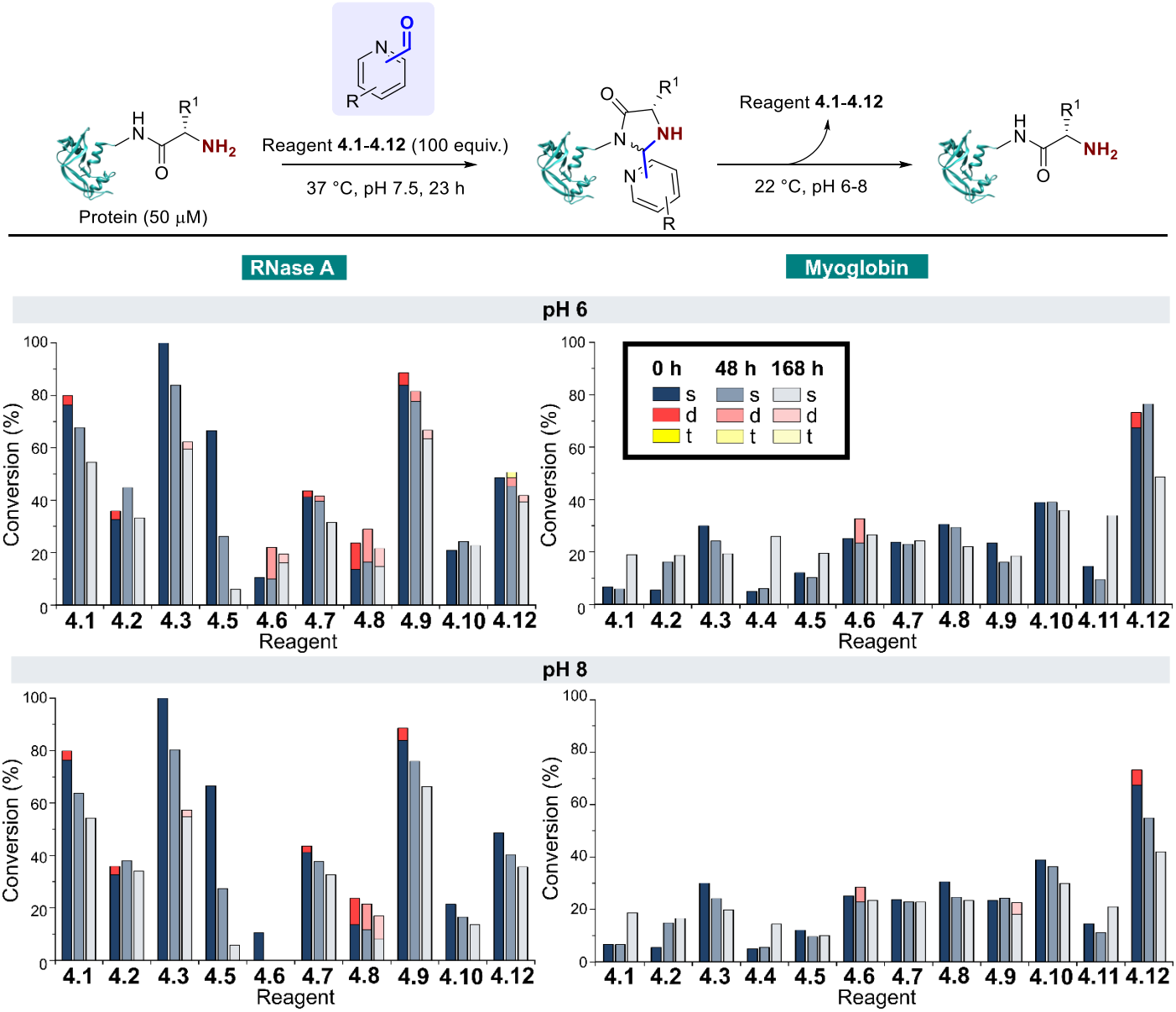


Figure 4.20. Stability of protein conjugates over time. After initial conjugation at 37 °C for 23 h, proteins were purified by dialysis at 4 °C. Samples were then incubated under the specified conditions at 22 °C and conjugation monitored over time by LC-MS (s = single, d = double, t = triple modification). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[5].

4.3 Conclusion

Our results highlight the complex influence of PCA functionalisation on N-terminal labelling, with a combination of factors influencing the interplay between hydration, imine formation, and imidazolidinone formation. By varying both the nature and location of substituents, we were able to isolate the different roles to identify PCA derivatives with increased rates of imidazolidinone formation for which a reversible model could not be fitted, as reflected by the highest conversions of model proteins. Our approach provides a valuable step forward in the goal to achieve a bioconjugation strategy applicable to a wide range of protein targets with high efficiency, selectivity, and stability.

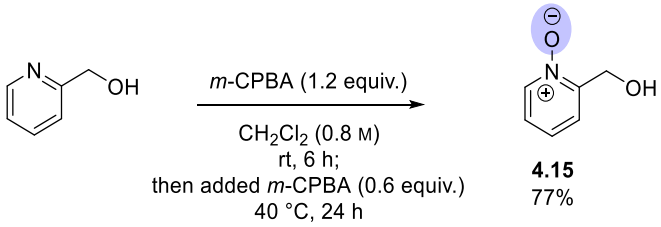
Variability in the PCA candidate with the highest protein labelling efficiency between protein targets reflected the protein dependence seen previously during protein modification screens (Chapter 2), and further emphasises the need to screen reagent libraries to identify the optimal reagent for a target protein. We have demonstrated that even subtle changes in reagent design between similar protein modification reagents can significantly impact reactivity and N-terminal targeting. We have identified PCAs **4.3** and **4.12**, both with hydrogen bonding acceptor potential, as key candidates to include in reagent libraries. However, the promising dipeptide modification kinetics for these reagents unfortunately did not translate in protein conjugate stability tests, and increased imine stabilisation with *N*-oxide **4.12** raised concerns about the stabilisation of off-target lysine side-chain modifications.

4.4 Experimental

4.4.1 Reagent synthesis

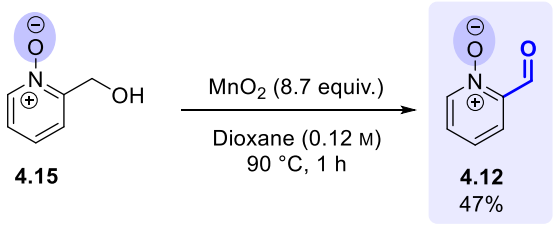
a) *PCA* ***4.12***

2-(hydroxymethyl)pyridine-1-oxide (4.15)



m-CPBA (0.96 g, 5.5 mmol, 1.2 equiv.) was added to a solution of 2-(hydroxymethyl)pyridine (0.44 mL, 4.6 mmol, 1.0 equiv.) in CH_2Cl_2 (6 mL, 0.8 M) and the reaction mixture was stirred at rt for 6 h. At this point, TLC analysis indicated that the reaction was incomplete, so a further portion of *m*-CPBA (0.47 g, 2.7 mmol, 0.59 equiv.) was added and the reaction mixture was stirred at 40 °C for 24 h. The reaction was then allowed to cool to rt, and concentrated under reduced pressure. The resulting off-white solid obtained was washed with Et_2O (100 mL) and purified by flash column chromatography (0-7.5% MeOH:CH_2Cl_2, R_f 0.46 in 10% MeOH:CH_2Cl_2). Pure fractions were concentrated under reduced pressure to afford the title compound (0.44 g, 3.5 mmol, 77%) as a white solid with spectroscopic data in accordance with the literature.[18] **^{1}H NMR** (400 MHz, $CDCl_3$) $\delta_{H:}$ 4.81 (2H, s, -C<u>H</u>$_2$), 7.24-7.42 (3H, m, Ar<u>H</u>), 8.25 (1H, ddd, *J* = 6.3, 0.9, 0.9 Hz, Ar<u>H</u>); **mp** 115-119 °C {Lit.[18] 130-132 °C}.

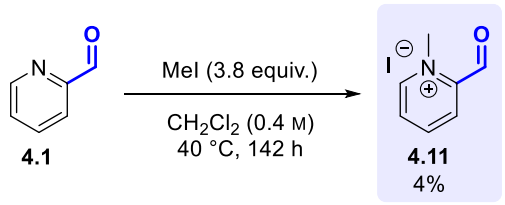
2-formylpyridine 1-oxide (4.12)



MnO_2 (0.91 g, 10.4 mmol, 8.7 equiv.) was added to a solution of compound **4.15** (0.15 g, 1.2 mmol, 1.0 equiv.) in 1,4-dioxane (10 mL, 0.12 M) and the reaction mixture was stirred at 90 °C for 1 h. The still hot reaction mixture was filtered through Celite and washed with hot 1,4-dioxane (100 mL). The filtrate was concentrated under reduced pressure and the resulting yellow solid was purified by flash column chromatography (0-5% MeOH:CH_2Cl_2, R_f 0.33 in 10% MeOH:CH_2Cl_2). Pure fractions were concentrated under reduced pressure to afford the title compound (69 mg, 0.6 mmol, 47%) as a pale-yellow solid with spectroscopic data in accordance with the literature.[19] **^{1}H NMR** (400 MHz, $CDCl_3$) $\delta_{H:}$ 7.33 (1H, dd, $J_1 = J_2$ = 7.7 Hz, Ar<u>H</u>5), 7.46 (1H, ddd, *J* = 7.7, 6.5, 2.2 Hz, Ar<u>H</u>4), 7.82 (1H, dd, *J* = 7.7, 2.2 Hz, Ar<u>H</u>6), 8.21 (1H, d, *J* = 6.5 Hz, Ar<u>H</u>3), 10.64 (1H, s, CO<u>H</u>); **mp** 54-57 °C {Lit.[20] 63-65 °C}.

b) *PCA* ***4.11***

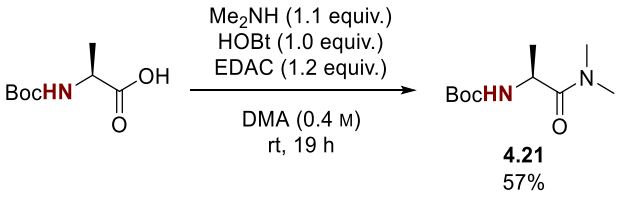
2-formyl-1-methylpyridin-1-ium iodide (4.11)



Methyl iodide (0.45 mL, 7.1 mmol, 3.8 equiv.) was added to a solution of 2-PCA **4.1** (0.18 mL, 1.9 mmol, 1.0 equiv.) in CH_2Cl_2 (5 mL, 0.4 M), and the reaction mixture was stirred at 40 °C for 142 h. After cooling to room temperature, the reaction mixture was filtered to obtain the title compound as a brown oil (18 mg, 0.07 mmol, 4%), which formed the hydrate on storage with spectroscopic data in accordance with the literature.[21] **^{1}H NMR** (400 MHz, D_2O) δ_H: 4.42 (3H, s, Me), 6.42 (1H, s, ArC<u>H</u>$(OH)_2$), 8.00 (1H, ddd, *J* = 7.8, 6.1, 1.6 Hz, Ar<u>H</u>5), 8.35 (1H, dd, *J* = 8.1, 1.6 Hz, Ar<u>H</u>3), 8.59 (1H, ddd, *J* = 8.1, 7.8, 1.4 Hz, Ar<u>H</u>4), 8.80 (1H, dd, *J* = 6.2, 1.4 Hz, Ar<u>H</u>6).

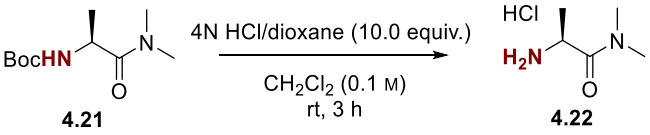
c) *Dimethylamide* ***4.20***

***Tert*-butyl-(*S*)-(1-(dimethylamino)-1-oxopropan-2-yl)carbamate (4.21)**



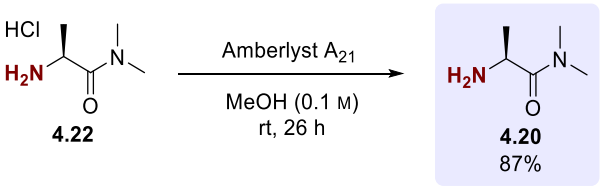
Dimethylamine (1.1 mL, 8.7 mmol, 1.1 equiv., 40 wt.% in H_2O) was added to a solution of Boc-Ala-OH (1.50 g, 7.9 mmol, 1.0 equiv.) in DMA (20 mL, 0.4 M). HOBt (1.07 g, 7.9 mmol, 1.0 equiv.) was added and the reaction mixture was cooled to 0 °C. EDAC (1.52 g, 9.8 mmol, 1.2 equiv.) was then added at 0 °C, and the reaction mixture was allowed to warm to rt and stirred for 19 h. The reaction mixture was then added to H_2O (20 mL) and extracted with EtOAc (3 × 40 mL). The organic layers were combined, washed with H_2O (20 mL) and saturated brine (20 mL), dried over $MgSO_4$, filtered, and concentrated. The resulting colourless liquid was purified by flash column chromatography (10-50% EtOAc in Petrol, R_f 0.24, visualised using ninhydrin stain) and pure fractions were concentrated under reduced pressure to obtain the title compound as a colourless liquid (0.98 g, 4.5 mmol, 57%) with spectroscopic data in accordance with the literature.[22] **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H: 1.26 (3H, d, *J* = 7.0 Hz, CH<u>Me</u>), 1.39 (9H, s, Boc), 2.93 (3H, s, N<u>Me</u>), 3.02 (3H, s, N<u>Me</u>), 4.59 (1H, dq, $J_1 = J_2$ = 7.0 Hz, C<u>H</u>Me), 5.49 (1H, d, *J* = 7.0 Hz, N<u>H</u>Boc).

(*S*)-2-amino-*N*,*N*-dimethylpropanamide hydrogen chloride (4.22)



Hydrochloric acid (5.8 mL, 23.1 mmol, 10.0 equiv., 4N in dioxane) was added to a solution of compound **4.21** (0.50 g, 2.3 mmol, 1.0 equiv.) in CH_2Cl_2 (23 mL, 0.1 M) and the reaction mixture was stirred at rt for 3 h. The solvent was removed under reduced pressure and the resulting residue was azeotroped with CH_2Cl_2 (25 mL). The white solid obtained (0.38 g) was used immediately in the subsequent reaction without characterisation.

(*S*)-2-amino-*N*,*N*-dimethylpropanamide (4.20)



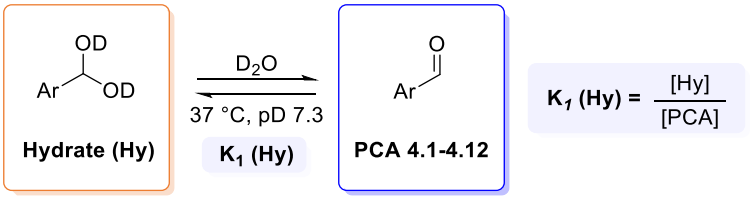
Amberlyst A_{21} (2.00 g) was added to a solution of compound **4.22** (0.21 g, 1.4 mmol, 1.0 equiv.) in MeOH (10 mL, 0.1 M) and the reaction mixture was stirred at rt for 26 h, followed by filtration through celite. To remove dissolved celite, the filtrate was concentrated under reduced pressure, dissolved in CH_2Cl_2, and filtered. The resulting filtrate was concentrated under reduced pressure to obtain the title compound (0.13 g, 1.1 mmol, 87%) as a yellow liquid. **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H: 1.36 (3H, d, *J* = 6.8 Hz, CHMe), 2.95 (3H, s, NMe), 3.04 (3H, s, NMe), 4.10 (1H, br s, CHMe), 4.35 (2H, br s, NH).

4.4.2 Hydrate, imine and imidazolidinone formation

4.4.2.1 Hydrate formation

^{1}H NMR spectra of PCAs **4.1**-**4.12** (50 mM in 100 mM deuterated Na phosphate buffer, pD 7.3) were recorded at 37 °C to determine the ratio between aldehyde and hydrate forms. It was assumed that equilibria were reached quickly before NMR spectra were recorded. Note: due to reduced quantities of material available, the data for 2-PCA **4.11** was collected at a reduced concentration of 15 mM. The concentration of aldehyde [**PCA**] and hydrate [**Hy**] were calculated *via* the relative integrals of diagnostic ^{1}H NMR signals outlined in *Fig. S4.3a*. The equilibrium constant for hydrate formation (**K_1 (Hy)**) was calculated using *Equation 4.1*:

$$\mathbf{K_1\,(Hy)} = \frac{[\mathbf{Hy}]}{[\mathbf{PCA}]} \qquad \textit{(Equation 4.1)}$$



PCA	Time / min	Aldehyde	Hydrate	**K_1 (Hy)**
4.1	75	69%	31%	0.44
4.2	105	97%	3%	0.03
4.3	60	76%	24%	0.31
4.4	110	99%	1%	0.01
4.5	35	94%	6%	0.06
4.6	50	71%	29%	0.41
4.7	135	53%	47%	0.90
4.8	20	90%	10%	0.11
4.9	5	57%	43%	0.75
4.10	80	24%	76%	3.17
4.11	35	<1%	100%	>1000
4.12	50	12%	88%	7.08

Table S4.1. Hydrate/aldehyde equilibria (initial PCA concentration: 50 mM; 15 mM for PCA **4.11**).

4.4.2.2 Imine formation

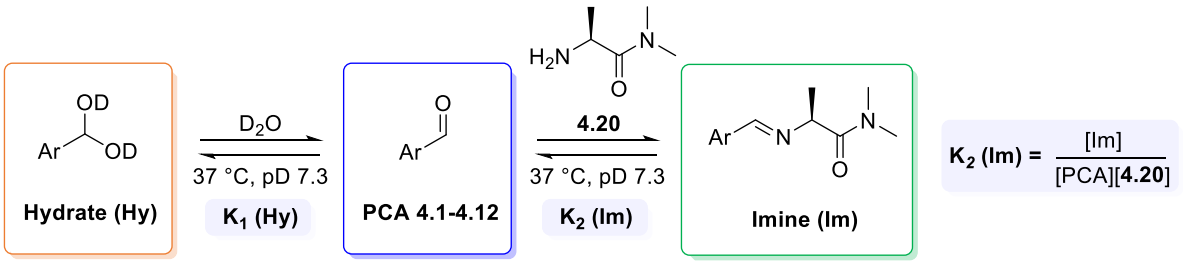
An aliquot of each PCA/hydrate solution (300 µL, 50 mM, 15 µmol, 1 equiv., in 100 mM deuterated Na phosphate buffer, pD 7.3) was added to a solution of dimethyl amide **4.20** (300 µL, 50 mM, 15 µmol, 1 equiv., in 100 mM deuterated Na phosphate buffer, pD 7.3) and the reaction mixtures were incubated at 37 °C for up to 48 h. ^{1}H NMR spectra were recorded at 37 °C, at the time-points indicated in *Table S4.2*. At this point, equilibrium had been reached in all cases. Note: due to the limited quantity of material available, the data for 2-PCA **4.11** was collected at a reduced concentration; an aliquot of the PCA **4.11**/hydrate solution (150 µL, 15 mM, 2.3 µmol, 1 equiv., in 100 mM deuterated Na phosphate buffer, pD 7.3) was added to dimethyl amide **4.20** (150 µL, 15 mM, 2.3 µmol, 1 equiv., in 100 mM deuterated Na phosphate buffer, pD 7.3).

The concentration of aldehyde [**PCA**], hydrate [**Hy**] and imine [**Im**] species were calculated *via* the relative integrals of diagnostic ^{1}H NMR signals outlined in *Fig. S4.3b*. The concentration of unreacted **4.20** was calculated using *Equation 4.2*, where $[\mathbf{4.20}]_{t=0}$ is the starting concentration of **4.20**.

$$[\mathbf{4.20}] = [\mathbf{4.20}]_{t=0} - [\mathbf{Im}] \qquad \textit{(Equation 4.2)}$$

The equilibrium constant for imine formation (**K_2 (Im)**) was then calculated using *Equation 4.3*:

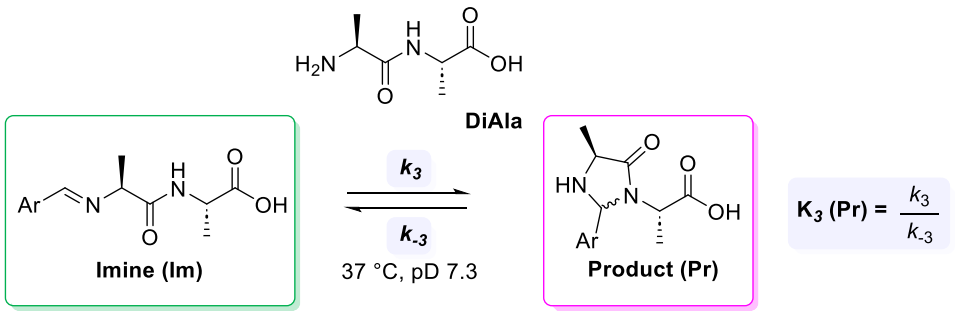
$$\mathbf{K_2\ (Im)} = \frac{\mathbf{[Im]}}{\mathbf{[PCA][4.20]}} \qquad \textit{(Equation 4.3)}$$



PCA	Time / h	Aldehyde	Hydrate	Imine	K_2 (Im) / M^{-1}
4.1	1	63%	28%	9%	6.28
4.2	1	77%	2%	22%	14.65
4.3	1	69%	23%	8%	5.04
4.4	1	99%	1%	<1%[b]	-
4.5	1	89%	5%	5%	2.37
4.6	1	61%	25%	14%	10.67
4.7	1	48%	40%	12%	11.36
4.8	17	82%	6%	12%	6.65
4.9	1	51%	37%	12%	10.70
4.10	1	19%	70%	11%	26.02
4.11	18	<1%[b]	100%	<1%[b]	-
4.12[a]	17	<1%[b]	85%	15%	8.30[a]

Table S4.2. Hydrate/aldehyde equilibria (initial PCA/**4.20** concentration: 25 mM; 7.5 mM for PCA **4.11**). [a] For PCA **4.12**, the absence of aldehyde made it impossible to calculate K_2 directly. Instead, K_{obs} is given for the equilibrium between hydrate and imine. [b] Levels were below the detection sensitivity of the measurement.

4.4.2.3 Imidazolidinone formation



A solution of DiAla (150 µL, 100 mM, 15 µmol, 1 equiv.) was added to solutions of reagents **4.1**-**4.12** (150 µL, 100 mM, 15 µmol, 1 equiv.), both in deuterated sodium phosphate buffer (100 mM, pD 7.3). The reactions were incubated at 37 °C for 16 h and conversion was followed by ^{1}H NMR spectroscopy at 30 min intervals.

At each timepoint ($t = x\ h$), the concentration of aldehyde [**PCA**], hydrate [**Hy**], imine [**Im**] and sum of imidazolidinone product [**Pr**] species were calculated *via* the relative integral ratios of the diagnostic ^{1}H NMR signals outlined in *Fig. S4.3c*. Concentration values were normalised using a correction factor (**CF**) in *Equation 4.4* to remove background noise from the imidazolidinone signals, based on the assumption that the initial rate was linear over the first 3 time-points collected, and that $[\mathbf{Pr}]_{t=0\ h} = 0$. Correction factors were calculated as the y-intercept of the linear regression line of the plot of the sum of integrals of product diastereomers over time (for the first 3 time-points collected), as demonstrated in *Fig. S4.1* for PCA **4.4**.

$$\mathbf{Conversion} = 100 \times \frac{\int \mathbf{Pr} - \mathbf{CF}}{\int \mathbf{PCA} + \int \mathbf{Hy} + \int \mathbf{Im} + \int \mathbf{Pr} - \mathbf{CF}} \qquad \textit{(Equation 4.4)}$$

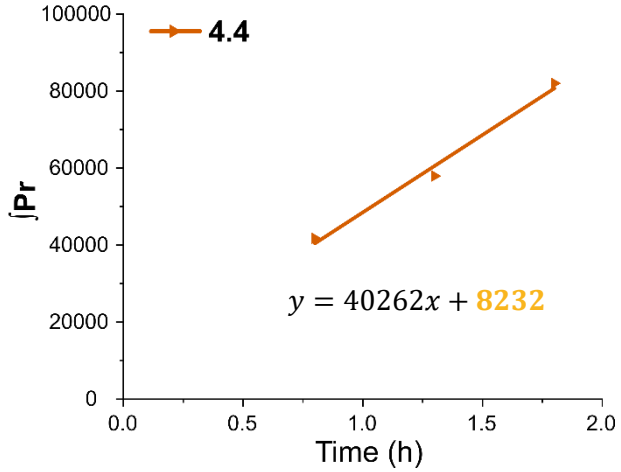
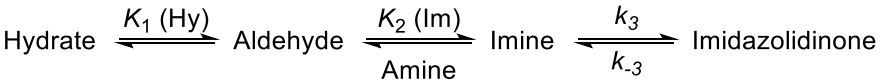


Figure S4.1. Example calculation of correction factor (**CF**) for PCA **4.4**. Fit is a linear regression, based on the initial rates model.

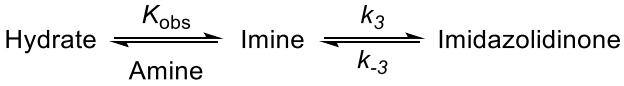
Data were fit to a multi-step reversible kinetic model in Copasi 4.34.251, encompassing hydrate, imine, and imidazolidinone formation, to estimate values of k_3 and k_{-3}.



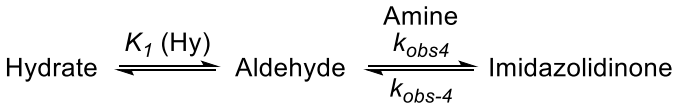
Since the rates of reversible hydrate and imine formation (k_1/k_{-1} and k_2/k_{-2} respectively) were >4-orders of magnitude greater than k_3 and k_{-3}, the exact values of these parameters does not affect imidazolidinone formation. Instead, the values of K_1 (Hy) and K_2 (Im) calculated above could be used to fix the ratio of the forward and back rates for each step since $K_x = k_x/k_{-x}$.

k_3 and k_{-3} 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_3 10^{-12}-10^5 s^{-1}; k_{-3} 10^{-12}-10^5 s^{-1}.

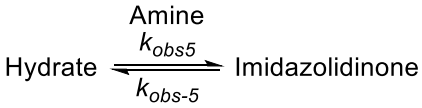
For PCA **4.12**, where no aldehyde was observed in the experiment outlined in section *4.4.2.2*, K_{obs} was calculated as a pseudo-equilibrium constant for the conversion of hydrate to imine. K_{obs} was built into the multi-step kinetic model described above, giving k_3 and k_{-3} values that were comparable to the other PCAs.

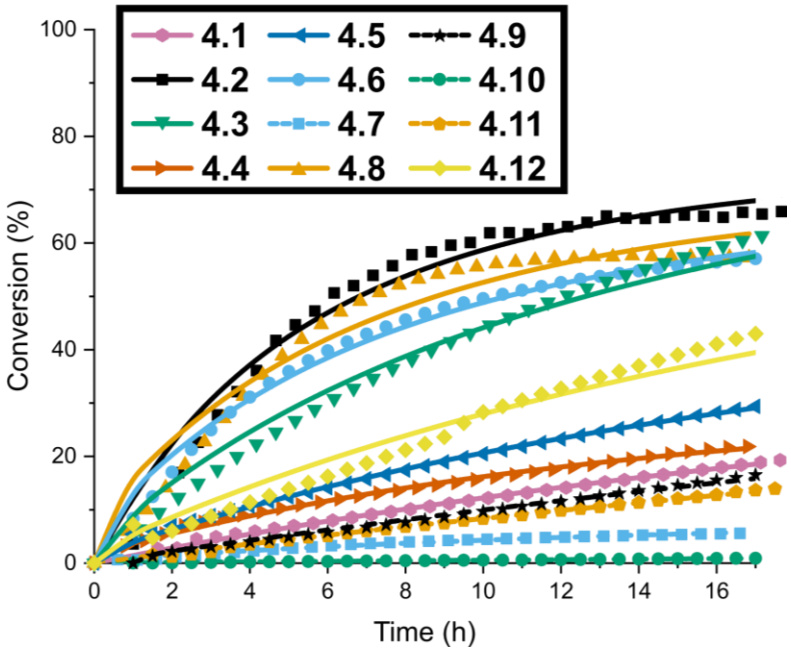


For PCA **4.4**, where no imine was observed in the experiment outlined in section *4.4.2.2*, a steady-state approximation was applied to the reaction whereby the concentration of imine was assumed to be constant throughout the reaction. The data was therefore fitted to give the *observed* second-order rate constants k_{obs4} and k_{obs-4} which cannot be compared to the other rate constants generated.



For PCA **4.11**, where no aldehyde or imine was observed in the experiments outlined in sections *4.4.2.1* and *4.4.2.2*, a steady-state approximation was applied to the reaction whereby the concentrations of aldehyde and imine were assumed to be constant throughout the reaction. The data was therefore fitted to give the *observed* second-order rate constants k_{obs5} and k_{obs-5} which cannot be compared to the other rate constants generated.



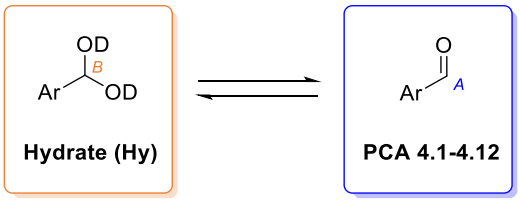


PCA	k_3 / h^{-1}	k_{-3} / h^{-1}	**K_3**
4.1	$9.3 \pm 0.09 \times 10^{-2}$	$1.2 \pm 0.1 \times 10^{-2}$	7.7
4.2	0.43 ± 0.01	$2.1 \pm 0.3 \times 10^{-2}$	19.5
4.3	0.49 ± 0.02	*	-
4.4[a]	$k_{obs4} = 0.44 \pm 0.005$ M^{-1} h^{-1}	$k_{obs-4} = 3.4 \pm 0.1 \times 10^{-2}$	12.9 M^{-1}
4.5	0.28 ± 0.02	$1.3 \pm 0.1 \times 10^{-2}$	21.6
4.6	0.42 ± 0.01	$2.2 \pm 0.2 \times 10^{-2}$	18.8
4.7	0.03 ± 0.0007	$5.4 \pm 0.4 \times 10^{-2}$	0.5
4.8	0.54 ± 0.02	$1.7 \pm 0.4 \times 10^{-2}$	31.4
4.9	$5.4 \pm 0.06 \times 10^{-2}$	*	-
4.10	$2.5 \pm 0.06 \times 10^{-3}$	*	-
4.11[a]	$k_{obs5} = 0.18 \pm 0.002$ M^{-1} h^{-1}	*	-
4.12[b]	0.14 ± 0.005	*	-

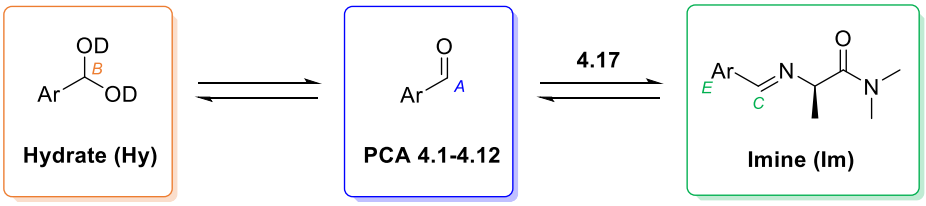
Table S4.3. Imidazolidinone formation (initial PCA/DiAla concentration: 50 mM). **Fitted* $k_{-3} < 10^{-12}$ s^{-1}; [a] For PCAs **4.4** and **4.11**, data were fitted to give second-order rate constants as detailed above; [b] For PCA **4.12** data were fitted to a simplified model as detailed above.

4.4.2.4 Representative ^{1}H NMR spectra

a)



b)



c)

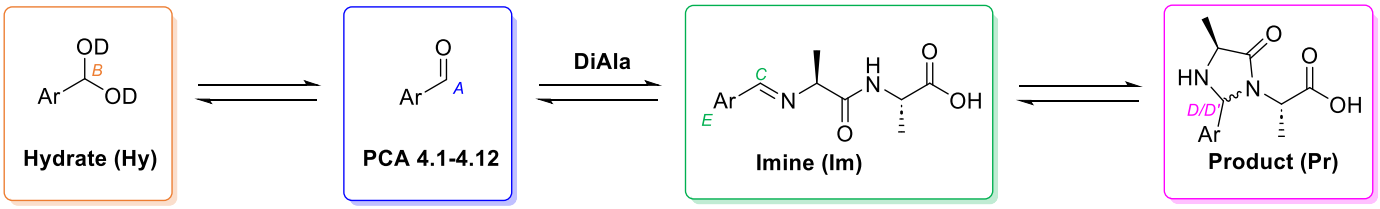
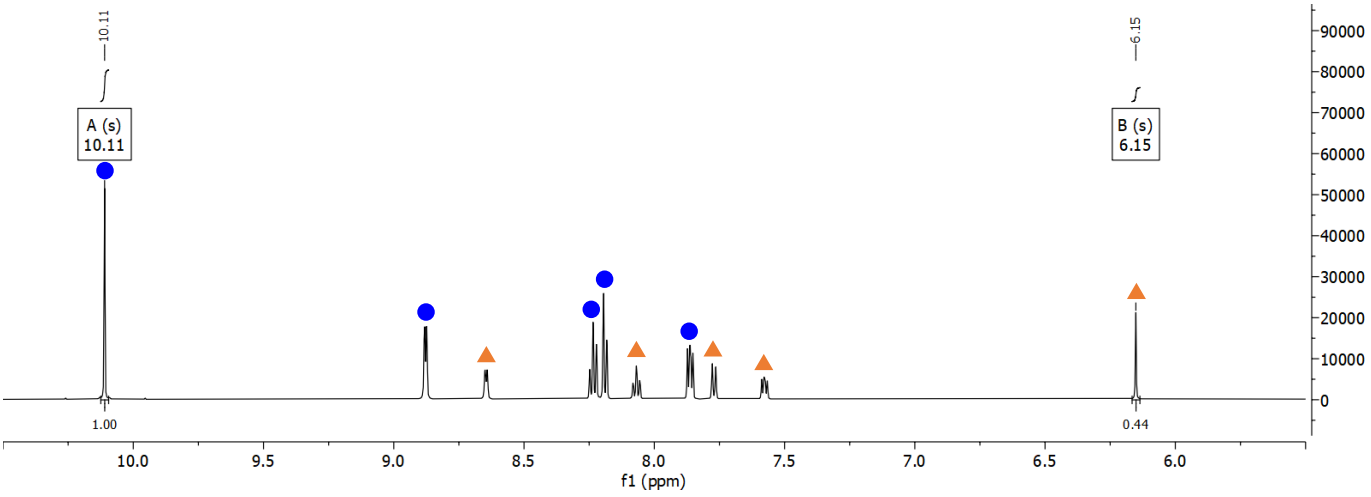


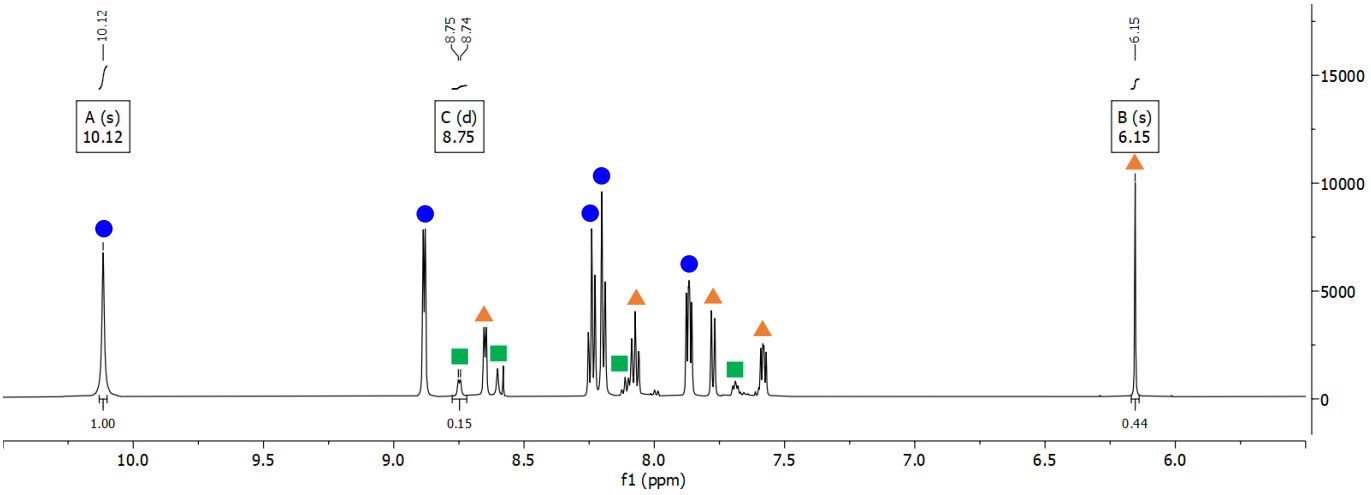
Figure S4.2. Overview of the diagnostic protons used to calculate: ***a)*** the equilibrium constant for hydrate formation; ***b)*** the equilibrium constant for imine formation; ***c)*** rate constants for imidazolidinone formation.

PCA 4.1

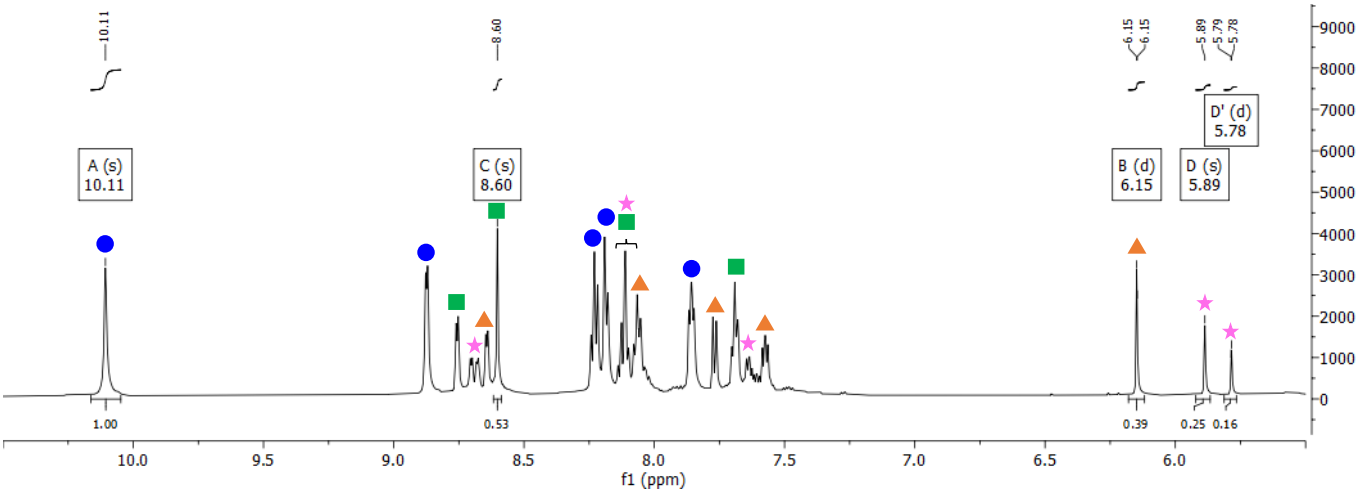
a)



b) (t = 1 h)

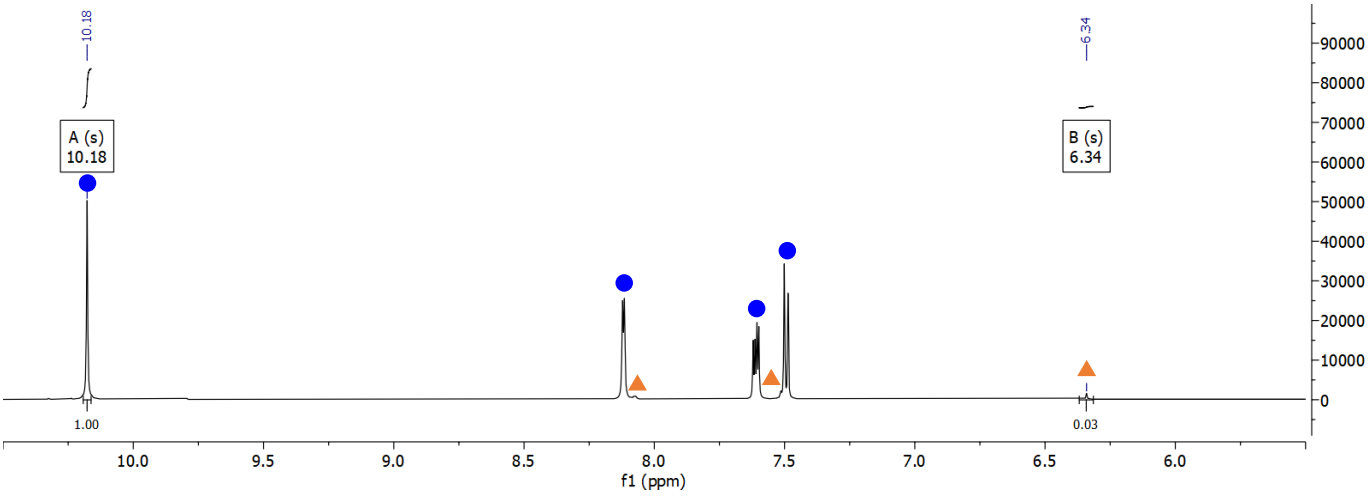


c) (t = 17 h)

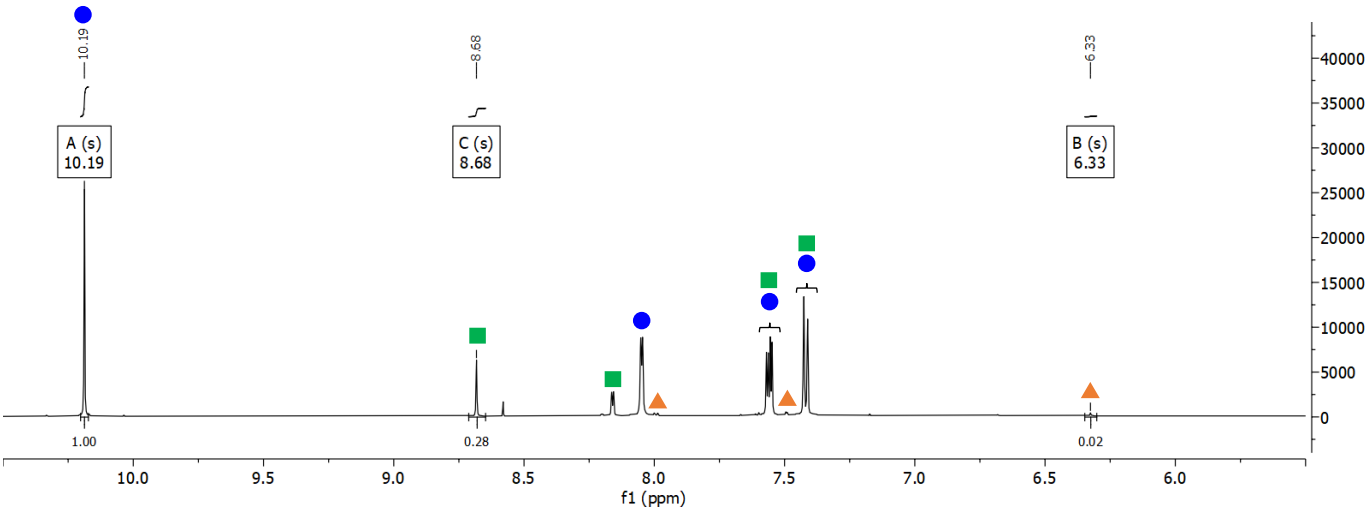


PCA 4.2

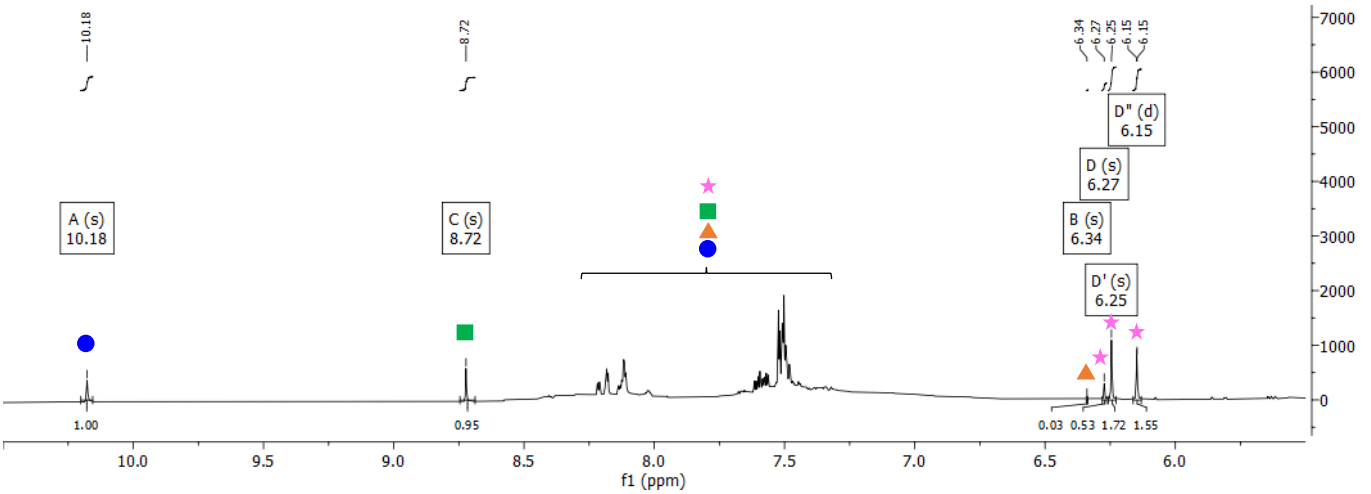
a)



b) (t = 1 h)

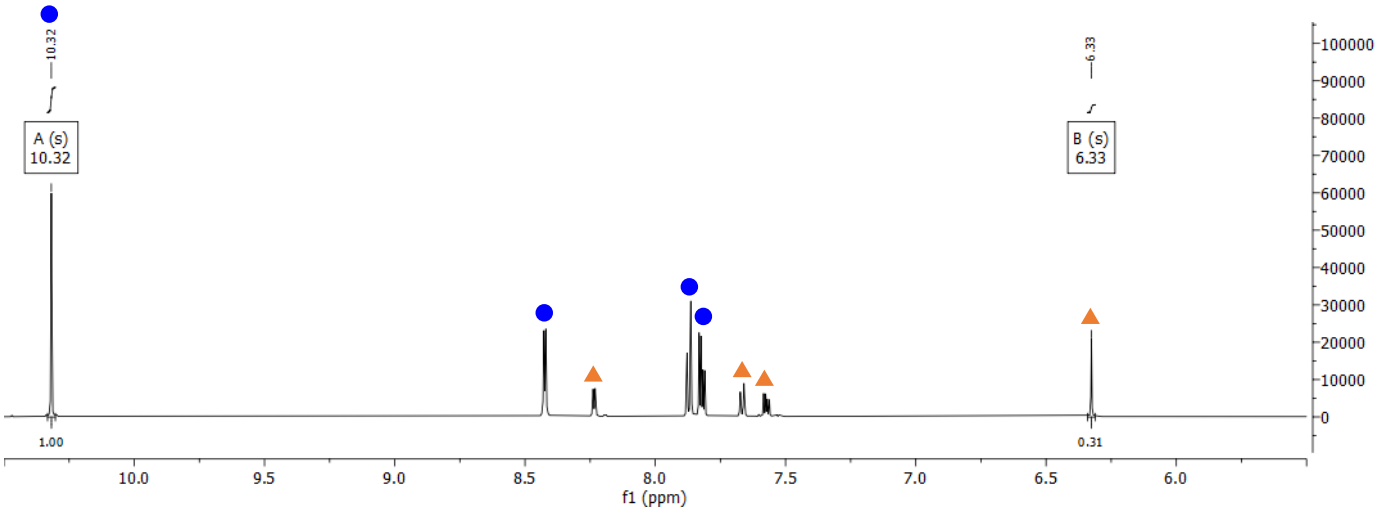


c) (t = 17 h)

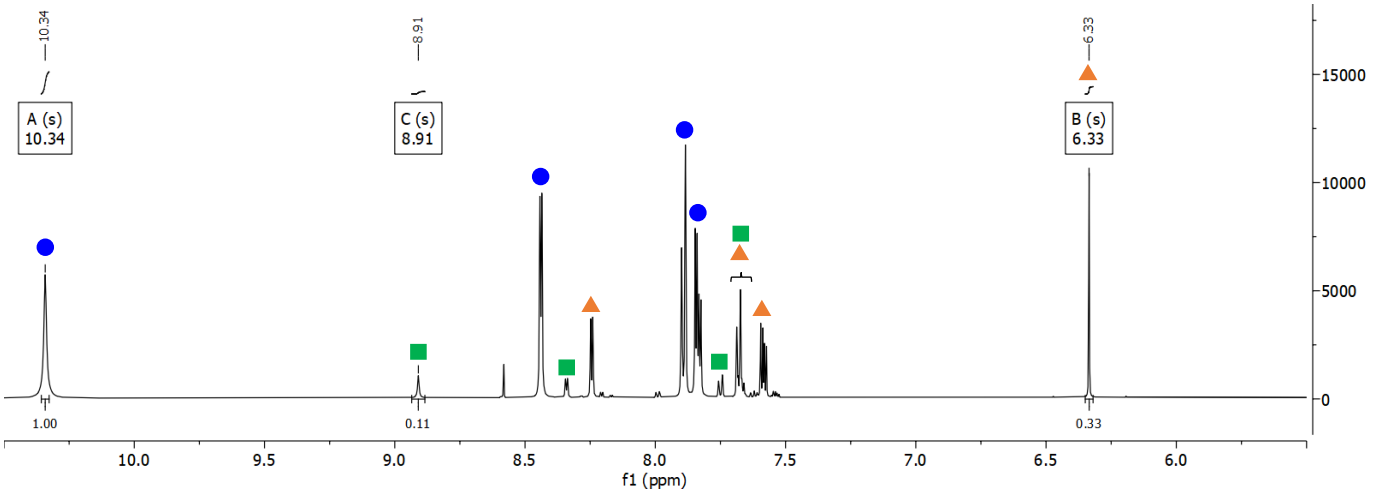


PCA 4.3

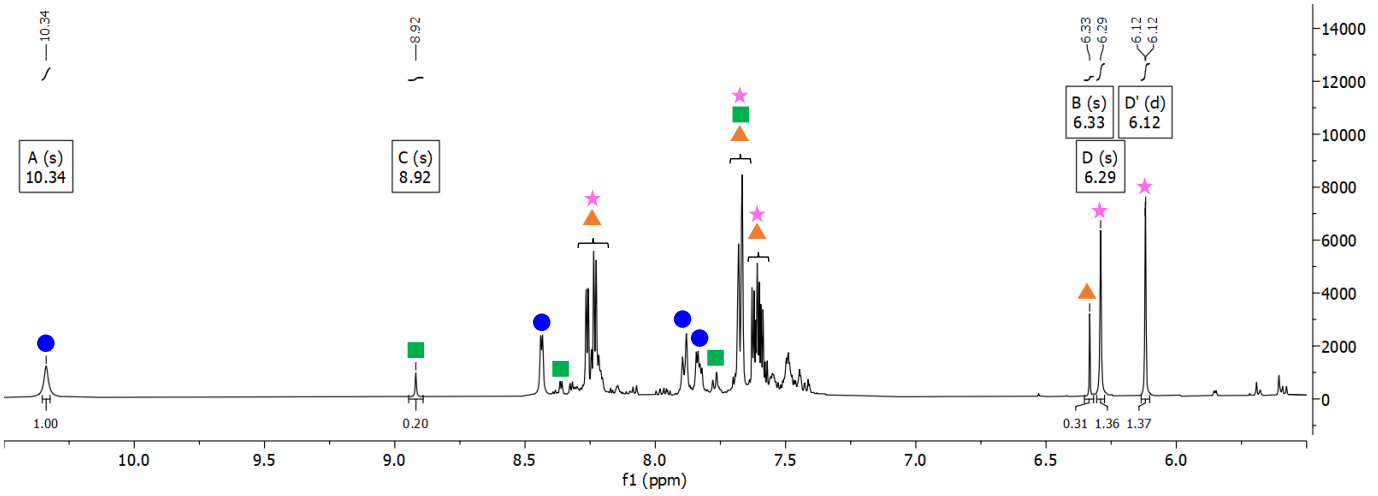
a)



b) (t = 1 h)

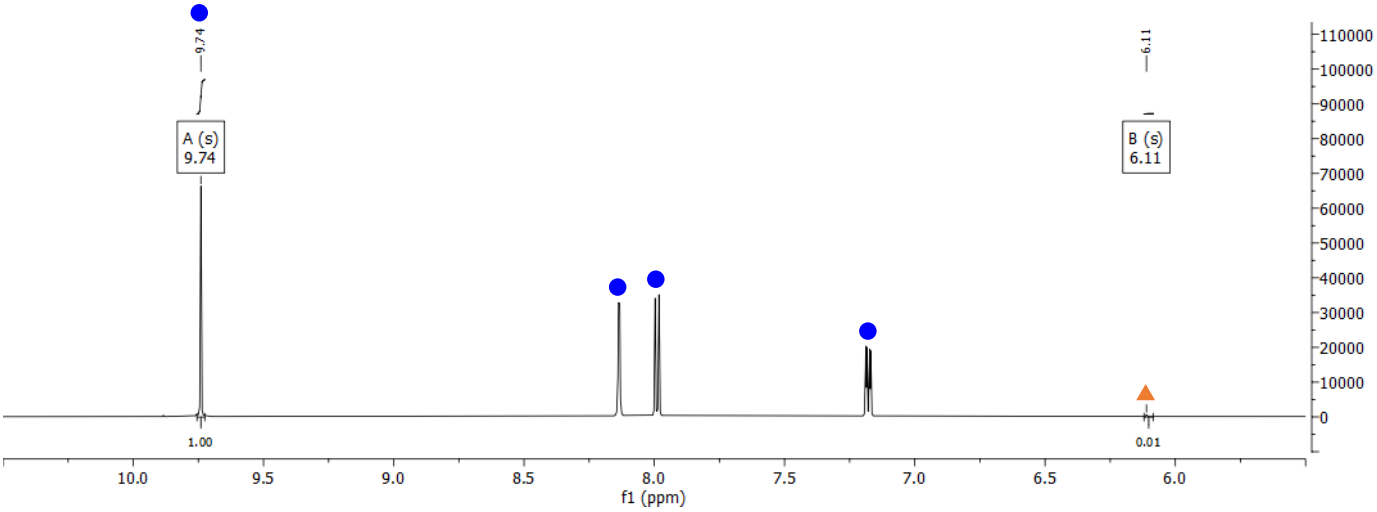


c) (t = 17 h)

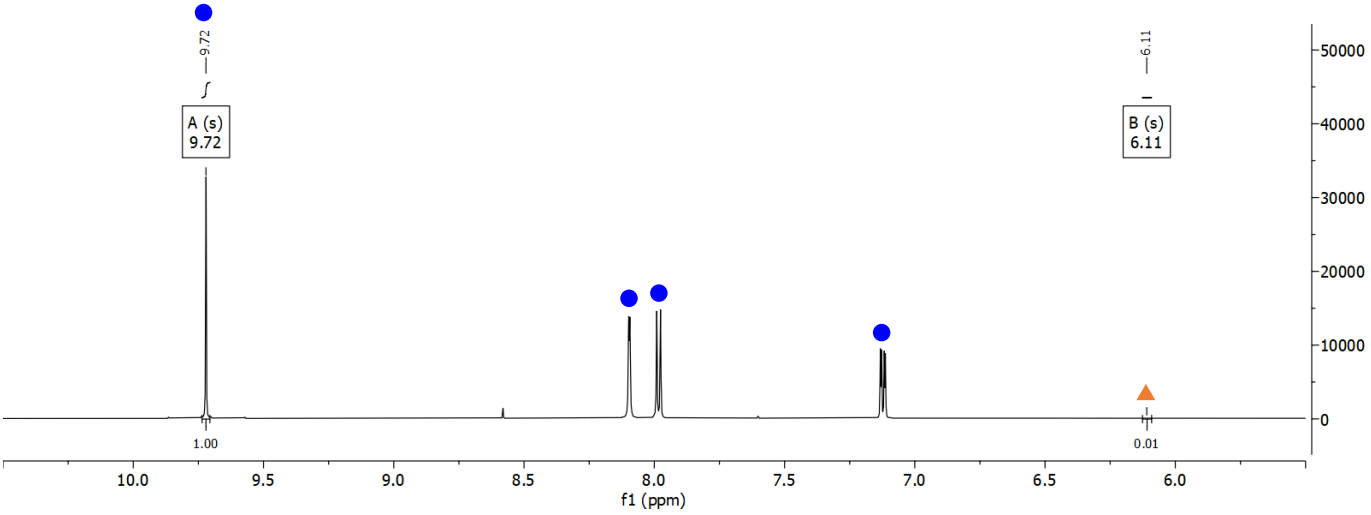


PCA 4.4

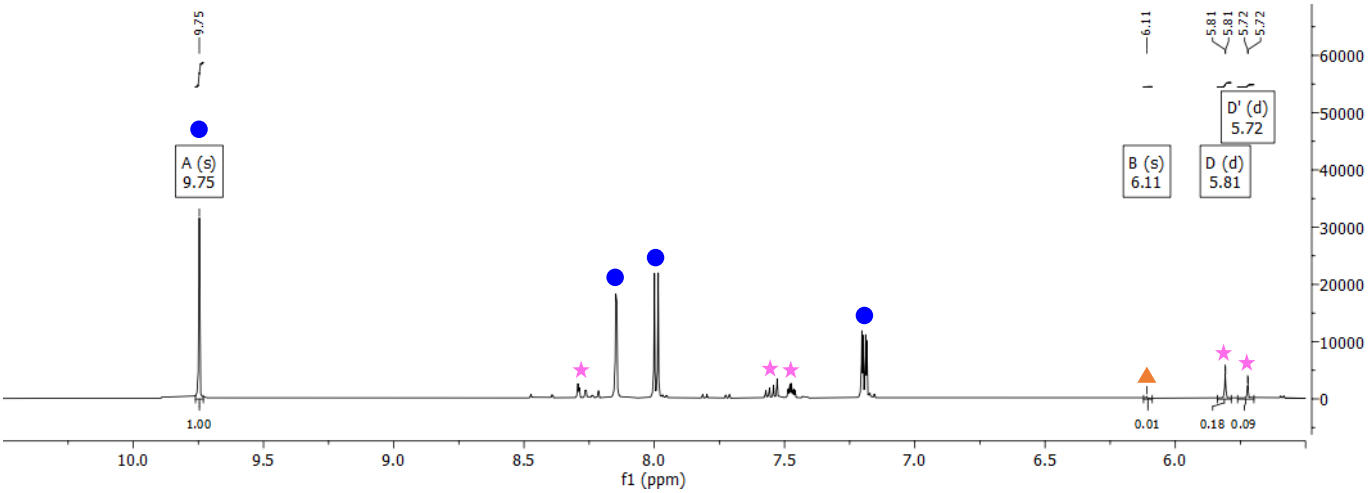
a)



b) (t = 1 h)

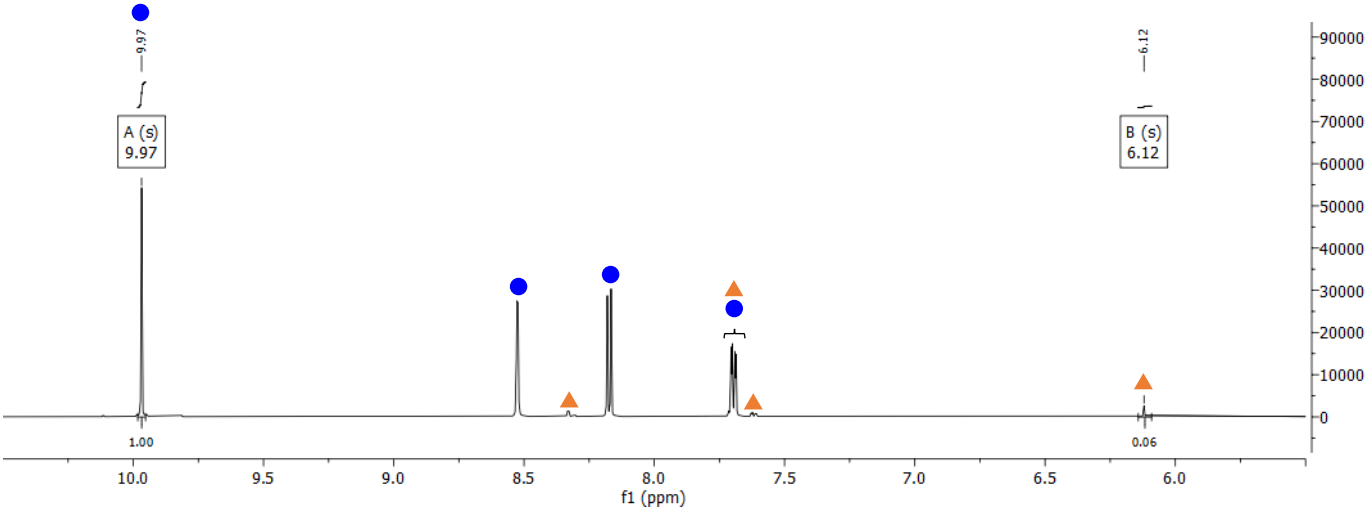


c) (t = 17 h)

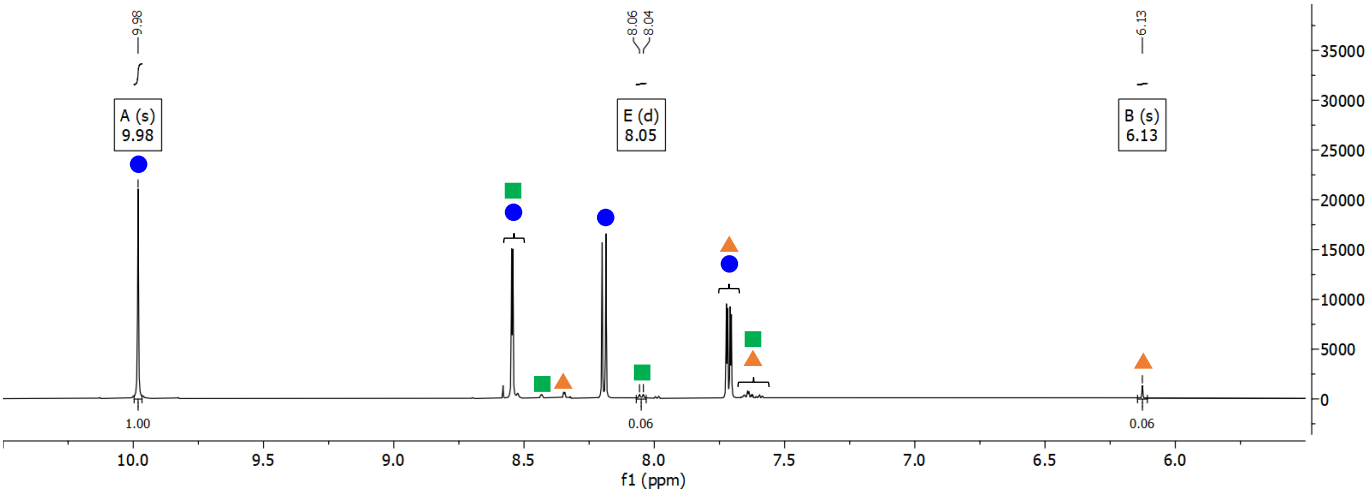


PCA 4.5

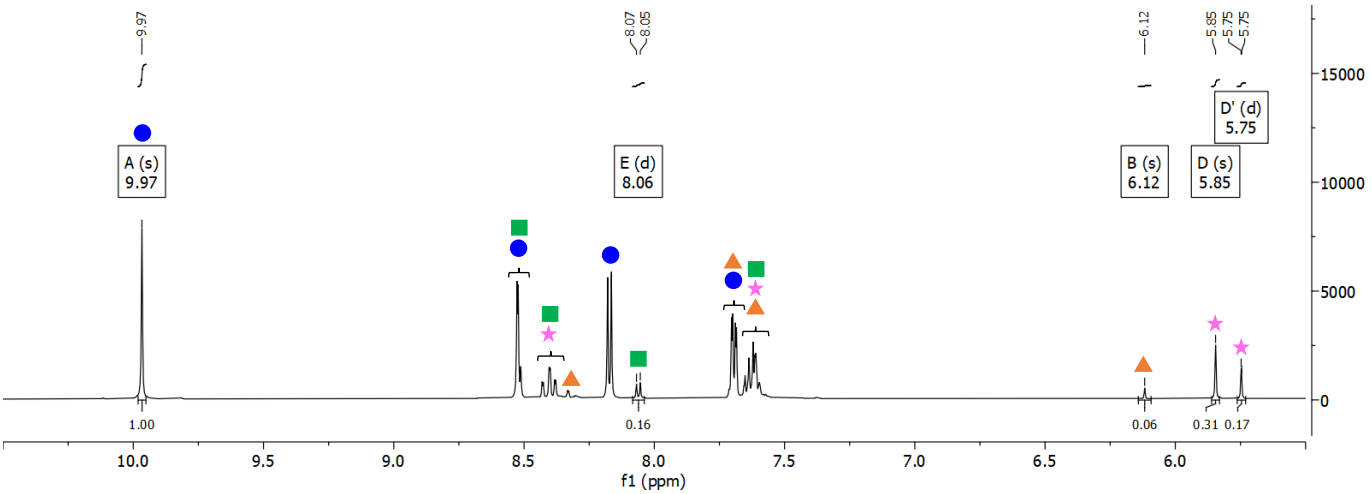
a)



b) (t = 1 h)

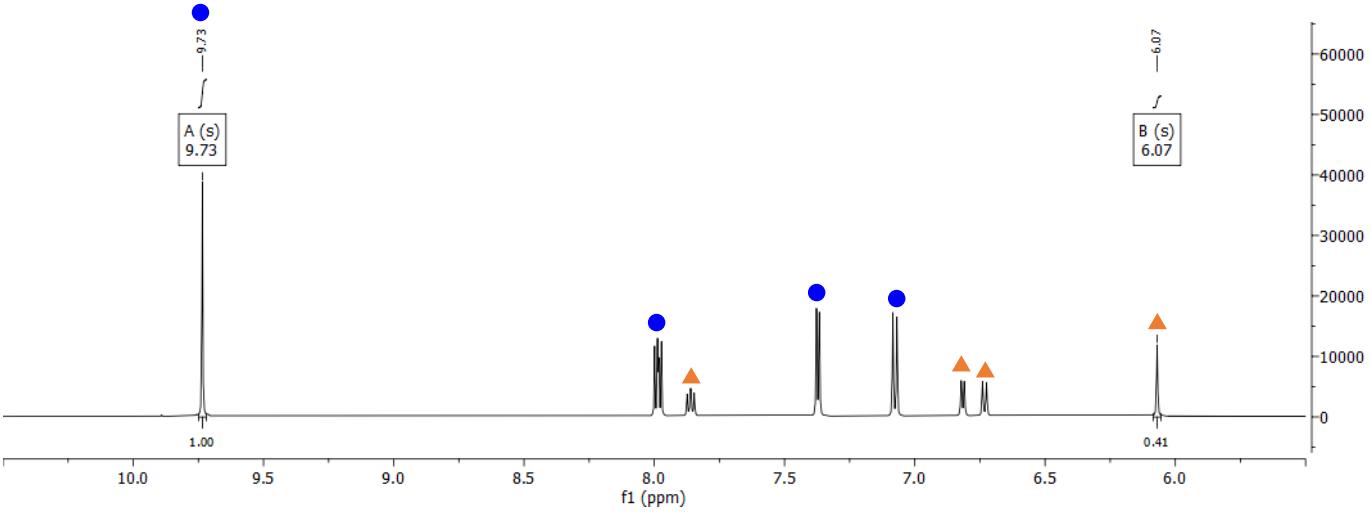


c) (t = 17 h)

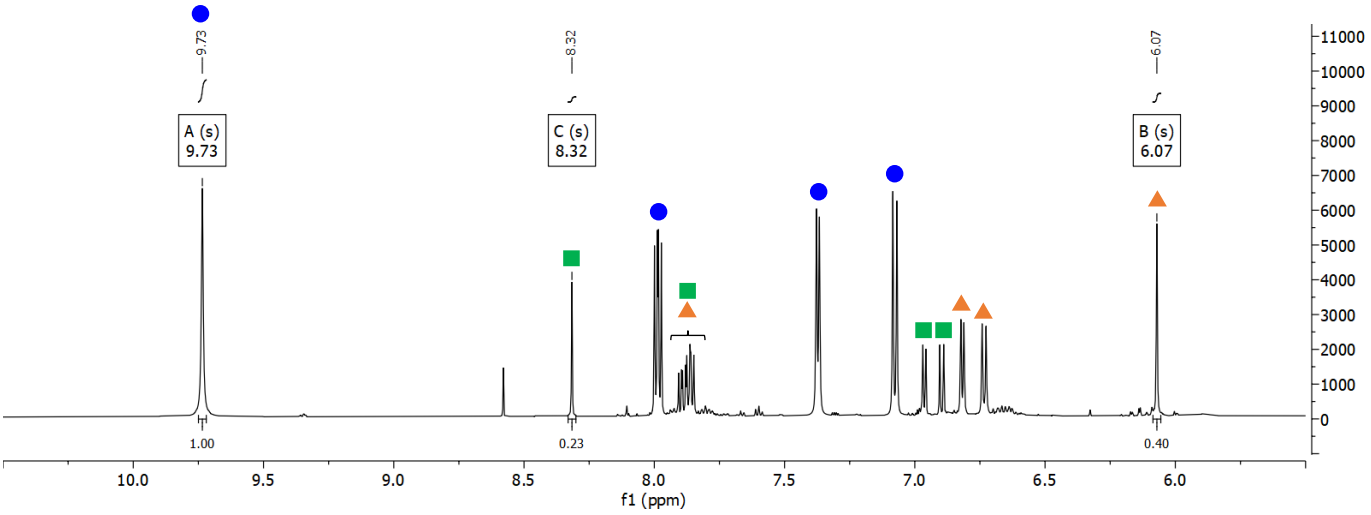


PCA 4.6

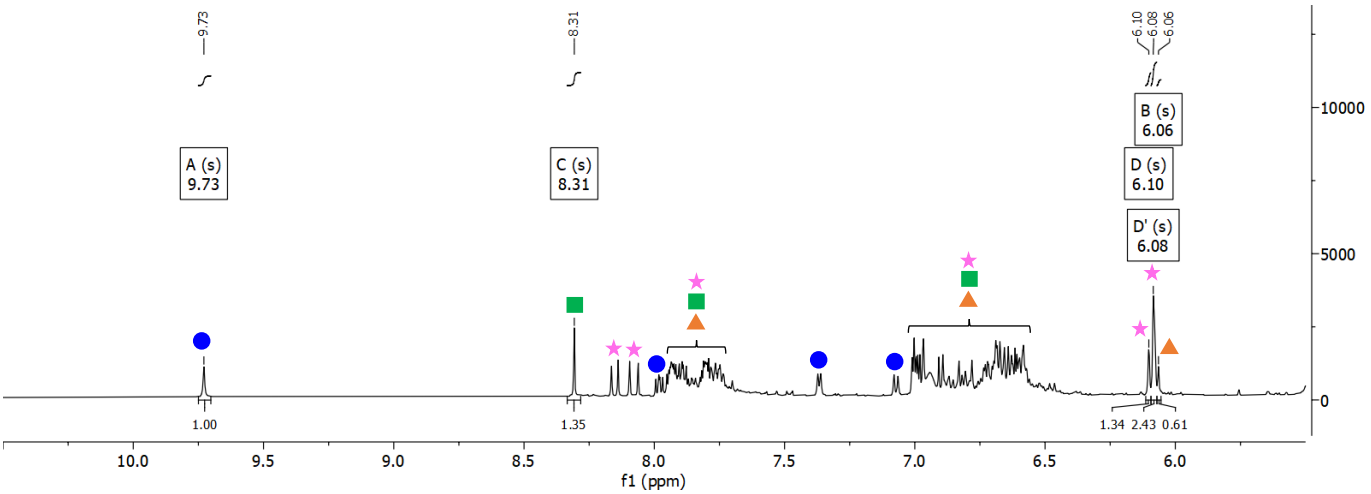
a)



b) (t = 1 h)

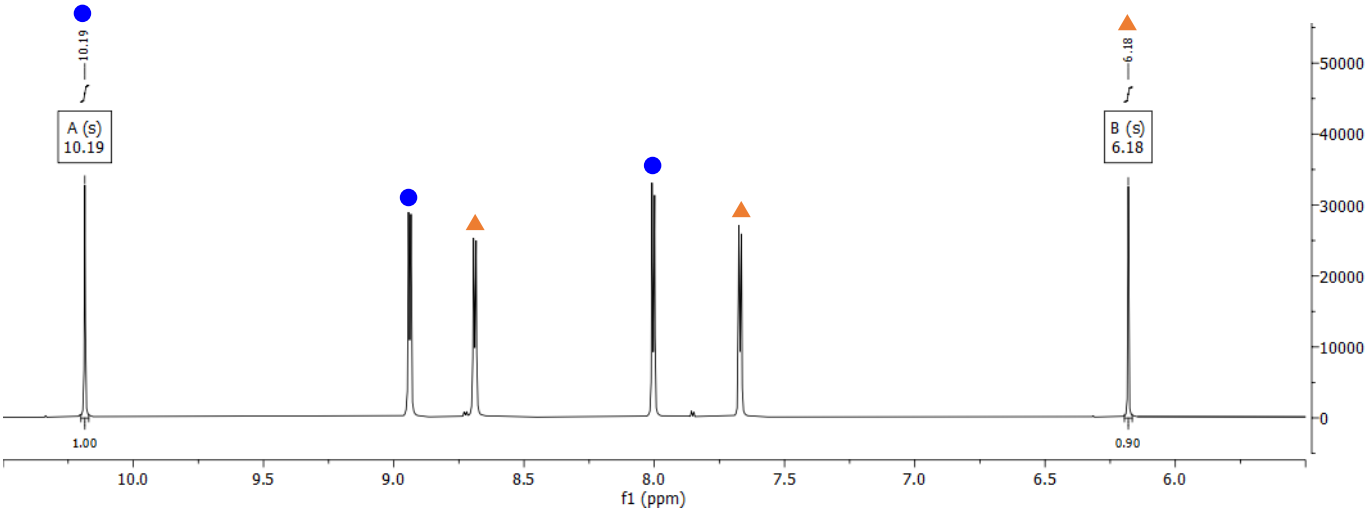


c) (t = 17 h)

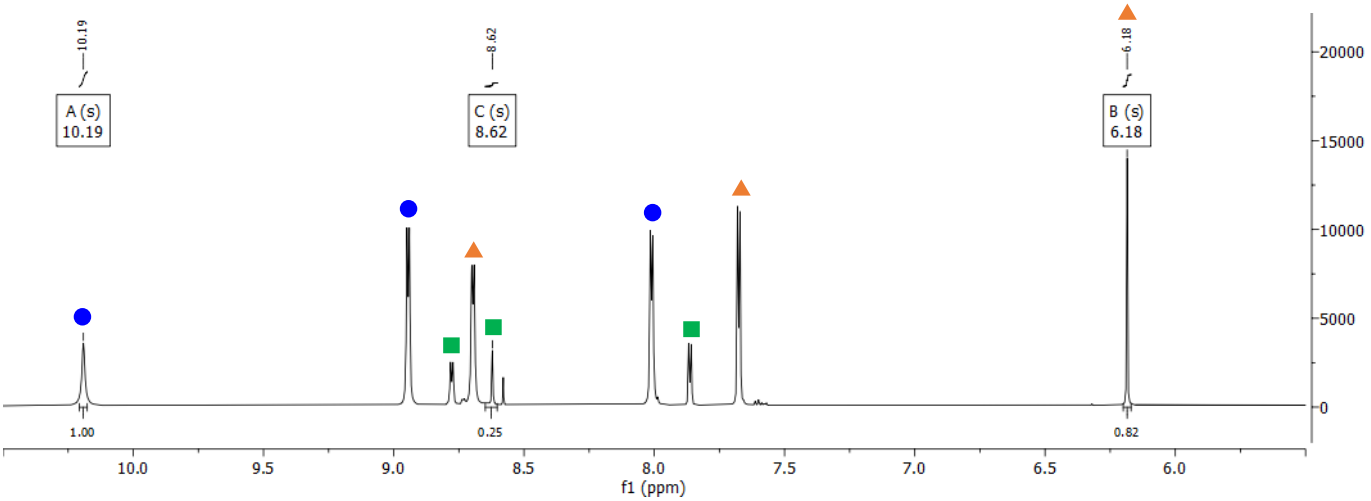


PCA 4.7

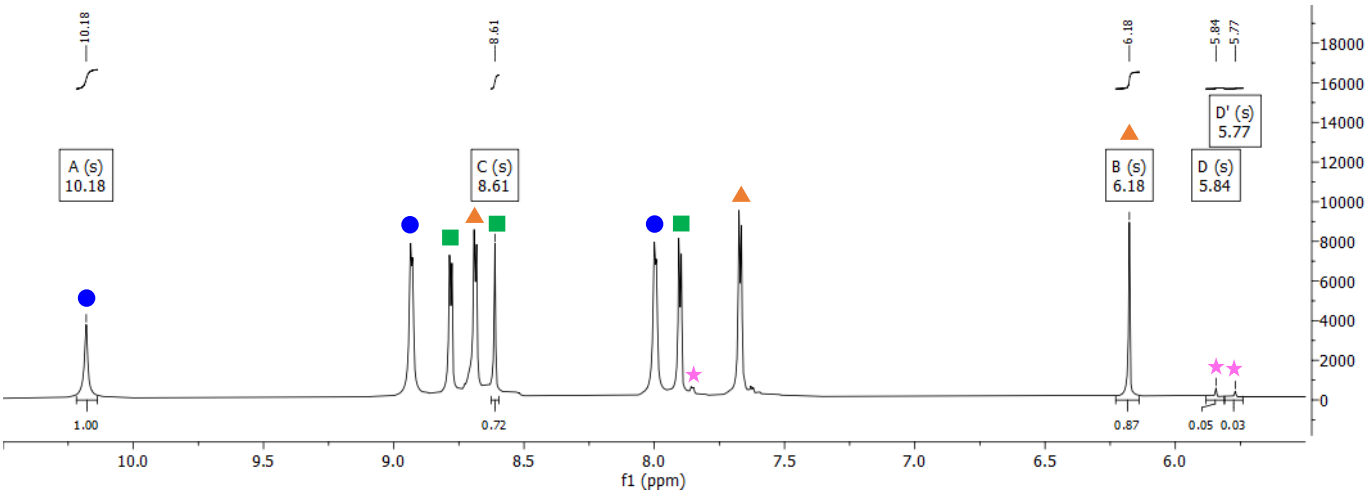
a)



b) (t = 1 h)

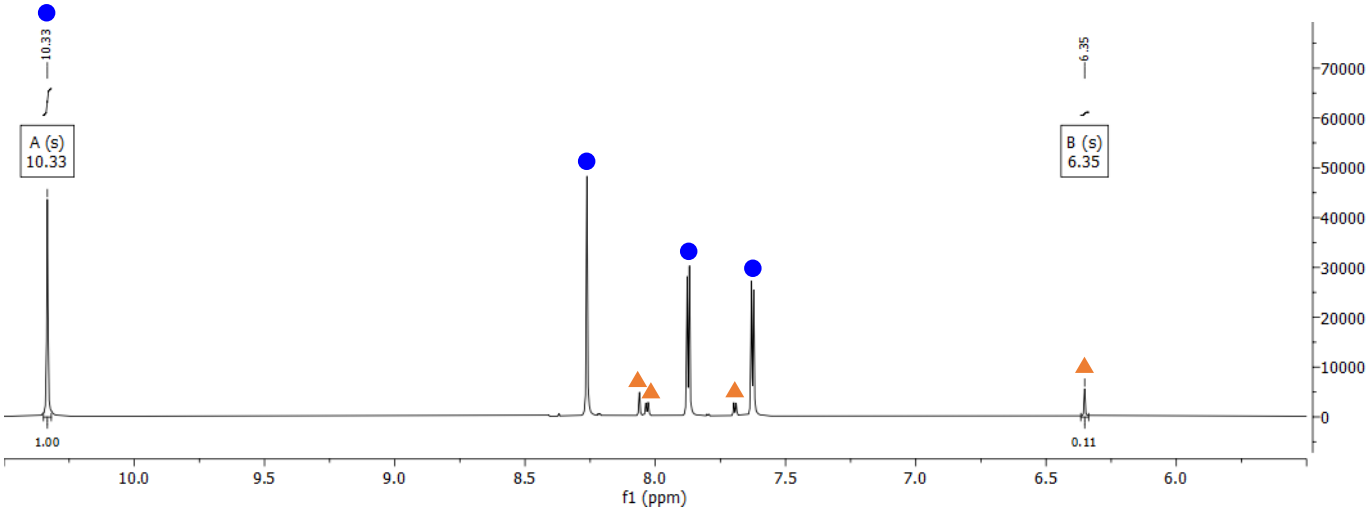


c) (t = 17 h)

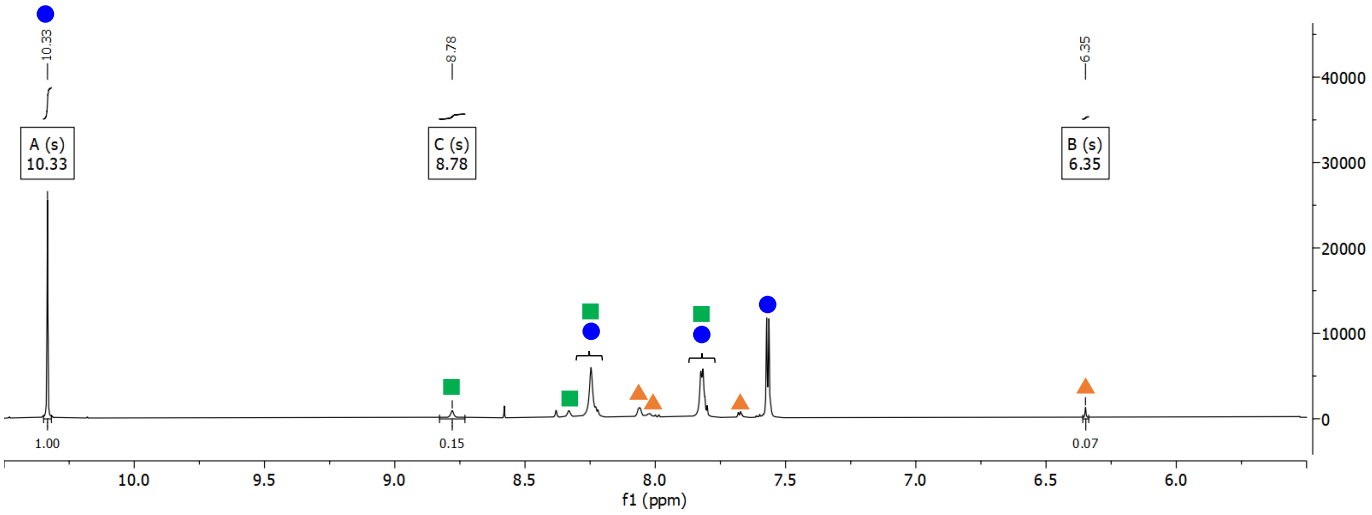


PCA 4.8

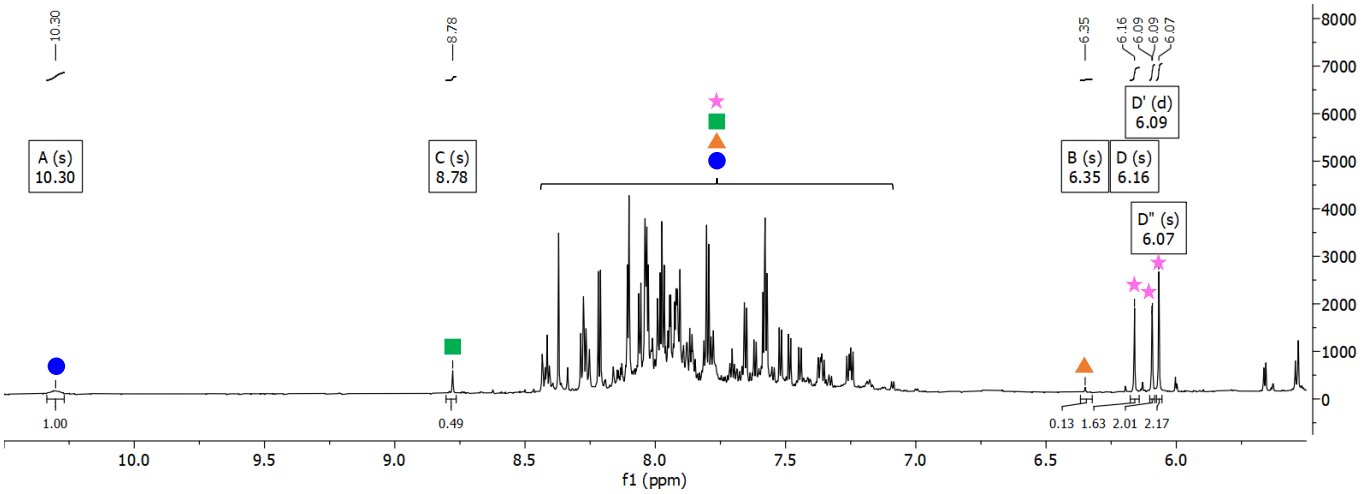
a)



b) (t = 17 h)

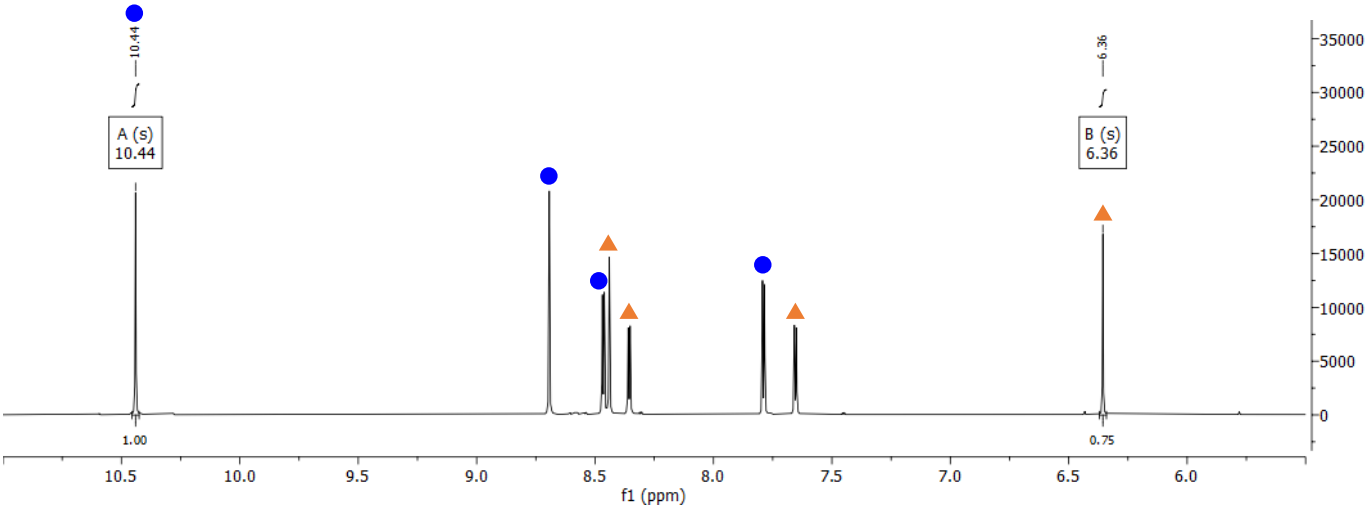


c) (t = 17 h)

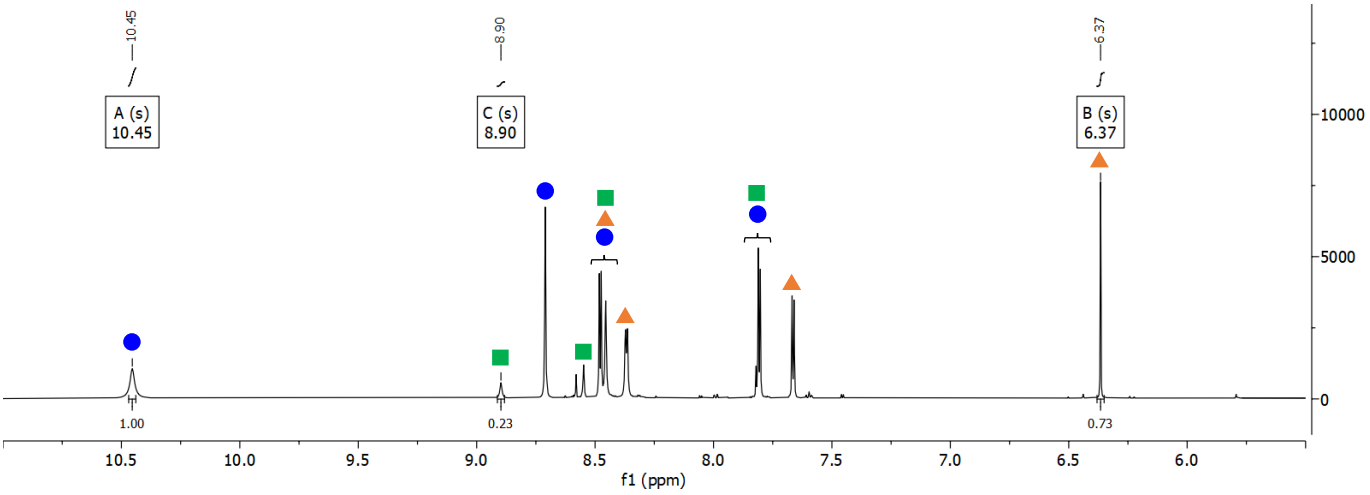


PCA 4.9

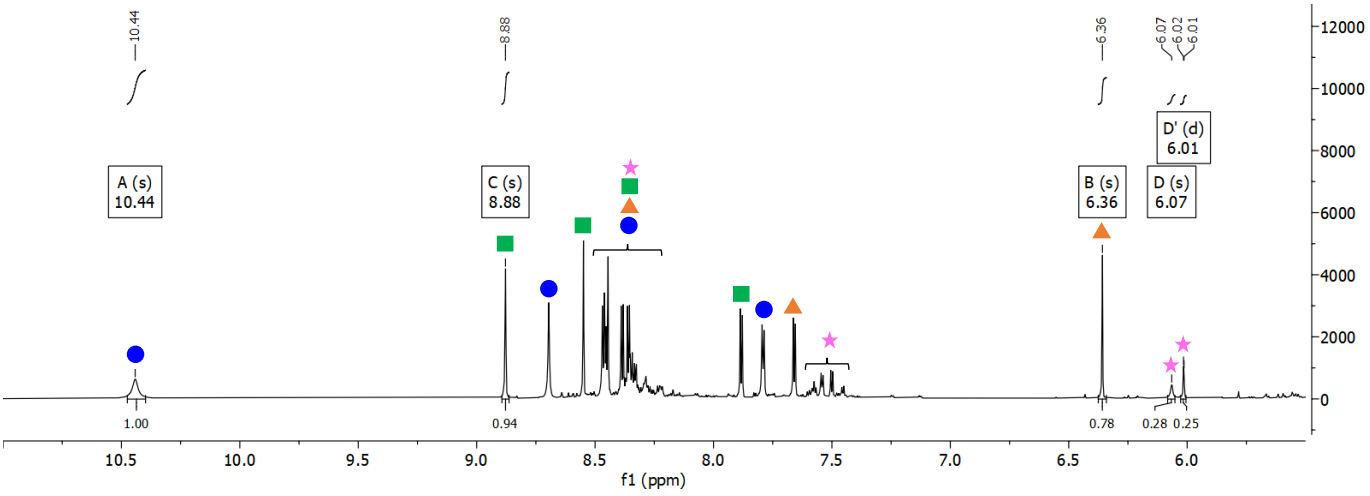
a)



b) (t = 1 h)

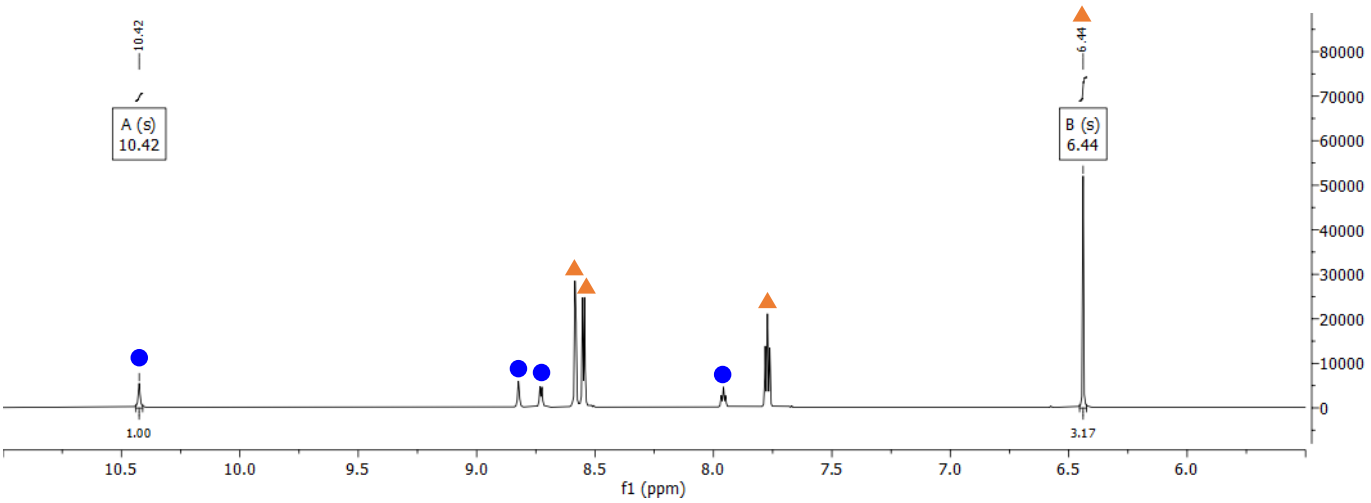


c) (t = 17 h)

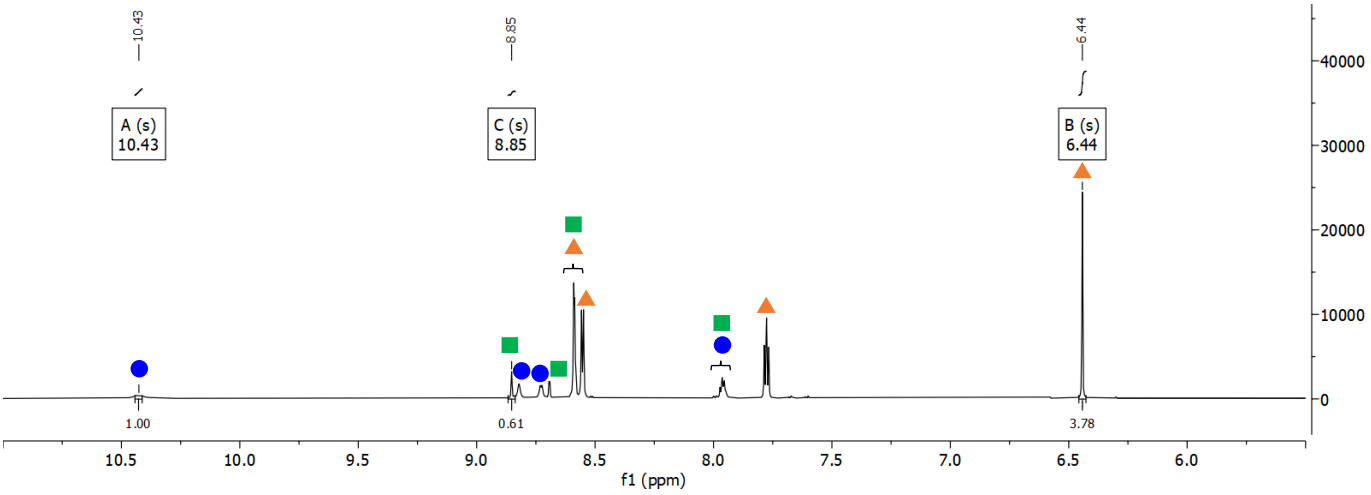


PCA 4.10

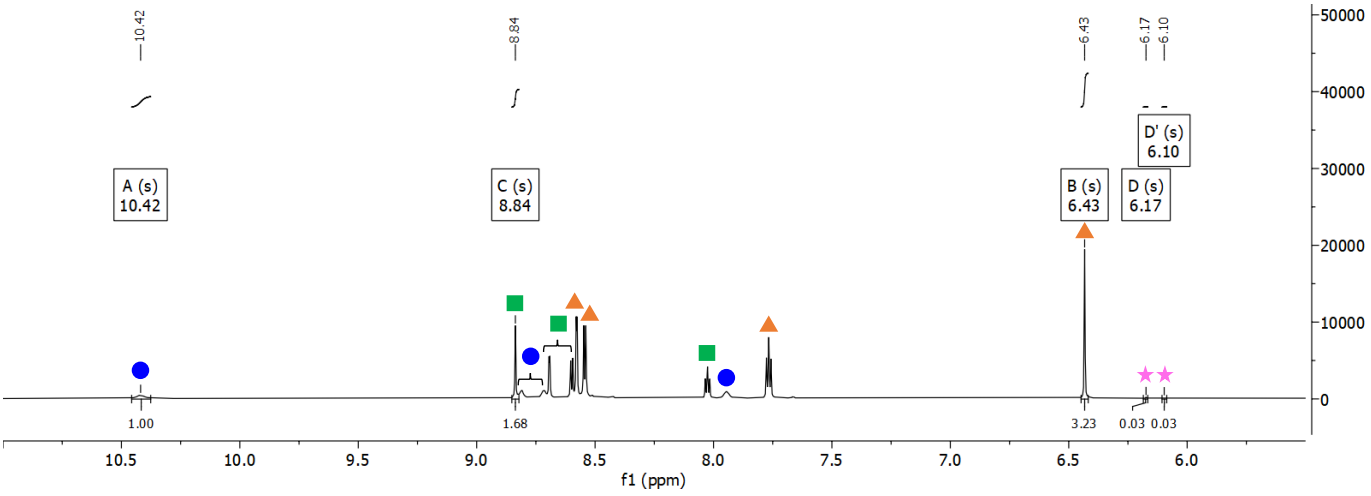
a)



b) (t = 1 h)

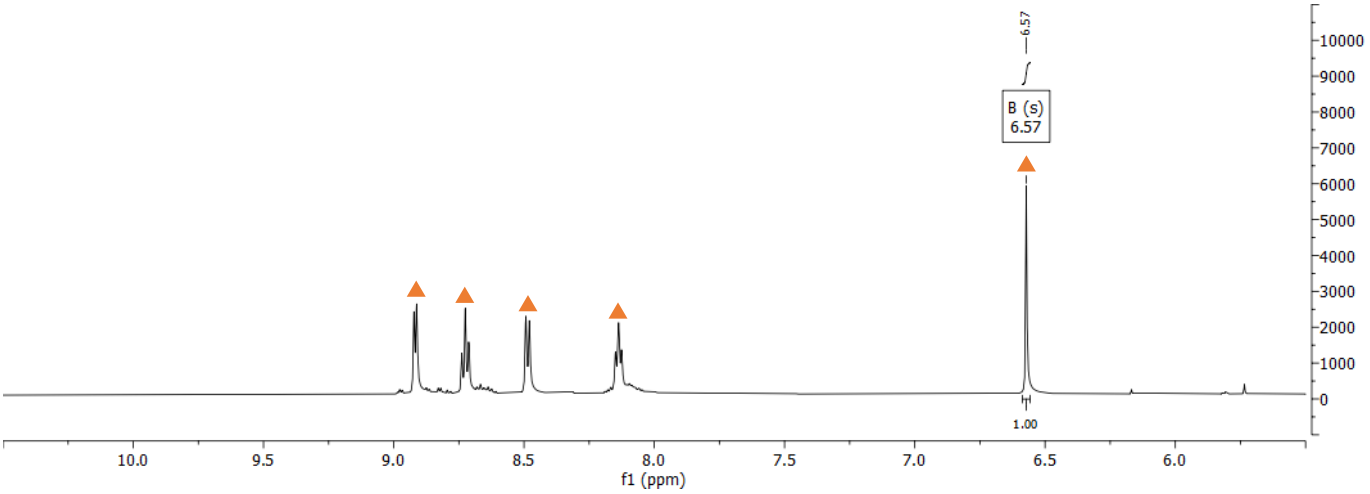


c) (t = 17 h)

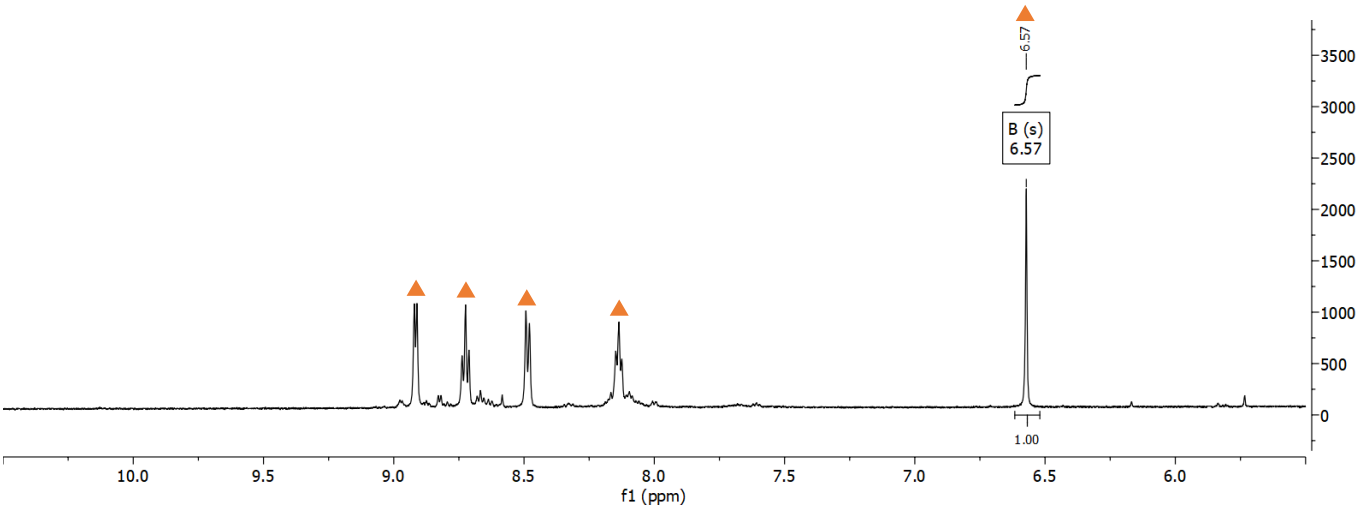


PCA 4.11

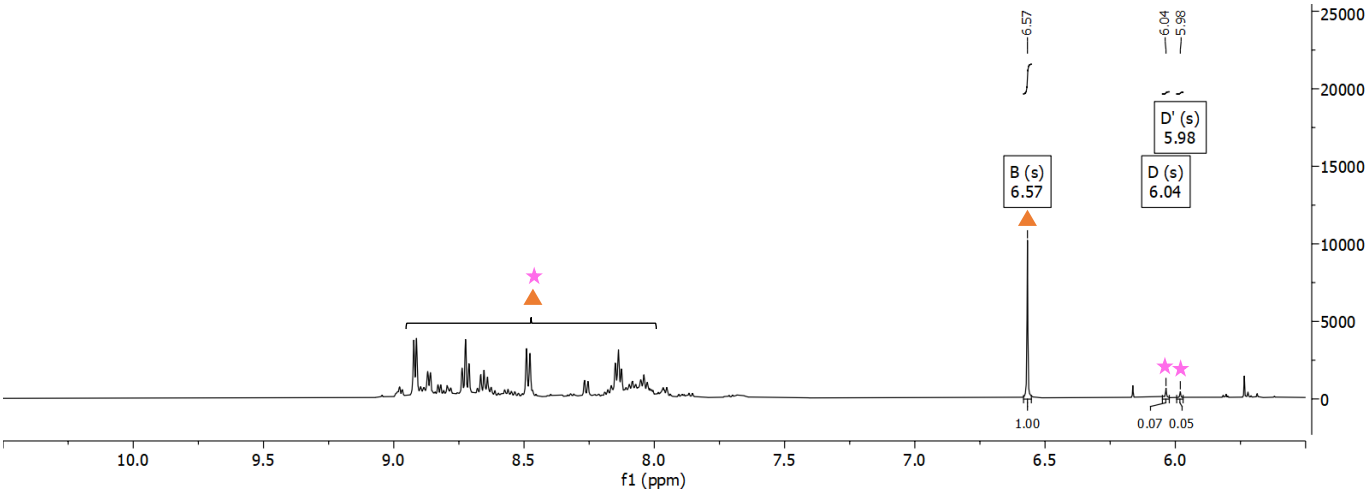
a)



b) (t = 18 h)

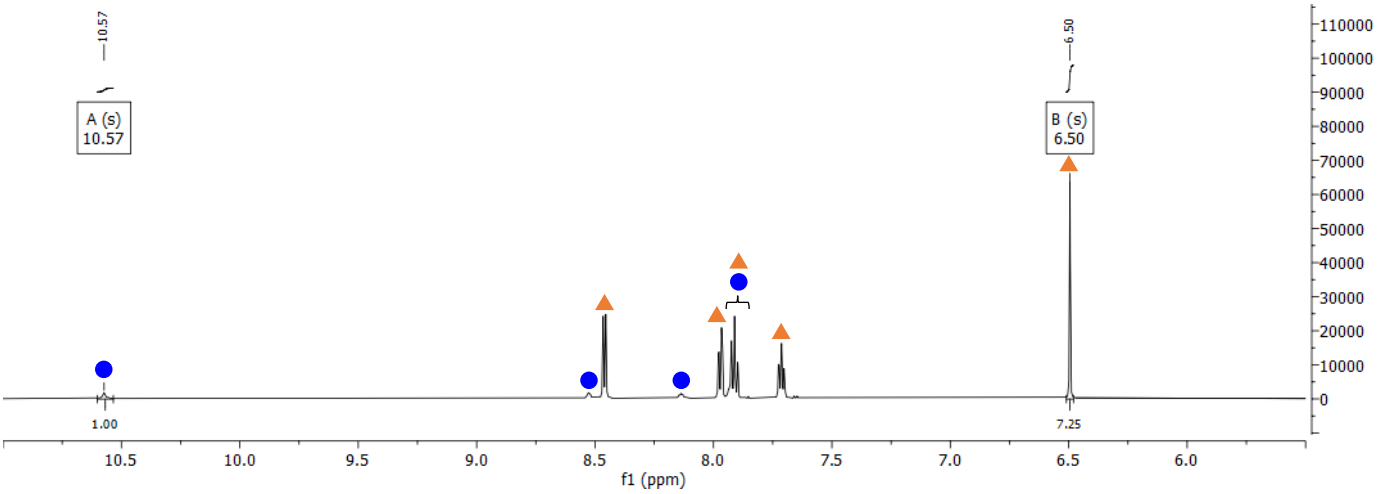


c) (t = 17 h)

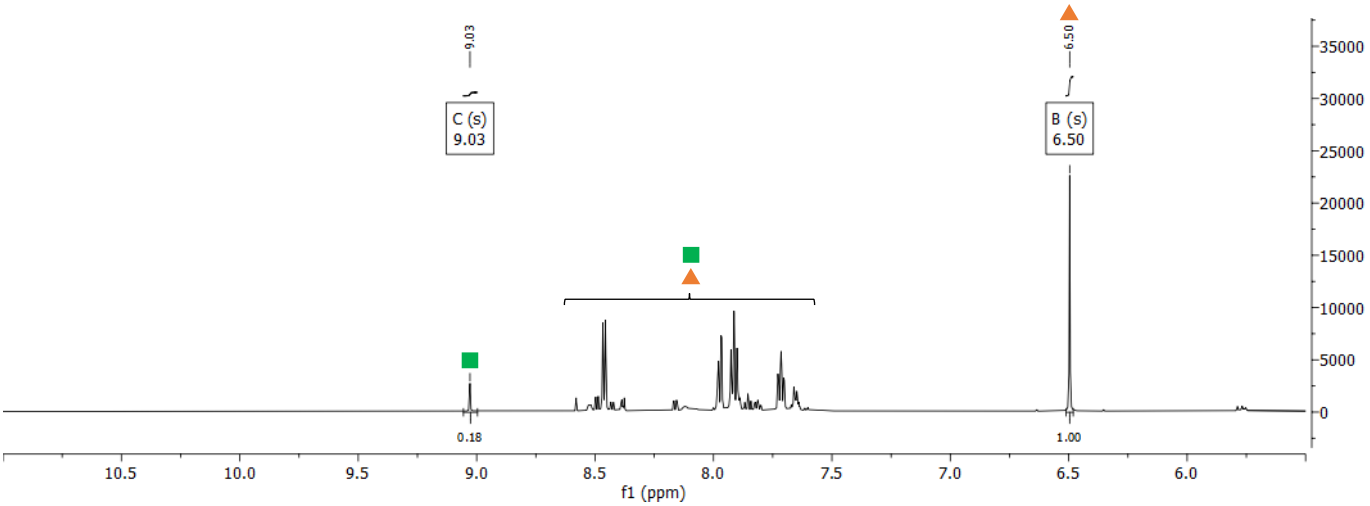


PCA 4.12

a)



b) (t = 17 h)



c) (t = 17 h)

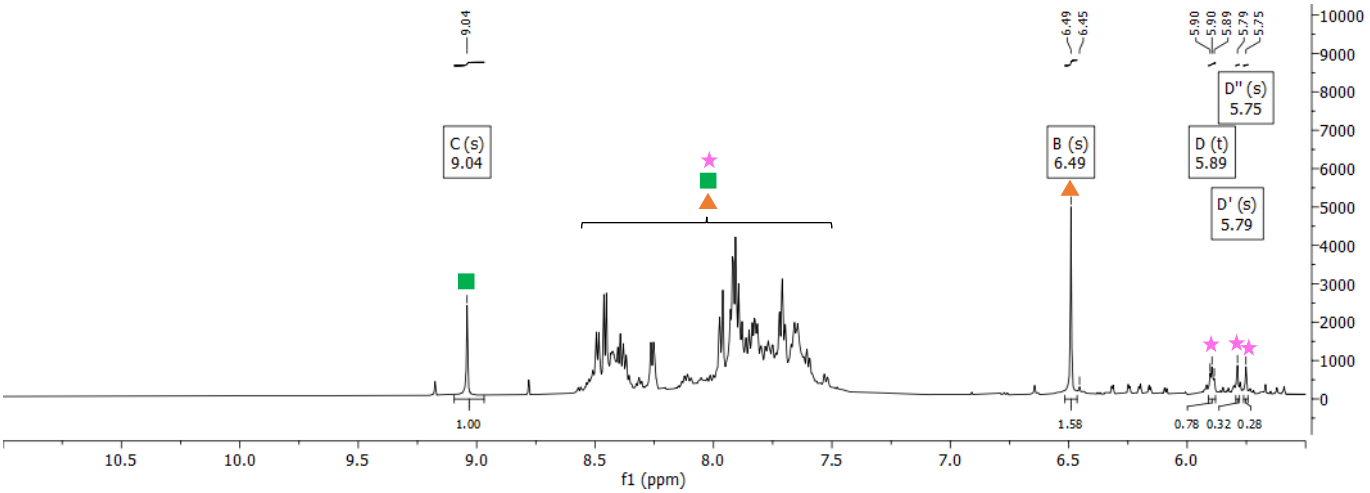
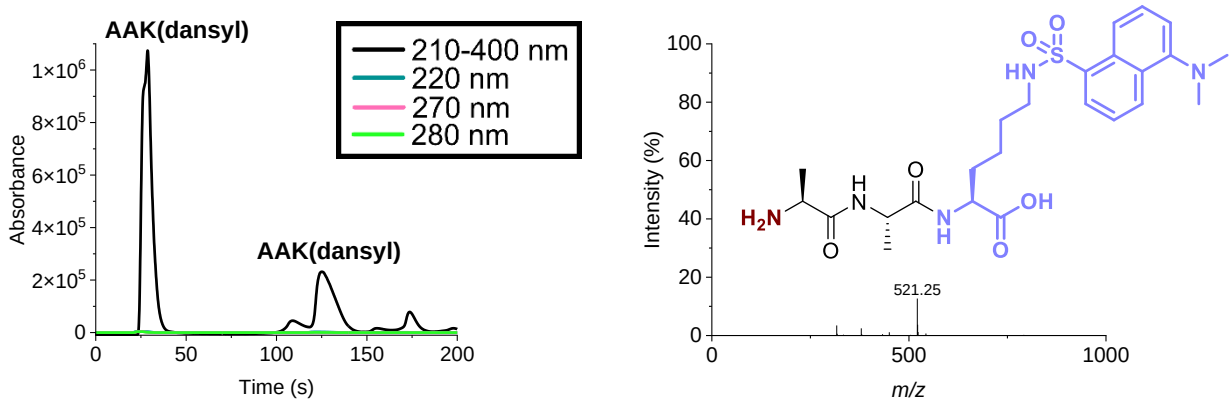


Figure S4.3. Representative ^{1}H NMR spectra of hydrate equilibria experiments showing the diagnostic signals used to calculate: ***a)*** the equilibrium constant for hydrate formation; ***b)*** the equilibrium constant for imine formation; ***c)*** rate constants for imidazolidinone formation (**triangle** = hydrate species, **circle** = aldehyde species, **square** = imine species, **star** = imidazolidinone species).

4.4.2.5 LC-MS Kinetics for PCA analogues with complex spectra

AAK(dansyl) was prepared by Julia Capar *via* solid-phase peptide synthesis (SPPS) and purified by reverse phase chromatography.

AAK(dansyl)



MS (ESI^{+}) $[AAK(dansyl)+H]^{+}$ found 521.25, calculated 522.64.

Figure S4.4. UV-vis 210-400, 220, 270 and 280 nm traces, and mass spectra of AAK(dansyl).

A stock solution of PCA **4.1** or **4.12** (50 µL, 100 mM, 5 µmol, 1 equiv., in 100 mM deuterated pH 7.5 sodium phosphate buffer) was added to a solution of AAK(dansyl) (50 µL, 100 mM, 5 µmol, 1 equiv., in 100 mM deuterated pH 7.5 sodium phosphate buffer), and the mixture incubated at 37 °C for 8 h with agitation (1000 rpm). Aliquots (2 µL) of the reaction mixture were diluted with water (198 µL) and analysed by LC-MS at 1 h (PCA **4.1**) or 1.5 h (PCA **4.12**) time-points. Conversion was determined by the relative integrals of the unmodified/modified peptide in the 325 nm UV trace from LC-MS analysis.

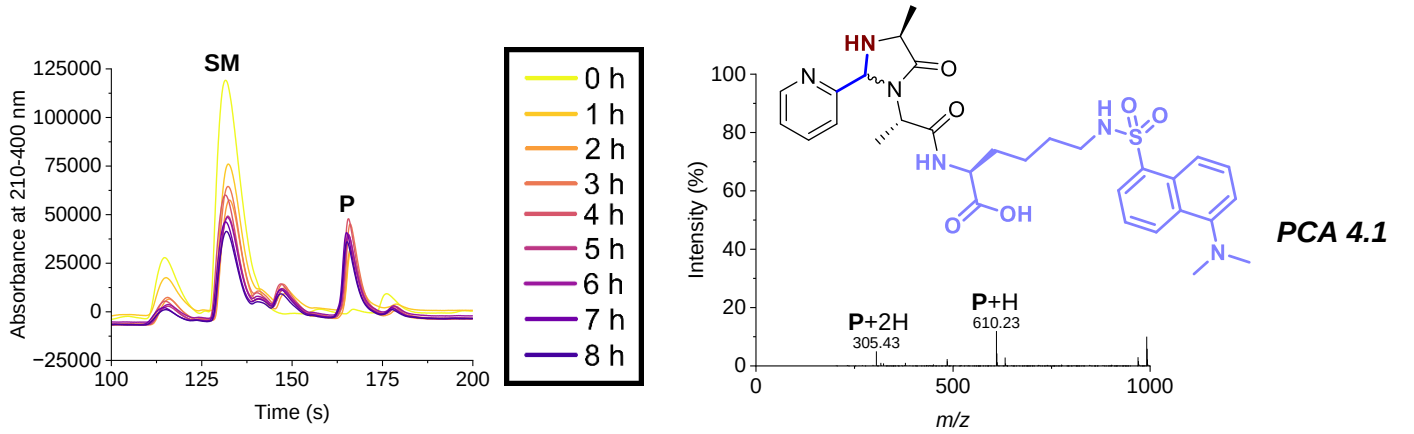
Conversion values were normalised using a correction factor (**CF**) in *Equation 4.5* to remove background noise from the conjugate signals, based on the assumption that the initial rate was linear over the first 3 time-points collected, and that $[\mathbf{Pr}]_{t=0\,h} = 0$. Correction factors were calculated as the y-intercept of the linear regression line of the plot of the integral of conjugates over time (for the first 3 time-points collected).

$$\mathbf{Conversion} = \mathbf{100} \times \frac{\int \mathbf{Pr} - \mathbf{CF}}{\int \mathbf{SM} + \int \mathbf{Pr} - \mathbf{CF}} \qquad \textit{(Equation 4.5)}$$

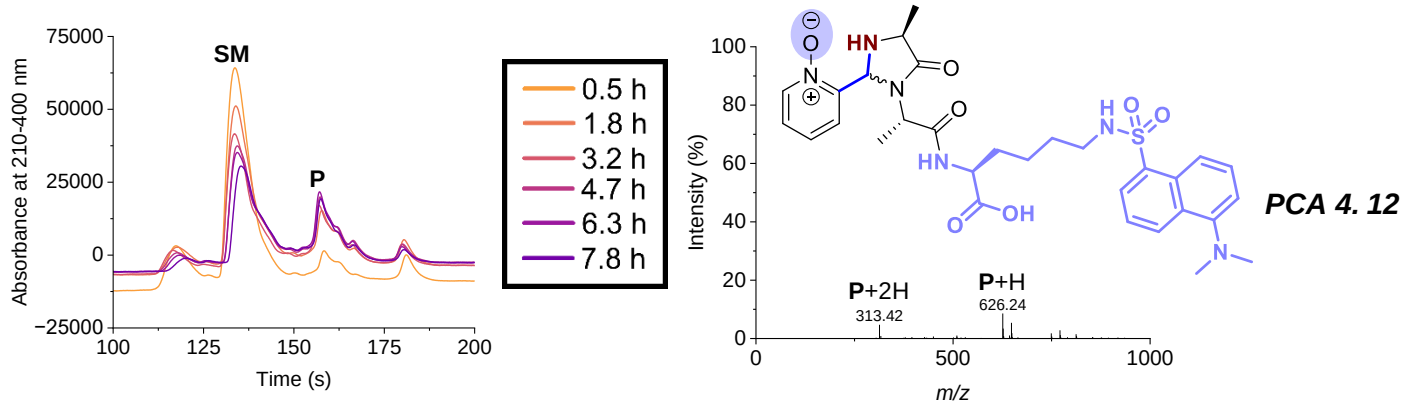
Data were fit to a second order reversible kinetic model in Copasi 4.34.251. A simplified model was used;

A + B = C;

in which **A** represented the concentration of AAK(dansyl) [**SM**], **B** represented the concentration of aldehyde [**PCA**], and **C** represented the concentration of imidazolidinone [**Pr**] conjugate. The model did not integrate hydrate and imine formation, neither of which were observed by LC-MS, and relied on the assumption that imidazolidinone formation was the rate-limiting step of the reaction. k_6 and k_{-6} 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_6 10^{-6}-10^{6} M^{-1} h^{-1}; k_{-6} 10^{-6}-10^{6} h^{-1}.



MS (ESI$^+$) [**P**+2H]$^+$ found 305.43, calculated 306.14; [**P**+H]$^+$ found 610.23, calculated 611.27.



MS (ESI$^+$) [**P**+2H]$^+$ found 313.42, calculated 314.13; [**P**+H]$^+$ found 626.24, calculated 627.26.

Figure S4.5. UV-vis 210-400 nm traces over time, and example conjugate mass spectra at the final time-point for the modification of AAK(dansyl) with reagents **4.1** and **4.12** (**SM** = unmodified AAK(dansyl), **P** = modified AAK(dansyl)). Note that absorbance traces at 210-400 nm are shown for illustrative purposes only: conversions were calculated from the extracted 325 nm UV chromatogram, which could not be exported from DataAnalysis but had a very similar profile to the 210-400 nm absorbance trace.

4.4.3 pK_a measurement (carried out by Dr. Chris Spicer)

Hydroxy-PCA analogues **4.2**, **4.4**, or **4.8** were dissolved in water (5 mL) at a concentration of 50 mM. The solution was then acidified by the addition of concentrated hydrochloric acid (5 µL). Aliquots of sodium hydroxide solution (0.1 M, 100 µL) were added and the pH of the solution measured after each addition. The pK_a of the hydroxyl group was calculated by determining the half equivalence point.

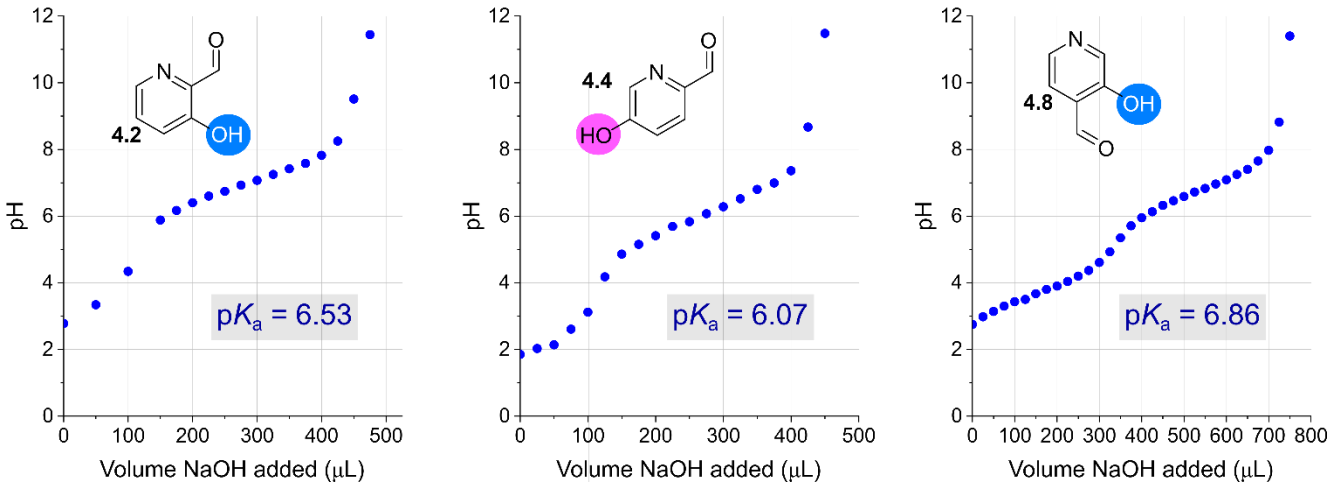
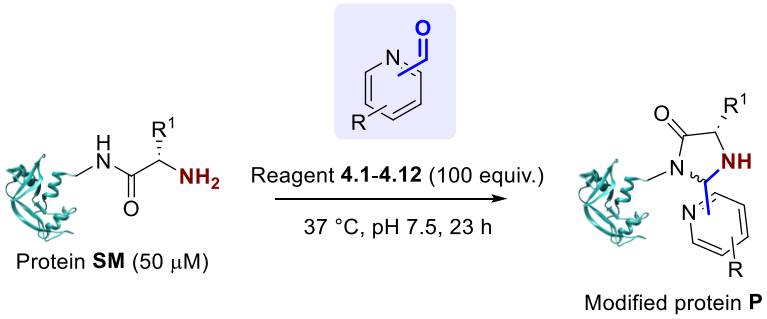


Figure S4.6. pK_a calculation, experimental work and figure preparation by Dr. Chris Spicer.

4.4.4 Protein modification

4.4.4.1 Screening reagents against a panel of model proteins



Procedure 4A: *Modification with reagents under conditions adapted from MacDonald et al.*[6] A stock solution of the modification reagent (85 µL, 10 mM, 850 nmol, 100 equiv., in 50 mM pH 7.5 sodium phosphate buffer) was added to a solution of protein (85 µL, 100 µM, 8.5 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer), and the mixture incubated at 37 °C for 23 h with agitation (1000 rpm).

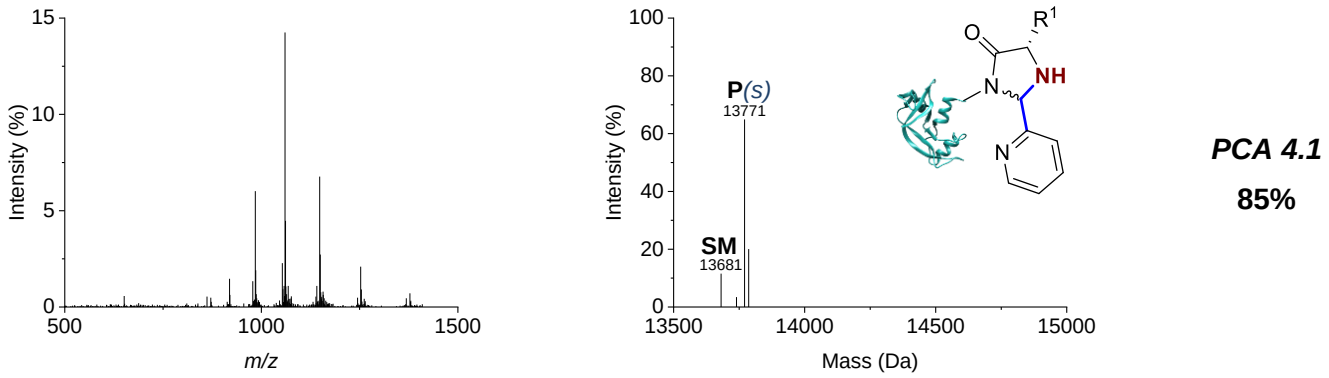
Reagents **4.1**-**4.12** were used to modify RNase A, myoglobin and HASPA G2S under the conditions outlined in **General Procedure 4A**. Conversion was determined by LC-MS analysis without purification (Crude). Protein conjugates **P** were then purified by dialysis to remove excess reagent (4 °C, 3.5 kDa MWCO; 1 × 50

mM pH 7 sodium phosphate buffer, 2 h; 1 × water, 3 h; 1 × water, 16 h; 1 × water, 4 h), and conversion was again determined by LC-MS analysis (Purified).

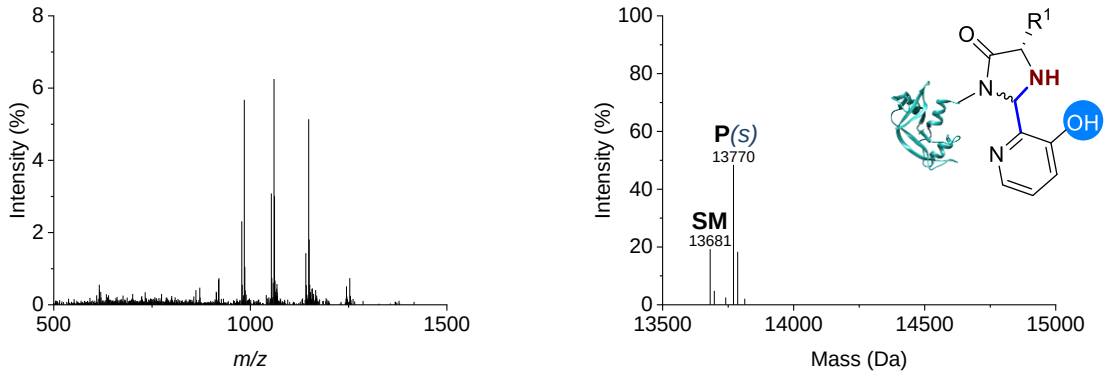
PCA	Conversion (%)					
	RNase A		Myoglobin		HASPA G2S	
	Crude	*Purified*	*Crude*	*Purified*	*Crude*	*Purified*
4.1	**85** *(s)*	**80** *76(s):3(d)*	**5** *(s)*	**9** *(s)*	**88**	**89**
4.2	**72** *(s)*	**36** *33(s):3(d)*	**11** *(s)*	**8** *(s)*	N/A *(missing data)*	
4.3	**100** *95(s):5(d)*	**100** *(s)*	**30** *(s)*	**30** *(s)*		
4.4	**14** *(s)*	**0**	**8** *(s)*	**10** *(s)*		
4.5	**77** *(s)*	**67** *(s)*	**13** *(s)*	**12** *(s)*		
4.6	**11** *(s)*	**10** *(s)*	**24** *(s)*	**27** *(s)*	**51** *42(s):9(d)*	**20**
4.7	**48** *44(s):4(d)*	**43** *41(s):2(d)*	**26** *(s)*	**21** *(s)*	**76** *57(s):18(d)*	**69** *54(s):15(d)*
4.8	**63** *51(s):12(d)*	**24** *14(s):10(d)*	**31** *(s)*	**30** *(s)*	**84** *67(s):18(d)*	**10** *0(s):10(d)*
4.9	**90** *84(s):6(d)*	**89** *84(s):5(d)*	**19** *(s)*	**23** *(s)*	N/A *(missing data)*	
4.10	**29** *(s)*	**21** *(s)*	**39%** *33(s):6(d)*	**37** *(s)*	**42** *26(s):15(d)*	**54** *30(s):23(d)*
4.11	**8** *(s)*	**0**	**16** *(s)*	**19** *(s)*	**0**	**0**
4.12	**51** *42(s):6(d):3(t)*	**49** *(s)*	**75** *65(s):10(d)*	**66** *59(s):7(d)*	**68** *30(s):38(d)*	**65** *26(s):39(d)*

Table S4.4. Conversions for the modification of RNase A, myoglobin, and HASPA G2S before (crude) and after (purified) dialysis at 4 °C. s = single, d = double, t = triple modification.

RNase A *(crude)*



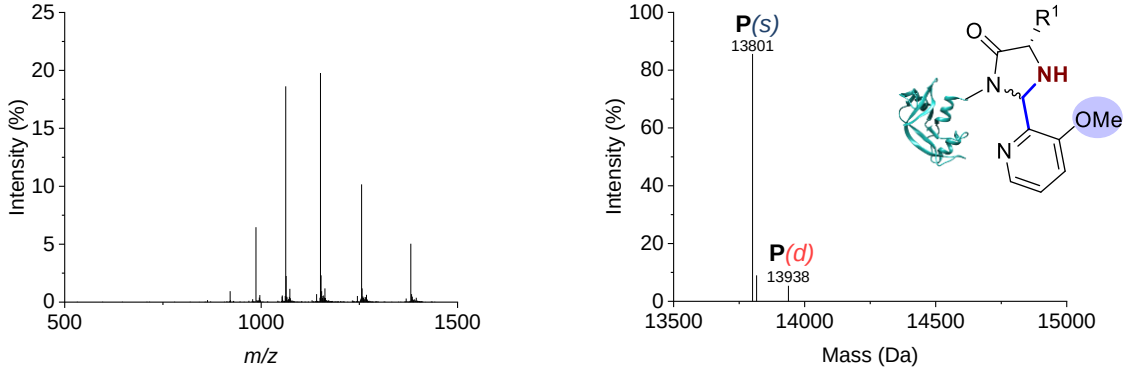
MS (ESI$^+$) [**SM**+H]$^+$ found 13681, calculated 13681; [**SM**+MeCN+NH_4]$^+$ found 13740, calculated 13739; [**P**(s)+H]$^+$ found 13771, calculated 13770; [**P**(s)+H_2O+H]$^+$ found 13786, calculated 13788.



PCA 4.2

72%

MS (ESI$^+$) [**SM**+H]$^+$ found 13681, calculated 13681; [**SM**+H_2O+H]$^+$ found 13697, calculated 13699; [**SM**+MeCN+NH_4]$^+$ found 13741, calculated 13739; [**P**(s)-H_2O+H]$^+$ found 13770, calculated 13768; [**P**(s)+H]$^+$ found 13786, calculated 13786; [**P**(s)-H_2O+MeCN+H]$^+$ found 13814, calculated 13809.

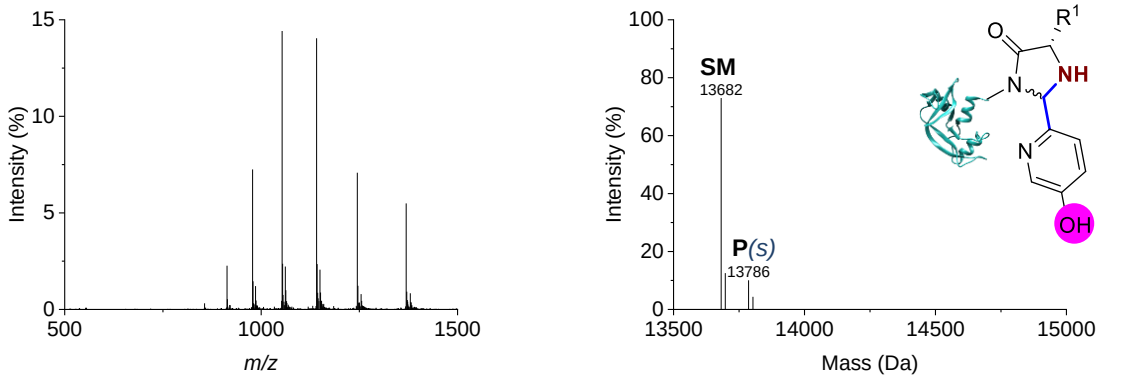


PCA 4.3

100%

95(s):5(d)

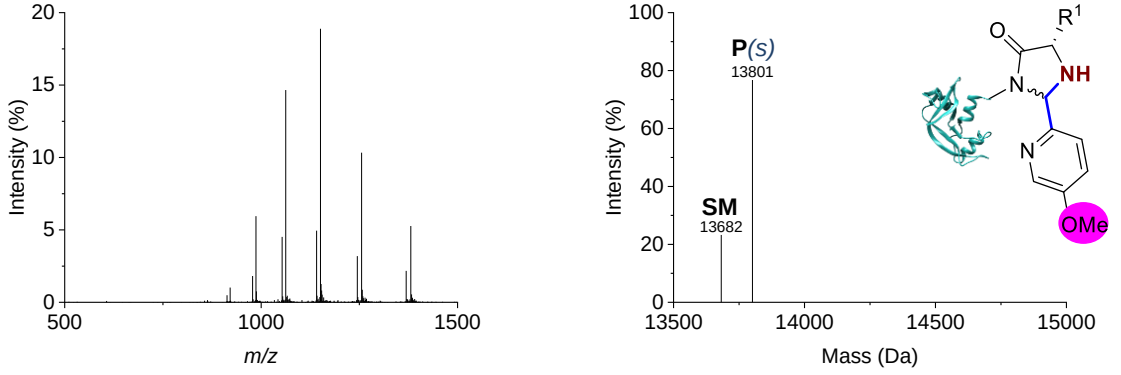
MS (ESI$^+$) [**P**(s)+H]$^+$ found 13801, calculated 13800; [**P**(s)+H_2O+H]$^+$ found 13817, calculated 13818; [**P**(d)+H_2O+H]$^+$ found 13938, calculated 13937.



PCA 4.4

14%

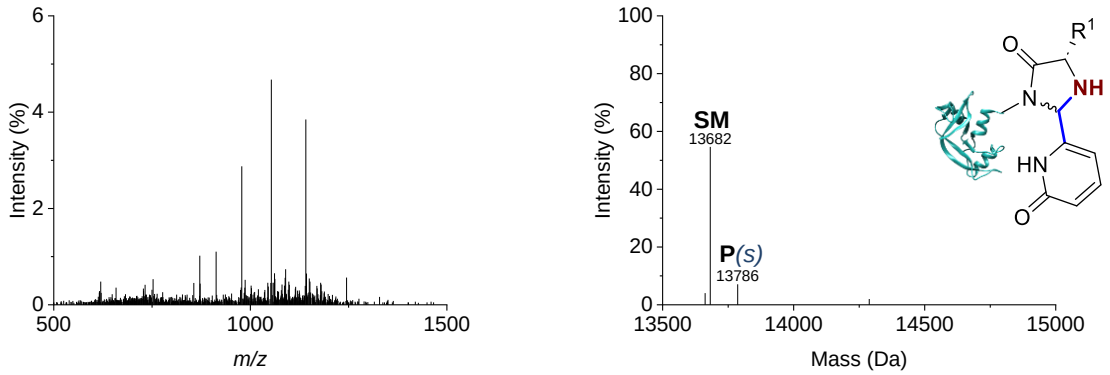
MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681; [**SM**+H_2O+H]$^+$ found 13698, calculated 13699; [**P**(s)+H]$^+$ found 13786, calculated 13786; [**P**(s)+H_2O+H]$^+$ found 13803, calculated 13804.



PCA 4.5

77%

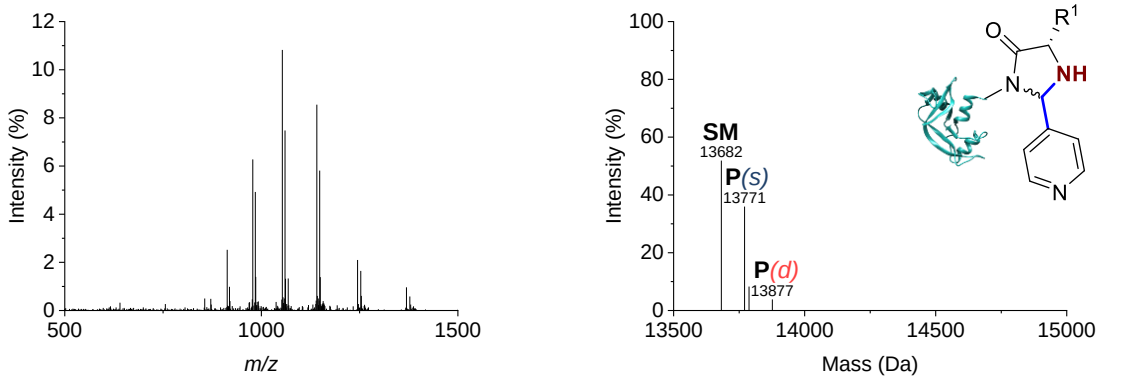
MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681; [**P**(s)+H]$^+$ found 13801, calculated 13800.



PCA 4.6

11%

MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 13662, calculated 13663; [**SM**+H]$^+$ found 13682, calculated 13681; [**P**(s)+H]$^+$ found 13786, calculated 13786.

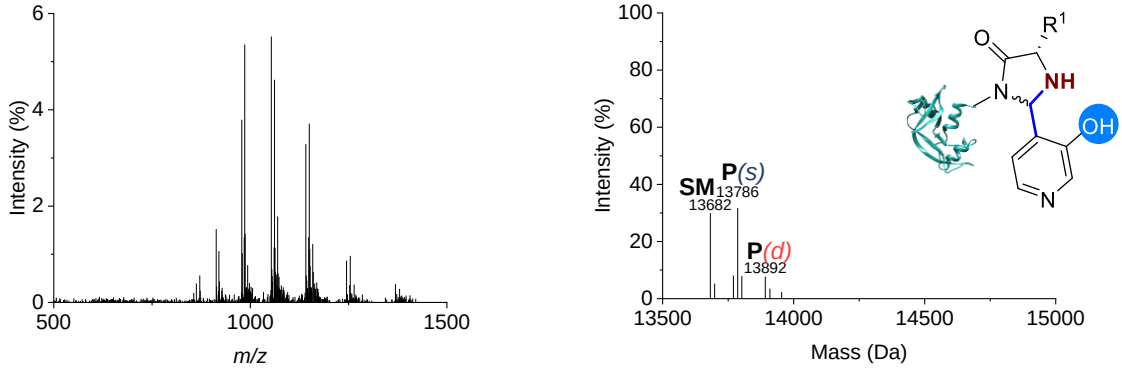


PCA 4.7

48%

44(s):4(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681; [**P**(s)+H]$^+$ found 13771, calculated 13770; [**P**(s)+H_2O+H]$^+$ found 13788, calculated 13788; [**P**(d)+H_2O+H]$^+$ found 13877, calculated 13877.

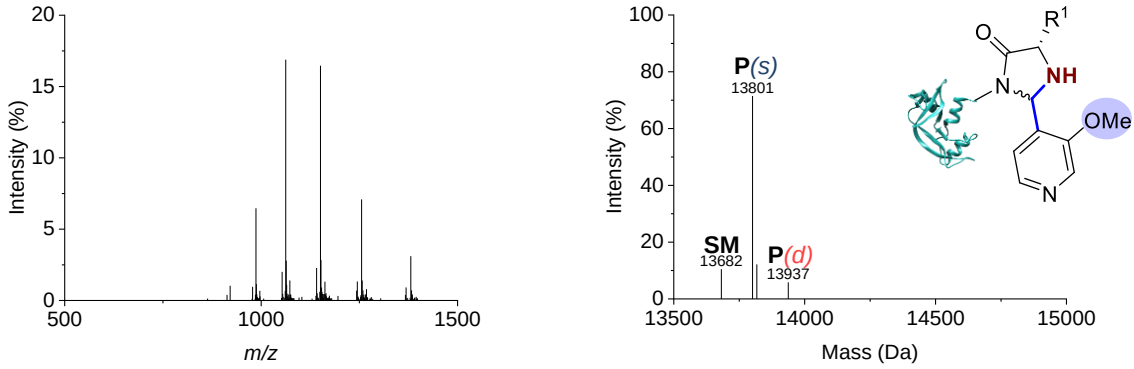


PCA 4.8

63%

51(s):12(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681; [**SM**+H_2O+H]$^+$ found 13699, calculated 13699; [**P**(s)-H_2O+H]$^+$ found 13770, calculated 13768; [**P**(s)+H]$^+$ found 13786, calculated 13786; [**P**(s)+H_2O+H]$^+$ found 13802, calculated 13804; [**P**(d)+H]$^+$ found 13892, calculated 13891; [**P**(d)+H_2O+H]$^+$ found 13909, calculated 13909.

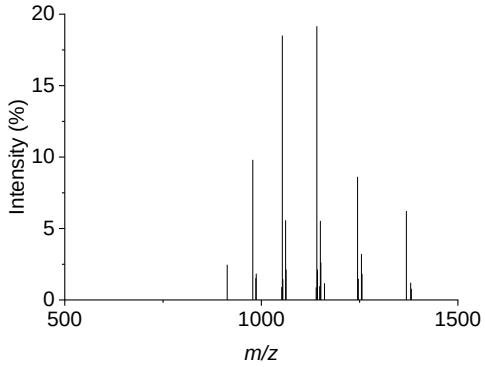


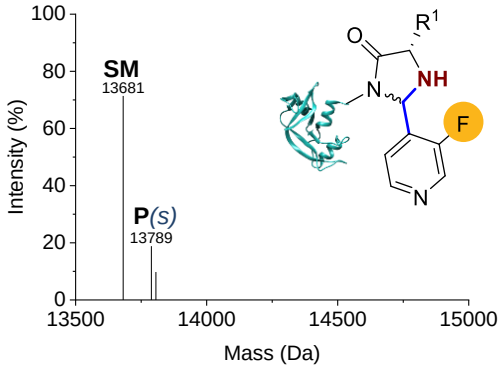
PCA 4.9

90%

84(s):6(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681; [**P**(s)+H]$^+$ found 13801, calculated 13800; [**P**(s)+H_2O+H]$^+$ found 13817, calculated 13818; [**P**(d)+H]$^+$ found 13937, calculated 13937.

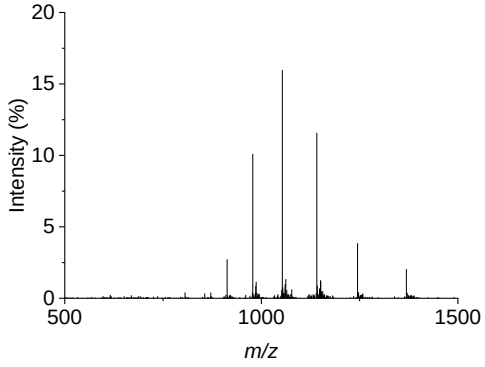


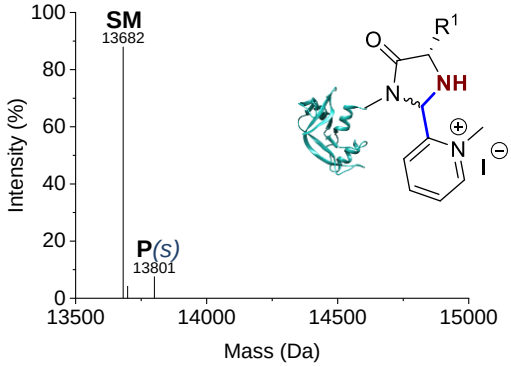


PCA 4.10

29%

MS (ESI$^+$) [**SM**+H]$^+$ found 13681, calculated 13681; [**P**(s)+H]$^+$ found 13789, calculated 13788; [**P**(s)+H_2O+H]$^+$ found 13806, calculated 13806.

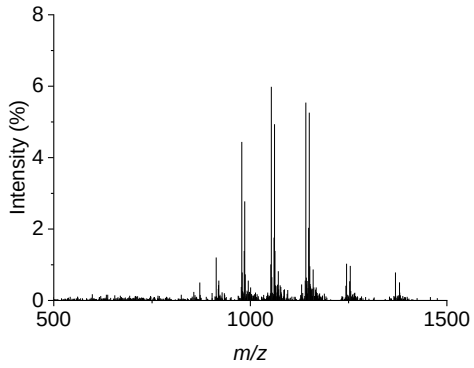


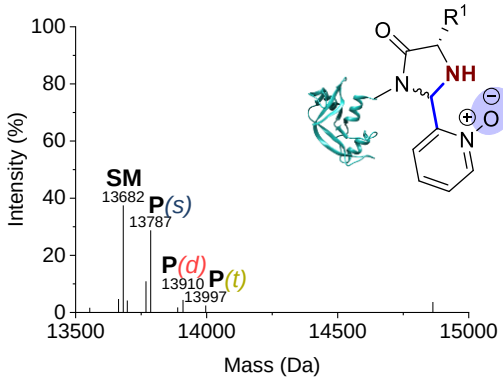


PCA 4.11

8%

MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681; [**SM**+H_2O+H]$^+$ found 13698, calculated 13699; **P**(s)+H_2O+H]$^+$ found 13801, calculated 13803.





PCA 4.12

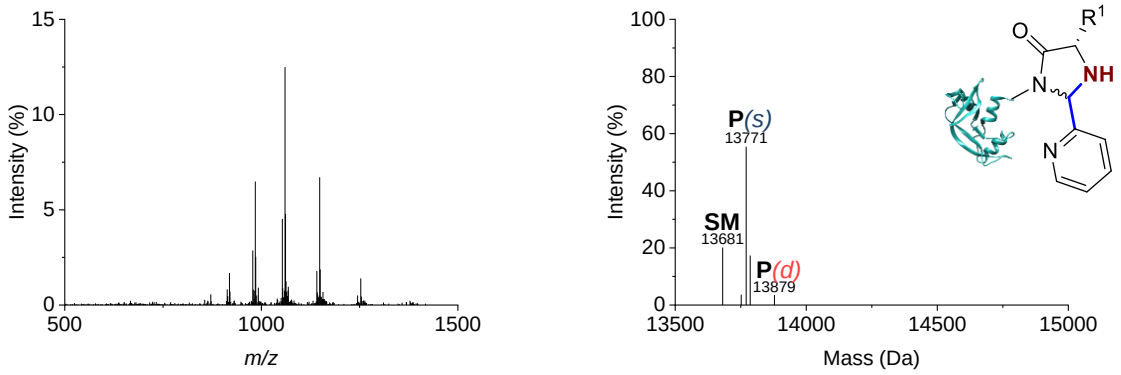
51%

42(s):

6(d):3(t)

MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 13664, calculated 13663; [**SM**+H]$^+$ found 13682, calculated 13681; [**SM**+H_2O+H]$^+$ found 13697, calculated 13699; [**P**(s)-H_2O+H]$^+$ found 13768, calculated 13768; [**P**(s)+H]$^+$ found 13787, calculated 13786; [**P**(d)+H]$^+$ found 13890, calculated 13891; [**P**(d)+H_2O+H]$^+$ found 13910, calculated 13909; [**P**(t)+H]$^+$ found 13997, calculated 13996.

RNase A *(purified)*

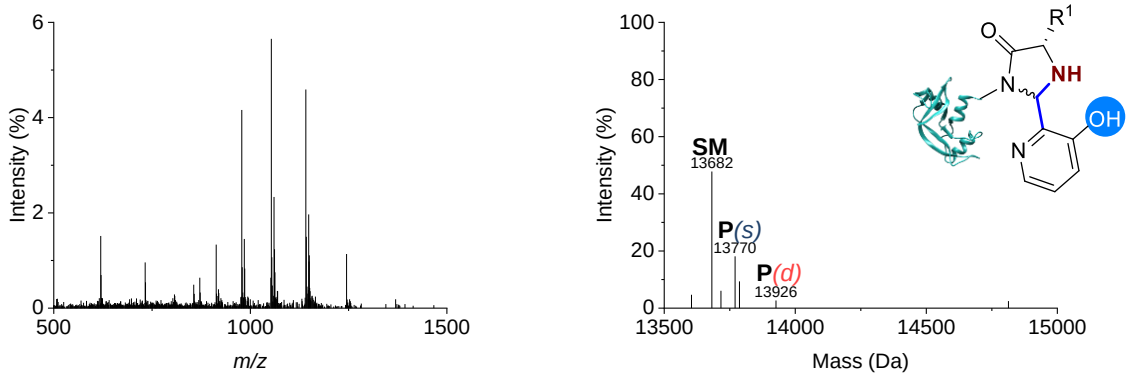


PCA 4.1

85%

76(s):3(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 13681, calculated 13681; [**P**(s)-H_2O+H]$^+$ found 13752, calculated 13752; [**P**(s)+H]$^+$ found 13771, calculated 13770; [**P**(s)+H_2O+H]$^+$ found 13786, calculated 13788; [**P**(d)+H_2O+H]$^+$ found 13879, calculated 13877.

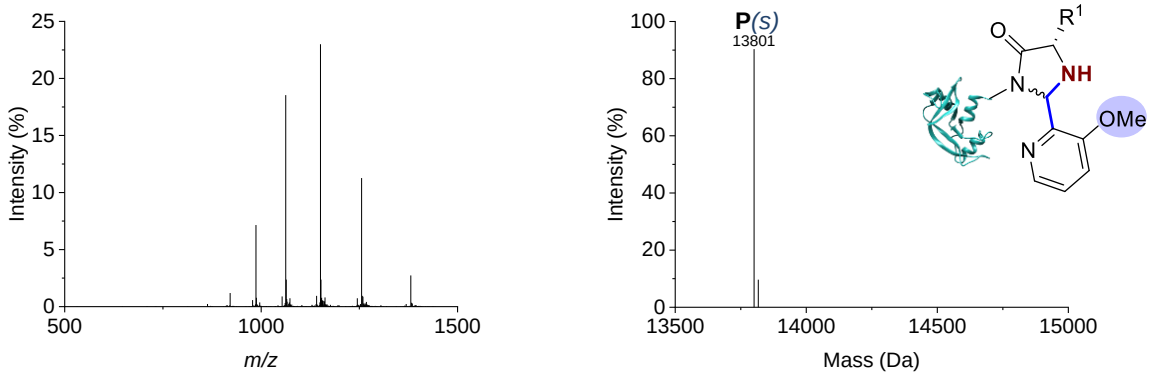


PCA 4.2

36%

33(s):3(d)

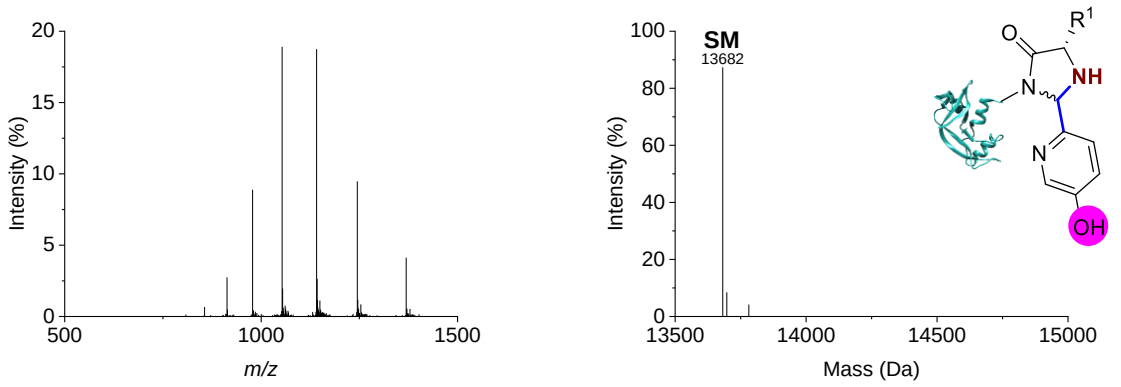
MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681; [**SM**+MeOH+H]$^+$ found 13717, calculated 13713; [**P**(s)-H_2O+H]$^+$ found 13770, calculated 13768; [**P**(s)+H]$^+$ found 13787, calculated 13786; [**P**(d)+2H_2O+H]$^+$ found 13926, calculated 13927.



PCA 4.3

100%

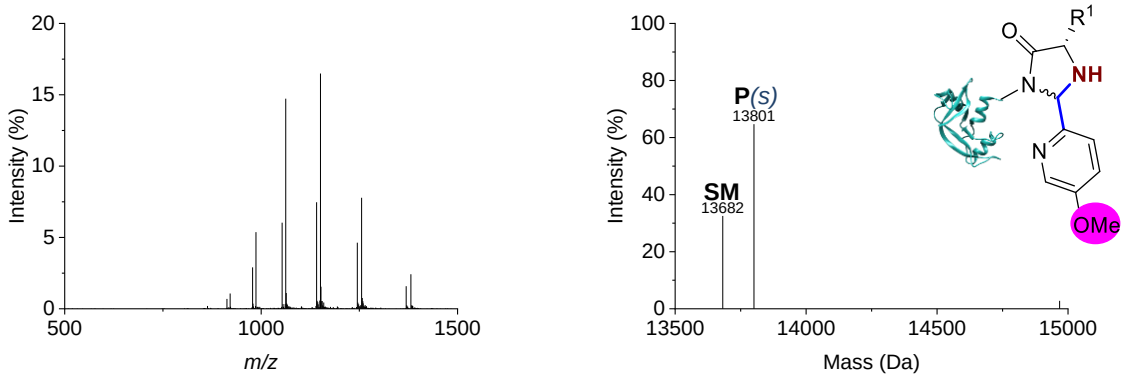
MS (ESI$^+$) [**P**(s)+H]$^+$ found 13801, calculated 13800; [**P**(s)+H_2O+H]$^+$ found 13817, calculated 13818.



PCA 4.4

0%

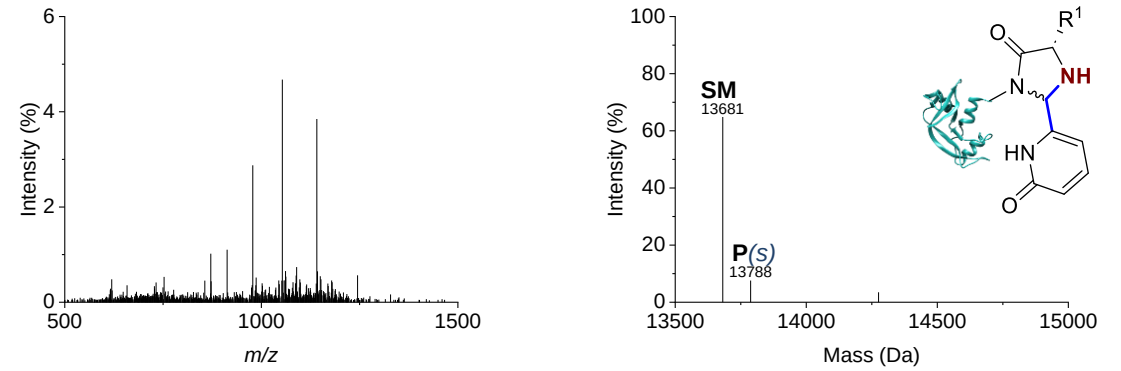
MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681; [**SM**+H_2O+H]$^+$ found 13697, calculated 13699; [**SM**+H_3PO_4+H]$^+$ found 13781, calculated 13779.



PCA 4.5

67%

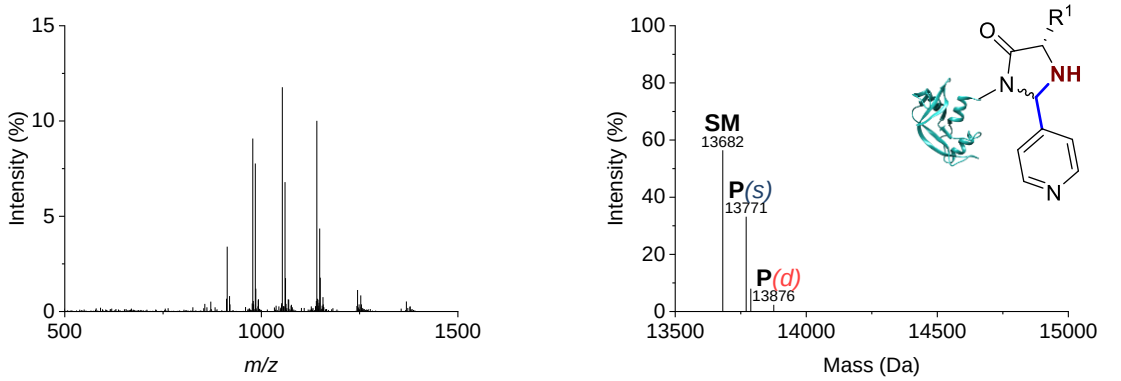
MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681; [**P**(s)+H]$^+$ found 13801, calculated 13800.



PCA 4.6

10%

MS (ESI$^+$) [**SM**+H]$^+$ found 13681, calculated 13681; [**P**(s)+H]$^+$ found 13788, calculated 13786.

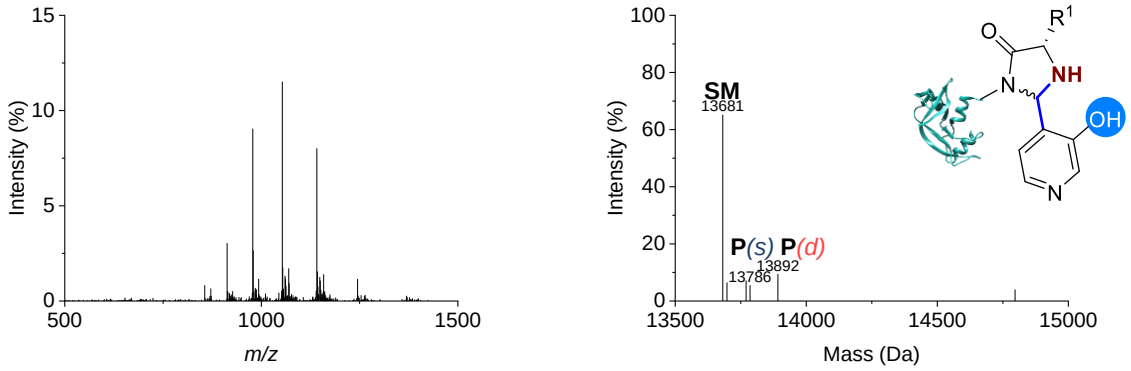


PCA 4.7

43%

41(s):2(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681; [**P**(s)+H]$^+$ found 13771, calculated 13770; [**P**(s)+H_2O+H]$^+$ found 13788, calculated 13788; [**P**(d)+H_2O+H]$^+$ found 13876, calculated 13877.

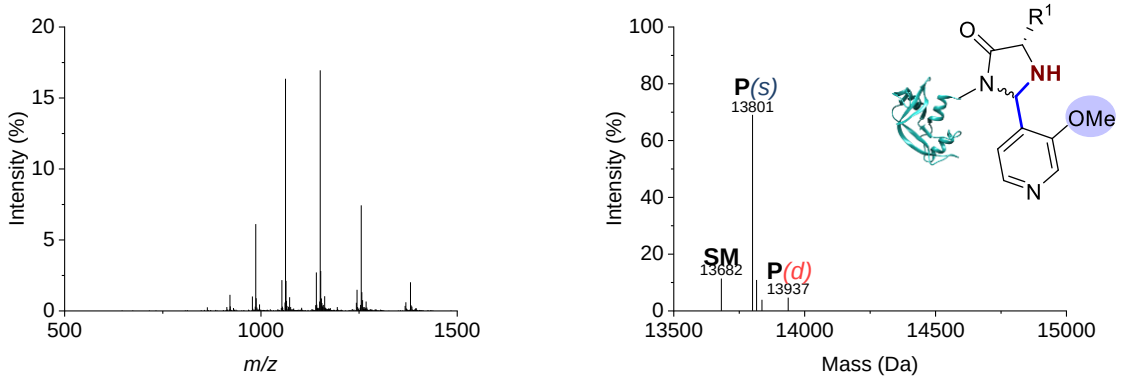


PCA 4.8

24%

14(s):10(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 13681, calculated 13681; [**SM**+H_2O+H]$^+$ found 13698, calculated 13699; [**P**(s)-H_2O+H]$^+$ found 13770, calculated 13768; [**P**(s)+H]$^+$ found 13786, calculated 13786; [**P**(d)+H]$^+$ found 13892, calculated 13891.

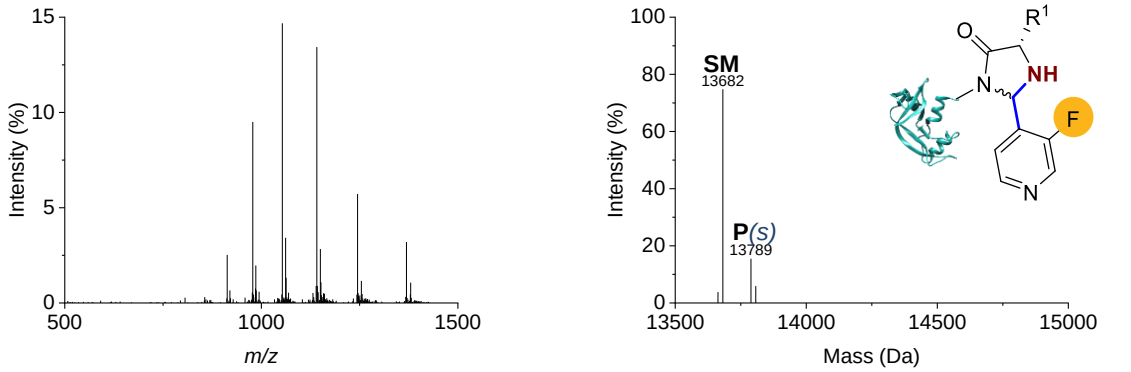


PCA 4.9

89%

84(s):5(d)

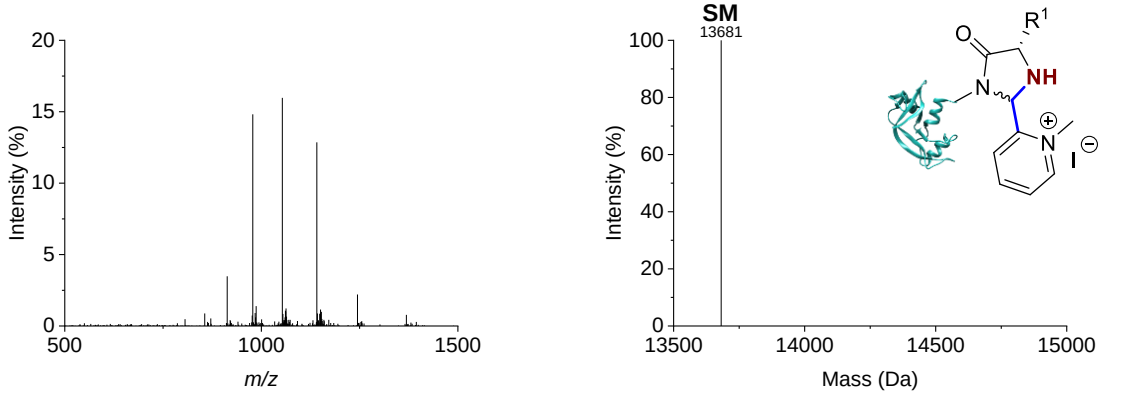
MS (ESI$^+$) [**SM**+H]$^+$ found 13682, calculated 13681; [**P**(s)+H]$^+$ found 13801, calculated 13800; [**P**(s)+H_2O+H]$^+$ found 13816, calculated 13818; [**P**(s)+MeOH+H]$^+$ found 13837, calculated 13832; [**P**(d)+H]$^+$ found 13937, calculated 13937.



PCA 4.10

21%

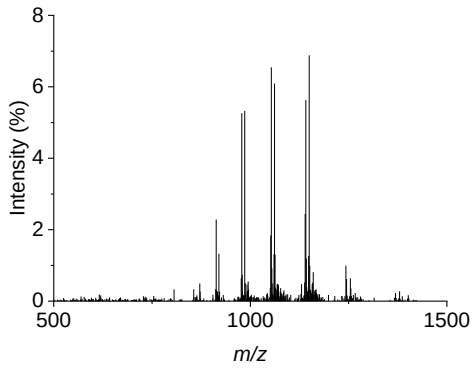
MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 13663, calculated 13663; [**SM**+H]$^+$ found 13682, calculated 13681; [**P**(s)+H]$^+$ found 13789, calculated 13788; [**P**(s)+H_2O+H]$^+$ found 13807, calculated 13806.

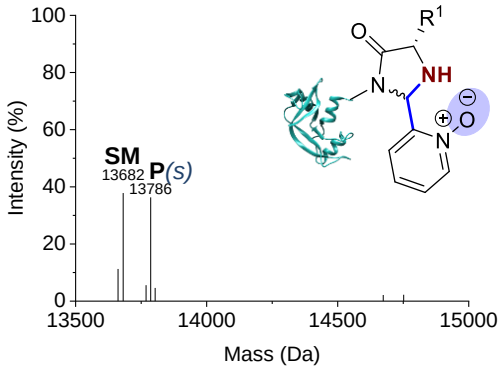


PCA 4.11

0%

MS (ESI$^+$) [**SM**+H]$^+$ found 13681, calculated 13681.



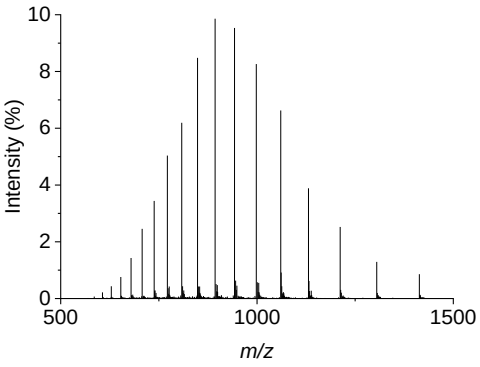


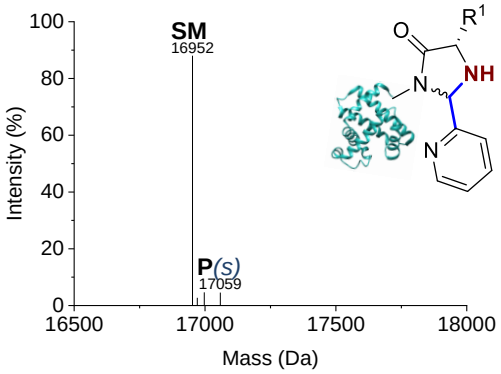
PCA 4.12

49%

MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 13662, calculated 13663; [**SM**+H]$^+$ found 13682, calculated 13681; [**P**(s)-H_2O+H]$^+$ found 13769, calculated 13768; [**P**(s)+H]$^+$ found 13786, calculated 13786; [**P**(s)+H_2O+H]$^+$ found 13804, calculated 13804.

Myoglobin *(crude)*

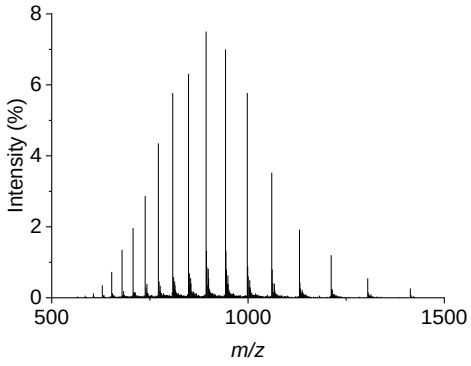


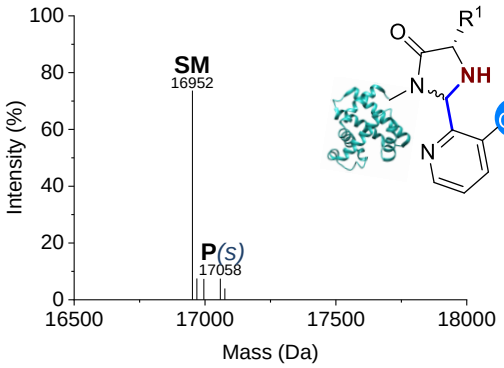


PCA 4.1

5%

MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951; [**SM**+H_2O+H]$^+$ found 16970, calculated 16969; [**SM**+MeCN+H]$^+$ found 16997, calculated 16992; [**P**(s)+H_2O+H]$^+$ found 17059, calculated 17058.

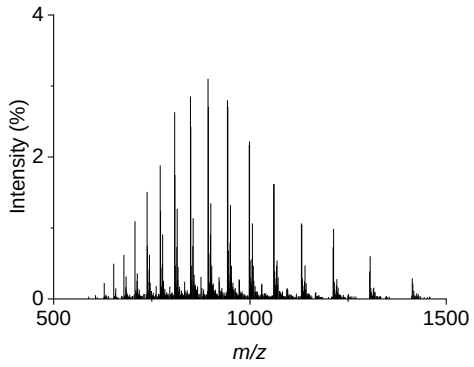


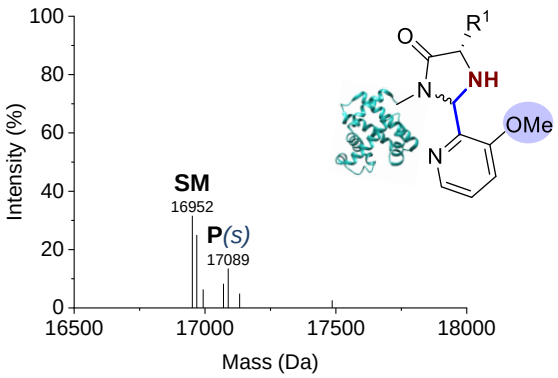


PCA 4.2

11%

MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951; [**SM**+H_2O+H]$^+$ found 16969, calculated 16969; [**SM**+MeCN+H]$^+$ found 16995, calculated 16992; [**P**(s)+H]$^+$ found 17058, calculated 17056; [**P**(s)+H_2O+H]$^+$ found 17076, calculated 17074.

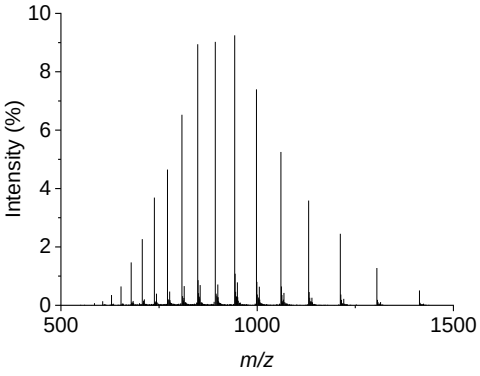


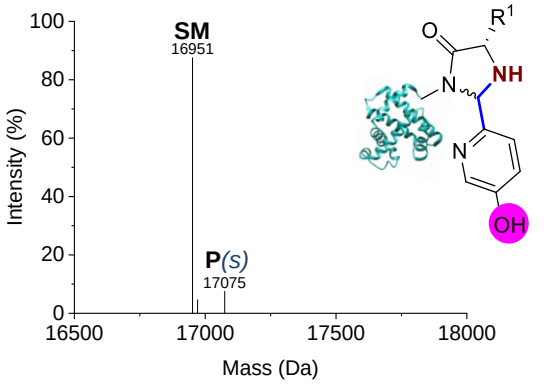


PCA 4.3

30%

MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951; [**SM**+H_2O+H]$^+$ found 16968, calculated 16969; [**SM**+MeCN+H]$^+$ found 16993, calculated 16992; [**P**(s)+H]$^+$ found 17071, calculated 17070; [**P**(s)+H_2O+H]$^+$ found 17089, calculated 17088; [**P**(s)+MeCN+NH_4]$^+$ found 17132, calculated 17128.

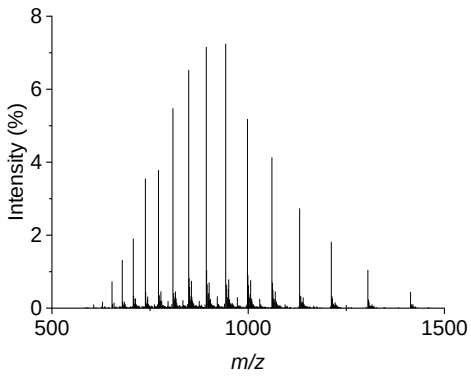


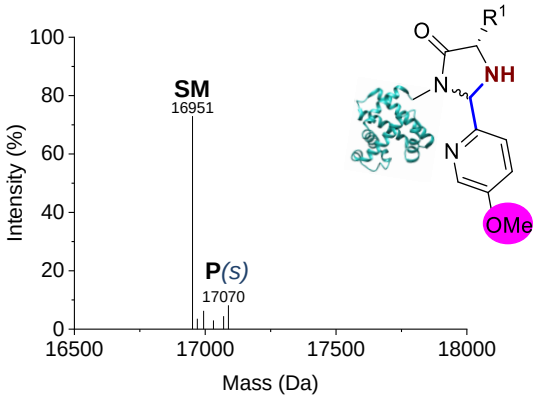


PCA 4.4

8%

MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16971, calculated 16969; [**P**(s)+H_2O+H]$^+$ found 17075, calculated 17074.

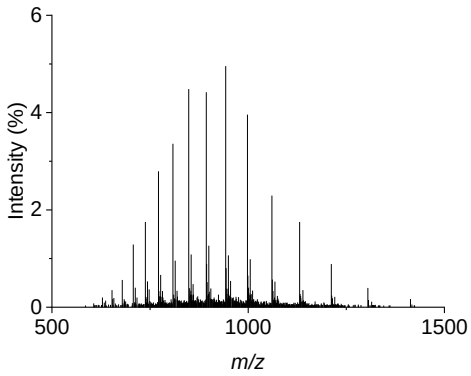


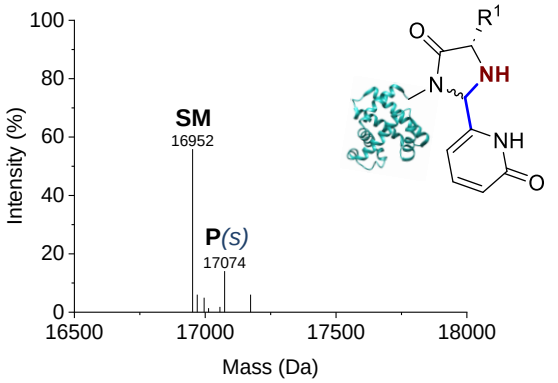


PCA 4.5

13%

MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16970, calculated 16969; [**SM**+MeCN+H]$^+$ found 16994, calculated 16992; [**SM**+2MeCN+H]$^+$ found 17032, calculated 17033; [**P**(s)+H]$^+$ found 17070, calculated 17070; [**P**(s)+H_2O+H]$^+$ found 17089, calculated 17088.

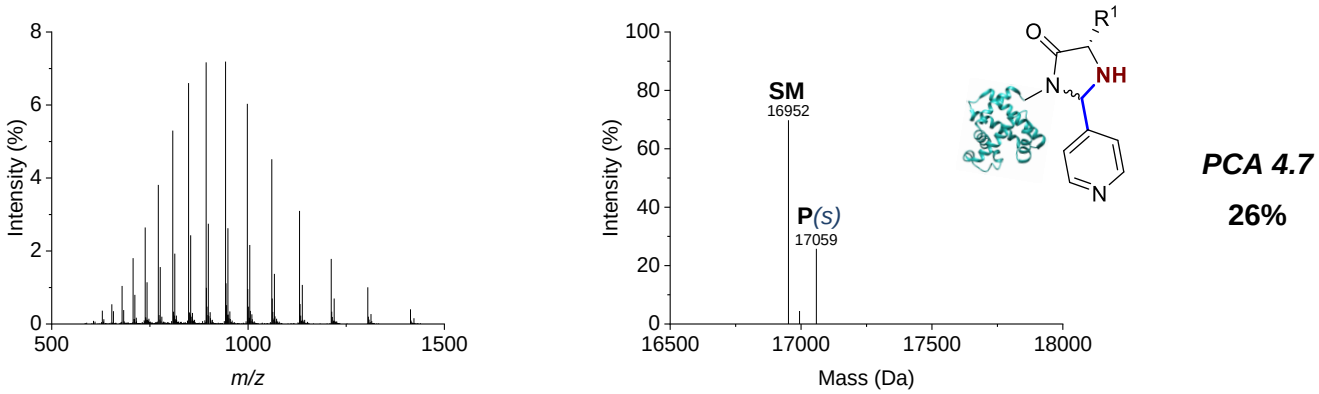




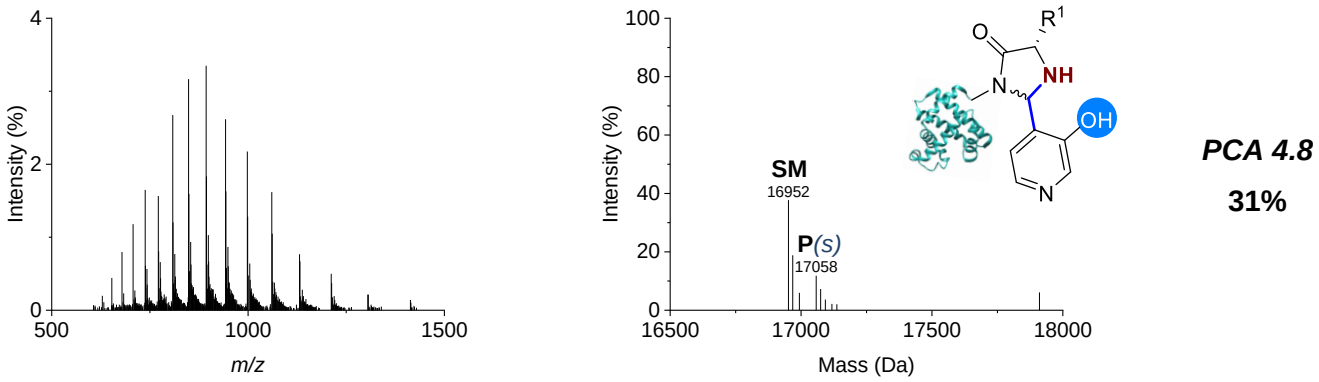
PCA 4.6

24%

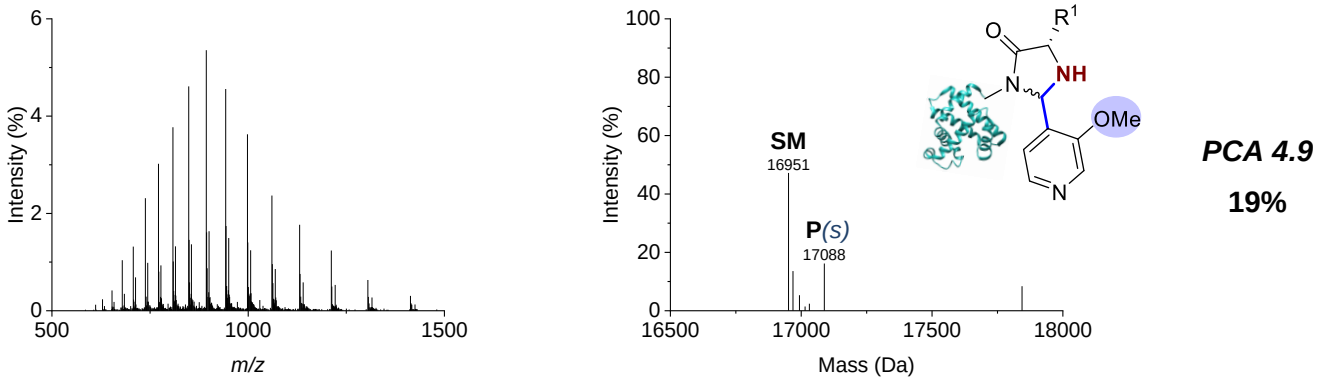
MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951; [**SM**+H_2O+H]$^+$ found 16969, calculated 16969; [**SM**+MeCN+H]$^+$ found 16996, calculated 16992; [**SM**+MeCN+Na]$^+$ found 17013, calculated 17014; [**P**(s)+H]$^+$ found 17056, calculated 17056; [**P**(s)+H_2O+H]$^+$ found 17074, calculated 17074; [**P**(s)+H_2O+H_3PO_4+H]$^+$ found 17174, calculated 17172.



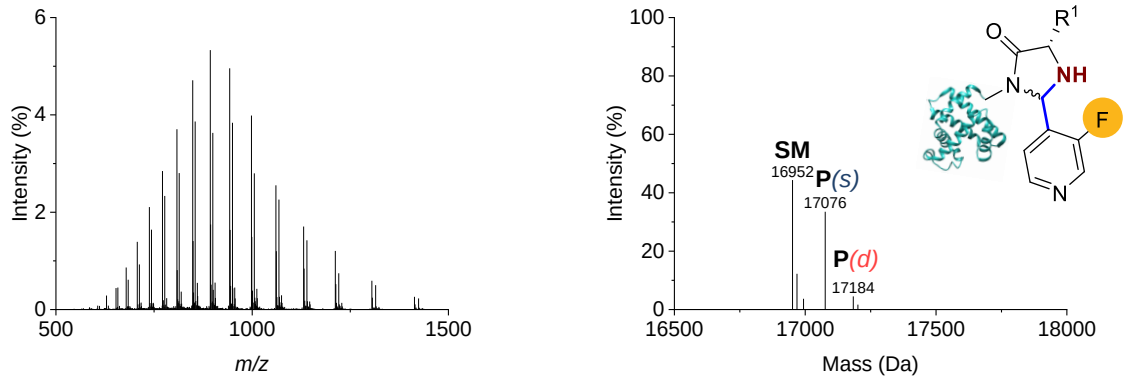
MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951; [**SM**+MeCN+H]$^+$ found 16994, calculated 16992; [**P**(s)+H_2O+H]$^+$ found 17059, calculated 17058.



MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951; [**SM**+H_2O+H]$^+$ found 16968, calculated 16969; [**SM**+MeCN+H]$^+$ found 16994, calculated 16992; [**P**(s)+H]$^+$ found 17059, calculated 17056; [**P**(s)+H_2O+H]$^+$ found 17074, calculated 17074; [**P**(s)+MeCN+Na]$^+$ found 17120, calculated 17119; [**P**(s)+DMSO+H]$^+$ found 17137, calculated 17134.



MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16969, calculated 16969; [**SM**+MeCN+H]$^+$ found 16993, calculated 16992; [**SM**+MeCN+Na]$^+$ found 17015, calculated 17014; [**SM**+2MeCN+H]$^+$ found 17031, calculated 17033; [**P**(s)+H_2O+H]$^+$ found 17088, calculated 17088.

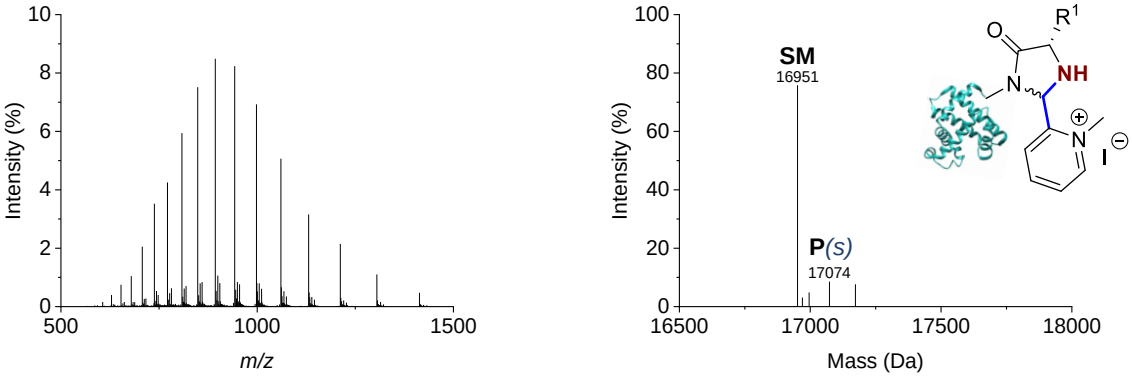


PCA 4.10

39%

33(s):6(d)

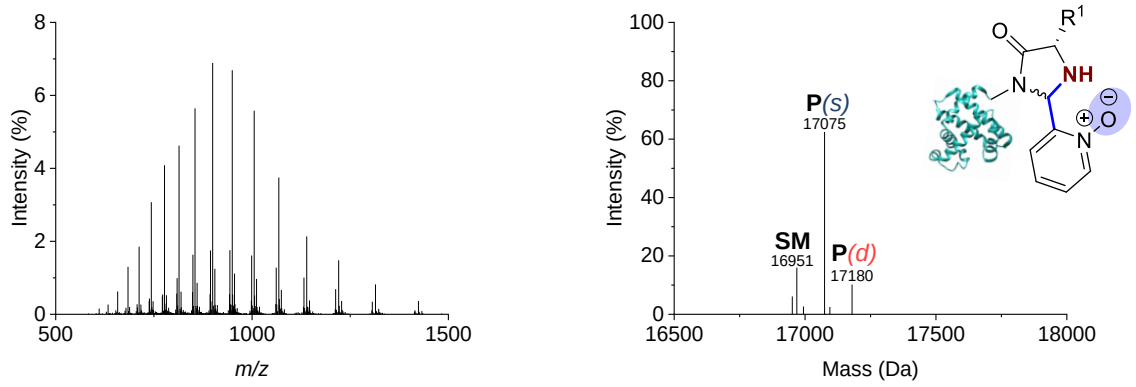
MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951; [**SM**+H_2O+H]$^+$ found 16968, calculated 16969; [**SM**+MeCN+H]$^+$ found 16993, calculated 16992; [**P**(s)+H_2O+H]$^+$ found 17076, calculated 17076; [**P**(d)+H_2O+H]$^+$ found 17184, calculated 17183; [**P**(d)+2H_2O+H]$^+$ found 17201, calculated 17201.



PCA 4.11

16%

MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16970, calculated 16969; [**SM**+MeCN+H]$^+$ found 16996, calculated 16992; [**P**(s)+H_2O+H]$^+$ found 17074, calculated 17073; [**P**(s)+H_2O+H_3PO_4+H]$^+$ found 17173, calculated 17171.



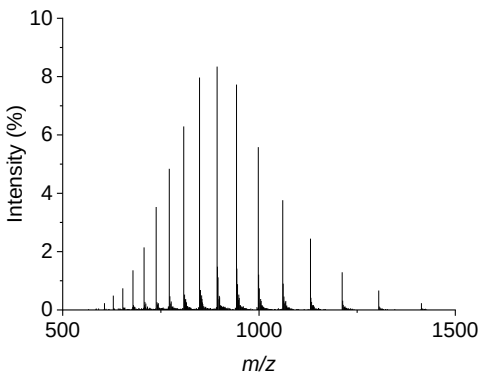
PCA 4.12

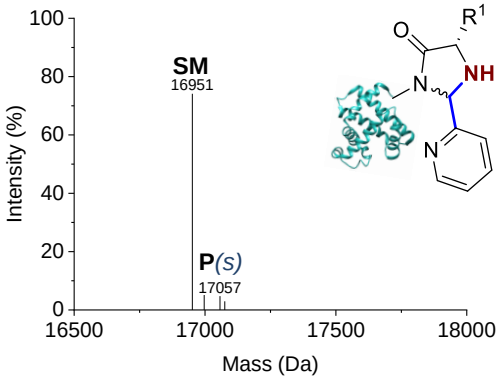
75%

65(s):10(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16968, calculated 16969; [**SM**+MeCN+H]$^+$ found 16994, calculated 16992; [**P**(s)+H_2O+H]$^+$ found 17075, calculated 17074; [**P**(s)+MeCN+H]$^+$ found 17095, calculated 17097; [**P**(d)+H_2O+H]$^+$ found 17180, calculated 17179.

Myoglobin *(purified)*

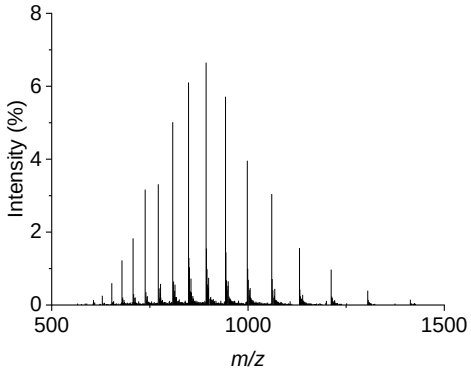


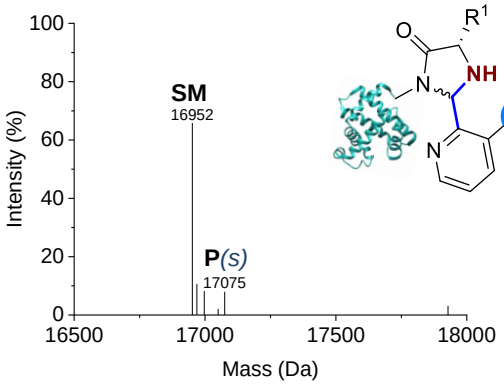


PCA 4.1

9%

MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+MeCN+H]$^+$ found 16997, calculated 16992; [**P**(s)+H_2O+H]$^+$ found 17057, calculated 17058; [**P**(s)+2H_2O+H]$^+$ found 17076, calculated 17076.

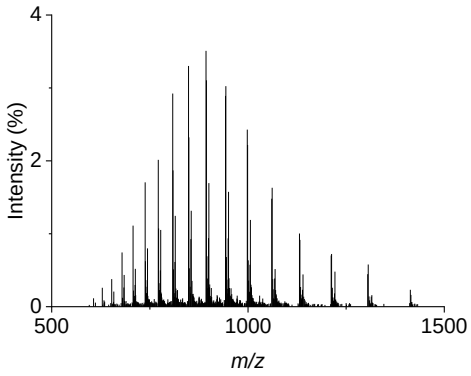


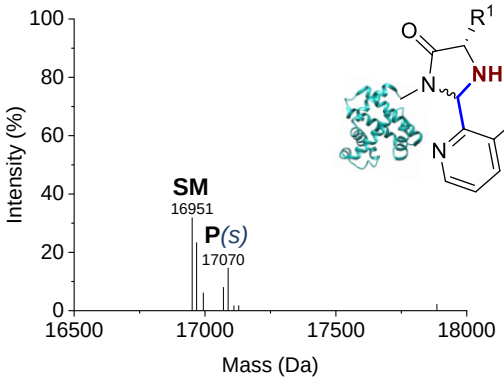


PCA 4.2

8%

MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951; [**SM**+H_2O+H]$^+$ found 16969, calculated 16969; [**SM**+MeCN+H]$^+$ found 16997, calculated 16992; [**SM**+H_3PO_4+H]$^+$ found 17050, calculated 17049; [**P**(s)+H_2O+H]$^+$ found 17075, calculated 17074.

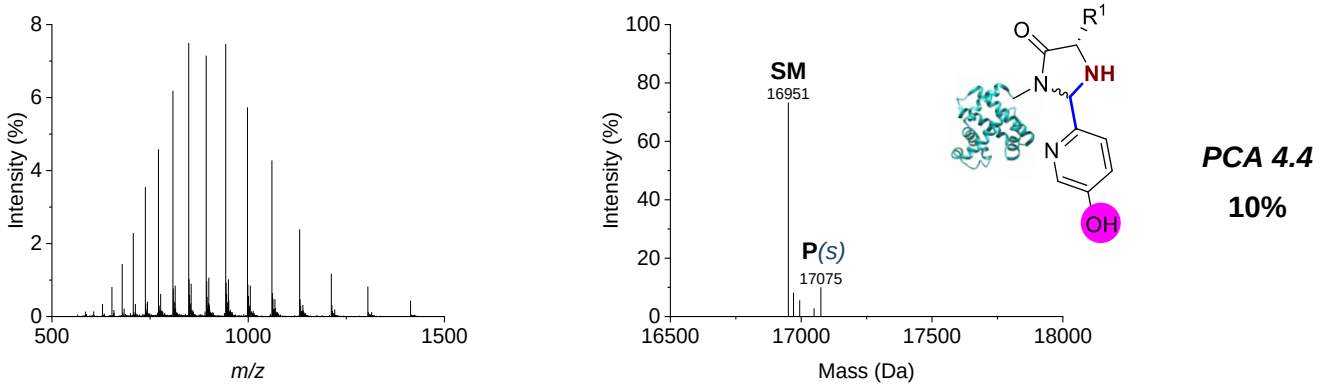




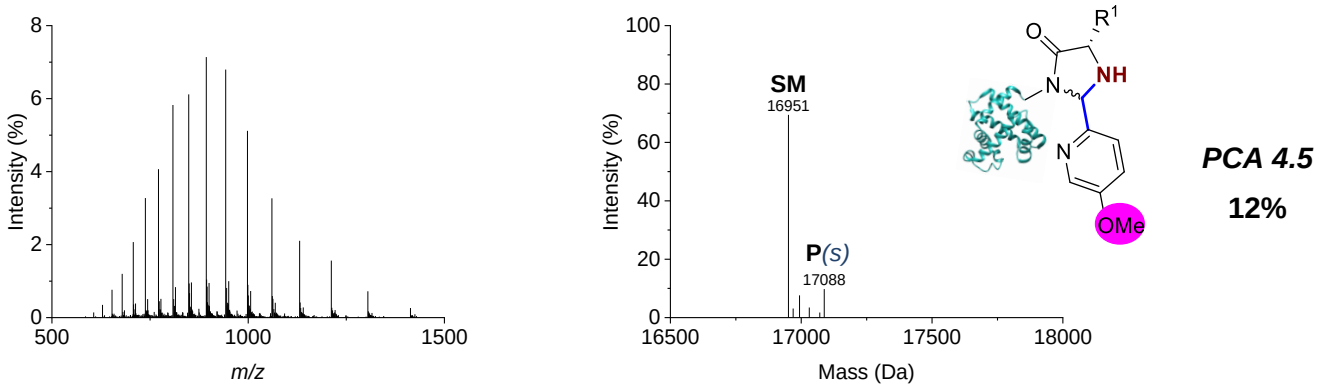
PCA 4.3

30%

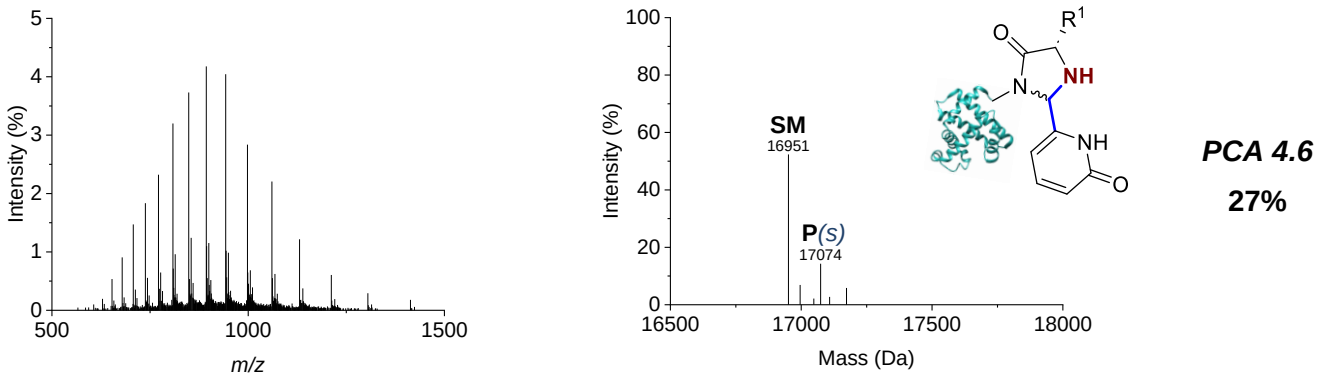
MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16968, calculated 16969; [**SM**+MeCN+H]$^+$ found 16993, calculated 16992; [**P**(s)+H]$^+$ found 17070, calculated 17070; [**P**(s)+H_2O+H]$^+$ found 17088, calculated 17088; [**P**(s)+MeCN+H]$^+$ found 17111, calculated 17111; [**P**(s)+MeCN+NH_4]$^+$ found 17129, calculated 17128.



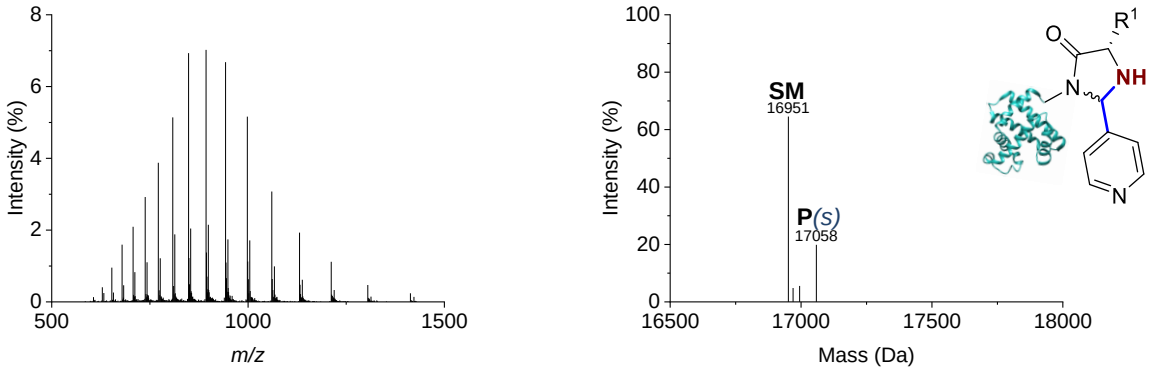
MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16971, calculated 16969; [**SM**+MeCN+H]$^+$ found 16994, calculated 16992; [**SM**+H_3PO_4+H]$^+$ found 17049, calculated 17049; [**P**(s)+H_2O+H]$^+$ found 17075, calculated 17074.



MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16969, calculated 16969; [**SM**+MeCN+H]$^+$ found 16993, calculated 16992; [**SM**+2MeCN+H]$^+$ found 17031, calculated 17033; [**P**(s)+H]$^+$ found 17071, calculated 17070; [**P**(s)+H_2O+H]$^+$ found 17088, calculated 17088.



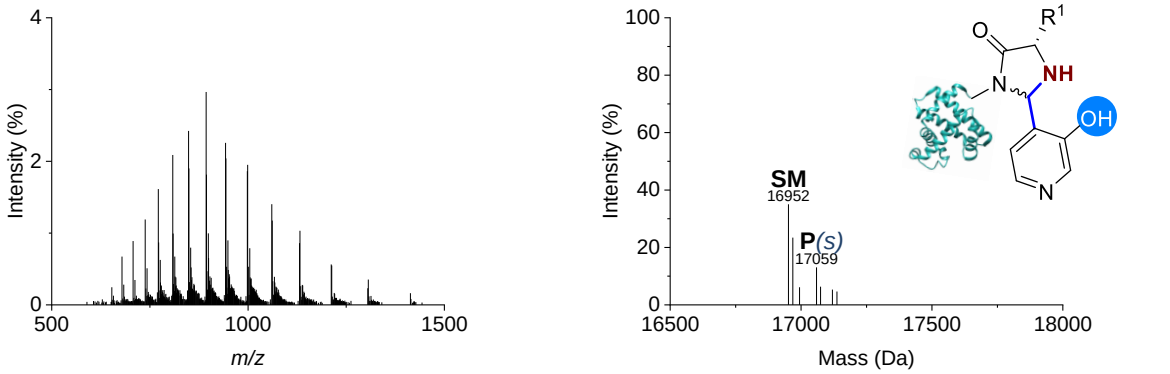
MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+MeCN+H]$^+$ found 16996, calculated 16992; [**SM**+H_3PO_4+H]$^+$ found 17048, calculated 17049; [**P**(s)+H_2O+H]$^+$ found 17074, calculated 17074; [**P**(s)+H_2O+MeOH+H]$^+$ found 17109, calculated 17106; [**P**(s)+H_2O+H_3PO_4+H]$^+$ found 17173, calculated 17172.



PCA 4.7

21%

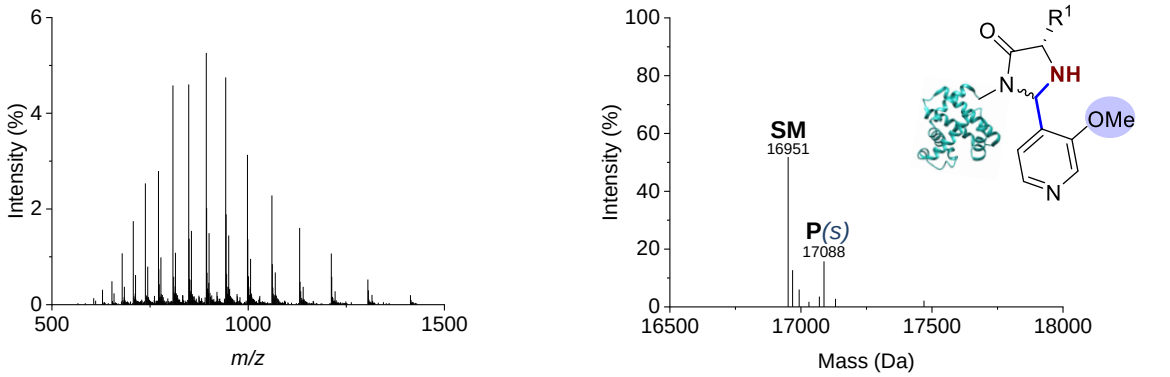
MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16970, calculated 16969; [**SM**+MeCN+H]$^+$ found 16995, calculated 16992; [**P**(s)+H_2O+H]$^+$ found 17058, calculated 17058.



PCA 4.8

30%

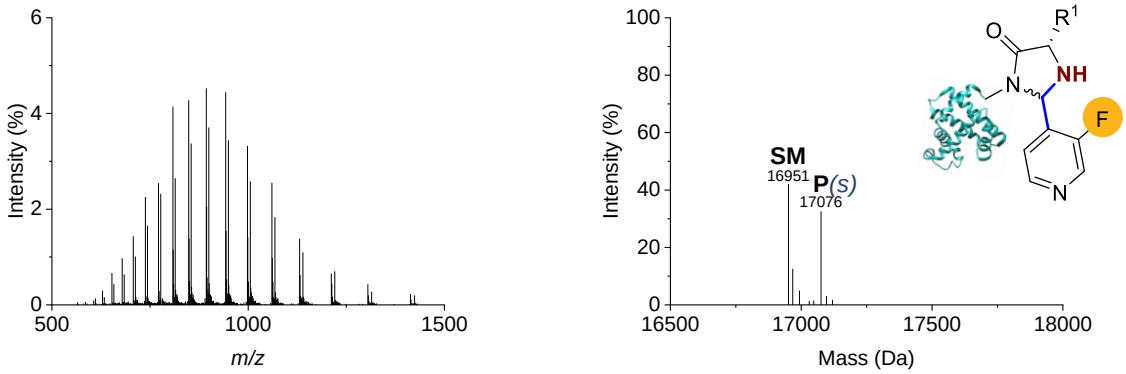
MS (ESI$^+$) [**SM**+H]$^+$ found 16952, calculated 16951; [**SM**+H_2O+H]$^+$ found 16968, calculated 16969; [**SM**+MeCN+H]$^+$ found 16994, calculated 16992; [**P**(s)+H]$^+$ found 17058, calculated 17056; [**P**(s)+H_2O+H]$^+$ found 17075, calculated 17074; [**P**(s)+MeOH+H]$^+$ found 17093, calculated 17088; [**P**(s)+MeCN+Na]$^+$ found 17118, calculated 17119; [**P**(s)+DMSO+H]$^+$ found 17137, calculated 17134.



PCA 4.9

23%

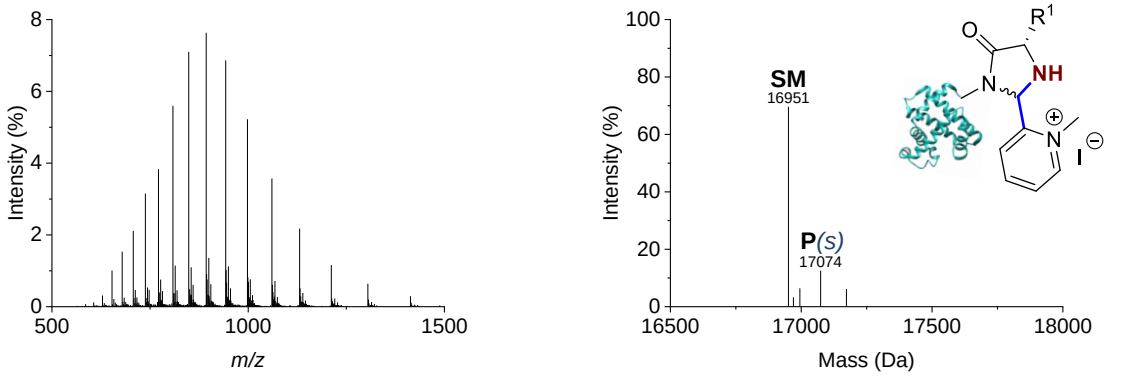
MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16968, calculated 16969; [**SM**+MeCN+H]$^+$ found 16993, calculated 16992; [**SM**+2MeCN+H]$^+$ found 17030, calculated 17033; [**P**(s)+H]$^+$ found 17070, calculated 17070; [**P**(s)+H_2O+H]$^+$ found 17088, calculated 17088; [**P**(s)+MeCN+Na]$^+$ found 17132, calculated 17133.



PCA 4.10

37%

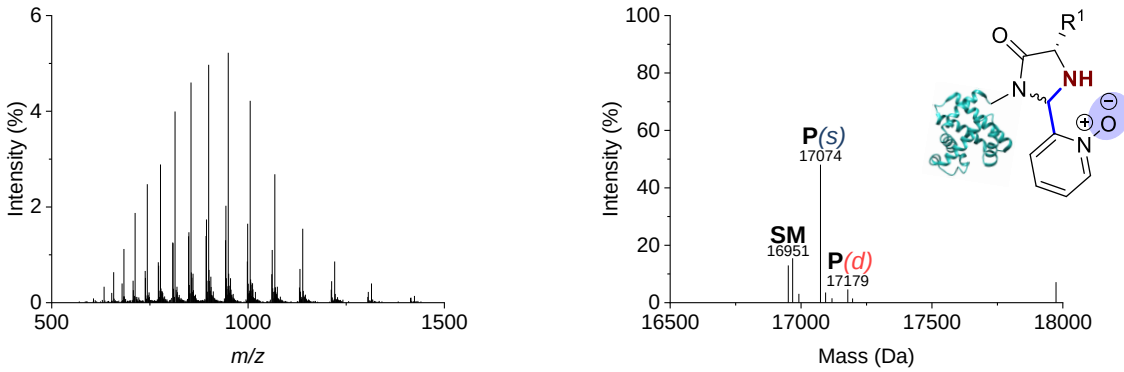
MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16968, calculated 16969; [**SM**+MeCN+H]$^+$ found 16993, calculated 16992; [**SM**+DMSO+H]$^+$ found 17031, calculated 17029; [**SM**+H_3PO_4+H]$^+$ found 17048, calculated 17049; [**P**(s)+H_2O+H]$^+$ found 17076, calculated 17076; [**P**(s)+MeCN+H]$^+$ found 17096, calculated 17099; [**P**(s)+MeCN+H_2O+H]$^+$ found 17120, calculated 17117.



PCA 4.11

19%

MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16970, calculated 16969; [**SM**+MeCN+H]$^+$ found 16995, calculated 16992; [**P**(s)+H_2O+H]$^+$ found 17074, calculated 17073; [**P**(s)+H_2O+H_3PO_4+H]$^+$ found 17173, calculated 17171.



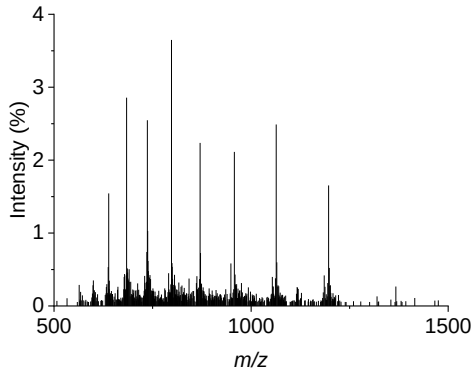
PCA 4.12

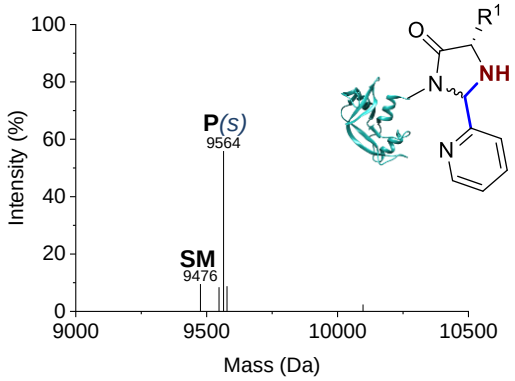
66%

59(s):7(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 16951, calculated 16951; [**SM**+H_2O+H]$^+$ found 16968, calculated 16969; [**SM**+MeCN+H]$^+$ found 16992, calculated 16992; [**P**(s)+H_2O+H]$^+$ found 17074, calculated 17074; [**P**(s)+MeCN+H]$^+$ found 17094, calculated 17097; [**P**(s)+MeCN+Na]$^+$ found 17118, calculated 17119; [**P**(d)+H_2O+H]$^+$ found 17179, calculated 17179; [**P**(d)+2H_2O+H]$^+$ found 17197, calculated 17197.

HASPA G2S *(crude)*

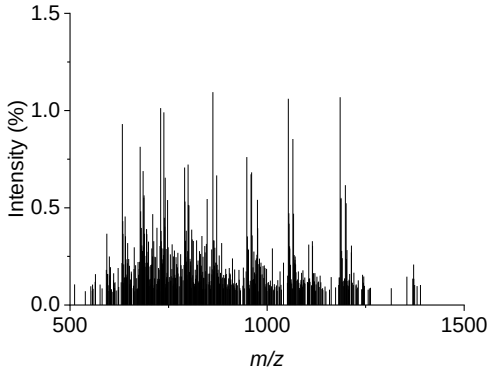


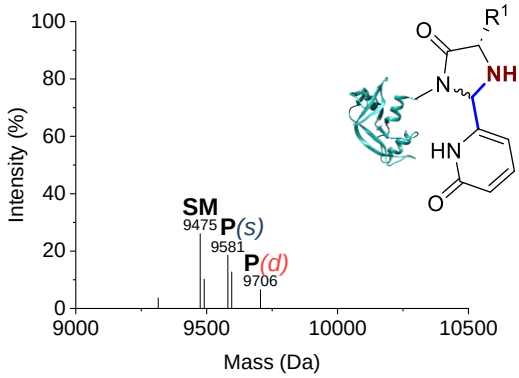


PCA 4.1

88%

MS (ESI$^+$) [**SM**+H]$^+$ found 9476, calculated 9476; [**P**(s)-H_2O+H]$^+$ found 9547, calculated 9547; [**P**(s)+H]$^+$ found 9564, calculated 9565; [**P**(s)+H_2O+H]$^+$ found 9578, calculated 9583.



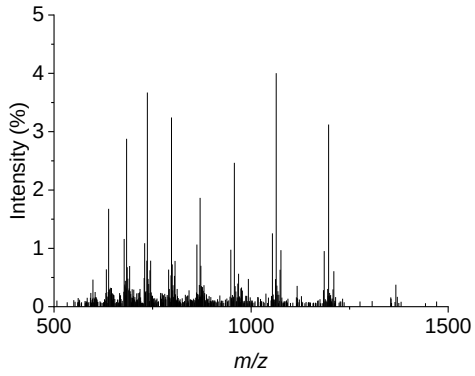


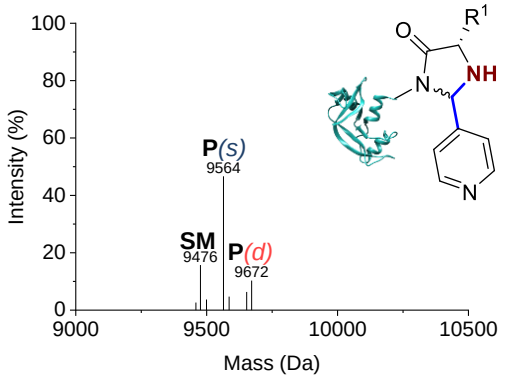
PCA 4.6

51%

42(s):9(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 9475, calculated 9476; [**SM**+H_2O+H]$^+$ found 9491, calculated 9494; [**P**(s)+H]$^+$ found 9581, calculated 9581; [**P**(s)+H_2O+H]$^+$ found 9596, calculated 9599; [**P**(d)+H_2O+H]$^+$ found 9706, calculated 9704.



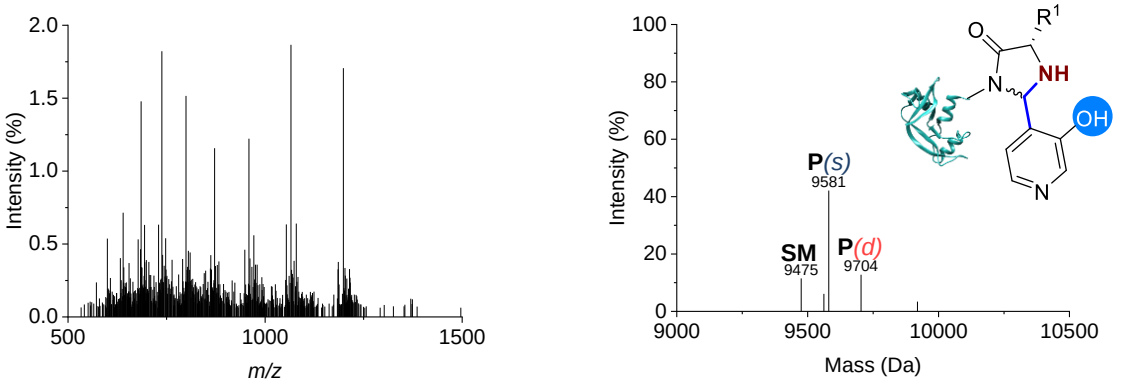


PCA 4.7

76%

57(s):18(d)

MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 9459, calculated 9458; [**SM**+H]$^+$ found 9476, calculated 9476; [**SM**+Na]$^+$ found 9499, calculated 9498; [**P**(s)+H]$^+$ found 9564, calculated 9565; [**P**(s)+H_2O+H]$^+$ found 9586, calculated 9583; [**P**(d)+H_2O+H]$^+$ found 9653, calculated 9564; [**P**(d)+2H_2O+H]$^+$ found 9672, calculated 9672.

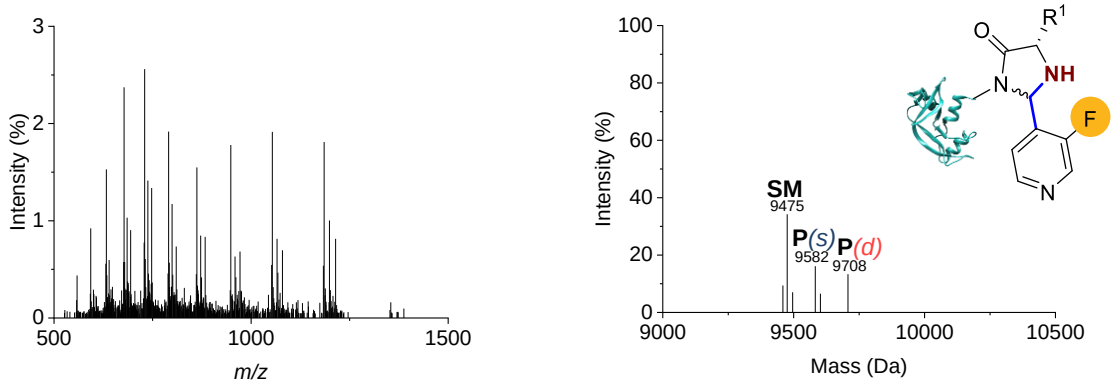


PCA 4.8

84%

67(s):18(d)

MS (ESI$^+$) [**SM**+H]$^+$ found 9475, calculated 9476; [**P**(s)-H_2O+H]$^+$ found 9562, calculated 9563; [**P**(s)+H]$^+$ found 9581, calculated 9581; [**P**(d)+H_2O+H]$^+$ found 9704, calculated 9704.

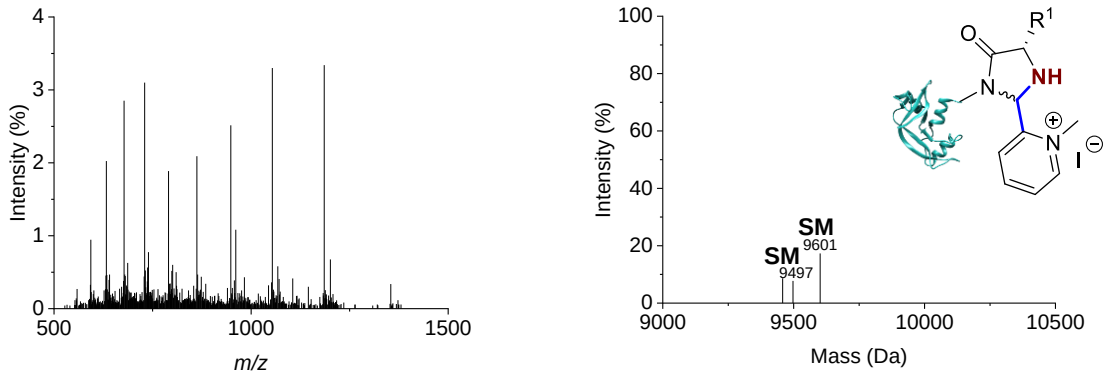


PCA 4.10

42%

26(s):15(d)

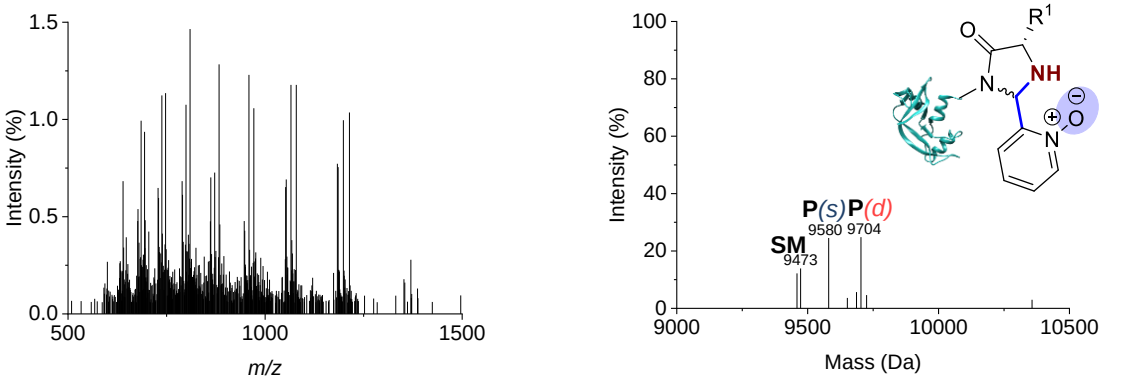
MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 9459, calculated 9458; [**SM**+H]$^+$ found 9475, calculated 9476; [**SM**+H_2O+H]$^+$ found 9496, calculated 9494; [**P**(s)+H]$^+$ found 9582, calculated 9583; [**P**(s)+H_2O+H]$^+$ found 9602, calculated 9601; [**P**(d)+H_2O+H]$^+$ found 9708, calculated 9708.



PCA 4.11

0%

MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 9458, calculated 9458; [**SM**+Na]$^+$ found 9497, calculated 9498; [**SM**+I+2H]$^+$ found 9601, calculated 9604.



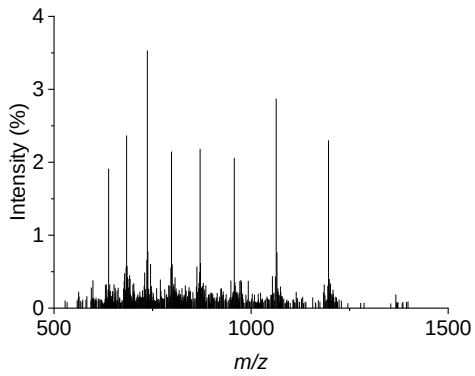
PCA 4.12

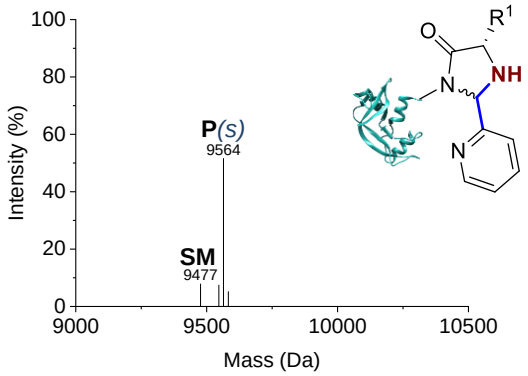
68%

30(s):38(d)

MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 9459, calculated 9458; [**SM**+H]$^+$ found 9473, calculated 9476; [**P**(s)+H]$^+$ found 9580, calculated 9581; [**P**(d)+H]$^+$ found 9686, calculated 9686; [**P**(d)+H_2O+H]$^+$ found 9704, calculated 9704.

HASPA G2S *(purified)*

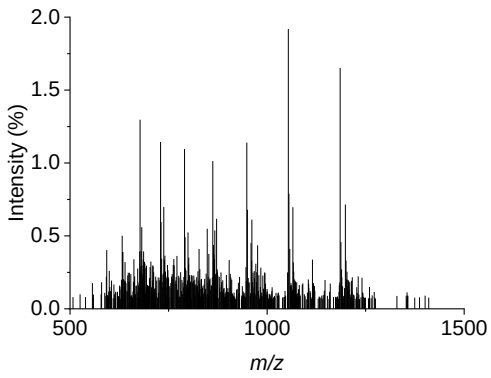


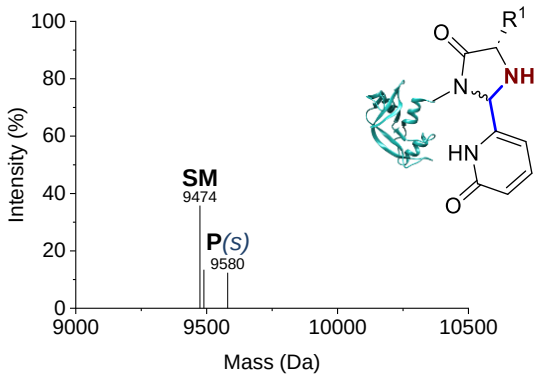


PCA 4.1

89%

MS (ESI$^+$) [**SM**+H]$^+$ found 9477, calculated 9476; [**P**(s)-H_2O+H]$^+$ found 9547, calculated 9547; [**P**(s)+H]$^+$ found 9564, calculated 9565; [**P**(s)+H_2O+H]$^+$ found 9583, calculated 9583.

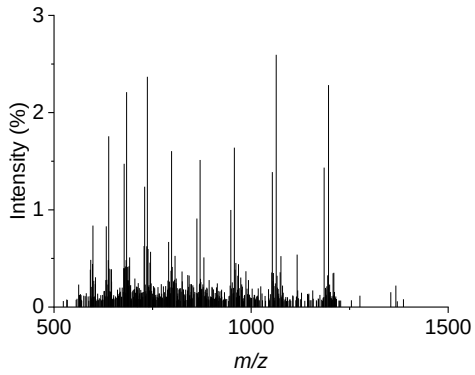


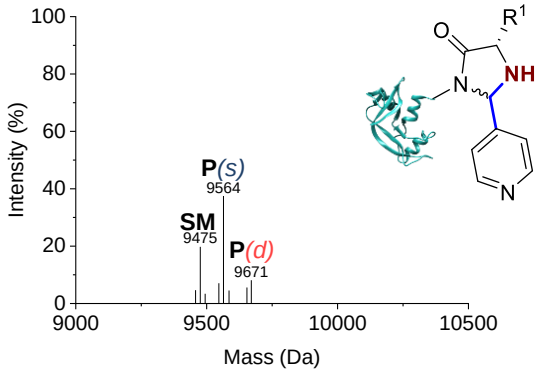


PCA 4.6

20%

MS (ESI$^+$) [**SM**+H]$^+$ found 9474, calculated 9476; [**SM**+H_2O+H]$^+$ found 9489, calculated 9494; [**P**(s)+H]$^+$ found 9580, calculated 9581.



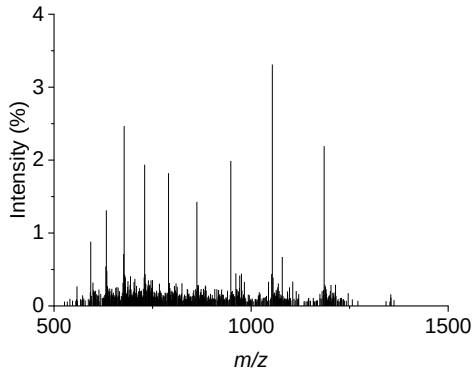


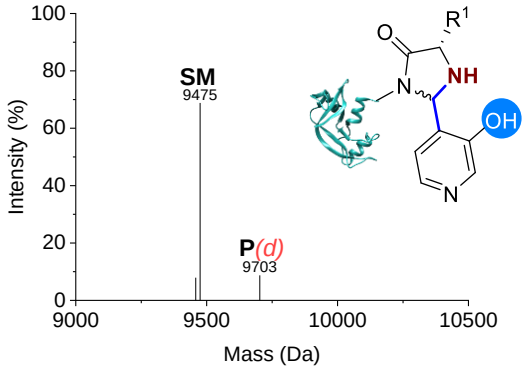
PCA 4.7

69%

54(s):15(d)

MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 9458, calculated 9458; [**SM**+H]$^+$ found 9475, calculated 9476; [**SM**+H_2O+H]$^+$ found 9494, calculated 9494; [**P**(s)-H_2O+H]$^+$ found 9547, calculated 9547; [**P**(s)+H]$^+$ found 9564, calculated 9565; [**P**(s)+H_2O+H]$^+$ found 9586, calculated 9583; [**P**(d)+H_2O+H]$^+$ found 9654, calculated 9564; [**P**(d)+2H_2O+H]$^+$ found 9671, calculated 9672.



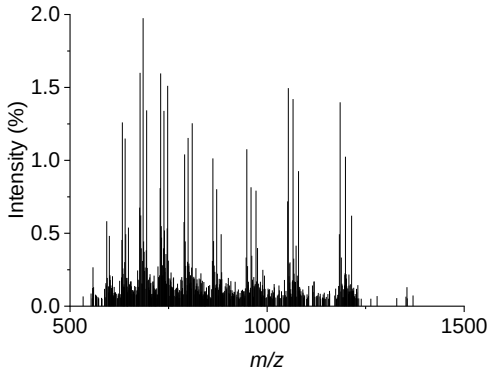


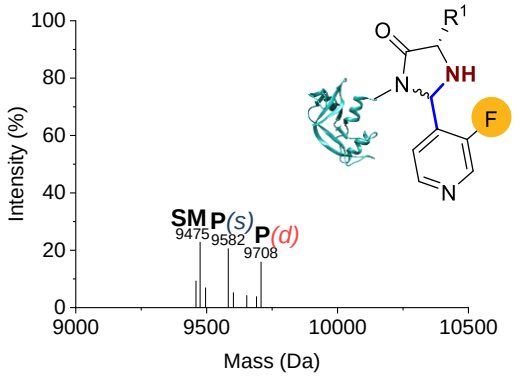
PCA 4.8

10%

0(s):10(d)

MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 9458, calculated 9458; [**SM**+H]$^+$ found 9475, calculated 9476; [**P**(d)+H_2O+H]$^+$ found 9703, calculated 9704.



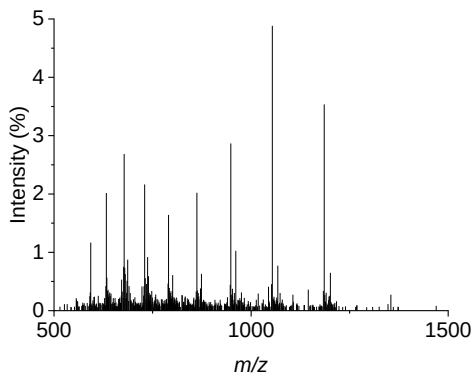


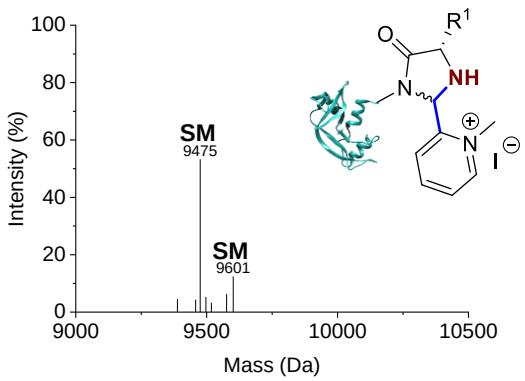
PCA 4.10

54%

30(s):23(d)

MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 9459, calculated 9458; [**SM**+H]$^+$ found 9475, calculated 9476; [**SM**+H_2O+H]$^+$ found 9496, calculated 9494; [**P**(s)+H]$^+$ found 9582, calculated 9583; [**P**(s)+H_2O+H]$^+$ found 9602, calculated 9601; [**P**(d)+H]$^+$ found 9691, calculated 9690; [**P**(d)+H_2O+H]$^+$ found 9708, calculated 9708.

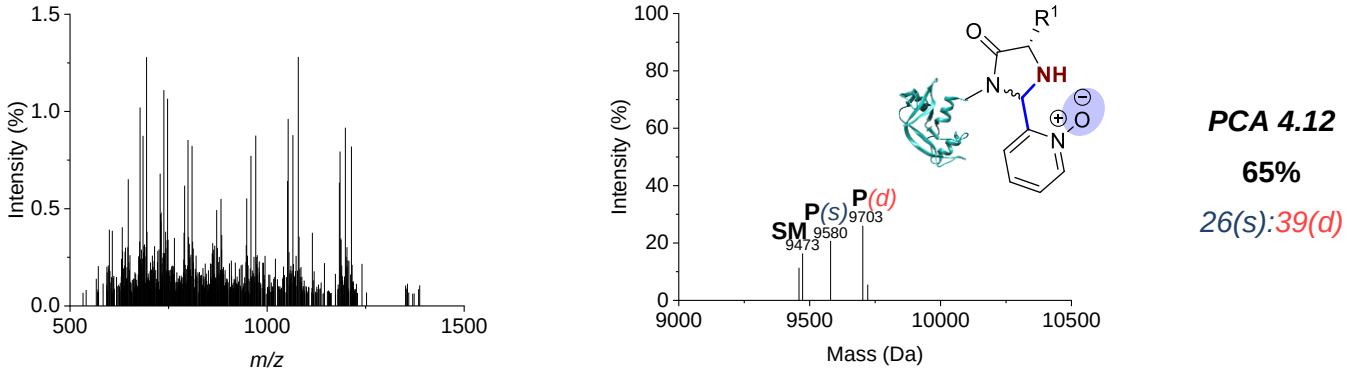




PCA 4.11

0%

MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 9458, calculated 9458; [**SM**+H]$^+$ found 9475, calculated 9476; [**SM**+Na]$^+$ found 9497, calculated 9498; [**SM**+MeCN+H]$^+$ found 9518, calculated 9517; [**SM**+I+2H]$^+$ found 9601, calculated 9604.



MS (ESI$^+$) [**SM**-H_2O+H]$^+$ found 9459, calculated 9458; [**SM**+H]$^+$ found 9473, calculated 9476; [**P**(s)+H]$^+$ found 9580, calculated 9581; [**P**(d)+H_2O+H]$^+$ found 9703, calculated 9704; [**P**(d)+2H_2O+H]$^+$ found 9722, calculated 9722.

Figure S4.7. Raw (left) and deconvoluted (right) mass spectra for RNase A, myoglobin and HASPA G2S conjugates. **SM** = unmodified modified protein; **P** = modified protein.

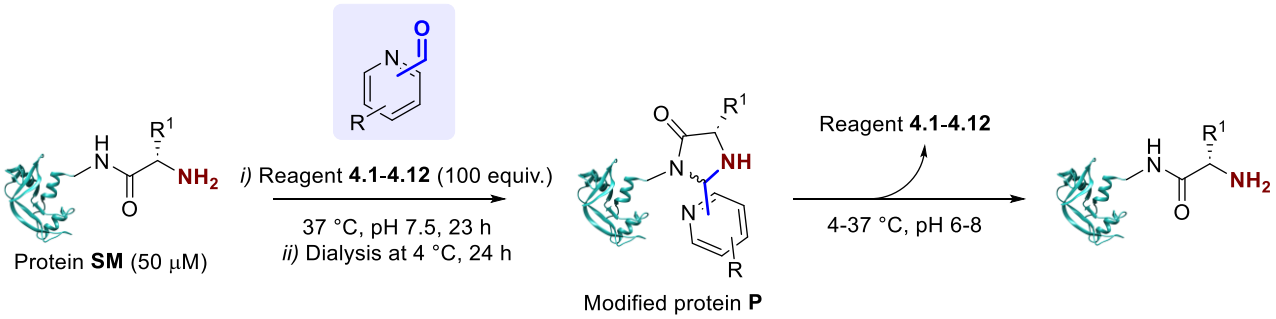
4.4.4.2 CjX183-D selectivity

Reagents **4.1**-**4.2**, **4.4**, **4.6**-**4.8** and **4.10**-**4.12** were used to modify CjX183-D WT and CjX183-D R51K under the conditions outlined in **General Procedure 4A**, on a 50 µL scale. Conversion was determined by LC-MS analysis without purification (Crude). Protein conjugates **P** were then purified by dialysis to remove excess reagent (4 °C, 3.5 kDa MWCO; 1 × 50 mM pH 7 sodium phosphate buffer, 5 h; 1 × water, 3 h; 1 × water, 16 h), and conversion was again determined by LC-MS analysis (Purified).

PCA	Conversion (%)			
	CjX183-D WT		CjX183-D R51K	
	Crude	*Purified*	*Crude*	*Purified*
4.1	**64** *54(s):10(d)*	**52** *43(s):9(d)*	**53** *47(s):6(d)*	**45** *41(s):5(d)*
4.2	**60** *23(s):20(d):17(t)*	**45**	**53** *19(s):21(d):13(t)*	**46** *37(s):9(d)*
4.4	**10**	**6**	**5**	**0**
4.6	**80** *37(s):35(d):8(t)*	**76** *43(s):25(d):8(t)*	**72** *30(s):42(d)*	**80** *34(s):45(d)*
4.7	**73** *57(s):14(d):2(t)*	**69** *59(s):8(d):2(t)*	**65** *54(s):11(d)*	**64** *60(s):4(d)*
4.8	**66** *21(s):39(d):6(t)*	**32** *27(s):5(d)*	**76** *25(s):43(d):8(t)*	**51** *46(s):5(d)*
4.10	**74** *57(s):15(d):2(t)*	**73** *68(s):2(d):3(t)*	**61** *46(s):15(d)*	**67** *60(s):6(d):1(t)*
4.11	**41** *13(s):29(d)*	**28** *9(s):20(d)*	**38** *9(s):28(d)*	**29** *9(s):19(d)*
4.12	**67** *29(s):30(d):7(t)*	**65** *30(s):29(d):5(t)*	**62** *31(s):23(d):7(t)*	**61** *29(s):23(d):9(t)*

Table S4.5. Conversions for the modification of CjX183-D WT and CjX183-D R51K before (crude) and after (purified) dialysis at 4 °C. s = single, d = double, t = triple modification.

4.4.4.3 pH and temperature stability



Following purification of conjugates prepared in section 4.4.3.1, aliquots of protein (20 µL) were diluted with water (80 µL), placed inside a 3.5 kDa MWCO dialysis membrane, and incubated in sodium phosphate buffer (40 mL, 50 mM) at either pH 6, 7, or 8, and at either 4°C, 22 °C, or 37 °C. Samples were taken from the dialysis membranes at *t* = 0, 48, and 168 h, and analysed via LC-MS.

Temp. /°C	pH	PCA	Conversion (%) RNase A 0 h	RNase A 48 h	RNase A 168 h	Myoglobin 0 h	Myoglobin 48 h	Myoglobin 168 h
4	7	**4.1**	**80** *76(s):3(d)*	**78**	**75**	**7**	**6**	**6**
		4.2	**36** *33(s):3(d)*	**49**	**65**	**5**	**17**	**11**
		4.3	**100**	**93** *90(s):3(d)*	**88** *84(s):4(d)*	**30**	**30** *28(s):2(d)*	**24**
		4.4	N/A *(missing data)*			**5**	**7**	**10**
		4.5	**67**	**57**	**37**	**12**	**12**	**12**
		4.6	**10**	**22** *10(s):7(d):5(t)*	**11**	**25**	**30**	**28** *23(s):5(d)*
		4.7	**44** *41(s):2(d)*	**46** *43(s):3(d)*	**41**	**24**	**23**	**21**
		4.8	**24** *14(s):10(d)*	**27** *14(s):13(d)*	**20** *9(s):11(d)*	**31**	**22**	**18**
		4.9	**89** *84(s):5(d)*	**85** *81(s):4(d)*	**85** *81(s):4(d)*	**23**	**23**	**23**
		4.10	**21**	**24**	**21**	**39**	**38**	**40** *39(s):1(d)*
		4.11	N/A *(missing data)*			**15**	**14**	**13**
		4.12	**49**	**45** *41(s):4(d)*	**42** *39(s):4(d)*	**73** *67(s):6(d)*	**74** *71(s):3(d)*	**72**
22	7	**4.1**	**80** *76(s):3(d)*	**75**	**61**	**7**	**6**	**8**

		4.2	**36** *33(s):3(d)*	**53**	**45**	**5**	**14**	**18**
		4.3	**100**	**84**	**60** *57(s):3(d)*	**30**	**26**	**20**
		4.4	N/A *(missing data)*			**5**	**6**	**8**
		4.5	**67**	**30**	**5**	**12**	**10**	**10**
		4.6	**10**	**24** *21(s):3(d)*	**17**	**25**	**26**	**42** *30(s):6(d):6(t)*
		4.7	**44** *41(s):2(d)*	**39**	**33**	**24**	**21**	**22**
		4.8	**24** *14(s):10(d)*	**22** *14(s):8(d)*	**22** *15(s):7(d)*	**31**	**24**	**17**
		4.9	**89** *84(s):5(d)*	**80** *76(s):4(d)*	**67** *65(s):2(d)*	**23**	**20**	**20** *17(s):3(d)*
		4.10	**21**	**18**	**16**	**39**	**36**	**34**
		4.11	N/A *(missing data)*			**15**	**14**	**15**
		4.12	**49**	**42** *38(s):4(d)*	**39**	**73** *67(s):6(d)*	**71**	**73**
		4.1	**80** *76(s):3(d)*	**39**	**21**	**7**	**6**	**12**
		4.2	**36** *33(s):3(d)*	**27**	**5**	**5**	**12**	**17**
		4.3	**100**	**29**	-	**30**	**18**	N/A
		4.4	N/A *(missing data)*			**5**	**5**	**17**
		4.5	**67**	**0**	**0**	**12**	**8**	**12**
		4.6	**10**	**0**	**0**	**25**	**39** *26(s):7(d):5(t)*	**27**
37	7	**4.7**	**44** *41(s):2(d)*	**27**	**17**	**24**	**20**	**21**
		4.8	**24** *14(s):10(d)*	**16** *9(s):8(d)*	**15** *6(s):8(d)*	**31**	**23**	**24**
		4.9	**89** *84(s):5(d)*	**54**	**26**	**23**	**16**	**19** *13(s):6(d)*
		4.10	**21**	**16**	**13**	**39**	**39** *38(s):1(d)*	**35**
		4.11	N/A *(missing data)*			**15**	**13**	**18**
		4.12	**49**	**36**	**27**	**73** *67(s):6(d)*	**68**	**72**
		4.1	**80** *76(s):3(d)*	**68**	**55**	**7**	**6**	**19**
		4.2	**36** *33(s):3(d)*	**45**	**33**	**5**	**16**	**19**
22	6	**4.3**	**100**	**84**	**62** *59(s):3(d)*	**30**	**24**	**19**
		4.4	N/A *(missing data)*			**5**	**6**	**26**
		4.5	**67**	**26**	**6**	**12**	**10**	**19**

		4.6	**10**	**22** *10(s):12(d)*	**20** *16(s):3(d)*	**25**	**33** *23(s):9(d)*	**27**
		4.7	**44** *41(s):2(d)*	**42** *40(s):2(d)*	**32**	**24**	**23**	**24**
		4.8	**24** *14(s):10(d)*	**29** *16(s):12(d)*	**22** *15(s):7(d)*	**31**	**29**	**22**
		4.9	**89** *84(s):5(d)*	**82** *78(s):4(d)*	**67** *63(s):3(d)*	**23**	**16**	**18**
		4.10	**21**	**24**	**23**	**39**	**39**	**36**
		4.11	N/A *(missing data)*			**15**	**10**	**34**
		4.12	**49**	**51** *45(s):3(d):2(t)*	**42** *39(s):2(d)*	**73** *67(s):6(d)*	**77**	**49**
22	8	**4.1**	**80** *76(s):3(d)*	**64**	**54**	**7**	**7**	**19**
		4.2	**36** *33(s):3(d)*	**38**	**34**	**5**	**15**	**17**
		4.3	**100**	**80**	**57** *55(s):3(d)*	**30**	**24**	**20**
		4.4	N/A *(missing data)*			**5**	**5**	**14**
		4.5	**67**	**27**	**6**	**12**	**10**	**10**
		4.6	**10**	**0**	**0**	**25**	**29** *23(s):6(d)*	**23**
		4.7	**44** *41(s):2(d)*	**38**	**33**	**24**	**23**	**23**
		4.8	**24** *14(s):10(d)*	**22** *12(s):10(d)*	**17** *8(s):9(d)*	**31**	**25**	**23**
		4.9	**89** *84(s):5(d)*	**76**	**66**	**23**	**24**	**23** *18(s):4(d)*
		4.10	**21**	**17**	**14**	**39**	**36**	**30**
		4.11	N/A *(missing data)*			**15**	**11**	**21**
		4.12	**49**	**40**	**36**	**73** *67(s):6(d)*	**55**	**42**

Table S4.5. Stability of protein conjugates to pH and temperature over time. Modification of RNase A and myoglobin with compounds **4.1**-**4.12** under conditions outlined in **General Procedure 4A**, and conversions observed over time upon exposure to a range of conditions (pH 6-8, 4-37 °C).

4.5 References

1. Quach, D.; Tang, G.; Anantharajan, J.; Baburajendran, N.; Poulsen, A.; Wee, J. L. K.; Retna, P.; Li, R.; Liu, B.; Tee, D. H.; Kwek, P. Z.; Joy, J. K.; Yang, W.-Q.; Zhang, C.-J.; Foo, K.; Keller, T. H.; Yao, S. Q. *Angew. Chem., Int. Ed.* **60**, 17131–17137 (2021).
2. Chen, P.; Sun, J.; Zhu, C.; Tang, G.; Wang, W.; Xu, M.; Xiang, M.; Zhang, C.-J.; Zhang, Z.-M.; Gao, L.; Yao, S. Q. *Angew. Chem., Int. Ed.* **61**, e202203878 (2022).
3. Bandyopadhyay, A.; Cambray, S.; Gao, J. *Chem. Sci.* **7**, 4589–4593 (2016).
4. Chen, P.; Tang, G.; Zhu, C.; Sun, J.; Wang, X.; Xiang, M.; Huang, H.; Wang, W.; Li, L.; Zhang, Z.-M.; Gao, L.; Yao, S. Q. *J. Am. Chem. Soc.* **145**, 3844–3849 (2023).
5. Chatani, E.; Hayashi, R.; Moriyama, H.; Ueki, T. *Protein Sci.* **11**, 72–81 (2002).
6. MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. *Nat. Chem. Biol.* **11**, 326–331 (2015).
7. Barman, S.; Diehl, K. L.; Anslyn, E. V. *RSC Adv.* **4**, 28893–28900 (2014).
8. Frank, J.; Katritzky, A. R. *J. Chem. Soc., Perkin Trans. 2* 1428–1431 (1976).
9. Oka, N.; Yamada, T.; Sajiki, H.; Akai, S.; Ikawa, T. *Org. Lett.* **24**, 3510–3514 (2022).
10. Ganley, J. M.; Christensen, M.; Lam, Y.; Peng, Z.; Angeles, A. R.; Yeung, C. S. *Org. Lett.* **20**, 5752–5756 (2018).
11. Cox, P. A.; Leach, A. G.; Campbell, A. D.; Lloyd-Jones, G. C. *J. Am. Chem. Soc.* **138**, 9145–9157 (2016).
12. Chmurzyriski, L. *J Heterocycl. Chem* **37**, 71 (2000).
13. Rosen, C. B.; Francis, M. B. *Nat. Chem. Biol.* **13**, 697–705 (2017).
14. Crugeiras, J.; Rios, A.; Riveiros, E.; Richard, J. P. *J. Am. Chem. Soc.* **131**, 15815–15824 (2009).
15. Barber, L. J.; Yates, N. D. J.; Fascione, M. A.; Parkin, A.; Hemsworth, G. R.; Genever, P. G.; Spicer, C. D. *RSC Chem. Biol.* **4**, 56–64 (2023).
16. Spears, R. J.; Brabham, R. L.; Budhadev, D.; Keenan, T.; McKenna, S.; Walton, J.; Brannigan, J. A.; Brzozowski, A. M.; Wilkinson, A. J.; Plevin, M.; Fascione, M. A. *Chem. Sci.* **9**, 5585–5593 (2018).
17. Grabarczyk, D. B.; Chappell, P. E.; Eisel, B.; Johnson, S.; Lea, S. M.; Berks, B. C. *J. Biol. Chem.* **290**, 9209–9221 (2015).
18. Limnios, D.; Kokotos, C. G. *Chem. - Eur. J.* **20**, 559–563 (2013).
19. Landa, A.; Minkkilä, A.; Blay, G.; Jørgensen, K. A. *Chem. - Eur. J.* **12**, 3472–3483 (2006).
20. Roudesly, F.; Veiros, L. F.; Oble, J.; Poli, G. *Org. Lett.* **20**, 2346–2350 (2018).
21. Dodin, G.; Bourliataud, B.; Cordier, C.; Blais, J.-C. *J. Org. Chem.* **61**, 2561–2563 (1996).
22. Pérez-Faginas, P.; Aranda, M. T.; García-López, M. T.; Snoeck, R.; Andrei, G.; Balzarini, J.; González-Muñiz, R. *Bioorg. Med. Chem.* **19**, 1155–1161 (2011).

Chapter 5

Enhancing the potency of Sonic Hedgehog

Chapter 5: Enhancing the potency of Sonic Hedgehog

Note: The experimental work of this chapter was conducted in collaboration with our industrial partners, Qkine. This chapter contains experimental data collected by Dr. Helen Bell (protein expression/purification and solubility studies), Dr. Kerry Price (Gli-luciferase assay) and Dr. Greg Hughes (thermal shift assay), as indicated in relevant text and figures.

5.1 Introduction

Human Sonic Hedgehog (Shh) is an extracellular signalling protein involved in developmental pathways dictating processes such as cell proliferation, cell specialisation, and the patterning of cellular differentiation to direct the formation of organs.[1,2] Shh has many applications in embryonic development and neural stem cell culture: for example, tissue models of the brain are essential to gain an understanding of the key processes involved in conditions such as neurodegenerative diseases and brain cancer. However, the unique and complex nature of neural tissue presents a challenge in the design of suitable model systems. Human induced pluripotent stem cells (hiPSCs) have emerged as useful tools to develop neural models as they can be directed, using growth factors such as Shh, to differentiate into many different neural cell types.[3]

Qkine, our industrial partners, manufacture proteins of high purity and bioactivity for research and bio-manufacturing, in an animal-free laboratory. Shh is one of the recombinantly-produced products expressed in *E. coli* available from Qkine. Whilst *E. coli* is the most widely utilised expression system due to fast growth kinetics and the extensive range of tools and protocols available,[4,5] the lack of the required processing enzymes (for example, acetyl-, methyl-, myristoyl-, and palmitoyltransferases) causes an absence of post-translational modifications often required for protein regulation and cell signalling in eukaryotic systems.[6]

In the biosynthetic pathway of Shh, a pre-protein is synthesised with an N-terminal signal sequence to allow entry into the endoplasmic reticulum. During natural processing of Shh in eukaryotic systems, the N-terminus of the precursor is then exposed at the 24th residue (cysteine (Cys), hereby referred to as C24) due to cleavage of the signal peptide. Autocleavage of the 45 kDa precursor is then facilitated by the enzymatic C-terminal region, to obtain the active 19 kDa N-terminal fraction, and a 26 kDa C-terminal fraction which is degraded. Post-translational lipidation of both termini of the 19 kDa fragment occurs to form the mature, active form of Shh: cholesterol is attached to the C-terminal glycine during autocleavage, and Hedgehog acetyltransferase (Hhat) catalyses the attachment of palmitate to the α-amine of the N-terminal C24 *via* amide linkage (*Fig. 5.1a*).[7,8]

Most post-translational palmitoylations of proteins occur *via* labile thioester linkages to Cys residues (i.e., S-palmitoylation) to allow removal/reattachment of palmitate: Hedgehog proteins are from a small class of N-palmitoylated proteins with linkages of high stability.[9] The mechanism of Hhat-mediated palmitoylation is unknown, and thought to be one of two possible routes: *i)* first *via* nucleophilic attack of the C24 thiol to palmitoyl coenzyme A (CoA) to form an intermediate thioester linkage, followed by an intramolecular S → N shift to transfer the fatty acid onto the α-amine; or *ii) via* direct nucleophilic attack of the α-amine.[10] The dual

lipidation of Shh plays an integral role in Shh signalling and increases the signalling potency: mutant forms of Shh associated with various genetic diseases often involve errors in this post-translational processing.[8,10,11] It is therefore critical that lipidation, required for activity of native human Shh, is mimicked when non-native expression systems are employed.

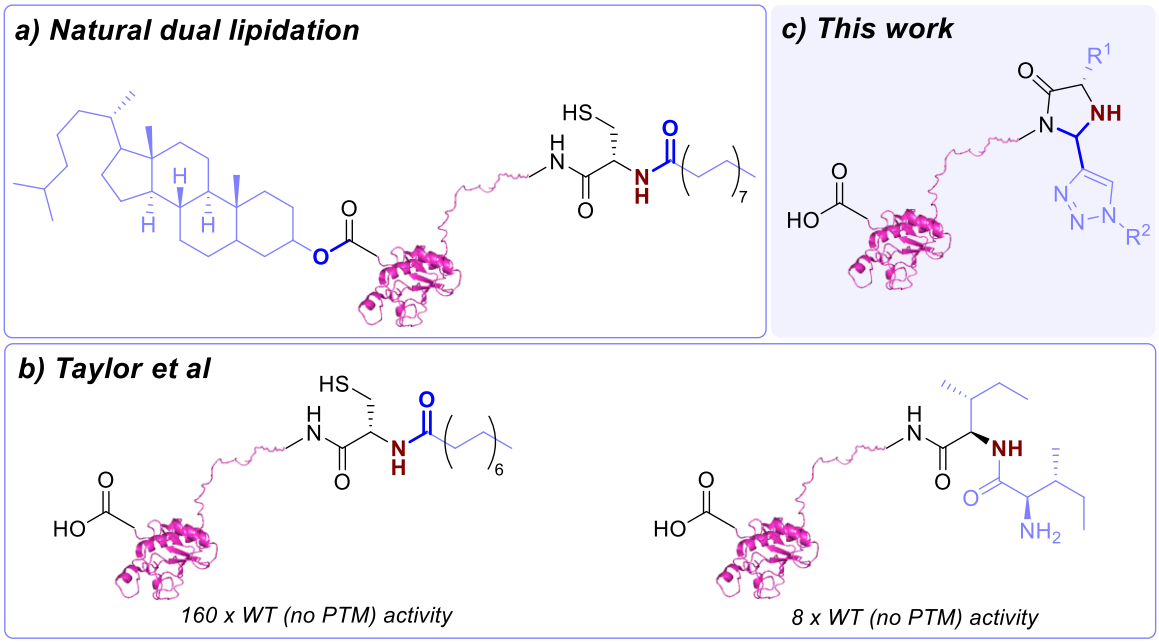


Figure 5.1. ***a)*** Natural dual lipidation of Shh from post-translational modification; ***b)*** Chemical and mutagenic N-terminal modifications employed by Taylor *et al.* (WT = wild-type expressed in *E. coli* with no post-translational modification (PTM));[12] ***c)*** this work: proposed N-terminal modification of Shh *via* TA4C conjugation. Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

In complex with its receptor Patched-1 (Ptch1), the N-terminus of Shh is surrounded by a hydrophobic cavity (illustrated in *Fig. 5.2*).[13] Increasing the N-terminal hydrophobicity *via* palmitoylation of Shh has no influence on the binding affinity,[12] instead modulating activity through cholesterol transport. The first step of Shh signalling is the binding of Shh to the Ptch1 receptor: this binding event relieves suppression of the downstream Smoothened (SMO) receptor, activating a signalling cascade. Ptch1 regulates SMO activity by controlling the availability of cholesterol: a 150 Å hydrophobic tunnel in Ptch1 transports cholesterol away, depleting cholesterol levels and inhibiting SMO activity. Upon Shh-Ptch1 binding, the N-terminal palmitate sits in a hydrophobic pocket (surrounding hydrophobic residues illustrated in *Fig. 5.2*) and, together with the C-terminal cholesterol, blocks the cholesterol transport tunnel. Blockage of the tunnel therefore restores the levels of cholesterol in the membrane, activates SMO and results in a downstream response (*Fig. 5.3*).[10,14,15]

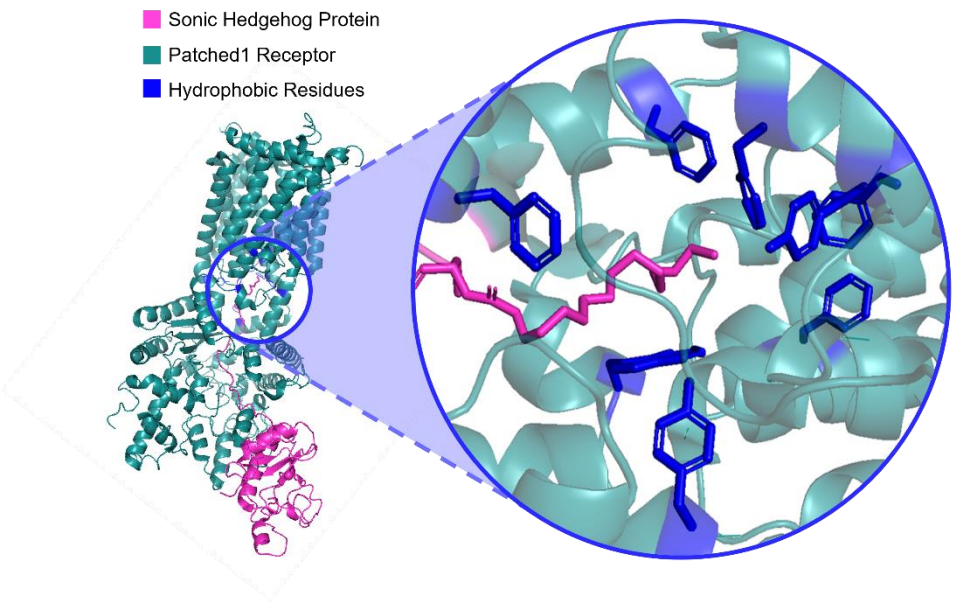


Figure 5.2. Shh-Patched1 binding, with surrounding hydrophobic residues highlighted. Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

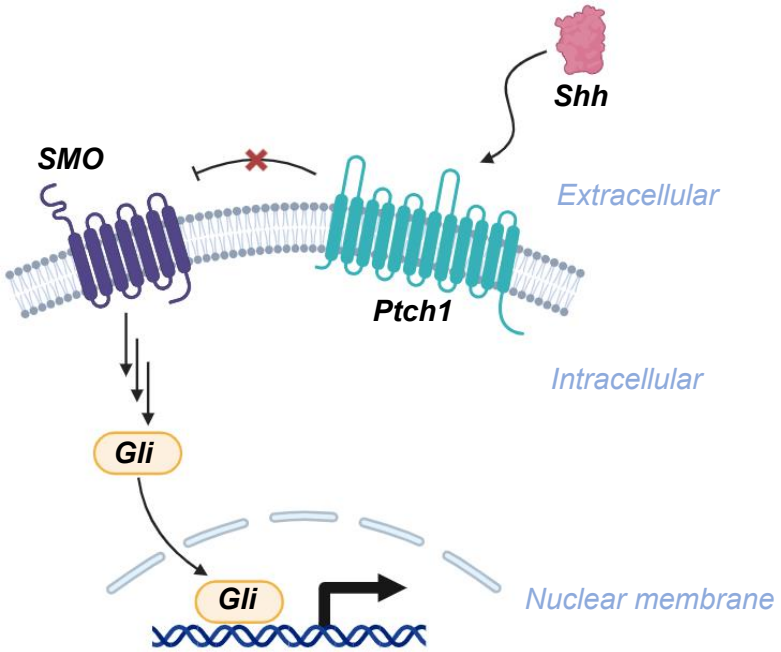


Figure 5.3. Simplified outline of the Shh signalling mechanism: binding of Shh to the Patched-1 (Ptch1) receptor relieves suppression of the downstream Smoothened (SMO) receptor, activating a signalling cascade which results in activation of Gli transcription factors and expression of target genes.[10] Figure created with BioRender.com.

As described earlier, the post-translational dual lipidation modifications in human Shh do not occur when recombinantly expressed in *E. coli*. One approach to solve this problem could be to co-express the recombinant protein with the required enzyme for the post-translational modification; this strategy has been reported for myristoylation and palmitoylation of other proteins in *E. coli*.[16,17] Note that the Shh N-terminal C24 residue is required for Hhat activity, with mutation of Cys to alanine (Ala) or serine (Ser) resulting in

no/partial palmitoylation respectively.[10] Alternatively, N-terminal lipidation can be mimicked using chemical and mutagenic modifications. Taylor *et al.* explored the effect of a range of hydrophobic and hydrophilic *N*-terminal modifications of Shh expressed in *E. coli*. They selectively introduced long-chain fatty acids (C_8-C_{16}) *via* reaction of the α-amine with fatty acid coenzyme A (CoA) esters (*Fig. 5.4a*), and utilised thiol-specific maleimide (*Fig. 5.4b*) and iodoacetyl (*Fig. 5.4c*) strategies to introduce a wide range of hydrophobic moieties, including heteroaromatics, aromatics, and branched hydrocarbon chains. Note that all chemical modifications required reaction of the N-terminal Cys, with fatty acid CoA ester modifications proceeding first *via* nucleophilic attack of the C24 thiol, followed by an intramolecular S → N shift to transfer the fatty acid onto the α-amine.[12]

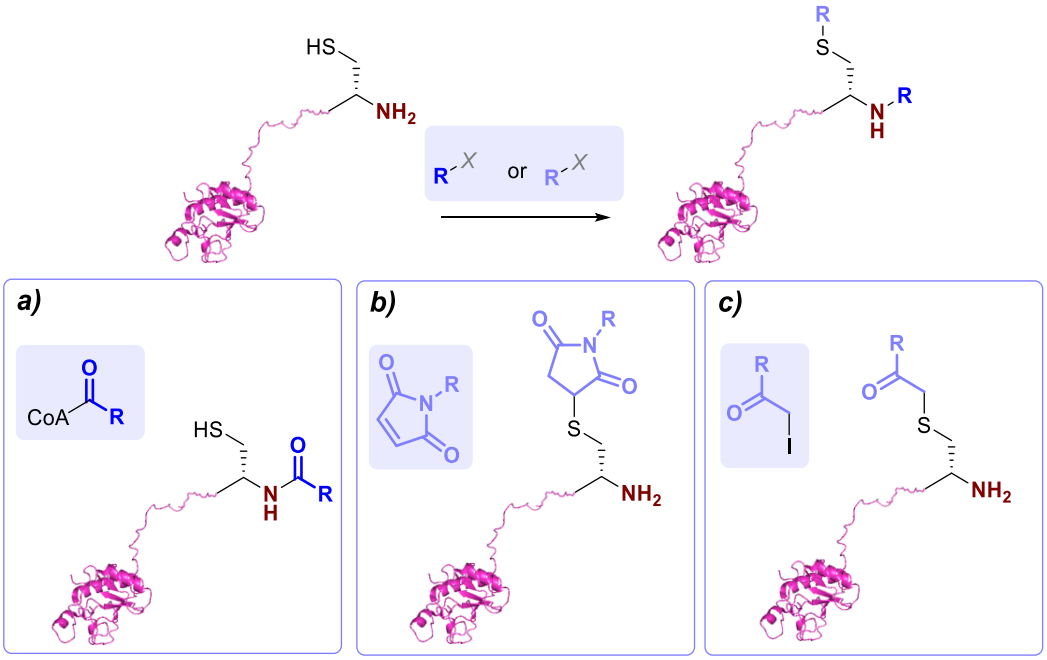


Figure 5.4. Chemical modification strategies employed by Taylor *et al.* to explore the scope of hydrophobic N-terminal modifications of Shh, *via* ***a)*** fatty acid CoA ester, ***b)*** maleimide; or ***c)*** iodoacetyl reactivity.[12] Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

Taylor *et al.* found that the potency of Shh increased with increasing hydrophobicity of the modification, for example with increasing alkyl chain length from C_8-C_{14}. Myristoylation (C_{14}) resulted in a 160-fold increase in potency in comparison to the wild-type protein expressed in *E. coli* with no PTMs, restoring the activity to a level comparable to the dually-lipidated natural eukaryotic protein. They also found that mutation of the N-terminal Cys to two isoleucine (Ile) residues (hydrophobic, branched hydrocarbons) gave an 8-fold increase in potency with respect to the WT (expressed in *E. coli* with no PTMs), however the introduction of any more additional hydrophobic amino acid residues at the N-terminus resulted in a decrease in activity: there are some spatial restrictions on the hydrophobic groups which can be effectively accommodated within the hydrophobic pocket of Ptch1 (*Fig. 5.1b*).[12]

Based on the work by Taylor *et al.*,[12] replacement of the C24 residue with two Ile residues has become a commercial strategy for the expression of bioactive Shh in *E. coli*. For example, a commercially available Shh C24II construct was developed in an *E. coli* expression system at Qkine. As much higher levels of bioactivity were obtained by Taylor *et al.* upon myristoylation than C24II mutation,[12] there is potential scope for further improvement of bioactivity of the product developed at Qkine. In collaboration with Qkine, we therefore aimed to improve the bioactivity of Shh, by mimicking natural palmitoylation *via* hydrophobic N-terminal chemical modification. As outlined above, co-expression of Shh with Hhat or the chemical modification strategies utilised by Taylor *et al.*[12] both rely on the presence of the N-terminal cysteine residue, providing complications with aggregation due to the formation of intermolecular disulfide bridges. We therefore aimed to employ new chemical modification strategies independent of the N-terminal amino acid to create novel conjugates without relying on the presence of an N-terminal cysteine (*Fig. 5.1c*).

5.2 Results and Discussion

5.2.1 Gli-luciferase assay

5.2.1.1 Overview

The bioactivity of Shh conjugates prepared throughout this chapter was evaluated using a Gli-luciferase assay, developed by Dr. Kerry Price at Qkine. The assay was conducted using a stable NIH3T3-derived cell line (murine epithelial cells) containing the Firefly luciferase gene under the control of Gli transcription factors. In the assay, the binding of Shh to Ptch1 relieves suppression of the downstream SMO receptor, activating a signalling cascade which results in activation of Gli transcription factors and expression of target genes, including Gli1, Ptch1, and Firefly luciferase (*Fig. 5.3*).[10] The light emitted upon oxidation of luciferin by the expressed Firefly luciferase can be quantified using a luminometer (*Fig. 5.5*).

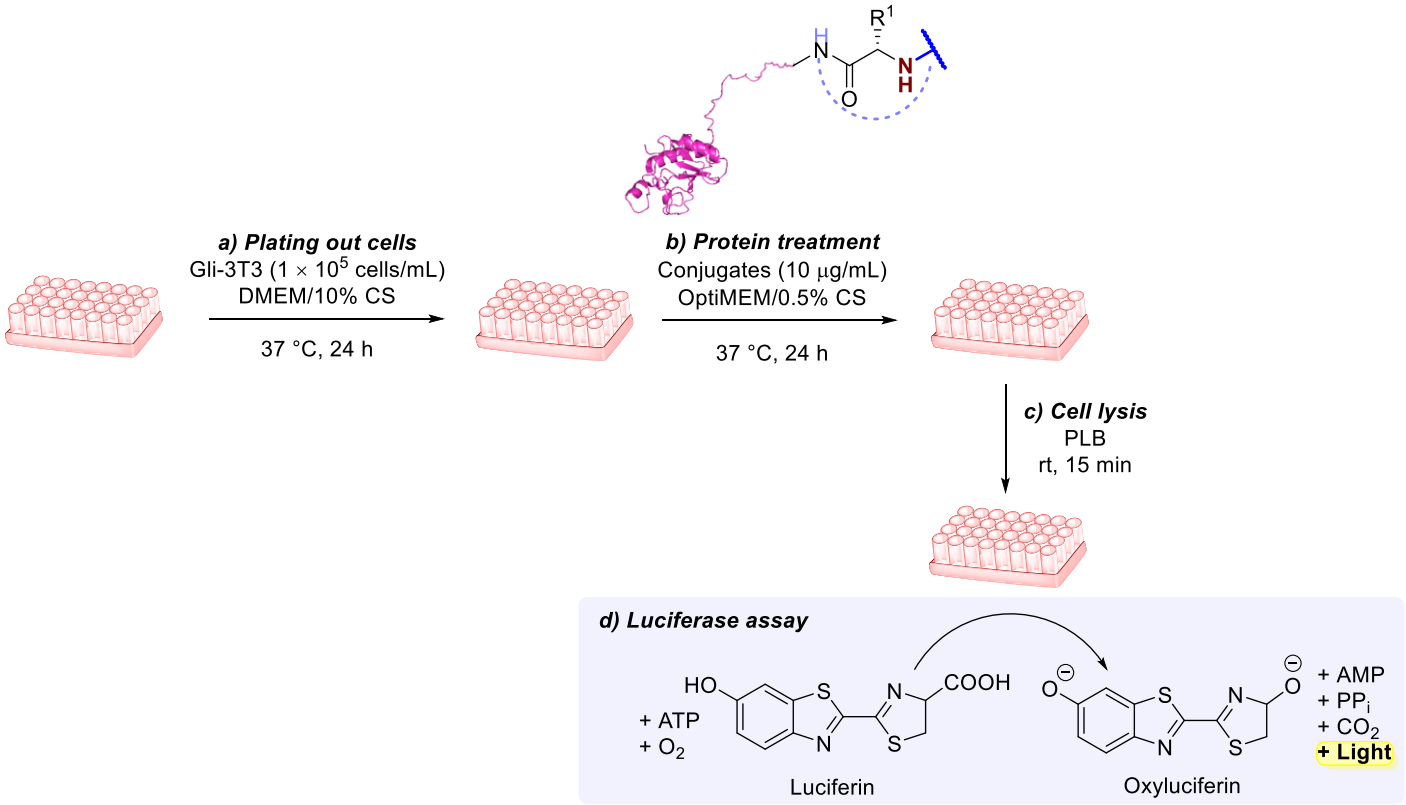


Figure 5.5. Luciferase assay protocol: ***a)*** Plating out Gli-3T3 cells on plates pre-coated with poly-L-lysine at a cell concentration of 1×10^5 cells/mL in DMEM/10% CS (DMEM = Dulbecco's Modified Eagle Medium; CS = calf serum), followed by incubation at 37 °C for 24 h; ***b)*** Protein treatment of the cells with a serial dilution of Shh conjugates, media change to OptiMEM/0.5% CS, followed by incubation at 37 °C for 24 h; ***c)*** Cell lysis using PLB (passive lysis buffer); ***d)*** Luciferase assay, using a luminometer to treat samples with luciferin and measure the luminescence. Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

During data analysis, luciferase assay data are displayed as dose-response curves, where the "dose" is the concentration of Shh conjugate, and the "response" is the luminescence reading. Despite the complex link between Shh binding and the luminescence, with multiple steps occurring between protein treatment of the

assay and the read-out of the response, dose-response curves have the same shape as receptor binding curves. Luciferase assay data can therefore be analysed by Hill fitting (*Equation 5.1*):

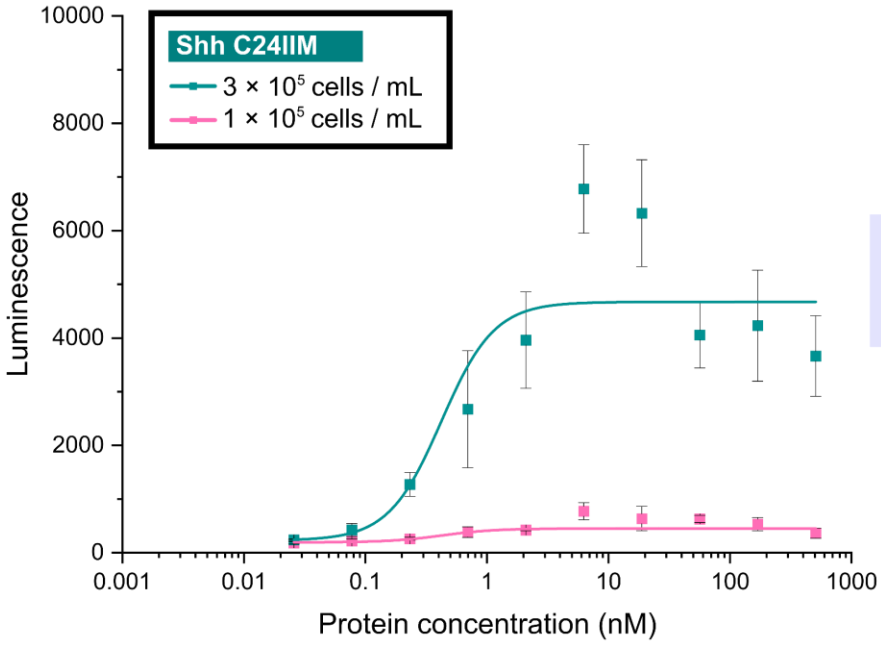
$$y = START + (END - START)\frac{x^n}{k^n + x^n} \qquad \textit{(Equation 5.1)}$$

where $(END - START)$ is the maximum luminescence (accounting for any background levels of luminescence), k is the Michaelis constant, x is the protein concentration, y is the luminescence, and n is the number of cooperative sites. During receptor binding, a single Shh protein engages with two different binding regions of two Ptch1 receptors, both of which are required for efficient signalling, so $n = 2$.[18] Potency was measured by calculation of EC_{50} values, i.e. the dose at which 50% of the maximal response was observed. In all Gli-luciferase assays discussed throughout this chapter, we included a freshly reconstituted Shh sample as a positive control of Shh activity.

5.2.1.2 Assay optimisation

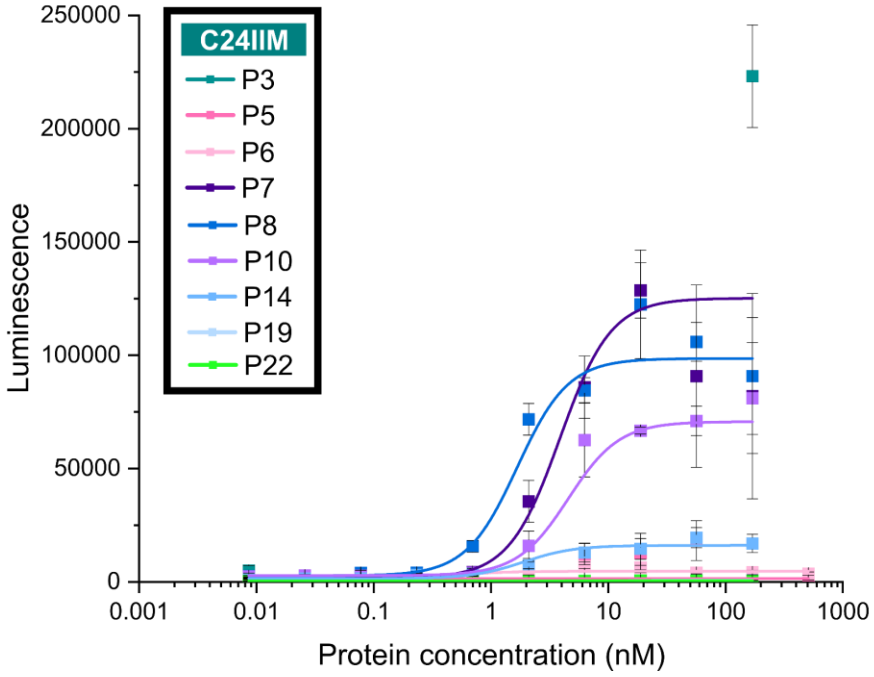
Gli-luciferase assays reported in sections 5.2.2 and 5.2.3 were carried out by Dr. Kerry Price at Qkine; following this, assays detailed in section 5.2.4 were carried out during my iCASE placement and required initial optimisation. We found both cell confluency and passage number to be critical factors, with an increased cell plating concentration (3×10^5 cells/mL, *Fig. 5.6*) and optimal passage window (7-10 passages, *Fig. 5.7*) required for the highest response levels: even within this, there was still irreproducibility between samples.

Luminescence can vary between samples for many reasons, such as cell count. Conventional protocol at Qkine is to co-transfect the experimental reporter with a control reporter encoding for Renilla luciferase. The luminescence of the experimental reporter can be normalised by comparison to the Renilla readings, thus accounting for sample variability. However, due to the use of a pre-transfected NIH3T3-derived cell line, the gene encoding Renilla luciferase was not incorporated: readings were therefore normalised by scaling the highest reading of each replicate to equal one. We found normalisation by scaling to one to be poorly representative, with many Hill fittings failing to converge. We therefore chose to display raw data for Luciferase assays, and all interpretations throughout have been made from analysis of raw data. Assay variability will be discussed further in section 5.3.



Cell conc. / cells/mL	EC_{50} / nM
$3*10^5$	0.4 ± 0.1
$1*10^5$	0.4 ± 0.2

Figure 5.6. Gli-luciferase reporter assay for analysis of Shh bioactivity at varying cell plating concentrations. *Passage number: 6.*



Passage	EC_{50} / nM
3	9×10^{-5} ± 0.7
5	0.1 ± 0.2
6	0.4 ± 0.1
7	3.9 ± 1.1
8	1.7 ± 0.3
10	4.6 ± 1.5
14	1.9 ± 0.8
19	8.6 ± 2.6
22	6.4 ± 8.5

Figure 5.7. Gli-luciferase reporter assay for analysis of Shh bioactivity of Shh C24IIM (plated at 3×10^5 cells/mL) at passage numbers 3-22 *(note: at passage 19, the weak luminescence values were also due to a lower luminescence gain setting of 2700, rather than 3600)*.

5.2.2 Optimisation of modification conditions

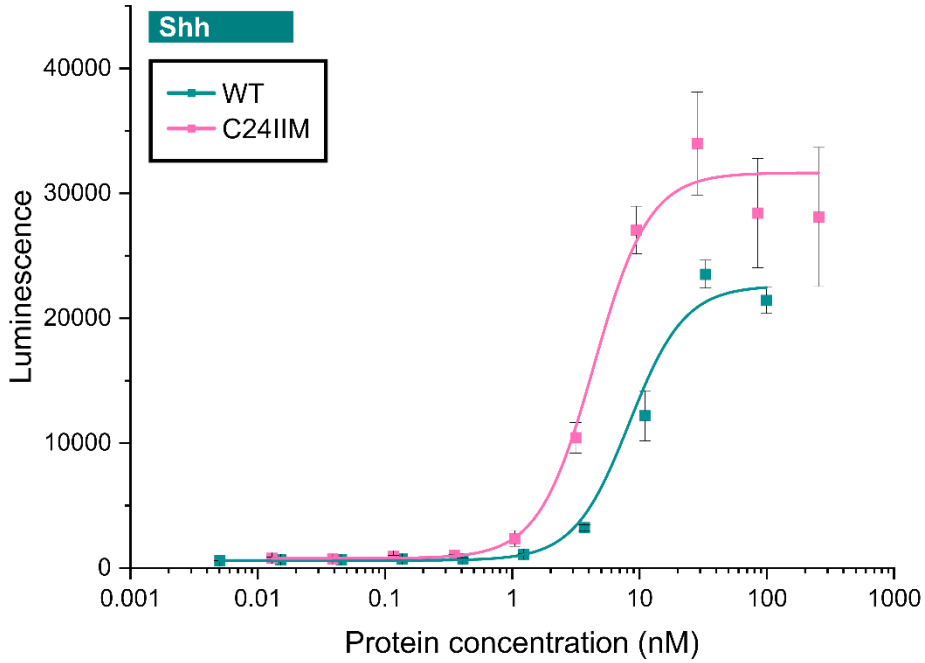
5.2.2.1 Initial reagent screening

Our previous work highlighted the importance of screening a panel of reagents for modification of a target protein, to ensure optimal conjugation is realised: this is due to the protein-dependent nature of N-terminal bioconjugation (Chapter 2).[19] We initially decided to: *i)* compare the bioactivity of Shh WT and the C24II mutant developed at Qkine to validate if the improvements observed by Taylor *et al.*[12] could be reproduced; *ii)* screen a panel of small molecule reagents for Shh modification; and *iii)* carry out a Gli-luciferase assay with the modified samples, to validate if Shh could be chemically modified and if these modifications were tolerated by the cell signalling system.

i) Comparison of Shh WT and Shh C24II

Qkine provided two protein samples: Shh wild-type (WT) and an N-terminal Shh mutant (C24II). Unfortunately, mass spectra obtained upon LC-MS analysis of Shh WT were of poor quality, with many additional unassigned signals (*Fig. S5.1*). We suspected these issues to be due to the N-terminal C24 residue. The Cys residue was not involved in intramolecular disulfide bridging; this left the residue free to form intermolecular disulfide bridges, resulting in dimerisation and visible protein aggregation. Due to poor protein solubility and poor-quality mass spectra obtained in preliminary modification studies, Shh WT was excluded from further studies. Interestingly, LC-MS analysis of the C24II mutant revealed a retained N-terminal methionine (Met) residue (herein referred to as C24IIM, *Fig. S5.2*). As discussed in Chapter 3.1.4, large sidechains (e.g. Ile) can block Met aminopeptidase activity, preventing Met cleavage at the N-terminus.[20] This issue was avoided by Taylor *et al.* through expression and purification of Shh as a six histidine-tagged fusion protein (His-tag), followed by His-tag cleavage with an enterokinase to expose the N-terminal C24 residue.[12] However, His-tags were not used at Qkine due to poor scalability and expense of nickel columns for His-tag purification.

To evaluate the impact of the residual N-terminal Met in combination with penultimate Ile residues on the bioactivity of the product developed at Qkine, a Gli-luciferase reporter assay was carried out by Dr. Kerry Price. The extension of Shh C24II with an N-terminal Met has not been previously explored, although Taylor *et al.* observed a 2-fold increase in potency of C24M in comparison to the unmodified WT (expressed in *E. coli* without PTMs) due to the hydrophobicity of Met, but a reduction in potency of C24II upon addition of further N-terminal Ile residues.[12] Our comparison of the unmodified Shh WT and Shh C24IIM samples *via* Gli-luciferase assay revealed an increase in maximum response levels and a 2-fold increase in potency of Shh C24IIM in comparison to the WT, despite the retained N-terminal Met of the Shh C24IIM mutant (*Fig. 5.8*). Whilst this was a less pronounced increase in potency than the 8-fold increase observed by Taylor *et al.*,[12] we cannot directly compare results due to the use of a different assay system: Taylor *et al.* used an alkaline phosphatase induction assay employed with the C3H10T1/2 cell line.[12]



Shh Protein	EC_{50} / nM
WT	8.3 ± 0.5
C24IIM	4.4 ± 0.4

Figure 5.8. Gli-luciferase reporter assay for analysis of Shh bioactivity of Shh WT and Shh C24IIM *(data collected by Dr. Kerry Price). Passage number: 6.*

ii) Protein modification

We next decided to screen a range of small molecule reagents readily available to us, under conditions normally employed in previous studies (Chapter 2): whilst these initial reagents were not hydrophobic, this preliminary study would provide an indication of the most suitable method for Shh modification. Shh C24IIM was modified with reagents **5.1**-**5.6**, and the conversion was determined prior to purification, and following removal of excess reagent via dialysis at 4 °C (*Fig. 5.9*, *S5.6*, *Table S5.1*). Modification of Shh C24IIM was successful, with approx. 80% modification after dialysis with 2-pyridinecarboxaldehyde (2-PCA) reagents **5.1** and **5.2**. Due to the high conversions with carboxaldehydes, we decided to use 2-PCA/TA4C-based modification strategies moving forwards. Whilst triazolecarbaldehyde (TA4C) **5.3** gave lower conversions than PCAs (41% modification of Shh C24IIM after dialysis), the TA4C reagents were also pursued due to ease of synthesis, as discussed below. Conversions obtained upon modification with oxazoline (Ox) **5.5** and benzaldehyde (BA) **5.6** were poor (<11%). As seen previously, protein signals could not be deconvoluted for 2-ethynylbenzaldehyde (2-EBA) **5.4** due to off-target reactivity leading to a mixture of heterogeneously modified proteins.[19]

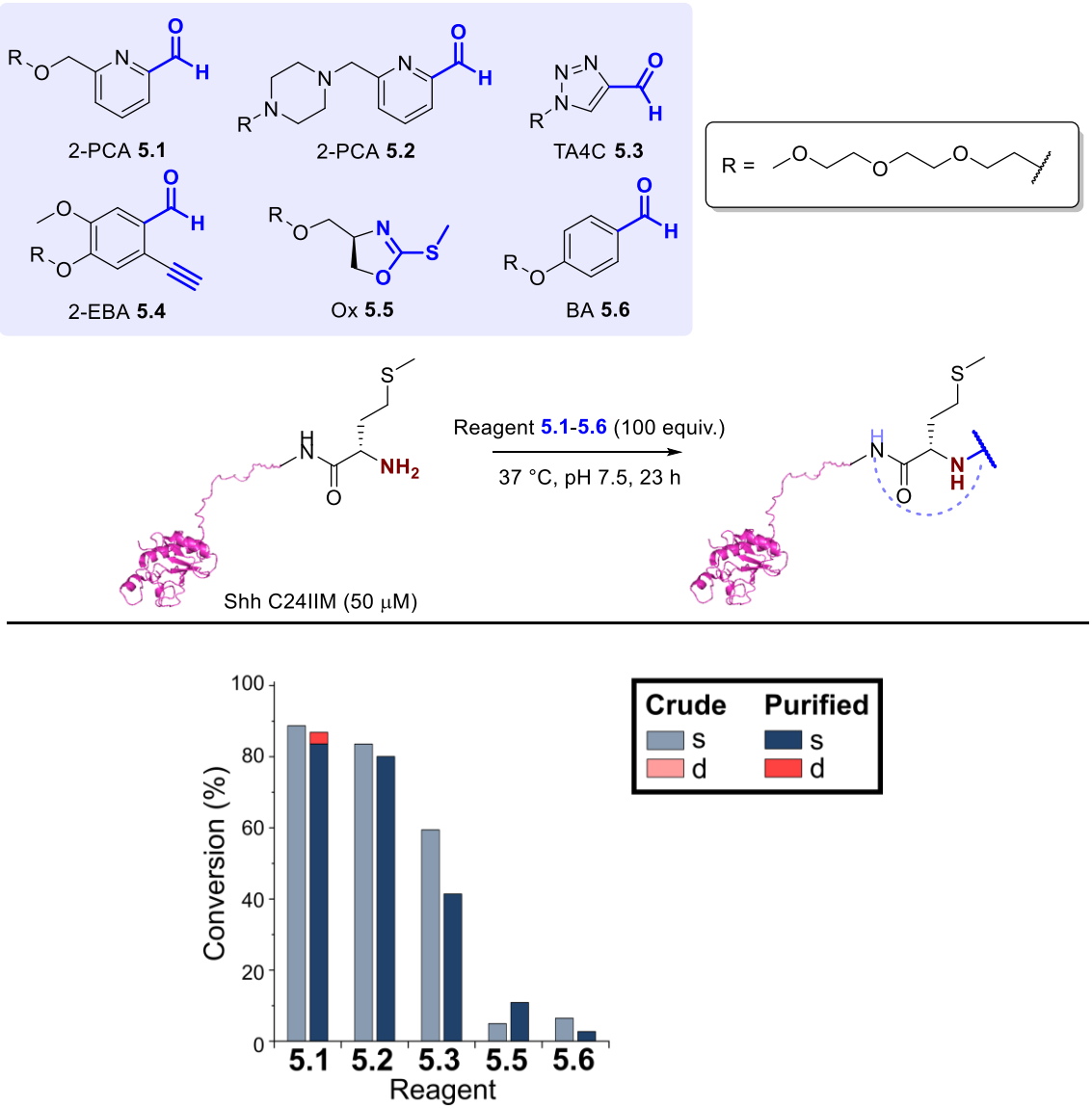


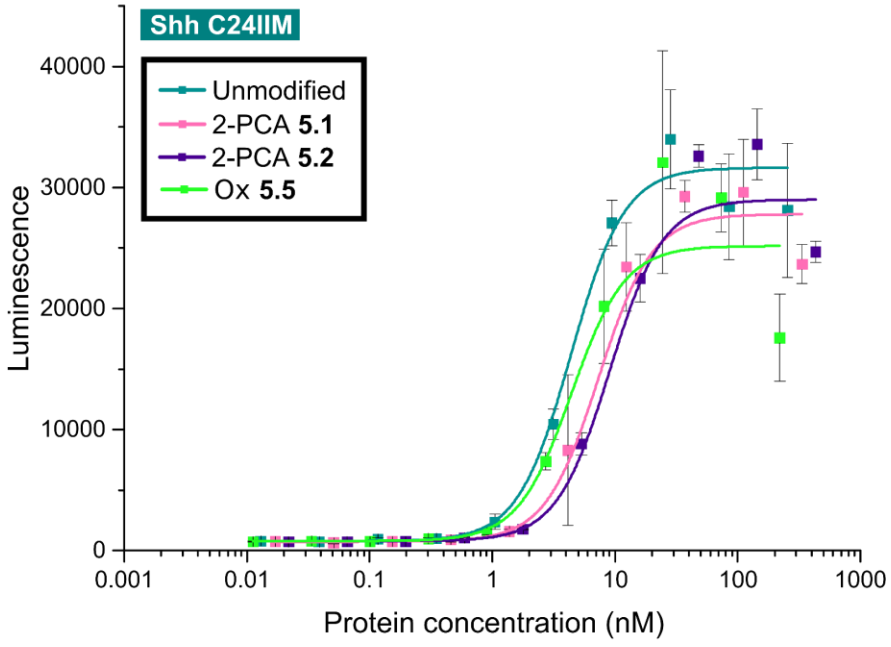
Figure 5.9. Screening reagents for the N-terminal modification of Shh C24IIM. Shh C24IIM (50 μM) was modified with reagents **5.1**-**5.6** (100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 23 h, and the conversion was determined by LC-MS before and after purification by dialysis (s = single, d = double modification). Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

iii) Gli-luciferase assay

To determine if the N-terminal modifications had any effect on biological activity, the modified Shh samples (from *Fig. 5.9*) were sent to Qkine for a Gli-luciferase reporter assay (*Fig. 5.10*). Purified conjugates were used as during preliminary work an excess of aromatic small molecules was found to introduce inaccuracy in the concentration determination of samples and cause interference in the assay. The luciferase assay indicated that the N-terminal modifications were tolerated, albeit with some loss of activity. However, issues with measurement of the protein concentration due to poor solubility may have influenced the results, preventing accurate measurement of activity. Concentrations of purified conjugates measured by UV-vis spectroscopy prior to the protein treatment stage of the assay were variable (22-43 μM, *Table S5.2*), and

mostly lower than expected (50 μM). Whilst this may have been partly due to the influence of chemical modifications on the molar extinction coefficient of the protein, the low concentrations were most likely due to precipitation of Shh during modification. Controls where Shh was treated identically in the absence of modification reagents also showed a reduction in protein concentration.

We suspected that the minimal differences in EC_{50} between modified and unmodified conjugates may have also been due to poor conjugate stability at 37 °C in the presence of competitors; during the protein treatment stage of the Gli-luciferase assay, the conjugates were incubated at 37 °C for 24 h in OptiMEM/0.5% calf serum (CS). Accessible N-termini from other components were present during the protein treatment stage, such as insulin and L-glutamine from OptiMEM media, and antibodies/growth factors from the CS supplementation. Our previous studies revealed the poor stability of RNase A conjugates to incubation at 37 °C, and dissociation *via* competitive cleavage in the presence of DiAla (Chapter 2): we expected competitive cleavage here to be less significant due to the inability of L-glutamine conjugates to cyclise. Whilst a potential drawback, throughout this chapter we have assumed that the conjugate stability was not an issue; see section 5.2.4.3 for further discussion.



Small molecule	EC_{50} / nM
Unmodified	4.4 ± 0.8
2-PCA **5.1**	7.3 ± 1.2
2-PCA **5.2**	9.1 ± 0.5
Ox **5.5**	4.4 ± 0.7

Figure 5.10. Raw data from the Gli-luciferase reporter assay for analysis of Shh bioactivity of C24IIM conjugates modified with reagents **5.1**, **5.2** and **5.5** *(data collected by Dr. Kerry Price). Passage number: 6.*

5.2.2.2 Solubility studies

Having encountered solubility issues during Shh modification, we next sought to explore the effect of the reaction parameters on the solubility of Shh C24IIM, with the aim to optimise reaction conditions that avoid

protein precipitation and increase the accuracy of the luciferase assay. Parameters tested included: *i)* 2-PCA concentration, *ii)* reaction time, and *iii)* protein concentration, buffer composition, and temperature.

i) 2-PCA concentration

We decided to investigate the effect of 2-PCA concentration on protein solubility, as we suspected that the large excess of small molecule may have caused the protein to precipitate (as seen previously in Chapter 3, with 2-PCA/oxime reagents inducing precipitation of myoglobin and RNase A). We also attempted modification at a lower protein concentration (25 µM rather than 50 µM). 2-PCA **5.2** was selected as a suitable modification reagent due to high conversions seen previously in the initial screening. Interestingly reduction in 2-PCA concentration from 5 mM to 1.25 mM, or protein concentration from 50 µM to 25 µM, resulted in very little decrease in conversion and enhanced the selectivity for single modification (*Fig. 5.11*, *Table S5.3*). We did not probe any further reduction in PCA/protein concentration as precipitation was still observed in all cases, including the controls in which no 2-PCA was used, and so we next looked to alter other conditions.

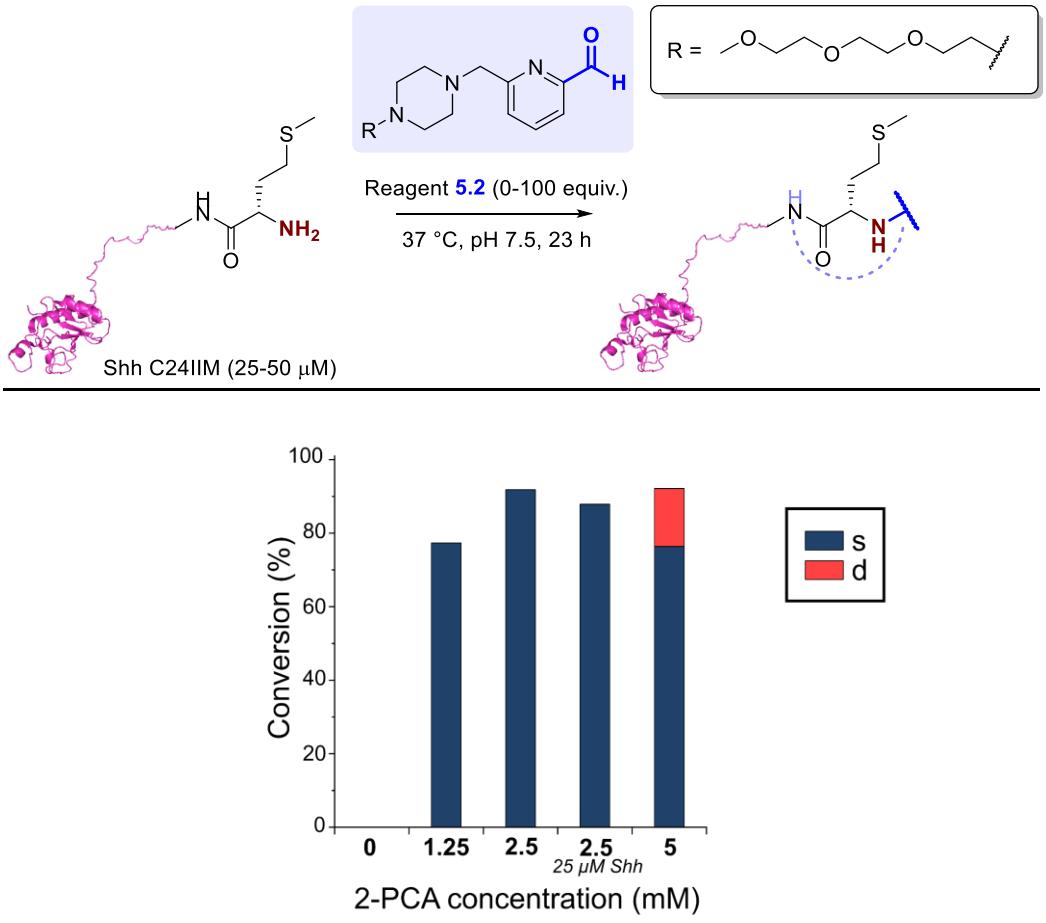
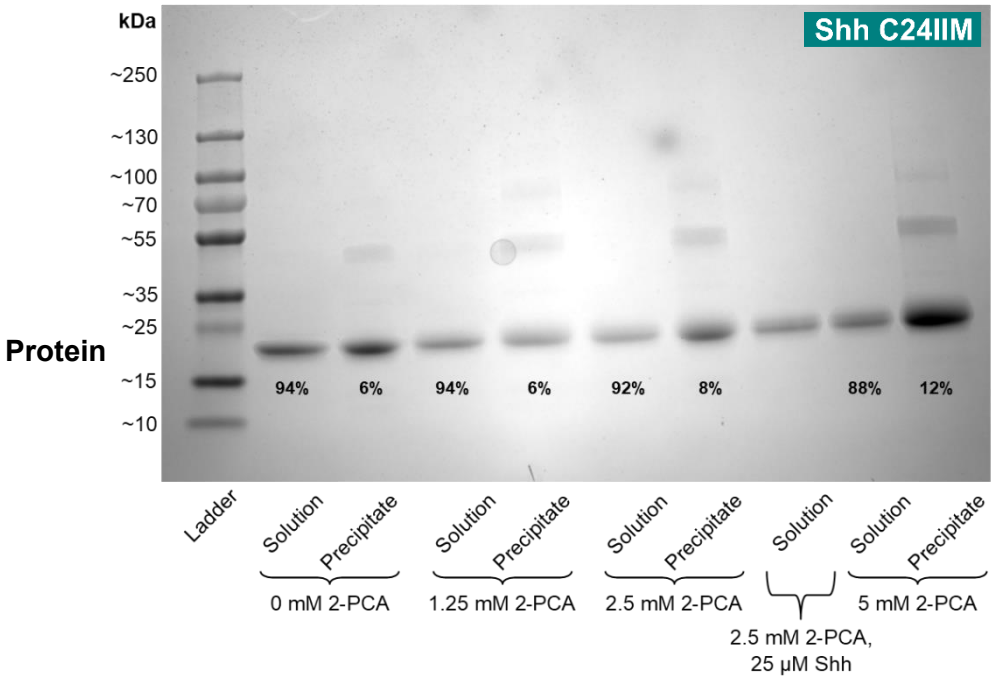


Figure 5.11. Conversions of Shh C24IIM after modification with a range of 2-PCA and protein concentrations. Shh C24IIM (25-50 µM, assume 50 µM protein concentration unless indicated) was modified with 2-PCA **5.2** (0-100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 23 h, and the conversion was determined by LC-MS analysis of the crude reaction mixture (s = single, d = double modification). Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

We ran SDS-PAGE gels of soluble fractions of the reactions and the precipitates formed to validate that protein precipitation occurred, rather than exclusively small molecule precipitation (*Fig. 5.12*). SDS-PAGE revealed that some protein precipitation occurred under all reaction conditions. An analogous experiment was carried out using Shh WT; as expected, based on the quality of mass spectra obtained, the level of protein precipitation was higher for Shh WT (*Fig. 5.12b*) than Shh C24IIM (*Fig. 5.12a*). For example, at 0 mM 2-PCA, 6% and 17% precipitation occurred for Shh C24IIM and WT respectively. For Shh C24IIM, precipitation increased with increasing 2-PCA concentration (from 6-12% precipitation at 0-5 mM 2-PCA), confirming that protein solubility was influenced by the 2-PCA concentration. However, protein precipitation in the absence of 2-PCA reagent indicated that the set of reaction parameters (50 µM protein, 23 h at 37 °C in 50 mM Na phosphate buffer, pH 7.5) used were unsuitable for this protein and required optimisation; the influence of changing 2-PCA concentration alone was insufficient to avoid protein precipitation.

a)



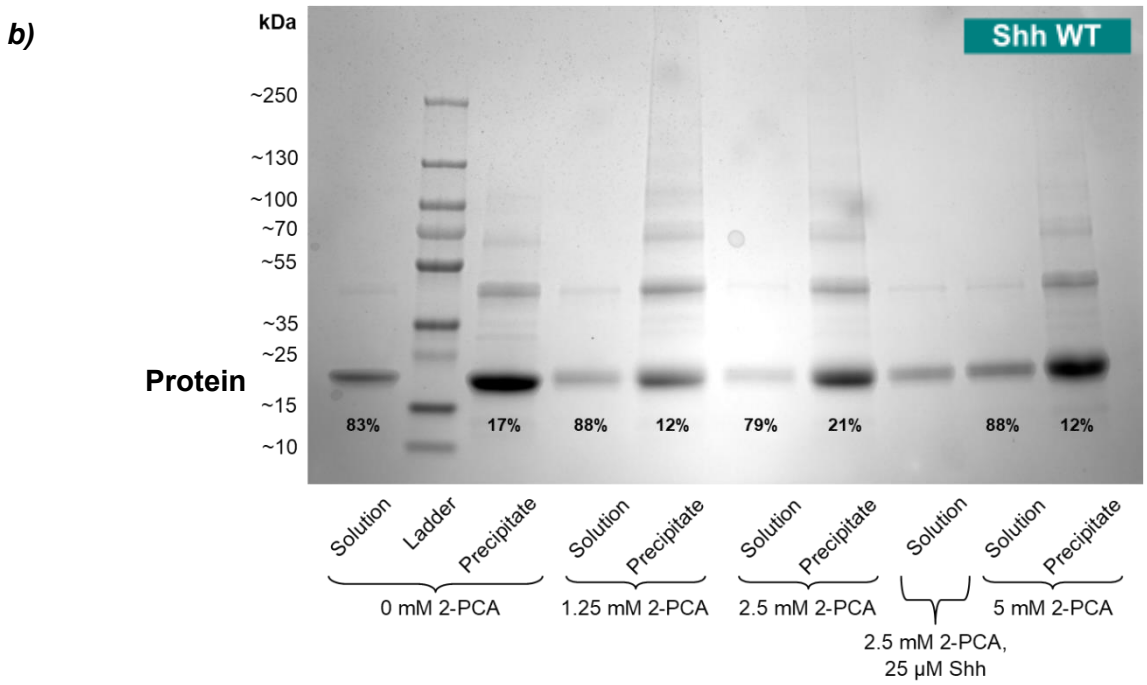


Figure 5.12. SDS-PAGE of reaction solutions and precipitates formed upon modification of ***a)*** Shh C24IIM; and ***b)*** Shh WT under a range of 2-PCA **5.2** and protein concentrations (assume 50 µM protein concentration unless indicated), in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 23 h. The precipitate bands were from resuspension of the total precipitate formed per reaction, whereas soluble bands were taken from a 5.11 µL aliquot of a 100 µL reaction mixture. The percentage of precipitation was calculated by comparison of the image density of the Shh monomer band of soluble/precipitate samples per 2-PCA concentration, measured using ImageJ software. Due to differences in the size of aliquots added for soluble/precipitate samples, image densities were extrapolated for soluble bands to the densities expected for the total reaction mixture volume (100 µL). *Note: The precipitate formed for the 2.5 mM 2-PCA **5.2**/25 µM protein sample was not included in these gels but was assumed to be protein precipitation too.*

ii) Reaction time

We next decided to carry out a kinetic study by following the modification of Shh C24IIM with 2-PCA **5.2** over time at 37 °C by LC-MS analysis. This allowed us to determine if modification was possible over a shorter timeframe than the 23 h used previously, in which we hoped the protein would stay in solution. Promisingly, 78% modification was observed over 5 h with no visible precipitation, indicating that high conversions could be achieved even with reduction of reaction time. Whilst conversion ultimately reached 100% after 29 h, double modification also occurred in addition to visible precipitation, indicating that longer reaction times are unsuitable (*Fig. 5.13*, *Table S5.4*).

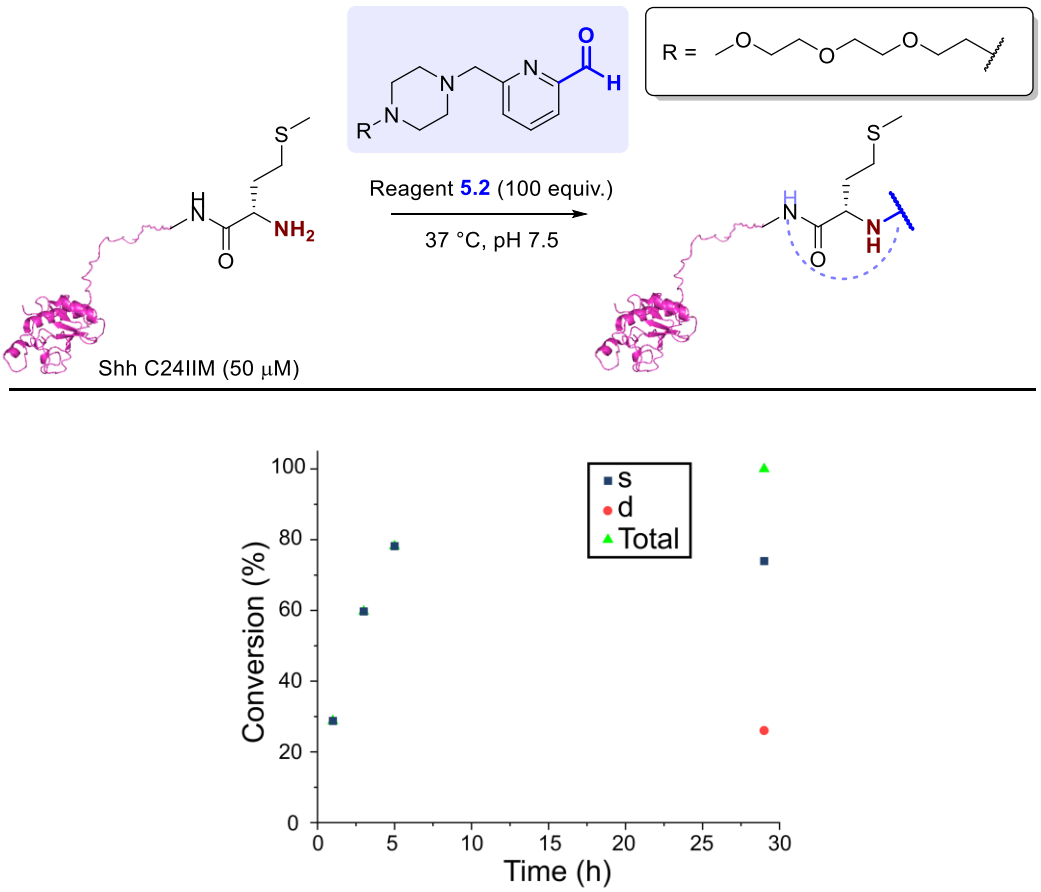


Figure 5.13. Conversion of Shh C24IIM over time upon modification with 2-PCA **5.2**. Shh C24IIM (50 μM) was modified with 2-PCA **5.2** (100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C, and the conversion was determined over time by LC-MS analysis of the crude reaction mixture (s = single, d = double modification). Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

Having achieved high conversion (78%) of Shh C24IIM with 2-PCA **5.2** after 5 h at 37 °C, we next screened a panel of 2-PCA-based reagents (5 mM) to determine if high conversions could be achieved using alternative reagents at reduced reaction times too (*Fig. 5.14*, *Table S5.5*); the design and synthesis of reagents **5.7** and **5.8** will be discussed later in Chapter 5.2.3.1. Whilst fair conversions of purified conjugates (>30%) were achieved for reagents **5.1**-**5.3** and **5.8**, no modification was observed with 2-PCA **5.7** due to poor reagent solubility, despite the use of 10% DMSO co-solvent in attempt to keep the reagent in solution.

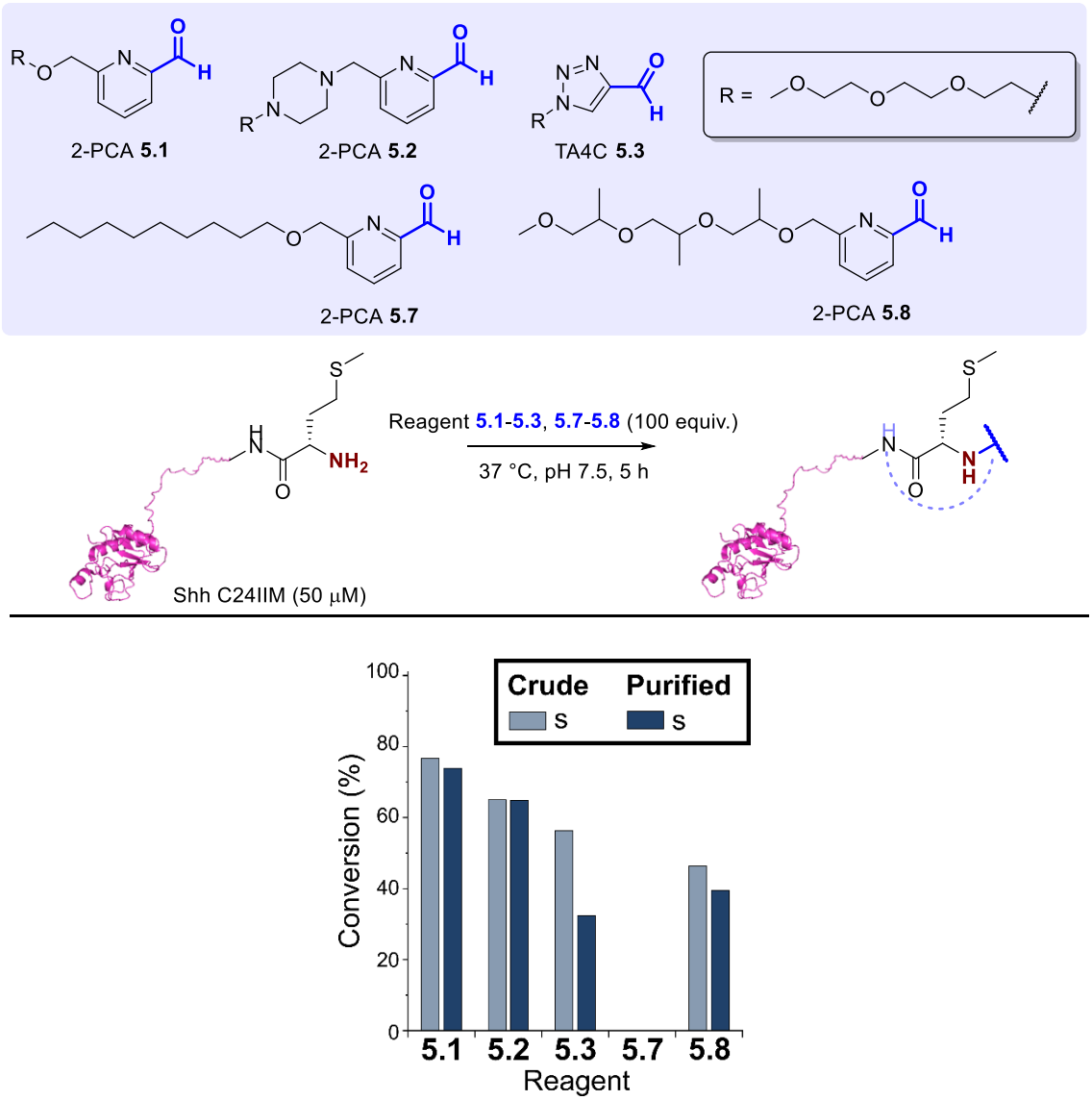


Figure 5.14. Screening reagents for the N-terminal modification of Shh C24IIM at 37 °C for 5 h. Shh C24IIM (50 μM) was modified with reagents **5.1**-**5.3** and **5.7**-**5.8** (100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 5 h, and the conversion was determined by LC-MS before and after purification by dialysis (s = single modification). Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

Unfortunately, despite no visibly observable precipitation after 5 h in the previous experiment, SDS-PAGE indicated 4-19% precipitation of protein in each sample (*Fig. 5.15*). Poor stability of Shh at 37 °C was confirmed by solubility experiments performed by Dr. Helen Bell at Qkine, in which the concentration of a Shh C24IIM solution was followed over time by UV-vis spectroscopy upon incubation at 37 °C (*Fig. 5.16*). These data indicated that whilst high levels of protein modification could be achieved upon reduction of reaction time to 5 h, Shh stability remained an issue under these conditions.

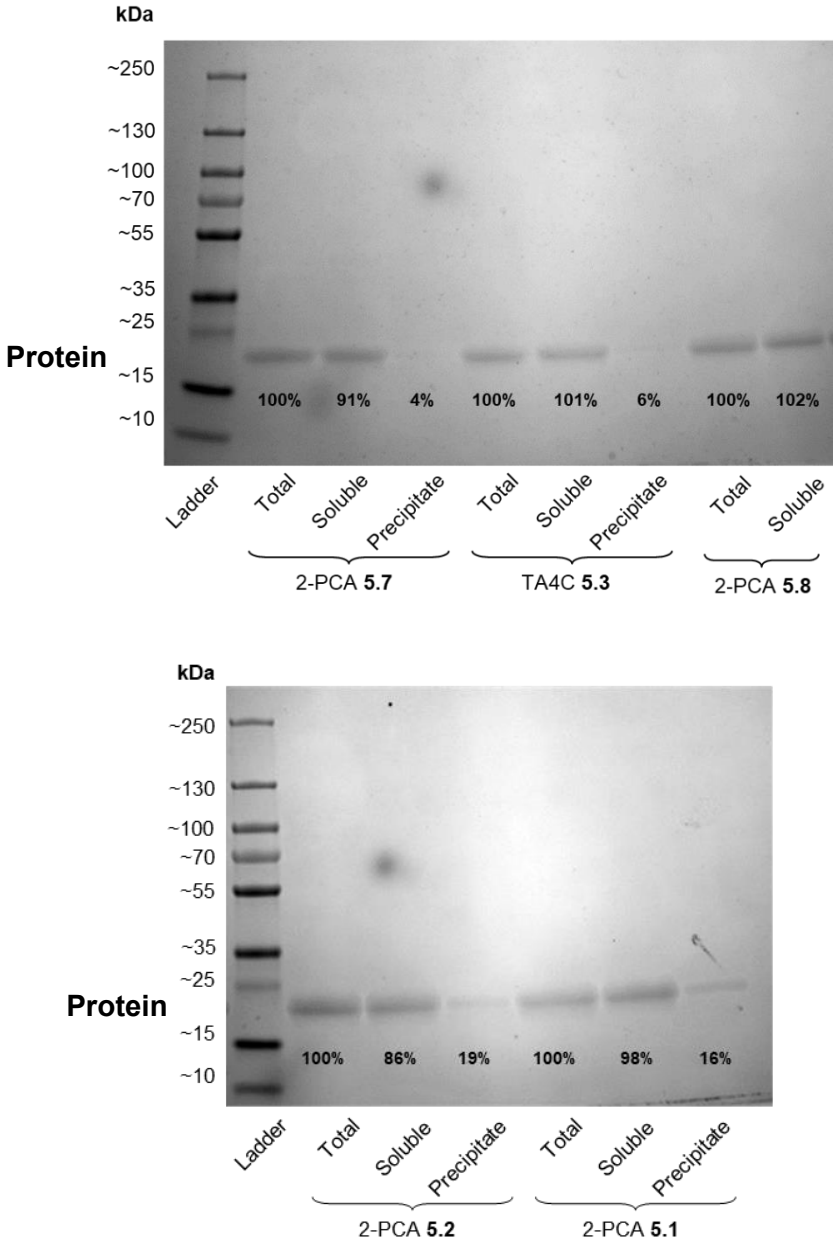


Figure 5.15. SDS-PAGE of reaction solutions and precipitates formed upon modification of Shh C24IIM (50 µM) with reagents **5.1**-**5.3** and **5.7**-**5.8** (100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 5 h. Before purification (after 5 h incubation), a 5.05 µL sample for SDS-PAGE was taken (total). A 5.05 µL sample was taken and centrifuged at 13000 rpm for 5 min; the supernatant was decanted (soluble) and the pellet was reconstituted in 5.05 µL PBS (precipitate). Total, soluble, and precipitate samples were analysed by 15% SDS-PAGE gels. The level of precipitation was calculated by comparison of soluble/insoluble protein bands as a percentage of the total protein band for each modification reagent.

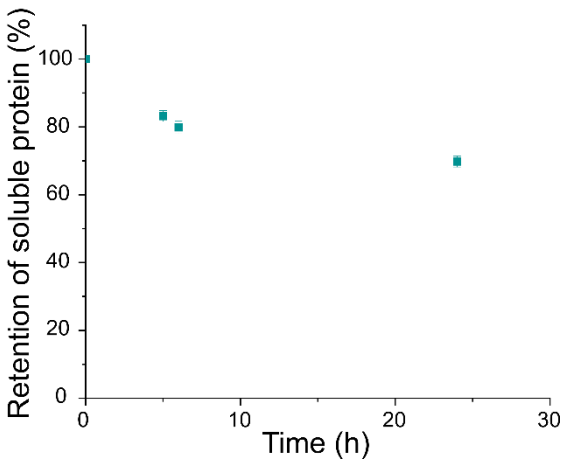


Figure 5.16. Retention of soluble Shh C24IIM over time at 37 °C *(data collected by Dr. Helen Bell)*.

iii) Protein concentration, buffer composition, and temperature

To optimise protein modification reaction conditions, Dr. Helen Bell at Qkine carried out a stability test in which the protein concentration of Shh C24IIM was measured by UV-vis spectroscopy 24 h after incubation under a range of conditions (*Fig. 5.17*). The stability test indicated that Shh was not stable at 37 °C but was stable when incubated at 25 °C for 24 h, with 99% retention of soluble protein. The reaction mixtures were analysed by SDS-PAGE after 24 h incubation, by Dr. Helen Bell (*Fig. 5.18*). Little difference was observed between sets of conditions, although the lowest level of precipitation was observed for the sample incubated at 25 °C, in agreement with the absorbance measurements. Two bands were observed due to dimerisation: however, due to the absence of an N-terminal cysteine residue in the C24IIM variant, the origin of the dimerisation is unknown. Further stability experiments were next carried out by Dr. Helen Bell and indicated <20% protein loss upon incubation at 25 °C for 1 week (*Fig. 5.19*), allowing the reaction times to be extended to over 24 h for any modification reagents with poor reactivity at reduced temperature.

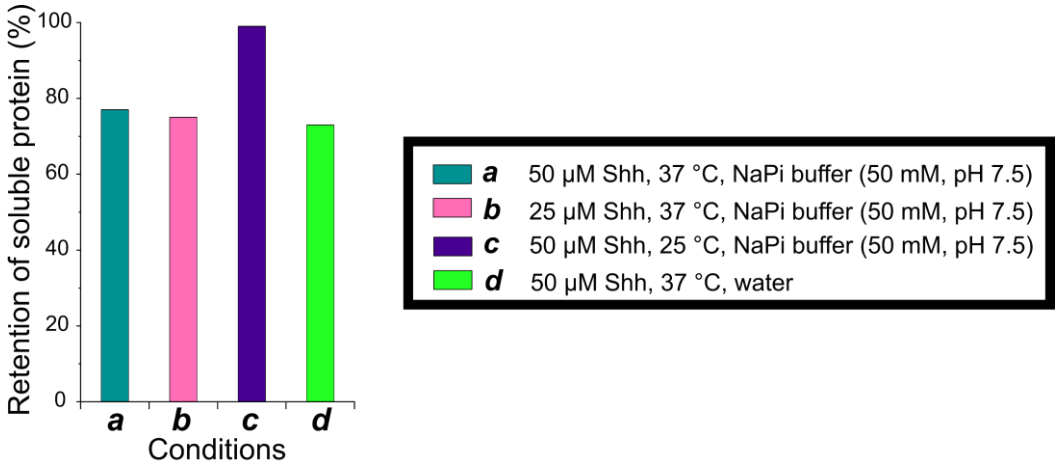


Figure 5.17. Retention of soluble Shh C24IIM after 24 h treatment under a range of conditions *(data collected by Dr. Helen Bell)*.

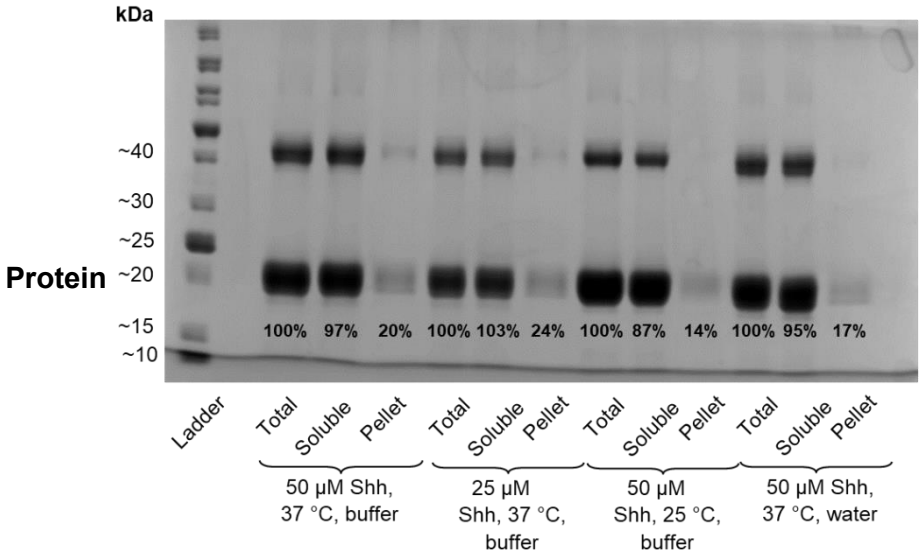


Figure 5.18. SDS-PAGE of stability samples of Shh C24IIM treated under a range of conditions *(data collected by Dr. Helen Bell)*. Before purification (after 24 h incubation), a 5 µL sample for SDS-PAGE was taken (total). A 5 µL sample was taken and centrifuged at 13000 rpm for 5 min; the supernatant was decanted (soluble) and the pellet was reconstituted in 5 µL PBS (precipitate). Total, soluble, and precipitate samples were analysed by 15% SDS-PAGE gels. The level of precipitation was calculated by comparison of soluble/insoluble protein bands as a percentage of the total protein band for each modification reagent.

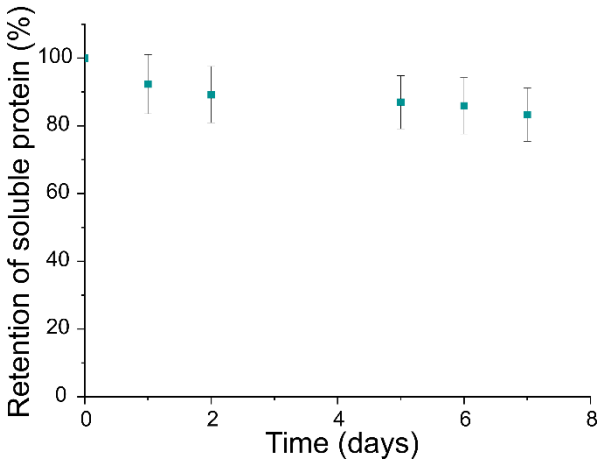


Figure 5.19. Retention of soluble Shh C24IIM over time at 25 °C *(data collected in triplicate by Dr. Helen Bell)*.

We next decided to follow the progress of modifications with 2-PCA **5.2**, TA4C **5.3** and 2-PCA **5.8** at 25 °C to determine if high conversions could be obtained at 25 °C, and to verify that the presence of modification reagents would not cause protein loss by precipitation. High conversions (up to 79%) were observed after 21 h (*Fig. 5.20*, *Table S5.6*), at which point the conversions plateaued: reaction times exceeding 21 h were not necessary. No visible precipitation was observed for 2-PCA **5.2** or TA4C **5.3**, although some white precipitate was formed with 2-PCA **5.8**. SDS-PAGE indicated <5% precipitation; the observed precipitation may have been 2-PCA **5.8** itself (*Fig. 5.21*), with slightly lower conversions observed than for 2-PCA **5.2** due to

increased reagent hydrophobicity. Due to the low levels of Shh precipitation, we decided to proceed with optimised reaction conditions at 25 °C, as both high conversions and high protein solubility were observed.

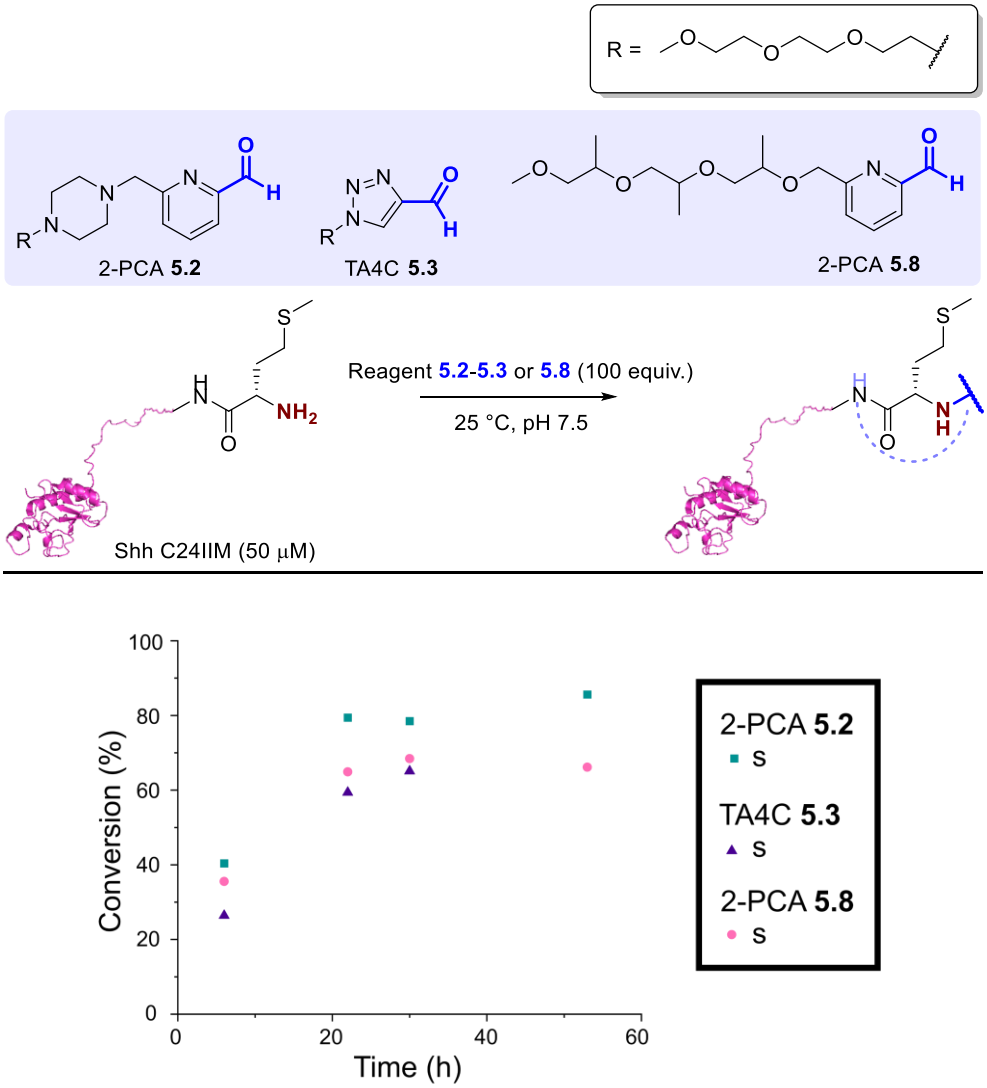


Figure 5.20. Conversion of Shh C24IIM over time upon modification with reagents **5.2**, **5.3** and **5.8** at 25 °C. Shh C24IIM (50 μM) was modified with reagents **5.2**, **5.3** and **5.8** (100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 25 °C, and the conversion was determined over time by LC-MS analysis of the crude reaction mixture (s = single modification). Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

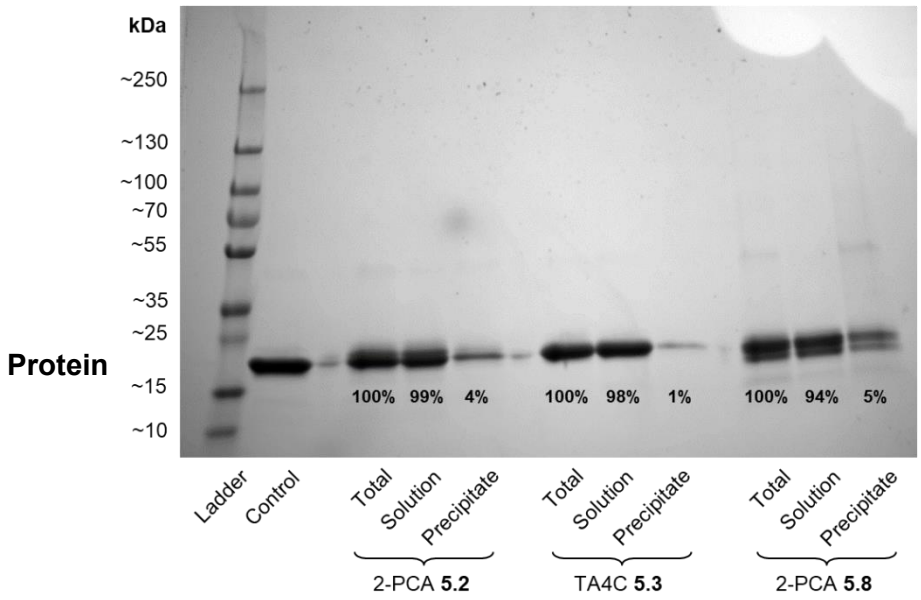
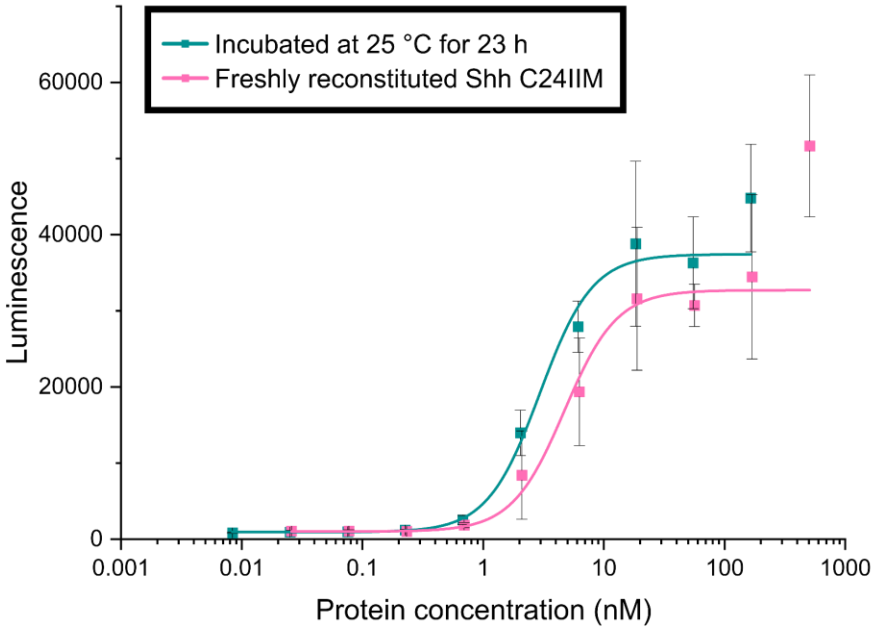


Figure 5.21. SDS-PAGE of reaction solutions and precipitates formed upon modification of Shh C24IIM with reagents **5.2**, **5.3** and **5.8** at 25 °C. Before purification (after 53 h incubation), a 6.73 µL sample for SDS-PAGE was taken (total) from the remaining 70 µL of reaction mixture. The remaining reaction mixture was centrifuged at 13000 rpm for 5 min; the supernatant was decanted and a 6.73 µL aliquot taken (soluble), and the pellet was reconstituted in 6.73 µL PBS (precipitate). Total, soluble, and precipitate samples were analysed by 15% SDS-PAGE gels. The percentage of soluble/precipitate was calculated by comparison of the image density of the Shh monomer band of soluble/precipitate samples to the total sample, measured using ImageJ software. Due to differences in the size of aliquots added for soluble/total samples in comparison to precipitate samples, image densities were extrapolated for precipitate bands to the densities expected for a 6.73 µL sample. *Note: this gel was of poor quality and used just for an indication of solubility: total/soluble bands had higher image densities than the control sample due to overloading of the gel causing protein to spill into neighbouring wells, and unequal band widths due to blanks not being loaded into the empty wells caused error in image analysis.*

A Gli-luciferase assay was carried out by Dr. Kerry Price at Qkine to validate the bioactivity of Shh C24IIM was retained after 23 h incubation at our proposed reaction conditions at 25 °C, followed by dialysis at 4 °C. Very similar EC_{50}s were calculated for both the incubated/dialysed sample and a freshly reconstituted C24IIM solution, indicating that the conditions for protein modification and conjugate purification had minimal effect on bioactivity (*Fig. 5.22*).



Treatment	EC_{50} / nM
Incubated at 25 °C	3.0 ± 0.6
Freshly reconstituted	4.7 ± 1.6

Figure 5.22. Gli-luciferase reporter assay for analysis of Shh bioactivity of Shh C24IIM incubated at 25 °C for 23 h in comparison to freshly reconstituted protein *(data collected by Dr. Kerry Price). Passage number: 3.*

5.2.3 Selection of a suitable modification protocol

With suitable conditions in hand, we next turned our attention to the design of hydrophobic moieties to attach to the Shh N-terminus. Possible strategies for reagent design included: *i)* a molecular docking approach to inform reagent design; or *ii)* high-throughput screening of a large range of different moieties to probe the structure-activity relationship. Docking would allow examination of the structure of the hydrophobic pocket, allowing us to calculate the chain length that would be tolerated, and areas where hydrophilicity could be introduced to improve solubility. Molecular docking requires high resolution structural information obtained using techniques such as x-ray crystallography, NMR, or cryo-electron microscopy.[21] However, due to the low resolution of Shh structures available in the PDB (e.g., 3.80 Å from electron microscopy),[13] molecular docking was not a viable solution. We therefore decided to proceed with the latter approach of library screening rather than targeted reagent design: ease of reagent synthesis was therefore an important factor to consider, to accelerate the study progress.

5.2.3.1 2-PCA reagents

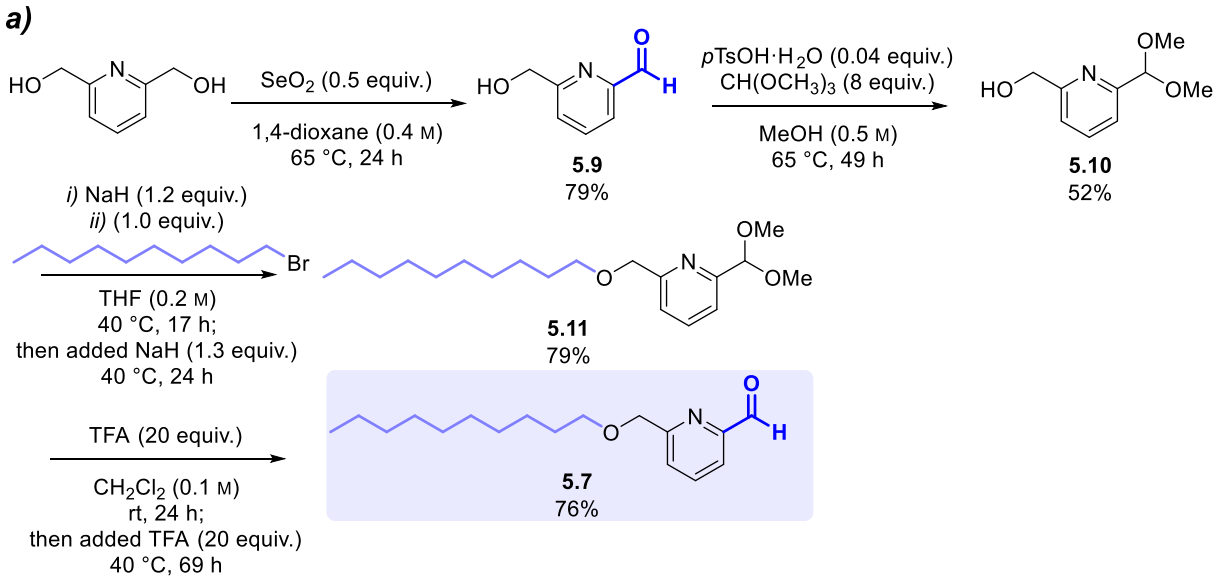
Having established 2-PCAs as effective reagents for Shh modification in preliminary studies, we first decided to explore the effects of chain structure *via*: *i)* the synthesis of alternative 2-PCA analogues, *ii)* comparison of the conversion and selectivity of the analogues upon protein modification, and *iii)* evaluation of bioactivity.

i) Synthesis of 2-PCAs

We first synthesised 2-PCA analogues **5.7** and **5.8** (*Fig. 5.23*). 2-PCA **5.7** was prepared with a hydrophobic alkyl chain (C_{10}): whilst we hoped to mimic palmitoylation (C_{16}) through use of a longer alkyl chain, we suspected long chain lengths would not be tolerated due to poor water-solubility. We hoped the branched structure of 2-PCA **5.8** would increase hydrophobicity (compared to 2-PCA **5.1**) by effectively providing additional hydrophobic contacts to balance the hydrophilic ether groups.

To synthesise 2-PCA **5.7**, we first used selenium dioxide to convert 2,6-pyridinedimethanol to mono-oxidised species **5.9** (79%) following the procedure reported by MacDonald *et al.* (*Fig. 5.23a*).[22] We next decided to protect the aldehyde as an acetal (52%), due to the incompatibility of the aldehyde group with *O*-alkylation conditions observed previously in the preparation of 2-PCA **2.1** (Chapter 2). *O*-alkylation under basic conditions afforded compound **5.11** (79%), with an extra portion of NaH required for full reactivity. The acetal was then deprotected to afford 2-PCA **5.7** (76%), with an extra portion of TFA and elevation of the temperature to 40 °C required for full reactivity.

To synthesise 2-PCA **5.8**, we first tosylated the branched chain (75%) to install a suitable leaving group for the following *O*-alkylation of acetal **5.10** (42%), again with an extra portion of NaH and an extended reaction time required for full reactivity. We then applied the deprotection conditions developed above to afford 2-PCA **5.8** (25%).



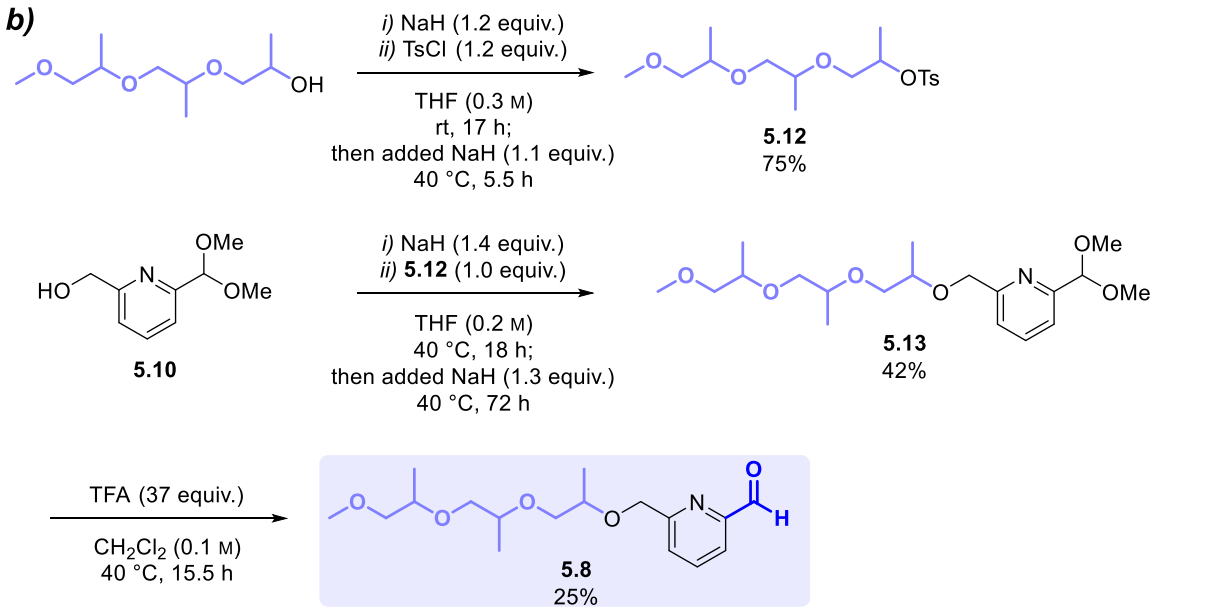


Figure 5.23. Synthesis of: ***a)*** 2-PCA **5.7**; and ***b)*** 2-PCA **5.8**.

ii) Protein modification

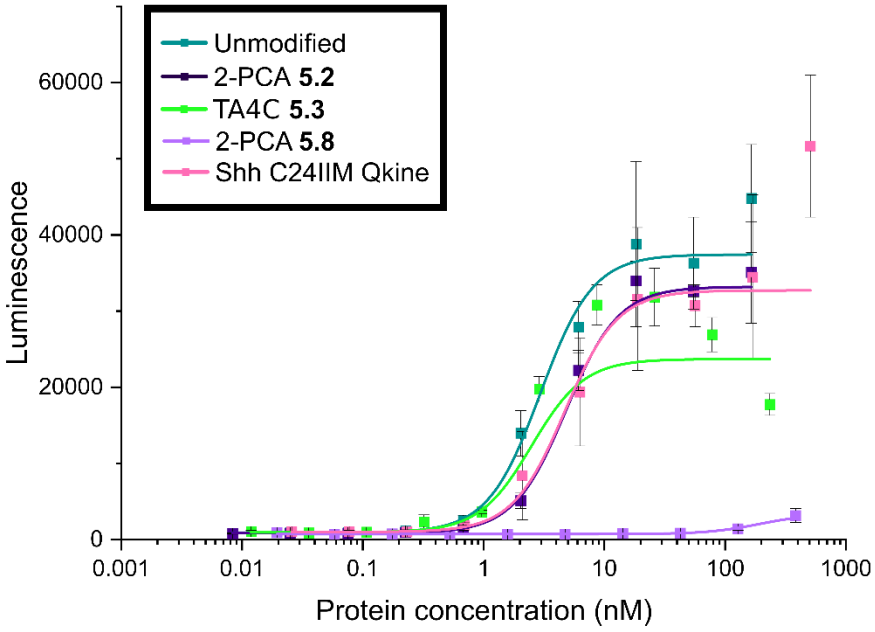
In previous solubility studies, we carried out the modification of Shh C24IIM with 2-PCAs **5.7** and **5.8** (*Fig. 5.14*, *5.20*). 2-PCA **5.7** was insoluble in water so gave no modification of Shh C24IIM (*Fig. 5.14*), whereas the ether chain of 2-PCA **5.8** provided enough water-solubility to allow protein modification up to approx. 60% conversion to the singly modified conjugate (*Fig. 5.14*, *5.20*).

iii) Gli-luciferase assay

Next, we decided to assess the bioactivity of the Shh / 2-PCA **5.8** conjugate in comparison to the non-branched 2-PCA and TA4C analogues (**5.2** and **5.3**). Note that 2-PCA **5.2** also contained a piperazine linker, and that TA4C **5.3** had an alternative aromatic ring structure, so this study was not a direct comparison of the effects of branching. Conjugated Shh C24IIM samples (prepared in *Fig. 5.20*) were purified by dialysis to remove excess reagents. Despite improved protein solubility upon modification at 25 °C, concentrations measured by UV-vis spectroscopy were still variable (16-38 μM, *Table S5.7*), presumably because purification by dialysis caused some protein loss. We expected these fluctuations to introduce error into the assay, so we decided to explore alternative options for protein conjugate purification in subsequent studies, as detailed in Section 5.2.3.3.

Gli-luciferase assays indicated that both 2-PCA **5.2** and TA4C **5.3** modifications were tolerated in terms of activity, so both were suitable for reagent development (*Fig. 5.24*). Due to the inherent variability of the assay, little conclusion can be drawn from differences in EC_{50} up to an order of magnitude. Interestingly, 2-PCA **5.8** was not tolerated: the large decrease in activity and potency may have been due an increase in steric

hindrance from the branched structure, preventing binding to the Ptch1 hydrophobic pocket. Note that the bell-shaped curve for TA4C **5.3** may have been due to cell death at higher protein concentrations due to overstimulation.



Small molecule	EC_{50} / nM
Unmodified	3.0 ± 0.6
2-PCA **5.2**	4.8 ± 0.5
TA4C **5.3**	2.5 ± 0.2
2-PCA **5.8**	192.4 ± 94.4
Shh C24IIM Qkine	*4.7 ± 1.6*

Figure 5.24. Gli-luciferase reporter assay for analysis of Shh bioactivity of purified C24IIM conjugates prepared at 25 °C with reagents **5.2**, **5.3** and **5.8** *(data collected by Dr. Kerry Price)*. *Passage number: 3.* Two controls were used: an unmodified sample treated identically to the conjugate samples in the absence of modification reagent *(unmodified)*, and a freshly reconstituted Shh C24IIM solution prepared by Dr. Kerry Price immediately before the protein treatment stage of the assay *(Shh C24IIM Qkine)*.

This study into 2-PCA structure highlighted the challenges in reagent design due to the requirement for both a hydrophobic tail and high water-solubility for the protein modification to occur with enhanced bioactivity; the Gli-luciferase assay indicated that TA4C reagents may be a more suitable starting point for reagent screening. A limitation of the use of 2-PCAs was the time-consuming multi-step synthesis required for each reagent, only to find poor solubility or reduced bioactivity. Whilst the use of a common intermediate **5.12** (*Fig. 5.23*) cut down on the number of steps required for the synthesis of each individual reagent, *O*-alkylation, acetyl deprotection, and preparation of a chain with a suitable leaving group were steps required in each case, with purification necessary for each step. We therefore decided to explore other reagent structures.

5.2.3.2 Phenol ester reagents

i) Synthesis of phenol esters

In attempt to overcome reagent solubility issues encountered with 2-PCAs, we synthesised activated phenol ester **5.14** (*Fig. 5.25*). Such esters have been developed by Mikkelsen *et al.*,[23] for N-terminal acylation.

Activated phenol esters with a sulfonic acid substituent have improved water solubility, allowing the activation of hydrophobic groups in aqueous conditions. We hoped this would enable N-terminal protein modification of Shh with hydrophobic moieties. Initially, we chose to prepare the sulfonamide rather than the sulfonic acid due to the higher selectivity of sulfonamides observed by Mikkelsen *et al*.[23] We incorporated the hydrophobic alkyl chain (C_{10}) as previously for 2-PCAs, with the aim to subsequently increase the chain length up to C_{16} to mimic palmitoylation.

To prepare activated phenol ester **5.14**, we first introduced a sulfonyl chloride substituent *via* electrophilic aromatic substitution to afford compound **5.15** (15%), followed by reaction with dimethylamine to afford compound **5.16** (83%). Compound **5.16** was then coupled with undecanoic acid, with extra portions of DCC and DMAP required to push the reaction towards completion and afford compound **5.14** (84%, *Fig. 5.25*).

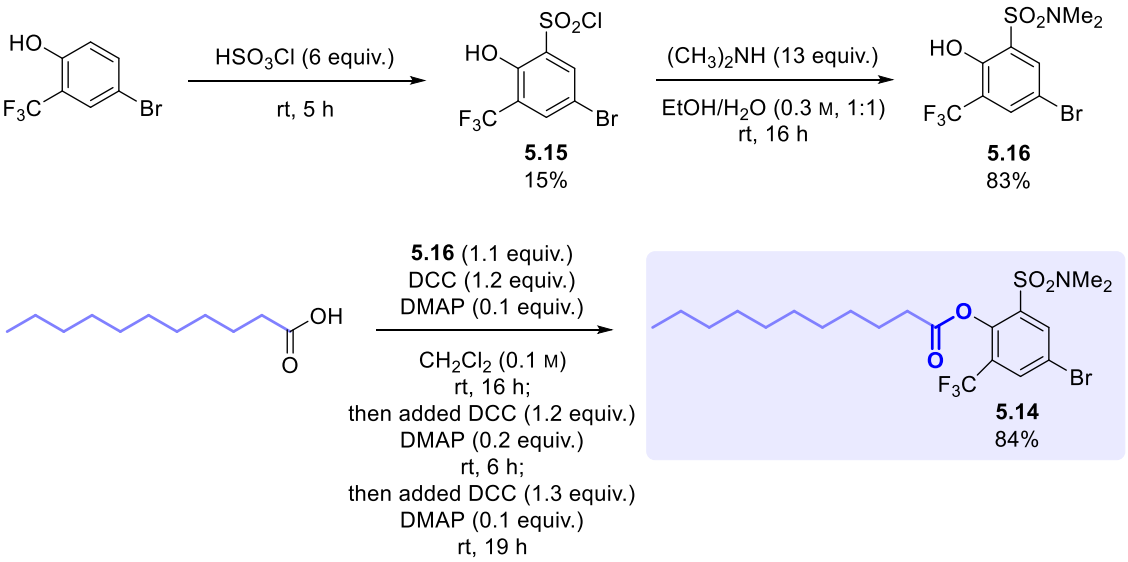


Figure 5.25. Synthesis of activated phenol ester **5.14**.

ii) Protein modification

With activated phenol ester **5.14** in hand, we next attempted the modification of Shh C24IIM (50 µM) in Na phosphate buffer (50 mM, pH 7.5) at 25 °C for 23 h. We used both 3 equiv. (stoichiometry as reported by Mikkelsen *et al.*),[23] and 100 equiv. (stoichiometry we had previously used for N-terminal modification) of **5.14**: we suspected the low number of equivalents of the modification reagent used in the prior report indicated poor selectivity. Unfortunately, reagent **5.14** was still not water-soluble so 10% DMSO was used as an additive, but no modified protein was observed with 3 or 100 equiv. reagent **5.14** (*Fig. 5.26*). Very little protein was detected by LC-MS, and visible precipitation was observed: the presence of the phenol ester may have resulted in protein precipitation.

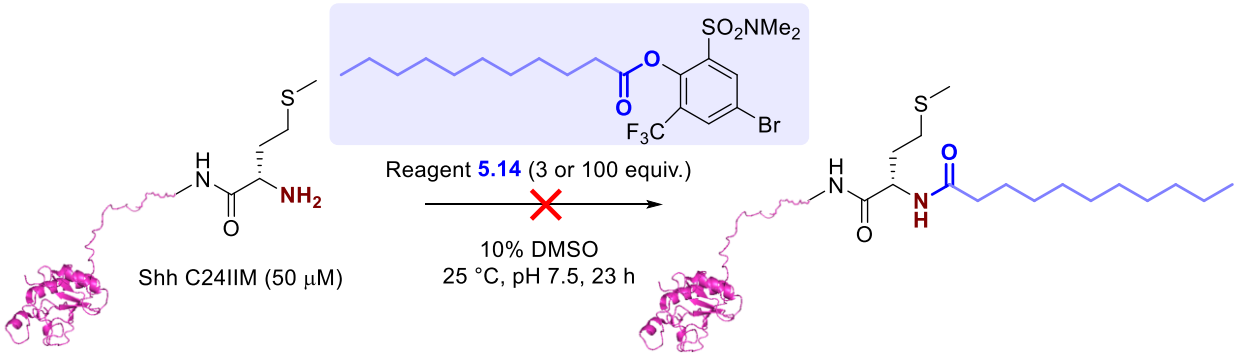


Figure 5.26. Modification of Shh C24IIM with activated phenol ester **5.14**. Shh C24IIM (50 μM) was modified with reagent **5.14** (3 or 100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) with 10% DMSO at 25 °C for 23 h, and the conversion was determined by LC-MS analysis of the crude reaction mixture. Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

We next sought to validate if alternative proteins, known to be less susceptible to precipitation than Shh, could be modified with reagent **5.14**. We attempted the modification of RNase A and equine myoglobin with reagent **5.14** under conditions directly replicated from those reported (*Fig. 5.27*).[23] No protein modification was observed in either case, indicating that reagent **5.14** was an unsuitable protein modification reagent. We suspected that the poor modification of proteins with phenol ester **5.14** was due to poor reagent solubility, as seen previously for 2-PCA **5.7** (also C_{10}). The high protein concentration reported for protein modification (0.73 mM) was also indicative of low reactivity, highlighting a potential limitation of the methodology developed by Mikkelsen *et al*.[23]

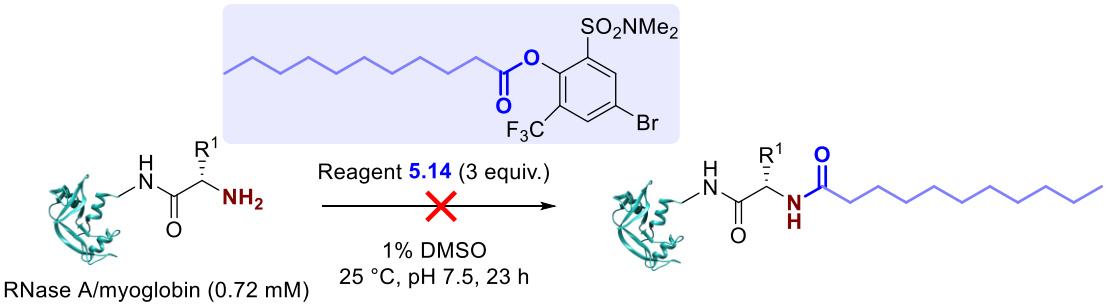


Figure 5.27. Modification of RNase A and equine myoglobin with activated phenol ester **5.14**. RNase A and myoglobin (0.73 mM) were modified with reagent **5.14** (3 equiv.) in HEPES buffer (200 mM, pH 7.4) with 1% DMSO at 25 °C for 23 h, and the conversion was determined by LC-MS analysis of the crude reaction mixture. Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[24].

In the design of new activated phenol esters, another future approach could be to introduce a sulfonic acid group rather than a sulfonamide, as this has been shown to result in increased solubility at neutral pH. However, this enhancement of water-solubility might come at the expense of selectivity, as demonstrated by Mikkelsen *et al*.[23] Shortening the alkyl chain, or using an ether chain with branched alkyl groups, may also improve the reagent solubility whilst still providing the hydrophobicity necessary for Ptch1 binding. Activated

phenol esters were also more suitable than 2-PCAs for high-throughput screening: preparation of reagents with alternative hydrophobic chains would require just one coupling step with mutual intermediate compound **5.16** (*Fig. 5.25*). However, due to poor preliminary results, the activated phenol ester approach was not prioritised.

5.2.3.3 TA4C reagents

We previously established that bioactivity was retained upon modification of Shh C24IIM with TA4C **5.3** (*Fig. 5.24*). TA4C reagents are a suitable strategy for high-throughput screening due to the ease of reagent preparation: Onoda *et al.* developed an *in-situ* Dimroth rearrangement protocol for straight-forward functionalisation of TA4Cs in DMSO, which could be directly added to the protein solution for N-terminal modification (*Fig. 5.28*).[25] This method provides a highly efficient route for the preparation of TA4Cs, allowing us to screen a large range of hydrophobic moieties with minimal synthesis required. We first sought to validate that the Dimroth rearrangement protocol was suitable for modification of model proteins with hydrophobic alkyl chains *(i)* and could be translated to the modification of Shh C24IIM *(ii)*; we then examined the influence of the modification protocol on bioactivity *(iii)*.

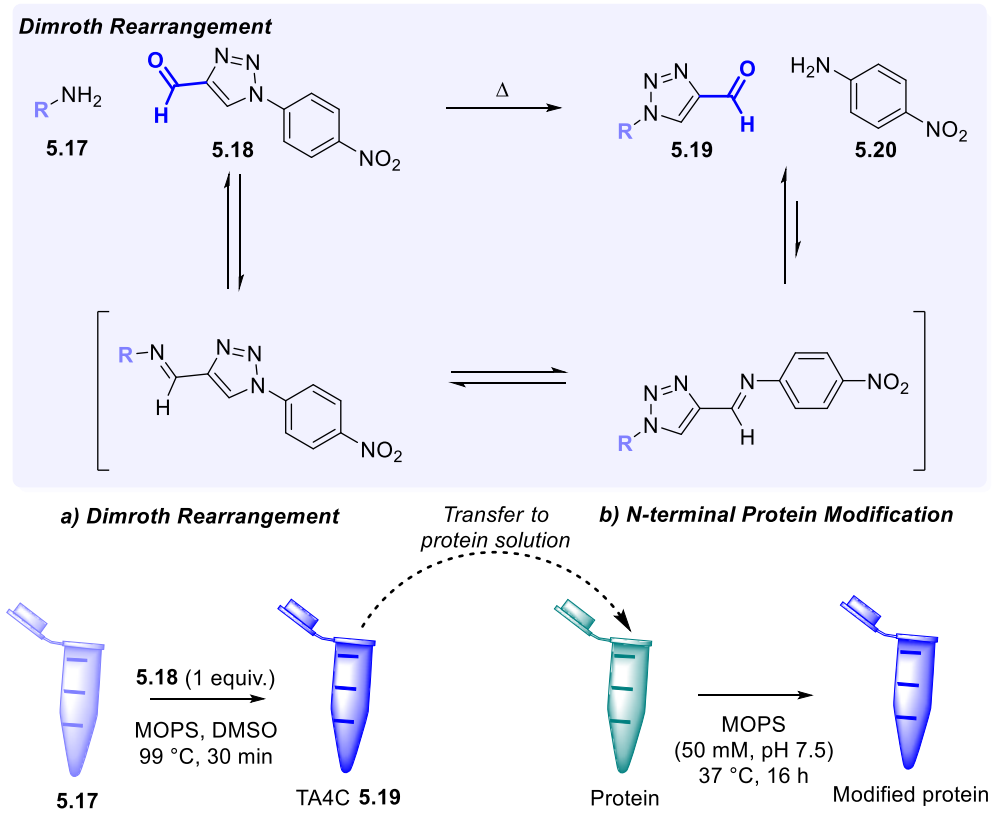


Figure 5.28. Dimroth Rearrangement of TA4C **5.18** to form N-functionalised TA4Cs for N-terminal protein modification (Figure adapted from Onoda *et al.*).[25]

i) Validation of Dimroth rearrangement protocol using RNase A

We first sought to validate if we could use Onoda's protocol[25] to prepare alkylated TA4Cs, followed by modification of RNase A as a model protein. Previously, Onoda *et al.* successfully used this methodology to modify RNase A with TA4Cs functionalised with benzyl, fluorescein, alkyne and azide moieties.[25] Firstly, we prepared TA4C **5.18** as the precursor for Dimroth rearrangement (*Fig. 5.29*). To prepare TA4C **5.18**, azide **5.21** was synthesised from 4-nitroaniline **5.20** *via* a diazonium salt intermediate, followed by copper-catalysed azide alkyne cycloaddition (61%).

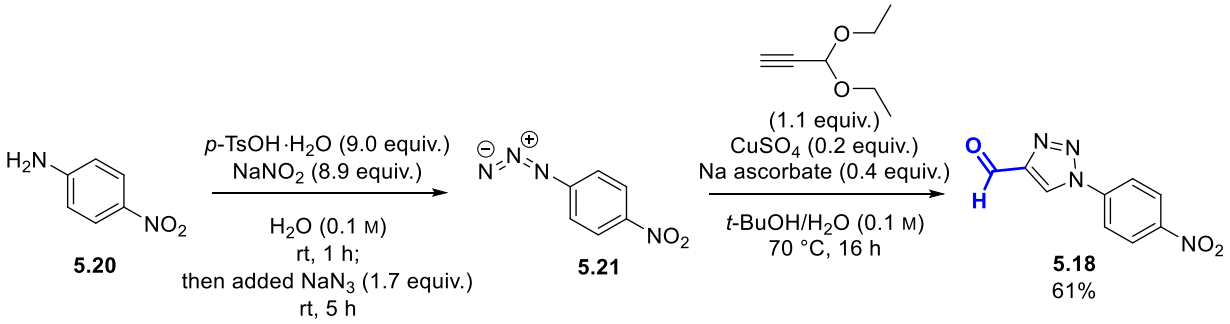


Figure 5.29. Synthesis of TA4C **5.18**.

We decided to compare Onoda's *in-situ* protocol for protein modification with conventional modification with purified TA4C reagent, to determine if the presence of the crude reaction mixture had an impact on the outcome of the protein modification. We chose to prepare *n*-butyl-alkylated TA4C **5.22** initially *via* Dimroth rearrangement (19%), with the hope that the short length of the alkyl chain (C_4) would provide a reagent with sufficient water solubility, having previously established that longer alkyl chains (C_{10}) tended to cause reagent precipitation (*Fig. 5.30*).

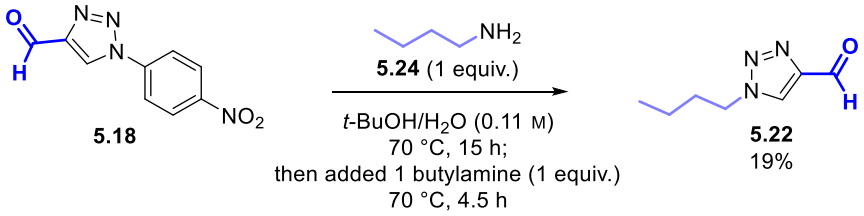


Figure 5.30. Synthesis of TA4C **5.22**.

Next, we prepared TA4C **5.22** *in-situ*, and then modified RNase A with the crude reaction mixture ***a)***, as well as with the purified TA4C under: ***b)*** literature conditions,[25] and ***c)*** our standard set of conditions (*Fig. 5.31*, *Table S5.8*). We also attempted the RNase A modification with TA4C **5.23** (prepared *via* the *in-situ* Dimroth rearrangement protocol ***a)***), comprising of a slightly longer pentyl chain. Promisingly, the level of modification with TA4C **5.22** was similar in the *in-situ* reaction ***a)*** to the modifications carried out with the isolated reagent (***b***, ***c***). However, using the *in-situ* approach ***a)***, 9% modification with TA4C **5.18** was also observed. Analysis

of the Dimroth rearrangement reaction mixtures, which used 1.0 equiv. of amine, by ^{1}H NMR prior to protein modifications indicated 79% conversion to TA4C **5.22**, and 94% conversion to TA4C **5.23** (*Fig. 5.32*). Evidently, pushing the Dimroth rearrangements to completion is important to ensure that TA4C **5.18** has all been consumed before proceeding with the protein modification, to avoid the undesired protein labelling. This study demonstrated that the protocol for *in-situ* preparation of TA4Cs developed by Onoda *et al.* can be applied to alkylated TA4C reagents,[25] but that extra care is required to ensure that the Dimroth rearrangement step has reached full conversion before proceeding with the protein modification step.

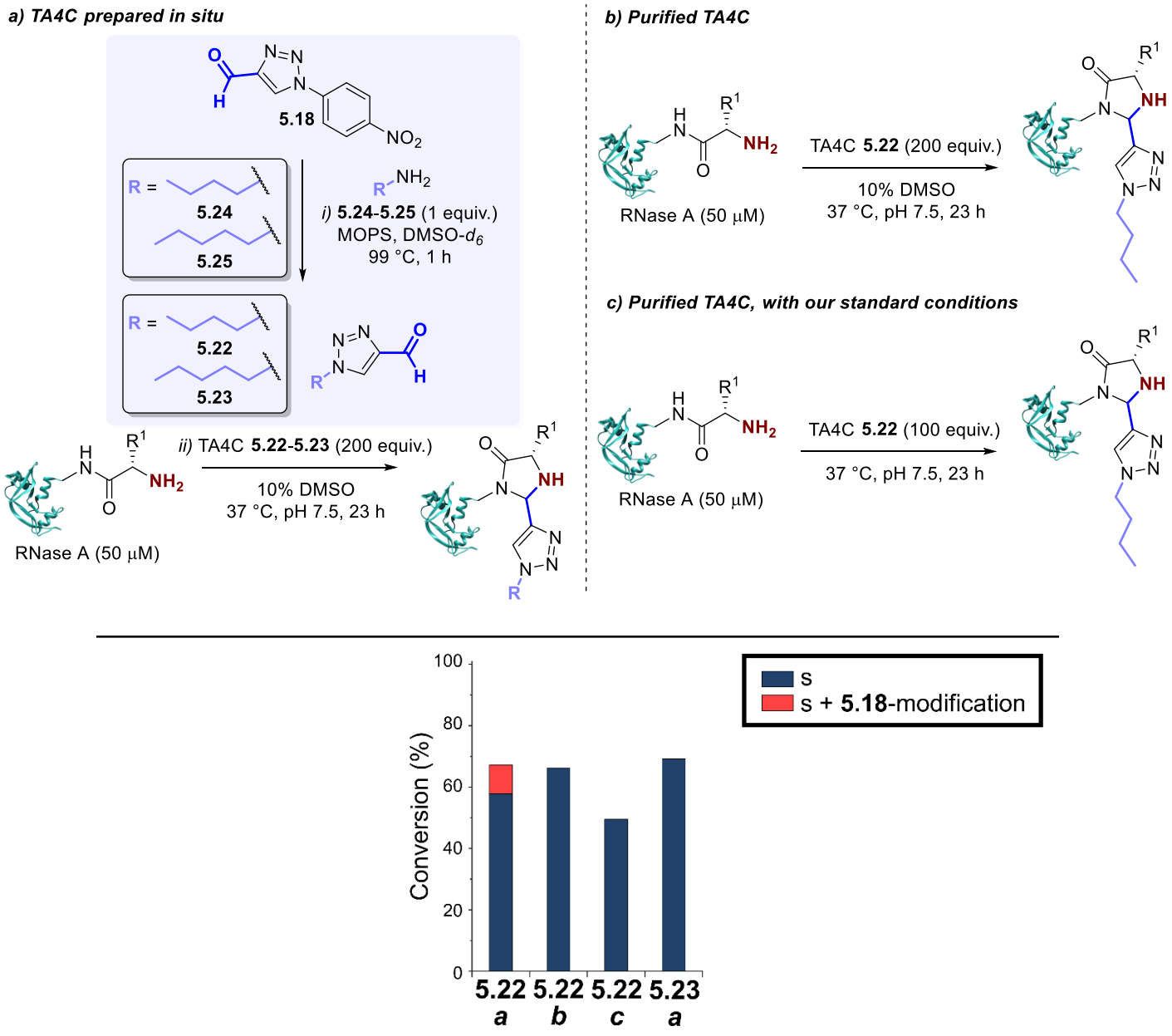


Figure 5.31. Modification of RNase A with TA4Cs **5.22** and **5.23** prepared both *in situ* and by chemical synthesis: ***a)*** *TA4C prepared in situ:* Amine **5.24** or **5.25** (1 equiv.) was added to TA4C **5.18** (1 equiv.) in DMSO-d_6 and the reaction mixture was stirred at 99 °C for 1 h. RNase A (50 μM) was modified with an aliquot of crude reagent **5.22** or **5.23** (200 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 23 h; ***b)*** *Literature conditions:*[25] RNase A (50 μM) was modified with reagent **5.22** (200 equiv.) in Na phosphate buffer (50 mM, pH 7.5) with 10% DMSO at 37 °C for 23 h; ***c)*** *Our standard conditions:* RNase A (50 μM) was modified with reagent **5.22** (100 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 37 °C for 23 h; the conversion was determined by LC-MS analysis of

the crude reaction mixtures (s = single modification, **5.18**-modification due to reaction of the N-terminus with remaining TA4C **5.18**). Figure built using structural data obtained by Chatani *et al.* (RNase A, PDB 1FS3)[24].

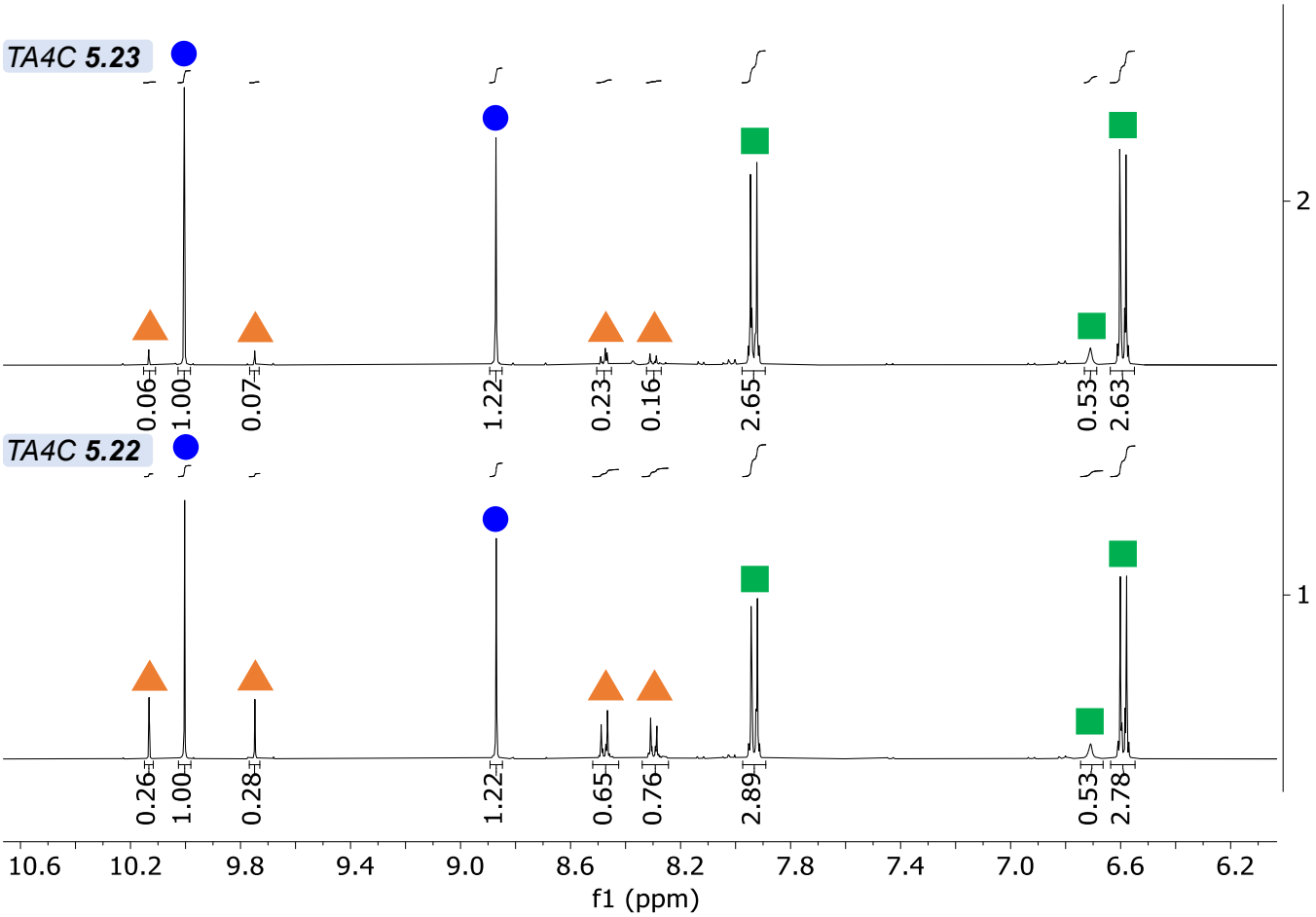


Figure 5.32. ^{1}H NMR of reaction mixtures after Dimroth rearrangement of TA4C **5.18** with amines **5.24** and **5.25** (**triangle** = unreacted TA4C **5.18**, **circle** = TA4C **5.22** or **5.23**, **square** = by-product **5.20**).

ii) Modification of Shh using Dimroth rearrangement protocol

We next applied the Dimroth rearrangement protocol to the modification of Shh C24IIM with TA4Cs **5.22** and **5.23**. We added a small excess of amine (1.5 equiv.) during the rearrangement and increased the reaction time to 2 h, to ensure TA4C **5.18** was all consumed; conversion of the rearrangement reaction mixtures increased to 93% and 100% for TA4Cs **5.22** and **5.23** respectively (*Fig. S5.7*). High levels of modification were observed for both TA4Cs **5.22** and **5.23** before and after purification (approx. 70% after purification), although a small level (10%) of undesired TA4C-**5.18** conjugation was still observed from the Shh modification with TA4C **5.22**. We also modified Shh C24IIM with TA4C **5.18** (200 equiv.), as a positive control of Shh activity for the bioassay (*Fig. 5.33*, *Table S5.9*).

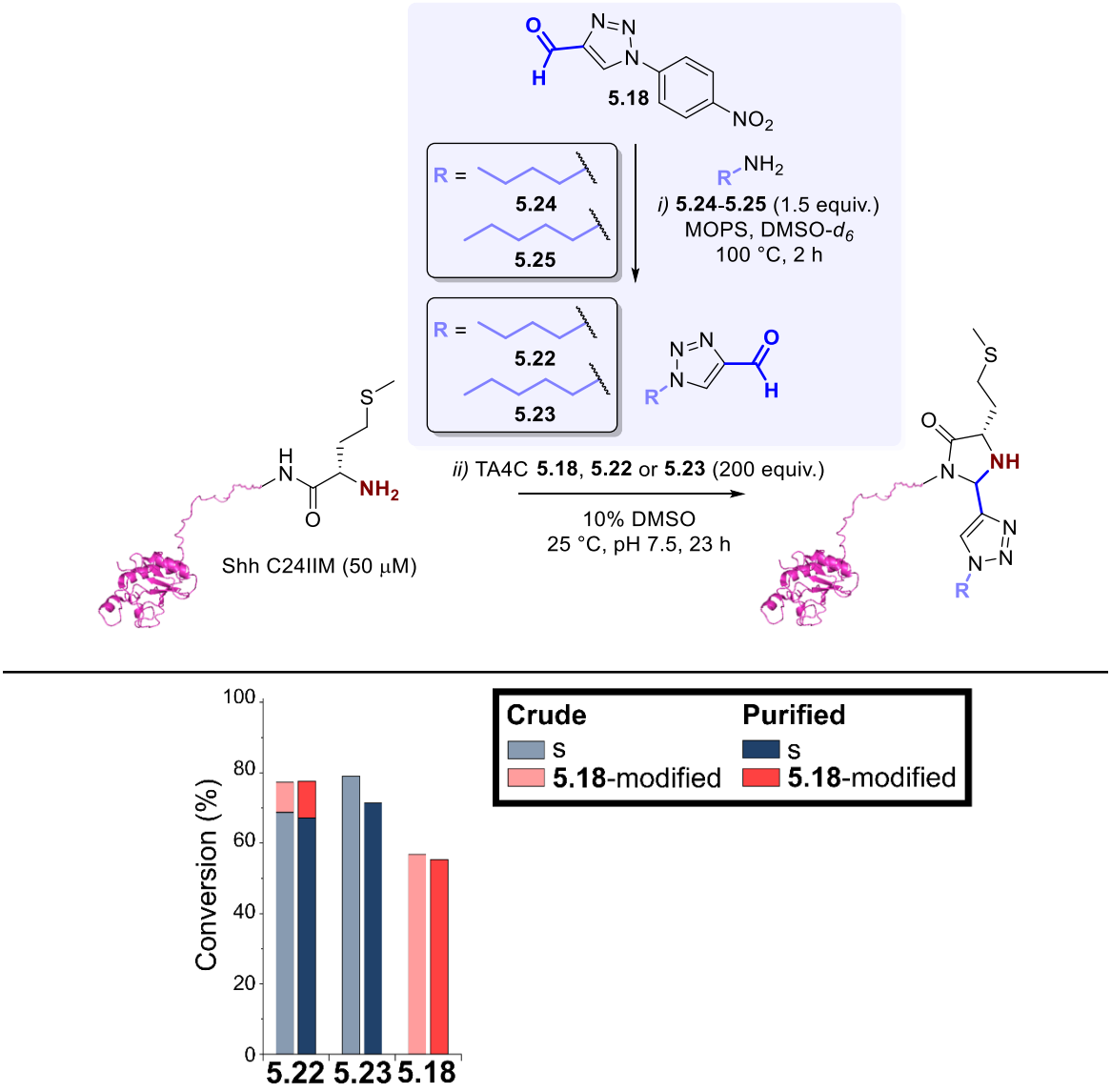


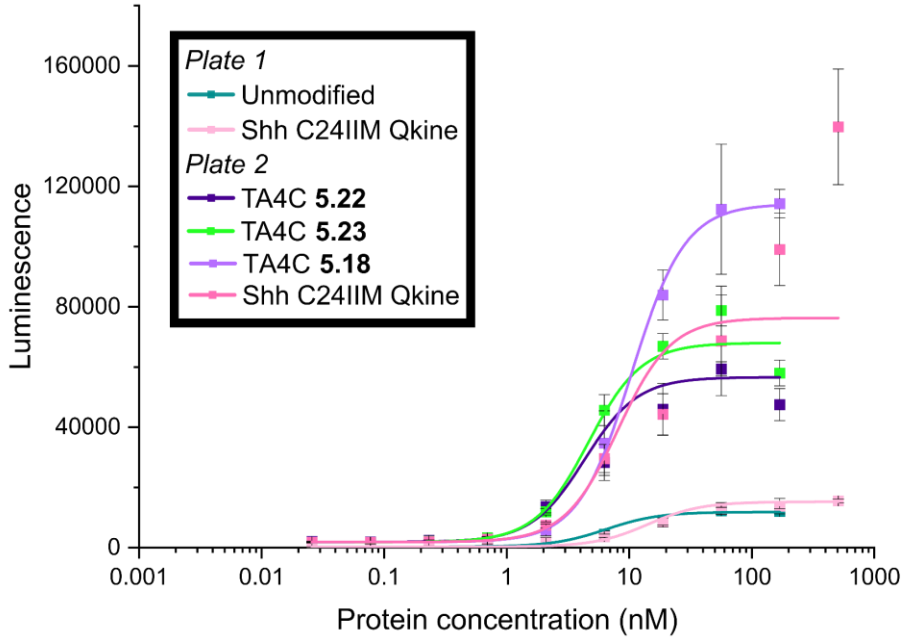
Figure 5.33. Modification of Shh C24IIM with TA4Cs **5.22** and **5.23**, prepared *in situ*, and with TA4C **5.18** (control). Amine **5.24** or **5.25** (1.5 equiv.) was added to TA4C **5.18** (1 equiv.) in DMSO-d_6 and the reaction mixture was stirred at 100 °C for 2 h. Shh C24IIM (50 μM) was modified with an aliquot of crude reagent **5.22** or **5.23**, or a solution of TA4C **5.18** (200 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 25 °C for 23 h, and the conversion was determined by LC-MS before and after purification by size exclusion chromatography (s = single modification). Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

iii) Influence of the modification protocol on bioactivity

To determine if the N-terminal modifications were tolerated and had any effect on biological activity, the modified Shh samples (from *Fig. 5.33*) were sent to Qkine for a Gli-luciferase reporter assay (*Fig. 5.34*). Note that the unmodified control sample (treated under identical conditions in the absence of modification reagent) and one of the freshly reconstituted Qkine Shh C24IIM samples (prepared directly before the assay protein treatment stage) were run on a different plate to the other samples. As discussed in section 5.2.1.2, as our data was not normalised using Renilla luciferase, plate-to-plate variability in maximum luminescence from differences such as cell count cannot be accounted for. Differences between Shh C24IIM control samples

run on different plates (EC_{50} 8.1 nM *vs.* 14.2 nM), with an approx. 4-fold difference in maximum luminescence, indicated that samples on separate plates could not be reliably compared.

Fortunately, Shh modified with TA4C **5.18** remained active, indicating that any undesired side-reaction from incomplete Dimroth rearrangement was unlikely to have a significant impact on bioactivity. The EC_{50} values showed that the modified Shh C24IIM conjugates were all active and had similar activities to the Shh C24IIM positive control, however the values were too similar to be confident that any one sample was the most active given the uncertainty in the concentrations, as discussed below.



Small molecule	EC_{50} / nM
Plate 1	
Unmodified	6.5 ± 0.9
Shh C24IIM Qkine	14.2 ± 3.2
Plate 2	
TA4C **5.22**	4.3 ± 0.7
TA4C **5.23**	4.6 ± 0.4
TA4C **5.18**	10.5 ± 2.1
Shh C24IIM Qkine	8.1 ± 1.5

Figure 5.34. Gli-luciferase reporter assay for analysis of Shh bioactivity of TA4C **5.22**, **5.23** and **5.18** conjugates, prepared *via* Dimroth rearrangement *(data collected by Dr. Kerry Price)*. *Passage number: 3.* Three controls were used: an unmodified sample treated identically to the conjugate samples in the absence of modification reagent *(unmodified)*, and a freshly reconstituted Shh C24IIM solution prepared by Dr. Kerry Price immediately before the protein treatment stage of the assay for each plate *(Shh C24IIM Qkine)*.

In the above experiment, we chose to purify the protein conjugates using PD MiniTrap G-25 columns after labelling instead of dialysis, as previously dialysis resulted in inconsistencies in protein concentration (see section 5.2.3.1) and was ineffective at removing species which were poorly water-soluble; in this case, TA4C **5.18** had poor water-solubility. However, the very low concentrations of the samples due to sample dilution during the size exclusion protocol (2-4 µM, *Table S5.10*) could not be determined accurately due to weak UV-vis absorbance. Due to the low protein concentrations, the ratio of OptiMEM culture media to buffer (4:1 in some cases) in the protein treatment step of the Gli-luciferase assay was low; this may have significantly affected the cell response for the data-points at the highest protein concentrations.

Moving forwards, we next attempted lyophilisation of the protein conjugates before the assay, to allow for reconstitution at a higher concentration and a higher level of confidence in the assay results. However, initial attempts at reconstitution in water by Dr. Kerry Price were unsuccessful, with incomplete solubilisation observed and low protein absorbance values measured by UV-vis spectroscopy. In later bioassays, samples were purified by size exclusion chromatography (PD MiniTrap G-25), with fractionation (50 µL) to minimise the sample dilution. Concentrations of all fractions collected were measured by UV-vis spectroscopy, and the most concentrated fractions were used for determination of conversion and bioactivity.

5.2.4 Exploring the N-terminal chemical space

Having established TA4Cs as a suitable reagent class for high-throughput screening, we next decided to explore the N-terminal chemical space with a wider scope of N-terminal mutants and TA4C structures.

5.2.4.1 N-terminal mutants

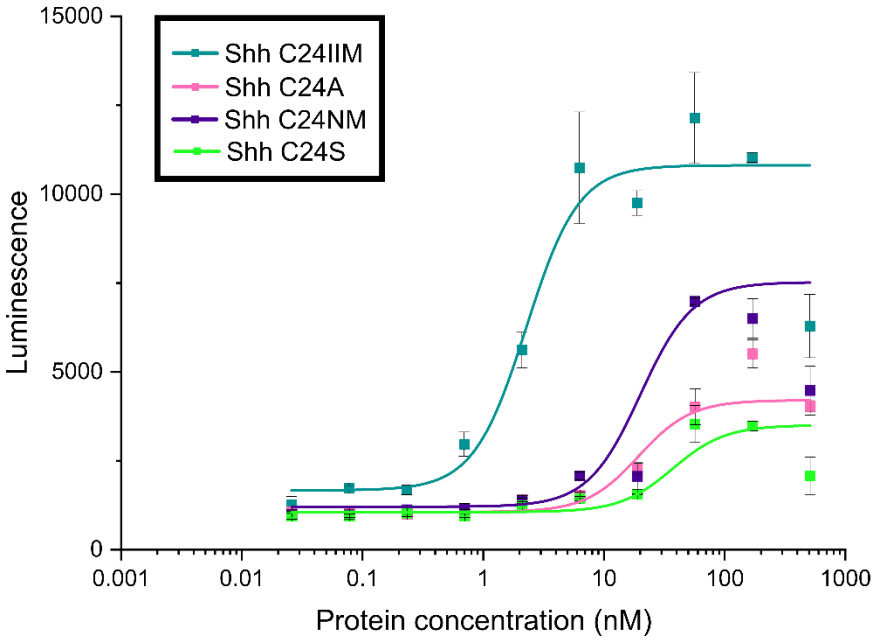
In previous studies, small changes in activity upon modification of Shh C24IIM could not be reliably detected due to issues with the accuracy of the Gli-luciferase assay. We decided to modify other, potentially less active, mutants of Shh in attempt to detect larger differences in activity. Previous work by Taylor *et al.* revealed low tolerance for structural variability in addition to the double Ile N-terminal mutation of Shh C24II, with a decrease in potency upon extension of the N-terminus with tryptophan (Trp) or Ile.[12] As discussed in section 5.2.2.1, we suspected that the Shh C24IIM mutant prepared by Qkine may also exhibit reduced potency (in comparison to the target Shh C24II mutant without an N-terminal Met residue) due to the retention of an N-terminal Met residue. We therefore hoped that there may be more scope for chemical modifications extending less far from the N-terminus, achievable with alternative Shh mutants.

i) Design, preparation, and initial assay of N-terminal mutants

Dr. Helen Bell prepared a range of N-terminal mutants (C24A, C24S and C24NM) to probe the effect of the N-terminal residue on bioactivity (*Fig. S5.3-S5.5*). We selected C24A and C24S as mutants with small N-terminal residues, to allow extra potential space for the structural modifications to be explored. We were also interested in the modification of Shh C24N with an N-terminal asparagine (Asn) residue, due to the ability of the Asn residue to participate in the PCA cyclisation reaction mechanism (as discussed in Chapter 3). However, LC-MS analysis indicated that the N-terminal Met of C24N was not successfully cleaved, with 11% cleaved C24N and 89% C24NM with the retained Met residue (*Fig. S5.4*). Conversion values reported for C24N are an average of the conversions observed for both cleaved/retained mutant, where “unmodified” represents the sum of unmodified C24N and C24NM, and “modified” represents the sum of modified C24N and C24NM. In general, the C24NM modified/unmodified signals dominated, with no modification of the C24N

mutant (with cleaved Met) observed. This suggests that Asn does not accelerate cyclisation, contrary to our predictions in Chapter 3, although further tests are required to confirm this.

Gli-luciferase assay by Dr. Kerry Price revealed that all mutants exhibited a lower maximum response and reduced activity than Shh C24IIM (*Fig. 5.35*). The activity of the C24S was the most reduced due to the hydrophilicity of Ser (EC_{50} = 37.2 nM). We expected no activity, as previously the C24S mutant was reported to be inactive, with <0.1 × activity of unmodified Shh (without PTMs). However, as highlighted previously in section 5.2.2.1, in the literature the bioactivity was assessed using an alternative assay (alkaline phosphatase induction) with a different cell line (C3H10T1/2) so cannot be directly compared.[12]



Mutant	EC_{50} / nM
Shh C24IIM	2.3 ± 0.6
Shh C24A	19.4 ± 5.1
Shh C24NM	20.1 ± 5.3
Shh C24S	37.2 ± 4.6

Figure 5.35. Gli-luciferase reporter assay for analysis of Shh bioactivity of N-terminal mutants *(data collected by Dr. Kerry Price)*. *Passage number: 3.*

ii) Protein Modification

We next modified each N-terminal mutant with TA4C **5.22**, to probe the influence of N-terminal amino acid on the conversion, selectivity, and stability of the conjugation (*Fig. 5.36*, *Table S5.11*). Previously, MacDonald *et al.* demonstrated excellent (>85%) conversion for X-ADSWAG peptides with 2-PCAs prior to purification, where the N-terminal amino acid (X) was A, M, or S.[22] Interestingly, whilst the crude conversion of each of the N-terminal Shh mutants was also similar (approx. 80%, with up to 7% double modification), conjugates showed large differences in selectivity and stability to purification. For example, the purified C24A conjugate showed high stability and the lowest N-terminal selectivity, with 18% double modification, whereas the C24S conjugate showed the lowest stability to purification, with the lowest conversion upon purification of 23%.

These data indicated the significant effect of N-terminal residue on reactivity of both the α-amine and off-target residues with TA4Cs and the resulting stability.

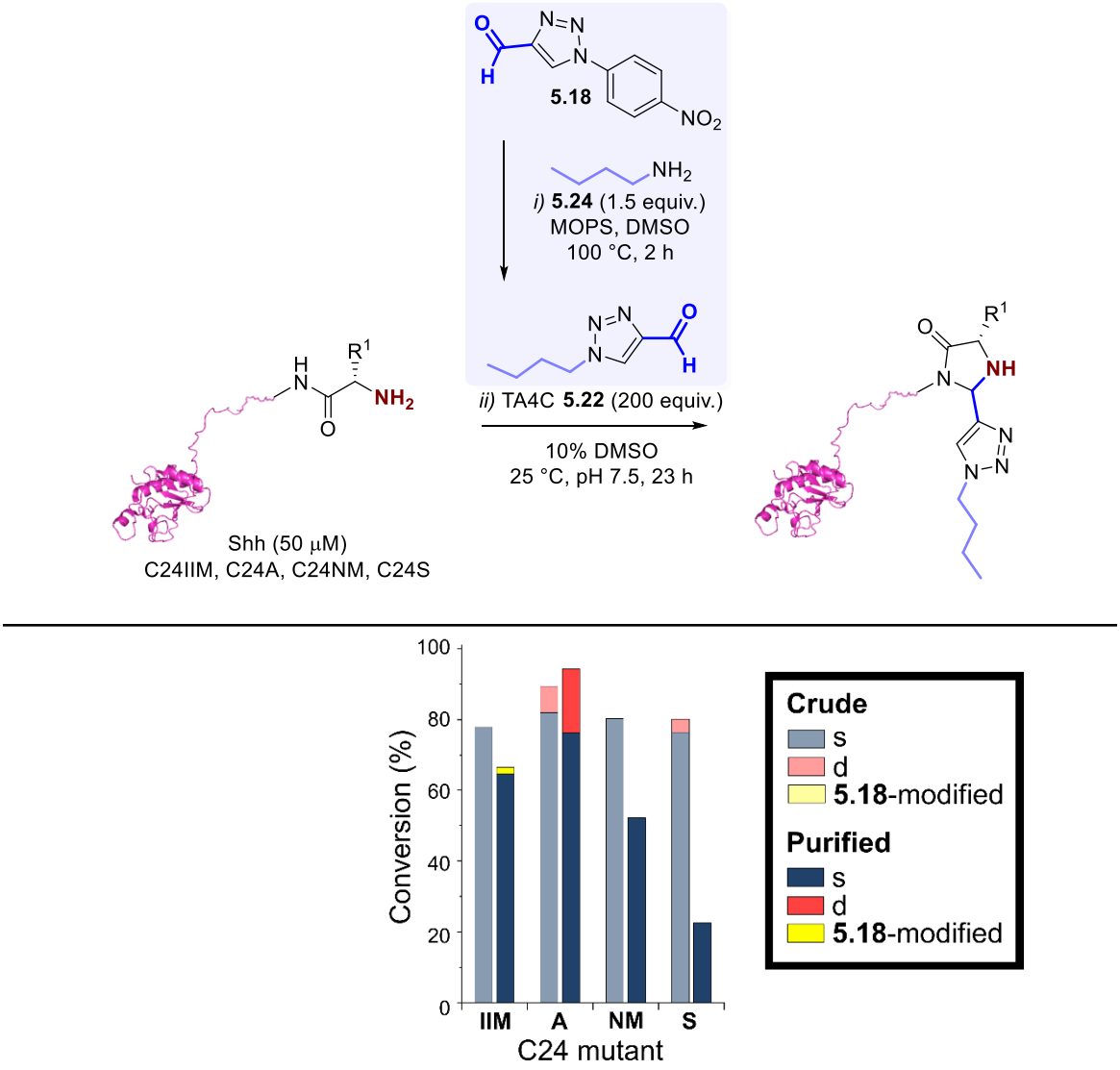


Figure 5.36. Modification of Shh *N*-terminal mutants with TA4C **5.22**, prepared *in situ*. Amine **5.24** (1.5 equiv.) was added to TA4C **5.18** (1 equiv.) in DMSO and the reaction mixture was stirred at 100 °C for 2 h. Shh C24IIM, C24A, C24NM and C24S (50 μM) were modified with an aliquot of crude reagent **5.22** (200 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at 25 °C for 23 h, and the conversion was determined by LC-MS before and after purification by size exclusion chromatography (s = single modification, d = double modification). Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

5.2.4.2 Alkyl chain length

Having demonstrated that all the N-terminal mutants prepared were suitable for modification with TA4Cs (albeit with variation in conversion, selectivity, and stability), we next decided to explore the scope of TA4C reagent structure. We carried out an initial screen of a range of alkyl chains for each of the N-terminal mutants *(i)*, assessed bioactivity, and tested the reproducibility of results for the most promising candidates *(ii)*.

i) Alkyl chain screen with a range of N-terminal mutants

TA4Cs **5.22**, **5.23**, and **5.26**-**5.30** were prepared *via* Dimroth rearrangement in DMSO to investigate the effect of alkyl chain length and branching/aromaticity on activity. The logarithm of the calculated partition coefficient (cLogP) of each compound in octanol (hydrophobic) and water (hydrophilic) was used as an estimate of hydrophobicity (*Fig. 5.37*).

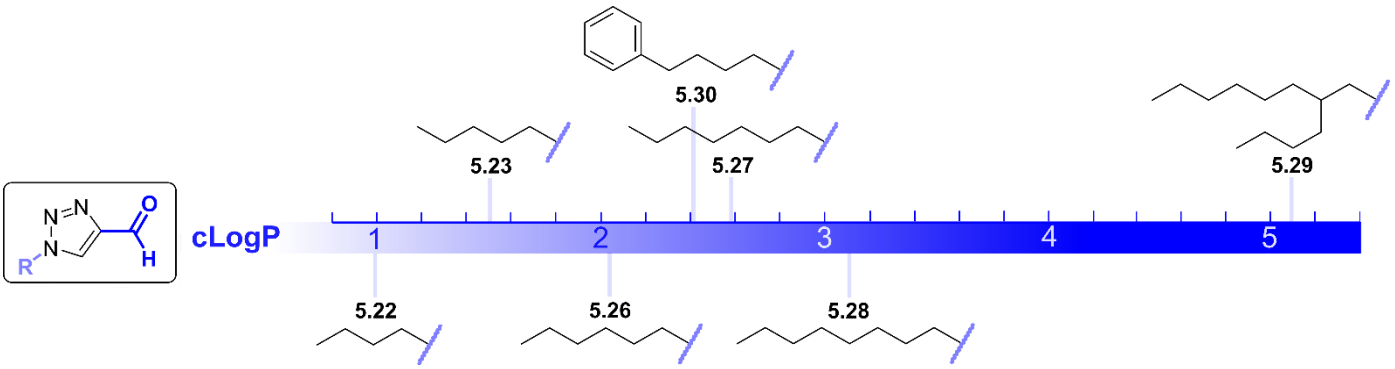
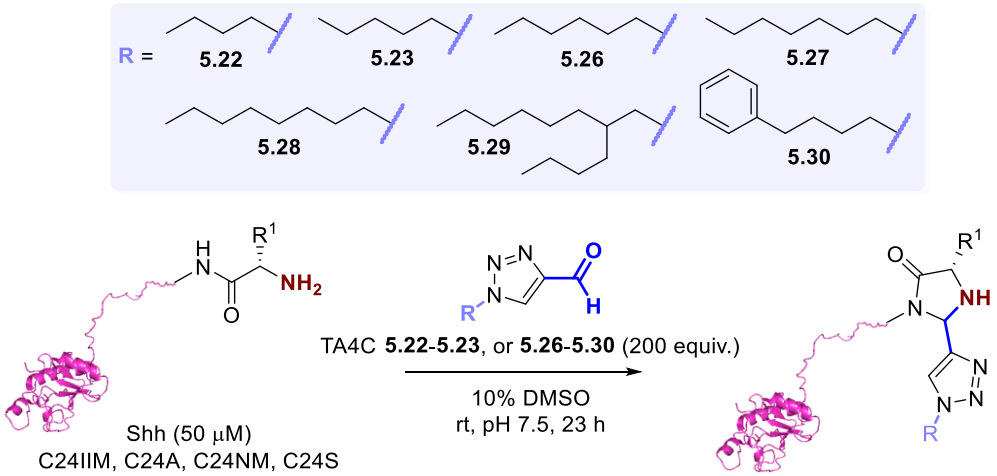


Figure 5.37. cLogP values of TA4Cs **5.22**-**5.23** and **5.26**-**5.30**.

Each of the Shh N-terminal mutants (in 50 mM Na phosphate buffer, pH 7.5) was modified with aliquots of the Dimroth rearrangement reaction mixtures of the TA4Cs (*Fig. 5.38*, *Table S5.12*). As expected, as the alkyl chain length increased from C_4-C_8, a decrease in conversion of each of the N-terminal mutants was observed due to a decrease in water solubility of modification reagent (cLogP increased from 0.99 to 3.1). No modification was observed for branched TA4C **5.29** due to steric hindrance and very low water solubility (cLogP = 5.1). Despite slightly higher predicted water solubility than TA4C **5.27** (cLogP = 2.4 *vs* 2.6 for TA4C **5.30** and **5.27** respectively), TA4C **5.30** showed slightly poorer reactivity than TA4C **5.27** with most of the mutants; we suspected this to be due to increased steric hindrance from the aromatic ring.

A general trend was observed across the TA4C reagents, where C24NM exhibited the highest levels of conversion, and C24S the lowest levels. The observed reactivity of N-terminal mutants was not in agreement with the pK_a values of the N-terminal ammonium of the respective amino acids: 9.69, 9.21 and 9.15 for Ala, Met, and Ser respectively.[26] Based on pK_a, we would expect the highest imine formation reactivity for C24S and the lowest reactivity for C24A. Despite having the same N-terminal Met residue, C24NM showed higher reactivity than C24IIM; extension of the N-terminus of C24IIM with an additional residue in comparison to C24NM may have exposed the N-terminus to a different local environment. This highlights that the influence of the surrounding protein environment on the modification extends beyond the nature of the N-terminal residue or pK_a alone, as implicated in our early study in Chapter 2. Interestingly, the variation in conversion and selectivity between N-terminal mutants upon modification with TA4C **5.22** was less pronounced than in previous LC-MS experiments (*Fig. 5.36*).



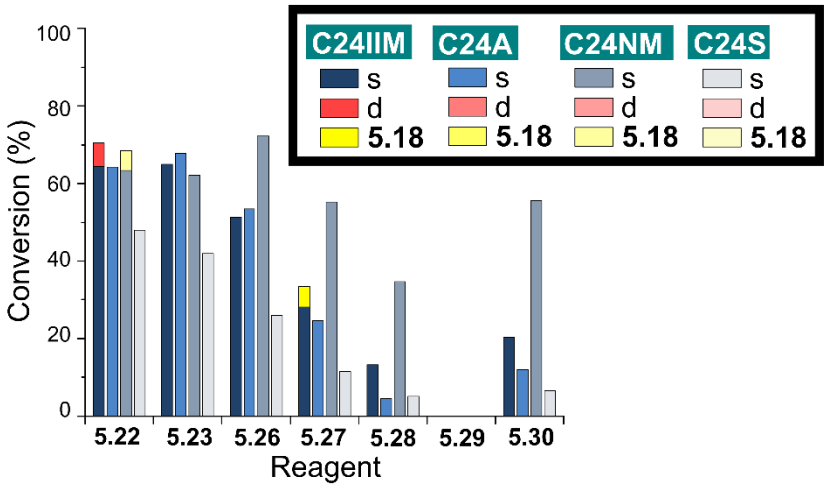
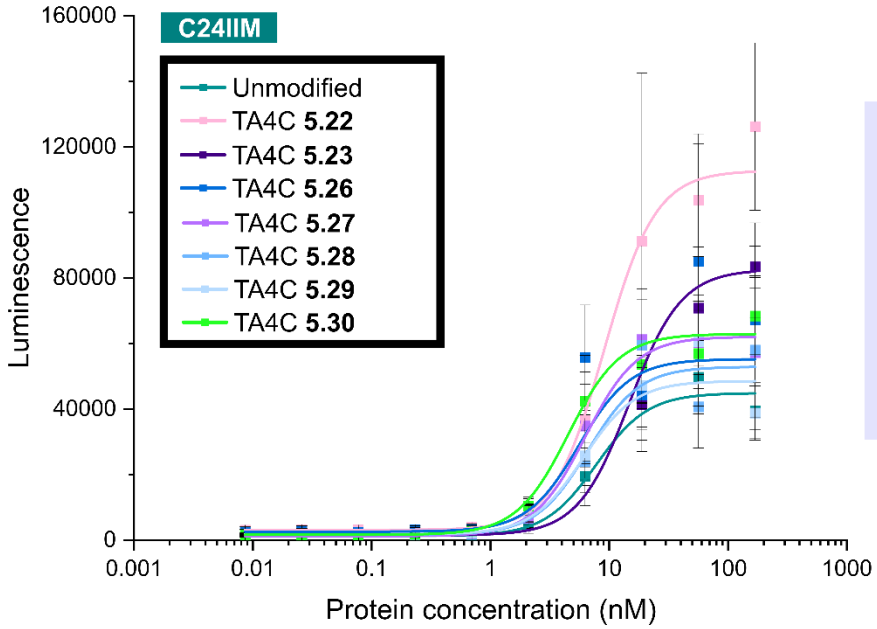


Figure 5.38. Modification of Shh *N*-terminal mutants with TA4Cs **5.22**-**5.23**, and **5.26**-**5.30**, prepared *in situ*. Shh C24IIM, C24A, C24NM and C24S (50 μM) were modified with an aliquot of crude reagent **5.22**, **5.23**, or **5.26**-**5.30** (200 equiv.) in Na phosphate buffer (50 mM, pH 7.5) at rt for 23 h, and the conversion was determined by LC-MS after purification by size exclusion chromatography (s = single modification, d = double modification). Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

ii) Gli-luciferase assays

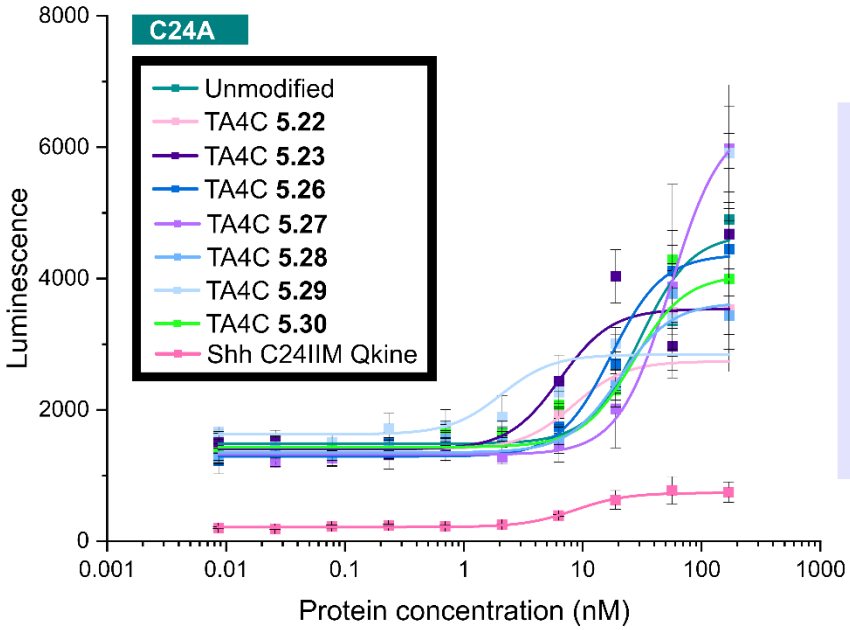
We next assessed the bioactivity of the conjugates *via* Gli-luciferase assays (*Fig. 5.39*). Initially, we hoped that an increase in Shh potency could be achieved by modification with long alkyl chains. Contrary to our predictions, an increase in chain length from TA4C **5.22**-**5.28** (C_4-C_8) gave minimal enhancement in potency: we suspected this to be due to poor conversions observed for longer chain lengths due to increased reagent hydrophobicity, and the use of these heterogeneous mixtures of modified/unmodified protein. Note that whilst conjugates were purified to remove excess modification reagent, unmodified protein was not removed, i.e. samples represented a mixture of modified and unmodified protein.

a)



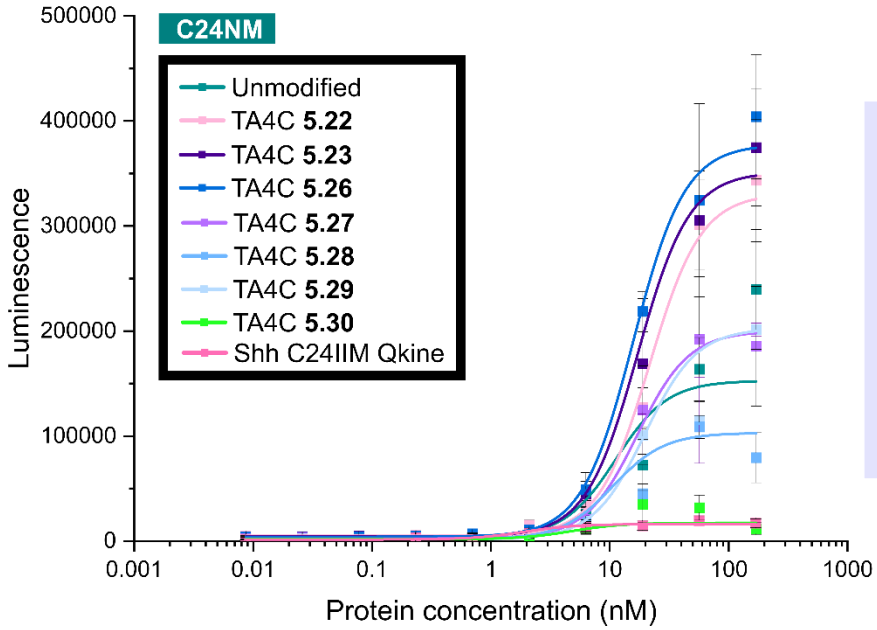
Small molecule	EC_{50} / nM
Unmodified Shh C24IIM	7.8 ± 1.3
TA4C **5.22**	9.0 ± 1.9
TA4C **5.23**	14.9 ± 5.2
TA4C **5.26**	5.6 ± 1.8
TA4C **5.27**	6.3 ± 0.6
TA4C **5.28**	6.6 ± 1.6
TA4C **5.29**	6.1 ± 0.9
TA4C **5.30**	4.5 ± 0.7

b)



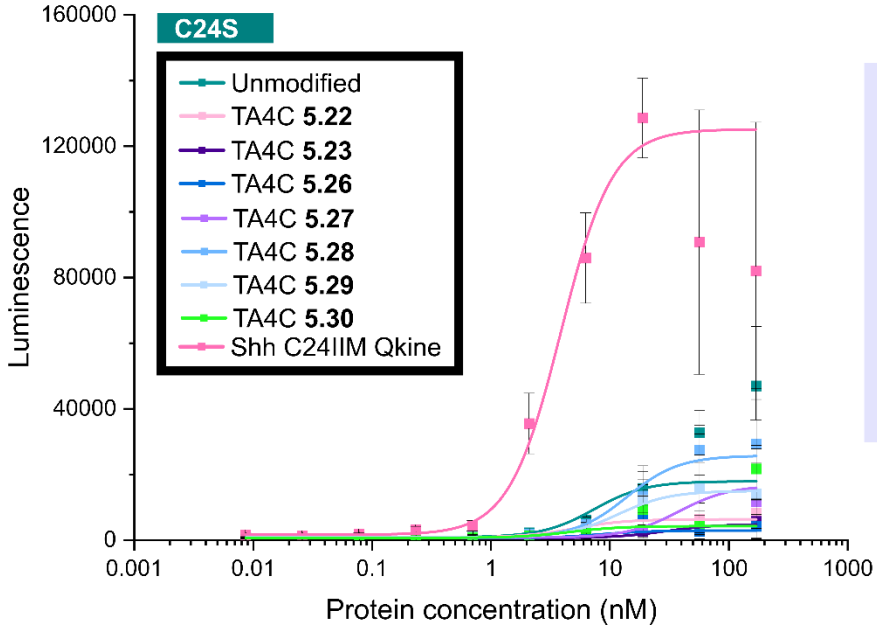
Small molecule	EC_{50} / nM
Unmodified Shh C24A	30.3 ± 14.3
TA4C **5.22**	8.4 ± 4.1
TA4C **5.23**	6.4 ± 2.7
TA4C **5.26**	17.4 ± 8.6
TA4C **5.27**	54.6 ± 26.3
TA4C **5.28**	20.5 ± 10.6
TA4C **5.29**	2.1 ± 0.4
TA4C **5.30**	26.7 ± 11.3
Shh C24IIM Qkine	*8.6 ± 2.2*

c)



Small molecule	EC_{50} / nM
Unmodified Shh C24NM	11.3 ± 1.7
TA4C **5.22**	21.6 ± 3.5
TA4C **5.23**	17.2 ± 2.4
TA4C **5.26**	15.8 ± 2.5
TA4C **5.27**	17.1 ± 5.4
TA4C **5.28**	10.6 ± 1.3
TA4C **5.29**	19.8 ± 2.4
TA4C **5.30**	4.8 ± 2.0
Shh C24IIM Qkine	*1.9 ± 0.5*

d)



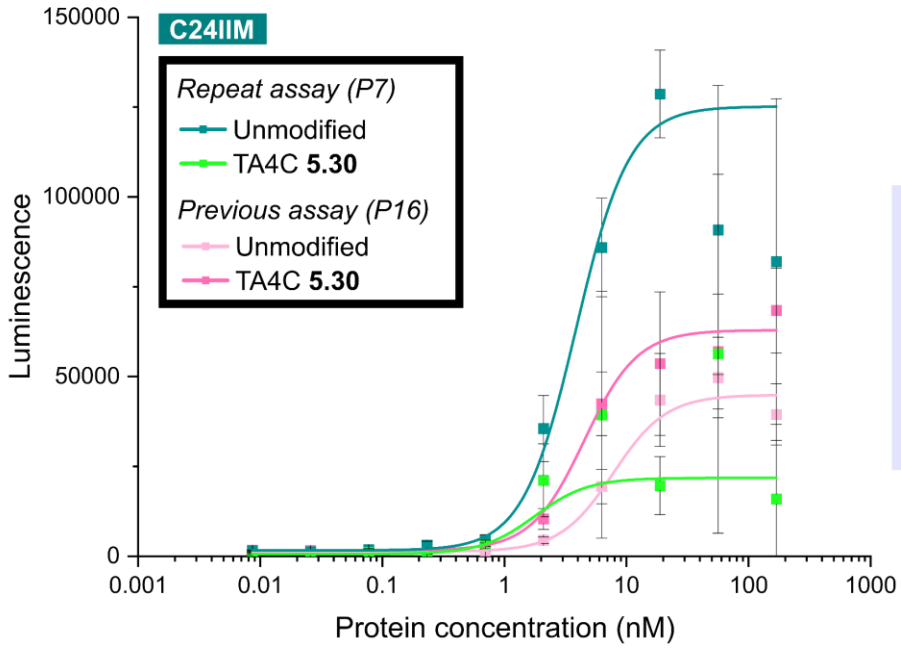
Small molecule	EC_{50} / nM
Unmodified Shh C24S	7.4 ± 1.4
TA4C **5.22**	5.0 ± 3.2
TA4C **5.23**	29.8 ± 47.9
TA4C **5.26**	8.2 ± 6.3
TA4C **5.27**	38.0 ± 28.6
TA4C **5.28**	14.0 ± 3.8
TA4C **5.29**	11.8 ± 9.1
TA4C **5.30**	4.6 ± 3.8
Shh C24IIM Qkine	*3.9 ± 0.9*

Figure 5.39. Gli-luciferase reporter assay for analysis of Shh bioactivity upon modification of: ***a)*** C24IIM (*passage number: 16*); ***b)*** C24A (*passage number: 19; weak luminescence values for Shh C24IIM as measured using 2700, rather than 3600, gain setting; poor quality data due to the high passage number of the Gli-3T3 cells resulting in low fluorescence readings*); ***c)*** C24NM (*passage number: 14*); ***d)*** C24S (*passage number: 7*); with TA4Cs **5.22**, **5.23**, or **5.26**-**5.30**.

As discussed previously, little conclusion can be drawn from differences in EC_{50} up to an order of magnitude. Note that plate-to-plate variability, seen previously in section 5.2.2.3, was not considered here, with only one overall control used per assay. We decided to repeat Gli-luciferase assays of several of the most promising reagent/mutant combinations in attempt to obtain reproducible results. Previously, modification of Shh

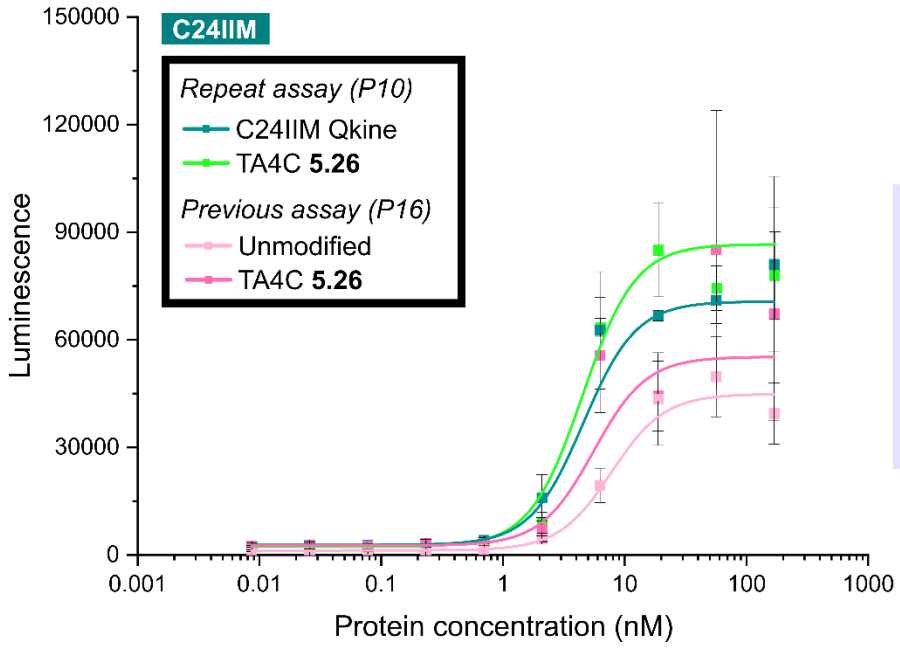
C24IIM with TA4C conjugates **5.26**-**5.30** resulted in up to a 1.7-fold increase in potency (*Fig. 5.39a*), and modification of Shh C24A with TA4C **5.23** also gave a 1.3-fold increase in potency with respect to C24IIM (*Fig. 5.39b*). Whilst the enhancement in potency was reproduced for the C24IIM / TA4C **5.30** conjugate (*Fig. 5.40a*), enhancements in potency for the C24IIM / TA4C **5.26** *(Fig. 5.40b)* and C24A / TA4C **5.23** (*Fig. 5.40c*) conjugates were not reproduced. Any enhancements observed were therefore likely due to assay variability.

a)



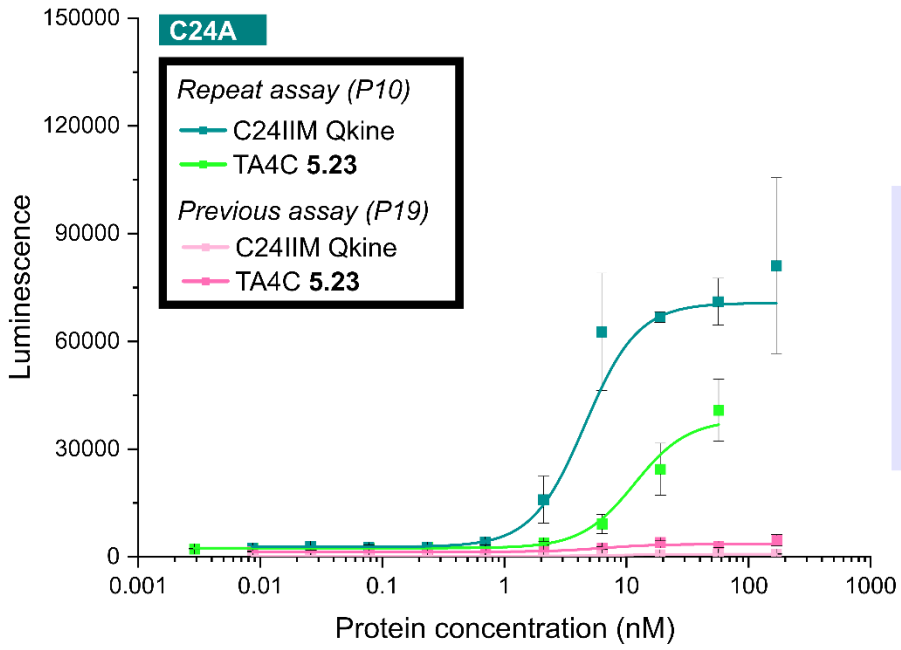
Small molecule	EC_{50} / nM
Repeat assay (P7)	
C24IIM Qkine	3.9 ± 0.7
TA4C **5.30**	1.8 ± 0.7
Previous assay (P16)	
Unmodified	7.8 ± 1.3
TA4C **5.30**	4.5 ± 0.7

b)



Small molecule	EC_{50} / nM
Repeat assay (P10)	
C24IIM Qkine	4.6 ± 1.0
TA4C **5.26**	4.5 ± 0.4
Previous assay (P16)	
Unmodified	7.8 ± 1.3
TA4C **5.26**	5.6 ± 1.8

c)



Small molecule	EC_{50} / nM
Repeat assay (P10)	
C24IIM Qkine	4.6 ± 1.0
TA4C **5.23**	6.4 ± 2.0
Previous assay (P19)	
C24IIM Qkine	8.6 ± 2.2
TA4C **5.23**	6.4 ± 2.7

Figure 5.40. Gli-luciferase reporter assay for analysis of Shh bioactivity of: ***a)*** C24IIM / TA4C **5.30** conjugates *(data from passage number 16 from previous assay in Fig. 5.39a; repeat assay passage number: 7*); ***b)*** C24IIM / TA4C **5.26** conjugates *(data from passage number 16 from previous assay in Fig. 5.39a; repeat assay passage number: 10*); ***c)*** C24A / TA4C **5.23** conjugates *(data from passage number 19 from previous assay in Fig. 5.39b; repeat assay passage number: 10*).

5.2.4.3 Alkyl chain solubility

Having observed a decrease in protein conversion with increasing reagent hydrophobicity, we next aimed to increase the conversions of Shh modification with the long-chain reagents through use of solubilising additives. We set out to: *i)* identify solubilising additives tolerated by Shh C24IIM; *ii)* optimise the level of additives required for high conversions, and *iii)* probe the influence of the increased conversion on bioactivity.

i) Identification of solubilising additives tolerated by Shh C24IIM

A preliminary thermal shift assay was conducted in triplicate by Dr. Greg Hughes at Qkine to measure the melting point (T_m) of Shh as an indication of stability, to determine if more reagent-solubilising conditions were tolerated by the protein (*Fig. 5.41*). Parameters explored included DMSO levels (0-10%), and the use of additives (0.1-10% glycerol, 0.01-1% tween-20). These additives were selected as tween-20 is a surfactant commonly used as a detergent; glycerol is amphiphilic and is known to inhibit protein aggregation through interactions with both the polar solvent, and with the hydrophobic surface regions of proteins; and the modification reagents were known to be soluble in DMSO.[27] Another option to study in future work could be refolding techniques using: “super-buffers” which contain multiple buffering components, e.g. MMT (L-malic acid, MES and Tris), SSG (succinic acid, sodium phosphate and glycine) or CHC (citric acid, HEPES and

CHES); urea as an additive, shown to inhibit aggregation and promote refolding at low concentrations; amino acid additives shown to inhibit aggregation, e.g. arginine, glycine or proline, or sugars to stabilise protein structures, e.g. dextrose, sucrose.[28]

A solution of Shh (5 µM) in sodium phosphate buffer (50 mM, pH 7.5) and MOPS (1 mM), in addition to solubilising additives/DMSO, was gradually heated in the presence of 5× SYPRO™ orange dye. At low temperatures, the protein was folded: the SYPRO™ orange molecules were surrounded by water molecules, resulting in fluorescence quenching. As the temperature increased, the protein began to unfold during melting, exposing hydrophobic regions of Shh. The dye bound to these exposed hydrophobic regions, leading to reduced fluorescence quenching. Melting temperature could therefore be calculated from the fluorescence readings obtained by differential scanning fluorimetry (DSF) with a PCR instrument.

Normalisation of melting curve values (MC) removed background fluorescence and any effects from interaction of buffer components with the dye, *via* subtraction of the value of a no-protein control from each of the data points of the curve. The first derivative of the curve (FDC) was calculated (*Equation 5.2*) to determine the point of inflection of the melting curve; the T_m was the maximum point of the FDC (*Equation 5.3*).[29]

$$FDC = \frac{MC(n+1) - MC(n-1)}{2 * \Delta t} \qquad \textit{(Equation 5.2)}$$

$$Tm = Temperature(maxFDC) \qquad \textit{(Equation 5.3)}$$

The thermal shift assay indicated little change in T_m in 10% DMSO (55.72 °C) in comparison to 1% DMSO (55.52 °C), confirming that the use of 10% DMSO under protein modification conditions was tolerated (*Fig. 5.41b*). The use of 10% glycerol was also tolerated, with minimal influence on T_m (56.08 °C, *Fig. 5.41c*). No conclusions can be made from tween-20: the lack of visible phase transition was presumably due to interference of the tween-20 additive with the SYPRO™ orange dye (*Fig. 5.41d*).

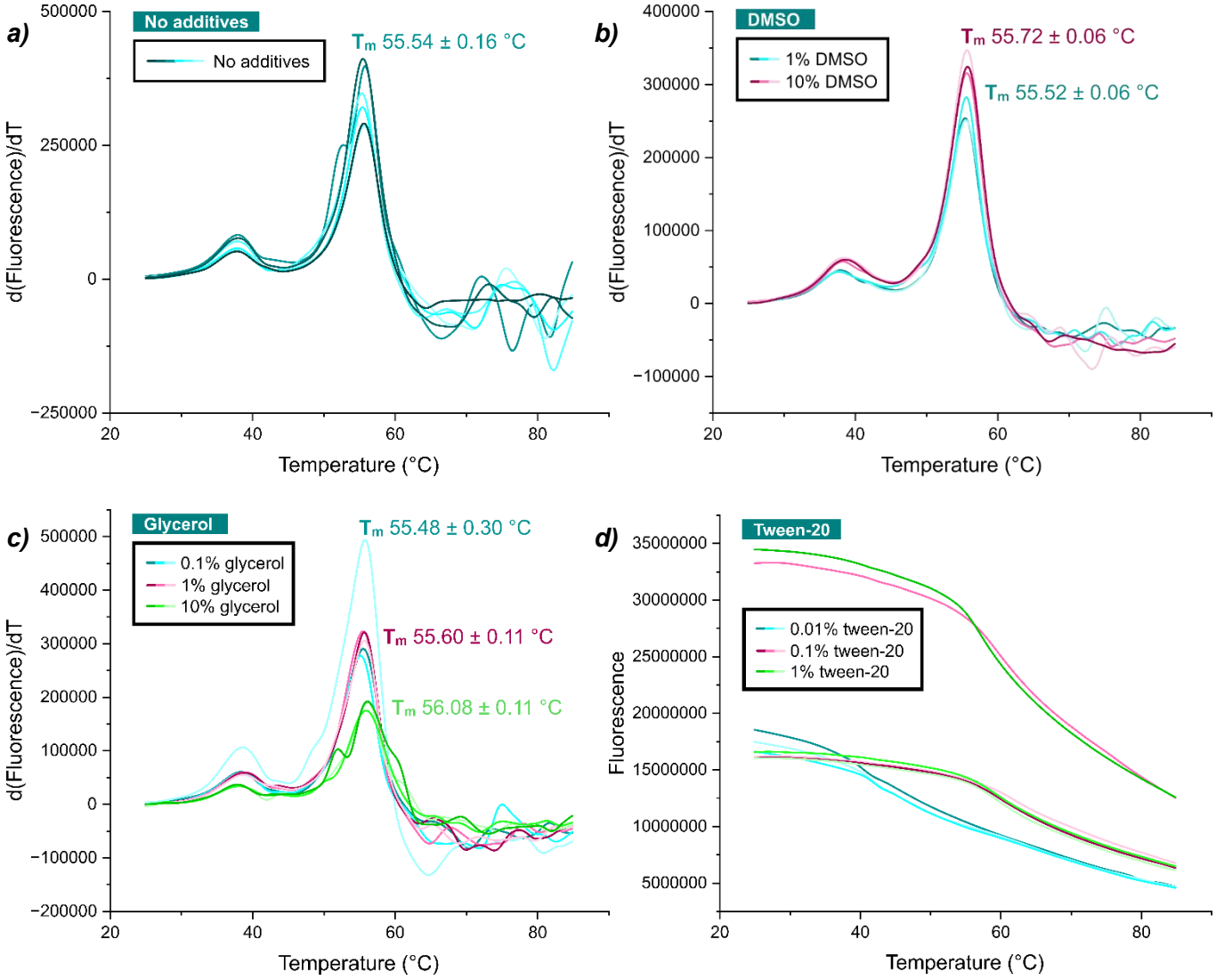


Figure 5.41. Thermal shift assay of Shh C24IIM with ***a)*** no additives; ***b)*** 1-10% DMSO; ***c)*** 0.1-10% glycerol; ***d)*** 0.01-1% tween-20 (data collected by Dr. Greg Hughes at Qkine).

ii) Optimisation of solubilising additive levels

Having established that 10% DMSO and glycerol were tolerated by Shh, we next modified Shh C24IIM (50 μM) with TA4C **5.28** (200 equiv.) under a range of solubilising conditions (10% DMSO with 0.1-10% glycerol or with 0.01-1% tween-20). TA4C **5.28** was selected as the reagent with the longest alkyl chain (and thus lowest conversions, not considering TA4C **5.29**) used so far (C_8), with the greatest scope for increase in conversion, and the most similarity to palmitoylation in terms of alkyl chain length. Concentrations of conjugates following purification by size exclusion chromatography were measured by UV-vis spectroscopy. Interestingly, no protein was detected at 0.1%-1% glycerol, or 1% tween-20 in 10% DMSO: we suspected that these alterations to the reaction conditions may have resulted in protein precipitation (*Table S5.15*). Conversion of purified C24IIM / TA4C **5.28** conjugates was determined by LC-MS analysis (*Fig. 5.42*, *Table S5.14*). Promisingly, increased conversions were observed using 10% glycerol, 0.01% tween-20, and 0.1% tween-20 in 10% DMSO. The largest increase in conversion from 13 to 31% was observed upon addition of

0.1% tween-20, so this was selected as the optimal additive level, despite unknown effects on protein stability from the thermal shift assay (*Fig. 5.41*).

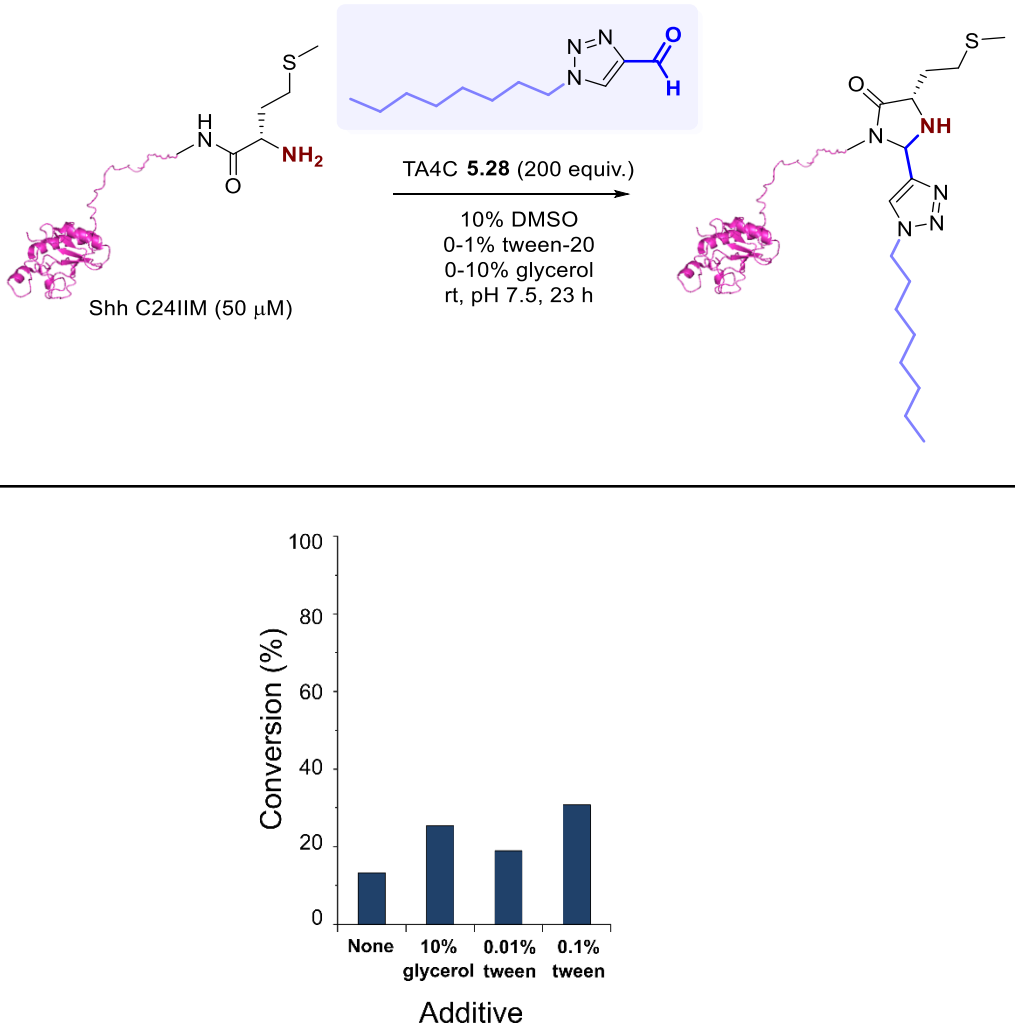


Figure 5.42. Modification of Shh C24IIM with TA4C **5.28**, prepared *in situ*. Shh C24IIM (50 μM) was modified with an aliquot of crude reagent **5.28** (200 equiv.) in Na phosphate buffer (50 mM, pH 7.5) with glycerol (0 or 10%) or tween-20 (0, 0.01% and 0.1%) at rt for 23 h, and the conversion was determined by LC-MS after purification by size exclusion chromatography. Only single modifications observed. Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

iii) Probing the influence of the increased conversion on bioactivity

With optimised modification conditions in hand for reagents with poor water-solubility, we next carried out the modification of all N-terminal Shh mutants with TA4Cs **5.28** and **5.30**, with 0.1% tween-20 (*Fig. 5.43*, *Table S5.16*). We chose to include TA4C **5.30**, as the reagent with the most promising effects previously observed on EC_{50} despite low conversion levels (*Fig. 5.40a*). The conversion was improved for the modification of all N-terminal mutants with TA4C **5.28**, and for the modification of C24IIM with TA4C **5.30** (from 20 to 33% with use of 0.1% tween-20).

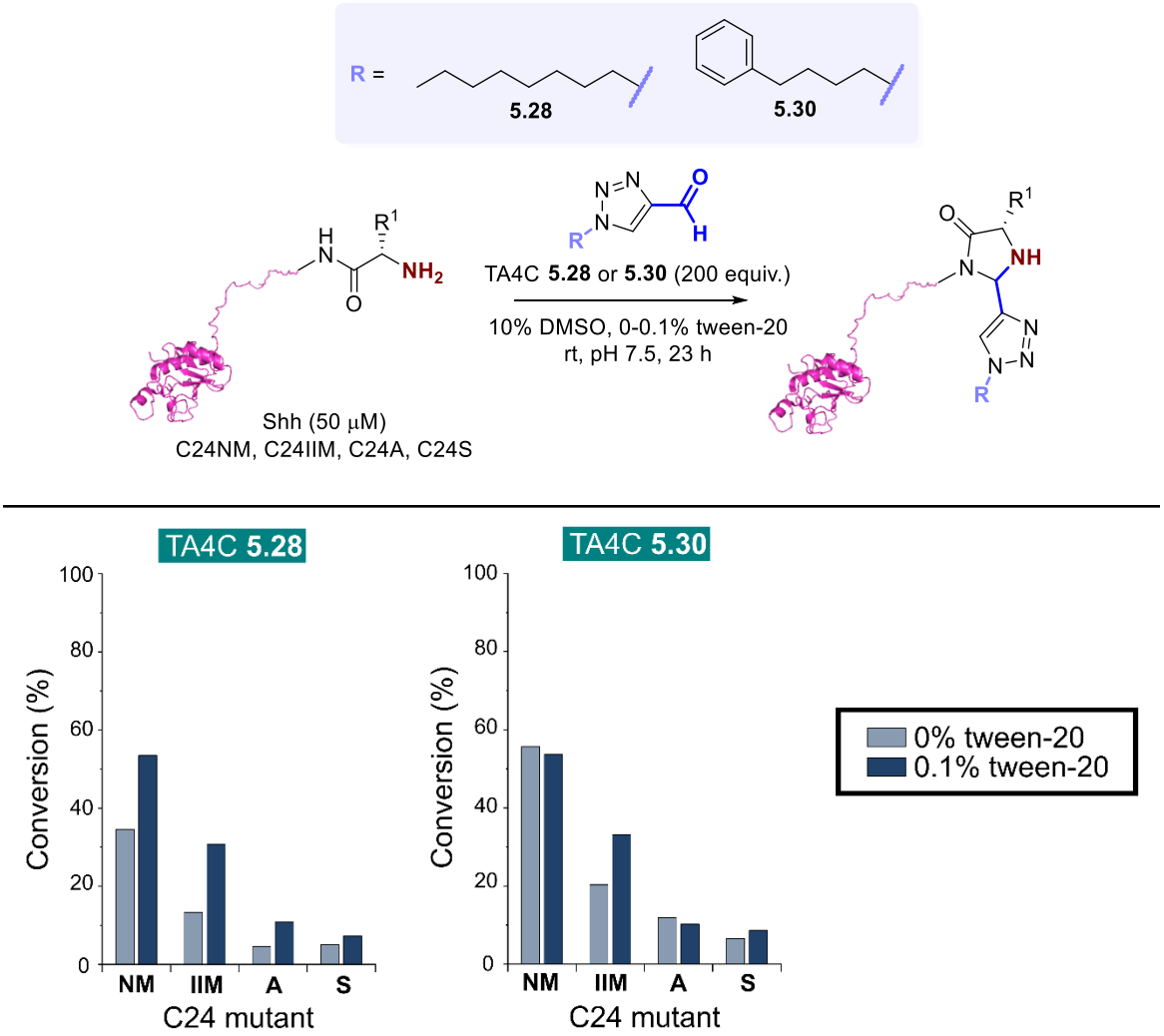
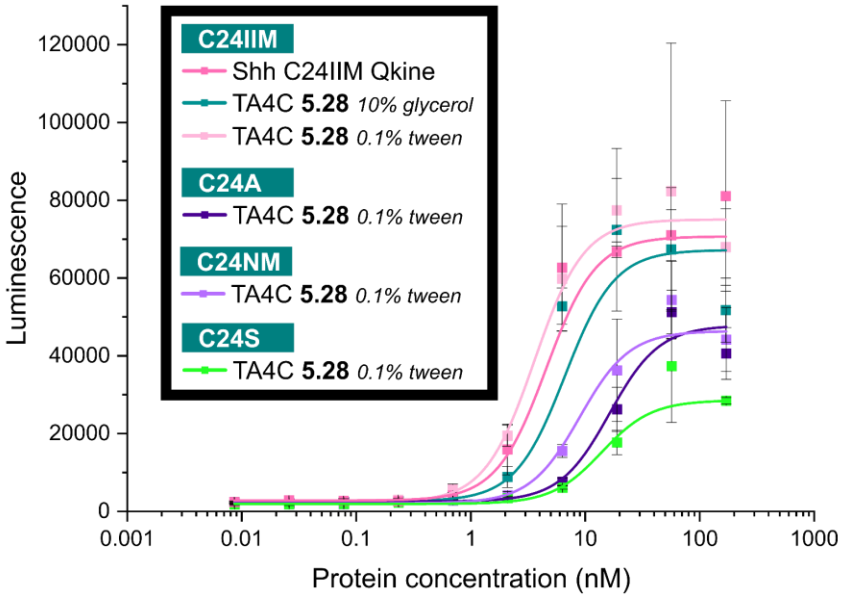
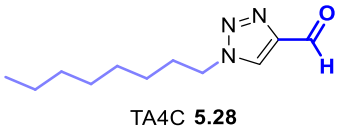


Figure 5.43. Application of optimised solubilising conditions to the modification of all *N*-terminal mutants with TA4C **5.28** and TA4C **5.30**, prepared *in situ*. Shh C24NM, C24IIM, C24A and C24S (50 µM) were modified with an aliquot of crude reagent **5.28** and **5.30** (200 equiv.) in Na phosphate buffer (50 mM, pH 7.5) with tween-20 (0 or 0.1%) at rt for 23 h, and the conversion was determined by LC-MS after purification by size exclusion chromatography. Only single modifications observed. Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

A Gli-luciferase assay was carried out to verify if the increase in conversions obtained through use of additives had any impact on the bioactivity (*Fig. 5.44*). Unfortunately, despite improved levels of modification, the influence of conjugation on bioactivity was minimal.

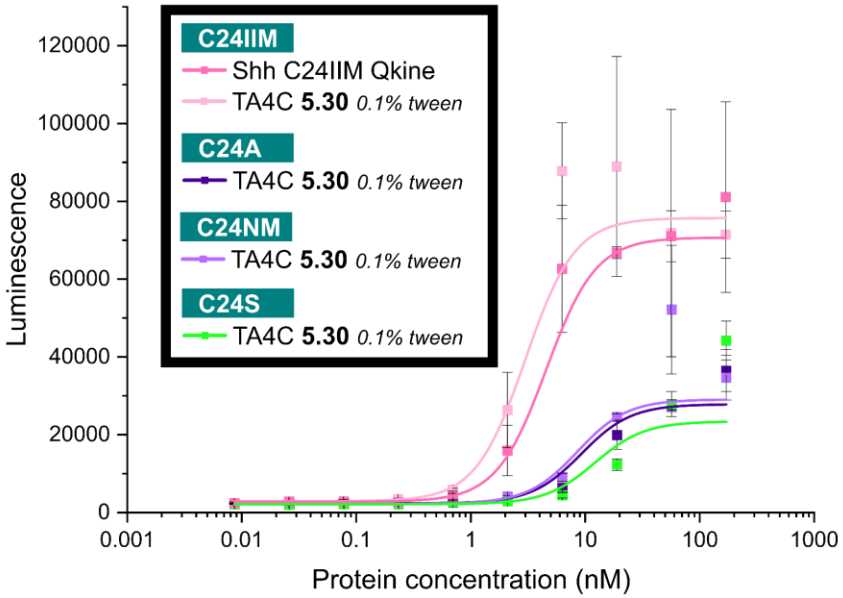
a)

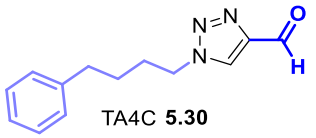




Mutant	TA4C	EC_{50} / nM
C24IIM	-	4.6 ± 1.7
C24IIM	**5.28** *(10% glycerol)*	6.7 ± 1.1
C24IIM	**5.28** *(0.1% tween)*	3.6 ± 0.7
C24A	**5.28** *(0.1% tween)*	16.3 ± 3.2
C24NM	**5.28** *(0.1% tween)*	9.0 ± 2.1
C24S	**5.28** *(0.1% tween)*	14.0 ± 1.4

b)





Mutant	TA4C	EC_{50} / nM
C24IIM	-	4.6 ± 1.7
C24IIM	**5.30** *(0.1% tween)*	3.0 ± 0.4
C24A	**5.30** *(0.1% tween)*	9.3 ± 1.0
C24NM	**5.30** *(0.1% tween)*	8.6 ± 1.1
C24S	**5.30** *(0.1% tween)*	12.2 ± 1.8

Figure 5.44. Gli-luciferase reporter assay for analysis of Shh bioactivity of TA4C: ***a)*** **5.28**; and ***b)*** **5.30** conjugates prepared with solubilising additives. *Passage number: 10*.

As outlined earlier in section 5.2.2.1, we suspected that the minimal differences in EC_{50} between modified and unmodified conjugates may have been due to poor conjugate stability at 37 °C in the presence of competitors. If this was the case, slightly increased conversion levels may have little impact on bioactivity. We therefore decided to test the stability of Shh/TA4C conjugates to temperature (37 °C) and competitive nucleophiles for 24 h to validate whether the modifications were robust enough to be tolerated for the duration of the Gli-luciferase assay.

As the TA4C with the most promising effect on activity, the stability of TA4C **5.30** conjugates was assessed. Due to the low conversions of Shh mutants with TA4C **5.30**, the stability of TA4C **5.22** conjugates was also probed, as an example of conjugates with higher conversions. Purified conjugates of Shh C24IIM, C24A, C24NM and C24S, prepared previously in *Fig. 5.38* and *5.43* for TA4Cs **5.22** and **5.30** respectively, were incubated at 37 °C for 24 h, both with and without 4 mM L-glutamine as an example of one of the competitive nucleophiles present in the cell culture medium (*Fig. 5.45*, *S5.18*). Initial conversions were lower than those observed in *Fig. 5.38*/*5.43*, perhaps due to storage and transfer of conjugates from Qkine (Cambridge) to York. Unfortunately, most TA4C **5.30** conjugates were unstable, with no/low levels of modified protein observed after 24 h. Stability of TA4C **5.22** conjugates was highly dependent on the N-terminal residue, with C24A and C24S conjugates showing much higher stability than C24IIM or C24NM (both N-terminal mutants with a retained Met).

Whilst stability experiments indicated that TA4C conjugates were not stable for the full duration of the protein treatment stage in the Gli-luciferase assay protocol (24 h, 37 °C), a study of the kinetics of the Shh response pathway by Wen *et al.* indicated that localisation of transcription factors upon activation of the Shh pathway occurred within 5 min.[30] This suggests that response elements further down the pathway, such as transcription, may be the steps with a longer duration, i.e. initial activation of the pathway is fast, so the stability of Shh conjugates for the full 24 h is neither relevant nor necessary. To probe this further, a kinetic study of the conjugate stability over time could be conducted: whilst not proven, we expect conjugate cleavage to be minimal over the first 5 min at 37 ° C.

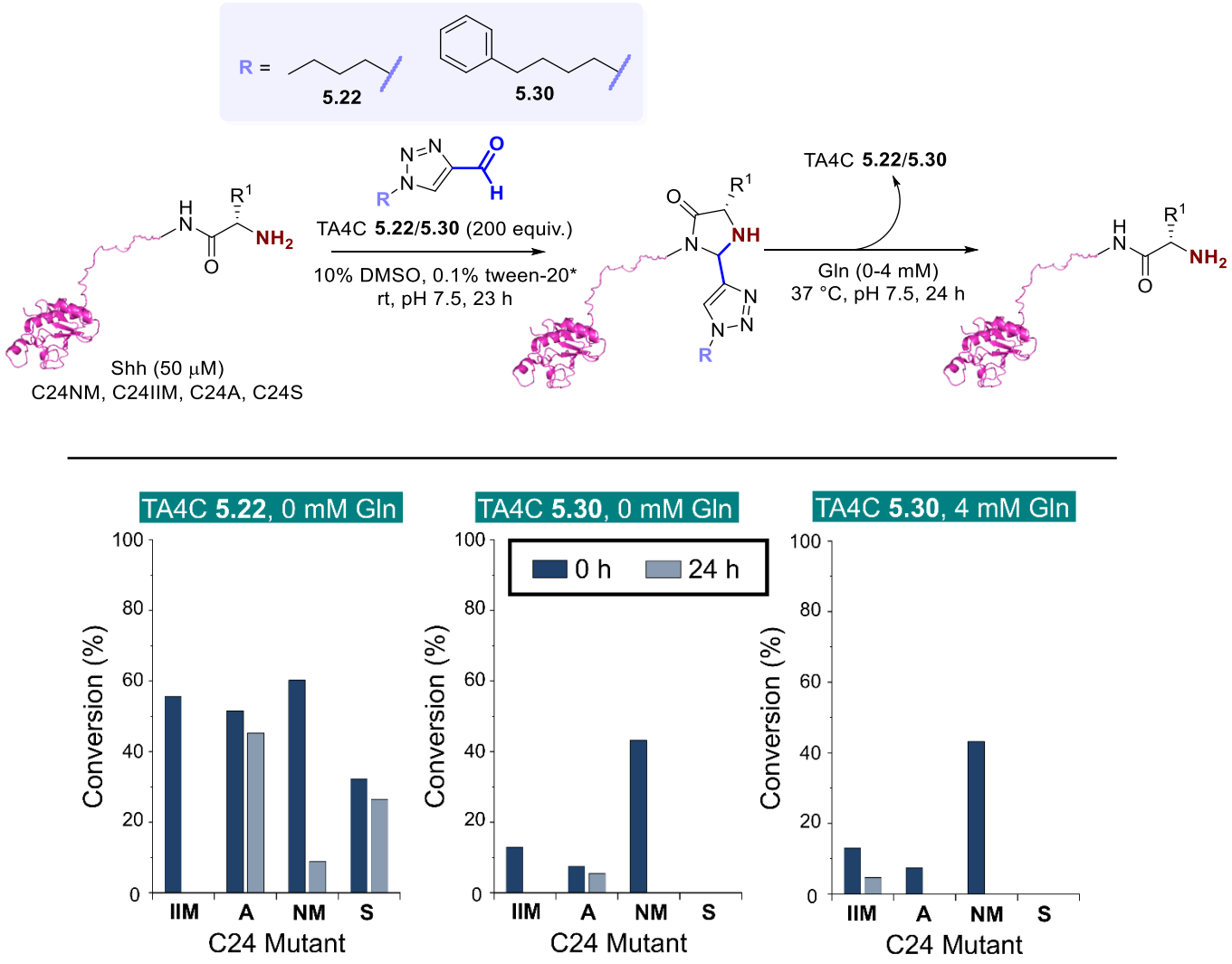


Figure 5.45. Stability of TA4C **5.22** and **5.30** conjugates to incubation at 37 °C with 0 and 4 mM L-glutamine. Shh C24NM, C24IIM, C24A and C24S (50 μM) were modified with an aliquot of crude reagent **5.22** and **5.30** (200 equiv.) in Na phosphate buffer (50 mM, pH 7.5) with tween-20 (0.1%, *for TA4C **5.30** only) at rt for 23 h, and the conversion was determined by LC-MS after purification by size exclusion chromatography (0 h). The conjugates were incubated at 37 °C with L-glutamine (0 and 4 mM), and conversion was determined by LC-MS after 24 h incubation. Only single modifications observed. Figure built using structural data obtained by Qi *et al.* (Shh, PDB 6OEV)[13].

5.3 Conclusion

Due to poor reproducibility of the Gli-luciferase assay, no conclusive interpretations of the bioactivity can be made. Any slight enhancements in activity cannot be consistently reproduced with confidence, and improvements on the level achieved by Taylor *et al.* (up to 160-fold increase in potency with respect to the WT without PTMs)[12] were not observed. In the continuation of this work, one of the first points to address is the cleavage of the N-terminal Met of the C24IIM mutant; minimal differences in EC_{50}s upon chemical modification of Shh C24IIM may have been due to spatial restrictions, as seen previously by Taylor *et al.* upon extension of the N-terminus of Shh C24II with additional hydrophobic amino acids.[12] As mentioned previously in Chapter 3, this could be attempted by co-expression of an engineered *E. coli* methionine aminopeptidase, developed by Liao *et al.*[31] Mutation of three residues in the binding pocket of the enzyme to smaller residues allows the engineered enzyme to cleave Met from proteins with bulky penultimate residues, such as Ile. With Ile exposed at the N-terminus, we hope for an enhancement in potency in the range observed by Taylor *et al.* (8-fold, rather than 2-fold)[12] in comparison to the WT (with no PTMs), hopefully large enough to be detected reproducibly by Gli-luciferase assay.

Our next priority is the development of a method to obtain reliable, reproducible readouts of conjugate bioactivity. The Gli-luciferase assay could be improved with transfection of the cells with a control reporter encoding for Renilla luciferase, to account for variability in cell count *via* normalisation. The passage protocol, in which digestive protease trypsin (250 µL) was diluted with DMEM/10% CS (9.75 mL), may have also contributed to assay variability as residual trypsin may have been able to degrade some of the Shh. In the development of future assays to evaluate Shh bioactivity, the passage protocol may be improved with a centrifugation step to ensure the removal of trypsin. Alternatively, detachment of cells with ethylenediaminetetraacetic acid (EDTA) could be attempted: EDTA is a calcium ion chelator and removes the calcium ions required for cell adhesion. An alternative assay could also be attempted, such as the alkaline phosphatase induction assay employed by Taylor *et al.* with the C3H10T1/2 cell line.[12]

The fundamental challenge limiting the impact of this work is the conflict between the aim to attach hydrophobic moieties and the requirement for the modifications to be water soluble, two concepts which are inherently contradictory. In future work, more extensive reagent design could be explored to include the incorporation of solubilising groups such as sulfonic acids. Enrichment of the levels of modified protein could also improve the bioactivity, for example through modification with a TA4C reagent with a cleavable biotin moiety to allow the separation of modified and unmodified protein.

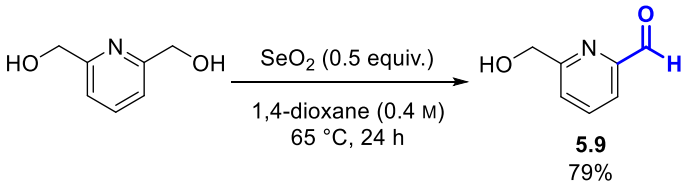
Whilst Onoda's Dimroth rearrangement protocol was a highly effective method for fast and efficient preparation of conjugates,[25] stability experiments indicated poor stability of conjugates at 37 °C. The suitability of Shh/TA4C conjugates for cell culture is therefore dependent on the length of time the conjugates are required to be kept under cell culture conditions: if activation time of the Shh pathway is fast, as suggested by Wen *et al.*,[30] prolonged stability of conjugates may not be a necessity.

5.4 Experimental

5.4.1 Synthesis

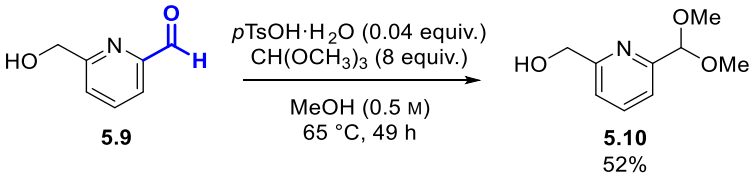
a) *2-PCA* ***5.7***

6-(Hydroxymethyl)-2-pyridinecarboxaldehyde (5.9)



Selenium dioxide (1.62 g, 14.6 mmol, 0.5 equiv.) was added to a solution of 2,6-pyridinedimethanol (4.00 g, 28.7 mmol, 1.0 equiv.) in 1,4-dioxane (80 mL, 0.4 M) at room temperature. The resulting mixture was sonicated for 2 min and then stirred at 65 °C for 24 h. After cooling to room temperature, CH_2Cl_2 (100 mL) was added, the resulting mixture was filtered through Celite, and the filtrate was concentrated under reduced pressure to afford a yellow oil. The residue was purified by flash column chromatography (2.5% MeOH:CH_2Cl_2, R_f 0.16), and pure fractions were concentrated under reduced pressure to afford the title compound (3.12 g, 22.8 mmol, 79%) as a yellow oil with spectroscopic data in accordance with the literature.[22] **^{1}H NMR** (300 MHz, $CDCl_3$) $\delta_{H:}$ 3.59 (1H, br s, -OH), 4.87 (2H, s, -CH_2-), 7.47-7.57 (1H, m, ArH4), 7.83-7.93 (2H, m, ArH3, ArH5), 10.08 (1H, s, -CHO).

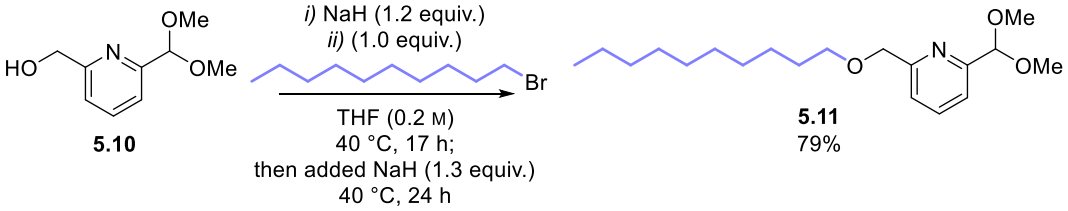
(6-(Dimethoxymethyl)pyridin-2-yl)methanol (5.10)



p-Toluenesulfonic acid monohydrate (174 mg, 0.9 mmol, 0.04 equiv.) and trimethyl orthoformate (20 mL, 182.2 mmol, 8.0 equiv.) were added to a solution of compound **5.9** (3.12 g, 22.8 mmol, 1.0 equiv.) in anhydrous methanol (40 mL, 0.5 M) under argon. The reaction mixture was heated under reflux for 49 h. After cooling to room temperature, the solvent was removed under reduced pressure to give a dark green liquid. The liquid was partitioned between dichloromethane (100 mL) and saturated aqueous $NaHCO_3$ solution (40 mL). The aqueous layer was extracted with dichloromethane (3 × 50 mL). The combined organic layers were dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The brown residue obtained was purified by flash column chromatography (50% EtOAc:Petrol, R_f 0.26) to afford the title compound (2.18 g, 11.9 mmol, 52%) as a pale pink liquid. **^{1}H NMR** (300 MHz, $CDCl_3$) $\delta_{H:}$ 3.42 (6H, s, OMe), 4.79 (2H, s, -CH_2-), 5.38 (1H, s, -CH(OMe)$_2$), 7.24 (1H, d, J = 7.7 Hz, ArH3), 7.48 (1H, d, J = 7.7 Hz, ArH5), 7.75 (1H, dd, J_1 = J_2 = 7.7 Hz, ArH4); **^{13}C NMR** (75 MHz, $CDCl_3$) $\delta_{C:}$ 53.89 (OMe), 64.29 (-CH_2-), 104.07 (-CH(OMe)$_2$), 119.96

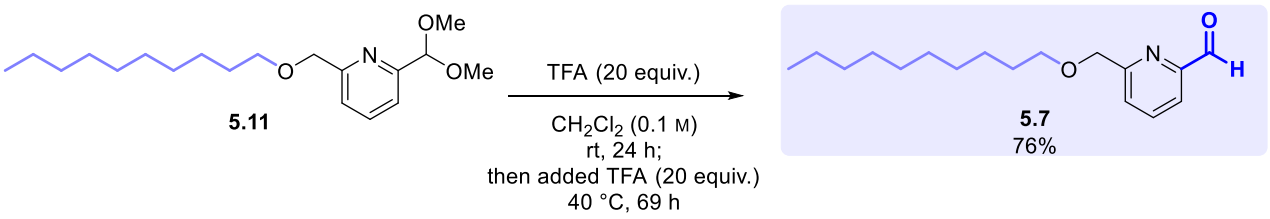
(ArC5), 120.58 (ArC3), 137.41 (ArC4), 156.30 (ArC6), 158.84 (ArC2); **λmax** (ATR, cm^{-1}) 2937, 1723, 1594, 1460, 1353, 1268, 1193, 1155, 1109, 1053, 989, 932, 898, 839, 764, 687, 644, 623, 562, 422; **HRMS** (ESI^+) $C_9H_{13}NNaO_3$ $[M+Na]^+$ found 206.0790, calculated 206.0788.

2-((Decyloxy)methyl)-6-(dimethoxymethyl)pyridine (5.11)



Sodium hydride (60%, 53 mg, 0.13 mmol, 1.2 equiv.) was added slowly to a solution of **5.10** (198 mg, 0.11 mmol, 1.0 equiv.) in anhydrous THF (5.5 mL, 0.2 M), and after 50 min 1-bromodecane (0.23 mL, 0.11 mmol, 1.0 equiv.) was added. The reaction mixture was then stirred at 40 °C for 17 h under nitrogen. At this point TLC indicated that the reaction was incomplete, so another portion of sodium hydride (60%, 58 mg, 0.14 mmol, 1.3 equiv.) was added, and the reaction mixture was stirred at 40 °C for 24 h. After cooling to room temperature, the reaction was quenched through careful dropwise addition of water (10 mL), and the aqueous extracted with ethyl acetate (3 × 30 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The crude product was purified by flash column chromatography (10% EtOAc:Petrol, R_f 0.93 in 66% EtOAc). Pure fractions were concentrated under reduced pressure to afford the title compound (280 mg, 0.86 mmol, 79%) as a colourless liquid. **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H: 0.87 (3H, t, *J* = 6.8 Hz, -CH_2CH_3), 1.22-1.32 (12H, m, -CH_2-), 1.34-1.42 (2H, m, $ArCH_2OCH_2CH_2CH_2$-), 1.64 (2H, dd, J_1 = J_2 = 6.6 Hz, $ArCH_2OCH_2CH_2$), 3.40 (6H, s, OMe), 3.55 (2H, t, *J* = 6.7 Hz, $ArCH_2OCH_2$-), 4.66 (2H, s, $ArCH_2O$-), 5.34 (1H, s, $ArCH(OMe)_2$), 7.44 (2H, m, ArH3 & ArH5), 7.74 (1H, t, *J* = 7.7 Hz, ArH4); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C: 14.26 (-CH_3), 22.82 (-CH_2-), 26.31 (-CH_2-), 29.46 (-CH_2-), 29.63 (-CH_2-), 29.71 (-CH_2-), 29.75 (-CH_2-), 29.90 (-CH_2-), 32.04 (-CH_2-), 53.91 (OMe), 71.40 ($ArCH_2OCH_2$-), 73.70 ($ArCH_2O$-), 104.25 ($ArCH(OMe)_2$), 119.84 (ArC5), 121.06 (ArC3), 137.44 (ArC4), 156.47 (ArC6), 158.84 (ArC2); **λmax** (ATR, cm^{-1}) 2925, 2855, 1722, 1593, 1459, 1347, 1304, 1193, 1111, 1060, 989, 765, 642; **HRMS** (ESI^+) $C_{19}H_{33}NNaO_3$ $[M+Na]^+$ found 346.2360, calculated 346.2353.

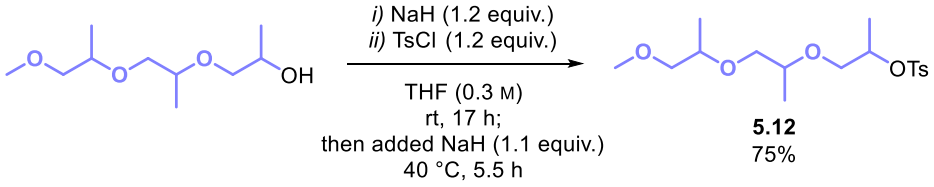
6-((Decyloxy)methyl)picolinaldehyde (5.7)



TFA (240 µL, 3.1 mmol, 20 equiv.) was added to a solution of compound **5.11** (50 mg, 0.2 mmol, 1.0 equiv.) in dichloromethane (2 mL, 0.1 M), and the reaction mixture was stirred at rt for 24 h. At this point TLC indicated that the reaction was incomplete, so another portion of TFA (240 µL, 3.1 mmol, 20 equiv.) was added, and the reaction mixture was stirred at 40 °C for 69 h. The reaction mixture was allowed to cool to room temperature, the solvent was removed under reduced pressure, and the remaining yellow liquid was azeotroped with dichloromethane (3 × 20 mL). The crude product was purified by flash column chromatography (20% EtOAc:Petrol, R_f 0.59). Pure fractions were concentrated under reduced pressure to afford the title compound (33 mg, 0.12 mmol, 76%) as a colourless liquid. **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H: 0.85-0.90 (3H, m, -C$\underline{H}_3$), 1.24-1.34 (12H, m, -C$\underline{H}_2$-), 1.34-1.44 (2H, m, ArCH$_2$OCH$_2$CH$_2$C$\underline{H}_2$-), 1.61-1.72 (2H, m, ArCH$_2$OCH$_2$C$\underline{H}_2$-), 3.59 (2H, t, *J* = 6.6 Hz, ArCH$_2$OC$\underline{H}_2$-), 4.71 (2H, s, ArC$\underline{H}_2$O-), 7.70 (1H, dd, *J* = 1.9, 7.0 Hz, Ar$\underline{H}$5), 7.82-7.93 (2H, m, Ar$\underline{H}$3 & Ar$\underline{H}$4), 10.05 (1H, s, CO$\underline{H}$); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C: 14.26 (-$\underline{C}$H$_3$), 22.82 (-$\underline{C}$H$_2$-), 26.30 (-ArCH$_2$OCH$_2$CH$_2$$\underline{C}H_2$-), 29.46 (-$\underline{C}H_2$-), 29.61 (-$\underline{C}H_2$-), 29.71 (-$\underline{C}H_2$-), 29.74 (-$\underline{C}H_2$-), 29.85 (-$\underline{C}H_2$-), 32.03 (-$\underline{C}H_2$-), 71.63 (ArCH$_2O\underline{C}H_2$-), 73.38 (Ar$\underline{C}H_2$O-), 120.52 (Ar$\underline{C}$3 or Ar$\underline{C}$4), 125.68 (Ar$\underline{C}$5), 137.73 (Ar$\underline{C}$3 or Ar$\underline{C}$4), 152.20 (Ar$\underline{C}$2), 160.16 (Ar$\underline{C}$6), 193.63 ($\underline{C}$OH); **$\lambda_{max}$** (ATR, cm^{-1}); 2922, 2852, 1713, 1594, 1454, 1349, 1265, 1210, 1122, 991, 784, 718, 665, 620; **HRMS** (ESI$^+$) $C_{17}H_{28}NO_2$ $[M+H]^+$ found 278.2120, calculated 278.2115.

b*) *2-PCA* *5.8

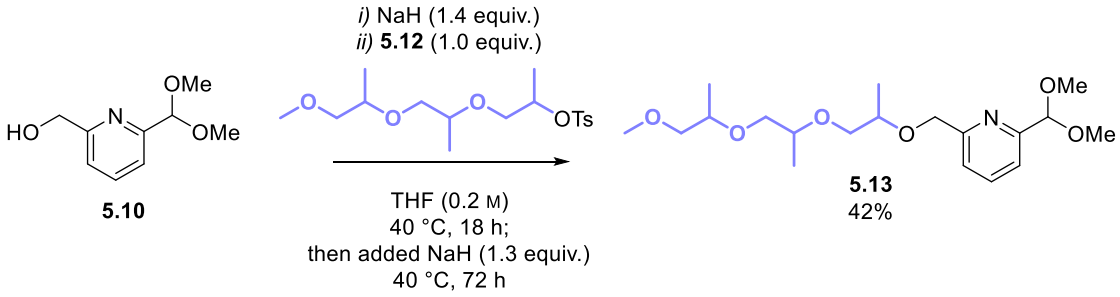
1-((1-((1-Methoxypropan-2-yl)oxy)propan-2-yl)oxy)propan-2-yl 4-methylbenzenesulfonate (5.12)



Sodium hydride (60%, 0.24 g, 5.9 mmol, 1.2 equiv.) was added slowly to a solution of 1-((1-((1-methoxypropan-2-yl)oxy)propan-2-yl)oxy)propan-2-ol (1.01 g, 4.9 mmol, 1.0 equiv.) in anhydrous THF (15 mL, 0.3 M), and after 40 min 4-toluenesulfonyl chloride (1.15 g, 6.0 mmol, 1.2 equiv.) was added. The reaction mixture was then stirred at room temperature for 17 h under nitrogen. At this point TLC indicated that the reaction was incomplete, so another portion of sodium hydride (60%, 0.21 g, 5.3 mmol, 1.1 equiv.) was added, and the reaction mixture was stirred at 40 °C for 5.5 h. After cooling to room temperature, the reaction was quenched through careful dropwise addition of water (15 mL), and the aqueous extracted with ethyl acetate (3 × 30 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The crude product was purified by flash column chromatography (25% EtOAc:Petrol, R_f 0.25). Pure fractions were concentrated under reduced pressure to afford the title compound (1.33 g, 3.7 mmol, 75%) as a colourless liquid. **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H: 0.99-1.07 (3H, m, -CH$\underline{Me}$), 1.07-1.19 (3H, m, -CH$\underline{Me}$), 1.27 (3H, dt, *J* = 6.4, 0.7 Hz, -CH$\underline{Me}$OTs), 2.44 (3H, s, Ar$\underline{Me}$), 3.26-3.47 (9H, m, 2 × -OC$\underline{H}_2$, 2 × -C$\underline{H}$Me, O$\underline{Me}$), 3.48-3.60 (2H, m, -C$\underline{H}_2$CHMeOTs), 4.60-4.72 (1H, m, -C$\underline{H}$MeOTs), 7.32 (2H, d, *J* = 8.3 Hz, -

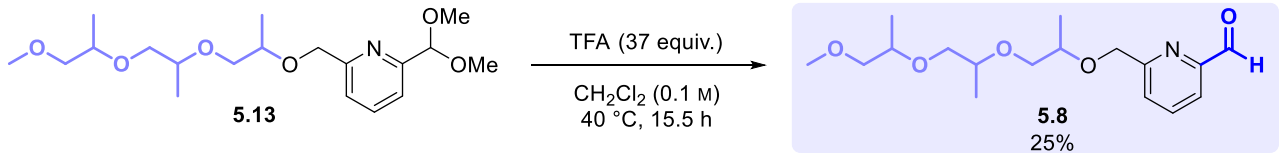
SO_2CCHCH-), 7.80 (2H, d, *J* = 8.3 Hz, -SO_2CCH-); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C: 16.39 (-CHMe), 17.03 (-CHMe), 17.76 (-CHMeOTs), 21.66 (ArMe), 56.74 (-CHMe or OMe), 59.18 (-CHMe or OMe), 71.64 (-CH_2CHMeOTs), 75.02 (-OCH_2), 75.45 (-OCH_2), 78.72 (-CHMeOTs), 127.89 (-SO_2CCH-), 129.66 (-SO_2CCHCH-), 134.39 (-SO_2C-), 144.46 (-$SO_2C(CH)_2$CMe); **λ_{max}** (ATR, cm^{-1}) 2976, 2878, 1599, 1453, 1354, 1189, 1175, 1097, 1020, 923, 901, 815, 769, 664, 555; **HRMS** (ESI^+) $C_{17}H_{28}NaO_6S$ $[M+Na]^+$ found 383.1485, calculated 383.1499.

2-(Dimethoxymethyl)-6-(3,6,9-trimethyl-2,5,8,11-tetraoxadodecyl)pyridine (5.13)



Sodium hydride (60%, 63 mg, 1.6 mmol, 1.4 equiv.) was added slowly to a solution of compound **5.10** (210 mg, 1.1 mmol, 1.0 equiv.) in anhydrous THF (5.5 mL, 0.2 M), and after 20 min compound **5.12** (418 mg, 1.2 mmol, 1.0 equiv.) was added. The reaction mixture was then stirred at 40 °C for 18 h under nitrogen. At this point TLC indicated that the reaction was incomplete, so another portion of sodium hydride (60%, 61 mg, 1.5 mmol, 1.3 equiv.) was added, and the reaction mixture was stirred at 40 °C for 72 h. After cooling to room temperature, the reaction was quenched through careful dropwise addition of water (15 mL), and the aqueous extracted with ethyl acetate (3 × 30 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The crude product was purified by flash column chromatography (80% EtOAc:Petrol, R_f 0.26 in 50% EtOAc:Petrol). Pure fractions were concentrated under reduced pressure to afford the title compound (178 mg, 0.5 mmol, 42%) as a colourless liquid. **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H: 1.07-1.18 (6H, m, -CHMeCH_2OCHMeCH_2OMe), 1.21 (3H, d, *J* = 6.3 Hz, -CHMeOCH_2Ar), 3.17-3.57 (15H, m, -CHMeCH_2OCHMeCH_2OMe and ArCH(OMe)$_2$), 3.58-3.69 (2H, m, -CH_2CHMeOCH_2Ar), 3.71-3.80 (1H, m, -CHMeOCH_2Ar), 4.68-4.83 (2H, m, -CH_2Ar), 5.35 (1H, s, ArCH(OMe)$_2$), 7.42 (1H, d, *J* = 7.7 Hz, ArH3), 7.52 (1H, dd, *J* = 7.7, 3.0 Hz, ArH5), 7.74 (1H, dd, J_1 = J_2 = 7.7 Hz, ArH4); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C: 16.45 (-CHMeCH_2OMe or -CHMeCH_2OCHMeCH_2OMe), 17.13 (CHMeOCH_2Ar), 17.17 (-CHMeCH_2OMe or -CHMeCH_2OCHMeCH_2OMe), 53.74 (OMe and -CHMeCH_2OMe, -CH_2OMe, or -CH_2OMe), 56.74 (-CHMeCH_2OMe, -CH_2OMe, or -CH_2OMe), 59.18 (-CHMeCH_2OMe, -CH_2OMe, or -CH_2OMe), 71.64 (-CH_2Ar), 72.99 (-CH_2CHMeOCH_2Ar), 75.11 (-CHMeOCH_2Ar), 75.35 (-CHMeCH_2OCHMeCH_2OMe), 75.88 (-CH_2OCHMeCH_2OMe), 103.93 (ArCH(OMe)$_2$), 119.74 (ArC3), 121.09 (ArC5), 137.40 (ArC4), 156.11 (ArC2 and ArC6); **λ_{max}** (ATR, cm^{-1}) 2912, 1725, 1591, 1456, 1350, 1292, 1220, 1107, 1058, 991, 934, 839, 765, 690, 643; **HRMS** (ESI^+) $C_{19}H_{33}NNaO_6$ $[M+Na]^+$ found 394.2205, calculated 394.2200.

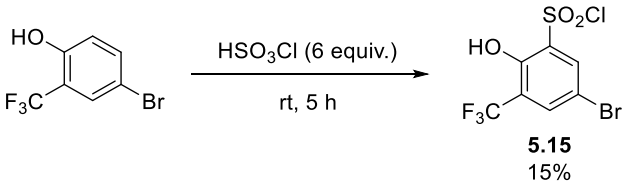
6-(3,6,9-Trimethyl-2,5,8,11-tetraoxadodecyl)picolinaldehyde (5.8)



TFA (200 µL, 2.6 mmol, 37 equiv.) was added to a solution of compound **5.13** (26 mg, 0.1 mmol, 1.0 equiv.) in dichloromethane (1.3 mL, 0.1 M), and the reaction mixture was stirred at 40 °C for 15.5 h. The reaction mixture was allowed to cool to room temperature, the solvent was removed under reduced pressure, and the remaining yellow liquid was azeotroped with dichloromethane (3 × 15 mL). The crude product was purified by flash column chromatography (50% EtOAc:Petrol, R_f 0.48). Pure fractions were concentrated under reduced pressure to afford the title compound (6 mg, 0.02 mmol, 25%) as a colourless liquid. **^{1}H NMR** (400 MHz, $CDCl_3$) $\delta_{H:}$ 1.08-1.19 (6H, m, 2 × Me), 1.20-1.37 (6H, m, 2 × Me), 3.22-3.59 (6H, m, -CHMeCH$_2$OCHMeCH$_2$OMe), 3.60-3.71 (2H, m, -CH$_2$CHMeOCH$_2$Ar), 3.74-3.86 (1H, m, -CHMeOCH$_2$Ar), 4.84 (2H, s, -CH$_2$Ar), 7.76-7.82 (1H, m, ArH), 7.83-7.91 (2H, m, 2 × ArH), 10.04 (1H, s, CHO); **λ_{max}** (ATR, cm^{-1}) 2930, 1755, 1459, 1348, 1166, 1067, 1035, 926, 847, 761, 658, 491; **HRMS** (ESI$^+$) $C_{17}H_{27}NNaO_5$ $[M+Na]^+$ found 348.1790, calculated 348.1781.

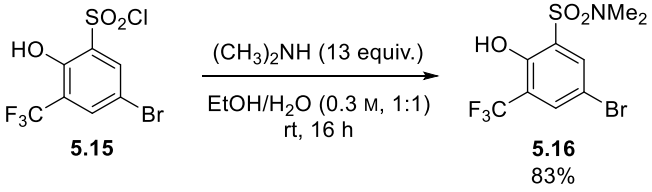
c) *Activated phenol ester* ***5.14***

5-Bromo-2-hydroxy-3-(trifluoromethyl)benzenesulfonyl chloride (5.15)



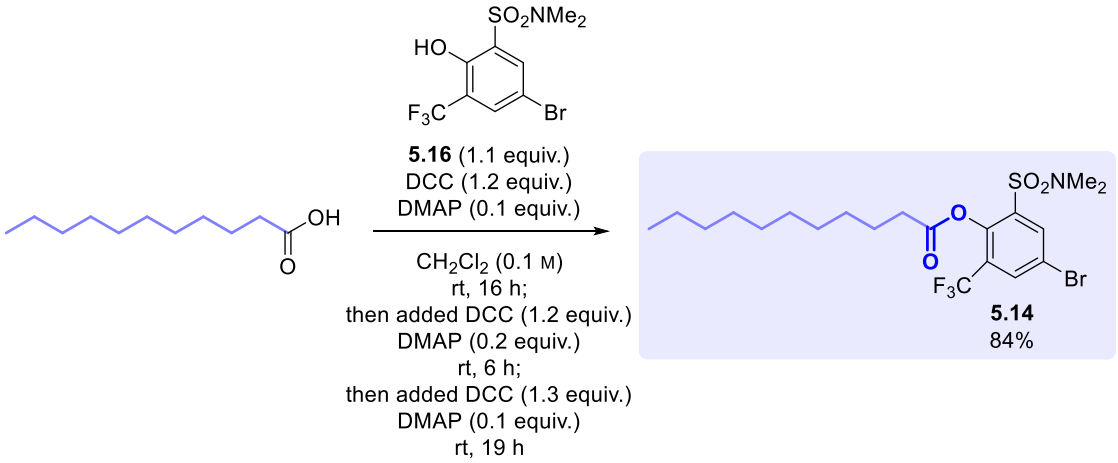
4-Bromo-2-(trifluoromethyl)phenol (2.01 g, 8.4 mmol, 1.0 equiv.) was added portion-wise to chlorosulfonic acid (3.3 mL, 49.8 mmol, 6.0 equiv.) at 0 °C. The reaction mixture was stirred at room temperature for 5 h, then poured onto crushed ice (50 mL). The resulting precipitate was collected by filtration, washed with cold water (3 × 15 mL), and dried under vacuum to afford the title compound (0.44 g, 1.3 mmol, 15%) as a white solid with spectroscopic data in accordance with the literature.[23] **^{1}H NMR** (400 MHz, DMSO-d_6) $\delta_{H:}$ 7.74 (1H, d, J = 2.5 Hz, ArH4 or ArH6), 7.78 (1H, d, J = 2.5 Hz, ArH4 or ArH6); **^{19}F NMR** (376 MHz, DMSO-d_6) $\delta_{F:}$ -61.18 (s, CF$_3$).

5-Bromo-2-hydroxy-N,N-dimethyl-3-(trifluoromethyl)benzenesulfonamide (5.16)



Compound **5.15** (400 mg, 1.2 mmol, 1.0 equiv.) was added in small portions to a solution of dimethylamine (40% in water, 2 mL, 15.8 mmol, 13.4 equiv.) in ethanol (2 mL, 0.3 M). The reaction mixture was stirred at room temperature for 16 h, then acidified to pH 1 using concentrated HCl, and the aqueous was extracted with ethyl acetate (50 mL). The organic layer was washed with brine (2 × 15 mL), dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The crude product was purified by flash column chromatography (30% EtOAc:Petrol, R_f 0.33). Pure fractions were concentrated under reduced pressure to afford the title compound (340 mg, 1.0 mmol, 83%) as an orange solid with spectroscopic data in accordance with the literature.[23] **^{1}H NMR** (400 MHz, DMSO-d_6) $\delta_{H:}$ 2.77 (6H, s, Me), 8.00 (1H, d, J = 2.5 Hz, ArH4 or ArH6), 8.07 (1H, d, J = 2.5 Hz, ArH4 or ArH6); **^{19}F NMR** (376 MHz, DMSO-d_6) $\delta_{F:}$ -60.70 (s, CF_3).

4-Bromo-2-(*N*,*N*-dimethylsulfamoyl)-6-(trifluoromethyl)phenyl undecanoate (5.14)

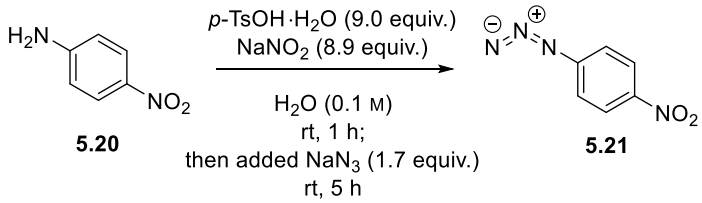


DCC (100 mg, 0.5 mmol, 1.2 equiv.) and DMAP (6 mg, 0.05 mmol, 0.1 equiv.) were added to a solution of undecanoic acid (74 mg, 0.4 mmol, 1.0 equiv.) in dichloromethane (4 mL, 0.1 M), and the reaction mixture was stirred at room temperature for 15 min. Compound **5.16** (150 mg, 0.4 mmol, 1.1 equiv.) was added in small portions, and the reaction mixture was stirred at room temperature for 16 h. At this point TLC indicated that the reaction was incomplete, so another portion of DCC (100 mg, 0.5 mmol, 1.2 equiv.) and DMAP (7 mg, 0.06 mmol, 0.2 equiv.) was added, and the reaction mixture was stirred at room temperature for 6 h. At this point TLC indicated that the reaction was incomplete, so another portion of DCC (102 mg, 0.5 mmol, 1.3 equiv.) and DMAP (5 mg, 0.04 mmol, 0.1 equiv.) was added, and the reaction mixture was stirred at room temperature for 19 h. The reaction mixture was filtered and the filtrate was concentrated under reduced

pressure. The crude product was purified by flash column chromatography (10% EtOAc:Petrol, R_f 0.26). Pure fractions were concentrated under reduced pressure to afford the title compound (172 mg, 0.3 mmol, 84%) as a colourless liquid. **^{1}H NMR** (400 MHz, $CDCl_3$) $\delta_{H:}$ 0.87 (3H, t, *J* = 7.0 Hz, -CH_2Me), 1.19-1.44 (14H, m, -$CH_2CH_2CH_2CH_2CH_2CH_2CH_2$Me), 1.72 (2H, tt, $J_1 = J_2$ = 7.3 Hz, -CH_2CH_2COOAr), 2.62 (2H, app td, *J* = 7.3, 2.5 Hz, -CH_2COOAr), 2.81 (6H, s, NMe_2), 7.99 (1H, d, *J* = 2.4 Hz, ArH3 or ArH5), 8.20 (1H, d, *J* = 2.4 Hz, ArH3 or ArH5); **^{13}C NMR** (101 MHz, $CDCl_3$) $\delta_{C:}$ 14.26 (-CH_2Me), 22.82 (-CH_2-), 24.33 (-CH_2CH_2COOAr), 29.04 (-$CH_2CH_2CH_2$COOAr), 29.35 (-CH_2-), 29.44 (-CH_2-), 29.56 (-CH_2-), 29.68 (-CH_2-), 32.03 (-CH_2-), 34.18 (-CH_2COOAr), 37.77 (NMe_2), 119.60 (ArC4), 134.53 (ArC3 or ArC5), 137.13 (ArC3 or ArC5), 145.84 (ArC1), 170.97 (-COOAr) *(note that CF_3, ArC6 and ArC2 signals were not observed)*; **^{19}F NMR** (376 MHz, $CDCl_3$) $\delta_{F:}$ -61.39 (s, CF_3); **λ_{max}** (ATR, cm^{-1}) 2924, 2854, 2119, 1788, 1702, 1445, 1352, 1305, 1263, 1224, 1144, 1075, 965, 905, 848, 821, 739, 700, 577, 521, 484; **HRMS** (ESI^+) $C_{20}H_{29}BrF_3NNaO_4S$ $[M+Na]^+$ found 538.0859, calculated 538.0845.

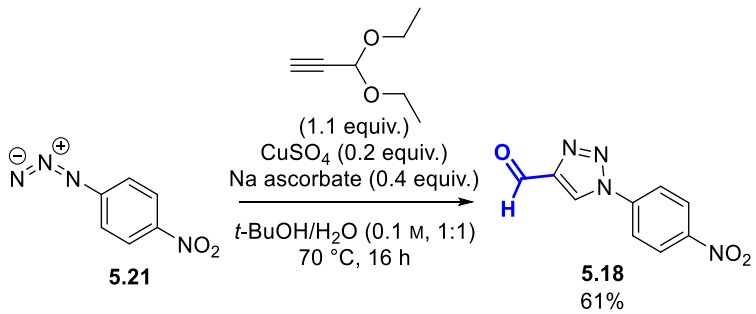
d) *TA4C* ***5.22***

1-Azido-4-nitrobenzene (5.21)



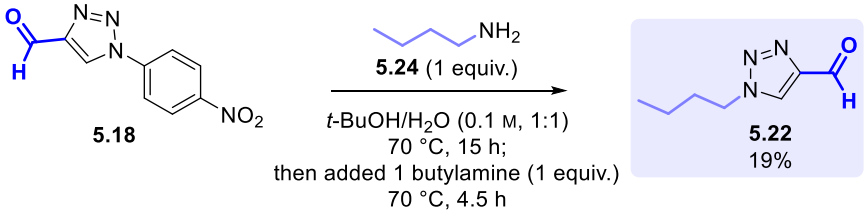
4-nitroaniline (502 mg, 3.6 mmol, 1.0 equiv.) was added to a solution of *p*-toluenesulfonic acid monohydrate (6.22 g, 32.7 mmol, 9.0 equiv.) in water (32 mL, 0.1 M), and the reaction mixture was stirred at room temperature for 1 min. Sodium nitrite (2.23 g, 32.3 mmol, 8.9 equiv.) was added over 5 min, and the reaction mixture was stirred at room temperature for 1 h. Sodium azide (0.40 g, 6.1 mmol, 1.7 equiv.) was added slowly, and the reaction mixture was stirred at room temperature for 30 min. The reaction mixture was filtered and the precipitate was washed with water (50 mL) and dried under reduced pressure. The pale-yellow solid obtained (0.54 g) was used immediately in the subsequent reaction without characterisation.

1-(4-Nitrophenyl)-1'*H*-1',2',3'-triazole-4'-carbaldehyde (5.18)



$CuSO_4$ (83 mg, 0.5 mmol, 0.2 equiv.), L-ascorbic acid sodium salt (196 mg, 1.0 mmol, 0.4 equiv.), compound **5.21** (403 mg, 2.5 mmol, 1.0 equiv.) and propargyl aldehyde diethyl acetal (400 µL, 2.8 mmol, 1.1 equiv.) were dissolved in a 1:1 mixture of *t*-BuOH and water (25 mL, 0.1 M), and the reaction mixture was stirred at 70 °C for 16 h. After cooling to rt, NH_4OH (25 mL, 2 M) was added, and the reaction mixture was extracted with dichloromethane (4 × 75 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The crude product was purified by flash column chromatography (0-100% EtOAc:Petrol, R_f 0.27 in 25% EtOAc:Petrol). Pure fractions were concentrated under reduced pressure to afford the title compound (328 mg, 1.5 mmol, 61%) as a yellow solid with spectroscopic data in accordance with the literature.[32] **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H: 7.99-8.08 (2H, m, ArH2, ArH6), 8.43-8.52 (2H, m, ArH3, ArH5), 8.65 (1H, s, ArH5'), 10.25 (1H, s, COH).

1-Butyl-1*H*-1,2,3-triazole-4-carbaldehyde (5.22)



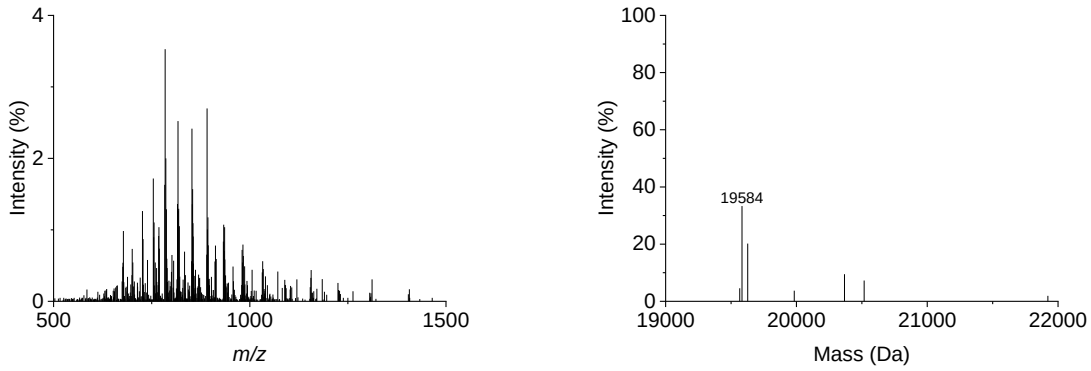
Butylamine **5.24** (23 µL, 0.2 mmol, 1.0 equiv.) was added to a solution of compound **5.18** (51 mg, 0.2 mmol, 1.0 equiv.) in a 1:1 mixture of *t*-BuOH and water (2 mL, 0.1 M), and the reaction mixture was stirred at 70 °C for 15 h under nitrogen. At this point TLC indicated that the reaction was incomplete, so another portion of butylamine **5.24** (23 µL, 0.2 mmol, 1.0 equiv.) was added, and the reaction mixture was stirred at 70 °C for 4.5 h. The reaction mixture was allowed to cool to room temperature, diluted with HCl (25 mL, 1 M), and extracted with ethyl acetate (3 × 20 mL). The combined organic layers were dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The crude product was purified by flash column chromatography (15% EtOAc:Petrol, R_f 0.28). Pure fractions were concentrated under reduced pressure to afford the title compound (7 mg, 0.04 mmol, 19%) as a yellow residue. **^{1}H NMR** (400 MHz, DMSO-d_6) δ_H: 0.89 (3H, t, J = 7.4 Hz, Me), 1.18-1.32 (4H, m, 2 × -CH$_2$-), 1.78-1.88 (2H, m, -CH$_2$-), 8.89 (1H, s, ArH), 10.01 (1H, s, CHO); **λ_{max}** (ATR, cm^{-1}) 2962, 1745, 1696, 1599, 1531, 1465, 1345, 1234, 1138, 1042, 854, 781, 752, 642, 600; **HRMS** (ESI$^+$) $C_7H_{11}N_3NaO$ $[M+Na]^+$ found 176.0796, calculated 176.0794.

5.4.2 Protein modification

Protein expression and purification of Shh WT, Shh C24IIM, Shh C24A, Shh C24NM and Shh C24S was carried out by Dr. Helen Bell at Qkine.

Shh WT

Theoretical M_w: 19560 Da



Shh WT

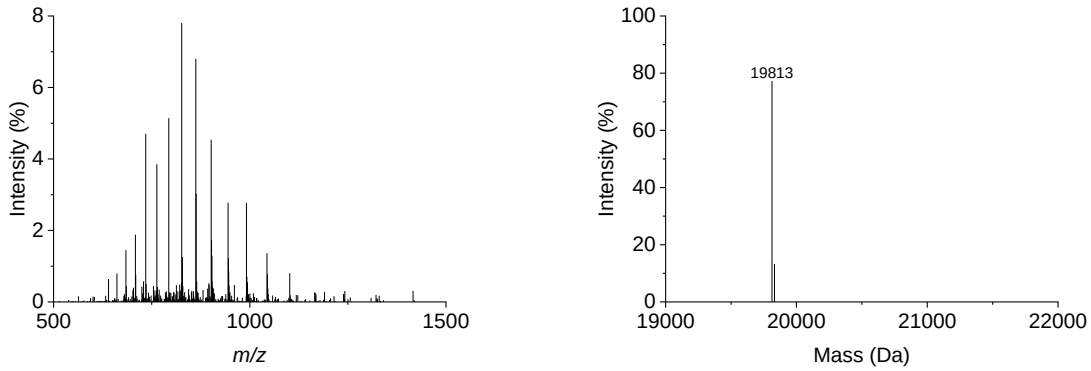
MS (ESI$^+$) [**WT**+Na]$^+$ found 19584, calculated 19584.

Figure S5.1. Raw (left) and deconvoluted (right) mass spectra for Shh WT from Qkine.

Shh C24II

Theoretical M_w: 19683 Da

Theoretical M_w (+ retained Methionine): 19814 Da



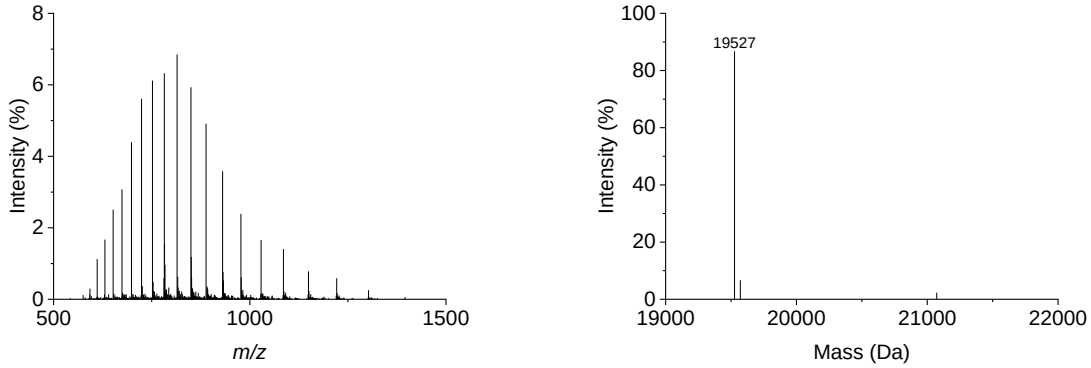
Shh C24IIM

MS (ESI$^+$) [**C24IIM**+H]$^+$ found 19813, calculated 19815; [**C24IIM**+H+H_2O]$^+$ found 19832, calculated 19833; [**C24IIM**+H+MeCN]$^+$ found 19856, calculated 19856.

Figure S5.2. Raw (left) and deconvoluted (right) mass spectra for Shh C24II from Qkine.

Shh C24A

Theoretical M_w: 19528 Da



Shh C24A

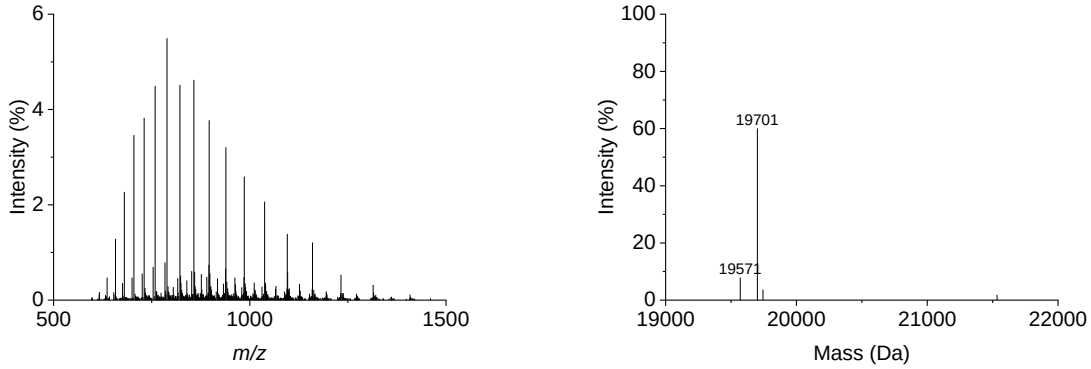
MS (ESI$^+$) [**C24A**+H]$^+$ found 19527, calculated 19529; [**C24A**+H+MeCN]$^+$ found 19570, calculated 19570.

Fig. S5.3. Raw (left) and deconvoluted (right) mass spectra for Shh C24A from Qkine.

Shh C24N

Theoretical M_w: 19571 Da

Theoretical M_w (+ retained Methionine): 19702 Da



Shh C24N

C24N (11%)

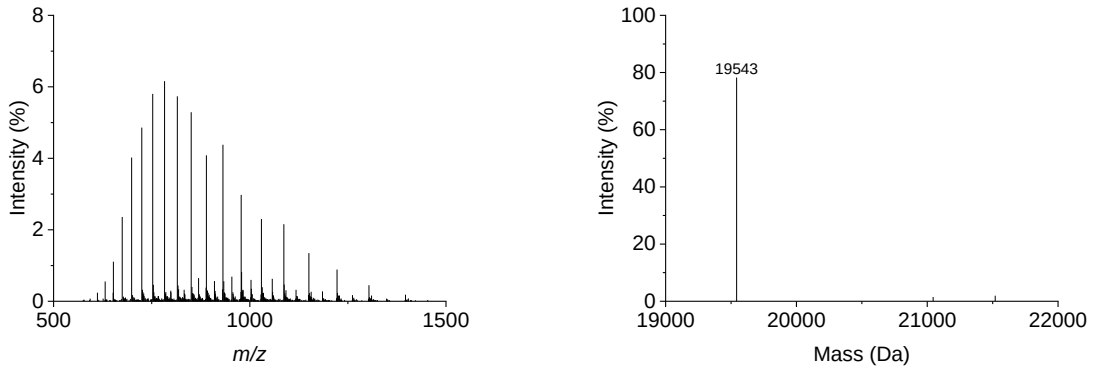
C24NM (89%)

MS (ESI$^+$) [**C24N**+H]$^+$ found 19571, calculated 19572; [**C24NM**+H]$^+$ found 19701, calculated 19703; [**C24NM**+H+MeCN]$^+$ found 19744, calculated 19744.

Figure S5.4. Raw (left) and deconvoluted (right) mass spectra for Shh C24N from Qkine.

Shh C24S

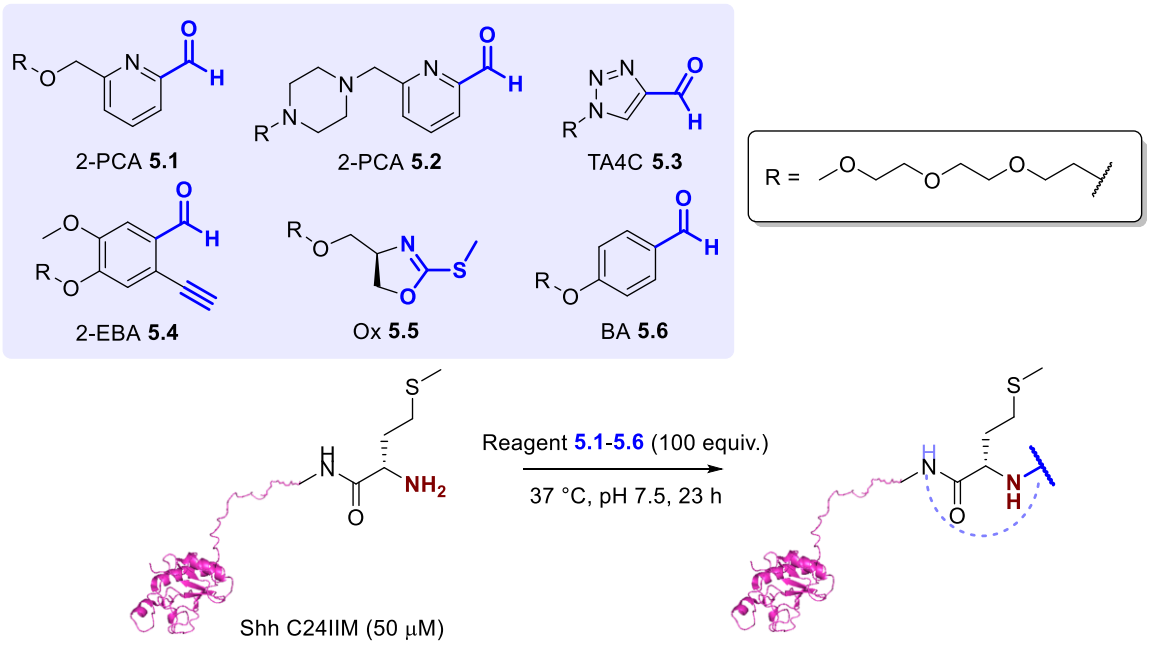
Theoretical M_w: 19544 Da



MS (ESI$^+$) [**C24S**+H]$^+$ found 19543, calculated 19545.

Figure S5.5. Raw (left) and deconvoluted (right) mass spectra for Shh C24S from Qkine.

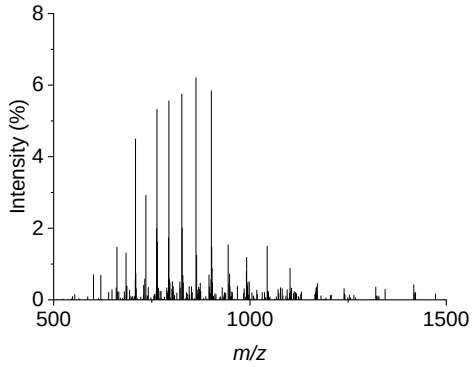
5.4.2.1 Initial reagent screening

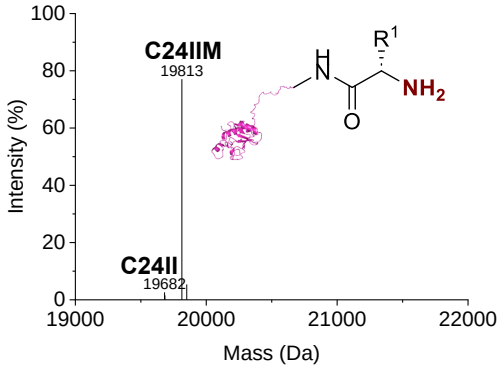


Procedure 5A: *Modification with reagents under conditions adapted from MacDonald et al.*[22] A stock solution of the modification reagent (50 µL, 10 mM, 500 nmol, 100 equiv., in 50 mM pH 7.5 sodium phosphate buffer) was added to a solution of protein (50 µL, 100 µM, 5 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer), and the mixture incubated at 37 °C for 23 h with agitation (1000 rpm). Conversion was determined by LC-MS analysis without purification.

Each of the reagents **5.1**-**5.6** were used to modify Shh C24IIM under the conditions outlined in **General Procedure 5A**. For compound **5.6**, 5 equiv. of sodium cyanoborohydride in 5 µL of water was also added, and compound **5.6** was added as a more concentrated solution (45 µL, 11 mM) to keep the total volume constant. Protein conjugates were purified by dialysis to remove excess reagent (4 °C, 3.5 kDa MWCO; 1 × 50 mM pH 7.4 sodium phosphate buffer, 5 h; 1 × water, 17 h; 1 × water, 2 h), and conversion of the purified

conjugates was determined by LC-MS analysis. Concentration was determined by UV-vis spectroscopy at 280 nm by Dr. Kerry Price. A control was also prepared for Shh C24IIM as outlined above, in which a buffer solution (50 µL, 50 mM pH 7.5 sodium phosphate buffer) was used instead of the modification reagent.





Shh C24IIM

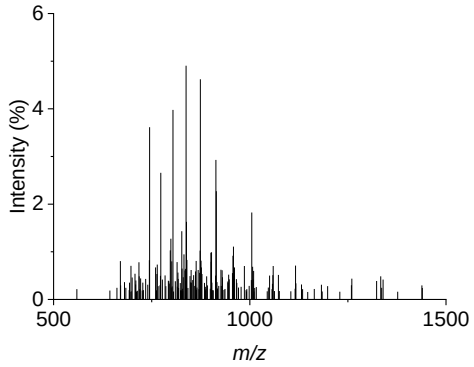
No reagent

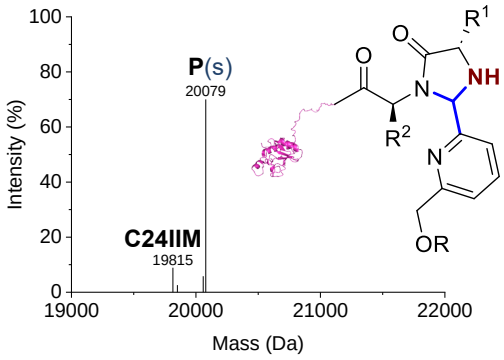
Purified

C24II (5%)

C24IIM (95%)

MS (ESI$^+$) [**C24II**+H]$^+$ found 19682, calculated 19684; [**C24IIM**+H]$^+$ found 19813, calculated 19815; [**C24IIM**+H+MeOH]$^+$ found 19851, calculated 19848.





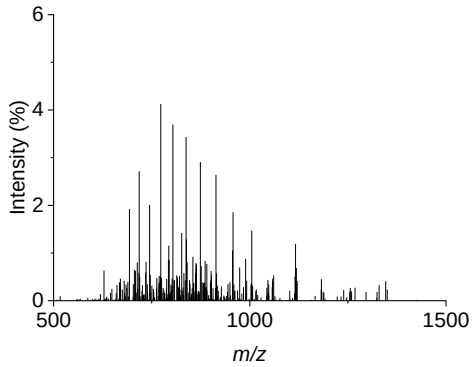
Shh C24IIM

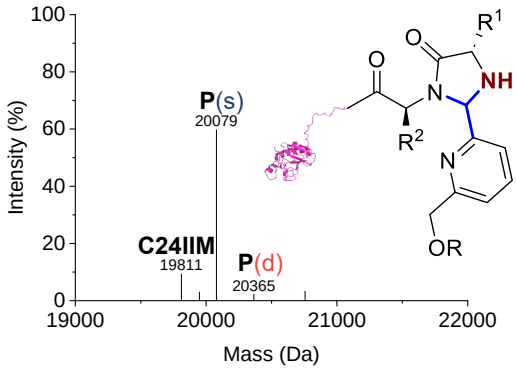
2-PCA **5.1**

Crude

89%

MS (ESI$^+$) [**C24IIM**+H]$^+$ found 19815, calculated 19815; [**C24IIM**+H+MeOH]$^+$ found 19851, calculated 19848; [**P**(s)$_{\mathbf{C24IIM}}$+H]$^+$ found 20079, calculated 20080.





Shh C24IIM

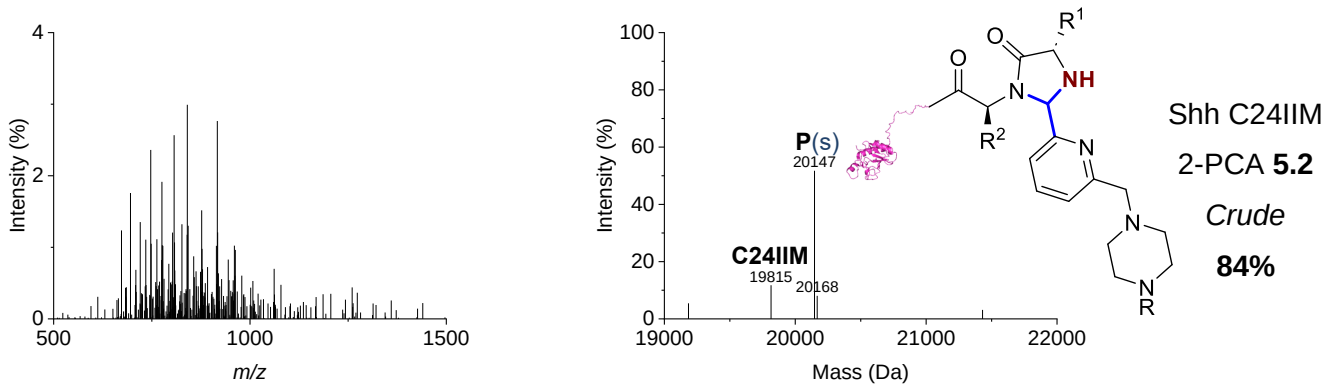
2-PCA **5.1**

Purified

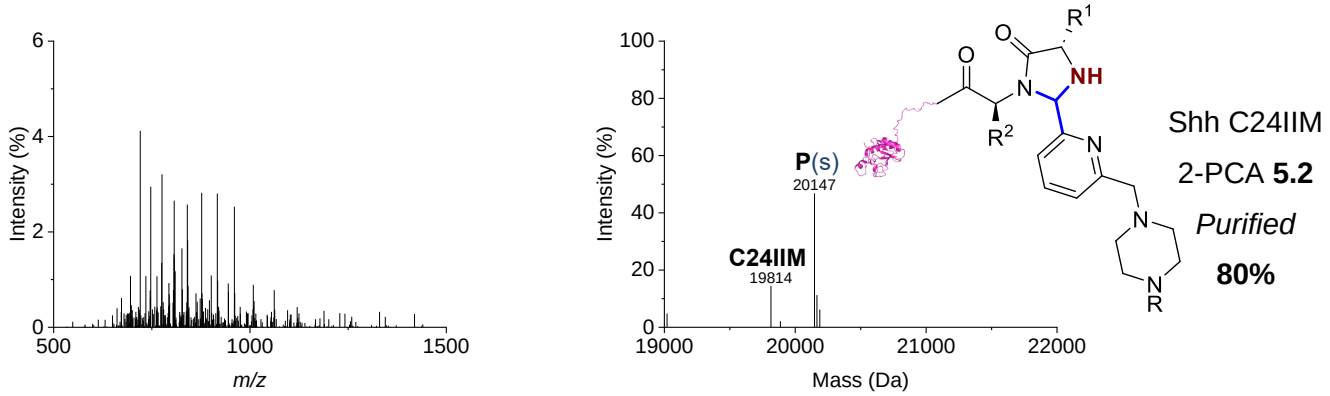
87%

84(s):3(d)

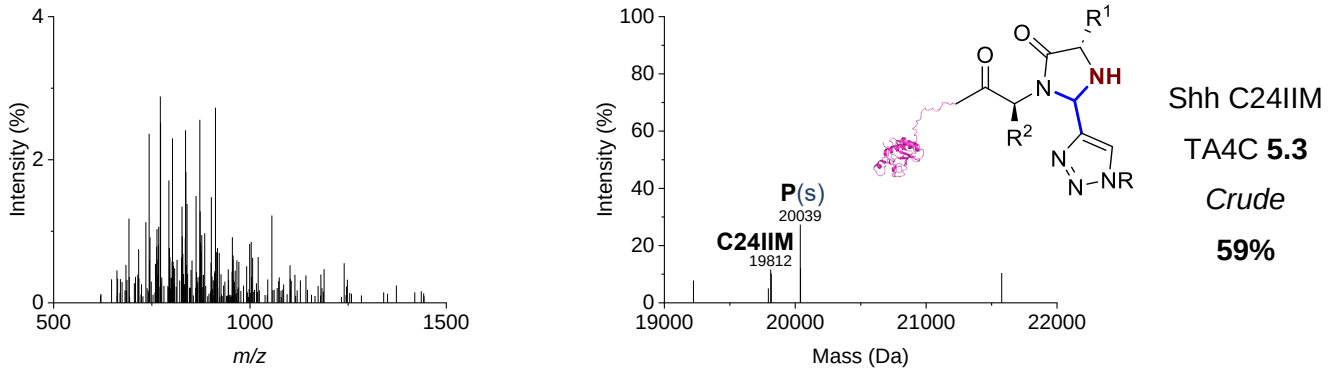
MS (ESI$^+$) [**C24IIM**+H]$^+$ found 19811, calculated 19815; [**P**(s)$_{\mathbf{C24IIM}}$+H]$^+$ found 20079, calculated 20080; [**P**(d)$_{\mathbf{C24IIM}}$+H+H$_2$O]$^+$ found 20365, calculated 20364.



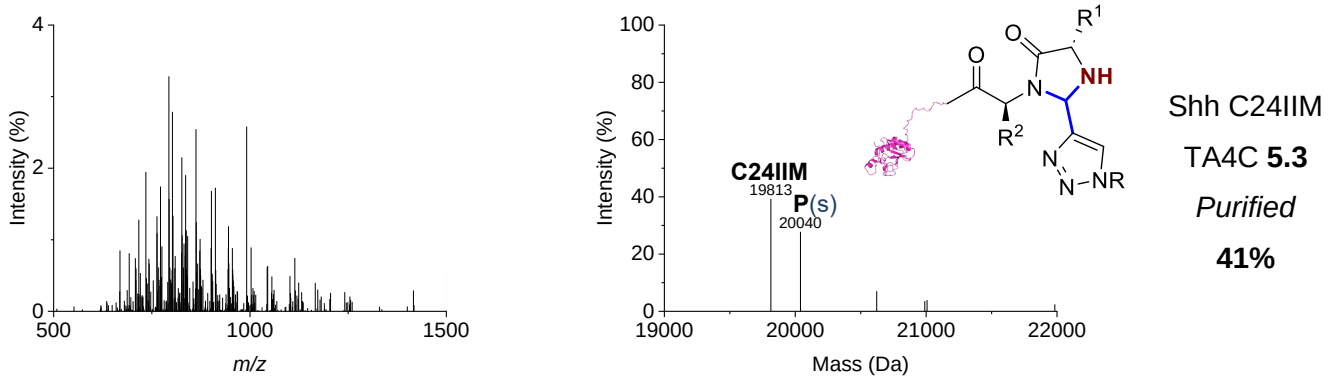
MS (ESI$^+$) [**C24IIM**+H]$^+$ found 19815, calculated 19815; [**P**(s)$_{\mathbf{C24IIM}}$+H]$^+$ found 20147, calculated 20148; [**P**(s)$_{\mathbf{C24IIM}}$+H+H_2O]$^+$ found 20168, calculated 20166.



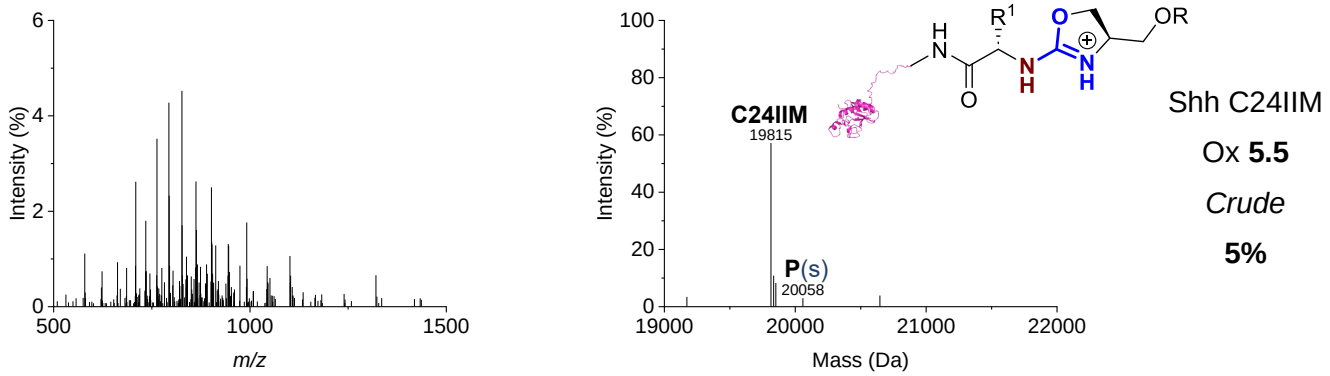
MS (ESI$^+$) [**C24IIM**+H]$^+$ found 19814, calculated 19815; [**P**(s)$_{\mathbf{C24IIM}}$+H]$^+$ found 20147, calculated 20148; [**P**(s)$_{\mathbf{C24IIM}}$+H+H_2O]$^+$ found 20166, calculated 20166; [**P**(s)$_{\mathbf{C24IIM}}$+H+MeCN]$^+$ found 20188, calculated 20189.



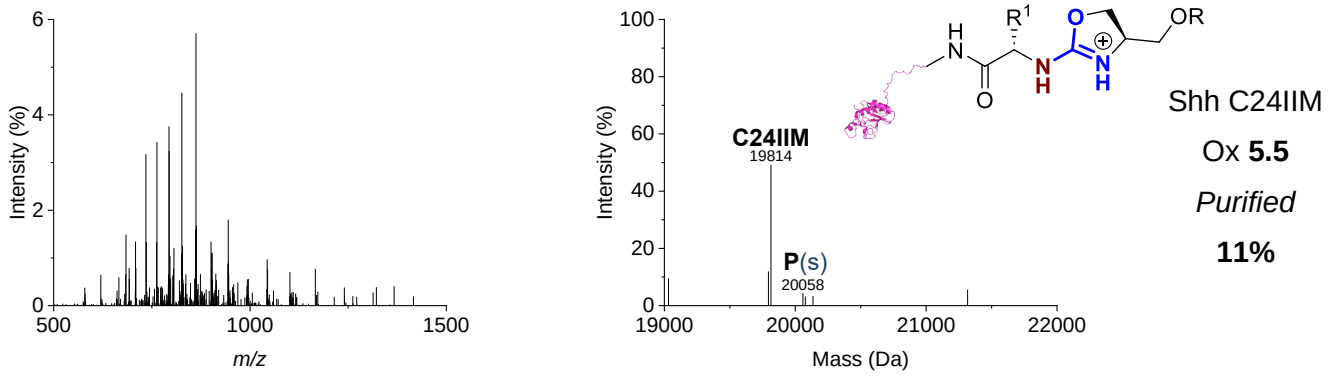
MS (ESI$^+$) [**C24IIM**–H_2O+H]$^+$ found 19794, calculated 19797; [**C24IIM**+H]$^+$ found 19812, calculated 19815; [**C24IIM**+H]$^+$ found 19817, calculated 19815; [**P**(s)$_{\mathbf{C24IIM}}$+H]$^+$ found 20039, calculated 20040; [**P**(s)$_{\mathbf{C24IIM}}$+H]$^+$ found 20041, calculated 20040.



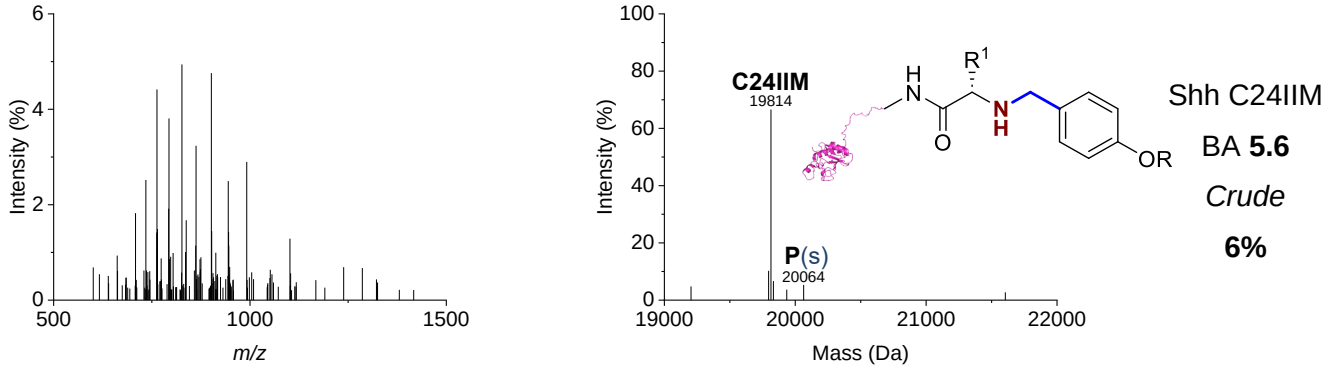
MS (ESI$^+$) [**C24IIM**+H]$^+$ found 19813, calculated 19815; [**P**(s)$_{\mathbf{C24IIM}}$+H]$^+$ found 20040, calculated 20040.



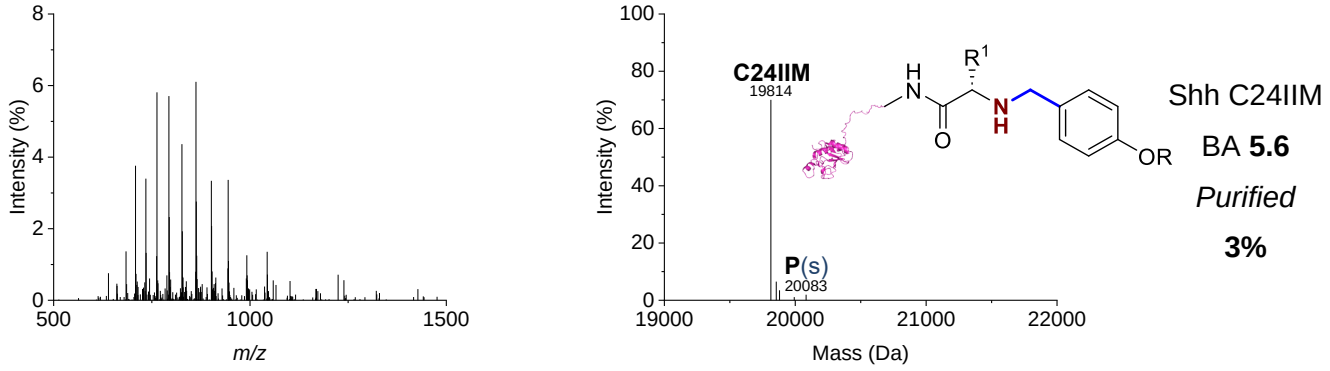
MS (ESI$^+$) [**C24IIM**+H]$^+$ found 19815, calculated 19815; [**P**(s)$_{\mathbf{C24IIM}}$+H]$^+$ found 20058, calculated 20061.



MS (ESI$^+$) [**C24IIM**–H_2O+H]$^+$ found 19795, calculated 19797; [**C24IIM**+H]$^+$ found 19814, calculated 19815; [**P**(s)$_{\mathbf{C24IIM}}$+H]$^+$ found 20058, calculated 20061; [**P**(s)$_{\mathbf{C24IIM}}$+H+H_2O]$^+$ found 20077, calculated 20079.



MS (ESI$^+$) [**C24IIM**–H_2O+H]$^+$ found 19796, calculated 19797; [**C24IIM**+H]$^+$ found 19814, calculated 19815; [**C24IIM**+H+H_2O]$^+$ found 19833, calculated 19833; [**P**(s)$_{\mathbf{C24IIM}}$+H]$^+$ found 20064, calculated 20067.



MS (ESI$^+$) [**C24IIM**+H]$^+$ found 19814, calculated 19815; [**C24IIM**+MeCN+H]$^+$ found 19855, calculated 19856; [**C24IIM**+2MeOH+H]$^+$ found 19880, calculated 19879; [**P**(s)$_{C24IIM}$+H_2O+H]$^+$ found 20083, calculated 20085.

Figure S5.6. Raw (left) and deconvoluted (right) mass spectra for Shh C24IIM conjugates. **P** = modified protein product; s = single modification; d = double modification.

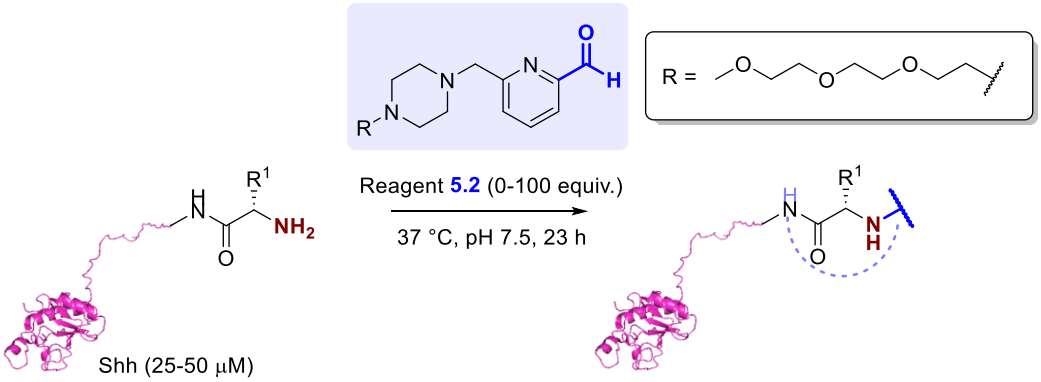
Small molecule	Conversion (%)	
	Shh C24IIM *(crude)*	Shh C24IIM *(purified)*
5.1	**89**	**87** *84(s):3(d)*
5.2	**84**	**80**
5.3	**59**	**41**
5.4	**-**	**-**
5.5	**5**	**11**
5.6	**6**	**3**

Table S5.1. Screening of reagents for the modification of Shh C24IIM. Modification of Shh C24IIM with compounds **5.1**-**5.6** under conditions outlined in General Procedure **5A**. s = single modification; d = double modification (*missing data due to poor quality mass spectra*).

Small molecule	Measured concentration (μM)
Unmodified control	25
2-PCA **5.1**	33
2-PCA **5.2**	43
Ox **5.5**	22

Table S5.2. Concentrations of Shh conjugates measured by UV-vis spectroscopy at 280 nm. Calculation of concentrations based on 1% extinction coefficients of 13.75 for Shh C24II *(data collected by Dr. Kerry Price at Qkine)*.

5.4.2.2 Solubility studies: 2-PCA concentration



To explore the influence of 2-PCA concentration on conversion and solubility, Shh WT and Shh C24IIM were modified with reagent **5.2** according to **General Procedure 5A** on a 50 μL scale, varying the equiv. of 2-PCA

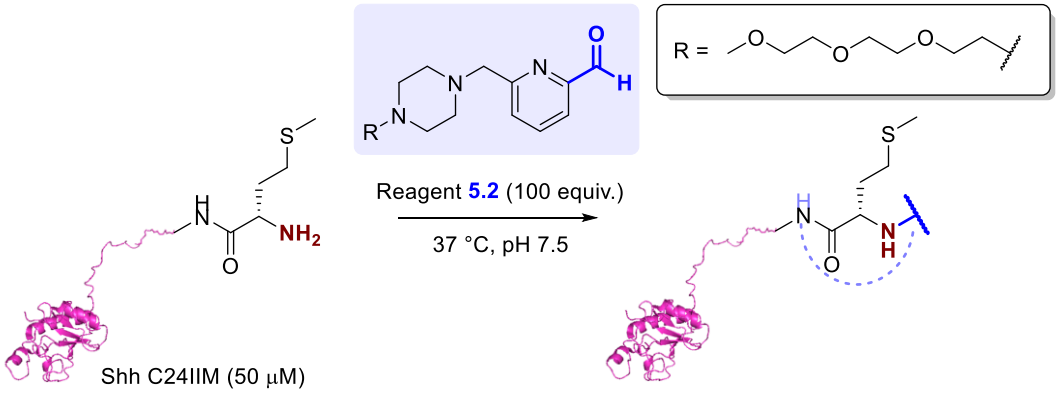
while maintaining all other conditions (0, 25, 50, 100 equiv.). Reagent **5.2** was also applied under the conditions outlined in **General Procedure 5A** on a 40 µL scale, with reduced protein concentration (25 µM; 100 equiv. 2-PCA).

2-PCA concentration (mM)	Protein concentration (µM)	Conversion (%)	
		Shh WT	Shh C24IIM
0	50	-	**0**
1.25	50	-	**77**
2.5	50	-	**92**
2.5	25	-	**88**
5	50	-	**92** *76(s):16(d)*

*Table S5.3. Variation of concentrations for the modification of Shh WT and Shh C24IIM with 2-PCA **5.2**.* Modification of Shh WT and Shh C24IIM with compound **5.2** under conditions adapted from General Procedure **5A**. s = single modification; d = double modification (*missing data due to poor quality mass spectra*).

Aliquots of the reaction mixtures (5.11 µL) and precipitates (resuspended in 5.11 µL PBS) were analysed by 15% SDS-PAGE gels (see *Fig. 5.12* for analysis details).

5.4.2.3 Solubility studies: reaction time



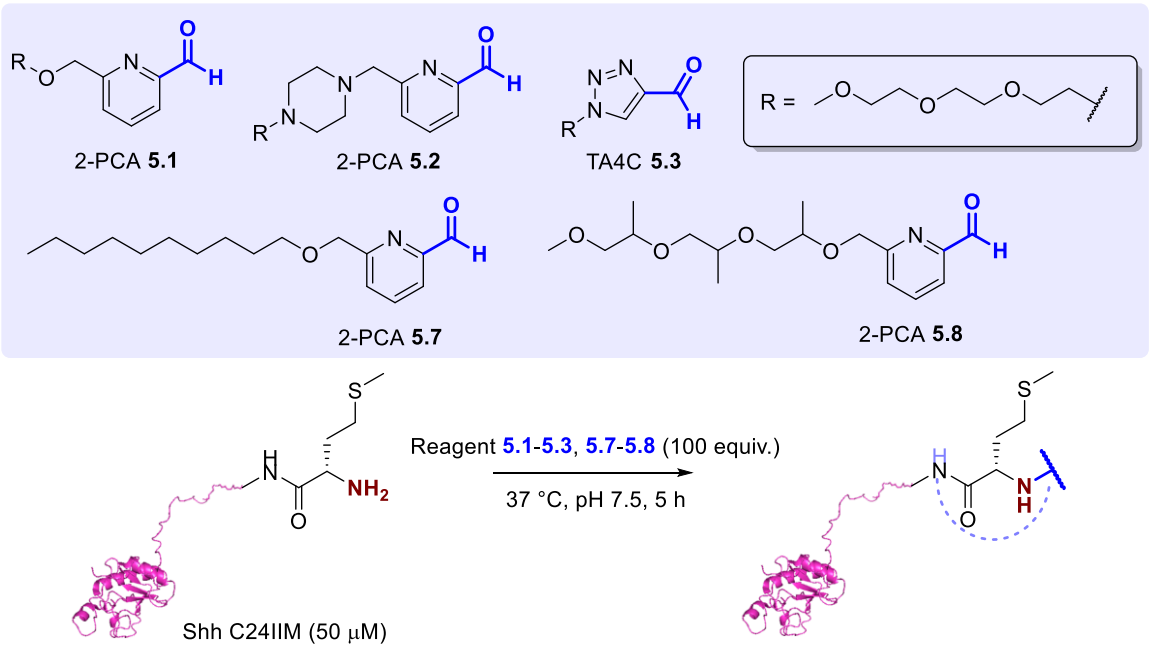
Shh C24IIM was modified with reagent **5.2** according to **General Procedure 5A**. Conversion was determined by LC-MS analysis without purification 1 h, 3 h, 5 h and 29 h after reagent addition. Note that these time-points are approximate, due to a 20 min delay between sample collection and LC-MS analysis, during which the sample was at rt. The timepoints indicated are the times at which the sample aliquots were taken for LC-MS analysis.

Time (h)	**Conversion (%)**
0	**29**
3	**60**
5	**78**

29	**100** *74(s):26(d)*

Table S5.4. Kinetic study of the modification of Shh C24IIM with 2-PCA ***5.2****.* Modification of Shh C24IIM with compound **5.2** under conditions outlined in General Procedure **5A**. s = single modification; d = double modification.

5.4.2.4 Reagent screening at 37 °C for 5 h



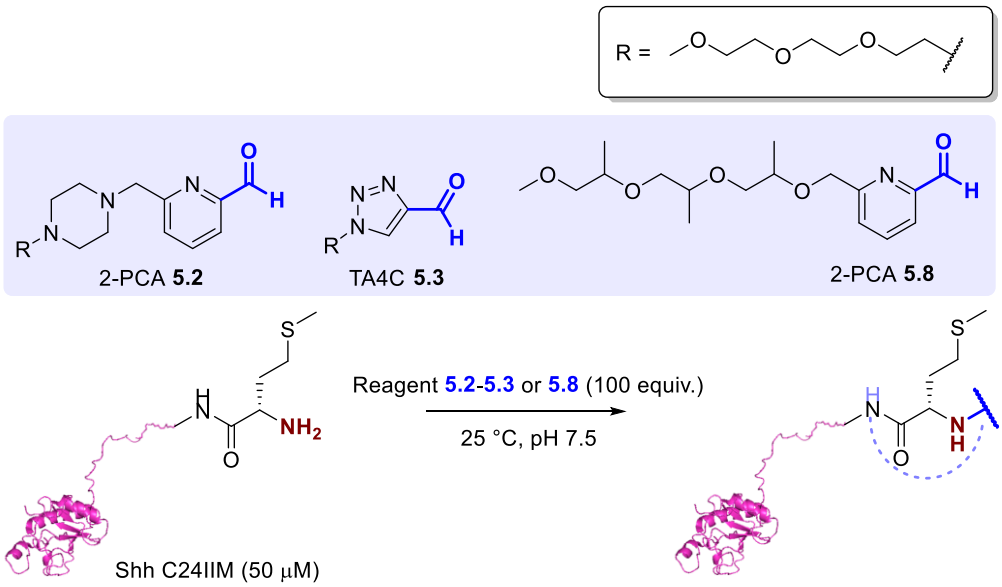
Shh C24IIM was modified with reagents **5.1**-**5.3** and **5.7**-**5.8** according to **General Procedure 5A**, with a reduced incubation time of 5 h at 37 °C. Reagent **5.7** was found to be insoluble in sodium phosphate buffer, so a stock solution of reagent **5.7** (10 µL, 50 mM, 500 nmol, 100 equiv., in DMSO) was added to a solution of protein (50 µL, 100 µM, 5 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer) and 50 mM pH 7.5 sodium phosphate buffer (40 µL), and the mixture incubated at 37 °C for 23 h with agitation (1000 rpm).

Protein conjugates were purified by dialysis to remove excess reagent (4 °C, 3.5 kDa MWCO; 1 × water, 17 h; 1 × water, 3 h; 2 × water, 2 h), and conversion of the purified conjugates was determined by LC-MS analysis. Total, soluble, and precipitate (reconstituted in 5.05 µL PBS) samples were analysed by 15% SDS-PAGE gels (see *Fig. 5.15* for analysis details).

Small molecule	**Conversion (%)**	
	Crude	*Purified*
5.1	**77**	**74**
5.2	**65**	**65**
5.3	**56**	**32**
5.7	**0**	**0**
5.8	**46**	**39**

Table S5.5. Reduction of reaction time to 5 h. Modification of Shh C24IIM with compounds **5.1**-**5.3** and **5.7**-**5.8** under conditions outlined in General Procedure **5A**, with a reduced incubation time of 5 h. s = single modification; d = double modification.

5.4.2.5 Reagent screening at 25 °C, over time



Shh C24IIM was modified with reagents **5.2**, **5.3** and **5.8** according to **General Procedure 5A** on a 150 µL scale, with a reduced temperature of 25 °C. Conversion was determined by LC-MS analysis without purification 6 h, 22 h, 30 h and 53 h after reagent addition. Total, soluble, and precipitate (reconstituted in 6.73 µL PBS) samples were analysed by 15% SDS-PAGE gels (see *Fig. 5.21* for analysis details).

Protein conjugates were purified by dialysis to remove excess reagent for the Gli-luciferase assay (4 °C, 3.5 kDa MWCO; 1 × water, 17 h; 1 × water, 2 h; 1 × water, 6.5 h; 1 × water, 15 h; 1 × water, 4.5 h; 1 × water, 2.5 h; 1 × 50 mM pH 7.5 sodium phosphate buffer, 2 h). Concentration was determined by UV-vis spectroscopy at 280 nm by Dr. Kerry Price. A control was prepared for Shh C24IIM as outlined above on a 36 µL scale without LC-MS analysis (and SDS-PAGE analysis of total only), in which a buffer solution (18 µL, 50 mM pH 7.5 sodium phosphate buffer) was used instead of the modification reagent.

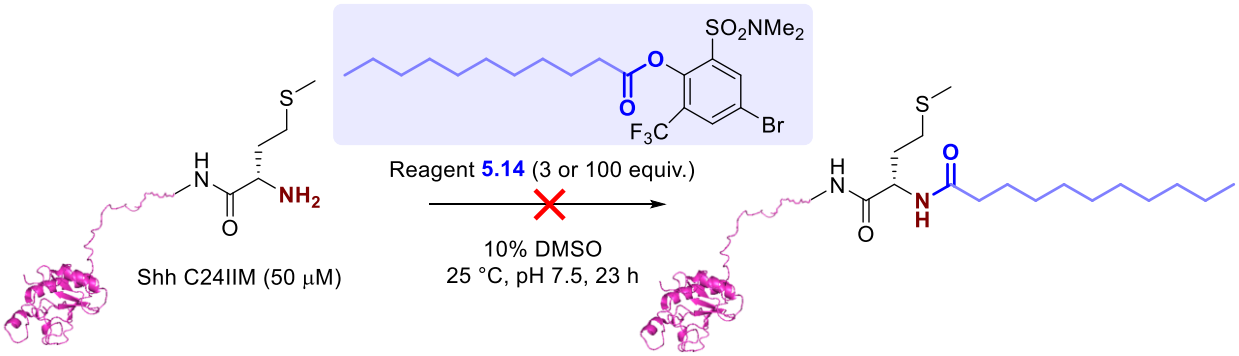
Small molecule	Conversion (%)			
	6 h	*22 h*	*30 h*	*53 h*
5.2	**40**	**79**	**78**	**86**
5.3	**36**	**65**	**68**	**66**
5.8	**26**	**59**	**65**	**-**

Table S5.6. Kinetic study of the modification at 25 °C. Modification of Shh C24IIM with compounds **5.2**, **5.3** and **5.8** under conditions outlined in General Procedure **5A**, with a reduced temperature of 25 °C. s = single modification; d = double modification (*missing data due to poor quality mass spectra*).

Small molecule	Measured concentration (μM)
Unmodified	17
2-PCA **5.2**	16
TA4C **5.3**	23
2-PCA **5.8**	38

Table S5.7. Concentrations of Shh C24IIM conjugates measured by UV-vis spectroscopy at 280 nm. Calculation of concentrations based on 1% extinction coefficient of 13.75 for Shh C24II *(data collected by Dr. Kerry Price at Qkine)*.

5.4.2.6 Modification of Shh C24IIM with activated phenol ester

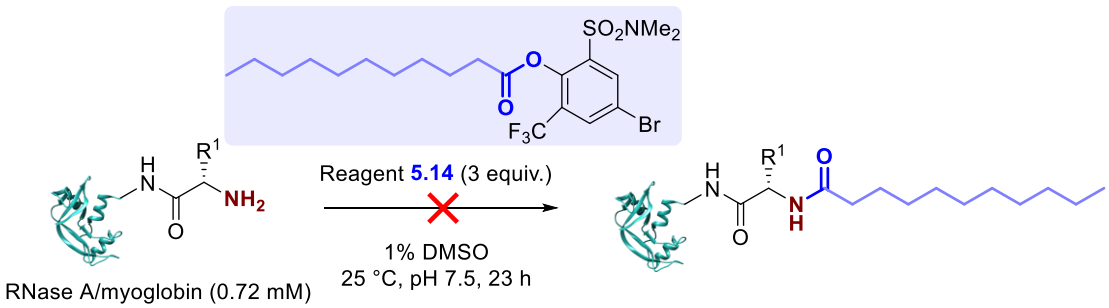


Stoichiometry as used previously in ***General Procedure 5A****.* A stock solution of reagent **5.14** (50 μL, 10 mM, 500 nmol, 100 equiv., in 20% DMSO in 50 mM pH 7.5 sodium phosphate buffer) was added to a solution of Shh C24IIM (50 μL, 100 μM, 5 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer).

Stoichiometry as reported by Mikkelsen et al.[23] A stock solution of reagent **5.14** (50 μL, 0.3 mM, 15 nmol, 3 equiv., in 20% DMSO in 50 mM pH 7.5 sodium phosphate buffer) was added to a solution of Shh C24IIM (50 μL, 100 μM, 5 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer).

Both reaction mixtures were incubated at 25 °C for 23 h with agitation (1000 rpm). Conversion was determined by LC-MS analysis without purification.

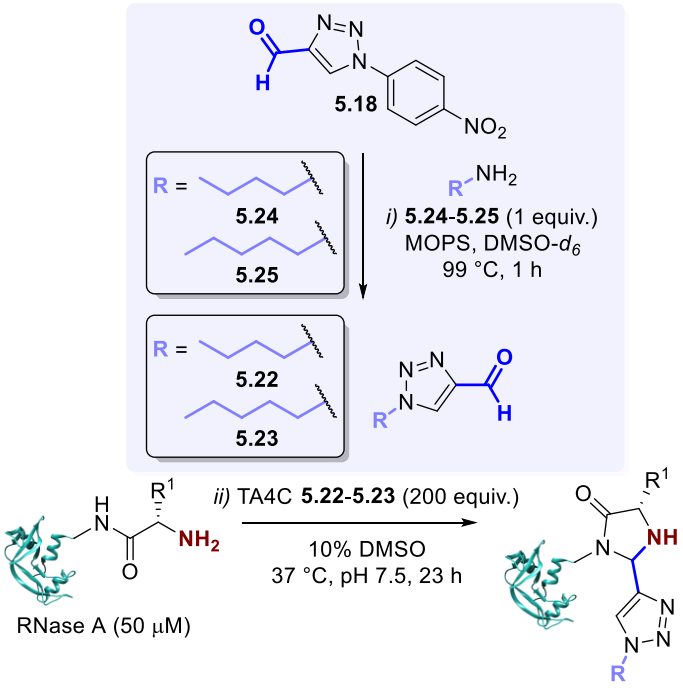
5.4.2.7 Modification of RNase A and myoglobin with activated phenol ester



Reagent **5.14** was used to modify RNase A and myoglobin under the conditions outlined in **General Procedure 5B**.

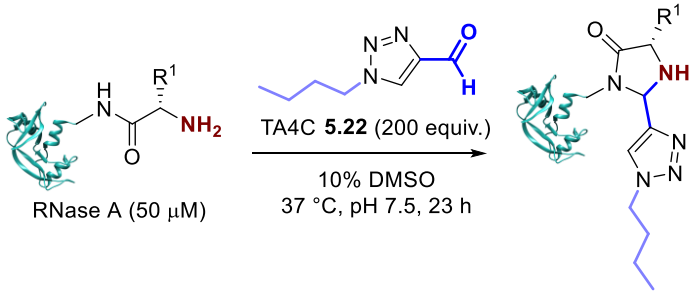
Procedure 5B: *Modification with reagents under conditions reported by Mikkelsen et al.*[23] A stock solution of the modification reagent (2 µL, 110 mM, 219 nmol, 3 equiv., in DMSO) was added to a solution of protein (100 µL, 731 µM, 73 nmol, 1 equiv., in 200 mM pH 7.5 HEPES buffer), and the mixture incubated at 25 °C for 23 h with agitation (1000 rpm). Conversion was determined by LC-MS analysis without purification.

5.4.2.8 Validation of Dimroth rearrangement protocol using RNase A

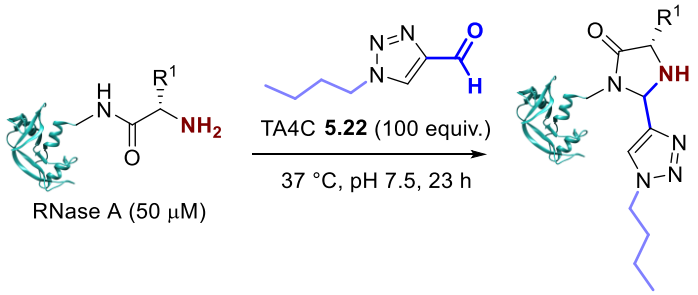


(a) Reagent **5.22** and reagent **5.23** were prepared using amines **5.24** and **5.25** respectively, and used to modify RNase A under the conditions outlined in **General Procedure 5C**.

Procedure 5C: *In-situ preparation of TA4C reagent and protein modification conditions reported by Onoda et al.*[25] *i)* A stock solution of the relevant amine (300 µL, 200 mM, 60 µmol, 1 equiv., in DMSO-d_6) and a solution of MOPS (15 µL, 400 mM, 6 µmol, 0.1 equiv., in D_2O) were added to a solution of TA4C **5.18** (300 µL, 200 mM, 60 µmol, 1 equiv., in DMSO-d_6), and the reaction mixture was stirred at 99 °C for 1 h. The conversion of the Dimroth rearrangement reaction was determined by ^{1}H NMR analysis of the crude reaction mixture. *ii)* A solution of protein (45 µL, 56 µM, 2.5 nm, 1 equiv., in 50 mM pH 7.5 Na phosphate buffer) was modified with an aliquot of the crude TA4C reagent from part *i)* (5 µL, 100 mM, 0.5 µmol, 200 equiv.) at 37 °C for 23 h. Conversion was determined by LC-MS analysis without purification.



(b) *Modification with isolated TA4C, stoichiometry as reported by Onoda et al.*[25] A stock solution of isolated reagent **5.22**, i.e. prepared by chemical synthesis (5 µL, 100 mM, 0.5 µmol, 200 equiv., in DMSO), was added to a solution of RNase A (45 µL, 56 µM, 2.5 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer), and the mixture incubated at 37 °C for 23 h with agitation (1000 rpm). Conversion was determined by LC-MS analysis without purification.

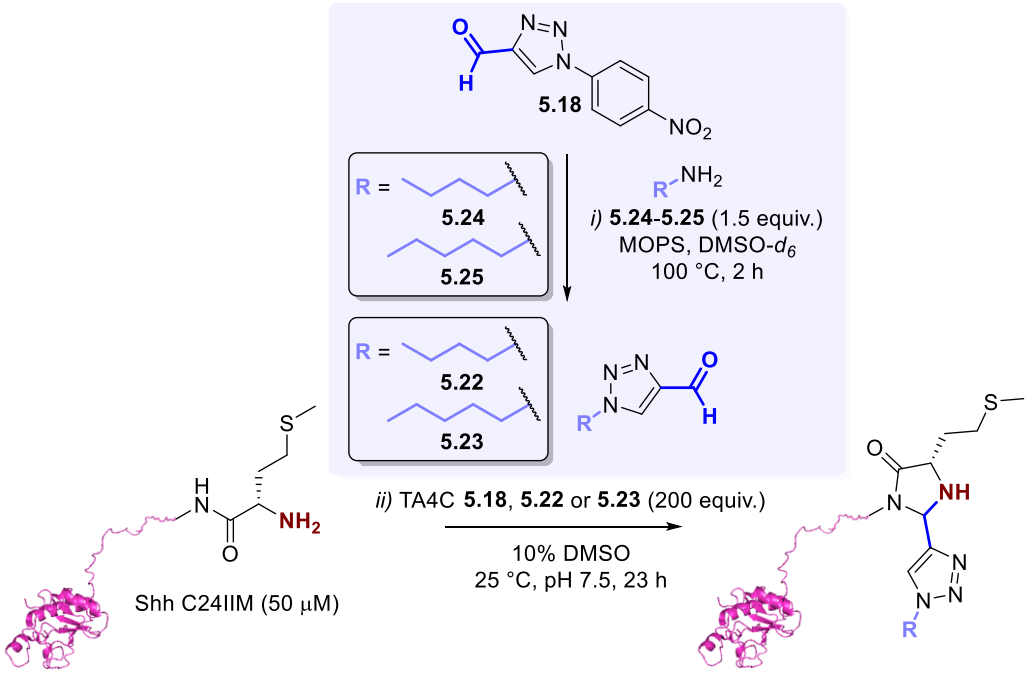


(c) *Modification with isolated TA4C, under standardised conditions.* A stock solution of isolated reagent **5.22**, i.e. prepared by chemical synthesis (5 µL, 50 mM, 0.25 µmol, 100 equiv., in 50 mM pH 7.5 sodium phosphate buffer), was added to a solution of RNase A (45 µL, 56 µM, 2.5 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer), and the mixture incubated at 37 °C for 23 h with agitation (1000 rpm). Conversion was determined by LC-MS analysis without purification.

Conditions	Small molecule	**Conversion (%)**	Off-target conversion with TA4C **5.18** (%)
(a)	**5.22**	**58**	9
(a)	**5.23**	**69**	0
(b)	**5.22**	**66**	0
(c)	**5.22**	**49**	0

Table S5.8. Validation of Dimroth rearrangement protocol. A comparison of RNase A / TA4C **5.22**/**5.23** conjugation using a Dimroth rearrangement protocol, to standard modification conditions with isolated reagent **5.22**.

5.4.2.9 Modification of Shh C24IIM using the Dimroth rearrangement protocol



Reagent **5.22** and reagent **5.23** were prepared using amines **5.24** and **5.25** respectively and used to modify Shh C24IIM under the conditions outlined in **General Procedure 5C**. Adapted parameters included the number of equivalents of amine (300 µL, 300 mM, 90 µmol, 1.5 equiv., in DMSO-d_6), reaction time for part *i)* (2 h), and temperature of the protein modification step *ii)* (25 °C).

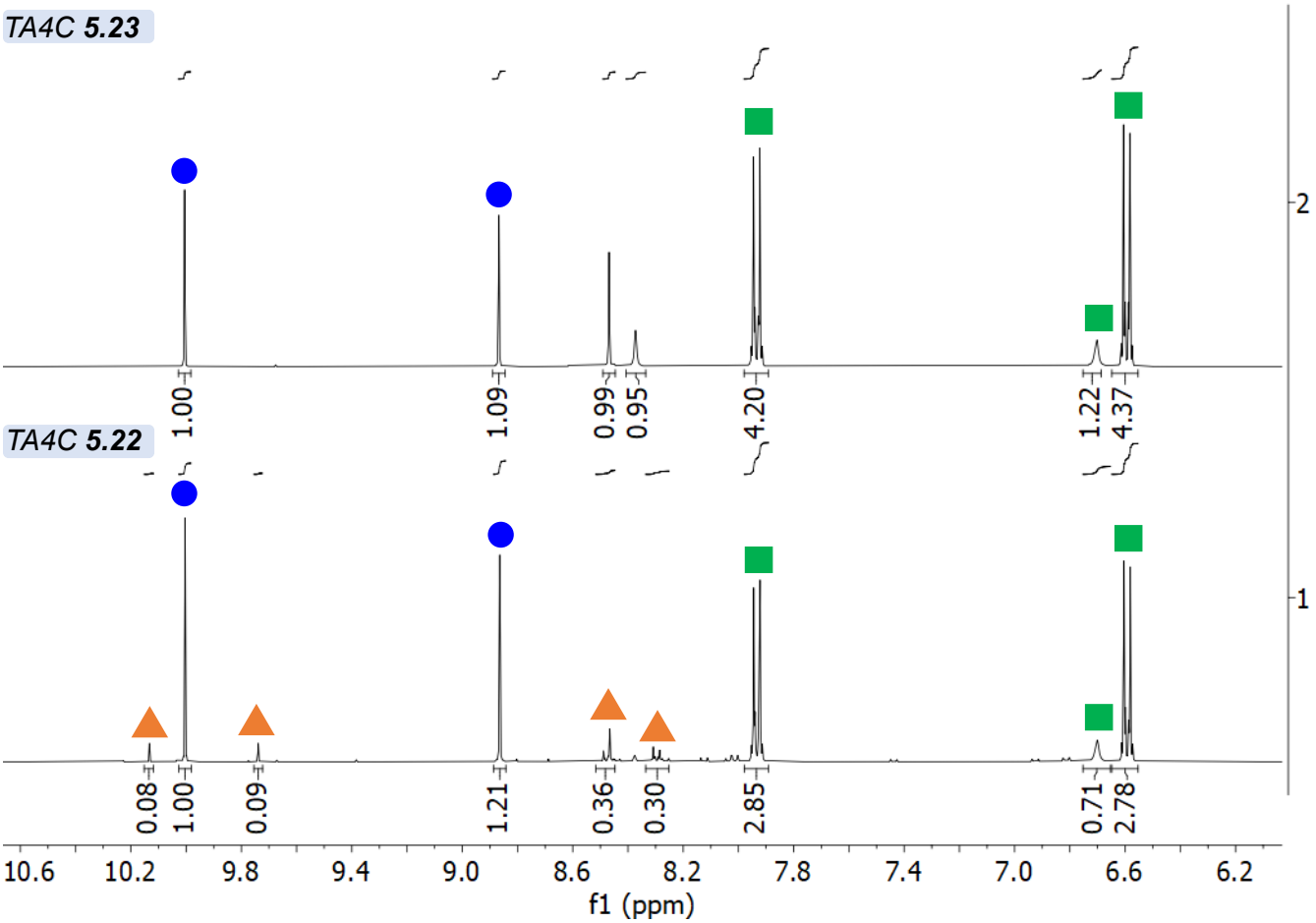


*Figure S5.7. ^{1}H NMR of reaction mixtures after Dimroth rearrangement of TA4C **5.18** with amines **5.24** and **5.25*** (**triangle** = unreacted TA4C **5.18**, **circle** = TA4C **5.22** or **5.23**, **square** = by-product **5.20**). *Note: unassigned signals for TA4C **5.23** may have been due to the intermediate imine.*

A control was prepared where a solution of reagent **5.18** (12 µL, 100 mM, 1.2 µmol, 200 equiv., in DMSO) was added to a solution of Shh C24IIM (108 µL, 56 µM, 6 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer), and the mixture incubated at 25 °C for 23 h with agitation (1000 rpm). Conversion was determined by LC-MS analysis without purification.

Modified Shh C24IIM samples were purified by size exclusion chromatography (PD MiniTrap G-25) following the manufacturer's instructions; a 100 µL sample was eluted with a 400 µL stacker volume (50 mM pH 7.5 sodium phosphate buffer), followed by 1 mL 50 mM pH 7.5 sodium phosphate buffer. Concentration was determined by UV-vis spectroscopy at 280 nm by Dr. Kerry Price, and conversion was determined by LC-MS analysis.

	Crude		*Purified*	
Small molecule	**Conversion (%)**	Conversion with TA4C **5.18** (%)	**Conversion (%)**	Conversion with TA4C **5.18** (%)
5.22	**69**	**9**	**67**	**10**
5.23	**79**	**0**	**71**	**0**
5.18	**57**	**57**	**55**	**55**

Table S5.9. Application of Dimroth rearrangement protocol to Shh C24IIM. Modification of Shh C24IIM with TA4Cs **5.22** and **5.23** using a Dimroth rearrangement protocol.

Small molecule	**Measured concentration (μM)**
Shh C24IIM Qkine	51
Unmodified control	4
TA4C **5.22**	2
TA4C **5.23**	2
TA4C **5.18**	3

Table S5.10. Concentrations of Shh C24IIM conjugates measured by UV-vis spectroscopy at 280 nm. Calculation of concentrations based on 1% extinction coefficient of 13.75 for Shh C24II *(data collected by Dr. Kerry Price at Qkine)*. "Shh C24II Qkine" was a freshly prepared solution of Shh C24IIM in water, reconstituted at Qkine by Dr. Kerry Price for Gli-luciferase assay.

*5.4.2.10 Modification of N-terminal mutants with TA4C **5.22***

Shh C24IIM, C24A, C24NM and C24S were modified with reagent **5.22**, according to **General Procedure 5D**. Conversion was determined by LC-MS analysis without purification. Unmodified controls were prepared under identical conditions, in the absence of TA4C reagent/amines.

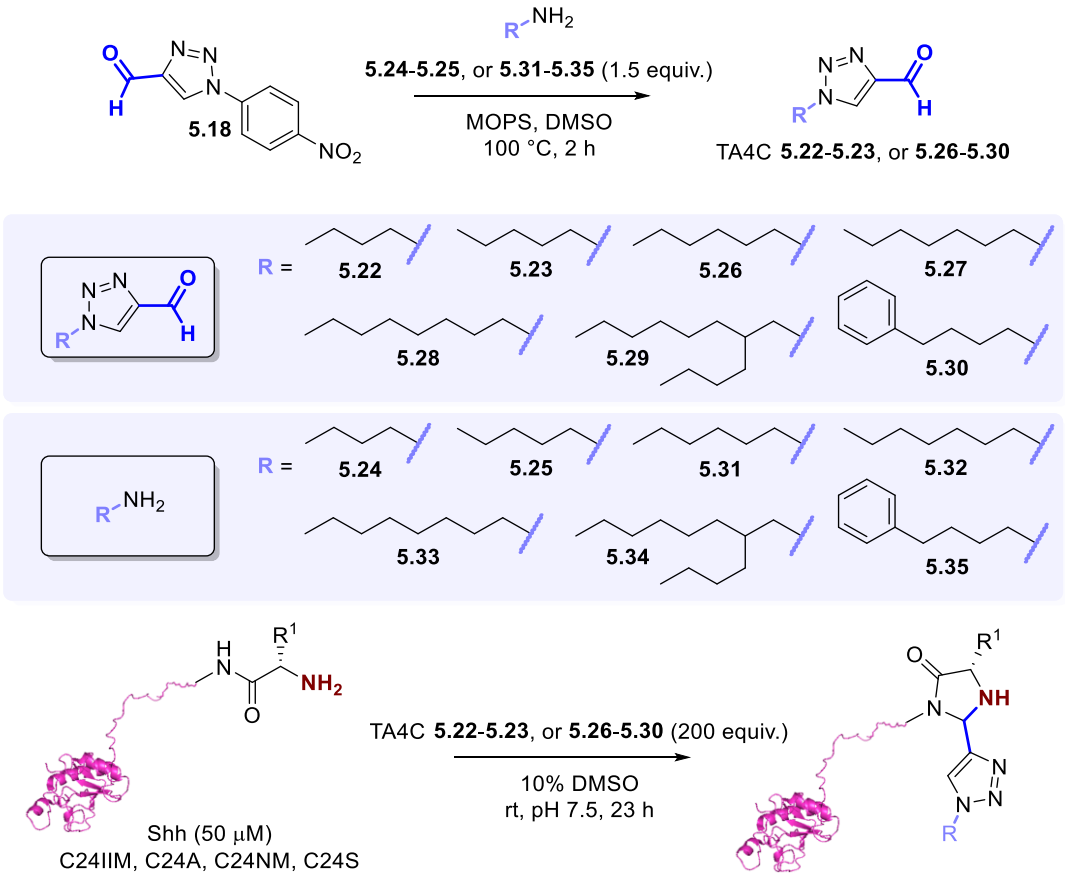
Procedure 5D: *In-situ preparation of TA4C reagent and protein modification conditions adapted for Shh modification. i)* A stock solution of the relevant amine (60 μL, 300 mM, 18 μmol, 1.5 equiv., in DMSO) and a solution of MOPS (3 μL, 400 mM, 1.2 μmol, 0.1 equiv., in H_2O) were added to a solution of TA4C **5.18** (60 μL, 200 mM, 12 μmol, 1 equiv., in DMSO), and the reaction mixture was stirred at 100 °C for 2 h. *ii)* A solution of protein (108 μL, 56 μM, 6 nmol, 1 equiv., in 50 mM pH 7.5 Na phosphate buffer) was modified with an aliquot of the crude TA4C reagent from part *i)* (12 μL, 100 mM, 1.2 μmol, 200 equiv.) at 25 °C for 23 h. Modified Shh samples were purified by size exclusion chromatography (PD MiniTrap G-25); a 100 μL sample was eluted with a 400 μL stacker volume (50 mM pH 7.5 sodium phosphate buffer), followed by 1 mL 50 mM pH 7.5 sodium phosphate buffer (500 μL fraction size). Concentration was measured by UV-vis spectroscopy at 280 nm, and conversion was determined by LC-MS analysis.

N-terminal mutant	*Crude*		*Purified*	
	Conversion (%)	Conversion with TA4C **5.18** (%)	**Conversion (%)**	Conversion with TA4C **5.18** (%)
C24IIM	**78**	0	**65**	2
C24A	**89** *82(s):7(d)*	0	**94** *76(s):18(d)*	0
C24NM	**80**	0	**52**	0
C24S	**80** *76(s):4(d)*	0	**23**	0

Table S5.11. Application of Dimroth rearrangement protocol to Shh N-terminal mutants. Modification of Shh C24IIM, C24A, C24NM and C24S with TA4C **5.22** using the Dimroth rearrangement protocol. s = single modification; d = double modification.

5.4.2.11 Alkyl chain screen with a range of N-terminal mutants

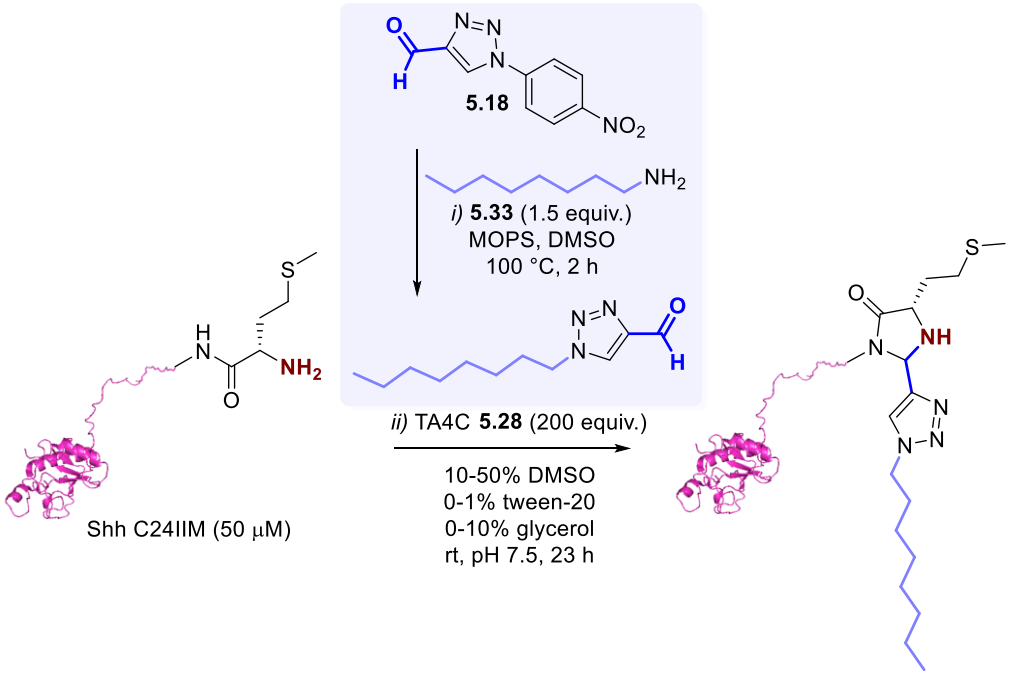
Shh C24IIM, C24A, C24NM and C24S were modified with reagents **5.22**, **5.23**, and **5.26**-**5.30**, according to **General Procedure 5D**, with size exclusion chromatography sample collection in 50 µL fractions. The Dimroth rearrangement step *i)* was carried out on a 61.5 µL scale (without stirring); the protein modification step *ii)* was carried out on a 100 uL scale, at room temperature.



N-terminal mutant	Small molecule	**Conversion (%)**	Off-target conversion with TA4C **5.20** (%)	Measured concentration (μM)
C24IIM	Unmodified	**-**	-	19
	5.22	**70** *65(s):6(d)*	0	14
	5.23	**65**	0	9
	5.26	**51**	0	11
	5.27	**28**	5	13
	5.28	**13**	0	13
	5.29	**0**	0	16
	5.30	**20**	0	13
C24A	Unmodified	**-**	-	19
	5.22	**64**	0	12
	5.23	**68**	0	7
	5.26	**53**	0	7
	5.27	**25**	0	13
	5.28	**5**	0	15
	5.29	**0**	0	17
	5.30	**12**	0	16
C24NM	Unmodified	**-**	-	20
	5.22	**63**	5	4
	5.23	**62**	0	7
	5.26	**72**	0	8
	5.27	**55**	0	10
	5.28	**35**	0	12
	5.29	**0**	0	15
	5.30	**56**	0	12
C24S	Unmodified	**-**	-	21
	5.22	**48**	0	13
	5.23	**42**	0	7
	5.26	**26**	0	8
	5.27	**12**	0	13
	5.28	**5**	0	14
	5.29	**0**	0	20
	5.30	**7**	0	16

Table S5.12. Alkyl chain screen with a range of Shh N-terminal mutants. Modification of Shh C24IIM, C24A, C24NM and C24S with TA4Cs **5.22**, **5.23** and **5.26**-**5.30** using the Dimroth rearrangement protocol. s = single modification; d = double modification. *Concentrations of Shh conjugates measured by UV-vis spectroscopy at 280 nm (most concentrated fractions shown in table)*. Calculation of concentrations based on 1% extinction coefficient of 13.75 for Shh C24II, 13.85 for C24A, 13.82 for C24N and 13.84 for C24S.

5.4.2.12 Optimisation of solubilising additive levels



Shh C24IIM was modified with reagent **5.28**, according to an adapted version of **General Procedure 5D:** *i)* A stock solution of octylamine **5.33** (50 µL, 300 mM, 15 µmol, 1.5 equiv., in DMSO) and a solution of MOPS (2.5 µL, 400 mM, 1 µmol, 0.1 equiv., in H_2O) were added to a solution of TA4C **5.18** (50 µL, 200 mM, 10 µmol, 1 equiv., in DMSO), and the reaction mixture was incubated at 100 °C for 2 h. *ii)* A solution of protein (45 µL, 111 µM, 5 nmol, 1 equiv., in 50 mM pH 7.5 Na phosphate buffer) was modified with an aliquot of the crude TA4C reagent from part *i)* (10 µL, 100 mM, 1 µmol, 200 equiv.) and an ***additive solution (45 µL)*** at rt for 23 h.

The reaction parameters used are outlined in the table below.

Conditions	Reaction Parameters		
	DMSO (%)	Glycerol (%)	Tween-20 (%)
(a)	10	0	0
(b)	20	0	0
(c)	50	0	0
(d)	10	0.1	0
(e)	10	1	0
(f)	10	10	0
(g)	10	0	0.01
(h)	10	0	0.1
(i)	10	0	1

Table S5.13. Reaction parameters for the optimisation of solubilising additive levels.

Modified Shh samples were purified by size exclusion chromatography (PD MiniTrap G-25) with sample collection in 50 µL fractions, and conversion was determined by LC-MS analysis.

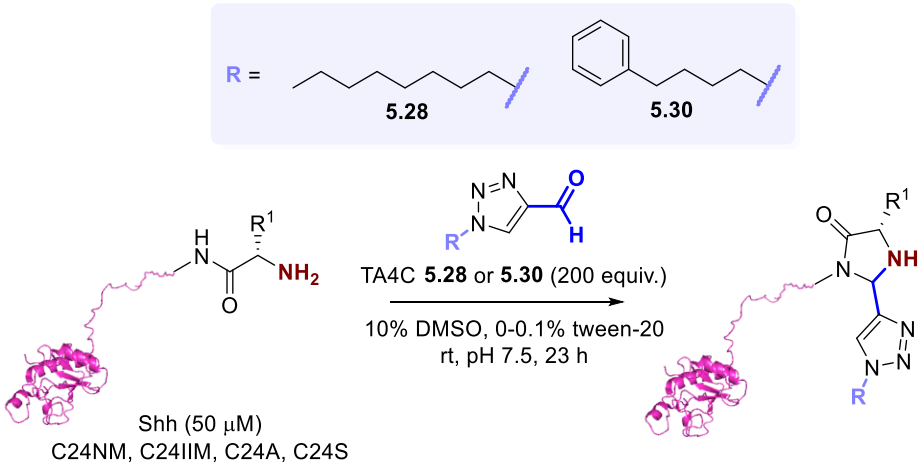
Conditions	**Conversion (%)**
(a)	**13**
(f)	**25**
(g)	**19**
(h)	**31**

Table S5.14. Optimisation of solubilising additive levels. Modification of Shh C24IIM with TA4C **5.28**, conjugates prepared with a range of solubilising additives.

DMSO level (%)	**Additive**	**Additive level (%)**	**Measured concentration (µM)**
10	-	-	13
10	Glycerol	0.1	0
10	Glycerol	1	0
10	Glycerol	10	10
10	Tween-20	0.01	11
10	Tween-20	0.1	10
10	Tween-20	1	0

Table S5.15. Concentrations of Shh conjugates measured by UV-vis spectroscopy at 280 nm (most concentrated fractions shown in table). Calculation of concentrations based on 1% extinction coefficient of 13.75 for Shh C24II.

5.4.2.13 Modification of N-terminal mutants at optimised tween-20 levels



Shh was modified with reagents **5.28** and **5.30**, according to an adapted version of **General Procedure 5D:** *i)* A stock solution of amine **5.33** or **5.35** (40 µL, 300 mM, 12 µmol, 1.5 equiv., in DMSO) and a solution of MOPS (2 µL, 400 mM, 0.8 µmol, 0.1 equiv., in H_2O) were added to a solution of TA4C **5.18** (40 µL, 200 mM, 8 µmol, 1 equiv., in DMSO), and the reaction mixture was incubated at 100 °C for 2 h. *ii)* A solution of protein (45 µL, 111 µM, 5 nmol, 1 equiv., in 50 mM pH 7.5 Na phosphate buffer) was modified with an aliquot of the

crude TA4C reagent from part *i)* (10 µL, 100 mM, 1 µmol, 200 equiv.), with 50 mM pH 7.5 Na phosphate buffer (35 µL) and a solution of tween-20 (10 µL, 1% in 50 mM pH 7.5 Na phosphate buffer) at rt for 23 h. Modified Shh samples were purified by size exclusion chromatography (PD MiniTrap G-25) with sample collection in 50 µL fractions, and conversion was determined by LC-MS analysis.

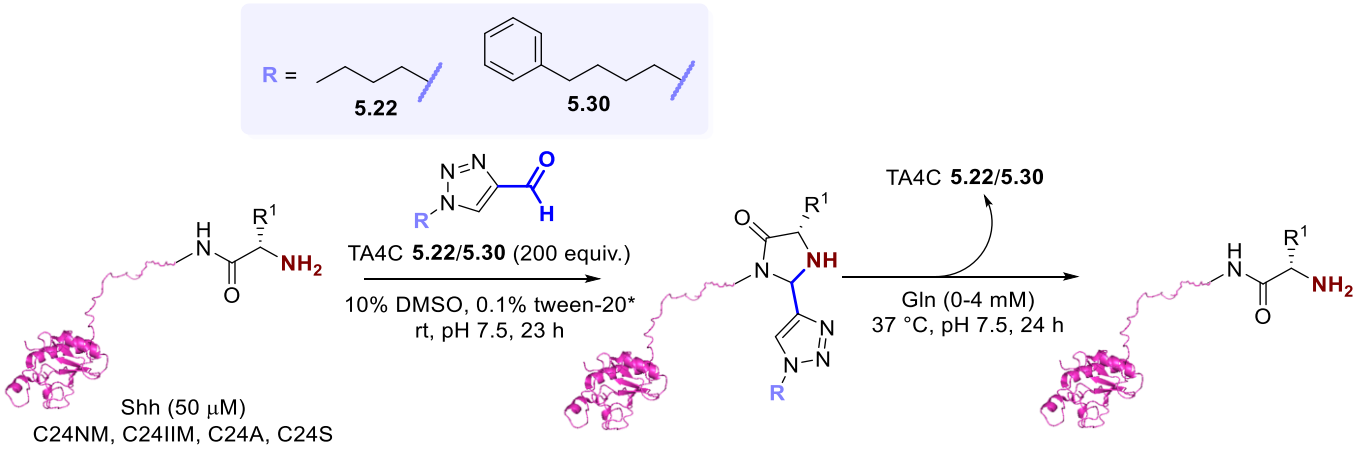
Small molecule	Conversion (%)			
	Shh C24NM	Shh C24IIM	Shh C24A	Shh C24S
5.28	**53**	**31**	**11**	**7**
5.30	**54**	**33**	**10**	**9**

*Table S5.16. Modification of Shh with TA4Cs **5.28** and **5.30** under solubilising conditions.* Modification of Shh C24IIM, C24NM, C24A and C24S with TA4Cs **5.28** and **5.30**, conjugates prepared with 0.1% tween-20.

Mutant	Small molecule	Measured concentration (µM)
C24NM	TA4C **5.28**	9
	TA4C **5.30**	13
C24IIM	TA4C **5.28**	10
	TA4C **5.30**	14
C24A	TA4C **5.28**	8
	TA4C **5.30**	12
C24S	TA4C **5.28**	9
	TA4C **5.30**	12

Table S5.17. Concentrations of Shh conjugates measured by UV-vis spectroscopy at 280 nm (most concentrated fractions shown in table). Calculation of concentrations based on 1% extinction coefficient of 13.75 for Shh C24II, 13.85 for C24A, 13.82 for C24N and 13.84 for C24S.

*5.4.2.14 Stability of TA4C **5.22** and **5.30** conjugates*



Stability was assessed using purified samples of C24NM, C24IIM, C24A and C24S conjugates modified with TA4Cs **5.22** and **5.30**, prepared previously in sections *5.4.2.11* and *5.4.2.13*. Aliquots of protein (25 µL, 50

µM, 1.25 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer)** were incubated at 37 °C for 24 h, and analysed via LC-MS after 24 h. Aliquots of protein (25 µL, 50 µM, 1.25 nmol, 1 equiv., in 50 mM pH 7.5 sodium phosphate buffer)** were also added to a solution of L-glutamine (25 µL, 8 mM, 0.2 µmol, 160 equiv.), incubated at 37 °C for 24 h, and analysed via LC-MS after 24 h.

0.1% tween-20 for TA4C **5.30 only*

***Expected concentrations assuming 100% retention of protein in solution following purification; in reality, the concentrations were lower, as reported in Figures S5.13 and S5.18.*

Gln concentration (mM)	Small molecule	N-terminal mutant	Conversion (%)	
			0 h	24 h
0	**5.22**	C24NM	**60**	**9**
		C24IIM	**56**	**0**
		C24A	**52**	**45**
		C24S	**32**	**26**
0	**5.30**	C24NM	**43**	**0**
		C24IIM	**13**	**0**
		C24A	**7**	**5**
		C24S	**0**	**0**
4	**5.22**	C24NM	**60**	**-***
		C24IIM	**56**	**-**
		C24A	**52**	**-**
		C24S	**32**	**-**
4	**5.30**	C24NM	**43**	**0**
		C24IIM	**13**	**5**
		C24A	**7**	**0**
		C24S	**0**	**0**

Table S5.18. Stability of TA4C conjugates to incubation at 37 °C with 0 and 4 mM L-glutamine. Modification of Shh C24IIM, C24NM, C24A and C24S with TA4Cs **5.22** and **5.30**, followed by incubation at 37 °C for 24 h with 0 and 4 mM L-glutamine. **Data quality too poor for analysis*.

5.4.3 Gli-luciferase assay

Note: assays in *Fig. 5.8, 5.10, 5.22, 5.24, 5.34,* and *5.35* were carried out by Dr. Kerry Price at Qkine.

For the assays in Fig. 5.6, 5.8, 5.10, 5.22, 5.24, 5.34, and 5.35. The Gli-luciferase assay was carried out according to **General Procedure 5E**.

For the assays in Fig. 5.6, 5.7, 5.39, 5.40 and 5.44. The Gli-luciferase assay was carried out according to an adapted version of **General Procedure 5E**, with a cell plating concentration of 3×10^5 cells/mL.

General Procedure 5E. The Gli Luciferase Reporter NIH3T3 Cell Line (Hedgehog pathway) was purchased from BPS Bioscience, and they were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented

with 10% Calf Serum (CS) and 500 µg/mL Geneticin in a 5% CO_2 incubator at 37 °C. 96-well plates were coated with poly-L-lysine (0.005%, 20 µL per well) at rt for 1 h. The wells were aspirated and allowed to dry for at least 1 h. Gli-3T3 cells were seeded into the pre-coated 96-well plates (1×10^5 cells/mL in 100 µL of medium per well) and cultured in a 5% CO_2 incubator at 37 °C for 24 h. The medium was aspirated, the cells were treated with a concentration gradient of Shh (0~10 µg/mL in OptiMEM supplemented with 0.5% Calf Serum) and cultured in a 5% CO_2 incubator at 37 °C for 24 h. The medium was aspirated, and the cells were treated with 1× passive lysis buffer (20 µL per well) on the microplate shaker for 15 min. Luciferase activity was measured using a FLUOstar Omega Microplate Reader, with cell lysate (5 µL) and firefly substrate (25 µL). To prepare firefly substrate, NaOH (2 M, 570 µL), followed by $(MgCO_3)_4 \cdot Mg(OH)_2 \cdot 5H_2O$ (50 mM, 1.21 mL), was added to a solution of tricine (0.817 g), $MgSO_4 \cdot 7H_2O$ (0.5 M, 1.21 mL), EDTA (0.5 M, pH 8.0, 46.5 µL), DTT (1.17 g), ATP (10 mM, 12 mL), acetyl coenzyme A (25 mg/mL, 1.89 mL) and firefly D-luciferin (30 mg) in MilliQ water (207 mL). “Unmodified” samples were prepared under identical conditions to the modified samples, in the absence of modification reagent. “Shh C24IIM Qkine” samples were freshly reconstituted Shh C24IIM samples prepared immediately before the protein treatment stage of the assay at Qkine.

5.5 References

1. Choudhry, Z.; Rikani, A. A.; Choudhry, A. M.; Tariq, S.; Zakaria, F.; Asghar, M. W.; Sarfraz, M. K.; Haider, K.; Shafiq, A. A.; Mobassarah, N. J. *Ann. Neurosci.* **21**, 28-31 (2014).
2. Yang, C.; Qi, Y.; Sun, Z. *Front. Mol. Biosci.* **8**, 711710 (2021).
3. Yan, Y.; Bejoy, J.; Xia, J.; Guan, J.; Zhou, Y.; Li, Y. *Acta Biomater.* **42**, 114–126 (2016).
4. Terol, G. L.; Gallego-Jara, J.; Martínez, R. A. S.; Vivancos, A. M.; Díaz, M. C.; Puente, T. D. *Front. Microbiol.* **12**, 682001 (2021).
5. Rosano, G. L.; Ceccarelli, E. A. *Front. Microbiol.* **5**, 172 (2014).
6. Varland, S.; Osberg, C.; Arnesen, T. *Proteomics* **15**, 2385–2401 (2015).
7. Farzan, S. F.; Singh, S.; Schilling, N. S.; Robbins, D. J. *Am. J. Physiol.: Gastrointest. Liver Physiol.* **294**, G844–G849 (2008).
8. Hardy, R. Y.; Resh, M. D. *J. Biol. Chem.* **287**, 42881–42889 (2012).
9. Buglino, J. A.; Resh, M. D., in *Hedgehog Signaling*, ed. Litwack, G., Academic Press, vol. 88, 2012, Chapter 10, 229–252.
10. Resh, M. D. *Open Biol.* **11**, 200414 (2021).
11. Ciepla, P.; Konitsiotis, A. D.; Serwa, R. A.; Masumoto, N.; Leong, W. P.; Dallman, M. J.; Magee, A. I.; Tate, E. W. *Chem. Sci.* **5**, 4249–4259 (2014).
12. Taylor, F. R.; Wen, D.; Garber, E. A.; Carmillo, A. N.; Baker, D. P.; Arduini, R. M.; Williams, K. P.; Weinreb, P. H.; Rayhorn, P.; Hronowski, X.; Whitty, A.; Day, E. S.; Boriack-Sjodin, A.; Shapiro, R. I.; Galdes, A.; Pepinsky, R. B. *Biochemistry* **40**, 4359–4371 (2001).
13. Qi, X.; Schmiege, P.; Coutavas, E.; Wang, J.; Li, X. *Nature* **560**, 128–132 (2018).
14. Zhang, Y.; Beachy, P. A. *Nat. Rev. Mol. Cell Biol.* **24**, 668–687 (2023).
15. Zhang, Y.; Bulkley, D. P.; Xin, Y.; Roberts, K. J.; Asarnow, D. E.; Sharma, A.; Myers, B. R.; Cho, W.; Cheng, Y. Beachy, P. A. *Cell* **175**, 1352–1364 (2018).
16. Jia, B.; Jeon, C. O. *Open Biol.* **6**, 160196 (2016).
17. Yadav, P.; Ayana, R.; Garg, S.; Jain, R.; Sah, R.; Joshi, N.; Pati, S.; Singh, S. *FEBS Open Bio* **9**, 248–264 (2019).
18. Qi, X.; Schmiege, P.; Coutavas, E.; Li, X. *Science* **362**, eaas8843 (2018).
19. Barber, L. J.; Yates, N. D. J.; Fascione, M. A.; Parkin, A.; Hemsworth, G. R.; Genever, P. G.; Spicer, C. D. *RSC Chem. Biol.* **4**, 56–64 (2023).
20. Sherman, F.; Stewart, J. W.; Tsunasawa, S. *BioEssays* **3**, 27–31 (1985).
21. Agu, P. C.; Afiukwa, C. A.; Orji, O. U.; Ezeh, E. M.; Ofoke, I. H.; Ogbu, C. O.; Ugwuja, E. I.; Aja, P. M. *Sci. Rep.* **13**, 13398 (2023).
22. MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. *Nat. Chem. Biol.* **11**, 326–331 (2015).
23. Mikkelsen, J. H.; Gustafsson, M. B. F.; Skrydstrup, T.; Jensen, K. B. *Bioconjugate Chem.* **33**, 625–633 (2022).
24. Chatani, E.; Hayashi, R.; Moriyama, H.; Ueki, T. *Protein Sci.* **11**, 72–81 (2002).

25. Onoda, A.; Inoue, N.; Sumiyoshi, E.; Hayashi, T. *ChemBioChem* **21**, 1274–1278 (2019).
26. Carey, F. A.; Giuliano, R. M. *Organic Chemistry*, McGraw-Hill, 2011.
27. Vagenende, V.; Yap, M. G. S.; Trout, B. L. *Biochemistry* **48**, 11084–11096 (2009).
28. Yamaguchi, H.; Miyazaki, M. *Biomolecules* **4**, 235–251 (2014).
29. Sun, C.; Li, Y.; Yates, E. A.; Fernig, D. G. *Protein Sci.* **29**, 19–27 (2020).
30. Wen, X.; Lai, C. K.; Evangelista, M.; Hongo, J.-A.; Sauvage, F. J.; Scales, S. J. *Mol. Cell. Biol.* **30**, 1910–1922 (2010).
31. Liao, Y.-D.; Jeng, J.-C.; Wang, C.-F.; Wang, S.-C.; Chang, S.-T. *Protein Sci.* **13**, 1802–1810 (2004).
32. Fletcher, J. T.; Christensen, J. A.; Villa, E. M. *Tetrahedron Lett.* **58**, 4450–4454 (2017).

Future Work: Functionalisation of PVA

In Chapter 4, we demonstrated that the kinetics and stability of imidazolidinone formation can be tuned using PCA derivatives with various substituents. Whilst the substrates explored did not exhibit levels of stability high enough for therapeutic application *in vivo*,[1] in future work we hope to use these tunable properties to our advantage. For example, the potential application of PCA derivatives to haematopoietic stem cell culture is outlined below.

6.1 Introduction

Haematopoietic stem cells (HSCs) are found in adult bone marrow and umbilical cord blood (UCB) during pregnancy and can differentiate into all blood cell types. Clinical applications include the renewal of the haematopoietic system following HSC transplantation, providing a curative therapy for haematological diseases such as leukaemia. Whilst HSCs can be obtained by collection of UCB (a less invasive approach than bone marrow biopsy), the cell numbers obtained are too low for effective and safe HSC transplantation, with delayed engraftment and risk of immune reconstitution inflammatory syndrome (IRIS). The maintenance and stable expansion of HSCs *ex vivo* could provide a key approach to obtain sufficient cell numbers for transplantation.[2,3]

In recent years, advances in the field have led to the development of cell culture systems capable of expanding mouse HSCs.[2–4] In initial work, HSC maintenance factors such as haematopoietic cytokines were identified through characterisation of the local tissue HSC microenvironments in bone marrow *in vivo*.[5] Common HSC cultures included cytokines in combination with human serum albumin (HSA) but were only suitable for short-term maintenance of HSCs, without expansion.[4] Despite the identification of culture constituents capable of maintaining HSCs, *ex vivo* expansion of HSCs has been previously unattainable due to media supplement impurities and poor optimisation of the levels of culture components.[2]

Wilkinson *et al.* developed a cell culture system in which the conventional HSA, required as a source of amino acids and as a “carrier molecule”, was replaced by polyvinyl alcohol (PVA). In addition to superior function for supporting HSC survival, maintenance and growth, replacement with PVA introduced fewer biological contaminants thought to be responsible for inflammation and cell differentiation into the culture system. Optimal levels of the growth factors thrombopoietin (TPO) and stem-cell factor (SCF) were identified, in combination with plate coating with fibronectin, an HSC maintenance factor. Fibronectin plate coating functioned to enable HSC attachment, to retain HSCs during complete medium changes.[2]

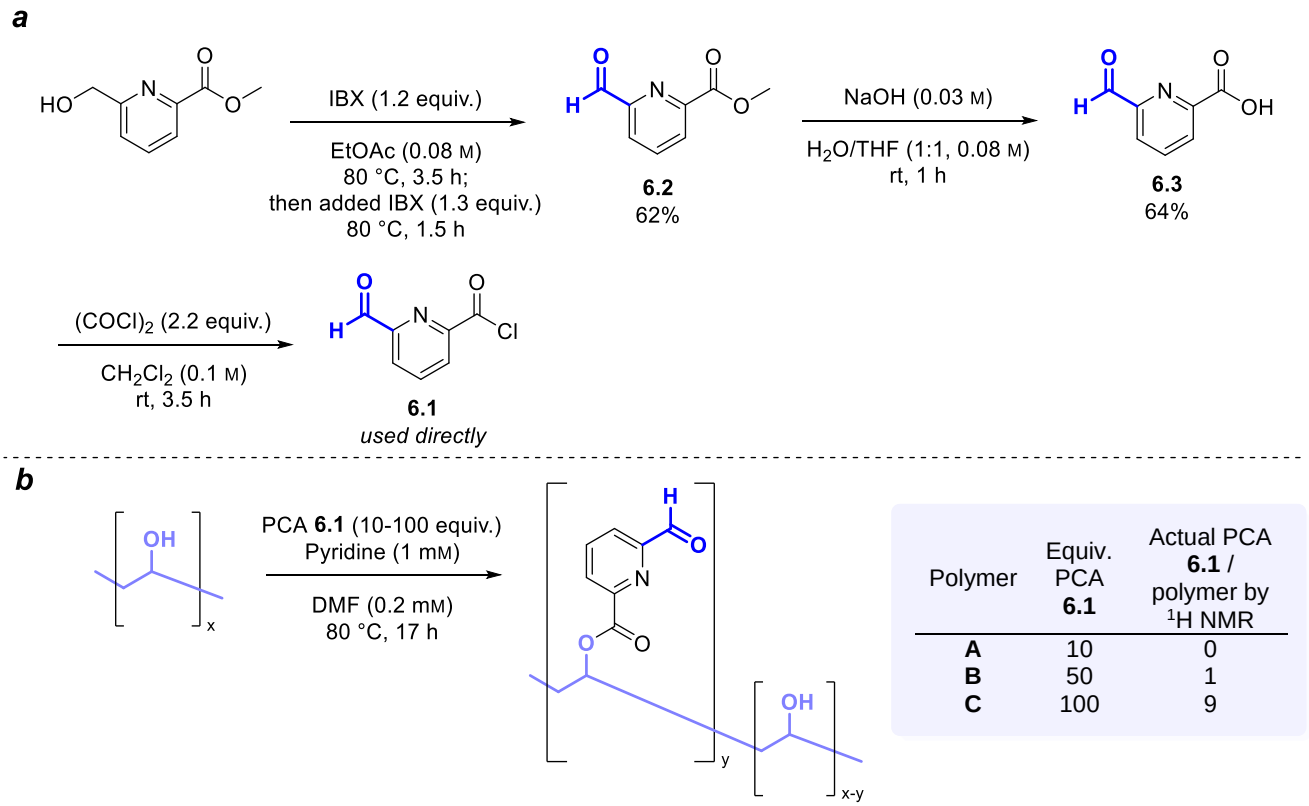
The relative levels of TPO and SCF were found to play an important role in the maintenance of HSCs, with high TPO levels and low SCF levels found to work synergistically.[2] We proposed that the functionalisation of PVA with 2-PCAs capable of N-terminal protein modification could enable the attachment of PVA to fibronectin or haematopoietic cytokines. In our previous work, we found that the introduction of substituents

onto 2-PCAs had an influence on kinetics and stability of protein modification reactions, allowing these properties to be tuned *via* substituent variation. Due to the tuneable stability of imidazolidinones under cell culture conditions (37 °C, with competitive nucleophiles), we proposed that the use of cytokine-PVA conjugates could allow the controlled release of different free cytokines over time into the HSC culture medium at tunable rates, allowing the relative levels of cytokines to be controlled.

6.2 Preliminary work

6.2.1 Polymer Synthesis

Firstly, 2-PCA **6.1** was synthesised *via* oxidation of methyl 6-(hydroxymethyl)pyridine-2-carboxylate, hydrolysis of ester **6.2**, and subsequent conversion of the resulting carboxylic acid **6.3** to acid chloride **6.1** (*Fig. 6.1a*). Due to poor bench stability of acid chlorides, crude PCA **6.1** was directly coupled with PVA immediately after production. We expected a variation in the level of 2-PCA functionalisation of PVA to influence the extent of protein-polymer bioconjugation and thus allow us to probe the influence of conjugation on the cell culture. We therefore functionalised PVA with loadings of 10, 50 and 100 equiv. 2-PCA per individual polymer strand: the level of functionalisation achieved was estimated to be low, with an average of 0, 1 and 9 2-PCA units per polymer respectively (*Fig. 6.1b*, by ^{1}H NMR spectroscopy; diagnostic signals highlighted in *Fig. S6.1*).



Polymer	Equiv. PCA **6.1**	Actual PCA **6.1** / polymer by ^{1}H NMR
A	10	0
B	50	1
C	100	9

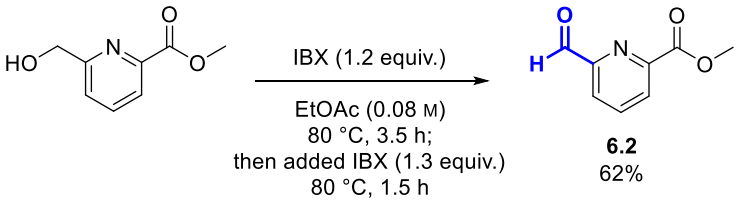
Figure 6.1. Functionalisation of PVA with 2-PCAs *via*: ***a)*** Synthesis of acid chloride-functionalised 2-PCA **6.1**; ***b)*** PVA/acid chloride coupling.

6.2.2 Next Steps

Having verified the levels of 2-PCA attachment to the PVA, future work in this project will involve validating that the 2-PCA moieties of the polymers are capable of protein modification, in order to prepare PVA-cytokine conjugates and ultimately apply these conjugates to cell culture systems.

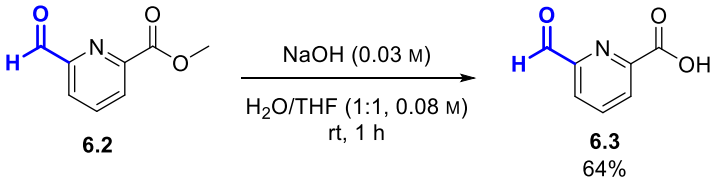
6.3 Experimental

Methyl 6-formylpicolinate (6.2)



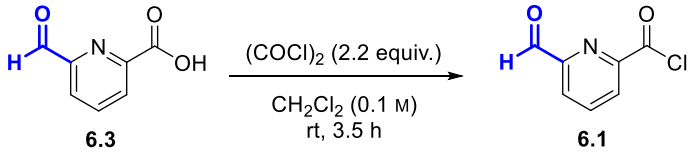
2-iodoxybenzoic acid (45 wt.%, 3.37 g, 5.4 mmol, 1.2 equiv.) was added to a solution of methyl 6-(hydroxymethyl)picolinate (0.75 g, 4.5 mmol, 1.0 equiv.) in ethyl acetate (57 mL, 0.08 M), and the reaction mixture was stirred at 80 °C for 3.5 h under nitrogen. At this point TLC indicated that the reaction was incomplete, so another portion of 2-iodoxybenzoic acid (45 wt.%, 3.37 g, 5.9 mmol, 1.3 equiv.) was added, and the reaction mixture was stirred at 80 °C for 1.5 h. After cooling to room temperature, the reaction mixture was filtered through Celite, and the filtrate was concentrated under reduced pressure to afford a white solid. The solid was purified by flash column chromatography (0-30% EtOAc:Petrol, R_f 0.58 in 50% EtOAc), and pure fractions were concentrated under reduced pressure to afford the title compound (0.46 g, 2.8 mmol, 62%) as a white solid with spectroscopic data in accordance with the literature.[6] **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H: 4.07 (3H, s, Me), 8.06 (1H, dd, $J_1 = J_2$ = 7.8 Hz, ArH4), 8.16 (1H, d, J = 7.8 Hz, ArH3 or ArH5), 8.36 (1H, d, J = 7.8 Hz, ArH3 or ArH5), 10.20 (1H, s, CHO).

6-formylpicolinic acid (6.3)



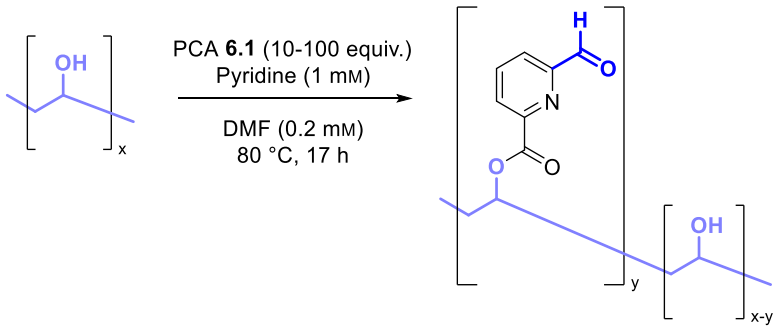
NaOH (1 M in water, 8 mL, 8 mol) was added to a solution of compound **6.2** (402 mg, 2.4 mmol, 1.0 equiv.) in 1:1 THF:H_2O (30 mL, 0.08 M), and the reaction mixture was stirred at room temperature for 1 h. The reaction mixture was concentrated under reduced pressure, and the residue obtained was acidified with HCl (1 M in water, 10 mL), and extracted with EtOAc (4 × 50 mL). The combined organics were dried over $MgSO_4$, filtered, and concentrated under reduced pressure to afford the title compound (238 mg, 1.6 mmol, 64%) as a yellow solid with spectroscopic data in accordance with the literature.[7] **^{1}H NMR** (400 MHz, DMSO-d_6) δ_H: 8.11 (1H, d, J = 7.7 Hz, ArH3 or ArH5), 8.22 (1H, dd, $J_1 = J_2$ = 7.7 Hz, ArH4), 8.30 (1H, d, J = 7.7 Hz, ArH3 or ArH5), 10.02 (1H, s, CHO).

6-formylpicolinoyl chloride (6.1)



Oxalyl chloride (63 µL, 0.7 mmol, 2.2 equiv.) was added dropwise to a solution of compound **6.3** (49 mg, 0.3 mmol, 1.0 equiv.) in dichloromethane (3 mL, 0.1 M) [WARNING: Gas evolved], and the reaction mixture was stirred at room temperature for 3.5 h. The reaction mixture was then concentrated under reduced pressure and azeotroped with CH_2Cl_2 (2 × 10 mL). The residue obtained (55 mg) was used immediately in the subsequent reaction without characterisation or purification due to poor bench stability of acid chlorides.

PCA-functionalised PVA



Procedure 6A: *Synthesis of 2-PCA-PVA.* PVA was purchased from Sigma-Aldrich (87-90% hydrolysed, average M_w 30-70 kDa). A stock solution of 2-PCA **6.1** (148 mM in DMF; see below for individual parameters) was added drop-wise to a solution of PVA (91 mg, 1.8 µmol, 1.0 equiv.) and pyridine (909 µL, 11.2 mmol) in DMF (9 mL, 0.2 mM), and the reaction mixture was stirred at 80 °C for 17 h. After cooling to room temperature, the solvent was removed under reduced pressure and the residue obtained was dissolved in water (10 mL). 2-PCA-PVAs were purified by dialysis to remove excess reagent (rt, 10 kDa MWCO; 1 × water, 16 h; 1 × water, 3 h; 1 × water, 22 h; 1 × water, 7 h), and lyophilised to afford the title compound as a white solid with 2-PCA moieties conjugated primarily in the hydrate form. The number of 2-PCA units per polymer was calculated by comparison of the intensities of -C<u>H</u>$_2$- and -ArC<u>H</u>(OH)$_2$ signals highlighted in *Fig. S6.1*, based on the assumption that the average number of monomer units per polymer strand was 1136. **^{1}H NMR** (400 MHz, D_2O) $\delta_{H:}$ 1.41-2.13 (2H, m, -C<u>H</u>$_2$-), 3.90-4.11 (1H, m, -C<u>H</u>-), 5.83 (<1H, s, -ArC<u>H</u>(OH)$_2$), 7.83 (<1H, s, Ar<u>H</u>), 8.07 (<1H, s, Ar<u>H</u>).

*For the polymer prepared using 10 equiv. 2-PCA **6.1**:* reagent **6.1** (123 µL, 148 mM, 18 µmol, 10 equiv., in DMF) was applied under the conditions outlined in **General Procedure 6A** to afford the title compound (48 mg, 1.0 µmol, 53%, *yield calculated assuming y = 0, as estimated by ^{1}H NMR*).

*For the polymer prepared using 50 equiv. 2-PCA **6.1**:* reagent **6.1** (610 µL, 148 mM, 90 µmol, 50 equiv., in DMF) was applied under the conditions outlined in **General Procedure 6A** to afford the title compound (55 mg, 1.1 µmol, 60%, *yield calculated assuming y = 1, as estimated by ^{1}H NMR*).

*For the polymer prepared using 100 equiv. 2-PCA **6.1**:* reagent **6.1** (1.23 mL, 148 mM, 182 µmol, 100 equiv., in DMF) was applied under the conditions outlined in **General Procedure 6A** to afford the title compound (55 mg, 1.1 µmol, 59%, *yield calculated assuming y = 9, as estimated by ^{1}H NMR*).

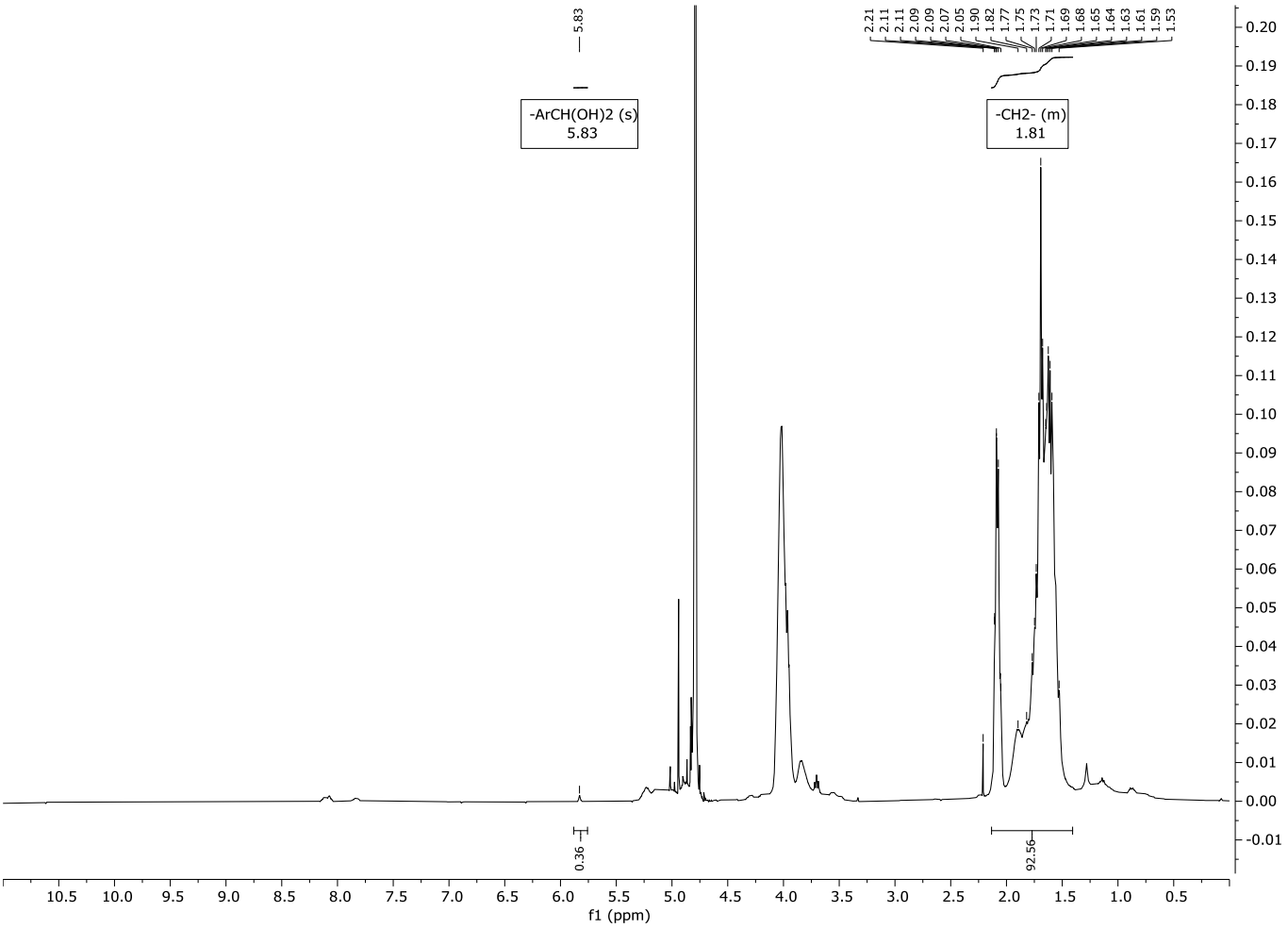


Figure S6.1. Example ^{1}H NMR spectrum of PCA-functionalised PVA (PCA loading: 100 equiv.) showing the diagnostic signals used to calculate the level of PCA functionalisation.

6.4 References

1. Sangsuwan, R.; Tachachartvanich, P.; Francis, M. B. *J. Am. Chem. Soc.* **141**, 2376–2383 (2019).
2. Wilkinson, A. C.; Ishida, R.; Kikuchi, M.; Sudo, K.; Morita, M.; Crisostomo, R. V.; Yamamoto, R.; Loh, K. M.; Nakamura, Y.; Watanabe, M.; Nakauchi, H.; Yamazaki, S. *Nature* **571**, 117–121 (2019).
3. Grey, W.; Chuahan, R.; Piganeau, M.; Encabo, H. H.; Garcia-Albornoz, M.; MacDonald, N. Q.; Bonnet, D. *Blood* **136**, 2535–2547 (2020).
4. Sakurai, M.; Ishitsuka, K.; Ito, R.; Wilkinson, A. C.; Kimura, T.; Mizutani, E.; Nishikii, H.; Sudo, K.; Becker, H. J.; Takemoto, H.; Sano, T.; Kataoka, K.; Takahashi, S.; Nakamura, Y.; Kent, D. G.; Iwama, A.; Chiba, S.; Okamoto, S.; Nakauchi, H.; Yamazaki, S. *Nature* **615**, 127–133 (2023).
5. Morrison, S. J.; Scadden, D. T. *Nature* **505**, 327–334 (2014).
6. Shi, X.-F.; Wang, M.-M.; Huang, S.-C.; Han, J.-X.; Chu, W.-C.; Ciao, C.; Zhang, E.; Qin, S. *Eur. J. Med. Chem.* **167**, 367–376 (2019).
7. Bush, J. T.; Leśniak, R. K.; Yeh, T.-L.; Belle, R.; Kramer, H.; Tumber, A.; Chowdbury, R.; Flashman, E.; Mecinović, J.; Schofield, C. J. *Chem. Commun.* **55**, 1020–1023 (2019).

Overall Conclusion

In this thesis, we initially identified a need to directly compare and elucidate more mechanistic and kinetic insight into the factors governing selectivity, reactivity, and stability of N-terminal labelling strategies developed over the last 10 years. Our comparative study detailed in Chapter 2 therefore provides a useful guide to researchers on how to select a suitable labelling strategy for a target protein, advocating for the screening of a panel of reagents to ensure optimal conjugation with suitable properties for the given protein/application is realised. This work also supports the need to develop new modification strategies addressing current limitations in the field.

One of the key challenges highlighted by our attempts to improve the kinetics and stability of N-terminal labelling is that the conjugate properties often reflect a balance between multiple different effects; often the improvement of one property for a given application may be at the expense of another. Our study on reagent functionalisation detailed in Chapter 4 demonstrated that the properties of N-terminal labelling using 2-pyridinecarboxaldehydes (2-PCAs) can be tuned for a given application *via* functionalisation. Promisingly, we identified reagents with intramolecular hydrogen bonding acceptor potential as key candidates to include in reagent libraries for enhanced protein conversion. Our approach therefore provides a valuable step forward in the goal to achieve a bioconjugation strategy applicable to a wide range of protein targets with high efficiency, selectivity, and stability. Our findings complement other recent advances in 2-PCA methodology, for example where Hanaya *et al.* improved conjugate stability *via* addition of a copper(II) salt, but with loss of control of the number of modification sites.[1]

An exciting recent development in N-terminal labelling is the design of reversible modification strategies with a switchable cage/decage process; for example, Lin *et al.* reported a strategy for reversible palladium-catalysed N-terminal cinnamylation, where palladium-catalysed deconjugation can be promoted by addition of 1,3-dimethylbarbituric acid.[2] Development of strategies with an “on”/”off” switch opens up new opportunities for applications such as temporary protection strategies in the semi-synthesis of proteins,[3] or the traceless enrichment of target proteins for proteomics analysis.[4] We demonstrated that the release rate of peptides/proteins from 2-PCA conjugates can also be controlled, either gradually over time through reagent functionalisation to tune the equilibria governing the labelling process (Chapter 4), or accelerated through addition of a competitive nucleophile such as dialanine (Chapter 2), thus expanding the possible applications of 2-PCAs.

In summary, this thesis has contributed to the field through *i)* identifying a reagent screening approach to select a suitable labelling strategy for a given application; *ii)* highlighting current limitations of leading methodologies and the challenges in developing new strategies with optimal properties; and *iii)* identifying key derivatives of PCA reagents to improve kinetics and conversion, adding to the toolbox of N-terminal modification reagents and advancing current modification strategies.

References

1. Hanaya, K.; Yamoto, K.; Taguchi, K.; Matsumoto, K.; Higashibayashi, S.; Sugai, T. *Chem. - Eur. J.* **28**, e202201677 (2022).
2. Lin, Z.; Liu. B.; Lu, M.; Wang, Y.; Ren, X.; Liu, Z.; Luo, C.; Shi, W.; Zou, X.; Song, X.; Tang, F.; Huang, H.; Huang, W. *J. Am. Chem. Soc.* **146**, 23752–23763 (2024).
3. Wang, S.; Zhou, Q.; Li, Y.; Wei, B.; Liu, X.; Zhao, J.; Ye, F.; Zhou, Z.; Ding, B.; Wang, P. *J. Am. Chem. Soc.* **144**, 1232–1242 (2022).
4. Cowell, J.; Buck, M.; Essa, A. H.; Clarke, R.; Vollmer, W.; Vollmer, D.; Hilkens, C. M.; Isaacs, J. D.; Hall, M. J.; Gray, J. *ChemBioChem* **18**, 1688–1691 (2017).