Development of pyrazole and triazole analogues of phosphohistidine for use in investigatory probes

Jake David Wheeler

Submitted in accordance with the requirement for the degree of Doctor of Philosophy

The University of Leeds

School of Chemistry

December 2025

The candidate confirms that the work submitted is their own and that appropriate credit has been given where reference has been made to the work of others.

Reference is made where the work of others has been included. The contributions of other authors to this work have been explicitly declared below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others. The following are to be acknowledged for their contributions:

Anti CTB 24AA affimer expression (Tomasz Kaminski);
Synthesis of EGFR probe (Nicholas Morris);
DIA-NN proteomics analysis - volcano plots (Brian Suarez Mantilla);
RP expression (Zhenlian Ling)
SNAC/PEGA-surugamide synthesis and cyclisation (Asif Fazal)

Acknowledgements

I am very grateful for the support given by my supervisors, Professor Michael Webb and Dr Megan Wright over the course of the last 4 years; their insights have helped me immeasurably in producing this work, their assistance has made a challenging project significantly more manageable.

I am indebted to the financial support given by the BBRSC White Rose Mechanistic Biology Doctoral Training Programme, without whom this work would not have been possible and for providing regular training opportunities over the course of the programme.

I would also like to extend my gratitude to all the members of Chemistry Lab 1.49, who have made my time working in the laboratory genuinely wonderful. I am extremely thankful for the support I received from Dr Tameryn Stringer and Dr Pablo Caramés-Méndez, who provided me with invaluable assistance in the laboratory in the early stages of my research. I would also like to thank Jacob Webb for being a trusty diazirine synthesis companion and for his welcome support during last minute challenges; Charlie Stevenson for her constant level-headed support, sensible advice and those much-needed coffee breaks and Chloe Shingler for contending with my leftover peptide mess and of course for our great conversations.

I have to reserve my deepest thank you to my family, who have constantly been supportive both before and during my research (and hopefully after). My parents have provided me with huge support and have even had to take me back under their roof again during my write-up period (sorry about that) - I cannot thank them enough. I am hugely grateful to my sisters for their support throughout.

I would like to dedicate this work in memory of my grandparents, who have supported me and wished me nothing but success in my academic pursuits and life.

Table of Contents

List of Abbreviations

acetyl-CoA	acetyl-coenzyme A
ACL/ACLY	ATP-citrate lyase
ADP	adenosine diphosphate
AEX	anion exchange
AKT	protein kinase B
AMP	adenosine monophosphate
ASK1	apoptosis signalling kinase 1
ATP	adenosine triphosphate
CAM	calmodulin
CD4	cluster of differentiation 4
CKII	casein kinase II
CoA	coenzyme A
CRAC	calcium release activated channels
CTB	cholera toxin B
DAST	diethylaminosulfur triflouride
DBA	dibromoadipamide
DCM	dichloromethane
DFT	density functional theory
DIA-NN	Data-Independent Acquisition by Neural Networks
DIC	N,N′-Diisopropylcarbodiimide
DIPEA	N,N-Diisopropylethylamine
DMF	N,N-Dimethylformamide
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
EDT	ethane dithiol
EDTA	ethylenediaminetetraacetic acid
EGF/EGFR	epidermal growth factor receptor
EIC	extracted ion chromatogram
ER	endoplasmic reticulum
ERH	enhancer of rudimentary homolog

ESI/ESMS	electrospray ionisation
ETT	5-(ethylthio)-1H-tetrazole
FDR	false discovery rate
FHA	forkhead-associated
FKBP1	FK506-binding protein 1
FP	fluorophosphonate
FPLC	fast protein liquid chromatography
FPR1	formyl peptide receptor 1
FUNDC1	FUN14 domain-containing 1
G3BP1	Ras GTPase-activating protein-binding protein 1
GAP	GTPase-activating protein
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GDP	guanosine diphosphate
GEF	guanine nucleotide exchange factor
GNB1	guanine nucleotide-binding protein subunit beta-1
GPCR	G-protein coupled receptors
GRK2	G-protein coupled receptor kinase 2
GRK5	G protein-coupled receptor kinase 5
GSK3	glycogen synthase kinase-3
GTP	guanosine triphosphate
HCTU	O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HFIP	hexafluoroisopropanol
HOBt	hydroxybenzotriazole
HOSA	hydroxylamine-O-sulfonic acid
HPLC	high-performance liquid chromatography
HRMS	high-resolution mass spectrometry
HSQC	heteronuclear single quantum coherence
IMAC	immobilised metal affinity chromatography
ITC	isothermal titration calorimetry
KCa3.1	calcium-activated potassium channel
KLH	keyhole limpet haemocyanin

LCMS	liquid chromatography-mass spectrometry
LFQ	label-free quantification
LHPP	phospholysine phosphohistidine inorganic pyrophosphate phosphatase
MALDI-TOF	matrix-assisted laser desorption ionization–time of flight
MAPK	mitogen-activated protein kinase
MOAC	metal oxide affinity chromatography
MS	mass spectrometry
NAD	nicotinamide adenine dinucleotide
NAMPT	nicotinamide phosphoribosyltransferase
NBS	N-bromosuccinimide
NDPK	nucleoside diphosphate kinase
NFAT	nuclear factor of activated T-cells
NME	nucleoside diphosphate kinase
NME1	nucleoside diphosphate kinase A
NME2	nucleoside diphosphate kinase B
NME4	nucleoside diphosphate kinase D
NMR	nuclear magnetic resonance
NRPS	non-ribosomal peptide synthetase
NTP	nucleoside triphosphate
OPA1	dynamin-like GTPase OPA1
PAG	photoactivatable groups
PAL	photoactivatable linker
PAP	prostatic acid phosphatase
PBS	phosphate-buffered saline
PCP	peptidyl carrier protein
PDRP	pyruvate orthophosphate dikinase
PEGA	polyethylene glycol acrylamide
PEP	phosphoenolpyruvate
PEPC	phosphoenolpyruvate carboxylase
PGAM	phosphoglycerate mutase
PGAM1	phosphoglycerate mutase 1
PGAM5	phosphoglycerate mutase 5

pHis	phosphohistidine
PHPT1	phosphistidine phosphatase 1
PLD	phospholipase D1
PPDK	pyruvate phosphate dikinase
PPIase	peptidyl prolyl isomerase
pSer	phosphoserine
PTB	phosphotyrosine binding
pThr	phosphothreonine
pTyr	phosphotyrosine
RP	regulatory protein
RT	room temperature
SAX	strong anion exchange
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SH2	Src homology 2
SH3	Src homology 3
SIMAC	sequential-immobilised metal affinity chromatography
SLB	sample loading buffer
SNAC	N-acetylcysteamine
SOS	son of sevenless
SPE	solid phase extraction
SPPS	solid phase peptide synthesis
SPR	surface plasmon resonance
SUCLG1	succinyl-COA ligase
TAMRA	tetramethylrhodamine
TBAF	tetrabutylammonium fluoride
TBTA	tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine
TCEP	tris(2-carboxyethyl)phosphine
TCS	two-component systems
TEAB	tetraethylammonium bromide
TFA	trifluoroacetic acid
THF	tetrahydrofuran

TIS/TIPS	triisopropylsilane/triisopropylsilyl
TLC	thin layer chromatography
TMS	trimethylsilyl
TRPC4	transient receptor potential channel 4
TRPV5	transient receptor potential cation channel subfamily V member 5
TS	transition state
UPAX	unbiased phosphopeptide strong anion exchange
UV	ultraviolet
WAX	weak anion exchange

Abstract

Phosphorylation is a ubiquitous cellular regulation process in prokaryotic and eukaryotic species. Many aspects of the phosphoproteome are well explored and characterised, but there is a severe lack of understanding of the role of histidine phosphorylation in more complex eukaryotic systems. The dearth of tools with which to explore histidine phosphorylation has precluded deeper understanding of the cellular processes that phosphohistidine might regulate; the development of tools to investigate phosphohistidine interactors would enable exploration of the histidine phosphoproteome and potentially uncover interesting novel binding protein partners and pathways.

This work documents the synthesis of a phosphonate pyrazole analogue of phosphohistidine in 29% yield via a two-step process, starting from 4-iodopyrazole and the subsequent testing of heterocycle addition to an affimer containing dehydroalanine as a means of producing proteins containing phosphohistidine analogues. Addition of heterocycles: imidazole, pyrazole, ethyl-1H-pyrazole-4-carboxylate and 1H-pyrazole-4-carboxylic-acid to an affimer with a dehydroalanine residue was tested. Imidazole showed around ~30% addition, while pyrazole addition was significantly less abundant and no substitution was observed for ethyl-1H-pyrazole-4-carboxylate and 1H-pyrazole-4-carboxylic-acid. Hence demonstrating the challenges of incorporating a phosphonate pyrazole analogue of phosphohistidine into proteins through substitution on dehydroalanine.

The synthesis of electrophilic triazole moieties as reactive phosphohistidine analogues is explored. Synthesis of a novel benzyl-ethyl ethynyl phosphonate is outlined, starting from commercially available TIPS-acetylene. The synthesis is performed via a Grignard substitution to prepare a phosphoramidite alkyne that can be subjected to tetrazole-mediated substitution with benzyl alcohol and ethanol. In controlled conditions a phosphonate with benzyl-ethyl mixed substitution is produced in a 23% yield. The benzyl-ethyl protected phosphonate alkyne is click-coupled with an azidoalanine residue in a peptide to produce a fluorophosphonate-precursor peptide in a 52% yield.

Additionally, photo-reactive peptide mimics of epidermal growth factor receptor (EGFR) and pyruvate phosphate dikinase (PPDK) are developed and tested for binding to their partner proteins Grb2-SH2 and regulatory protein (RP) respectively, as a proof-of-concept for discovering novel phosphoprotein interactions. The peptides contain photo-activatable

diazirine groups, enabling covalent binding between both the probes and their respective partner proteins. The proteins show inconsistent binding to their isolated partner proteins when characterised by gel electrophoresis and subsequent proteomic analysis did not show statistically significant binding of the EGFR photoprobe to Grb2-SH2, highlighting challenges associated with using photo-activatable peptides to validate these binding interactions.

1 Introduction

Phosphorylation is a critical protein modification process in cells. Phosphate acts as a molecular switch regulating the function of proteins in cells by stimulating confirmational change and hence triggering binding interactions. Given the significance of phosphorylation on cellular processes, there is a vast amount of research documenting phosphoprotein interactions. However, investigating histidine phosphorylation has been significantly more challenging than other for many other amino acids.

Steady progress has been achieved in the study of the function and significance of phosphohistidine in eukaryotic cells. The following section addresses what is known to date with respect to phosphohistidine structure and function with regards to: the eukaryotic cellular processes involving phosphohistidine signalling; the kinase and phosphatases that regulate histidine phosphorylation; the chemical stability of phosphohistidine; research regarding detection and isolation of phosphohistidine peptides; and, how the desirability of immunoprecipitation enrichment methods led to a strong focus on phosphohistidine analogue development.

Subsequently, this work aimed to develop reactive modalities that were compatible with known phosphohistidine analogues. Development of reactive versions of these analogues would enable production of reactive probes containing phosphohistidine analogues to enable eventual identification of binding partner proteins in phosphohistidine-based signalling. Identification of proteins that interact with phosphohistidine would therefore reveal unknown aspects of the phosphoproteome so that the significance and functions of phosphohistidine signalling can be more deeply understood.

1.1 What is phosphorylation and why is phosphohistidine important?

Some estimates suggest that up to 30% of human proteins are phosphorylated and dephosphorylated by approximately 1000 different kinases and 500 different phosphatases (1). Protein phosphorylation is a vital component of signal transduction in all cells. Despite many years of research, there are still aspects of phosphorylation that are unexplored (2, 3). Research into phosphorylation has largely centred on the phosphate esters formed by serine, threonine and tyrosine displayed in **Figure 1** (pSer **1**, pThr **2** and pTyr **3** respectively) these are all hydrolytically stable and, consequently, are the prevalent forms of phosphorylation and accounts for more than a third of all annotated eukaryotic phosphorylated protein residues, 86% (pSer), 12% (pThr) and 2% (pTyr) (4, 5).

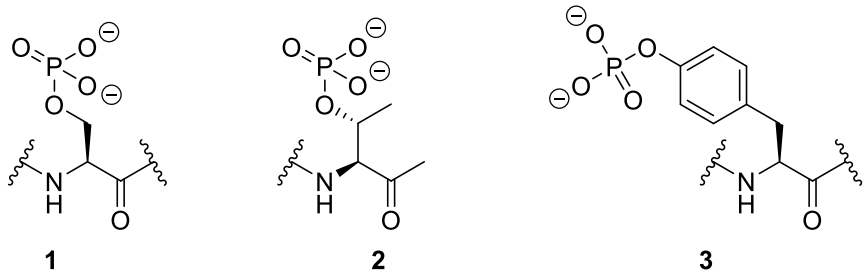


Figure 1: Structures of phosphoserine **1**, phosphothreonine **2** and phosphotyrosine **3**.

Phosphorylation is the migration of a phosphate group from a phosphate anhydride, typically the γ-phosphate group of adenosine triphosphate (ATP), onto a protein substrate (6). The free energies of hydrolysis for phosphonate esters are similar to that of the ATP β-γ linkage, [27-40 $kJmol^{-1}$ (pSer-pTyr) (7, 8)] and [21-40 $kJmol^{-1}$ (ATP) (9)]; and to prevent excessive equilibration and promote phosphorylation cells tend to maintain a high ATP: ADP ratio (6). Phosphorylation is mediated by the interplay of three other species a binding partner, kinase and phosphatase. In epigenetics, these are known as the reader, writer and eraser proteins respectively (10). **Figure 2** illustrates the process of phosphorylation for a prototypical pTyr system.

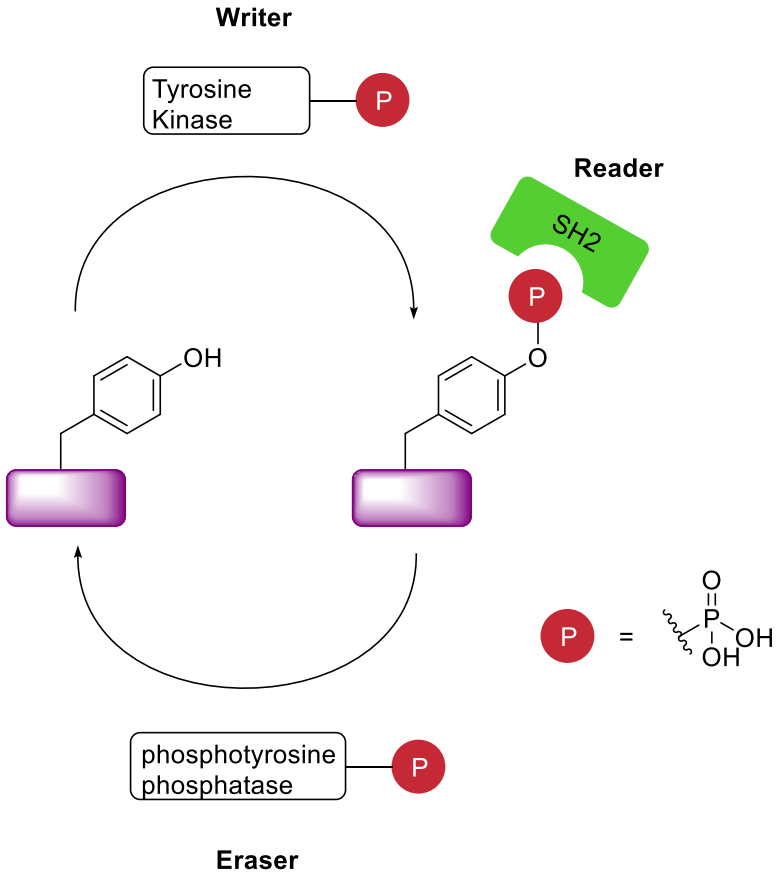


Figure 2: Phosphorylation cycle for a tyrosine-containing protein. Tyrosine-containing protein is phosphorylated by a pTyr kinase (**Writer**), the resulting pTyr-containing protein binds to src homology 2 (SH2) (**Reader**) that elicits a cellular response (further binding/a direct effect on the transcription machinery). pTyr phosphatases remove the phosphate (**Eraser**) from the pTyr-containing protein to regulate the cellular response. (11)

Phosphorylation perturbs the biophysical state of a protein, through the electronic shift that is caused by the addition of a -2 charge after modification. This charge can result in conformational changes that alter the intermolecular and intramolecular protein interactions of a substrate (6). Nature's selection of phosphate is interesting, the -2 charge shift gives rise to unique binding ability compared to other negatively charged amino acids, aspartate and glutamate, which only have a single negative charge. Phosphoamino acids therefore have greater shell hydration and hydrogen-bond-accepting propensity, due to their greater charge density when compared with other amino acids at physiological pH (6, 12). This unique characteristic of phosphoproteins enables phospho-dependent protein-protein interactions to occur, resulting in the formation of multiprotein complexes that enable substrate binding, such as pTyr generating complexes through interaction with SH2 and PTB domains (13). pSer and pThr largely regulate processes through allosteric adjustment, though some evidence indicates

that they may mediate protein-protein interactions to generate binding sites recognised by WW and forkhead-associated (FHA) domains (14).

Despite the prevailing focus of research being on the stable phosphate ester amino acids, it is also possible to phosphorylate arginine, lysine, histidine, cysteine, aspartate and glutamate, and some of these modifications have been shown to occur in biological systems (6). Histidine phosphorylation, for example, was first observed by Boyer *et al.* (1962) where it was reputedly involved in an oxidative phosphorylation process with ATP (15). Since the discovery of phosphohistidine, researchers have been eager to identify what processes it regulates, how significant it is with respect to the wider study of phosphoproteomics - some early estimates attribute 6% of protein phosphorylation in eukaryotes to histidine kinases - and if it is possible to inhibit histidine phosphorylation to prevent disease states (as has been done with phosphotyrosine) (16, 17). Bacterial signalling relies on two-component systems (TCS), in which a histidine kinase is autophosphorylated and subsequently transfers the phosphate group of phosphohistidine to an aspartate on a response regulator (18). Bacterial signalling is significantly less complex than mammalian phosphorylation and is consequently extensively studied (19), as such this discussion is limited to what is known to date about eukaryotic histidine phosphorylation (20, 21).

1.2 The chemistry of phosphohistidine: relative stability of phospho-amino acids?

The instability of phosphohistidine provides a fundamental challenge to its study using proteomics approaches. This instability is the inherent issue plaguing phosphohistidine proteomics, especially when compared to the amino acids serine, threonine and tyrosine. The structures of the two isoforms of phosphohistidine τ-phosphohistidine **4** and π-phosphohistidine **5** are shown in **Figure 3**.

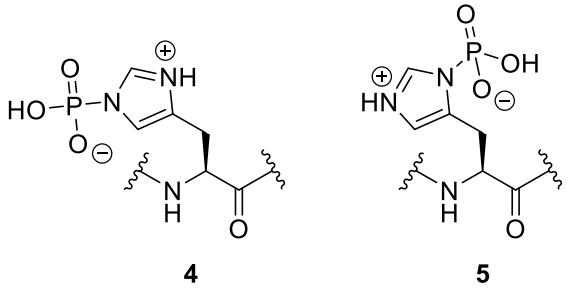


Figure 3: Delocalisation of the phosphoramidate nitrogen lone pair into the imidazole ring of τ-phosphohistidine **4** and π-phosphohistidine **5**, the lack of electrons in the phosphoramidate bond results in its higher energy and aforementioned tendency to hydrolyse (Adapted from reference (8)).

The instability of phosphohistidine can be attributed to the nature of the phosphoramidate (P-N) bond. The nitrogen lone pair, in phosphohistidine, is invested in the π-system of the imidazole ring and therefore does not effectively interact with the orbitals of the P-O phosphoramidate system, which is substantially higher in energy – involving a higher energy 3d-shell (22), giving rise to a high energy of hydrolysis, $\Delta G^0 = -RT\ln K_d$, for phosphohistidine, in proteins, of -50 to -59 kJ mol^{-1} (accounting for the two isoforms), whereas for phosphoserine and phosphotyrosine the energy of hydrolysis is -27 and -40 kJ mol^{-1} respectively (7, 8).

The nitrogen in the imidazole ring that is not phosphorylated is basic, for τ- and π-phosphohistidine the pK_a values are 6.4 (23) and 7.3 (24) respectively, thus they are partially protonated and exist as a zwitterionic form at physiological conditions, pH 7. As a result, the phosphoramidate bond in phosphohistidine is readily hydrolysed *in vivo*. This tendency for hydrolysis, acting as a source of phosphate, allows phosphohistidine to perform its biological function in cell-signalling as a phospho-relay more expediently (8). Despite clear differences in their thermodynamic stability, the mechanistic ramifications of τ- or π-phosphohistidine formation have not been elucidated; whilst some genes such as NDPK-A and B appear to favour π-phosphohistidine and, as stated earlier, LHPP favours τ-phosphohistidine no discernible conclusions can be made about the functional consequences of these regioisomers.

Of the two isoforms, τ-phosphohistidine **4** is more thermodynamically stable relative to π-phosphohistidine **5**, consequently this isoform is observed in analytical samples more often; but, may not be representative of the relative abundance *in vivo* (25).

In a protein, the free α-amino group, which is regarded as a strong contributor to free phosphohistidine hydrolysis in both an acidic and less-acidic environment, would be bound in a peptide bond having neither the basicity nor the nucleophilicity to facilitate hydrolysis. In protein structures the hydrolysis of phosphohistidine is dependent on the electrostatic interactions of neighbouring residues (8, 16). Density-functional theory (DFT) and transition state (TS) studies regarding prostatic acid phosphatase (PAP) show that the positive charges on neighbouring arginine residues help to stabilise a TS leading to the hydrolysis of phosphohistidine and a neighbouring Asp258 residue in the active site of the enzyme transfers a proton to the anionic phosphate moiety being ejected at the end of the phosphohistidine hydrolysis TS (26, 27).

Phosphohistidine in proteins is likely to participate in hydrogen bond interactions to the protein backbone (28), resulting in greater hydrolytic stability than free phosphohistidine, as shown in **Figure 4**.

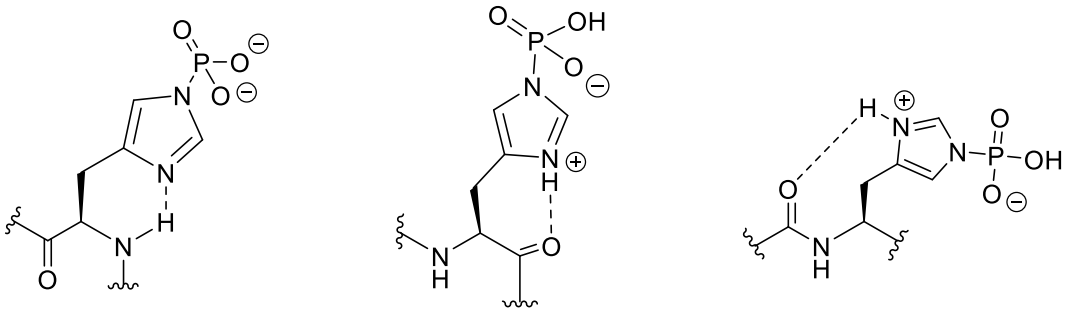


Figure 4: Possible structures for protein backbone stabilisation of a 6-membered ring in phosphohistidine **6**, unlikely transition states for phosphohistidine acid-catalysed hydrolysis stabilised by less stable 7- and 8-membered ring structure **7** and **8** respectively.

1.3 Phosphohistidine in biological systems

The phosphohistidine post-translational modification has been implicated in a range of mammalian biological processes including chromatin biology (29), endocytosis (2, 30), KCa3.1 and TRPV5 channel regulation (31, 32) and G-protein signalling (33-35).

The genes Hunter *et al.* associated with phosphohistidine isoforms and their biological function are summarised below (**Table 1**) (2).

Gene	Uniprot ID	Biological function	Isoform
NDPK-A/NME1	P15531	NDPK Histidine Kinase	π-pHis
NDPK-B/NME2	P22392	NDPK Histidine Kinase	π-pHis
NDPK-D/NME4	O00746	NDPK Histidine Kinase (mitochondrial localised)	π-pHis
PGAM1	P18669	Glycolysis	τ-pHis
PGAM 5	Q96HS1	pHis phosphatase	τ-pHis
SUCLG1	P53597	Succinyl-COA ligase	τ-pHis
ACLY	P53396	ATP-citrate synthase	τ-pHis
PHPT1	Q9NRX4	pHis phosphatase	-
LHPP	Q9H008	pHis phosphatase	-
GNB1	P62873	GPCR signal transduction	τ-pHis
KCa3.1	015554	Calcium-activated potassium channel	τ-pHis
TRPV5	Q9NQA5	Calcium channel	τ-pHis
HIST1H4A	P62805	Chromatin regulation	τ-pHis/ π-pHis
PLD	Q13393	Phospholipid metabolism	?
NAMPT	P43490	NAD+ biosynthesis	π-pHis

Table 1: Genes associated with phosphohistidine, their Uniprot ID code, biological function and the observed isoform. Adapted from Fuhs *et al.* 2017 (2)

1.3.1 Histone H4 (Chromatin)

Histone H4 is one of the core histone proteins. Histones are the proteins around which DNA is wound to form nucleosomes (36). Octameric nucleosome cores are comprised of histones: H2A, H2B, H3 and H4 (37). Histone proteins therefore play a significant role in processes involving DNA regulation, where post-translational medication of histones will have implications on transcription and replication (38).

In 1973 Smith *et al.* discovered a protein phosphate in histone proteins whose presence was dependent on pH, indicating the presence of an acid-labile phosphoprotein (29). Two novel histone kinases had been found that were capable of forming acid-labile phosphorylated-histone at differing pH optima and acting on different histone substrates, giving credence to the idea that acid-labile phosphates may be more widespread (29). The later discovery of a histone H4 kinase from porcine thymus that specifically phosphorylates histidine supported the idea that histidine phosphorylation is present in significant biological pathways (reputedly up to four kinases were present in the porcine thymus sample; but some histone H4 kinases have been shown to phosphorylate serine and threonine and the method used did not discriminate between acid and alkali-stable phosphates, so only one of the identified kinases may have been a histidine kinase) (39, 40).

The acid-labile phosphates on histone H4 were found at higher abundance in regenerative rat tissue than in normal tissue. Having already been discovered in Walker-256 carcinosarcomas, histone H4 kinases and pHis appear to have some association with cell proliferation (41). Later the histone H4 phosphate (from Walker-256 carcinosarcomas) was identified as τ-phosphohistidine **4** (**Figure 5**) through Phosphorus-31 NMR (^{31}P NMR). Interestingly the isoform of phosphohistidine formed on the histone H4 protein was tissue-source dependent, kinases in rat liver gave rise to the π-phosphohistidine **5** structure (**Figure 5**); despite this isoform specificity, any structure-function ramifications associated with the isoforms are unclear (42).

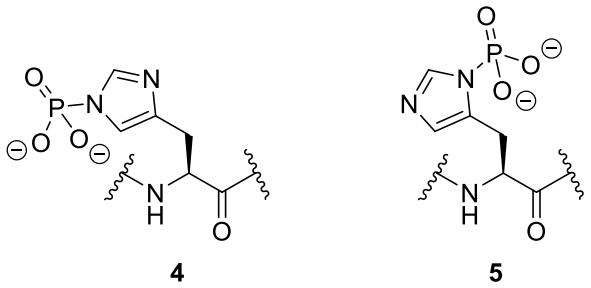


Figure 5: Structures of isoforms: τ-phosphohistidine **4** and π-phosphohistidine **5**

1.3.2 Acetyl coenzyme A synthesis

Catabolism of protein, fats and carbohydrates lead to the formation of acetyl coenzyme A (acetyl-CoA) (43). Glycolysis of cytoplasmic glucose produces pyruvate and subsequently the carboxyl group of pyruvate is transferred to CoA through thiamine pyrophosphate and acyl lipoyllysine intermediates producing acetyl-CoA in the mitochondria (**Figure 6**) (44).

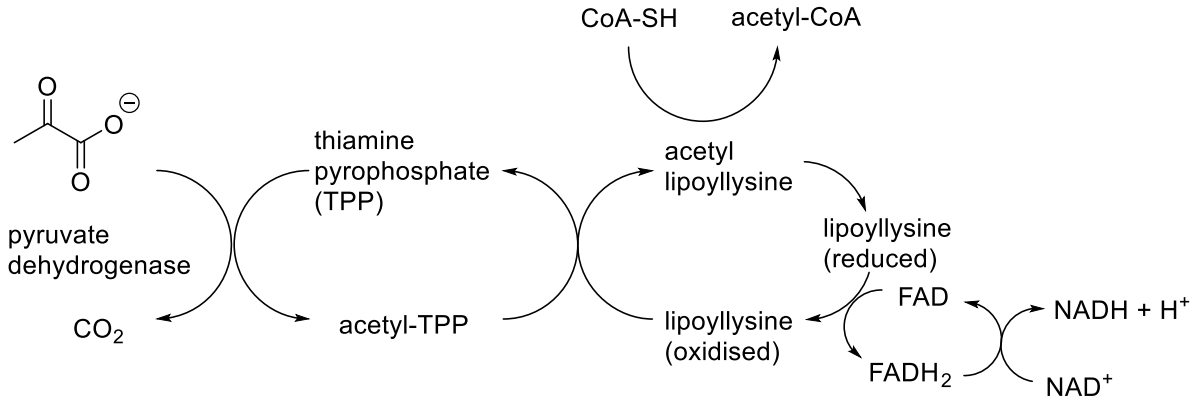


Figure 6: Mitochondrial biosynthetic pathway for acetyl-CoA production from pyruvate. (45)

In the cytosol, acetyl-CoA is a critical metabolic intermediate for energy production (Krebs cycle) through donation of an acetyl moiety to oxaloacetate to form citrate (44). Cytosolic acetyl-CoA availability is also a key regulator of lipid synthesis with acetyl-CoA conversion to malonyl-CoA by acetyl-CoA carboxylase being the initial step (46).

ATP-citrate lyase (ACLY) is an enzyme responsible for cytosolic production of acetyl-CoA from citrate (**Figure 7**) (47). Increased ACLY action leads to increased levels of acetyl-CoA in the cytosol and therefore increases fatty acid synthesis as well as increasing the availability of acetyl-CoA for acetylation of histones (48, 49).

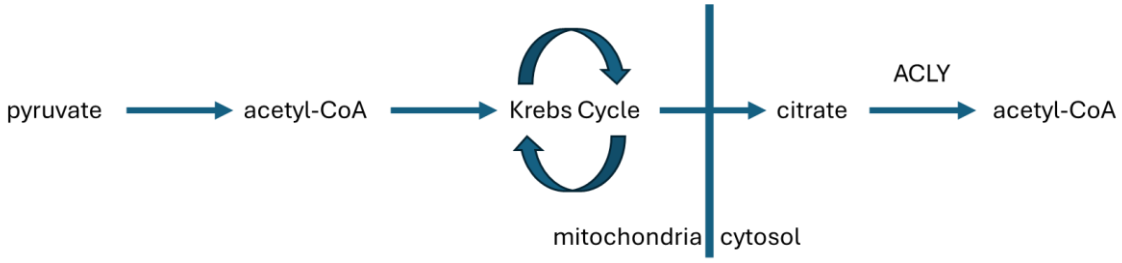


Figure 7: Mitochondrial citrate as a product of the Krebs Cycle migrates to the cytoplasm and is converted to acetyl-CoA by ACLY. (50)

ACLY plays a significant role in the regulation of acetyl-CoA levels and thus has a knock-on significance on Krebs cycle function, fatty acid synthesis and histone acetylation. ACLY is autophosphorylated by ATP on a catalytic histidine residue (His760) and Fan *et al.* show that H760A results in deactivation of ACLY function (51). Nucleoside Diphosphate Kinase A (NDPK-A), a mammalian histidine kinase, may also be involved in phosphorylation of ACLY (52). The phosphorylation of His760 on ATP-citrate lyase initiates the formation of acetyl-CoA from citrate and CoA (52). ACLY upregulation is observed in many malignant tumour cells, with inhibition of ACLY function resulting in reduction in cell proliferation. Evidently, histidine phosphorylation plays a key role in activation of ACLY function and its subsequent cellular response.

1.3.3 KCa3.1 Ion Channel Regulation

Ca^{2+} influx through calcium-release-activated channels (CRAC) is crucial for cellular immune response, Ca^{2+} ion levels are a crucial trigger for the production of antigen-responsive B-lymphocytes (53) as well as T-lymphocyte production by maintaining a favourable membrane potential (54). T-cell receptor activation depletes Ca^{2+} in the endoplasmic reticulum (ER) and consequently results in further calcium influx required for T-cell activation (55, 56). The influx of Ca^{2+} will eventually reduce the potential of the cell thus K^+ channels are required for the efflux of K^+ to restore the potential of Ca^{2+} to influx and activate lymphocytes (57).

Intracellular calcium-activated potassium channel (KCa3.1) aids the cellular modulation of Ca^{2+} influx (31). KCa3.1 is associated with calmodulin (CaM), a calcium binding protein, in the presence of excess free Ca^{2+}, CaM binds Ca^{2+} leading to KCa3.1 channel activation (58, 59). KCa3.1 expends the excess free Ca^{2+} promoting K^+ efflux and consequently aids in cellular homeostasis by driving towards a negative potential (54). KCa3.1 plays a vital role in cell proliferation as part of the cellular ecosystem controlling calcium-dependent protein interactions (60).

Nucleotide diphosphate kinase B (NDPK-B) is documented as a KCa3.1 regulator; phosphorylation of a KCa3.1 histidine residue (His358) (**Figure 8**) appears to correlate with channel activation by reducing copper inhibition of the channel, the exact mechanism by which this activation operates is unknown; but, evidence for high channel activity and histidine phosphorylation in copper-deficient cluster of differentiation 4 ($CD4^+$) T cells, may indicate pHis involvement in this process despite the absence of an exact mechanism (61). PHPT1, a phosphohistidine phosphatase, is capable of dephosphorylating the histidine residue (His358) in KCa3.1 and this dephosphorylation has been shown to coincide with KCa3.1 channel inhibition, emphasising the importance of histidine phosphorylation in this process (**Figure 8**) (62). Given the downstream importance of KCa3.1 in regulating ion concentration for lymphocyte activation, dysregulation of pHis pathways involved in KCa3.1 activation/inhibition would have a knock-on effect on immune activation.

A second phosphatase inhibiting potassium channel activation has also been identified, however in this case the phosphatase, PGAM5, dephosphorylates the His118 intermediate in the kinase NDPK-B rather than removing phosphohistidine from KCa3.1 (**Figure 8**) (63). Inhibiting KCa3.1 function, and NDPK-B function, provides a means of treating auto-immune

diseases. Additionally, phosphorylation of G-protein subunits has been observed to increase in the presence of overexpressed NDPK-B, resulting in G-protein activation (34), further emphasising the significance of histidine phosphorylation *in vivo* (31, 64).

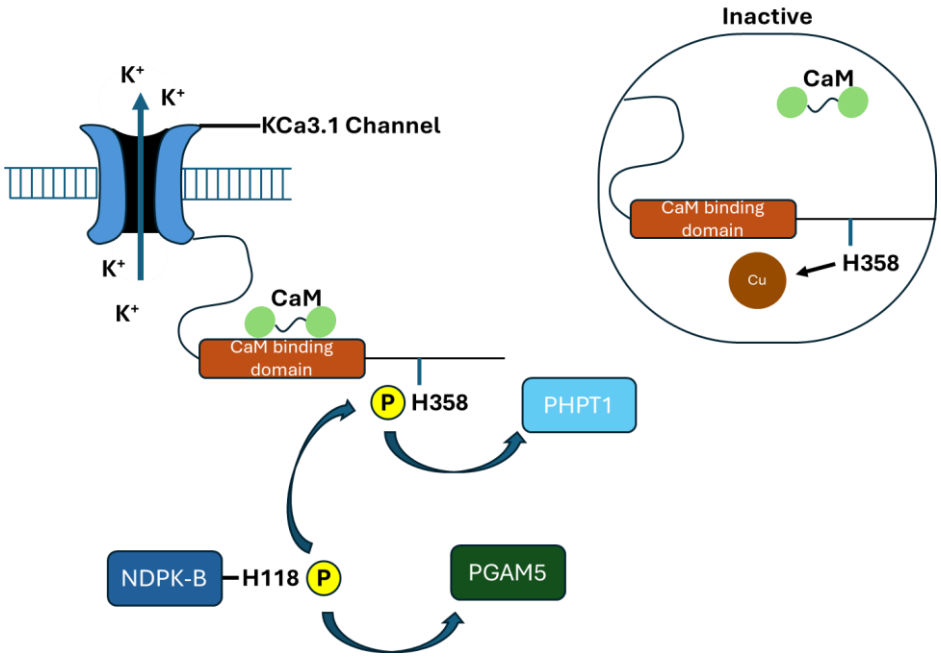


Figure 8: Phosphorylation of KCa3.1 protein prevents copper inhibition. NDPK-B phosphorylates KCa3.1 (H358) resulting in calmodulin (CaM) binding and subsequent high channel activity. The phosphatases, PHPT1 and PGAM5, dephosphorylate their respective substrates (KCa3.1 and NDPK-B) and consequently reduce channel activity (31, 61-63).

1.3.4 TRPV5 ion regulation

Transient receptor potential vanilloid 5 (TRPV5) channels are vital for calcium reabsorption, controlling levels of Ca^{2+} in the blood prior to epithelial trafficking of Ca^{2+} (65). TRPV5 is mainly located in the renal tubular epithelial cells where 99% of calcium absorption is occurring (~ 15% calcium absorption from TRPV5) (66-68). As with other calcium channels, TRPV5 is responsible for maintaining calcium homeostasis which is critical for cellular health (65). The presence of hypercalciuria and bone mineral loss in TRPV5 knockout mice, demonstrates the significant function of calcium homeostasis (as regulated by TRPV5) for mammalian health (69). Moreover, there is evidence that issues with calcium homeostasis can cause cardiovascular diseases (70), and in the presence of a mutant TRPV5, an increased risk of kidney stone formation (71).

Histidine phosphorylation plays a crucial role in the regulation of calcium ion excretion by TRPV5; akin to potassium channel regulation, phosphorylation of His711 in TRPV5 by NDPK-B leads to an increase in calcium channel activation that, as observed in KCa3.1, can be inhibited with phosphohistidine phosphatase (PHPT1), further establishing histidine phosphorylation as an important mediator in mammalian biology (**Figure 9**). Cai *et al.* measured ion channel activity from HEK293 cells transfected with wild-type TRPV5, TRPV5 (H711N), TRPV5 + NDPK-B and TRPV5 (NDPK-B knockdown). In cells transfected with TRPV5 and NDPK-B, calcium currents dropped significantly, indicating a functional significance for NDPK-B on the ion channel activity of TRPV5. The wild-type, histidine mutant (H711N) and NDPK-B knockdown cell transfections showed no significant difference in their ion channel activity based on calcium currents. This would indicate that histidine itself has no significant effect on basal ion channel activity but a modification by NDPK-B when present modifies ion-channel activity. The same study additionally demonstrates that the addition of PHPT1 can reverse the channel activity promoted by NDPK-B. This interplay between NDPK-B and PHPT1 signifies that the phosphorylation and dephosphorylation of a histidine residue is responsible for channel activity modulation (32). Cai *et al.* by extension, argue that His711 phosphorylation prevents binding of calmodulin to the C-terminal binding domain on TRPV5 and thus facilitates channel high channel activity by preventing calmodulin feedback required for channel closure (32). This is logically coherent with previous studies on TRPV5 activity by de Groot *et al.* who showed that phosphorylation of Thr709 in the

calmodulin binding domain of TRPV5 led to higher channel activity by attenuating the channel closing effect of calcium-driven calmodulin binding (72).

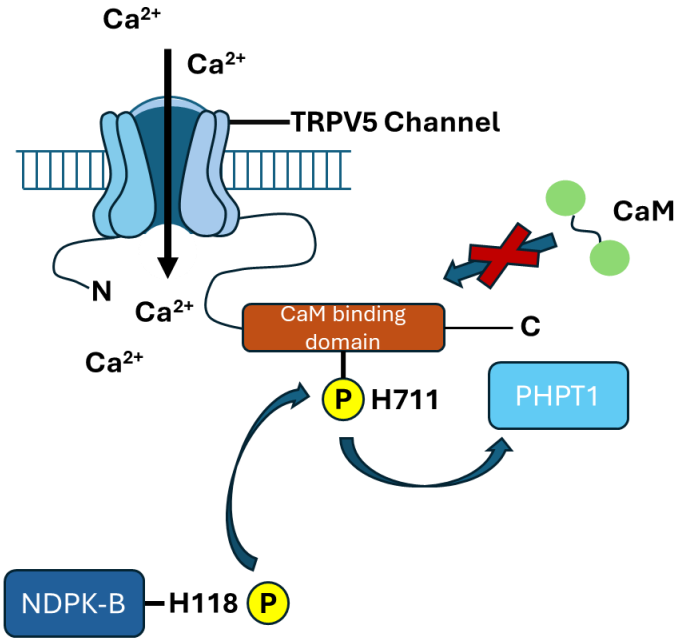


Figure 9: TRPV5 channel activated by phosphorylation of His711 in the C-terminal calmodulin binding domain. pHis711 prevents calmodulin (CaM) binding which in turn inhibits channel closure by CaM binding to calcium. His711 is phosphorylated and dephosphorylated by NDPK-B and PHPT1 respectively. (32, 73)

1.3.5 TRPC4 Ion Channels

TRPC4 (Transient receptor potential channel 4) is a calcium permeable cationic ion channel promoting Ca^{2+} influx into the cell upon activation (74). TRPC4 is implicated in processes including endothelial permeability (75), cell proliferation (ovarian and renal cancer) (76, 77), carcinoma inhibition (78) and vascular homeostasis (79).

Studies suggests that a histidine residue (H912) in the C-terminal loop of TRPC4 is dephosphorylated by PHPT1 (phosphohistidine phosphatase 1); in cells lacking PHPT1 pHis cannot be dephosphorylated and TRPC4 activity is reduced (**Figure 10**) (80). Mutation of H912 in TRPC4 to asparagine restored TRPC4 activity with leptin stimulation. Additionally, where wild-type PHPT1 was tested against a mutant PHPT1 (where H53 was mutated to A53), TRPC4 only showed increased activity in the presence of the wild-type PHPT1 (80), demonstrating the direct significance of histidine phospho-regulation in channel function.

Therefore, pHis acts as an inhibitor to TRPC4 function with PHPT1 being directly responsible for enabling activity through dephosphorylation. With TRPC4 function being critical for the aforementioned biological processes, the regulation of histidine phosphorylation plays a critical role in regulating many cellular responses.

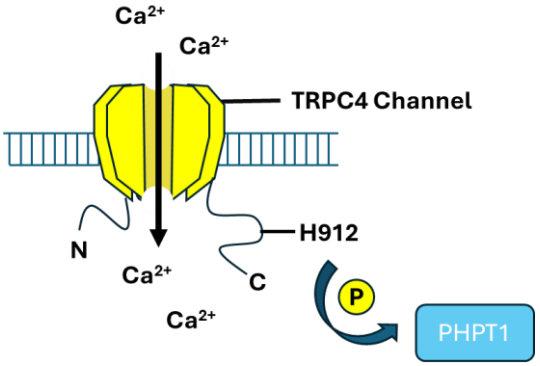


Figure 10: Leptin-activated TRPC4 channel activity associated with dephosphorylation of His912 by PHPT1 in the C-terminal domain. (80)

1.3.6 G-protein signalling

G-protein mediated signalling is a ubiquitous mammalian cellular process for transduction of extracellular signals across cell membranes (81, 82). GPCRs (G-protein coupled receptors) are large structures that bind smaller G-proteins, which are typically heterotrimeric - comprised of smaller protein sub-units (Gα, Gβ and Gγ) (83). G-protein function is regulated by the GTP cycle, a shifting of phosphate resulting in a cycle between GTP/GDP (guanosine triphosphate/guanosine diphosphate) (83, 84). Once GDP is phosphorylated to form GTP it binds to the α-subunit of the membrane-bound G-protein substructure enabling signalling to downstream messenger proteins that migrate throughout the cell initiating a cascade of cellular processes dependent on the G-protein subtype (84).

In the heart, G-protein signalling is responsible for activating processes controlling heart rate, contraction forces, relaxation and more (81). As such, dysregulated G-protein signalling is associated with various heart diseases (81). G-protein coupled receptor kinase 2 (GRK2), a Ser/Thr kinase, is a key regulator of cardiac function by initiating a cascade promoting protein kinase B (AKT) signalling and inhibiting GSK3β signalling which leads to decreased regulation of NFAT (a key transcription factor promoting hypertrophy, phosphorylation of NFAT leads to its deactivation) (85). GRK2 has been shown to induce cardiac hypertrophy when overexpressed as well as hypertrophy attenuation in its absence in GRK2-knockout mice (85). Elevated levels of GRK2 is also associated with hypertension (high-blood pressure), blood pressure is regulated by the size/diameter of blood vessels changes to vessels size will have a profound effect on arterial pressures. Upregulation of GRK2 as well as GRK5 (G-protein coupled receptor kinase 5) is associated with hypertension (86, 87).

NDPK-B has been discussed for its role in regulating ion channel activation; the histidine kinase is also involved in G-protein signalling pathways (**Figure 11**).

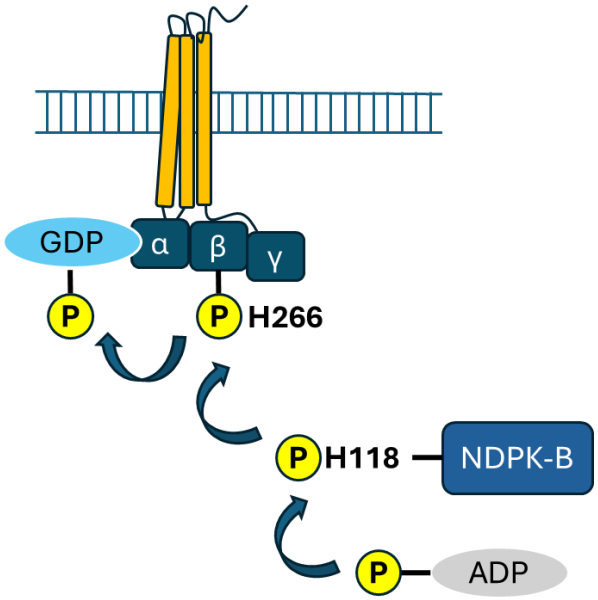


Figure 11: GPCR activation via phospho-relay involving shuttling of phosphate from γ-phosphate of ATP through NDPK-B H118. NDPK-B transfers phosphate to His266 of β-subunit of G-protein; the high energy phosphate on the β-subunit transfers to GDP and rebound as GTP activating the G-protein. (88, 89)

NDPK-B modulates the cycle between GDP/GTP by regulating the availability of phosphate from high-energy ATP. The γ-phosphate of ATP is transferred by autophosphorylation onto NDPK-B (His118) and subsequently shifted to His266 on the Gβ subunit of the G-protein (88). Given that Gα binds GDP prior to activation it is theorised that the high energy phosphate on His266 is transferred to GDP in the Gα subunit leading to GTP formation and G-protein activation in lieu of agonist-mediated activation (89). One study suggests that cardiac adenylyl cyclase, a critical heart function regulating enzyme, is activated by the interaction of GDP and NDPK-B in the absence of GTP formation and α-subunit localisation (90), however, the exact mechanism by which NDPK-B activates G-protein pathways and bypasses agonist-activation is largely still debated. Critically there is a clear influence of the histidine kinase, NDPK-B, on activation of cardiac adenylyl cyclase which promotes cAMP (cyclic adenosine monophosphate) production, cAMP is a critical component of heart electrophysiology (91).

1.4 Identified Kinases and Phosphatases

Identifying and studying kinases will be essential to any biological study of histidine phosphorylation and its consequences; studying phosphatases is also crucial. The aforementioned examples, PHPT1 and PGAM5, are phosphatases for which corresponding substrates and kinases have been identified and their roles elucidated.

Another phosphatase, LHPP, has also been identified as a phosphohistidine phosphatase, but no specific substrates for the phosphatase have been identified (2). LHPP has been documented as a phosphatase broadly acting on the τ-phosphohistidine isoform in peptide substrates (92). LHPP has a reputed role in tumour suppression supported by its downregulation in proliferating cells, coupled with its ability for dephosphorylating substrates, the upregulated presence of NDPK-A and B in tumour cells and previous evidence for phosphohistidine prevalence in proliferating cells (histone H4) (92, 93).

So far, a small subset of enzymes has been identified as being associated with phosphohistidine incorporation and removal in human cells. NDPK-A, NDPK-B and histone H4 kinases are understood to be responsible for histidine phosphorylation and PHPT1, PGAM5 and LHPP for its dephosphorylation. However, there are likely to be many more as-yet-unidentified kinases and phosphatases but the lack of tools with which to study phosphohistidine, as a result of its instability relative to phosphate esters, has precluded research progress (3, 93, 94).

The aforementioned kinases and phosphatases, for which structures exist, are summarised in **Table 2**:

Enzyme Role	**Gene**	**Uniprot ID**	**Isoform**	**Residues Flanking Active Histidine**
Kinase	NDPK-A	P15531	π-pHis	FCIQVGRNII**H**GSDSVESAEK
	NDPK-B	P22392	π-pHis	FCIQVGRNII**H**GSDSVKSAEK
	NDPK-D	O00746	π-pHis	FSVHISRNVI**H**ASDSVEGAQR
Phosphatase	PGAM5	Q96HS1	τ-pHis	KATRHIFLIR**H**SQYHVDGSLE
	PHPT1	Q9NRX4	-	IVRGYKWAEY**H**ADIYDKVSGD
	LHPP	Q9H008	-	-

Table 2: Summary of known histidine kinases and phosphatases.

1.4.1 Known Histidine Kinases (NDPK/NME)

A histidine kinase is an enzyme that transfers phosphate onto a histidine in a target substrate. NDPK-A and NDPK-B are the two most prevalent histidine kinases; they appear to be responsible for phosphorylating target histidine residues in the majority of investigated biological systems.

The structures of NDPK-A and NDPK-B are very similar (**Figure 12**), both kinases have been identified bearing the π-pHis isoform and whilst the sequences share significant universal homology, the residues flanking the phosphorylated histidine (H118) show greater similarity than the rest of the sequence with only a substitution between E124 to K124 between NDPK-A and NDPK-B respectively in the 20 amino acid residues enveloping the active histidine residue (**Figure 13**).

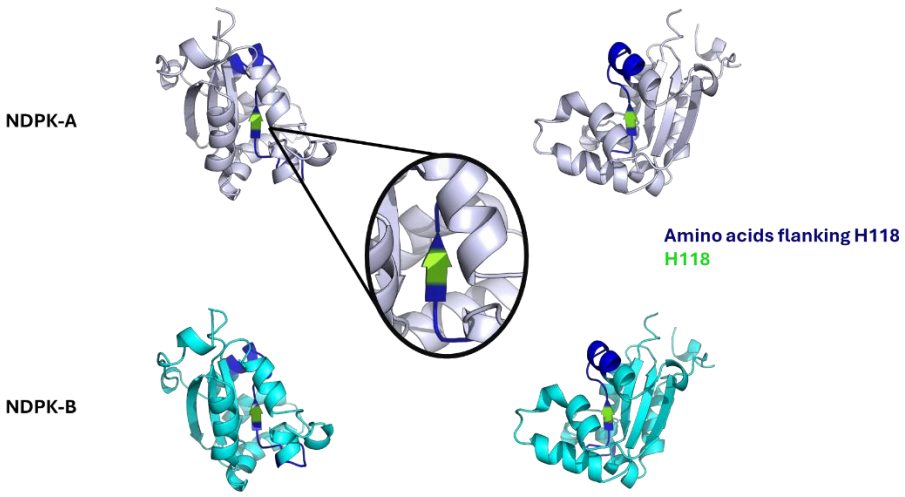


Figure 12: NDPK-A and NDPK-B monomer structures, dark blue highlights indicate position of FCIQVGRNIIHGSDSVESAEK and FCIQVGRNIIHGSDSVKSAEK for NDPK-A and NDPK-B respectively. With light green indicates the position of the catalytic histidine (H118). In both structures clear structural homology exists with the histidine being present in the lower portion of a centralised β-sheet enveloped by external α-helices. [Images created in PyMOL with AlphaFold PDB files: AF-P15531-F1-model_v4 (NDPK-A) and AF-P22392-F1-model_v4 (NDPK-B)]

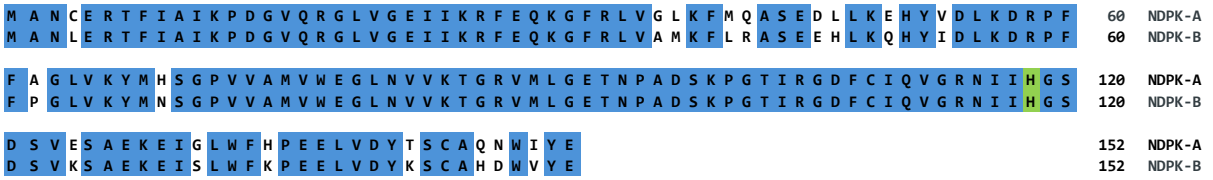


Figure 13: Sequence alignment for NDPK-A and NDPK-B, showing homologous residues (blue) and catalytic H118 (green).

Despite the structural similarity between NDPK-A and NDPK-B the two kinases are relatively functionally discrete. NDPK-A and NDPK-B are both involved in direct conversion of GDP to GTP driving dynamin-mediated endocytosis (**Figure 14**). Endocytosis is the internalisation of extracellular content by plasma membrane deformation and subsequent severance of the deformed plasma producing an intracellular vesicle. Dynamin is a GTPase that promotes vesicle fission from the membrane (95). Boissan *et al.* show that NDPK-A and NDPK-B enhance GTPase activity of dynamin-1 and dynamin-2. In the absence of the catalytic histidine (H118N) of NDPK-A dynamin GTPase activity was not enhanced. Additionally, NDPK-A and NDPK-B when in the presence of GDP showed similar dynamin GTPase activation to GTP alone, suggesting functional significance in NDPK conversion of GDP to GTP for dynamin activation. They postulate that NDPK-A and NDPK-B play a role in maintaining a high localised GTP:GDP ratio to promote dynamin activity (96).

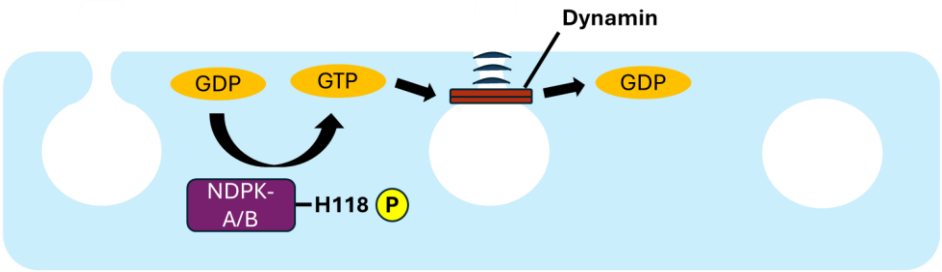


Figure 14: Dynamin-mediated vesicle fission. NDPK-A and NDPK-B convert GDP to GTP required for dynamin function. Polymerised dynamin GTPase domains perform GTP hydrolysis reforming GDP and finalising vesicle fission (95).

As histidine kinases, NDPK-A and NDPK-B show significant structural homology and an analogous His118 catalytic domain, but each isoform is largely involved in specific processes without cross-functionality. NDPK-A may play a supplementary role to ATP in acetyl-CoA formation by phosphorylating ACLY (H760) which leads to lysis of citrate into acetyl-CoA and oxaloacetate (52). Conversely, ion channel regulation appears to be exclusively regulated by NDPK-B, this is observed for KCa3.1 (31) and TRPV5 (32). NDPK-B has also been shown

to phosphorylate His266 in the Gβ subunit of a G-protein enabling GTP formation and G-protein activation (88, 89). These processes and the respective kinases are outlined in **Table 3**.

Kinase	**Processes**
NDPK-A/NME1/nm23-H1	Dynamin endocytosis, NTP homeostasis, ACLY
NDPK-B/NME2/nm23-H2	Dynamin endocytosis, NTP homeostasis, KCa3.1, TRPV5, GPCR (Gβ-subunit)

Table 3: Summary of NDPK-A, B processes involving catalytic phosphohistidine (97).

NDPK-D is also part of the nucleoside diphosphate kinase enzyme group. Despite belonging to the same family of enzymes as NDPK-A and NDPK-B, to date NDPK-D is implicated in fewer phosphohistidine regulated processes. NDPK-D appears to be more localised to mitochondrial processes and is therefore less functionally ubiquitous than the other NDPK enzymes. NDPK-D is a metastasis inhibitor indicated by a clear correlation between NDPK-D downregulation and increased tumour activity (98). In analogy with NDPK-A and NDPK-B in dynamin-mediated endocytosis, NDPK-D is responsible for generating GTP for promoting OPA1 (dynamin-related GTPase) function at the inner membrane (96, 98).

NDPK-D shares significant sequence homology with other members of the NDPK/NME enzyme family, while the overlap in sequence similarity between NDPK-A and B is around 88% (99), which explains their ability to perform analogous roles in dynamin endocytosis and NTP homeostasis, NDPK-D still contains some remarkable structural similarity (**Figure 15**).

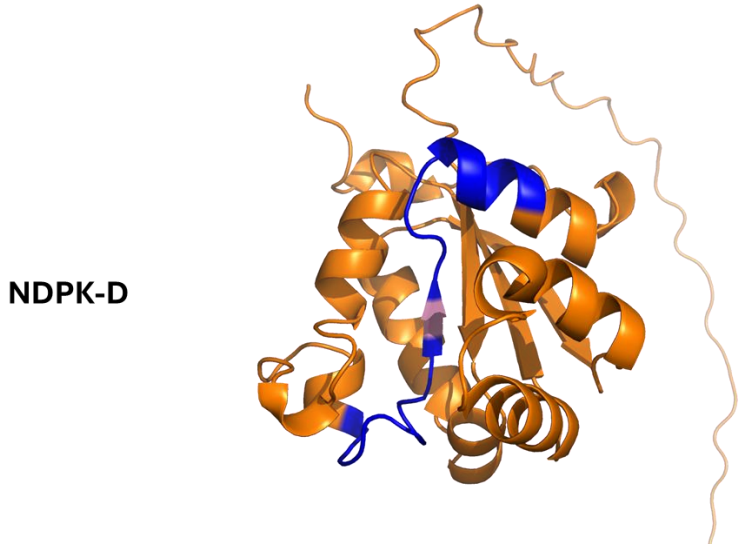


Figure 15: Structure of NDPK-D monomer, catalytic histidine H151 (100) (highlighted in pink), is located in the lower part of a centralised β-sheet, the central structure is enveloped in a conformation of α-helices. The location of the catalytic histidine residue is structurally analogous with NDPK-A and B despite lesser overall sequence homology, displaying some level of sequence homology and conservation at the catalytic histidine residue. [Image created in PyMOL with AlphaFold PDB file: AF-O00746-F1-model_v4]

1.4.2 Known Phosphohistidine Phosphatases

PHPT1 (phosphohistidine phosphatase 1) displays the most prevalent pHis dephosphorylation activity across cellular processes, dephosphorylating the histidine kinase: NDPK-B (H118); ion channel proteins: KCa3.1 (H358) (62), TRPV5 (H711) (32), TRPC4 (H912) (80); lipid enzyme: ACLY (H760) (101) and G-protein β-subunit (H266) (102).

The structure of PHPT1 is displayed in **Figure 16**:

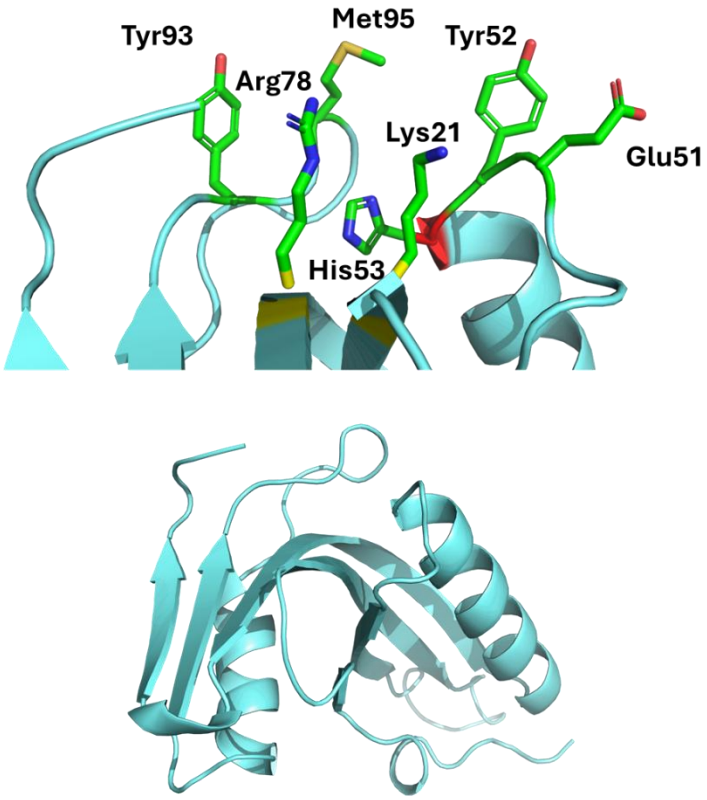


Figure 16: Structure of PHPT1, catalytic histidine residue H53 is surrounded by residues identified as being conserved in related protein structures and are thus hypothesised to be significant in stabilising the phosphate transfer activity of PHPT1, these residues are Lys21, Glu51, Tyr52, Arg78, Tyr93, Met95 (103). Below the active site residues is an overview of the overall protein structure, six β-strands flanked by two α-helices. [Images produced in PyMOL with AlphaFold PDB file: AF-Q9NRX4-F1-model_v4]

Studies into PHPT1 activity indicate that mutant R78A leads to a 30% decrease in PHPT1 activity and H53A completely eliminates PHPT1 activity (104). The residues labelled in **Figure 16** have been assessed by molecular modelling as forming a binding pocket for substrates where His53, Arg78 and Tyr93 form key hydrogen bonds to ligands representing the β-subunit of G-protein transducin (103). PHPT1 has also been shown to dephosphorylate ACLY (H760) and the domain housing histidine is conserved across human ACL, as such

docking of a succinyl-CoA peptide to PHPT1 has been proposed for binding to the His53 active site (103). An additional histidine site (His102) appears to display some significance in PHPT1 function, PHPT1 mutant H102A (104); however, no further structural discussion appears to explain the functional significance of His102. **Figure 17**, shows the spatial relationship between the two histidine residues.

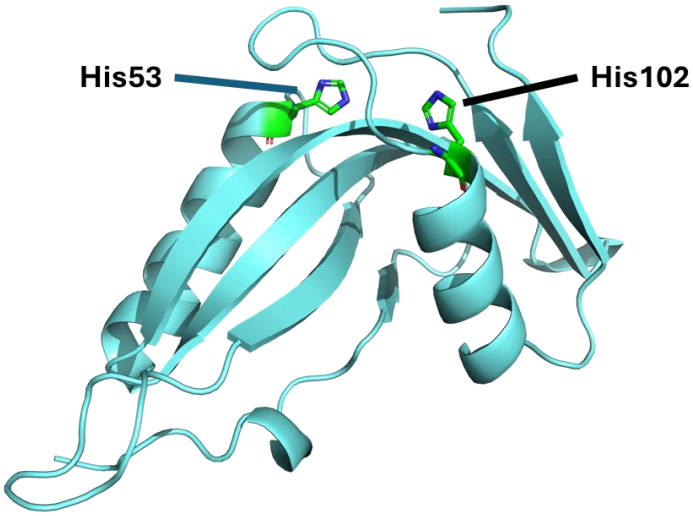


Figure 17: Structure of PHPT1, H53 and H102 are critical for PHPT1 phosphatase activity. The histidine residues are positioned with a degree of vertical symmetry across the protein at the termini of their respective α-helices. [Image produced in PyMOL with AlphaFold PDB file: AF-Q9NRX4-F1-model_v4]

Whilst PHPT1 appears to be the most ubiquitous pHis phosphatase, accounting for activity across multiple biological processes; another protein, phosphoglycerate mutase family 5 (PGAM5) also displays pHis phosphatase activity. PGAM5 specifically dephosphorylates NDPK-B (His118) and thus inhibits the phosphorylation and activation of the KCa3.1 response negatively regulating the cytokine producing $CD4^+$ T cell response (63). In contrast to the broad-acting phosphatase PHPT1, PGAM5 appears to be localised to the inner mitochondrial membrane (105, 106) and shows some promiscuity as a serine/threonine phosphatase dephosphorylating a serine residue on a mitophagy receptor, FUNDC1, which consequently results in degradation/destruction of mitochondria (mitophagy) (106)

The phosphoglycerate mutase family (PGAM) typically catalyse the conversion of 3-phosphoglycerate to 2-phosphoglycerate but PGAM5 appears to differ by partaking in pSer/Thr and pHis phosphatase activity specifically in the mitochondria instead of mutase activity (107). PGAM5 contains a conserved histidine residue, His105, that appears to be critical for Ser/Thr phosphatase activity (**Figure 18**) with mutation removing His105 PGAM5

no longer exhibits phosphatase activity with apoptosis signalling kinase 1 (ASK1) and phosphothreonine peptides (107).

Similarly, PGAM5 is reputedly responsible for removal of phosphate from His118 in NDPK-B, transferring the phosphate to His105 and consequently attenuating the KCa3.1 channel response which is activated by NDPK-B phosphorylation of the ion channel (63). No phosphatase activity was observed in PGAM5 mutant (H105A), however binding was still observed despite lack of enzymatic activity showing a specific interaction between PGAM5 and NDPK-B that is not observed with other kinases including NDPK-A (63).

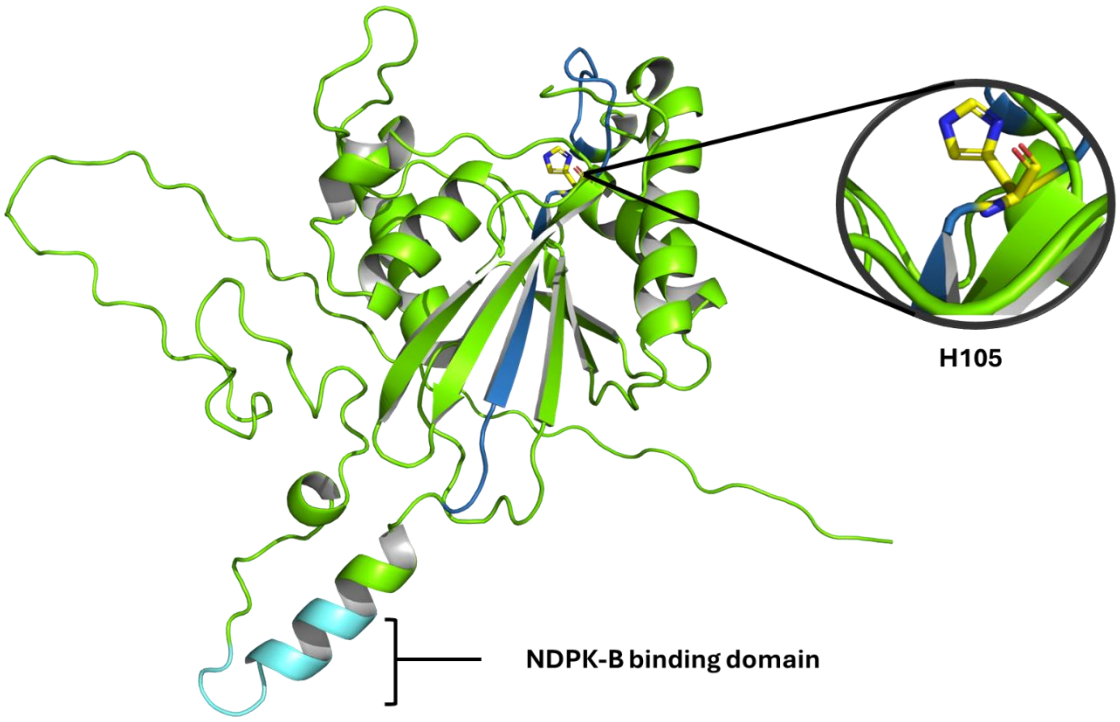


Figure 18: PGAM5 structure, central β-sheet enveloped by four α-helices forming the phosphatase domain. The central portion of the β-sheet contains primary sequence flanking the catalytic histidine residue H105. An α-helix (pale blue) separate from the main phosphatase domain contains amino acids 77-88 which are essential for NDPK-B binding prior to dephosphorylation. [Image produced in PyMOL with AlphaFold PDB file: AF-Q96HS1-F1-model_v4]

In contrast to PHPT1 with more exposed histidine residues (His53 and His102) as displayed in **Figure 17**, PGAM5 has a more enveloped histidine residue which may explain why, in addition to its localisation to the mitochondrial cellular domain, this phosphatase exhibits more specific binding to its interactors (e.g. preference for NDPK-B and no binding to the closely related isoform NDPK-A despite sharing significant sequence homology and an analogous histidine residue (His118) (63)).

Additionally, binding of PGAM5 to NDPK-B is observed even in the PGAM5 mutant (H105A) indicating that binding of the two species is essential for dephosphorylation (63). This binding is reliant on the amino acids 77-88 in the PGAM5 sequence which are present in a terminal α-helix (63), indicated in **Figure 18**.

The structure of the PGAM5 active site is displayed in **Figure 19**. There is clear empirical evidence for the activity of PGAM5 being contingent on the presence of H105. Looking at the active histidine residue in a monomer of the protein structure also provides interesting structural discussion. The active histidine appears to be surrounded by a pair of arginine residues (Arg104 and Arg152), which is common in phosphatases and phosphotransferases as the positive charges of these residues stabilises the phosphate intermediate during transfer from one species to another. Additionally polar residues are present to support the active site; in PHPT1, for example, tyrosine 52 and 93 (Tyr52 and Tyr93) are present in the active site and in PGAM5 a tyrosine residue (Tyr108) overlooks the catalytic histidine residue.

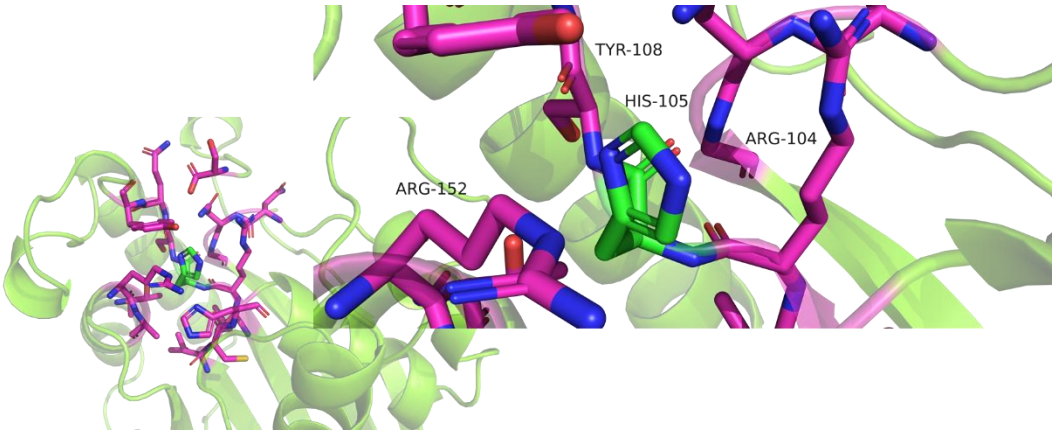


Figure 19: PGAM5 active site structure centred on catalytic histidine (His105). His105 is surrounded by two arginine residues (Arg104, Arg152) providing positive charge stabilisation for phosphate transfer. Arginine and Lysine are commonly observed nearby active residues for the purpose of stabilising this transition. Tyrosine (Tyr108) sits above the catalytic histidine residue and is also found in the active site of PHPT1. [Images produced in PyMOL with AlphaFold PDB file: AF-Q96HS1-F1-model_v4]

PHPT1 and PGAM5 are better characterised phosphatases; PHPT1 shows an almost ubiquitous role as a phosphohistidine phosphatase, while PGAM5 shows specificity as a phosphohistidine phosphatase for NDPK-B and acts more broadly as a Ser/Thr phosphatase as well. An additional phosphatase LHPP also acts as a phosphohistidine phosphatase, although no specific substrates have so far been identified and elucidated, and no structural discussion appears to be available.

Looking at the structures of the known phosphohistidine phosphatases PHPT1 and PGAM5, they share ~18% commonality (percentage identity matrix) and no clear motifs are apparent in the primary sequences to shows any correspondence between primary structure and function or conserved functionality. So far, both phosphatases rely on a catalytic histidine residue His52 and His105 in PHPT1 and PGAM5 respectively and each of these residues is flanked by loops of positively charged amino acids, such as lysine and arginine, as well as polar residues serine and tyrosine.

Although speculative, it is worth looking at the histidine residues present in LHPP and assessing its potential as a histidine phosphatase from a rudimentary overview of its structure. Whilst not as empirically thorough as excising segments of protein or mutating histidine residues for alanine, based on what has been shown for known phosphatases there should be at least one histidine residue enveloped in a loop of unstructured residues that are capable of stabilising a phosphotransfer (a histidine with proximal arginine/lysine and polar amino acids serines/tyrosine).

The LHPP structure contains 5 histidine residues. One histidine residue His103 sits within an unstructured region on a primary chain between a β-sheet and α-helix and on a primary sequence YLLIHD which shares some analogy with a sequence found in PHPT1 - YVLIRV (potential YXLI motif). LHPP His103 is surrounded by aspartate residues (Asp104 and Asp125) which are important for coordinating metal ions in phosphotransfer reactions. Additionally, there is a proximal glutamine residue (Gln58) that may well stabilise water residues during phosphatase activity. The main issues with His103 as a target histidine residue is the lack of surrounding arginine and lysine residues which appear to be crucial and numerous around the active sites of the other known histidine phosphatases, PHPT1 and PGAM5. Additionally, whilst the presence of negatively charged Asp may enable co-ordination of metal ions, the presence of negative charge would destabilise a phosphate group. **Figure 20** (**a**) shows His103 in a model structure.

The second histidine that may be a possible catalytic residue is His245, on an exposed unstructured region between two α-helices. His245 is overlooked by an unstructured loop (LGKGRY - amino acids: 153-158). This loop contains two positively charged residue arginine (Arg157) and lysine (Lys155) that are typically observed around phosphotransfer active sites for stabilising the negative charge on phosphate. Also, the LGKGR sequence in LHPP shares analogy with LGGGR observed in PHPT1, where the Arg78 residue in LGGGR is depicted as

a key stabilising catalytic residue for His53 in PHPT1. It is also surrounded by an aspartate residue (Asp242) and glutamate residues (Glu243, Glu247), which may well provide co-ordination to metal ions during phosphorylation. However, it is also worth noting again that a region of negative charge could destabilise a phosphate group. **Figure 20** (**b**) depicts this potential active site.

His267 in LHPP sits at the base of an α-helix and is in proximity of a lysine (Lys270) and a glutamine (Gln266) that may stabilise the phosphate and water molecule respectively during enzymatic dephosphorylation. These sites are illustrated in **Figure 20** (**c**).

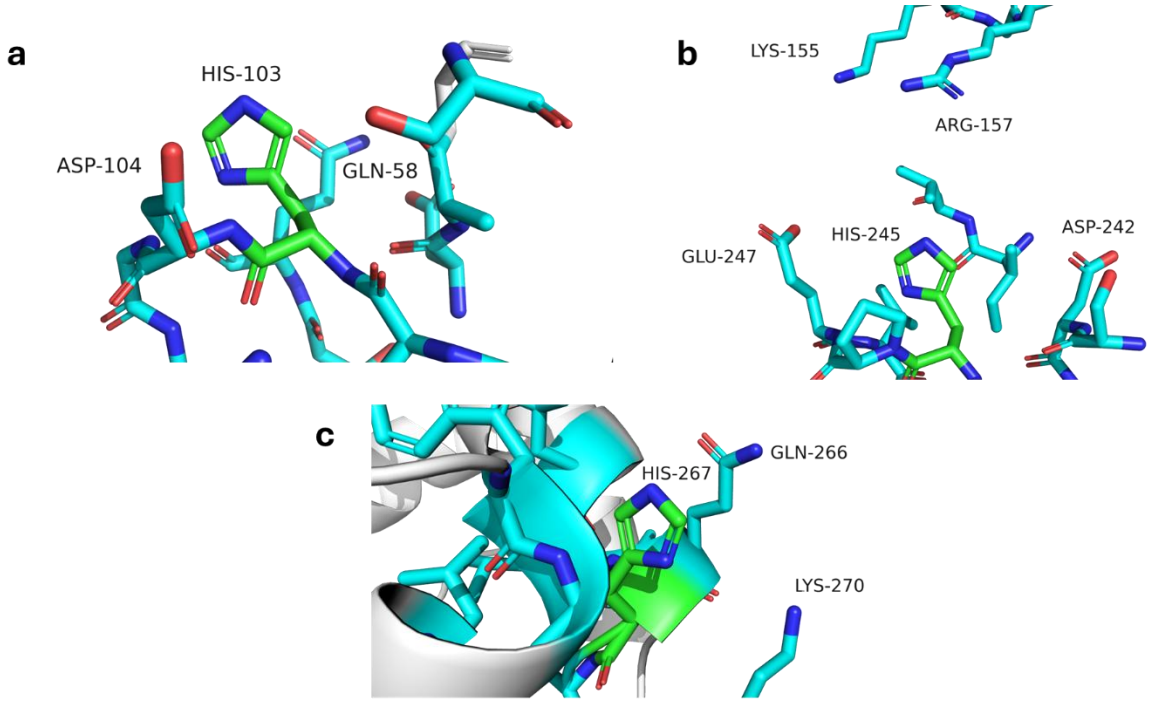


Figure 20: Potential histidine catalytic residues identified on models of LHPP. His103 flanked by Gln58 and Asp104 (**a**). His245 active site surrounded by Asp242 and Glu247, His245 is beneath overhanging Lys155 and Arg157 on an LGKGR chain which shows analogy with an PHPT1 LGGGR sequence (**b**). His267 at the outside of the LHPP structure on the end of an α-helix, in proximity of Gln266 and Lys270 (**c**). [Images produced in PyMOL with AlphaFold PDB file: AF-Q9H008-F1-model_v4]

Until an exact pHis substrate is identified that interacts with LHPP, the phosphohistidine phosphatase activity of LHPP is hard to discuss on a structural basis. However, based on the overview of LHPP structural models, it is clear that there are potential histidine sites that may be capable of phosphatase activity.

1.5 Detection and enrichment of phosphohistidine

Understanding of the phosphoproteome is centred on empirical evidence for the existence of proteins containing phosphorylated amino acids. For a phosphorylated protein to be identified methods must be developed for selecting and analysing these post-translationally modified proteins. The following sub-chapter addresses the workflows performed to prepare protein mixtures for terminal analysis and the methods employed to enhance the proportion of the desired phosphoprotein in the analytical sample.

1.5.1 Distinguishing between phosphoproteins and phosphopeptides

The detection of proteins containing phosphate esters (pSer/pThr/pTyr) is typically carried out with mass spectrometry (MS) after an extensive proteomics preparation workflow involving enrichment steps to amplify the presence of the desired protein (108), a typical workflow is outlined in **Figure 21**. Phosphoproteins already tend to occur in low abundance relative to other species and will occur in conjunction with their unphosphorylated counterpart which complicates downstream MS analysis (108). A method for mitigating this complication and improving the detection of the desired phosphoprotein in MS analysis is to perform enrichment steps which remove undesired proteins and therefore reduce "background noise" on MS analysis, consequently improving phosphoprotein identification.

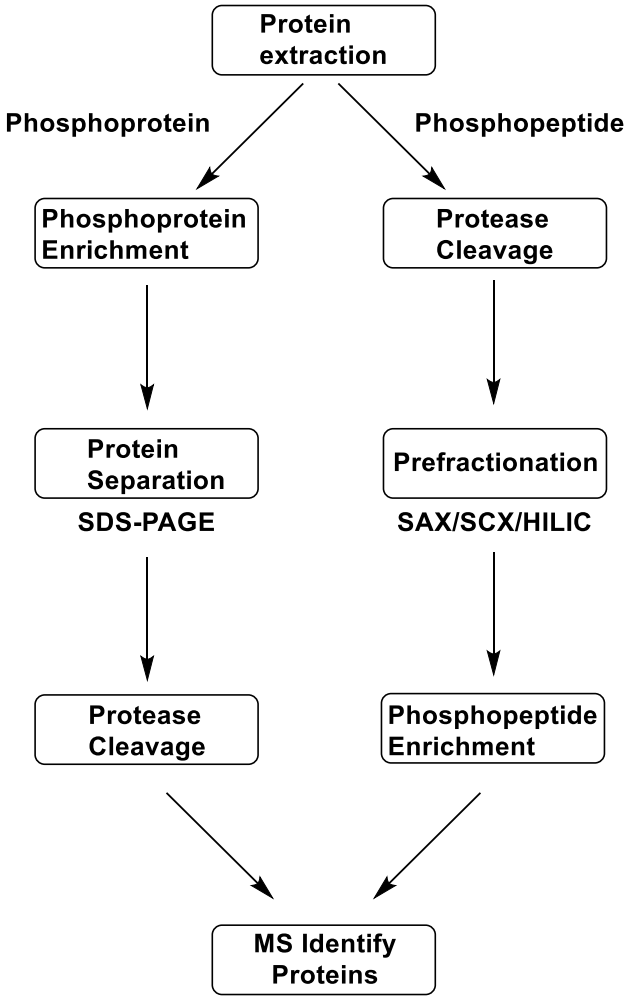


Figure 21: Workflow from protein extraction to identification of phosphoproteins. The left workflow isolates and purifies phosphoproteins with a terminal protease cleavage producing phosphopeptides. The right workflow opens with protease cleavage and purifies crude phosphopeptides in preparation for MS analysis (108).

As outlined in **Figure 21**, a proteomics workflow for phosphoprotein identification can focus on purifying phosphoproteins with subsequent protease cleavage to form phosphopeptides for MS analysis or phosphoproteins can be immediately digested to produce phosphopeptides for subsequent purification and analysis.

1.5.2 An overview of enrichment methods

Enrichment, in this context, is the removal of undesired protein or peptide substrates from a complex assortment of proteins or peptides respectively, to leave behind a larger proportion of the target phosphoprotein/phosphopeptide. From the aforementioned workflow, **Figure 21**, enrichment is a significant step in the processing of both phosphoproteins and phosphopeptides for subsequent analysis. There are many ways to enrich a protein sample to improve the quality of the obtained sample data: IMAC, MOAC, AEX and immunoprecipitation are the methods that shall be outlined to aid in the upcoming discussion of pHis-phosphopeptide enrichment.

Immobilised Metal Affinity Chromatography (IMAC) is a method of enrichment utilising the affinity of phosphate for positively charged metal ions to separate desired phosphate compounds from a mixture. A tightly packed column consisting of a solid-phase resin binds a positively charged metal ion such as: Fe^{3+} (109, 110), Cu^{2+} (111), Al^{3+} (112) and Ga^{3+} (113). The negative charge density of phosphate enables strong interactions between phosphoprotein/phosphopeptides and the resin-bound metals while the unphosphorylated material elutes from the column. After washes with solvent to ensure excess material is removed, a terminal wash with an eluting buffer releases the phosphoprotein/peptide from the matrix (108). In some cases sequential-IMAC columns (SIMAC) can be utilised to enhance separation if allowing for separation of differing phosphates particularly where mixtures involve a mixture of mono- and multiply phosphorylated peptides (114).

Metal Oxide Affinity Chromatography (MOAC), in principle, is similar to IMAC; metal oxides (TiO_2 (115), ZrO_2 (116), Ga_2O_3 (117), Nb_2O_3 (118)) are used as binders for the phosphate in phosphoproteins/phosphopeptides, the most common of the metal oxides used is TiO_2 (119, 120). The workflow requires acidification of the crude protein/peptide sample to protonate acidic residues on a protein/peptide to prevent off-target binding of these residues to the metal oxide (119). After incubation of the acidified protein/peptide solution with the metal oxide, an alkaline solution can be used to free the phosphoproteins/phosphopeptides from the oxide binding into solution. One study, for example, found that sequential basic washes with NH_4OH (5%), piperidine sol. (5%) and pyrrolidine sol. (5%) significantly improved total peak area of phosphopeptides in final MS analysis (121).

Anion Exchange Chromatography (AEX) is a method for separating negatively charged phosphoproteins through interaction of the negative phosphate charge with a polyethylenimine-

silica material solid phase (122). There are two types commonly discussed, weak anion exchange (WAX) and strong anion exchange (SAX), the difference is based on the material on the stationary phase: WAX relies on primary, secondary and tertiary amines and binding improves at lower pH values, while SAX is a similar structure which predominantly consists of quaternary ammonium salts; thus, SAX binds phosphopeptides at higher pH values due to the presence of cations at physiological pH (122). The AEX column is then flushed with increasing concentrations of salt which will displace the bound molecules from the stationary phase based on affinity for the stationary phase (123). More negatively charged species (such as phosphoproteins/phosphopeptides) will be eluted later than less negatively charged species, creating an elution gradient, enabling separation of species in a complex mixture (124). Elution can also be achieved with pH alteration as pH decreases more negatively charged molecules will be released by the cationic stationary phase creating a separation gradient (123).

Immunoprecipitation relies on the binding specificity of an antibody for its corresponding antigen (108). Unlike the aforementioned methods, that rely on binding through opposite charge attraction, immunoprecipitation relies on a solid-phase bound antibody that specifically recognises a feature of the desired compound. In the context of precipitating phosphoproteins/phosphopeptides an antibody would specifically recognise a motif in the phosphoprotein/phosphopeptide (typically the phosphorylated amino acid), thus in immunoprecipitation the solid phase is specific for the desired phosphoprotein/phosphopeptide which theoretically enables precise enrichment of a protein/peptide mixture (125). In immunoprecipitation the target phosphoprotein/phosphopeptide interact with solid-bound antibodies (antibodies linked to magnetic/agarose beads), the beads can then be separated from the heterogeneous sample of proteins/peptides. Washing the beads with increasing concentrations of salt solution disrupts the binding of the phosphoprotein/phosphopeptide from the antibody before analysis by MS (126).

1.5.3 Progress towards enrichment and analysis of phosphohistidine peptides

Unfortunately, given phosphohistidine's (pHis) propensity for hydrolysis at mildly acidic conditions, enrichment of phosphohistidine proteins in preparation for MS detection is significantly more challenging than for canonical phosphoamino acids (127). One of the most common methods of enrichment is IMAC, using metal cations to bind the phosphate on the phosphoprotein/phosphopeptide (**Figure 22**). Work by Napper *et al.* suggests that of: Ga^{3+}, Cu^{2+} and Fe^{3+}, only Cu^{2+} showed binding capability to pHis-peptides (111), optimisation of metal selection during IMAC may enable isolation of pHis peptides in the presence of other phosphopeptides (pSer peptides showed specific binding to Ga^{3+}) (111). However, an issue raised is the natural affinity of histidine for immobilised metal ions, thus unphosphorylated histidine peptides will be bound by the immobilised metal ions and appear in the final analytical mixture.

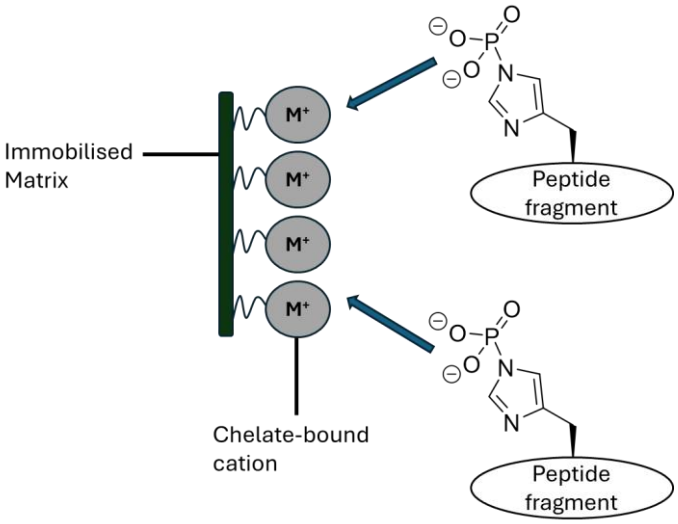


Figure 22: Cationic metal ions bound to immobilised matrix bind to negative charges on phosphate. Phosphohistidine in peptide fragments will be concentrated in the column leading to amplification of those peptides in a final sample (enrichment).

To reduce the quantity of histidine peptides, Napper *et al.* discuss the possibility of mildly pre-acidifying the mixture of peptides such that the basic histidine imidazole nitrogen (pK_a 6.5) is protonated and thus repelled by the immobilised metal cations (111). However, as noted by Attwood *et al.*; proteins rich in acidic side chains (Asp and Glu) with a carboxyl side chain pK_a of approximately 4.4, will not be deprotonated in mildly acidic conditions and still bind to the immobilised metal complicating enrichment (8). Napper *et al.* acidified their mixture of peptides to pH ~ 3.5 (0.1% acetic acid) to prevent the imidazole of unphosphorylated histidine binding to the immobilised metal ions (111). However, a point not discussed with regards to

the protein mixture acidification, the phosphohistidine residue will be dephosphorylated by the acidic conditions used to eliminate histidine contamination resulting in loss of phosphorylated histidine by the same mechanism used to eliminate histidine impurities. pHis isoforms, τ-pHis and π-pHis, have pKa values of 6.4 (at 25 ˚C) (23) and 7.3 (at 46 ˚C) (24) respectively, therefore even at pH 6, pHis will be present as a greater proportion of protonated isoform to conjugate base with the ratio of protonated pHis increasing by orders of magnitude for every integer decrease in pH value; at pH 3.5 used by Napper *et al.* (111) pHis peptides would be susceptible to huge levels of hydrolysis. This is the most fundamental issue with IMAC, for enrichment of pHis peptides, stringent pH control is required for the optimal preservation of the modification for analysis. This hydrolysis/loss of pHis peptides is reflected in relative abundances of pHis detection in final ESI-MS by positive ion mode where peptides appear in 1:25 (pHis: His) ratio (8). Attwood *et al.* also briefly alludes to the degradation of pHis in MALDI-TOF during both positive and negative ion modes once again emphasising the issues for confirming the presence and abundance of the pHis modification (8). The reliance of IMAC enrichment and subsequent analytical methods such as ESI-MS and MALDI-TOF on using acidic conditions in workflows makes the analysis of pHis peptides very challenging (given pHis has a half-life of approximately 30 minutes at pH 3) (128).

If enrichment were to be achieved for pHis peptides it would require a method at neutral or basic conditions to mitigate the effect of hydrolysis. One method discussed is use of polymeric material (Strata-X) under basic conditions (129, 130) to capture canonical phosphopeptides from tryptic digests. Strata-X binds preferentially to more hydrophobic residues, so inherently hydrophilic peptides such as phosphopeptides will be eluted in the flowthrough (129). Analysis of the flowthrough and retained material will enable phosphopeptide to be identified by the discrepancy in spectral data; in one study, the peak for a pHis peptide (ALLMAGVApHDLR) was significantly more intense on the flowthrough spectrum compared to the retained material eluted subsequently (129). This highlights the potential for new materials or chromatographic methods to improve enrichment methods. Additionally, the same study investigates the potential for identifying phosphohistidine in MS/MS using varied conditions, for an optimal intensity signal a short gradient on LC-MS in acidic conditions is best to enable separation but capture data before pHis has time to decompose significantly (129).

Strong-anion exchange chromatography (SAX) has been used for pHis-peptide enrichment and fractionation (127). SAX has already been outlined as a chromatography method with a quaternary amine salt stationary phase for capable of binding the double negative charges of

phosphate at pH values greater than 6 (122, 131). Therefore, binding can occur at higher pH conditions which is beneficial for pHis due to its inherent instability in acidic conditions. Whilst SAX enables enrichment at higher pH values, there are some disadvantages to this approach; acidic residues such as Asp and Glu will be deprotonated at higher pH values and thus bind to the stationary phase compromising the quality enrichment process (8). Elution on a pH gradient would not be suitable for pHis-peptides given their hydrolysis profile at lowering pH, therefore removal of pHis-peptides from the stationary phase can be achieved, according to Hardman *et al.*, with a change in ionic gradient with a triethylammonium phosphate buffer in increasing concentrations based on a technique utilised by Alpert *et al.* (122, 127). This process of using SAX at a higher pH and utilising an ionic gradient has been termed UPAX (unbiased phosphopeptide anion exchange) chromatography (127). Effectively, UPAX is indiscriminate between canonical and non-canonical fractionation because it operates at a higher pH, all phosphate species should show a greater affinity for the column than unphosphorylated peptides, enabling "unbiased" phosphopeptide separation (127). Hardman *et al.* test fractionation at pH (6, 6.8 and 8) and determine that pH 6.8 offers optimal preservation and fractionation of the pHis modification in peptides of myoglobin cell lysates (127). UPAX performed under these conditions led to the identification of 83 unique pHis phosphosites at high confidence in HeLa cell lines (127).

One method for enriching is generating antibodies to the desired hapten and using those antibodies to specifically capture a residue or motif of interest (immunoprecipitation). Adam *et al.* have utilised immobilised τ- and π-pHis monoclonal antibodies (mAbs), generated from pHis analogues, to enable enrichment of some pHis residues in substrates (132). Fuhs *et al.* generated isoform specific mAbs for pHis by synthesising libraries of peptide sequences containing two isoforms of pTza: 1- and 3- corresponding to the authentic isoforms π- and τ- respectively bound to keyhole limpet hemocyanin (KLH) (25), which is a strong immunogen promoting a lymphocyte response to the peptide library. The 1-pTza library and 3-pTza library are injected into separate groups of rabbits which produce antibodies in response to the antigen, the splenocytes are fused with myelomas to produce hybridomas. The hybridoma cell lines produce the desired antibody on a larger scale for use in assays (25). Fuhs *et al.* have used τ- and π-pHis mAbs in immunoprecipitation assays, performing LCMS on tryptic peptides, showing enrichment of known pHis proteins (NME1/2, PGAM, ACLY and histone H4)(25) as well as another set of pHis proteins found by Lapek *et al.* such as tubulin beta-5, lactate dehydrogenase, GAPDH and more (25, 133). Fuhs *et al.* enriched a total of 786 proteins with

their τ- and π- mAbs, identifying a significant number of potential pHis-containing proteins (25). Oslund *et al.* have generated pan-pHis antibodies which have aided in MS studies of pHis-peptides, enabling immunoprecipitation-based enrichment of pHis tryptic peptides prior to MS analysis (134, 135). 15 pHis sites were identified from 14 *E. coli* proteins as a result of this enrichment study (134). This work has led to the identification of a triplet dissociation pattern in MS/MS (Δ80 (HPO_3), Δ98 ($HPO_3 + H_2O$) and Δ116 ($HPO_3 + 2H_2O$) Da) of neutral losses in collision-induced dissociation (CID) that can act as a tandem MS fingerprint for pHis (134). Loss of Δ80 Da is particularly indicative of pHis, the mass loss is consistent with the breaking of the labile phosphoramidate (P-N) bond (134).

A final method for identifying pHis in peptides is Phosphorus-31 NMR (^{31}P NMR). ^{31}P NMR should be capable of providing a chemical shift value in ppm for the τ- and π-isoforms to assist in the identification of pHis tryptic peptides. ^{31}P NMR has already been mentioned, having been used to identify shifts for τ- and π-phosphohistidine (4.9 and 5.3 ppm respectively) measured from rat liver and Walker-256 carcinosarcomas (42). However, identification of novel pHis substrates with ^{31}P NMR may be complicated by the variance of its chemical shift in differing contexts, pHis shift varies dependent on its presence in a substrate and which isoform is present. Shifts of pHis in NDPK from *Dictyostelium discoidium* showed a π-pHis shift of -2.74 ppm (native NDPK) to -4.20 ppm (denatured NDPK), indicating the discrepancy between native substrate shifts and denatured shifts (136). Additionally, comparison between pHis amino acid standards and phosphorylated NDPK was not diagnostic for determining the isoform of pHis present in NDPK; a synthetic pHis phosphorylated peptide (Glu-pHis-Gly) was required for confirmation that the isoform of pHis in NDPK is π-pHis (136). As such, pHis amino acid standards may not be suitable as references for identifying novel pHis shifts. The chemical shift of pHis in ^{31}P NMR is pH dependent, each isoform has unique shifts that vary with pH (137); the shift is more than likely a consequence of the three forms of phosphoryl group: the phosphonic acid, dianion and monoanion (8). Work by Himmel *et al.* showed that isoforms of pHis can be distinguished through NMR analysis by combining hydrogen-nitrogen HSQC and nitrogen-phosphorus HSQC; these plots can be combined to form correlations referred to as HNP correlation patterns for specific isoform identification (138). Additionally, they acknowledged the pH dependence of J coupling values between ^{15}N-^{31}P nuclei and plotted the impact of this to characterise pHis isoform shift at varying pH (138). Absolute ^{31}P NMR shifts for each isoform appear to vary slightly between studies for τ- and π-pHis, ~4.7 ppm and ~5.5 ppm respectively (137); ~ -3.9 ppm and ~ -4.9 ppm respectively (139); ~ -5.0 ppm and ~

-5.9 ppm respectively (138); -4.99 ppm and -5.76 ppm respectively (140). Whilst work by Lecroisey *et al.* demonstrated that caution is required before assigning a ^{31}P NMR pHis isoform shift based on amino acid standards (136), Makwana *et al.* use a denatured myoglobin-pHis shift of ~ -4.94 ppm to assign the isoform as τ-pHis based on amino acid standards (140). Evidently, there is considerable research regarding pHis identification with ^{31}P NMR; but, as with MS analysis there is considerable challenge in verifying pHis isoform and presence by ^{31}P NMR alone, especially given the discrepancy of NMR shifts observed between studies. Where mixed isoforms are present, it is frequently determined that τ-pHis has a less negative chemical shift than the π-pHis isoform, which may well be diagnostic in conjunction with careful use of amino acid standards.

1.5.4 Scepticism towards identified phosphohistidine phosphosites

Recent research by Leijten *et al.* has challenged the idea that pHis is a widespread modification in mammalian cells; claiming that approximately 0.4% of all the original pHis phosphosites in mammalian cells are legitimate (141). This contradicts the seemingly ubiquitous belief that, up until now, pHis likely has an undiscovered but significantly widespread function in the human proteome. The study uses Fe^{3+} IMAC to isolate proteins from various cell lines and filters out all but class I phosphosites (sites with high confidence in phosphate localisation to a single residue) (142, 143). The study claims that, due to a greater reduction in pHis than pSer, pThr and pTyr sites (~80% pSer conserved while ~70-80% pHis lost) (141); the veracity of the initially identified pHis sites must be questionable. Leijten *et al.* filtered additionally by considering the appearance of these phosphosites in multiple cell lines (A431, HEK293T, HeLa and PC9), finding only 10% of these pHis sites were present in all 4 cell lines, further reduced to 5% when accounting for the class I filter (141). This stringent filtering only preserved 4 already documented pHis sites: pHis 118 (NME1/2), pHis 11 (PGAM1), pHis 254 (PFKFB3), pHis 299 (SUCLG1) (141).

The work by Leijten *et al.* (141) indicates that sole pHis sites are not particularly common, but the filtering for class I phosphosites does provide room for interpretation. Perhaps pHis operates in a different way to pSer, so its phosphosites are typically neighboured by other potential phosphosites – rationalising the observed sharp drop with class I filtering. For example, pHis may act as an intramolecular phospho-relay, shuttling phosphate from one residue to another, thus examples of suitably isolated pHis, in line with class I filtering, would be comparatively small. Altogether, this study perhaps indicates something about the nature of pHis sites in proteins and in and of itself does not, necessarily, indicate that pHis is an insignificant modification. It does, perhaps, introduce a healthy amount of scepticism as to the role of pHis in mammalian cells; perhaps there is a lower abundance of as-yet-undiscovered and phosphohistidine-specific machinery in the eukaryotic cell than previously believed.

1.6 Antibody generation and the development of stable analogues of Phosphohistidine

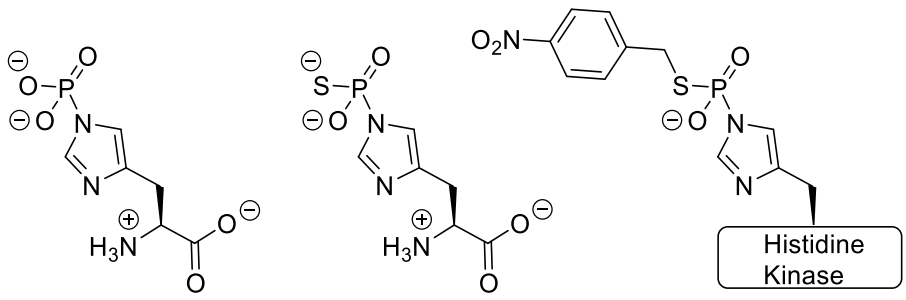
Access to phosphohistidine (pHis) specific antibodies has enabled huge advances in phosphoprotein enrichment and identifying phosphosites (25).

Antibodies can be extremely specific for their hapten, or analogues of the immunogen; phosphotyrosine (pTyr), for example, has a variety of antibodies that are specific for its structure over the unphosphorylated form or other phosphorylated amino acids (phosphoserine (pSer)/phosphothreonine (pThr) specifically) (144-146). However, antibodies can be specific for phosphorylated residues in specific protein substructures, and therefore cannot be used as a global identifier of a specific amino acid in various different structures; this is certainly the case for many pSer and pThr antibodies, therefore adding great challenge for pSer and pThr identification in multiple environments (144).

Given the structural instability of pHis; most immunogenic sources would be hydrolysed far too rapidly for antibodies to be generated against it as a hapten, thus it is challenging to isolate those antibodies for use as a pHis detection reagent, consequently precluding understanding of the pHis proteome especially in comparison to the more stable and prevalent phosphoserine, threonine and tyrosine (2, 147). Whilst antibodies may be challenging to raise against the native phosphohistidine, a major discovery gave credence to the idea that it may be possible to raise antibodies to pHis; pHis in ATP citrate-lyase can be bound by an antibody that was designed to predominantly bind pTyr (2, 148).

Given the benefit of generating antibodies and the proof of principle demonstrated by ATP-citrate-lyase binding, researchers were keen to develop stable analogues of pHis to which they could successfully raise antibodies. An early attempt to generate analogues was made by Lasker *et al.* (1999); they produced τ-thiophosphohistidine (thiopHis) **9** by incubating histidine with $PSCl_3$, the product differed from τ-phosphohistidine **4** only through the replacement of a sigma-bonded oxygen for a sulphur atom. The sulphur is less electronegative than oxygen and thus the electron density in the phosphoramidate bond is higher than in the native pHis; the thiopHis has a lower energy phosphoramidate bond than native pHis resulting in hydrolytic stability. The stability of the analogue was verified in acidic conditions at pH 1 and 0, pHis had hydrolysed up to 50% and completely hydrolysed respectively whereas thiopHis is stable at pH 1 and shows 29.1% hydrolysis at pH 0 (149). The structural similarity coupled with hydrolytic stability seemed like promising properties for this analogue as a stable hapten; but

no attempt was made to generate antibodies against this immunogen. There are no recorded attempts to generate antibodies against thiopHis; later analogues with a similar substructure may give some insight into the likely outcome of attempting to generate antibodies to thiopHis. A similar analogue to thiopHis, a thioester with a p-nitrobenzyl moiety attached **10**, was used as a means of probing phosphorylation; the idea was to use an ATPγS kinase that thiophosphorylates its respective substrates, with a p-nitrobenzyl derivative, and consequently, having generated antibodies to the thiophosphorylated amino acid, leads to detection of phosphorylation sites through the use of thiophosphoryl amino acid specific antibodies (150). The technique was used to study phosphohistidine; the antibodies raised were specific for the thioester component and thus independent of amino acid, but they could not distinguish between the phosphorylated amino acids, which makes the technique obsolete for studying novel phosphorylation – where the modified amino acid needs to be determined; the need to use a bio-orthogonal kinase to study the phosphorylation precludes detection of phosphorylation in novel scenarios – where the kinase is not known and thus a bio-orthogonal equivalent cannot be engineered (151).



4 **9** **10**

Figure 23: Structures of τ-phosphohistidine **4**, thiophosphohistidine **9** and thiophosphohistidine (p-nitro substituted) **10**.

Schenkels *et al.* developed a (4'-phospho-2'-furyl) alanine analogue **11** of τ-phosphohistidine. In contrast to Lasker *et al.*, Schenkels *et al.* replaced the phosphoramidate with a more stable carbon-phosphorus bond to prevent hydrolysis of the analogue and replaced the unsubstituted nitrogen, which would be phosphorylated in π-phosphohistidine, with an oxygen in an attempt to mimic the electronic surface of τ-phosphohistidine (sp^2 plane, hydrogen bond acceptor site). Schenkels *et al.* also provided the structure of phosphopyrrole, though there is clear scepticism as to the success this would have as an analogue, the pyrrole analogue would have an sp^2 H-bond donor nitrogen, inferring that authentic mimicry would be undermined by this electronic surface discrepancy (152). There also appears to be no attempt to generate antibodies to the

furyl analogue **11**; but an attempt to raise antibodies to an electronically similar analogue may give credence to the rationale that pHis specific antibodies would not be raised to Schenkels' furyl analogue. Lilley (2014), attempted to raise antibodies against an analogue **13** of pHis, similar to Schenkels' analogue. The analogue raised antibodies, but they were strongly reactive towards phosphotyrosine and showed no activity towards τ-phosphohistidine (153). Consequently, it may be appropriate to assume that the electronically analogous (4'-phospho-2'-furyl) alanine would likely act as an epitope for phosphotyrosine selective antibodies. Attwood (2007), attempted to generate polyclonal antibodies to the phosphopyrrole analogue **12**, proposed by Schenkels, and found that the antibodies were specific for the epitope and did not recognise phosphohistidine, one might argue that this supports Schenkels' suggestion of the need for a hydrogen bond accepting sp^2 lone pair and consequent dismissal of phosphopyrrole, though no binding studies have shown that this lone pair is essential for binding (8, 152).

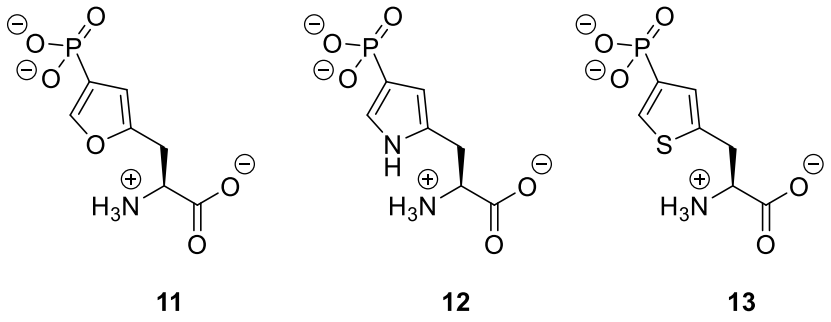
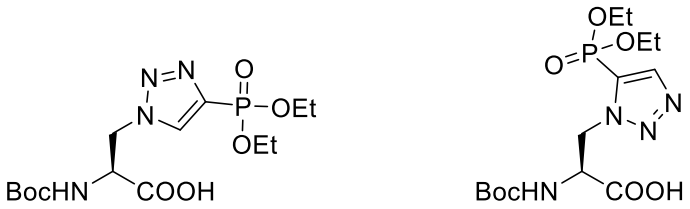


Figure 24: Structures of (4'-phospho-2'-furyl) alanine **11**, phosphopyrrole analogue **12** and phosphothiophene analogue **13**.

Kee *et al.* (2010) created analogues of pHis, phosphoryltriazolylalanine (pTza), as both a τ-phosphohistidine **14** and π-phosphohistidine **15** isomer. The analogues contain a triazole heterocycle as opposed to the imidazole heterocycle in histidine and contain non-hydrolysable phosphonate (P-C) bonds in contrast to phosphohistidine's hydrolysable phosphoramidate (P-N) bond. The electronic conformation of the analogues closely resembles native phosphohistidine, except for the additional surface potential around the extra nitrogen and its lone pair of electrons which, if the shortcomings of the phosphopyrrole analogue are to be considered, lacking an sp^2 plane ring donor relative to phosphohistidine, would likely preclude successful mimicry (8, 154). Kee *et al.* incorporated these analogues as Boc-protected amino acids into proteins using SPPS. They created two peptides, a τ- and π- isomer based structure, corresponding to the N-terminal tail of the histone H4 protein (replacing His18 for the analogous residue). The peptides, incorporating analogues, were bound to keyhole limpet

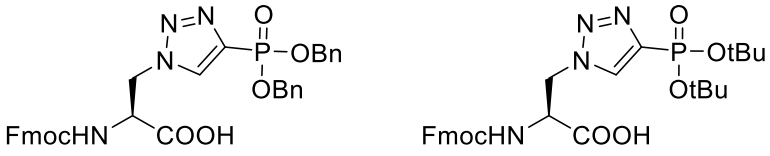
haemocyanin (KLH) and used to generate antibodies; the antibodies appeared to have strong selectivity for His18, as demonstrated by Western blot analysis. However, the antibody generated (Ab-3pHis) shows a degree of cross-reactivity with pTyr (147). The antibody generated therefore is not quite as sequence-independent as would be desired for eukaryotic proteomic analysis, where pTyr is particularly abundant. histone H4 contains two phosphorylated residues, His18 and His75, although the designs by Kee *et al.* centred around generating antibodies to pTza in the place of His18, the Ab-3pHis were only capable of identifying His18 in the native substrate and not His75 (147).



14 **15**

Figure 25: Structure of Boc-protected pTza analogues of pHis, τ-pTza **14** and π-pTza **15**

McAllister *et al.* (2011), having observed the potential for incorporating pTza analogues into degenerate peptide libraries, developed a synthetic route to form the pTza analogues **16** and **17** (**Figure 26**) in a manner compatible with Fmoc-SPPS. Moreover, McAllister presented DFT calculations to rationalise Kee's observations regarding antibody specificity for His18 in histone H4 as opposed to His75; the sequence-dependence of the antibody is due to the aforementioned additional nitrogen electronic surface potential that is not present in phosphohistidine; pTza will bind preferentially through dominant hydrogen bonds to the additional nitrogen, precluding true pHis mimicry (147, 154).



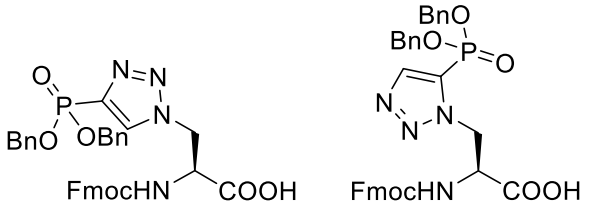
16 **17**

Figure 26: Structures of Fmoc-protected τ-pTza analogues with benzyl-protected phosphonate **16** and OtBu protected phosphonate **17**.

McAllister *et al.* (2014) later assessed the ability of pTza to bind to the SH2 binding domain of the Grb2 protein – a well characterised, prototypical pTyr binding protein. The intention was to prove that the binding domain in Grb2, which appeared to have promiscuous activity with

respect to binding phosphotyrosine and phosphohistidine, could be used to bind pHis in vivo. Inadvertently, this demonstrated that pTza is unsuitable as a specific pHis mimic, as supported not only by DFT calculation but by the cross-reactivity for pHis and pTyr of the antibodies raised to pTza as a hapten by Kee *et al.* in histone H4 (147, 154, 155).

Despite DFT calculations and evidence of binding promiscuity regarding pTza, Fuhs *et al.* (2015) incorporated the pTza analogues, τ-pTza **16** and π-pTza **18** (**Figure 27**), into a degenerate peptide library (peptides consisting of the analogue flanked by a varying assortment of alanine and glycine - so as to avoid any immunogenic epitopes except pTza). They obtained isoform specific τ-pHis and π-pHis mAbs that showed no cross-reactivity with pTyr peptides. The antibodies raised to τ-pTza peptides recognised the τ-pHis in PGAM, whilst the π-pTza antibodies recognised NME1. This suggested that mAbs had been developed with isoform and sequence-independence; however, the study was limited in its use of epitopes to fully support an assertion of sequence-independence, NME1/2 were used to demonstrate the specificity of the π-pHis mAbs and PGAM for the τ-pHis mAbs. However, with such a limited substrate assortment, it is possible that the mAbs had a sequence bias that could not be verified due to a lack of mammalian substrates. This study proved that potentially sequence-independent and isoform specific antibodies could be formed to pHis from stable analogues, a tremendous progression from the sequence-dependent antibodies generated by Kee (2010) to His18 in histone H4; but, the possibility for cross-reactivity with pTyr precludes pTza's use as a phosphomimetic, even if it is capable of generating isoform specific mAbs (25).



16 **18**

Figure 27: Structures of Fmoc-compatible Boc-protected τ-pTza **16** and π-pTza **18** used to generate isoform specific mAbs (after side-chain deprotection).

pTza showed promise as a phosphohistidine analogue but determined to generate more specific antibodies for pHis, Kee *et al.* (2015) used a phosphopyrazole ethylamine (pPye) analogue of pHis **19;** this may have been encouraged by their early research, which noted that phosphotriazolyl ethylamine (pTze) **20** had better performance as a hapten than the corresponding amino acid (pTza). pTze generated antibodies capable of binding to His18 and

His75 in histone H4 and not unphosphorylated histidine, a significant improvement over the sequence-dependent His18 specific antibodies they initially generated against pTza. The antibodies generated against pTze, whilst pan-specific and somewhat selective, did show some cross-reactivity with pTyr; hence the need to further investigate analogues (135, 147). pPye was an ethylamine hapten, which had shown promise previously, with remarkable structural and electronic similarity to the native τ-phosphohistidine, as shown by DFT calculation (156). pPye was capable of producing sequence-independent antibodies capable of binding to pHis selectively over pTyr; some pTyr binding is observed but the antibodies show 1:20 selectivity over 1:10 pPye generated and pTze generated antibodies respectively, a marked improvement in selectivity (156). Furthermore, pPye generated antibodies were shown to be able to bind His18 and His75 in histone H4 as well as Pts1 and PGAM1, demonstrating the sequence independence of the antibodies generated by pPye to detect phosphohistidine *in vitro*.

The work of Lilley (2015) corroborated the suitability of the pyrazole motif to act as a pHis analogue. Using a phosphopyrazole alanine analogue (pPza) **21** of τ-phosphohistidine to generate pHis selective antibodies; through dot blot analysis, they observed strong selectivity for pHis over non-phosphorylated histidine and pTyr, in agreement with observations by Kee *et al*. with the pPye analogue (156, 157). The detection of phosphohistidine in NME1/2 and Gβ proteins with the antibodies, and finding strong selectivity for pHis over other epitopes, confirmed the pyrazole structure was a suitable analogue of histidine regardless of its presence as an amino acid or amine (157).

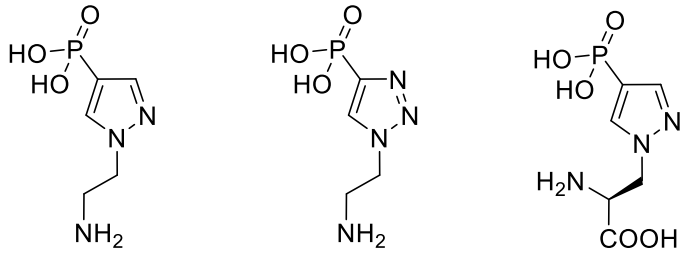


Figure 28: Structures of phosphopyrazole ethylamine **19**, phosphotriazolyl ethylamine **20**, phosphopyrazole alanine **21**.

The generation of pan-specific sequence-independent antibodies should provide a means of selectively detecting pHis, allowing for isolation methods using antibodies to selectively enrich pHis in the presence of other phosphoamino acids.

1.7 Structural analysis of pHis analogues

Of the stable analogues developed for antibody generation, triazole and pyrazole moieties proved the most effective at generating antibodies that cross-reacted with the wild-type authentic pHis (156). Therefore, triazole and pyrazole act as suitable mimics of pHis and could arguably be substituted for pHis in a mimic sequence for use in phosphoproteomics.

Density functional theory (DFT) was often discussed in regard to analogue structure; when recognising the analogues of pHis, the immune system will develop antibodies that are specific for the steric and electronic properties of the epitope; analogues are developed to mimic these qualities (156). The key to developing a pHis analogue is to produce a mimic with similar shape and electronic distribution; as such, phosphonopyrazole (**4**) and phosphonotriazole (**3**) (**Figure 29**) are apt structural motifs for a pHis mimic, both consisting of 5-membered heterocycles with the requisite negative charge density of the phosphate.

The triazole motif generated antibodies that interacted with both the authentic pHis and cross-reacted with pTyr (135). Cross-reaction with pTyr is not ideal because antibodies generated may bind to background sources of pTyr convoluting any analysis (156). Antibodies generated to the pyrazole motif showed significantly more selectivity for the natural pHis modification than the triazole antibodies. Therefore, it could be argued, phosphonopyrazole (**4**) acts as a better mimic for pHis than phosphonotriazole (**3**). DFT may provide a structural explanation for the observed discrepancy in mimicry between phosphonotriazole and phosphonopyrazole as compared to the authentic pHis.

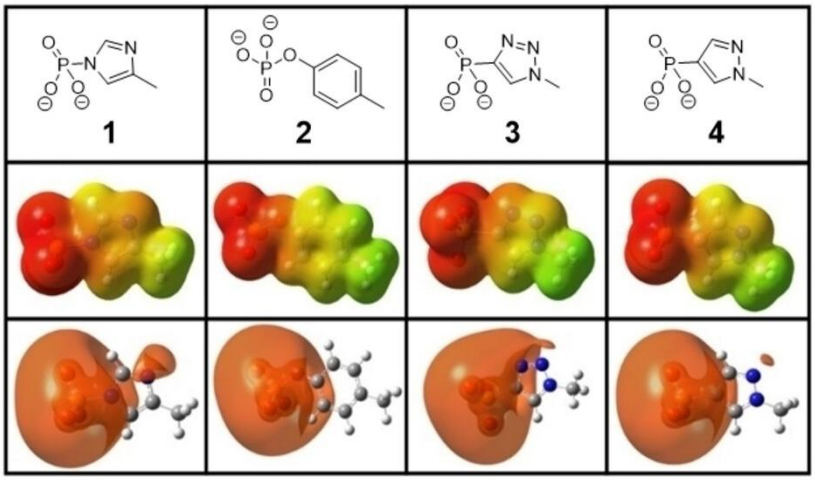


Figure 29: DFT modelling distribution of electron density across pHis (**1**), pTyr (**2**), phosphonotriazole (**3**) and phosphonopyrazole (**4**). (Copied directly from work by Makwana *et al.*) (158)

Crucially, **Figure 29** shows that pHis has a field of electron density over the phosphate and the the free nitrogen in the imidazole ring. pTyr differs sterically as a 6-membered ring and is also not a heterocycle and as such lacks the electron density and hydrogen bonding potentially of the free nitrogen in pHis. pTyr is also critically a phosphate and the additional oxygen adds steric bulk that is not present in the phosphoramidate bond of pHis. The triazole side chain contains an additional basic nitrogen compared to pHis (triazole and imidazole based respectively), thus the additional nitrogen acts as another hydrogen bond acceptor in pTza that is not present in pHis enabling unique interactions (precluding exact mimicry). The pyrazole motif is sterically very similar to pHis and, in common with pHis, has only one free basic nitrogen in the heterocyclic 5-membered ring.

Structurally, pyrazole is a more analogous heterocycle than triazole for imidazole-based phosphohistidine, based on the direction of hydrogen bonding from the lone pair on the pyrazole nitrogen being the same as those in the authentic phosphohistidine; and, this is empirically reflected by pyrazole-generated antibodies showing a greater cross reactivity for pHis than triazole-generated antibodies (156).

1.8 Aims of this work

The phosphohistidine (pHis) post-translational modification has been documented as a key regulator of chromatin processes, ion channel regulation and G-protein signalling. There are histidine kinases (NME1/NME2/NME4) and pHis phosphatases (PHPT1/PGAM5/LHPP) which is evidence for cellular machinery operating to regulate the pHis modification specifically. Evidence for the pHis modification in a variety of distinct biological processes and the presence of specialised kinases and phosphatases for the pHis modification, would indicate that pHis is a ubiquitous cell signalling modification with the potential to be implicated in many as-yet undiscovered biological processes. The pHis phosphatase LHPP strongly exemplifies this investigation; it is a reputed pHis phosphatase with no known specific substrates, there may be many other enzymes regulating pHis that have not been identified.

The challenges of identifying pHis proteins with standard proteomics techniques, due to pHis instability, justifies the slow pace of research with regards to this protein modification. To circumvent enrichment challenges, researchers developed antibodies against stable analogues of pHis to improve the enrichment of pHis proteins through immunoprecipitation assays. The most analogous of these structures, as outlined by density functional theory (DFT) discussion, are based on pyrazole and triazole structures, displayed in **Figure 30**.

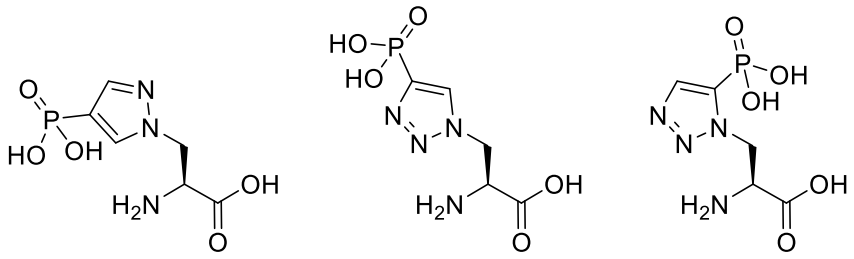


Figure 30: Structures of phosphohistidine analogues based on heterocycles: phosphonopyrazole **21**, τ-phosphonotriazole **22** and π-phosphonotriazole **23**.

This work focusses on developing the pyrazole **21** and triazole-based **22** analogues of phosphohistidine to create reactive probe structures capable of identifying proteins that interact with the pHis modification (such as novel phosphatases). This work will investigate methods for the incorporation of pHis analogues into protein and peptide substrates for producing pHis-mimic proteins or peptides that are capable of being recognised by a cognate partner. Additionally, this work will explore methods for generating covalent binding on these

analogues such that binding partners will be bound to the probe enabling downstream identification by proteomics methods. **Figure 31** is a basic schematic outlining the use of analogues in prior research against how this work aims to adapt and utilise the pHis analogues.

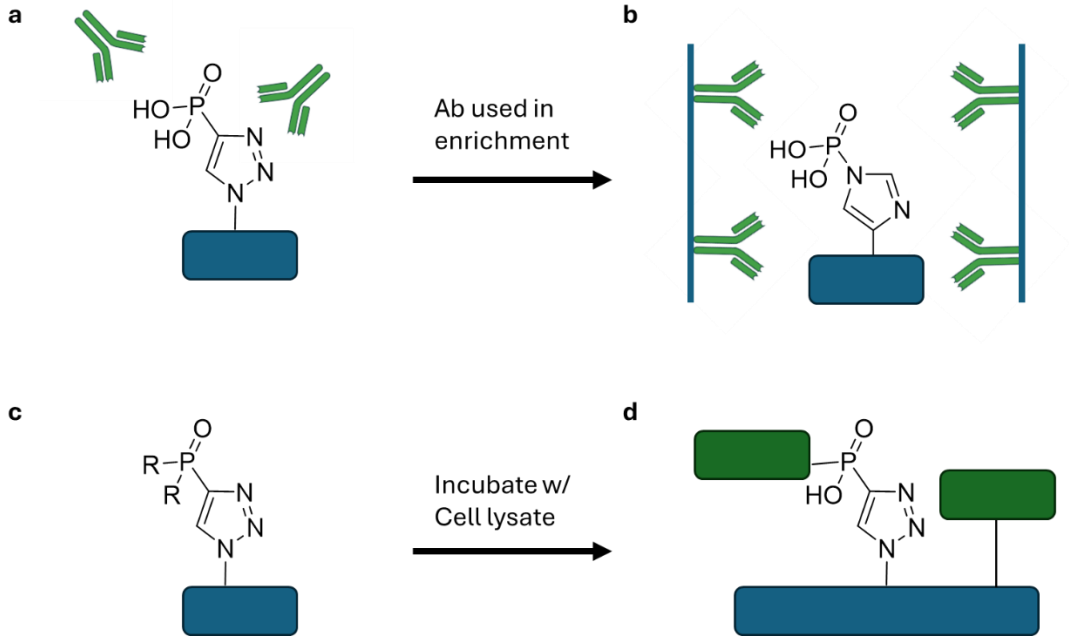


Figure 31: Antibodies generated against pHis analogue (**a**) used in enrichment protocols to bind to authentic pHis modification to enable identification of legitimate pHis phosphosites (**b**). This research will aim to modify a pHis analogue with reactive capability (represented by R) (**c**) or the protein/probe structure itself to enable binding to partner proteins that recognise the authentic pHis (**d**).

2 Towards the incorporation of phosphoryl pyrazole into peptides and proteins

Previous research has focussed on the identification of candidate phosphohistidine (pHis) sites in proteins. A number of pHis analogues have been synthesised (147, 159) with the aim of producing antibodies that could recognise phosphohistidine in protein sequences (25, 156). Such antibodies would theoretically enable isolation of these proteins and subsequent identification. Identifying proteins containing phosphohistidine would give some indication of the role of phosphohistidine within the cell and thus the processes that it regulates. Their endeavours have been successful with Srivastava *et al.* identifying τ-pHis isoform specificity in KCa3.1 regulation (His358) (61). Khan *et al.* used π-phosphohistidine specific mAbs to assess the significance of NDPK-A on tumour metastasis suppression (160). Hindupur *et al.* used pHis mAbs which enabled immunoprecipitation of pHis proteins allowing for investigation of LHPP as a broad acting τ-phosphohistidine phosphatase and tumour suppressor (92). Panda *et al.* used anti-π-phosphohistidine antibodies to verify wild-type PGAM5 dephosphorylation of NDPK-B (63).

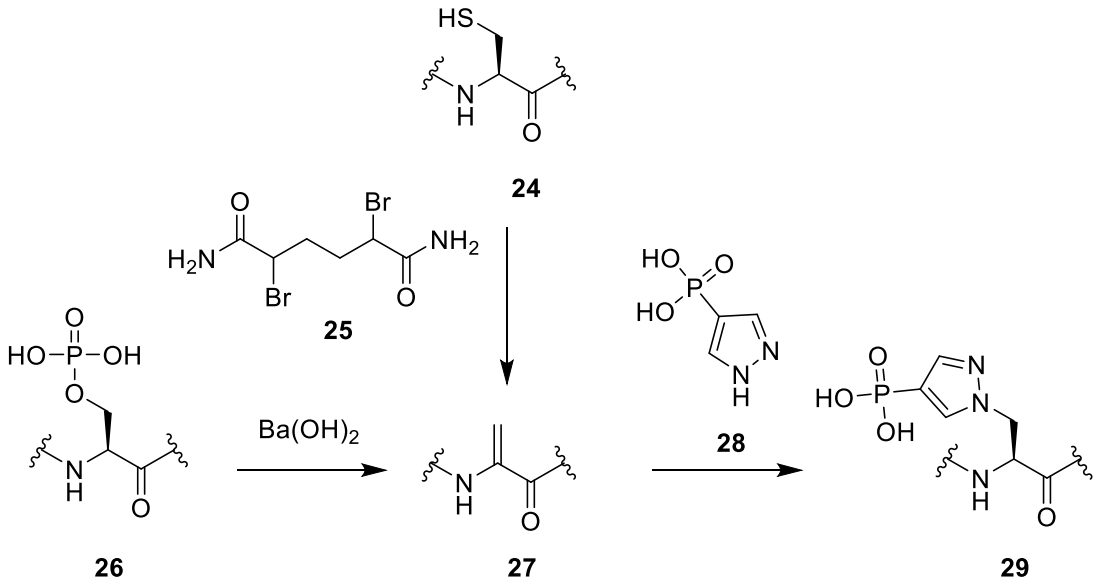
Whilst antibodies have provided a vast amount of information about pHis-containing proteins and their isoform specificity, there are may well be many unidentified systems and regulatory pathways. For example, the aforementioned phosphatase LHPP has been identified as a broad spectrum phosphatase with the use of antibodies (92), however no specific substrates have been identified for this phosphatase to date.

The stable analogues of phosphohistidine, produced to generate antibodies (triazole/pyrazole side chain), could themselves be ideal substitutes for the unstable phosphohistidine in a protein context. Indeed, Horner *et al.* originally proposed the incorporation of the homologue of pTza using homoazidoalanine and subsequent reaction with phosphonate alkynes to generate an effective analogue of phosphotyrosine (155). Incorporation of homoazidoalanine into proteins relies on methionine surrogacy, essentially if the aminoacyl-tRNA synthetase that would typically recognise methionine (methionine tRNA synthetase) is capable of accepting an unnatural amino acid instead, that amino acid will be incorporated into a protein sequence during translation (161). To produce the pTza histidine mimic in protein sequences by this method would require azidoalanine incorporation and the feasibility of this by methionine surrogacy is not well supported (162, 163). Therefore, production of protein probes containing the triazole analogue of pHis presented clear challenges. An alternative approach would be to

generate a pyrazole analogue of phosphohistidine and then identify conditions to incorporate the analogue by amber suppression for unnatural amino acid incorporation. This would require the complexity of an engineered tRNA for insertion of the pyrazole amino acid. Lilley *et al.* have a synthetic route for the pyrazolyl alanine (pPza) which could be incorporated into a protein by amber suppression if this were to be pursued (157).

Additionally, based on reduced cross-reactivity and the selectivity of the generated antibodies to the pyrazole-based analogues of phosphohistidine as shown through work by Kee (156) and Lilley (157), pyrazole is a more suitable analogue for phosphohistidine than triazole. Therefore, it seemed fitting to investigate routes for the insertion of pyrazole in peptide and proteins.

Given the phosphoryl pyrazole structure it was thought that incorporation of the phosphohistidine analogue into proteins in a reaction with dehydroalanine (Dha) was feasible based on literature precedent for heterocycle addition to Dha (164) (**Scheme 1**).

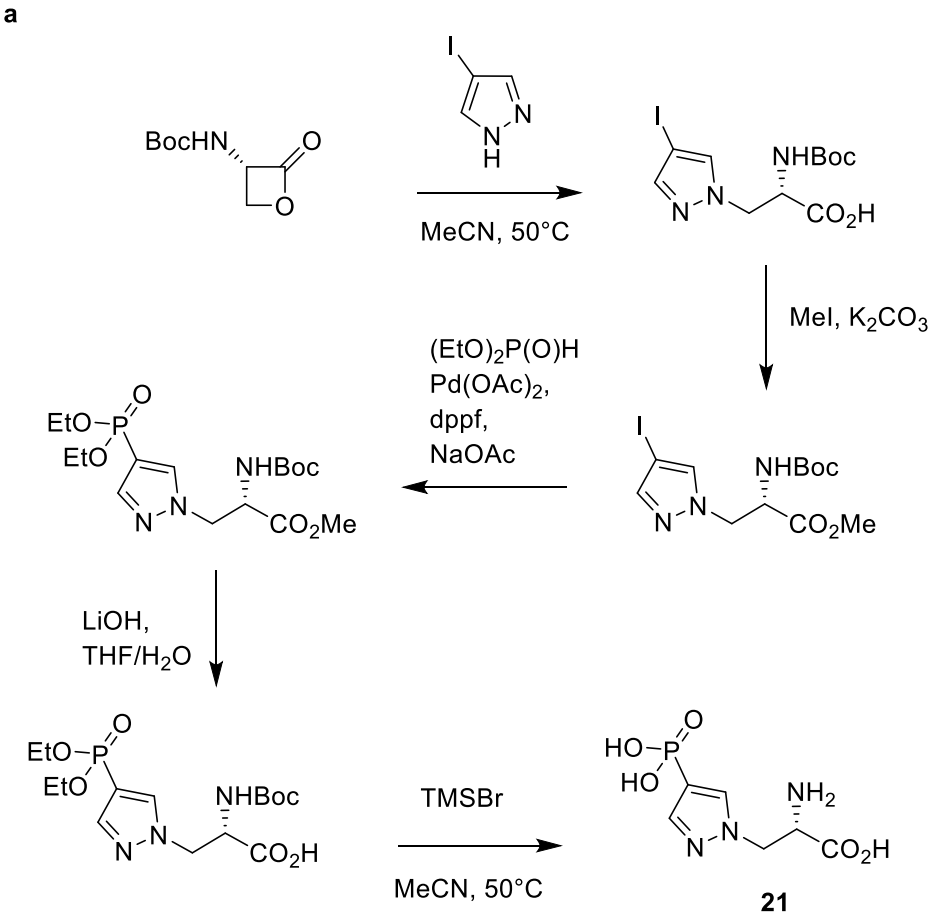


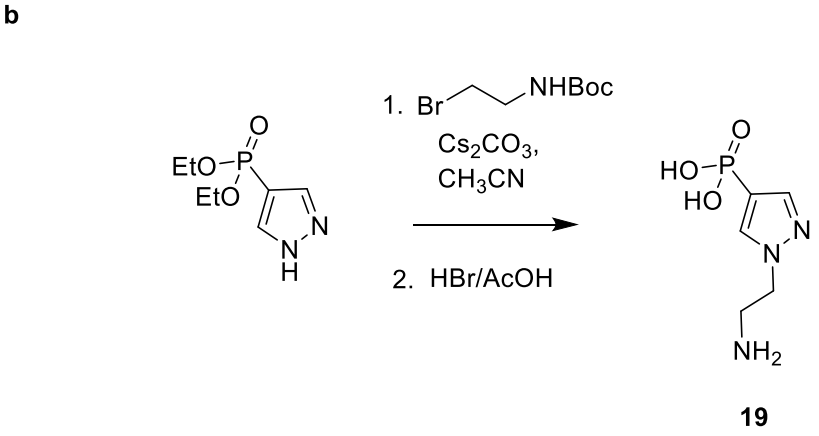
Scheme 1: General scheme for generation of Dha on a peptide substrate containing phosphoserine **26** or cysteine **24**, using barium hydroxide (165) or 2,5-dibromoadipamide **25** (166) respectively. Incubating the peptide containing Dha with phosphonate pyrazole **28** produces peptide probes **29**.

Davis *et al.* have previously shown that their strategy to modify dehydroalanine (Dha) residues can be adapted via addition of imidazole to generate isohistidine-proteins (164). It was hypothesized that reaction of a phosphonopyrazole would introduce a pyrazole analogue of phosphohistidine into a peptide sequence with a similar approach. For this methodology to work, dehydroalanine needed to be produced in the specific location of a histidine in a native sequence. Yang *et al.*, demonstrated a site-specific means of incorporating pSer and converting this to dehydroalanine (165). For model studies it is also possible to generate the

dehydroalanine through the presence of cysteine and subsequent conversion of cysteine to dehydroalanine. This conversion can be achieved with multiple different reagents such as 2,5-dibromovalerate, for sequences with multiple cysteine residues to avoid formation of stapled by-products (166). The more commonly used 2,5-dibromoadipamide **25** was selected given that work by Chalker *et al.* suggested that **25** shows the most scope for conversion of cysteine to Dha across multiple substrates (167).

Having outlined a general approach for producing pyrazole-containing peptides, a route to synthesise the pyrazole analogue of phosphohistidine as a free heterocycle such that it could be incubated with Dha and incorporated into a protein sequence was required. Prior synthesis, such as those by Kee (156) and Lilley (157), produced ethylamine **19** and amino acid **21** derivatives of the analogue which would be unhelpful for the intended nucleophilic incubation with Dha (**Scheme 2**).

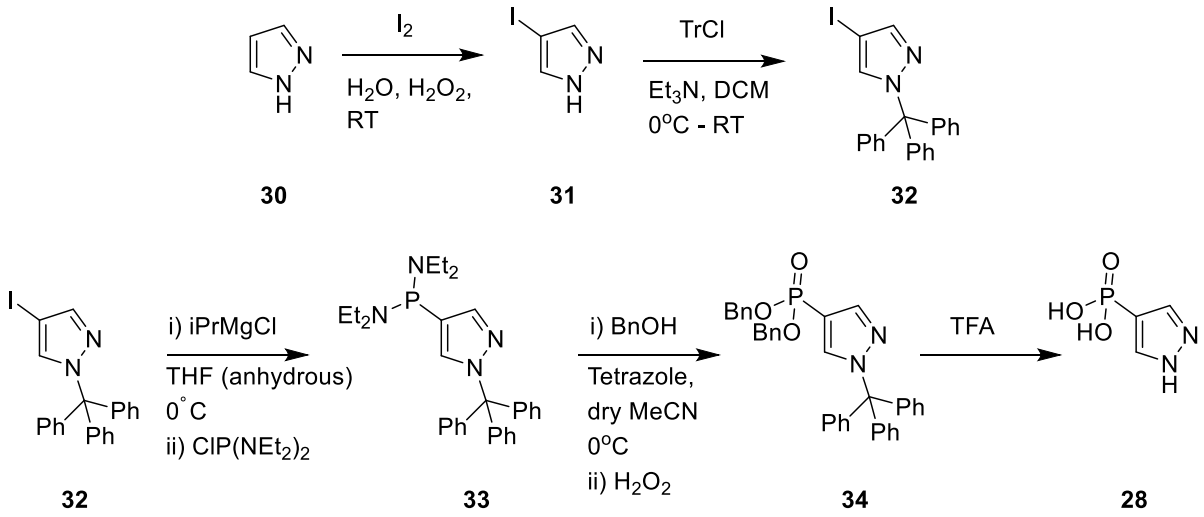




Scheme 2: Synthetic route reported by Lilley *et al.* for synthesis of pyrazole analogue of phosphohistidine (pPya) (**a**) (157) and route reported by Kee *et al.* for production of pyrazole ethylamine (pPye) (**b**) (156).

2.1 Strategy for production of phosphoryl pyrazole

The initial proposed approach (**Scheme 3**) was to react *bis*(diethylamino)chlorophosphine with a pyrazole Grignard reagent at the C-4 position, enabling subsequent benzyl alcohol substitution to produce a benzyl-protected precursor of phosphonate pyrazole **34**. This method was chosen because it mirrored similar methods employed by McAllister *et al.* (159) to successfully produce Fmoc-SPPS compatible pTza precursors for analogues of phosphohistidine. Halogen-magnesium exchange on protected iodopyrazole **32** and subsequent reaction with a *bis*(diethylamino)chlorophosphine would generate **33** and from this tetrazole-mediated benzyl alcohol substitution would be possible. Oxidation of P(III) to P(V) and subsequent deprotection of the benzyl groups would produce the desired phosphoryl pyrazole analogue of phosphohistidine **28**.

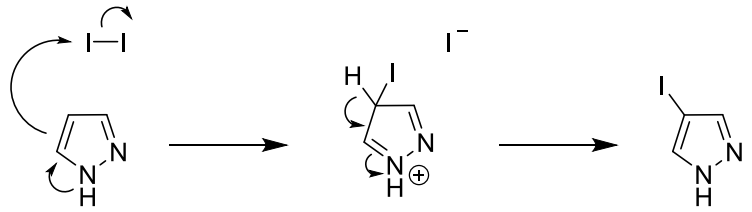


Scheme 3: Initial strategy for synthesis of phosphonate pyrazole **28** from pyrazole **30**.

2.1.1 4-iodopyrazole synthesis

Initially, 4-iodopyrazole was generated from pyrazole and iodine in water in the presence of 30 wt% H_2O_2 for 2 nights followed by quenching with sodium thiosulfate (**Scheme 5**). A cream/white solid was obtained. Recrystallisation, from H_2O, produced yellow-tinged white crystals of 4-iodopyrazole in an 19% yield. Following this exact procedure Lilley *et al*. (153) obtained a 54% yield. Subsequent repeats of this reaction gave inconsistent yields with the causes being removal of excess solid iodine or loss of product due to hot filtration steps.

This method of iodination uses I_2 as a weak electrophile enabling I^+ substitution onto pyrazole with catalytic H_2O_2 oxidising byproduct I^- to regenerate I_2 (**Scheme 4**). Any excess I_2 would have to be removed later during purification. Increasing I_2 equivalents to promote iodopyrazole formation would only exacerbate the issue of excess I_2 for removal.



Scheme 4: Mechanism for iodination of pyrazole

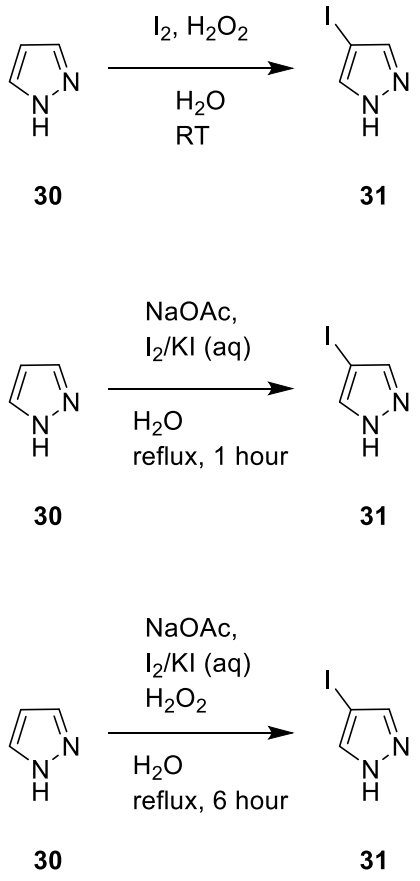
A method was sought that minimised I_2 impurity and improved yields by minimising filtration steps. In a new method, **Scheme 5**, pyrazole and sodium acetate were added to water and the mixture was heated under reflux followed by dropwise addition of KI/I_2 solution with stirring for an hour. The reaction was made basic with NaOH (0.1 M), to convert excess I_2 to sodium hypoiodite and sodium iodide facilitating easier removal of iodine impurities by aqueous extraction. Extensive aqueous back-extraction to remove excess sodium acetate salts yielded the target product in 11% overall yield.

The final method combined NaOAc and H_2O_2, **Scheme 5**. Recrystallisation from H_2O produced white crystals of 4-iodopyrazole in an 18% yield. A repeat resulted in the isolation of 4-iodopyrazole in 22% yield. On this repeat solvation of the white precipitate with each addition of NaOH was observed, leaving the unreacted iodine suspended. This provided an opportunity to purify the 4-iodopyrazole by intentionally dissolving it in excess water and separating the insoluble iodine.

In a final optimisation, excess NaOH was used in a more dilute reaction mixture to eliminate I_2 impurities (**Equation 1**). This conversion was leveraged to convert the solid iodine to soluble salts. Any remaining unreacted iodine would be filtered off prior to recrystallisation of the filtrate resulting in a 4-iodopyrazole yield of 30%.

$$I_2 + 2NaOH \rightarrow NaIO + NaI + H_2O$$

Equation 1: Reaction of iodine with sodium hydroxide produces soluble sodium hypoiodite, sodium iodide and water.



Scheme 5: Methods for synthesising 4-iodopyrazole **31** from pyrazole **30**.

The attempted methods and yields are summarised in the table below:

Method	Reagents	Yield (%)
1	I_2 (0.5 eq.) /H_2O_2 (0.9 eq.)	19
	I_2 (0.5 eq.)/ H_2O_2 (0.9 eq.) (excess hot filtration)	1
2	NaOAc (2 eq.)/KI (6 eq.)/ I_2 (2 eq.) (reflux for 1 hour)	11
3	NaOAc (2 eq.)/KI (2 eq.)/ I_2 (0.6 eq.) /H_2O_2 (0.5 eq.) (reflux for 6 hours)	18
	NaOAc (2 eq.)/NaI (2 eq.)/ I_2 (0.6 eq.) /H_2O_2 (0.5 eq.) (reflux for 6 hours) (solvent reduced w/ aim to improve recrystallisation)	22
	NaOAc (2 eq.)/NaI (2 eq.)/ I_2 (0.6 eq.) /H_2O_2 (0.5 eq.) (reflux for 6 hours) (Excess NaOH to quench I_2)	30

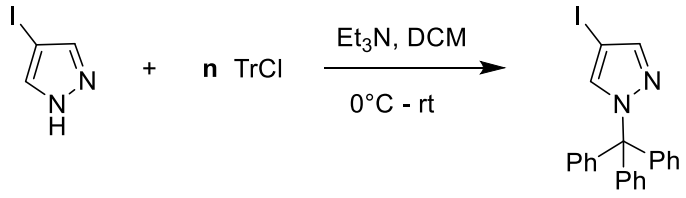
Table 4: Summary of methods used by reagents to produce 4-iodopyrazole with corresponding yield (%). Equivalents (eq.) given relative to pyrazole.

2.1.2 4-iodopyrazole protection reaction

The synthesised 4-iodopyrazole was reacted with trityl chloride to remove the acidic proton from the pyrazole for subsequent reaction.

Initially 4-iodo-1-trityl-1H-pyrazole was generated by reaction of 4-iodopyrazole **31** with an excess of trityl chloride in the presence of Et_3N. Excess trityl alcohol eluted with the product in column chromatography (hexane: ethyl acetate (9:1)) and the target compound was isolated as an off-white powder in a yield of only 17%. The TLC conditions were altered slightly in a repeat attempt using hexane: ethyl acetate (95:5), the product $R_f = 0.26$ and the trityl chloride impurity was present at $R_f = 0.13$, with no movement of 4-iodopyrazole from the baseline. The altered column chromatography solution yielded 26% of the trityl-protected 4-iodopyrazole.

The low mobility of 4-iodopyrazole in the TLC conditions, was exploited and the reaction was therefore repeated with stoichiometric trityl chloride (0.9 eq.) to minimise residual trityl alcohol during purification. TLC indicated complete conversion of trityl chloride in these conditions and the target compound was obtained in 32% yield following column chromatography with only 1% trityl alcohol impurity (determined by NMR).

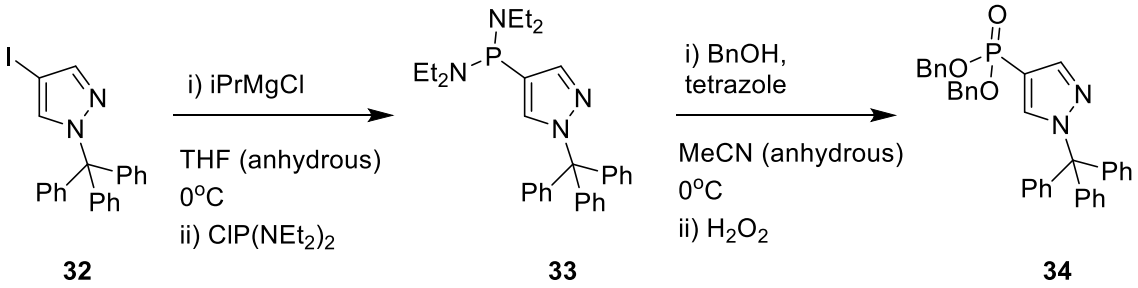


31 **32**

Scheme 6: Approach for synthesis of 4-iodo-1-trityl-1H-pyrazole **32** from 4-iodopyrazole **31**. In initial attempt **n** = 1.1, after reaction alteration equivalence changed to **n** = 0.9

2.1.3 Grignard-halogen exchange reaction

Using the 4-iodo-1H-trityl-pyrazole, a halogen-Grignard exchange was performed to enable reaction with *bis*(diethylamino)chlorophosphine to generate a phosphoamidite for subsequent alcohol substitutions (**Scheme 7**).



32 **33** **34**

Scheme 7: Synthesis of benzyl-protected phosphonate pyrazole **34** through formation of *bis*(diethylamino)phosphine pyrazole **33**.

LCMS showed the presence of intermediate **33** at $[M+H]^{1+}$ = 485 as a minor component, but no peak for product **34** was observed. Despite this, NMR analysis showed a pair of doublets at δ = 7.00 and 6.93 ppm, **Figure 32**. The integral and splitting corresponded with aromatic protons of the pyrazole ring coupling to phosphorus indicating some level of C-P bond formation.

The integral of the aromatic protons on *bis*(diethylamino)phosphoryl pyrazole **33** relative to other components was low. The triplet at 1.12 ppm that corresponded to the ethyl CH_3 showed comparable intensity to the aromatic intensity, if the reaction had been successful the aromatic region would have an intensity larger than that of the ethyl triplet. This indicated the presence of large quantities of unreacted *bis*(diethylamino)chlorophosphine raw material. The halogen-Grignard exchange was either unsuccessful or extremely low yielding. Given the expense of *bis*(diethylamino)chlorophosphine, an alternative route was sought for that would provide a desired phosphoryl pyrazole.

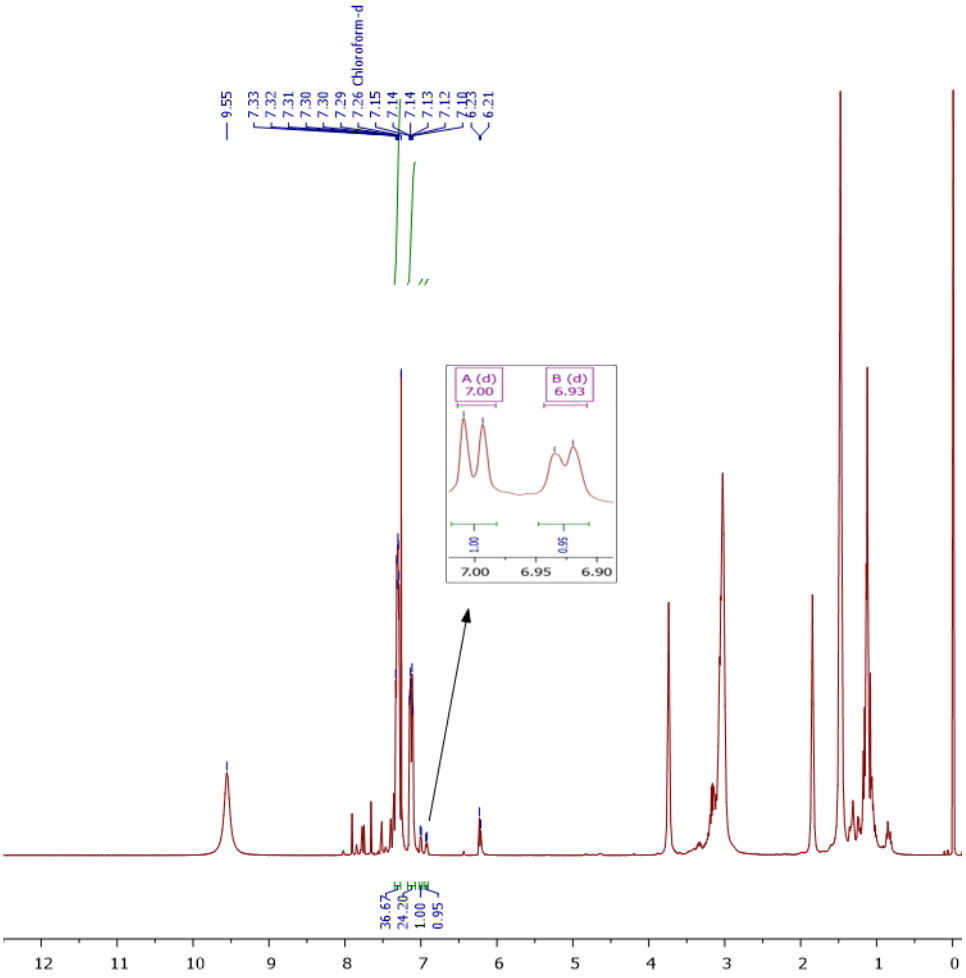


Figure 32: ^{1}H NMR spectrum representing the crude mixture of *bis*(diethylamino)phosphoryl pyrazole **33**. The pair of split doublets at 7.00 and 6.93 ppm may indicate the presence of phosphorylated pyrazole. These doublets are in low concentration in comparison to the trityl peak region at 7.33-7.10, indicating a low conversion.

2.2 Alternative pathway to phosphonate pyrazole

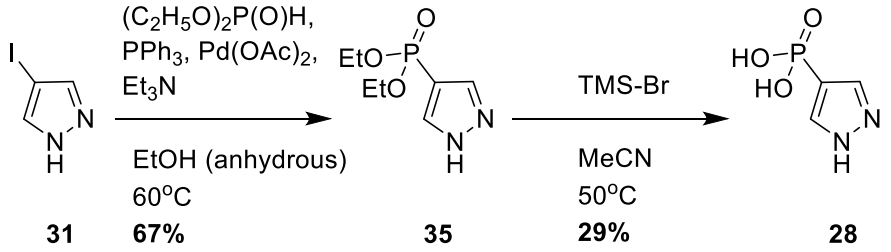
One possible route was a reported palladium-catalysed Hirao phosphorylation (158, 168). This was initially avoided due to the limited number of protecting groups available for the final phosphate product. The initial plan was to generate proteins and peptides containing the protected phosphonopyrazole prior to deprotection and it was also planned to incorporate this building block into an amino acid. Benzyl groups afford more opportunity for deprotection in a peptide context; they can be deprotected in standard Fmoc-SPPS cleavage conditions - TFA(95%):H_2O(2.5%):TIS(2.5%) - and may then facilitate generation of Fmoc-protected phosphonopyrazoles for applications in peptide synthesis.

In contrast, the ethyl-protected phosphonate pyrazole **35** would require both the standard cleavage cocktail and prolonged reaction with TMS-Br to afford the phosphonate pyrazole **28** on a peptide. Moreover, McAllister *et al.* had found ethyl-protected phosphonates troublesome in deprotection reactions, providing a mixture of mono-protected phosphonates and unprotected phosphonates even after substantial exposure to TMS-Br (159).

Multiple attempts were made to obtain the ethyl-protected phosphonate product of the Hirao reaction **35**. In the first instance, 4-iodopyrazole **31**, diethyl phosphite, triethylamine and triphenylphosphine were combined in anhydrous ethanol and heated to 60˚C with subsequent addition of catalytic amounts of palladium acetate. After heating at reflux with stirring for 12 hours, the solution was extracted in ethyl acetate (EtOAc) and purification by column chromatography in EtOAc was attempted. Isolated fractions contained only triphenylphosphine and diethyl phosphite. On a subsequent repeat the reaction was left to mix for 17 hours under reflux with minimal evidence of product formation and only triphenylphosphine, diethyl phosphite and a small quantity of 4-iodopyrazole being recovered. The third attempt at synthesis showed evidence of the product by TLC (EtOAc: R_f = 0.12). The success of the reaction was dependent on assiduous vacuum degassing and backfilling with nitrogen multiple times before heating at 60˚C for 22 hours. Isolation with column chromatography was achieved by sequential elution with ethyl acetate: hexane (1:9) to remove impurities and final elution with EtOAc to obtain the product as a brown solid in a 67% yield.

Deprotection of **35** in acetonitrile (MeCN) with TMS-Br at 50˚C overnight produced the phosphonate pyrazole **28** (**Scheme 8**). Purification was achieved via an initial trituration in water: isopropanol, however only small quantities of product were recovered. The trituration

in water: isopropanol gave a white precipitate product which was insoluble in MeCN. The insolubility of the phosphonate pyrazole in MeCN was leveraged isolate the precipitate through recrystallisation, giving a 29% yield. A comparison of the starting material **35** and product **28** in D_2O shows loss of ethyl groups by ^1H NMR (**Figure 33**) and a shift of the phosphorus signal in ^{31}P NMR (**Figure 34**).



Scheme 8: Revised approach for synthesis of phosphonate pyrazole **28** from 4-iodopyrazole **31**. Palladium catalysed Hirao reaction affords a one-step route to obtain ethyl-protected phosphonate pyrazole **35** and subsequent reaction with TMS-Br affords the desired product.

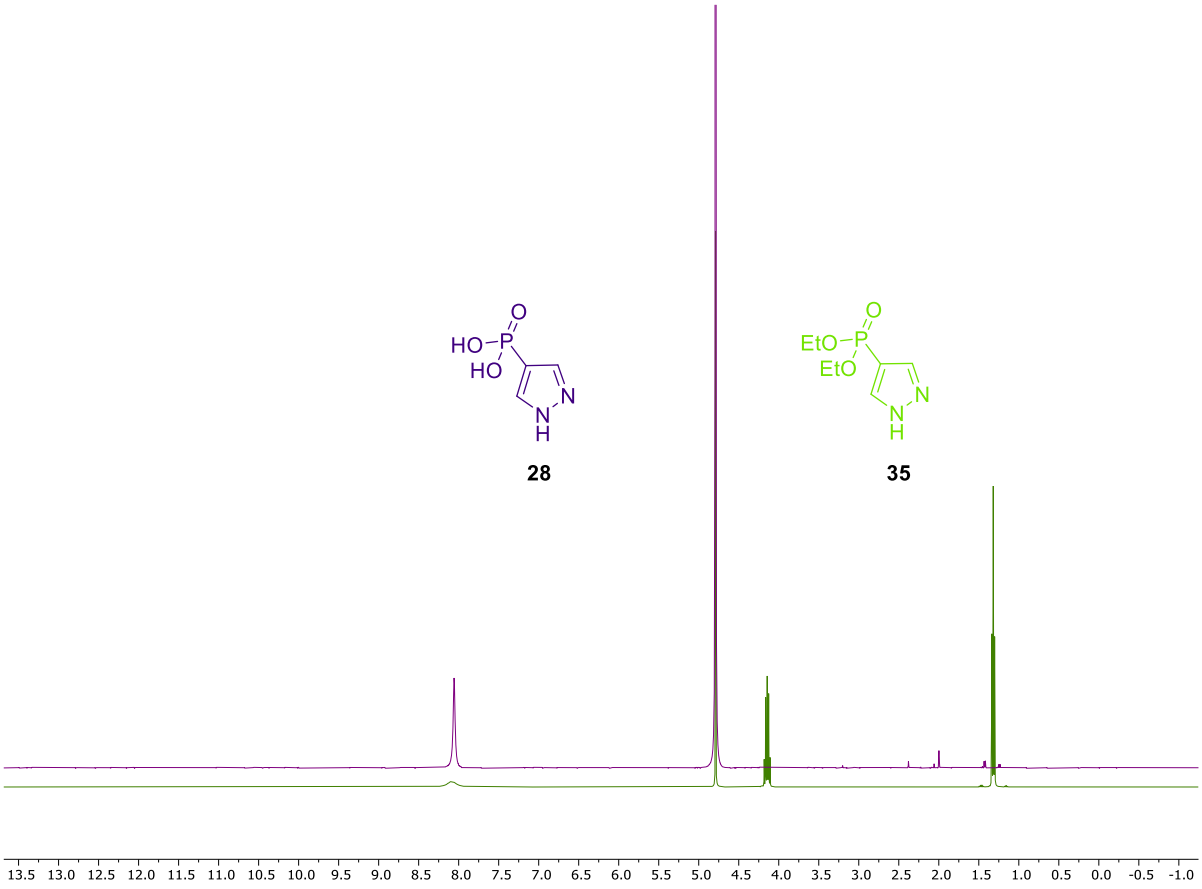


Figure 33: ^{1}H NMR in D_2O, product phosphonate pyrazole **28** spectrum shows no ethyl peaks at 4.18-4.11 ppm (CH_2) and 1.32 ppm (CH_3). Pyrazole protons appear at 8.06 ppm and 8.08 ppm for phosphoryl pyrazole **28** and ethyl-protected phosphonate pyrazole **35** respectively.

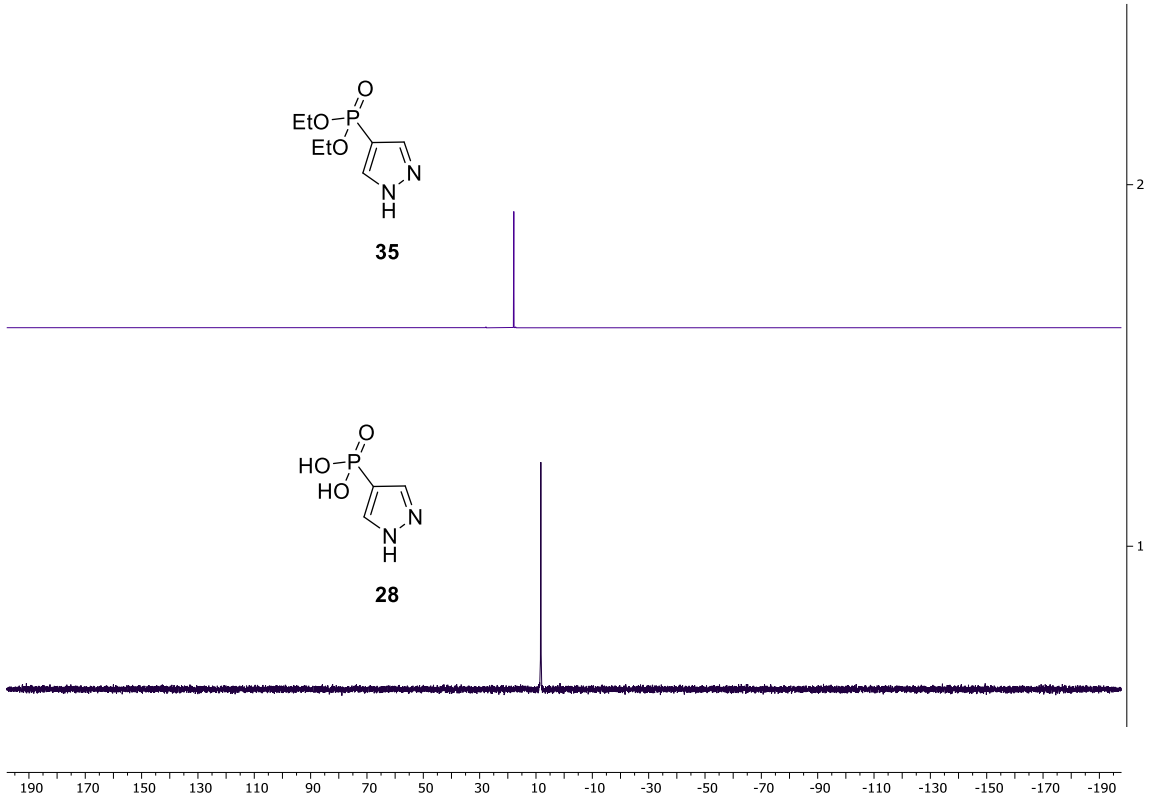


Figure 34: ^{31}P NMR in D_2O, comparison of phosphorus shift for ethyl-protected pyrazole **35** (17.87 ppm) and deprotected phosphoryl pyrazole **28** (8.37 ppm).

2.3 Generation of Dha-containing peptides

2.3.1 Generation from phosphoserine (pSer)

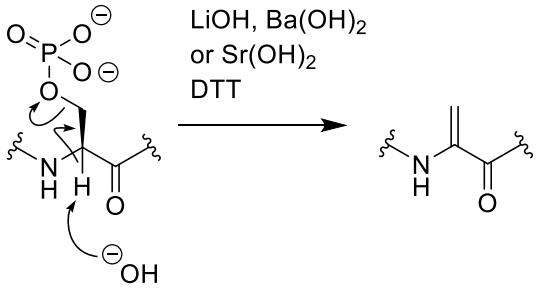
Having synthesised a phosphonate pyrazole analogue of phosphohistidine, generation of dehydroalanine (Dha) on a model peptide substrate and protein was required to test the potential for the analogue to perform nucleophilic addition in these contexts leading to production of an analogue-containing peptide and protein.

The initial plan was to form Dha from either phosphoserine (pSer) or cysteine in an expressed protein. For more complicated proteins (potentially containing multiple cysteine residues), chemoselective formation of Dha from a specific cysteine residue using DBA would not be possible in the presence of multiple cysteine residues. To affect a specific substitution, all other cysteine residues would have to be mutated out of the expressed protein, and only then could Dha be generated followed by the desired substitution. Potentially forming Dha from pSer circumvents this challenge but instead requires site-specific incorporation of pSer at the histidine site.

There is precedent for this synthetic pathway, work by Rogerson *et al.* has shown it is possible to incorporate phosphoserine and its non-hydrolysable analogue into a protein using an orthogonal synthetase/tRNA pair that complement a stop codon (169). This operates through use of an O-phosphoseryl-tRNA synthetase (SepRS) associated with the formation of a cysteine tRNA. In methanogenic archaea such a pSer-$tRNA^{Cys}_{GCA}$ is converted to Cys $tRNA^{Cys}_{GCA}$ by an enzyme Sep-tRNA-Cys-tRNA synthase (SepCysS) (170). In other organisms, the SepCysS enzyme is not present and therefore pSer-tRNA will not be converted to Cys-tRNA enabling incorporation of pSer (170). The specific aminoacylation of the $tRNA^{Cys}_{GCA}$ with pSer by SepRS was adapted by alteration of the tRNA sequence anticodon (GCA transplanted for CUA) to direct pSer insertion (169) in place of the Amber stop codon. Using this system, substitution of requisite histidine codons with the stop codon would enable site-specific incorporation of pSer at the desired sites which following conversion to Dha would enable insertion of the stable analogue.

The second step of this potential scheme has been exemplified by Yang *et al.* who demonstrated pSer conversion to Dha *in vitro* with a modified histone H3 and an alkaline source (LiOH, $Ba(OH)_2$, $Sr(OH)_2$) to assist with phosphate lability (165). Under the alkaline conditions, the

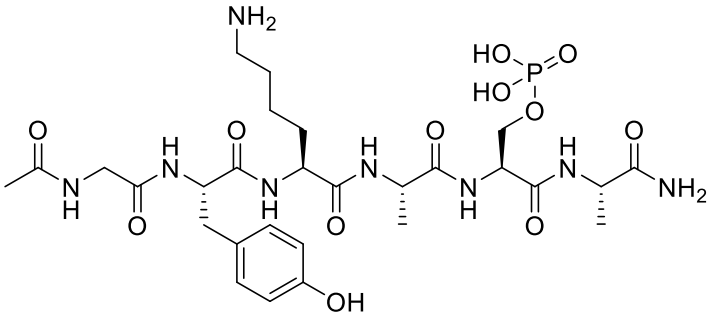
α-carbon proton is extracted leading to double bond formation and elimination of the phosphate group from the β-carbon (**Scheme 9**).



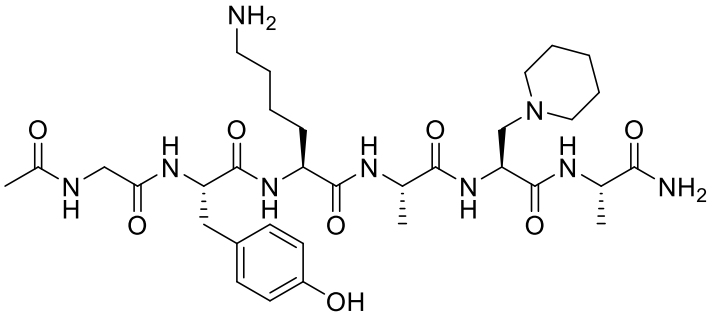
Scheme 9: Conversion of pSer-containing protein to Dha-containing protein (165).

An initial simple peptide sequence **36** was used as a test pSer peptide sequence for synthesis comprised of mainly alanine residues with a lysine to assist with solubility and a tyrosine residue to aid in HPLC purification.

Initial synthesis of this pSer-peptide was attempted with a peptide synthesiser using DIC and Oxyma Pure as coupling reagents. This presented clear challenges, while LCMS analysis indicated the presence of the desired pSer-peptide product **36**, this was accompanied by a major contaminant in the form of the piperidinyl alanine peptide **37**, formed by microwave-promoted substitution of piperidine for phosphate during peptide synthesis (171).



36



37

Figure 35: Structure of test substrate for converting pSer to Dha on a peptide, Ac-GYKApSA-NH_2 **36** and β-elimination by-product **37**.

Piperidine catalyses Fmoc-removal during Fmoc-SPPS via dual-action as both a base and nucleophile, deprotonating the fluorenyl ring of the Fmoc group leads to subsequent deprotection of the N-terminal amine and sequesters the dibenzofulvene byproduct. Piperidine's ability to abstract weakly acidic protons is problematic during synthesis of phosphopeptides, since piperdine can remove a proton from the α-carbon of the pSer component of the peptide leading to a Dha residue; and, in an analogous manner to the scavenging of dibenzofulvene, then act as a nucleophile to Dha to generate piperidinylalanine (**Figure 36**).

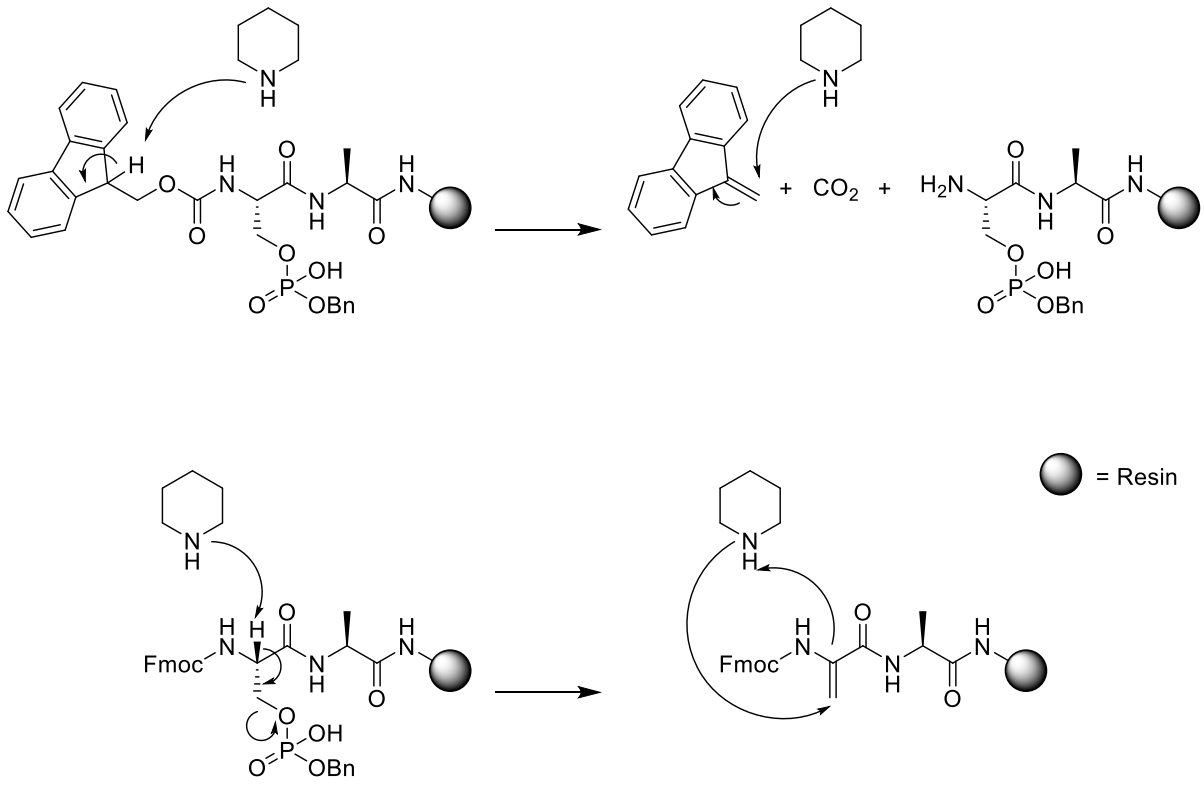
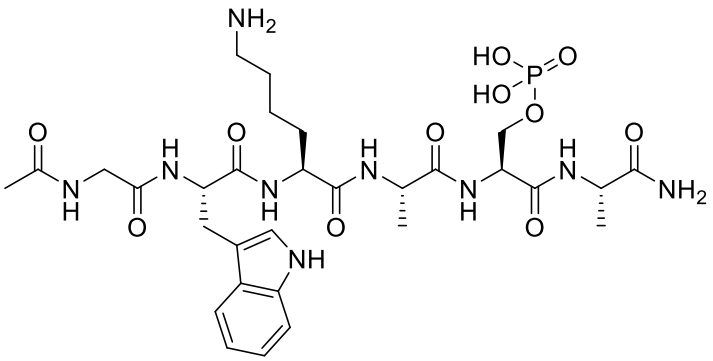


Figure 36: Mechanism for deprotection of Fmoc by piperidine in Fmoc-SPPS and removal of β-elimination on phosphoserine to generate dehydroalanine followed by addition of piperidine. (172-174)

Use of the monobenzyl-protected pSer building block (Fmoc-Ser(PO(OBzl)OH)-OH) is reported to reduce the effect of this β-elimination mechanism (175), however it appeared that caution was still advised where microwave-mediated coupling was involved as this promoted the unwanted side reaction (176, 177); which may explain the observed substitution in the microwave-assisted peptide synthesis with this building block. The monobenzyl building block also presented additional challenges, while the free hydroxyl group on the phosphonate could be deprotonated by piperidine which reduced the favourability of the β-elimination by increasing the charge density on the phosphate (173), its deprotonation leads to formation of a piperidinium salt that is not removed during washing leading to reaction of piperidine with the next activated amino acid (178). A subsequent attempt to generate a similar peptide substrate **38** was carried out using a manual synthesis protocol to avoid the microwave-promoted β-elimination byproduct; HCTU and DIPEA were also used to effect the pSer coupling in line with a literature study showing more complete conversion in non-DIC mediated coupling (178), all other non-phosphoamino acid couplings in the peptide sequence were performed using DIC and Oxyma Pure. In anticipation of challenges producing a pure peptide, due to potential side reaction, tyrosine was replaced with tryptophan to enable easier identification of the product on UV-HPLC due to the higher extinction coefficient of tryptophan.

In this repeat, evidence of piperidinyl-alanine formation was not apparent by HRMS analysis which showed a characteristic $[M-H]^-$ = 738.297 peak corresponding to the desired peptide. However, HPLC analysis indicated only 37% purity with the presence of impurities including a major impurity with a mass of 610.201 corresponding to the peptide sequence **38** without a lysine residue. The formation of a piperidinium salt may have explained a missing alanine residue adjacent to the pSer residue, but the presence of a significant portion of the peptide missing a lysine residue is harder to rationalise but might be due to co-ordination of the alanine N-terminus with a negatively charged phosphate group at this step of the reaction.



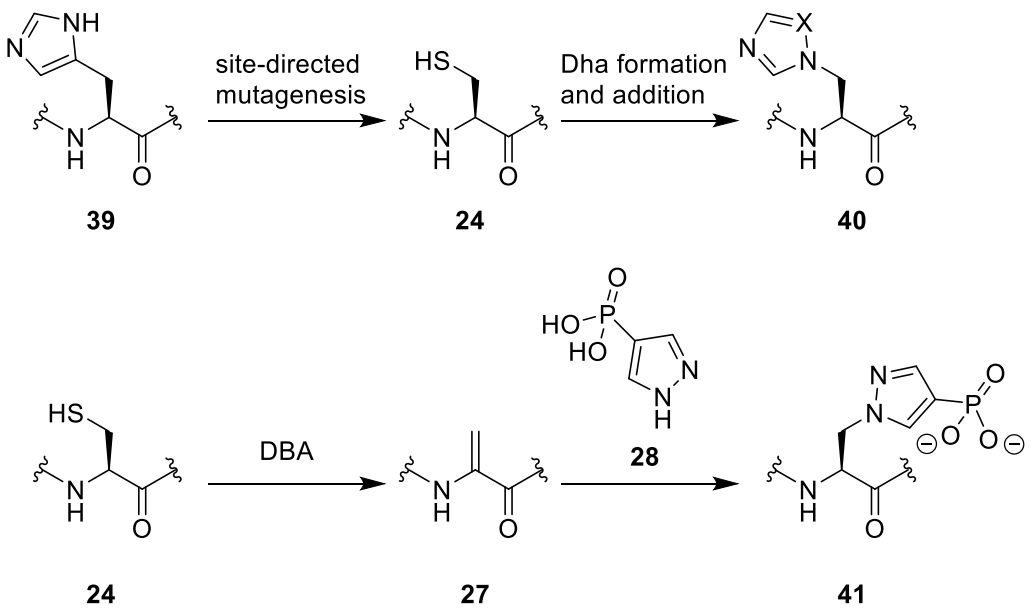
38

Figure 37: Structure of Ac-GWKApSA-NH_2 **38** the second substrate synthesis attempt for producing Dha from pSer on a peptide substrate.

Synthesis of pSer peptides provided significant challenges and whilst the potential of the pSer-Dha pathway as a means to produce probes containing the phosphonate pyrazole **28** was very interesting, parallel work on the production of Dha on cysteine in peptides proved more successful so cysteine-based peptides were prioritised in lieu of working with pSer.

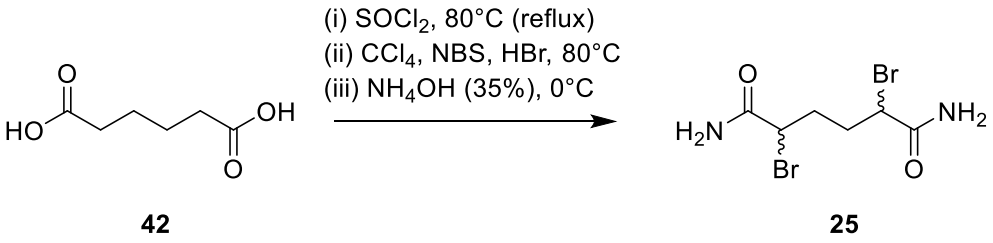
2.3.2 Synthesis of cysteine-containing peptides

The strategy for incorporating the phosphonate pyrazole moiety into a cysteine-containing protein was based on a method by Davis *et al.* for functionalising proteins through site-directed mutagenesis to exchange histidine **39** with cysteine **24** followed by generation of a dehydroalanine residue and subsequent addition of imidazole and [1,2,4]-triazole to form iso-histidine and azaiso-histidine **40** respectively (164). Reproduction of this method on a degenerate peptide sequence containing cysteine by forming Dha and incubating with the phosphonate pyrazole analogue **28** was theorised to produce a peptide containing a phosphohistidine analogue **41**.



Scheme 10: Formation of protein containing isohistidine **40** from mutated protein containing cysteine **24** – method by Davis *et al.* (164). The approach in this work emulated the method by Davis *et al.* but aimed to substitute phosphonate pyrazole **28** into the protein to form probes **41**, 2,5-dibromoadipamide (DBA) was the reagent of choice for conversion of cysteine to Dha in keeping with work by Morrison *et al.* (166).

Generation of Dha on a protein sequence containing cysteine requires dibromoadipamide **25** (DBA) which was synthesised by the method shown in **Scheme 11** as previously reported by Morrison *et al.* (166). Briefly, the adipic acid was converted to a *bis*-acid chloride intermediate, which was subsequently α-brominated before quenching of the intermediate with ammonium hydroxide to form the target diamide. ^{1}H NMR analysis indicated the presence of dibromoadipamide diastereomers, in keeping with the literature the grey solid was used in subsequent attempts to generate Dha (166).



Scheme 11: Reaction procedure for the synthesis of 2,5 dibromoadipamide **25** (DBA) from adipic acid **42**.

Two test peptide sequences containing cysteine residues, which could be incubated with 2,5 – dibromoadipamide **25** to test formation of Dha were generated via solid phase peptide synthesis. The first peptide sequence was Ac-GYKACA-NH_2 **43**. Following confirmation by presence of a single species $[M+H]^{1+}$ = 653 by LCMS analysis, a second peptide **44**, based on the sequence flanking His18 in histone H4 (**Figure 38**) was generated to test Dha formation in a more complex context, the sequence was similar to a substrate used by Kee *et al.* to produce pTza histone H4 probes (147).

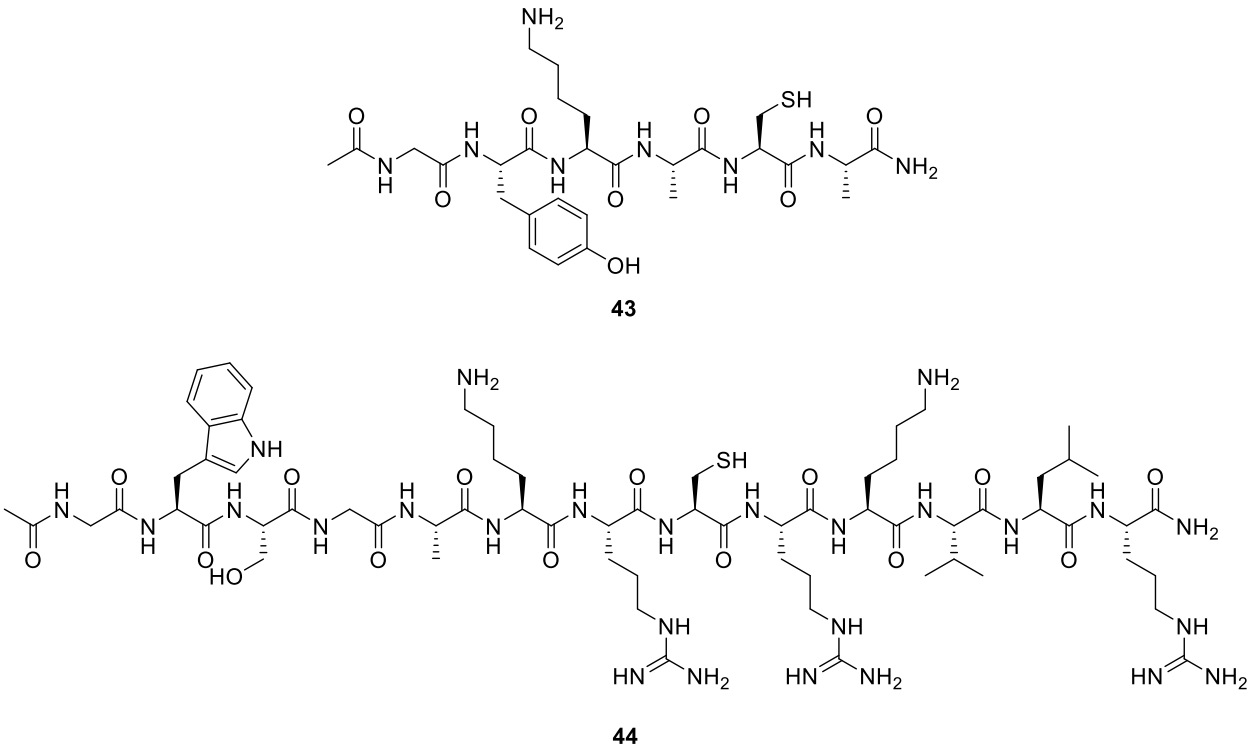
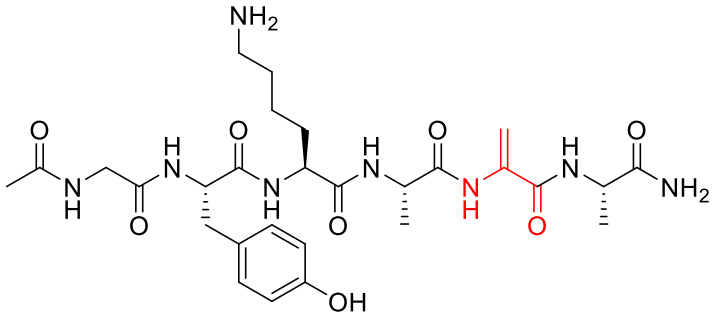


Figure 38: Structure of test peptides, Ac-GYKACA-NH_2 **43** and a peptide segment of the histone H4 sequence **44**, with additional N-terminal acetyl-glycine and tryptophan, serine replaces the position of a cysteine (to prevent cyclisation product formation and chemoselectivity issues with DBA) and cysteine replaces histidine. Based on sequences used by Kee et al (147).

Having successfully produced a test peptide sequence **43** and a peptide sequence analogous with a sequence of histone H4 **44**, Dha formation was attempted.

Ac-GYKACA-NH_2 **43** was incubated with TCEP in water, to reduce cystines; and the reduced peptide was subsequently mixed with DBA **25** in DMF and K_2CO_3 was added. The reaction was heated to 37˚C with shaking at 200 rpm and conversion measured at hourly intervals by LCMS, after 2 hours LCMS confirmed presence of $[M+H]^{1+}$ = 619 corresponding to the Dha-containing peptide displayed in **Figure 39**. The quantities used at this stage to achieve the conversion of cysteine to dehydroalanine were relatively large, however subsequent attempts to generate Dha on a smaller scale were unsuccessful, and the product $[M+H]^{+}$ = 619 peak was not observed.



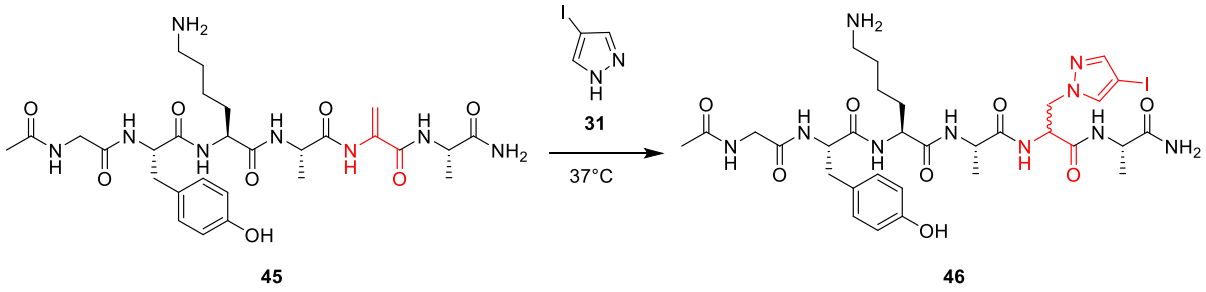
45

Figure 39: Ac-GYKACA-NH_2 **43** with DBA **25** on a 37˚C incubator at 200 rpm had shown complete conversion to Ac-GYKADhaA-NH_2 **45** as indicated by the presence of the product $[M+H]^{1+}$ = 619 peak.

Attempts were also made to generate Dha on the histone H4 analogue **44**, but conversion to the Dha-containing species was not apparent, after 3 hours at 37˚C, 200 rpm, the cysteine containing peptide was the predominant species as indicated by $[M+4H]^{4+}$ = 390 and $[M+3H]^{3+}$ = 520. Extracted ion chromatograms did indicated some level of conversion to the Dha-containing species, but its presence could not be substantiated by HRMS analysis.

2.4 Addition of heterocycles to Dha-containing peptides

Following the successful production of Ac-GYKADhaA-NH_2 **45**, addition of 4-iodopyrazole **31** to this Dha-containing peptide was conducted by incubation of equimolar amounts of 4-iodopyrazole with peptide **45** in sodium phosphate (50 mM, pH 8) at 37˚C, as shown in **Scheme 12**.



Scheme 12: Nucleophilic addition of 4-iodopyrazole **31** onto Dha in Ac-GYKADhaA-NH_2 **45**.

Unfortunately, this substitution did not succeed, initially 4-iodopyrazole (75 mM) was incubated with the peptide solution (75 mM) but after 4 hours, LCMS analysis indicated only the $[M+H]^{1+}$ = 619 peak, 4-iodopyrazole had not reacted. In an attempt to force the reaction an additional quantity of 4-iodopyrazole (1.05M), was added and the incubation continued for 1 hour, but the target mass could not be detected.

As stated above, production of the Dha-containing peptide was inconsistent. To simplify the analysis by ESMS (and to reduce potential variation in ionisation as a result of the chemical changes) a model protein was instead used as a substrate for Dha formation and a model protein (an affimer) containing a single cysteine residue was available in the laboratory for this purpose (provided by T. Kaminski). DBA and K_2CO_3 were prepared as 4.06 mM solutions. They were mixed with the affimer (40.6 uM, 150 uL, 1 eq.) and DBA (150 uL, 100 eq.) and K_2CO_3 (15 uL, 10 eq.) and incubated at 37˚C. After an hour, the affimer solution was analysed by HRMS to check for conversion of a cysteine residue to Dha. The presence of a 17181 Da peak indicated complete conversion to the Dha-containing affimer. The Dha-containing affimer was buffer exchanged into sodium phosphate (50 mM, pH 8) and subsequently spin-concentrated followed by incubation with heterocyclic compounds to test for substitution.

Five independent reactions (200 µL) consisted of imidazole, pyrazole, diethyl phosphonate pyrazole, ethyl-1H-pyrazole-4-carboxylate and 1H-pyrazole-4-carboxylic-acid (500mM with the Dha-containing affimer (50 µM)) were incubated at 25˚C for 3 hours.

Imidazole and pyrazole both added to the Dha-containing affimer in alignment with work by Davis *et al.* (164). The substituted pyrazoles were used as benchmarks to assess the likelihood of addition to the Dha-affimer by phosphoryl pyrazoles. In all cases, addition was not observed by the substituted pyrazoles. Overall conversion of the dehydroalanine to substituted molecules on this timescale was poor (**Figure 40**). Imidazole added onto the Dha affimer to generate ~30% product on a 3-hour timescale, but pyrazole added much more slowly (<10% conversion). While it was hoped that the electron withdrawing groups in the substituted pyrazoles would promote addition onto the Dha-containing affimer, diethyl phosphoryl pyrazole showed no addition and neither did the other pyrazoles tested as additional substrates: ethyl-1H-pyrazole-4-carboxylate and 1H-pyrazole-4-carboxylic-acid. Given the predicted extended times required for reaction, this suggested that the proposed approach to generation of phosphohistidine analogues would not be feasible for implementation to incorporate pHis analogues as planned.

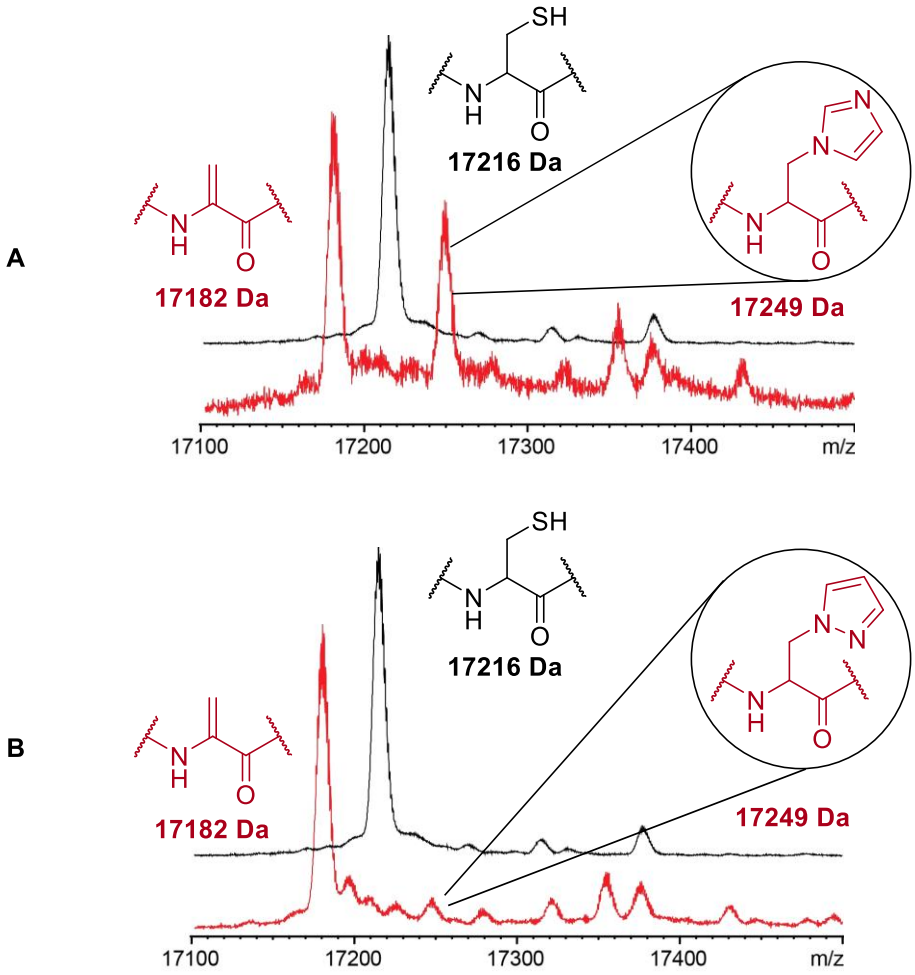


Figure 40: Anti-CTB affimer 3A2 with mass 17216 Da is modified by reaction with DBA to generate modified anti-CTB affimer 3A2 with Dha replacing a cysteine residue with mass 17182 Da. Subsequent

incubation of the Dha-containing affimer (buffer exchanged into sodium phosphate 50 mM, pH 8) with imidazole (**A**) and pyrazole (**B**) generate substituted affimer species with 17249 Da.

2.5 Conclusion and Future Outlook

Pyrazole was a more suitable mimic for phosphohistidine than triazole based on DFT calculation and antibody cross-reactivity. As such, investigation into how pyrazole may be incorporated into peptide and protein structures was a worthy pursuit to produce pHis-mimic proteins and peptides. Synthesis of phosphoryl pyrazole proved challenging, the main goal lay in synthesising the analogue in a way such that incorporation into a peptide or protein would be straightforward, particularly in regard to removing any protecting groups on the phosphonate. This led to a focus on benzyl-protected phosphonate routes, but given the challenges faced here, the Hirao phosphorylation was resorted to, even if this method did leave the challenge of removing ethyl protecting groups from the phosphonyl pyrazole with TMS-Br.

Despite challenges with the synthesis of phosphoryl pyrazole, the synthesis was eventually achieved. To meet the aims of this work, the phosphoryl pyrazole needed to be incorporated into a protein/peptide through incubation with a Dha residue in accordance with methodology by Davis *et al.* (164). Unfortunately, despite successfully generating a Dha-containing peptide (modified CTB affimer 3A2), addition to this peptide was only significantly observed with imidazole to generate an isohistidine affimer and to a lesser extent with pyrazole. The substituted pyrazoles showed no evidence of addition to the modified affimer.

While incorporation of phosphoryl pyrazole into protein structures seemed potentially feasible. The slow rate of addition of these pyrazoles to Dha-containing peptides led to the termination of this part of the project at this point. Without the ability to incorporate the pyrazole analogue of pHis into a structure the hope of generating these proteins was eliminated. This investigation into pyrazole acts as a cautionary tale against attempting to incorporate substituted small molecule analogues of phosphohistidine on Dha sites as a future means of making chemical probes.

Although small molecule incorporation via incubation with Dha was not successful in this work, it is possible that through testing of a wider range of temperature conditions, pH optima and use of different buffer systems and bases, the rate of pyrazole addition to the Dha affimer could be expedited.

Another path that would be interesting to pursue in the future would be the production of reactive phosphonate pyrazolyl alanine-containing peptide probes. If the pyrazole analogue

were incorporated into peptides with reactive handles, it may serve as a useful tool to investigate cell extracts.

3 Generation of reactive pTza analogues of phosphohistidine

The widespread identification of phosphohistidine sites in proteins in the last 15 years was contingent on synthesis of stable analogues of histidine which mimicked the phosphate post-translational modification (147, 159). To a large extent, this was since these stable analogues enabled the production of selective antibodies for phosphohistidine-containing peptides and proteins (25, 156). The aim of this part of the project was to create reactive probes that could be used to identify phosphohistidine (pHis) binding partners. The initial objective was to adapt previously produced stable analogues of pHis, such as phosphonate pyrazole and triazole, to generate reactive analogues. While the pyrazole motif appeared to more accurately mimic the structure of the histidine imidazole than the triazole heterocycle (based on antibody cross-reactivity) the synthesis of this motif provided a significant challenge. Therefore, whilst exploring the potential for production of pyrazole-based probes, described in Chapter 2, a means of producing a reactive triazole pHis mimic (pTza) was sought.

The basis for this analogue design was the presence of a catalytic histidine residue in the active site of the known phosphohistidine phosphatases: PHPT1 (H53) and PGAM-5 (H105). In the characterised phosphohistidine phosphatases evidence showed that phosphate is transferred to histidine residues. Theoretically, if an analogue were to be designed that interacted with the nucleophilic histidine on a phosphohistidine phosphatase, and the analogue contained a reactive electrophilic group, that analogue could covalently bind the phosphatase which may enable downstream identification through a proteomics workflow. Therefore, a means of generating electrophilic binding capability was sought for a triazole-based phosphohistidine analogue.

Inspiration for creating reactivity on phosphonate emerged from work by Cravatt *et al.* (179) who developed electrophilic fluorophosphonate probes for capturing serine hydrolases (**Figure 41**).

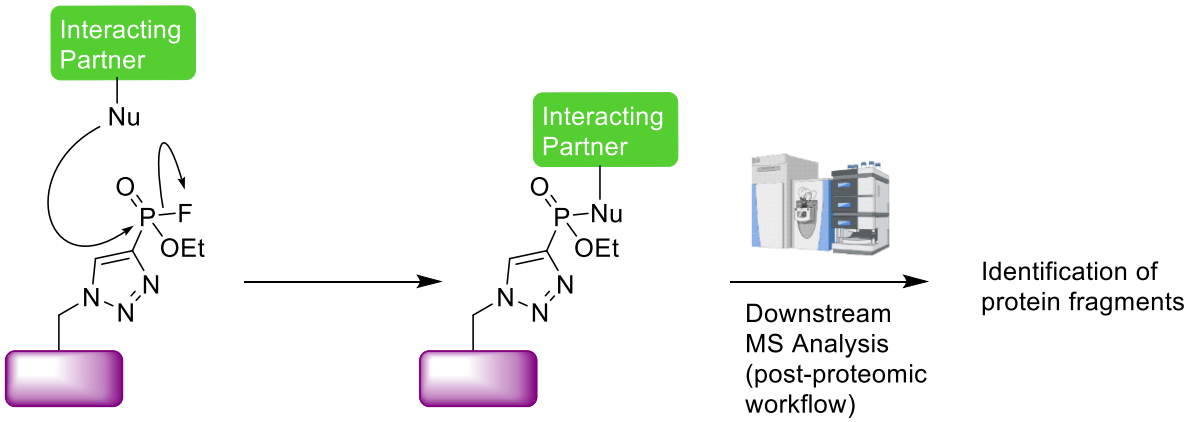


Figure 41: Example of a pTza-containing reactive probe peptide, with fluorophosphonate functionality, the electrophilic group can be used to selectively bind nucleophilic (Nu) residues on binding partners that interact with the specific peptide substrate. Subsequent MS fragment analysis could lead to identification of binding partners.

One of the initial proposals was the production of reactive pTza probes. It was recognised that to generate the fluorophosphonate functionality in the Cravatt-style fluorophosphonate probe (179), an orthogonally protected alkyne phosphonate precursor would be needed. The existing literature covers methods for producing disubstituted alkyne phosphonates (**Figure 42**), phosphonates with mixed substitution were targeted to facilitate formation of the fluorophosphonate group.

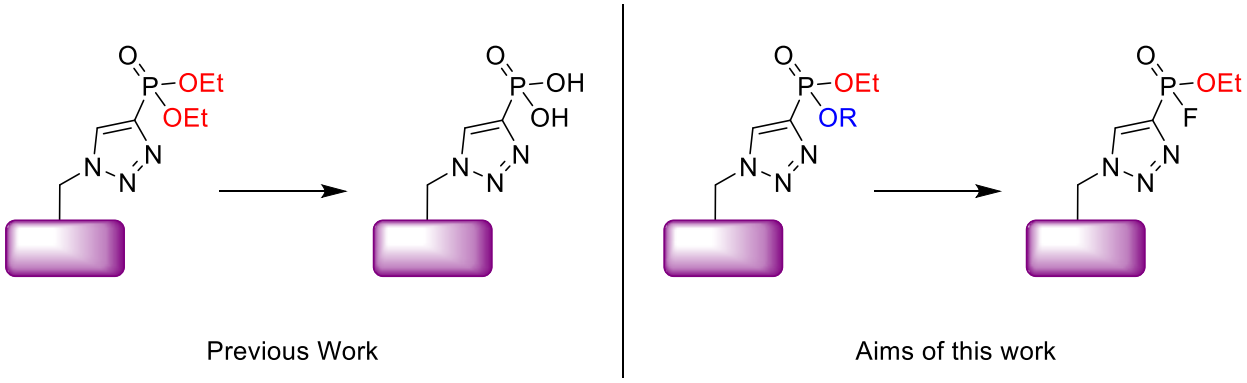


Figure 42: Outline of the aims of this work against previous work, adapting pre-existing analogues. Diethyl-protected pTza incorporated into a peptide is used as an example of a diethyl substituted pTza used in prior work by Kee *et al.* (147) In that work, the diethyl pTza was a precursor to the 3-pTza analogue of pHis and subsequently used for antibody generation in a histone H4 sequence (147). This work aimed to produce mixed phosphonates as precursors to fluorophosphonate functionality which could be used to bind nucleophilic interactors in the same fashion as Cravatt (179).

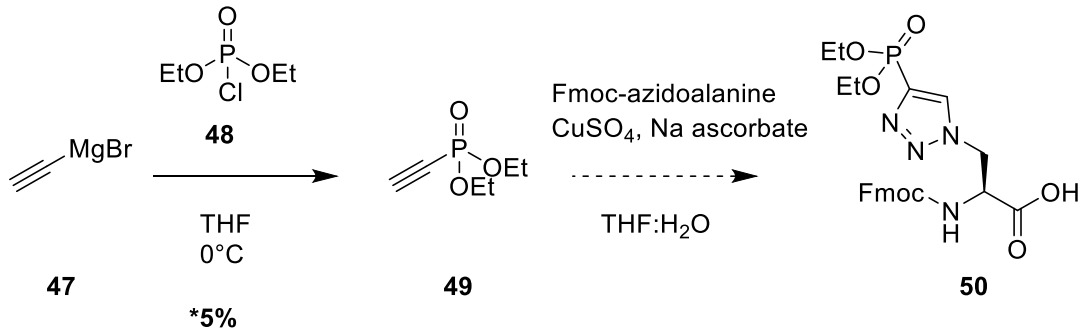
3.1 Generation of novel mixed phosphonate alkyne as a pTza precursor

Previously, routes have been reported both for the synthesis of diethyl-protected pTza (**Scheme 13**) and dibenzyl-protected pTza (**Scheme 15**).

Diethyl ethynyl phosphonate **49** can be produced by via a Grignard reaction of P(V) diethyl chlorophosphate **48** with ethylene magnesium bromide **47**. Subsequent copper-click cycloaddition generates an ethyl-protected pTza analogue of phosphohistidine **50**. The reliance on existing chlorophosphates, such as diethyl chlorophosphate **48** limits the scope of this reaction. Additionally, McAllister *et al.* found that ethyl protection on the analogue **50** greatly limited the scope for incorporation into a peptide given that ethyl protection is not compatible with typical cleavage conditions and would require additional orthogonal deprotection with TMS-Br (159).

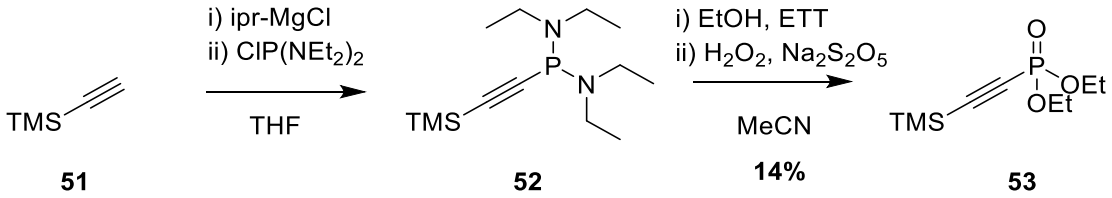
Attempts to replicate the synthesis of diethyl ethynyl phosphonate **49** generally resulted in low yield of the desired products. Addition of ethynyl magnesium bromide **47** to initial clear solutions of diethylchlorophosphate to generate **49** as outlined in **Scheme 13**, would yield reactions mixtures which darkened from light brown to black with some solid particulate suspension. Isolation of crude oil showed the presence of a 162 $[M+H]^+$ signal, corresponding to the desired diethyl ethynyl phosphonate **49**, as well as the observed presence of ethyl peaks and a large splitting corresponding to the phosphorus coupled alkynyl proton. However, column chromatography (hexane: ethyl acetate (1:1)), Rf = 0.22 (consistent with literature) yielded the product as a colourless oil in only 5% yield (based on ^{31}P NMR analysis). It was hypothesized that polymerisation of the reagents (due to disproportionation of the ethynyl Grignard reagent to acetylene and the *bis*-Grignard reagent) was limiting reaction. However,

dilution of the reaction conditions (by 4-fold) led to lower overall reaction and no product was isolated from this attempt.



Scheme 13: Synthetic route for diethyl-protected pTza analogue **50** of pHis based on work by McAllister *et al.* (154) *yield based on ^{31}P analysis of crude product mixture in attempt at replicating this synthesis.

Synthesis of **49** via the TMS-protected alkyne **51** was also attempted. In this case, **52** was obtained from TMS-protected acetylene via deprotonation with ^{i}pr-MgCl and addition of *bis*(diethylamino)chlorophosphine. Two equivalents of ethanol and ETT were added and after an hour of stirring, the mixture was oxidised in H_2O_2 and extracted in ethyl acetate. Column chromatography in hexane: ethyl acetate (2:3) (R_f = 0.44) gave the TMS-protected compound **53** in a yield of 0.55 mmol, ~14% yield based on impurities remaining in ^{1}H NMR. The TMS-protected alkyne **53**, was deprotected by mixing with tetrabutylammonium fluoride (TBAF) at 0˚C in THF: H_2O (1:1). After 3 hours the crude mixture was extracted in ether and the ethereal layer was concentrated *in vacuo*. The crude oil (44 mg) was however lost during column chromatography in hexane: ethyl acetate (1:1) (R_f = 0.2), likely due to its relative volatility leading to loss via co-evaporation with solvent, a common problem throughout this synthetic campaign. Therefore, more reliable methods for producing stable pTza precursor alkynes were sought.

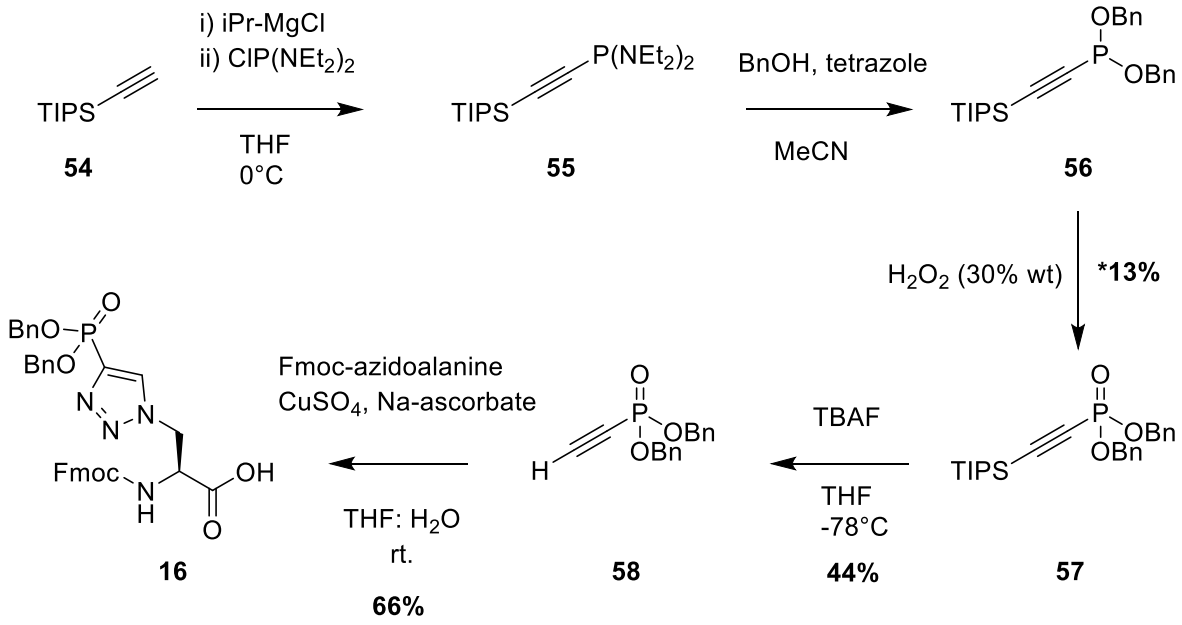


Scheme 14: Synthesis of a TMS-protected diethyl ethynyl phosphonate **53** from TMS acetylene **51** via tetrazole-mediated substitution on a *bis*(diethylamino) intermediate **52**.

Dibenzyl-protected pTza **16** was generated via the dibenzyl ethynyl precursor **58** (**Scheme 15**). This route was based on work by McAllister *et al.* where a dibenzyl ethynyl species was

produced and reacted with azidoalanine to form the benzyl-protected pTza analogue of phosphohistidine (159). Under an N_2 atmosphere, TIPS-acetylene was deprotonated by the addition of ^{i}Pr-MgCl before addition of *bis*(diethylamino)chlorophosphine to produce the P(III) intermediate **55**. Following removal of the solvent (THF) *in vacuo* and addition of MeCN, the P(III) intermediate was converted to the P(V) intermediate **57** by addition of tetrazole and benzyl alcohol, followed by oxidation with H_2O_2. TIPS-protected dibenzyl phosphonate **57** was produced in 13% yield (from TIPS-acetylene **54**). The TIPS group was removed with tetrabutyl ammonium fluoride (TBAF) to obtain compound **58** in 44% yield.

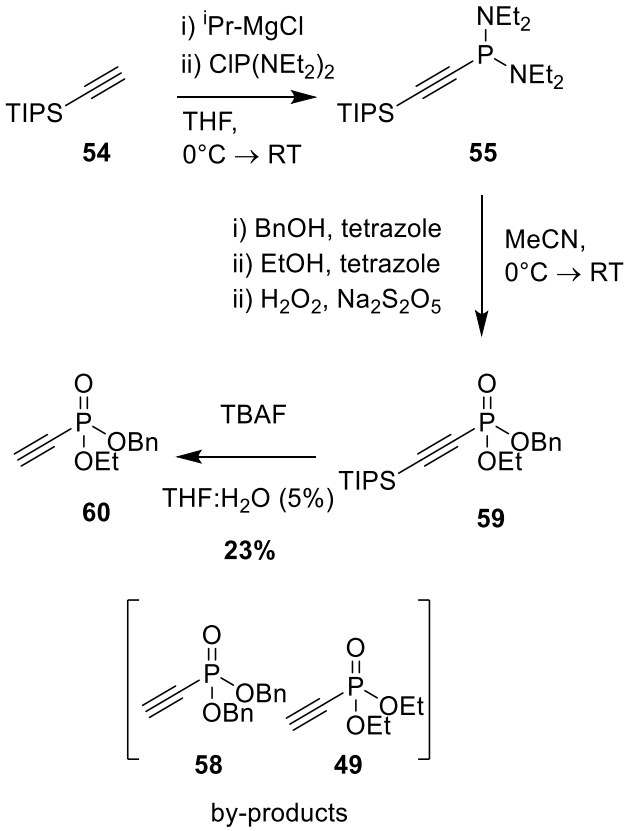
The dibenzyl ethynyl phosphonate **58** (1.2 eq.) was subsequently reacted with Fmoc-azidoalanine (1 eq.) in a copper-catalysed click reaction with $CuSO_4$ (0.1 eq.) and sodium ascorbate (0.3 eq.) in THF: H_2O (1:1). The reaction was stirred at room temperature overnight. The solvent was removed *in vacuo* and the residue was washed with Na_2CO_3 (10%) and ethereal washes. The aqueous layer was acidified to pH 1 with HCl (conc.) and washed with EtOAc. Concentration of the organic layer provided the dibenzyl-protected pTza analogue of phosphohistidine **16** in 66% yield.



Scheme 15: Synthetic route for dibenzyl pTza analogue of phosphohistidine **16**. Production of a *bis*(diethylamino) species **55** enables tetrazole-mediated substitution of alcohols onto P(III) species before oxidation to form protected phosphonate precursor **57**. Subsequent copper-click addition of the dibenzyl ethynyl phosphonate **58** with azidoalanine produces dibenzyl-protected pTza analogue of pHis **16**. The yields represent quantities obtained when reproducing these routes. *yield based on starting limiting reagent $ClP(NEt_2)_2$ from the initial step.

Having reproduced the described synthetic routes for ethyl-protected (**Scheme 13**) and benzyl-protected ethynyl phosphonates (**Scheme 15**), production of a mixed benzyl-ethyl protected alkyne phosphonates was attempted (**Scheme 16**). Given the successful synthesis of the dibenzyl ethynyl phosphonate **58** using the tetrazole-mediated substitution, the method was adapted to enable mixed phosphonate production. This was first attempted through addition of 1 eq. alcohol and 1 eq. tetrazole to the P(III) intermediate **55**, followed by 1 eq. of a distinct alcohol to the intermediate **55** to produce the mixed phosphate ester.

As before, TIPS-acetylene **54** was deprotonated using ^{i}Pr-MgCl and $ClP(NEt_2)_2$ was added to generate the P(III) species **55**. This species was not isolated, concerned about possible degradation or premature oxidation if exposed to the atmosphere. Therefore, THF was removed using a Schlenk line and dry acetonitrile was added. Benzyl alcohol (1 eq. to TIPS-acetylene) and tetrazole (1 eq.) were added and the reaction allowed to stir under N_2 at room temperature for an hour, before the addition of an additional equivalent each of ethanol and tetrazole followed by stirring for an hour. The TIPS-protected benzyl-ethyl compound **59** was isolated, together with TIPS-protected dibenzyl **58** and diethyl ethynyl phosphonates **49** in a 0.6:1:0.3 ratio (as determined by analysis of the mixture by ^{31}P NMR, **Figure 43**). The three species were inseparable by TLC, however since the free dibenzyl ethynyl phosphonate and diethyl ethynyl phosphonate are separable the mixture was incubated with TBAF to yield the free alkynes **60**, **58** and **49.** Column chromatography in hexane: ethyl acetate (3:7) yielded three distinct species $R_f = 0.35$, $R_f = 0.43$ and $R_f = 0.52$ corresponding to the diethyl phosphonate **49**, benzyl-ethyl phosphonate **60** and dibenzyl ethynyl phosphonate **58** compounds respectively. With the desired mixed phosphonate alkyne **60** being obtained in 23% yield.



Scheme 16: Pathway for the production of alkynyl phosphonates from TIPS-acetylene **54** generating a phosphoramidite intermediate **55** that can be substituted with alcohols (BnOH and EtOH) to generate a protected benzyl-ethyl ethynyl phosphonate **59**. Subsequent silyl deprotection with TBAF generates the benzyl-ethyl ethynyl phosphonate **60**.

Curiosity regarding the preference of phosphonate formation for dibenzyl, diethyl or ethyl-benzyl species led to investigation into the mechanism of the reaction. The initial hypothesis was a statistical argument: that a monobenzyl phosphoramidite intermediate would form on addition of benzyl alcohol, with the *bis*(dialkylamino) species being more reactive and therefore more likely to convert to a monobenzyl, monodialkylamino intermediate. When ethanol was added the monobenzyl would then form the mixed benzyl-ethyl phosphonite leading to a majority of the mixed phosphonate product. To test this hypothesis, the reaction was performed with concurrent addition of benzyl alcohol and ethanol to complement this initial sequential reaction.

Following the concurrent addition of benzyl alcohol: ethanol (1:1) to the P(III) intermediate **55**, ^{31}P NMR analysis showed predominant formation of the statistically favoured mixed product. Higher quantities of diethyl phosphonate were present, and the minor product appeared to be dibenzyl ethynylphosphonate (**Figure 43**). This is in contrast to the sequential addition where a majority of the dibenzyl-substituted product was formed.

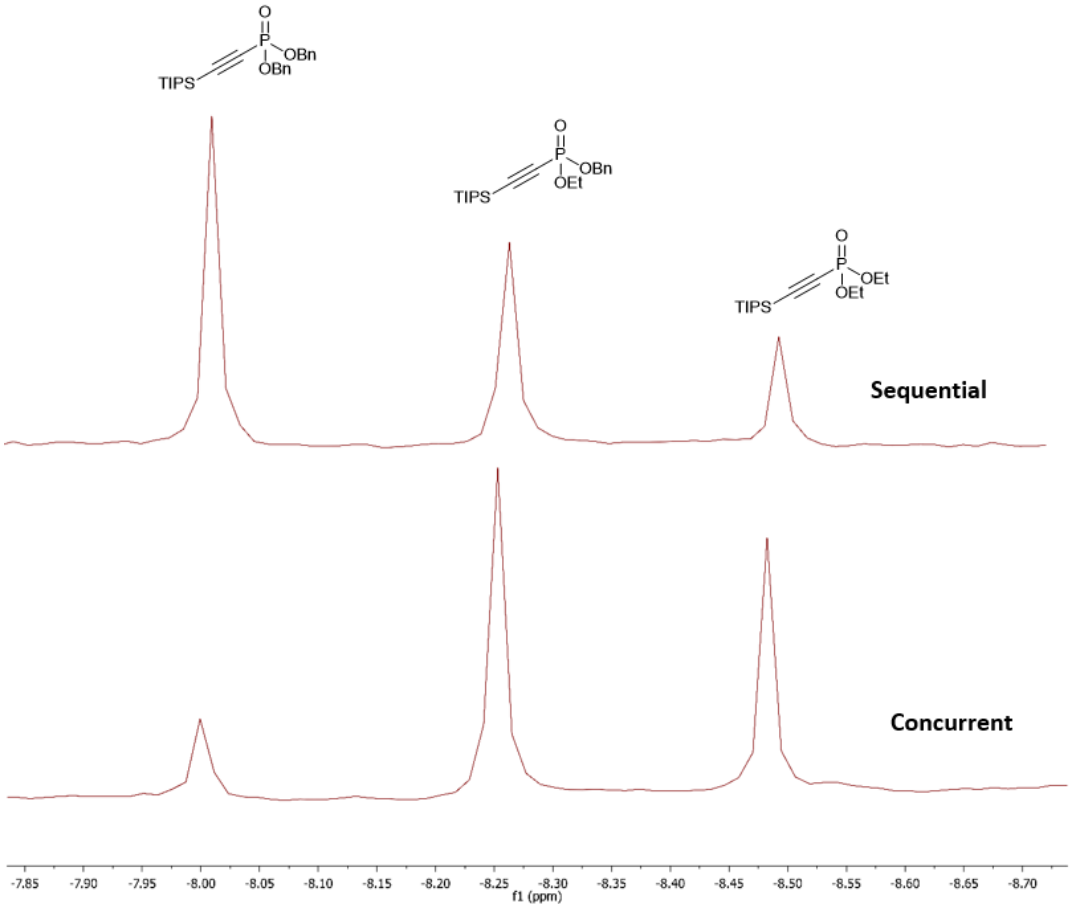


Figure 43: ^{31}P NMR of crude benzyl-ethyl ethynylphosphonate synthesis mixture. Displaying the intensities of TIPS-protected dialkyl alkynyl phosphonates in the samples after sequential addition of benzyl alcohol and ethanol and concurrent addition of both alcohols.

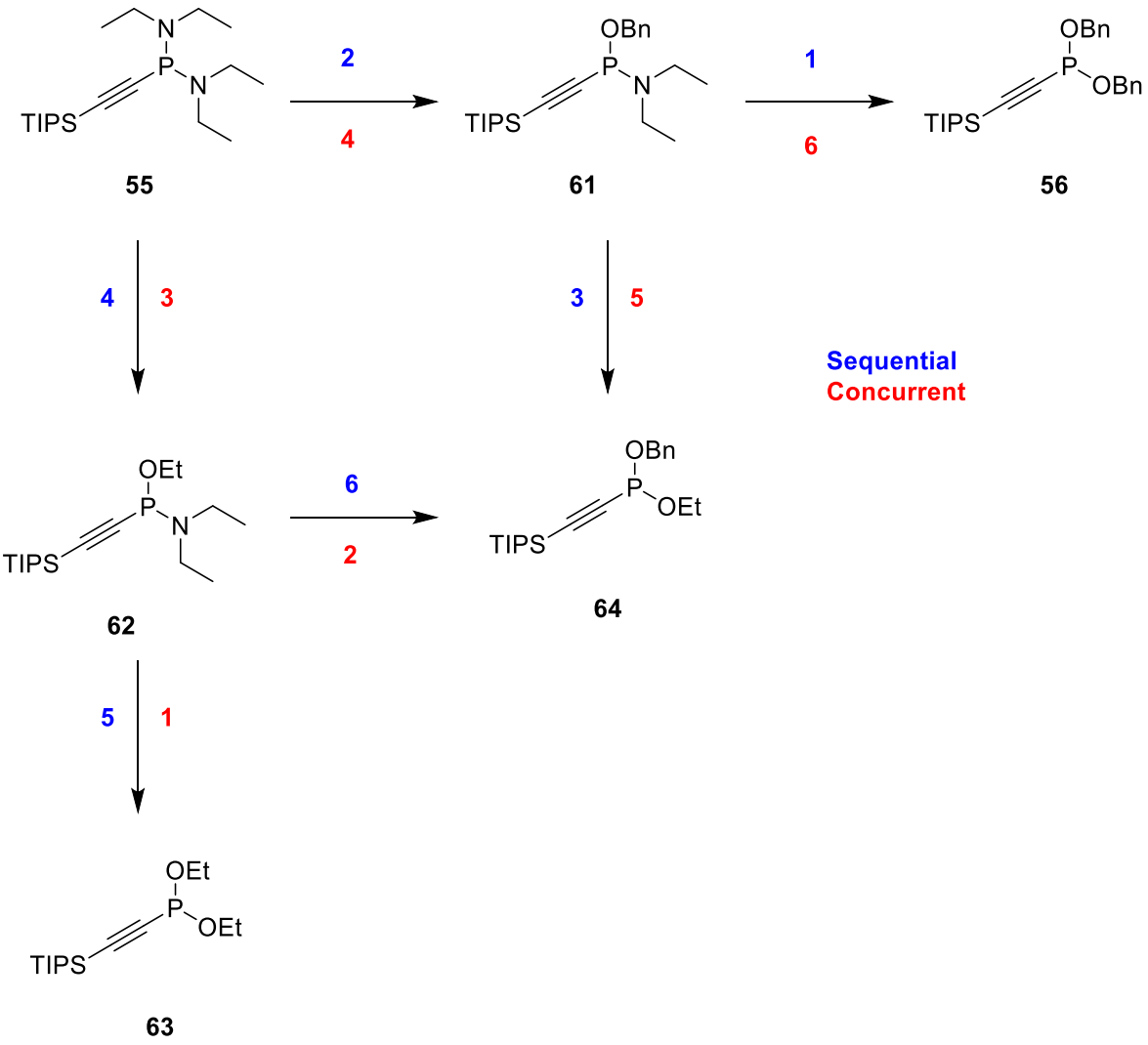


Figure 44: Outline for conversion of diamino intermediate **55** to benzyl-ethyl ethynyl phosphonite **64** through monoalkyl intermediates **61** and **62**. The numbers above the arrows (blue, sequential; red, concurrent) represent the theorised favourability of the product formation (1 = most favoured conversion).

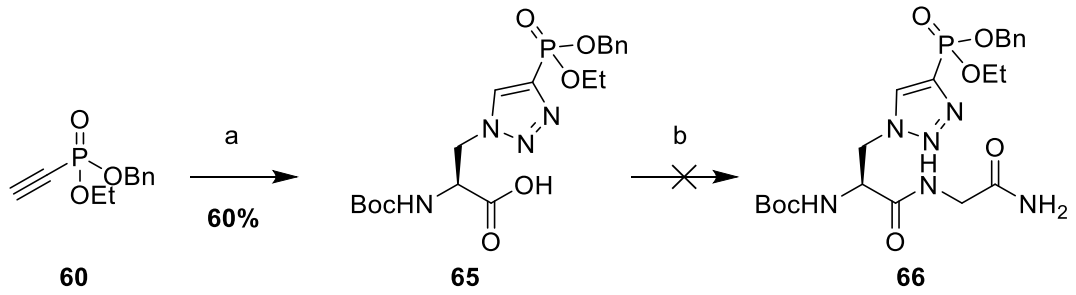
The two experiments suggest that the monosubstituted phosphoramidite intermediate is more reactive than the starting material and that ethyl substitution is favoured over benzyl substitution. The monosubstituted phosphoramidite is likely more reactive due to the steric accessibility of the phosphorus atom; the ethyl or benzyl group is smaller than the diethylamino groups of the starting material and this likely makes the monosubstituted phosphoramidite more available for further substitution than the diphosphoramidite. Sterically, the argument also suffices that ethyl is a smaller side chain than benzyl and thus a monoethylphosphoramidite would react more rapidly than a monobenzylphosphoramidite, hence diethyl substitution was more probable than dibenzyl substitution when both possibilities were available. An electronics rationale may be available by assuming the pKa of the diethylamine conjugate acid (~11) models the diethylamino groups in diphosphoramidite and

monophosphoramidite; and the pKa of ethanol (~16) and benzyl alcohol (~15.5) accurately depict their respective groups (ethyl and benzyl). The amino groups of the diphosphoramidite starting material and monosubstituted intermediates were more acidic than the alcohol groups, as such, substitution with alcohol was favoured with ethyl substitution being more stable than benzyl. The steric and electronic favourability of ethyl substitution provides a challenge for optimising the yield of a benzyl-ethyl mixed substitution product. This emphasised the significance of controlled addition of the alcohol reagents to optimise monobenzyl formation prior to addition of ethanol to maximise benzyl-ethyl yields.

3.2 Generation of orthogonally protected phosphonotriazoles

Despite not being able to generate 100% yield of the desired mixed benzyl/ethyl ethynyl phosphonate it was possible to readily isolate the desired mixed species. Incorporation of this precursor to generate a phosphohistidine analogue was then investigated. Boc-pTza(OBnOEt)-OH **66** was generated via reaction of benzyl-ethyl ethynylphosphonate **60** with Boc-protected azidoalanine via Cu-catalysed click reaction with $CuSO_4$ and sodium ascorbate in 60% yield. The produced pTza compound **65** was coupled to glycinamide with Oxyma Pure, DIC and DIPEA, outlined in **Scheme 17**. The objective was then to biotinylate this structure for use as a pulldown probe.

Unfortunately, the polar coupled compound **66** was not isolable by column chromatography and could only be eluted from silica using high concentrations of methanol >10% to salvage the coupled product leading to contamination of the product.

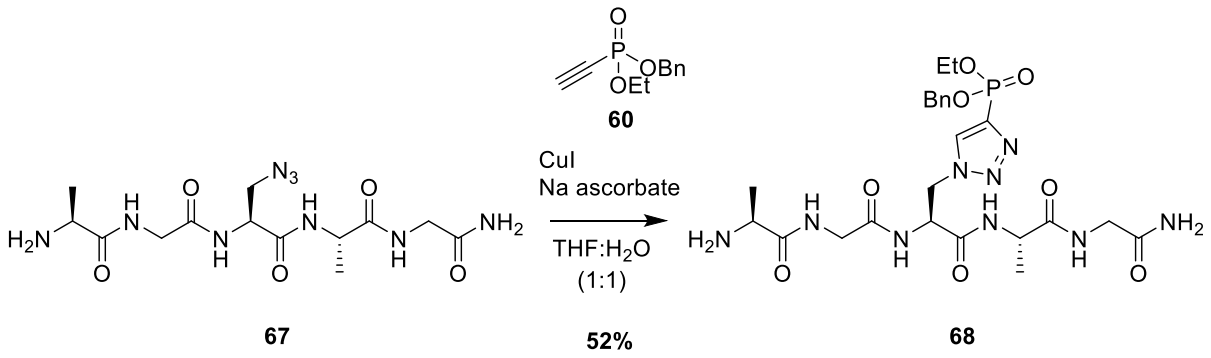


60 **65** **66**

Scheme 17: Conversion of benzyl-ethyl ethynyl phosphonate **60** to Boc-pTza(OBnOEt)-OH **65** via Cu catalysed click with Boc-azidoalanine, $CuSO_4$ and sodium ascorbate (**a**). Compound **65** was used as a building block for coupling to glycinamide with DIC and Oxyma Pure (b) to generate a larger compound **66**.

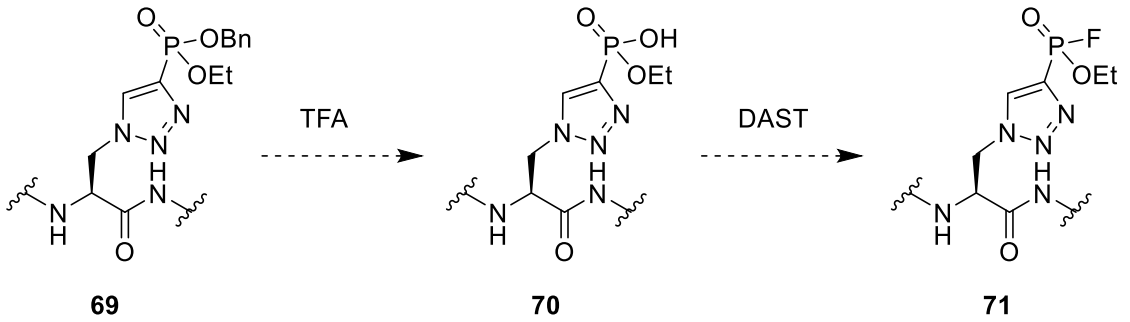
After loss of the Boc-protected benzyl-ethyl building block **65** an alternative possibility for copper-catalysed click of the benzyl-ethyl alkyne **60** into an azidoalanine-containing peptide was considered. An azidoalanine-containing peptide **67** (82 mg, 0.21 mmol) was produced by manual Fmoc-SPPS in 97% yield (purity unknown) and a diagnostic $[M+H]^+$ = 386.04 peak was detected by LCMS. Peptide **67** (1 eq.) was mixed with benzyl-ethyl alkyne **60** (1.2 eq.) with CuI and sodium ascorbate in THF: H_2O (1:1) overnight. The volatiles were removed *in vacuo* and extraction was performed in ether. The ethereal layer was concentrated, and the peptide was resuspended in water and lyophilised. Peptide **68** was obtained (68 mg, 0.11 mmol)

in a 52% yield (purity unknown). Analysis by mass spectrometry showed a $[M+H]^+ = 610.2501$ peak corresponding to the copper-click product peptide **68**.



Scheme 18: Copper-catalysed click, combining benzyl-ethyl ethynyl phosphonate **60** with a simple azide containing peptide **67** to produce a pTza-containing peptide **68**.

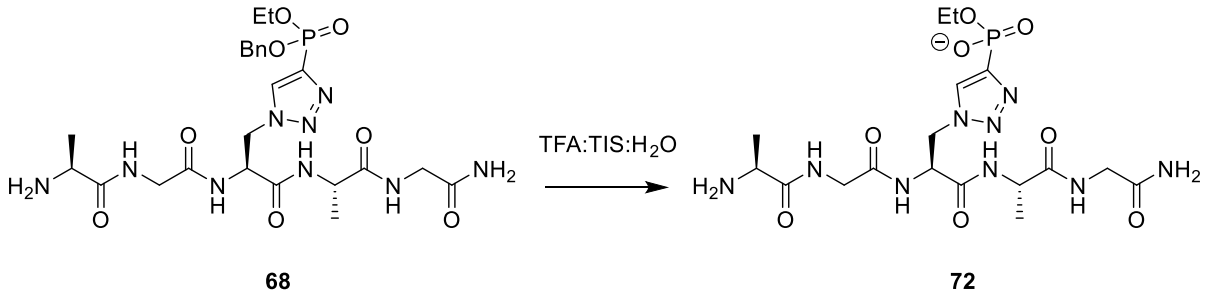
The original strategy would have been removal of the benzyl group on **68** to generate a monoethylphosphonate, this compound would then have been converted to an ethyl fluorophosphonate as reported by Cravatt *et al.* using diethylaminosulphur triflouride (DAST) (**Scheme 19**) (179).



Scheme 19: Proposed route for conversion of a benzyl-ethyl pTza peptide **69** to a monoethyl pTza peptide **70** and subsequent conversion to a fluorophosphonate pTza peptide **71** using DAST.

To generate a monoethyl phosphonate pTza, peptide **68** was treated with a standard deprotection cocktail of TFA: TIS: H_2O (95:2.5:2.5) for 3 hours at room temperature. The volatiles were removed *in vacuo* and the residual oil was purified by anion exchange chromatography (Q-Sepharose FF) in an increasing gradient of ammonium bicarbonate (0-50 mM) the eluent was assessed by mass spectrometry until eluting fractions showed no further evidence of the monoethyl pTza peptide **72**. The monoethyl pTza peptide **72** was detected at $[M+NH_4]^+ = 536.2$ but there was significant co-elution throughout the gradient with

another unidentified impurity (687.1 Da). The purest fractions from the ion exchange column, from higher ammonium concentration elutions, resulted in a low mass of peptide **72** (>1 mg).

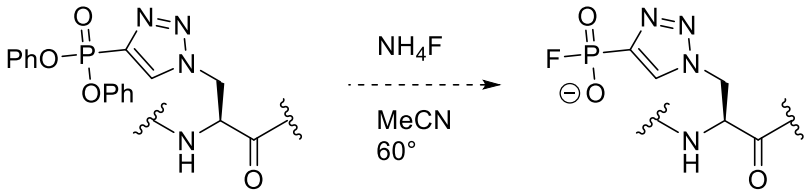


Scheme 20: Conversion of benzyl-ethyl pTza peptide **68** to monoethyl pTza peptide **72** as precursor to fluorophosphonate.

In addition to the low yield of peptide, it was determined that this route was not suitable for further derivatisation to generate a pulldown probe due to the challenges in generating the required fluoro-derivative. The concern here was the limitation of producing the fluorophosphonate functionality with DAST which may have fluorinated the hydroxyl on a deprotected phosphonate as well as deprotected serine, threonine or tyrosine, which would have created off-target reactivity in a probe peptide. Production of the fluorophosphonate (FP) functionality on the analogue prior to incorporation in a peptide was not feasible due to the instability/reactivity of the FP functionality, functionality needed to be generated, and the probe used immediately. Therefore, this route was never taken any further due to the realisation of challenges associated with potential indiscriminate fluorination by DAST.

3.2.1 Generation of Phenyl-modified pTza analogues

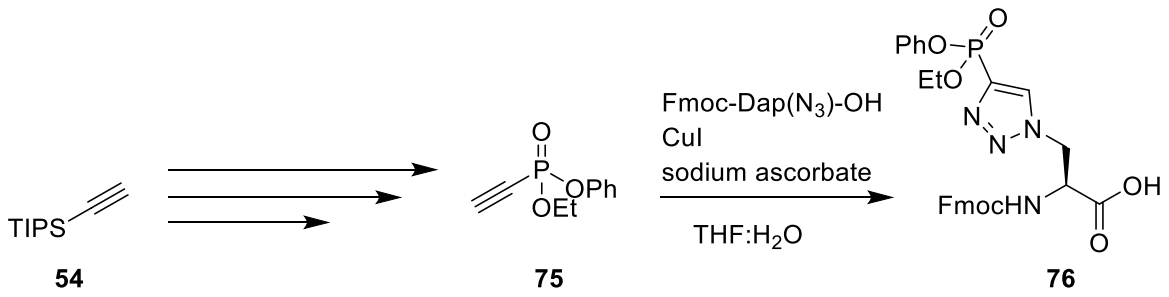
A potential alternative to the direct conversion of the ethyl phosphonate monoester is to instead generate the fluorophosphonate from phenyl substituted phosphonates (**Scheme 21**). A similar reaction was reported by d'Andrea and Townsend (180). The milder conditions of this reaction (NH_4OH in acetonitrile at 60˚) may have been more compatible with the functional groups found a typical peptide structure unlike the planned conversion of monoethylphosphonates using DAST. Implementation of the strategy would be dependent upon access to the peptides containing a diphenyl-protected pTza analogue of phosphohistidine. Alternatively, whilst d'Andrea and Townsend had presented a conversion of diphenyl phosphonate to a fluorophosphonate (180), we anticipated that the monoethyl fluorophosphonates would be more effective. This functional group has proven ability as electrophilic capture probes as shown by Cravatt *et al.* (179), therefore production of an asymmetric phenyl-ethyl ethynyl phosphonate as a pTza fluorophosphonate precursor was prioritised.



73 **74**

Scheme 21: Conversion of diphenyl pTza **73** to fluorophosphonate pTza **74** by fluoride source, R represents an alkyl group, as outlined by d'Andrea (180)

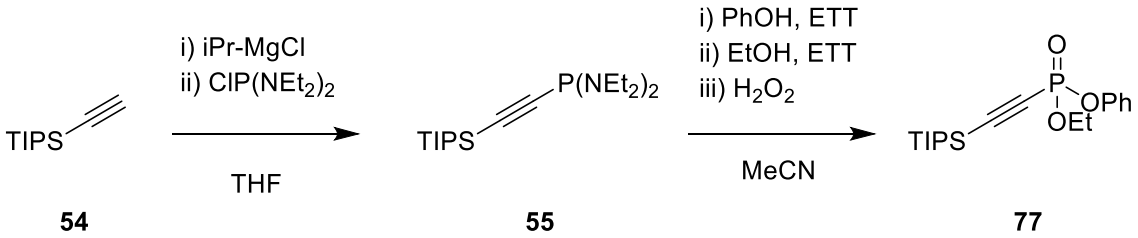
The planned route for generation of a phenyl-ethyl protected pTza analogue of pHis was via synthesis of a phenyl-ethyl ethynyl phosphonate **75** and subsequent Cu-catalysed click with Fmoc-azidoalanine (**Scheme 22**). Effectively utilising the asymmetric phosphonate substitution performed to produce benzyl-ethyl phosphonates and Cu-catalysed click chemistry to produce the functionalised Fmoc-pTza which would be SPPS compatible.



Scheme 22: Overview for conversion of TIPS-acetylene **54** into phenyl-ethyl ethynyl phosphonate **75** following methodology described in **Scheme 16**, replacing one equivalent of alcohol with phenol. Cu-catalysed click would convert alkyne **75** into pTza FP-precursor **76** capable of incorporation into peptide sequences through Fmoc-SPPS.

Synthesis of the TIPS-protected precursor **75** was performed by deprotonation of TIPS-acetylene by isopropyl-magnesium chloride in THF followed by addition of *bis*(diethylamino)chlorophosphine to generate intermediate **55**. The solvent was removed *in vacuo* and exchanged for MeCN. An equivalent of phenol and 5-ethythiotetrazole (ETT) was added followed by another equivalent of ethanol and ETT. After an hour of stirring, the alkyne **77** was extracted in EtOAc and the organic extracts were oxidised with H_2O_2. This initial attempt did not show evidence of product **77**. It is thought that this was due to phenol not dissolving fully in the MeCN reaction solvent.

To remedy this, on a subsequent attempt, an equivalent of phenol was pre-dissolved in MeCN (0.5M) and ETT in MeCN (0.5M) followed by an equivalent of ethanol and ETT in MeCN (0.5M) were added to intermediate **55**. The crude product **77** was washed with ethyl acetate and oxidised in H_2O_2. The ethyl acetate washings were reduced *in vacuo* and the crude oil was purified by column chromatography (hexane: ethyl acetate (3:7)). The colourless oil obtained by column chromatography (93 mg) was analysed by LCMS which showed a $[M+H]^+$ 366 peak, indicative of the TIPS-protected phenyl-ethyl ethynyl phosphonate **77** in large proportion (~80%) however, there were impurities present which were not separable by column chromatography, including masses associated with triphenyl phosphonate and diphenyl phosphonate. The production of **77** was hard to corroborate through NMR analysis of the crude oil due to the presence of impurities, particularly of phenyl-containing species which made assignment impossible.



Scheme 23: Synthesis of TIPS-protected phenyl-ethyl phosphonate alkyne **77**.

Based on asymmetric synthesis of benzyl-ethyl phosphonate alkynes outlined earlier in this chapter (**Scheme 16**), the possibility of contamination with side-products of diphenyl and diethyl substituted phosphonates makes isolation challenging without deprotection of the TIPS-protecting group and purification.

Silyl deprotection of these alkyne species (diphenyl, phenyl-ethyl and diethyl) would have typically been performed with a source of fluoride, which presented a problem; the formation of fluorophosphonates as outlined by d'Andrea and Townsend was also reliant on a source of fluoride (180), so deprotection of the alkyne would have generated a, potentially volatile, fluorophosphonate. The reactivity of the fluorophosphonate alkyne species would have made formation of the pTza analogue impossible.

Additionally, the potential for accidental formation of a small volatile fluorophosphonate was avoided due to the potential for creating an extremely deadly compound. Fluorophosphonates are known motifs in chemical warfare nerve agents (**Figure 45**) (181). The same biological activity that makes fluorophosphonate compounds excellent tools for covalent molecular biology is responsible for their activity as toxins, for example, as acetylcholinesterase inhibitors (181, 182). Fluorophosphonate nerve agents work by irreversibly binding to a serine residue in the active site of acetylcholinesterase, this inhibits the function of the enzyme leading to excessive activity of acetylcholine receptors and subsequent death by asphyxiation (181, 182). As such, the importance of avoiding the production of these compounds, particularly inadvertently, was paramount.

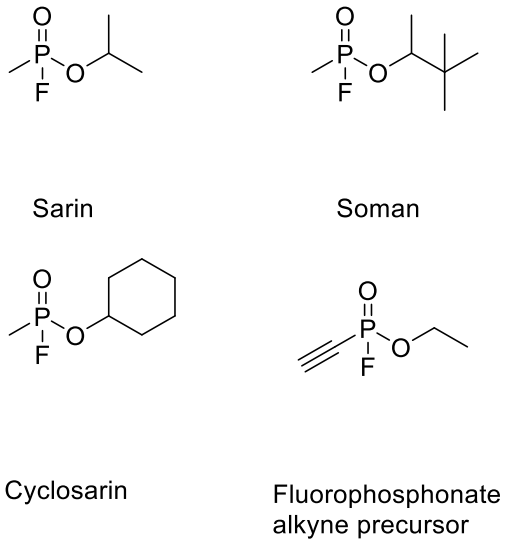
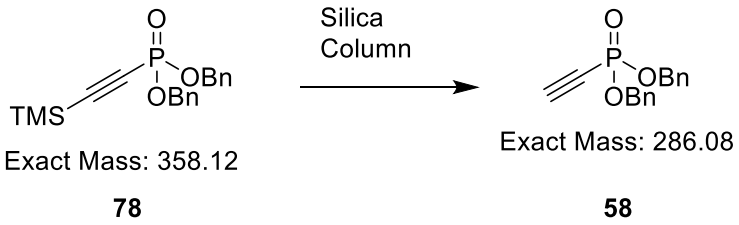


Figure 45: Structures of known nerve agents with fluorophosphonate functionality and the alkyne precursor of pTza if prematurely fluorinated (181).

The nerve agent structures in **Figure 45** show remarkable similarity to the fluorinated pTza precursor alkyne. Therefore, exposure of phenyl-ethyl ethynyl phosphonate to fluoride sources was diligently avoided. Silyl deprotection from an alkyne is almost universally performed with a source of fluoride, most commonly TBAF (183-185), but silver fluoride (186) and caesium fluoride (187) are also reported. As such, use of the TIPS protecting group was an obstruction to safely producing the desired alkyne and needed to be avoided. If a fluorophosphonate pTza analogue were to be made, a phenyl phosphonate alkyne needed to be produced either by finding alternative protecting groups or circumventing deprotection altogether.

One alternative was use of the TMS protecting group as an alternative. In parallel with work on TIPS-protected alkynes, work had been attempted with TMS-protected dibenzyl ethynyl phosphonate **78**. TMS-acetylene was deprotonated with ^{i}pr-MgCl in ice-cold THF and subsequently mixed with *bis*(diethylamino)chlorophosphine. The intermediate was immediately mixed with 2 equivalents of benzyl alcohol and 2 equivalents of ETT, followed by extraction in ethyl acetate and oxidation with hydrogen peroxide. The crude oil (770 mg) was purified by column chromatography in hexane: ethyl acetate (3: 2) (R_f = 0.26) to yield 510 mg of a colourless oil. However, after chromatography, LCMS analysis showed $[M+H]^+$ 286 and $[2M+H]^+$ = 572.90 in the isolated product, having showed $[M+H]^+$ 358 in the crude product prior to silica chromatography (**Scheme 24**). NMR analysis was more challenging, the only diagnostic would have been comparison of the alkyne protons to the benzyl integrals, and these did not match up. If column chromatography removed TMS protection from the alkyne, this would have avoided the need for sources of fluoride for

deprotection of the silyl-protected phenyl ethynyl phosphonate. Interestingly this observation appeared to have precedent in literature where Dutremez *et al.* had obtained a mixture of deprotected and TMS-protected diethyl ethynyl phosphonate after silica gel column chromatography (188).



Scheme 24: TMS-protected dibenzyl ethynyl phosphonate **78** loses TMS-protection to generate free alkyne **58** after silica column.

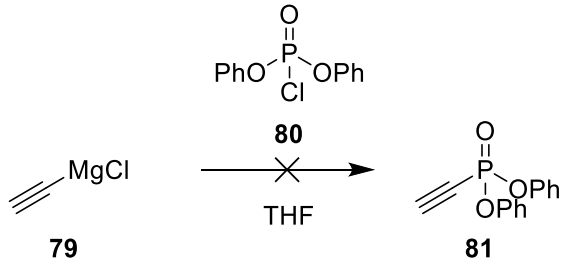
To test the repeatability of this observation, synthesis of diethyl ethynyl phosphonate was repeated, as earlier described with TMS-protection (**Scheme 14**). Purification of the alkyne by column chromatography did not produce the deprotected alkyne. LCMS analysis indicated $[M+H]^+$ 234, corresponding to the TMS-protected diethyl ethynyl phosphonate **53**.

Nonetheless, the synthesis of the TMS-protected diphenyl ethynyl phosphonate was performed to assuage any doubt. TMS acetylene was deprotonated with ^{i}pr-MgCl and mixed with *bis*(diethylamino)chlorophosphine in THF. After a solvent exchange to MeCN, the intermediate was mixed with 2 equivalents of phenol (1M in MeCN) and 2 equivalents of ETT. After extraction in ethyl acetate and oxidation in H_2O_2, the organic layer was concentrated *in vacuo*. After column chromatography in hexane: ethyl acetate (3:2) the isolated fractions were analysed by LCMS which showed peaks at $[M+H]^+$ 330 and $[2M + H]^+$ 661.13, corresponding to the TMS-protected diphenyl ethynyl phosphonate. NMR analysis indicated the presence of impurities, likely from phenyl phosphonates that complicated confirmation of the synthesis. However, column chromatography had not removed the TMS protection from the alkyne in this instance either.

An alternate route to synthesise phosphonate alkynes by mixing chlorophosphates with alkynyl-Grignard reagents exists in the literature (**Scheme 13**) (154). This route was discussed earlier as an initial means of producing diethyl ethynyl phosphonate that was disregarded in favour of tetrazole-mediated substitution. Additionally, this route was initially disregarded given that synthetic pathways would be limited by the commercial availability of varied chlorophosphates. However, the late-stage fluorination described by d'Andrea and Townsend (180) relied on diphenylphosphonates unlike other fluorination methods which appear to use more asymmetrically-substituted phosphonates, for example, Dutra *et al.* (189) using p-nitrophenol-ethanol phosphonates or the monoethyl phosphonate by Cravatt *et al.* (179).

Diphenyl chlorophosphite and ethynyl magnesium bromide were commercially available compounds. Therefore, combination of these two compounds ought to have produced diphenyl ethynylphosphonate (**Scheme 25**), This may have provided a simpler route to the desired diphenylphosphonate alkyne **81**, given the literature precedence for synthesising the diethyl ethynyl phosphonate by this route (154).

Synthesis of diphenyl ethynyl phosphonate **81** was attempted using ethynyl magnesium chloride (0.5M in THF) **79** and diphenylchlorophosphate **80** (**Scheme 25**). The reaction underwent a slow colour change from orange to black. During ethereal work-up the black colour appeared to be a tar-like solid, it was insoluble in the organic solvents tested (ethyl acetate, hexane, DCM and diethyl ether) as well as aqueous solutions. The collected organic layers after extraction were dried and gave a residual black, orange-tinged oil. LCMS analysis gave two peaks one at $[M+H]^+ = 250$ and another at $[M+H]^+ = 326$, corresponding to the masses of diphenylphosphate and triphenylphosphate respectively.



Scheme 25: Conversion of ethynyl magnesium chloride **79** to diphenyl ethynyl phosphonate **81** through reaction with commercially available P(V) species: diphenyl chlorophosphonate **80**.

An explanation for the formation of triphenylphosphate and diphenylphosphate was sought. Conversion of diphenyl chlorophosphate **80** to diphenyl phosphate **83** by hydrolysis was a feasible route, there was precedent for this reaction in THF (190) which indicated potential degradation of the raw material chlorophosphate. The mechanism by which triphenylphosphate **82** was formed was less clear, but it was the most likely structure justifying the presence of a $[M+H]^+$ 326 peak (**Figure 46**).

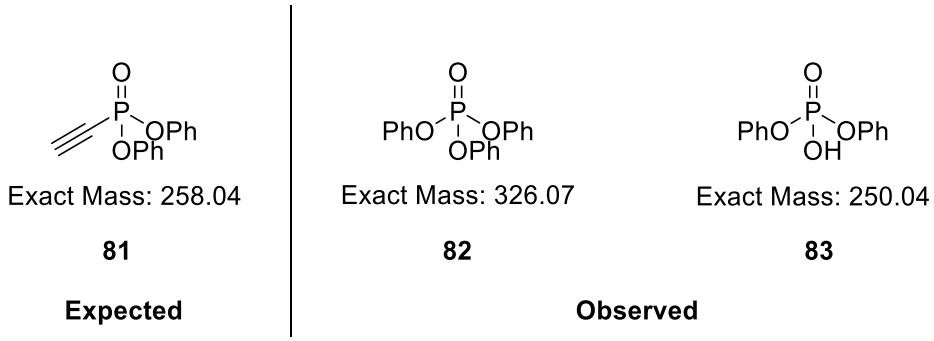
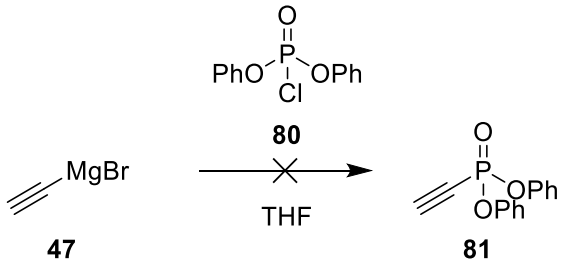


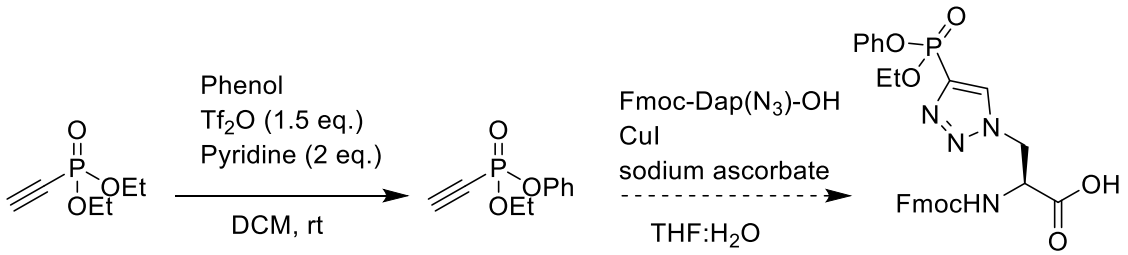
Figure 46: Expected masses for desired diphenyl ethynyl phosphonate **81** against observed masses in LCMS analysis corresponding to triphenylphosphonate **82** and diphenylphosphonate **83**.

Literature sources tended to list the use of ethynyl magnesium bromide **47** to generate diethyl ethynyl phosphonate **49** (154, 191) therefore the reaction was reattempted with ethynyl magnesium bromide **47** in place of ethynyl magnesium chloride **79**, Reattempting the reaction with ethynyl magnesium bromide (**Scheme 26**), also gave rise to the $[M+H]^+$ 250 and 326 in LCMS analysis.



Scheme 26: Ethynyl magnesium bromide **47** in place of ethynyl magnesium chloride **79** to generate diphenyl ethynyl phosphonate **81**.

Literature precedent existed for the conversion of diethyl phosphonate to a phenyl-ethyl phosphonate using Tf_2O and pyridine (192), (**Scheme 27**).



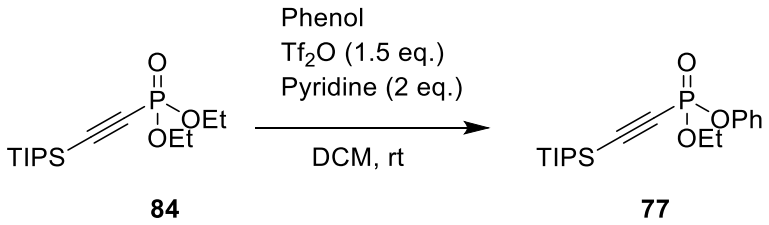
49 **75** **76**

Scheme 27: Feasible route to produce a reactive pTza analogue **76**. Starting with diethyl ethynyl phosphonate **49**, which can be synthesised according to methods described by McAllister *et al.* (159) and outlined in **Scheme 13**. Production of a phenyl-ethyl ethynyl phosphonate **75** can be achieved by the procedure outlined by Huang *et al.* (192) and Cu-catalysed cycloaddition of the mixed alkyne **75** directly onto Fmoc-Dap(N_3)-OH would produce the pTza analogue **76** for insertion into peptide sequences.

Diethyl ethynyl phosphonate (1 eq.) in DCM was mixed with triflic anhydride (1.5 eq.), pyridine (2 eq.) was added, and the mixture was stirred for 10 minutes. Phenol (4.4 eq.) was added, and the reaction was stirred at room temperature for an hour. After an hour, undissolved phenol remained and TLC showed starting material in hexane: ethyl acetate (4:1) (R_f = 0.24) but no evidence of product formation. LCMS analysis of the crude reaction mixture showed $[M+H]^+$ = 163, indicating no change from the initial diethyl phosphonate alkyne **49**.

Given the difficulty in synthesising the diethyl ethynyl phosphonate **49** in sufficient quantity, the TIPS-protected diethyl ethynyl phosphonate was used to test the Tf_2O reaction in **Scheme 28**.

TIPS-protected diethyl ethynyl phosphonate (1 eq.) was dissolved in DCM and triflic anhydride (1.5 eq.) and pyridine (2 eq.) were added. After stirring for 10 minutes, phenol (2.7 eq.) was added as a 0.3M solution pre-dissolved in DCM to promote solubility in the reaction mixture. After 30 minutes, TLC in hexane: ethyl acetate (4:1) showed loss of starting alkyne at R_f = 0.24 and formation of a spot at R_f = 0.3. The reaction mixture was concentrated *in vacuo* and purified by column chromatography in hexane: ethyl acetate (4:1), fractions containing the compound corresponding with R_f = 0.3 were isolated. The isolated fractions were analysed by LCMS showing $[M+H]^+$ = 367.2, aligning with the TIPS-protected phenyl-ethyl ethynyl phosphonate. NMR and LCMS analysis showed co-elution with TIPS-protected diphenyl ethynyl phosphonate, which happened to be the dominant species (82% by area). Frustratingly, this demonstrated the reaction conditions outlined in the work by Huang *et al.* (192) were not as selective for the mixed product as claimed.



Scheme 28: Formation of TIPS-protect phenyl-ethyl ethynylphosphonate **77** from TIPS-protected diethyl ethynylphosphonate **84**, in accordance with methodology by Huang *et al.* (192)

Ultimately, production of a di/mono phenyl ethynyl phosphonate appeared to be challenging. A more effective way to produce the compound would have been synthesis of the diethyl pTza or monoethyl pTza (achieved through deprotection of benzyl-ethyl pTza) and subsequent treatment with Tf_2O to produce phenyl-substituted phosphonate pTza analogues, therefore avoiding silyl deprotection altogether. The amino acids could then be incorporated into peptide sequences for use as probes.

3.3 Conclusions and Future Work

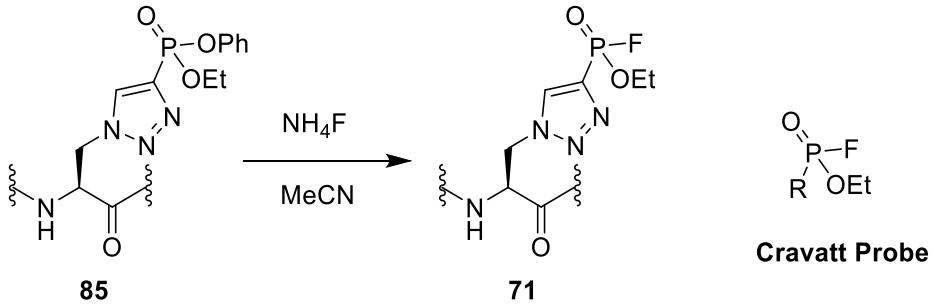
The production of fluorophosphonate functionality on a pTza analogue, either prior to or post insertion into a peptide, appeared untenable. No pathway had been produced to reach a suitable fluorophosphonate precursor alkyne or FP-capable pTza analogue.

The possibility of generating asymmetrically substituted phosphonate alkynes had been explored and attempts had been made to understand how addition order impacted the selectivity of the resulting alkynes. This could have been explored further, and in doing so a reliable means of producing asymmetrically substituted phosphonates may be found that are of use for other synthetic purposes.

A novel benzyl-ethyl alkyne phosphonate had been successfully incorporated into an amino acid (**Scheme 17**), and a peptide sequence (**Scheme 18**) successfully.

Unfortunately, benzyl-ethyl functionality presented no immediate use beyond the initial use as a precursor to FP functionality via DAST. Apprehension and scepticism over the selectivity of generating FP functionality with DAST on a peptide sequence with hydroxyl groups, coupled with the substrate selection limitations that would be presented by avoiding peptides containing hydroxyls (serine, threonine and tyrosine) resulted in the abandonment of this route.

There may have been a route to produce a molecule analogous to the Cravatt probe (**Scheme 29**) (179) by using a phenyl-ethyl phosphonate alkyne as a precursor and achieving fluorophosphonate functionality with a source of fluoride, circumventing the need for DAST, for example, d'Andrea utilised ammonium fluoride (180).



Scheme 29: Conversion of phenyl-ethyl pTza peptide **85** to a fluorophosphonate containing pTza **71** with analogy to the Cravatt fluorophosphonate (179), using ammonium fluoride, inspired by pathway described by d'Andrea (180)

This route was hampered by the requirement for silyl deprotection of the TIPS acetylene which was reliant on sources of fluoride. Sources of fluoride used on a TIPS-protected phenyl-ethyl

ethynyl phosphonate would have resulted in the production of a potentially dangerous volatile fluorophosphonate which was a major safety concern. Additionally, premature generation of the fluorophosphonate, would have produced an unstable and reactive compound which would not have been of use for incorporation into peptide sequences.

Production of a diphenyl ethynyl phosphonate by direct reaction of ethynyl magnesium bromide with diphenyl chlorophosphate, (**Scheme 26**), was a promising route to obtain diphenyl ethynyl phosphonate. Unfortunately this route did not produce the desired compound in analogy to route producing diethyl ethynyl phosphonate (154). Despite the existing literature precedent for production of the diethyl ethynyl phosphonate by these means, there was no success producing the product.

A putative route for conversion of diethyl ethynyl phosphonate to phenyl-ethyl ethynyl phosphonate using Tf_2O and pyridine (**Scheme 27**), showed strong literature precedence (192). There was evidence that phenyl-ethyl phosphonates served the same function as fluorophosphonates when used in the capture of serine hydrolases (193) and hence may be just as useful as electrophilic capture probes in pHis-partner interaction assays. However, this route was less selective for the mixed phosphonate than displayed in the literature (192). Provided the diethyl ethynyl phosphonate was produced in sufficiently high yields, this route may lead to sufficient quantities of alkyne that could be isolated despite the lack of selectivity and potentially implemented into pTza analogues of pHis.

Once produced, a phenyl-ethyl ethynyl/ diphenyl phosphonate could, by Cu-catalysed click, be inserted into an Fmoc-azidoalanine-OH to produce a pTza analogue for insertion into a peptide or the alkyne could be directly inserted into a peptide by cycloaddition, this work demonstrates methods with a peptide substrate and the benzyl-ethyl ethynyl phosphonate; this method was also used by Hunter *et al.* to generate immunogenic peptide substrates (25).

4 Photoreactive peptides as investigative tools

Given the challenges encountered on generating fluorophosphonate containing pTza analogues of phosphohistidine (pHis) outlined in Chapter 3, an alternate strategy for studying phosphoprotein interactions was investigated.

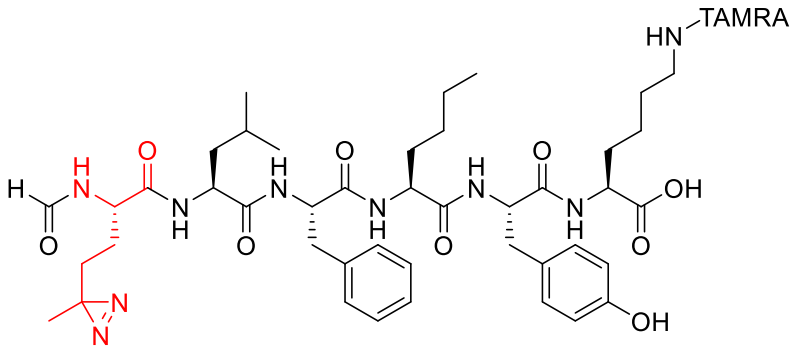
The initial aims of this work were to identify enzymes that interacted with pHis-containing proteins by creating probes with reactive analogues of pHis (pyrazole or triazole based). Attempts to incorporate pyrazole into proteins was unsuccessful and attempts to generate reactivity on the phosphonate of the triazole analogue were largely unsuccessful due to instability or incompatibility of one aspect of probe synthesis with another. However, generating reactivity directly on the pHis analogue was only one approach; methods were sought out for covalently binding interactor proteins to probes by creating reactivity elsewhere in a probe structure.

4.1 Inspiration for photoactivation approach

The new approach was based on the use of diazirine photoaffinity labelling (PAL) to capture reactive partner proteins. PAL is a common method for the covalent binding of transient interactors of proteins. There are multiple photoactivatable groups (benzophenone, aryl azide and diazirine); diazirine have been reported as a highly specific photoactivatable group with room temperature stability requiring short durations and shorter wavelengths for irradiation than alternatives (194). The advantages and capabilities of diazirine crosslinking is exemplified by Yang *et al.* who used diazirine-functionalised peptides to identify 'reader' and 'eraser' enzymes capable of post-translational modifying histone H3 (195). Lin *et al.* likewise explore histone H3 interactors and develop a tri-functional amino acid with a diazirine functional group enabling photo-activated crosslinking to interactors as well as a cleavable linker (disulphide group) to reduce the complexity of MS analysis; showing the versatility of diazirine functionalised peptide probes (196).

Work performed adjacent to the phosphohistidine investigation inspired this approach. This work focussed on producing an established fluorescent labelling protein **86** for the Formyl peptide receptor 1 (FPR1). FPR1 are expressed by leukocytes and play a key role in the regulation of the host cell immune response (197), recognising the bacterial gene coding for N-formyl peptides during bacterial and mitochondrial protein synthesis (198, 199). Therefore, there is interest in developing diagnostic ligands that can bind to FPR1, a G-protein coupled receptor (GPCR), for investigation into their role in immune response modulation (200).

The fluorescent peptide agonist of FPR1 synthesised is displayed in **Figure 47** and has been reported previously (200).



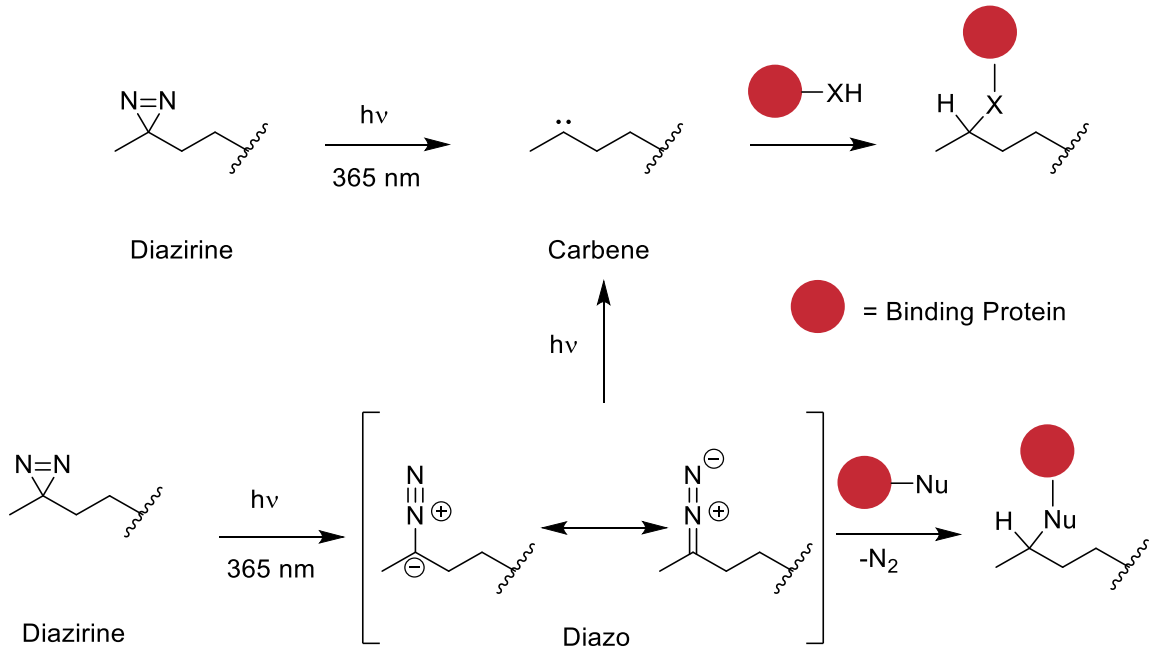
86

Figure 47: Structure of FPL1 agonist peptide **86** containing a diazirine moiety (highlighted in red) and TAMRA fluorescent dye for visualisation on SDS-PAGE.

The FPL1 agonist peptide contains a diazirine moiety which enables crosslinking to interacting peptides upon exposure to UV radiation this is outlined in **Scheme 30**. For the agonist peptide this enables binding and subsequent visualisation of bound FPL1 protein for investigating the role of GPCRs in immune regulation.

The FPL1 photoprobe was recognised by FPR1 even in competition with known binding ligands (200); ligand binding was not significantly reduced by the presence of a diazirine group. Whilst it does need to be considered that FPR1 has a broad tolerance for bacterial/mitochondrial sequence variety provided the peptide substrate is formylated (198); the potential for a receptor to bind a peptide containing a diazirine residue with specificity inspired the idea of attempting this approach with phosphopeptides and their corresponding protein regulators.

To enable this approach, the overall aim would be synthesis of peptides containing the well-established Fmoc-pTza analogue (159) and reactivity would be provided by the photoactivatable groups (PAG), diazirine for photocrosslinking. The diazirine covalent modification has a track-record of application in the analysis of cell lysates (201) the irradiation of the photoreactive diazirine group covalently captures target proteins as outlined in **Scheme 30**. The method also requires a probe substrate to be equipped with a bio-orthogonal handle such as an alkyne for copper-click mediated addition of a dye or biotin-azide tag (202) for downstream analysis by either SDS-PAGE or mass spectrometry (MS) respectively.



Scheme 30: General overview of diazirine crosslinking with peptide/proteins. The diazirine containing peptide or protein forms a reactive carbene/diazo compound when irradiated with UV light. The reactive carbene/diazo moiety can react with sites on a proximal partner protein (194, 203).

Additionally, the analysis of protein mass spectrometry data obtained by photo crosslinkers can be challenging due to complex crosslinking patterns. One method used to improve analysis is the incorporation of cleavable crosslinkers to peptide substrates (**Figure 48**). Müller *et al.* report synthesis of a urea-based cleavable linker used for identification of a peptide substrate based on "fingerprint" mass shifts produced by dissociation and modification of the cleavable linker (204). Korovesis *et al.* provide insight into a variety of cleavable linkers fragmentation patterns in tandem mass spectrometry, they assessed the fragmentation patterns of sulfoxide and urea-based cleavable linkers that are cleaved by collision-induced dissociation. Azobenzene, disulphide and 2-acyl-dimedone linkers were also tested. They proceed to test cleavable linkers in conjunction with photoaffinity labelling (diazirine groups), producing diazirine peptides with cleavable disulphide and 2-acyl-dimedone groups. These peptides were used to investigate site specific interactions between melittin (disulfide-diazirine probe) and calmodulin as well as glucagon (dimedone-diazirine probe) and GLP-1R. They demonstrate the ability to site-specifically identify binding sites using cleavable photo-crosslinkers (205).

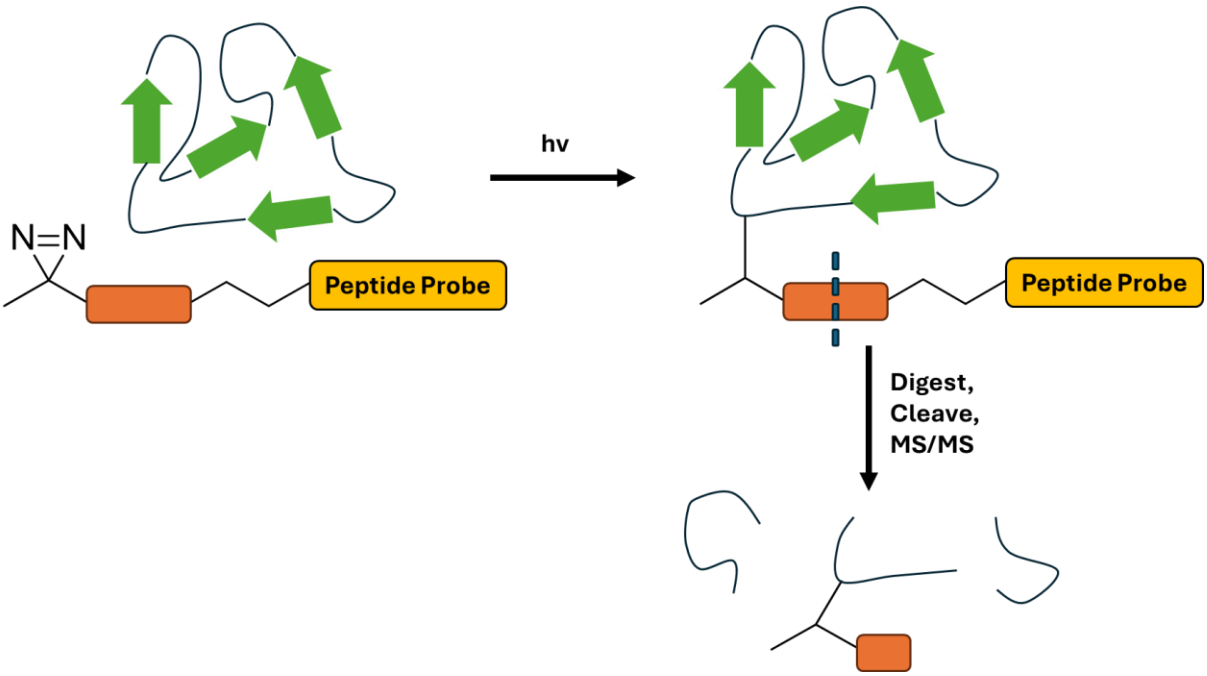


Figure 48: Overview of diazirine crosslinking with cleavable crosslinker. UV irradiation enables covalent binding through diazirine photo-crosslinker and subsequent digest and MS/MS analysis provides peptide segments with cleavable linker fragments for site identification (205).

Since no well-characterised phosphohistidine-containing peptide/protein interactions have been reported, creating pTza-containing peptides as analogues of pHis peptides straightaway and applying these as test substrates would present a challenge. For example, if no binding was observed on SDS-PAGE would that indicate the absence of peptide/protein interactions or an issue with the methodology? It seemed prudent to test the crosslinking methodology on well-established systems to refine the crosslinking methodology before application to pHis-based systems.

4.2 EGFR and Grb2

A known phospho-amino acid binding interaction between the SH2 (Src homology 2) domain of growth factor receptor bound protein 2 (Grb2) (a phosphotyrosine (pY) binding domain) and portions of the epidermal growth factor receptor (EGFR) cytoplasmic tail, provide a suitable model for testing photoactivatable crosslinkers in a phosphoprotein context.

EGFR dimerises and undergoes autophosphorylation on tyrosine residues as a response to extracellular stimuli, resulting in the binding of adaptor protein Grb2 (206, 207). The Grb2-SH2 domain recognises the phosphotyrosine sequence on EGFR enabling binding when this signal is produced. Grb2 also contains SH3 binding domains that recognise proline rich residues, such as those in Son of Sevenless (SOS) – a guanine nucleotide exchange factor (GEF) which activates Ras, another downstream protein (206-208). These interactions are the initial steps of the well-documented mitogen activated protein kinase (MAPK) pathway (**Figure 49**).

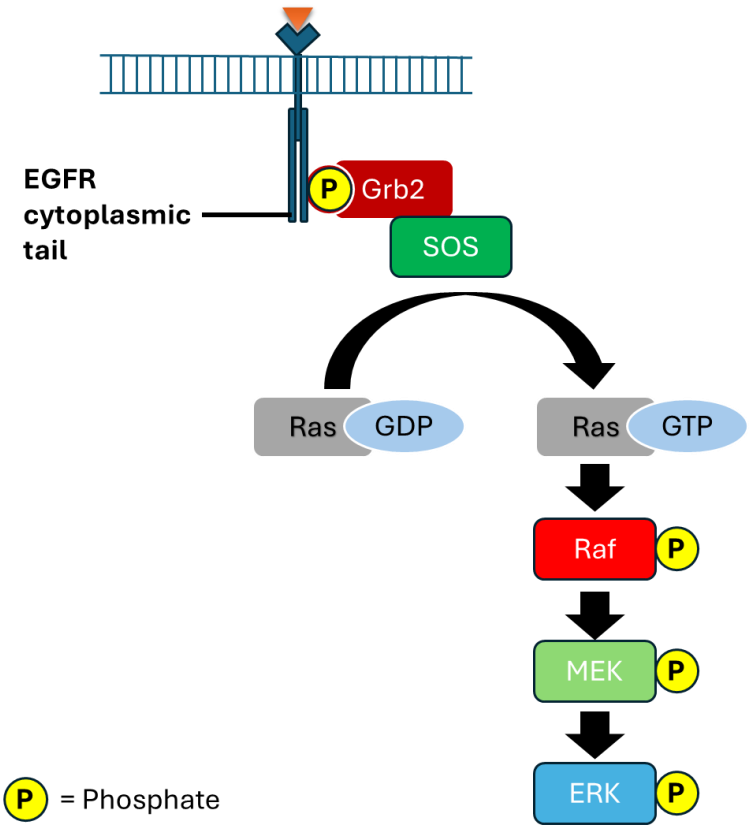


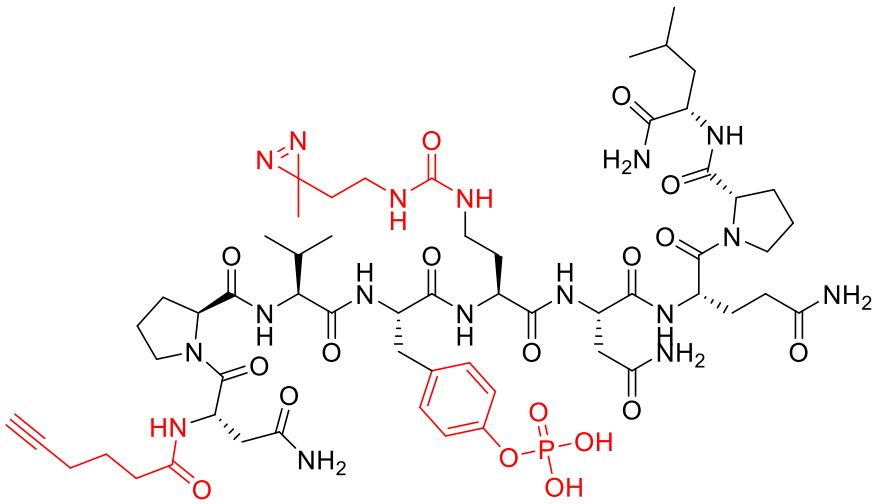
Figure 49: MAPK cell signalling pathway overview. Phosphorylated EGFR dimer binds Grb2-SH2 which subsequently interacts with SOS. SOS promotes Ras GDP-GTP exchange thereby enabling the signalling cascade relaying phosphate to partner proteins: Raf, MEK, ERK. (209, 210)

The crux of EGFR-Grb2 binding is the affinity of the Grb2-SH2 domain for specific phosphotyrosine containing sequences in EGFR. Grb2 binds preferentially at Y1068 and to a lesser extent to Y1086 based on EGFR binding attenuation in the presence of phosphopeptides containing pY1068 and pY1086 (211, 212). Additionally, use of phosphopeptides has shown the ability for pY1114 in the EGFR C-terminal tail to bind Grb2 (212). Grb2 has also been shown to bind indirectly to EGFR via Shc (signalling adaptor proteins) which binds at two separate tyrosine sites (Y1173 and Y992) (211). Jadwin *et al.* describe a significant pYXN motif in sequences binding Grb2 (where X represents tolerance for many different residues) (213), which provides opportune placement for a photoactivatable group such as a diazirine without disrupting the natural binding capability of a peptide sequence for Grb2-SH2. Aside from the tolerance of a recognised motif and the well documented binding of EGFR to Grb2, the dependence of the binding of these two species on phosphotyrosine is interesting because of the cross reactivity of pTyr interactors (pY antibodies) with pHis (214). Researchers have previously shown that substitution of pTza for pY, in a pYVN sequence (pTzaVN), does not eliminate the binding interaction (155). Therefore, the EGFR-Grb2 interaction would be an ideal system to test pTza probe substrates for their ability to bind the same interactors as the phosphorylated residue, prior to making pHis sequence probes.

4.3 Investigating binding of Grb2-SH2 to an EGFR phosphopeptide

The interaction between phosphotyrosine-containing peptides and Grb2-SH2 is well-characterised (215). Therefore, testing the binding between pY-diazirine peptides and Grb2-SH2, would provide a good baseline test for the methodology and also provide a novel method of confirmation for the otherwise well-characterised binding of EGFR and Grb2.

An EGFR phosphopeptide probe (synthesised by Nicholas Morris) (**Figure 50**) was designed around the pY1086 binding site, given the tolerance for varied amino acids at the X position in the pYXN motif, the phosphopeptide was designed with a diazirine photocrosslinker in this tolerant position, enabling crosslinking functionality whilst avoid binding disruption. An alkyne handle was introduced at the N-terminus to allow Cu-catalysed cycloaddition of TAMRA-N_3 and biotin-N_3 for fluorescent visualisation and enrichment for proteomics workflows respectively. Additionally, the diazirine functionality is linked to the photoprobe by a urea-based cleavable linker. The cleavable linker may reduce the complexity of proteomic analysis by providing a "tag" to search for during data analysis as a specific modification to covalently linked proteins.



87

Figure 50: EGFR peptide sequence (Alkyne-NPVpYXNQPL-NH_2) **87**. Highlighted in red from left to right: The probe contains an N-terminal alkyne for Cu-catalysed crosslinking to a fluorescent dye, TAMRA-N_3 in subsequent reactions; Phosphotyrosine as a key motif for binding to Grb2-SH2 and a urea-based cleavable photocrosslinker (diazirine) for binding to protein partners when incubated with isolated protein or cell lysate. (Synthesised by Nicholas Morris)

As such, the Grb2-SH2 domain as a His-tagged fusion (155) was incubated with a peptide probe of the pY-EGFR sequence as a proof of concept for the diazirine probes detecting a known binding (in this case EGFR-Grb2).

EGFR probe **87** was incubated at various concentrations (0.5-5 μM) with isolated and purified Grb2-SH2 protein at a constant concentration (5 μM). This was an initial test to determine what concentration of probe was required to see any binding activity. Alongside the EGFR probe test samples were two DMSO controls containing no EGFR peptide.

In each case the samples were incubated on a shaker-incubator and subsequently irradiated with UV light to activate reaction of the photo-reactive diazirine group to the incubated protein partner (Grb2-SH2). The probe was then modified using a click reaction with a fluorescent dye containing an azide group, in this case TAMRA-N_3 was used, enabling visualisation of the cross-linked protein on a gel after SDS-PAGE with the presence of a fluorescent rhodamine stain. **Figure 51** shows the resulting SDS-PAGE gel from the incubation of EGFR probe with Grb2.

Illumination at 5 μM and 2.5 μM of EGFR photoprobe showed that there was binding of the probe with isolated Grb2 which was promising. There was a reduction in illumination intensity as the concentration of EGFR photoprobe decreased, which was expected given that low concentration samples would make less crosslinks to Grb2 during irradiation. The lack of Coomassie staining in **Figure 51** was unfortunate, because there was no loading control to assess the fluorescence against. While the illumination follows the expected trend, the lack of validation with Coomassie staining means there is no confirmation that the trend is affinity-based rather than an issue with sample loading. Additionally, the observed presence of dual banding was due to curtailed segments from the Grb2 vector.

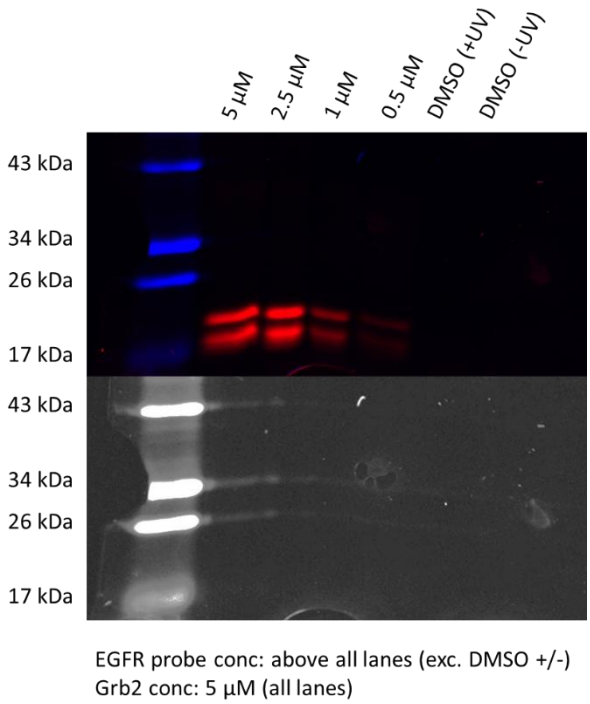
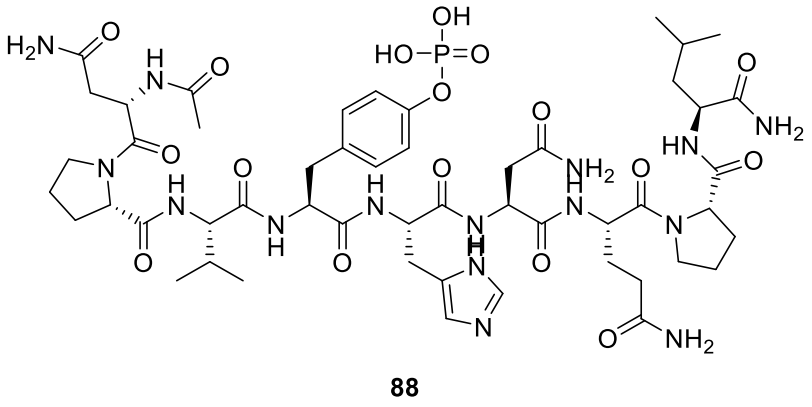


Figure 51: Rhodamine-coomassie composite gel. Rhodamine gel displaying fluorescent bands corresponding with bands of curtailed Grb2-SH2 protein. Red bands from left to right correspond with EGFR probe concentrations of 5 μM, 2.5 μM, 1 μM and 0.5 μM, Grb2-SH2 present at 5 μM in all lanes (top). Coomassie gel shows no bands for protein loading (bottom).

Figure 51 demonstrated potential for the EGFR probe **87** to bind to Grb2-SH2. In order to validate this binding, competitor testing was undertaken, if the observed binding interactions in **Figure 51** were legitimate a comparable peptide sequence to the photoprobe would be capable of attenuating or eliminating the observed fluorescence.

Therefore, an analogous peptide of the EGFR probe was synthesised bearing a phosphotyrosine (pY) amino acid but no photocrosslinker **88** (**Figure 52**).



88

Figure 52: Structure of the EGFR competitor sequence (Ac-NPVpYHNQPL-NH_2) **88**

Incubation experiments combining the EGFR photoprobe **87** with Grb2-SH2 were repeated but adding varying quantities of pY competitor **88** to assess the binding affinity. It was expected that increased quantities of pY competitor would result in lower binding affinity between the photoprobe and Grb2-SH2, demonstrated by lower levels of illumination on a rhodamine gel.

Given that similar illuminations were observed between 5 μM and 2.5 μM of EGFR photoprobe concentration the initial gel in **Figure 51**; and in the interest of conserving the photoprobe, the next reaction was performed with a constant photoprobe concentration of 2.5 μM and the competitor pY probe **88** was introduced at varying concentrations to see if it would have the intended effect and compete off the EGFR photoprobe. **Figure 53** shows the resulting SDS-PAGE gel from this competition assay.

Clearly there was some binding interaction between the EGFR photoprobe **87** and the Grb2 protein; and this binding was disrupted by the presence of an analogous pY-containing peptide sequence **88**, which supports the specificity of the probe for binding. One unexpected aspect of this experiment, however, was the sharp drop off in binding of the photoprobe; the binding would be expected to slowly decline with increased concentrations of competitor probe **88**. A plausible explanation may well be that the concentration difference between the first two lanes was too large, and the critical concentration lay somewhere between 0.5 – 1 μM.

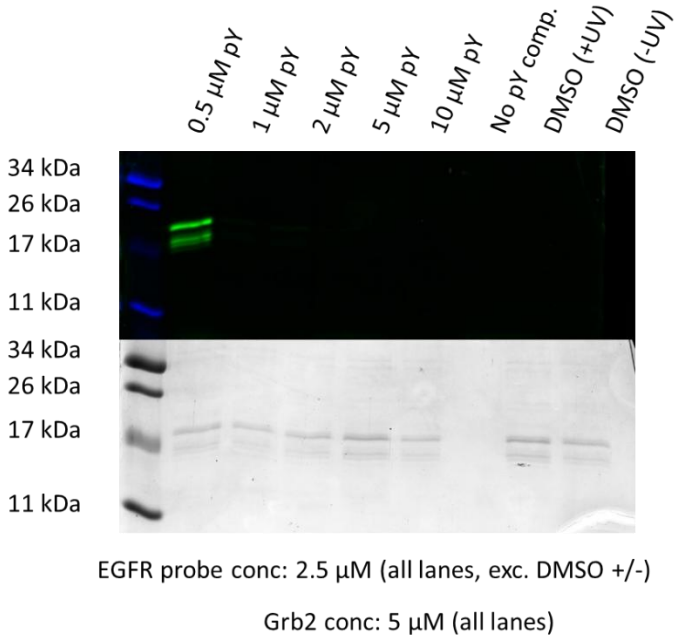


Figure 53: Rhodamine-coomassie composite gel image, displaying standard protein ladder and lowest concentration of pY competitor (0.5 µM) (top). Coomassie gel image, displaying the loaded protein without rhodamine stain (unbound) (bottom). The concentration of competitor pY increases along the lanes from left to right (0.5 µM – 10 µM).

The experiment was repeated with a narrower range of competitor concentrations, to observe the gradual reduction in binding between the EGFR probe **87** and Grb2 and eliminate the sharp drop off in rhodamine illumination observed in **Figure 53**.

Incubation of EGFR photoprobe **87** with Grb2 and pY competitor **88** (0.4 – 1 µM) was attempted (**Figure 54**). A steady reduction in luminescence was expected from the rhodamine gel with the increased concentration of pY competitor up to 1 µM, where little to no binding was expected to be seen.

Unfortunately, all the lanes showed some level of binding of EGFR photoprobe to Grb2-SH2 (**Figure 54**), even at 1 µM competitor concentration which contradicted the observation in the earlier experiment depicted in **Figure 51**, which had showed no rhodamine binding at a pY competitor concentration of 1 µM. There was also no clear trend from left to right; as the concentration of pY competitor increased from 0.4 to 1 µM, the level of rhodamine illumination detected was inconsistent. Based on gel (**Figure 54**) It would seem binding between the photoprobe and Grb2 was, in some cases, more prevalent in the presence of higher levels of competitor. In **Figure 54**, 1 µM competitor appeared to result in greater binding of the EGFR sequence, based on rhodamine illumination, than 0.5 µM competitor. The coomassie stained gel also seemed to suggest there was little discrepancy with protein loading to justify

inconsistent results. The assay was repeated on this concentration range again, but the results were just as sporadic and suggested that there was either an issue with the methodology or the samples themselves.

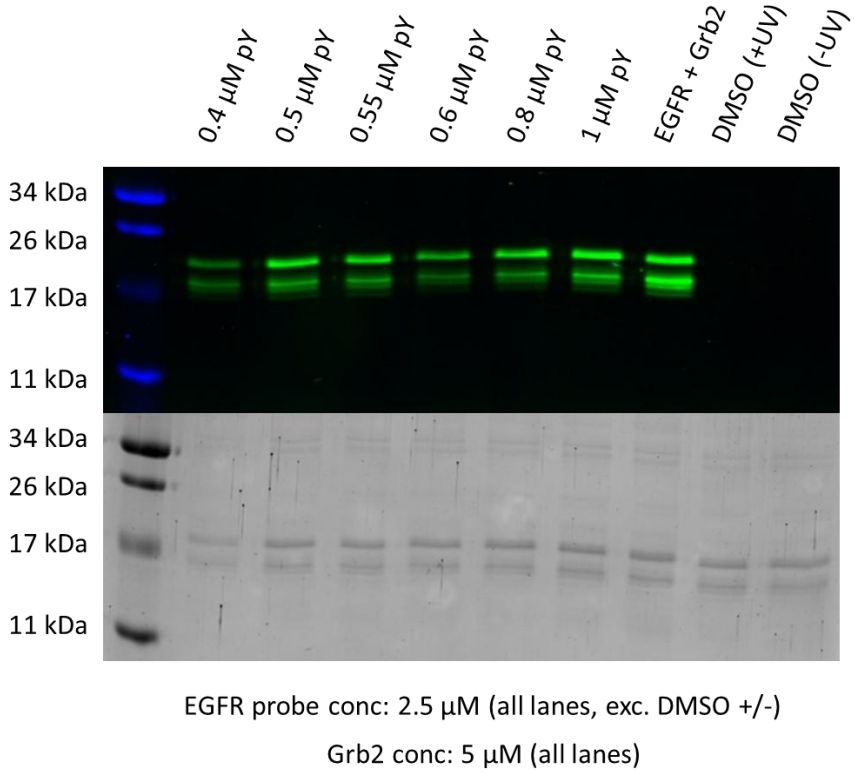


Figure 54: Rhodamine-coomassie composite gel displaying luminescent lanes for EGFR probe **87** to Grb2 interactions from left to right at increasing pY competitor **88** concentrations from 0.4 – 1 µM and a non-competition sample with just EGFR photoprobe (2.5 µM) and Grb2 (5 µM) for comparison (top). Coomassie gel displaying protein loading of Grb2 in each lane including the additional DMSO +/- control samples (bottom).

Clearly the photoprobe was capable of binding to the expressed protein based on rhodamine illumination. Concerns regarding the purity of the protein sample led to tests of the photoprobe in a different environment. The photoprobe pulldown experiments were repeated on a HeLa cell lysate. This provided another route to interrogate EGFR binding to Grb2-SH2 and potentially reveal other binding proteins for investigation. The methodology was much the same as with the isolated protein; EGFR photoprobe (2.5 µM) was incubated at a constant concentration with increasing concentrations of HeLa cell lysate, irradiated and a rhodamine tag was clicked onto the protein for imaging. Photoprobe-lysate incubation gel (**Figure 55**) showed protein binding was largely the same at all concentrations of HeLa lysate. There was a stronger binding for higher mass proteins, the intensity of the bands faded as the protein mass

diminished. The photoprobe didn't appear to be very selective but it did bind to some proteins in the HeLa lysate compared to the DMSO controls.

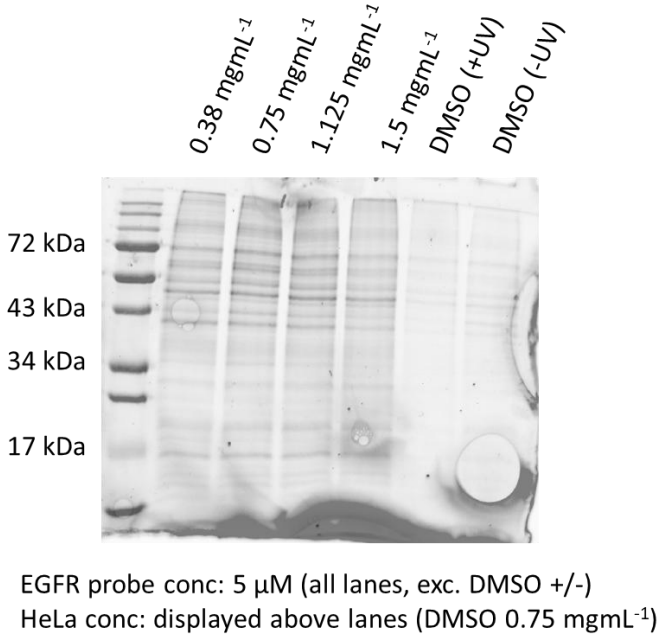


Figure 55: Rhodamine gel, displaying standard protein ladder and incubations of EGFR photoprobe with increasing concentrations of HeLa lysate (0.38 – 0.75 $mgmL^{-1}$) and two DMSO controls with irradiation (+) and without (-) respectively (from left to right).

Subsequent efforts were made to bind EGFR photoprobe **87** with HeLa lysate with increasing concentrations of photoprobe and repeat experiments where multiple lanes were produced on the same gel with equal concentrations of EGFR photoprobe and equal concentrations of HeLa cell lysate. All subsequent efforts suffered the same inconsistency as the previous assays with isolated protein.

It was very likely that the inconsistency in the gel data was due to instrumental issues. The probe lanes and DMSO control consistently showed a different binding intensity, if the probe was not binding to the isolated protein or proteins in the cell lysate then both the probe lanes and controls would look the same. Evidently, the EGFR probe **87** was binding but not with the trend/consistency expected on a concentration gradient. It was possible that if there were issues with the UV irradiation lamp, the samples would all have been irradiated inconsistently and therefore some samples will have formed more crosslinks than others. A low concentration of photoprobe irradiated for longer than a high concentration photoprobe may result in a similar

level of binding fluorescence on a gel for two distinct concentrations based on irradiation variance.

*Later mass spectrometric analysis indicated that the dual-bands observed on SDS-PAGE appear to originate from curtailments of the desired protein from the construct. While the EGFR probe **87** appeared to show binding interactions, it was possible that degradation compromised the binding affinity of **87** over time, hence the inconsistency in the observed data. Proteomic analysis would confirm binding in the presence of other proteins to the authentic Grb2 to support the observed binding on SDS-PAGE.*

.

4.3.1 Proteomic analysis of EGFR-Grb2 interaction

SDS-PAGE analysis did not give consistent binding data for the EGFR probe **87** and the isolated Grb2 protein interaction. Data from HeLa cell lysates appeared even more inconclusive than the gel data from EGFR probe **87** interacting with isolated Grb2. One option to explore the binding of **87** with Grb2 was MS analysis of a cell lysate, a proteomics approach. Given that the EGFR probe **87** contained an alkyne handle this enabled the option of using click chemistry to bind biotin azide for a pulldown as opposed to rhodamine azide for fluorescent gel visualisation.

The binding of some proteins in HeLa cell lysate as visualised in **Figure 55** showed that the photoprobe **87** had some affinity for proteins in the HeLa cell line that was worth exploring, particularly if the proteins were related to the EGFR-Grb2 signalling network.

The EGFR probe **87** was incubated with HeLa cell lysate at 2 μM and 10 μM concentrations of probe **87**. The samples were prepared in triplicate (i.e. three samples at 2 μM and three at 10 μM). After irradiation with UV light for 30 seconds to enable conversion of the diazirine to the reactive diazo intermediate required for crosslinking the probe **87** to interacting proteins in the HeLa lysate, the samples were biotinylated via click chemistry with biotin azide on to the alkyne handle of the probe. Biontinylation enabled enrichment and sample preparation prior to analysis by mass spectrometry. The biotinylated proteins were incubated with avidin beads, and the beads were washed to remove impurities. The proteins were then digested into peptides with trypsin and prepared in buffer suitable for MS analysis.

For proteomics analysis, the two samples at two concentrations were measured against a control sample (DMSO control). If the samples containing biotinylated probe peptide **87** showed a statistically significant uptick in a protein relative to the DMSO control, that protein is labelled. Volcano plot, **Figure 56,** illustrates the binding of lysate protein to probe **87** (2 μM).

Proteomic analysis was performed with data-independent acquisition (DIA-NN) standalone software enabling comparison of raw MS data with theoretical trypsin-digested ions from the human proteome database (Uniprot). Label-free quantification (LFQ) was generated by DIA-NN measuring relative intensities of precursor ions across samples. The probability of the protein observation being real between the samples and control gives the probability of observing the protein by random chance (p-value). The p-values are false-discovery rate (FDR) adjusted, i.e. an FDR-adjusted p-value of 0.01 would mean 1% of identified sites were false

positives. Therefore, a low FDR-adjusted p-value would correspond with a higher confidence in that observation. In the volcano plot, these FDR-adjusted p-values were transformed logarithmically ($-\log_{10}$(p-values)) such that low FDR-adjusted p-values appear higher on the y-axis of a volcano plot. The fold change on the x-axis of the volcano plot represents the magnitude of the difference in protein abundance between the sample and control. The proteins in the top right of the volcano plot were statistically significant (p-value < 0.05). They showed a presence greater than the threshold (fold change > 1) relative to the DMSO control, thus proteins observed more towards the top right of the volcano plot (above those thresholds) were more likely to be candidate interactors of the probe **87** based on these t-test conditions. (DIA-NN analysis and volcano plot produced by Dr. Brian Suarez Mantilla)

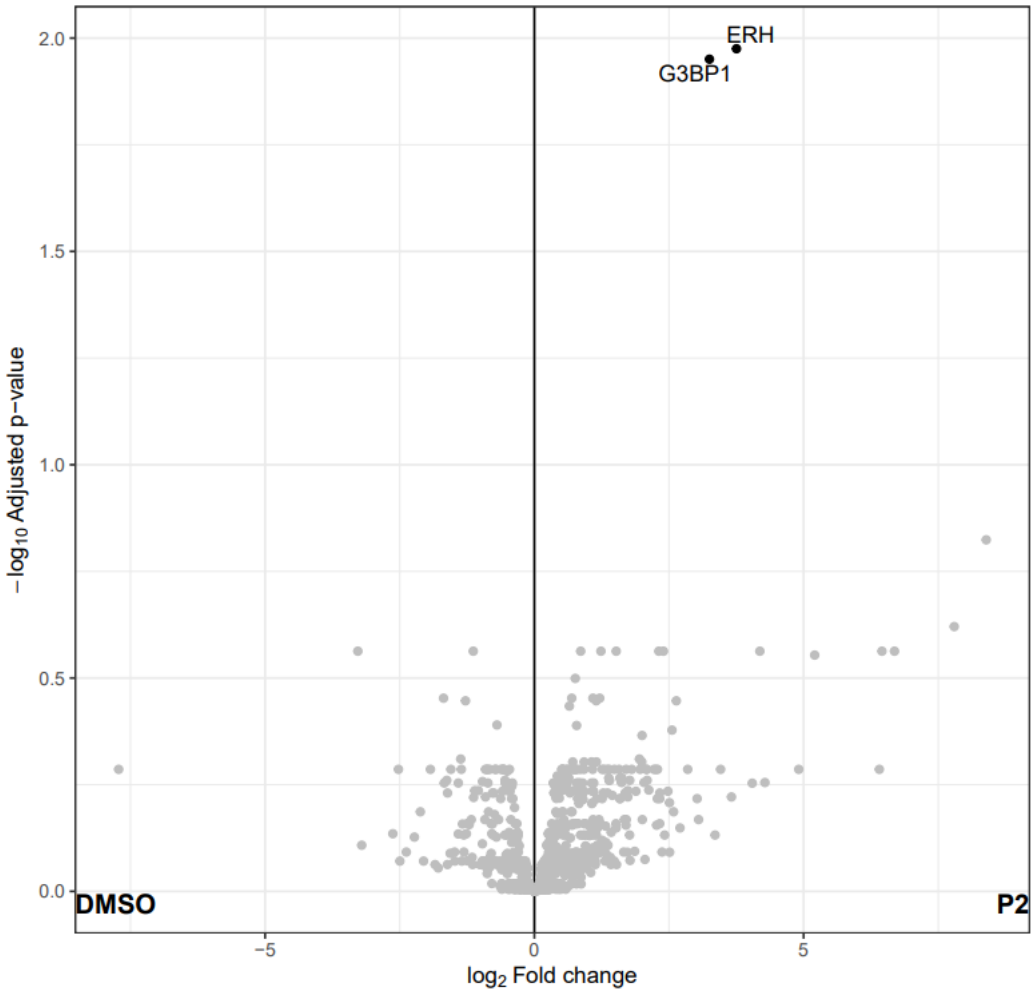


Figure 56: Volcano plot for EGFR probe **87** (2 μM) binding to HeLa cell lysate. Visual representation of t-test with the statistical significance of observed protein (p-value) plotted as $-\log_{10}$ (p-value) on the y-axis and the magnitude of the difference between the control and the sample (fold change) in the x-axis. Data based on label-free quantification of a trypsin digest via DIA-NN processing. (volcano plot produced by Dr. Brian Suarez Mantilla)

The two statistically significant proteins extracted were: enhancer of rudimentary homolog (ERH) and Ras GTPase-activating Src homology 3 binding protein 1 (G3BP1). ERH is a protein responsible for regulating the cell cycle and transcriptional inhibition as well as pyrimidine biosynthesis (RNA precursor synthesis) (216, 217). There is evidence to suggest that ERH in fruit flies (*Drosophila melanogaster*) is phosphorylated by casein kinase II (CKII) (218), whether ERH is phosphorylated in mammalian cells has not been established (219) but given the high degree of conserved sequence homology of the ERH protein across eukaryotic cells, it is likely that the protein serves a similar purpose across eukaryotes (220). The interaction of a phosphorylated EGFR probe **87** with ERH was not expected.

The other protein bound by the EGFR probe **87** was G3BP1 which is a protein that bind Ras GTPase-activating protein (RasGAP) through association at an SH3 domain (221). RasGAP regulate the activity of Ras by accelerating the conversion of the active GTP state of Ras to the inactive GDP state (222). G3BP1 is a protein that appears to co-precipitate with RasGAP, however studies suggest that G3BP1 does not regulate GTPase activity but has some implication in cell proliferation and metastasis (223). Ras is a downstream partner in the EGF signalling pathway, as such, there is an indirect association between EGFR and downstream Ras interactors (such as G3BP1) (210) but a direct binding between EGFR and a binding protein of a Ras regulator would not be expected.

In proteomic analysis of EGFR 10 μM samples, both ERH and G3BP1 were detected but an additional protein also showed statistically significant binding, FKBP1A (**Figure 57**)

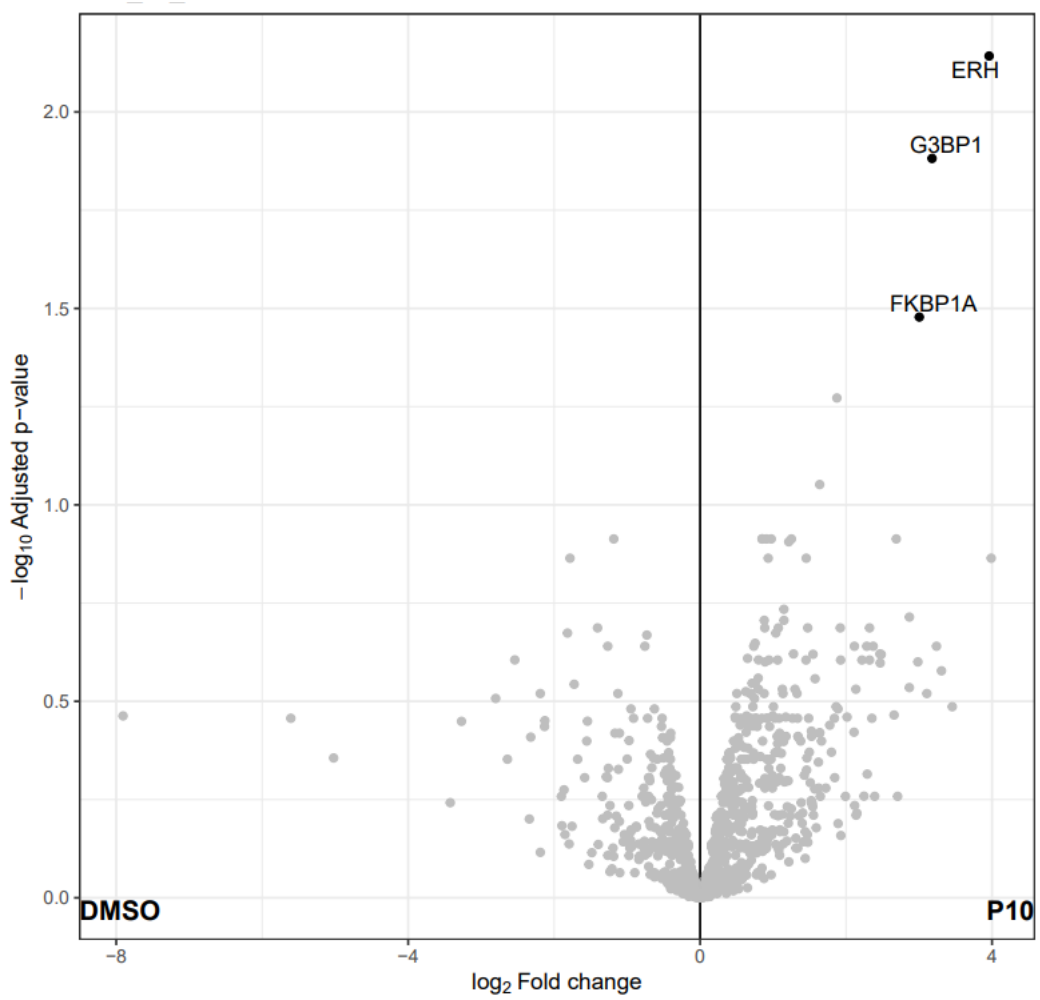


Figure 57: Volcano plot for EGFR probe **87** (10 μM) binding to HeLa cell lysate. Visual representation of t-test with the statistical significance of observed protein (p-value) plotted as -$\log_{10}$ (p-value) on the y-axis and the magnitude of the difference between the control and the sample (fold change) in the x-axis. Data based on label-free quantification of a trypsin digest via DIA-NN processing. (volcano plot produced by Dr. Brian Suarez Mantilla)

FKBP1A is a small protein (~11 kDa) of the FK506-binding proteins family (224). FKBP1 is a cis-trans peptidyl prolyl isomerase (PPIase) and has roles in immunoregulation, signal transduction modulation and protein conformation control (225). As a result of widespread involvement in a broad spectrum of processes, proteins in the FKBP family have been implicated in cancer (224). The function of FKBP1A as a prolyl isomerase and its abundance in the cell may well explain the binding interaction between the EGFR probe **87** as a proline rich peptide substrate and FKBP1A.

Additionally, there is literature precedent for an observed interaction between EGFR and FKBP1A in A431 cells. EGFR phosphorylation regulation is reputedly driven by modulation of tyrosine kinase activity by FKBP1A (226), but other modes of action may exist such as phosphotyrosine phosphatase inhibition by FKBP1A (227). EGFR autophosphorylation

suppression is reduced in the presence of rapamycin and FK506, which binds the active site of FKBP1A and thus inhibits PPIase activity (226, 227). Therefore, the PPIase activity of FKBP1A may play some role in EGFR phosphorylation modulation through direct binding to EGFR (226).

EGFR probe **87** when incubated with a HeLa cell lysate at 2 μM and 10 μM indicated binding to ERH, G3BP1 and FKBP1A proteins, in the absence of well-known binding proteins such as Grb2 or SOS which would have been expected to demonstrate binding to an EGFR mimic substrate. Isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR) studies suggested that the K_d of an EGFR phosphopeptide binding Grb2 is 400 nM (206). Another study determined the K_d of EGFR for Grb2 to be 97, 340 and 650 nM of EGFR for 1, 10 and 100 nM of Grb2 respectively (228). Evidently, the EGFR-Grb2 binding interaction is a relatively high affinity nanomolar interaction; as such, it was curious that this proteomics work did not observe any binding of an EGFR mimic **87** for Grb2. Grb2 and SOS were not observed in the dataset, including in the control samples, the absence of Grb2 from the identified proteins may have indicated an issue with the cell sample rather than the probe itself. It would have been possible to identify Grb2 if it was not present in the original cell sample, or it was present in low abundance. The lack of EGFR binding to membrane bound proteins and the high affinity in this binding data for cytosolic proteins was unexpected. In a HeLa quality control sample, Grb2 was only discovered by one peptide (data not shown). Despite the difference in preparation, the quality control sample was provided by a manufacturer whereas the proteomics HeLa was prepared in-house, it may be speculated that Grb2 is present in low abundance in the tested HeLa lysate.

4.4 Investigating PPDK to RP binding

A different system was sought out to test the methodology: the well-documented binding between pyruvate orthophosphate dikinase (PPDK) and its regulatory protein (PDRP) (229-231). PPDK is a photosynthetic enzyme, responsible for modulating the conversion of pyruvate to phosphoenolpyruvate (PEP) (231). PEP is utilised for carbon fixation in the C_4 photosynthetic cycle, as it is converted to oxaloacetate by phosphoenolpyruvate carboxylase (PEPC) and CO_2. Oxaloacetate is converted to malate, which is transported from the mesophyll cells to bundle sheath cells where it is decarboxylated to provide CO_2 for the Calvin cycle, resulting in glucose formation as an energy source for the C_4 plant (**Figure 58**) (232).

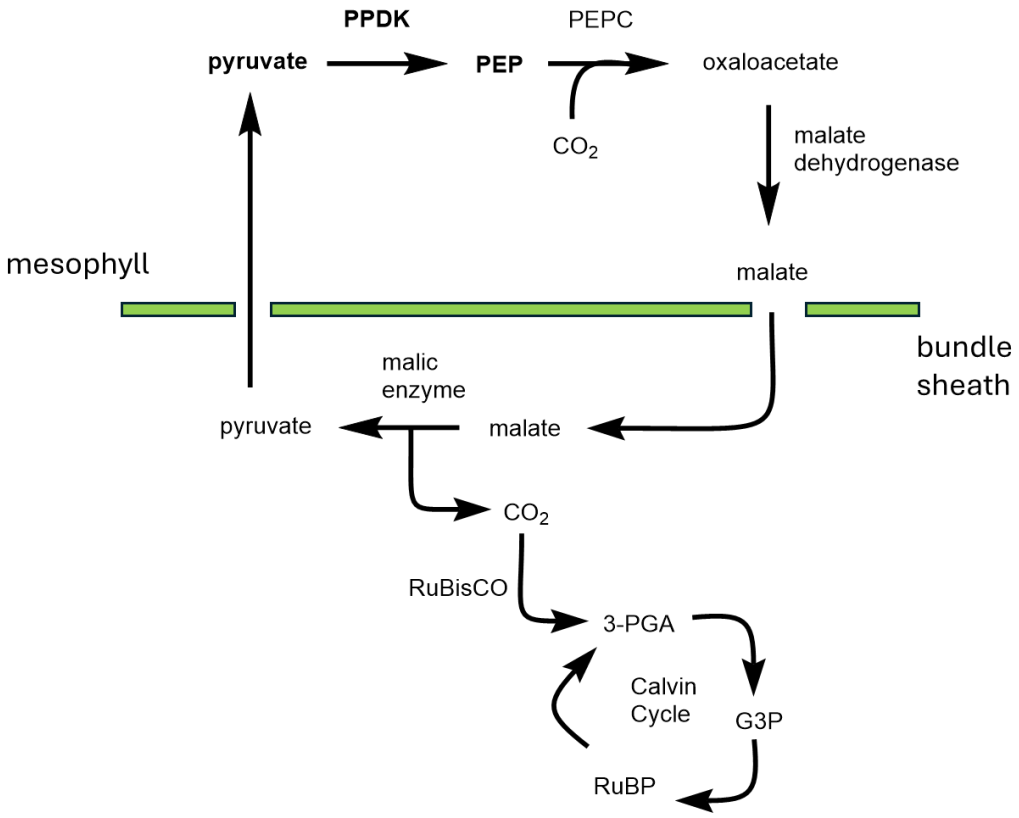


Figure 58: C_4 cycle overview, PPDK plays a role in the energy generation cycle by catalysing the production of PEP from pyruvate in the mesophyll cells which in turn produces key intermediates needed for CO_2 provision in the Calvin cycle in bundle sheath cells (232).

PPDK contains a conserved threonine residue which is phosphorylated and dephosphorylated by RP (**Figure 59**). PPDK was of particular interest not only because of its well-documented phosphothreonine regulation pathway, but also the presence of a proximal phosphohistidine residue in the catalytic domain (233).

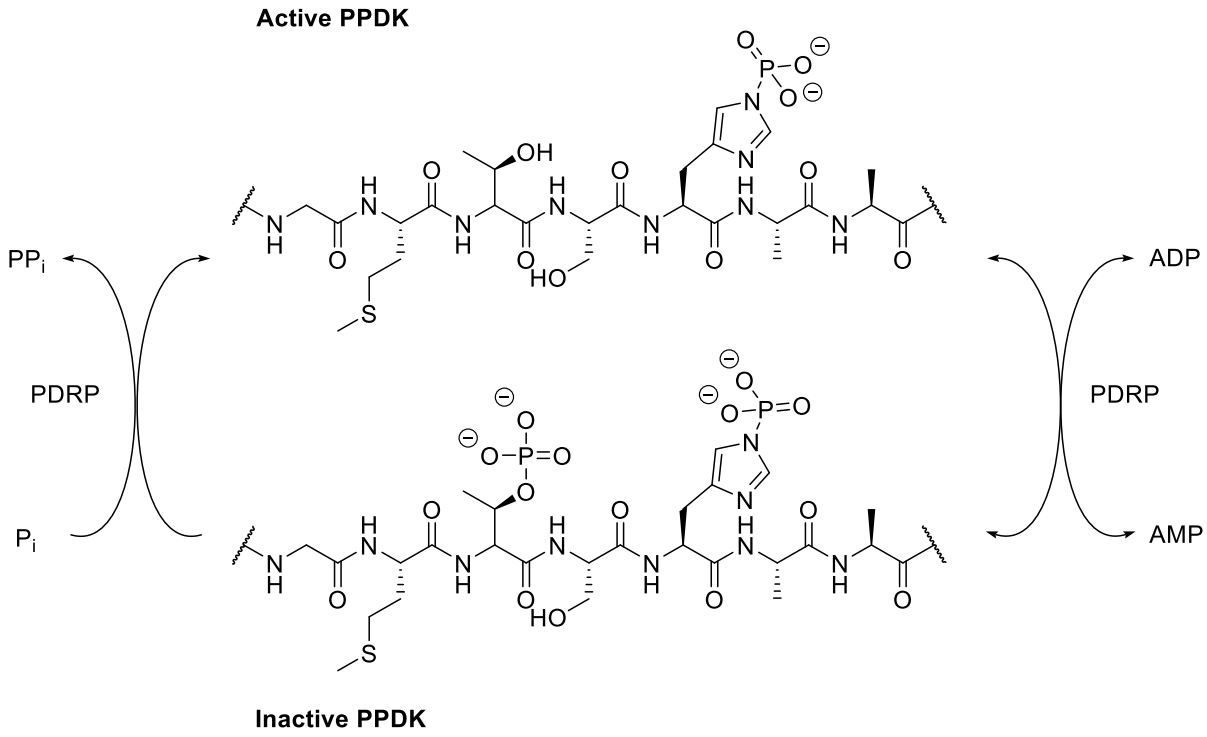
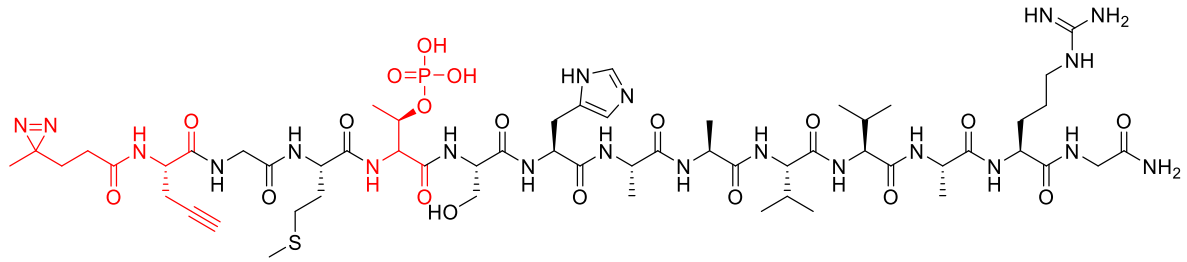


Figure 59: PDRP regulation cycle, PDRP utilises ADP to phosphorylate threonine (inactive form) and catalyses the phosphotransfer reaction shuttling phosphate from phosphothreonine to generate pyrophosphate (active form). Peptide based on a portion of the catalytic conserved domain (GMTSHA). Adapted from figure by Chastain *et al.* (234)

A PPDK photoprobe design based around phosphothreonine would provide a novel route for demonstrating PPDK-PDRP binding, if the methodology were to be optimised on phosphothreonine based probes, later investigation with phosphohistidine probes would be feasible. Given that phosphothreonine was a relatively stable post-translational modification and there were commercially available Fmoc-amino acid building blocks for phosphothreonine, these probes were thought to be more accessible than phosphohistidine equivalents. As with investigation of EGFR-Grb2, production of a PPDK sequence utilising Fmoc-pTza to install a phosphohistidine analogue would be feasible, but any identified interaction would be speculative without some methodological control (e.g. identifying known bindings with a PPDK photoprobe) before asserting novel observations are representative. Therefore, phosphothreonine probes were prioritised to assess the ability of a photoprobe peptide of PPDK to bind PDRP.

The synthetic photo-probe peptide **89** of the PPDK conserved catalytic domain for investigating PPDK-RP binding (**Figure 60**) was synthesised by Fmoc-SPPS (peptide synthesiser) on a rink amide MBHA resin. Phosphothreonine was installed as an Fmoc-protected amino acid with monobenzyl side chain protection. Fmoc-propargylglycine was

installed at the N-terminus of the peptide for convenient placement of an alkyne handle which would be used for Cu-catalysed click with TAMRA-N_3 for gel visualisation in SDS-PAGE. Fmoc-propargyl glycine was then manually deprotected and coupled with a diazirine derivatised from levulinic acid. The alkyne and diazirine were installed for synthetic convenience and to minimise interruption to the core sequence of the peptide. The peptide was cleaved from the rink amide resin with a standard cleavage cocktail with TFA: EDT: H_2O: TIS (94%: 2.5%: 2.5%: 1%) and after ethereal washing was lyophilised to provide 16 µmol, 16% yield.



89

Figure 60: PPDK sequence peptide photoprobe **89** based on PPDK1_MAIZE (G525-R535) catalytic domain residues including T527. Highlighted in red are the photoreactive diazirine, alkyne handle and phosphothreonine (T527) groups (from left to right). The probe also contains the catalytic histidine residue (H529).

In the initial assay (**Figure 61**), samples of the PPDK peptide were incubated with RP (Provided by Z. Ling), irradiated with UV light and derivatised with TAMRA-azide for rhodamine imaging, exactly as had been done with the EGFR-Grb2 system.

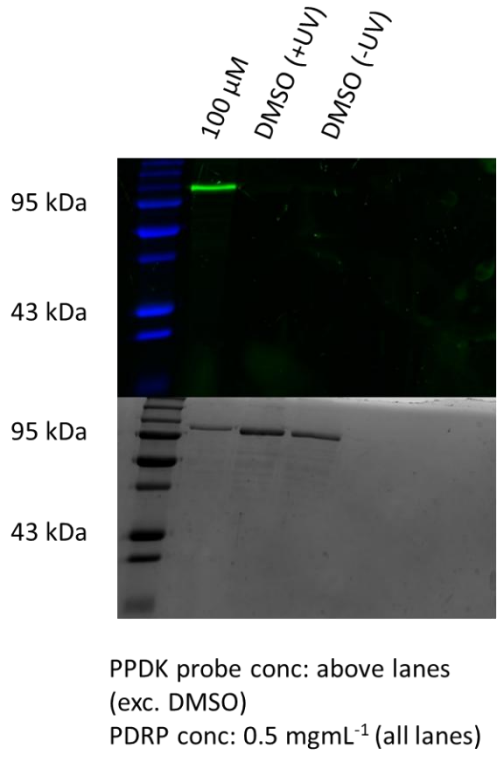


Figure 61: Rhodamine-stained gel (top), displaying standard protein ladder, PPDK photoprobe-RP binding and two DMSO controls, with and without irradiation. A clear luminescent band corresponds to the lane for PPDK-RP and no bands are present for the DMSO control lanes. Coomassie-stained gel (bottom) displays protein loading of RP in each lane (one PPDK-RP and two DMSO controls).

The presence of a luminescent band for the PPDK photoprobe binding to RP was initially promising, indicating binding of the PPDK mimic sequence to the regulatory protein. However, this needed to be optimised across a larger concentration range and also tested with subsequent repeats for consistency.

Much the same as with EGFR-Grb2 interaction, a lower concentration range was tested to assess the binding concentration of the PPDK photoprobe **89** for RP. A range of 0.1 - 2 µM PPDK photoprobe was tested against RP.

Given the observation of a band at 100 kDa in the initial attempt but no observed bands at a lower concentration range of 0.1 – 2 µM and the clear presence of enough protein in the coomassie gel, the crosslinking reaction was repeated with the PPDK peptide at an increased concentration range of 10 – 100 µM peptide (**Figure 62**).

Observation of a gradual increase in binding starting at little to no binding for a PPDK peptide concentration of 10 µM up to a strong binding represented by strong rhodamine illumination for PPDK 100 µM was expected, in keeping with the strong band seen in **Figure 61**. Clearly,

the corresponding coomassie gel shows inconsistent protein loading per lane, which would explain inconsistent fluorescence between lanes that is not in keeping with the expected illumination for each concentration. In any case, even the lanes with strong Coomassie-stained bands appear to have extremely weak fluorescent-peptide binding indicated by the rhodamine gel. Unfortunately, as with the EGFR-Grb2 system, results were inconsistent on successive repeats.

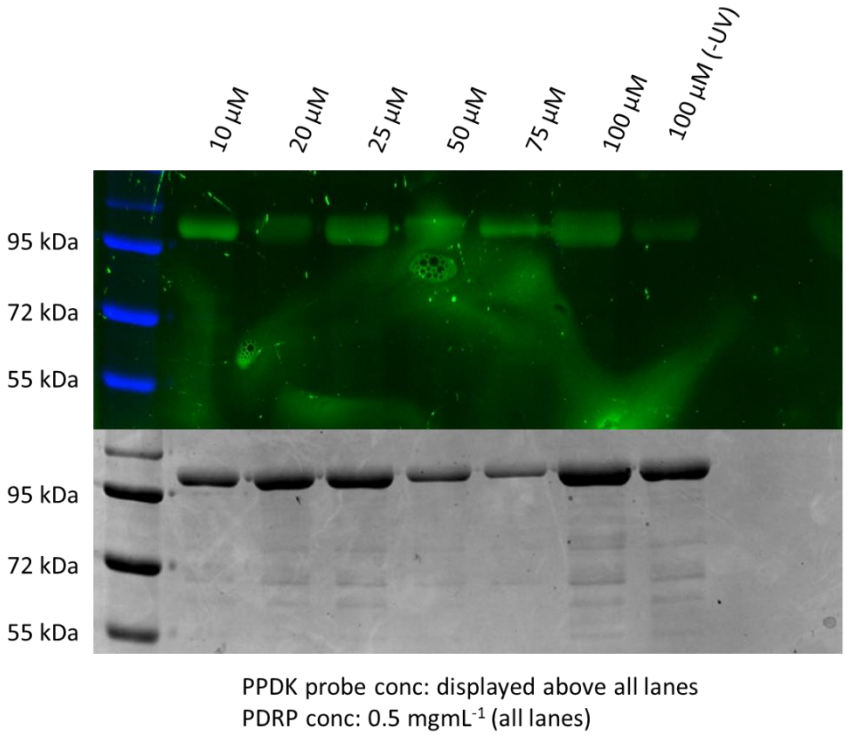


Figure 62: PPDK-RP binding assay (PPDK photoprobe: 10 – 100 µM): rhodamine composite gel (top). The corresponding Coomassie gel (bottom) displays protein loading per lane.

PPDK-PDRP binding does have unusual characteristics, whilst PDRP appears to be a Ser/Thr kinase, it contains no recognised homology with known kinases; PDRP appears also to be a dual kinase and phosphatase and there appears to be some significance with the presence of ADP for P-loop activation (235). It is entirely possible that a short linear photoreactive peptide does not suffice the mimic the binding of PPDK-PDRP because the binding is contingent on interaction from multiple domains. Perhaps future target selection requires a more significant study into protein structure to understand binding mechanisms prior to investigation with linear peptide probes.

4.5 Conclusion and Future Work

The intention of this work was to produce photoprobe pTza-sequences that interact with cognate partner proteins as a basis for investigating phosphohistidine. Photo-reactive groups being used to bind partner compounds has a basis in the literature; for example, Yang *et al.* have used diazirine photoprobes to "trap" transient "reader/eraser" interactors of post-translational modifications on histone H3 (195). Photo-crosslinking has not been directly used to study phosphohistidine and thus may have provided a unique angle for investigation of pHis protein sequences through covalent capture of transient partner proteins and identification via SDS-PAGE or downstream proteomic MS analysis.

The photo probe **87** used to assess EGFR to Grb2-SH2 binding showed evidence of an ability to bind isolated Grb2-SH2 throughout the experiments; it was a struggle to find a consistent concentration gradient for the binding between the probe **87** and Grb2-SH2 and this was further exacerbated by the presence of a pY competitor probe **88**, where the binding became more convoluted and uncertain.

Proteomic analysis of the EGFR-Grb2 interaction had the potential to clarify uncertainty presented by earlier SDS-PAGE analysis, however the EGFR photoprobe **87** showed no statistically significant binding interactions to Grb2 in the workflow performed. The potential interactors identified were unexpected binders: ERH, G3BP1, FKBP1A. Grb2 was only found to bind one peptide in a HeLa quality control digest, whilst the HeLa sample used in this work was prepared differently, it could be speculated that Grb2 was present in low abundance and hence enrichment and identification in those circumstances was ambitious. Given the observation of binding in SDS-PAGE between the EGFR photoprobe **87** and Grb2, it may be possible to spike the HeLa cell lysate with isolated protein to artificially increase the abundance of the Grb2 in the sample. If the EGFR photoprobe **87** is a binder of Grb2 then the crosslinked complex will be amplified through enrichment resulting in Grb2 identification during analysis. Alternatively, it would be feasible to test the proteomics workflow with the EGFR photoprobe **87** on multiple cell lines where Grb2 may be present in greater abundance.

The issues with assessing EGFR-Grb2 binding with photoprobes were also reflected in the work assessing PPDK-RP binding. The initial incubation of the photo probe PPDK sequence **89** with isolated RP gave a clear illuminating band – indicative of probe binding to RP. However, subsequent experiments showed weak binding resolution and inconsistent trends. It

is likely that linear photoprobe peptide sequences are insufficient as direct mimics of authentic sequences in some instances, in particular PPDK-PDRP, where phosphorylation appears to be regulated by a fairly complex mechanism involving multiple domains. This may explain limited success in observing PPDK-PDRP interaction when compared with EGFR-Grb2 interaction. PDRP is a much more complicated Ser/Thr kinases-phosphatase hybrid, with some ADP dependence. In order to successfully produce photoprobes for these systems more studies need to be performed into structural analysis of these proteins (e.g. X-ray crystallographic analysis).

In any case, this initial attempt to identify phosphoprotein interaction with known partners using photoprobe peptides, indicates the challenges in demonstrating known interactions between two proteins: EGFR-Grb2 and PPDK-PDRP. To demonstrate a legitimate pHis peptide interaction with an unknown binding partner would be even more challenging, especially if the structural interaction is not understood.

One proposal for investigating pHis interaction with photoprobes in the future would be to produce peptide probes based on reputed pHis-sites reported in literature sources and corroborating the nature of the phosphosite (structured/unstructured) using AlphaFold predictions and attempt to use these unstructured interactions as a basis for investigating pHis interactions with pTza photoprobe mimic sequences. Structures in **Table 5** could act as a basis for photoreactive peptide probes of phosphohistidine in future work.

Protein	**Accession Number**	**pHis Site**	**Sequence-containing histidine pX**
Activity-dependent neuroprotector box protein	Q9H2P0	H950	PTKLM**H**NASDSE
Sp110 nuclear body protein	Q9HB58	H243	QEMP**H**SPLG
Integrator complex subunit 10	Q9NVR2	H220	IENQ**H**QGA
Lamin-B1	P20700	H571	EELF**H**QQG

SAFB-like transcription modulator	Q9NWH9	H556	SKSPG**H**MVILD

Table 5: Proteins identified through antibody immunoprecipitation by Fuhs *et al.* (25) and adaptation of table data compiled by Hardman *et al.* (127) on phosphosite prediction with UPAX (filtered for high confidence phosphosites). The sequences were observed in alphafold (236, 237) and chosen based on probability of the histidine site being located in an unstructured region

5 Investigating protein cyclisation (SurE – surugamide thioester cyclase)

Independent of the studies to generate reactive probes to study protein-protein interactions, a separate study to understand the role of peptide-modifying enzymes from natural product biosynthetic pathways was also undertaken. This study necessitated the optimisation of modified peptide synthesis methodology which fed into the remainder of the previously described work.

5.1 NRPS (non-ribosomal peptide synthetase) and surugamide biosynthesis

Surugamides are a class of peptidic natural products with activity against cathepsin B. Their biosynthesis is catalysed by a non-ribosomal peptide synthetase (NRPS), a large multi-modular enzyme. These NRPSs generate polypeptide structures by binding each amino acid building block to a phosphopantetheinyl-modified side chain, modifying these intermediates (e.g. by epimerisation) and subsequently catalysing amide-bond formation by condensation with an amino acid on an adjacent subunit to extend the peptide sequence (238-240). In general, condensation of amino acid subunits is continued to build a polypeptide until a terminal thioesterase domain is reached which cleaves the polypeptide from the peptidyl carrier protein (PCP) domain, in some systems additional modification occurs with cleavage, such as an *in-situ* cyclisation (238, 240).

NRPS produce specific peptide substrates from amino acid residues in a sequence in an assembly line, building until a thioesterase domain cleaves the sequence (238). NRPS systems often contain Type I or *cis*-thioesterase domains which are fused to the C-terminal domain within the assembly line infrastructure and directly cleave the produced peptide from the NRPS architecture (241-243). However, in surugamide biosynthesis a trans-acting thioesterase (SurE) performs the terminal cleavage from the NRPS (240, 244). SurE hydrolyses the peptide from the NRPS framework with concomitant cyclisation, the SurE cyclase shows a broad tolerance for amino acid components resulting in the production of multiple cyclic surugamide variants from a single NRPS architecture. In the surugamide NRPS architecture, four gene products (surA-D) are responsible for the production of surugamide polypeptide precursors. surA/D produce 5 related octapeptide sequences which are cleaved and cyclised by SurE thioesterase (surugamide A-E) and surB/C are responsible for the production of the sixth linear polypeptide,

surugamide F, and a seventh sequence cyclosurugamide F (244). The assembly line architecture of the surugamide biosynthetic system is displayed in **Figure 63** (based on diagram by Kuranaga *et al.* (244)). The final peptide sequence representing a linear surugamide A would be cleaved and cyclised from the PCP arm at the terminal domain of the SurD gene by the thioesterase protein SurE.

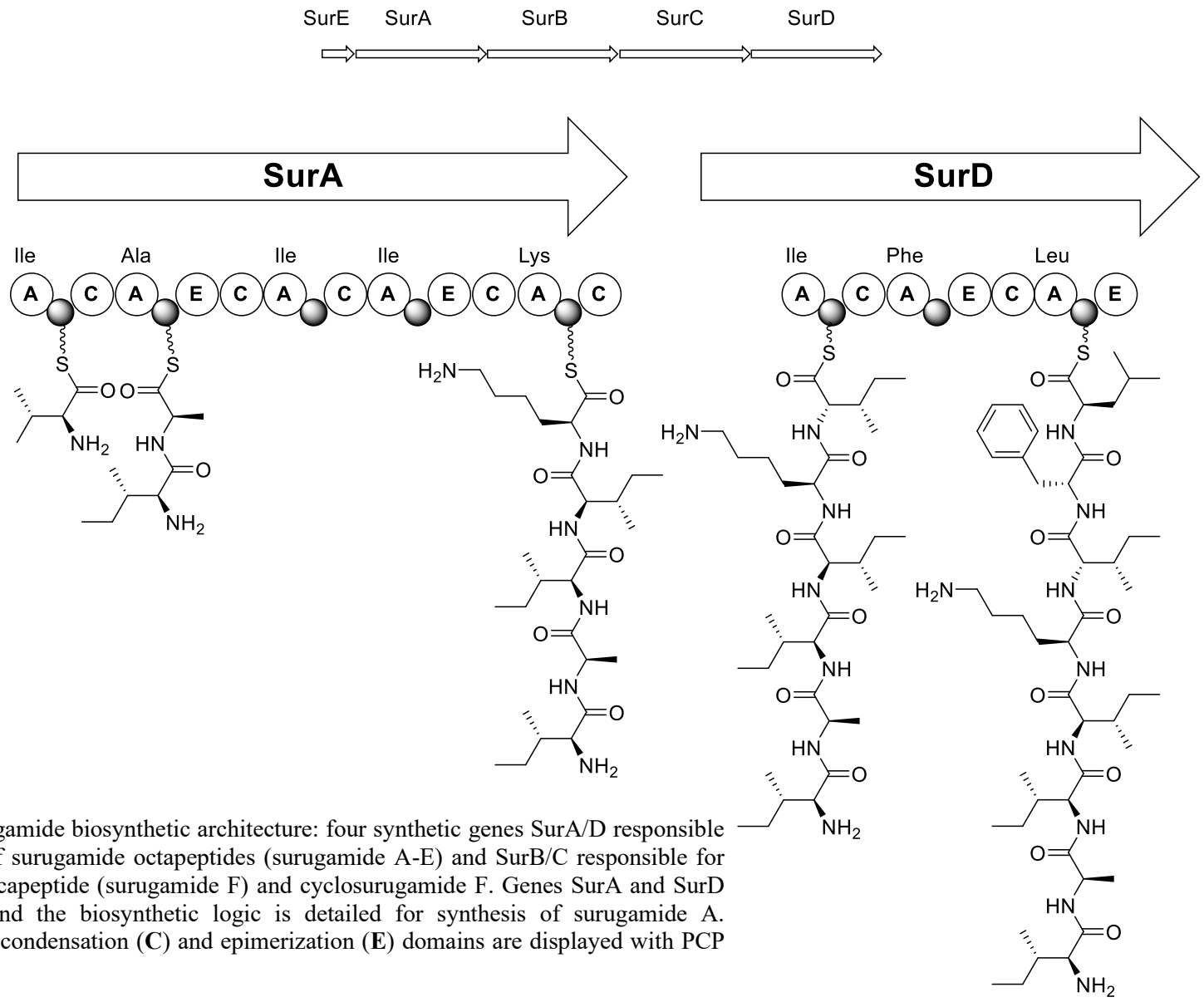


Figure 63: Surugamide biosynthetic architecture: four synthetic genes SurA/D responsible for production of surugamide octapeptides (surugamide A-E) and SurB/C responsible for production of decapeptide (surugamide F) and cyclosurugamide F. Genes SurA and SurD are expanded, and the biosynthetic logic is detailed for synthesis of surugamide A. adenylation (**A**), condensation (**C**) and epimerization (**E**) domains are displayed with PCP domain in grey.

Since SurE can cyclise a range of different octapeptide surugamide sequences and a decapeptide, there may be many different substrates that SurE is capable of cyclising. That SurE promiscuity towards varied substrates for cyclisation has led to strong interest in the use of SurE for the production of cyclic peptides *in vitro* perhaps with the aim of novel bioactive/antimicrobial compounds.

In this project, SurE promiscuity was investigated using peptidyl-SNAC (N-acetylcysteamine) thioester substrates that act as a suitable *in vitro* mimic for the *in vivo* binding of the surugamide sequence to the NRPS infrastructure by the PCP (4' phosphopantetheinyl cofactor). Previous studies demonstrate the high tolerance of SurE for amino acid variation in the central portion of the peptide sequence and D-isomer specificity at the C-terminus - where the SNAC or pantetheinyl cofactor bind *in vitro* and *in vivo* respectively (245) .

Whilst modest characterisation of SurE specificity is a great advancement, further characterisation would be required to fully elucidate the scope for SurE to act as a broad-spectrum cyclase for producing bioactive cyclic compounds. The individual production of SNAC-surugamides is a laborious process and purification of individual substrates by HPLC hampers the mass production of peptide-SNAC substrates for SurE cyclisation tests. As such, a screening method was sought for that would enable mass screening of SurE cyclisation tolerance without the bottleneck of individual SNAC-peptide purification by HPLC.

5.2 Aims of this work

The aim of this work was to hasten characterisation of SurE cyclisation tolerance by circumventing the need for peptidyl-SNAC substrates, which are time consuming to individually produce and purify. Production of peptidyl-SNAC substrates requires a solid-phase peptide synthesis (SPPS) workflow followed by cleavage from the solid resin and subsequent coupling with N-acetylcysteamine before subsequent purification by HPLC. A means for directly cyclising peptides from the resin would significantly expedite the process.

An analogous precedence existed for this process, where Kohli *et al.* showed TycC (a macrocyclising thioesterase) can recognise and perform its function on a solid-support bound peptidomimetic of tyrocidine A (246). They create a symmetry between resin-bound peptides and NRPS-bound proteins, whereby a thioesterase could catalyse release and cyclisation of the linear peptide substrate from the resin exhibiting biomimicry with the *in vivo* release and cyclisation of the protein from the NRPS machinery.

This research aimed to assess the suitability of PEGA-surugamide substrates for characterisation of SurE in analogy with the PEGA-tyrocidine A and TycC system, entrenching this method as an accurate means of assaying enzyme activity in NRPS systems. The results of PEGA-surugamide cyclisation with SurE would need to be validated against the established method with SNAC-surugamide substrates to confirm the novel results align with results from the pre-existing SNAC-surugamide data.

Additionally, this research aimed to clarify SurE tolerance for varied N-terminal residues on a surugamide substrate. Previous research has probed the tolerance of SurE for surugamides with N-terminal alloisoleucine, tryphtophan and proline (245); presumably to broadly probe the N-terminal specificity of SurE for isomers, larger hydrophobic moieties and the effect of the proline side chain on the ability of SurE to cyclise in the absence of the typical N-terminal amide. However, three residues alone was not a comprehensive scope of this facet of SurE specificity.

5.3 Synthesis and purification of SNAC-surugamide substrates

To synthesise surugamide-SNAC peptides, typical Fmoc-SPPS conditions were performed to produce a Boc-protected peptide on resin, the workflow is displayed in **Figure 64**.

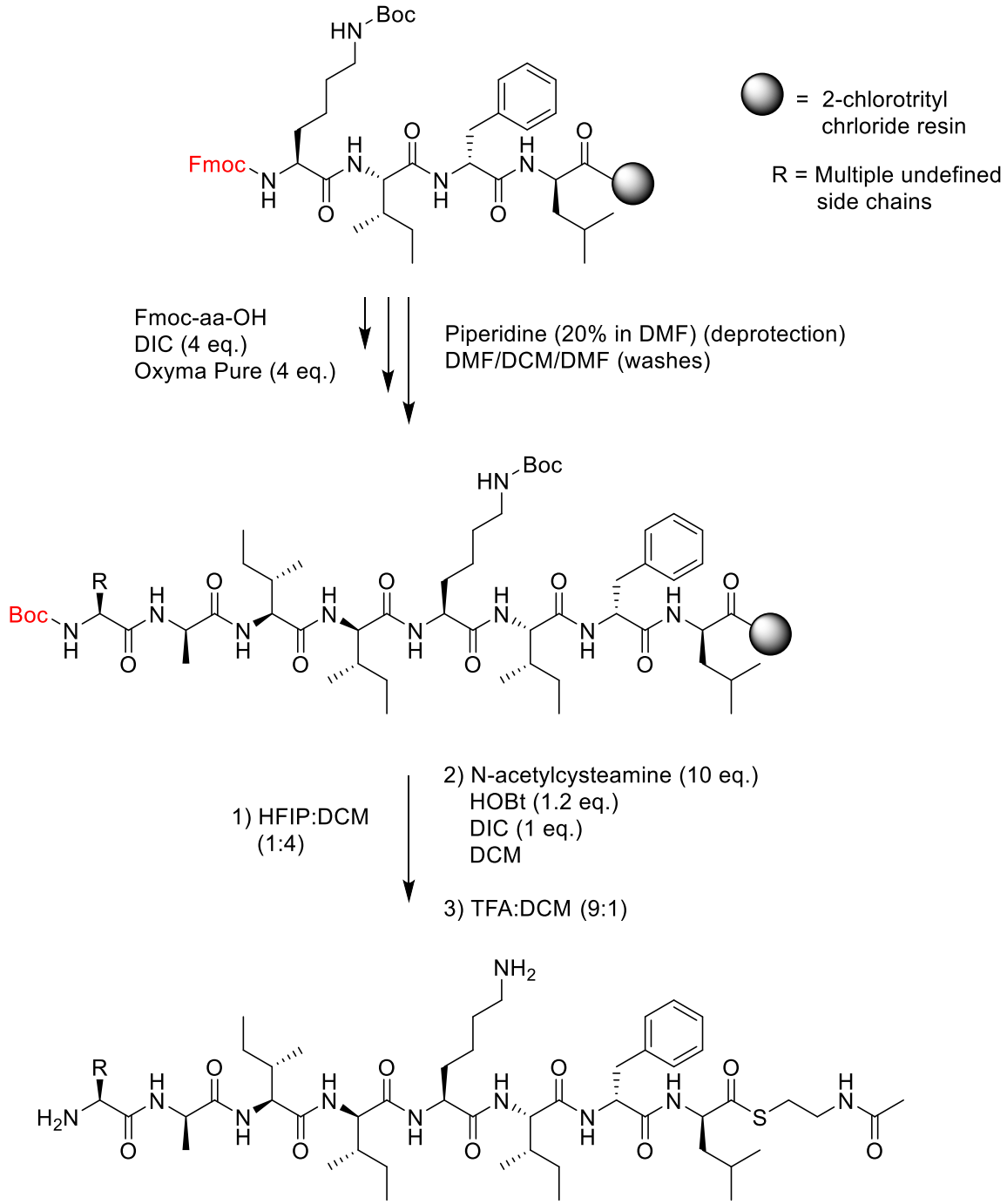


Figure 64: Workflow to produce surugamide-SNAC compounds through Fmoc-SPPS. The main chain of the peptide is generated by addition of Fmoc-amino acids (Fmoc-aa-OH) and the chain is terminated with an N-terminally Boc-protected amino acid (red). The Boc-amino acid is cleaved from the 2-chlorotrityl chloride resin with HFIP: DCM (1:4) and coupled to N-acetylcysteamine at the C-terminus. The surugamide-SNAC compounds are then globally deprotected with TFA: DCM (9:1).

2-chlorotrityl chloride resin provided some advantages compared with other resins, firstly it cleaves to provide an acidic C-terminus required for formation of the thioester bond needed in SNAC-surugamide formation. With 2-chlorotrityl chloride resin, a peptide can be cleaved in mild conditions (HFIP: DCM 1:4) retaining the orthologous protecting groups on the side chains of the amino acids. This opened the C-terminus for coupling to N-acetylcysteamine without exposing reactive side chain groups or the N-terminal amine.

In the NRPS synthesis, displayed in **Figure 63**, surugamide peptides are produced from the N-terminus to the C-terminus. As such, the NRPS synthetase ends with an epimerase domain to provide the C-terminal D-Leucine. In chemical synthesis, the peptides are synthesised from the C-terminus to the N-terminus. As such, the resin would need to be loaded with D-leucine. 2-chlorotrityl chloride resins are not commercially available with D-isomer amino acids preloaded; as such, the resin would need to be loaded manually to produce the surugamide sequence. 2-chlorotrityl chloride is relatively unstable in atmospheric conditions so loading under an N_2 atmosphere and using dry solvents is necessary. After the initial coupling of D-leucine onto the resin, Fmoc-SPPS can be perfomed under very standard conditions to extend the chain regardless of the stereochemistry of the building blocks.

Boc-protection was used on the N-terminus of the peptide to conceal the amine of the primary chain in the surugamide sequences to prevent any cross-reaction during subsequent N-acetylcysteamine coupling. Additionally, the Boc-protecting group could be easily removed with TFA: DCM (9:1) in a global deprotection and evaporated off eliminating additional purification steps for less volatile reagents.

In the standard surugamide octapeptides, surugamides A-D are sequences with an N-terminal isoleucine; surugamide E has an N-terminal valine and surugamide F (a distinct decapeptide) terminates with tryptophan. Therefore, valine was chosen for the terminus of one of the peptide substrates as a control, given that wild-type surugamide E ends in valine, the SurE thioester should cyclise a SNAC-surugamide with N-terminal valine. Glycine and alanine were chosen to test the ability for SurE to cyclise peptides with no side chain on the N-terminal amino acid and a sterically small and hydrophobic terminal residue respectively. Lysine would assess the ability for SurE to cyclise in the presence of a longer side chain with hydrophilic properties. Threonine, differing from valine by the presence of an OH in place of a CH_3, would also highlight any more distinct specificity between valine (which was expected to cyclise based on

the wild-type cyclisation of surugamide E) with its hydrophobic side chain and the sterically similar but more polar side chain of threonine. Cyclosurugamide F contains N-terminal tryptophan, additionally previous studies have shown successful cyclisation of an octapeptide surugamide with N-terminal tryptophan (245), so success with this cyclisation would act as a cross-research confirmation of observation adding credibility to novel observations.

Surugamide-SNACs were produced by first producing a base heptapeptide sequence without the terminal amino acid. Initial loading onto the resin was promoted by swelling the chlorotrityl resin in anhydrous DCM for 30 minutes. The initial Fmoc-D-Leu-OH was loaded with DIPEA (4 eq.) in anhydrous DCM and the solution was stirred for 40 minutes. The resin was endcapped with a wash of DCM:MeOH:DIPEA (17:2:1). Following Fmoc-SPPS, the resin-bound heptapeptides were split prior to the final coupling to form 6 N-terminally distinct surugamide sequences. The resin-bound surugamide heptapeptide sequences were terminally coupled with (Boc-(Gly, Ala, Lys(Boc), Thr(tBu), Trp(Boc), Val)-OH). Following coupling, the Boc-protected peptides were removed from the resin with hexafluoroisopropanol: DCM (1:4) and individually coupled to N-acetylcysteamine (10 eq.) in DCM with HOBt (1-1.5 eq.). The 6 SNAC-surugamide peptides were globally deprotected by stirring in TFA: DCM (9:1) for 3 hours and the residual volatiles were evaporated off.

The produced sequences are listed in **Table 6** below:

SNAC-Surugamide Sequence
H_2N-**Ala**IleDAlaIleDIleLysIleDPheDLeu-SNAC
H_2N-**Gly**IleDAlaIleDIleLysIleDPheDLeu-SNAC
H_2N-**Thr**IleDAlaIleDIleLysIleDPheDLeu-SNAC
H_2N-**Lys**IleDAlaIleDIleLysIleDPheDLeu-SNAC
H_2N-**Val**IleDAlaIleDIleLysIleDPheDLeu-SNAC
H_2N-**Trp**IleDAlaIleDIleLysIleDPheDLeu-SNAC

Table 6: SNAC-surugamide sequences synthesised (N-terminal amino acids emboldened).

Crude peptides were purified by high-pressure liquid chromatography (HPLC). Avoiding all purification steps until a final HPLC helped to preserve yield by minimising product loss through apparatus or transfer during multiple chromatography steps; however, the

disadvantage was that undesirable peptides (such as sequence curtailment variants of the desired sequence) may have been present as SNAC-peptides requiring removal at the end of synthesis. **Figure 65** displays the high-resolution mass spectrometry (HRMS) result and HPLC data for the Gly N-terminal SNAC-surugamide peptide.

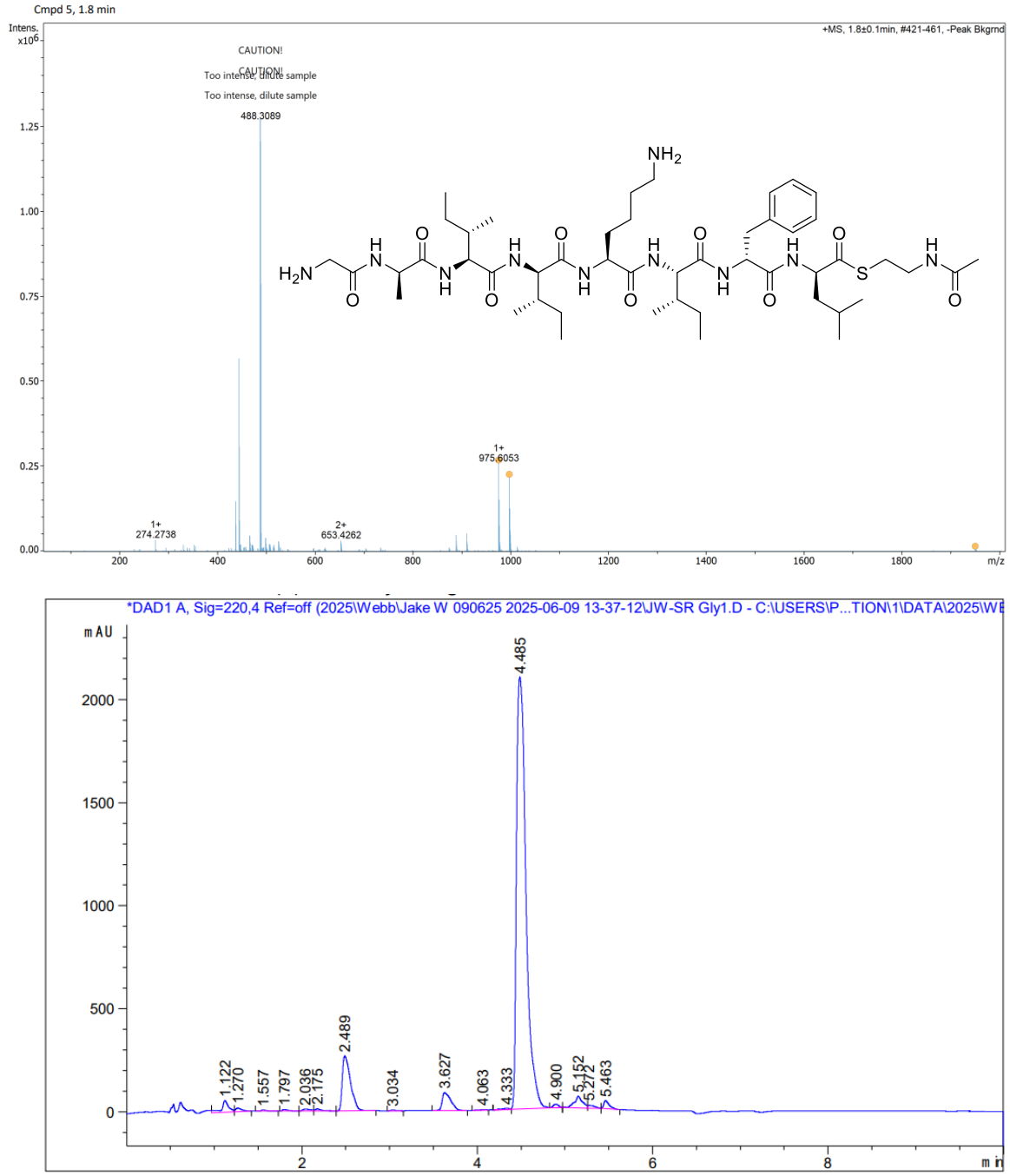


Figure 65: HRMS and HPLC of SNAC-surugamide Gly. Expected $[M+H]^+$ and $[M+2H]^{2+}$ observed in HRMS spectrum and 4.485 minute peak for SNAC-surugamide Gly, showing predominance of desired substrate by HPLC (81% purity by area).

HPLC purification and HRMS analysis was performed for each of the surugamide substrates to confirm the successful production of the desired substrates. The HRMS and HPLC data for all six substrates is shown in the Appendix.

Each of the SNAC-surugamide substrates was incubated with SurE and cyclisation was assessed by observing mass shifts in HRMS. In addition, the sequences were produced on a PEGA resin and cyclised by SurE. (These experiments were conducted by Dr. Asif Fazal in collaboration with the candidate) The results of this cyclisation are outlined in **Table 7**.

N-terminal residue	SNAC cyclised	PEGA cyclised
Ala	Yes	Yes
Gly	Yes	Yes
Lys	Yes	Yes
Trp	Yes	Yes
Thr	Yes	Yes
Val	Yes	Yes
Gln	No	No
Leu	Yes	Yes
Ile	Yes	Yes
Orn	Yes	Yes
Arg	-	No
Asn	-	No
Asp	-	No
Cys	-	No
Glu	-	No
His	-	No
Met	-	No
Phe	-	No
Pro	-	No
Ser	-	No
Tyr	-	No

Table 7: Summary of SNAC-peptide cyclisation by SurE. All surugamide-SNAC substrates produced were cyclised by SurE and the PEGA equivalents were also cyclised. The six SNAC substrates made

in this work are listed first in addition to 4 additional SNAC-substrates that were tested in this work by Fazal *et al.* (247), thus they are also listed to elucidate the scope of the work. (-) represented an untested SNAC substrate where only a PEGA resin result was determined.

Cyclisation of SNAC-surugamide substrates with N-terminal isoleucine would be expected as this is the native surugamide A substrate and cyclisation of N-terminal valine would be expected likewise due to being comparable to native surugamide E. Additionally N-terminal tryptophan has been assessed previously in work by Matsuda *et al.* (245) so a positive result in alignment with this prior work adds credibility to the novel observations.

For all the PEGA bound substrates with corresponding SNAC-surugamide substrates, SurE demonstrated an ability to cyclise both substrates or neither. Given the broad scope of SNAC-surugamide substrates produced and the consistency of the data, any further observations on PEGA resin were concluded to be sufficient to probe SurE N-terminal specificity in the absence of the corresponding SNAC-substrates. (SNAC-peptides with terminal Gln, Leu, Ile, Orn; PEGA-substrates synthesis and cyclisation were performed by Dr. Asif Fazal)

5.4 Conclusion

The SNAC-surugamide substrates produced were a significant contribution to confirming the activity of the SurE cyclase. The ability for SurE to cyclise SNAC-surugamide substrates with N-terminal: Ala, Gly, Lys, Thr and Val were a novel discovery having not been reported prior to this work and therefore further elucidates understanding of SurE cyclase promiscuity and tolerance for residue variance at the N-terminus. N-terminal Trp has been shown to cyclise in surugamide octapeptides in prior work (245), but positive cyclisation in this work was critical to show comparable results between this work and prior discovery. In addition to the SNAC-surugamides produced in this work, Fazal *et al.* (247) contributed 4 SNAC-surugamide substrates with N-terminal: Gln, Leu, Ile and Orn. Of all the SNAC-substrates cyclised in the study by SurE, only the substrate with N-terminal Gln did not cyclise when incubated with SurE and this was reflected by a lack of cyclisation from the PEGA-bound surugamide with N-terminal Gln.

Additionally, cyclisation of SNAC-substrates was corroborated by cyclisation of its corresponding PEGA substrate in all cases. Therefore, this work demonstrates that PEGA resin bound peptides can be utilised to successfully assess the activity of SurE on surugamide substrates. In a broader context, this work provides further credibility to the hypothesis that PEGA resin bound peptides may be suitable for assessing various thioesterase enzymes from NRPS systems therefore circumventing the need to produce SNAC-substrates as suitable mimics of the PCP domain; similar exploration of TycC on tyrocidine A by Kohli *et al.* showed that PEGA bound substrates can be successfully used to assess promiscuous thioesterase activity. As such, this work provides further evidence for utilising PEGA-bound peptides to expedite the process of exploring NRPS thioesterase activity.

6 Experimental Methods

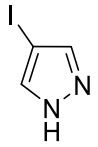
6.1 Reagents and Equipment

All reagents and starting materials were obtained from commercial suppliers, if further purification of the starting material was performed this will be outlined in the corresponding experimental procedure. Where reactions were performed in a dry environment, the glassware was been pre-dried in a glassware oven, exposed to vacuum and backflushed with a continuous flow of N_2. TLC was performed on aluminium sheets top with Merck silica gel 60Å (medium pore diameter) F_{254}. Column chromatography was performed on Merck Supelco silica gel 60 Å (mean pore size), 40-63 μm particle size.

All peptides were made in a polypropylene column with a filter and an attached stopcock. The tubes were mounted on a Supelco Visiprep™ SPE Vacuum Manifold, PTFE Free. All solvents used in peptide synthesis were commercially available and used direct from supplier with no further drying or purification. Some peptides were synthesised on a peptide synthesiser - CEM Liberty Blue; where this is the case, the experimental procedure will state use of the synthesiser explicitly. Centrifugation of peptides prior between wash steps was performed on an Eppendorf centrifuge 5810. Lyophilisation was performed on a Virtis Benchtop freeze dryer.

NMR spectrometry was performed on a Bruker AV3HD 9.4 T (400 MHz 1H) and Bruker AV4 NEO 11.75 T (500 MHz 1H, cryoprobe). Chemical shifts are reported in ppm with the multiplicity, integral and assignment for ^{1}H NMR analysis: δ = shift in ppm (multiplicity, coupling constant, integral, assignment). Where assignment is more complex, the atom being assigned will be listed with a numbered atom corresponding to a diagram e.g. two protons on H2 would be transcribed δ = shift in ppm (multiplicity, coupling constant, 2H, H2). Multiplicities will be recorded as: s = singlet, d = doublet, t = triplet, q = quartet, dd = doublet of doublets, td = triplet of doublets, dt = doublet of triplets, dq = doublet of quartets, m = multiplet. ^{31}P NMR is proton decoupled where multiplicities are not recorded and thus only singlets are observed unless explicitly stated otherwise. LCMS is obtained on a Bruker amaZon speed spectrometer and HRMS was obtained by a Bruker Maxis Impact.

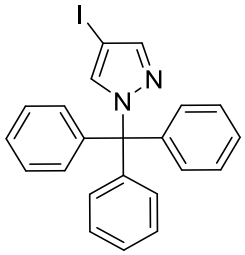
6.2 4-Iodo-1H-pyrazole (31)



Pyrazole (3.90 g, 57.3 mmol, 1.00 eq.) was suspended in deionised water (30 mL) to which NaOAc (9.84 g, 120 mmol, 2.09 eq.) was added. The solution was heated to 60˚C until limpid, before the addition of a solution of NaI (17.1 g, 114 mmol, 1.99 eq.) and I_2 (9.15 g, 36.1 mmol, 0.630 eq.) in deionised water – dropwise through an addition funnel. H_2O_2 (30 wt%) (11.4 mL, 3.79 g, 111 mmol, 1.93 eq.) was added dropwise and the reaction was heated to reflux for 6 hours. After reflux, the reaction was left to cool and sodium hydroxide pellets were added; addition of sodium hydroxide pellets resulted in white precipitate formation, pellets were added until the precipitate dissolved. The solution was then filtered and the filtrate placed in an ice bath to form the title compound as white crystals; the crystals were subsequently washed with ice-cold deionised water and dried in vacuo. In accordance with methodology by Padial *et al.* (248),

4-iodo-1H-pyrazole (3.37 g, 17.4 mmol, 30%), ^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.64 (s, 2H), 11.17 (s, 1H, N-H); ^{13}C NMR (126 MHz, $CDCl_3$): δ = 138.8, 56.7. Spectral data in alignment with Mezei *et al.* (249)

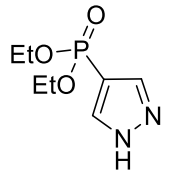
6.3 4-Iodo-1-trityl-1H-pyrazole (32)



Under N_2 atmosphere, 4-iodopyrazole (500 mg, 2.58 mmol, 1.00 eq.) was dissolved in anhydrous DCM (6 mL) and triethylamine (750 µL, 5.38 mmol, 2.08 eq.) was added. Trityl chloride (666 mg, 2.39 mmol, 0.89 eq.) was dissolved in anhydrous DCM (4 mL) and syringed into the reaction mixture over ice. After addition of the trityl chloride solution, the ice bath was removed, and the reaction was left to stir overnight. The crude solution was poured into water (8 mL) and saturated $NaHCO_3$ (8 mL) was added slowly. The crude product was extracted in DCM (4 × 5 mL), dried over $MgSO_4$ and the solvent was removed in vacuo. The crude product was subsequently purified by flash chromatography (9:1 hexane: ethyl acetate). In accordance with methodology from Zhang *et al.* (250)

4-Iodo-1-trityl-1H-pyrazole (376 mg, 0.862 mmol, 33%), R_f (Hex: EtOAc 9:1) = 0.35, ^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.66 (s, 1H), 7.41 (s, 1H), 7.34-7.28 (m, 11H), 7.15-7.09 (m, 7H); ^{13}C NMR (126 MHz, $CDCl_3$): δ = 144.7, 142.7, 136.5, 130.1, 129.7, 127.96, 127.94, 127.9, 127.8, 127.7, 127.3, 79.3. In accordance with data by Elguero *et al.* (251) and Zhang *et al.* (250)

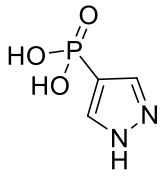
6.4 Diethyl-1H-pyrazole-4-ylphosphonate (35)



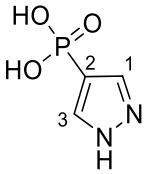
4-iodopyrazole (499 mg, 2.57 mmol, 1.00 eq.), PPh_3 (206 mg, 0.785 mmol, 0.305 eq.) and $Pd(OAc)_2$ (59 mg, 0.263 mmol, 0.102 eq.) were combined in a round-bottomed flask equipped with a condenser. The flask was evacuated under vacuum and flushed with N_2; this was repeated 3 times. Ethanol (9 mL), Et_3N (0.8 mL, 6 mmol, 2 eq.), $HPO(OEt)_2$ (0.8 mL, 6 mmol, 2 eq.) were added to the flask and the reaction was heated at reflux for 22 hours before being left to cool to room temperature. Saturated NH_4Cl (20 mL) was added. The solution was extracted in EtOAc (4 × 20 mL) and the combined organic layers were subsequently washed with brine (1 × 20 mL) before being dried over $MgSO_4$. The solvent was removed *in vacuo*, and the residue was purified by column chromatography with a 1:9 ethyl acetate: hexane flush followed by elution of the product spot in ethyl acetate. The product fractions were combined and the solvent removed in vacuo, producing the title compound as a brown solid. In accordance with methodology by Makwana *et al.* (158)

Diethyl-1H-pyrazole-4-ylphosphonate (354 mg, 1.73 mmol, 67%), R_f (EtOAc) = 0.12, ^1H NMR (400 MHz, $CDCl_3$): δ = 7.84 (d, J = 1 Hz, 2H, Ar-H), 4.05-3.99 (m, 4H, CH_2), 1.24 (t, J = 7.1 Hz, CH_3); ^{31}P NMR (162 MHz, $CDCl_3$): δ = 15.33. In accordance with data by Makwana *et al.* (158)

6.5 1H-pyrazole-4-yl phosphonate (28)

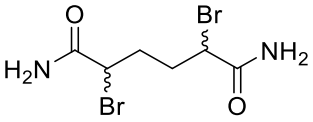


Diethyl (1H-pyrazol-4-yl) phosphonate (100 mg, 0.490 mmol, 1.00 eq.) was dissolved in MeCN (5 mL), followed by the addition of TMS-Br (750 mg, 635 µL, 4.90 mmol, 10.0 eq.). The reaction was heated to 50˚C and left overnight. The reaction was cooled to room temperature before the volatiles were removed *in vacuo*. The residue was triturated in water: MeCN, the supernatant was decanted, and the precipitate was centrifuged with two further washes in MeCN, to produce the title compound as a dry white powder.



1H-pyrazole-4-yl phosphonate (21 mg, 0.14 mmol, 29%), ^{1}H NMR (500 MHz, D_2O): δ = 8.02 (d, J = 2.3 Hz, 2H, H1/H3); ^{13}C NMR (126 MHz, D_2O): δ = 136.8 (d, J = 18.5 Hz, C1, C3), 112.6 (d, J = 209.0 Hz, C2); ^{31}P NMR (162 MHz, D_2O): δ = 8.27.

6.6 2,5-dibromohexanediamide (25)



Recrystallisation of N-Bromosuccinimide:

N-Bromosuccinimide (NBS) (25.63 g, 144.0 mmol) was dissolved in 250 mL of water. Hot filtration was performed, the filtrate was submerged in an ice bath for 2 hours before vacuum filtration to isolate NBS (16.19 g, 90.96 mmol, 63%).

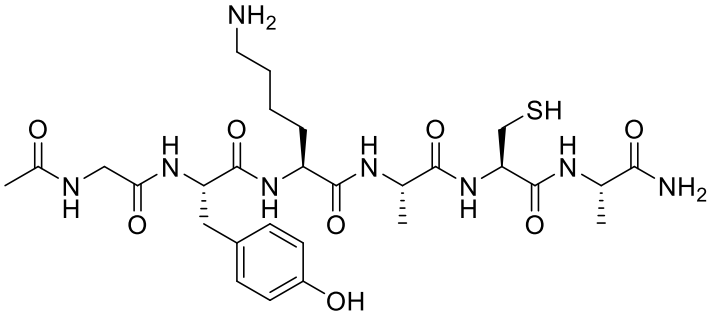
Synthesis of 2,5-dibromoadipamide

Adipic acid (5.23 g, 35.8 mmol) was dissolved in thionyl chloride ($SOCl_2$) (15 mL, 207 mmol, 5.8 eq.) and heated to 81˚C under reflux conditions for 1.5 hours. After which time, the solution was left to cool to room temperature overnight. CCl_4 (20 mL) was added, followed by, the freshly recrystallised, NBS (16.19 g, 90.96 mmol, 2.540 eq.). The solution was stirred vigorously followed by addition of HBr (33 wt% in acetic acid) (1 drop). A guard tube, containing sodium hydroxide pellets, was affixed to the top of the condenser before the mixture was heated to 80˚C for 4 hours, during which time, a colour change from red to dark red/black was observed. The reaction was cooled to 0˚C and the precipitate collected by filtration. The flask was washed with ether (50 mL). The filtrate was concentrated in vacuo to give a crude dark-red oil. Ammonium hydroxide solution (35%, 40 mL) was cooled to 0˚C before the dropwise addition of the crude dark-red oil. The solution, which had turned light blue, was stirred for an hour at 0˚C before a pale-blue chalky solid was isolated by filtration. Subsequent recrystallisation in MeOH: H_2O (1:1) at 60˚C for 30 minutes gave the title compound a grey solid. The filtrate was placed into an ice-bath where formation of the title compound as an off-white/grey crystal product was observed. Method in accordance with Chalker *et al.* (252)

2,5-dibromohexanediamide (2.76 g, 9.14 mmol, 25%) (off-white/grey crystals), ^{1}H NMR (500 MHz, DMSO-d_6): δ = 7.70 (s, 2H, N-H), 7.31 (s, 2H, N-H), 4.35-4.32 (m, 2H, CH), 2.06-1.81 (m, 4H, CH_2); ^{13}C NMR (126 MHz, DMSO-d_6): δ = 170.3, 170.25, 49.0, 48.7, 33.0, 32.9.

Analytical data in alignment with Morrison *et al.* (166)

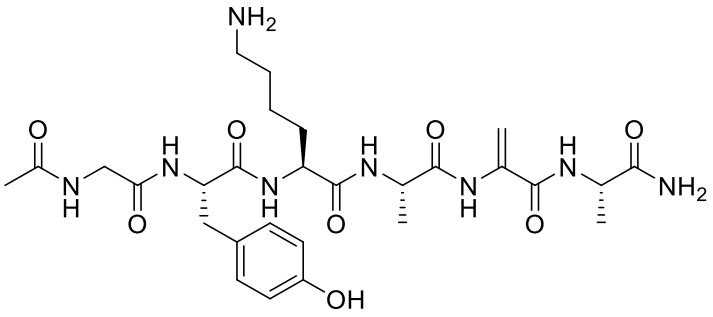
6.7 Ac-GYKACA-NH_2 (43)



Rink amide MBHA resin (0.51 mmol/g loading, 297 mg, 0.15 mmol) was swollen in DMF (1 mL, 1 hour). After filtering off the DMF, the resin was deprotected with a solution of piperidine (20% in DMF, 10 minutes). Piperidine was filtered off and followed by an additional wash of piperidine (20% in DMF, 5 minutes). Piperidine was filtered again and the resin washed with DMF (5 × 2 mL). A solution of Fmoc-Ala-OH (121 mg, 0.39 mmol, 2.6 eq.) with Oxyma Pure (57 mg, 0.40 mmol, 2.7 eq.) and DIC (23.7 μL, 0.15 mmol, 1 eq.) in DMF was added to the resin and agitated for 4 hours at room temperature. The amino acid solution was filtered and a solution of acetic anhydride: pyridine (3:2) was added and agitated for 30 minutes to endcap the resin. The loaded resin was washed with DMF (3 × 2 mL), DCM (3 × 2 mL) and MeOH (3 × 2 mL). The resin-bound Fmoc-protected alanine was deprotected with piperidine (20% in DMF, 2 × 5 minutes) and washed with DCM (2 × 2 mL) and DMF (2 × 2 mL). Deprotection of the loaded alanine was followed by addition of a premixed solution of: Fmoc-Cys(Trt)-OH (387 mg, 0.66 mmol, 4.4 eq.), Oxyma Pure (86 mg, 0.60 mmol, 4 eq.) and DIC (96 μL, 0.61 mmol, 4.1 eq.) in DMF (2 mL). After an hour of agitation, the amino acid cocktail was decanted and the resin was washed with DMF (3 x 2 mL), DCM (3 x 2 mL), piperidine (2 × 2 mL, 5 minutes each), DMF (3 × 2 mL) and DCM (3 × 2 mL). This process was repeated with subsequent amino acids: Fmoc-Ala-OH (190 mg, 0.61 mmol, 4 eq.), Fmoc-Lys(Boc)-OH (295 mg, 0.63 mmol, 4.2 eq.), Fmoc-Tyr(tBu)-OH (289 mg, 0.63 mmol, 4.2 eq.), Acetyl-Gly-OH (75 mg, 0.64 mmol, 4.3 eq.). The final coupling was followed by washes of DMF (3 × 2 mL), DCM (3 × 2 mL) and MeOH (3 × 2 mL). The resin was dried over a vacuum manifold and a stream of N_2 was passed over the resin until dry. A cocktail of TFA (94%): water (2.5%): EDT (2.5%): TIPS (1%) (2 mL) was added to the resin and agitated for 3 hours. The solution was then filtered into ice-cold Et_2O (20 mL) and centrifuged at 4000 rpm for 10 minutes. The Et_2O was decanted and the precipitate washed twice more with Et_2O, followed by centrifugation and decantation of residual ether. The residual solid was resuspended in deionised water (7 mL) and lyophilised.

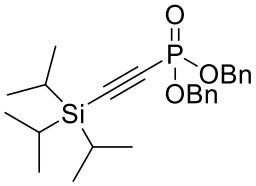
Peptide (Ac-GYKACA-NH_2) (27 mg, 0.04 mmol, 27%) was obtained as a flocculent white solid. LCMS analysis: observed $[M+H]^+ = 653.2$, expected $[M+H]^+ = 653.3$

6.8 Ac-GYKADhaA-NH_2 (45)

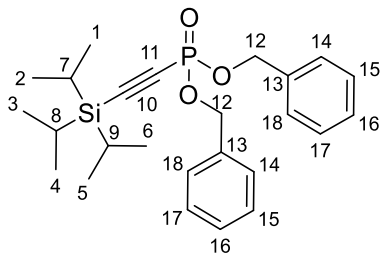


Peptide (Ac-GYKACA-NH_2) (19 mg, 0.03 mmol) was dissolved in water (0.4 mL) and TCEP (4 mg, 0.01 mmol, 0.5 eq.) was added. A solution of 2,5-dibromoadipamide (90 mg, 0.3 mmol, 10 eq.) in DMF (0.2 mL) was added. The solution was added to a vial containing K_2CO_3 (21 mg, 0.15 mmol, 5 eq.). The vial was placed in a falcon tube on a 37˚C, 200 rpm incubator. LCMS analysis was performed at hour intervals. After 4 hours, LCMS analysis: observed $[M+H]^+$ 619.3, expected $[M+H]^+$ = 619.3

6.9 Dibenzyl [(triisopropylsilyl)ethynyl]phosphonate (57)



TIPS-acetylene (1.4 mL, 6.2 mmol, 1.2 eq.) was added to dry THF (40 mL) and the solution cooled to 0˚C in an ice bath. ^{i}pr-MgCl (2M in THF, 2.6 mL, 5.2 mmol, 1.0 eq.) was added in a dropwise manner. The mixture was stirred for 10 minutes then warmed to room temperature for 1 hour. The solution was cooled in an ice bath and $ClP(NEt_2)_2$ (1 mL, 5 mmol, 0.9 eq.) was added dropwise before heating to room temperature and being left to stir overnight. The solvent was removed in vacuo before the addition of anhydrous MeCN (16 mL) and the solution was stirred over ice. Anhydrous benzyl alcohol (3.4 mL, 33 mmol, 5.3 eq.) was added in one portion followed by tetrazole (0.45M in MeCN, 23 mL, 10 mmol, 1.9 eq.) and the solution left to stir overnight. Saturated NH_4Cl (15 mL) was added, and the volatiles were removed in vacuo. The biphasic solution was washed with EtOAc (3 × 20 mL) and the combined organic layers washed with brine (1 × 20 mL). The organic layer was dried over $MgSO_4$, filtered and the solvent removed *in vacuo*. The crude oil (4.86 g) was analysed by LCMS displaying the diagnostic $[M+H]^{1+}$ 427 peak. The crude oil was dissolved in EtOAc (20 mL) and subsequently washed with water (2 × 10 mL), $Na_2S_2O_5$ (5%) (3 × 10 mL), water (1 × 10 mL) and brine (1 × 10 mL). The collected organic layer was dried with $MgSO_4$, filtered and the volatiles removed in vacuo. The colourless oil (1.90 g) was purified by column chromatography, a flush of hexane: ethyl acetate 9:1 followed by elution of the product in hexane: ethyl acetate 6:4. Removal of the solvent in vacuo gave a colourless oil. In accordance with methodology by McAllister *et al.* (159)

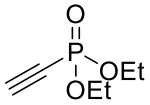


Dibenzyl [(triisopropylsilyl)ethynyl]phosphonate (294 mg, 0.66 mmol, 13%) (colourless oil), R_f (Hex: EtOAc 3:2) = 0.8, ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.34 (m, 10H, H14-18), 5.11 (d, J = 8.5 Hz, 4H, H12), 1.08 (m, 21H, H1-9); ^{13}C NMR (126 MHz, $CDCl_3$): δ = 135.5 (d, J = 7.6 Hz, C13), 128.56 (C15, C17), 128.50 (C16), 127.9 (C14, C18), 107.8 (d, J = 7.6 Hz, C10),

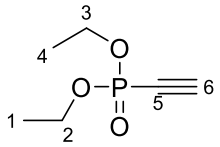
95.8 (d, J = 274.7 Hz, C11), 68.5 (d, J = 5.2 Hz, C12), 18.4 (C1-6), 10.8 (C7-9); ^{31}P NMR (162 MHz, $CDCl_3$): δ = -8.00.

Analytical data in alignment with McAllister *et al.* (159)

6.10 Diethyl ethynyl phosphonate (49)



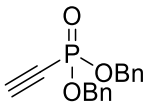
Diethyl chlorophosphate (700 µL, 4.84 mmol) was suspended in anhydrous THF (20 mL). Ethynyl magnesium bromide (10 mL, 5 mmol, 1 eq.) was added dropwise to the ice-cold diethyl chlorophosphate solution. After 3 hours of stirring at room temperature, saturated NH_4Cl (10 mL) was added, and the solution was extracted in Et_2O (3 × 10 mL). The combined organic extracts were dried over $MgSO_4$ and the solvent removed *in vacuo*. Column chromatography in hexane: ethyl acetate (1:1) R_f = 0.22 yielded 60 mg of an impure product. In alignment with methodology by McAllister *et al.* (154)



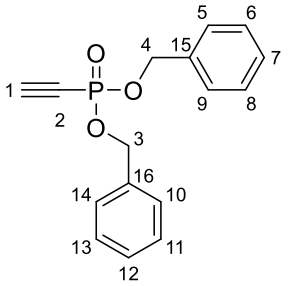
Diethyl chlorophosphate (40 mg, 0.24 mmol, 4.8% - based on ^{31}P NMR analysis), R_f (Hex: EtOAc 1:1) = 0.22, ^{1}H NMR (400 MHz, $CDCl_3$): δ = 4.22-4.15 (m, 4H, H2-3), 2.89 (d, J = 13.2 Hz, 1H, H6), 1.38 (td, J = 7.1, 0.8 Hz, 6H, H1 & H4); ^{31}P NMR (162 MHz, $CDCl_3$): δ = -8.37, LCMS analysis: $[M+H]^+$ = 162.4, expected: $[M+H]^+$ = 162.0

Analysis corresponds with McAllister *et al.* (154)

6.11 Dibenzyl ethynyl phosphonate (58)



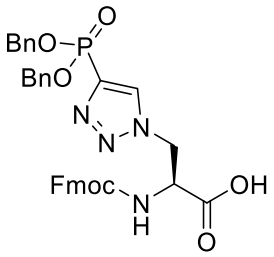
Dibenzyl [(triisopropylsilyl)ethynyl] phosphonate (2.81 g, 6.35 mmol, 1.00 eq.) was dissolved in THF (300 mL) and H_2O (50 mL). TBAF (1M in THF) (7 mL, 7 mmol, 1 eq.) was added and the mixture was stirred over ice for 2 hours. Saturated NH_4Cl (20 mL) was added, and the solution was extracted in Et_2O (4 × 50 mL), the combined organic extracts were dried over $MgSO_4$ and reduced *in vacuo*. The crude oil was purified by column chromatography in hexane: ethyl acetate (3:2). In accordance with methodology by McAllister *et al.* (159)



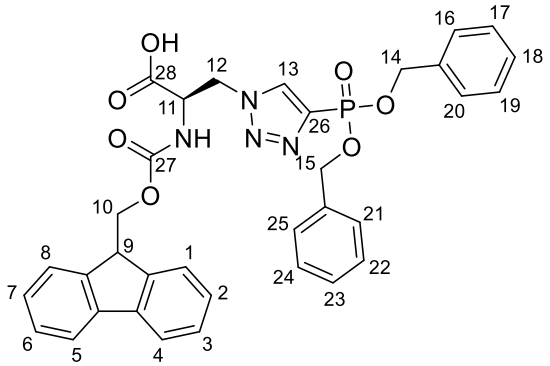
Dibenzyl ethynyl phosphonate (810 mg, 2.82 mmol, 44%) (colourless oil), R_f (Hex: EtOAc 3:2) = 0.26, ^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.36 (m, 10H, H5-14), 5.12 (d, J = 8.55 Hz, 4H, H3-4), 2.91 (d, J = 13.5 Hz 1H, H1); ^{13}C NMR (126 MHz, $CDCl_3$): 135.3 (d, J = 7.5 Hz, C15-16), 128.8 (s), 128.7 (s), 128.1 (s), 88.5 (d, J = 51.7 Hz, C2), 74.1 (d, J = 295.1 Hz, C1), 68.9 (d, J = 5.3 Hz, C3-4), ^{31}P NMR (162 MHz, $CDCl_3$): δ = -7.88. LCMS analysis shows $[2M+H]^+$ = 572.89, expected: $[2M+H]^+$ = 573.16

In alignment with data by McAllister *et al.* (159)

6.12 Fmoc-pTza(OBn)$_2$-OH (16)



Dibenzyl ethynyl phosphonate (810 mg, 2.82 mmol, 1.2 eq.) was dissolved in THF (32 mL). CuI (44 mg, 0.23 mmol, 0.1 eq.) and sodium ascorbate (140 mg, 0.7 mmol, 0.3 eq.) were dissolved in H_2O (16 mL) and added to the THF solution. Fmoc-Dap(N_3)-OH (828 mg, 2.35 mmol, 1 eq.) was dissolved in THF: H2O (1:1) (32 mL) and added to the alkyne solution. The reaction was stirred for 2 hours at room temperature. The volatiles were removed *in vacuo*, and the residue was washed with Na_2CO_3 (10%) (100 mL) and Et_2O (3 × 60 mL). The aqueous layer was acidified to pH 1 with HCl (conc.) and washed with EtOAc (4 × 60 mL). The combined organic extracts were dried over $MgSO_4$ and concentrated. In accordance with methodology by McAllister *et al.* (159)

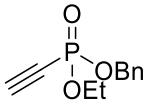


Fmoc-pTza(OBn)$_2$-OH was obtained (976 mg, 1.55 mmol, 66%), ^{1}H NMR (400 MHz, $CDCl_3$): δ = 8.05 (s, H13), 7.74 (d, 2H, J = 7.5 Hz, Fmoc-H), 7.55 (m, 2H, Fmoc-H), 7.40-7.25 (m, 14H, Fmoc-H and H16-25), 6.27 (d, 1H, NH), 5.22 (d, J = 8.2 Hz, 2H), 5.12 (dd, J = 8.4, 2.0 Hz, 4H, H14-15), 4.90-4.75 (m, 3H, H11-12), 4.41 (m, 1H, H10), 4.26 (m, 1H, H10), 4.14 (t, 7.2 Hz, 1H, H9); ^{13}C NMR (126 MHz, $CDCl_3$): δ = 171.7 (s, C28), 155.8 (s, C27), 143.6 (d, J = 20.2 Hz,

Fmoc/Ph C), 141.3 (d, J = 6.9 Hz , Fmoc/Ph C), 135.2 (d, 35.1 Hz, C26), 132.1 (d, 34.5 Hz, C13), 128.6 (s, Fmoc/Ph C), 128.6 (s, Fmoc/Ph C), 128.5 (s, Fmoc/Ph C), 128.1 (d, J = 10.9 Hz, Fmoc/Ph C), 127.7 (s, Fmoc/Ph C), 127.1 (d, 8.6 Hz, Fmoc/Ph C), 125.2 (d, 10.8 Hz, Fmoc-C), 119.9 (s, Fmoc-C), 69.3 (d, J = 5.2 Hz, C14-15), 69.07 (d, J = 5.4, C14-15), 67.3 (s, C10), 53.8 (s, C11), 51.1 (s, C12), 47.0 (s, C9); ^{31}P NMR (162 MHz, $CDCl_3$): δ = 8.48 (m), LCMS analysis $[M+H]^+$ = 639.3, expected: $[M+H]^+$ = 639.2

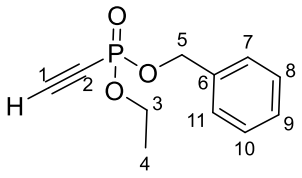
Analytical data in alignment with McAllister *et al.* (159)

6.13 Benzyl-ethyl ethynyl phosphonate (60)



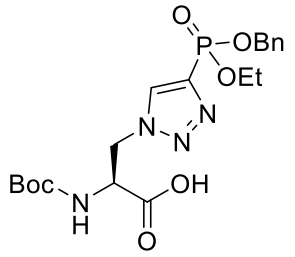
TIPS-acetylene (1.4 mL, 6.2 mmol, 1.24 eq.) was added to ice-cold anhydrous THF (40 mL), followed by dropwise addition of ^{i}pr-MgCl (2M in THF, 3 mL, 6 mmol, 1.2 eq.) over 10 minutes. Once added, the solution was left to stir at room temperature for 1 hour. The solution was cooled in an ice bath and $ClP(NEt_2)_2$ (1 mL, 5 mmol, 1 eq.) was added dropwise before heating to room temperature and being left to stir overnight. The solvent was removed *in vacuo* and the residual waxy solid was suspended in anhydrous MeCN (20 mL) and stirred over ice. Anhydrous benzyl alcohol (0.5 mL, 5 mmol, 1 eq.) was added in one portion followed by slow dropwise addition of tetrazole (11.4 mL, 5.15 mmol, 1.03 eq.) and stirred over ice for 10 minutes before heating to room temperature. After 1 hour, the solution was cooled in an ice bath and anhydrous ethanol (0.3 mL, 5 mmol, 1 eq.) was added followed by slow dropwise addition of tetrazole (11.4 mL, 5.15 mmol, 1.03 eq.), the reaction stirred over ice for 10 minutes before heating to room temperature and being left to stir overnight. Saturated NH_4Cl (20 mL) was added, and the volatile were removed *in vacuo*. The biphasic solution was washed with EtOAc (3 × 20 mL), the combined organic layers were washed with brine (1 × 20 mL) and dried over $MgSO_4$. The volatiles were removed *in vacuo* producing a colourless oil (2.53 g), LCMS analysis: $[M+H]^{1+} = 365$. The oil was resuspended in EtOAc (20 mL) and washed with H_2O_2 (30% wt, 10 mL), water (2 × 10 mL), sodium metabisulphite (5%) (3 × 10 mL), water (1 × 10 mL), brine (1 × 10 mL) and dried over $MgSO_4$. The volatiles were removed *in vacuo* producing a colourless oil (1.75 g), LCMS analysis: $[M+H]^{1+} = 381$.

The crude colourless oil (1.75 g) was dissolved in THF (120 mL) and H_2O (10 mL). TBAF (1M in THF with 5% H_2O, 6 mL, 6 mmol, 1.2 eq.) was added in one portion, the solution displayed a red-tinge and was stirred for 25 minutes. The mixture was poured into Et_2O (100 mL) and saturated NH_4Cl (50 mL). The mixture was shaken and the aqueous and organic phases collected. The aqueous phase was subsequently washed with Et_2O (4 × 50 mL), the combined organic phases were dried over $MgSO_4$, and the volatiles were removed *in vacuo*. LCMS analysis: $[M+NH_4]^{1+} = 242$, $[2M + H]^{1+} = 449$. Column chromatography was performed in hexane: ethyl acetate (3:7) and the title compound was isolated as a dark-purple oil.

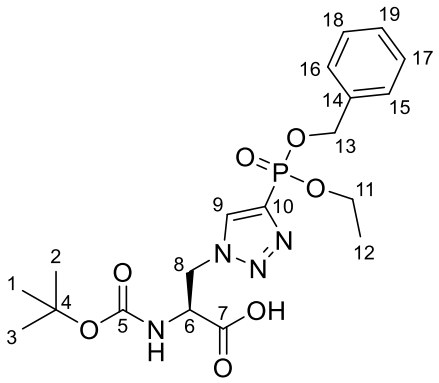


Benzyl-ethyl ethynyl phosphonate (260 mg, 1.16 mmol, 23%), R_f (Hex: EtOAc 7:3) = 0.43, ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.43 – 7.32 (m, 5H, H6-11), 5.16 (dd, J = 11.8, 8.4 Hz, 1H, H5), 5.12 (dd, J = 11.8, 8.5 Hz, 1H, H5), 4.18 (dq, J = 8.6, 7.1 Hz, 2H, H3), 2.90 (d, J = 13.4 Hz, 1H, H1), 1.36 (td, J = 7.1, 0.7 Hz, 3H, H4); ^{13}C NMR (126 MHz, $CDCl_3$): δ = 135.5 (d, J = 7.5 Hz, C6), 128.80 (C9), 128.76 (C8, C10), 128.15 (C7, C11), 88.1 (d, J = 51.2 Hz, C1), 74.34 (d, J = 291.7 Hz, C2), 68.8 (d, J = 5.2 Hz, C5), 63.9 (d, J = 5.7 Hz, C3), 16.2 (d, J = 51.2 Hz, C4); ^{31}P NMR (162 MHz, $CDCl_3$): δ = -8.13. HRMS analysis: observed $[M+Na]^+$ = 247.050, $[2M+Na]^+$ = 471.110; expected $[M+Na]^+$ = 247.049, $[2M+Na]^+$ = 471.109

6.14 Boc-protected benzyl-ethyl phosphonotriazolylalanine (pTza) (65)

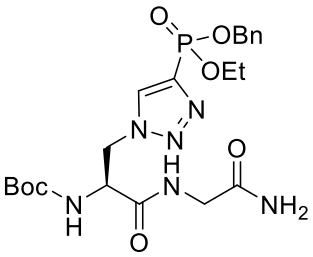


Benzyl-ethyl ethynyl phosphonate (110 mg, 0.491 mmol, 1.09 eq.) was dissolved in THF (3 mL). A solution of $CuSO_4$ (8 mg, 0.05 mmol, 0.11 eq.) and sodium ascorbate (28 mg, 0.14 mmol, 0.29 eq.) in H_2O (6 mL) was added. Boc-3-azido-alanine-OH (dicyclohexylammonium salt) (185 mg, 0.450 mmol, 1 eq.) in THF (3 mL) was added and the reaction was left to stir at room temperature. After 24 hours, the volatiles were removed *in vacuo*, the remaining solution was acidified to pH 1 in 1M HCl. The product was extracted in EtOAc (3 × 15 mL), HCl 1M (1 × 15 mL) and dried over $MgSO_4$. The volatiles were removed *in vacuo*. The crude material was purified by column chromatography eluting in Hex: EtOAc (3: 7) up to EtOAc (100%), giving the title compound a brown oil.



Boc-protected benzyl-ethyl protected phosphonotriazolylalanine (122 mg, 0.268 mmol, 55%), R_f = , ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.38-7.31 (m, 5H, H15-19), δ = 5.48 (d, 1H, H6), δ = 5.16 (m, 2H, H13), δ = 4.95 (s, 2H, H8), 4.76 (s, 1H, H9), 4.18 (t, 2H, H11), 1.45 (s, 9H, H1-3), 1.31 (t, 3H, H12); ^{13}C NMR (126 MHz, $CDCl_3$): δ = 170.4 (s, C7), 155.4 (s, C5), 135.3 (s, C14/C19), 128.8 (d, C17-18), 128.3 (d, C15-C16), 80.9 (s, C4), 69.0 (s, C13), 64.2 (s, C11), 53.8 (s, C6), 51.4 (s, C8), 28.4 (s, C1-3), 16.3 (s, C12); ^{31}P NMR (162 MHz, $CDCl_3$): δ = 7.14. LCMS analysis: observed $[M+H]^+$ = 455.11, expected $[M+H]^+$ = 455.16

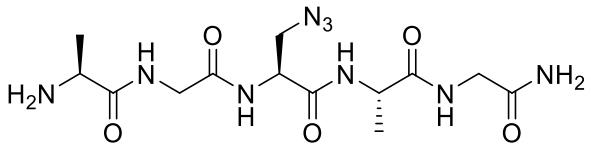
6.15 Boc-protected benzyl-ethyl phosphonotriazolylalanine glycinamide (66)



Boc-protected benzyl-ethyl phosphonotriazolylalanine (122 mg, 0.268 mmol, 1 eq.) was dissolved in anhydrous DCM (5 mL). Oxyma pure (54 mg, 0.38 mmol, 1.4 eq.) and DIC (52 µL, 0.33 mmol, 1.2 eq.) were added and the solution was stirred for one hour. A separate solution of glycinamide hydrochloride (37 mg, 0.33 mmol, 1.2 eq.) and DIPEA (94 µL, 0.54 mmol, 2.5 eq.) was prepared. With slow-dropwise addition, the pTza-oxyma solution was added slowly to the glycinamide-DIPEA mixture. The solution was stirred for two hours at room temperature. LCMS analysis showed the presence of $[M+H]^+$ = 511.12 along with some remaining $[M+H]^+$ = 455.11. An additional aliquot of glycinamide hydrochloride (37 mg, 0.33 mmol, 1.2 eq.) was added with DIC (52 µL, 0.33 mmol, 1.2 eq.). After 48 hours, LCMS analysis showed predominant presence of $[M+H]^+$ = 511.13.

The title compound was never isolated by column chromatography.

6.16 H-AlaGlyDap(N_3)AlaGly-NH_2 (67)

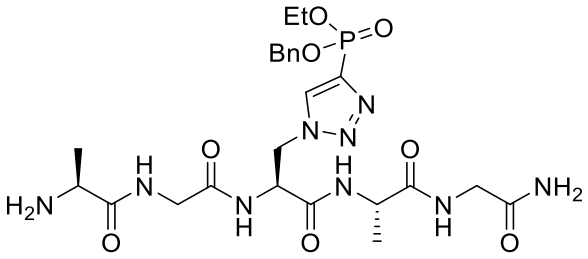


Rink amide MBHA resin (0.610 mmol/g loading, 353 mg, 0.215 mmol) was swollen in DMF (1 mL) for 1 hour, in a polyprep SPPS tube. DMF was filtered and the resin deprotected with a solution of piperidine (20% in DMF) for 10 minutes. The piperidine was filtered and an additional wash with piperidine (20% in DMF) was performed for 5 minutes. The piperidine was filtered and the resin washed with DMF (5 × 2 mL). A solution of Fmoc-Gly-OH (149 mg, 0.501 mmol, 2.5 eq.), Oxyma Pure (72 mg, 0.50 mmol, 2.5 eq.) and DIC (73 μL, 0.47 mmol, 2.35 eq.) was added to the resin and agitated for 4 hours. The amino acid solution was filtered off and a solution of acetic anhydride: pyridine (3:2) was added to the resin and agitated for 30 minutes before being filtered off. The resin was washed with DMF (3 × 2 mL), DCM (3 × 2 mL) and MeOH (3 × 2 mL). The loaded glycine residue was deprotected with piperidine (20% in DMF) (2 × 2 mL × 5 minutes), followed by DMF (1 × 2 mL), DCM (1 × 2 mL) and DMF (1 × 2 mL) washes. A solution of Fmoc-Ala-OH (250 mg, 0.804 mmol, 4 eq.), Oxyma Pure (114 mg, 0.802 mmol, 4.01 eq.) and DIC (124 μL, 0.792 mmol, 3.96 eq.) in DMF was added to the resin and the polyprep tube was agitated for 1 hour. The amino acid solution was filtered off and the resin was washed with DMF (1 × 2 mL), DCM (1 × 2 mL), DMF (1 × 2 mL). The resin was washed with piperidine (20% in DMF) (2 × 2 mL × 5 minutes), the piperidine solution was filtered off and the resin was washed with DMF (1 × 2 mL), DCM (1 × 2 mL) and DMF (1 × 2 mL). This procedure was repeated for the following amino acids: Fmoc-Dap(N_3)-OH (204 mg, 0.579 mmol, 2.90 eq.), Fmoc-Gly-OH (238 mg, 0.801 mmol, 4 eq.) and Fmoc-Ala-OH (250 mg, 0.804 mmol, 4.02 eq.). After the final coupling, the resin was washed with DMF (3 × 2 mL), DCM (3 × 2 mL) and Et_2O (3 × 2 mL) and left to dry on a vacuum manifold. The dried resin was cleaved in a cocktail of TFA: H_2O: TIS (95%: 2.5%: 2.5%) (2 mL) for 2 hours. The filtered solution was decanted into ice-cold Et_2O (20 mL). The solution was centrifuged at 4000 rpm for 10 minutes, the supernatant was decanted and an additional volume of Et_2O (20 mL) was added, the precipitate was agitated and centrifuged at 4000 rpm for 10 minutes. The

supernatant was discarded and the precipitate dried over a stream of N_2. The crude peptide was suspended in H_2O and lyophilised.

The title compound was obtained as a crude white powder (82 mg, 0.21 mmol, 97%) without further purification. LCMS analysis: observed $[M+H]^+$ = 386.04, expected $[M+H]^+$ = 386.18

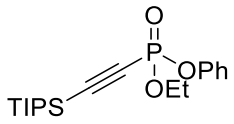
6.17 H-AlaGlypTza(OBnOEt)AlaGly-NH_2 (68)



Benzyl-ethyl ethynyl phosphonate (66 mg, 0.29 mmol, 1.38 eq.) was dissolved in THF (6 mL). A solution of $CuSO_4$ (5 mg, 0.03 mmol, 0.14 eq.) and sodium ascorbate (16 mg, 0.080 mmol, 0.38 eq.) in H_2O (3 mL) was added. Crude H-AlaGlyDap(N_3)AlaGly-NH_2 (82 mg, 0.21 mmol, 1 eq.) was dissolved in THF:H_2O (1:1) (6 mL) and added to the reaction mixture. The reaction was left to stir overnight. The reaction changed from clear to cloudy. The volatiles were removed *in vacuo*. The solution was washed with Et_2O (3 × 30 mL) and the aqueous layer was lyophilised to yield

H-AlaGlypTza(OBnOEt)AlaGly-NH_2 (68 mg, 0.11 mmol, 52%) without further purification. HRMS analysis observed: $[M+H]^+$ = 610.2501, expected: $[M+H]^+$ = 610.2492

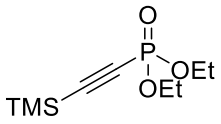
6.18 Phenyl-ethyl [(triisopropylsilyl)ethynyl] phosphonate (77)



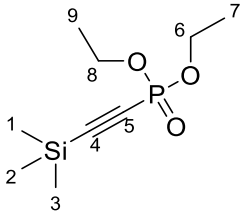
TIPS-acetylene (500 µL, 2.23 mmol, 1.34 eq.) was added to ice-cold anhydrous THF (40 mL), followed by dropwise addition of isopropyl magnesium chloride (2M in THF) (940 µL, 1.88 mmol, 1.13 eq.). The solution was stirred at room temperature for one hour. The solution was cooled over ice and $ClP(NEt_2)_2$ (350 µL, 1.66 mmol, 1 eq.) was added slowly dropwise. The reaction was left to stir for one hour. The solvent was removed *in vacuo* before resuspension in anhydrous acetonitrile (20 mL) and cooled over ice. Phenol (0.5M in MeCN) (156 mg, 3.30 mL, 1.65 mmol, 0.99 eq.) was added followed by dropwise addition of 5-(ethylthio)-1H-tetrazole (ETT) (0.5M in MeCN) (216 mg, 3.30 mL, 1.65 mmol, 0.99 eq.), the reaction was allowed to stir at room temperature for one hour. The solution was cooled over ice and anhydrous ethanol (100 µL, 1.66 mmol, 1 eq.) was added followed by dropwise addition of ETT (0.5M in MeCN) (216 mg, 3.30 mL, 1.65 mmol, 0.99 eq.), the reaction was stirred overnight at room temperature. Saturated NH_4Cl (20 mL) was added to the reaction mixture, and the volatiles were removed *in vacuo*. The biphasic solution was washed with EtOAc (3 × 20 mL), brine (1 × 20 mL), H_2O_2 (30 wt%) (10 mL), H_2O (2 × 10 mL), sodium metabisulphite (5%) (3 × 10 mL), H_2O (1 × 10 mL), brine (1 × 10 mL) and dried over $MgSO_4$. The organic phase was concentrated *in vacuo*. An attempt at purification by column chromatography in Hex: EtOAc (3:7) gave a clear-colourless oil, analysis showed remaining impurities.

The title compound was obtained as a clear-colourless oil (93 mg, 0.25 mmol, 15%), LCMS analysis $[M+H]^+$ = 366.68 and $[2M + H]^+$ = 733.34. NMR analysis is inconclusive; phenyl regions do not integrate and match with ethyl regions so cannot be assigned clearly.

6.19 Diethyl [(trimethylsilyl)ethynyl] phosphonate (53)



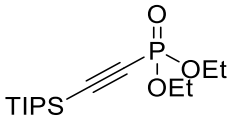
TMS-acetylene (500 µL, 3.83 mmol, 1.34 eq.) was dissolved in ice-cold anhydrous THF (25 mL). iPr-MgCl (2M in THF, 1.6 mL, 3.2 mmol, 1.12 eq.) was added dropwise to the solution and stirred at room temperature for 1 hour. The solution was cooled and $ClP(NEt_2)_2$ (600 µL, 2.86 mmol, 1 eq.) was added dropwise. The solution was stirred at room temperature for 1 hour before the solvent was removed *in vacuo*. Anhydrous ethanol (340 µL, 5.82 mmol, 2 eq.) was added followed by dropwise addition of 5-(Ethylthio)-1H-tetrazole (0.5M in MeCN, 5.72 mmol, 11 mL) to the ice-cold solution. The mixture was stirred overnight at room temperature. Saturated NH_4Cl (20 mL) was added to the solution, and the crude mixture was extracted in EtOAc (3 × 20 mL), H_2O_2 (30 wt%, 1 × 10 mL), H_2O (2 × 10 mL), sodium metabisulfite (5%) (3 × 10 mL), H_2O (2 × 10 mL), brine (1 × 20 mL). The combined organic extracts were dried over $MgSO_4$ and concentrated *in vacuo*. Column chromatography in Hex: EtOAc (2:3) (R_f = 0.44) did not give pure product, some impurity remained.



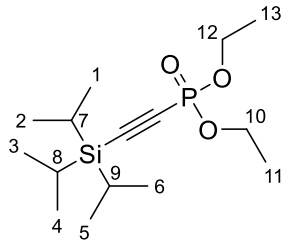
Diethyl [(trimethylsilyl)ethynyl]phosphonate (130 mg, 0.55 mmol, *14%), R_f (Hex:EtOAc 2:3) = 0.44, ^{1}H NMR (400 MHz, $CDCl_3$): δ = 4.18 (m, 4H, H6 & H8), 1.38 (t, J = 7.1, 6H, H7 & H9), 0.24 (s, 9H, H1-3); ^{13}C NMR (126 MHz, $CDCl_3$): δ = 63.5 (d, J = 5.6 Hz, C6/C8), 16.2 (d, J = 6.8 Hz, C7/C9), -0.8 (s, C1-3); ^{31}P NMR (162 MHz, $CDCl_3$): δ = -8.34, LCMS Analysis: $[M+H]^+$ = 234.3, $[2M+H]^+$ = 468.7, expected: $[M+H]^+$ = 235.0, $[2M+H]^+$ = 469.2. * Impurity still evident by ^{1}H NMR analysis.

Analytical data consistent with Acheson *et al.* (253) and Dutremez *et al.* (188)

6.20 Diethyl [triisopropylsilyl)ethynyl] phosphonate (84)



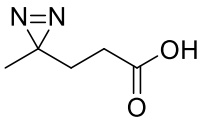
TIPS-acetylene (1 mL, 4.5 mmol, 1.34 eq.) was dissolved in anhydrous THF (50 mL). iPr-MgCl (2M in THF, 1.9 mL, 3.8 mmol, 1.13 eq.), was added dropwise to the ice-cold solution. After stirring at room temperature for 1 hour, $ClP(N^iPr_2)_2$ (894 mg, 3.35 mmol, 1 eq.) was suspended in anhydrous THF (15 mL) and added dropwise to the ice-cold reaction mixture. The mixture was stirred at room temperature overnight. The solvent was removed *in vacuo* and the mixture was resuspended in MeCN (20 mL). Anhydrous EtOH (0.4 mL, 6.7 mmol, 2 eq.) was added. The reaction was cooled over ice and ETT (875 mg, 6.7 mmol, 2 eq.) was added dropwise as a solution in MeCN (13 mL). The mixture was stirred for 1 hour. Saturated NH_4Cl (20 mL) was added and the mixture was extracted in EtOAc (3 × 20 mL), H_2O_2 (1 × 10 mL), H_2O (2 × 10 mL), sodium metabisulfite (5% w/v) (3 × 10 mL), H_2O (2 × 10 mL). The combined organic extracts were dried over $MgSO_4$ and concentrated *in vacuo*. The crude pale-yellow oil (0.94 g, 2.95 mmol) was purified by column chromatography, R_f = (Hex: EtOAc (4:1)) = 0.24.



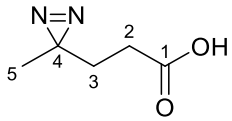
Diethyl [(triisopropylsilyl)ethynyl] phosphonate (0.15 g, 0.47 mmol, 14%), R_f (Hex: EtOAc 4:1) = 0.44, ^1H NMR (400 MHz, $CDCl_3$): δ = 4.22-4.12 (m, 4H, H10/H12), 1.37 (td, 6H, H11/H13), 1.10 (m, 21H, H1-9); ^{31}P NMR (162 MHz, $CDCl_3$): δ = -8.48; LCMS analysis: $[M+H]^+$ = 319.2, $[2M+H]^+$ = 637.2, expected: $[M+H]^+$ = 319.18, $[2M+H]^+$ = 637.3

Analytical data in alignment with Perez *et al.* (254)

6.21 3-(3-methyl-3H-diaziren-3-yl)propanoic acid



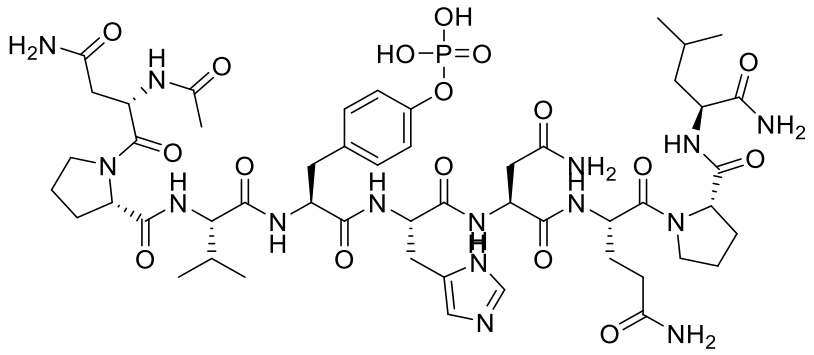
Levulinic acid (1.0 g, 8.6 mmol) was dissolved in liquid NH_3 (30 mL). The solution was stirred with the condenser kept at -78˚C. HOSA (1.13g, 10 mmol) was added dropwise as a solution in MeOH (15 mL). The reaction was stirred overnight at room temperature N_2 was flushed through the apparatus to ensure remaining NH_3 was expelled. The solution was then filtered and washed with MeOH; the volatiles were removed *in vacuo*. The oil was dissolved in anhydrous MeOH (15 mL) and cooled to 0˚C. Et_3N (1.6 mL, 11 mmol) was added dropwise to the solution. Iodine (3.8 g, 15 mmol) in anhydrous MeOH (20 mL) was added dropwise until the solution maintained a dark brown colour, the reaction was stirred for 1 hour at 0˚C. After suspending in EtOAc (15 mL), the solution was washed with 1M HCl (15 mL), 10% sodium thiosulphate (15 mL) and brine (15 mL). The organic phase was dried over $MgSO_4$ and concentrated *in vacuo*. In accordance with methodology by Kambe *et al.* (255)



3-(3-methyl-3H-diaziren-3-yl)propanoic acid (210 mg, 1.64 mmol, 19%), ^1H NMR (400 MHz, $CDCl_3$): δ = 2.24 (t, 2H, 7.7 Hz), 1.72 (t, 2H, 7.7 Hz), 1.05 (s, 3H); ^{13}C NMR (126 MHz, $CDCl_3$): δ = 177.5 (s, C1), 29.8 (s, C4), 29.5 (s, C3), 28.5 (s, C2), 19.8 (s, C1).

Analytical data in alignment with Wright *et al.* (256)

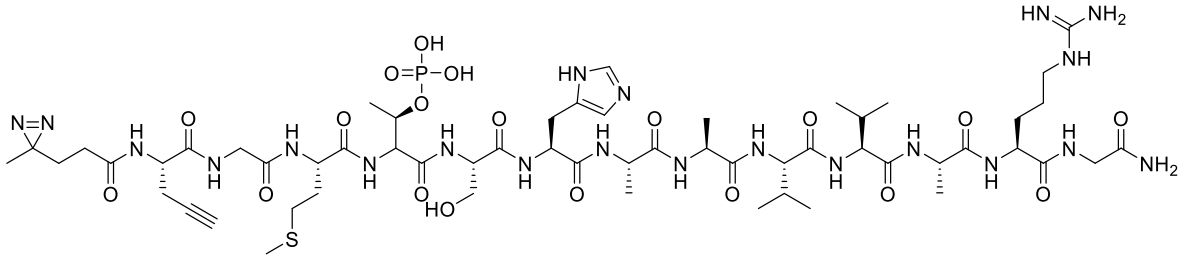
6.22 Ac-NPVpYHNQPL-NH_2 (88)



Synthesised on CEM Liberty Blue: Rink amide MBHA resin (0.55 mmol/g loading, 182 mg, 0.100 mmol), Fmoc-Leu-OH (0.39 g, 1.1 mmol, 5.5 mL), Fmoc-Pro-OH (0.74 g, 2.2 mmol, 11 mL), Fmoc-Glu(OtBu)-OH (0.47 g, 1.1 mmol, 5.5 mL), Fmoc-Asn(Trt)-OH (1.31 g, 2.2 mmol, 11 mL), Fmoc-His(Trt)-OH (0.68 g, 1.1 mmol, 5.5 mL), Fmoc-Tyr(PO(OBzl)OH)-OH (0.56 g, 1.1 mmol, 5.5 mL), Fmoc-Valine (0.39 g, 1.1 mmol, 5.5 mL). The N-terminal Fmoc loaded resin was deprotected with piperidine (20% in DMF) (2 × 2 mL) and washed with DMF (5 × 2 mL). N-terminal acetylation was performed with a solution Ac_2O and DIPEA (3:2) (5 mL). After 30 minutes, the acetylated peptide was washed with DMF (3 × 2 mL), DCM (3 × 2 mL) and Et_2O (3 × 2 mL). The resin was left to dry over vacuum manifold. The dry resin was suspended in TFA: TIPS: H_2O (95%: 2.5%: 2.5%) (3 mL) and agitated for 2 hours before decantation into ice-cold Et_2O (30 mL). The solution was centrifuged and the supernatant decanted, Et_2O (15 mL) was added, and the solution was centrifuged (× 2). The peptide was suspended in deionised water (10 mL) and lyophilised.

The title peptide was obtained (30 mg, 0.024 mmol, 24%) – analysis by HPLC showed 87% purity by area. HRMS analysis indicates $[M+H]^+$ = 1202.534, $[M+2H]^{2+}$ = 601.771, expected: $[M+H]^+$ = 1202.535, $[M+2H]^{2+}$ = 601.771.

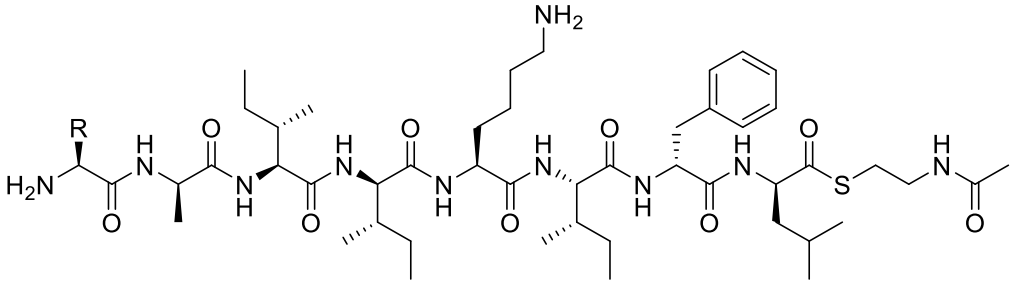
6.23 Diazirine-propargylglycine-GMpTSHAAVVARG-NH_2 (89)



Synthesised on CEM Liberty Blue: Rink amide MBHA resin (0.55 mmol/g loading, 182 mg, 0.100 mmol), Fmoc-Gly-OH (0.66 g, 2.2 mmol, 11 mL), Fmoc-Arg(Pbf)-OH (0.72 g, 1.1 mmol, 5.5 mL), Fmoc-Ala-OH (1.03 g, 3.31 mmol, 16.5 mL), Fmoc-Val-OH (0.75 g, 2.2 mmol, 11 mL), Fmoc-His(Trt)-OH (0.69 g, 1.1 mmol, 5.5 mL), Fmoc-Ser(OtBu)-OH (0.43 g, 1.1 mmol, 5.5 mL), Fmoc-Thr(PO(OBzl)OH)-OH (0.57 g, 1.1 mmol, 5.5 mL), Fmoc-Met-OH (0.41 g, 1.1 mmol, 5.5 mL), Fmoc-propargyl-Gly-OH (0.36 g, 1.1 mmol, 5.5 mL). The Fmoc-protected resin was deprotected manually in a polyprep tube with piperidine (20% in DMF) (2 × 2 mL) (5 minutes) followed by DMF (5 × 2 mL). Diazirine (3-(3-methyl-3H-diaziren-3-yl)propanoic acid) (25 mg, 0.20 mmol, 2 eq.) was dissolved in DMF (3 mL), to which Oxyma Pure (36 mg, 0.25 mmol, 2.5 eq.) and DIC (39 μL, 0.25 mmol, 2.5 eq.) were added. The activated diazirine solution was added to the resin and agitated for 2 hours. The solution was decanted, and the resin was washed with DMF (5 × 2 mL), DCM (2 × 2 mL) and Et_2O (2 × 2 mL). The dried resin was cleaved from the resin with TFA: EDT: H_2O: TIS (94%: 2.5%: 2.5%: 1%) (3 mL). After agitating the solution for 3 hours, the solution was decanted into ice-cold Et_2O (30 mL) and centrifuged. The supernatant was decanted, and the peptide was resuspended in Et_2O (15 mL) and centrifuged (×2). The peptide was resuspended in deionised water (10 mL) and lyophilised.

The title compound was obtained (23 mg, 0.016 mmol, 16%) - analysis by HPLC showed 85% purity by area. HRMS analysis observed: $[M+H]^+$ = 1440.663, $[M+2H]^{2+}$ = 720.832, $[M+2Na]^{2+}$ = 742.814, $[M+2H]^{2+}$ = 720.832, $[M+3H]^{3+}$ = 480.890; expected: $[M+H]^+$ = 1440.660, $[M+2Na]^{2+}$ = 742.814, $[M+2H]^{2+}$ = 720.832, $[M+3H]^{3+}$ = 480.890.

6.24 SNAC-Surugamide sequences (SNAC-DLeuDPheIleLysDIleIleDAlaX-NH_2)



2-chlorotrityl chloride resin (204 mg, 0.300 mmol – based on 1.47 mmol/g loading) was swollen in anhydrous DCM (4 mL) for 30 minutes. A premixed solution of Fmoc-D-Leu-OH (129 mg, 0.365 mmol, 1.21 eq.), DIPEA (247 µL, 1.42 mmol, 3.88 eq. – relative to amino acid) in dry DCM (2.04 mL) was added to the resin solution and stirred for 40 minutes. The resin was subsequently washed with DCM/MeOH/DIPEA (17:2:1) (3 × 2 mL), DCM (3 × 2 mL), DMF (2 × 2 mL) and DCM (2 × 2 mL). The resin was washed with piperidine (20% in DMF) (2 × 2 mL) (5 minutes), DMF (2 × 2 mL), DCM (2 × 2 mL). Fmoc-D-Phe-OH (466 mg, 1.20 mmol, 4.01 eq.), Oxyma Pure (172 mg, 1.21 mmol, 4.05 eq.) and DIC (185 µL, 1.18 mmol, 3.95 eq.). were added to the resin and agitated for 1 hour. The resin was washed with DMF (2 × 2 mL), DCM (2 × 2 mL), piperidine (20% in DMF) (2 × 2 mL) (5 minutes), DMF (2 × 2 mL), DCM (2 × 2 mL). This process was repeated for all subsequent amino acids: Fmoc-Ile-OH (431 mg, 1.22 mmol, 4.07 eq.), Fmoc-Lys(Boc)-OH (568 mg, 1.21 mmol, 4.05 eq.), Fmoc-D-Ile-OH (440 mg, 1.25 mmol, 4.16 eq.), Fmoc-Ile-OH (427 mg, 1.21 mmol, 4.04 eq.) and Fmoc-D-Ala-OH (379 mg, 1.22 mmol, 4.07 eq.). This process was repeated to produce another resin with 0.300 mmol of surugamide heptapeptide.

The peptide loaded resins were divided into six aliquots and each resin was henceforth treated separately. A solution of Boc-Val-OH (90.2 mg, 0.415 mmol, 4.15 eq.), Oxyma Pure (58 mg, 0.41 mmol, 4.1 eq.) and DIC (62.6 µL, 0.400 mmol, 4.00 eq.) in DMF was added to one resin. A solution of Boc-Trp(Boc)-OH (164 mg, 0.405 mmol, 4.05 eq.), Oxyma Pure (61 mg, 0.43 mmol, 4.3 eq.) and DIC (62.6 µL, 0.400 mmol, 4.00 eq.) in DMF was added to the second resin. A solution of Boc-Thr(tBu)-OH (113 mg, 0.410 mmol, 4.10 eq.), Oxyma Pure (57 mg, 0.40 mmol, 4.0 eq.) and DIC (62.6 µL, 0.400 mmol, 4.00 eq.) in DMF was added to the third resin. Boc-Lys(Boc)-OH (142 mg, 0.401 mmol, 4.01 eq.), Oxyma Pure (58 mg, 0.41 mmol, 4.1 eq.) and DIC (62.6 µL, 0.400 mmol, 4.00 eq.) in DMF was added to the fourth resin aliquot. Boc-Ala-OH (81 mg, 0.43 mmol, 4.3 eq.), Oxyma Pure (64 mg, 0.45 mmol, 4.5 eq.) and DIC (62.6 µL, 0.400 mmol, 4.00 eq.) in DMF was added to the fifth resin. A DMF solution of Boc-

Gly-OH (76 mg, 0.43 mmol, 4.3 eq.), Oxyma Pure (61 mg, 0.43 mmol, 4.3 eq.) and DIC (62.6 µL, 0.400 mmol, 4.00 eq.) was added to the six resin aliquots.

The resins were washed with DMF (2 × 2 mL) and DCM (2 × 2 mL) followed by suspension in HFIP: DCM (1:4) (3 mL) for 30 minutes. The eluent was collected and dried to leave crude peptides with N-terminal Boc-valine, Boc-tryptophan and Boc-threonine, Boc-lysine, Boc-alanine and Boc-glycine without further purification.

Crude surugamide-Val (40 mg, 0.039 mmol) was dissolved in DCM (2 mL) to which DIC (7.7 µL, 0.050 mmol, 1.3 eq.) was added. HOBt (7 mg, 0.05 mmol, 1 eq.) was added followed by N-acetylcysteamine (43 µL, 0.40 mmol, 10 eq.)

Crude surugamide-Trp (70 mg, 0.058 mmol) was dissolved in DCM (2 mL) to which DIC (9.4 µL, 0.061 mmol, 1.1 eq.) was added. HOBt (8 mg, 0.06 mmol, 1. eq.) was added followed by N-acetylcysteamine (61 µL, 0.57 mmol, 9.8 eq.)

Crude surugamide-Thr (60 mg, 0.056 mmol) was dissolved in DCM (2 mL) to which DIC (10.4 µL, 0.0671 mmol, 1.20 eq.) was added. HOBt (9 mg, 0.07 mmol, 1 eq.) was added followed by N-acetylcysteamine (59 µL, 0.55 mmol, 9.8 eq.)

Crude surugamide-Lys (250 mg, 0.218 mmol) was dissolved in DCM (2 mL) to which DIC (40.4 µL, 0.261 mmol, 1.20 eq.) was added. HOBt (34 mg, 0.25 mmol, 1.1 eq.) was added followed by N-acetylcysteamine (231 µL, 2.17 mmol, 9.95 eq.)

Crude surugamide-Gly (120 mg, 0.123 mmol) was dissolved in DCM (2 mL) to which DIC (22.7 µL, 0.147 mmol, 1.20 eq.) was added. HOBt (21 mg, 0.16 mmol, 1.3 eq.) was added followed by N-acetylcysteamine (131 µL, 1.23 mmol, 10.0 eq.)

Crude surugamide-Ala (304 mg, 0.308 mmol) was dissolved in DCM (2 mL) to which DIC (57.3 µL, 0.370 mmol, 1.20 eq.) was added. HOBt (53 mg, 0.39 mmol, 1.3 eq.) was added followed by N-acetylcysteamine (327 µL, 3.08 mmol, 10.0 eq.)

The samples were concentrated *in vacuo* and subsequently resuspended in TFA: DCM (9:1) (2 mL). The solutions were stirred for 3 hours at room temperature. After 3 hours, the solvent was removed with an N_2 airflow. The SNAC-peptides were purified by preparative HPLC.

SNAC-surugamide-Val (3 mg, 0.003 mmol, 3%) – analysis by HPLC showed 45% purity by area, HRMS analysis observed: $[M+H]^+$ = 1017.652; $[M+Na]^+$ = 1039.634, $[M+2H]^{2+}$ = 509.330; expected: $[M+H]^+$ = 1017.653, $[M+Na]^+$ = 1039.635, $[M+2H]^{2+}$ = 509.330.

SNAC-surugamide-Trp (6 mg, 0.005 mmol, 5%) – analysis by HPLC showed 63% purity by area, HRMS analysis observed: $[M+H]^+$ = 1104.663 , $[M+Na]^+$ = 1126.644; $[M+2H]^{2+}$ = 552.837; expected: $[M+H]^+$ = 1104.664, $[M+Na]^+$ = 1126.646, $[M+2H]^{2+}$ = 552.835.

SNAC-surugamide-Thr (12 mg, 0.012 mmol, 12%) – analysis by HPLC showed 29% purity by area, HRMS analysis observed: $[M+H]^+$ = 1019.632, $[M+Na]^+$ = 1041.614, $[M+2H]^{2+}$ = 510.319; expected: $[M+H]^+$ = 1019.632, $[M+Na]^+$ = 1041.614, $[M+2H]^{2+}$ = 510.319.

SNAC-surugamide-Lys (8 mg, 0.008 mmol, 8%) – analysis by HPLC showed 63% purity by area, HRMS analysis observed: $[M+H]^+$ = 1046.679, $[M+Na]^+$ = 1068.661, $[M+2H]^{2+}$ = 523.843, $[M+3H]^{3+}$ = 349.564; expected: $[M+H]^+$ = 1046.634, $[M+Na]^+$ = 1068.661, $[M+2H]^{2+}$ = 523.820, $[M+3H]^{3+}$ = 349.564.

SNAC-surugamide-Gly (14 mg, 0.014 mmol, 14%) – analysis by HPLC showed 81% purity by area, HRMS analysis observed: $[M+H]^+$ = 975.605, $[M+Na]^+$ = 1011.603, $[M+2H]^{2+}$ = 488.308; expected: $[M+H]^+$ = 975.606, $[M+Na]^+$ = 1011.603, $[M+2H]^{2+}$ = 488.306.

SNAC-surugamide-Ala (33 mg, 0.033 mmol, 33%) – analysis by HPLC showed 77% purity by area, HRMS analysis observed: $[M+H]^+$ = 989.621, $[M+Na]^+$ = 1011.603, $[2M+H]^+$ = 1978.238, $[M+2H]^{2+}$ = 495.317; expected: $[M+H]^+$ = 989.621, $[M+Na]^+$ = 1011.603, $[2M+H]^+$ = 1978.235, $[M+2H]^{2+}$ = 495.314.

7 General Methods

7.1 Generating Dha on affimer and incubating with heterocycles

Anti CTB affimer 3A2, 40.6 μM (150 μL) was mixed with Dibromoadipamide in DMF (4.06 mM) (150 μL) followed by K_2CO_3 (4.06 mM, 15 μL) and the solution was incubated at 200 rpm, 37˚C.

Anti CTB affimer 3A2 was measured by HRMS every hour for evidence of conversion to Modified anti CTB affimer 3A2 (containing dehydroalanine) – appearance of 17182 Da peak, typically after 3 hours.

The crude solution was buffer exchanged in a G25 Miditrap™ PD Column (Cytiva) with sodium phosphate buffer (50 mM, pH 8 at 25˚C). The resulting 1 mL sample was spin concentrated to 100 μL.

A 1M solution of the heterocycle (imidazole/pyrazole) was produced by solvation in sodium phosphate buffer (50 mM, pH 8 at 25˚C) and incubation at room temperature. The solution was monitored every hour by HRMS to assess conversion.

7.2 UV Crosslinking Assays

Samples of protein and peptide probe were mixed with PBS along with DMSO control solutions. The mixtures were incubated at 25˚C, 700 rpm for 30 minutes. The mixtures were irradiated with UV light (365 nm) for 30 seconds, except the DMSO UV negative control. If a competitor was added, this was done prior to irradiation.

After irradiation 10% SDS (1.5 µL) was added to each Eppendorf followed by click stock (3.6 µL)

Click Stock

Ingredient	**Quantity (µL)**
Tamra-N_3 (10 mM in DMSO)	0.6
$CuSO_4$ (50 mM in H_2O)	1.2
TCEP (50 mM in H_2O)	1.2
TBTA (10 mM in DMSO)	0.6
Total Volume	3.6 (per Eppendorf)

The Eppendorf tubes were subsequently incubated at 25˚C, 700 rpm for 1 hour. After incubation EDTA (0.5M) (0.5 µL) was added to each Eppendorf, the samples were precipitated in ice-cold acetone (100 µL). The samples are vortexed and left in a -20˚C freezer overnight.

The following day, the samples were centrifuged at 4˚C, 13’000g (10 minutes). The pellets were washed with ice-cold methanol (2 × 100 µL). The pellets were left to air-dry (no more than 30 minutes) and subsequently suspended in 2% SDS in PBS (12.5 µL) and 2x sample loading buffer (SLB) (12.5 µL). The samples were boiled for 2 minutes at 95˚C and then centrifuged at 4˚C, 13’000g for 2 minutes.

The samples were loaded onto a gel for SDS-PAGE electrophoresis. A pre-stained protein ladder (10 µL) was loaded alongside the samples in each lane (10 µL). The SDS-PAGE was run at 180V, 400 mA for 50 minutes.

Once run, the gel was imaged for rhodamine staining and subsequently soaked in Coomassie blue stain. The Coomassie blue destained gel can subsequently be compared to the rhodamine gel to assess crosslinking and protein concentration.

7.3 Proteomics Protocol

Peptide probe (20 μM, 25 μL, 10% DMSO) was mixed with HeLa Lysate (2.5 $mgmL^{-1}$, 100 μL) in PBS buffer (125 μL) to form 250 μL probe sample - Probe sample (probe 2 μM, protein 1 $mgmL^{-1}$, PBS buffer (1% DMSO)).

Peptide probe (100 μM, 25 μL, 10% DMSO) was mixed with HeLa Lysate (2.5 $mgmL^{-1}$, 100 μL) in PBS buffer (125 μL) to form 250 μL probe sample - Probe sample (probe 10 μM, protein 1 $mgmL^{-1}$, PBS buffer (1% DMSO)).

Sample	Probe (2 μM)	Probe (10 μM)	DMSO Control
Probe - volume from 10x stock (μL)	25	25	N/A (25 μL of 10% DMSO in PBS stock)
HeLa Lysate (2.5 $mgmL^{=1}$)	100	100	100
PBS (μL)	125	125	125

The samples were incubated (25 ˚C, 1 hour, 750 rpm). Following incubation, the sample were UV-irradiated (365 nm, 1 minute). 10% SDS (12.5 μL) was added to the samples. A biotinylating azide click mix was added to the samples (15 μL).

Click Stock

Compound	**Volume (μL)**
Biotin-N_3 (10 mM in DMSO)	20
$CuSO_4$ (50 mM in H_2O)	40
TCEP (50 mM in H_2O)	40
TBTA (10 mM in DMSO)	20

The samples with click stock were incubated (25 ˚C, 1 hour, 900 rpm) followed by addition of ice-cold acetone (1 mL) and the samples were left in a freezer overnight. The samples were centrifuged (4 ˚C, 20 minutes, 13,000 × g) and the supernatant was removed and ice-cold MeOH (1 mL) was added. The residual pellet was sonicated and centrifuged (4 ˚C, 10 minutes, 13,000 × g). The ice-cold MeOH supernatant was removed, and the process was repeated with

another volume of ice-cold MeOH (1 mL) twice more. After the final washing, the methanol supernatant was removed, and the samples were air-dried for 45 minutes.

Pellets were resuspended in SDS (2 %) in PBS (25 µL) and DTT (0.5M) (0.5 µL). PBS (225 µL) was added to the samples. Samples were centrifuged (25 ˚C, 20 minutes, 13,000 × g) and the supernatants were transferred to new vials. The samples were mixed with avidin resin (30 µL) – avidin resin pre-washed with 0.2 % SDS in PBS (5 × 500 µL). The samples were incubated (25 ˚C, 2 hours, 900 rpm) and the avidin beads were subsequently washed with SDS (0.2%) in PBS (2 × 500 µL), Urea (6M) in PBS (2 × 500 µL) and TEAB (50 mM) (2 × 500 µL) – centrifugation (25 ˚C, 2 minutes, 2000 × g) between every wash step. DTT (1M) (2 µL) was added, and the samples were incubated (55 ˚C, 30 minutes, 950 rpm) followed by washing with TEAB (50 mM) (1 × 500 µL). Iodoacetamide (500 mM, 2 µL) was added to the samples were incubated (30 ˚C, 30 minutes, 950 rpm) (protected from light). The samples were subsequently washed with TEAB (50 mM) (2 × 500 µL) and trypsin in acetic acid (0.2 $mgmL^{-1}$) (5 µL) was added to the samples. The samples were incubated (37 ˚C, 16 hours, 950 rpm) followed by the addition of formic acid (0.5 µL) and TEAB (50 mM) (50 µL). The samples were agitated and then centrifuged (25 ˚C, 5 minutes, 2000 × g). The supernatant was transferred to a clean Eppendorf vial. The residual beads were washed with TEAB (50 mM) (1 × 50 µL), and the combined supernatant layers were dried in an Eppendorf concentrator (3 hours).

7.4 Desalting Protocol

Polyprep tubes were placed on a vacuum manifold (Supelco Visiprep™ SPE). The tubes were washed with methanol: deionised water: formic acid (90: 10: 0.1%) (3 × 1 mL) and deionised water (formic acid 0.1%) (2 × 1 mL). The samples were resuspended in deionised water (formic acid 0.1%) (250 µL) and transferred to the polyprep tubes. Desalting was performed by washing samples with deionised water: methanol (95: 5%) (1 × 1 mL). Clean Eppendorf tubes were placed under the vacuum manifold outlets, and the samples were eluted with a solution of acetonitrile: water: formic acid (50: 50: ~0.1%) (1 × 1 mL). Samples were dried in an Eppendorf concentrator (3 hours). Samples were resuspended in deionised water: acetonitrile: formic acid (98: 2: ~0.1%) (50 µL).

7.5 Cell culture protocol

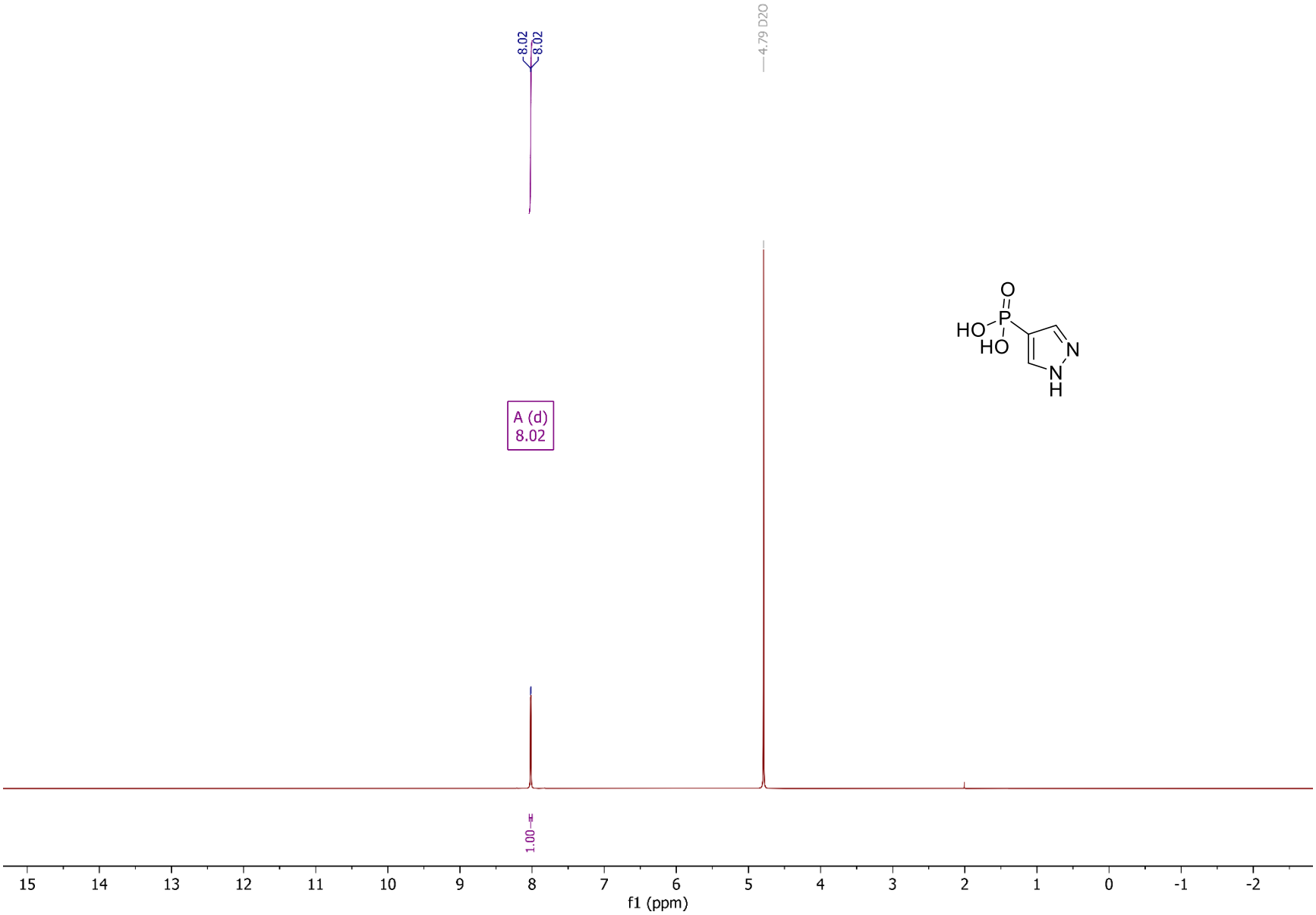
7.5.1 HeLa cell passaging

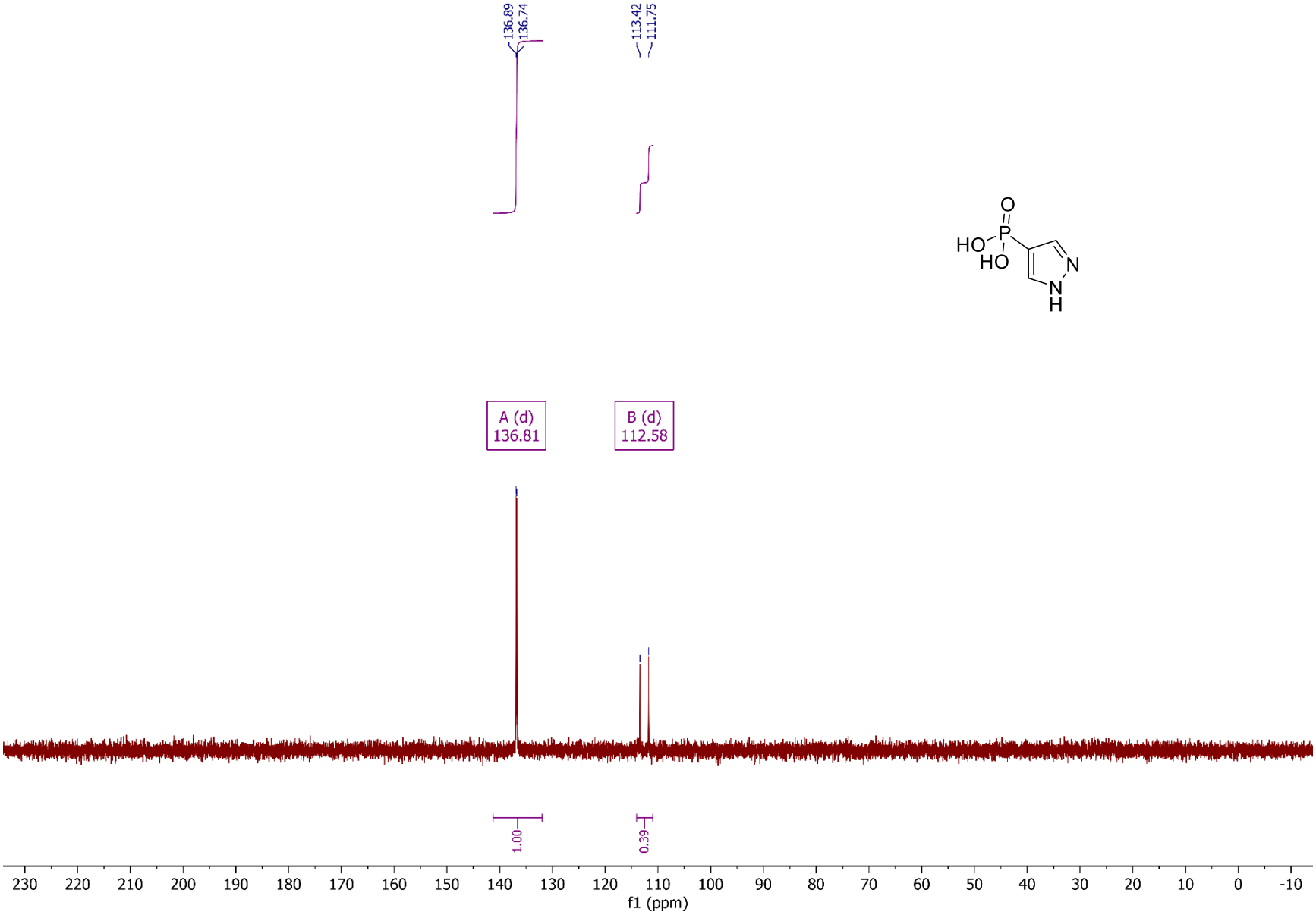
All equipment was sterilised with ethanol except for HeLa culture dish. The old cell culture was removed with a serological pipette and discarded. Using a fresh serological pipette, the cells were washed with PBS (3 mL) and the PBS was removed and discarded. Trypsin/EDTA (1.5 mL) was added to the cells, and they were incubated (37 ˚C, 5 minutes) to detach adherent cells. The enzymatic reaction was stopped by addition of growth medium (Dulbecco's Modified Eagle Medium - supplemented with 10% foetal bovine serum and 1% penicillin/streptomycin) (3 mL). The dish surface was rinsed by aspirating and dispensing the cells a few times before transferring the cells to a fresh falcon tube. The cells were centrifuged (500 × g, 20 ˚C, 5 minutes) and the supernatant was discarded. The cell pellet was reconstituted in growth medium (1 mL). Cell culture (100 µL) was added to growth medium (10 mL) in a fresh dish. The cells were incubated (37 ˚C, 5% CO_2).

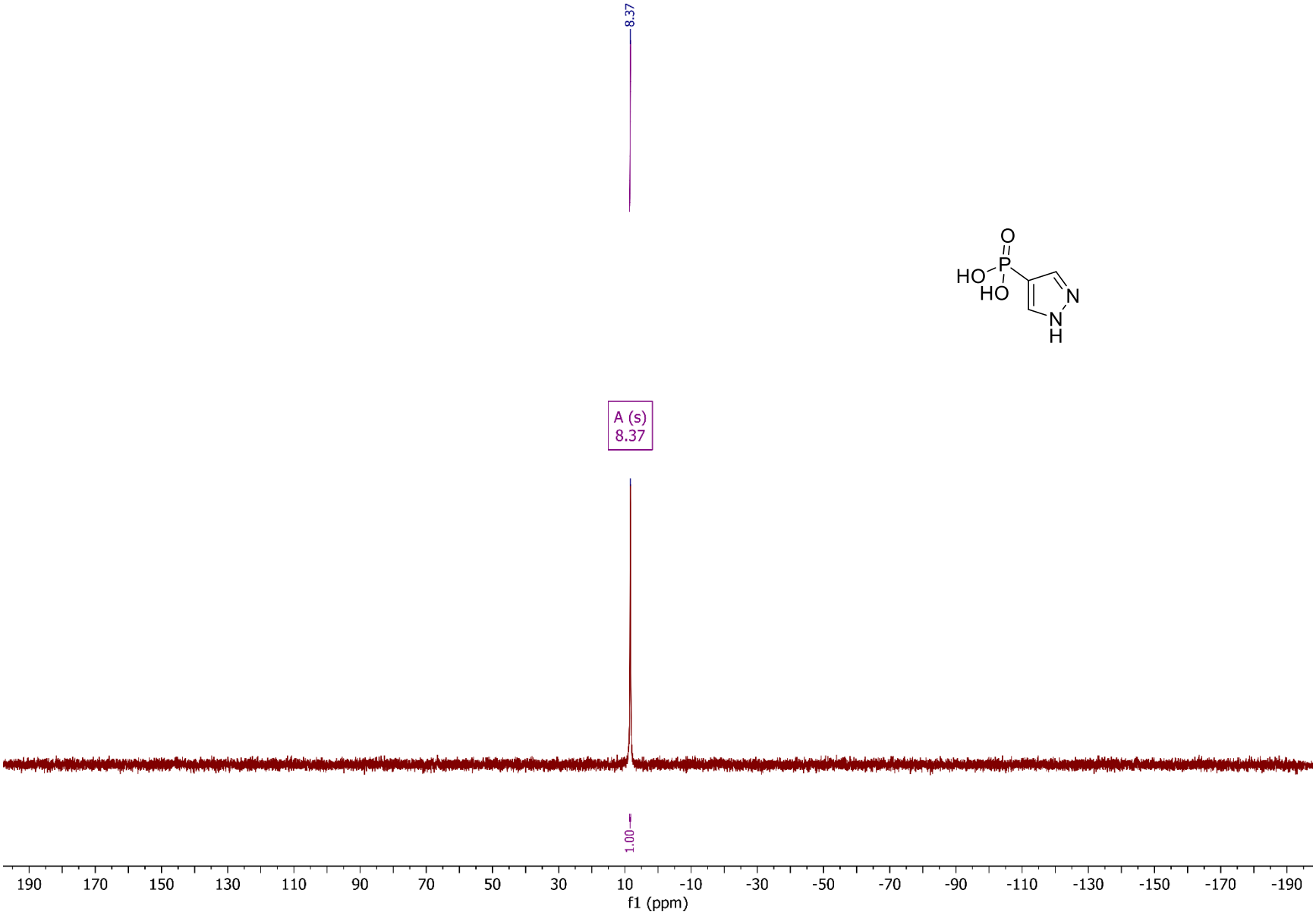
7.5.2 HeLa solution preparation

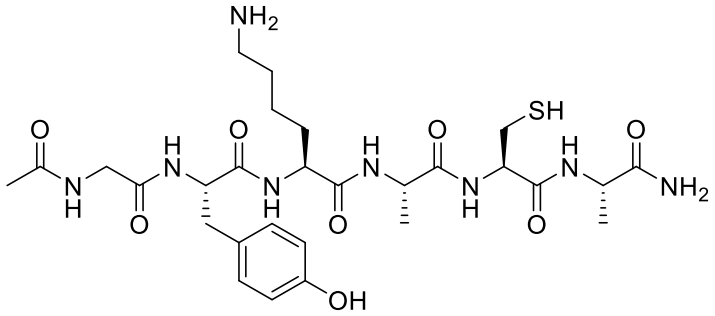
One phosphatase inhibitor tablet (PhosSTOP™, Roche) and one protease inhibitor tablet (cOmplete™ Protease Inhibitor Cocktail, Roche) were diluted to 1x working concentration in a Tris buffer (25 mM, pH 8, 25 ˚C). HeLa cell pellet was suspended in the Tris buffer (600 µL) and the suspended pellet was sonicated over ice. The solution was centrifuged (500 × g, 5 minutes, 4 ˚C) and the supernatant was collected. Protein concentration was determined by Bradford assay. The HeLa protein was aliquoted, flash frozen and subsequently stored in a freezer (-80 ˚C).

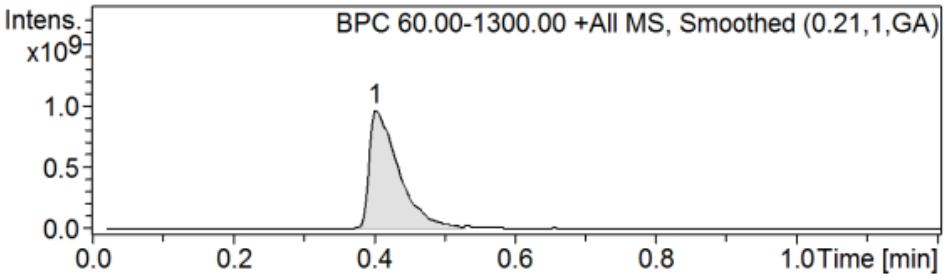
8 Appendix

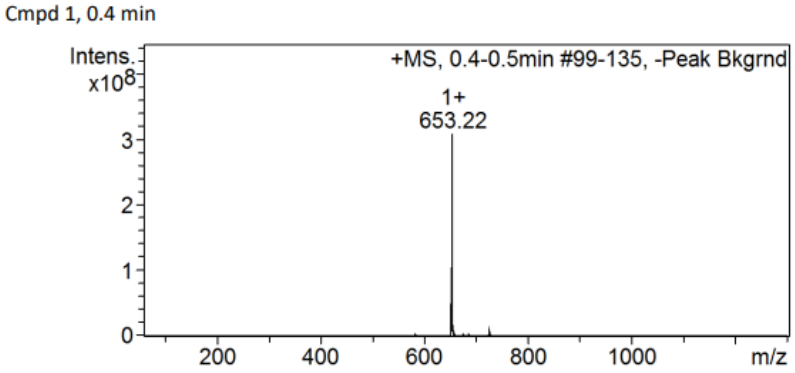

8.02
8.02
4.79 D2O
A (d)
8.02
1.00
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
-2
f1 (ppm)

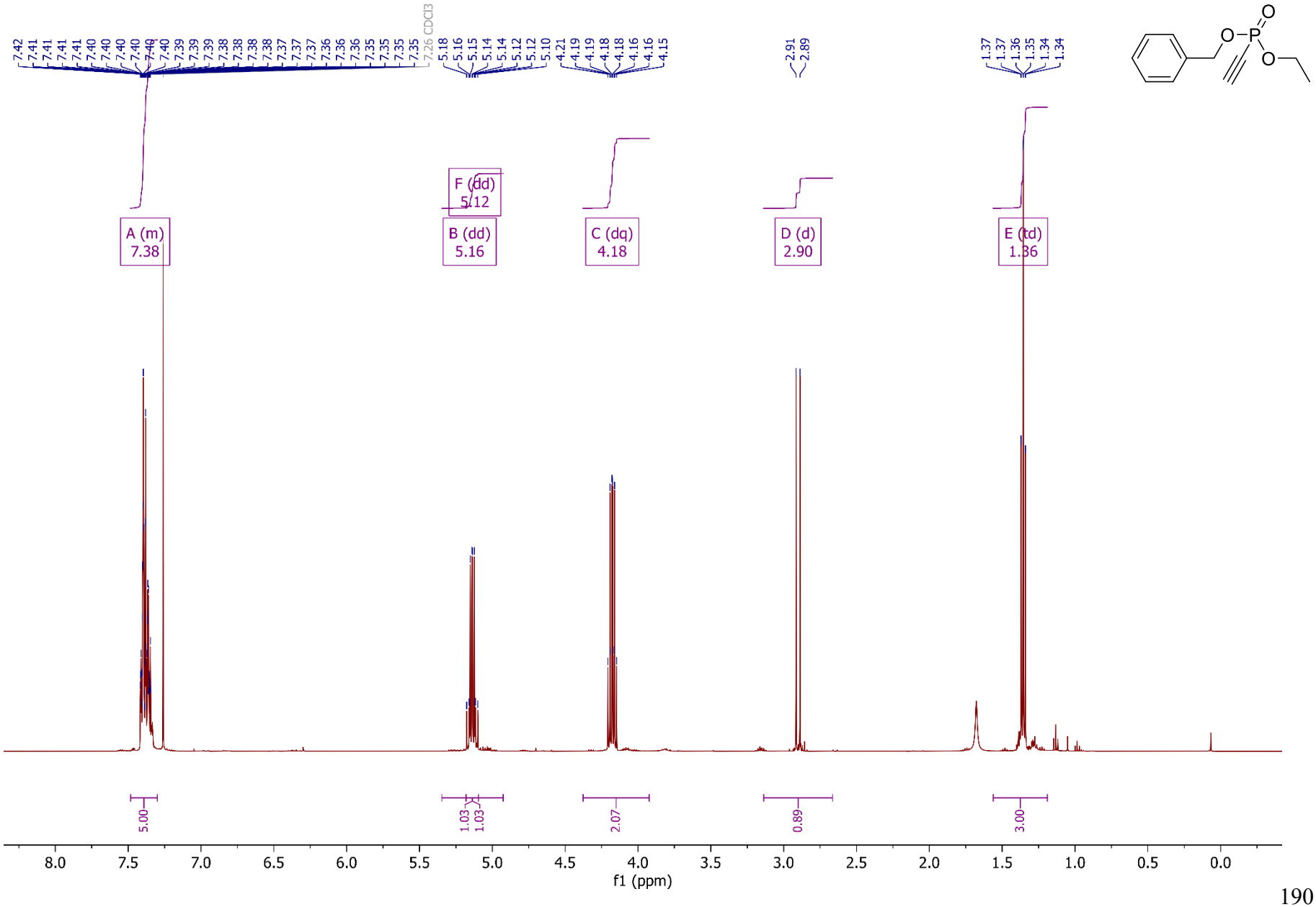

136.89
136.74
113.42
111.75
A (d)
136.81
B (d)
112.58
1.00
0.39
230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

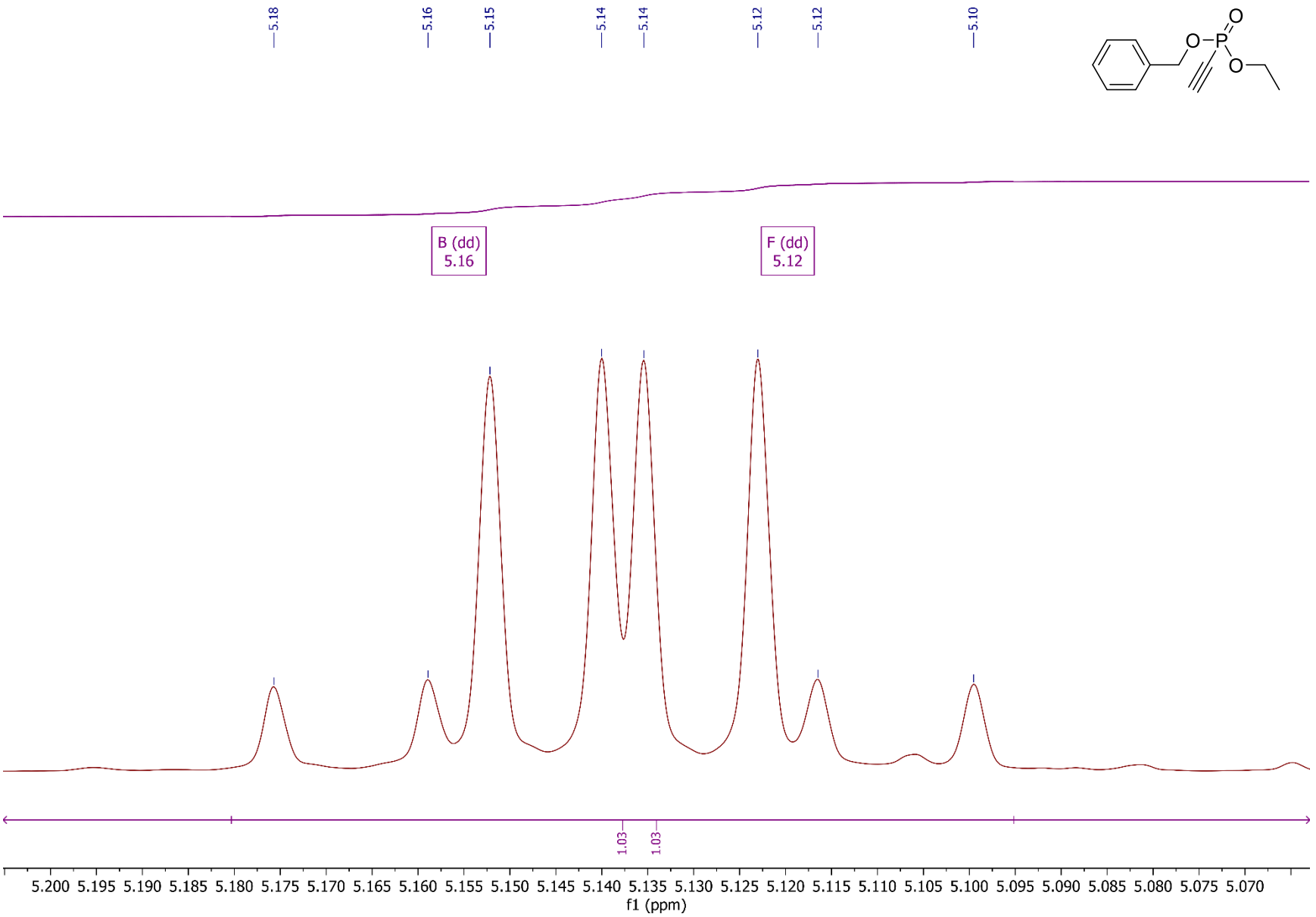

8.37
A (s)
8.37
1.00
190 170 150 130 110 90 70 50 30 10 -10 -30 -50 -70 -90 -110 -130 -150 -170 -190
f1 (ppm)

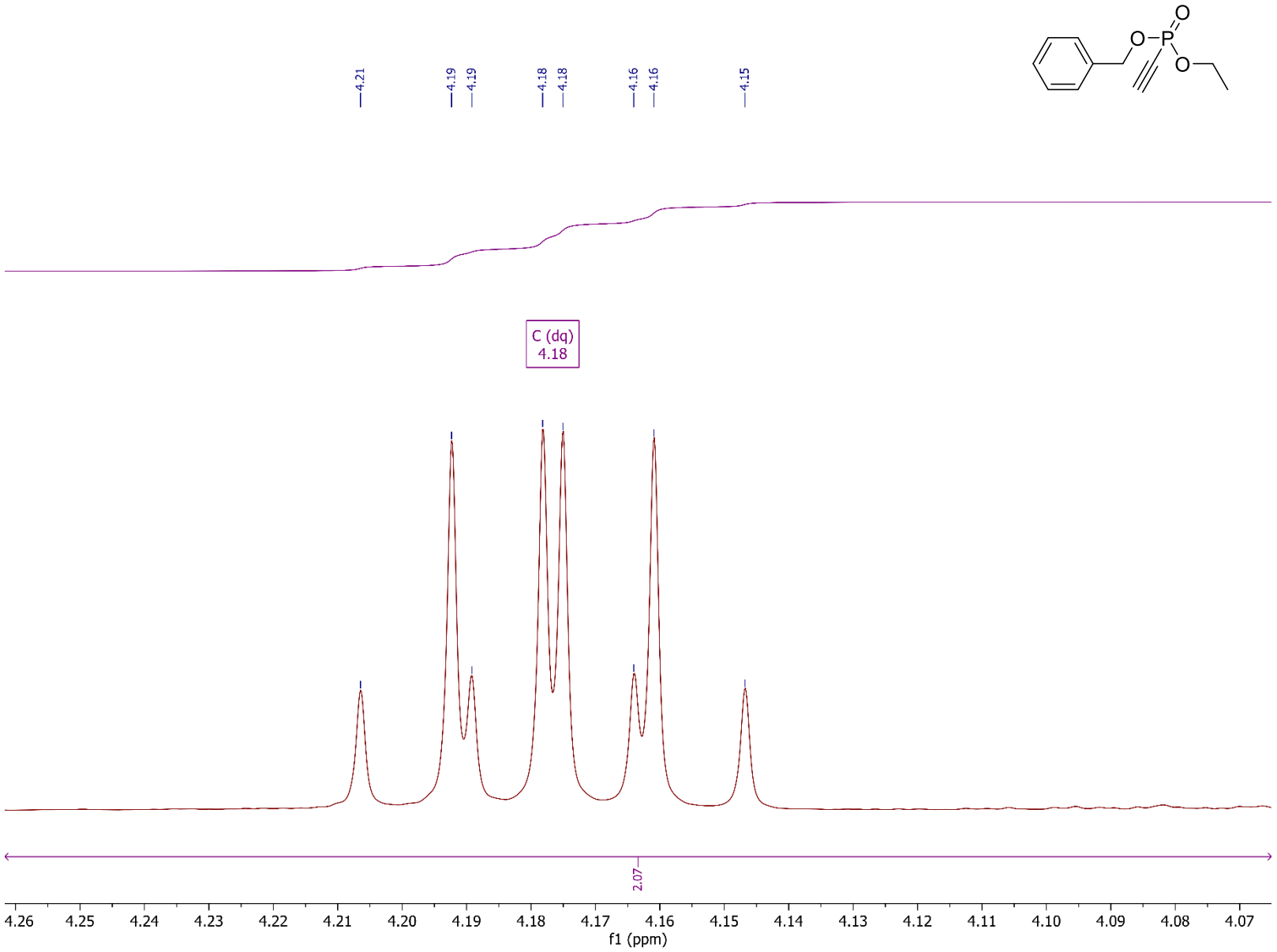

NH2
SH
O
H
N
O
H
N
O
H
N
O
H
N
O
NH2
OH

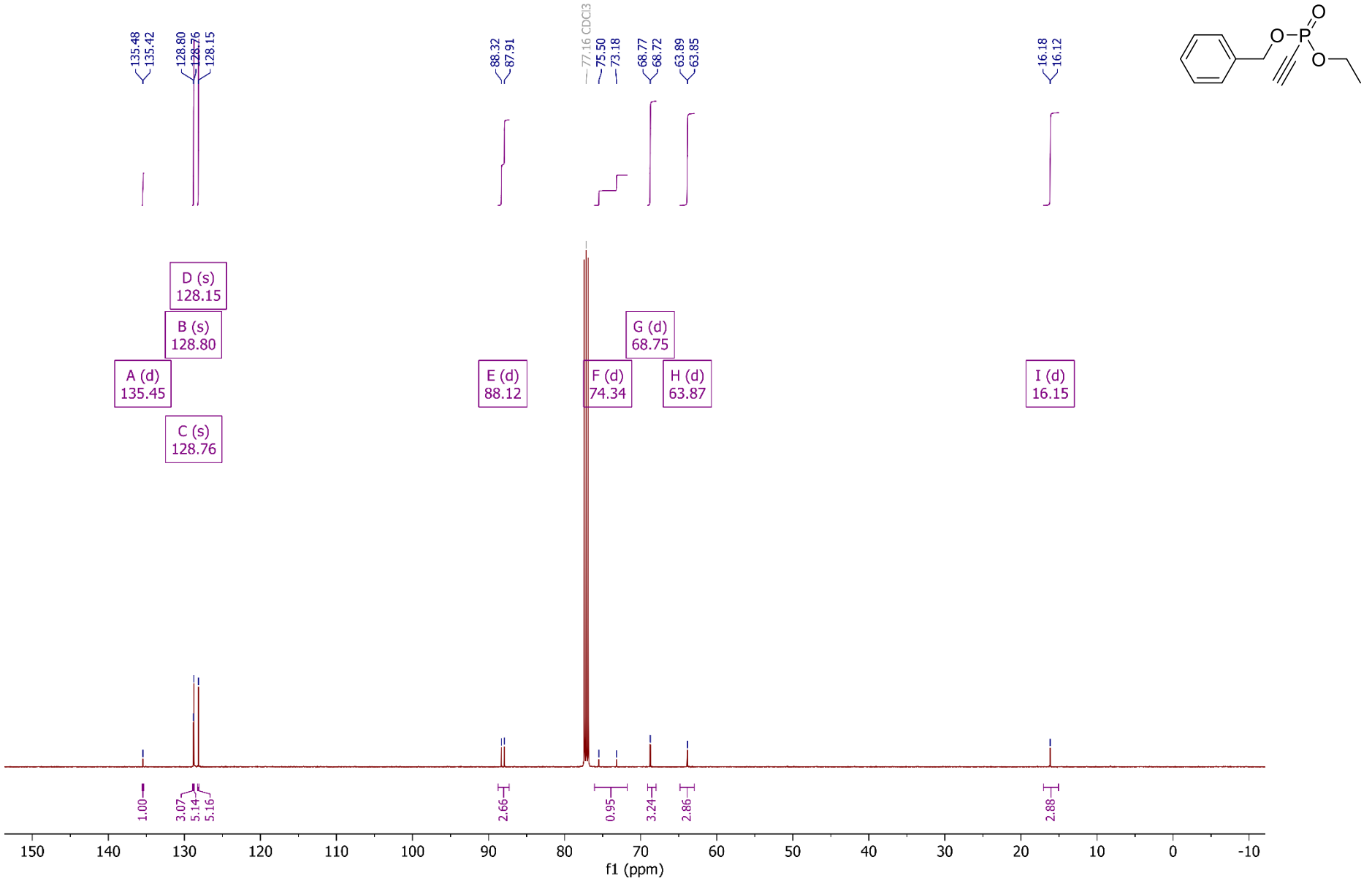

Intens.
x10^9
BPC 60.00-1300.00 +All MS, Smoothed (0.21,1,GA)
1
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Time [min]

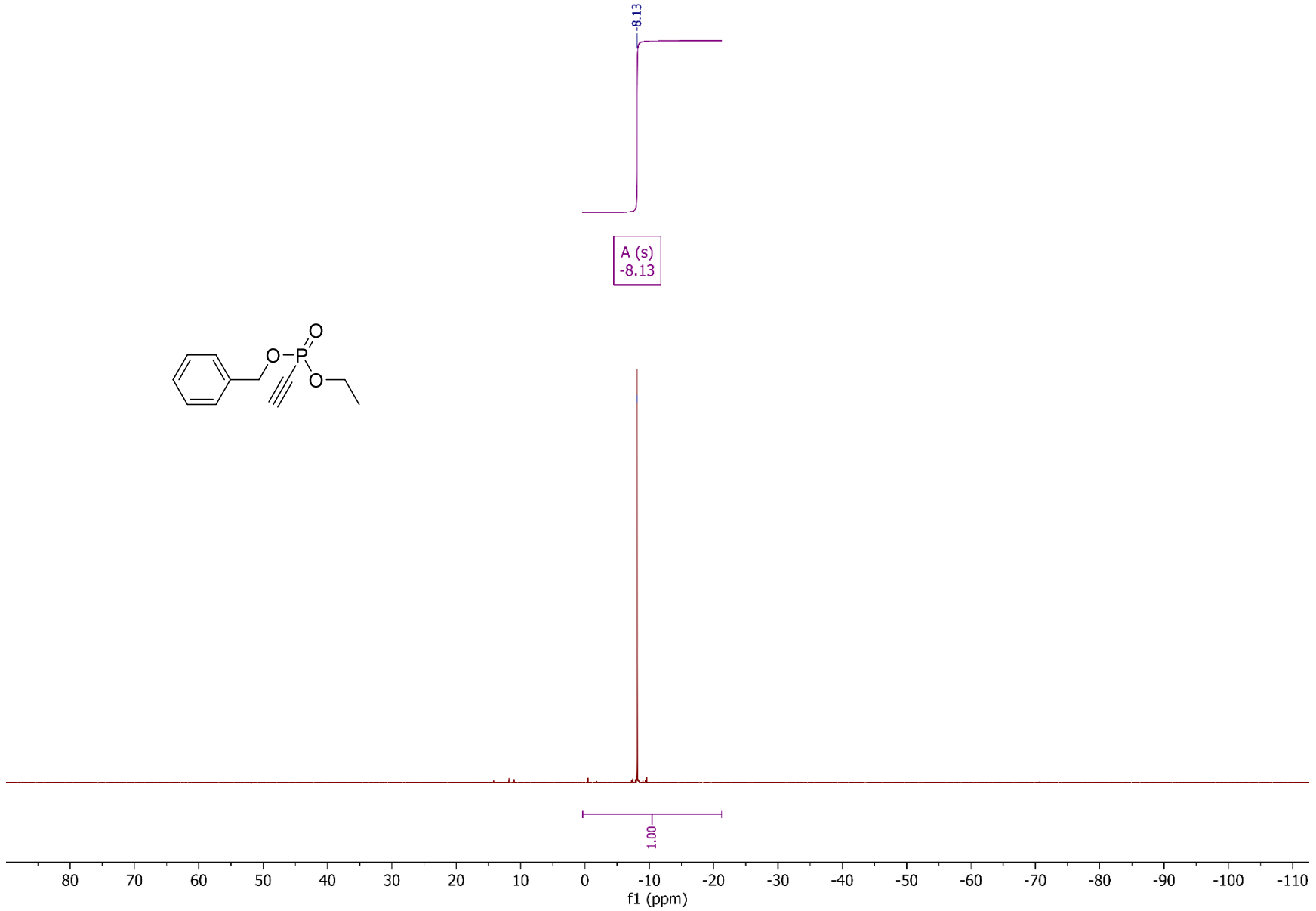

Cmpd 1, 0.4 min
Intens.
x10^8
+MS, 0.4-0.5min #99-135, -Peak Bkgrnd
1+
653.22
3
2
1
0
200
400
600
800
1000
m/z

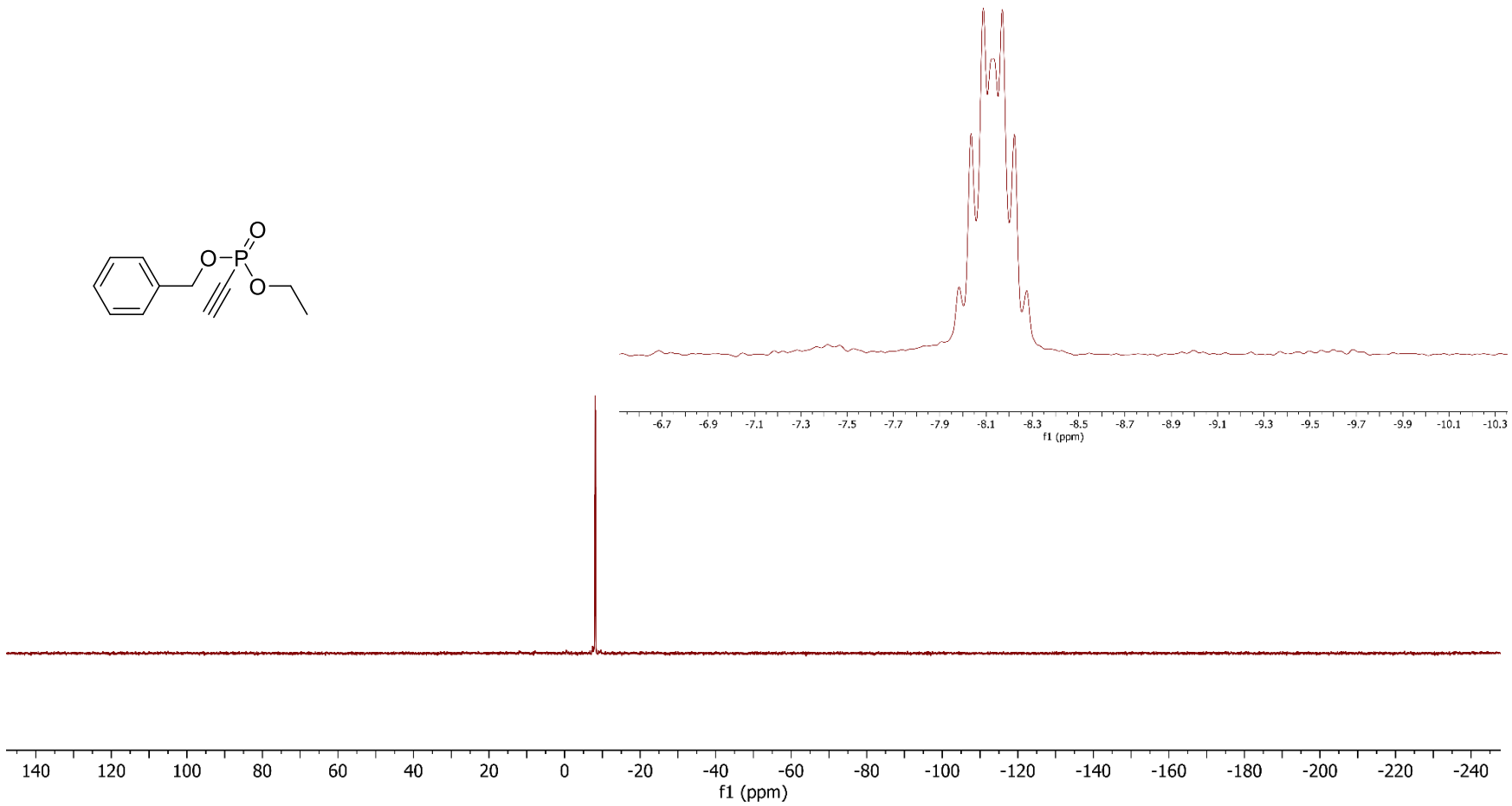

7.42 7.41 7.41 7.41 7.41 7.40 7.40 7.40 7.40 7.40 7.40 7.39 7.39 7.39 7.38 7.38 7.38 7.38 7.37 7.37 7.37 7.36 7.36 7.36 7.35 7.35 7.35 7.35
7.26 CDCl3
5.18 5.16 5.15 5.14 5.14 5.12 5.12 5.10
4.21 4.19 4.19 4.18 4.18 4.16 4.16 4.15
2.91 2.89
1.37 1.37 1.36 1.35 1.34 1.34
A (m) 7.38
B (dd) 5.16
F (dd) 5.12
C (dq) 4.18
D (d) 2.90
E (td) 1.36
5.00
1.03
1.03
2.07
0.89
3.00
8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)

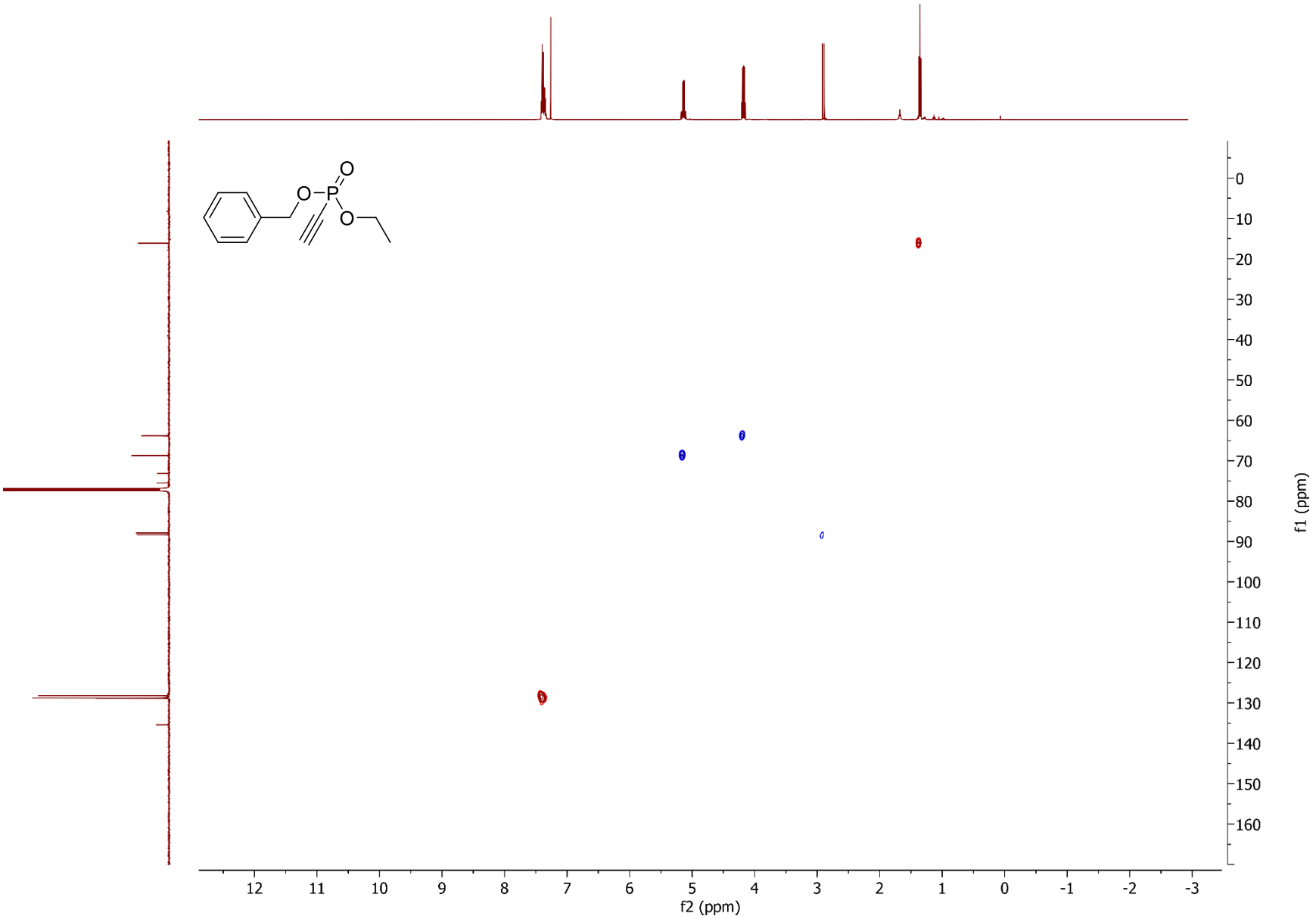

5.18
5.16
5.15
5.14
5.14
5.12
5.12
5.10
O
O
P
O
B (dd)
5.16
F (dd)
5.12
1.03
1.03
5.200 5.195 5.190 5.185 5.180 5.175 5.170 5.165 5.160 5.155 5.150 5.145 5.140 5.135 5.130 5.125 5.120 5.115 5.110 5.105 5.100 5.095 5.090 5.085 5.080 5.075 5.070
f1 (ppm)

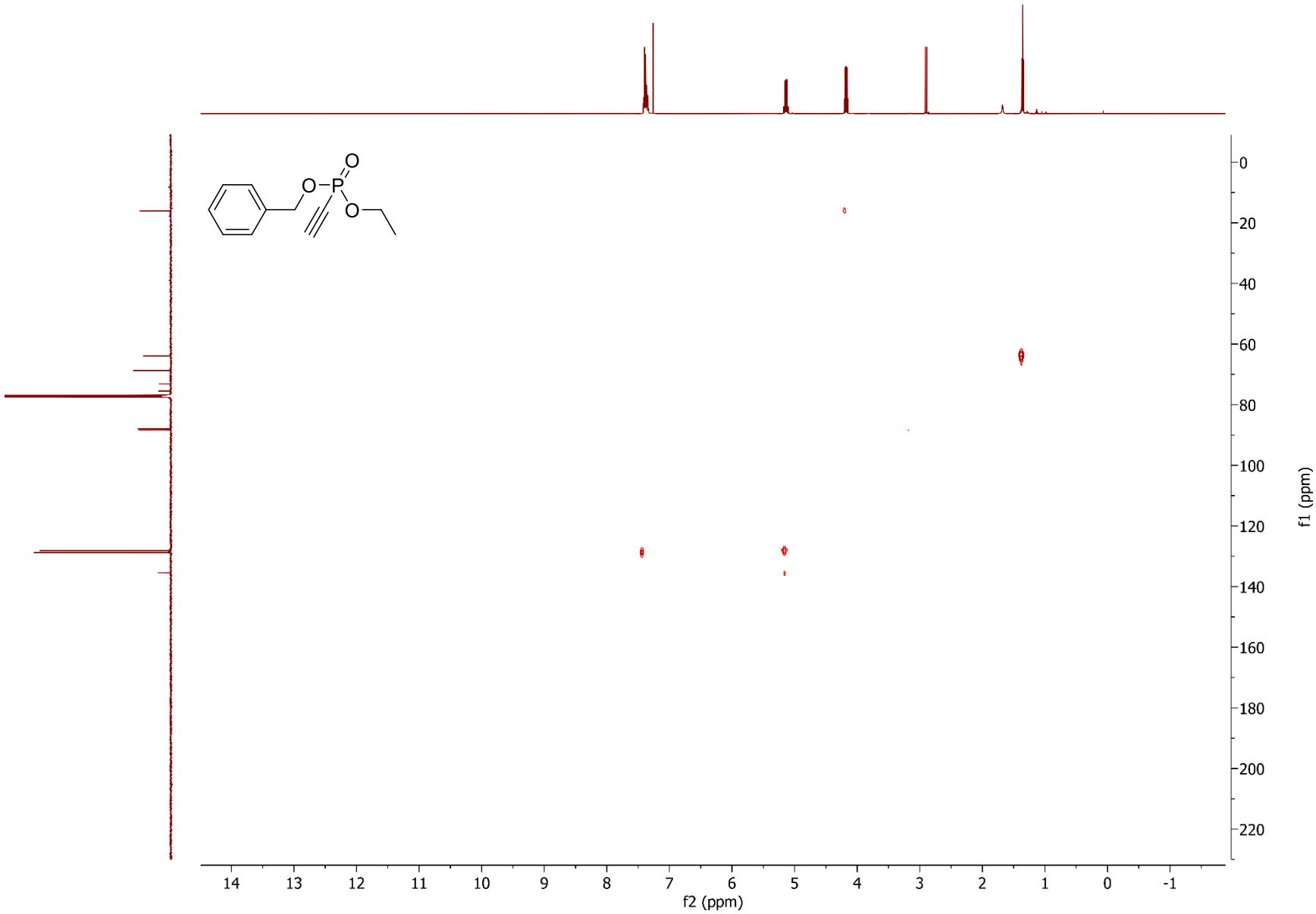

4.21
4.19
4.19
4.18
4.18
4.16
4.16
4.15
C (dq)
4.18
2.07
4.26 4.25 4.24 4.23 4.22 4.21 4.20 4.19 4.18 4.17 4.16 4.15 4.14 4.13 4.12 4.11 4.10 4.09 4.08 4.07
f1 (ppm)

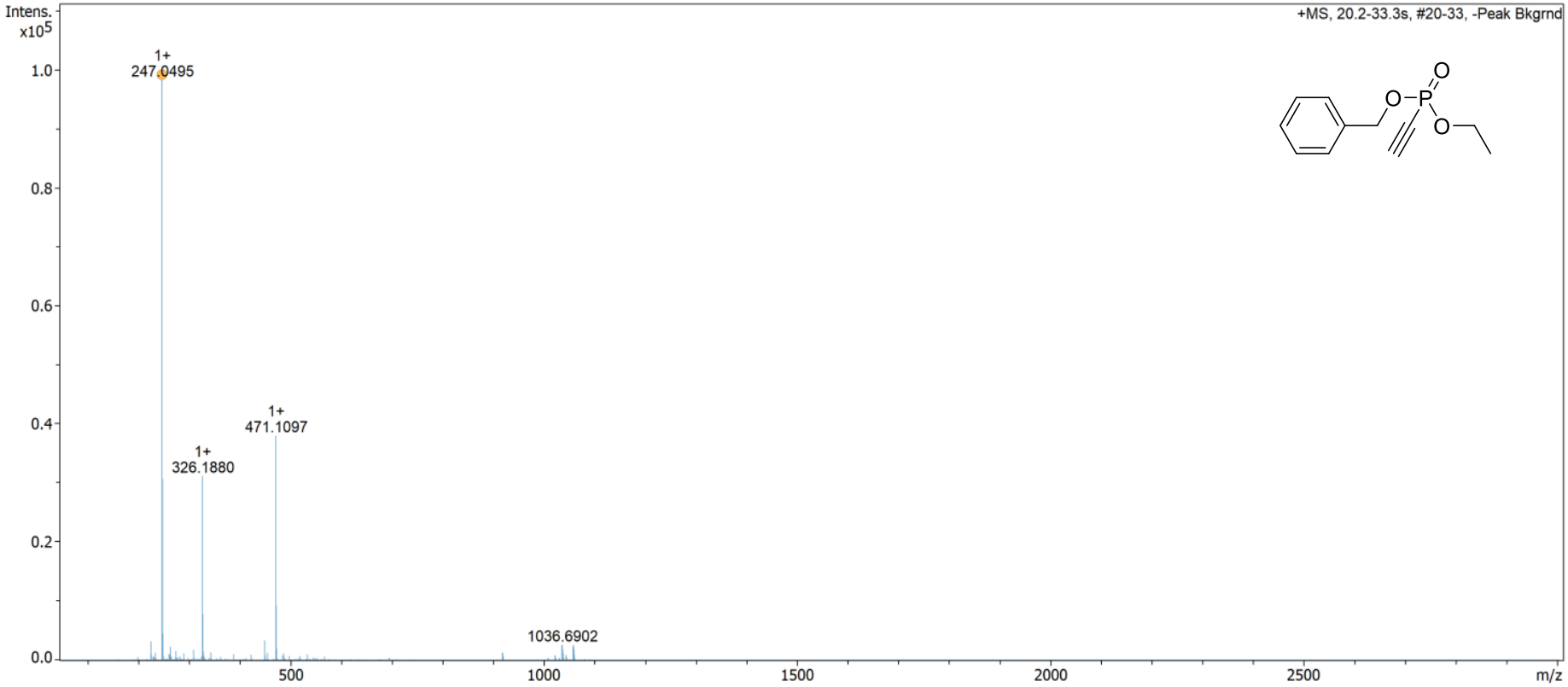

135.48
135.42
128.80
128.76
128.15
88.32
87.91
77.16 CDCl3
75.50
73.18
68.77
68.72
63.89
63.85
16.18
16.12
D (s)
128.15
B (s)
128.80
A (d)
135.45
C (s)
128.76
E (d)
88.12
G (d)
68.75
F (d)
74.34
H (d)
63.87
I (d)
16.15
1.00
3.07
5.14
5.16
2.66
0.95
3.24
2.86
2.88
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)

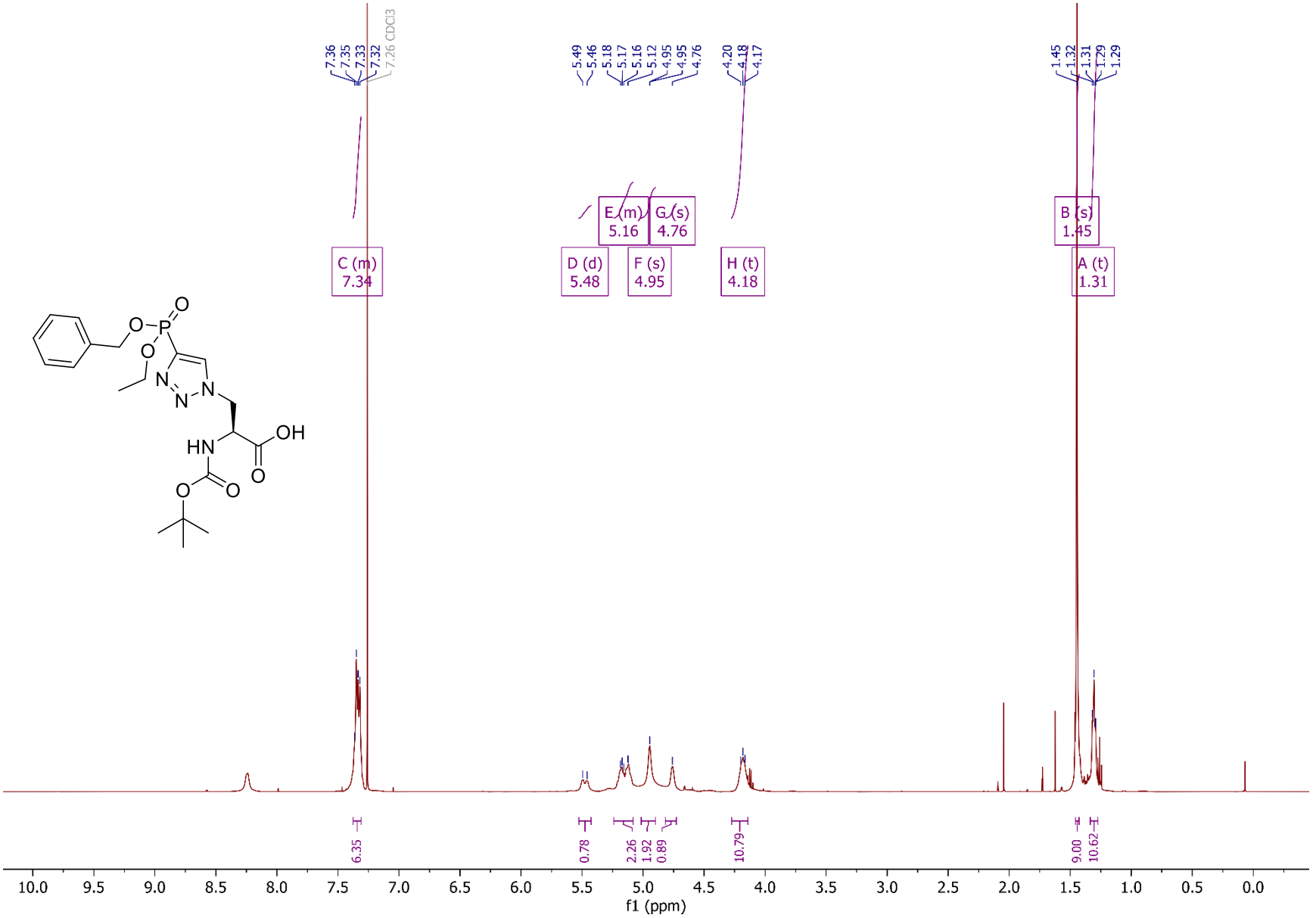

-8.13
A (s)
-8.13
O
O
P
O
O
1.00
80
70
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
-110
f1 (ppm)

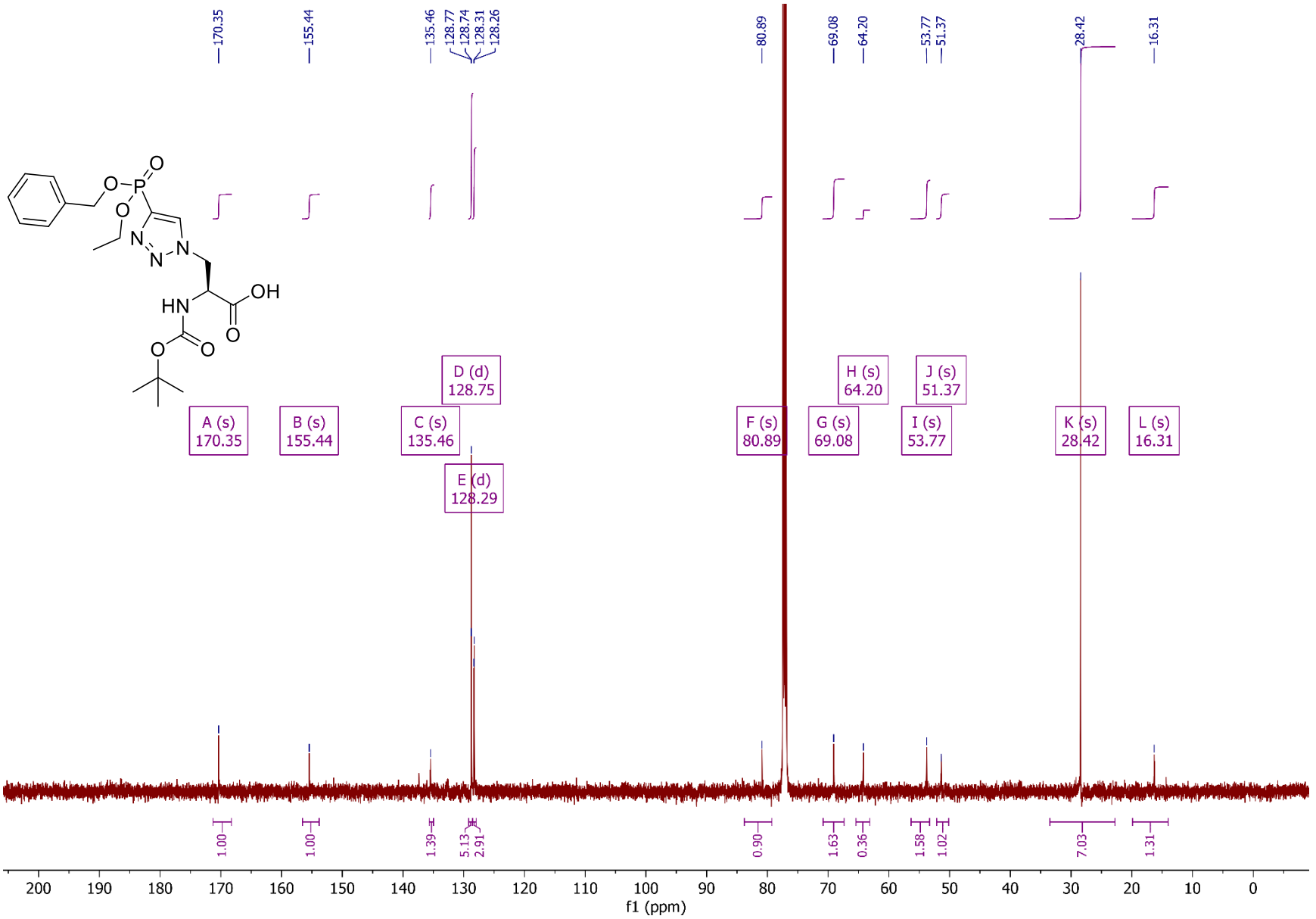

O
O
P
O
-6.7 -6.9 -7.1 -7.3 -7.5 -7.7 -7.9 -8.1 -8.3 -8.5 -8.7 -8.9 -9.1 -9.3 -9.5 -9.7 -9.9 -10.1 -10.3
f1 (ppm)
140 120 100 80 60 40 20 0 -20 -40 -60 -80 -100 -120 -140 -160 -180 -200 -220 -240
f1 (ppm)

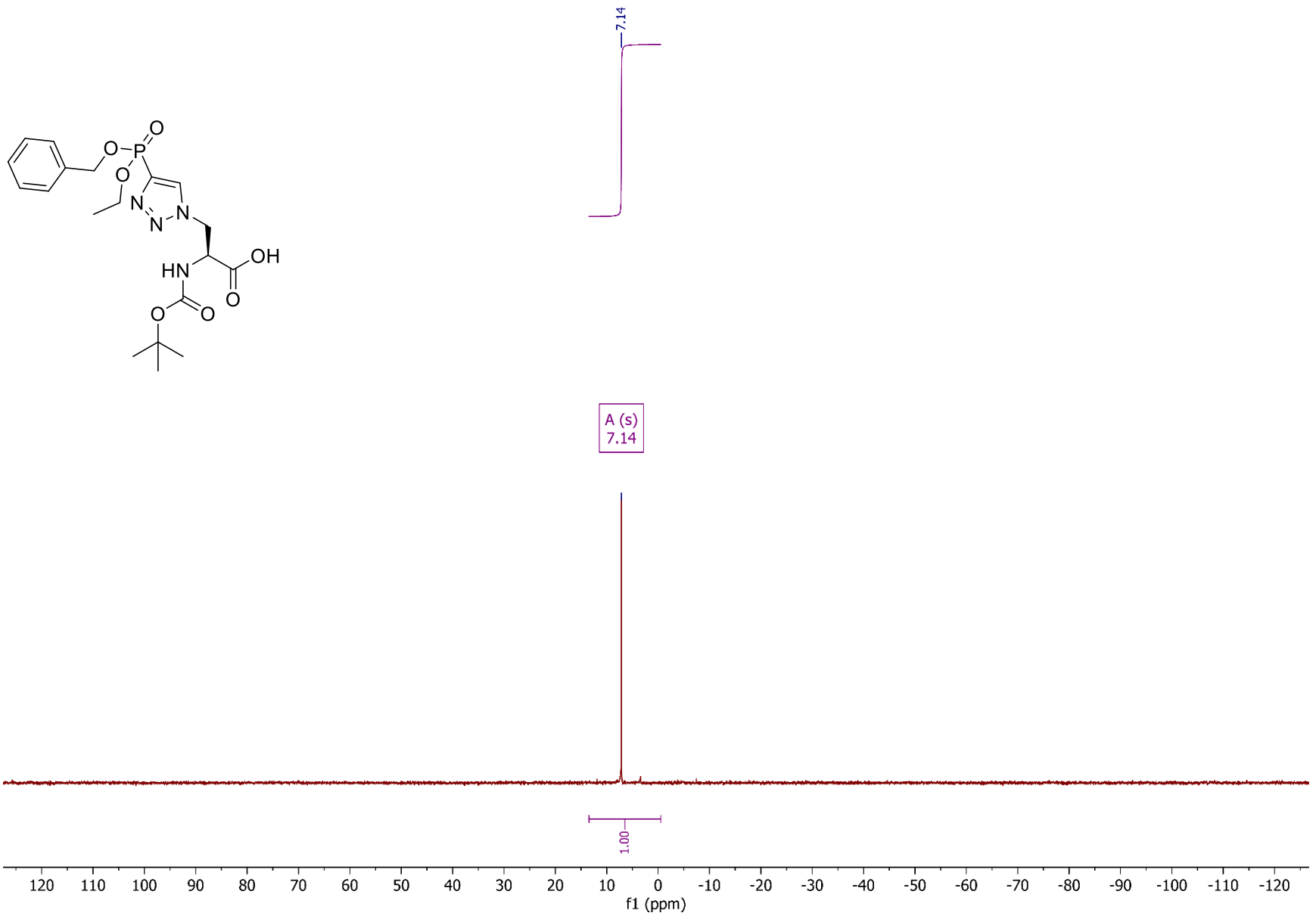

O
O
P
O
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
-2
-3
f2 (ppm)
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
f1 (ppm)

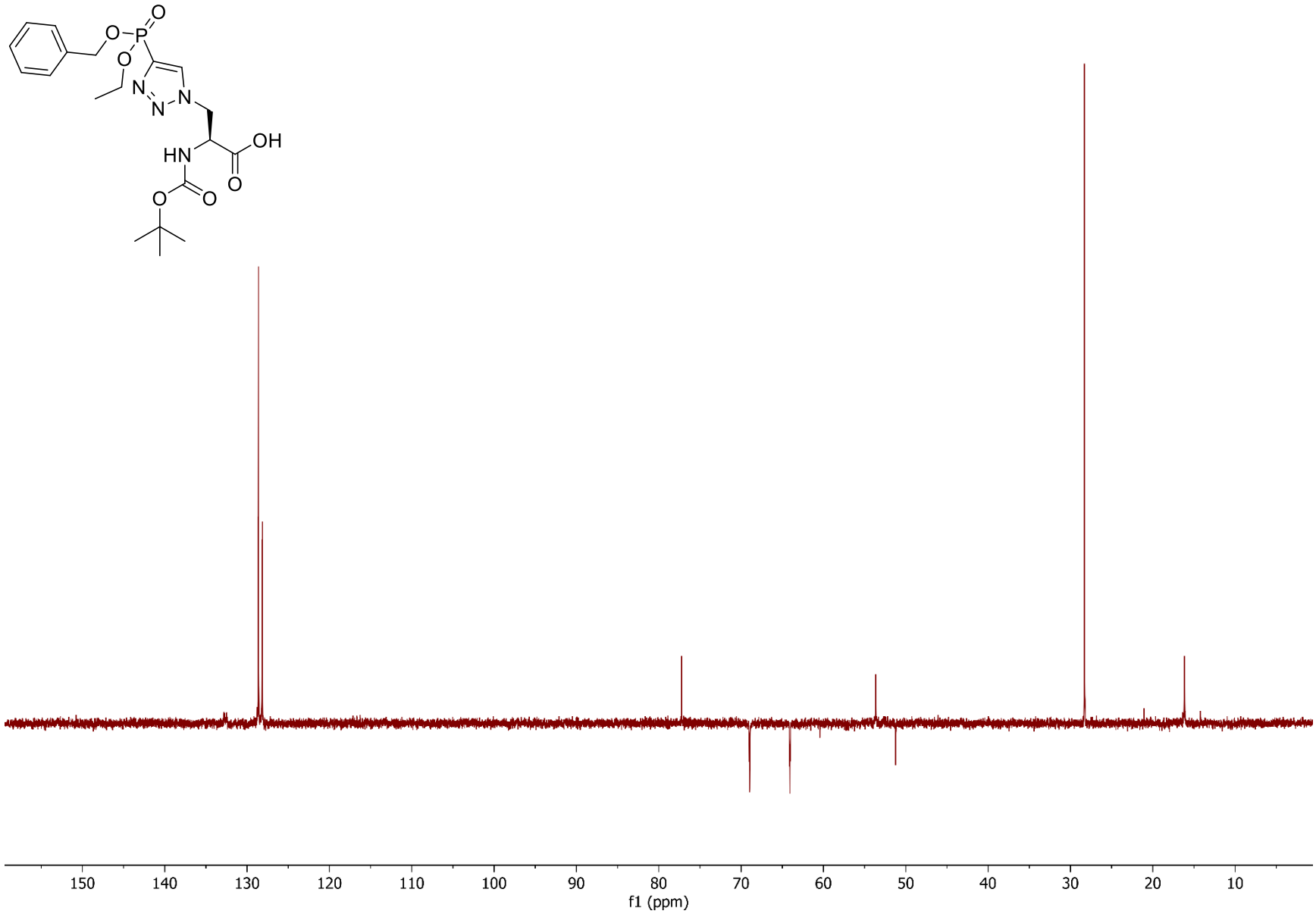

O
O
P
O
O
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
f2 (ppm)
0
20
40
60
80
100
120
140
160
180
200
220
f1 (ppm)

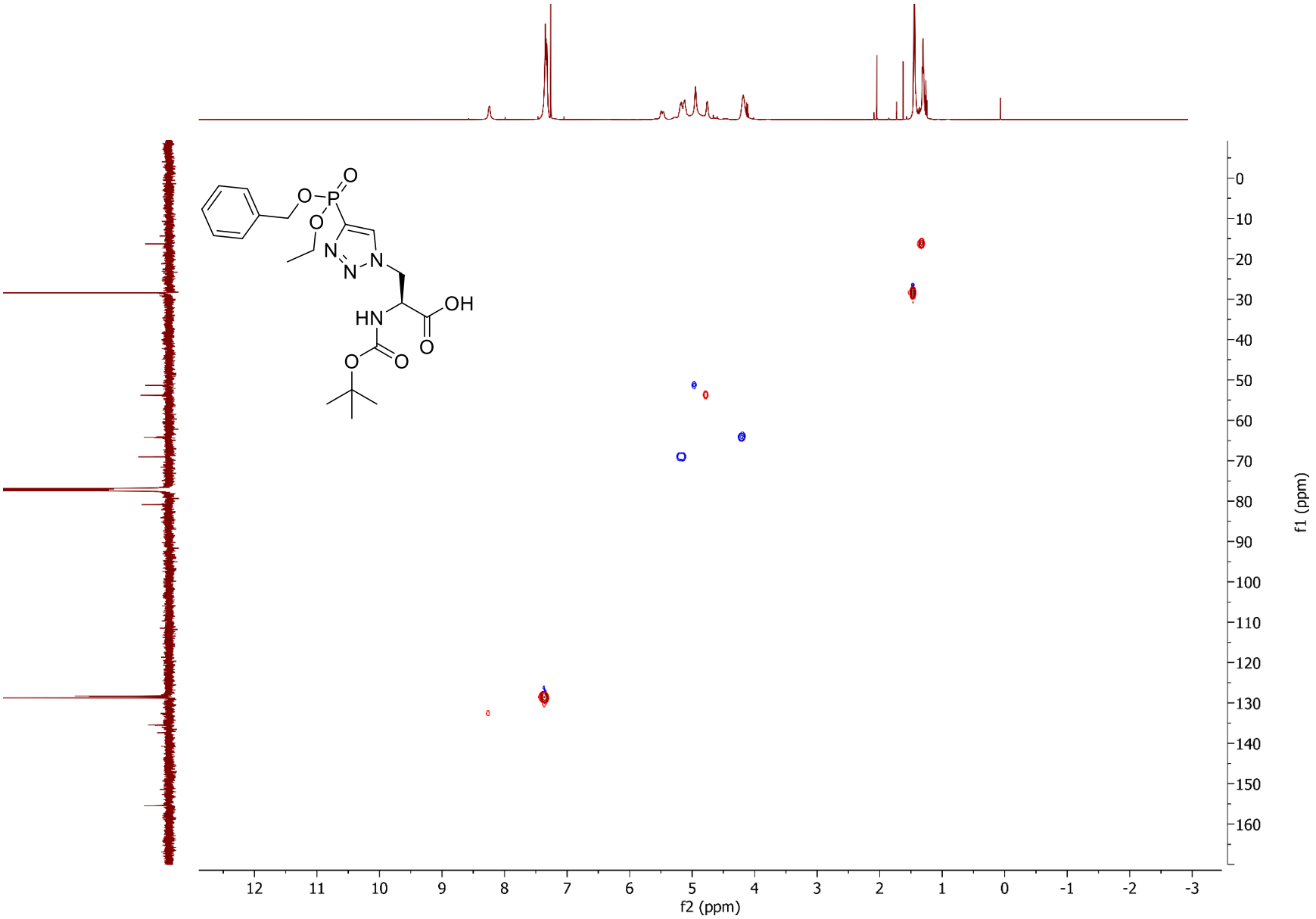

Intens.
x10^5
1.0
0.8
0.6
0.4
0.2
0.0
1+
247.0495
1+
326.1880
1+
471.1097
1036.6902
500
1000
1500
2000
2500
m/z
+MS, 20.2-33.3s, #20-33, -Peak Bkgrnd
O
O
P
O

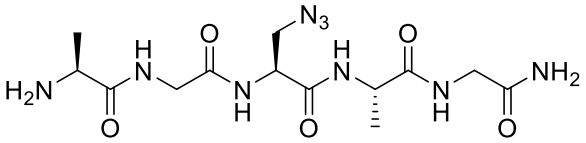

7.36
7.35
7.33
7.32
7.26 CDCl3
5.49
5.46
5.18
5.17
5.16
5.12
4.95
4.95
4.76
4.20
4.18
4.17
1.45
1.32
1.31
1.29
1.29
C (m)
7.34
D (d)
5.48
E (m)
5.16
F (s)
4.95
G (s)
4.76
H (t)
4.18
B (s)
1.45
A (t)
1.31
6.35
0.78
2.26
1.92
0.89
10.79
9.00
10.62
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

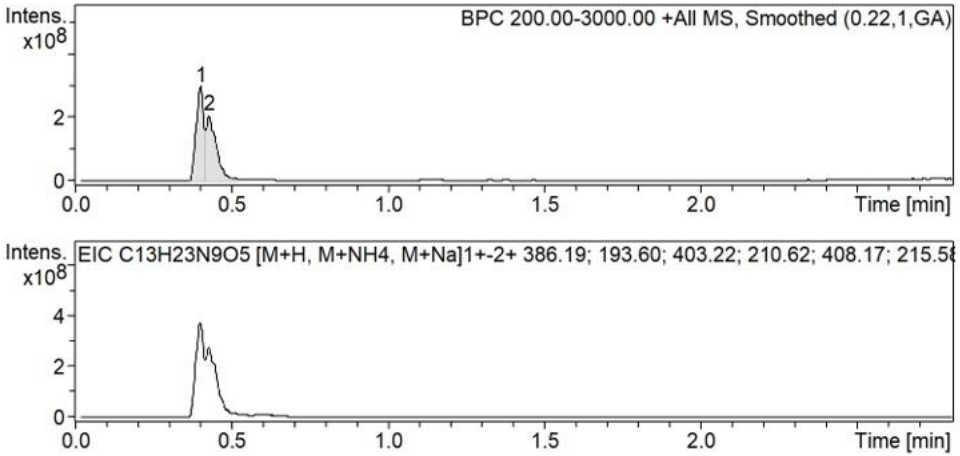

170.35
155.44
135.46
128.77
128.74
128.31
128.26
80.89
69.08
64.20
53.77
51.37
28.42
16.31
A (s) 170.35
B (s) 155.44
C (s) 135.46
D (d) 128.75
E (d) 128.29
F (s) 80.89
G (s) 69.08
H (s) 64.20
I (s) 53.77
J (s) 51.37
K (s) 28.42
L (s) 16.31
1.00
1.00
1.39
5.13
2.91
0.90
1.63
0.36
1.58
1.02
7.03
1.31
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

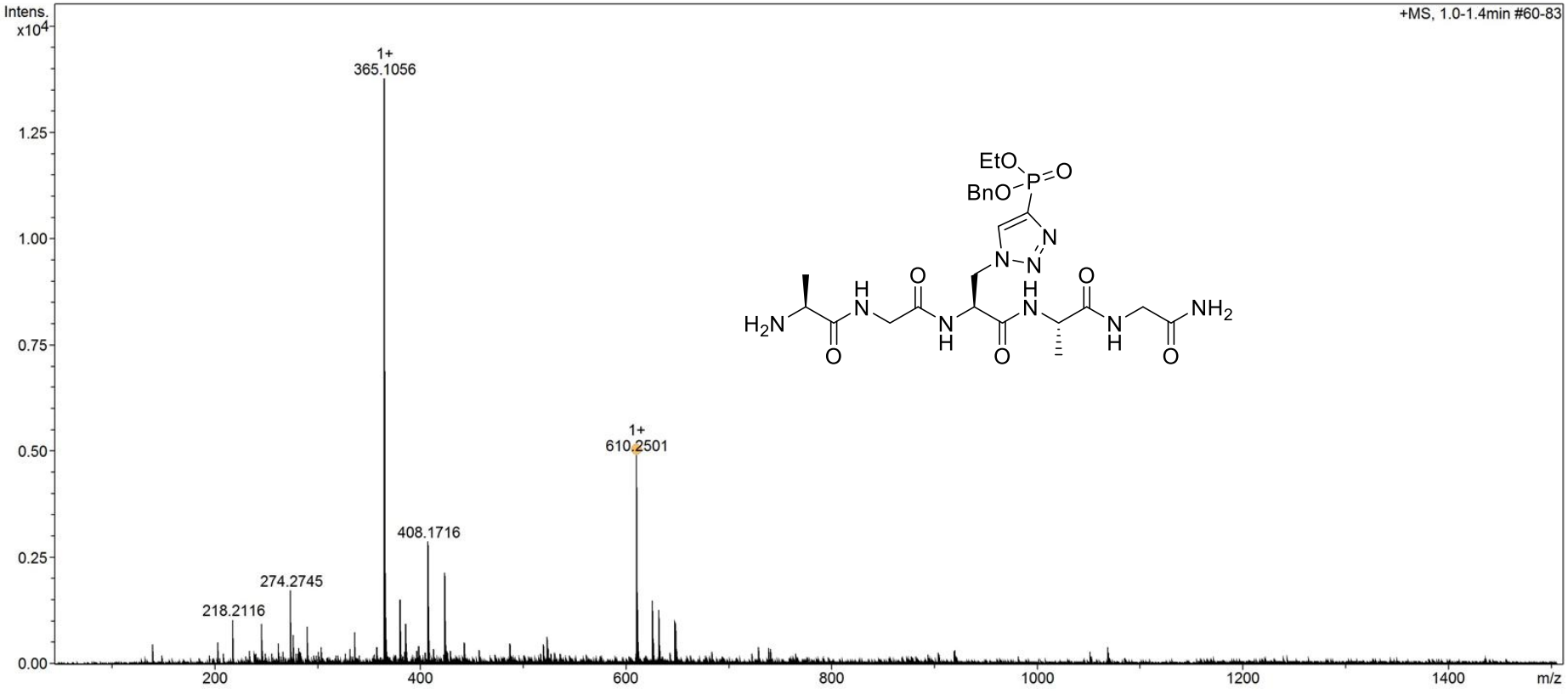

7.14
A (s)
7.14
1.00
120 110 100 90 80 70 60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60 -70 -80 -90 -100 -110 -120
f1 (ppm)

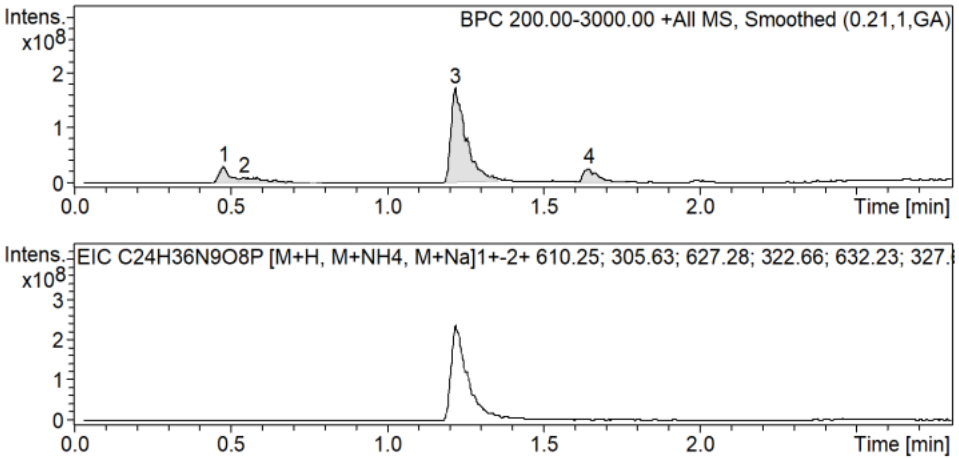

O
O
P
O
O
N
N
N
HN
OH
O
O
O
O
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
f1 (ppm)

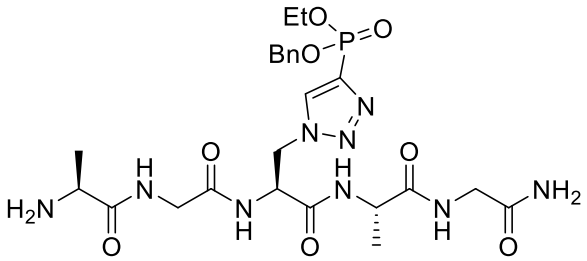

O
O
P
O
O
N
N
N
HN
OH
O
O
O
f2 (ppm)
f1 (ppm)

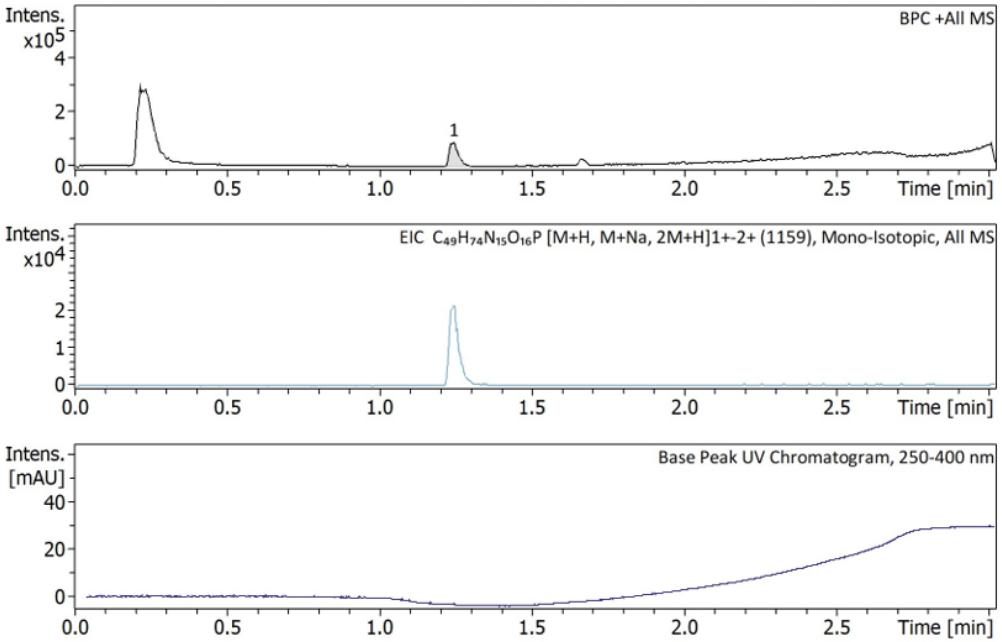


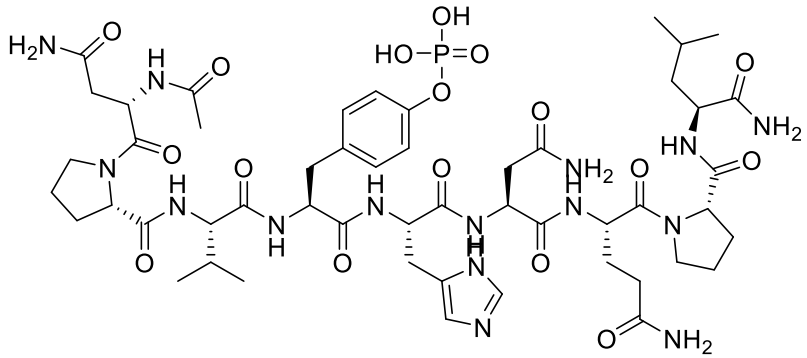

Intens.
x10^8
BPC 200.00-3000.00 +All MS, Smoothed (0.22,1,GA)
1
2
0
2
0.0
0.5
1.0
1.5
2.0
Time [min]
Intens.
x10^8
EIC C13H23N9O5 [M+H, M+NH4, M+Na]1+-2+ 386.19; 193.60; 403.22; 210.62; 408.17; 215.58
0
2
4
0.0
0.5
1.0
1.5
2.0
Time [min]

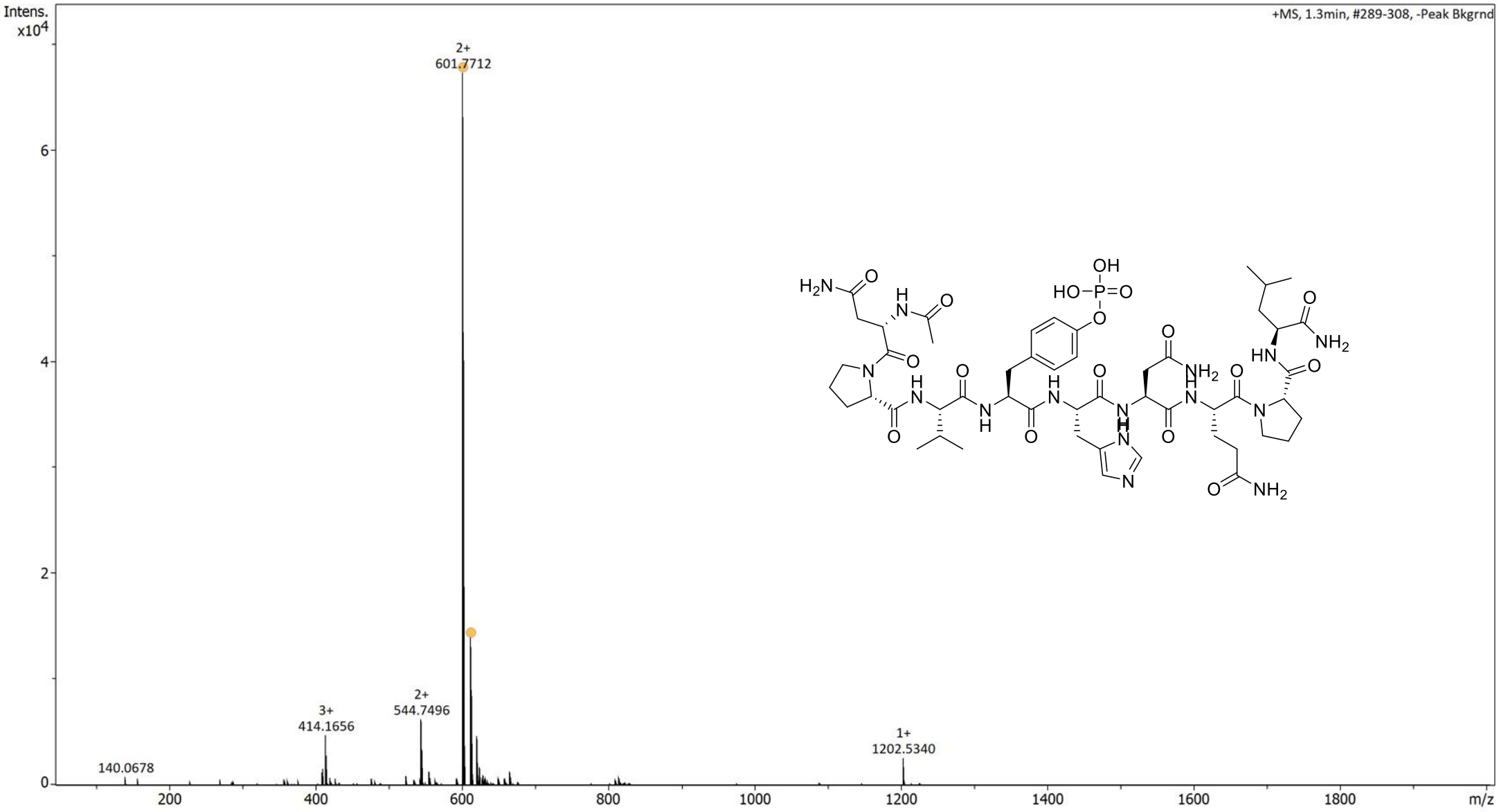

Intens.
x10^4
+MS, 1.0-1.4min #60-83
1+
365.1056
1+
610.2501
408.1716
274.2745
218.2116
1.25
1.00
0.75
0.50
0.25
0.00
200
400
600
800
1000
1200
1400
m/z
EtO
BnO
P
O
N
H
NH2

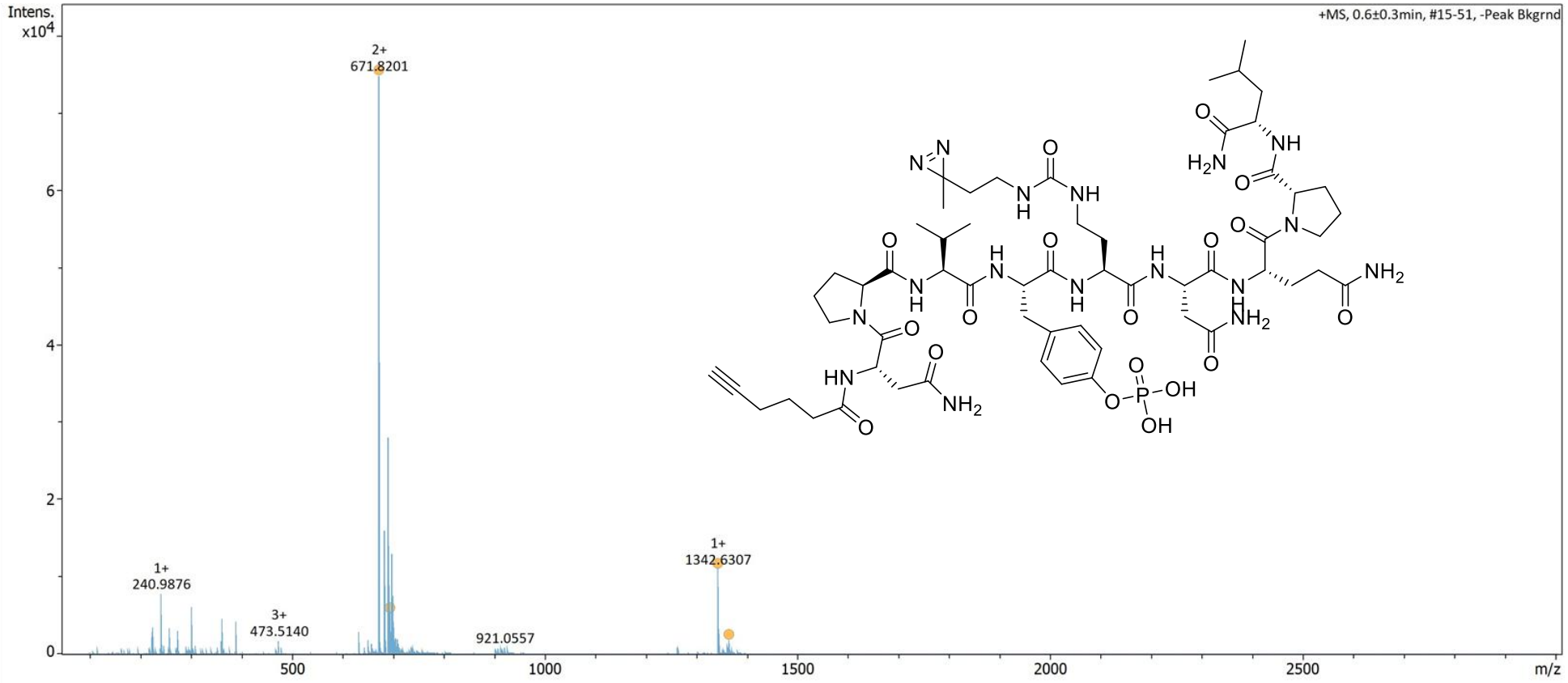

Intens.
x10^8
BPC 200.00-3000.00 +All MS, Smoothed (0.21,1,GA)
1
2
3
4
Time [min]
EIC C24H36N9O8P [M+H, M+NH4, M+Na]1+-2+ 610.25; 305.63; 627.28; 322.66; 632.23; 327.

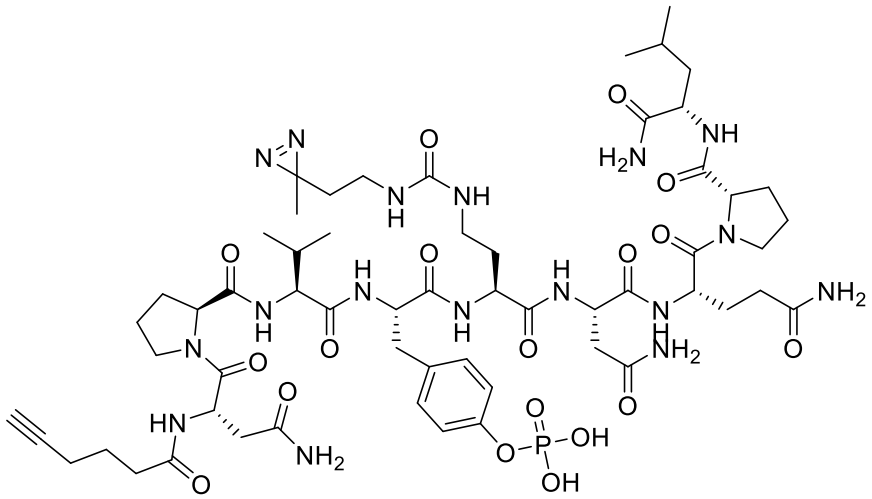

EtO
BnO
P
O
N
N
N
H
N
O
H2N
NH2

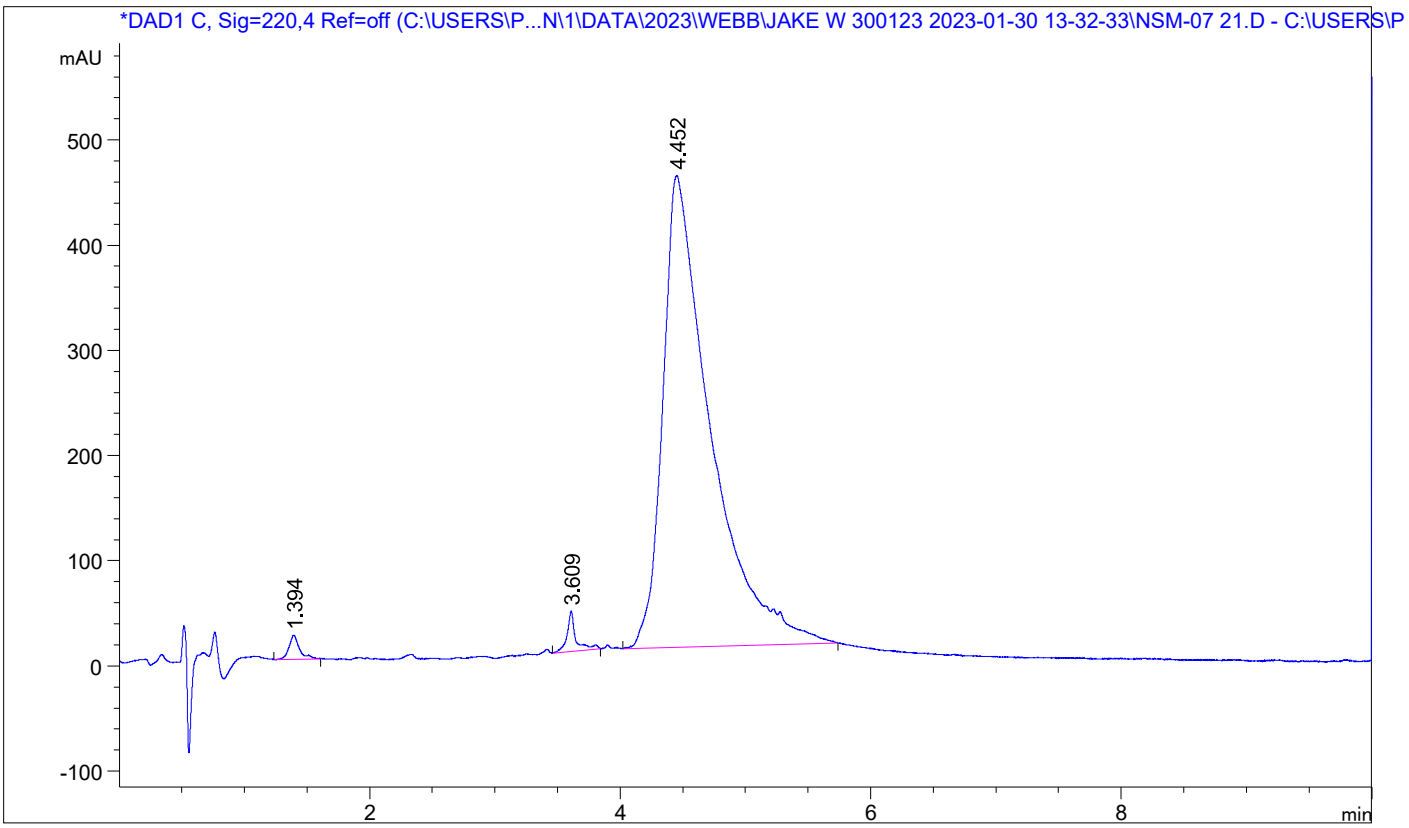

Intens.
x10^5
4
2
0
BPC +All MS
1
0.0
0.5
1.0
1.5
2.0
2.5
Time [min]
Intens.
x10^4
2
1
0
EIC C49H74N15O16P [M+H, M+Na, 2M+H]1+-2+ (1159), Mono-Isotopic, All MS
Intens.
[mAU]
40
20
0
Base Peak UV Chromatogram, 250-400 nm

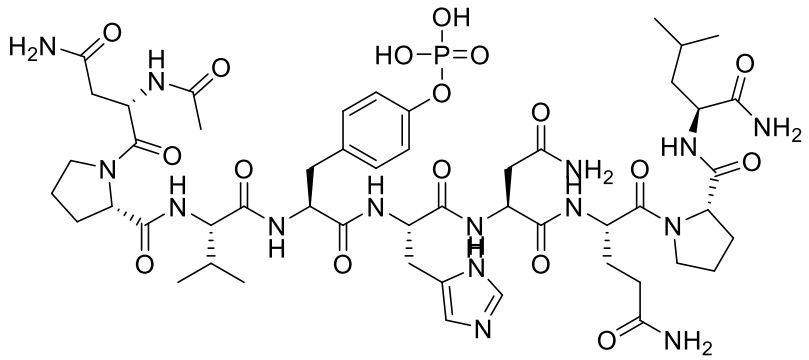


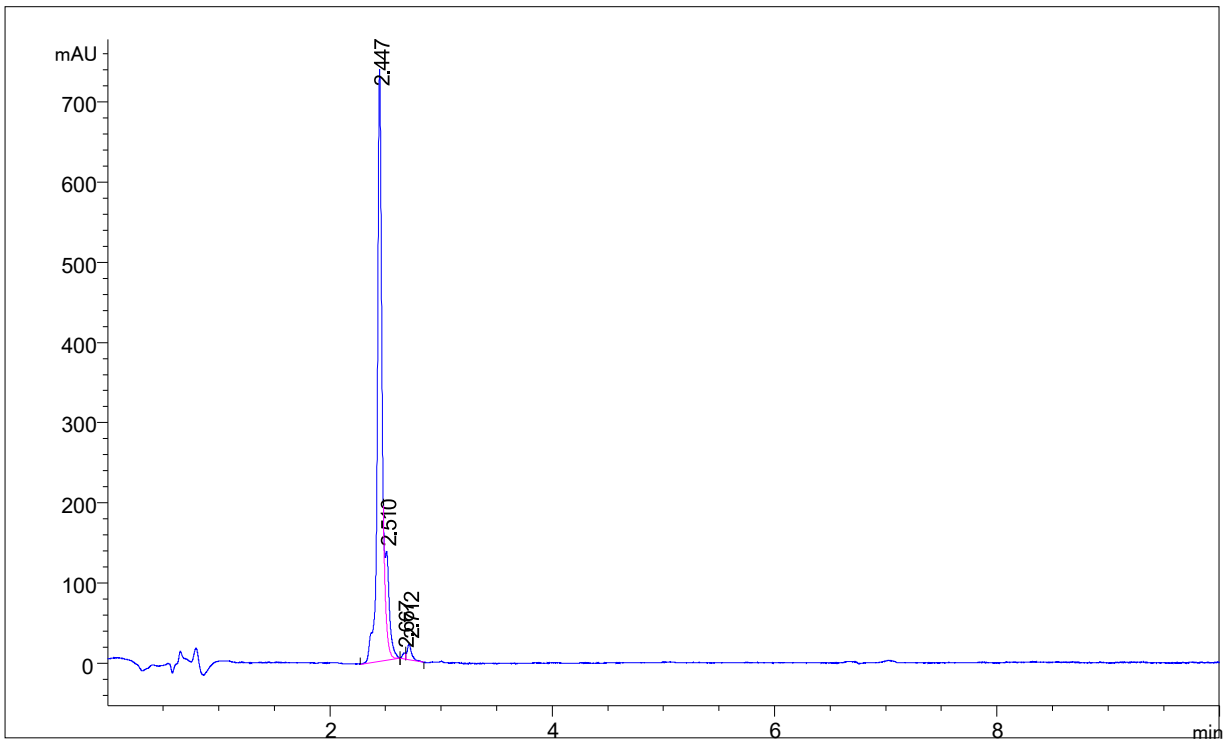

Intens.
x10^4
+MS, 1.3min, #289-308, -Peak Bkgrnd
2+
601.7712
3+
414.1656
2+
544.7496
1+
1202.5340
140.0678
0
2
4
6
200
400
600
800
1000
1200
1400
1600
1800
m/z

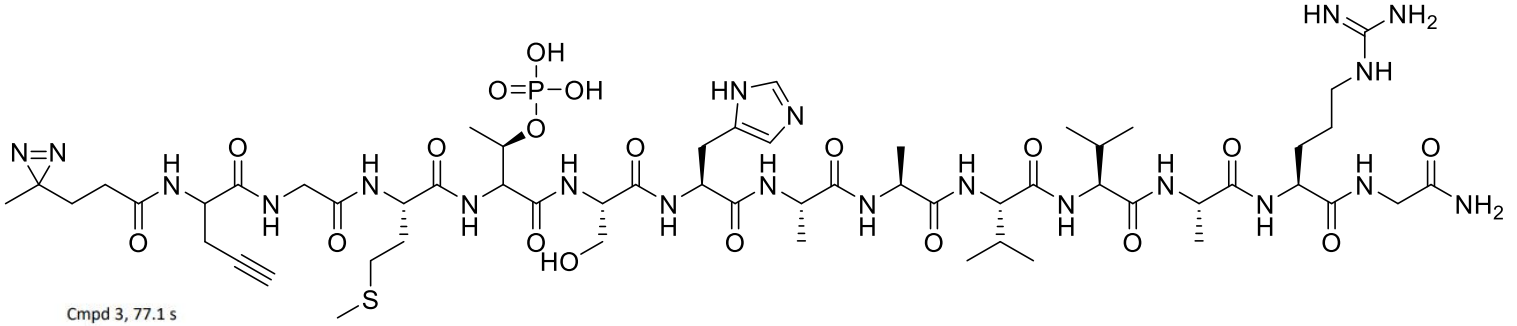

Intens.
x10^4
+MS, 0.6±0.3min, #15-51, -Peak Bkgrnd
2+
671.8201
1+
240.9876
3+
473.5140
921.0557
1+
1342.6307
0
2
4
6
500
1000
1500
2000
2500
m/z

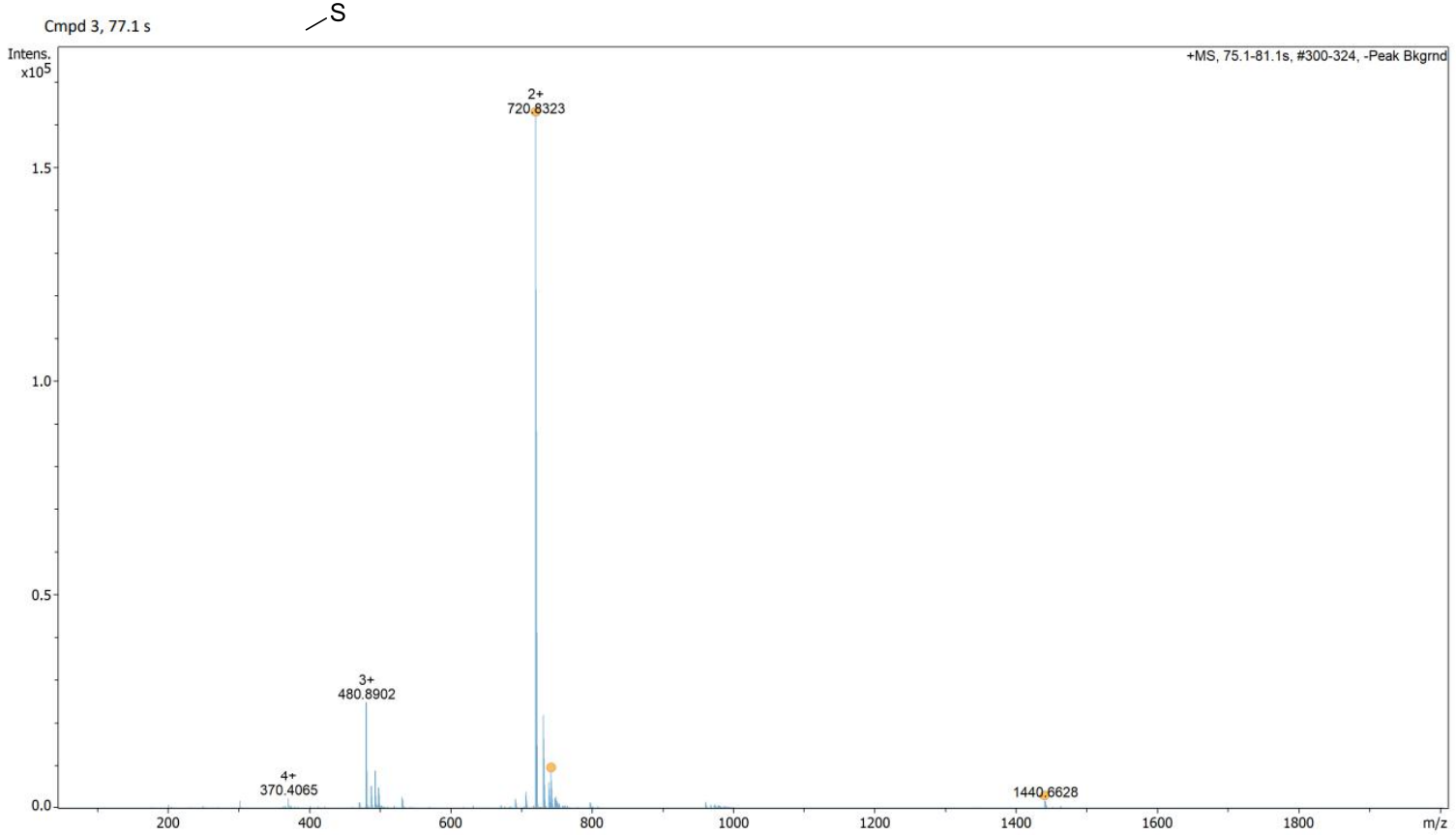


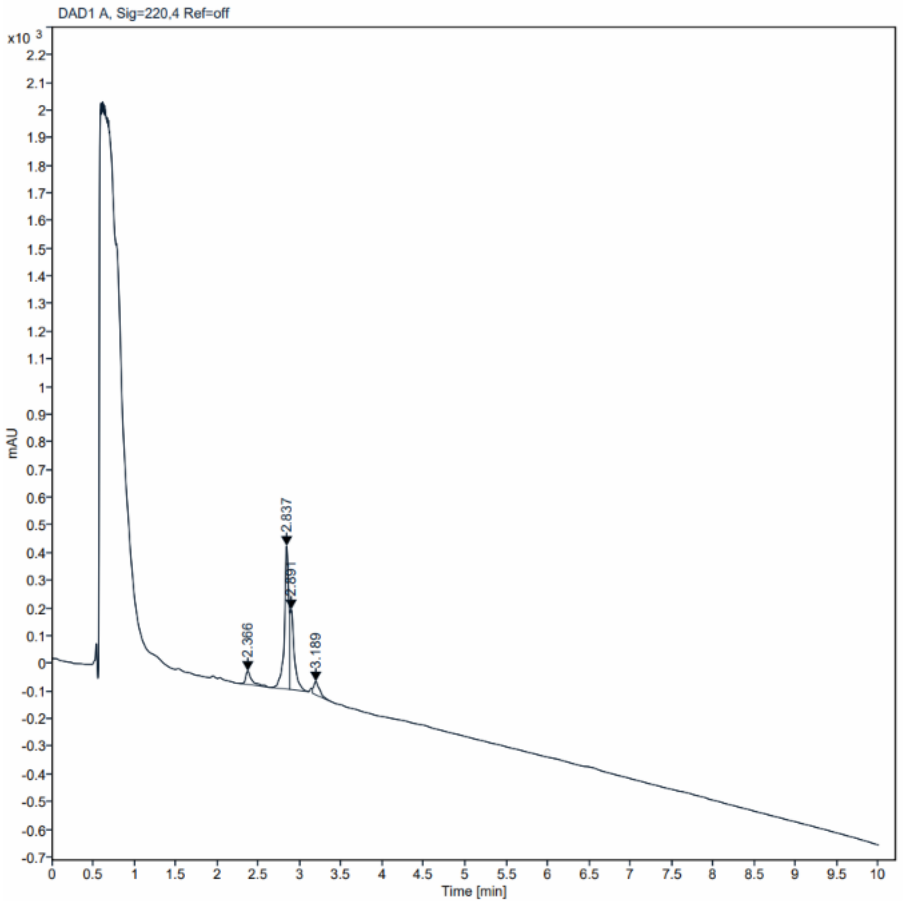

*DAD1 C, Sig=220,4 Ref=off (C:\USERS\P...N\1\DATA\2023\WEBB\JAKE W 300123 2023-01-30 13-32-33\NSM-07 21.D - C:\USERS\P
mAU
500
400
300
200
100
0
-100
1.394
3.609
4.452
2
4
6
8
min

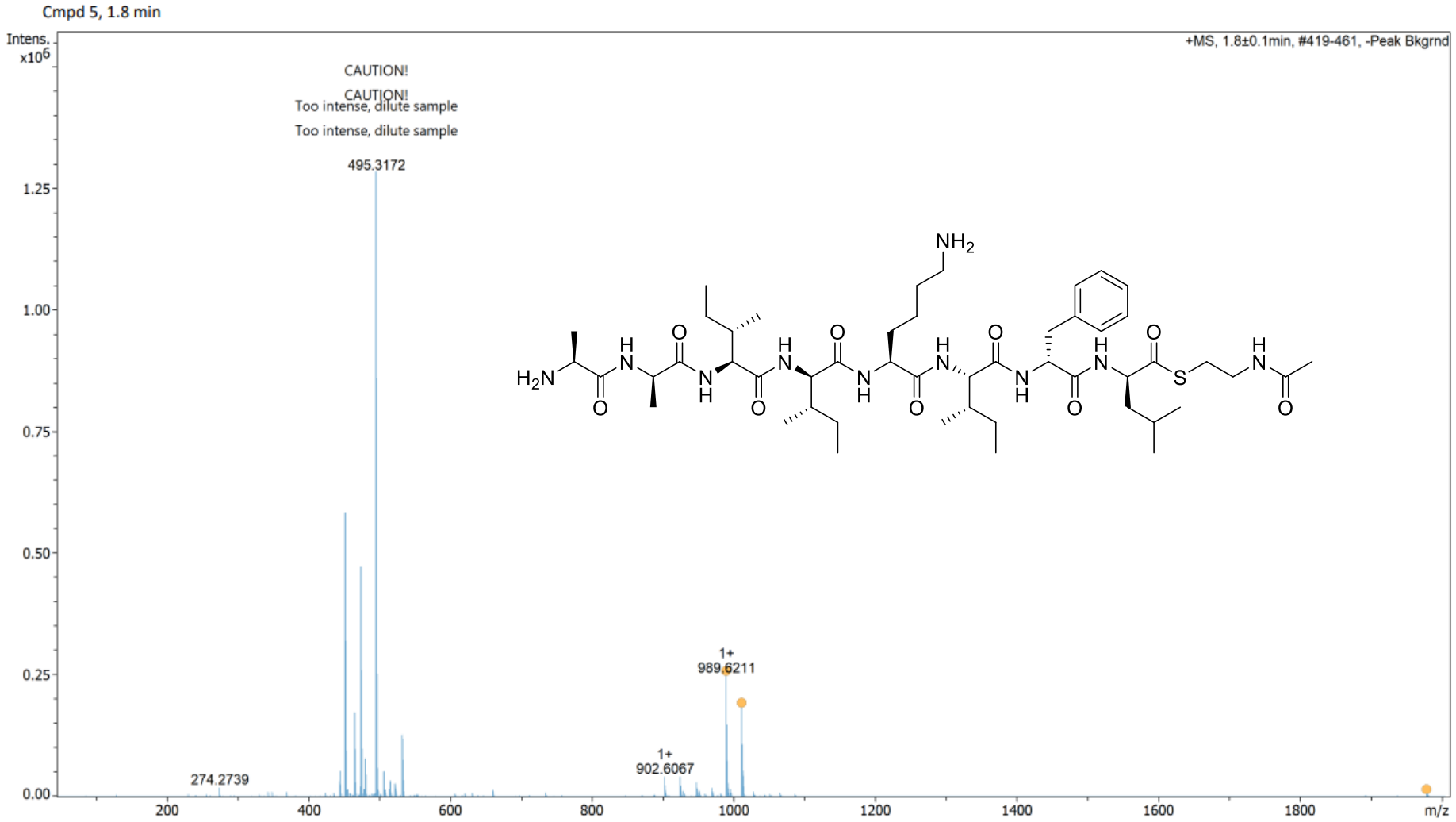


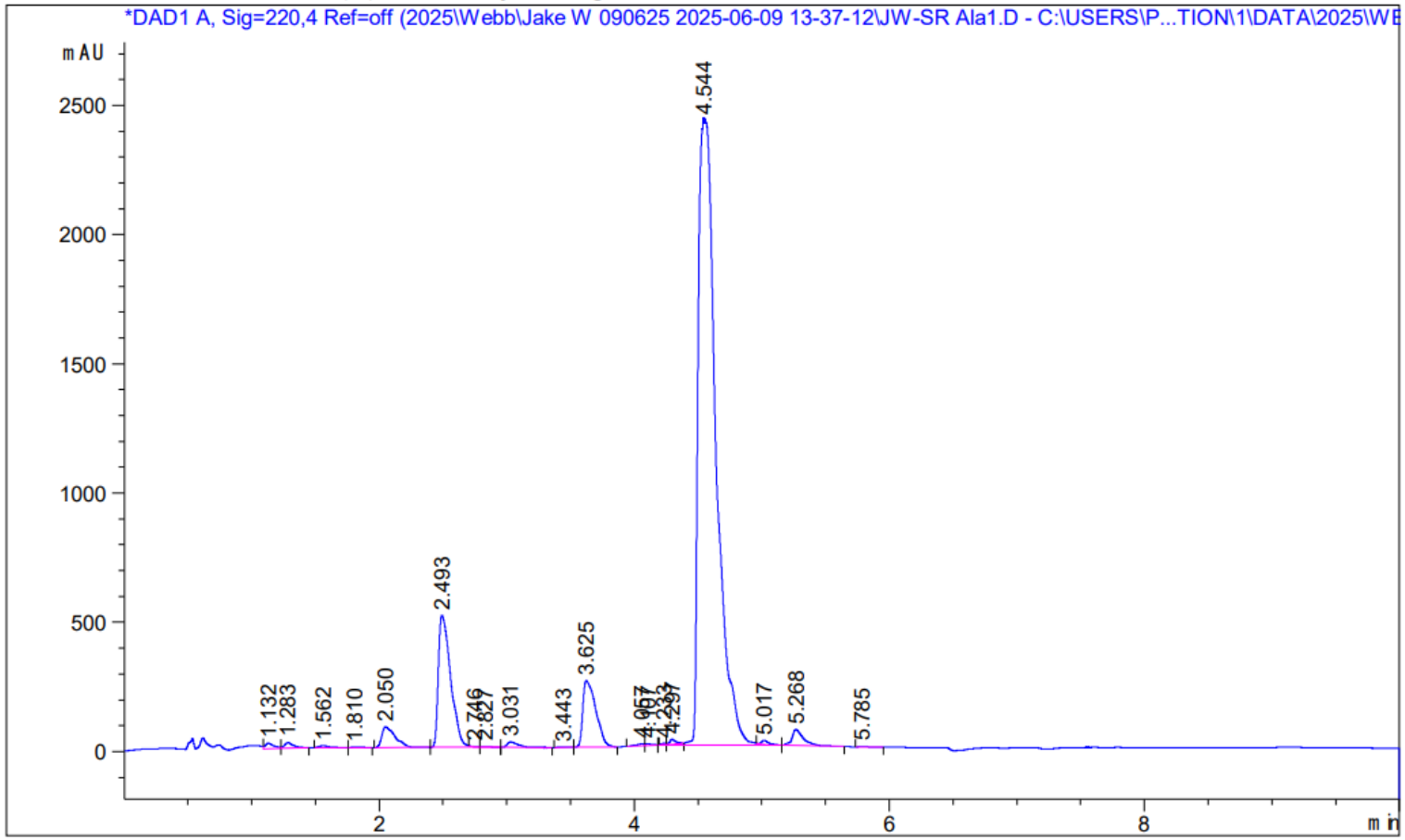

mAU
700
600
500
400
300
200
100
0
2.447
2.510
2.667
2.712
2
4
6
8
min

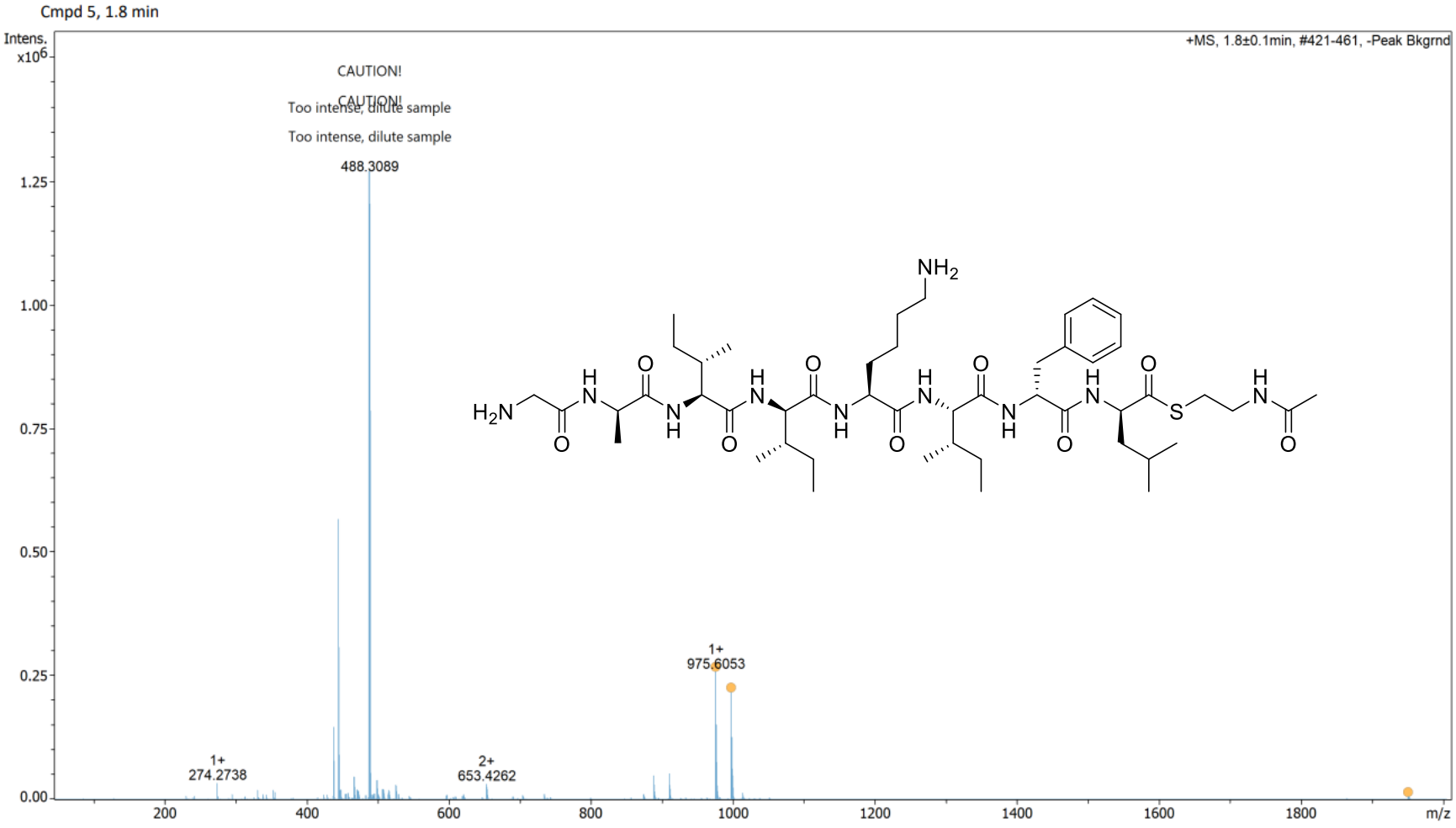


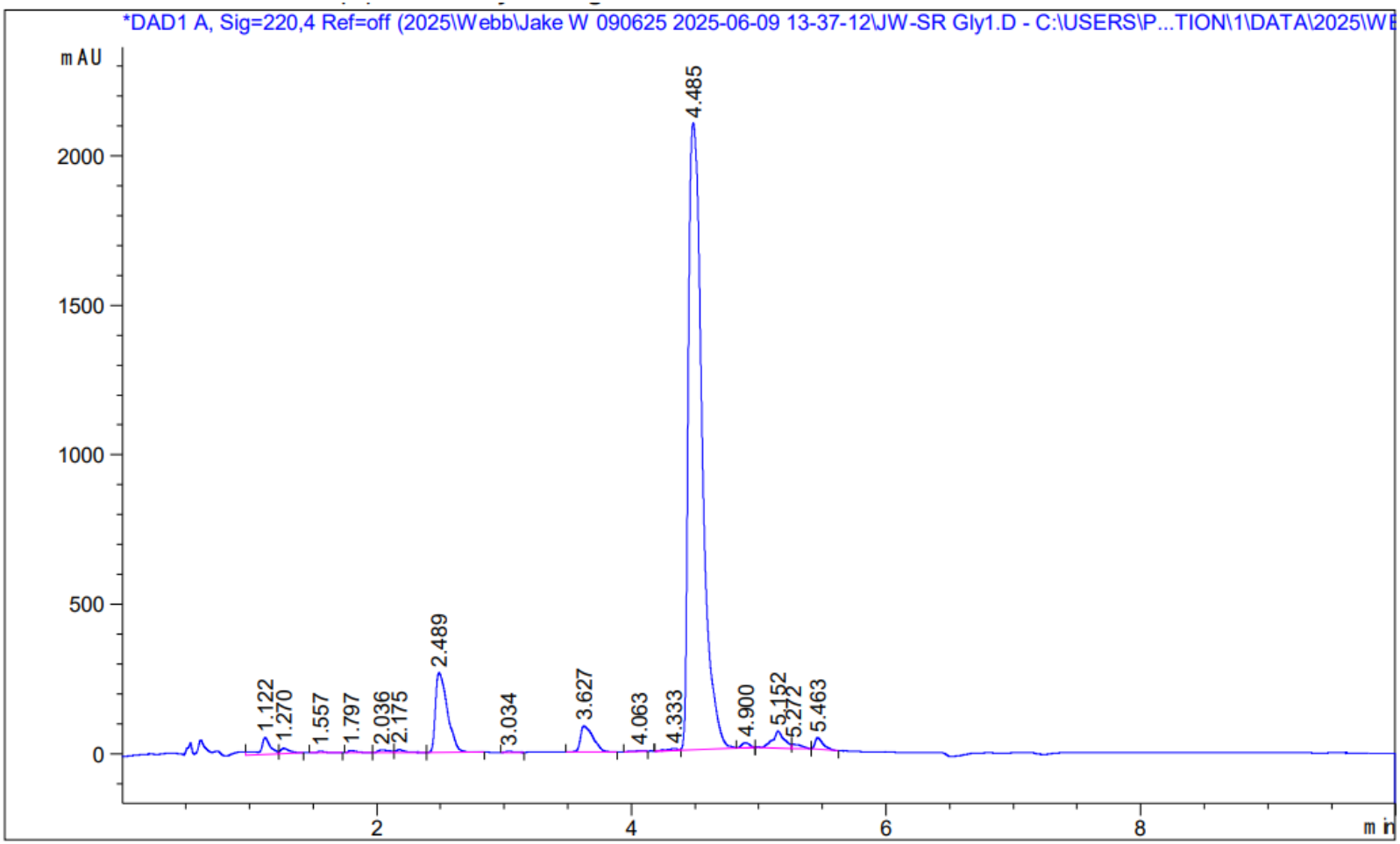

Cmpd 3, 77.1 s
+MS, 75.1-81.1s, #300-324, -Peak Bkgrnd
Intens.
x10^5
2+
720.8323
3+
480.8902
4+
370.4065
1440.6628
m/z

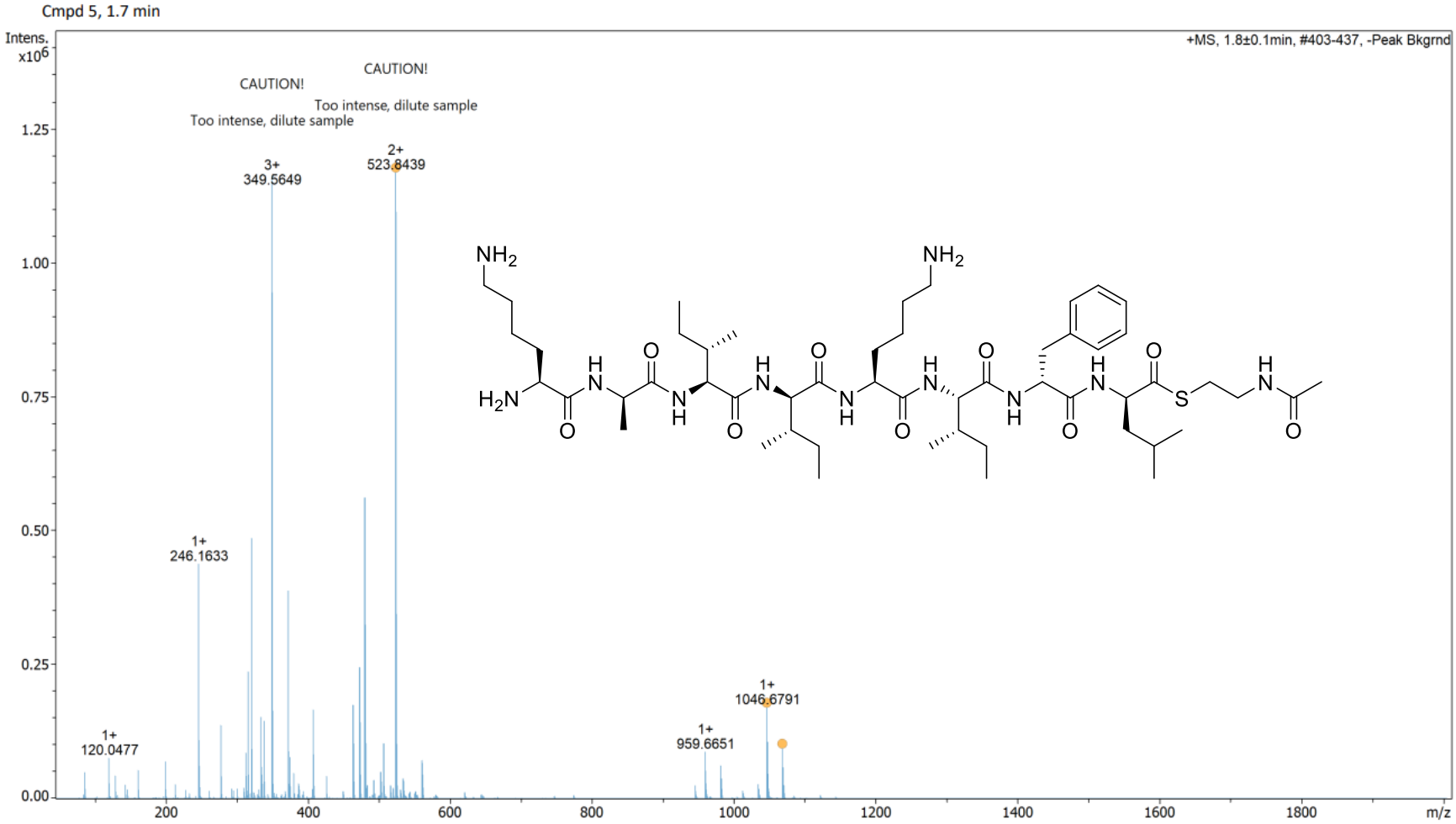

DAD1 A, Sig=220,4 Ref=off
x10 3
mAU
2.366
2.837
2.891
3.189
Time [min]

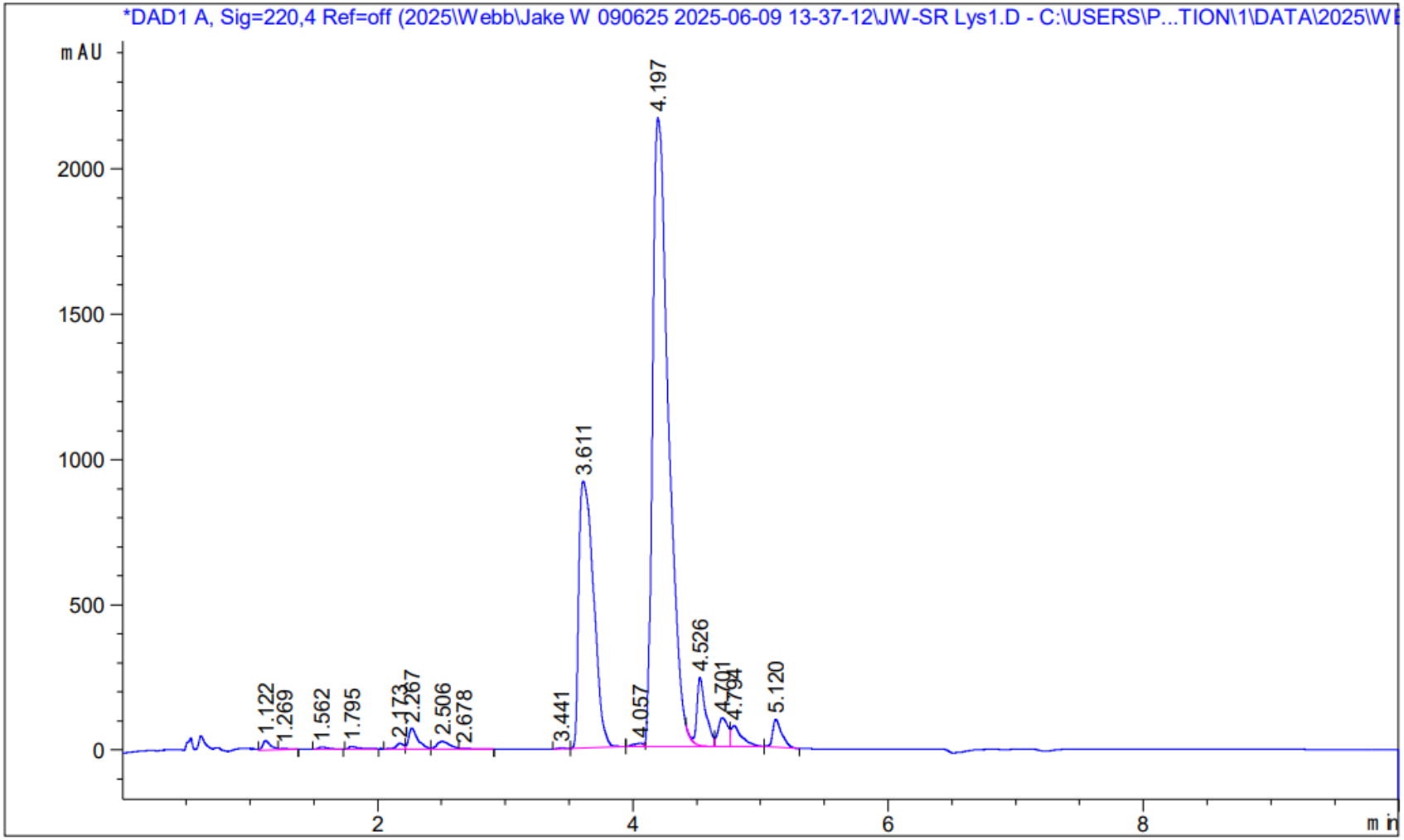

Cmpd 5, 1.8 min
+MS, 1.8±0.1min, #419-461, -Peak Bkgrnd
Intens.
x10^6
CAUTION!
Too intense, dilute sample
CAUTION!
Too intense, dilute sample
495.3172
1.25
1.00
0.75
0.50
0.25
0.00
1+
989.6211
1+
902.6067
274.2739
200
400
600
800
1000
1200
1400
1600
1800
m/z

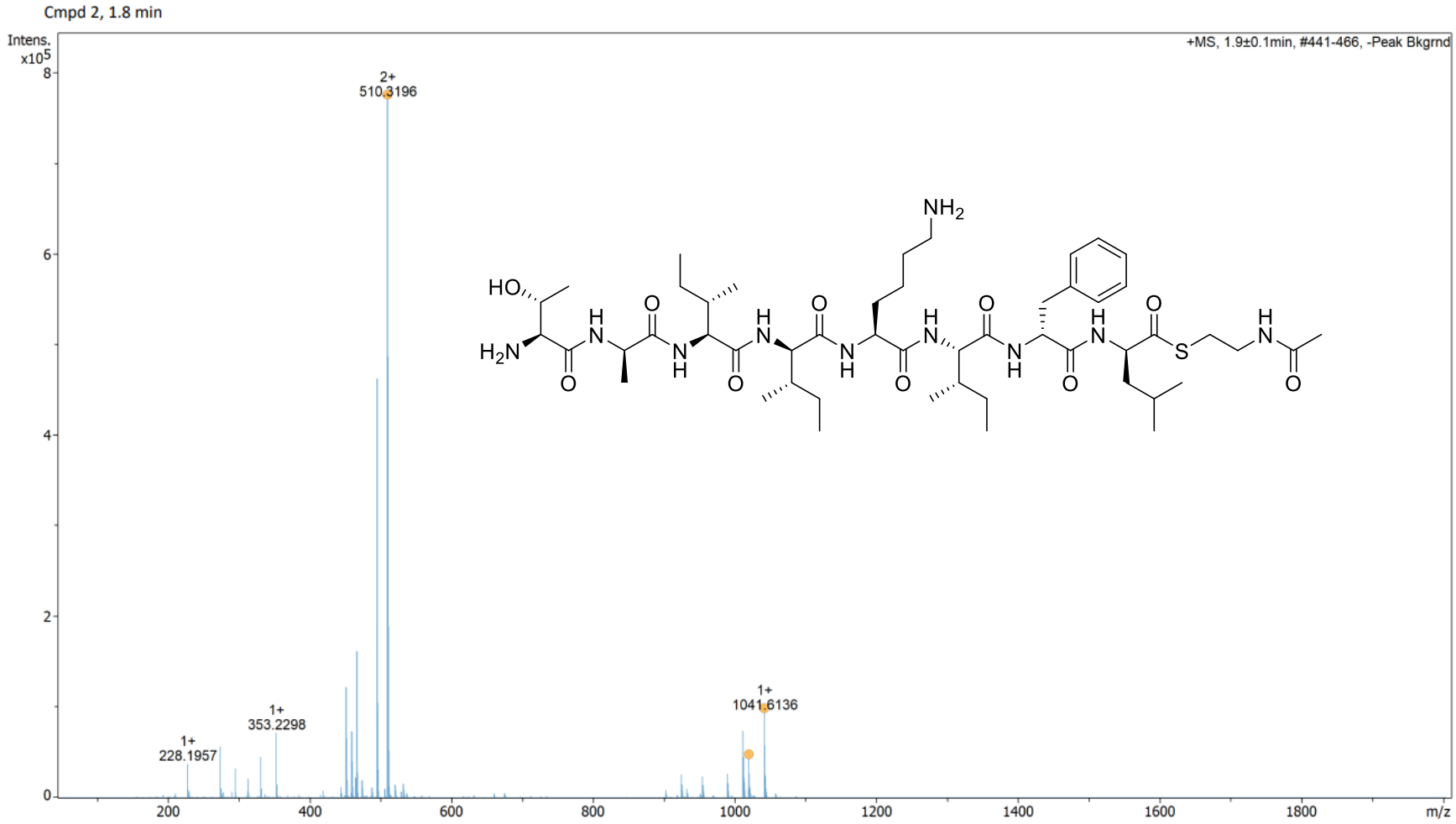

*DAD1 A, Sig=220,4 Ref=off (2025\Webb\Jake W 090625 2025-06-09 13-37-12\JW-SR Ala1.D - C:\USERS\P...TION\1\DATA\2025\WE
mAU
2500
2000
1500
1000
500
0
1.132
1.283
1.562
1.810
2.050
2.493
2.746
2.827
3.031
3.443
3.625
4.544
5.017
5.268
5.785
2
4
6
8
min

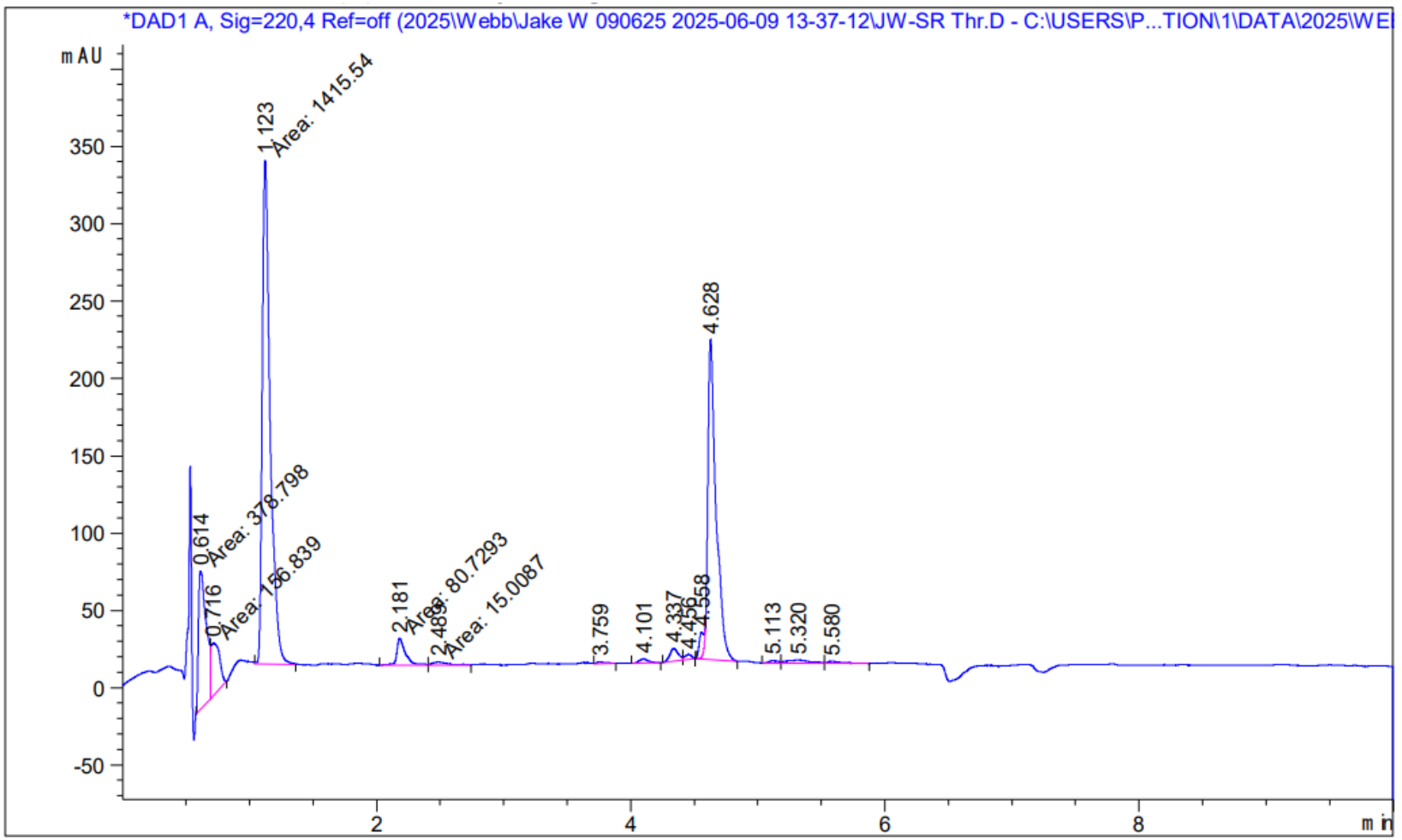

Cmpd 5, 1.8 min
+MS, 1.8±0.1min, #421-461, -Peak Bkgrnd
Intens.
x10^6
CAUTION!
CAUTION!
Too intense, dilute sample
Too intense, dilute sample
488.3089
1+
274.2738
2+
653.4262
1+
975.6053
1.25
1.00
0.75
0.50
0.25
0.00
200
400
600
800
1000
1200
1400
1600
1800
m/z

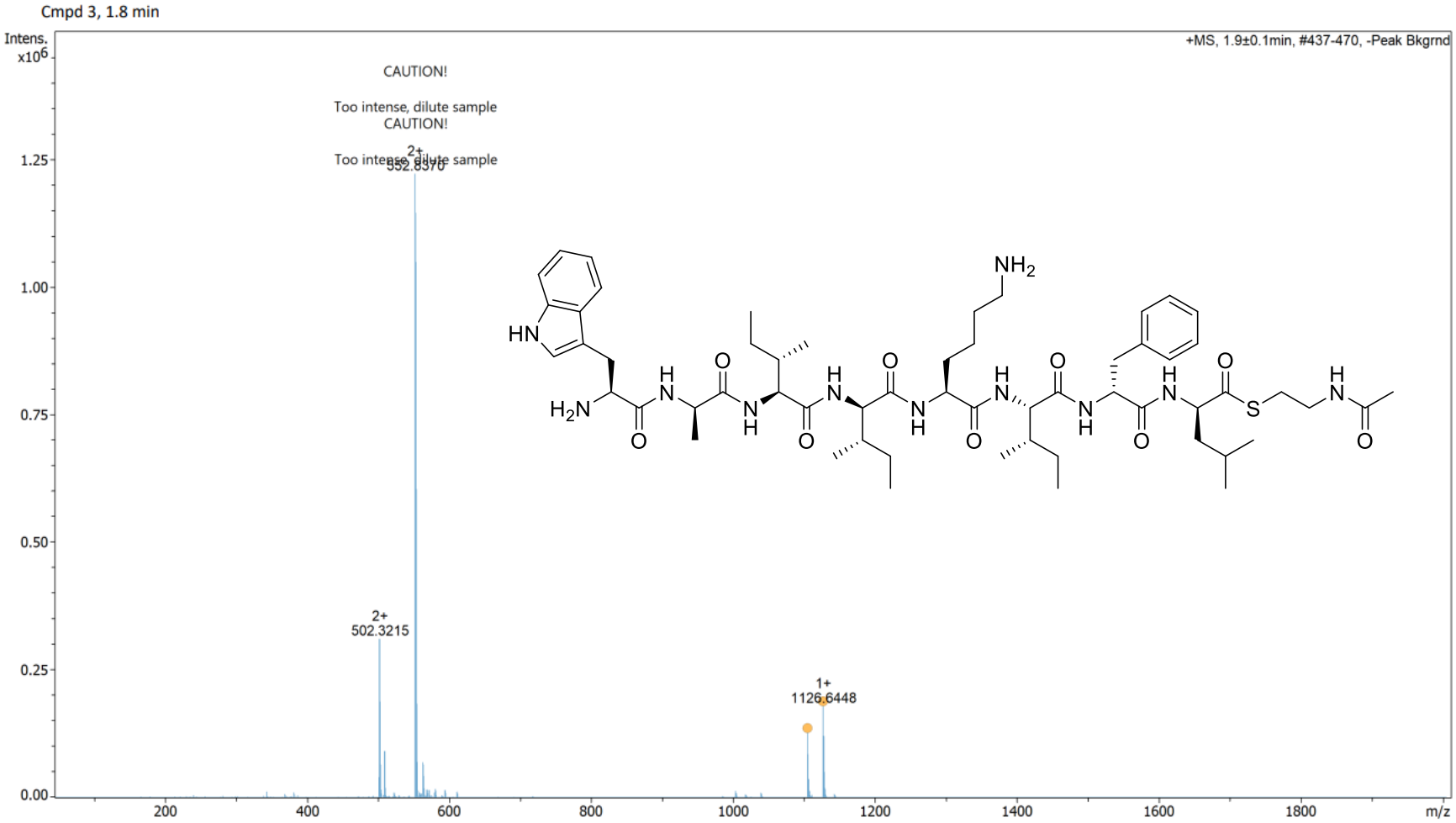

*DAD1 A, Sig=220,4 Ref=off (2025\Webb\Jake W 090625 2025-06-09 13-37-12\JW-SR Gly1.D - C:\USERS\P...TION\1\DATA\2025\WE
mAU
2000
1500
1000
500
0
1.122
1.270
1.557
1.797
2.036
2.175
2.489
3.034
3.627
4.063
4.333
4.485
4.900
5.152
5.272
5.463
2
4
6
8
min

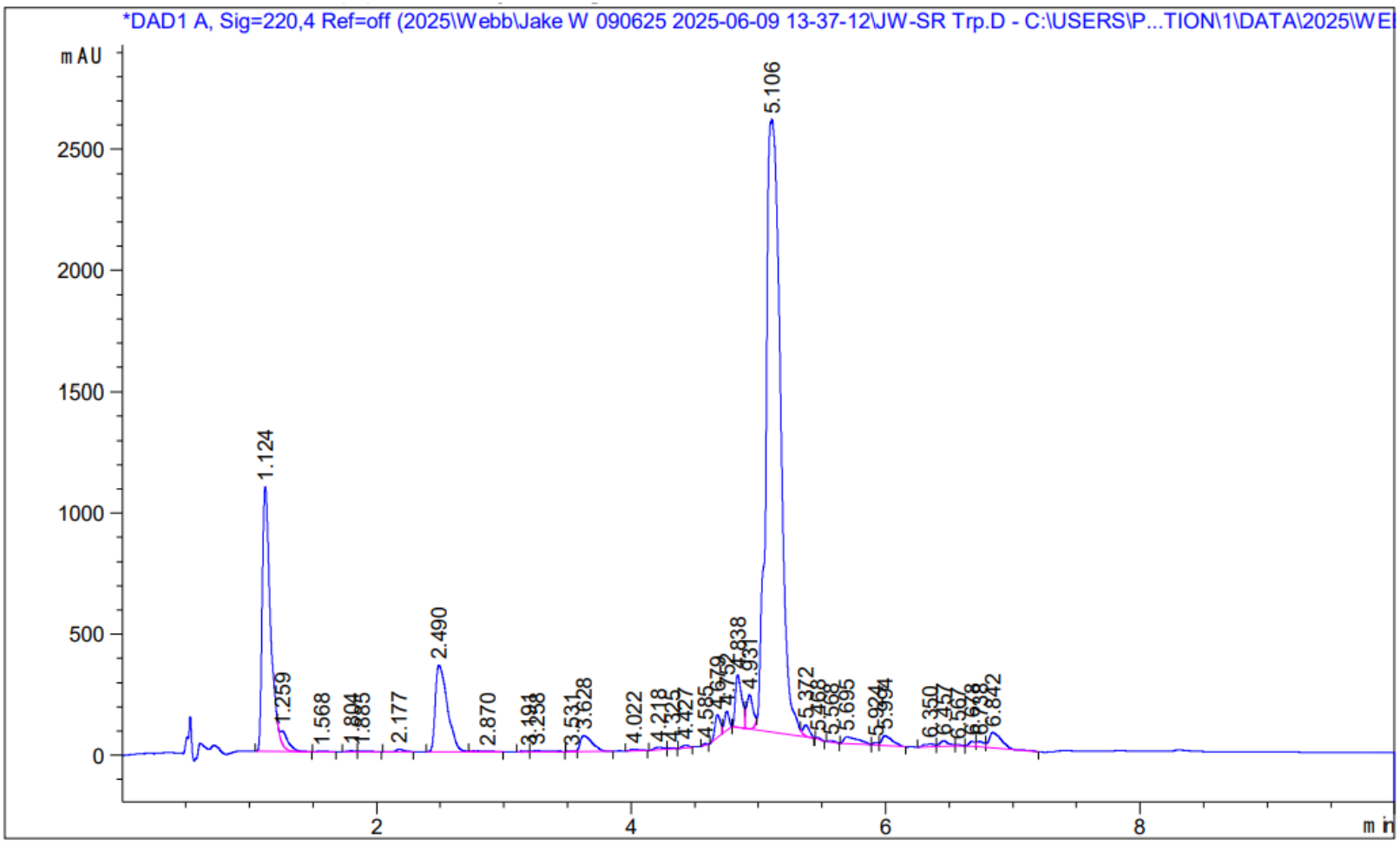

Cmpd 5, 1.7 min
+MS, 1.8±0.1min, #403-437, -Peak Bkgrnd
Intens.
x10^6
CAUTION!
CAUTION!
Too intense, dilute sample
Too intense, dilute sample
3+
349.5649
2+
523.8439
1+
246.1633
1+
120.0477
1+
959.6651
1+
1046.6791
1.25
1.00
0.75
0.50
0.25
0.00
200
400
600
800
1000
1200
1400
1600
1800
m/z

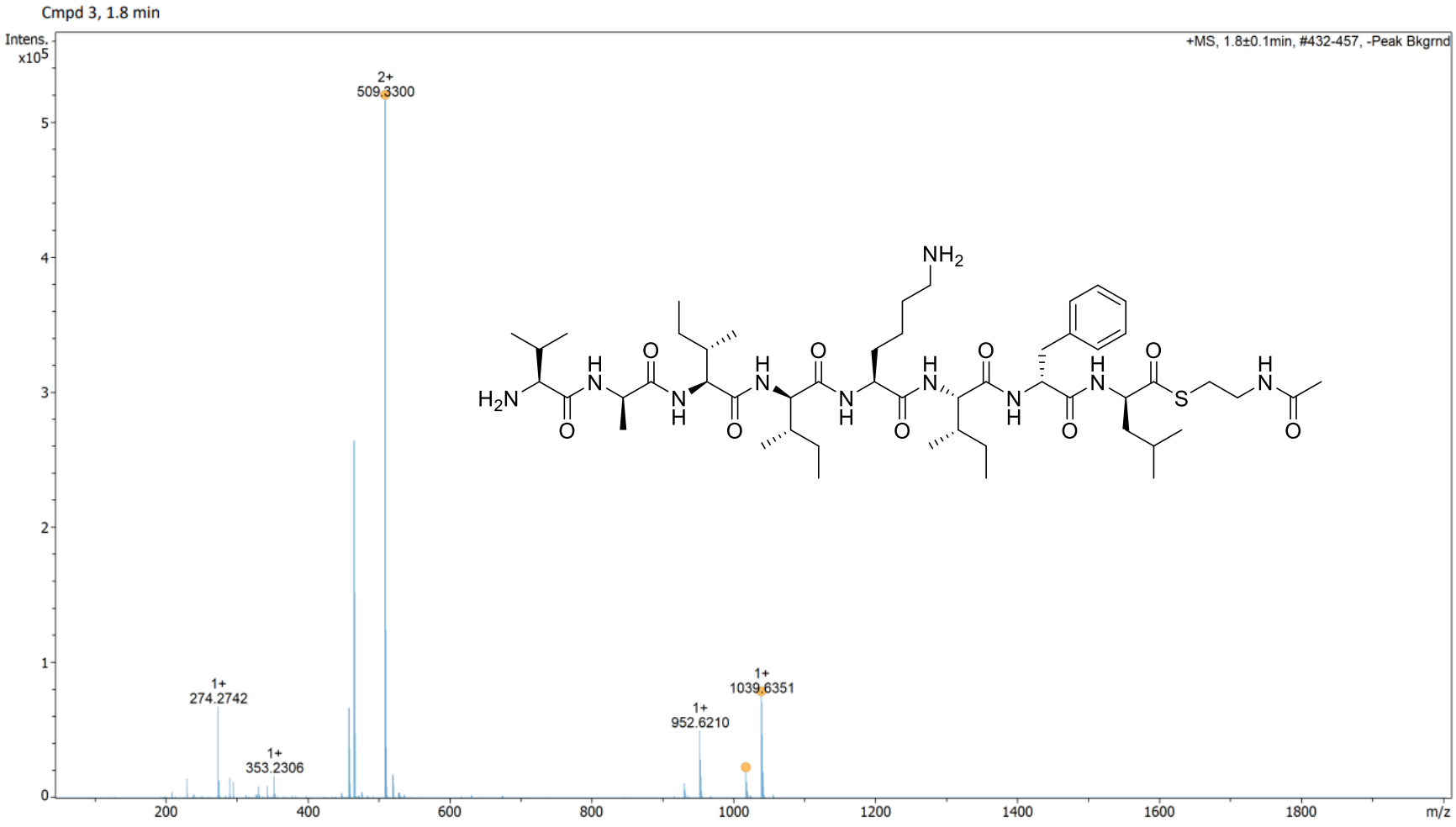

*DAD1 A, Sig=220,4 Ref=off (2025\Webb\Jake W 090625 2025-06-09 13-37-12\JW-SR Lys1.D - C:\USERS\P...TION\1\DATA\2025\W
mAU
2000
1500
1000
500
0
1.122
1.269
1.562
1.795
2.173
2.267
2.506
2.678
3.441
3.611
4.057
4.197
4.526
4.701
4.794
5.120
2
4
6
8
min

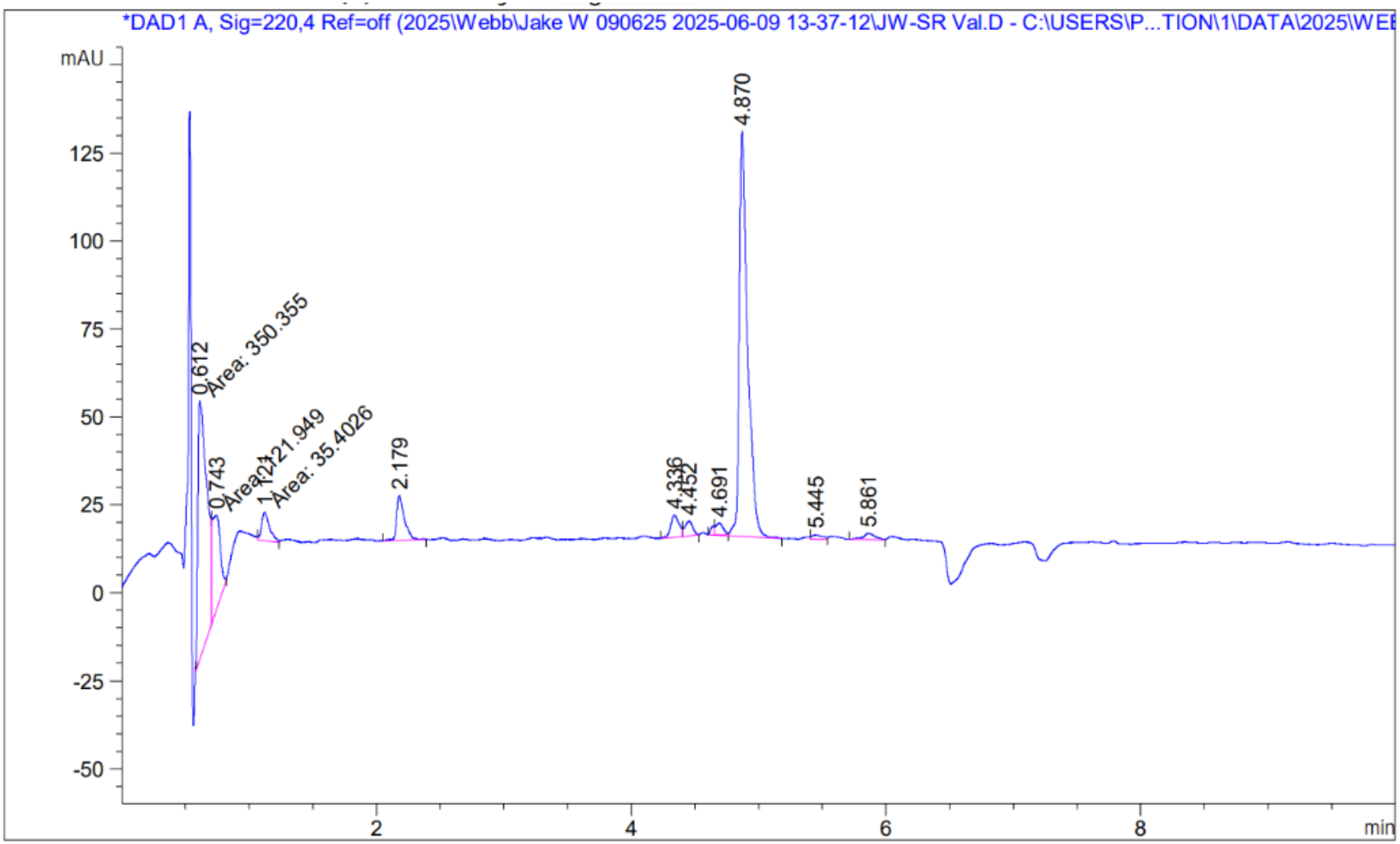

Cmpd 2, 1.8 min
Intens.
x10^5
+MS, 1.9±0.1min, #441-466, -Peak Bkgrnd
2+
510.3196
1+
353.2298
1+
228.1957
1+
1041.6136
m/z


*DAD1 A, Sig=220,4 Ref=off (2025\Webb\Jake W 090625 2025-06-09 13-37-12\JW-SR Thr.D - C:\USERS\P...TION\1\DATA\2025\WE
mAU
350
300
250
200
150
100
50
0
-50
1.123
Area: 1415.54
0.614
Area: 378.798
0.716
Area: 156.839
2.181
Area: 80.7293
2.489
Area: 15.0087
3.759
4.101
4.337
4.628
4.558
5.113
5.320
5.580
2
4
6
8
min

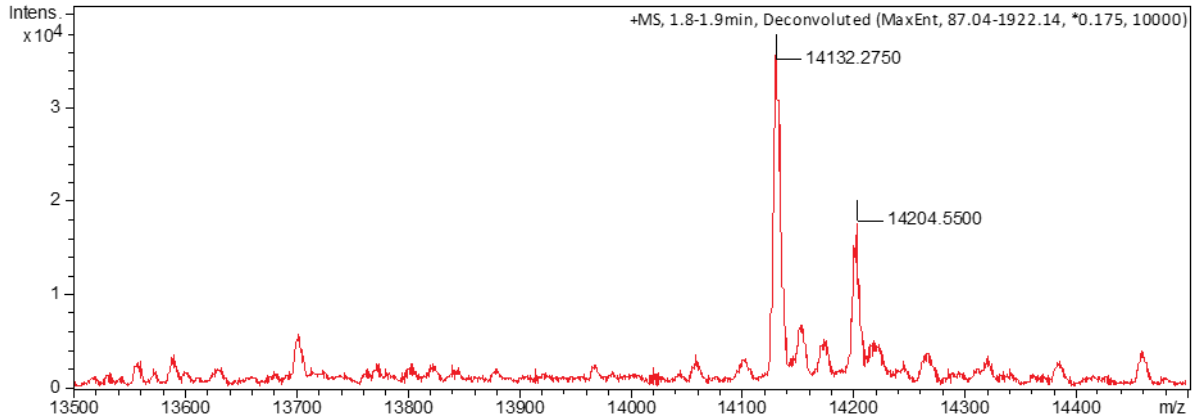

Cmpd 3, 1.8 min
+MS, 1.9±0.1min, #437-470, -Peak Bkgrnd
Intens.
x10^6
CAUTION!
Too intense, dilute sample
CAUTION!
Too intense, dilute sample
2+
552.8970
2+
502.3215
1+
1126.6448
1.25
1.00
0.75
0.50
0.25
0.00
200
400
600
800
1000
1200
1400
1600
1800
m/z
HN
H2N
NH2
O
N
H
S


*DAD1 A, Sig=220,4 Ref=off (2025\Webb\Jake W 090625 2025-06-09 13-37-12\JW-SR Trp.D - C:\USERS\P...TION\1\DATA\2025\WE
mAU
2500
2000
1500
1000
500
0
1.124
1.259
1.568
2.177
2.490
2.870
3.531
3.628
4.022
4.218
4.585
4.838
4.931
5.106
5.695
6.350
6.842
2
4
6
8
min


Cmpd 3, 1.8 min
Intens.
x10^5
+MS, 1.8±0.1min, #432-457, -Peak Bkgrnd
2+
509.3300
1+
274.2742
1+
353.2306
1+
952.6210
1+
1039.6351
m/z

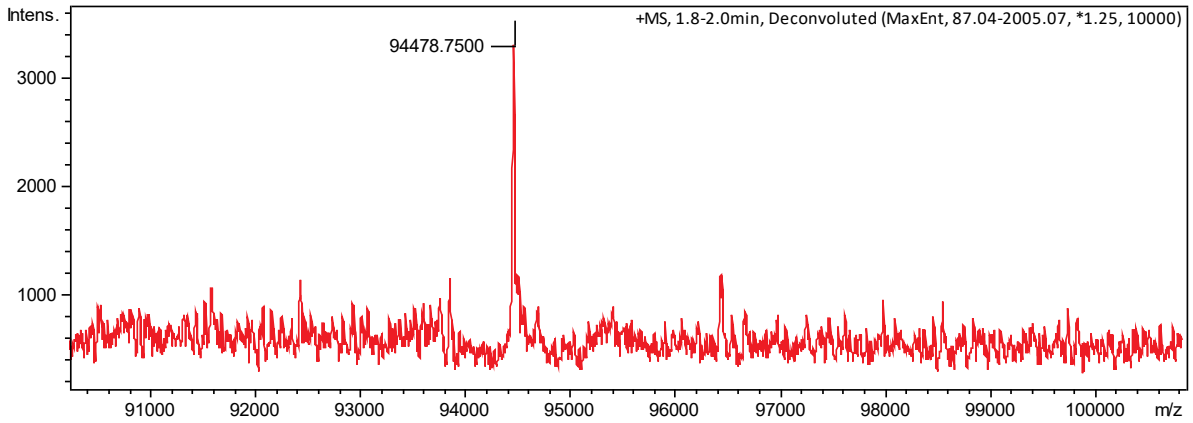

*DAD1 A, Sig=220,4 Ref=off (2025\Webb\Jake W 090625 2025-06-09 13-37-12\JW-SR Val.D - C:\USERS\P...TION\1\DATA\2025\WEE
mAU
125
100
75
50
25
0
-25
-50
0.612
Area: 350.355
0.743
Area: 21.949
1.121
Area: 35.4026
2.179
4.336
4.452
4.691
4.870
5.445
5.861
2
4
6
8
min

Growth factor receptor-bound protein (Grb2)



MGSSHHHHHHSSGRR/A/SVENLYFQGSGRAMKPHPWFFGKIPRAKAEEMLSKQRHDGAFLIRESESAPGDFSLSVKFGNDVQHFKVL
RDGAGKYFLWVVKFNSLNELVDYHRSTSVSRNQQIFLRDIEQVPQQPLINEF

Regulatory Protein (RP)



>pMAL-p5X [length=5752] [version=20-AUG-2009] [topology=circular] Cloning vector pMAL-p5X, complete sequence.
MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQD
KLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGK
YDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVL
SAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINA
ASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIEGRISHMAVESPGQSQSDDAPPPRSGEAASSLAPRASSHLDRWSRSRALRSGHRP
ALNRAALSSASVSAPPVIKSPRPEDAAVAAEDGEDDDVCEAERDAAAGKAIYIVSDGTGWTAEHSVNAALGQFENCLADRGCAVNTH
LFSLIDDMDRLIEVIKQAAKEGXLVLYTLADPSMAEATKKACDFWGVPCTDVLRPTVEAIASHIGVAPSGIPRSFPSRNGRLSEDYFQRI
DAIDFTIKQDDGVLPQNFYRADIVLAGVSRTGKTPLSIYLAQKGYKVANVPIVMGVALPKSLFEINQDKVFGLTINPAILQGIRKTRAKT
LGFDGRQSNYAEMDHVRQELVHANQIFVQNPWWPVIAVTGKAIEETAAVILGILHDRKQKCSMPRISKRY*

1. Cohen, P. The regulation of protein function by multisite phosphorylation – a 25 year update. *Trends in Biochemical Sciences.* 2000, **25**(12), pp.596-601.
2. Fuhs, S.R. and Hunter, T. pHisphorylation: the emergence of histidine phosphorylation as a reversible regulatory modification. *Current opinion in cell biology.* 2017, **45**, pp.8-16.
3. Kee, J.-M. and Muir, T.W. Chasing Phosphohistidine, an Elusive Sibling in the Phosphoamino Acid Family. *ACS Chemical Biology.* 2012, **7**(1), pp.44-51.
4. Makwana, M.V., Muimo, R. and Jackson, R.F.W. Advances in development of new tools for the study of phosphohistidine. *Laboratory Investigation.* 2018, **98**(3), pp.291-303.
5. McIntosh, L.P., Kang, H.-S., Okon, M., Nelson, M.L., Graves, B.J. and Brutscher, B. Detection and assignment of phosphoserine and phosphothreonine residues by 13C–31P spin-echo difference NMR spectroscopy. *Journal of Biomolecular NMR.* 2009, **43**(1), pp.31-37.
6. Hunter, T. Why nature chose phosphate to modify proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences.* 2012, **367**(1602), pp.2513-2516.
7. Stock, J.B., Stock, A.M. and Mottonen, J.M. Signal transduction in bacteria. *Nature.* 1990, **344**(6265), pp.395-400.
8. Attwood, P.V., Piggott, M.J., Zu, X.L. and Besant, P.G. Focus on phosphohistidine. *Amino Acids.* 2007, **32**(1), pp.145-156.
9. Harrison, C.B. and Schulten, K. Quantum and Classical Dynamics Simulations of ATP Hydrolysis in Solution. *Journal of Chemical Theory and Computation.* 2012, **8**(7), pp.2328-2335.
10. Gillette, T.G. and Hill, J.A. Readers, writers, and erasers: chromatin as the whiteboard of heart disease. *Circ Res.* 2015, **116**(7), pp.1245-1253.
11. Lim, W.A. and Pawson, T. Phosphotyrosine Signaling: Evolving a New Cellular Communication System. *Cell.* 2010, **142**(5), pp.661-667.
12. Pearlman, Samuel M., Serber, Z. and Ferrell, James E. A Mechanism for the Evolution of Phosphorylation Sites. *Cell.* 2011, **147**(4), pp.934-946.
13. Schlessinger, J. and Lemmon, M.A. SH2 and PTB Domains in Tyrosine Kinase Signaling. *Science's STKE.* 2003, **2003**(191), p.re12.
14. Yaffe, M.B. and Smerdon, S.J. PhosphoSerine/Threonine Binding Domains: You Can't pSERious? *Structure.* 2001, **9**(3), pp.R33-R38.
15. Boyer, P.D., DeLuca, M., Ebner, K.E., Hultquist, D.E. and Peter, J.B. Identification of Phosphohistidine in Digests from a Probable Intermediate of Oxidative Phosphorylation. *Journal of Biological Chemistry.* 1962, **237**(10), pp.PC3306-PC3308.
16. Matthews, H.R. Protein kinases and phosphatases that act on histidine, lysine, or arginine residues in eukaryotic proteins: A possible regulator of the mitogen-activated protein kinase cascade. *Pharmacology & Therapeutics.* 1995, **67**(3), pp.323-350.
17. Druker, B.J. STI571 (Gleevec™) as a paradigm for cancer therapy. *Trends in Molecular Medicine.* 2002, **8**(4 SUPPL.), pp.S14-S18.
18. Swanson, R.V., Alex, L.A. and Simon, M.I. Histidine and aspartate phosphorylation: two-component systems and the limits of homology. *Trends in Biochemical Sciences.* 1994, **19**(11), pp.485-490.
19. Mijakovic, I., Grangeasse, C. and Turgay, K. Exploring the diversity of protein modifications: special bacterial phosphorylation systems. *FEMS Microbiol Rev.* 2016, **40**(3), pp.398-417.

20. Khorchid, A. and Ikura, M. Bacterial histidine kinase as signal sensor and transducer. *The International Journal of Biochemistry & Cell Biology.* 2006, **38**(3), pp.307-312.
21. Perry, J., Koteva, K. and Wright, G. Receptor domains of two-component signal transduction systems. *Molecular BioSystems.* 2011, **7**(5), pp.1388-1398.
22. Modro, T.A., Lawry, M.A. and Murphy, E. Phosphoric amides. 1. Phosphorus-nitrogen vs. nitrogen-carbon bond cleavage in acidic solvolysis of N-alkyl phosphoramidates and phosphinamidates. *The Journal of Organic Chemistry.* 1978, **43**(26), pp.5000-5006.
23. Hultquist, D.E., Moyer, R.W. and Boyer, P.D. The Preparation and Characterization of 1-Phosphohistidine and 3-Phosphohistidine*. *Biochemistry.* 1966, **5**(1), pp.322-331.
24. Hultquist, D.E. The preparation and characterization of phosphorylated derivatives of histidine. *Biochimica et Biophysica Acta (BBA) - Bioenergetics.* 1968, **153**(2), pp.329-340.
25. Fuhs, Stephen R., Meisenhelder, J., Aslanian, A., Ma, L., Zagorska, A., Stankova, M., Binnie, A., Al-Obeidi, F., Mauger, J., Lemke, G., Yates, John R. and Hunter, T. Monoclonal 1- and 3-Phosphohistidine Antibodies: New Tools to Study Histidine Phosphorylation. *Cell.* 2015, **162**(1), pp.198-210.
26. Sharma, S. and Juffer, A.H. Hydrolysis of phosphohistidine in water and in prostatic acid phosphatase. *Chemical Communications.* 2009, (42), pp.6385-6387.
27. Sharma, S., Rauk, A. and Juffer, A.H. A DFT Study on the Formation of a Phosphohistidine Intermediate in Prostatic Acid Phosphatase. *Journal of the American Chemical Society.* 2008, **130**(30), pp.9708-9716.
28. Krishna Deepak, R.N.V. and Sankararamakrishnan, R. N–H···N Hydrogen Bonds Involving Histidine Imidazole Nitrogen Atoms: A New Structural Role for Histidine Residues in Proteins. *Biochemistry.* 2016, **55**(27), pp.3774-3783.
29. Smith, D.L., Bruegger, B.B., Halpern, R.M. and Smith, R.A. New Histone Kinases in Nuclei of Rat Tissues. *Nature.* 1973, **246**(5428), pp.103-104.
30. Rochdi, M.D., Laroche, G., Dupré, É., Giguère, P., Lebel, A., Watier, V., Hamelin, É., Lépine, M.-C., Dupuis, G. and Parent, J.-L. Nm23-H2 Interacts with a G Protein-coupled Receptor to Regulate Its Endocytosis through an Rac1-dependent Mechanism *. *Journal of Biological Chemistry.* 2004, **279**(18), pp.18981-18989.
31. Srivastava, S., Li, Z., Ko, K., Choudhury, P., Albaqumi, M., Johnson, A.K., Yan, Y., Backer, J.M., Unutmaz, D., Coetzee, W.A. and Skolnik, E.Y. Histidine Phosphorylation of the Potassium Channel KCa3.1 by Nucleoside Diphosphate Kinase B Is Required for Activation of KCa3.1 and CD4 T Cells. *Molecular Cell.* 2006, **24**(5), pp.665-675.
32. Cai, X., Srivastava, S., Surindran, S., Li, Z. and Skolnik, E.Y. Regulation of the epithelial Ca^{2+} channel TRPV5 by reversible histidine phosphorylation mediated by NDPK-B and PHPT1. *Mol Biol Cell.* 2014, **25**(8), pp.1244-1250.
33. Kowluru, A., Seavey, S.E., Rhodes, C.J. and Metz, S.A. A novel regulatory mechanism for trimeric GTP-binding proteins in the membrane and secretory granule fractions of human and rodent β cells. *Biochemical Journal.* 1996, **313**(1), pp.97-107.
34. Hippe, H.-J., Wolf, N.M., Abu-Taha, I., Mehringer, R., Just, S., Lutz, S., Niroomand, F., Postel, E.H., Katus, H.A., Rottbauer, W. and Wieland, T. The interaction of nucleoside diphosphate kinase B with Gβγ dimers controls heterotrimeric G protein function. *Proceedings of the National Academy of Sciences.* 2009, **106**(38), pp.16269-16274.
35. Wieland, T., Nürnberg, B., Ulibarri, I., Kaldenberg-Stasch, S., Schultz, G. and Jakobs, K.H. Guanine nucleotide-specific phosphate transfer by guanine nucleotide-binding

regulatory protein beta-subunits. Characterization of the phosphorylated amino acid. *Journal of Biological Chemistry.* 1993, **268**(24), pp.18111-18118.
36. Kornberg, R.D. Structure of Chromatin. *Annual Review of Biochemistry.* 1977, **46**(Volume 46, 1977), pp.931-954.
37. McGhee, J.D. and Felsenfeld, G. Nucleosome Structure. *Annual Review of Biochemistry.* 1980, **49**(Volume 49, 1980), pp.1115-1156.
38. Fischle, W., Wang, Y. and Allis, C.D. Histone and chromatin cross-talk. *Current opinion in cell biology.* 2003, **15**(2), pp.172-183.
39. Masaracchia, R.A., Kemp, B.E. and Walsh, D.A. Histone 4 phosphotransferase activities in proliferating lymphocytes. Partial purification and characterization of an enzyme specific for Ser-47. *Journal of Biological Chemistry.* 1977, **252**(20), pp.7109-7117.
40. Besant, P.G. and Attwood, P.V. Detection of a mammalian histone H4 kinase that has yeast histidine kinase-like enzymic activity. *The International Journal of Biochemistry & Cell Biology.* 2000, **32**(2), pp.243-253.
41. Chen, C.-C., Smith, D.L., Bruegger, B.B., Halpern, R.M. and Smith, R.A. Occurrence and distribution of acid-labile histone phosphates in regenerating rat liver. *Biochemistry.* 1974, **13**(18), pp.3785-3789.
42. Fujitaki, J.M., Fung, G., Oh, E.Y. and Smith, R.A. Characterization of chemical and enzymic acid-labile phosphorylation of histone H4 using phosphorus-31 nuclear magnetic resonance. *Biochemistry.* 1981, **20**(12), pp.3658-3664.
43. Guertin, D.A. and Wellen, K.E. Acetyl-CoA metabolism in cancer. *Nat Rev Cancer.* 2023, **23**(3), pp.156-172.
44. Assmann, N., O'Brien, K.L., Donnelly, R.P., Dyck, L., Zaiatz-Bittencourt, V., Loftus, R.M., Heinrich, P., Oefner, P.J., Lynch, L., Gardiner, C.M., Dettmer, K. and Finlay, D.K. Srebp-controlled glucose metabolism is essential for NK cell functional responses. *Nature Immunology.* 2017, **18**(11), pp.1197-1206.
45. Patel, M.S. and Roche, T.E. Molecular biology and biochemistry of pyruvate dehydrogenase complexes. *The FASEB Journal.* 1990, **4**(14), pp.3224-3233.
46. Dawson, A.G. Oxidation of cytosolic NADH formed during aerobic metabolism in mammalian cells. *Trends in Biochemical Sciences.* 1979, **4**(8), pp.171-176.
47. Watson, J.A., Fang, M. and Lowenstein, J.M. Tricarballylate and hydroxycitrate: substrate and inhibitor of ATP: citrate oxaloacetate lyase. *Arch Biochem Biophys.* 1969, **135**(1), pp.209-217.
48. Bradshaw, P.C. Acetyl-CoA Metabolism and Histone Acetylation in the Regulation of Aging and Lifespan. *Antioxidants (Basel).* 2021, **10**(4).
49. Davis, M.S., Solbiati, J. and Cronan, J.E., Jr. Overproduction of Acetyl-CoA Carboxylase Activity Increases the Rate of Fatty Acid Biosynthesis in Escherichia coli *. *Journal of Biological Chemistry.* 2000, **275**(37), pp.28593-28598.
50. Shi, L. and Tu, B.P. Acetyl-CoA and the regulation of metabolism: mechanisms and consequences. *Current opinion in cell biology.* 2015, **33**, pp.125-131.
51. Fan, F., Williams, H.J., Boyer, J.G., Graham, T.L., Zhao, H., Lehr, R., Qi, H., Schwartz, B., Raushel, F.M. and Meek, T.D. On the Catalytic Mechanism of Human ATP Citrate Lyase. *Biochemistry.* 2012, **51**(25), pp.5198-5211.
52. Wagner, P.D. and Vu, N.-D. Phosphorylation of ATP-Citrate Lyase by Nucleoside Diphosphate Kinase. *Journal of Biological Chemistry.* 1995, **270**(37), pp.21758-21764.
53. Baba, Y. and Kurosaki, T. Role of Calcium Signaling in B Cell Activation and Biology. *Curr Top Microbiol Immunol.* 2016, **393**, pp.143-174.

54. Di, L., Srivastava, S., Zhdanova, O., Ding, Y., Li, Z., Wulff, H., Lafaille, M. and Skolnik, E.Y. Inhibition of the K+ channel KCa3.1 ameliorates T cell–mediated colitis. *Proceedings of the National Academy of Sciences.* 2010, **107**(4), pp.1541-1546.
55. Winslow, M.M. and Crabtree, G.R. Decoding Calcium Signaling. *Science.* 2005, **307**(5706), pp.56-57.
56. Feske, S. Calcium signalling in lymphocyte activation and disease. *Nature Reviews Immunology.* 2007, **7**(9), pp.690-702.
57. Beeton, C. and Chandy, K.G. Potassium channels, memory T cells, and multiple sclerosis. *Neuroscientist.* 2005, **11**(6), pp.550-562.
58. Maher, A.D. and Kuchel, P.W. The Gárdos channel: a review of the Ca2+-activated K+ channel in human erythrocytes. *Int J Biochem Cell Biol.* 2003, **35**(8), pp.1182-1197.
59. Fanger, C.M., Ghanshani, S., Logsdon, N.J., Rauer, H., Kalman, K., Zhou, J., Beckingham, K., Chandy, K.G., Cahalan, M.D. and Aiyar, J. Calmodulin mediates calcium-dependent activation of the intermediate conductance KCa channel, IKCa1. *J Biol Chem.* 1999, **274**(9), pp.5746-5754.
60. Pinto, M.C.X., Kihara, A.H., Goulart, V.A.M., Tonelli, F.M.P., Gomes, K.N., Ulrich, H. and Resende, R.R. Calcium signaling and cell proliferation. *Cellular Signalling.* 2015, **27**(11), pp.2139-2149.
61. Srivastava, S., Panda, S., Li, Z., Fuhs, S.R., Hunter, T., Thiele, D.J., Hubbard, S.R. and Skolnik, E.Y. Histidine phosphorylation relieves copper inhibition in the mammalian potassium channel KCa3.1. *eLife.* 2016, **5**, p.e16093.
62. Srivastava, S., Zhdanova, O., Di, L., Li, Z., Albaqumi, M., Wulff, H. and Skolnik, E.Y. Protein histidine phosphatase 1 negatively regulates CD4 T cells by inhibiting the K⁺ channel KCa3.1. *Proceedings of the National Academy of Sciences.* 2008, **105**(38), pp.14442-14446.
63. Panda, S., Srivastava, S., Li, Z., Vaeth, M., Fuhs, S.R., Hunter, T. and Skolnik, E.Y. Identification of PGAM5 as a Mammalian Protein Histidine Phosphatase that Plays a Central Role to Negatively Regulate CD4(+) T Cells. *Molecular Cell.* 2016, **63**(3), pp.457-469.
64. Di, L., Srivastava, S., Zhdanova, O., Sun, Y., Li, Z. and Skolnik, E.Y. Nucleoside diphosphate kinase B knock-out mice have impaired activation of the K+ channel KCa3.1, resulting in defective T cell activation. *J Biol Chem.* 2010, **285**(50), pp.38765-38771.
65. Na, T. and Peng, J.-B. TRPV5: A Ca2+ Channel for the Fine-Tuning of Ca2+ Reabsorption. In: Nilius, B. and Flockerzi, V. eds. *Mammalian Transient Receptor Potential (TRP) Cation Channels: Volume I.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp.321-357.
66. Zhou, Y. and Greka, A. Calcium-permeable ion channels in the kidney. *American Journal of Physiology-Renal Physiology.* 2016, **310**(11), pp.F1157-F1167.
67. Hughes, T.E.T., Pumroy, R.A., Yazici, A.T., Kasimova, M.A., Fluck, E.C., Huynh, K.W., Samanta, A., Molugu, S.K., Zhou, Z.H., Carnevale, V., Rohacs, T. and Moiseenkova-Bell, V.Y. Structural insights on TRPV5 gating by endogenous modulators. *Nature Communications.* 2018, **9**(1), p.4198.
68. Boros, S., Bindels, R.J. and Hoenderop, J.G. Active Ca(2+) reabsorption in the connecting tubule. *Pflugers Arch.* 2009, **458**(1), pp.99-109.
69. Hoenderop, J.G., van Leeuwen, J.P., van der Eerden, B.C., Kersten, F.F., van der Kemp, A.W., Mérillat, A.M., Waarsing, J.H., Rossier, B.C., Vallon, V., Hummler, E. and Bindels, R.J. Renal Ca2+ wasting, hyperabsorption, and reduced bone thickness in mice lacking TRPV5. *J Clin Invest.* 2003, **112**(12), pp.1906-1914.

70. Reid, I.R., Birstow, S.M. and Bolland, M.J. Calcium and Cardiovascular Disease. *Endocrinol Metab (Seoul).* 2017, **32**(3), pp.339-349.
71. Oddsson, A., Sulem, P., Helgason, H., Edvardsson, V.O., Thorleifsson, G., Sveinbjörnsson, G., Haraldsdottir, E., Eyjolfsson, G.I., Sigurdardottir, O., Olafsson, I., Masson, G., Holm, H., Gudbjartsson, D.F., Thorsteinsdottir, U., Indridason, O.S., Palsson, R. and Stefansson, K. Common and rare variants associated with kidney stones and biochemical traits. *Nat Commun.* 2015, **6**, p.7975.
72. de Groot, T., Kovalevskaya, N.V., Verkaart, S., Schilderink, N., Felici, M., van der Hagen, E.A.E., Bindels, R.J.M., Vuister, G.W. and Hoenderop, J.G. Molecular Mechanisms of Calmodulin Action on TRPV5 and Modulation by Parathyroid Hormone. *Molecular and Cellular Biology.* 2011, **31**(14), pp.2845-2853.
73. van der Wijst, J., van Goor, M.K., Schreuder, M.F. and Hoenderop, J.G. TRPV5 in renal tubular calcium handling and its potential relevance for nephrolithiasis. *Kidney International.* 2019, **96**(6), pp.1283-1291.
74. Clapham, D.E., Runnels, L.W. and Strübing, C. The trp ion channel family. *Nature Reviews Neuroscience.* 2001, **2**(6), pp.387-396.
75. Tiruppathi, C., Ahmmed, G.U., Vogel, S.M. and Malik, A.B. Ca2+ signaling, TRP channels, and endothelial permeability. *Microcirculation.* 2006, **13**(8), pp.693-708.
76. Zeng, B., Yuan, C., Yang, X., Atkin, S.L. and Xu, S.Z. TRPC channels and their splice variants are essential for promoting human ovarian cancer cell proliferation and tumorigenesis. *Curr Cancer Drug Targets.* 2013, **13**(1), pp.103-116.
77. Cheung, S.Y., Henrot, M., Al-Saad, M., Baumann, M., Muller, H., Unger, A., Rubaiy, H.N., Mathar, I., Dinkel, K., Nussbaumer, P., Klebl, B., Freichel, M., Rode, B., Trainor, S., Clapcote, S.J., Christmann, M., Waldmann, H., Abbas, S.K., Beech, D.J. and Vasudev, N.S. TRPC4/TRPC5 channels mediate adverse reaction to the cancer cell cytotoxic agent (-)-Englerin A. *Oncotarget.* 2018, **9**(51), pp.29634-29643.
78. Carson, C., Raman, P., Tullai, J., Xu, L., Henault, M., Thomas, E., Yeola, S., Lao, J., McPate, M., Verkuyl, J.M., Marsh, G., Sarber, J., Amaral, A., Bailey, S., Lubicka, D., Pham, H., Miranda, N., Ding, J., Tang, H.-M., Ju, H., Tranter, P., Ji, N., Krastel, P., Jain, R.K., Schumacher, A.M., Loureiro, J.J., George, E., Berellini, G., Ross, N.T., Bushell, S.M., Erdemli, G. and Solomon, J.M. Englerin A Agonizes the TRPC4/C5 Cation Channels to Inhibit Tumor Cell Line Proliferation. *PLOS ONE.* 2015, **10**(6), p.e0127498.
79. Moccia, F., Brunetti, V., Perna, A., Guerra, G., Soda, T. and Berra-Romani, R. The Molecular Heterogeneity of Store-Operated Ca(2+) Entry in Vascular Endothelial Cells: The Different roles of Orai1 and TRPC1/TRPC4 Channels in the Transition from Ca(2+)-Selective to Non-Selective Cation Currents. *Int J Mol Sci.* 2023, **24**(4).
80. Srivastava, S., Li, Z., Soomro, I., Sun, Y., Wang, J., Bao, L., Coetzee, W.A., Stanley, C.A., Li, C. and Skolnik, E.Y. Regulation of K(ATP) Channel Trafficking in Pancreatic β-Cells by Protein Histidine Phosphorylation. *Diabetes.* 2018, **67**(5), pp.849-860.
81. Schnabel, P. and Böhm, M. Heterotrimeric G proteins in heart disease. *Cellular Signalling.* 1996, **8**(6), pp.413-423.
82. Eschenhagen, T. G proteins and the heart. *Cell Biology International.* 1993, **17**(8), pp.723-749.
83. Tan, L., Yan, W., McCorvy, J.D. and Cheng, J. Biased Ligands of G Protein-Coupled Receptors (GPCRs): Structure–Functional Selectivity Relationships (SFSRs) and Therapeutic Potential. *Journal of Medicinal Chemistry.* 2018, **61**(22), pp.9841-9878.
84. Felline, A., Gentile, S. and Fanelli, F. psnGPCRdb: The Structure-network Database of G Protein Coupled Receptors. *Journal of Molecular Biology.* 2023, **435**(14), p.167950.

85. Schlegel, P., Reinkober, J., Meinhardt, E., Tscheschner, H., Gao, E., Schumacher, S.M., Yuan, A., Backs, J., Most, P., Wieland, T., Koch, W.J., Katus, H.A. and Raake, P.W. G protein-coupled receptor kinase 2 promotes cardiac hypertrophy. *PLOS ONE.* 2017, **12**(7), p.e0182110.
86. Eckhart, A.D., Ozaki, T., Tevaearai, H., Rockman, H.A. and Koch, W.J. Vascular-targeted overexpression of G protein-coupled receptor kinase-2 in transgenic mice attenuates beta-adrenergic receptor signaling and increases resting blood pressure. *Mol Pharmacol.* 2002, **61**(4), pp.749-758.
87. Ishizaka, N., Alexander, R.W., Laursen, J.B., Kai, H., Fukui, T., Oppermann, M., Lefkowitz, R.J., Lyons, P.R. and Griendling, K.K. G Protein-coupled Receptor Kinase 5 in Cultured Vascular Smooth Muscle Cells and Rat Aorta: REGULATION BY ANGIOTENSIN II AND HYPERTENSION *. *Journal of Biological Chemistry.* 1997, **272**(51), pp.32482-32488.
88. Cuello, F., Schulze, R.A., Heemeyer, F., Meyer, H.E., Lutz, S., Jakobs, K.H., Niroomand, F. and Wieland, T. Activation of Heterotrimeric G Proteins by a High Energy Phosphate Transfer via Nucleoside Diphosphate Kinase (NDPK) B and Gβ Subunits: COMPLEX FORMATION OF NDPK B WITH Gβγ DIMERS AND PHOSPHORYLATION OF His-266 IN Gβ *. *Journal of Biological Chemistry.* 2003, **278**(9), pp.7220-7226.
89. Abu-Taha, I.H., Heijman, J., Feng, Y., Vettel, C., Dobrev, D. and Wieland, T. Regulation of heterotrimeric G-protein signaling by NDPK/NME proteins and caveolins: an update. *Laboratory Investigation.* 2018, **98**(2), pp.190-197.
90. Niroomand, F., Mura, R., Jakobs, K.H., Rauch, B. and Kübler, W. Receptor-Independent Activation of Cardiac Adenylyl Cyclase by GDP and Membrane-Associated Nucleoside Diphosphate Kinase. A New Cardiotonic Mechanism? *Journal of Molecular and Cellular Cardiology.* 1997, **29**(5), pp.1479-1486.
91. Hippe, H.-J., Luedde, M., Lutz, S., Koehler, H., Eschenhagen, T., Frey, N., Katus, H.A., Wieland, T. and Niroomand, F. Regulation of Cardiac cAMP Synthesis and Contractility by Nucleoside Diphosphate Kinase B/G Protein βγ Dimer Complexes. *Circulation Research.* 2007, **100**(8), pp.1191-1199.
92. Hindupur, S.K., Colombi, M., Fuhs, S.R., Matter, M.S., Guri, Y., Adam, K., Cornu, M., Piscuoglio, S., Ng, C.K.Y., Betz, C., Liko, D., Quagliata, L., Moes, S., Jenoe, P., Terracciano, L.M., Heim, M.H., Hunter, T. and Hall, M.N. The protein histidine phosphatase LHPP is a tumour suppressor. *Nature.* 2018, **555**(7698), pp.678-682.
93. Steeg, P.S., Palmieri, D., Ouatas, T. and Salerno, M. Histidine kinases and histidine phosphorylated proteins in mammalian cell biology, signal transduction and cancer. *Cancer Letters.* 2003, **190**(1), pp.1-12.
94. Adam, K., Ning, J., Reina, J. and Hunter, T. NME/NM23/NDPK and Histidine Phosphorylation. *Int J Mol Sci.* 2020, **21**(16).
95. Ferguson, S.M. and De Camilli, P. Dynamin, a membrane-remodelling GTPase. *Nature Reviews Molecular Cell Biology.* 2012, **13**(2), pp.75-88.
96. Boissan, M., Montagnac, G., Shen, Q., Griparic, L., Guitton, J., Romao, M., Sauvonnet, N., Lagache, T., Lascu, I., Raposo, G., Desbourdes, C., Schlattner, U., Lacombe, M.-L., Polo, S., van der Bliek, A.M., Roux, A. and Chavrier, P. Nucleoside diphosphate kinases fuel dynamin superfamily proteins with GTP for membrane remodeling. *Science.* 2014, **344**(6191), pp.1510-1515.
97. Hunter, T. A journey from phosphotyrosine to phosphohistidine and beyond. *Mol Cell.* 2022, **82**(12), pp.2190-2200.
98. Prunier, C., Chavrier, P. and Boissan, M. Mechanisms of action of NME metastasis suppressors - a family affair. *Cancer Metastasis Rev.* 2023, **42**(4), pp.1155-1167.

99. Boissan, M., Schlattner, U. and Lacombe, M.-L. The NDPK/NME superfamily: state of the art. *Laboratory Investigation.* 2018, **98**(2), pp.164-174.
100. Lacombe, M.-L., Tokarska-Schlattner, M., Boissan, M. and Schlattner, U. The mitochondrial nucleoside diphosphate kinase (NDPK-D/NME4), a moonlighting protein for cell homeostasis. *Laboratory Investigation.* 2018, **98**(5), pp.582-588.
101. Klumpp, S., Bechmann, G., Mäurer, A., Selke, D. and Krieglstein, J. ATP-citrate lyase as a substrate of protein histidine phosphatase in vertebrates. *Biochemical and Biophysical Research Communications.* 2003, **306**(1), pp.110-115.
102. Mäurer, A., Wieland, T., Meissl, F., Niroomand, F., Mehringer, R., Krieglstein, J. and Klumpp, S. The β-subunit of G proteins is a substrate of protein histidine phosphatase. *Biochemical and Biophysical Research Communications.* 2005, **334**(4), pp.1115-1120.
103. Busam, R.D., Thorsell, A.-G., Flores, A., Hammarström, M., Persson, C. and Hallberg, B.M. First Structure of a Eukaryotic Phosphohistidine Phosphatase *. *Journal of Biological Chemistry.* 2006, **281**(45), pp.33830-33834.
104. Ma, R., Kanders, E., Sundh, U.B., Geng, M., Ek, P., Zetterqvist, Ö. and Li, J.-P. Mutational study of human phosphohistidine phosphatase: Effect on enzymatic activity. *Biochemical and Biophysical Research Communications.* 2005, **337**(3), pp.887-891.
105. Sekine, S., Kanamaru, Y., Koike, M., Nishihara, A., Okada, M., Kinoshita, H., Kamiyama, M., Maruyama, J., Uchiyama, Y., Ishihara, N., Takeda, K. and Ichijo, H. Rhomboid Protease PARL Mediates the Mitochondrial Membrane Potential Loss-induced Cleavage of PGAM5 *. *Journal of Biological Chemistry.* 2012, **287**(41), pp.34635-34645.
106. Chen, G., Han, Z., Feng, D., Chen, Y., Chen, L., Wu, H., Huang, L., Zhou, C., Cai, X., Fu, C., Duan, L., Wang, X., Liu, L., Liu, X., Shen, Y., Zhu, Y. and Chen, Q. A Regulatory Signaling Loop Comprising the PGAM5 Phosphatase and CK2 Controls Receptor-Mediated Mitophagy. *Molecular Cell.* 2014, **54**(3), pp.362-377.
107. Takeda, K., Komuro, Y., Hayakawa, T., Oguchi, H., Ishida, Y., Murakami, S., Noguchi, T., Kinoshita, H., Sekine, Y., Iemura, S., Natsume, T. and Ichijo, H. Mitochondrial phosphoglycerate mutase 5 uses alternate catalytic activity as a protein serine/threonine phosphatase to activate ASK1. *Proc Natl Acad Sci U S A.* 2009, **106**(30), pp.12301-12305.
108. Fíla, J. and Honys, D. Enrichment techniques employed in phosphoproteomics. *Amino Acids.* 2012, **43**(3), pp.1025-1047.
109. Andersson, L. and Porath, J. Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. *Anal Biochem.* 1986, **154**(1), pp.250-254.
110. Muimo, R., Hornickova, Z., Riemen, C.E., Gerke, V., Matthews, H. and Mehta, A. Histidine Phosphorylation of Annexin I in Airway Epithelia *. *Journal of Biological Chemistry.* 2000, **275**(47), pp.36632-36636.
111. Napper, S., Kindrachuk, J., Olson, D.J.H., Ambrose, S.J., Dereniwsky, C. and Ross, A.R.S. Selective Extraction and Characterization of a Histidine-Phosphorylated Peptide Using Immobilized Copper(II) Ion Affinity Chromatography and Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Analytical Chemistry.* 2003, **75**(7), pp.1741-1747.
112. Andersson, L. Recognition of phosphate groups by immobilized aluminium(III) ions. *Journal of Chromatography A.* 1991, **539**(2), pp.327-334.
113. Posewitz, M.C. and Tempst, P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem.* 1999, **71**(14), pp.2883-2892.
114. Thingholm, T.E., Jensen, O.N., Robinson, P.J. and Larsen, M.R. SIMAC (Sequential Elution from IMAC), a Phosphoproteomics Strategy for the Rapid Separation of

Monophosphorylated from Multiply Phosphorylated Peptides *. *Molecular & Cellular Proteomics.* 2008, **7**(4), pp.661-671.
115. Liang, S.-S., Makamba, H., Huang, S.-Y. and Chen, S.-H. Nano-titanium dioxide composites for the enrichment of phosphopeptides. *Journal of Chromatography A.* 2006, **1116**(1), pp.38-45.
116. Kweon, H.K. and Håkansson, K. Selective Zirconium Dioxide-Based Enrichment of Phosphorylated Peptides for Mass Spectrometric Analysis. *Analytical Chemistry.* 2006, **78**(6), pp.1743-1749.
117. Li, Y., Lin, H., Deng, C., Yang, P. and Zhang, X. Highly selective and rapid enrichment of phosphorylated peptides using gallium oxide-coated magnetic microspheres for MALDI-TOF-MS and nano-LC-ESI-MS/MS/MS analysis. *PROTEOMICS.* 2008, **8**(2), pp.238-249.
118. Ficarro, S.B., Parikh, J.R., Blank, N.C. and Marto, J.A. Niobium(V) Oxide (Nb2O5): Application to Phosphoproteomics. *Analytical Chemistry.* 2008, **80**(12), pp.4606-4613.
119. Leitner, A. Phosphopeptide enrichment using metal oxide affinity chromatography. *TrAC Trends in Analytical Chemistry.* 2010, **29**(2), pp.177-185.
120. Pinkse, M.W.H., Uitto, P.M., Hilhorst, M.J., Ooms, B. and Heck, A.J.R. Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS/MS and Titanium Oxide Precolumns. *Analytical Chemistry.* 2004, **76**(14), pp.3935-3943.
121. Kyono, Y., Sugiyama, N., Imami, K., Tomita, M. and Ishihama, Y. Successive and Selective Release of Phosphorylated Peptides Captured by Hydroxy Acid-Modified Metal Oxide Chromatography. *Journal of Proteome Research.* 2008, **7**(10), pp.4585-4593.
122. Alpert, A.J., Hudecz, O. and Mechtler, K. Anion-Exchange Chromatography of Phosphopeptides: Weak Anion Exchange versus Strong Anion Exchange and Anion-Exchange Chromatography versus Electrostatic Repulsion–Hydrophilic Interaction Chromatography. *Analytical Chemistry.* 2015, **87**(9), pp.4704-4711.
123. Soto Pérez, J. and Frey, D.D. Behavior of the Inadvertent pH Transient Formed by a Salt Gradient in the Ion-Exchange Chromatography of Proteins. *Biotechnology Progress.* 2005, **21**(3), pp.902-910.
124. Dai, J., Wang, L.S., Wu, Y.B., Sheng, Q.H., Wu, J.R., Shieh, C.H. and Zeng, R. Fully automatic separation and identification of phosphopeptides by continuous pH-gradient anion exchange online coupled with reversed-phase liquid chromatography mass spectrometry. *J Proteome Res.* 2009, **8**(1), pp.133-141.
125. Grønborg, M., Kristiansen, T.Z., Stensballe, A., Andersen, J.S., Ohara, O., Mann, M., Jensen, O.N. and Pandey, A. A Mass Spectrometry-based Proteomic Approach for Identification of Serine/Threonine-phosphorylated Proteins by Enrichment with Phospho-specific Antibodies: Identification of a Novel Protein, Frigg, as a Protein Kinase A Substrate *. *Molecular & Cellular Proteomics.* 2002, **1**(7), pp.517-527.
126. Lombardi, B., Rendell, N., Edwards, M., Katan, M. and Zimmermann, J.G. Evaluation of phosphopeptide enrichment strategies for quantitative TMT analysis of complex network dynamics in cancer-associated cell signalling. *EuPA Open Proteomics.* 2015, **6**, pp.10-15.
127. Hardman, G., Perkins, S., Brownridge, P.J., Clarke, C.J., Byrne, D.P., Campbell, A.E., Kalyuzhnyy, A., Myall, A., Eyers, P.A., Jones, A.R. and Eyers, C.E. Strong anion exchange-mediated phosphoproteomics reveals extensive human non-canonical phosphorylation. *The EMBO journal.* 2019, **38**(21), pp.e100847-e100847.
128. Sickmann, A. and Meyer, H.E. Phosphoamino acid analysis. *PROTEOMICS.* 2001, **1**(2), pp.200-206.

129. Kleinnijenhuis, A.J., Kjeldsen, F., Kallipolitis, B., Haselmann, K.F. and Jensen, O.N. Analysis of Histidine Phosphorylation Using Tandem MS and Ion−Electron Reactions. *Analytical Chemistry.* 2007, **79**(19), pp.7450-7456.
130. Kapková, P., Lattová, E. and Perreault, H. Nonretentive Solid-Phase Extraction of Phosphorylated Peptides from Complex Peptide Mixtures for Detection by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Analytical Chemistry.* 2006, **78**(19), pp.7027-7033.
131. Han, G., Ye, M., Zhou, H., Jiang, X., Feng, S., Jiang, X., Tian, R., Wan, D., Zou, H. and Gu, J. Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. *PROTEOMICS.* 2008, **8**(7), pp.1346-1361.
132. Adam, K., Fuhs, S., Meisenhelder, J., Aslanian, A., Diedrich, J., Moresco, J., La Clair, J., Yates, J.R. and Hunter, T. A non-acidic method using hydroxyapatite and phosphohistidine monoclonal antibodies allows enrichment of phosphopeptides containing non-conventional phosphorylations for mass spectrometry analysis. *bioRxiv.* 2019, p.691352.
133. Lapek, J.D., Tombline, G., Kellersberger, K.A., Friedman, M.R. and Friedman, A.E. Evidence of histidine and aspartic acid phosphorylation in human prostate cancer cells. *Naunyn-Schmiedeberg's Archives of Pharmacology.* 2015, **388**(2), pp.161-173.
134. Oslund, R.C., Kee, J.-M., Couvillon, A.D., Bhatia, V.N., Perlman, D.H. and Muir, T.W. A phosphohistidine proteomics strategy based on elucidation of a unique gas-phase phosphopeptide fragmentation mechanism. *Journal of the American Chemical Society.* 2014, **136**(37), pp.12899-12911.
135. Kee, J.-M., Oslund, R.C., Perlman, D.H. and Muir, T.W. A pan-specific antibody for direct detection of protein histidine phosphorylation. *Nature Chemical Biology.* 2013, **9**(7), pp.416-421.
136. Lecroisey, A., Lascu, I., Bominaar, A., Veron, M. and Delepierre, M. Phosphorylation Mechanism of Nucleoside Diphosphate Kinase: 31P-Nuclear Magnetic Resonance Studies. *Biochemistry.* 1995, **34**(38), pp.12445-12450.
137. Gassner, M., Stehlik, D., Schrecker, O., Hengstenberg, W., Maurer, W. and Rüterjans, H. The phosphoenolpyruvate-dependent phosphotransferase system of Staphylococcus aureus. 2. 1H and 31P-nuclear-magnetic-resonance studies on the phosphocarrier protein HPr, phosphohistidines and phosphorylated HPr. *Eur J Biochem.* 1977, **75**(1), pp.287-296.
138. Himmel, S., Wolff, S., Becker, S., Lee, D. and Griesinger, C. Detection and Identification of Protein-Phosphorylation Sites in Histidines through HNP Correlation Patterns. *Angewandte Chemie International Edition.* 2010, **49**(47), pp.8971-8974.
139. Besant, P.G., Byrne, L., Thomas, G. and Attwood, P.V. A Chromatographic Method for the Preparative Separation of Phosphohistidines. *Analytical Biochemistry.* 1998, **258**(2), pp.372-375.
140. Makwana, M.V., Williamson, M.P., Jackson, R.F.W. and Muimo, R. Quantitation of phosphohistidine in proteins in a mammalian cell line by 31P NMR. *PLOS ONE.* 2022, **17**(9), p.e0273797.
141. Leijten, N.M., Heck, A.J.R. and Lemeer, S. Histidine phosphorylation in human cells; a needle or phantom in the haystack? *Nature Methods.* 2022, **19**(7), pp.827-828.
142. Sampadi, B., Pines, A., Munk, S., Mišovic, B., de Groot, A.J., van de Water, B., Olsen, J.V., Mullenders, L.H.F. and Vrieling, H. Quantitative phosphoproteomics to unravel the cellular response to chemical stressors with different modes of action. *Archives of Toxicology.* 2020, **94**(5), pp.1655-1671.

143. Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P. and Mann, M. Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell.* 2006, **127**(3), pp.635-648.
144. Reinders, J. and Sickmann, A. State-of-the-art in phosphoproteomics. *Proteomics* 2005, **5**(16), pp.4052-4061.
145. Ignatoski, K.M. Immunoprecipitation and western blotting of phosphotyrosine-containing proteins. *Methods in molecular biology (Clifton, N.J.).* 2001, **124**, pp.39-48.
146. Izaguirre, G., Aguirre, L., Hu, Y.-P., Lee, H.Y., Schlaepfer, D.D., Aneskievich, B.J. and Haimovich, B. The Cytoskeletal/Non-muscle Isoform of α-Actinin Is Phosphorylated on Its Actin-binding Domain by the Focal Adhesion Kinase. *Journal of Biological Chemistry.* 2001, **276**(31), pp.28676-28685.
147. Kee, J.-M., Villani, B., Carpenter, L.R. and Muir, T.W. Development of Stable Phosphohistidine Analogues. *Journal of the American Chemical Society.* 2010, **132**(41), pp.14327-14329.
148. Frackelton, A.R., Ross, A.H. and Eisen, H.N. Characterization and use of monoclonal antibodies for isolation of phosphotyrosyl proteins from retrovirus-transformed cells and growth factor-stimulated cells. *Molecular and Cellular Biology.* 1983, **3**(8), p.1343.
149. Lasker, M., Bui, C.D., Besant, P.G., Sugawara, K., Thai, P., Medzihradszky, G. and Turck, C.W. Protein histidine phosphorylation: increased stability of thiophosphohistidine. *Protein science : a publication of the Protein Society.* 1999, **8**(10), pp.2177-2185.
150. Allen, J.J., Li, M., Brinkworth, C.S., Paulson, J.L., Wang, D., Hübner, A., Chou, W.-H., Davis, R.J., Burlingame, A.L., Messing, R.O., Katayama, C.D., Hedrick, S.M. and Shokat, K.M. A semisynthetic epitope for kinase substrates. *Nature Methods.* 2007, **4**(6), pp.511-516.
151. Carlson, H.K., Plate, L., Price, M.S., Allen, J.J., Shokat, K.M. and Marletta, M.A. Use of a semisynthetic epitope to probe histidine kinase activity and regulation. *Analytical Biochemistry.* 2010, **397**(2), pp.139-143.
152. Schenkels, C., Erni, B. and Reymond, J.-L. Phosphofurylalanine, a stable analog of phosphohistidine. *Bioorganic & Medicinal Chemistry Letters.* 1999, **9**(10), pp.1443-1446.
153. Lilley, M., Mambwe, B., Jackson, R.F.W. and Muimo, R. 4-Phosphothiophen-2-yl alanine: a new 5-membered analogue of phosphotyrosine. *Chemical Communications.* 2014, **50**(66), pp.9343-9345.
154. McAllister, T.E., Nix, M.G. and Webb, M.E. Fmoc-chemistry of a stable phosphohistidine analogue. *Chemical Communications.* 2011, **47**(4), pp.1297-1299.
155. McAllister, T.E., Horner, K.A. and Webb, M.E. Evaluation of the Interaction between Phosphohistidine Analogues and Phosphotyrosine Binding Domains. *ChemBioChem.* 2014, **15**(8), pp.1088-1091.
156. Kee, J.-M., Oslund, R.C., Couvillon, A.D. and Muir, T.W. A Second-Generation Phosphohistidine Analog for Production of Phosphohistidine Antibodies. *Organic Letters.* 2015, **17**(2), pp.187-189.
157. Lilley, M., Mambwe, B., Thompson, M.J., Jackson, R.F.W. and Muimo, R. 4-Phosphopyrazol-2-yl alanine: a non-hydrolysable analogue of phosphohistidine. *Chemical Communications.* 2015, **51**(34), pp.7305-7308.
158. Makwana, M.V., dos Santos Souza, C., Pickup, B.T., Thompson, M.J., Lomada, S.K., Feng, Y., Wieland, T., Jackson, R.F.W. and Muimo, R. Chemical Tools for Studying Phosphohistidine: Generation of Selective τ-Phosphohistidine and π-Phosphohistidine Antibodies. *ChemBioChem.* 2023, **24**(16), p.e202300182.

159. McAllister, T.E. and Webb, M.E. Triazole phosphohistidine analogues compatible with the Fmoc-strategy. *Organic & Biomolecular Chemistry.* 2012, **10**(20), pp.4043-4049.
160. Khan, I. and Steeg, P.S. The relationship of NM23 (NME) metastasis suppressor histidine phosphorylation to its nucleoside diphosphate kinase, histidine protein kinase and motility suppression activities. *Oncotarget.* 2017, **9**(12).
161. Kiick, K.L., van Hest, J.C.M. and Tirrell, D.A. Expanding the Scope of Protein Biosynthesis by Altering the Methionyl-tRNA Synthetase Activity of a Bacterial Expression Host. *Angewandte Chemie International Edition.* 2000, **39**(12), pp.2148-2152.
162. Kiick, K.L., Saxon, E., Tirrell, D.A. and Bertozzi, C.R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proceedings of the National Academy of Sciences.* 2002, **99**(1), pp.19-24.
163. Link, A.J., Vink, M.K.S. and Tirrell, D.A. Presentation and Detection of Azide Functionality in Bacterial Cell Surface Proteins. *Journal of the American Chemical Society.* 2004, **126**(34), pp.10598-10602.
164. Dadová, J., Wu, K.-J., Isenegger, P.G., Errey, J.C., Bernardes, G.J.L., Chalker, J.M., Raich, L., Rovira, C. and Davis, B.G. Precise Probing of Residue Roles by Post-Translational β,γ-C,N Aza-Michael Mutagenesis in Enzyme Active Sites. *ACS central science.* 2017, **3**(11), pp.1168-1173.
165. Yang, A., Ha, S., Ahn, J., Kim, R., Kim, S., Lee, Y., Kim, J., Söll, D., Lee, H.-Y. and Park, H.-S. A chemical biology route to site-specific authentic protein modifications. *Science.* 2016, **354**(6312), p.623.
166. Morrison, P.M., Foley, P.J., Warriner, S.L. and Webb, M.E. Chemical generation and modification of peptides containing multiple dehydroalanines. *Chemical Communications.* 2015, **51**(70), pp.13470-13473.
167. Chalker, J.M., Gunnoo, S.B., Boutureira, O., Gerstberger, S.C., Fernández-González, M., Bernardes, G.J.L., Griffin, L., Hailu, H., Schofield, C.J. and Davis, B.G. Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chemical Science.* 2011, **2**(9), pp.1666-1676.
168. Bessmertnykh, A., Douaihy, C.M., Muniappan, S. and Guilard, R. Efficient Palladium-Catalyzed Synthesis of Aminopyridyl Phosphonates from Bromopyridines and Diethyl Phosphite. *Synthesis.* 2008, **2008**(10), pp.1575-1579.
169. Rogerson, D.T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaite, A., Hancock, S.M., Huguenin-Dezot, N., Muqit, M.M.K., Fry, A.M., Bayliss, R. and Chin, J.W. Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nature Chemical Biology.* 2015, **11**(7), pp.496-503.
170. Fukunaga, R. and Yokoyama, S. Structural insights into the first step of RNA-dependent cysteine biosynthesis in archaea. *Nature Structural & Molecular Biology.* 2007, **14**(4), pp.272-279.
171. Reddy, Kishorkumar M. and Mugesh, G. Application of dehydroalanine as a building block for the synthesis of selenocysteine-containing peptides. *RSC Advances.* 2019, **9**(1), pp.34-43.
172. Lukszo, J., Patterson, D., Albericio, F. and Kates, S.A. 3-(1-Piperidinyl)alanine formation during the preparation ofC-terminal cysteine peptides with the Fmoc/t-Bu strategy. *Letters in Peptide Science.* 1996, **3**(3), pp.157-166.
173. Attard, T.J., O'Brien-Simpson, N. and Reynolds, E.C. Synthesis of Phosphopeptides in the Fmoc Mode. *International Journal of Peptide Research and Therapeutics.* 2007, **13**(4), pp.447-468.

174. Bilbrough, T., Piemontese, E. and Seitz, O. Dissecting the role of protein phosphorylation: a chemical biology toolbox. *Chemical Society Reviews.* 2022, **51**(13), pp.5691-5730.
175. Wakamiya, T., Saruta, K., Yasuoka, J.-i. and Kusumoto, S. An Efficient Procedure for Solid-Phase Synthesis of Phosphopeptides by the Fmoc Strategy. *Chemistry Letters.* 1994, **23**(6), pp.1099-1102.
176. Harris, P., Williams, G., Shepherd, P. and Brimble, M. The Synthesis of Phosphopeptides Using Microwave-assisted Solid Phase Peptide Synthesis. *International Journal of Peptide Research and Therapeutics.* 2008, **14**, pp.387-392.
177. Behrendt, R., White, P. and Offer, J. Advances in Fmoc solid-phase peptide synthesis. *J Pept Sci.* 2016, **22**(1), pp.4-27.
178. Perich, J.W., Ede, N.J., Eagle, S. and Bray, A.M. Synthesis of phosphopeptides by the multipin method: Evaluation of coupling methods for the incorporation of Fmoc-Tyr(PO3Bzl,H)-OH, Fmoc-Ser(PO3Bzl,H)-OH and Fmoc-Thr(PO3Bzl,H)-OH. *Letters in Peptide Science.* 1999, **6**(2), pp.91-97.
179. Liu, Y., Patricelli, M.P. and Cravatt, B.F. Activity-based protein profiling: the serine hydrolases. *Proc Natl Acad Sci U S A.* 1999, **96**(26), pp.14694-14699.
180. d'Andrea, F.B. and Townsend, C.A. Late-Stage Conversion of Diphenylphosphonate to Fluorophosphonate Probes for the Investigation of Serine Hydrolases. *Cell Chemical Biology.* 2019, **26**(6), pp.878-884.e878.
181. Mukherjee, S. and Gupta, R.D. Organophosphorus Nerve Agents: Types, Toxicity, and Treatments. *Journal of Toxicology.* 2020, **2020**(1), p.3007984.
182. Chappell, W.P., Schur, N., Vogel, J.A., Sammis, G.M., Melvin, P.R. and Ball, N.D. Poison to Promise: The Resurgence of Organophosphorus Fluoride Chemistry. *Chem.* 2024, **10**(6), pp.1644-1654.
183. Trost, B.M., Machacek, M.R. and Faulk, B.D. Sequential Ru−Pd Catalysis: A Two-Catalyst One-Pot Protocol for the Synthesis of N- and O-Heterocycles. *Journal of the American Chemical Society.* 2006, **128**(20), pp.6745-6754.
184. Aechtner, T., Barry, D.A., David, E., Ghellamallah, C., Harvey, D.F., de la Houpliere, A., Knopp, M., Malaska, M.J., Pérez, D., Schärer, K.A., Siesel, B.A., Vollhardt, K.P.C. and Zitterbart, R. Cobalt-Mediated [2+2+2] Cycloadditions of Alkynes to Benzo--[b]furans and Benzo[b]thiophenes: A Potential Route toward Morphanoids. *Synthesis.* 2018, **50**(05), pp.1053-1089.
185. Péréz, D., Siesel, B.A., Malaska, M.J., David, E. and Vollhardt, K.P.C. Stereoselective Cobalt-Mediated [2+2+2]Cycloadditions to the Benzofuran Nucleus: A Novel Strategy Towards the Synthesis of Morphinoids. *Synlett.* 2000, **2000**(03), pp.306-310.
186. Kim, S., Kim, B. and In, J. Facile Deprotection of Bulky (Trialkylsilyl)acetylenes with Silver Fluoride. *Synthesis.* 2009, **2009**(12), pp.1963-1968.
187. Chalifoux, W.A. and Tykwinski, R.R. Synthesis of polyynes to model the sp-carbon allotrope carbyne. *Nature Chemistry.* 2010, **2**(11), pp.967-971.
188. Dutremez, S.G., Guerin, C., Henner, B.J.L. and Tomberli, V. DIETHYL ETHYNYLPHOSPHONATE: A VERSATILE SYNTHON FOR THE PREPARATION OF 1-ALKYNYL- AND 1,3-BUTADIY NE-1,4-DIYLPHOSPHONATES. *Phosphorus, Sulfur, and Silicon and the Related Elements.* 2000, **160**(1), pp.251-269.
189. Dutra, J.K., Foley, T.L., Huang, Z., Fisher, E.L., Lachapelle, E.A., Mahapatra, S., Ogilvie, K., Butler, T.W., Bellenger, J., Devraj Majmudar, J. and am Ende, C.W. Fluorophosphonates on-Demand: A General and Simplified Approach toward Fluorophosphonate Synthesis. *ChemBioChem.* 2021, **22**(10), pp.1769-1774.

190. Lacasse, M.-C., Poulard, C. and Charette, A.B. Iodomethylzinc Phosphates: Powerful Reagents for the Cyclopropanation of Alkenes. *Journal of the American Chemical Society.* 2005, **127**(36), pp.12440-12441.
191. Wanat, P., Walczak, S., Wojtczak, B.A., Nowakowska, M., Jemielity, J. and Kowalska, J. Ethynyl, 2-Propynyl, and 3-Butynyl C-Phosphonate Analogues of Nucleoside Di- and Triphosphates: Synthesis and Reactivity in CuAAC. *Organic Letters.* 2015, **17**(12), pp.3062-3065.
192. Huang, H., Denne, J., Yang, C.-H., Wang, H. and Kang, J.Y. Direct Aryloxylation/Alkyloxylation of Dialkyl Phosphonates for the Synthesis of Mixed Phosphonates. *Angewandte Chemie International Edition.* 2018, **57**(22), pp.6624-6628.
193. Wang, C., Abegg, D., Dwyer, B.G. and Adibekian, A. Discovery and Evaluation of New Activity-Based Probes for Serine Hydrolases. *ChemBioChem.* 2019, **20**(17), pp.2212-2216.
194. Wozniak, J.M., Li, W. and Parker, C.G. Chemical proteomic mapping of reversible small molecule binding sites in native systems. *Trends in Pharmacological Sciences.* 2024, **45**(11), pp.969-981.
195. Yang, T., Liu, Z. and Li, X.D. Developing diazirine-based chemical probes to identify histone modification 'readers' and 'erasers'. *Chemical Science.* 2015, **6**(2), pp.1011-1017.
196. Lin, J., Bao, X. and Li, X.D. A tri-functional amino acid enables mapping of binding sites for posttranslational-modification-mediated protein-protein interactions. *Molecular Cell.* 2021, **81**(12), pp.2669-2681.e2669.
197. Leslie, J., Millar, B.J., Del Carpio Pons, A., Burgoyne, R.A., Frost, J.D., Barksby, B.S., Luli, S., Scott, J., Simpson, A.J., Gauldie, J., Murray, L.A., Finch, D.K., Carruthers, A.M., Ferguson, J., Sleeman, M.A., Rider, D., Howarth, R., Fox, C., Oakley, F., Fisher, A.J., Mann, D.A. and Borthwick, L.A. FPR-1 is an important regulator of neutrophil recruitment and a tissue-specific driver of pulmonary fibrosis. *JCI Insight.* 2020, **5**(4).
198. Chen, G., Wang, X., Liao, Q., Ge, Y., Jiao, H., Chen, Q., Liu, Y., Lyu, W., Zhu, L., van Zundert, G.C.P., Robertson, M.J., Skiniotis, G., Du, Y., Hu, H. and Ye, R.D. Structural basis for recognition of N-formyl peptides as pathogen-associated molecular patterns. *Nature Communications.* 2022, **13**(1), p.5232.
199. Weiß, E. and Kretschmer, D. Formyl-Peptide Receptors in Infection, Inflammation, and Cancer. *Trends Immunol.* 2018, **39**(10), pp.815-829.
200. Field, D.H., White, J.S., Warriner, S.L. and Wright, M.H. A fluorescent photoaffinity probe for formyl peptide receptor 1 labelling in living cells. *RSC Chem Biol.* 2023, **4**(3), pp.216-222.
201. Halloran, M.W. and Lumb, J.-P. Recent Applications of Diazirines in Chemical Proteomics. *Chemistry – A European Journal.* 2019, **25**(19), pp.4885-4898.
202. Porta, E.O.J. and Steel, P.G. Activity-based protein profiling: A graphical review. *Curr Res Pharmacol Drug Discov.* 2023, **5**, p.100164.
203. Mackinnon, A.L. and Taunton, J. Target Identification by Diazirine Photo-Cross-linking and Click Chemistry. *Curr Protoc Chem Biol.* 2009, **1**, pp.55-73.
204. Müller, M.Q., Dreiocker, F., Ihling, C.H., Schäfer, M. and Sinz, A. Cleavable Cross-Linker for Protein Structure Analysis: Reliable Identification of Cross-Linking Products by Tandem MS. *Analytical Chemistry.* 2010, **82**(16), pp.6958-6968.
205. Korovesis, D., Gaspar, V.P., Beard, H.A., Chen, S., Zahédi, R.P. and Verhelst, S.H.L. Mapping Peptide–Protein Interactions by Amine-Reactive Cleavable Photoaffinity Reagents. *ACS Omega.* 2023, **8**(28), pp.25487-25495.

206. Lemmon, M.A., Ladbury, J.E., Mandiyan, V., Zhou, M. and Schlessinger, J. Independent binding of peptide ligands to the SH2 and SH3 domains of Grb2. *Journal of Biological Chemistry.* 1994, **269**(50), pp.31653-31658.
207. Chook, Y.M., Gish, G.D., Kay, C.M., Pai, E.F. and Pawson, T. The Grb2-mSos1 Complex Binds Phosphopeptides with Higher Affinity than Grb2*. *Journal of Biological Chemistry.* 1996, **271**(48), pp.30472-30478.
208. Gale, N.W., Kaplan, S., Lowenstein, E.J., Schlessinger, J. and Bar-Sagi, D. Grb2 mediates the EGF-dependent activation of guanine nucleotide exchange on Ras. *Nature.* 1993, **363**(6424), pp.88-92.
209. Zhang, W. and Liu, H.T. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Research.* 2002, **12**(1), pp.9-18.
210. Ge, Y., Jin, J., Li, J., Ye, M. and Jin, X. The roles of G3BP1 in human diseases (review). *Gene.* 2022, **821**, p.146294.
211. Batzer, A.G., Rotin, D., Ureña, J.M., Skolnik, E.Y. and Schlessinger, J. Hierarchy of binding sites for Grb2 and Shc on the epidermal growth factor receptor. *Mol Cell Biol.* 1994, **14**(8), pp.5192-5201.
212. Ward, C.W., Gough, K.H., Rashke, M., Wan, S.S., Tribbick, G. and Wang, J.-x. Systematic Mapping of Potential Binding Sites for Shc and Grb2 SH2 Domains on Insulin Receptor Substrate-1 and the Receptors for Insulin, Epidermal Growth Factor, Platelet-derived Growth Factor, and Fibroblast Growth Factor (∗). *Journal of Biological Chemistry.* 1996, **271**(10), pp.5603-5609.
213. Jadwin, J.A., Curran, T.G., Lafontaine, A.T., White, F.M. and Mayer, B.J. Src homology 2 domains enhance tyrosine phosphorylation in vivo by protecting binding sites in their target proteins from dephosphorylation. *Journal of Biological Chemistry.* 2018, **293**(2), pp.623-637.
214. Klumpp, S. and Krieglstein, J. Phosphorylation and dephosphorylation of histidine residues in proteins. *European Journal of Biochemistry.* 2002, **269**(4), pp.1067-1071.
215. McNemar, C., Snow, M.E., Windsor, W.T., Prongay, A., Mui, P., Zhang, R., Durkin, J., Le, H.V. and Weber, P.C. Thermodynamic and Structural Analysis of Phosphotyrosine Polypeptide Binding to Grb2-SH2. *Biochemistry.* 1997, **36**(33), pp.10006-10014.
216. Pang, K., Li, M.L., Hao, L., Shi, Z.D., Feng, H., Chen, B., Ma, Y.Y., Xu, H., Pan, D., Chen, Z.S. and Han, C.H. ERH Gene and Its Role in Cancer Cells. *Front Oncol.* 2022, **12**, p.900496.
217. Jin, T., Guo, F., Serebriiskii, I.G., Howard, A. and Zhang, Y.-Z. A 1.55 Å resolution X-ray crystal structure of HEF2/ERH and insights into its transcriptional and cell-cycle interaction networks. *Proteins: Structure, Function, and Bioinformatics.* 2007, **68**(2), pp.427-437.
218. Gelsthorpe, M., Pulumati, M., McCallum, C., Dang-Vu, K. and Tsubota, S.I. The putative cell cycle gene, enhancer of rudimentary, encodes a highly conserved protein found in plants and animals. *Gene.* 1997, **186**(2), pp.189-195.
219. Weng, M.-T. and Luo, J. The enigmatic ERH protein: its role in cell cycle, RNA splicing and cancer. *Protein & Cell.* 2013, **4**(11), pp.807-812.
220. Weng, M.T., Lee, J.H., Wei, S.C., Li, Q., Shahamatdar, S., Hsu, D., Schetter, A.J., Swatkoski, S., Mannan, P., Garfield, S., Gucek, M., Kim, M.K., Annunziata, C.M., Creighton, C.J., Emanuele, M.J., Harris, C.C., Sheu, J.C., Giaccone, G. and Luo, J. Evolutionarily conserved protein ERH controls CENP-E mRNA splicing and is required for the survival of KRAS mutant cancer cells. *Proc Natl Acad Sci U S A.* 2012, **109**(52), pp.E3659-3667.

221. Parker, F., Maurier, F., Delumeau, I., Duchesne, M., Faucher, D., Debussche, L., Dugue, A., Schweighoffer, F. and Tocque, B. A Ras-GTPase-Activating Protein SH3-Domain-Binding Protein. *Molecular and Cellular Biology.* 1996, **16**(6), pp.2561-2569.
222. Simanshu, D.K., Nissley, D.V. and McCormick, F. RAS Proteins and Their Regulators in Human Disease. *Cell.* 2017, **170**(1), pp.17-33.
223. Wang, Y., Fu, D., Chen, Y., Su, J., Wang, Y., Li, X., Zhai, W., Niu, Y., Yue, D. and Geng, H. G3BP1 promotes tumor progression and metastasis through IL-6/G3BP1/STAT3 signaling axis in renal cell carcinomas. *Cell Death & Disease.* 2018, **9**(5), p.501.
224. Romano, S., D'Angelillo, A. and Romano, M.F. Pleiotropic roles in cancer biology for multifaceted proteins FKBPs. *Biochimica et Biophysica Acta (BBA) - General Subjects.* 2015, **1850**(10), pp.2061-2068.
225. Somarelli, J.A., Lee, S.Y., Skolnick, J. and Herrera, R.J. Structure-based classification of 45 FK506-binding proteins. *Proteins: Structure, Function, and Bioinformatics.* 2008, **72**(1), pp.197-208.
226. Mathea, S., Li, S., Schierhorn, A., Jahreis, G. and Schiene-Fischer, C. Suppression of EGFR Autophosphorylation by FKBP12. *Biochemistry.* 2011, **50**(50), pp.10844-10850.
227. Lopez-Ilasaca, M., Schiene, C., Küllertz, G., Tradler, T., Fischer, G. and Wetzker, R. Effects of FK506-binding Protein 12 and FK506 on Autophosphorylation of Epidermal Growth Factor Receptor *. *Journal of Biological Chemistry.* 1998, **273**(16), pp.9430-9434.
228. Morimatsu, M., Takagi, H., Ota, K.G., Iwamoto, R., Yanagida, T. and Sako, Y. Multiple-state reactions between the epidermal growth factor receptor and Grb2 as observed by using single-molecule analysis. *Proc Natl Acad Sci U S A.* 2007, **104**(46), pp.18013-18018.
229. Astley, H.M., Parsley, K., Aubry, S., Chastain, C.J., Burnell, J.N., Webb, M.E. and Hibberd, J.M. The pyruvate, orthophosphate dikinase regulatory proteins of Arabidopsis are both bifunctional and interact with the catalytic and nucleotide-binding domains of pyruvate, orthophosphate dikinase. *Plant J.* 2011, **68**(6), pp.1070-1080.
230. Chastain, C.J. and Chollet, R. Regulation of pyruvate, orthophosphate dikinase by ADP-/Pi-dependent reversible phosphorylation in C3 and C4 plants. *Plant Physiology and Biochemistry.* 2003, **41**(6), pp.523-532.
231. Chastain, C.J., Fries, J.P., Vogel, J.A., Randklev, C.L., Vossen, A.P., Dittmer, S.K., Watkins, E.E., Fiedler, L.J., Wacker, S.A., Meinhover, K.C., Sarath, G. and Chollet, R. Pyruvate,orthophosphate dikinase in leaves and chloroplasts of C(3) plants undergoes light-/dark-induced reversible phosphorylation. *Plant Physiol.* 2002, **128**(4), pp.1368-1378.
232. Om, K., Arias, N.N., Jambor, C.C., MacGregor, A., Rezachek, A.N., Haugrud, C., Kunz, H.H., Wang, Z., Huang, P., Zhang, Q., Rosnow, J., Brutnell, T.P., Cousins, A.B. and Chastain, C.J. Pyruvate, phosphate dikinase regulatory protein impacts light response of C4 photosynthesis in Setaria viridis. *Plant Physiol.* 2022, **190**(2), pp.1117-1133.
233. Burnell, J.N. and Hatch, M.D. Regulation of C4 photosynthesis: Identification of a catalytically important histidine residue and its role in the regulation of pyruvate,Pi dikinase. *Archives of Biochemistry and Biophysics.* 1984, **231**(1), pp.175-182.
234. Chastain, C.J., Baird, L.M., Walker, M.T., Bergman, C.C., Novbatova, G.T., Mamani-Quispe, C.S. and Burnell, J.N. Maize leaf PPDK regulatory protein isoform-2 is specific to bundle sheath chloroplasts and paradoxically lacks a Pi-dependent PPDK activation activity. *Journal of Experimental Botany.* 2017, **69**(5), pp.1171-1181.

235. Jiang, L., Chen, Y.-b., Zheng, J., Chen, Z., Liu, Y., Tao, Y., Wu, W., Chen, Z. and Wang, B.-c. Structural Basis of Reversible Phosphorylation by Maize Pyruvate Orthophosphate Dikinase Regulatory Protein1 *Plant Physiology.* 2015, **170**(2), pp.732-741.
236. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P. and Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021, **596**(7873), pp.583-589.
237. Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J., Kovalevskiy, O., Tunyasuvunakool, K., Laydon, A., Žídek, A., Tomlinson, H., Hariharan, D., Abrahamson, J., Green, T., Jumper, J., Birney, E., Steinegger, M., Hassabis, D. and Velankar, S. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research.* 2024, **52**(D1), pp.D368-D375.
238. Fischbach, M.A. and Walsh, C.T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chemical Reviews.* 2006, **106**(8), pp.3468-3496.
239. Zhou, Y., Lin, X., Xu, C., Shen, Y., Wang, S.P., Liao, H., Li, L., Deng, H. and Lin, H.W. Investigation of Penicillin Binding Protein (PBP)-like Peptide Cyclase and Hydrolase in Surugamide Non-ribosomal Peptide Biosynthesis. *Cell Chem Biol.* 2019, **26**(5), pp.737-744 e734.
240. Thankachan, D., Fazal, A., Francis, D., Song, L., Webb, M.E. and Seipke, R.F. A trans-Acting Cyclase Offloading Strategy for Nonribosomal Peptide Synthetases. *ACS Chemical Biology.* 2019, **14**(5), pp.845-849.
241. Marahiel, M.A., Stachelhaus, T. and Mootz, H.D. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chemical Reviews.* 1997, **97**(7), pp.2651-2674.
242. Trauger, J.W., Kohli, R.M., Mootz, H.D., Marahiel, M.A. and Walsh, C.T. Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature.* 2000, **407**(6801), pp.215-218.
243. Kotaka, M., Kong, R., Qureshi, I., Ho, Q.S., Sun, H., Liew, C.W., Goh, L.P., Cheung, P., Mu, Y., Lescar, J. and Liang, Z.X. Structure and catalytic mechanism of the thioesterase CalE7 in enediyne biosynthesis. *J Biol Chem.* 2009, **284**(23), pp.15739-15749.
244. Kuranaga, T., Matsuda, K., Sano, A., Kobayashi, M., Ninomiya, A., Takada, K., Matsunaga, S. and Wakimoto, T. Total Synthesis of the Nonribosomal Peptide Surugamide B and Identification of a New Offloading Cyclase Family. *Angewandte Chemie International Edition.* 2018.
245. Matsuda, K., Zhai, R., Mori, T., Kobayashi, M., Sano, A., Abe, I. and Wakimoto, T. Heterochiral coupling in non-ribosomal peptide macrolactamization. *Nature Catalysis.* 2020, **3**(6), pp.507-515.
246. Kohli, R.M., Takagi, J. and Walsh, C.T. The thioesterase domain from a nonribosomal peptide synthetase as a cyclization catalyst for integrin binding peptides. *Proceedings of the National Academy of Sciences.* 2002, **99**(3), pp.1247-1252.
247. Fazal, A., Wheeler, J., Webb, M.E. and Seipke, R.F. The N-terminal substrate specificity of the SurE peptide cyclase. *Organic & Biomolecular Chemistry.* 2022, **20**(36), pp.7232-7235.

248. Padial, N.M., Quartapelle Procopio, E., Montoro, C., López, E., Oltra, J.E., Colombo, V., Maspero, A., Masciocchi, N., Galli, S., Senkovska, I., Kaskel, S., Barea, E. and Navarro, J.A.R. Highly Hydrophobic Isoreticular Porous Metal–Organic Frameworks for the Capture of Harmful Volatile Organic Compounds. *Angewandte Chemie International Edition.* 2013, **52**(32), pp.8290-8294.
249. Mezei, G. and Raptis, R.G. Effect of pyrazole-substitution on the structure and nuclearity of Cu(II)-pyrazolato complexes. *Inorganica Chimica Acta.* 2004, **357**(11), pp.3279-3288.
250. Zhang, Z., Lieu, T., Wu, C.-H., Wang, X., Wu, J.I., Daugulis, O. and Miljanić, O.Š. Solvation-dependent switching of solid-state luminescence of a fluorinated aromatic tetrapyrazole. *Chemical Communications.* 2019, **55**(63), pp.9387-9390.
251. Elguero, J., Jaramillo, C. and Pardo, C. Palladium(0)-Catalyzed Preparation of 4-Arylpyrazoles. *Synthesis.* 1997, **1997**(05), pp.563-566.
252. Chalker, J.M., Lercher, L., Rose, N.R., Schofield, C.J. and Davis, B.G. Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post-Translational Modifications. *Angewandte Chemie International Edition.* 2012, **51**(8), pp.1835-1839.
253. Acheson, R.M.A., Peter J.; Murray, Jane R. Addition Reactions of Heterocyclic Compounds. Part 82. The Synthesis and Reactions of Diethyl Ethynylphosphonate and Tetraethyl Ethynyldiphosphonate. *Journal of Chemical Research, Miniprint.* 1986, (10), p.3001.
254. Perez, V., Fadel, A. and Rabasso, N. Synthesis of N-Sulfonyl Ynamido-Phosphonates: Valuable Partners for Cycloadditions. *Synthesis.* 2017, **49**(17), pp.4035-4044.
255. Kambe, T., Correia, B.E., Niphakis, M.J. and Cravatt, B.F. Mapping the Protein Interaction Landscape for Fully Functionalized Small-Molecule Probes in Human Cells. *Journal of the American Chemical Society.* 2014, **136**(30), pp.10777-10782.
256. Wright, M.H., Fetzer, C. and Sieber, S.A. Chemical Probes Unravel an Antimicrobial Defense Response Triggered by Binding of the Human Opioid Dynorphin to a Bacterial Sensor Kinase. *Journal of the American Chemical Society.* 2017, **139**(17), pp.6152-6159.