Imperial College London

THE FRANCIS CRICK INSTITUTE

PhotoRaPID: Development of a trillion-member discovery platform to identify *de novo* light-responsive cyclic peptides

Thomas Benjamin Jackson

BSc. Hons, MRes

Under the supervision of:

Dr Louise J. Walport,

Prof. Matthew J. Fuchter & Prof. Edward W. Tate

In partial fulfilment of the requirements for the degree of:

Doctor of Philosophy

Oct 2019 – Oct 2023

Department of Chemistry, Imperial College London & The Francis Crick Institute

To my parents, who have supported me endlessly.

“How do you eat an elephant…one bite at a time. You know it’s the same for a PhD!”

- Mum and Dad

DECLARATION

This dissertation is the result of my own work, conducted under the supervision of Dr Louise J. Walport[†‡], Prof. Matthew J. Fuchter[†], and Prof. Edward W. Tate[†] across both the Department of Chemistry at Imperial College London and the Francis Crick Institute. It includes nothing, which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification.

In accordance with Imperial College London guidelines, this thesis has been completed.

Date: Friday 31 st October 2023.

Thomas Benjamin Jackson BSc, MRes.

London.

† - Department of Chemistry

Molecular Sciences Research Hub (MSRH),

Imperial College London,

White City Campus,

Wood Lane, London

W12 0BZ, UK.

‡ - Francis Crick Institute,

4SW - 1 Midland Road, London,

NW1 1AT, UK.

COPYRIGHT

SUMMARY/ ABSTRACT

PhotoRaPID: Development of a trillion-member discovery platform to identify *de novo* light-responsive cyclic peptides

Cyclic peptides are an emerging class of exciting chemical tools, bridging the gap between small molecule therapeutics and larger biologics. They have recently gained traction probing classically undruggable protein targets of interest. Many powerful methods to identify novel cyclic peptides have been developed, including mRNA display-based RaPID (Random Non-standard Peptide Integrated Discovery). Such platforms enable the screening of >1 trillion-member cyclic peptide libraries incorporating non-proteogenic amino acids, against any protein of interest.

Despite the development of cyclic peptides as novel tools to answer questions in Chemical Biology, the fine-tuned spatiotemporal druggability of complex biological pathways remains elusive. Light is an incredibly powerful non-invasive stimulus to introduce temporal control within a system. Photoswitches, which harness light as a stimulus, are small molecules that can undergo isomerisation when irradiated at a specific wavelength of light. This project focuses on the integration of light controllable 'photoswitches' within every macrocycle of the trillion-member library, via genetic code re-programming, inducing global conformational changes within the macrocyclic peptides structure to give switchable binding.

As a proof-of-concept, methodology for the identification of photoswitchable macrocyclic peptides has been developed against human Protein Arginine Deiminase II. Identified peptides show nanomolar $_{Z}$KDs and up to 29-fold differential binding and inhibition of PADI2 between each of the two light-responsive (*E*/*Z*) macrocyclic isomers. Further work to streamline the methodology was also undertaken. Preliminary data suggests a general correlation between differential sequence enrichment of the two light-responsive isomers binding to PADI2 during the selection and their respective binding affinities determined via SPR. This will potentially streamline the transition from light-responsive candidate peptides to validated hits, a bottleneck of high-throughput screening platforms.

In summary, this work has developed a methodology to allow for the robust identification of light-responsive photoswitchable macrocyclic peptides against any protein target of interest and offers the potential to interrogate complex biological systems in novel and exciting ways.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my three PhD supervisors, Louise, Matt, and Ed. Their continuous support and excellent guidance throughout these last four years have made me not only a better scientist but also a more well-rounded individual. I could not have asked for better mentors. Thank you for also choosing me to pursue such an exciting multidisciplinary research project, I have thoroughly enjoyed my time and learnt so much. I would like to thank Louise especially; it is a momentous occasion for us both as I will be your first PhD student to finish. Your attention and mentorship have been especially valuable to me as I have navigated this PhD. You have read this thesis, chapter by chapter, and have done so with such dedication and attention to detail. I cannot thank you enough.

To the Walport, Fuchter and Tate groups, I also thank you for welcoming me with open arms, despite not working with most of you day-to-day. You are all amazing people and have broadened my horizons both personally and professionally, I have made some friends for life. To the Walports in particular, whom I have worked closely with for the last three years, I could not have asked for a better group of people to support me through this PhD journey. You have all made the lab such a lovely place to work, I hope I have given to you, half as much, as you have given to me.

Now to the shoutouts, firstly to Tessa, Ewen, Cat and Jemima, all who battled through each chapter's first draft and have taught me so much throughout this PhD journey. You all put so much effort into making this the thesis it is today, I cannot thank you enough. Especially Tessa, who ended up with more chapters than she bargained for, who read them so quickly and annotated them so thoroughly, each time without complaint. To Ana, who has been the one there throughout it all. I could not ask for a better friend, and I hope my advice has been even half as valuable as yours has been to me. To Jack, who has been so supportive and has been cheering me on whilst we finish our PhDs together. To Maggie, who I started this PhD journey with back in White City and has been continuously supportive, despite me abandoning you to move to the Crick. To Vincent, who is always ready for a much-needed tea break and always has an interesting new story to tell. Also, Issy for letting me into the fridge, Jack for your nonstop jokes and especially Cecilie, for putting up with the explosion of thesis material always making its way onto your side of the desk.

Thank you to my family and friends for supporting me, not only throughout this PhD but also for the growth I've had over these amazing last four years. Finally, to my parents who have always been there. I hope I have made you proud.

CONTENTS

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

A (Ala)	**Alanine**
AA	Amino acid(s)
AAPF	Arylazopyrazole-modified phenylalanine
ABT	Amino derived benzyl thioester
AcOH	Acetic acid
AMP	Adenosine monophosphate
APS	Ammonium persulfate
Aq.	Aqueous
Aq. sat. $NaHCO_3$	Aqueous saturated sodium bicarbonate solution
ATP	Adenosine triphosphate
BirA	*E.coli* biotin ligase
BPB	Bromophenol blue
BSA	Bovine serum albumin
C (Cys)	Cysteine
CBT	Chloro benzyl mercaptan ester
CDI	1-Ethyl-3-(3- dimethylaminopropyl)carbodimide hydrochloride
ClAc	Chloroacetyl
CME	Cyanomethyl ester
CT	Cycle threshold
D (Asp)	Aspartic acid
DBE	Dinitro benzyl ester
DCM	Dichloromethane
DIPEA	N,N'-diisopropylethylamine
DMAP	Dimethylaminopyridine
DMF	Dimethylformamide

DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
E (Glu)	Glutamic acid
E.coli	*Escherichia coli*
EDC.HCl	1-ethyl-3-(3-dimethylaminopropyl)carbodimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
EF-Tu	Elongation factor thermos unstable
Eq.	Equivalents
EtOH	Ethanol
EtOH	Ethanol
F (Phe)	Phenylalanine
f.c.	Final concentration
FA	Formic acid
FAM	4-carboxy fluorescein
FIT	Flexible in vitro Translation system
FMet tRNA	Formylated initiator methionine tRNA
Fx	Flexizyme
G (Gly)	Glycine
H (His)	Histidine
H_2O	Water
HPLC	High-performance liquid chromatography
Hypoth.	Hypothesis
I (Iso)	Isoleucine
IC_{50}	The Inhibitory concentration at which 50% of the enzyme in inhibited.

IMAC	Immobilised metal affinity chromatography
IniMet tRNA	Initiator methionine tRNA
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
IrrC	Irradiation condition
IVT	In vitro translation
IVTT	In vitro transcription and translation
K (Lys)	Lysine
L (Leu)	Leucine
LC	Liquid chromatography
LiCl	Lithium chloride
$LiOH.H_2O$	Lithium hydroxide hydrate
M (Met)	Methionine
MALDI	Matrix-assisted laser desorption/ ionization
Max theo.	Maximum theoretical yield
MeCN	Acetonitrile
MiHx	Microhelix
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
N (Asn)	Asparagine
NaOAc	Sodium acetate
NGS	Next generation sequencing
ns-AA	Non-standard amino acid
ns-AA	Non-standard amino acid(s)
NTP	Nucleoside triphosphate
ORF	Open reading Frame
P (Pro)	Proline

PADI	Peptidyl arginine deiminase
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PBST	Phosphate-buffered saline plus 0.1% tween
PCR	Polymerase chain reaction
PROTAC	Proteolysis targeting chimera
PSS	Photostationary state
PURE system	Protein synthesis using recombinant elements system
PYBOP	Benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate
Q (Gln)	Glutamine
qt.PCR	Real-time quantitative polymerase chain reaction
R (Arg)	Arginine
R1	Round 1
RaPID	Random non-standard peptides integrated discovery
RNA	Ribonucleuic acid
rpm	Revolutions per minute
rt	Room temperature
RT	Reverse transcriptase
rt.	Retention time
RU	Response unit
Sat.	Saturated
S (Ser)	Serine
SDS	Sodium dodecyl sulfate
SEC	Size exclusion column chromatography
Sn	Tin
SPR	Surface plasmon resonance

sssDNA	Sheered salmon sperm deoxyribonucleic acid
Strep-negative	The negative streptavidin counter selection.
T (Thr)	Threonine
TAE	Tris base, acetic acid, EDTA
TBE	Tris, borate, EDTA
TCEP	Tris(2-carboxyethyl)phosphine
TEA	Triethylamine
TEMED	Tetramethylethylenediamine
TES	Top enriched sequences
TES-500	Top 500 enriched sequences from the NGS data
TFA	Trifluoroacetic acid
tRNA	Transfer ribonucleic acid
TUS	Top unique sequences
Tween	Polysorbate 20
T*Z/E*-500	Top 500 ranked peptides with the highest NGS *Z/E* scores
UHPLC	Ultra-high-performance liquid chromatography
V (Val)	Valine
W (Trp)	Tryptophan
Y (Tyr)	Tyrosine
Z/E Score	Fold change between the enrichment of a peptide between each isomer.
Z/N Score	Fold change between the enrichment of a peptide (Z-positive) vs negative beads
Zn	Zinc
α-CHCA	A-cyano-4-hydroxycinnamic acid

1 Introduction

1.1 Manipulating systems with light:

Light is a powerful non-invasive stimulus to modulate biological systems. It is widely harnessed throughout the natural world and is essential for many processes including vision and photosynthesis.[1,2] Many chemical systems that are influenced by light can be defined as photochromic. Photochromism is the reversible transformation between two forms of a chemical species, caused by the absorbance of light, where the two forms have different absorption spectra.[3] Applications of photochromic systems include optogenetics (genetic manipulation via optical control) and photopharmacology (targeted therapeutic action using light).[4,5] Within photopharmacology applications, the light-responsive chemical compounds are largely classified as either photochromic molecular photoswitches or non-photochromic light-responsive photocages.[6]

Molecular photoswitches are compounds that can exist in two or more isomeric forms and are interconvertible upon the absorption of photons (photochromism).[7] Photoswitches can be further characterised into two broad categories relating to the methods of interconversion between each isomeric form. P-type photoswitches have two stable isomer states and can only switch back to the other state when exposed to light of a different wavelength.[8] In contrast, T-type photoswitches are a type of photoswitch that return to their thermodynamically stable isomer state over a given period.[8] This is termed as the thermal relaxation profile of the photoswitch (Figure 1).[7] The rate of thermal relaxation is determined by the physicochemical properties of the photoswitch and environment.[7] Upon light irradiation, this equilibrium is shifted towards either a single isomer population or mixed population depending on how distinct the absorption profile of each isomer of the molecular photoswitch is. In cases where the absorption profile of both isomers overlap at a given wavelength, irradiation would result in the interconversion of the isomers between each other until a steady-state mixed population is reached.

Molecular photoswitches differ primarily from non-photochromic caged systems as photo-decaging is the irreversible switching of a chemical compound to a different state, typically from pro-drug to active compound following light irradiation. Some applications combine both approaches within a single system to impart multi-layered light-responsive control.[7,9,10]

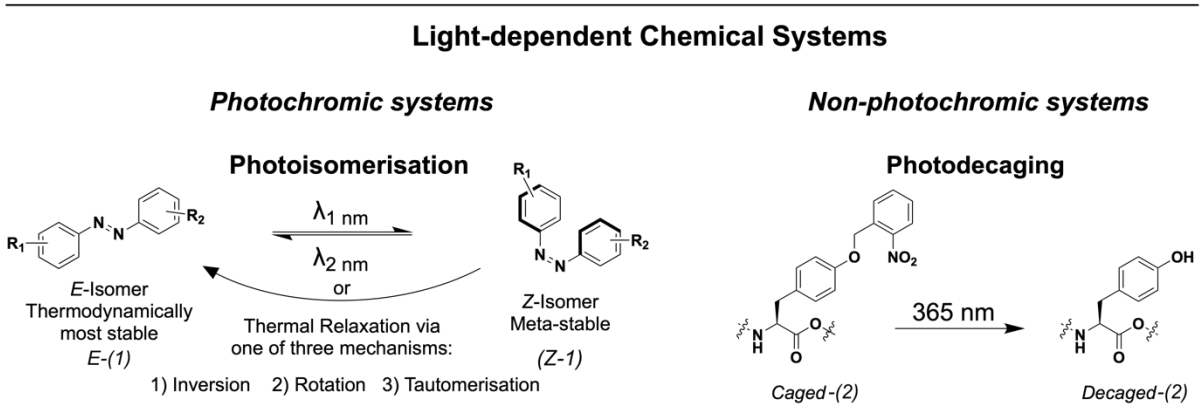


Figure 1: Photochromic and non-photochromic systems: photoswitches and photocages. Photoswitches (1) undergo reversible photoisomerisation and thermal reversion via one of three main mechanisms depending upon the functional groups decorating the photoswitch (mechanisms available at Appendix 1). Photocages (2) undergo photodecaging which is irreversible.

1.2 Photopharmacology:

Photopharmacology is a rapidly emerging field where chemical ligands (therapeutics) are developed to be activated or deactivated by light within a biological setting. Conformational changes are induced within the chemical structure of the ligand, generating differences in biological activity between the light-responsive states. This enables precise, spatial, and temporal control over a given biological target, in a non-invasive manner.[5] Several classes of molecular photoswitches are found within the literature for use as photopharmacology agents. These include: azobenzenes, stilbenes, spiropyrans, dithienylethenes, thiophenefulgides and hemithioindigos (Figure 2).[11–14]

All classes display distinct photochemical properties and were largely discovered through serendipity, typically as dyes.[15] The azo-based class of molecular photoswitches remains the most prominent class for research and applications.[15] Today, derivatives are synthesised with specific photochemical properties in mind, although predicting these structure-related properties remains challenging.[7] To describe a molecular photoswitch for photopharmacology with idealised characteristics, it would have a combination of three key photochemical properties: (1) Red-shifted switching, (2) Quantitative photoconversion between the isomers of the photoswitch and (3) A biologically relevant thermal relaxation.[15]

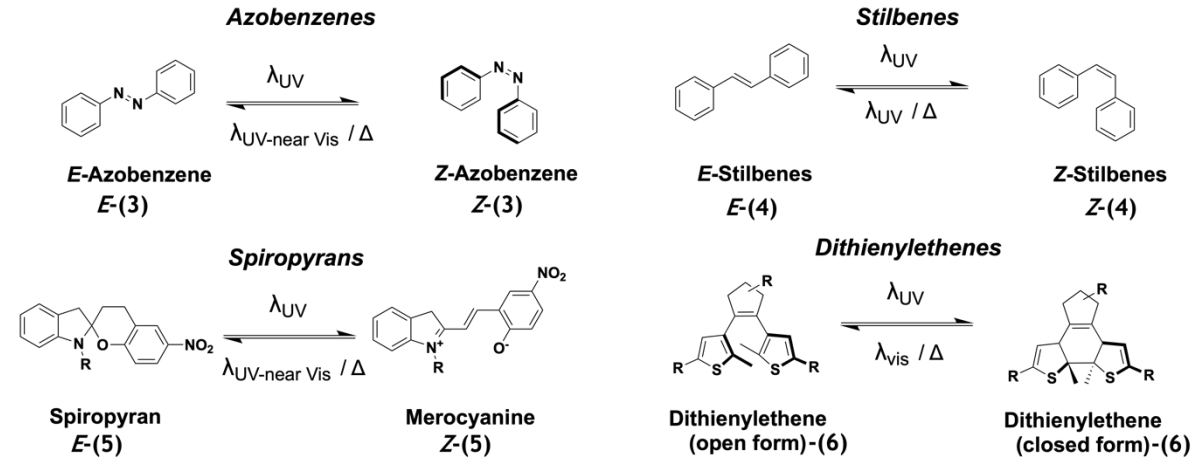


Figure 2: Representative molecular photoswitches used in photopharmacology applications. Four of the major classes are azobenzene (3), stilbene (4), spiropyran (5) and dithienylethene (6).

(1) Red-shifted switching is sought after in photopharmacology, as longer-wavelength switching enables greater tissue depth penetration within biological applications. Ultraviolet (UV) light (100-400 nm) has minimal tissue penetration (<1 mm), contrasting with red or near visible light (400–700 nm), which is much greater (1-5mm).[16] This is because the skin consists of a range of chromophores, with absorption coefficients largely in the UV region. As such, irradiating light-responsive agents at any significant tissue depth becomes very challenging as scattering increases proportionally.[17–19] As such, the development of molecular photoswitches with red-shifted switching enables access to a wider range of photopharmacology applications. The use of red-to-visible light also mitigates the risk of UV-associated toxicity in biological samples.[16] Prolonged UV exposure, mainly UVB (280-315 nm), is well known to result in adverse effects on genome stability via mutagenic and cytotoxic DNA (deoxyribonucleic acid) lesions.[20] Therefore, when developing photopharmacology agents, the furthest red-shifted wavelengths possible are used whilst still maintaining the desired photochemical effect.

(2) Quantitative photoconversion between the isomers of the photoswitch is also highly desirable. This enables the complete isolation of a single photoswitch isomer following light irradiation. In photopharmacology, full activation or deactivation of ligands is essential to ensure tight regulation of a biological target in a precise, spatial, and temporal manner.

(3) A biologically relevant thermal relaxation is also important, as it determines the applications that the photoswitch can be used for. Traditionally, biologically relevant thermal relaxation times, i.e. time for isomerisation from the active to inactive isomer, are on the order of magnitude between hours-to-days.[15] This is to prevent constitutive activation of the photopharmacological agent following irradiation. The primary reason for this is that it is not always possible to deactivate the therapeutic following dosing. Having a therapeutic which deactivates over a suitable timeframe enables greater control over the biological activity governed.[15]

1.2.1 Prerequisite key terms and definitions in photochemistry:

These key terms are (1) photostationary state (PSS) and (2) thermal half-life. Both are crucial for work within the photochemistry and photopharmacology fields and are defined below.

Key Term (1) – Photoconversion efficiency at a given photostationary state (PSS):

The proportion of a given isomer population present at a specific wavelength of light is described as the photoconversion efficiency of the isomerisation at the photostationary state (PSS).[7,21] The PSS is the steady state reached within a chemical reaction where the rates of both the forward (A→B) and reverse isomerisation (B→A) become equal under continuous irradiation at a specific wavelength of light (Eq.1.2.1-1).[7,21]

$$PSS(\lambda) = \frac{A \rightarrow B}{B \rightarrow A} \qquad \text{(Eq.1.2.1-1)}$$

Describing this in greater detail, the photoisomerisation rates are equivalent to the quantum yield ($\Phi_{A\rightarrow B}$ or $\Phi_{B\rightarrow A}$) multiplied by the extinction coefficient (ε_A or ε_B) at the wavelength of irradiation λ (Eq.1.2.1-2). Quantum yield (Φ) refers to the efficiency of a light-triggered process x at a given wavelength λ. Specifically for photoswitches, the $\Phi_x(\lambda)$ is equal to the number of photoisomerisation events occurred n_x divided by the amount of photons n_p supplied to the system with wavelength λ. The molar extinction coefficient is a measure of how strongly a chemical species or substance absorbs light at a particular wavelength.[7,21]

$$PSS(\lambda) = \frac{\phi_{A\rightarrow B}\, \varepsilon_A(\lambda)}{\phi_{B\rightarrow A}\, \varepsilon_B(\lambda)} \qquad \text{(Eq.1.2.1-2)}$$

Photoconversion efficiency at a given photostationary state is experimentally determined primarily via either UV-Vis spectrophotometry or NMR Spectroscopy. To achieve high photoconversions at a given PSS, absorption band separation is essential such that irradiation at a specific wavelength drives the conversion of one isomer to the other, solely (Figure 3). [7,21]

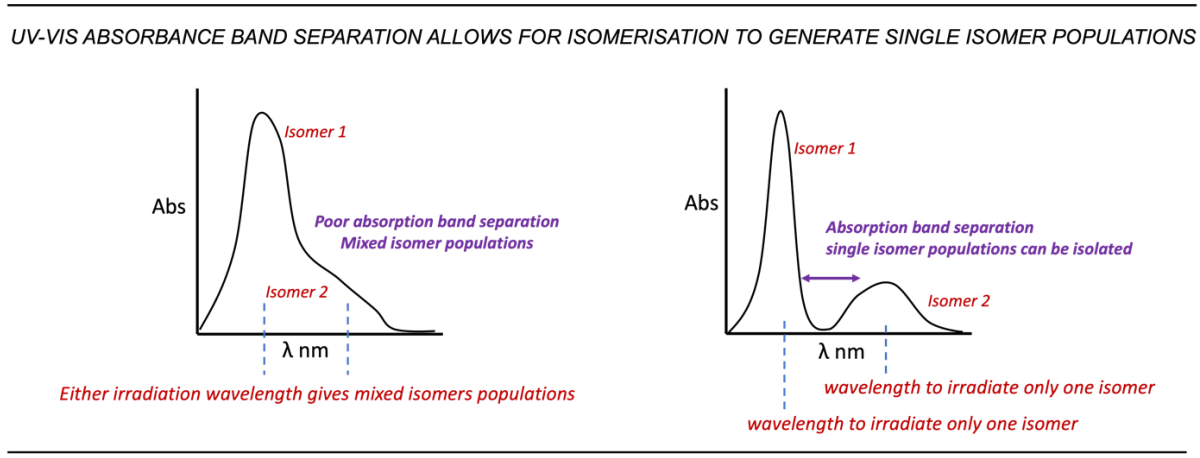


Figure 3:Example UV-Vis absorption spectra of a given photoswitch which photoisomerises between two isomer populations. If there is absorption band overlap at a given wavelength a mixed isomer PSS at that wavelength (λ) will be seen. In contrast, if good absorption band separation is seen, single isomer PSS populations can be isolated by irradiating at each isomers corresponding λ.

Key Term (2) – Thermal relaxation – Thermal half-life ($\tau_{1/2}$):

Thermal relaxation is experimentally defined as the thermal half-life of a molecular photoswitch. It is used to define how long the molecular photoswitch remains in its metastable isomer. The thermal half-life of photoswitch represents the time needed for the concentration of the metastable isomer to decrease to one-half of its original value, interconverting to the thermodynamic isomer in the process. As thermal isomerisation typically follows first-order kinetics it can be calculated using Eq.1.2.1-4 where k equals the thermal isomerisation rate constant.

$$\tau_{1/2} = \frac{ln2}{k} \qquad \text{(Eq.1.2.1-4)}$$

Specifically for this work, $\tau_{1/2}$ is defined as the rate at which half of the metastable *Z*-isomer has interconverted back to the *E*-isomer, which is the thermodynamic state. To monitor this, the increasing absorbance at 365 nm over time is tracked via UV-vis spectrophotometry following irradiation at a specific wavelength.

1.2.2 Classical azobenzenes and their derivatives:

Classical azobenzenes are considered the archetypal compounds of the azo-based class of molecular photoswitches. First prepared by Mitscherlich in 1834,[22] azobenzene (3) was developed as a novel synthetic dye to replace natural pigments. Since then, derivatives have undergone continuous development for numerous applications such as textiles, cosmetics, food, energy storage and medicine.[22] In Photopharmacology, most applications use the change in end-to-end distance of the phenyl rings (*E*-isomer to *Z*-isomer, 0.90 nm to 0.55 nm),[23] occurring upon photoisomerization, as the mechanism for driving functional change in the targeted biomolecule (Figure 4(a)).[15,24]

Photoisomerisation is achieved in a photochromic manner as azobenzenes show two key characteristic absorption bands for each isomer in the UV–vis absorption spectra (200-800 nm). The absorption bands correspond to π-π* transitions (~340 nm) and n-π* transitions (~450 nm) about the azo bridge (N=N (Figure 4(b/c)).[23] For the planar *E*-azobenzene (3), due to symmetry considerations, it shows strong π-π* transitions (excitation of an electron from the azo bonding π orbital to an antibonding π* orbital) but weak n-π* transitions (excitation of an electron from the nitrogen non-bonding orbital to an antibonding π* orbital). This is because the n-π* transition is considered symmetry forbidden (C_{2h} point group S_0 to S_1 transition is forbidden).[25] Only a small absorption band is seen due to vibronic processes resulting in a partial break in symmetry. In contrast for Z-azobenzene (3), due to the non-symmetrical twisted shape of Z-azobenzene (3) the n-π* transitions become symmetry allowed showing greater absorbance whilst the π-π* absorption band is shifted to shorter wavelengths (hypsochromic effect) and decreases significantly in intensity.[23] The exact mechanism of switching remains debated in the literature but is believed to proceed via rotation, inversion, or a combination of the two (Figure 4(d)).[26] There are also further variations proposed, concerted inversion and inversion-assisted (not shown), but they remain contested.[26]

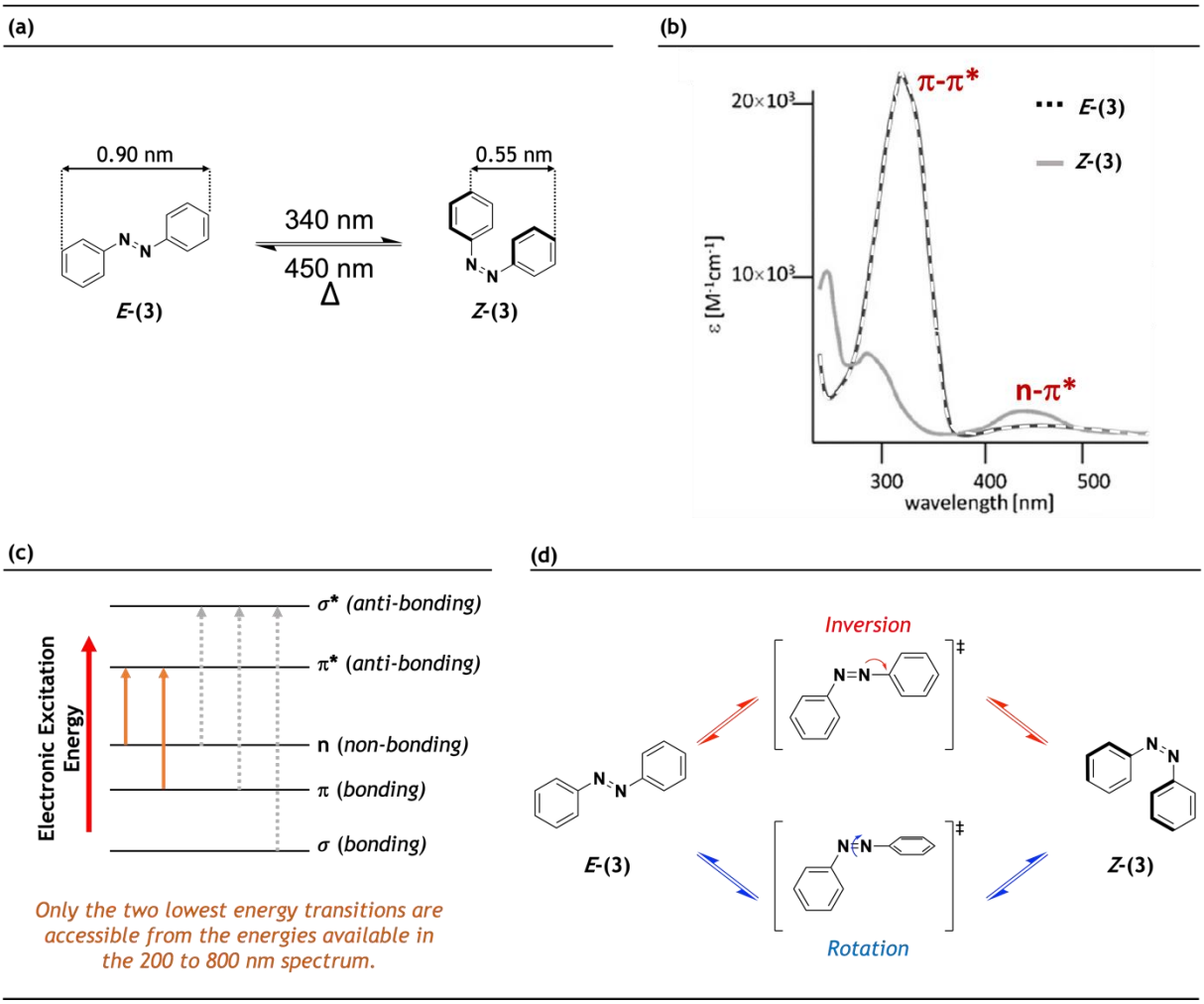


Figure 4: Classical azobenzenes, their structure, photochemical properties, and isomerisation mechanism. (a) E-isomer and Z-isomer of classical azobenzene (3) with phenyl end-to-end distance shown. (b) Photochemical profile of (3) across the UV-Visible spectrum with molecular orbitals of the azo bridge referenced – adapted.[24,27] *(c) Energies in UV-Vis spectrum (143-36 kcal/mole, 200-800 nm) are only sufficient to promote electronic excitation to the two lowest transitions (n-π* and π-π*), out of the six possible energy transitions – adapted.*[28] *(d) Proposed mechanisms for azobenzene isomerisation.*

1.2.3 The photochemical properties of azobenzenes making them ideal candidates for photopharmacology:

The ubiquitous nature of azobenzenes in photopharmacology is due to their largely ideal photochemical properties. The three key photochemical properties in the context of azobenzene photoswitches are described below:

(1) Towards red-shifted switching: Isomerisation of azobenzene between each isomer (*E*-isomer and *Z*-isomer) uses wavelengths of light between 360-500 nm, depending on the substituent groups upon the phenyl rings. Although not fully red-shifted, these wavelengths are sufficient for biological applications without causing a high mutational burden.[22]

(2) Near quantitative photoconversion between the isomers of the photoswitch: Undecorated azobenzenes display near but non-quantitative isomerisation for the *E*→*Z* π-π* transition (~90% at PSSs between 290-350 nm) (Figure 4(b)).[7] For the *Z*→*E* n-π* transition however, incomplete photoconversion (~79% at PSS of 420-450 nm) is seen.[7] This is because the nπ* bands of both isomers significantly overlap, with the λmax for *E* and *Z* isomers being 445 nm and 435 nm, respectively. Despite this overlap, the reason the considerable photoconversion yield at ~79%, is due to the difference in the molar extinction coefficients of the nπ* bands of both isomers (*E*: 390 L mol^{-1} cm^{-1}, Z: 1100 L mol^{-1} cm^{-1}).[7] This difference in molar absorption is attributed to the planar structure of *E*-azobenzene, where the nπ* transition is symmetry forbidden, in contrast to the slightly distorted structure of the Z-isomer.[7,29] The non-quantitative back-isomerisation (*Z*→*E* n-π* transition) is the greatest drawback of the class as a mixed population exists upon switching which is not ideal for photopharmacology.

(3) Thermal Half-life: A classical azobenzene has a thermal half-life of $\tau_{1/2}$ of 2-4 days depending on the solvent and completion of switching to the *Z*-isomer population.[30] This falls within the desired magnitude for photopharmacological applications.

Rational design programmes have also been undertaken to improve the key photochemical properties of this class where possible. Particularly interesting is the work of Hecht *et al.*,[31] with the development of *o*-fluoroazobenzenes, an analogue containing two ortho-fluorine atoms adjacent and an ethyl ester para- to the azo bridge on each phenyl ring. This derivative has superior key photochemical properties, exhibiting selective excitation of either isomer with visible light, resulting in high (90% *E*→*Z* using a PSS λ> 500 nm) to almost complete (97% Z→E at PSS of 410 nm) photoconversion. Despite the advantages in photoconversion, these *o-fluoroazobenzenes* display extended thermal relaxations at 700 days for (*tetra*-fluoro-*o*-azobenzene). This is too long for biological applications as following irradiation the therapeutic would remain active long beyond its duration of action.

1.2.4 Applications of azobenzenes in photopharmacology:

As azobenzenes are considered the archetypal class of photoswitches, they have been extensively used in photopharmacology applications with significant but limited success. Many examples can be found from the extensive work of Trauner and co-workers in this field.[32–36] Recent examples include the development of Photochemically Targeting Chimeras (PHOTACS) to target the Ca2+/calmodulin-dependent protein kinase II alpha (CaMKIIα),[36] where CaMKIIa was removed from the mouse hippocampus only within 25 μm of the illuminated brain surface. The optically controlled degradation decreased synaptic function within minutes of light activation despite the incomplete switching displayed by the azobenzene-based molecular photoswitch.

One further example relates to the use of azobenzenes for the optical control of translation with a Puromycin Photoswitch (Puroswitch).[37] Puromycin is an aminoacyl tRNA (transfer ribonucleic acid) mimetic causing ribosomal stalling, leading to premature chain termination. Activation of the puroswitch (Z-puroswitch) led to suppression of ribosomal translation in a dependent manner. For both applications presented, the impact of limited back-switching (inactivation of the photopharmacological agent) is minimised as the systems are designed to be triggered upon activation without needing specific deactivation.

Most excitingly, an azobenzene-based photoswitch (KIO-301 – Kiora Pharmaceuticals) is currently undergoing its first-in-human phase I/ II clinical trials, aiming to restore functional vision in patients with retinitis pigmentosa.[38,39] KIO-301 is taken up selectively by viable retinal ganglion cells downstream of degenerated photoreceptors. KIO-301 inhibits voltage-gated ion channels following light-activated inhibition (*Z*-active). Ion efflux (K+) is prevented, resulting in a charge gradient triggering photo-transductive cell activation, and signalling to the brain.[40]

These examples, in combination, highlight the successful ways the intrinsic limiting photochemical properties related to azobenzenes can be overcome. But does mean that there is further scope to improve the photoswitches used in photopharmacological applications. Particularly in settings where tightly controlled activation and deactivation is desired in a highly precise, spatial, and temporal manner.

1.2.5 Azopyrazoles – improved intrinsic photochemical properties:

Expanding on classical and derivatised azobenzenes, an emerging sub-class of azo-based molecular photoswitches is that of arylazoheteroarenes, more specifically azopyrazoles. First developed by Fuchter and co-workers, these photoswitches contain a pyrazole and phenyl ring bridged by an azo moiety.[41–46] In particular, the azopyrazole derivative, 4pzMe identified by Weston *et al*.,[44,46] displays quantitative bi-directional photoconversion between the *E*- and *Z*-isomers and displays a thermal relaxation ($\tau_{1/2}$) of 10 days (Figure 5(a/b)).[44,46] Similarly to azobenzenes, azopyrazoles also display large changes in the pyrazole-to-phenyl end-to-end distance and have related but distinct isomerisation mechanisms.[46] Undecorated azopyrazoles also adopt a non-planar Z-isomer conformation which is T-shaped.[46] This is due to the less sterically encumbered pyrazole, in contrast to the larger phenyl ring of azobenzenes which adopts a slightly twisted *Z*-conformation.

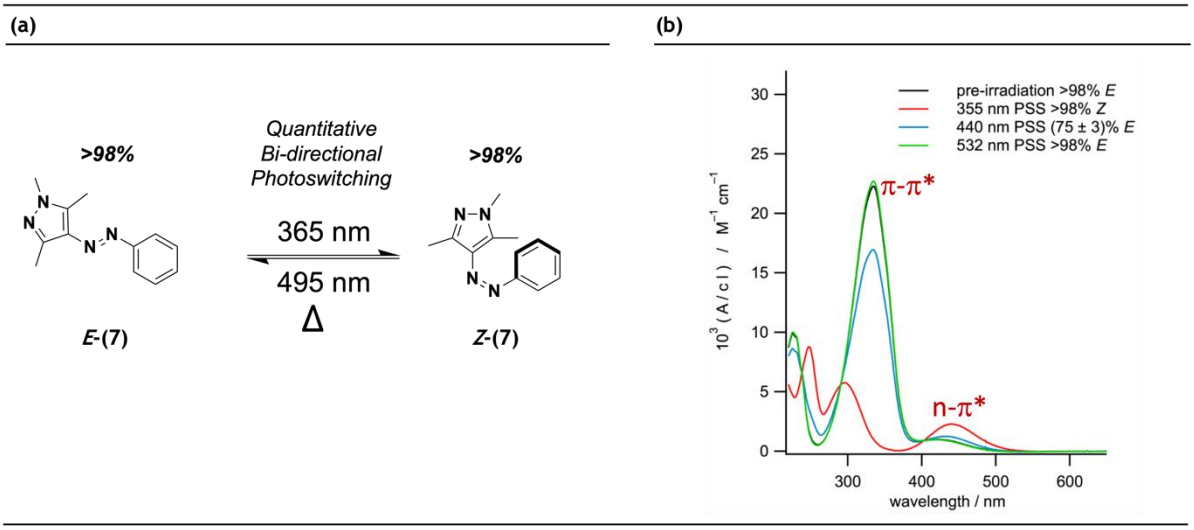


Figure 5: 4pzMe an azopyrazole photoswitch, structure and photochemical properties. (a) E-isomer and Z-isomer of azopyrazole 4pzMe (7). (b) Photochemical profile of (7) with electronic excitation bands relating to the molecular orbitals of the azo bridge shown. Photochemical profile adapted from Weston et al.[44,46]

The 4pzMe azopyrazole derivative (7) is particularly important as it also largely fulfils the key photochemical requirements of a good molecular photoswitch for photopharmacology applications. Most prominently, it equals or surpasses the photochemical profile of classical azobenzenes and is gaining traction in the literature. **(1)** It uses wavelengths for irradiation that are comparable to the classical azobenzenes, although not fully red-shifted. Improving the overall redshift of these photoswitches is a priority for the field but remains challenging to match predicted properties to experimental outcomes.[47] **(2)** It offers quantitative photoconversion between the two isomers of the molecular photoswitch which is superior to undecorated azobenzenes and **(3)** it offers thermal

relaxation rates that are almost optimal for biological applications, whilst maintaining the desired quantitative switching (not seen e.g., *O*-fluoroazobenzenes). In summary, 4pzMe appears strongly suited for applications within photopharmacology surpassing classical azobenzenes.

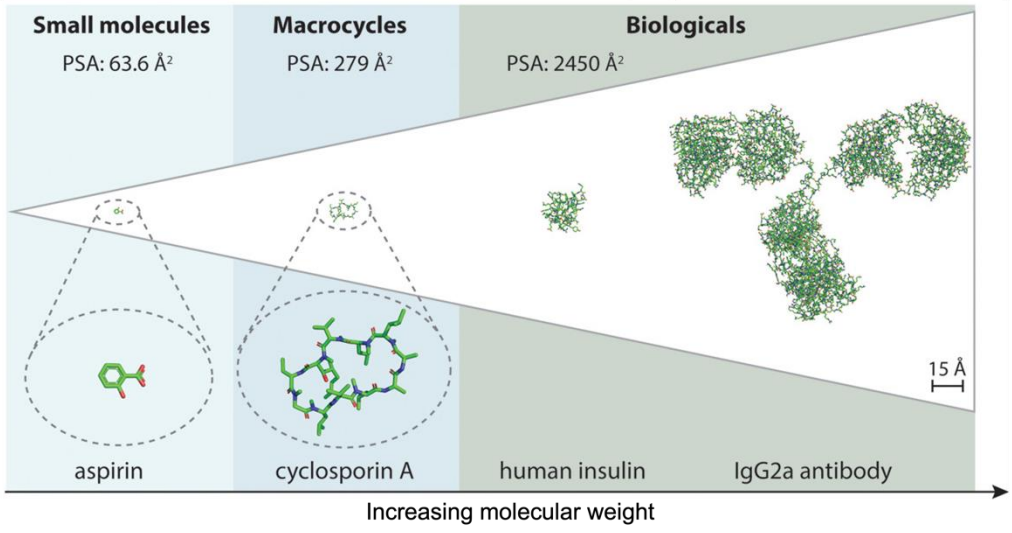
1.2.6 Applications of azopyrazoles in photopharmacology:

The use of azopyrazole molecular photoswitches within photopharmacology is emerging due to the acknowledgement of their photochemical properties by the field. Lam *et al.*,[48] developed photo-regulated control of the transient receptor potential ankyrin 1 (TRPA1) channel in larval zebrafish hearts exogenously expressing zebrafish TRPA1, such that the zebrafish heartbeat could be controlled *in vivo* using an azopyrazole-based TRPswitch and light. The azopyrazole-based TRPswitch allowed for reversible, repeatable, and nearly quantitative light-induced activation and deactivation of the vertebrate TRPA1 channel with violet and green light, respectively. This system highlighted the need for high-efficiency bidirectional optical switching for certain photopharmacology applications. This is something the azopyrazole class can achieve in a superior manner to the classical azobenzenes.

1.3 Cyclic peptides as emerging tool compounds:

Cyclic peptides are an emerging class of exciting chemical tools, bridging the gap between small molecule ligands and larger biologics.[49] This is due to their intermediary, size and physicochemical properties, placing them within the 'Goldilocks' region of chemical space (Figure 6(a)).[50] This emergence has also been reflected in the pharmaceutical pipeline with 18 cyclic peptides approved for clinical use within the past two decades.[51]

(a)



(b)

Peptide Sequence:	Small Molecules:	Macrocycles:	Larger Biologics:
Target Potency	+	++	+++
Target Selectivity	+	++	+++
Metabolic Stability	+++	++	+
Size of binding interaction	+	++	+++
Oral Bioavailability	+++	++	+
Toxicity	+++	++	+
Manufacturing Cost	+	++	+++

Figure 6: Cyclic peptides, the sweet spot between small molecules and larger biologics. (a) The chemical space of bioactive modalities, small molecules, macrocycles, and larger biologics and the relative sizes of each modality. (b) Table showing the characteristics of each of the bioactive modalities where macrocycles possess the intermediary properties of small molecules and larger biologics. (a and b) – adapted with permission (permission available at Appendix 43).[49]

Due to their properties (Figure 6(b)), cyclic peptides are excellent tools for probing classically undruggable protein targets of interest, such as proteins with shallow surface topologies, or to disrupt protein-protein interactions. These are areas traditionally challenging for small molecule therapeutics

and as such are currently dominated in the clinic by larger biologics (e.g. antibody therapies) (Figure 6(b)). Similarly, to larger biologics, cyclic peptides have increased target selectivity and tight binding affinity. However, advantageously, cyclic peptides also offer more advantageous pharmacokinetic and pharmacodynamic properties when compared to larger biologics. This is because larger biologics are often prone to poor bioavailability, premature degradation, and immunogenicity over time, whilst also being expensive to manufacture on scale.[49,52]

1.3.1 Advantages of cyclic over linear peptides:

Cyclic peptides have distinct advantages over linear peptides for therapeutic or chemical biology applications. Firstly, when cyclised, the increased conformational rigidity can result in increased binding affinity and often selectivity towards the target. This is because the increased conformational strain results in a smaller change in entropy (ΔS), i.e., a lower entropic cost contributing to Gibbs free energy ($\Delta G = \Delta H - T\Delta S$), when the cyclic peptide is engaged in a large binding surface.[53,54]

Cyclic peptides also display greater resistance to peptidases when compared to their linear counterparts.[54] Cyclisation most dramatically increases resistance to exopeptidases, such as carboxypeptidases or aminopeptidases. Carboxypeptidases and aminopeptidases cleave single amino acid (AA) residues from the C- or N- termini respectively, if exposed. Cyclisation often masks one (head-to-side chain cyclisation) or both termini (head-to-tail cyclisation) preventing degradation via these enzymes.[55,56] Cyclic peptides may also impart some resistance to endopeptidases (i.e. pepsin, trypsin, chymotrypsin), reducing cleavage at specific motifs within peptide chains due to their conformational rigidity.[57]

1.3.2 Advantages of introducing ns-AAs into peptidic macrocycles:

Incorporation of non-standard amino acids, particularly *N*-methylated residues, also offers greater resistance to endopeptidases significantly increasing the serum stability of non-standard amino acid (ns-AA) containing macrocyclic peptides.[58] Incorporation of ns-AAs also aids in diversifying the chemical space cyclic peptides occupy within drug discovery.[59,60] This is especially seen in natural product peptidic macrocycles which have undergone extensive post-translational modification (or non-ribosomal synthesis) to achieve a final bioactive state.[61] Importantly, the incorporation of ns-AAs can also impart a host of advantageous characteristics, including, greater target potency, specificity, macrocycle rigidity, resistance to degradation and premature clearance, and importantly cell penetration.[62]

1.4 Discovery platforms to identify novel cyclic peptide ligands:

Due to their emergence as chemical tools, methods to discover novel cyclic peptide ligands have rapidly expanded. A range of *in vivo* and *in vitro* display systems based on DNA-encoded peptide libraries have been developed in the literature.[49] All can be divided into two main components (1) A peptide generation and display step, which is then followed by (2) An affinity panning step to identify novel cyclic peptide binders against an immobilised protein of interest. The primary difference between the techniques is the method to achieve peptide generation and display, (1).

1.4.1 *In vivo* display systems:

The most popular *in vivo* display system currently used is phage display, although other notable methods include, yeast display, SICLOPPS ('split intein-mediated circular ligation of peptides and proteins') and bacterial cell surface display.[63–65] Phage display is an affinity panning approach using filamentous bacteriophage (phage) as the *in vivo* host for generating the peptide libraries (Figure 7).[49,66] Phage are viruses which infect bacteria (*E.coli typically used for phage display*), hijacking the translation machinery to replicate. As part of the replication process phage coat proteins are synthesised. Phage coat proteins are endogenously displayed on the outer surface of the phage and this is exploited in phage display. For use as a discovery platform, the peptide library is encoded within the phage coat protein gene of the bacteriophage (recombinant phage). As such, following the translation of the phage genes required for replication (via *E.coli* translation machinery), the encoded cyclic peptide is displayed on the phage surface as part of the phage coat protein. The genetic barcode for the presented peptide is retained within the phage genome – preserving the genotype-phenotype link. The peptides can then be cyclised through a variety of stapling strategies if desired.[67] The (cyclic) peptide presenting phage are then ready to be panned against an immobilised protein target of interest. Enriched hits are decoded after several rounds of enrichment via high throughput sequencing (e.g. Next Generation Sequencing (NGS)) of the phage coat gene. This reveals the sequences of the binding peptides to the target of interest.

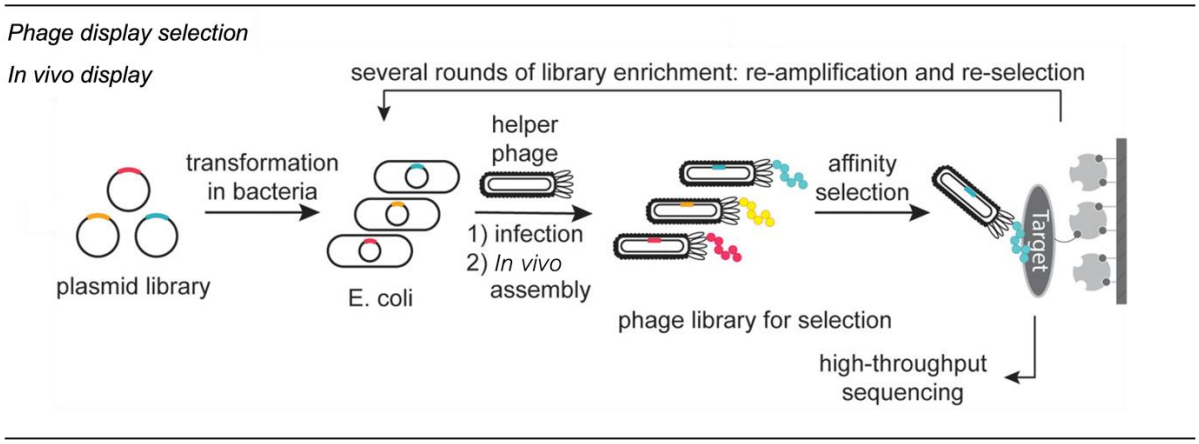


Figure 7: Phage display in vivo selection method. A plasmid library each encoding for phage containing a cyclic peptide sequence (modified coat gene) is transformed into bacteria. Cyclic peptide library is presented on phage and panned against immobilised protein of interest. Several rounds of library enrichment occur to find tight binding sequences. Figure adapted from Huang et al. (permission available at Appendix 43).[49]

Despite their popularity, all *in vivo* display systems require transformation of the DNA libraries into the *in vivo* host (transformation/ transfection).[49] This remains the limiting factor for peptide library diversity, e.g. for phage display a maximum of $<10^9$ unique library members can be generated due to the incomplete efficiency of plasmid uptake by the organism (e.g., bacteria).[49] Furthermore, incorrect *in vivo* folding due to a lack of the correct eukaryotic chaperones within bacteria, peptide toxicity and failed membrane insertion of the cyclic peptide-coat protein fusion, typically lead to biased and further diminished sequence diversity.[49] This led to the development of *in vitro* cell-free display methods which circumvent this.

1.4.2 *In vitro* display systems:

An alternative to *in vivo* display systems is the use of *in vitro* cell-free display methods. These methods differ from *in vivo* display systems as they do not require a cellular host to generate and present the mRNA-peptide library. They rely on the direct linkage of the peptide sequence with its encoding mRNA, achieved via differing methods. *In vitro* display systems include ribosomal display and mRNA display. These *in vitro* systems require greater technical stringency (i.e., nuclease-free environments), but offer significantly increased library diversity (10^{12}-10^{15} unique library members) when compared to yeast or phage display (10^7-10^9).[49] This is because these methods circumvent the need to introduce the library DNA into a cellular host, as the DNA libraries can be directly translated by a reconstituted translation system (once transcribed to mRNA). The *in vitro* cell-free systems are accessible via the

translation machinery of cell extracts (e.g., rabbit reticulocyte lysate) or purified translation components (PURE system).[68,69] The PURE System (protein synthesis using recombinant elements) first developed by Ueda *et al.* in 2001,[69] is a completely reconstituted cell-free system comprising 36 enzymes that are involved in transcription and translation, as well as highly purified 70S ribosomes.[70,71] The PURE system enables greater system precision and manipulability due to its modular composition and is used predominantly for Ribosome and mRNA display.

1.4.3 Ribosomal display:

Ribosomal display couples the genotype (mRNA) to the phenotype (nascent peptide) through a non-covalent peptide-ribosome-mRNA ternary complex (Figure 8(a)).[49] This complex is stabilised via stop codon mutation, Mg^{2+} addition and precise temperature control (4 ^{o}C).[72] Following stabilisation of the peptide-ribosome-DNA complex, similarly to phage display, the translated cyclic peptides are panned against an immobilised protein of interest. Sequence deconvolution is then achieved via high throughput sequencing (e.g. NGS). The greatest limitation to ribosomal display is the magnesium-rich and low-temperature conditions, which can result in a loss of potential binders.[49] To circumvent these issues, mRNA display was subsequently developed, largely surpassing this technique for cyclic peptide discovery applications.

1.4.4 mRNA display:

mRNA display (Figure 8(b)), differs from ribosomal display primarily through the method of genotype-phenotype linkage, the way to ensure that the genotype-phenotype (encoding DNA-cyclic peptide) link is maintained. As mentioned, ribosomal display uses a non-covalent (mildly stable) peptide-ribosome-mRNA ternary complex. In contrast, mRNA display uses a covalent ternary complex approach of peptide-puromycin-mRNA first developed by Yanagawa and Szostak independently in 1997 (described in greater detail herein).[49,73,74] The workflow of mRNA display to generate the peptide-puromycin-mRNA complex formation is as follows:

(1) Transcription of the DNA library to mRNA.
(2) Puromycin ligation to mRNA.
(3) Translation of the encoded peptides: peptide-puromycin-mRNA complex formation.
(4) Reverse transcription of the mRNA to mRNA/ DNA hybrid.
(5) A peptide-mRNA/ DNA hybrid screening library is generated.

It is important to note here that other derivatives of mRNA display which enable direct *in vitro* transcription, translation and puromycin ligation in a 'one-pot' method do exist e.g., TRAP display (Transcription−translation coupled with association of puromycin linker).[75] This system, however, does not follow the mRNA display workflow most similar to RaPID-related approaches and as such will not be discussed in further detail.

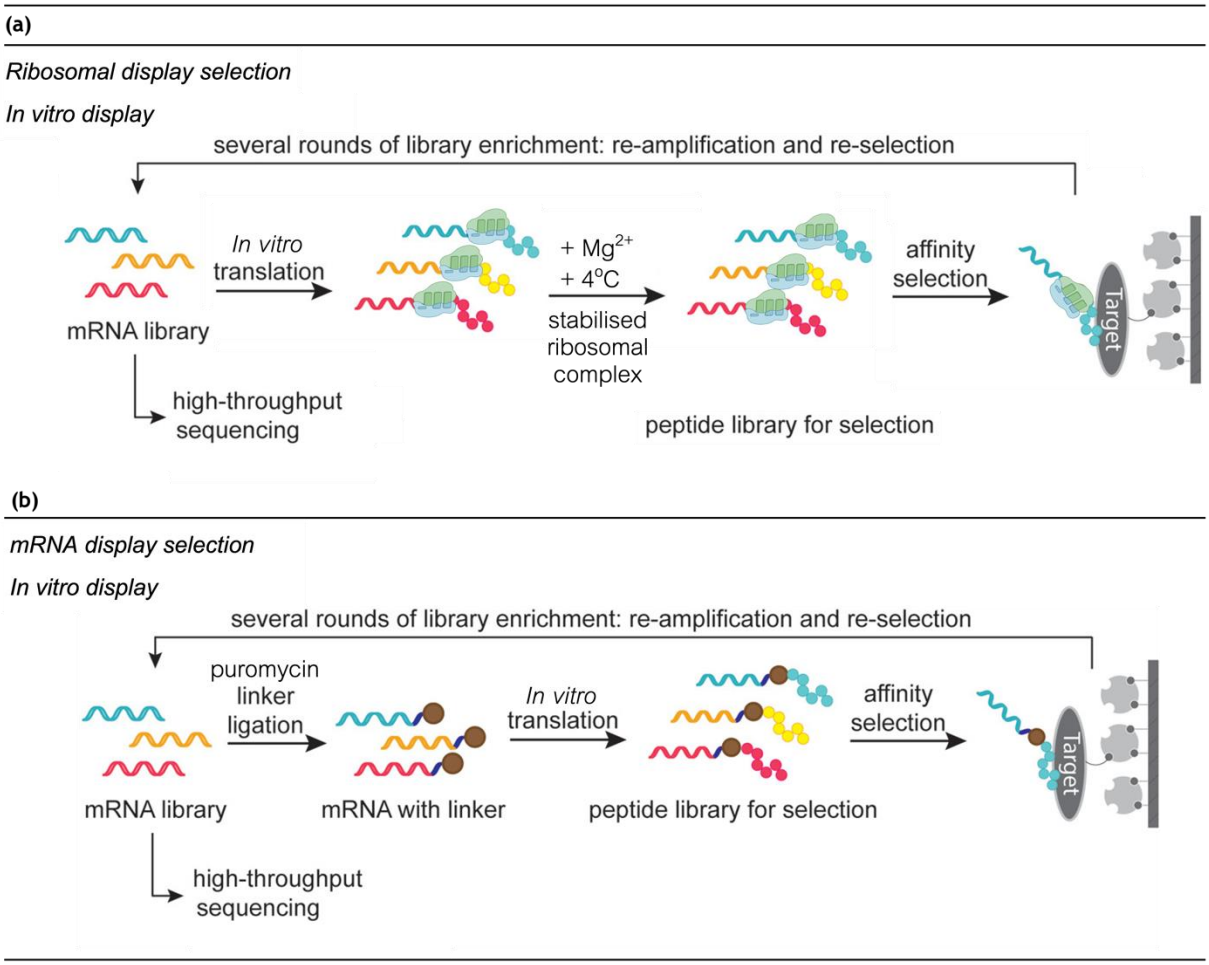


Figure 8: In vitro display selections: (a) Ribosomal display: An mRNA library undergoes in vitro translation. Upon reaching the stop codon, stabilisation of the peptide-ribosome-mRNA ternary complex ensures that the peptide sequence remains linked to its encoding mRNA – this complex is relatively unstable, requiring precise maintenance of magnesium-rich and low-temperature conditions to prevent ribosome disassembly. (b) mRNA display: An mRNA library is ligated to a synthetic oligo linker with terminal puromycin. Translated peptides using a reconstituted ribosome are conjugated to their encoding mRNA via puromycin. Following reverse transcription (mRNA/ DNA hybrid) the cyclic peptide library is panned against an immobilised protein of interest. Several rounds of library enrichment occur to find tight binding sequences. Figure adapted from Huang et al. (permission available at Appendix 43).[49]

1.4.5 DNA transcription to mRNA and puromycin ligation:

Following transcription of the DNA library using T7 RNA polymerase, at the 3' end of the mRNA, a synthetic linker (DNA-PEG-Puro linker) is conjugated (Figure 9). The DNA-PEG-Puro linker is composed of 16 DNA nucleotide bases, followed by a polyethylene glycol (PEG) spacer (5x *sp18 spacer*). Two additional cytosine nucleotides then follow, ending with puromycin, a naturally occurring amino-adenosine nucleoside antibiotic that inhibits protein synthesis by preventing chain elongation.[76] In the case of mRNA display though, puromycin is not used as an antibiotic. It is repurposed to enable the direct conjugation of this linker to the encoded nascent peptide chain. The purpose of the adenosine and two cytosine nucleotides proceeding the puromycin is to mimic the CCA tail of a tRNA to promote entry of the terminal puromycin (an amino-adenosine nucleoside mimetic - CCA) into the A site of the ribosome.

One of the most common methods of mRNA-puromycin ligation is Y-loop ligation. The first three nucleotide bases of the linker are non-complementary (5'-Phos-CTC-3') to the mRNA library (-AAA-3') (Figure 9), followed by the remaining 13 nucleotides which are complementary to promote annealing. The terminal non-complementary DNA results in a bulged Y-loop, angling the DNA strands for T4 RNA-DNA Y-loop ligation using a T4 RNA polymerase in the presence of ATP. Finally, the presence of the 5'-phosphorylated cytosine is also required, providing the basis for the phosphodiester backbone following T4 RNA-DNA ligation.

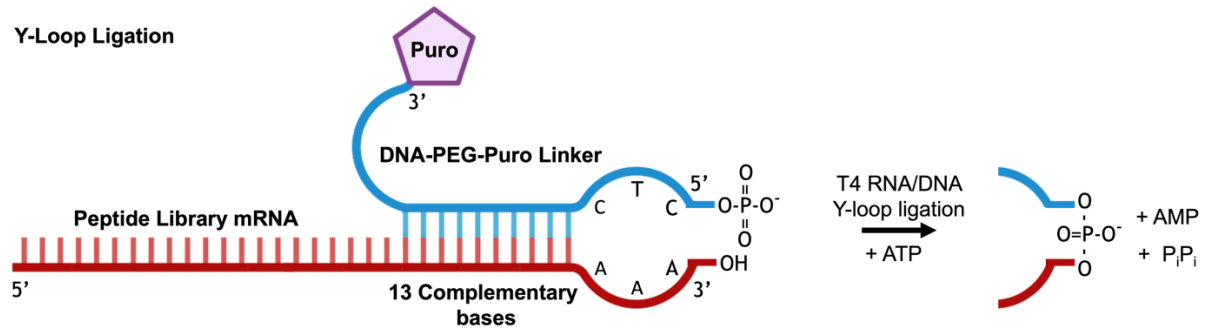


Figure 9: Puromycin ligation step: A synthetic oligo (blue) containing 16 DNA-nucleotide bases (1-3 bases mismatch to the peptide library mRNA (red), 4-16 bases complementary, A flexible PEG spacer, terminating (3') with puromycin. Following annealing, T4-RNA/DNA Y-loop ligation occurs between the 5' phosphorylated cytosine in the DNA PEG-Puro linker and the 3' hydroxyl of the ribose sugar backbone of the peptide library mRNA via T4 RNA ligase and ATP.

1.4.5.1 Ribosomal translation:

Classical prokaryotic ribosomal translation proceeds via the ribosome 'reading' the mRNA in the 5'-to-3' direction (Figure 10).[77] It is initiated by the small ribosomal subunit interacting with initiation factors (IF1 2, 3), the mRNA to be translated, and initiator Met-tRNA (^{Ini}Met-tRNA). Recruitment of the large ribosomal subunit (50S) then complexes with the small subunit around the mRNA. This places the ^{Ini}Met-tRNA within the peptidyl (P) site of the ribosome complex.[78]

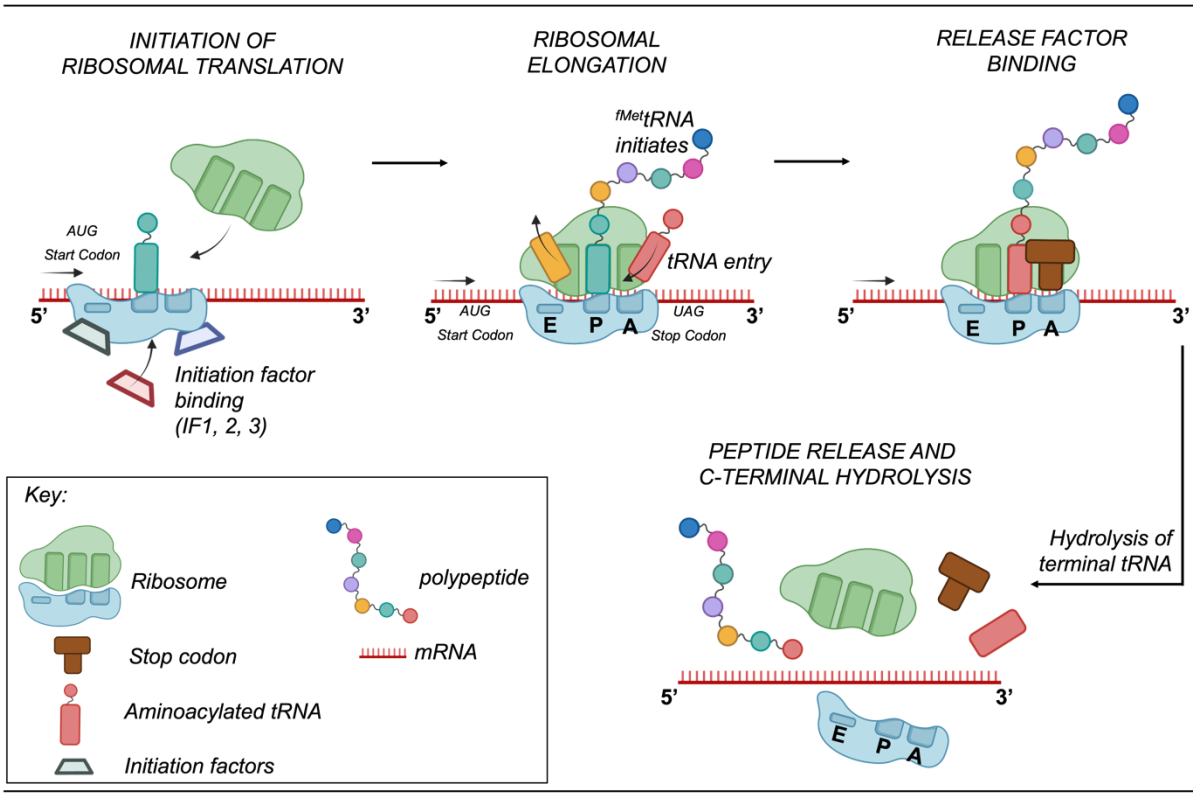


Figure 10: Ribosomal translation: ribosomal translation occurs via initiation with Met-tRNA and subsequent complementary tRNA binding to the A site of the ribosome. This is per the encoding mRNA sequence. Amide bond formation occurs via nucleophilic attack of the incoming amino acid at the ester carbonyl of the amino acyl-tRNA. Following amide coupling, the hydrolysed tRNA exits and the nascent peptide-tRNA transfers to the P site ready for subsequent coupling. Release factors (brown) enter the A site upon reaching a stop codon and simultaneously release the ribosomal subunits whist promoting water catalysis to hydrolyse the terminal amino acid from its cognate tRNA. Peptide release with free carboxyl terminus completes ribosomal translation.

Note: Methionine is universally used as the initiation amino acid for biological systems. In prokaryotic systems, it is endogenously formylated (^{f}Met) while in eukaryotic systems it remains unformylated.[79] The structure of ^{Ini}Met-tRNA is shown in Figure 11.

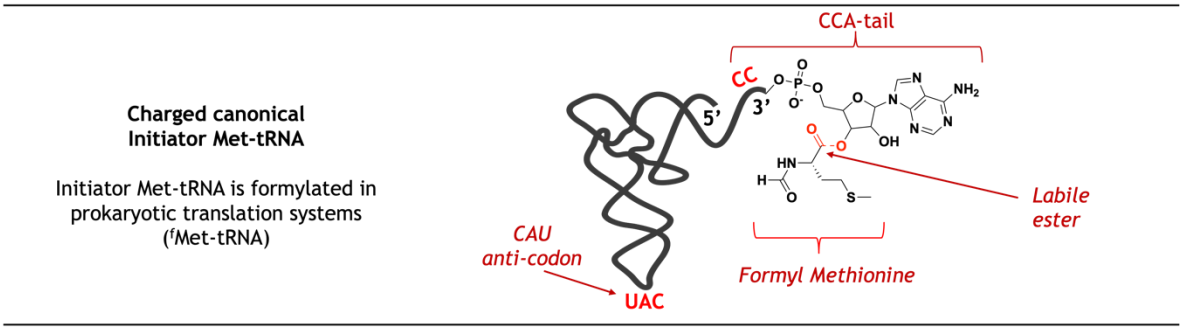


Figure 11: Structure of a prokaryotic initiator methionine aminoacylated tRNA bearing the anticodon CAU. Note the formylated methionine, indicative of a bacterial system and labile ester conjugating the AA and the adenosine nucleoside of the conserved CCA tail of the tRNA.

Following initiation, translocation of the large ribosomal subunit followed by the small ribosomal subunit then occurs. As a result of the two translocation steps, the ribosome complex moves three nucleotides along the mRNA. With the next mRNA codon exposed, the growing peptide chain is supplied with the cognate elongation aminoacyl-tRNA (aa-tRNA), corresponding to the encoding mRNA sequence, which enters via the aminoacyl (A) site within the ribosome. [77]

> Note: An aa-tRNA comprises of the AA conjugated at the terminal 3' hydroxyl adenosine of the cognate tRNA (CCA tail of tRNA). Critically an ester bond conjugates the amino acid to the tRNA (Figure 12). This bond is labile enough to promote amide formation via peptidyl transferase catalysis within the ribosome. The mechanism consists of an S_N2 attack (nucleophilic substitution) from the amino group of the incoming aa-tRNA (e.g.Tyr-tRNA) bound to the A site when brought into proximity with the nascent peptide chain in the P site by the quaternary structure of the ribosomal complex.

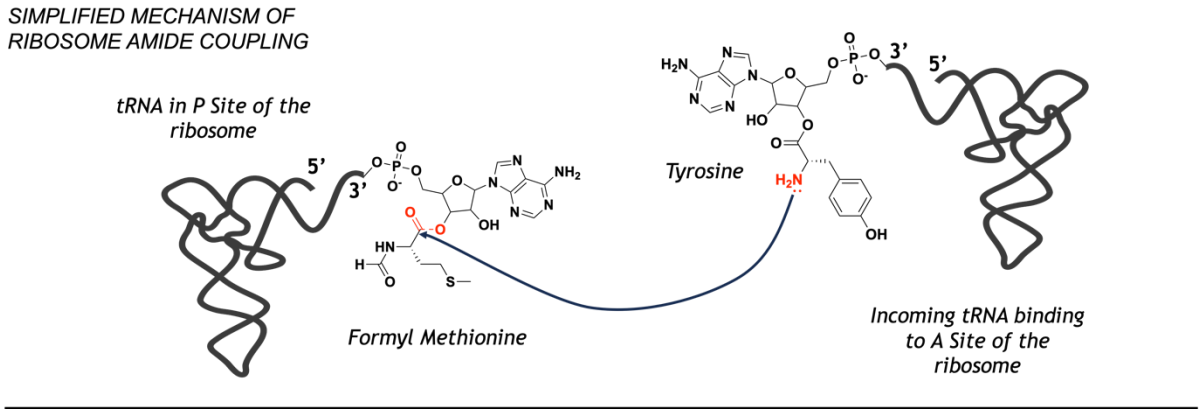


Figure 12: Ribosomal Amide Coupling occurring between the $^{ini\,f}$Met-tRNA in the P site of the Ribosomal complex and an incoming tRNA entering the A site.

Following the first peptide elongation, the dipeptide-tRNA (e.g., ^{f}Met-Tyr-tRNA) is then transferred into the P-site ready for sequential amino acid conjugation further growing the nascent peptide chain following the directing mRNA sequence.[77]

Translation ends upon the presence of a stop codon (bacterial stop mRNA codons: UAG, UAA, or UGA) in the mRNA sequence at which point release factors (RF1, 2 and 3) bind to the ribosome, releasing the nascent peptide-tRNA (Figure 10).

> Note: Release factors fall within two classes, class 1 and class 2. Class 1 release factors (RF1 or RF2) recognize stop codons and bind at the A site of the ribosome (brown cartoon (Figure 10). They mimic tRNA and release the nascent polypeptide via ribosome subunit disassembly.[80] Class 2 release factors (RF3) are GTPases that enhance the activity of class 1 release factors.[80] In the process of release factor (RF1 or RF2) entry into the A site to promote ribosome dissociation, water is also forced into the site. The presence of water enables hydrolysis of the final tRNA off the nascent peptide chain (again via the peptidyl transferase enzyme within the ribosome). This releases the peptide with a free carboxyl C-terminus leaving via the exit tunnel of the ribosome (E site) and completing the translation.[77]

1.4.5.2 Ribosomal translation of puromycin ligated mRNA – Genotype-phenotype linkage:

In the case of mRNA display, upon reaching the stop codon, i.e., completion of the nascent peptide chain, the newly ligated DNA-PEG-Puro linker is reached (Figure 13). As an *in vitro* purified translation component system is used, the RF1, 2 and 3 are omitted from the mixture. As mentioned, the RFs control the release of the ribosome from the mRNA and as such with their omission, ribosomal stalling results. This provides time for puromycin, at the 3'-terminus of the flexible DNA-PEG-Puro linker, to enter the ribosome A site. The conjugated puromycin mimics the 3' adenosine of a tRNA charged with a modified tyrosine (Figure 13(a/b)).[81] Puromycin then forms an amide bond with the nascent peptide chain from the *O*-methylated tyrosine amino group. The critical aspect of puromycin however, is that whilst typical tRNAs have an ester bond between the amino acid and their cognate tRNA, puromycin has an amide bond between the *O*-methylated tyrosine and the adenosine-tRNA mimetic (the nucleoside). Importantly, as this amide bond is resistant to hydrolysis, the nascent peptide chain is covalently attached to the DNA-PEG puromycin linker. The nascent peptide-puromycin ligated mRNA is finally released upon the addition of EDTA, which chelates ribosomally-bound magnesium resulting in ribosome dissociation (Figure 13(b)).

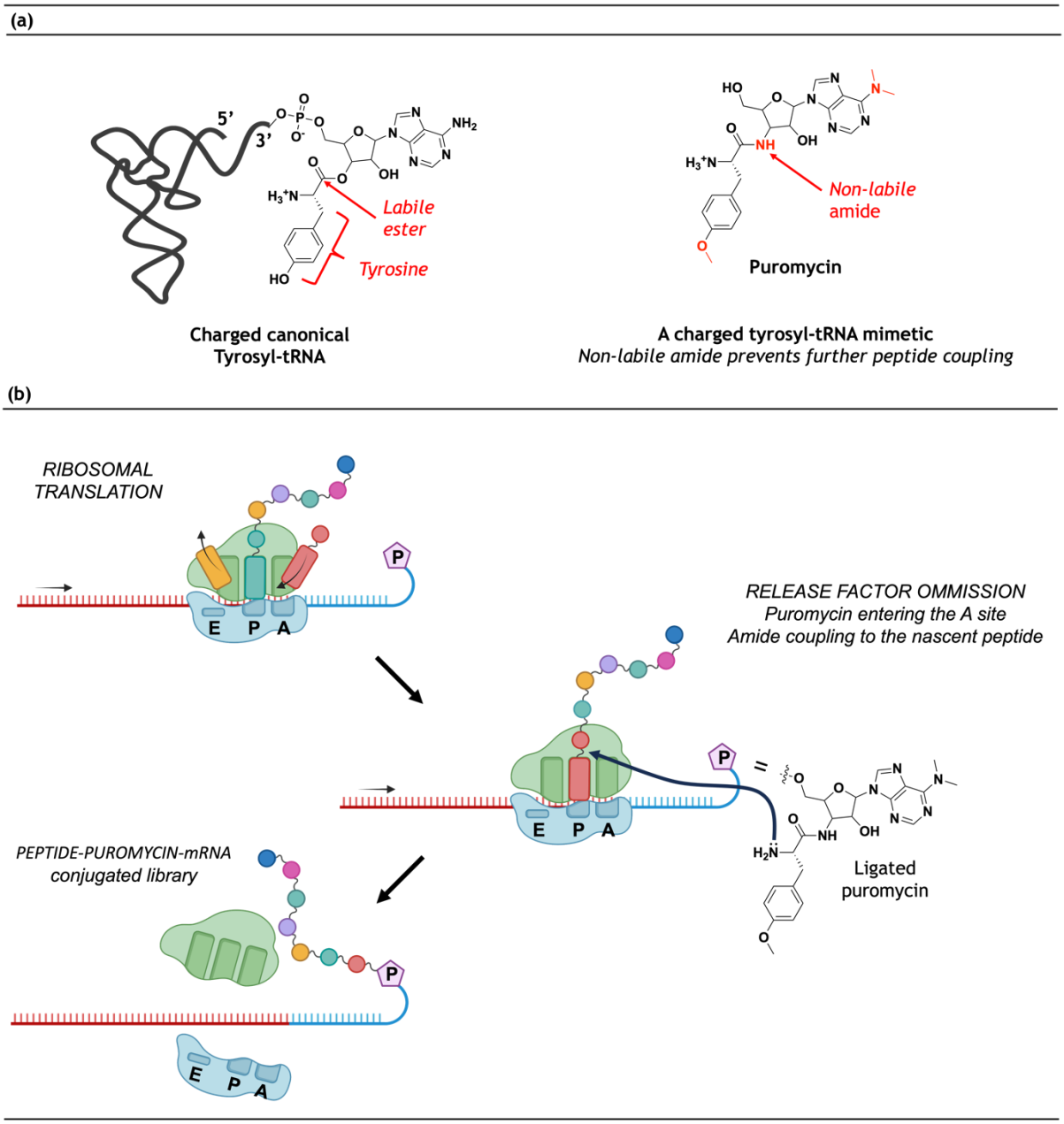


Figure 13: The use of Puromycin-linked mRNA to generate peptide-mRNA conjugates during ribosomal translation. (a) A charged canonical tyrosyl-tRNA (left) - Note the labile ester linking the amino acid and the adenosine nucleoside of the CCA tRNA tail. This ester carbonyl is critical to allow for amide bond formation of the next amino acid coupling. Puromycin (right) - A tyrosyl mimetic aminonucleoside comprising an O-methylated tyrosine, ribose sugar and adenosine. Note the non-labile amide bond, critical to prevent peptide chain elongation, or in the case of mRNA display to provide a covalent link between the peptide and its encoding mRNA. (b) Ribosomal translation of Puro-DNA ligated mRNA: As a reconstituted translation system is used, release factors are omitted so the puromycin aminonucleoside can form an amide bond with the nascent peptide chain. This forms a covalent link between the peptide and its encoding mRNA.

With the peptide translated and conjugated to its encoding mRNA (genotype-phenotype linkage), the mRNA is then reverse-transcribed to an mRNA/ DNA hybrid. This is to make the RNA more resistant to degradation (RNases are ubiquitous), but primarily to ensure that PCR amplification of the DNA recovered post-selection can occur. PCR amplification ensures that sufficient material is recovered for repeated rounds of selection and enables high-throughput sequencing techniques (e.g., NGS) to deconvolute the DNA to identify the binding peptide sequences. With this undertaken, the cyclic peptide-mRNA/ DNA hybrid library is generated ready for affinity panning against a protein target of interest (Figure 7(c)). Affinity panning, also known as biopanning, is a process which isolates binders against a specific immobilised target(s), using a compound library of interest, washing away non-binding or weak affinity interactions. Following affinity panning, the binders are recovered and amplified via PCR due to the prior reverse transcription step. At this point a single cycle is complete. In most cases, the cycle is repeated several times using the recovered peptide-binding DNA from the previous round as the input library for the next round. This promotes the competitive enrichment of binding sequences which drives the recovery of tightly binding hits post-selection.

1.4.6 Cyclisation methods for mRNA display and the other *in vitro* techniques:

mRNA display is traditionally undertaken using linear peptides; however, methods of cyclisation are also routinely implemented. It is important to note that these cyclisation methods are also applicable to ribosomal display. Cyclisation can occur via N- to C-terminal linkages, terminal or side chain linkages or disulfide bridges.[82] Most commonly, bis-electrophile alkylation is used to cyclise, reacting the bis-electrophile with two nucleophilic amino acid residues within the linear peptide. Within bis-electrophile alkylation, two cysteine residues are most usually bridged. Depending on the design of the initial DNA library, the cysteine residues may be fixed at one or both positions generating either a library of macrocyclic peptides with a single or randomised ring size.[82]

1.5 Incorporating ns-AA into ribosomally translated peptides:

Although the power of mRNA display has been highlighted thus far in this introduction, one major limitation exists. ns-AAs cannot not be integrated within the encoded peptides when using the original PURE translation system. To achieve this, Suga and co-workers developed a method, independent of *in vitro* display technologies, to integrate ns-AAs within ribosomally translated peptides/ proteins - the flexible *in vitro* translation (FIT) system.

1.5.1 The flexible *in vitro* translation system (FIT System):

The FIT system (Figure 14), developed by Suga and co-workers,[83] is a customised reconstituted *in vitro* translation (IVT) system combining bacterial ribosomal translation machinery with Flexizyme technology. The FIT system was first implemented to achieve the ribosomal synthesis of mRNA-dependent polyesters in 2007.[84] It is based on initial work modifying the PURE system (initially termed by Suga as wPURE (withdrawn PURE)) to allow for aggressive ns-AA reprogramming, something not envisioned in the base PURE system.[83]

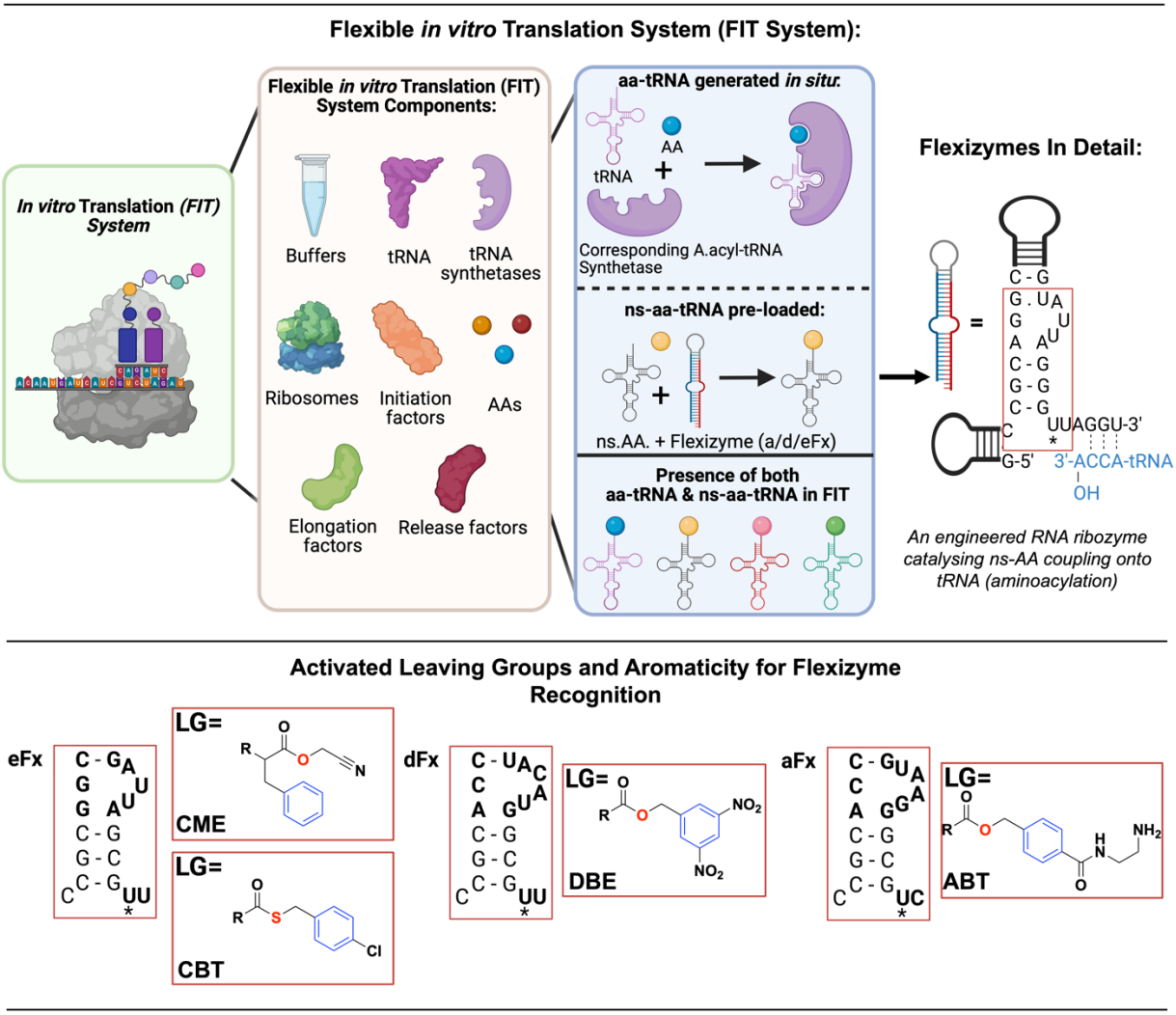


Figure 14: Flexible in vitro Translation System (FIT): A customised reconstituted in vitro translation system that enables ribosomal translation of peptides containing both canonical and ns-AA (non-standard) amino acids. Translation components are individually added as required with AAs and/or their corresponding tRNA and synthetase removed, depending on the genetic code reprogramming required. Canonical amino acids are loaded onto their cognate tRNAs in situ during translation using the endogenous aminoacyl-tRNA synthetase approach. ns-AA loading is achieved using Flexizyme technology which requires recognition motifs and activated leaving groups corresponding to the Flexizyme derivative used. Red ester/thioesters indicate activated leaving groups and blue phenyl rings indicate aromaticity recognition motifs.

Components of the bacterial ribosomal translation machinery include ribosomes, initiation factors, elongation factors, release factors, tRNAs, AA, and aminoacyl tRNA synthetases .[84–90] Due to the re-constituted nature of the machinery, some of the components such as aminoacyl tRNA synthetases, AA, methionyl-tRNA formyl transferase, 10-formyl-tetrahydrofolate, and/or release factors can be included or excluded depending on the desired codons to be reprogrammed or desired application.[91] Furthermore, other enzymes, (e.g., peptide deformylase, methionine aminopeptidase) can be added to the FIT system to enable post-translational modification of the peptides generated.[92] Each customised FIT is supplemented with the required aa-tRNAs and ns-aa-tRNAs, via two methods. The canonical aa-tRNAs are generated in an *in situ* manner, by including the unique cognate aminoacyl tRNA synthetase to that AA within the FIT mix. In contrast, as mentioned, the ns-aa-tRNAs are prepared by Flexizyme technology, occupying the vacant codons generated via genetic code reprogramming. Both methods to generate the required aa-tRNAs (aminoacyl syntheses) and ns-aa-tRNAs (Flexizyme technology) are discussed below.

(1) Aminoacyl tRNA synthetases – Generation of loaded canonical aa-tRNAs for FIT:

Generating aa-tRNA with a cognate canonical amino acid is achieved through the method used endogenously, via aminoacyl-tRNA synthetases. For each canonical amino acid, the corresponding tRNA and aminoacyl-tRNA synthetase is present in the translation mixture.[77] The synthetase loads the amino acid of interest via a two-step mechanism. (i) ATP hydrolysis at the synthetase esterifies the amino acid with adenosine monophosphate (AMP), generating the corresponding adenylated amino acid. Without leaving the synthetase the adenylated-amino acid is then (ii) transesterified to the 3'-hydroxyl group of the ribose sugar of the CCA tail of the cognate tRNA also bound to the synthetase (Figure 15).[77]

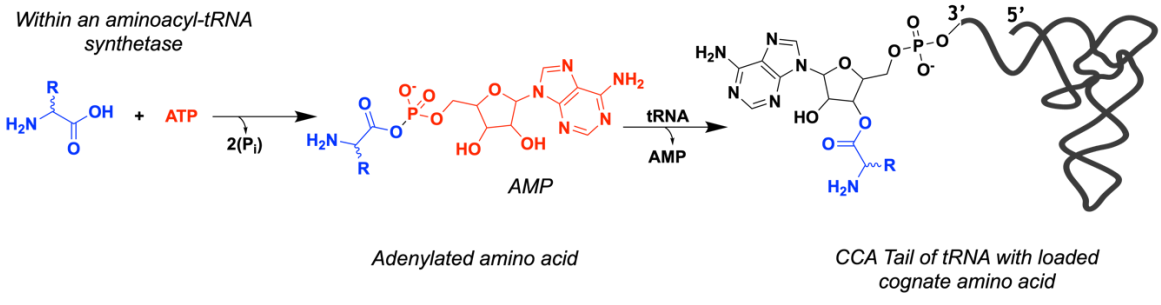


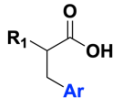
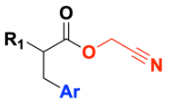
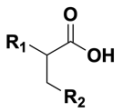
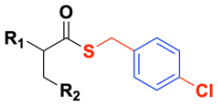
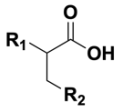
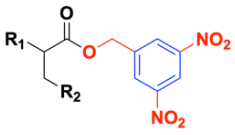
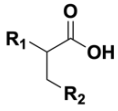
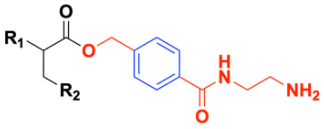
Figure 15: Loading of canonical amino acids by their corresponding aminoacyl-tRNA synthetase via an adenylated intermediate.

(2) Flexizyme technology – Generation of loaded ns-aa-tRNAs for FIT:

Non-standard amino acids can be loaded onto any tRNA of interest using Flexizyme (Fx) technology (Figure 14). Flexizymes are short RNA hairpins and are a type of ribozyme, a form of RNA with enzymatic activity. Flexizymes catalyse the aminoacylation of an activated ns-AA onto a chosen tRNA of interest. To achieve loading, aromaticity recognition motifs are required within the ns-AA. The aromaticity within the ns-AA interacts with the engineered Fx, placing the ns-AA in proximity to the annealed tRNA CCA tail. Activated leaving groups ($_{Act.}$LG) are also required within the ns-AA scaffold. This is to generate the intermediate ester bond between ns-AA and the tRNA which is thermodynamically unfavourable (prone to hydrolysis in aqueous environments). Therefore, to promote this aminoacylation onto tRNA, a more thermodynamically unstable $_{Act.}$LG (esters, thioesters, containing electron-withdrawing substituents) must be used for displacement. More specifically, ns-AA loading occurs via the transesterification of the ns-AA-$_{Act.}$LG upon the 3'-hydroxyl group of the tRNA CCA tail, mirroring the product of standard tRNA aminoacylation. As mentioned, any tRNA can be recognised, this is because only the conserved ACCA tail is recognised for complementary annealing by Fx (Figure 14 – *tRNA annealing in blue*).

Enhanced Flexizyme (eFx), was the first of the contemporary Fx designed beyond initial prototypes (Fx3).[93,94] eFx requires intrinsic aromaticity within the ns-AA and requires an activated cyanomethyl ester (CME). As mentioned, the aromaticity acts as a recognition point for the Fx (Figure 14 – *Red boxes, blue chemical structures*), while the activated CME promotes the aminoacylation thermodynamically. Further expansion to load ns-AA with non-aromatic side chains was then achieved by inducing subtle changes in the active site of Fx3. This led to the identification of the aromatic $_{Act.}$LG chloro-benzyl mercaptan thioester (CBT), which is also accepted by eFx, and a novel pair, dFx (dinitrobenzyl ester (DBE) accepting Fx) with the $_{Act.}$LG DBE.[90] With these successes, the loading of ns-AAs with much greater structural diversity, particularly those with hydrophobic side chains was desired. This was achieved by further engineering Fx3 to generate aFx (amino-derived benzyl thioester (ABT) accepting Fx) which recognises the $_{Act.}$LG ABT.[95] It is typical upon identifying the loading conditions of the ns-AA upon the tRNA of interest, to confirm that it is also suitable for ribosomal translation. As many ns-AA have now been loaded, Goto et al.,[91] grouped the previously loaded ns-AAs by their general physiochemical properties to aid the loading of novel ns-AAs. These general ns-AA characteristics include an absence of aromatic side chains, bulky side chains, exotic side chains and poor water solubility. Lee et al., also developed design rules for Flexizyme-catalysed aminoacylation.[96] Novel ns-AAs are therefore advised to be loaded as per their findings (Table 1).[91,96]

Table 1: A summary of ns-AAs with various structural characteristics each with an associated activated ester for recognition to a corresponding Fx derivative engineered to accommodate these characteristics. Red indicates the activated ester or thioester used to promote the aminoacylation (transesterification reaction). Blue indicates the aromaticity motif used for recognition with the Fx derivatives. Adapted from Goto et al..[91]

ns-AA	Characteristics of ns-AA	Compatible activated ester/ thioester	Compatible Flexizyme
	R_1 = NH_2, RCNH, RNH, H-$(AA)_n$-NH/OH Ar = Various aromatic side chains	 **CME** *(cyano methyl ester)*	eFx
	R_1 = RCONH, H-$(AA)_n$-NH, NH_2, RNH, OH R_2 = Various side chains (especially bulky ones)	 **CBT** *(chloro-benzyl mercaptan thioester)*	eFx
	R_1 = NH_2, RNH, OH R_2 = Various side chains	 **DBE** *(Dinitrobenzyl ester)*	dFx
	R_1 = RCONH, H-$(AA)_n$-NH, NH_2, RNH, -NH, OH R_2 = Various side chains (especially hydrophobic ones)	 **ABT** *(Amino-derived benzyl thioester)*	aFx

From this foundation, the general consensus is that if poor ns-AA loading is seen using the Flexizyme technology, most commonly changing the activated leaving group is the usual first course of action. Following this, changing the position of the aromatic recognition moiety relative to the activated leaving group is also often trialled. This is often undertaken as a last resort as core structural changes are required in the desired ns-AA to be incorporated.

Furthermore, if poor translation efficiency within the ribosome but good Flexizyme loading is seen, modification of the tRNA stem may be required. Suga and Iwane et al.,[97] developed modified T-stems (1-4) of the synthetic tRNA stem $^{AsnE2}tRNA_{NNN}$, where the engineered tRNA stems have been designed to enhance affinity to the elongation factor thermal unstable (EF-Tu). EF-Tu mediates the accommodation of the aminoacyl-tRNAs into the ribosome. As the designer tRNA stems have been designed to have enhanced affinity to EF-Tu, the increased affinity has been shown to promote increased ribosomal accommodation, entry, and successful incorporation of the ns-AA into the desired nascent peptide chain. This is because ns-aa-tRNA typically have weaker EF-Tu affinity compared to proteogenic aa-tRNAs.[97] The approach, however, is only compatible with elongation reprogramming

as initiator tRNA is the only tRNA to bind directly to the P site of the ribosome during the translational cycle. It is also one of the only tRNAs that must avoid binding to elongation factor Tu as a result.[98]

1.5.2 The FIT system under genetic code reprogramming:

With the ability to generate ns-aa-tRNAs, the FIT system requires customisation under genetic code reprogramming, depending on what specific tRNA is loaded with the ns-AA. This customisation is highly specific to each FIT undertaken to achieve the desired incorporated ns-AA application. Genetic code reprogramming is the removal of a cognate aa-tRNA from the translation system, replacing this free codon with the corresponding ns-aa-tRNA of interest, generated via the Flexizyme technology just described (Figure 16).[91] This is easily achieved due to the reconstituted nature of the FIT system, as specific components can be added or removed at will. In addition, multiple codons can be reprogrammed at once, leading to libraries with multiple ns-AA substitutions.

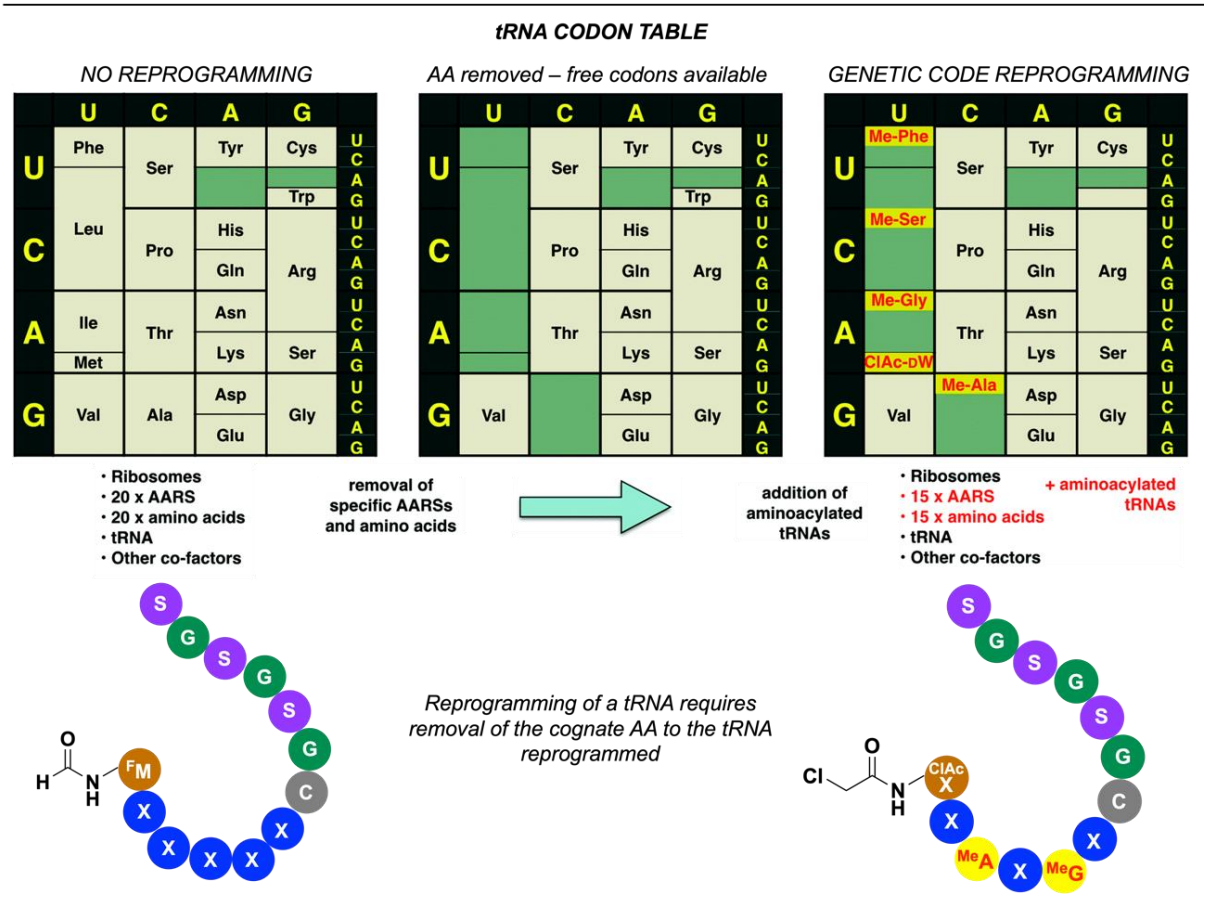


Figure 16: Genetic code reprogramming via FIT. The canonical codon table for bacterial ribosomal translation (left) is shown. The first base of the codon is shown on the left side of the table, the second base of the codon is indicated along the top and the third base of the codon is indicated on the right side of the table. Reprogramming requires the removal of a canonical AA 'freeing' the codon position for a reprogrammed ns-AA to take its place. Adapted from Passioura et al. (permission available at Appendix 43).[99]

In a quirk of biology, reprogramming of the Met codon (CAU) enables the reprogramming of two ns-AAs for the 'price' of one amino acid replacement (Met).[77] This is because endogenously, Met has two independent tRNAs, both with the same anticodon (CAU). One of the tNRAs is for Met used in protein translation initiation (initiation position ^{ini}Met-tRNA^{CAU}) whilst the other, is for Met appearing during mid-translation (elongation position ^{En}Met-tRNA^{CAU}). In all other cases, one canonical amino acid must be removed to free one codon for genetic code reprogramming. To practically achieve this, if initiation reprogramming is required, Fx-mediated aminoacylation occurs onto the endogenous ^{ini}tRNA^{CAU} tRNA (which is synthetically made). Synthetic hybrid tRNAs are used if the elongator positions are reprogrammed. These hybrid tRNAs comprise the base of the tRNA stem derived from an *Escherichia coli* (*E.coli*) asparaginyl tRNA$^{Asn}_{NNN}$ or glutaminyl tRNA$^{Glu}_{NNN}$ with the corresponding anticodon triplet (NNN) mutated as required to whichever canonical AA anticodon is being reprogrammed.[91]

1.5.3 Alternative approaches to generate ns-AA incorporation in peptides/ proteins:

Orthogonal aminoacyl-tRNA synthetases (ORSs), are an alternative approach used to incorporate ns-AAs using genetic code expansion and amber stop codon suppression. It is traditionally used to introduce ns-AA within a cellular environment, something not possible with the FIT Flexizyme technology.[100–102] To incorporate novel ns-AAs, the generation of specific ns-AA-ORS pairs is required. They must be highly specific to each other without cross-reacting with other added ns-AAs, cellular AAs, tRNAs, or aa-tRNAs. This is because the ns-AA is loaded onto a tRNA *in situ* by its paired ORS. It is a powerful approach, as ORSs circumvent the need for preactivated ns-AAs, pre-acylation and most prominently, allow for in-cell integration of ns-AA within proteins. However, extensive directed evolution is required to generate novel functional ns-AA ORS pairs. Otherwise only very limited diversification of existing orthogonal ns-AA-ORS pairs is tolerated. Largely, FIT dominates ns-AA incorporation within mRNA display due to its vast scope and easy accessibility of the ns-aa-tRNAs. Nevertheless, use of both these approaches within a single mRNA display system has been presented in the literature. This highlights the orthogonality of these approaches for specific applications.[102]

1.6 The RaPID System – Integrating mRNA display with the FIT system:

The Random nonstandard Peptides Integrated Discovery system (RaPID) was first developed by Hiroaki Suga and co-workers at the University of Tokyo, following the success of the FIT system.[103] RaPID expands on standard mRNA display as it replaces the PURE system, substituting it for the modified FIT system under genetic code reprogramming. This enables the incorporation of ns-AA into the ribosomally translated trillion-member peptide libraries. The first reported example of RaPID was published by Suga and co-workers in 2011,[103] where they aimed to develop natural product-like macrocyclic *N*-methyl-peptide Inhibitors against a Ubiquitin Ligase.

> "Here, we report a new means of synthesizing a *de novo* library of ''natural product-like'' macrocyclic *N*-methyl-peptides using translation machinery under the reprogrammed genetic code, which is coupled with an *in vitro* display technique, referred to as RaPID (random nonstandard peptides integrated discovery) system"- Suga *et al.*[103]

An overview of the RaPID cycle:

As a combination of mRNA display with the FIT system, RaPID shares many common features of the Ribosomal and mRNA *in vitro* display technologies. The RaPID system (Figure 17),[83,103] begins with a semi-randomised DNA library, encoding for $>10^{13}$ cyclic peptides . Critically as ns-AA can be integrated, the design of the peptide sequences defined by the encoding DNA library often reflect this (e.g. ns-AA induced cyclisation, fixed regions etc.) and are described further below.

With the generated library, it is then transcribed via T7 RNA polymerase to generate the required mRNA template for translation. Upon generating the mRNA, the library design allows for annealing and ligation of the DNA-PEG-Puro linker. This puromycin ligation step enables covalent genotype-phenotype linkage of the peptide-puromycin-mRNA, ensuring the genotype-phenotype link necessary for *in vitro* display platforms.

With the puromycin-ligated mRNA in hand, translation via the FIT system can then proceed. The exact composition of the FIT-for-RaPID components differs slightly from stand-alone FIT, e.g., tRNA reprogramming and release factor omission (also described further below). It is important to note that the peptide libraries are almost exclusively generated as macrocyclic peptides due to their advantageous characteristics over linear peptides (Chapter 1.3.1).

With the macrocyclic peptide library translated and conjugated to its encoding mRNA, reverse transcription then follows similarly to mRNA display. The reverse transcription occurs using reverse transcriptase, producing a final macrocyclic peptide library puromycin conjugated to its encoding mRNA/ DNA hybrid ready for affinity panning against any immobilised target of interest.

Following panning, in the same fashion as mRNA display, the binding peptides undergo several rounds of library enrichment through re-amplification (via PCR) and re-selection. This repeats until an enriched library (where multiple copies of multiple sequences are present in the library) is seen. Upon this, the DNA is subjected to high-throughput sequencing (NGS) to deconvolute the enriched binding peptide sequences.

The key differences from mRNA display are the FIT-for-RaPID system, the range ns-AAs that can therefore be introduced, and the DNA library design for RaPID selections which are all discussed in further detail below.

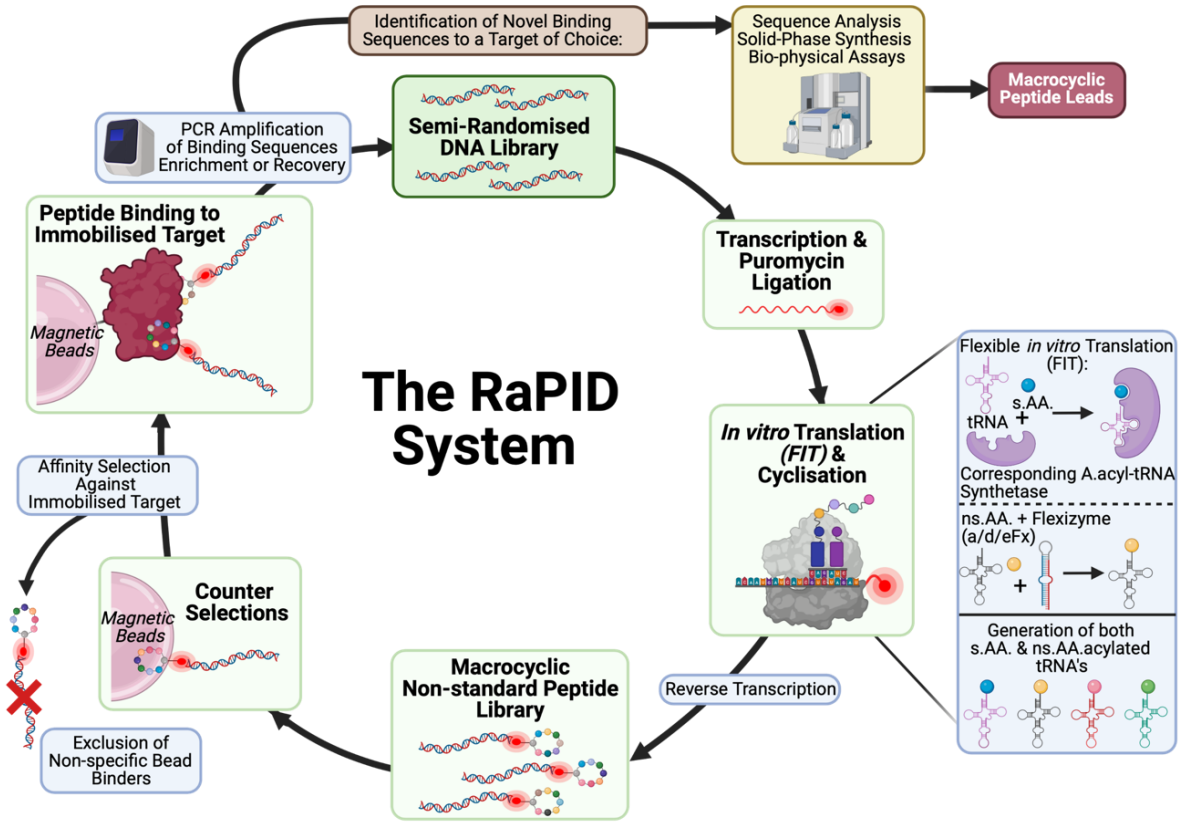


Figure 17: The RaPID System: An expansion on the in vitro mRNA display platform, using the FIT system to enable both canonical and ns-AA to be integrated into the macrocyclic peptide libraries. This is achieved using Flexizyme

technology. Upon generation of the macrocyclic non-standard peptide library, affinity panning against any recombinantly expressed protein of interest can be undertaken. Several repeat rounds of the selection cycle are undertaken to enrich tightly binding peptide sequences. Following sufficient library enrichment, sequence deconvolution is achieved via high-throughput sequencing (e.g. NGS) to identify the exact sequence of the binding hits.

1.6.1 Modified FIT system for compatibility with RaPID – FIT-for-RaPID:

Recall back to the general FIT system (Chapter 1.5.1) and mechanism of puromycin ligation during ribosomal translation (Chapter 1.4.5.2). To generate peptides which are conjugated to their encoding mRNA via puromycin, the release factors must be omitted regardless of whether the peptides contain ns-AA. This is in the same fashion as PURE based mRNA display. Removal of the release factors promotes ribosomal stalling, holding the nascent peptide chain within the P site of the ribosome whilst puromycin enters the A site, conjugating peptide-to-mRNA.

More specifically, the FIT-for-RaPID system undertaken within the Walport lab uses a commercial system (PURExpress – *in vitro Protein Synthesis ΔRF123kit*), as the principal source of the translation components. This kit enables quick access to the reconstituted bacterial translation components and the key aspect of this commercial kit is that the release factors are omitted as standard. It is separated into two mixes Sol A, and Sol B (Figure 18). Critically though, as it is designed for translations without ns-AA incorporation, pre-mixed within the Sol A, are all 20 AAs and cognate tRNAs. As such, homemade Sol A (h.Sol A) is required, replacing the commercial Sol A by separating the buffer components and tRNAs (synthetically made) from the AAs. Therefore, during RaPID, depending on which cognate tRNA has been reprogrammed with the ns-AA of choice, the corresponding cognate AA is removed from the separated AA mix (e.g., Met tRNA reprogrammed, AA mix contains 19 AA -Met).

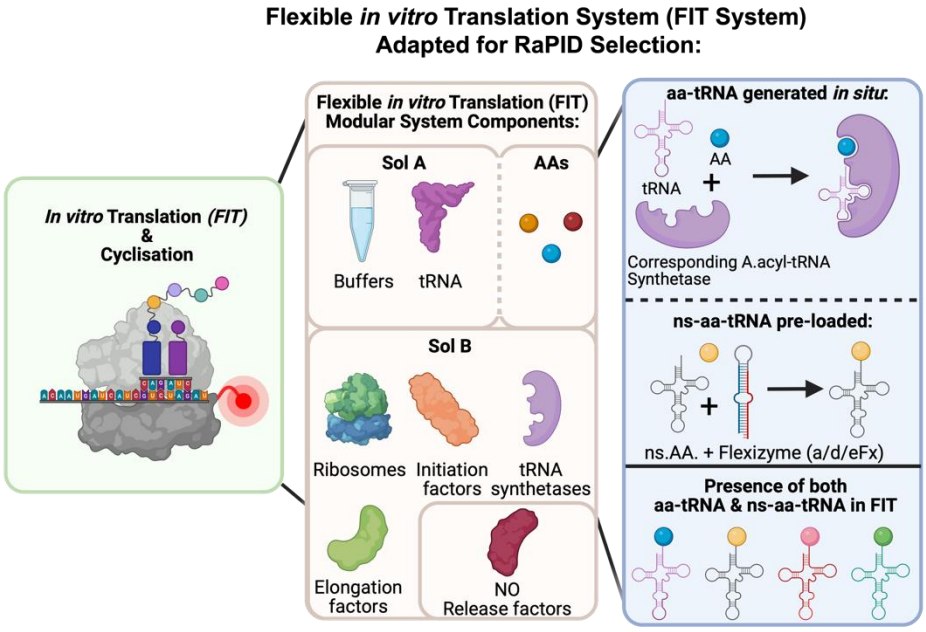


Figure 18: Flexible in vitro Translation System (FIT): A customised reconstituted in vitro translation that enables ribosomal translation of peptides containing both canonical and ns-AA (non-standard) amino acids. Translation components are stored pre-mixed in three components – AA mix is separate (dashed line) in homemade Sol A solutions to enable reprogramming. Release factors are not included for RaPID selections to enable puromycin ligation. Canonical amino acids are loaded onto their cognate tRNAs in situ during translation using the endogenous aminoacyl-tRNA synthetase approach. Ns-AA loading is achieved using Flexizyme technology.

1.6.2 Uses of genetic code reprogramming for FIT RaPID Selections:

A major application of genetic code reprogramming in FIT for RaPID selections is the use of ns-AA to generate cyclic peptide libraries. Reprogramming for cyclisation involves the inclusion of ns-AAs which contain a reactive electrophilic chloroacetyl moiety (ClAc).[86] Cyclisation is achieved via alkylation of the ClAc-ns-AA with a fixed cysteine residue downstream in the translated peptide sequence (Figure 19(a)). These ClAc-ns-AA typically are either chloroacetyl-$^{L/D}$tyrosine (ClAc-Y), chloroacetyl-$^{L/D}$phenylalanine (ClAc-F) or chloroacetyl-$^{L/D}$tryptophan (ClAc-W) in traditional RaPID selections.[86,104] They are generated via Flexizyme catalysis onto the corresponding freed initiator tRNA codon (^{ini}Met-tRNA^{CAU}). Either or both enantiomers of ClAc-Y, ClAc-F or ClAc-W are used (and can be undertaken in parallel selections). This is because the differences in stereochemistry promote opposite ring cyclisation due to opposite chirality. As such distinct peptide libraries are generated depending on which enantiomer for cyclisation is used.

Alternative uses of genetic code reprogramming in RaPID selections aim to promote specific library functionality or to increase chemical diversity (Figure 19(b)). Literature examples of ns-AA incorporated within RaPID cyclic peptide libraries at the initiation position include various ns-AA initiators, *D*-AA, amphotericin-like AA and exotic peptides such as dipeptides or foldamers. Other ns-AA have also been incorporated into elongation positions. These include ns-$^{L}\alpha$-AA with applications for alkylation, photo-crosslinking, or as biotinylated pull-down agents. Others include *N*-methyl or *N*-alkyl AA, ^{D}AA, β-AA as well as hydroxy and amino carbothioic AAs, but this is by no means an exhaustive list.[84–86,89,103,105–112] The ability to incorporate such a wide range of ns-AA within RaPID selections is one of the most powerful aspects of the platform.

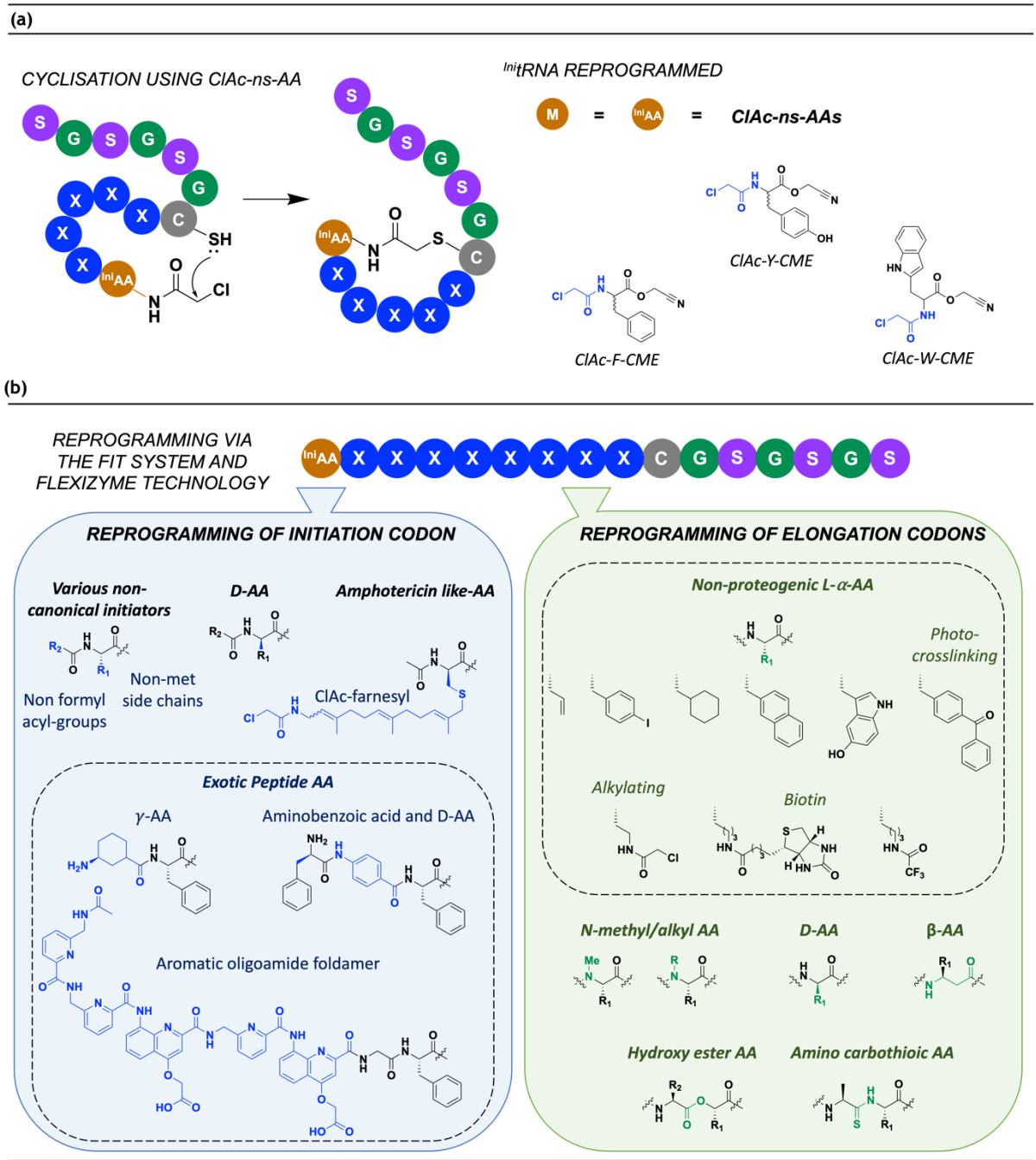


Figure 19. The most commonly used reprogrammed ClAc-ns-AAs to enable RaPID library cyclisation and an overview of the types of ns-AA incorporated within RaPID within the literature. (a) Use of ClAc-ns-AA to promote RaPID library cyclisation via initiator reprogramming to typically either ClAc-Y, ClAc-F or ClAc-W. (b) A non-exhaustive overview of the range of ns-AA that have been incorporated in the literature at both the initiator and elongator positions of peptides within RaPID selection libraries. [84–86,89,103,105–112] *Adapted from Goto et al.*[112]

1.6.3 DNA library design for RaPID selections:

Traditional RaPID DNA peptide libraries are designed such that they contain both fixed and randomised encoding regions. They follow a general feature consensus as described below (Figure 20). [103] The library DNA is generated from commercially bought DNA primers, synthesised via split and pool synthesis, and uses NNK nucleotide restriction for the randomised encoding region.

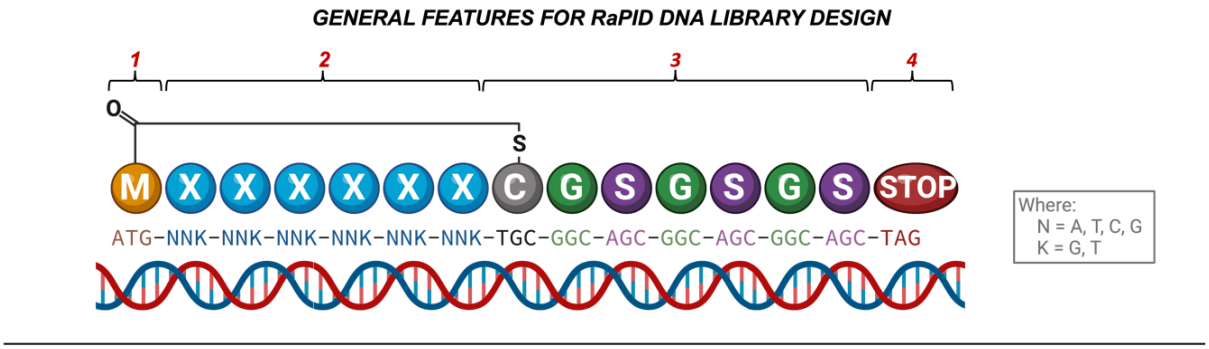


Figure 20: General features of RaPID DNA library design to ensure that macrocyclic peptide libraries are generated via head-to-side chain N-chloroacetyl thioether linkage. Four key features dominate, (1) fixed initiation codon, (2) randomised AA encoding region, (3) fixed CGSGSGS encoding region and (4) translation termination. Randomised AA encoding region is governed by NNK library design to reduce AA codon degeneracy.

(1) Fixed Initiation codon:

As all bacterial translation systems require translation initiation using ^{Ini}Met-tRNA^{CAU}, this codon (DNA = ATG, mRNA = AUG) is therefore also naturally fixed within all translated peptide libraries designed. ^{Ini}Met^{CAU} is the predominant codon reprogrammed in RaPID, via FIT, replacing the translated methionine with the desired ClAc-ns-AA of choice (as described previously).

(2) Randomised AA encoding region:

Following initiation and warhead installation into the peptide, a semi-randomised region of between 4-14 amino acids is then encoded within the library DNA. This is to enable library diversity as these positions can be any of the canonical AA or ns-AAs reprogrammed into a elongation tRNA position. Depending on the application, certain fixed AA regions can also be installed within the randomised region, particularly if further reprogramming is desired or if binding motifs to the target are already known. Randomised AA are encoded using an NNK-library configuration to prevent sequence bias (described in detail below).

(3) Fixed CGSGSGS encoding region:

After this, a short fixed repeating linker of a cysteine followed by three glycine serine repeats (CGSGSGS) is then typically encoded within the library DNA. The initial cysteine is used as the fixed downstream point to cyclise the entire cyclic peptide library with the ClAc-ns-AA of choice. It is also important to note that randomised cysteine residues are also possible as cyclisation points. The GS repeats are included to distance the peptide from the puromycin linker.

(4) Translation termination:

Finally, the translation ends using an amber-stop codon (TAG) encoded within the DNA library.

NNK libraries to reduce sequence bias in RaPID DNA Libraries:

One issue with mRNA display in general, is amino acid bias within the encoding DNA libraries. In the semi-randomised encoding region, minimising this AA bias is critical. A codon is a sequence of three nucleotides (RNA or DNA) which form a unit of genetic information encoding for a specific amino acid or signalling the termination of protein synthesis (stop codon).[77,113] There are 20 canonical amino acids but due to the three-letter nature of codons and the four possible nucleotides (mRNA: A, U, C, G/ DNA: A, T, C, G), there are 64 different codon combinations. To prevent over two-thirds of the codon combinations from being non-functional (i.e., non-coding for an amino acid) within endogenous mRNA translation, the phenomenon of codon redundancy (codon degeneracy) is seen during translation.[113] This means that multiple codons can encode a single amino acid (Figure 21).

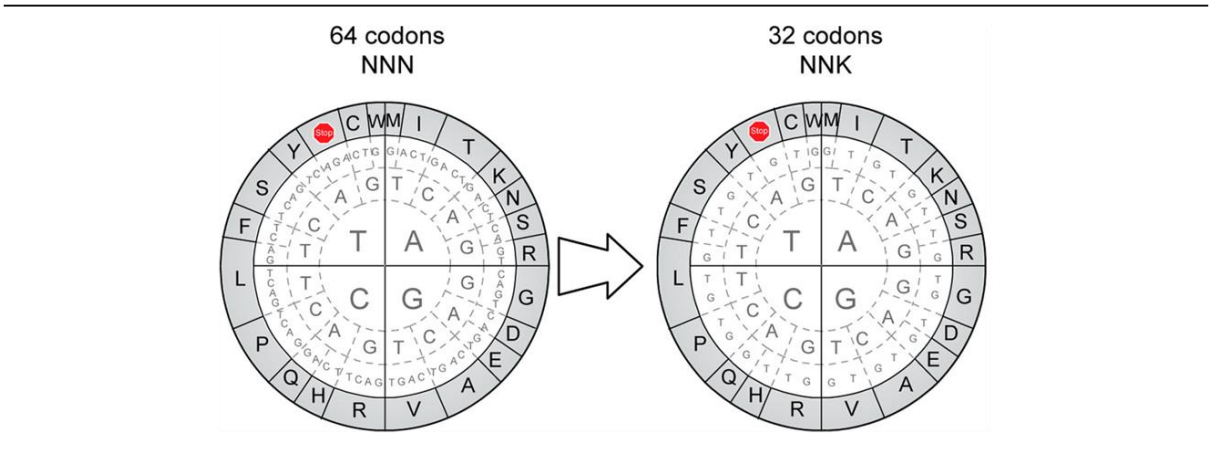


Figure 21: A DNA codon wheel indicating the triplet nucleotide sequences encoding for a single proteogenic AA. Read from inside to out, with the corresponding AA encoded shown on the outermost ring. The codon degeneracy seen in the final nucleotide position (wobble position) enables multiple codons to encode for a single AA. Introducing NNK codon restrictions during RaPID mRNA DNA library synthesis, significantly, but not completely

reduces, this codon degeneracy. This reduces AA bias within mRNA display libraries. NNK = where N = A, T, C or G, whilst K is restricted to only G or T. Adapted from Kille et al (permission available at Appendix 43).[113]

When designing peptide libraries, AA degeneracy can lead to biases in the probability that a specific amino acid will appear in the library which is undesirable. 'NNK' site-directed codon saturation mutagenesis is an approach to mitigate this by restricting the 'randomness' of the final codon nucleotide during split-and-pool synthesis to reduce degeneracy. As such in RaPID DNA libraries the codon is represented by NNK with N being any nucleotide (A, T, C, G), and K being only (T, G).[113] This reduces the codon redundancy by reducing the number of possible codons from 64 to 32 and as such, reduces the number of codons which encode for each AA. [113] It is important to state that the NNK approach does not fully prevent codon degeneracy (NNK degeneracy still contains three codons for arginine, leucine, and serine and two codons for the five amino acids alanine, glycine, proline, threonine, and valine). However, due to practical considerations,[113] NNK remains the preferred method to reduce amino acid bias in DNA libraries and nucleotide restraints are considered too restrictive.

1.7 Current methods of developing novel light-responsive ligands and their limitations:

As highlighted at the beginning of this introduction, due to the high spatial and temporal control that photopharmacology imparts, it has become an ideal candidate to interrogate particularly complex or dynamic biological systems. Widespread application of light-responsive chemical biology, however, has been limited. This is because the incorporation of molecular photoswitches within existing chemical tools has proved challenging. Particularly as methods to identify novel photoswitchable ligands are limited.[114]

The most common method of developing novel light-responsive ligands is via azologisation (Figure 22).[115,116] Azologisation involves the retrofitting of photoswitchable moieties within existing bioactive ligand scaffolds. It typically involves the replacement of one part of the core scaffold with the azo moiety. Another, but similar method, is called azo-extension (Figure 22),[116–118] which aims to append the photoswitch off one end of the bioactive ligand without partial scaffold replacement. Both approaches have led to the successful development of light-responsive ligands. An example of the azologisation approach is the development of Azo-fomocaine, a photochromic channel blocker

providing light-responsive anaesthetic effects.[115] For the azo-extension, the successful development of azo-propofols, photochromic potentiators of GABA(A) receptors are also reported.[118] However, for most applications, both azologified or azo-extended derivatives often adversely impact both the pharmacodynamic/ pharmacokinetic properties of the original lead (e.g., potency) and/or the photochemical properties of the switch itself (e.g., unpredictable changes in *E*/*Z* switching efficiency and/ or $\tau_{1/2}$ acceleration). Due to an often-limited number of photoswitch integration sites within or off the original scaffold, the azologified or azo-extended tool compounds also often face similarities in biological activity between the isomeric forms, limiting the degree of spatial and temporal control that can be achieved against a biological target.

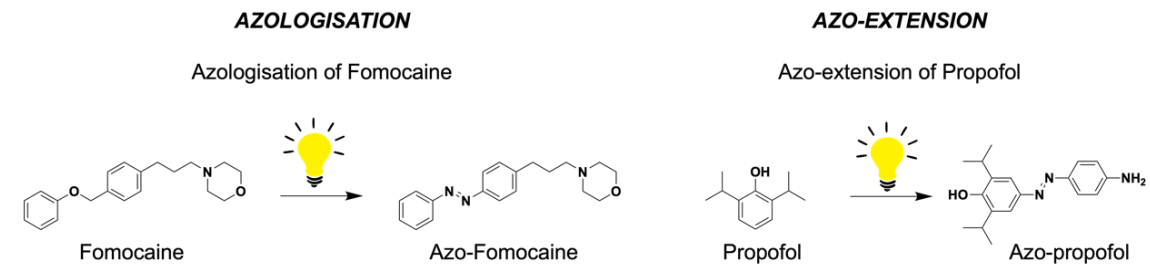


Figure 22: Examples of azologisation and azo-extension approaches to develop novel light-responsive ligands. Fomocaine is an example of successful azologisation whilst Azo-propofol is an example of successful azo-extension, both developed by Trauner and co-workers.[115,118]

Both azologification and azo-extension, however, rely on pre-existing ligands to generate the light-responsive tools. The ability to identify *de novo* photopharmacological tool compounds is less frequently reported in the literature. The main example of *de novo* identification involves a rational fragment-based approach.[119] This aims to build target specificity around a core but generic photoswitch scaffold. Some success has been seen with this approach, namely the work of Feringa and co-workers, on the structure-based design of light-regulated eDHFR inhibitors.[119] For more general applicability, however, this approach has also proved very challenging. This is because the photoswitch (a largely flat hydrophobic surface) is not primed for substantial protein-ligand interactions to begin the fragment-based process. Achieving suitable target potency and specificity against novel targets has been especially difficult.

Therefore, methods to identify *de novo* light-responsive chemical biology tools are greatly needed. To achieve this, libraries of photoswitchable compounds are required. Such libraries would ideally be

large, easily generated and have a photoswitch directly integrated during the affinity enrichment process. This is to prevent the need for retrofitting, whilst promoting potency, specificity, and light-responsive differential binding for any target of interest.

1.8 Current literature – Evolution of *in vitro* light-responsive peptide display platforms to identify *de novo* photopharmacological tools:

Recognising the power of integrating these two areas, three examples of evolving *de novo* peptide-based light-responsive ligands using *in vitro* discovery platforms have been recorded in the literature. These are the works of Ito and co-workers, Derda and co-workers, and Heinis and co-workers.[120–122]

Ito and co-workers – *In vitro* selection of a photo-responsive peptide aptamer:

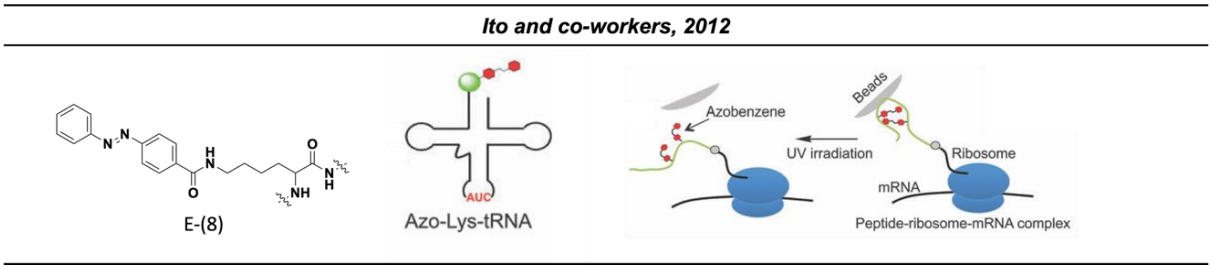


Figure 23: Ito and co-workers, demonstrated the first proof-of-principle to evolve a photo-responsive linear peptide aptamer using in vitro Ribosomal display and amber suppression reprogramming. Adapted from Liu et al.[120]

Ito and co-workers,[120] in 2012, demonstrated the first proof-of-principle to evolve a photo-responsive linear 'peptide aptamer' using *in vitro* Ribosomal display and amber (UAG) suppression reprogramming (Figure 23). A ε-(azobenzoyl)-lysine (8) was ligated with truncated tRNA carrying the amber anticodon (AUC). With this reprogrammed stop codon, the semi-randomised DNA aptamer library was then translated using a recombinant IVT system (PURE system). The *E*-aptamer peptide library generated was affinity panned (*E*-isomer active binders) against immobilised streptavidin as a model target. The best binding aptamer isolated (LA81), had a modest ${}_{E}K_{D}$ of 6.31 µM. Unfortunately, differential light-responsive binding could not be identified for LA81. This was because only <30% photoconversion efficiency (PSS at 365 nm) for the *E*→*Z* isomerisation was seen. This was in comparison to ε-(azobenzoyl)-lysine alone, which demonstrated a typical *E*→*Z* azobenzene isomerisation efficiency of 70-80% (PSS at 365 nm). The poor photoconversion efficiency for LA81 was rationalised by the authors, that the isomerization of azobenzene was more difficult when incorporated into the aptamer due to the stereo-inhibition of adjacent amino acids.

Derda and co-workers – Light responsive ligands through screening of a Light responsive genetically encoded library:

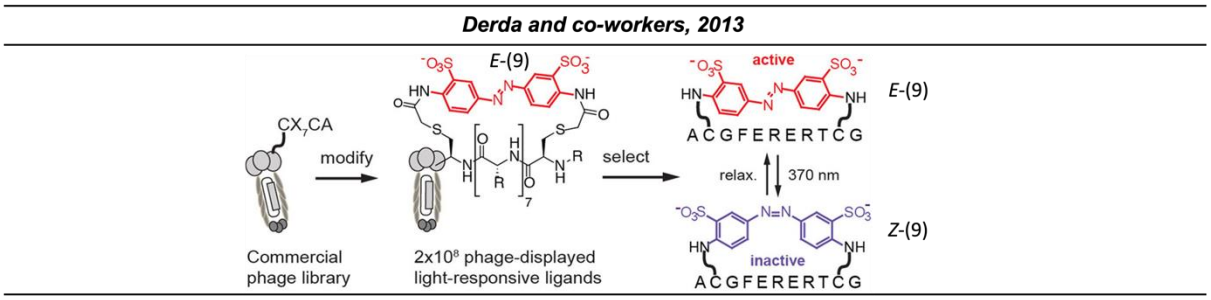


Figure 24: Derda and co-workers, aimed to generate light-responsive peptide ligands using a phage display and photoswitchable lynchpin approach. E-isomer active binders were developed. Adapted from Jafari et al.[121]

Following on from Ito and co-workers, Derda and co-workers,[121] in 2013, sought to generate Light responsive ligands using phage display (Figure 24). A lynchpin strategy was implemented to integrate a photoswitchable moiety using 3,3'-bis(sulfonato)-4,4'-bis(chloroacetamido)azobenzene (9) to cyclise the phage peptide library. Using the same model target as Ito and co-workers, the generated light-responsive phage library was panned against immobilised streptavidin, and again, *E*-isomer active binders were preferentially enriched. The best binder identified (L36), had a $_{E}K_{D}$ of 452 ± 63 µM and >4.5-fold weaker binding ($_{Z}K_{D}$ = >2000 µM) when irradiated to the *Z*-isomer. Once again as an azobenzene derivative was used, it displayed incomplete photoconversion (85% PSS at 370 nm) for the *E*-to-*Z*-isomer isomerisation.

Heinis and co-workers – Phage selection of photoswitchable peptide ligands:

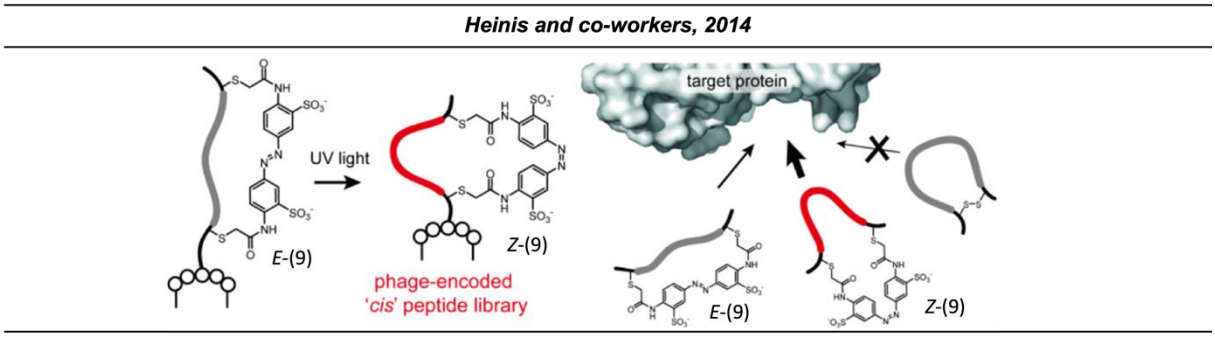


Figure 25: Heinis and co-workers, generated a Light responsive phage display platform with Z-active peptide screening using the lynchpin derivative similar to Derda and co-workers.. Adapted from Belotto et al.[122]

Finally, the work of Heinis and co-workers in 2014, completes the current literature in this area. Similarly, to Derda and co-workers, Heinis and co-workers,[122] sought to generate Light responsive ligands using phage display via a lynchpin strategy against immobilised streptavidin as the model target (Figure 25). The photoswitchable moiety (once integrated) was also the same 3,3'-bis(sulfonato)-4,4'-bis(acetamido) azobenzene (9), although differing electrophiles were used (dichloro, instead of Ito and co-workers who used dibromo electrophiles) to promote the bridging. In another divergence from Derda and co-workers, Heinis and co-workers sought to develop a strategy to evolve peptide ligands that are activated by light (*Z*-active) rather than inactivated (*E*-active). The advantage of a Z-active binding peptide strategy is that ~99.99% of the base peptide population would remain in the in-active *E*-isomer before light irradiation (i.e., the identified peptide hits would not be constitutively active against streptavidin upon dosing).

Initial hits identified found that the central loop region (C-X_7-C) was not disturbed upon photoswitch irradiation and as such, the peptide found did not demonstrate switchable binding. This may be an intrinsic limitation to integrating the photoswitch as a lynchpin, due to its distance away from the binding region and as flexible regions are required to ensure it can act as a bridging linker. Using smaller cyclic peptides (C-X_5-C), produced their best hit (SA149), which had a ${}_{Z}K_D$ of 2.2 ± 0 μM and >3-fold weaker binding (${}_{E}K_D$ = 6.7 ± 2 μM) when irradiated to the *E*-isomer. In agreement with the previous work, as Heinis and co-workers used the same azobenzene derivative as Derda and co-workers and the same azobenzoyl core as Ito and co-workers, their peptides also displayed incomplete photoconversion (~60% *Z*-isomer PSS at 365 nm) for *E*-to-*Z*-isomerisation and a $\tau_{1/2}$ of 130 mins. The implications for this on the *Z*-active binding strategy, however, were not as severe as Derda and co-workers's *E*-active binding strategy. This is because for Heinis and co-workers, as the peptide ligands are constitutively inactive, following incomplete *E*→*Z* isomerisation (to activate) a small population of inactivated peptide ligands would remain. This is less of an issue than for the constitutively active *E*-binding peptides of Derda and co-workers, and Ito and co-workers, where following incomplete *E*→*Z* isomerisation (to deactivate) the active tool ligand could not be fully 'switched-off'.

Other notable literature:

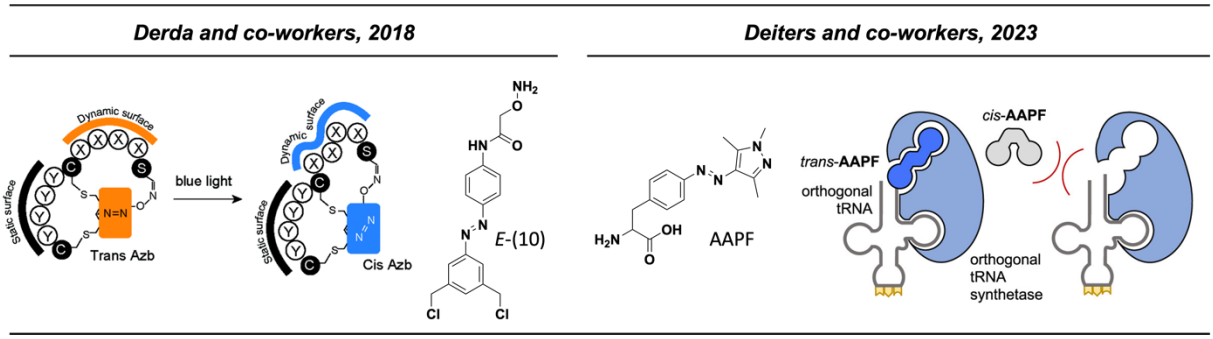


Figure 26: The work of Derda and co-workers, (left) and Dieters et al., (right). Derda and co-workers sought to develop a novel light-responsive bicyclic lynchpin strategy. Deiter and co-workers used an azopyrazole-based ns-AA to enable light-responsive ribosomal translation through the light-responsive control of ns-AA aminoacylation via an orthogonal tRNA synthetase.

Other notable literature includes further work from Derda and co-workers in 2018, on the generation of light-responsive bicyclic peptides (Figure 26).[123] This was through the development of a bicyclic photoswitchable lynchpin (10). The rationale for this work relates to the added advantages of bicyclic topologies, such as improved binding and stability when compared to monocyclic peptides. The paper primarily describes an accessible synthetic route to generate such bicyclic lynchpins but was not included as part of a light-responsive peptide display platform. However, as the photoswitch scaffold was based upon an azobenzene derivative, this lynchpin still shows incomplete photoconversion (71% *E*→*Z* PSS at 365nm). Interestingly, the photoswitch thermal relaxation appeared extended when introduced within a peptide (switch alone $\tau_{1/2}$ = 43 mins, integrated with the peptide $\tau_{1/2}$ = <4 hrs) which was attributed to *Z*-isomer ring strain causing the slower reversion.

Furthermore, Deiter and co-workers most recently in 2023,[124] developed the first example of a genetically encoded azopyrazole-based Phenylalanine (arylazopyrazole-modified phenylalanine (AAPF) to impart optical control over translation (Figure 26).[124] Complete selectivity for the *E*-isomer photoswitch by the engineered pyrrolysyl tRNA synthetase was developed. This demonstrated the light-responsive expression of a gene of interest selectively controlled via light exposure. Interestingly, due to the near quantitative nature of the azopyrazole-based derivative switching (98% bi-directional switching at PSS 355 nm and 532 nm, respectively), complete elimination of any observed background mRNA translation response was seen when fully switching to the *Z*-isomer (*Z*-conformation blocks entry to the synthetase). This contrasts with the strong mRNA translation response seen when irradiated to the active *E*-isomer. The work of Deiter and co-workers, although unrelated to the development of *de novo in vitro* light-responsive cyclic peptide discovery platforms, demonstrates the

utility of complete quantitative bi-directional light-responsive switching for photopharmacological and chemical biology applications.

1.9 Aims of this PhD:

1.9.1 Development of a next-generation *in vitro* light-responsive peptide display platform (PhotoRaPID):

From the current literature, precedence was set for the integration of photopharmacology with *in vitro* light-responsive peptide display platforms. However, limitations such as modest target potency and incomplete photoswitching using an azobenzene scaffold limited these approaches. As such this PhD work aimed to develop a next-generation photoswitchable macrocyclic peptide discovery platform (Figure 27). The project focused on the direct integration of an azopyrazole-based photoswitch within every macrocycle of a trillion-member RaPID library, via FIT genetic code re-programming. Workflows were developed to promote the evolution and rational discovery of cyclic peptides with light-dependent global conformational change of the macrocyclic peptides structure. This produced a next-generation platform to identify *de novo* tool compounds with potent, switchable binding against any recombinantly expressed protein of interest.

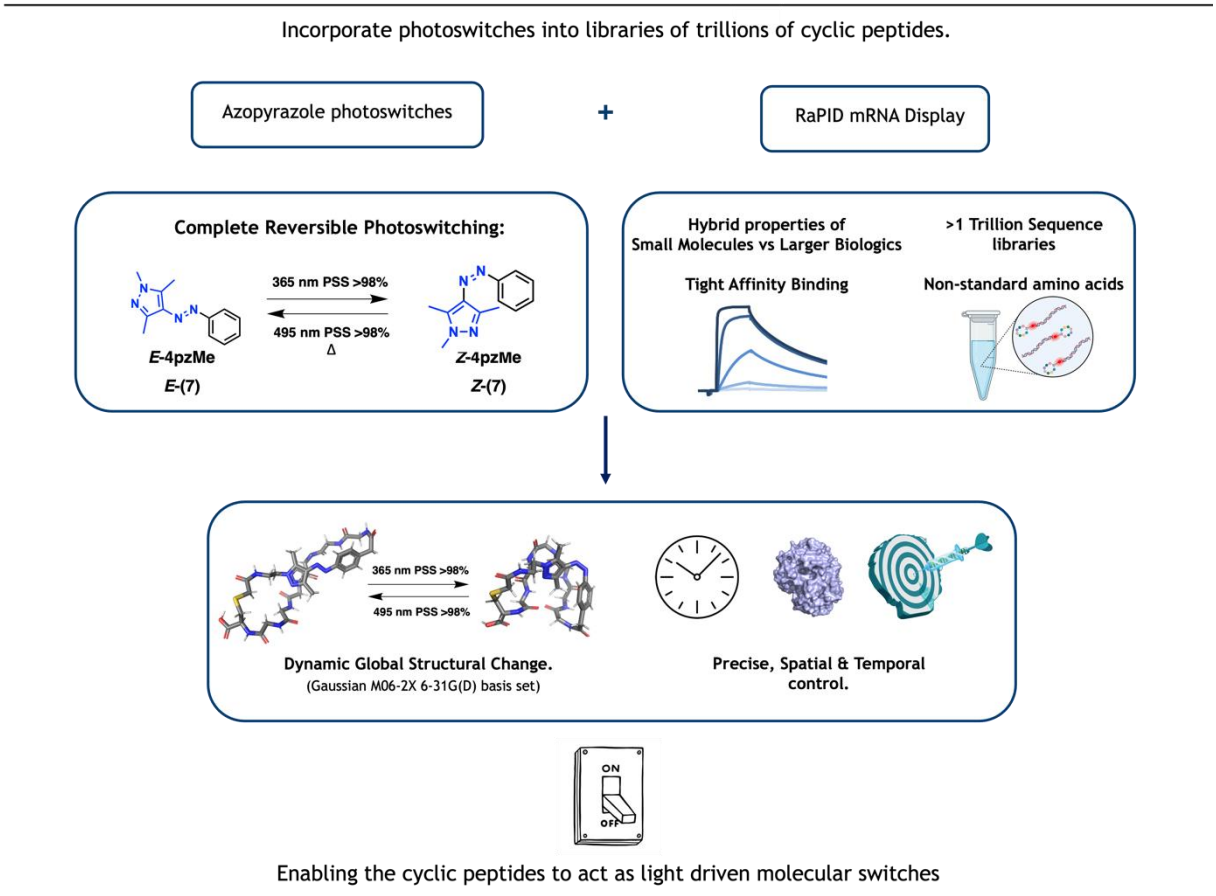


Figure 27: PhotoRaPID in vitro evolution platform will enable the discovery of light-responsive cyclic peptides against any recombinantly expressed protein target of interest.

1.9.2 Benefits of using azopyrazole-based photoswitches and RaPID mRNA display to achieve this next-generation in vitro light-responsive peptide display platform:

Using azopyrazole-based photoswitches:

The use of azopyrazole photoswitches builds upon the current literature on light-responsive peptide display platforms. This is because azopyrazoles enable quantitative bi-directional switching between each isomer. The complete photoconversion will enable distinct peptide isomer populations to be isolated, something not currently achieved by the azobenzene-based approaches previously discussed. This is a particularly important aspect of photopharmacology applications, where enabling full activation or deactivation of the light-responsive tool compound is critical for precise functionality. Additionally, certain azopyrazoles have thermal relaxation timescales (hours-to-days) appropriate for chemical biology applications, making them ideal to use as light-responsive tool compounds to interrogate complex and dynamic biological systems.

Using cyclic peptide RaPID mRNA display:

RaPID mRNA display will enable the evolution of light-responsive cyclic peptides from trillion-member DNA libraries. The power of the FIT system will enable genetic encoding of the photoswitch moiety and direct integration within the macrocyclic peptide ring. This is particularly important to ensure that the maximal global structural change can be imparted upon light irradiation – the key prerequisite to large dynamic switchable binding as highlighted by Heinis and co-workers,[122] previously. The use of cyclic peptides surrounding the integrated photoswitch will provide precise specificity and potency against any recombinantly expressed target of interest not currently possible with azologisation, azo-extension or fragment-based design approaches.

Integrating the two areas together:

By integrating azopyrazole-based photoswitches within trillions of cyclic peptides, this next-generation PhotoRaPID *in vitro* evolution platform will enable the discovery of light-responsive cyclic peptides with complete reversible photoswitching. This will enable the identification of novel cyclic peptides which act as light-driven molecular switches, liganding traditionally challenging targets and imparting precise spatial and temporal control over any biological system of interest (Figure 27).

1.9.2.1 Detailed aims of the PhD Work – Outlining the chapters of this thesis:

Detailed aims of the project are outlined as follows:

1. Synthesis of photoswitchable azopyrazole derivatives and testing for integration within the FIT system.
2. Preparation and expansion of the RaPID platform to evolve potent switchable binding peptides (PhotoRaPID system).
3. Testing of the PhotoRaPID system on a real protein target of interest (beyond a streptavidin model system).
4. Further optimisation of the PhotoRaPID workflows to ensure the robustness of the platform, streamline the hit deconvolution process post-selection, and further enhance the discovery of *de novo* evolved peptide ligands with light-responsive differential binding.

2 SYNTHESIS OF PHOTOSWITCH ANALOGUES FOR GENETIC CODE REPROGRAMMING

Chapter Premise:

This chapter will discuss work aimed towards the integration of photoswitches within the RaPID system. Analogues of an azopyrazole-based photoswitch were synthesised and their loading on a model tRNA was assessed. FIT under genetic code reprogramming was then carried out to demonstrate the successful translation of peptides containing a photoswitchable ns-AA. This validated the approach for the generation of photoswitchable responsive cyclic peptide libraries via RaPID. The outline of this chapter is as follows:

- Design and structural modifications required of an azopyrazole photoswitch to enable RaPID integration under genetic code reprogramming.
- Sequential synthesis of the compatible photoswitchable analogues.
- Testing of each analogue to identify loading, if any, onto a model tRNA.
- FIT under genetic code reprogramming and MALDI of a known peptide sequence to confirm the successful incorporation of a photoswitchable ns-AA.

The chapter workflow is as follows, (Figure 28).

Chapter Workflow:

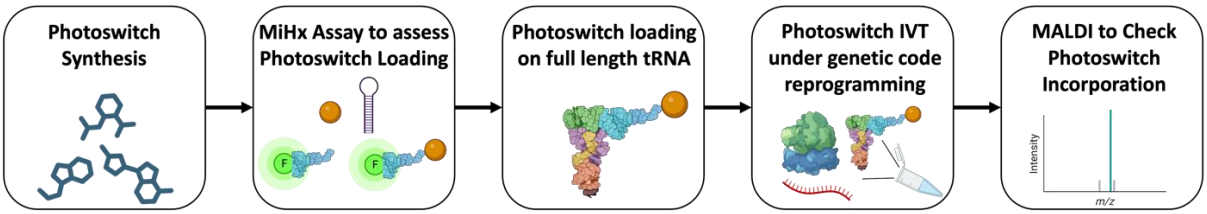


Figure 28: Workflow of the Chapter.

2.1 Photoswitch design for compatibility with genetic code reprogramming:

2.1.1 Structural modifications of parent photoswitch (4pzMe (7)) to enable genetic code reprogramming:

Following the work of Weston et al.,[46] highlighted in the introduction, an azopyrazole photoswitch, 4pzMe (7) (Figure 29(a)), was first chosen as the archetype for this work. 4pzMe (7) displays quantitative *E/Z* photoisomerisation and a thermal half-life ($\tau_{1/2}$) of 10 days, making this photoswitch advantageous for chemical biology/ photopharmacology applications. To enable compatibility with the RaPID system under genetic code reprogramming, structural modification of the photoswitch core were required.

The first key pre-requisite of the project was that the photoswitch must appear in every macrocyclic within the library. Because of this, it was a logical starting point to reprogramme a fixed codon within the RaPID library to ensure the permanent presence of the photoswitch. $^{ini}Met_{CAU}$ was the tRNA was chosen, reprogramming Methionine for the ns-AA photoswitch. This was because $^{ini}Met_{CAU}$ tRNA universally initiates all ribosomal translation and as such is naturally the fixed first amino acid codon in the RaPID DNA libraries (Figure 29(b)) (described in detail in Chapter 1.6.3). By reprogramming the initiator AA, however, as all the library members in RaPID need to be cyclic peptides, further modification of the photoswitch was also required (Figure 29(c)). *N*-alkylation of the pyrazole ring with an *N*-chloroacetyl ethylamine linker was decided, to enable covalent head-to-side chain cyclisation with a fixed downstream cysteine. This was to ensure that all members of the peptide library were cyclic when screening. In contrast, compatibility of the photoswitch with the Flexizyme technology (Chapter 1.5), required the integration of the activated leaving groups ($_{Act}$.LGs – CME, CBT, DBE, ABT), to enable loading of the photoswitch onto the tRNA of interest. This functionality was placed on the opposite side of the photoswitch core to the chloroacetyl moiety.

Following the guidance of previous literature:

Due to the previous work of Díaz-Lobo et al.,[125] where acceleration of the photoswitch thermal-relaxation ($\tau_{1/2}$) was seen when conjugating functionality directly off the photoswitch core, spacing groups were also introduced on either side of the key motifs required (an *N*-pyrazole ethylamine linker and an acetyl moiety off the phenyl ring). The use of saturated sp^3 hybridisation on either side of the core photoswitch motif was purposefully chosen to ensure deconjugation of the photoswitch electronics from any additional motifs required for compatibility. This was because the functionalised

photoswitch of Díaz-Lobo et al. contained conjugated π-electron-deficient groups adjacent to their photoswitch core. The acetyl spacer off the phenyl ring was also chosen following the work of Lee et al.[96] to help jump-start the successful loading of the photoswitch. This was because they compared the loading efficiency of various phenyl derivatives including phenylacetic-CME (40% loading), which looked structurally like the phenyl portion of the photoswitch. As deconjugation of the photoswitch with an acetyl spacer was already required, mimicking the phenylacetic-CME, but with a photoswitch replacing the base phenyl ring became the logical first analogue to synthesise.

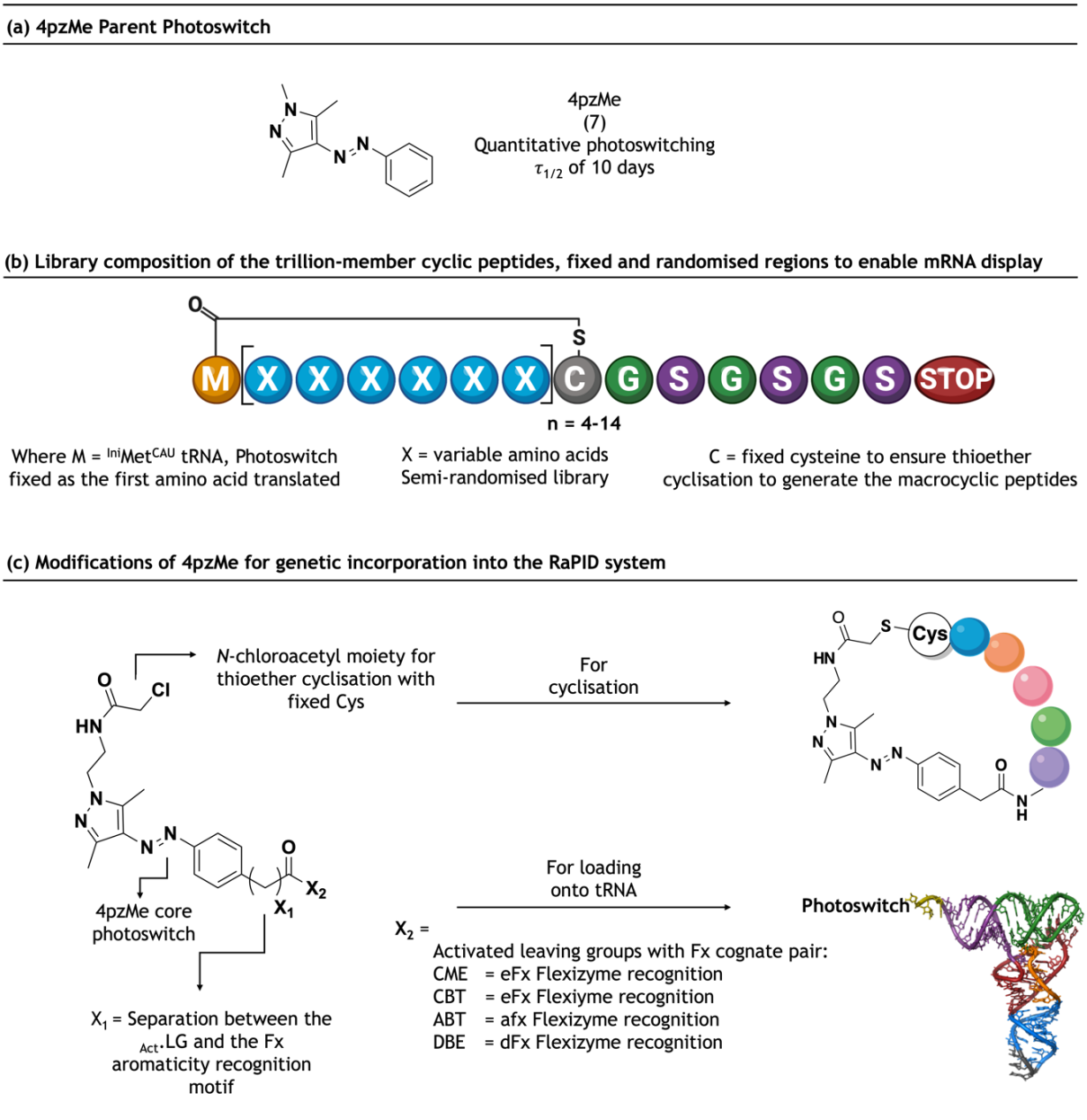


Figure 29: The parent photoswitch and its derivatisation to enable integration within RaPID libraries. (a) The parent photoswitch 4pzMe (7). (a) The 4pzMe (7) photoswitch core. (b) General structure of all library members with a fixed ^{ini}Met tRNA codon, (c) $_{act.}$LG required for Flexizyme recognition and aminoacylation onto the tRNA of interest.

2.2 Synthesis of photoswitch analogues and assessment of aminoacylation efficiency via MiHx assays:

2.2.1 Synthesis of ClAc-Photo-CME (11):

The first genetic-code-reprogramming compatible photoswitch analogue designed, ClAc-Photo-CME (11) (Figure 30), was synthesised with an activated cyanomethyl ester (CME). The CME ester is recognised by the eFx Flexizyme variant and requires intrinsic aromaticity within the non-standard amino acid for recognition. As the photoswitch core contains a phenyl ring, it was hoped that the aromaticity within the phenyl ring would have sufficient spatial distance from the activated CME group for adequate loading. Guidance for this rationale was taken from Lee et al. (mentioned previously).[96]

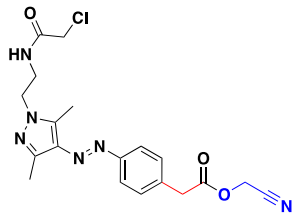
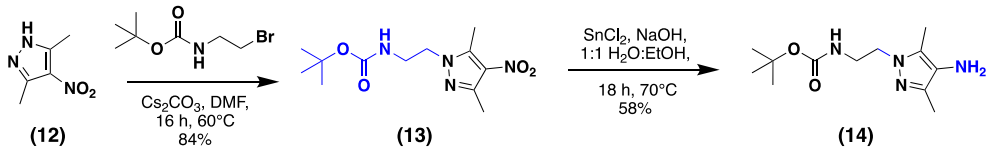


Figure 30: ClAc-Photo-CME (11).

The synthesis began with *N*-alkylation of 3,5-dimethyl-4-nitro-1*H*-pyrazole (12) with *tert*-butyl (2-bromoethyl)carbamate, under basic conditions (Scheme 1). Formation of substituted pyrazole (13) proceeded cleanly in robust yields after changing solvent to DMF from acetonitrile to improve the solubility of the starting materials.



Scheme 1: First two steps of the synthetic route to ClAc-Photo-CME (11). N-alkylation of the pyrazole ring to produce (13), followed by nitro reduction to the corresponding primary amine (14).

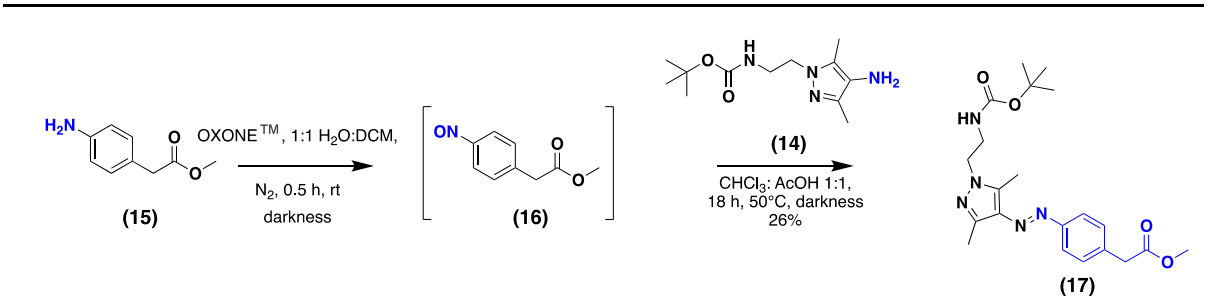
Reduction of nitro pyrazole (13) yielded the corresponding amine (14), although significant synthetic optimisation was required. Hydrogenation of the nitro was initially attempted with various palladium

or palladium hydroxide on carbon catalysts, at either atmospheric or 2 bar hydrogen pressure using a Parr Shaker (rows 1–3, Table 2). Minimal reduction was observed by TLC and only starting material (SM) or decomposition products were observed by ^{1}H NMR spectroscopy. Reaction with $SnCl_2$ successfully reduced the nitro pyrazole (13) to its corresponding amine (14) (row 4, Table 2), however, the workup was particularly challenging. Liquid-liquid extraction of the reaction mixture yielded significant tin contamination which could be removed via flash-chromatography but with difficulty. This procedure, however, was not ideal for reaction scaling and produced a maximum isolated yield of 58%. Alternative approaches using iron powder or sodium dithionite were also attempted to move away from tin reagents (rows 5&6, Table 2: Optimisation of pyrazole nitro reduction conditions.). However, these led to decomposition/ pyrazole ring opening or yielded only starting material. As such, tin(II) chloride was used to generate ClAc-Photo-CME (11).

Table 2: Optimisation of pyrazole nitro reduction conditions.

Entry:	Reagents/Conditions:	Yield (%):	Observations:
1	10% Pd/C, H_2, N_2, 1 bar, rt, 18 h	-	Minimal reduction/ decomposition.
2	10% Pd/C, H_2, N_2, 2 bar, rt, 18 h	-	Minimal reduction/ decomposition.
3	20% $Pd(OH)_2$/C, H_2, N_2, 1 bar, rt, 18 h	-	Minimal reduction/ decomposition.
4	$SnCl_2$, NaOH, H_2O: EtOH 1:1 70 °C, 18 h	58	Sufficient reduction was identified, however, problematic workup.
5	Fe powder, $CaCl_2$, EtOH: H_2O 20:1, 90 °C, 0.5 h	-	Decomposition of the pyrazole ring/ possible ring opening.
6	$Na_2S_2O_4$, EtOH, 50 °C, 4 h	-	Return of the starting material.

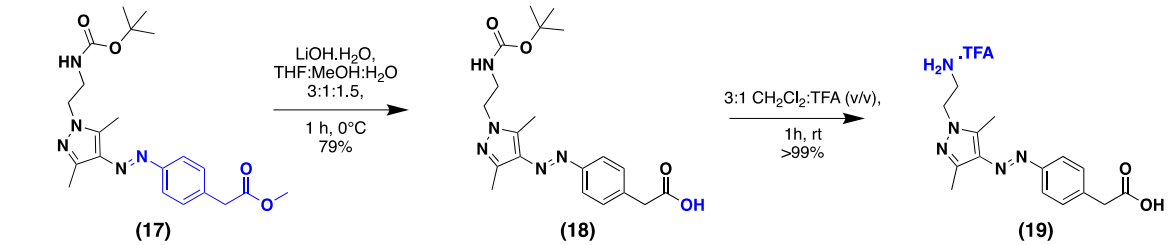
Generation of the azo bridge, producing the azopyrazole (17), was achieved using a Bæyer-Mills oxidative coupling approach (Scheme 2). Methyl 2-(4-aminophenyl)acetate (15) was oxidised to the corresponding nitroso intermediate (16) using a bi-phasic solution containing OXONE™. Following oxidation, the nitroso (16) was immediately added to a pre-stirring solution of (14).



Scheme 2: Third step, generation of the Nitroso intermediate and subsequent Bæyer-Mills coupling.

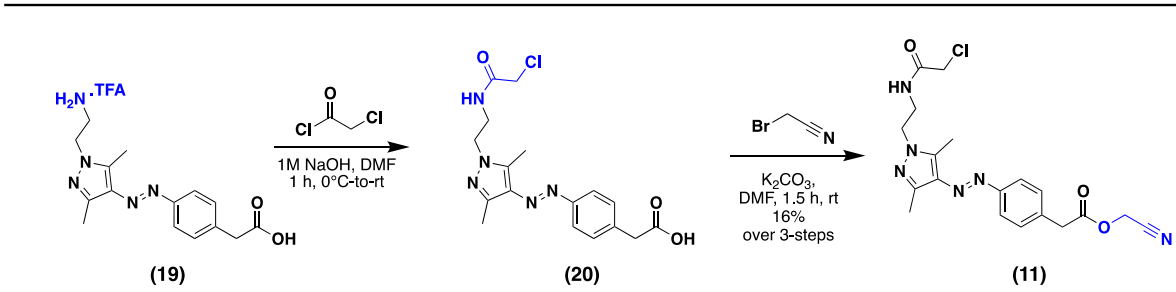
The reaction proceeded well but with a poor yield, however, as the focus was to test the incorporation of a photoswitchable ns-AA, this process was only optimised at a later point (Chapter 2.2.7).

Following the formation of the photoswitch azo-bridge, methyl ester hydrolysis using lithium hydroxide swiftly revealed the corresponding carboxylic acid (18), without the need for further purification (Scheme 3). Boc deprotection followed with quantitative yield to give (19).



Scheme 3: Fourth step and fifth steps, methyl ester hydrolysis and subsequent Boc-deprotection.

With the free amine (19) in hand, the chloroacetyl moiety could be installed with chloroacetyl chloride (Scheme 4). This produced the corresponding chloroacetylated product (20), which could be used without purification. The final product, ClAc-Photo-CME (11), was obtained by coupling the acid with bromoacetonitrile under basic conditions in a 16% yield over 3 steps (equivalent to 55% per step).



Scheme 4: Sixth and final seventh step, chloroacetylation of the corresponding free amine and esterification of the free carboxylic acid to generate the final product ClAc-Photo-CME (11).

2.2.2 Assessing aminoacylation efficiency of ClAc-Photo-CME (11) via a MiHx assay:

With ClAc-Photo-CME (11) in hand, the compatibility of this analogue with Flexizyme technology was assessed. To test the efficiency of Flexizyme-mediated loading (aminoacylation) of (11) onto a small synthetic tRNA mimetic, a Microhelix (MiHx) assay panel was used. This uses a truncated 23-base tRNA mimetic conjugated to 4-carboxy-fluorescein (FAM), termed FAM-23b. The FAM-23b truncated tRNA mimetic is used to enhance the molecular weight shift seen when the ns-AA is conjugated, whilst the presence of the fluorophore within FAM-23b allows for in-gel fluorescence. These features of the MiHx make it easy to visualise if successful loading is achieved (Figure 31). Once suitable loading conditions onto the truncated tRNA are found, loading of the photoswitch onto a full-length tRNA of interest using the pre-determined loading conditions then occurs. The efficiency of this loading onto full-length tRNA is then qualitatively determined via FIT (under genetic code reprogramming) of the loaded tRNA using a known test mRNA sequence and MALDI-TOF (matrix-assisted laser desorption/ ionization time-of-flight) mass spectrometry. MALDI enables a qualitative assessment of the efficiency of integration of that ns-AA within ribosomally translated peptides and to see if ribosomal stalling or miss incorporation is seen with the new ns-AA, i.e., the readout is a combination of the efficiency of tRNA loading and ribosomal tolerance of the ns-AA.

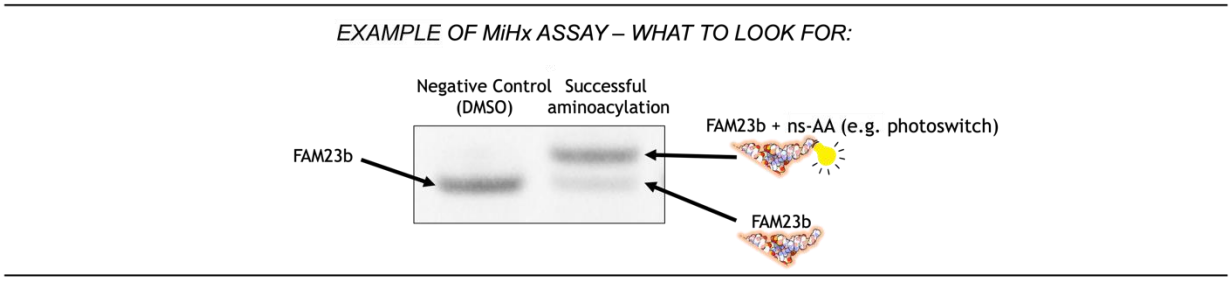


Figure 31: Example MiHx assay in-gel fluorescence, distinguishing the non-loaded and ns-AA loaded synthetic truncated tRNA FAM-23b.

Aminoacylation conditions for loading ClAc-Photo-CME (11) were assessed in an aqueous buffer on ice for between 1–24 hours, at either pH 7.5 or 9.0 (Lanes 6–15, Figure 32). Neat DMSO was used as a negative control (Lanes 1&2, Figure 32), whilst ClAc-^{L}Y-CME was used as the positive control for the reaction (Lanes 3&4, Figure 32). Aminoacylation of the positive control gave reassurance that the flexizyme was folded correctly as a second shifted upper band was seen. ClAc-Photo-CME (11) in contrast, could not be loaded as an upper-band shift was not seen. The upper-band shift indicates the increased molecular weight of the photoswitch being loaded onto the truncated MiHx.

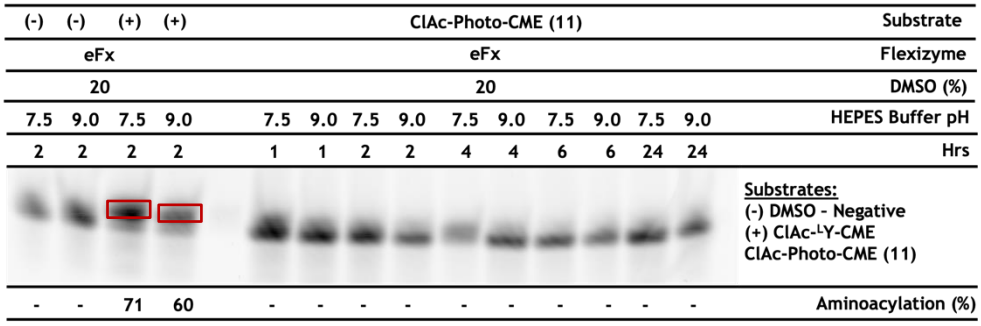


Figure 32: Aminoacylation efficiency of ClAc-Photo-CME (11) with a truncated tRNA (23b-FAM). Red boxes indicate the succesful positive controls. Following the MiHx assay panel of conditions ClAc-Photo-CME (11) could not be loaded successfully.

Additional screening conditions were also trialled (Figure 33). Firstly, an increased incubation to 30 h at 4 °C was trialled however, no aminoacylation was seen. Further effort to promote aminoacylation via incubating the reaction at room temperature or using an increased final concentration of DMSO of 40% were also attempted. This was intended to improve the solubility of (11) as some precipitation was seen during the initial screen. Unfortunately, all conditions trialled were unable to successfully aminoacylate ClAc-Photo-CME (11) onto FAM-23b.

With ClAc-Photo-CME (11) unable to be loaded, there were two likely causes for this lack of reactivity that needed investigating to generate a FIT-compatible photoswitch. The first was whether the photoswitches Flexizyme recognition motif (aromaticity with an activated CME ester) was not suitably organised for recognition. Alternatively, the second hypothesis was that the solubility of (11) was limiting aminoacylation. Regarding recognition, when cross-referencing against Lee et al.,[96] as mentioned prior, their much simpler phenylacetic-CME could only achieve an aminoacylation efficiency of 40%. Therefore, it was quite possible that by modifying the phenylacetic-CME further, to the photoswitch (11) derivative, led to even poorer aminoacylation. In contrast, a lack of solubility also appeared as a distinct possibility, as some precipitation of (11) was routinely observed after 1–2 hours. It is important to stress, however, that precipitation of ns-AA's within the highly aqueous buffer is not uncommon. tRNA loading typically solubilises the ns-AA; at which point the interplay between the rate of aminoacylation vs precipitation may determine the overall aminoacylation efficiency. With these potential concerns, testing these hypotheses with other activating motifs was explored.

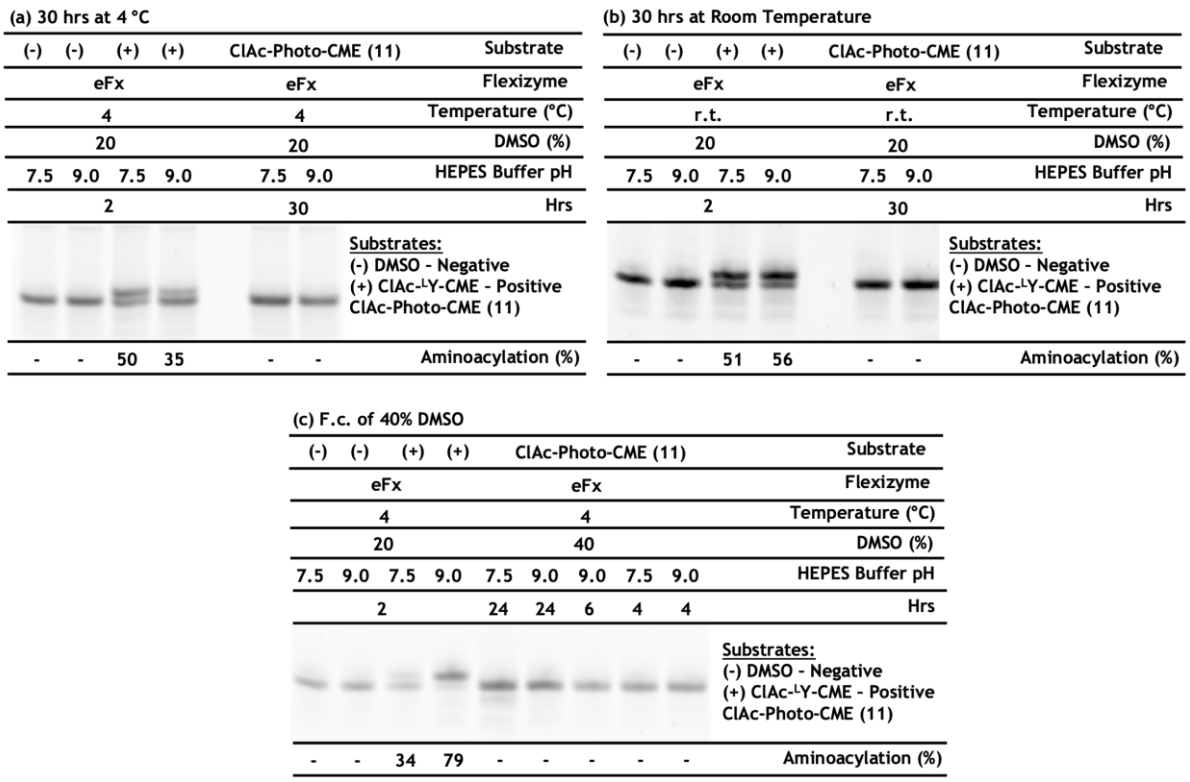


Figure 33: Combined panel of ClAc-Photo-CME (11) screening conditions. (a) Longer duration at 4 °C, (b) Longer duration at room temperature, (c) Increased DMSO at 40% f.c..

2.2.3 Synthesis of ClAc-Photo-CBT (21):

To address the potential poor recognition of the photoswitch core by the Flexizyme, ClAc-Photo-CBT (21) (Figure 34), was chosen as the second genetic-code-reprogramming compatible photoswitch to be synthesised.

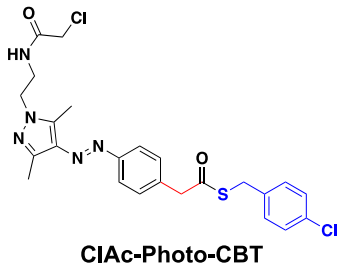
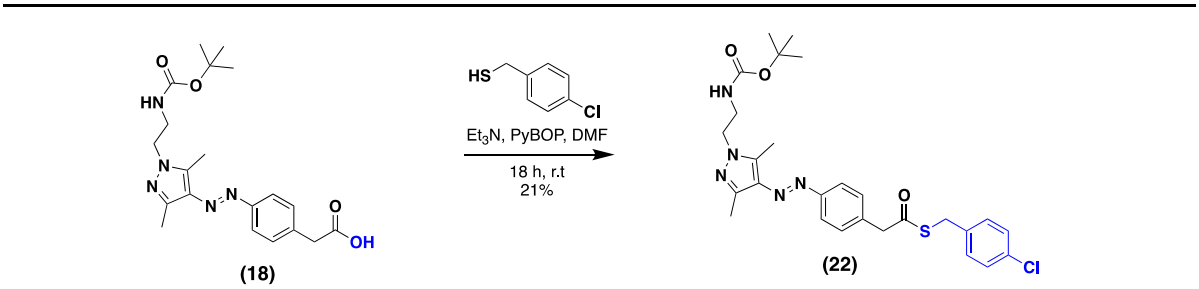


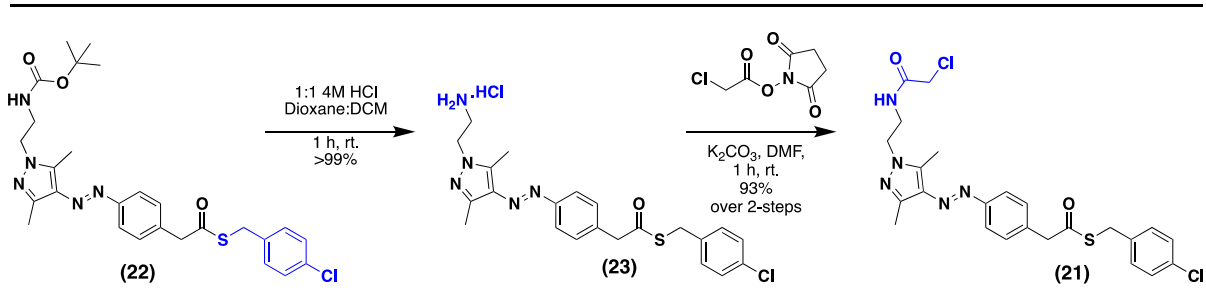
Figure 34: ClAc-Photo-CBT (21)

ClAc-Photo-CBT (21) contains a chlorobenzyl thioester (CBT) $_{act.}$LG which is intended to increase Flexizyme-mediated aminoacylation. CBT offers intrinsic aromatic recognition from the phenyl ring. The purpose of using this $_{act.}$LG was to circumvent any issues in Flexizyme recognition from the photoswitch core, if there were any. Simultaneously, the presence of the electron-withdrawing para-chlorine, in combination, with the thioester linkage also results in CBT being a more promiscuous leaving group. This increased reactivity (towards aminoacylation) aimed to promote an increased rate of tRNA loading over precipitation, thus, addressing the other potential issue of solubility associated with ClAc-Photo-CME (11).



Scheme 5: Synthetic route towards ClAc-Photo-CBT, thioesterification of common intermediate (18).

Starting from acid intermediate (18) the CBT thioester was installed using 4-chlorobenzyl-mercaptan and PyBOP (Scheme 5). Following this (Scheme 6), the Boc group was removed with 1:1 4 M hydrochloric acid: dioxane. This produced the desired hydrochloride salt. Finally, the chloroacetyl moiety was installed using chloroacetoxy succinimide. This was also an improvement over using chloroacetyl chloride due to easier handling and led to the generation of the final product (21) in excellent yields over the final two steps.



Scheme 6: Boc deprotection of (22) followed by chloroacetylation to generate final compound ClAc-Photo-CBT (21).

2.2.4 Aminoacylation efficiency of ClAc-Photo-CBT (21) via a MiHx assay:

After assessing the aminoacylation efficiency of ClAc-Photo-CBT (21) across both the standard screening conditions (Figure 35(a)) and increased incubation temperature (Figure 35(b)), it was determined that the additional recognition provided by the activated thioester provided no benefit.

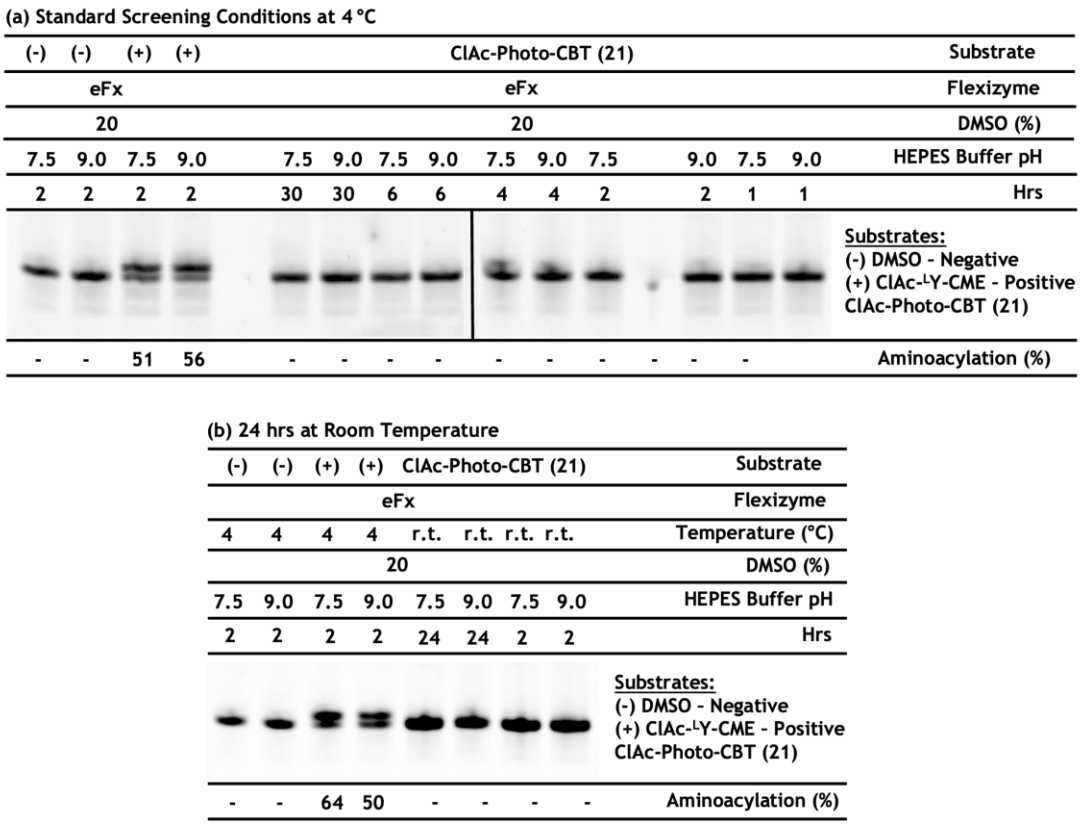


Figure 35: Combined panel of ClAc-Photo-CBT (21) screening conditions. (a) Standard screening panel at 4 °C, (b) 24 h duration at room temperature. ClAc-Photo-CBT (21) could not be loaded in the MiHx assay panel of conditions tested.

Since no aminoacylation was seen, alternative approaches to promote aminoacylation were trialled with the current analogues before moving on. Photoswitches can adopt two different isomers (*E/Z*) and it has been shown that the *Z*-isomer typically adopts a conformation with a greater dipole moment across the azo-bridge.[46] As such, it was hypothesised that irradiating ClAc-Photo-CME (11) and ClAc-Photo-CBT (21) to their more polar Z-isomers may provide additional solubility in the aqueous HEPES buffer, to promote aminoacylation. Despite this hypothesis, irradiation to the *Z*-isomer (using 365 nm light) unfortunately did not result in any noticeably increased aminoacylation (Figure 36(a)). A further truncated MiHx, with only 5 bases and the FAM fluorophore (FAM-5b) was also trialled, in the hope

that the addition of the photoswitch only resulted in a very small shift when coupled, and thus could not be seen with the standard FAM-23b (Figure 36(b)). ClAc-Photo-CME (11) and ClAc-Photo-CBT (21) produce an increased molecular weight of 360.82 $gmol^{-1}$ onto FAM-23b. Addition of molecular weights such as this typically do provide distinct band separation (>~100 $gmol^{-1}$ is typically sufficient for band separation) in the gel. However, due to other considerations when running through the gel (e.g. charge, photoisomerisation when in the gel) it was reasonable to confirm this. This was as it was unknown what effect the photoswitch had when running through the gel. As shown in Figure 36, both additional MiHx screening panels did not reveal any meaningful data to suggest successful aminoacylation, and as such an additional analogue ClAc-Photo-DiMeABT was subsequently generated.

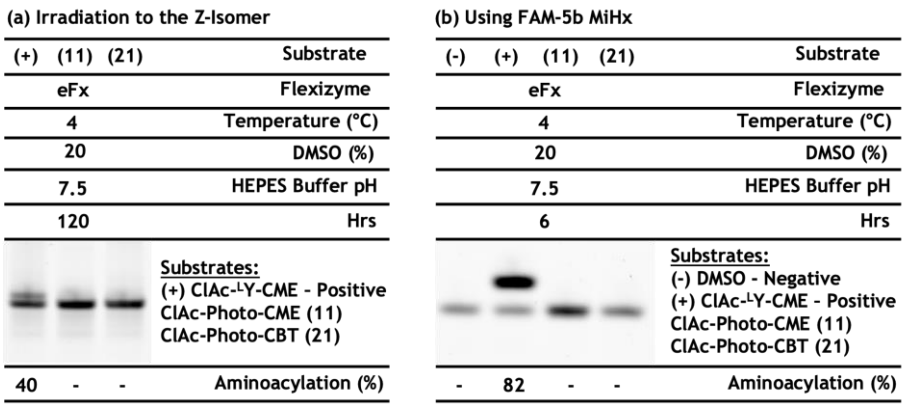
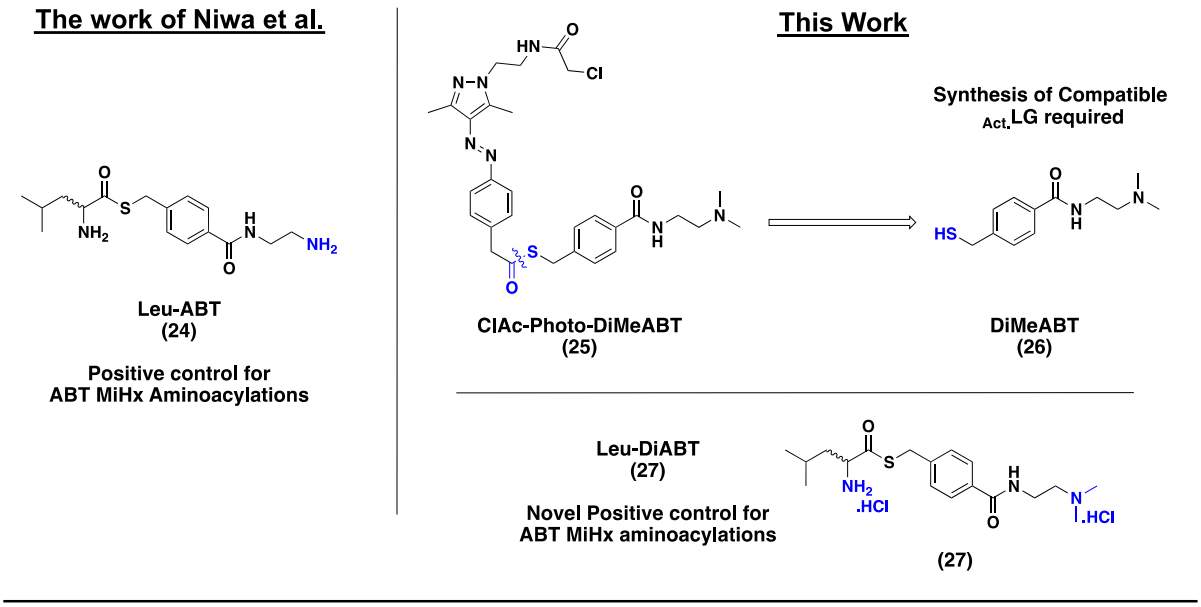


Figure 36: Combined panel of ClAc-Photo-CME (11) and ClAc-Photo-CBT (21) additional screening conditions. (a) Aminoacylation Efficiencies when irradiated to Z-isomer, (b) Visualising the Aminoacylation Efficiency using the smaller FAM-5b MiHx. All panels tested, unfortunately, did not indicate successful photoswitch aminoacylation.

2.2.5 Synthesis of ClAc-Photo-DiMeABT (25):

To further test the hypothesis that poor aminoacylation efficiency resulted from poor aqueous solubility, a more soluble activating group was desirable. Looking at the literature, an aminobenzyl thioester (ABT) activated leaving group had been used to promote the loading of hydrophobic ns-AAs onto tRNA with the engineered aFx derivative by Niwa et al. (Leu-ABT (24)).[95] This is because the primary amine within the $_{act.}$LG ABT provided increased solubility for the analogues reported. Therefore, ClAc-Photo-DiMeABT (25), was designed based on inspiration from the prior aminobenzyl thioester (ABT) activated leaving group. For our applications, however, the presence of the free primary amine would be problematic during the addition of the chloroacetyl moiety. As such a

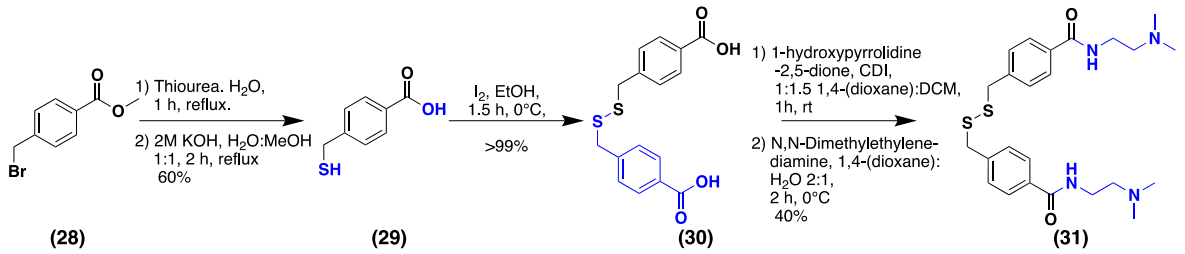
derivative of ABT, the tertiary amine (DiMeABT) was proposed. This novel $_{act.}$LG aimed to enable improved photoswitch solubility via the quaternary ammonium cation in aq. buffers, whilst not resulting in inter/intra molecular chloroacetyl-NH_2-Photoswitch polymerisation during synthesis. As DiMeABT was a novel and untested $_{act.}$LG, a suitable positive control Leu-DiMeABT was also synthesised in parallel, as its ABT counterpart was generated Niwa et al. (Scheme 7).[95]



Scheme 7: Synthesis of ClAc-Photo-DiMeABT based on the work of Niwa et al.[95]

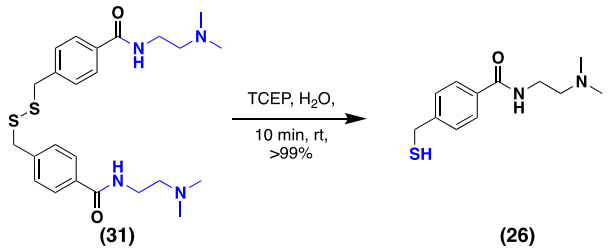
2.2.5.1 The DiMeABT thiol precursor for thioesterification:

To generate ClAc-Photo-DiMeABT (25), firstly the $_{act.}$LG DiMeABT required synthesis. Synthesis of the novel DiMeABT thiol (26) (Scheme 8 and Scheme 9), was carried out following a modification to the synthetic route published by Niwa et al.[95] 4-Bromobenzoic acid (28) was reacted with thiourea to generate the corresponding isothiouronium salt intermediate. This was subsequently hydrolysed using 2 M KOH to produce (29). The thiol (29) was then oxidatively dimerised using iodine in ethanol at 0 °C to quantitatively produce (30). With the disulfide in hand, succinimide activation of the carboxylic acids was achieved using N-hydroxysuccinimide and CDI. This activated intermediate was then immediately reacted with *N,N*-dimethylethylenediamine to produce the alkylated disulfide (31) in modest yields (40%).



Scheme 8: Synthetic route to the dimerised analogue of the DiMeABT thiol. DiMeABT was used to generate the activated thioester in ClAc-Photo-DiMeABT (25).

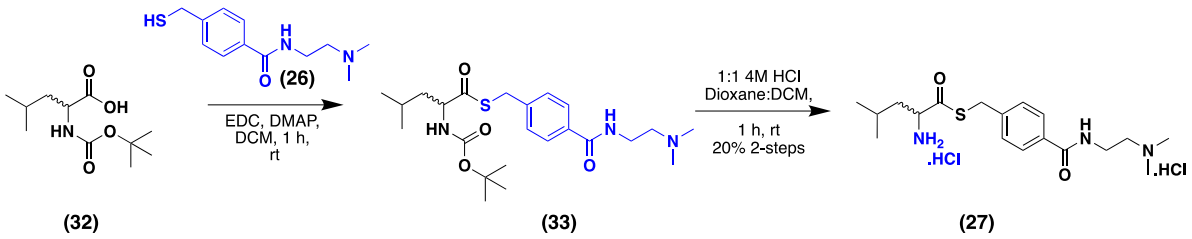
The intermediate (31), was stored at -20 °C and only reduced to the thiol (26) immediately before its use as the thioesterification reagent (Scheme 9). For this reduction, Niwa et al.,[95] used dithiothreitol (DTT) as the reducing agent, however, tris(2-carboxyethyl)phosphine (TCEP) proved faster and more efficacious.



Scheme 9: Disulfide reduction to generate the final precursor DiMeABT required for the synthesis of ClAc-Photo-DiMeABT (25).

2.2.5.2 Leu-DiMeABT (27):

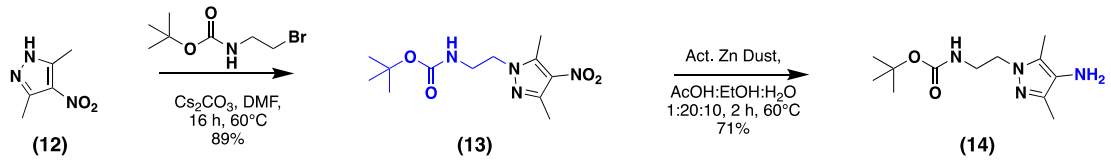
With the $_{act.}$LG in hand, synthesis of the positive control (Leu-DiMeABT (27)) was required to assess the aminoacylation efficiency of this novel $_{act.}$LG. Racemic Leu-DiMeABT (27) was synthesised (Scheme 10), again mirroring the work of Niwa et al.,[95] with confirmed aminoacylation of their Leu-ABT (24) derivative published. Leu-DiMeABT (27) was synthesised by reacting racemic Boc-protected leucine (32) with the free thiol DiMeABT intermediate (26), producing the protected product (33). This intermediate (33) was then Boc-deprotected using the standard conditions to produce the control compound Leu-DiMeABT (27) as the di-hydrochloride salt in a poor but sufficient yield.



Scheme 10: Synthesis of positive control Leu-DiMeABT (27).

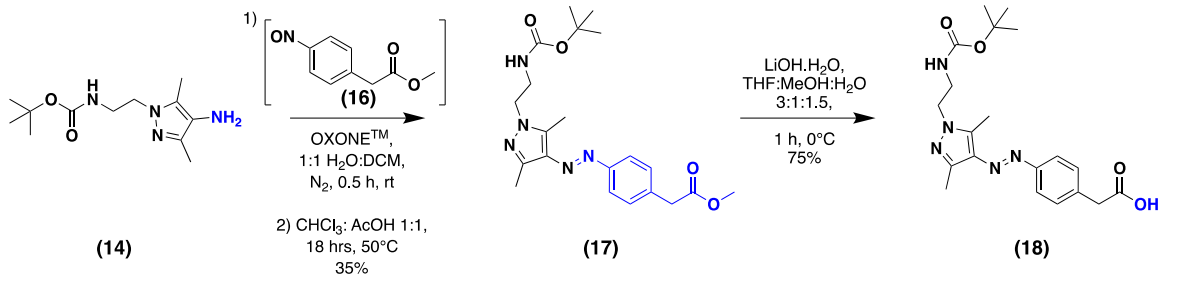
2.2.5.3 ClAc-Photo-DiMeABT (25):

ClAc-Photo-DiMeABT (25) was synthesised by coupling DiMeABT (26) with the common photoswitch intermediate (18). At this point in the project, synthesis of the common photoswitch intermediate (18), was modified with the discovery of a novel set of nitro reduction conditions (Scheme 11). An optimised reduction was found using activated zinc (Zn) dust (<10 µm) which led to robust and efficient reduction (71%). Most importantly the required workup was significantly easier and as such it replaced the $SnCl_2$ reduction conditions previously employed. This was one of the largest improvements in synthetic optimisation for these photoswitch analogues.



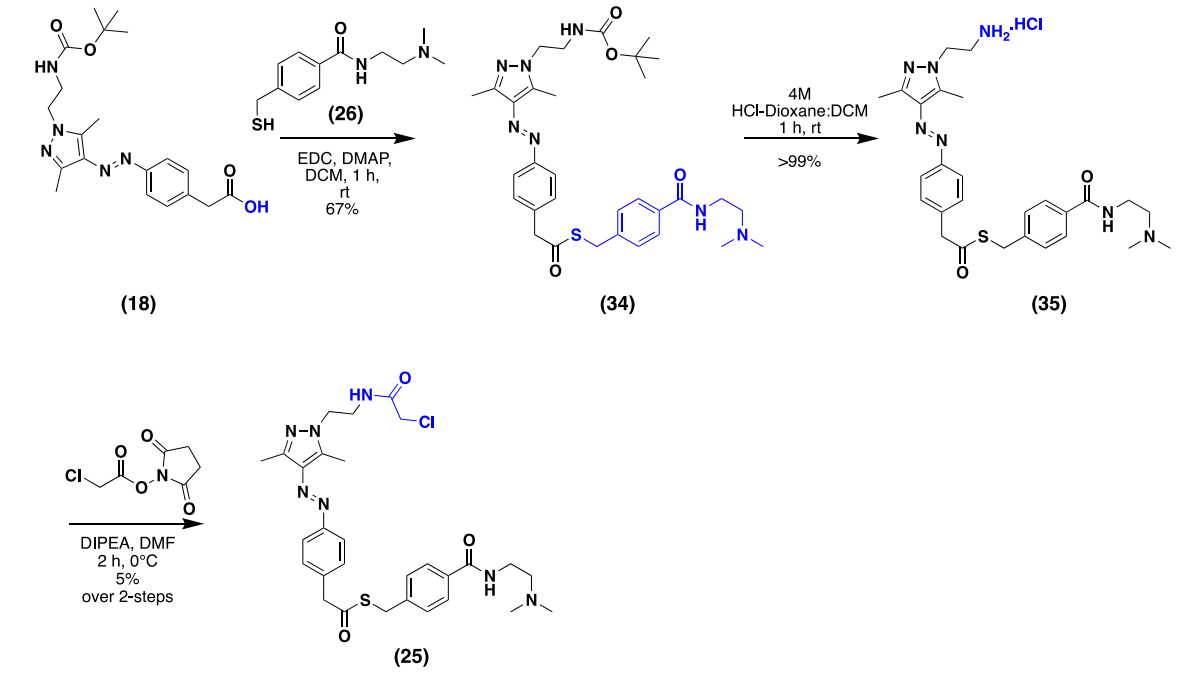
Scheme 11: Improved nitro reduction conditions using activated Zn.

Subsequent Bæyer-Mills coupling then proceeded as usual and the methyl ester hydrolysis generated the common photoswitch intermediate (18), as before (Scheme 12).



Scheme 12: Synthesis of the common photoswitch intermediate (18), using (14) which was generated from the improved Zn nitro reduction conditions (Scheme 11).

The thioester (34), was generated by reacting the photoswitch intermediate (18) with (26) (immediately after disulfide reduction of (31)), using the standard coupling conditions of EDC/DMAP. Boc deprotection and subsequent chloroacetylation were then carried out producing the final desired product ClAc-Photo-DiMeABT (25), in very low but sufficient yields (Scheme 13).



Scheme 13: Generation of the final analogue ClAc-Photo-DiMeABT (25).

2.2.6 Aminoacylation efficiency of ClAc-Photo-DiMeABT (25) via a MiHx assay:

Successful aminoacylation of the positive control Leu-DiMeABT (27) was seen at 73% (Figure 37). This was very encouraging as the use of the DiMeABT $_{act.}$LG could potentially be very useful for scope beyond this project. Despite this success, however, no aminoacylation of ClAc-Photo-DiMeABT (25) was identified, despite the 24 h incubation. It was very encouraging however, that very little precipitation of ClAc-Photo-DiMeABT (25) was seen, indicating qualitatively that the addition of the novel DiMeABT-activated leaving group provided increased solubility as desired. From this result, it was concluded that photoswitch solubility was not the critical limitation to successful aminoacylation as initially suggested. Therefore, pursuing structural modification of the photoswitch, aiming to improve the recognition of the analogues by Flexizymes was then attempted.

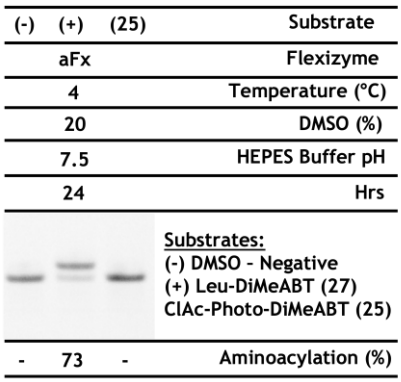
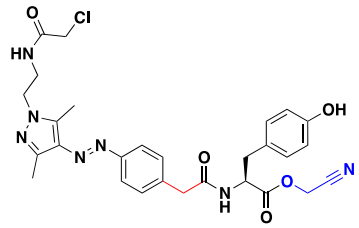


Figure 37: MiHx Assay for ClAc-Photo-DiMeABT (25) using Leu-DiMeABT (27) as a positive control. Novel activating group DiMeABT loads hydrophobic amino acids in high yield (Leu-DiMeABT) but fails to improve ClAc-Photo-DiMeABT (25) loading.

2.2.7 Synthesis of ClAc-Photo-^{L}Y-CME (36):

The next analogue designed, ClAc-Photo-^{L}Y-CME (36) (Figure 38), incorporated the standard photoswitch motif but with an *L*-tyrosine and a CME-activated ester. ^{L}tyrosine was used as it is an aromatic amino acid which was well characterised for recognition by eFx.[126] This provided confidence that this analogue should be readily aminoacylated and loaded. As the ^{L}tyrosine contains an aromatic phenyl sidechain, a CME ester was used as the $_{act.}$LG. ClAc-Photo-^{L}Y-CME (36), using tyrosine, was not initially attempted, as despite the increased chance of successful aminoacylation, this would mean that all macrocycles in the library would have a fixed tyrosine adjacent to the photoswitch. Due to the loading challenges seen thus far though, it became a suitable alternative to explore.

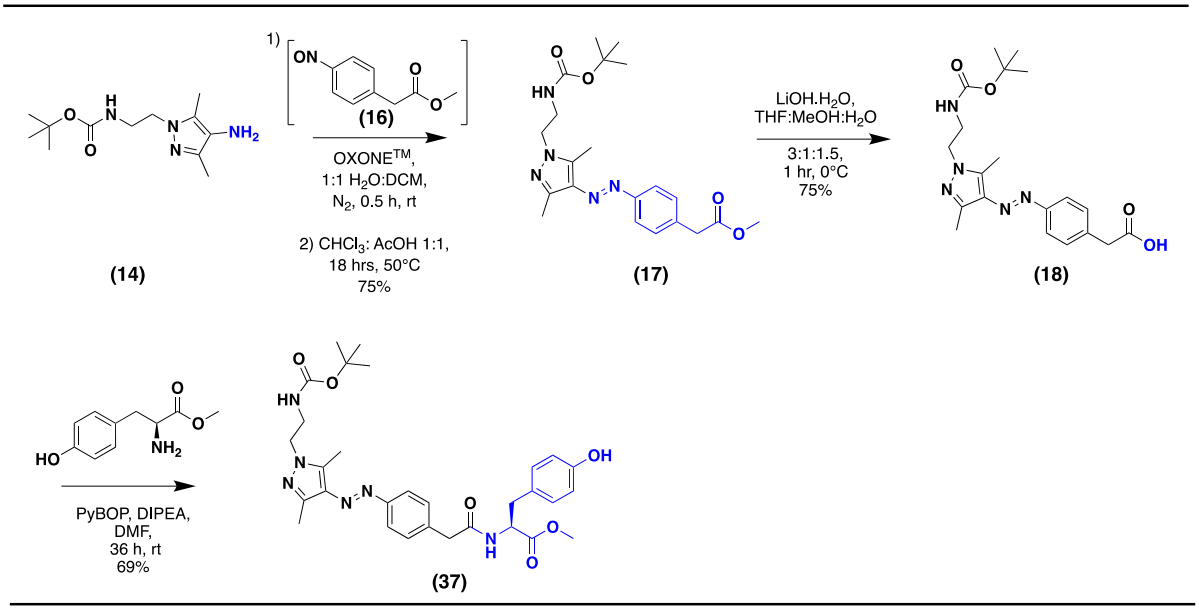


ClAc-Photo-^{L}Y-CME (36)

Figure 38: ClAc-Photo-^{L}Y-CME (36).

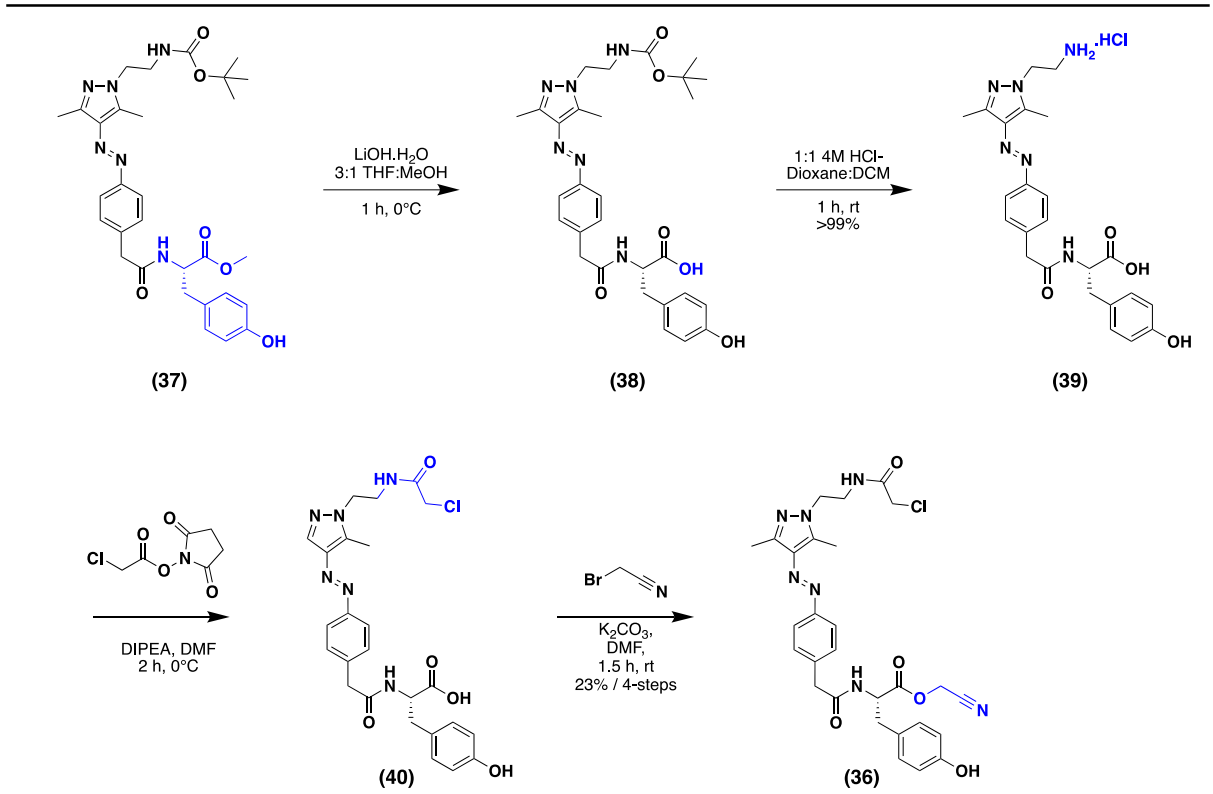
The synthetic route to ClAc-Photo-^{L}Y-CME (36) demonstrates the most optimised conditions to arrive at the common photoswitch intermediate (18), as developed throughout this chapter. *N*-alkylation and reduction of the nitropyrazole were achieved under the basic alkylation and the improved activated Zn reduction conditions as described previously. This generated the desired product (14), in respectable yields with a clean and facile workup (71%). Most prominently, further optimisation of the Bæyer-Mills coupling conditions (Scheme 14), specifically in the oxidation step, was achieved during this iteration of the synthesis. The use of an increased excess of the nitroso precursor (from 1.2 eq. to 2 eq.), whilst in overhead darkness, use of aluminium foil around the reaction vessel, nitrogen atmospheres for both steps, and working promptly to extract the oxidised nitroso from the OXONE™ solution using DCM, produced (17) in one of the highest yields to date (75%). The combination of these modifications appeared to increase and stabilise the amount of nitroso intermediate available for coupling. This is because nitroso groups are typically prone to over oxidation and or photodegradation with air and light.

The methyl ester of (17) was then cleanly hydrolysed to produce the common photoswitch intermediate (18) robustly (Scheme 14). Tailored specifically for this analogue, methyl ʟtyrosinate was then coupled using the amide coupling reagent PYBOP with DIPEA in DMF for 36 h at room temp. This produced (37) with a yield of 69%.



Scheme 14: Synthetic route toward ClAc-Photo-ᴸY-CME (36).

Once again facile methyl ester hydrolysis and Boc deprotection using the standard reagents revealed the deprotected analogue (39) (Scheme 15), in assumed quantitative yields as followed by UHPLC-MS. (39), which was then chloroacetylated using activated chloroacetoxy succinimide with DIPEA in dry DMF for 2 h at 0 °C. This was carried through crude to the final activated esterification step, furnishing the CME group, producing the final product ClAc-Photo-ᴸY-CME (36), in a modest yield of 23% over four steps.



Scheme 15: Continued synthetic route to the final product ClAc-Photo-^{L}Y-CME (36).

2.2.8 Synthesis of ClAc-Photo-Propyl-CME (41):

In parallel, a phenyl propanoic analogue (ClAc-Photo-Propyl-CME (41)) was chosen as an alternative scaffold for eFx Flexizyme recognition (Figure 39).

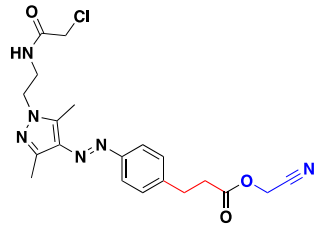
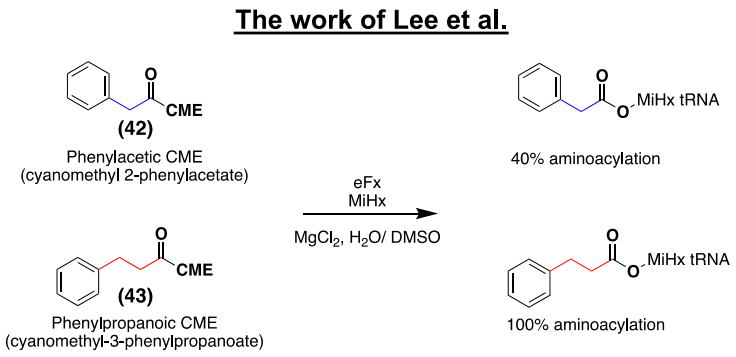


Figure 39: ClAc-Photo-Propyl-CME (41).

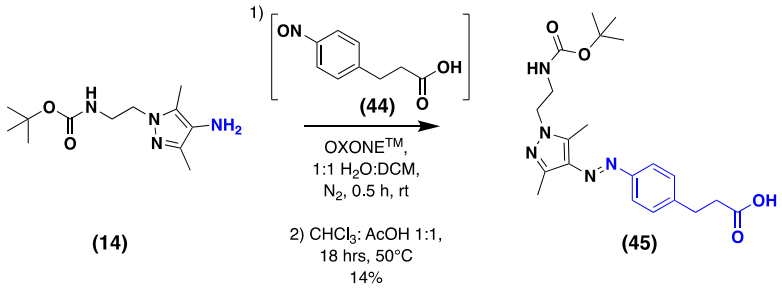
Once again using guidance from Lee et al.,[96] Lee et al. showed significantly improved aminoacylation yields of 100% with this phenyl propanoic base scaffold in comparison to the base phenylacetic scaffold used thus far in the project. The aminoacylation efficiency seen (Scheme 16), between the cyanomethyl 2-phenylacetate (42) (40%) and cyanomethyl 3-phenylpropanoate (43) (100%) is quite striking. It was initially rationalised that despite the lower efficiency of the cyanomethyl 2-phenylacetate (42), the one-fewer CH_2 might be more advantageous for causing maximal structural change on photoswitching. However, with the challenges seen thus far, this propyl analogue (43) if compatible, would be very desirable in comparison to ClAc-Photo-^{L}Y-CME (36) which is much bulkier. This would make ClAc-Photo-Propyl-CME more structurally like the desired but unsuccessful first analogue of ClAc-Photo-CME (11).



Scheme 16: Aminoacylation yield between Phenylacetic (42) and phenylpropanoic (43) cyanomethyl ester loading on MiHx FAM-23b tRNA identified by Lee et al..[96]

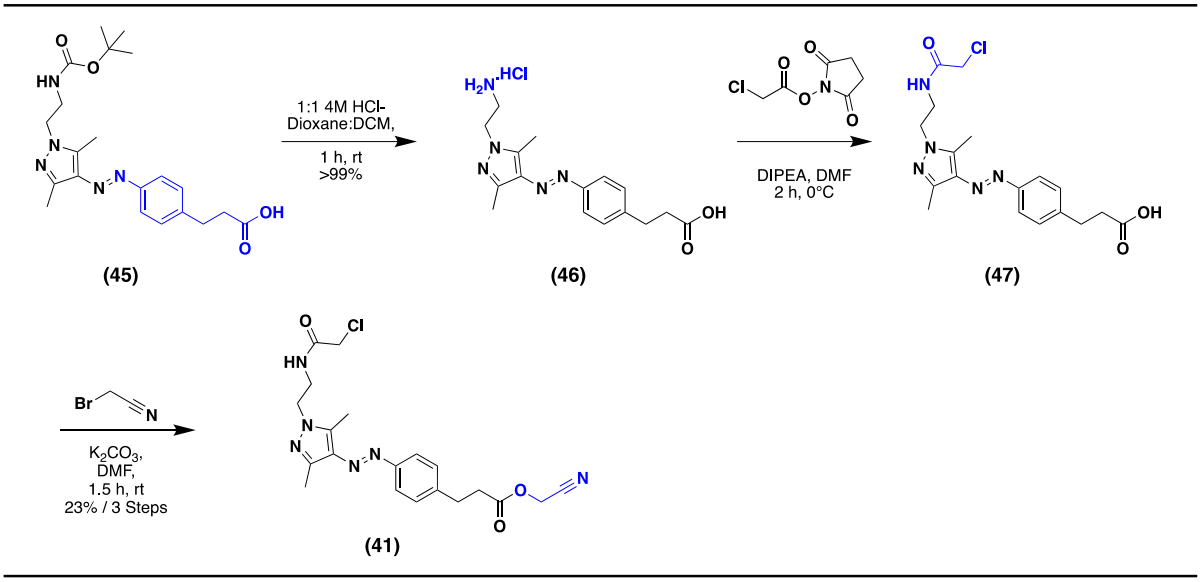
ClAc-Photo-Propyl-CME (41) was synthesised using a similar approach to the previous analogues. However, in the Bæyer-Mills coupling, the precursor to the nitroso methyl ester (16) used for the acetyl analogues, was substituted for its nitroso-propyl equivalent (44), 3-(4-aminophenyl)propanoic acid (Scheme 17).

The Bæyer-Mills coupling using this propyl derivative generated (45) in a significantly lower yield than previously seen with the optimised conditions for the acetyl counterpart (14% vs 75% yield for ClAc-Photo-CME (36) (Scheme 14)). As the introduction of an additional CH_2 would most likely not explain this drop in coupling efficiency, the lower yield was attributed to the free carboxylic acid. This is because it was most likely more soluble in the aq. phase during nitroso formation, resulting in increased over-oxidation to the nitro and reduced coupling to produce the desired diazo compound (45). The 3-(4-aminophenyl)propanoic acid (amine precursor to (44)) was used to speed up the synthetic route, however, it appears that the methyl ester is particularly beneficial for generating a high-yielding coupling of the azo bridge. If ClAc-Photo-Propyl-CME (41) was required to be re-synthesised in bulk, the methyl ester of the reagent (methyl ester, amine precursor to (44)) for this Bæyer-Mills coupling step would be used, as it would most likely give much higher yields.



Scheme 17: Bæyer-Mills coupling towards the synthesis of ClAc-Photo-Propyl-CME (41). Lower yields were identified (14% vs 75% acetyl), attributed to the presence of the free carboxyl acid of the nitroso (44) which led to over-oxidation.

Following this, Boc-deprotection, chloroacetylation and cyanomethyl esterification were performed under standard conditions to generate ClAc-Photo-Propyl-CME (41) in a final yield of 23% over three steps (Scheme 18).



Scheme 18: Final synthetic steps to generate the propyl photoswitchable analogue ClAc-Photo-Propyl-CME (41).

2.2.9 Aminoacylation efficiency of ClAc-Photo-^{L}Y-CME (36) and ClAc-Photo-Propyl-CME (41) via MiHx assays:

With ClAc-Photo-^{L}Y-CME (36) and ClAc-Photo-Propyl-CME (41) in hand, aminoacylation screening was carried out to assess the aminoacylation efficiency of each analogue. Excitingly, ClAc-Photo-^{L}Y-CME (36), Figure 40(a), was the first example of the successful loading of a photoswitch analogue onto a MiHx. Several of the trialled conditions were able to achieve comparable aminoacylation efficiencies, all being in the low 50% range (lanes 3-8, Figure 40). Pleasingly the ClAc-Photo-Propyl-CME (41) analogue was also successfully loaded in up to a 41% yield (lanes 4, 7, Figure 40(b)).

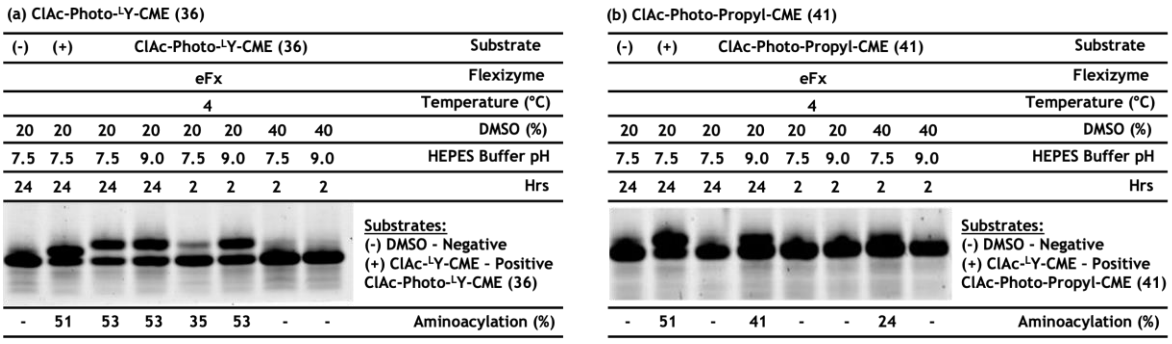


Figure 40: Combined panel MiHx Assay conditions for a) ClAc-Photo-^{L}Y-CME (36) and b) ClAc-Photo-Propyl-CME (41) where both gels indicate successful aminoacylation onto a truncated FAM-23b synthetic tRNA.

These gels demonstrated the first successful loading of two related photoswitch analogues onto tRNA, with ClAc-Photo-^{L}Y-CME (36) displaying slightly enhanced loading efficiency. The loading efficiencies of 53% and 41% respectively, are sufficient for the progression of these loaded tRNAs into an *in vitro* translation system such as FIT. This is as the loaded tRNAs are kept in excess to the rest of the translation mix and any unloaded tRNA (i.e., the remaining 47%/ 59%) remains largely inactive during translation.

2.3 Translation and MALDI of a known peptide sequence using photoswitch loaded $^{Ini}tRNA_{CAU}$:

With successful aminoacylation of both ClAc-Photo-^{L}Y-CME (36) and ClAc-Photo-Propyl-CME (41) onto a truncated tRNA (FAM-23b), aminoacylation onto a synthetic full-length tRNA was then required so that this aminoacylated tRNA could be used for FIT under genetic code reprogramming. As such, $^{Ini}tRNA^{Glu}_{CAU}$ was the tRNA chosen for reprogramming. $^{Ini}tRNA^{Glu}_{CAU}$ is a synthetic Glu tRNA stem with engineering of the anti-codon to CAU (as mentioned in Chapter 1.5).[127,128] CAU is the reverse complement anti-codon to the mRNA codon AUG which encodes for formyl Methionine translation initiation. This is important as the RaPID libraries, and consequently, the photoswitch analogues were designed such that they will always act as the initiation amino acid; promoting cyclisation downstream using the chloroacetyl moiety with cysteine to form head-to-side chain macrocycles. An alternative stem to $^{Ini}tRNA^{Glu}_{CAU}$ is $^{Ini}tRNA^{Asn}_{CAU}$, which is an engineered tRNA originally from an asparagine tRNA stem. These stems are often used interchangeably in the field, with the Glu stem routinely used as the primary stem in the Walport lab. As such $^{Ini}tRNA^{Glu}_{CAU}$ was arbitrarily chosen for the project without issue.

Aminoacylation of ClAc-Photo-^{L}Y-CME (36) and ClAc-Photo-Propyl-CME (41) onto the $^{Ini}tRNA^{Glu}_{CAU}$ tRNA stem was undertaken using the optimised loading conditions identified from the MiHx assay panel. This loaded tRNA was then introduced into the FIT mix and provided with mRNA encoding a known peptide sequence under genetic code reprogramming. Finding conditions to promote successful aminoacylation of the ns-AA onto the tRNA is often a challenge, however, ensuring that the ns-AA can be successfully translated by the ribosome (i.e. that the ns-AA does not block the ribosome tRNA entry channel) is also a critical reason for performing the FIT and MALDI. However, we were confident that the photoswitch would be readily accepted by the ribosome as large motifs such as helical aromatic peptide foldamers, have been successfully loaded and translated in the literature.[107]

Pleasingly, both ClAc-Photo-^{L}Y-CME (36) and ClAc-Photo-Propyl-CME (41) could be successfully translated by the ribosome into a known peptide sequence. This was confirmed by the identification of peaks corresponding to the observed mass $[M+H]^+$ of 2374.10 for ClAc-Photo-^{L}Y-CME (36) and $[M+H]^+$ of 2225.20 for ClAc-Photo-Propyl-CME (41) using MALDI-TOF mass spectroscopy (MS) (Figure 41). ClAc-Photo-^{L}Y-CME (36) was translated as the major species with the other major peak representing non-incorporation (replacement of the Photoswitch with fMet) (Figure 41(a)). This formyl methionine-initiated peptide can be observed as, despite removing methionine from the amino acid mix, there remains trace amounts within the IVT mix as it is purified from *E.coli*. In comparison, when ClAc-Photo-Propyl-CME (41) was translated, it was seen as the minor peak relative to the formyl methionine peak (Figure 41(b)). This qualitatively demonstrates that despite the comparable aminoacylation loading efficiencies between the two analogues (Figure 40), ClAc-Photo-^{L}Y-CME (36) could be successfully translated by the ribosome with greater efficiency. The reasoning for this is not immediately clear.

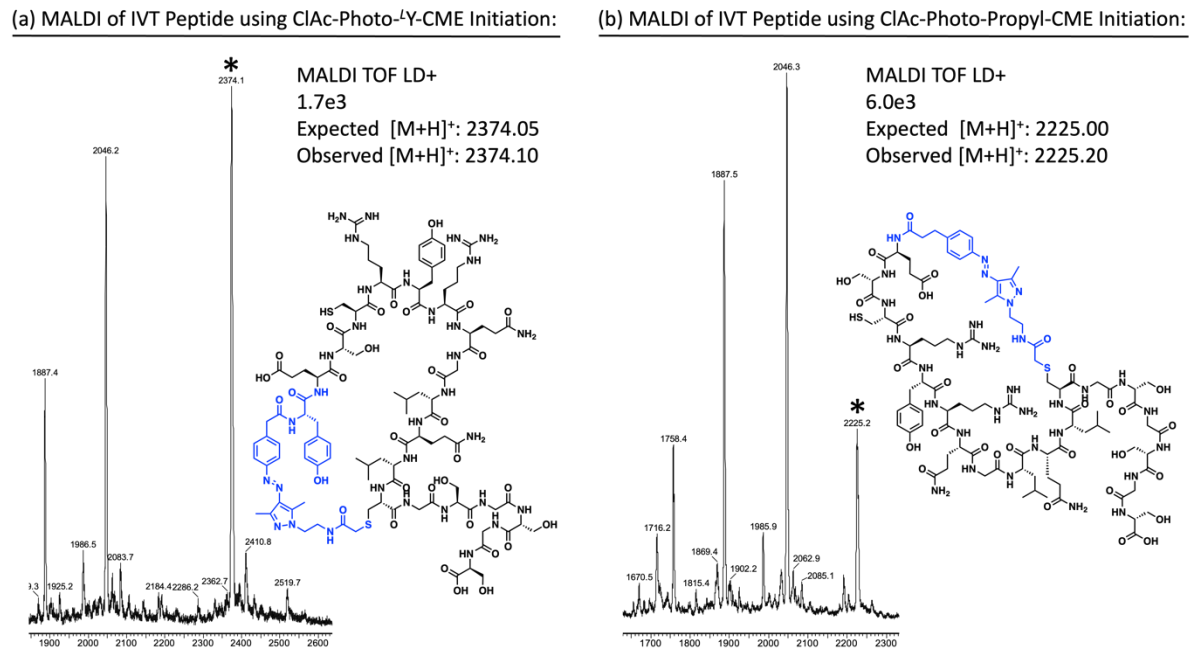


Figure 41: Successful translation of a photoswitch-loaded tRNA into a model peptide. a) Using ClAc-Photo-^{L}Y-CME (36) or b) ClAc-Photo-Propyl-CME (41) for translation initiation.

Both MALDI's also contain a mass corresponding to the first position *N*-terminal truncation of the formyl methionine/Photoswitch (1887.4, 1887.5 respectively). Typically, failure to initiate leads to no

peptide translation, however second site initiation has been seen by Suga and co-workers, referring to it as amino-terminal drop-off-reinitiation.[129,130] Within the spectra, the combination this with the formyl methionine incorporation observed, indicated that translation efficiency is substantially reduced compared to translation with proteogenic amino acids. This, however, is not uncommon for genetic code reprogramming. As these photoswitchable ns-AA were able to be successfully incorporated, even if that incorporation is not always the dominant species in every case, it was enough for use within RaPID selections. Particularly as, the counter screening measures discussed in the next chapter, aimed to circumvent any non-light-responsive binding peptides from enriching.

2.4 Conclusions

In conclusion, novel synthetic routes were developed to generate genetic-code-reprogramming compatible photoswitches. Using a convergent synthetic route enabled facile derivatisation to generate all the analogues, with substantial optimisation of the pyrazole nitro-reduction step as well as the Bæyer-Mills coupling. This resulted in procedures that were robust, reproducible, and of good yield. Five photoswitchable amino acid analogues were generated and two, ClAc-Photo-^{L}Y-CME (36) and ClAc-Photo-Propyl-CME (41), demonstrated substantial aminoacylation efficiencies with truncated tRNA (FAM-23b). Noticeable differences in translation efficiency by the ribosome were, however, highlighted. In the case of ClAc-Photo-Propyl-CME (41), relating to other work by Suga and Iwane et al.,[97] (Chapter 1.5.1), developing alternative affinity stems for initiation ($^{Ini}Met_{CAU}$), in a similar fashion to the varying elongation affinity tRNA stems developed to EF-Tu, would be ideal to see if the translation efficiency could be improved. This optimisation would be required against the ribosomes P-site directly though (if possible), as EF-Tu does not mediate initiation tRNA entry to the ribosome as previously discussed. This could be ideal for future work to enable progression further with the ClAc-Photo-Propyl-CME (41) analogue but remains unprecedented for initiator tRNA. As such, ClAc-Photo-^{L}Y-CME (36) was chosen as the optimal analogue to progress with, as it demonstrated significantly improved translational efficiency in recombinant FIT when qualitatively assessed using MALDI-TOF MS. This represents the first reported example of genetically encoded photoswitch compatible within mRNA display.

3 PHOTORAPID V1.0: DEVELOPMENT AND TESTING OF THE SYSTEM

Chapter Aims:

This chapter describes work aimed at developing and testing a next-generation platform for the discovery of *de novo* light responsive-cyclic peptides (PhotoRaPID), evolving enriched peptide hits with switchable binding between the two isomers, *E* and *Z*, of the molecular photoswitch. Human Peptidyl Arginine Deiminase 2 (PADI2) was the protein selected as a proof-of-concept target, with a view to expand the methodology to other targets in the future.

Key aspects of the chapter:

- Light irradiation and its impact on the encoded peptide mRNA/ DNA hybrid library.
- The relevance of PADI2 as the target of interest, the expression and purification of the protein, and confirmation that it is suitable for RaPID screening.
- The three irradiation workflows developed, to identify the best for enriching light-responsive peptide binders against PADI2.
- The first PhotoRaPID (v1.0) selection performed and identification of the enriched binding peptides.
- Synthesis of the enriched binding peptides by solid phase peptide synthesis and characterisation of the hits by biophysical, biochemical, and photochemical approaches.

A visual representation of the chapter workflow is shown below (Figure 42).

Chapter Workflow:

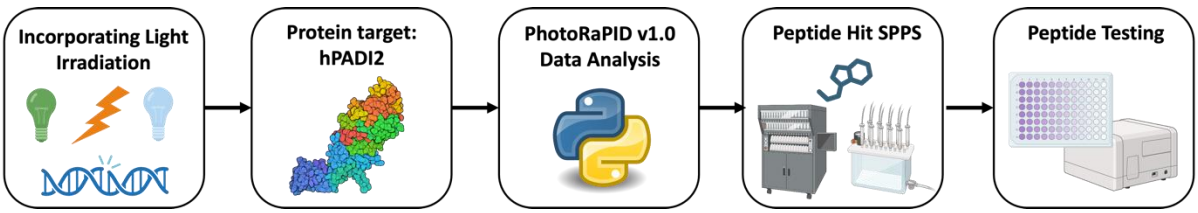


Figure 42: Workflow of the Chapter.

3.1 Developing the components of the PhotoRaPID v1.0 system:

With the successful integration of a compatible photoswitch within the FIT system, methodology to enable access to light-responsive cyclic peptide binders could then be developed as per the chapter aims.

3.1.1 Assessing the effect of UV-irradiation on the mRNA/ DNA hybrid RaPID library:

As mentioned in the introduction (Chapter 1.2.5), arylazoheteroarene photoswitches are small molecules that can isomerise with light to either *E* or *Z* geometry about their azo bridge. The wavelengths used to promote the transition between these two states for the azopyrazole class fall within the UV range, using 365 nm for the *Z*-to-*E* (π-π^*) transition or near-UV range, and 495 nm for the *E*-to-*Z* (n-π^*) transition. It is well known that UV-irradiation for long periods of time or with high intensity can result in DNA damage, resulting in nucleotide mutation.[20] It is critical within an mRNA display platform that mutation of the encoding nucleic acid sequences does not occur, as the DNA 'barcodes' are used to deconvolute the peptide sequences post-selection.

To confirm this, an mRNA sequence encoding for a single peptide was reverse transcribed to an mRNA/ DNA hybrid using the standard reverse transcription protocol for RaPID. The mRNA/ DNA hybrid was then subjected to irradiation at 365 nm for 0, 0.5, 1, 3, 5 and 15 minutes, taking aliquots at each time point (Figure 43(a)). 365 nm was chosen over 495 nm, as the shorter wavelength, with higher frequency of energy photon emissions was more likely to cause DNA damage. Following this, the mRNA/ DNA hybrid sequences were amplified by PCR using Taq Polymerase and sent for NGS. 365 nm UV-irradiation had a negligible effect on the percentage recovery of the known mRNA/ DNA hybrid sequence for up to 15 minutes of irradiation (Figure 43(b)). This data suggests that UV-irradiating the RaPID library would not be detrimental to mRNA display function. This now enabled the design of light irradiation workflows as a selection pressure to generate light-responsive cyclic peptides.

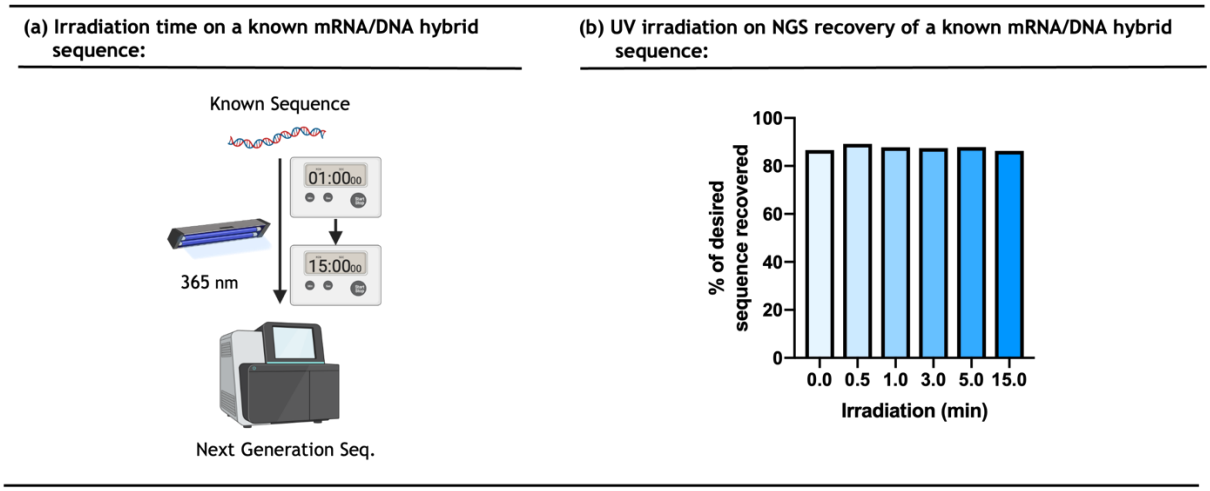


Figure 43: Assessing the effect of UV-irradiation on the mRNA/ DNA hybrid RaPID Library: (a) Workflow irradiating a known the mRNA/ DNA hybrid encoding peptide sequence for various time points up to 15 mins. Sequence deconvolution was achieved vis NGS. (b) NGS sequencing data to confirm that UV-irradiation did not cause mutation of a known mRNA/ DNA hybrid sequence.

3.1.2 Macrocycle library design for PhotoRaPID v1.0:

Initial library design was also an important factor when attempting to identify cyclic peptides with differential binding between the two light-responsive isomers of the Molecular photoswitch. For this work, Library design relates to (1) the fixed position of (unnatural) amino acids within the library and (2) the mixed macrocycle ring size libraries used for the selection.

(1) The fixed position of (unnatural) amino acids within the library: As mentioned previously (Chapter 1.6.3), the RaPID-compatible Photoswitch was designed to be introduced by reprogramming into the fixed initiator $^{f}Met_{CAU}$ position, as the first amino acid translated in every macrocycle in the library. Therefore, to enable full cyclisation of the library, the library was designed such that Cys was always fixed as the final amino acid after the randomised region to ensure cyclisation of all peptides.

(2) The mixed macrocycle ring size libraries used for the selection: As switchable binding is desired between the two light-responsive isomers of the molecular photoswitch, the impact of macrocycle ring size on the global structural dynamics of the ring upon irradiation must be considered. If too small, this may result in macrocycles that are either too constrained to switch or may lock or reverse the thermodynamic profile of the photoswitch. This is seen within severely constrained small molecule bridged azobenzene derivative 5,6-dihydrodibenzo[c,g][1,2]diazocine photoswitches in the literature.[131] By contrast, for these larger RaPID derived peptidic macrocycles, if too large, the

structural changes imposed by the photoswitch may not affect the structure of the whole of the ring as seen by Heinis and co-workers.[122] This may result in poorer switchable binding as the binding confirmation could be partially retained even after irradiation. RaPID libraries typically encode between 4 and 14 amino acids in the randomised region of the peptide and are designated NNK4-14 to represent this.[103,113] Based on these considerations and the assumption that the influence of the photoswitch upon the macrocycles was presumed to be highly sequence dependent, for this initial PhotoRaPID v1.0 selection, a mixed library of macrocycles containing variable positions ranging between NNK6-12 was chosen. The use a mixed library of NNK6-12 such as this has also been effectively implemented for previous novel targets by the group.[108]

3.1.3 The proof-of-concept target PADI2:

For the model protein target of interest, human peptidyl arginine deiminase type II (PADI2) was chosen.[132] PADI2 is a member of the PADI (peptidyl arginine deiminase) family of proteins, a family of enzymes that catalyse the reaction of peptidyl arginine (positively charged at physiological pH), to citrulline (neutral at physiological pH) within substrate proteins (Figure 44(a)).

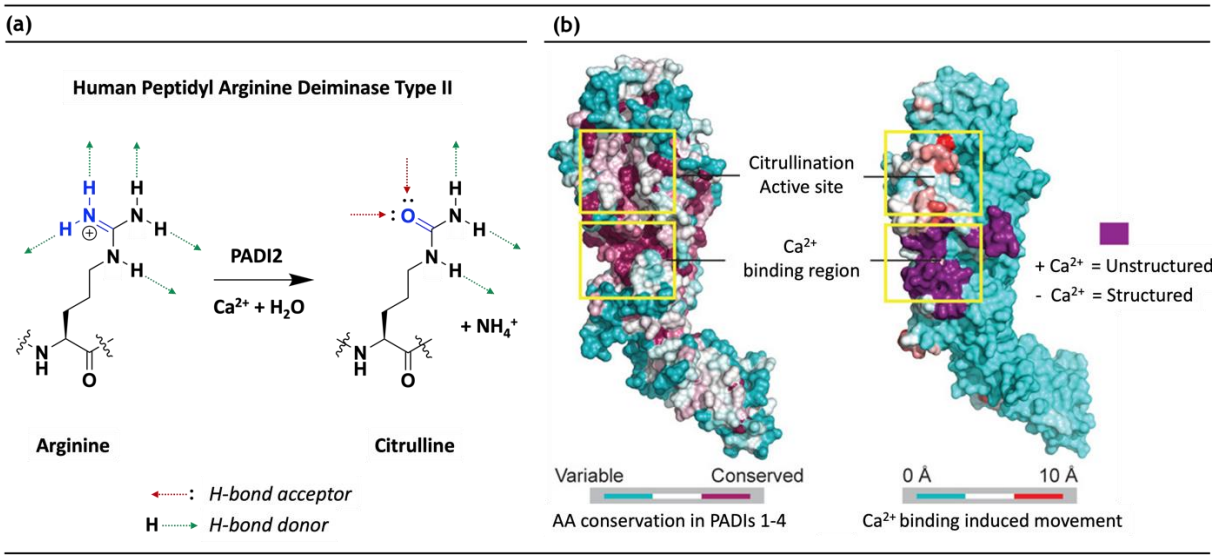


Figure 44: Human peptidyl arginine deiminase type II enzyme. (a) Post translation modification of peptidyl arginine to citrulline, with distinct changes in the number of hydrogen bond donors and acceptors. (b) Overlayed crystal structures of the PADI family members PADI1-4 highlighting the conserved active sites and calcium binding regions. Adapted from Thompson et al. (open access publication).[133]

This post-translational modification is called citrullination. In mammals, there are five identified isoforms (PADI 1, 2, 3, 4 and 6), sharing ~50% sequence homology.[134] PADI2 is the most widely

expressed family member, showing fairly ubiquitous expression with bias towards the colon, brain, salivary gland, and bone marrow tissues.[135] Not much is known about the endogenous activation mechanism, however, the binding of calcium, which is required for activation of PADI1-4, has been well studied.[133] Different PADIs require different coordination's of calcium binding for activation, and in the case of PADI2, 6 calcium atoms are required for activity (Figure 44(b)).[133,134] Citrullination results in a net change in charge and as such leads to differences in hydrogen bonding. This is because the guanidinium group within arginine has no H-bond acceptors and 4 H-bond donors, whilst the uronium group of citrulline has 2 H-bond acceptors and 3 H-bond donors. These changes in hydrogen bonding can cause distinct structural changes (changes in protein folding) of those protein substrates post-translationally modified by PADIs.

The endogenous substrate scope of PADIs remains challenging to probe as there is only a single mass change (+1 Da) between arginine and citrulline.[136] However, PADI dysregulation has been identified in several disease profiles.[136] For PADI2 specifically, these include neurodegenerative disorders such as Alzheimer's disease,[137] rheumatoid arthritis,[138] tamoxifen-resistant breast cancer,[139] transcriptional dysregulation in cancer,[140] as well as myelin loss in the central nervous system.[141]To date, very limited PADI2 specific chemical tools are available. One covalent selective PADI2 inhibitor and one non-covalent activator have identified within the literature. [142,143] Development of a photoswitchable pan-PADI inhibitor by Thompson and co-workers (Chapter 3.1.3.2) has also been undertaken.[144] Due to these considerations, the interest of the Walport lab on developing chemical tools to help understand the biological relevance of PADIs, as well as the availability in the lab of reagents and specific downstream biochemical assays, PADI2 was chosen as a proof-of concept Photo-RaPID target.

3.1.3.1 COLDER assay to measure PADI activity:

To assess PADI2 enzymatic function, a pre-established COLDER (Colour Development Reagent) assay developed by Vašák et al. has been reported in the literature. [145] The COLDER assay is a colourimetric plate-based assay used to assess PADI citrullination activity. The assay involves an acid-catalysed modification of the urea moiety within citrulline to produce a pink colour (Figure 45). A model arginine substrate BAEE (Nα-Benzoyl-L-arginine ethyl ester hydrochloride), is used as the substrate for this assay. Following conversion of BAEE, if any, to the citrulline derivative by the enzyme, it is further incubated with diacetyl monooxime in the presence of thiosemicarbazide, ammonium iron (III) sulfate and using phosphoric and sulfuric acids at high temperatures (95 °C). This generates a pink

colourimetric fluorescence.[145,146] The pink colour is quantified via a 96-well plate reader measuring absorbance at 540 nm.

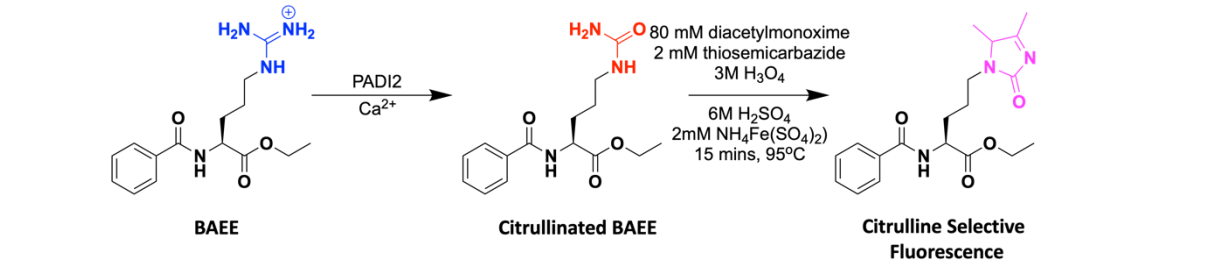


Figure 45: Reaction underpinning the COLDER assay. Citrullination of arginine substrate mimetic BAEE and subsequent visualisation using diacetylmonoxime, thiosemicarbazide, thiosemicarbazide, ammonium iron (III) sulfate and using phosphoric and sulfuric acids.

3.1.3.2 Current literature – Photoswitchable covalent PADI2 inhibitor:

Precedence has been set for the development of photoswitchable inhibitors against PADI2, with the azologisation of biphenyl,benzimidazole-chloro-amidine (BB-Cl-amidine), developed by Thompson and co-workers (Figure 46).[144]

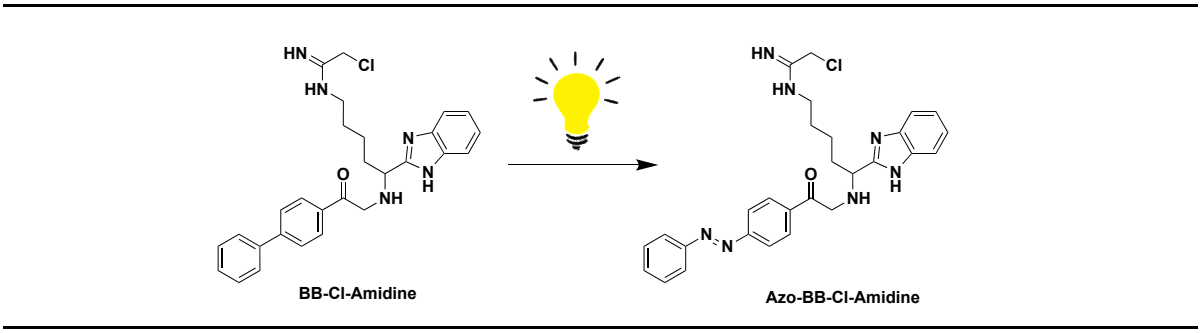


Figure 46: Photoswitchable inhibitor of PADI2, based on the azologisation of a known pan-PADI inhibitor.

BB-Cl-Amidine is a covalent pan-PADI inhibitor of the PAD family with a reported EC_{50} of 8.8 ± 0.6 µM against PADI2.[147] The best azologified derivative displayed an $_{Z}IC_{50}$ of 9.1 µM and an $_{E}IC_{50}$ of >100 µM, indicating a 10-fold inhibition differential against PADI2. Variations in potency and selectivity were also seen for the other PADI family members. For this PhD, this work highlights the potential for the development of photopharmacological agents against PADI2. However, due modest the potency, pan-activity, and covalent nature of this derivative, the photopharmacological applications are intrinsically limited (e.g., only one-time photoswitching (activation) is possible before covalent inhibition).

3.1.4 Protein expression and purification of PADI2:

To undertake a PhotoRaPID screen against PADI2, expression and purification of the protein was required. The plasmid construct used encoded for full length *PADI2* preceded by an *N*-terminal Avi tag and 6-His tag (*N-term*-Avi-6His-PADI2). An Avi tag is a 15 amino acid recognition motif (Gly-Leu-Asn-Asp-Ile-Phe-Glu-Ala-Gln-Lys-Ile-Glu-Trp-His-Glu) which is recognised and enzymatically biotinylated via the *E.coli* biotin ligase BirA.[148] This ligase enables the covalent attachment of biotin to the lysine (underlined) within a AviTag peptide. A 6xHis tag is a peptide motif of six sequential histidine residues which allows purification via immobilised metal affinity chromatography (IMAC). Transition metals such as Ni^{2+} are immobilised within a resin column matrix. His tags have high affinity binding to the hexa-His motif to the transition metal and enable capture of the desired protein of interest from the bacterial lysate. Elution of the bound protein is achieved through competition with imidazole (a mimetic of histidine).

PADI2 was first expressed using BirA-transformed chemically competent *Escherichia coli* BL21 (DE3) cells. This strain was chosen as it contains an IPTG-inducible (Isopropyl ß-D-1-thiogalactopyranoside) BirA expression plasmid to enable biotinylation of the desired PADI2 protein using the *N*-terminal Avi Tag, upon biotin supplementation. Following expression, the bacterial lysate was subjected to an IMAC nickel column, retaining the desired PADI2 protein indicated by the red box in Figure 47(i), at the expected mass of ~79 kDa. Size exclusion column chromatography (SEC) then proceeded, providing fractions of approximately the correct molecular weight (Figure 47(aii)). As the collected fractions correlated to two peaks with proteins of approximately the correct size (Coomassie staining), to identify PADI2, they were pooled separately based on molecular weight and activity assays were performed. Fractions 'B8, B9' and 'B12, C1, C2' were separately pooled, as indicated by the red box, and subjected to an *in vitro* COLDER citrullination assay (described above) (Figure 47(b)). Both pooled fractions were tested at three varying concentrations (250 nM, 500 nM, 1 µM) for citrullination activity. The 'B12, C1, C2' pooled fraction displayed far greater citrullination activity, indicating that this was most likely the desired fraction of interest, and that the protein was folded correctly for activity. In combination with the citrullination assay, high resolution TOF MS ES+ mass spectrometry was undertaken (Figure 47(c)). The mass identified (79,245.0 Da) differed from the expected mass (79,249.6 Da) by only -4.6 Da and as such it was concluded that this was indeed PADI2. Finally, the biotinylation status of the desired PADI2 was specifically confirmed via western blotting (Figure 47(d)). To selectively visualise the biotin on PADI2, a streptavidin-HRP (horseradish peroxidase enzyme) conjugate was used. A single chemiluminescent band at 79 kDa provided reassurance that only PADI2 would be captured by the RaPID selection magnetic streptavidin beads and during later biophysical

assays (e.g. surface plasmon resonance (SPR)). This was because despite two steps of purification, several impurities remained within the sample.

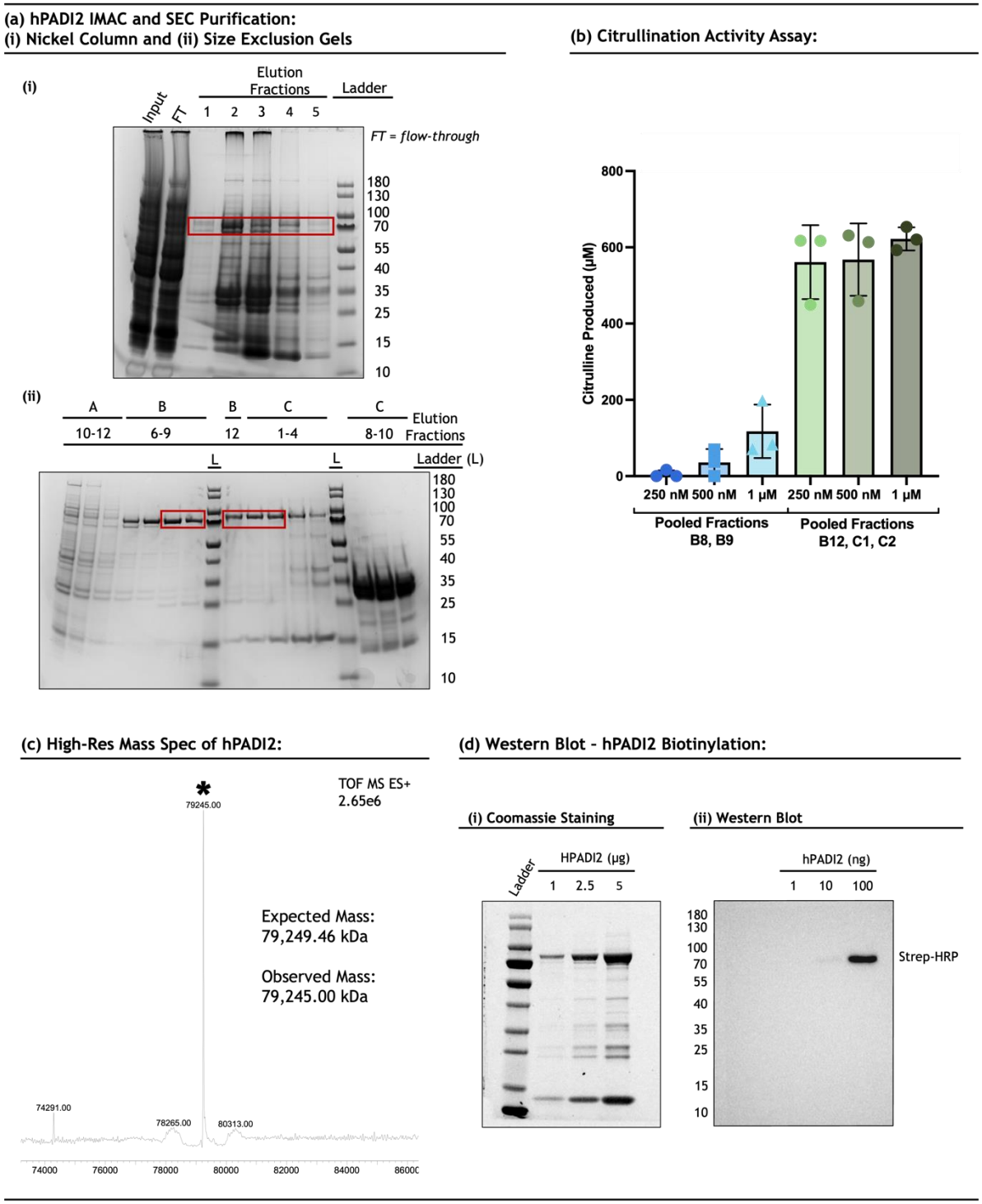


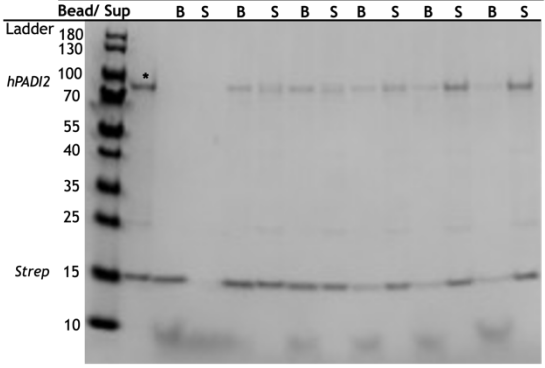
Figure 47: Expression and purification of PADI2 (a) Purification of biotinylated PADI2. Gels of fractions after (i) IMAC and (ii) SEC. (b) Citrullination activity of the pooled fractions after IMAC and SEC at differing concentrations of the pooled fractions of PADI2, (c) High-res mass spectrometry of PADI2 (full construct DNA sequence confirmed by nanopore sequencing), (d) Final biotinylation status and purity of PADI2, (i) Coomassie staining and (ii) western blot using Strep-HRP conjugate antibody.

3.1.5 Protein compatibility with RaPID:

3.1.5.1 PADI2 protein loading:

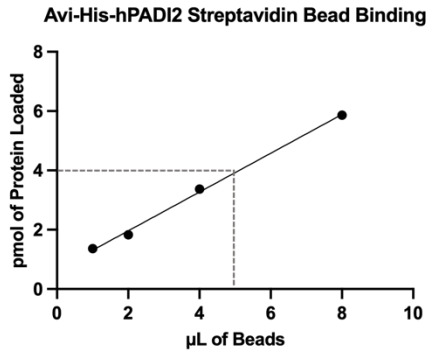
With PADI2 in hand, identifying the loading efficiency of the biotinylated protein onto the magnetic streptavidin beads via a bead binding assay was then undertaken. This is to ensure that the correct concentration of PADI2 is used when exposed to RaPID library to allow for peptide enrichment. As shown (Figure 48), 10 µL of 1 µM PADI2 was incubated with varying volumes of beads (dynabeads M-280) ranging from 1-16 µL prior to removal of the supernatant which was kept. Following washing, the protein bound beads were then boiled to recover the protein captured. The efficiency of protein loading was quantified via Coomassie stained in-gel imaging, assessing the proportion of the PADI2 recovered from both the beads and in the supernatant (Figure 48(a)).

(a) hPADI2 bead binding gel



(b) Tabulated/ graphical bead binding data

Volume of beads added (µL):	hPADI2 added (pmol):	% of hPADI2 on beads (fluorescence intensity):	Bead loading of hPADI2 (pmol/ x µL initial beads):
16	0	0	0
16	10	53	5.3
8	10	59	5.9
4	10	34	3.4
2	10	18	1.8
1	10	14	1.4



(c) Scaling for RaPID selection conditions

Selection round:	RaPID library volume (µL):	Required conc. of hPADI2 in RaPID library (nM):	Amount of hPADI2 required (pmol):	Amount of beads needed (µL):
1	~520	200	104	~130
2 onwards	~20	200	4	~5

Figure 48: (a) Gel of recovered PADI2 previously bound to the magnetic streptavidin beads, (b) quantification of the PADI2 loading efficiency based on in gel fluorescence of gel (a).

Plotting this data (Figure 48(b)), the top volume of beads (16 µL) was omitted as it appeared less efficient than the 8 µL beads (see tabulated data) and was assumed as general experimental error. In any case, plotting revealed that 5 µL of beads loads ~4 pmol of protein when using 10 µL of 1 µM PADI2 stock. As seen in Figure 48(c), the bead loading for 4 pmol of protein was chosen, as for every

round of the selection (excluding R1), 4 pmols of protein is required to ensure that the protein is at a f.c. of 200 nM in the library volume when incubated (R2-onwards library volume is ~20 µL). For R1, a much larger library volume is required to ensure sufficient sequence coverage. Therefore, the loading efficiency was then assumed to be linear and scaled accordingly.

3.1.6 Intrinsic PADI2-nucleic acid binding:

With PADI2 successfully loaded onto magnetic beads, testing the intrinsic nucleic acid binding of PADI2 was then undertaken. The purpose of this is to ensure that PADI2 does not bind non-specifically to the mRNA/ DNA hybrid library and only interacts with the displayed macrocyclic peptides. As such, a PADI2-nucleic acid binding assay was undertaken to see PADI2 binds to nucleic acids and if so, then to trial various conditions to reduce this non-specific library binding. A reverse transcribed mRNA/ DNA hybrid library was preincubated with bovine serum albumin (BSA) and or a combination of BSA with sheared salmon sperm DNA (s-ssDNA), prior to library mixing with immobilised PADI2 magnetic beads. BSA and s-ssDNA are known additives to reduce non-specific binding of nucleic acid and are typically used within mRNA display platforms, often described as the protein blocking solution.[149] After incubation, any non-specifically bound library was boiled off immobilised PADI2 and the proportion recovered relative to the initial input library was assessed via real-time quantitative polymerase chain reaction (qRT.PCR).

Table 3: Binding of selection library DNA to immobilised PADI2 under differing Blocking Solution compositions.

Blocking Solution composition (f.c.):	Sample:	Molecules of DNA recovered:	DNA recovered (%):
0.2 mg/mL (0.2%) BSA	Library Input	1.4×10^{11}	
	Negative: Strep beads	1.4×10^{6}	0.00104
	Positive: PADI2 Immobilised	5.3×10^{6}	0.00389
0.2% BSA, 2 mg/mL s-ssDNA	Library Input	2.3×10^{12}	
	Negative: Strep beads	1.1×10^{6}	0.00005
	Positive: PADI2 Immobilised	5.2×10^{7}	0.00230

Incubation of the libraries with only the streptavidin coated magnetic beads resulted in very low intrinsic binding as expected for both conditions (Table 3). Incubation of the library with PADI2 and either a blocking solution containing solely BSA (0.2 mg/mL (0.2%) BSA) or BSA with s-ssDNA (0.2%

BSA, 2 mg/mL s-ssDNA) resulted in increased recovery compared to the negative, but both being very low at 0.00389% and 0.00230% respectively. This highlighted a low intrinsic DNA binding of PADI2. Regardless, it is beneficial to reduce this binding as much as possible to prevent background DNA noise amplification during the selections. As such, because use of the s-ssDNA with BSA gave a very slight reduction in the non-specific binding of the library to PADI2 over BSA alone, it was chosen as the blocking solution for use during the upcoming photo-selections.

3.1.7 Designing the light irradiation RaPID workflows:

After assessing the effect of UV-irradiation on the mRNA/ DNA hybrid RaPID Library and ensuring compatibility of the other components of the RaPID system (Library design and PADI2 compatibility), irradiation selection workflows could then be considered. The aim of the irradiation condition workflows was to promote the enrichment of peptide binders against PADI2 with the largest switchable binding differentials possible. More specifically this binding differential is between the two light-responsive isomers of the molecular photoswitch integrated within the macrocycles.

Before the irradiation conditions could be devised, deciding which isomer would be the 'active' binding conformation for enrichment was necessary. Arguably, either isomer could have been chosen as seen in the previous literature (Chapter 1.8). For our applications, the *Z*-isomer was selected, as this form is the metastable conformation for azopyrazoles (Figure 49). This meant that the peptides would only be in their active conformation following irradiation and as such would not be constitutively 'switched-on' upon dosing. As the *Z*-metastable conformation the peptides also thermally revert to the ground *E*-state, the active peptides would therefore switch off over time following irradiation at a specific time of choosing. These factors mean that for downstream applications, fine-tuned temporal control can be more easily established during assays. This is by selectively activating and deactivating the light-responsive ligands in a precise, spatial, and temporal manner.

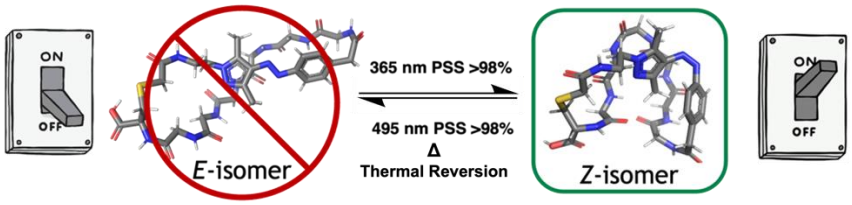


Figure 49: Choosing the isomer to be enriched for 'active' binding to the protein of interest.

As shown in Figure 50, three different irradiation workflows were proposed, all designed to promote enrichment of Z-state binding peptides within the selection library. A reference 'standard' RaPID workflow,[103] is also shown to allow for comparison to the modified irradiations conditions designed.

Irradiation Condition 1 was the least stringent condition planned, as it was designed to be most like the standard RaPID protocol. The selection library was first irradiated to the *Z*-isomer (365 nm, 3x 800 mW at 100% output, 3 mins) and three consecutive negative counter selections were performed against streptavidin magnetic beads loaded with ~50% of the sites occupied with biotin to remove non-specific bead binding peptides. Following this, the remaining library was recovered and re-irradiated to the *Z*-isomer to ensure that the Z-isomer PSS was reached again (to ensure that the library was fully in the *Z*-isomer). The *Z*-library was panned against immobilised PADI2 for 30 minutes. Following this the library was eluted from the PADI2-immobilised beads. The negative counter selection and *Z*-state positive selection binders were recovered after resuspending the beads in PCR mix and incubating at 95 °C for 5 minutes. The recovered DNA was then quantified by qRT-PCR, amplified by PCR and the *Z*-positive DNA was used to repeat the cycle or for preparation of NGS sequencing (negative DNA amplified only for NGS).

Irradiation Condition 2 was considered a more stringent approach by introducing a light-responsive counter selection as well. Condition 2 opted for no initial negative counter selections as it was assumed that screening and discarding the *E*-isomer library first would also deplete any non-specific binding peptides. The light-responsive counter selection was first undertaken by irradiating the library to the *E*-isomer (495 nm, 3x 750 mW at 100% output, 3 mins) and the *E*-library was panned against immobilised PADI2 for 30 minutes. After this, the library solution (library members not binding to PADI2) was then recovered and irradiated to the *Z*-isomer (365 nm, 3x 800 mW at 100% output, 3 mins). The light-responsive selection was then undertaken by incubating the *Z*-rich library for 30 minutes with a fresh set of PADI2-immobilised magnetic beads. Beads were washed and non-binding library members were then discarded, and the two sets of beads (the *E*-counter selection (*E*-negative), and the *Z*-positive selection) were resuspended in PCR mix and incubated at 95 °C for 5 minutes. The recovered DNA was then amplified by PCR and the *Z*-positive DNA was used to repeat the cycle or for preparation of NGS sequencing (*E*-negative counter selection amplified only for NGS).

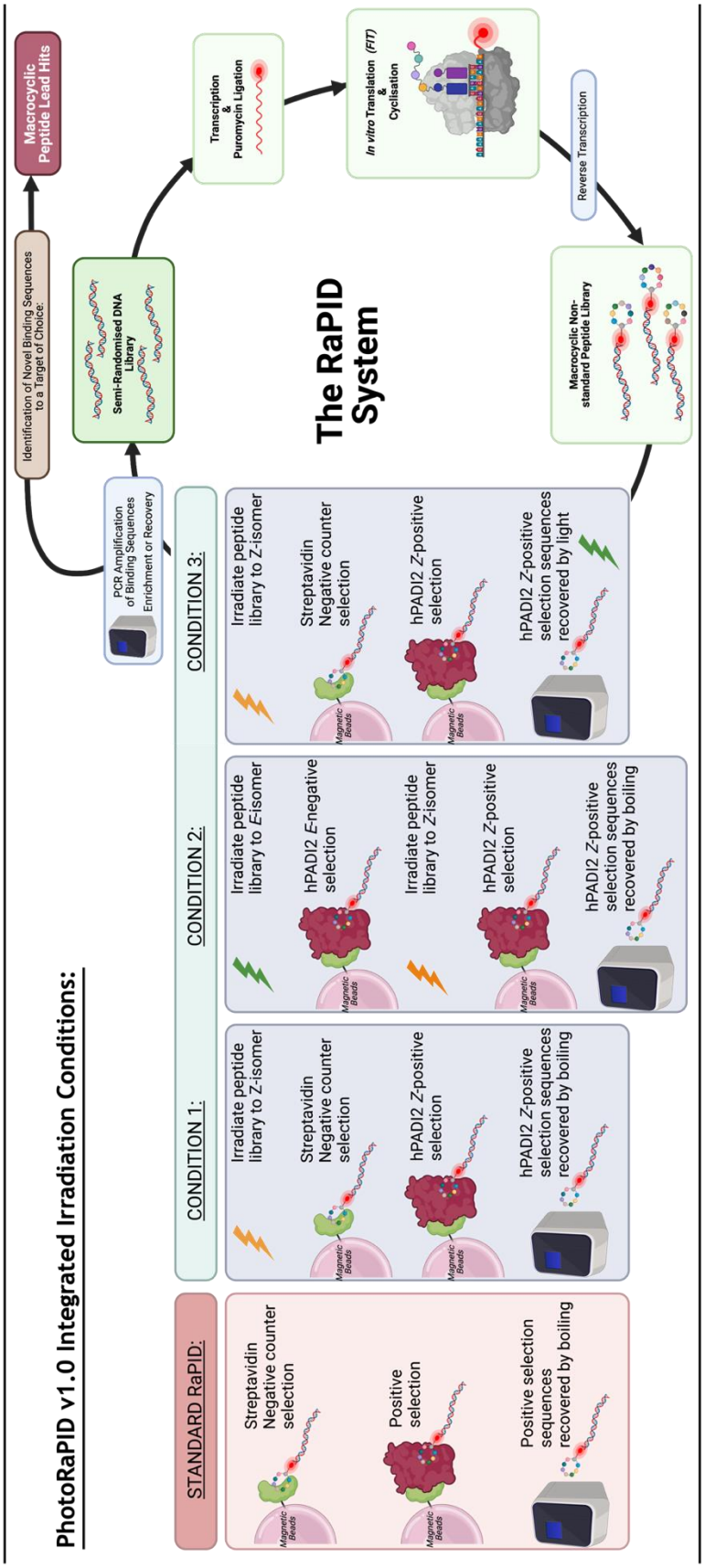


Figure 50: PhotoRaPID v1.0 Proposed irradiation conditions. Three irradiation conditions were trialled to generate light-responsive differentially binding peptide hits. A 'standard' RaPID workflow is also shown as a reference.

Irradiation Condition 3 was considered the most stringent approach by attempting identify selective peptides by irradiating binding peptides off the immobilised PADI2 magnetic beads by photoswitching. It was not known at the time if it was possible to switch small molecule photoswitches or photoswitchable macrocycles when bound to their target. This has since been shown through the work of Wranik et al.,[150] for a small molecule photoswitch and now provides support for this approach. In a near identical initial protocol to Irradiation Condition 1, the selection library was first irradiated to the *Z*-isomer (365 nm, 3x 800 mW at 100% output, 3 mins) and three consecutive negative counter selections were undertaken using streptavidin magnetic beads loaded with ~50% of the sites on the bead occupied with biotin. The remaining library was recovered and re-irradiated to the *Z*-isomer to ensure that the *Z*-isomer PSS was maintained. The *Z*-library was then panned against immobilised PADI2 for 30 minutes. Following this, the remaining library solution was discarded and to remove residual library, the PADI2-immobilised beads were then washed. The beads were resuspended in PCR mix. The beads in the PCR mix were then irradiated to the *E*-isomer (495 nm, 3x 750 mW at 100% output, 3 mins), mixed around and were removed, leaving only the PCR mix containing any peptides that had eluted from the protein following 495 nm irradiation. The 3x combined negative counter selection DNA was recovered via the standard boiling method. The recovered DNA was then amplified by PCR and the *Z*-positive DNA was used to repeat the cycle or for preparation of NGS submission (once again, the negative DNA was amplified only for NGS).

3.2 PhotoRaPID v1.0:

3.2.1 PhotoRaPID v1.0 selection – DNA recovery round-by-round:

Photo-RaPID v1.0 was undertaken, screening each irradiation condition (IrrC) 1-3 in parallel for six consecutive rounds of enrichment (Figure 51). To ensure sufficient library diversity the first selection round was performed on a large scale (520 µL). As a large volume of photoswitch conjugated-tRNA, initial DNA library, PADI2 and translation mix was therefore required, making this costly, only IrrC.1 was screened for Round 1 (R1). The recovered library from R1 (bold bars) was split into the three equal selection libraries for use each IrrC from R2 onwards. Each recovered library was kept isolated from one another during the remainder of the selection with each subjected to the same irradiation condition selection pressure during the entire course of the selection. For R1 a negative selection was also not performed. This is a typical strategy to ensure that the full library diversity is first panned against the protein of interest (PADI2).

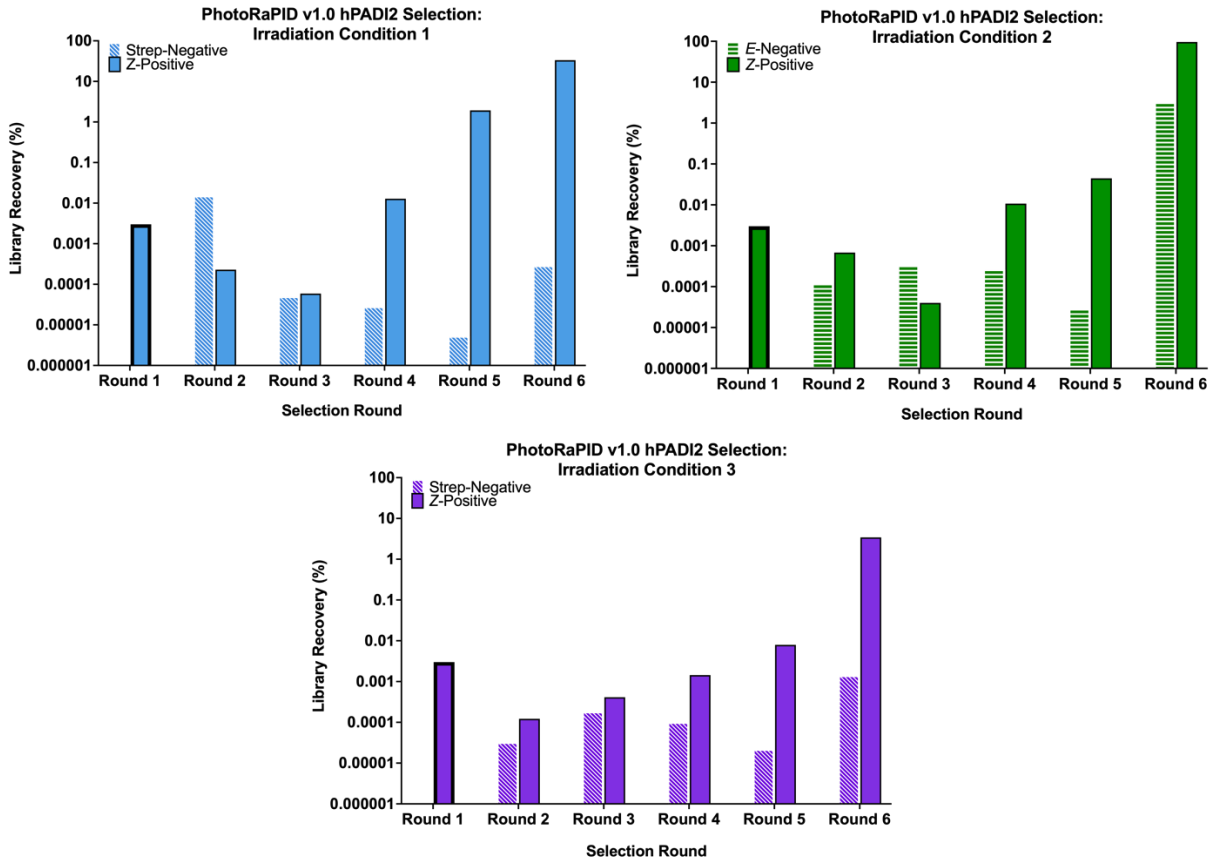


Figure 51: Combined panel of peptide enrichment against PADI2 assessed via DNA recovery round-by-round during the PhotoRaPID v1.0 selection. All three irradiation conditions were undertaken in parallel with the first round (bold) being the initial starting sample for all three conditions.

Following R1, only a small fraction of the total library DNA was recovered as desired at ~0.003% (Figure 51). This indicated that the selection library was not too promiscuous against the PADI2, and that no major non-specific binding contribution was initially occurring. This also indicated that the UV irradiation had not had some immediate and unforeseen negative impact on recovery. At R2, addition of the streptavidin negative counter selections resulted in further reductions in library recovery across all three conditions. This was also to be expected as by introducing the negative selections, further depletion of non-specific binding sequences from each of the enriching libraries occurred. By R3-4 all conditions begin to enrich indicating that binding peptide sequences were beginning to dominate the libraries, and thus be recovered to a greater extent round-by-round. Importantly, during these rounds, the negative control recoveries remained both low in terms of absolute value but also were consistently lower than the *Z*-positive enriching libraries. At R5, IrrC.1 had a substantial *Z*-positive DNA recovery at almost 2% of the library, whilst IrrC.2 and IrrC.3 each had only slight enrichment. This was thought to be due to IrrC.1 being the least stringent light-responsive selection pressure employed. By

R6, in IrrC.1, 33% of the *Z*-positive library was recovered post-round, which was considered quite high. In contrast, at R6, IrrC.2 was predicted to have a recovery of over 96% of the library which should not be possible as the puromycin ligation efficiency is only ~50% efficient. Only 50% of the input library should be able to be recovered. This recovery value indicated two possible issues, either that some combination of non-specific peptide/ DNA binding was occurring within the library or that the library input was under sampled. If assuming that the library was under sampled, as IrrC.2 had a 10-fold lower input that for IrrC.1 and IrrC.3 (Figure 51), this would mean that ~9.6% of the library was therefore recovered. This seemed more logical at the time. Repeat qPCR of the R6 samples in triplicate, still resulted in similar values of recovered DNA and as such it was assumed that IrrC.2 might produce questionable NGS data. In any case, with enrichment seen against PADI2, each round was prepared for NGS and submitted to identify the most enriched peptide sequences from each irradiation condition.

3.2.2 Post-PhotoRaPID v1.0 – NGS data analysis:

After undertaking the selection, the enriching peptide sequences recovered from each round were submitted for NGS sequencing. Previously developed in-house python scripts from the Suga group were used to translate the ORF of the recovered nucleotide sequences to reveal the encoded peptide sequences. The degree of peptide enrichment within the recovered libraries was provided by the NGS sequencing calculated using Eq. 3.2.2-1.

$$\left(\frac{number\ of\ NGS\ reads\ of\ a\ specific\ sequence}{total\ number\ of\ NGS\ reads}\right) * 100 \qquad \text{Eq. 3.2.2-1}$$

Following this, the libraries of enriched peptide sequences were then further processed using personally developed python scripts via the workflow as described in Figure 52(a). The peptide sequences were first loaded into a python dataframe and then filtered to remove any sequences that did not terminate in the fixed 'GSGSGSamber' motif. The remaining sequences were further filtered to remove any sequences that contained internal stop codons. Peptides containing internal stop codons do occur within the initial NNK library due to its design but can also result from mis-priming. In either case they are generally considered as non-specific binding sequences. The fixed 'GSGSGSamber' motif was then truncated as it only acts as a small spacer between the translated peptide and the conjugated DNA and typically does not contribute to binding. The sequence datasets were then condensed to the top 100 most enriched peptide sequences. In a new file, all peptides from

each condition were then cross-referenced against one another to identify unique sequences belonging to either IrrC.1, 2 or 3. The six processed datasets were then aligned using pair-wise sequence alignment (sequence homology analysis using QIAGEN CLC Sequence Viewer v8.0) to see if family sequence motifs were present for the most enriched peptide candidates (For scripts see Appendix 40).

As shown in Figure 52(b), six data sets of enriched peptides were generated. Shown, are the top 10 enriched sequences (TES) and the top 10 unique sequences (TUS) for each irradiation condition aligned by sequence homology. From this, and the top 100 most enriched peptide sequences, three main motifs across each of the three-irradiation conditions were qualitatively identified as 'WMD', 'HHW' and to a lesser extent 'IRL'. Sequence family motifs are where a highly enriched sequence also shares partial homology with other peptides dividing the sequences into distinct clusters. Presence of sequence family motifs is typically a good sign that this family has been enriching round-by-round and as such provides confidence that this may be a genuine hit prior to synthesis via SPPS. Initially, it was thought that the differing irradiation conditions might result in different motif clusters being identified across each irradiation condition. This was not seen in the actual sequencing data.

More surprisingly, even the 'unique' sequences (Figure 52(b) TUS sequences) appeared to be the result of single amino acid substitutions around the three family motifs previously described, rather than wholly novel motifs being recovered. On further reflection, it is perhaps logical that similar family motifs were recovered for all three conditions. This is because PADI2 only has a certain number of accessible binding surfaces. From this, as the core motif of the peptides remains largely consistent, it appears that the auxiliary amino acids (AA) may be responsible for any differences (if any) in the degree of light-responsive differential binding seen, rather than from contributions from the whole peptide motif.

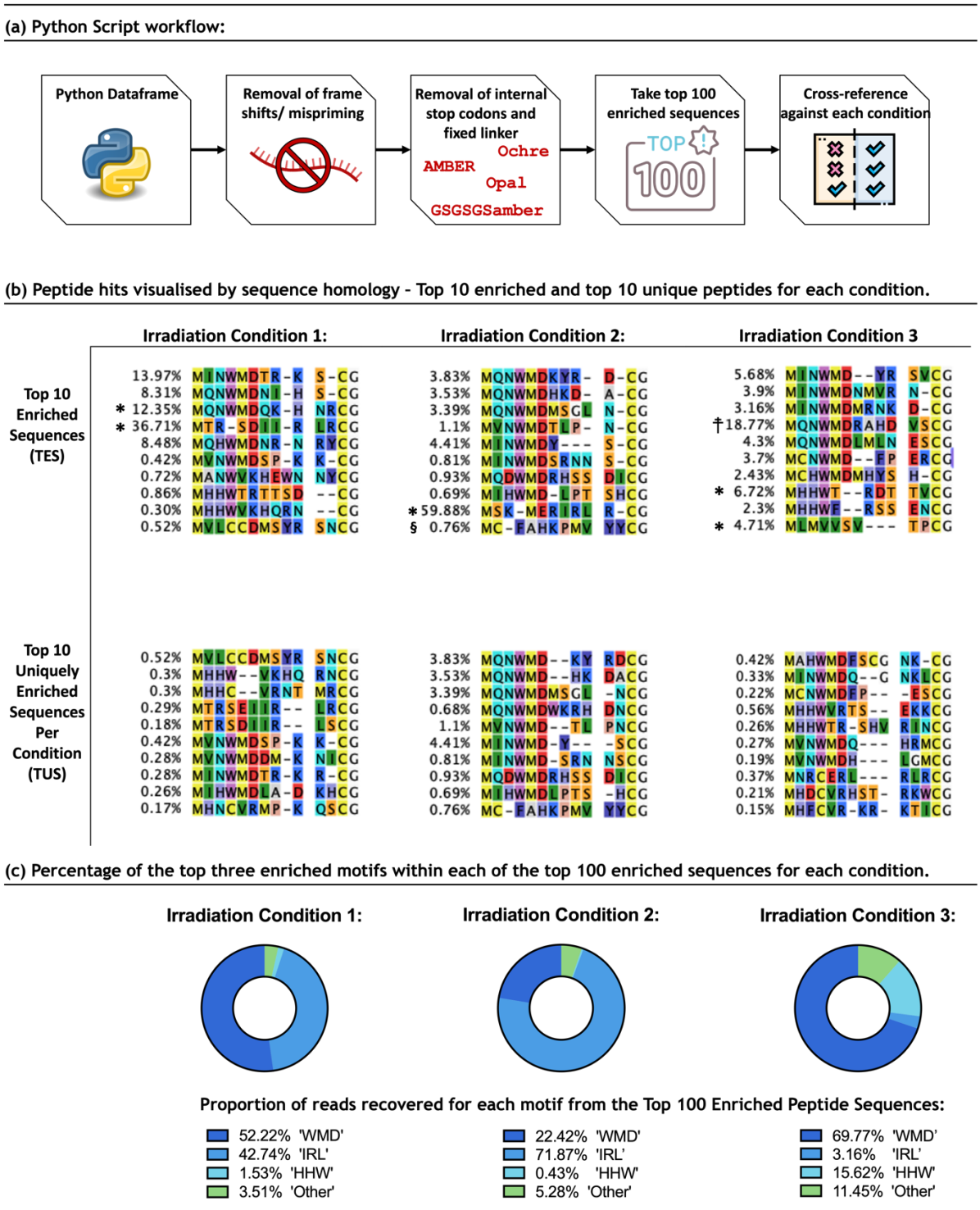


Figure 52: PhotoRaPID v1.0 Post-selection sequencing data: (a) Description of the data processing workflow to deconvolute the peptide sequences enriched from the NGS data. (b) Analysis of the enriched peptide hits identified across each of the three irradiation conditions during the first PhotoRaPID v1.0 selection. M is reprogrammed to Photo-^{L}Y via the FIT system during the selection. - peptides with an asterisk (*) were synthesised on SPPS as described in the next chapter. (c) The percentage of the top three enriched family motifs within each of the top 100 enriched sequences for each irradiation condition. 'WMD', HHW' and 'IRL' make up ~89-96% of the top 100 enriched peptides in each irradiation condition.

With the three dominant motifs qualitatively identified, the data was then interrogated in a more quantitative manner. This was to identify how representative these 'motifs' truly were across a 10-fold larger dataset. Pie charts were generated (Figure 52(c)), showing the proportion of reads recovered for each family motif within the top 100 Enriched Peptide Sequence datasets for each irradiation condition. To do so, further python scripts were generated, and the proportion of reads recovered for each motif were calculated as follows (Eq. 3.2.2-2):

$$For\ extracted\ sequences\ containing\ motif\ x\ in\ a\ new\ dataframe$$

$$\left(\frac{\Sigma \text{ of the reads of peptides containing } x}{total\ no.of reads\ in\ the\ top\ 100\ TES}\right) * 100 \qquad \text{Eq. 3.2.2-2}$$

$$Where\ x = 'WMD',\ 'IRL',\ 'HHW'$$

From the pie charts generated, it was confirmed that combined, the three dominant motifs qualitatively identified make up ~89-96% of the top 100 enriched peptides in each irradiation condition. The 'WMD' motif was the most prominent motif seen for IrrC.1 and IrrC3 accounting for 52.22% and 69.77% respectively, whilst 'IRL' dominated IrrC2 accounting 71.87% of the total reads. Preferential enrichment of 'HHW' within IrrC.3 (15.62%) was another key disparity, when compared to IrrC.1 (1.53%) and IrrC.2 (0.43%). Additionally, IrrC.3 also had a significantly reduced number of reads for peptides containing the 'IRL' motif, at only 3.16%.

3.3 Deciding and synthesising the peptides for further characterisation:

3.3.1 Choosing the specific peptides for further characterisation from the NGS data:

With the importance of these three family motifs solidified, selecting peptides for characterisation was then the priority. This was assessed by looking at the exact enrichment score of individual peptides within each family motif. Looking back at the top 10 TES peptides in more detail (Figure 52(b)), certain individual peptides could be identified as having high enrichment values. Those marked with an asterisk (*) (Figure 52(b)) were chosen for solid-phase-peptide synthesis (SPPS). The peptides (P001- P006) chosen are shown in greater detail (Figure 53(a/b)) and the reason for their choosing is described below.

Typically, the peptide chosen (*), was the most enriched family member within its respective motif. Peptides P001 and P003 were identified from IrrC.1 with individual peptide enrichments of 37% and 12% from the NGS data, respectively. P001 was the most enriched peptide and contained the 'IRL' family motif. P003 was also a dominantly enriched peptide within IrrC.1 and contained the 'WMD' family motif. Peptide P002 was chosen from IrrC.2 where its enrichment dominated, accounting for 60% of all sequences identified. An additional peptide P004, from IrrC.2, was also chosen (§). This was to promote sequence motif diversity in the chosen hits despite it being relatively poorly enriched (0.76%). Finally, P005 and P006 were both recovered from IrrC.3 where they had modest enrichment in this library at 5% and 7% respectively. P005 was chosen over the most dominantly enriched peptide (†) within that condition (cyclo-Photo-^{L}Y-QNWMDRAHDVSC(S-)G at 19% enrichment). This was because (†) contained the 'WMD' motif already being brought through as a candidate from IrrC.1 and in a similar fashion to P004, sequence motif diversity was prioritised. P006 was chosen as it was the second most enriched sequence from IrrC.3 and contain the dominant family motif 'HHW'.

No peptides were chosen from any of the irradiation conditions from the top 10 TUS. This reasoning was two-fold; firstly, despite them being classified as unique, almost all the peptides still retained one of the top three family motifs. This 'uniqueness' as stated earlier, arose from single amino acid substitutions and thus, were not particularly interesting to sample. Secondly, the individual peptide sequences were very poorly enriched when compared to the 10 TES.

(a) Peptide hits enriched against hPADI2 synthesised via SPPS from the deconvoluted NGS data (tabulated):

Peptide ID:	Peptide Sequence:	Total Ring Size:	NNK Ring Size:	Irradiation Condition:	Enrichment in Library (%):	Expected Mass: $[M+H]^+$ / $[M/2]^+$	Observed Mass: $[M/2]^+$
P001	Photo^{L}Y-TRSDIIRLRC(S-)G-NH_2	12	9	1	37	1775.07 888.55	888.6
P002	Photo^{L}Y-SKMERIRLRC(S-)G-NH_2	12	9	2	60	1835.21 918.11	918.0
P003	Photo^{L}Y-QNWMDQKHNRC(S-)G-NH_2	13	10	1	12	2003.24 1002.13	1002.5
P004	Photo^{L}Y-CFAHKPMVYYC(S-)G-NH_2	13	10	2	0.76	1902.46 953.23	Not observed
P005	Photo^{L}Y-LMVVSVTPC(S-)G-NH_2	11	8	3	5	1491.82 746.42	746.3
P006	Photo^{L}Y-HHWTRDTTVC(S-)G-NH_2	12	9	3	7	1798.99 900.00	899.9

(b) Peptide hits enriched against hPADI2 synthesised via SPPS from the deconvoluted NGS data (chemical structures):

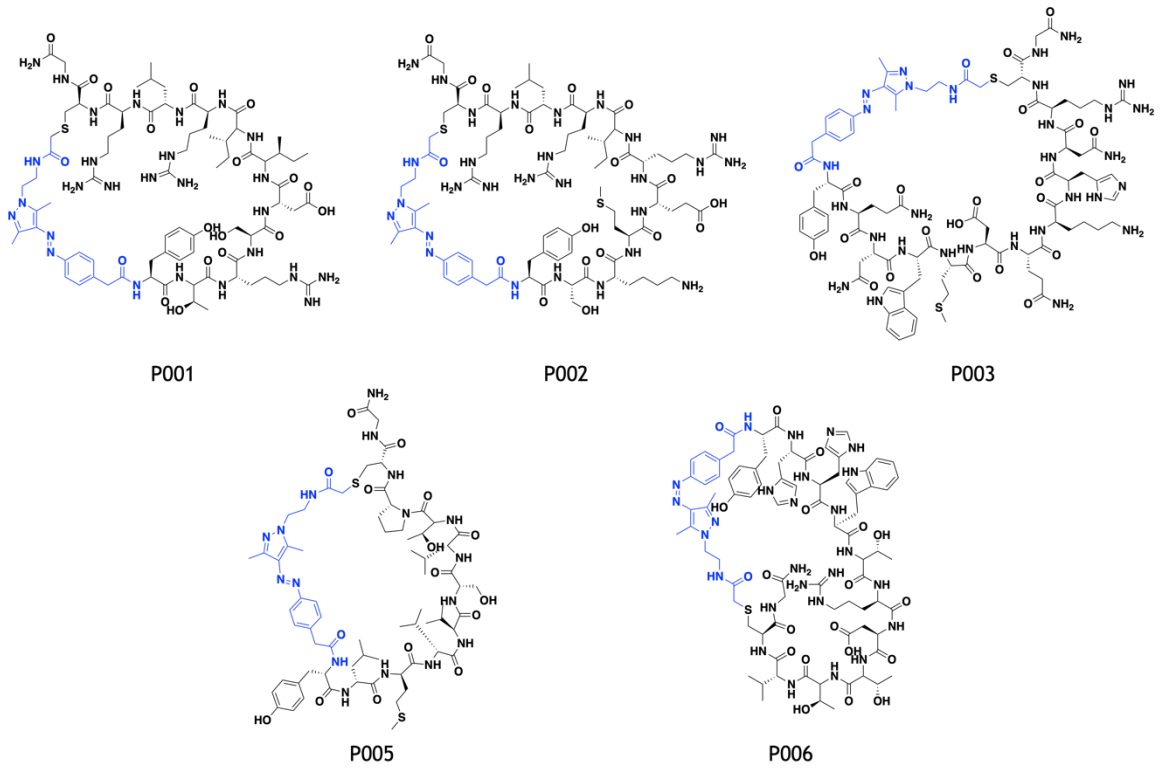
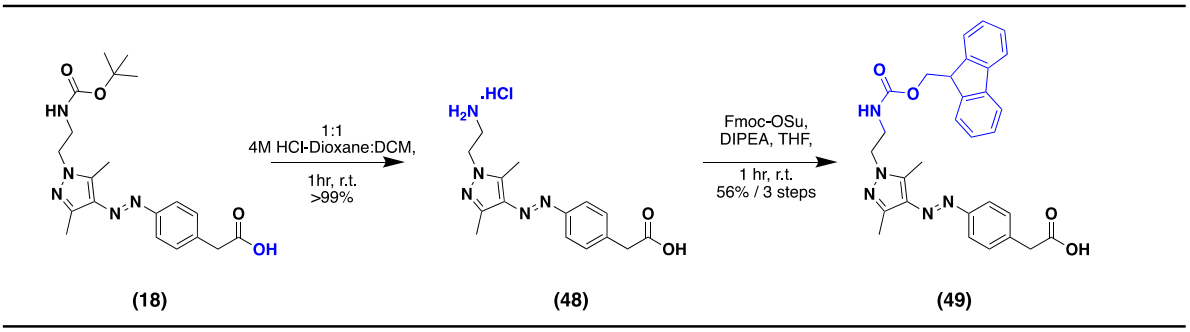


Figure 53: (a) Synthesised peptides selected from each of the three PhotoRaPID v1.0 irradiation conditions, (b) Expanded structures of the peptide macrocycles synthesised for biochemical testing. Bond angles are not represented accurately to enable cyclisation of the macrocycles in the 2D representations.

3.3.2 Designing a compatible photoswitch for SPPS:

To begin synthesis of the peptide hits via SPPS however, an SPPS-compatible analogue of the photoswitch ClAc-Photo-^{L}Y-CME (36), was needed. When looking at the structure of ClAc-Photo-^{L}Y-CME (36), for SPPS, several components were no longer required or could be added separately during SPPS. This was whilst still generating the same final photoswitch structure as in the PhotoRaPID library. Stripping away the reactive chloroacetyl moiety (which can be added separately during SPPS), activated CME ester (displaced upon Fx loading onto tRNA) and tyrosine (tyrosine fixed as final *N*-terminal amino acid on SPPS synthesiser), this left the core photoswitch of the common intermediate (18). Substitution of the Boc-protecting group for Fmoc was essential. This is because SPPS requires orthogonal base labile *N*-terminal amine protecting groups and acid-labile AA side chain protecting groups to prevent unwanted reactivity. This resulted in the novel analogue, Fmoc-Photo-Acetic Acid (49) (Scheme 19).



Scheme 19: Synthesis of Fmoc-Photo-Acetic Acid (49) from the common intermediate (18).

Starting from the common intermediate (18) (synthesis described in detail in Chapter 2) which was produced in the highest yields to date via the optimised Bæyer-Mills coupling (77%) and methyl ester deprotection (18). Standard de-Boc conditions were used to generate the free amino hydrochloride salt (48). Fmoc was then appended using activated Fmoc succinimide with DIPEA in THF at rt for 1 h. This generated the final SPPS Fmoc-Photo-Acetic Acid (49) in a satisfactory yield of 56% over 3-steps.

3.3.2.1 Synthesising the peptides for further characterisation:

As discussed earlier the six hit peptides (P001-P006) were selected for synthesis via SPPS. Rink amide resin was used to generate a *C*-terminal amide as not to install an extra charge into the peptides. The SPPS peptides were generated with a single Gly residue at the *C*-terminus. The remainder of the GSGSGS spacer used during the selection, was not synthesised as in general these spacer residues do not contribute to target binding. Standard SPPS conditions using the coupling reagent HATU were used

to synthesise each peptide sequence up to the penultimate tyrosine. With the peptides generated by SPPS, all with a *N*-terminal tyrosine, they were then manually coupled to Fmoc-Photo-Acetic Acid (49) to install the photoswitch using HATU once more. The chloroacetyl moiety was installed thereafter using chloroacetoxy-succinimide. The peptides were cleaved off the resin using a trifluoroacetic acid (TFA) cocktail, resuspended in DMSO and cyclised by raising the pH to ~pH 8 using triethylamine (TEA). Finally, they were purified via reverse-phase column chromatography. All the peptides excluding P004 were successfully synthesised, with their identity and purity confirmed by LCMS (Figure 53(a)) (traces available at Appendix 23). Issues in synthesis (truncations on the synthesiser) prevented synthesis of P004. As P004 was more poorly enriched, and only included to promote sequence diversity, in the interests of time it was discarded as the characterisation of the successfully synthesised hits remained the priority.

3.4 Biophysical, photochemical, and biochemical analysis of peptide hits:

3.4.1 Biophysical analysis – Surface plasmon resonance:

With the five peptides synthesised via SPPS, the first aim was to analyse the binding affinity and isomer selectivity of the SPPS synthesised hits against PADI2. To do so, surface plasmon resonance (SPR) was chosen as the biophysical technique. It measures the rate of association (k_a) and rate of dissociation (k_d) of the peptides binding to immobilised PADI2 on the surface of a chip. From this, the rate of dissociation is divided by the rate of association to generate an overall equilibrium dissociation constant (K_D), as a representation of binding affinity ($K_D = k_d/k_a$). This experiment aimed to determine if there were any differences in the K_D of the peptide hits between their respective *E* and *Z* isomers. To undertake the experiments, 300 nM biotinylated PADI2 was immobilised on a Biacore CAP chip (Biotin CAPture kit, series S Cytiva), for 120 seconds with a target density of ~1000-1500 RU. The peptides (P001, P002, P003, P005) were flowed using a single cycle protocol of increasing peptide concentrations at 25°C. Following determination of the K_Ds for each isomer, the degree of differential binding was calculated to produce *Z/E* binding differentials.

As shown in Figure 54(a and b), the *Z*-isomers of P001, P003 and P006 all bound to PADI2 with low nanomolar binding affinities at 11.1 ± 3.6 nM, 3.7 ± 2.9 nM and 41.9 ± 37 nM. Both P002 and P005 were quickly discarded as they both displayed substantial binding to the streptavidin reference channel, shown through the negative Response Unit (RU) values seen (Figure 55). Intriguingly, although P001 appeared to have strong affinity towards PADI2 in the *Z*-isomer (Figure 54(a/b)), it also repeatedly displayed a lack of clear *E*-binding, showing a combination of either significant bulk drift or partial streptavidin binding at the top two *E*-isomer concentrations (333.3 nM and 1 µM) (Figure 54(a/b)). This was indicated by the baseline or negative RU values which were seen. It remains unclear if this is just an artifact of no binding to PADI2 in the *E*-isomer or streptavidin binding as qualified above. As such, before drawing any decisive conclusions, additional work must be undertaken, e.g., immobilising streptavidin in the active reference channel to identify if streptavidin binding is actually seen. P001 does look like a selective *Z*-isomer binder against PADI2 and therefore may demonstrate very large differential binding if significantly reduced *E*-isomer binding can be quantified. What is interesting however, is that P001 and P002 share the family motif of ('IRLR') and P002 also appeared to be a streptavidin reference channel binder as mentioned. Identifying streptavidin binders from RaPID selections, however, is not uncommon due to the nature of the affinity enrichment process with streptavidin coated magnetic beads and is not particularly interesting for this work.

(a) SPR K_D values for each isomer of the peptides and calculated *E* / *Z* binding differential (tabulated):

Peptide ID:	*Z*-k_a (Ms^{-1})	*Z*-k_d (s^{-1})	*Z*-K_D (nM):	*E*-k_a (Ms^{-1})	*E*-k_d (s^{-1})	*E*-K_D (nM):	*Z*/*E* Binding differential:
P001	$1.8 \pm 0.4 \times 10^{6}$	$1.9 \pm 0.1 \times 10^{-2}$	11.1 ± 3.6	N/A	N/A	*	*
P002	N/A	N/A	Strep Binder	N/A	N/A	Strep Binder	N/A
P003	$4.6 \pm 4 \times 10^{5}$	$1.1 \pm 0.4 \times 10^{-3}$	3.7 ± 2.9	$1.9 \pm 0.8 \times 10^{5}$	$3.5 \pm 0.8 \times 10^{-3}$	21.6 ± 14	6-fold
P005	N/A	N/A	Strep Binder	N/A	N/A	Strep Binder	N/A
P006	$5.4 \pm 4.6 \times 10^{5}$	$1.2 \pm 0.9 \times 10^{-3}$	41.9 ± 37	$3.7 \pm 0.7 \times 10^{4}$	$4.4 \pm 0.1 \times 10^{-2}$	1210 ± 212	29-fold

(b) Representative SPR sensorgrams for each isomer of the peptides:

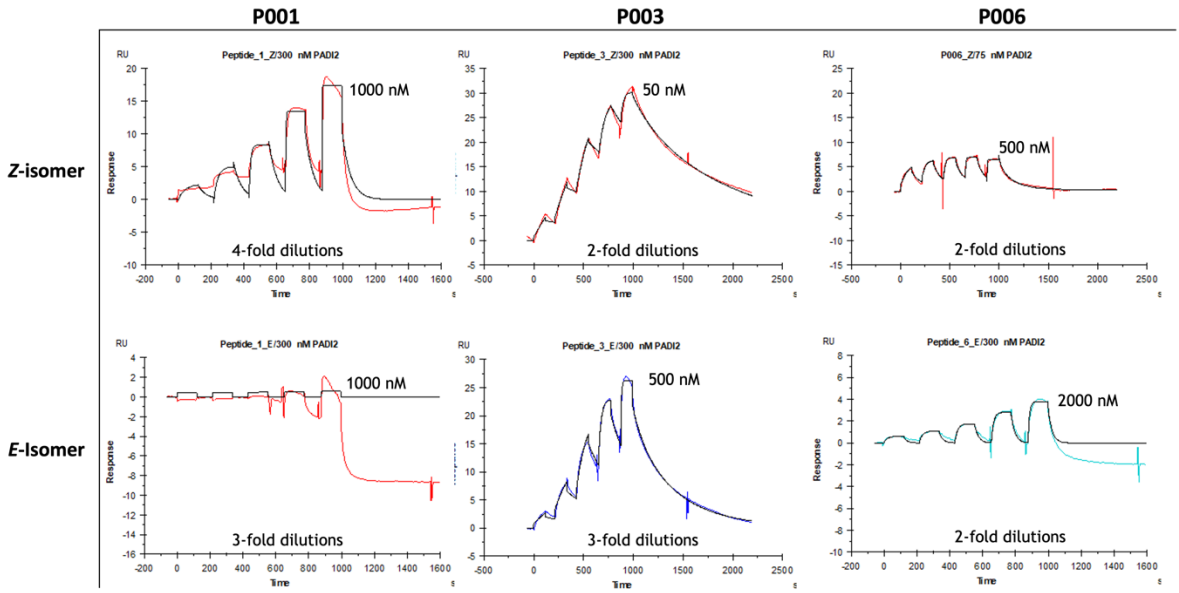


Figure 54: Surface plasmon Resonance (SPR) to determine the equilibrium dissociation constants (K_Ds) of peptides P001, P002, P003, P005 and P006 against immobilised PADI2. This is to identify any differential binding between the two light-responsive E-/ Z-isomers. (a) Tabulated averaged kinetics data to calculate the K_D for each isomer and calculated Z/E binding differential indicated where possible. (b) Representative sensograms of the binding peptides P001, P003 and P006. Note the significant reference channel binding of E-P001 – E-K_D value and Z/E differential not calculated (*). (a/b) Kinetics data and associated sensograms shown are representative of a minimum of 2 independent experiments for each isomer of each peptide tested (replicates at Appendix 38).

More excitingly for this project, when irradiated to the *E*-isomer, P003 and P006 also showed a reduction in affinity to ${}_{E}K_D$ 21.6 ± 14 nM, and 1210 ± 212 nM respectively. This demonstrated the first quantifiable examples of light-controlled differentially binding cyclic peptides against immobilised PADI2. With the differential binding for P003 and P006 identified, *Z/E* binding differentials could then be calculated. P003 demonstrated an *E/Z* binding differential of 6-fold and even more excitingly, a 29-fold differential binding was seen for P006. This was particularly exciting data as it validated the

approach to identify *de novo* photoswitchable cyclic peptides via the PhotoRaPID v1.0 methodology, far surpassing the current literature previous discussed in the introduction (Chapter 1.8).

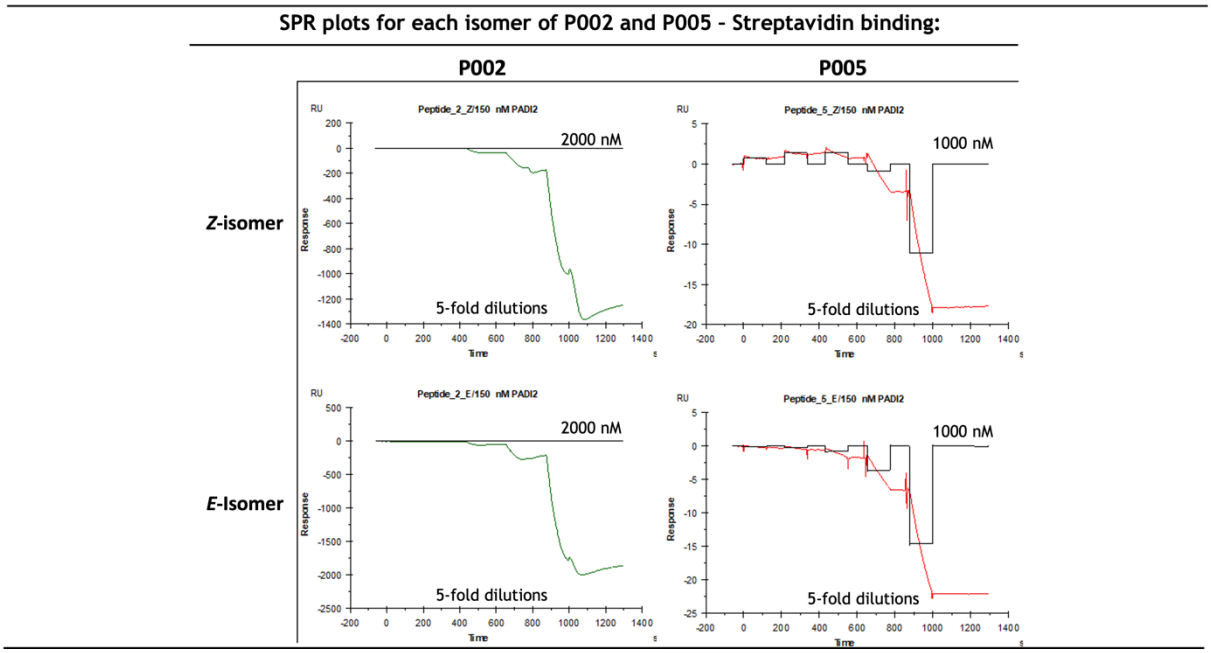


Figure 55: SPR of peptides P002 and P005 showing non-specific streptavidin binding. Associated sensograms shown are representative of only a single independent experiment for each isomer.

3.4.2 Photochemical analysis – UV-Vis spectrophotometry to assess photostationary state (PSS) and thermal half-life ($\tau_{1/2}$):

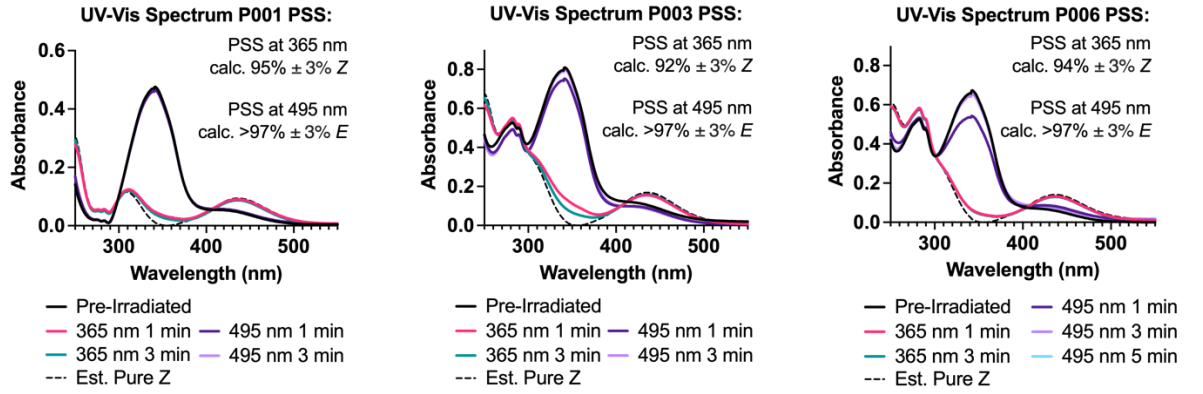
Having confirmed binding to PADI2, peptides P001, P003 and P006 were characterised by UV-Vis spectroscopy to explore the individual photochemical properties of each of the photoswitchable macrocycles. To identify the photoconversion efficiency, the peptides were first diluted to ~20 µM in H_2O (~2% DMSO). Irradiation at each wavelength (365 nm or 495 nm) for varying lengths of time (1-5 min), proceeded until the respective PSS was reached. The absorption profile of the photoswitch within the macrocycles was then assessed across 200-800 nm to identify the proportion of *Z* or *E* isomer at each PSS wavelength. Peaks around 200-300 nm result from residual DMSO absorption in the buffer and absorption from tyrosine and or tryptophan residues in the peptides, if present (280 nm).[151]

To indicate quantitative bi-directional switching within the UV-Vis spectra, two key changes in peak shape should be seen. Upon irradiation at 365 nm, when the PSS is reached, for quantitative forward switching (*E*→*Z*), the major peak at ~340 nm (π- π*) should be fully abrogated with an increase in the peak intensity at ~490 nm. This indicates a complete absence of the *E*-isomer population and thus full conversion to the Z-isomer (n-π*). In contrast, to highlight quantitative back switching (*Z*→*E*), following irradiation at 495 nm (n-π*) until reaching the PSS, the major peak at ~340 nm (π- π*) should be fully returned with a decrease in the peak intensity at ~490 nm (n-π*). Photoconversion efficiency is quantitatively determined by comparing the absorption profile of the (π- π*) peak in both a pure *E* sample and a sample where the maximal *Z*-isomer population has been reached at the PSS. Incomplete photostationary conversions result from an overlap of the absorption peaks between each isomer at a given wavelength. Therefore, upon irradiation at a wavelength where both isomers absorb, a mixed equilibrium population of both isomers is established interconverting back-and-forth. This is in contrast to the desired effect, where irradiation is undertaken at a specific wavelength unique to each isomer to generation single isomer populations in solution (i.e. quantitative photoconversion). The separation of the absorption peaks (absence of absorption band overlap) is an intrinsic photochemical property of the photoswitches structure. Thus, changing the electronics of the system can disrupt separation of the absorption peaks (Chapter 1.2.1 and Chapter 1.2.5).

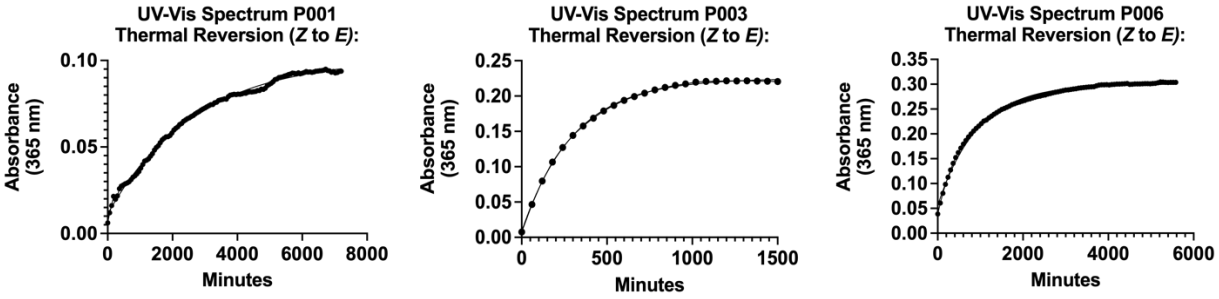
As seen in Figure 56(a), upon reaching the PSS at 365 nm for P001, P003 and P006, all peptides displayed populations containing $\geq$92-95 ± 3% of the *Z*-isomer. In contrast, upon reaching the PSS at 495 nm, 97± 3% *E*-isomer populations were seen. As such, P001, P003 and P006 all displayed near

quantitative PSS distributions (quantitative photoswitching) between both the *E*→*Z* transition (365 nm) and the reverse *Z*→*E* transition (495 nm). This was following irradiation at either 365 nm for ≤3 mins (LED 3x 800 mW at 100% output) or irradiation at 495 nm for ≤5 mins (LED 3x 750 mW at 100% output), depending on the exact peptide.

(a) UV-Vis spectra of Peptides P001, P003 and P006 identifying their Isomer populations at each photostationary state (PSS) - 365nm and 495 nm:



(b) Thermal half-life measurements of Peptides P001, P003, P006 assessed via thermal reversion from *Z* to *E* at 365 nm:



One Phase Decay	
Y=(Y0 - Plateau)*exp(-K*X) + Plateau	
Y0	0.01125
Plateau	0.1014
K	0.0003710
R^2	0.9961
Half-Life	**31.2 hrs**

One Phase Decay	
Y=(Y0 - Plateau)*exp(-K*X) + Plateau	
Y0	0.008279
Plateau	0.2249
K	0.003282
R^2	0.9995
Half-Life	**3.5 hrs**

One Phase Decay	
Y=(Y0 - Plateau)*exp(-K*X) + Plateau	
Y0	0.05574
Plateau	0.3004
K	0.001069
R^2	0.9948
Half-Life	**10.8 hrs**

Figure 56: UV-Vis Spectroscopy of P001, P003 and P006 peptides to assess their photochemical properties. (a) UV-vis spectra of the peptides scanning 200-800 nm following increasing durations of irradiation until the PSS was reached at both 365 nm and 495 nm respectively. (b) Thermal half-life ($\tau_{1/2}$) measurements of the peptides, tracking the increasing absorbance at 365 nm to indicate thermal reversion from the Z-to-E-isomer. $\tau_{1/2}$ is calculated by fitting the one-phase exponential decay. UV-Vis plots and $\tau_{1/2}$ are representative of a single experiment. (a/b) Peptides were diluted into H_2O (at ~20 µM f.c. ~2% DMSO). Pre-irradiated is defined as the pure E-isomer.

These findings were excellent, as the peptide-integrated photoswitches largely retained the photochemical PSS distribution of the parent model photoswitch 4pzMe (7), as intended (*E*/*Z* ≥98%).[44,46] This is likely because electronic deconjugation of the photoswitch from the remainder of the macrocycle was prioritised when designing the genetic-code-reprogramming compatible photoswitch analogues (Chapter 2). This absorption profile was one of the key reasons why 4pzMe (7) was initially chosen due to its excellent intrinsic band separation which provides its quantitative photoswitching capacity. Importantly, this has been a limitation within applications of classical azobenzenes, as they show incomplete switching due to absorption band overlap, limiting the effectiveness of their applications (Chapter 1.2.3).[152,153] It is also important to highlight that it was unknown if the specific amino acids within the macrocycles would adopt conformations that could result in spatial electronic interactions. As mentioned, such interactions could also affect the photoswitch electronics, despite ensuring that the core of 4pzMe (7) was electronically deconjugated. As such it was very validating that the parent photoswitch profile in the macrocycles was retained following integration into the peptides.

Additionally, thermal half-life ($\tau_{1/2}$) measurements were also undertaken (Figure 56(b)). As mentioned in the introduction (Chapter 1.2.1), $\tau_{1/2}$ (one half-life), represents the time needed for the concentration of the *Z*-isomer to decrease to one-half of its original value thermally reverting to the *E*-isomer.[154] Specifically for this work, this interconversion can be measured by monitoring the increase in absorbance at 365 nm over time. The $\tau_{1/2}$ is calculated by fitting a one-phase exponential decay to the plots upon the absorbance at 365nm reaching plateau ($\tau_{1/2}$ equation in Chapter 1.2.1).

Assessing the $\tau_{1/2}$ of each macrocycle was necessary for this work as the PhotoRaPID v1.0 Selection was designed to enrich for peptides that bind to PADI2 when in the metastable Z-isomer. As such, the rate of thermal reversion indicates the duration of time that the photoswitch remains within its 'active' conformation following 365 nm light irradiation. P001 displayed the longest $\tau_{1/2}$ at 31.2 hrs, whilst P003 and P006 resulted in more accelerated $\tau_{1/2}$ values of 3.5 and 10.8 hrs respectively. These $\tau_{1/2}$ values are substantially quicker (up to 69-fold acceleration) than the $\tau_{1/2}$ of 4pzMe (7), the parent photoswitch, at 10 days (Figure 57).[44] This is not directly comparable though as the 4pzMe (7) measurements were undertaken in pure DMSO whereas the P001, P003 and P006 were undertaken in water (pH ~7.5) with a significantly reduced DMSO content at (~2%) which an can lead to thermal acceleration due to protonation of the azo-bridge. These aqueous conditions were chosen as

identifying the thermal-half life in a biologically relevant setting was preferred, as peptides identified from this platform could become ligands for chemical biology/ photopharmacology applications.

Peptide ID:	$\tau_{1/2}$ (hrs):	Δ $\tau_{1/2}$ vs 4pzMe
4pzMe	240	-
P001	31.2	7.7-fold
P003	3.5	69-fold
P006	10.8	22-fold

Figure 57: Fold change acceleration in thermal half-life ($\tau_{1/2}$) of the photoswitchable macrocyclic peptides in comparison to the parent photoswitch 4pzMe (7).

Although the up to 69-fold acceleration in $\tau_{1/2}$ was disappointing photochemically, as these peptides still had $\tau_{1/2}$ values on the order-of-magnitude of hours, this still made them suitable for chemical biology applications which was gratifying. Two main rationales could begin to explain the acceleration seen. Firstly, AAs within the macrocycles could adopt conformations that could result in the acceleration of the *Z-E* transition. This could result from AA side chains that promote local protonation of the photoswitch azo bridge, which in the literature has been shown to result in decreased double bond character and increased thermal acceleration.[155] In this case, the difference in solvent (DMSO vs H_2O) would contribute little to the accelerated thermal reversions seen. In contrast, in aqueous conditions, there is precedent for accelerated thermal relaxation in the literature for other arylazoheteroarenes. Weston et al.,[44] identified that the thermal isomerisation rate of their tested azopyrroles were found to be dependent on the water (or methanol) content of the sample. They identified that the addition of 0.5 M water within their specific azopyrrole sample resulted in an approximately ten-fold increase in the rate of thermal isomerisation. This was not seen explicitly investigated for the azopyrazole class by Weston et al., though. Additionally, Walther et al.,[156] identified that the thermal half-life of a single water-soluble azopyrazole derivative could be tuned by more than five orders of magnitude (seconds to weeks) via the adjustment of the pH of aqueous buffers. As such it is possible that the increased thermal reversions seen within the photoswitchable macrocycles could be a combination of both factors presented and this is something that should be investigated further. Critically though, as mentioned, despite the acceleration seen, the $\tau_{1/2}$ values (at the assumed pH of 7.0) determined for these peptides were still suitable for use in chemical biology applications which was the primary end goal of the work when designing the methodologies.

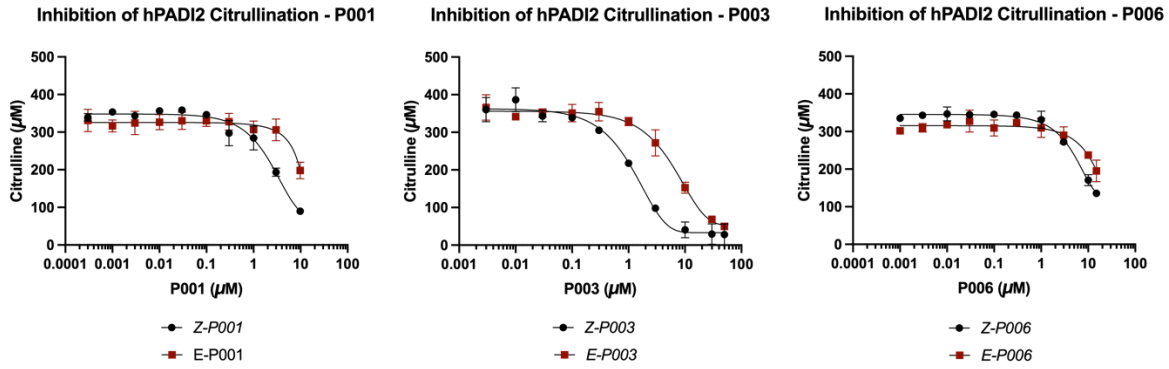
3.4.3 Biochemical analysis – *In vitro* enzymatic inhibition assays:

To examine peptide binding effects on PADI2 enzymatic function, as mentioned at the beginning of the chapter, the pre-established COLDER assay, developed by Vašák et al., was used. [145] This colourimetric plate-based assay is used to assess PADI citrullination activity and was used to measure the inhibitory activity, if any, of peptides P001, P003 and P006 in both isomers. The assay involves an acid-catalysed modification of the urea moiety within citrulline to produce a pink colour as citrulline is the product of enzymatic conversion of arginine by PADIs. If successful inhibition of PADI mediated citrullination is seen, a weaker dose dependent pink colour would be produced (indicating reduced citrullination activity of PADI2). This pink colour was quantified via a 96-well plate reader measuring absorbance at 540 nm.

It is important to note that traditional PhotoRaPID selections are designed to generate tight binding peptides, which may not necessarily be inhibitory.[103] This is because the RaPID methodology used promoted the enrichment of peptides against any surface-accessible region of PADI2, not specifically the active site.

P001, P003 and P006 all displayed mild inhibitory activity against enzymatic PADI2 citrullination with *Z*-IC_{50} values of 5.4, 1.2 ± 0.6 and 5.8 µM respectively (Figure 58). Encouragingly, all peptides also appeared to display light-responsive differential inhibitory activity between isomers to some extent. This is most clearly seen for P003, with 6-fold weaker inhibitory activity in the *E*-isomer compared to Z-isomer. Unfortunately, complete IC_{50} curves could not be generated for peptides P001 and P006 as the stock solutions were not concentrated enough. It was not possible to make more concentrated stocks at the time as increasing the f.c. of DMSO >0.1% lead to greater variation in PADI2 activity. However, from the current results, a slight difference in activity between the P001 and P006 *E* and *Z* isomers could be observed preliminarily, which was consistent with the different K_D values determined for each isomer. Interestingly, the *Z*-IC_{50} values obtained in the COLDER assay were all significantly weaker than the *Z*-K_D values achieved by SPR up to a maximum of <486-fold weaker (Figure 58). Some variation may arise due to the different assay conditions used, but this weakened inhibitory activity may imply that the peptides do not bind fully in the active site of PADI2. As such the inhibitory activity seen may be the result of displacing the BAEE substrate allosterically or promoting changes in the active site conformation, reducing citrullination catalytic efficiency. Investigation of the peptide binding sites with structural biology would be of future interest but was beyond the scope of this methodology development project.

***In vitro* Citrullination - IC_{50} values for each isomer of the peptides and calculated *E*/*Z* IC_{50} differential:**



Peptide ID:	*Z*-IC_{50} (μM):	*E*-IC_{50} (μM):	*E*/*Z* IC_{50} Differential:	SPR *Z*-K_D (nM):	*Z*-K_D / *Z*-IC_{50}:
P001	5.4	>9.8	>2-fold	11.1 ± 3.6	486-fold
P003	1.2 ± 0.6	7.0 ± 5.3	6-fold	3.7 ± 2.9	324-fold
P006	5.8	>15	>3-fold	41.9 ± 3.7	138-fold

Figure 58: In vitro enzymatic colourimetric assay (COLDER assay) to assess the inhibitory activity of P001, P003 and P006. 100 nM PADI2 was preincubated with varying concentrations of each isomer of each peptide, individually, for 10 mins before addition of the BAEE substrate (f.c. 10 mM total BAEE, f.c. 0.1% DMSO from peptide, volume = 50 µL). The reaction proceeded for 30 mins at 37 °C and then visualised. Graphs are representative of one independent experiment with three technical replicates for each isomer of each peptide. For P001 and P006 IC_{50} standard deviations could not be accurately calculated due to the incomplete inhibition.

3.5 Conclusions

In conclusion, it was first confirmed that the UV light irradiation employed did not result in adverse effects on DNA stability within the selection library. This was a critical first step to ensure that a light-responsive screening platform could be feasibly developed. If DNA mutation was occurring, then the genotype-phenotype between the cyclic peptide and its conjugated encoding DNA would be lost. This would have meant that it wouldn't have been possible to incorporate light into mRNA display using the wavelengths or duration of time required for optimal photoswitching. Following this, PADI2 was chosen as the proof-of-concept protein target. This was because it is involved in several important disease profiles and as such cyclic peptide ligands against this target would be useful in the future regardless of their switchable binding efficiency. Biotinylated His-Avi-PADI2 expression, loading efficiency of the protein onto magnetic streptavidin beads, and assessment of the intrinsic DNA binding affinity of PADI2 all ensured that PADI2 could be readily incorporated into the RaPID system for screening.

Following the first PhotoRaPID v1.0 selection, each of the three light irradiation conditions tested produced enriched pools of cyclic peptides, despite the uncharacteristically high DNA recoveries seen within R6 during the selection. Analysis of this data revealed three distinct family motifs, 'WMD', 'HHW' and 'IRL'. Interestingly, these family motifs made up ~89-96% of the top 100 enriched peptide sequences in each of the three irradiation conditions. The 'WMD' motif remain the most prominent motif seen for IrrC.1 and IrrC.3 whilst 'IRL' dominated IrrC.2. Most interestingly, IrrC.3 had a significant reduction in the number of reads for peptides containing the 'IRL' motif, whilst the 'HHW' motif was overrepresented within this irradiation condition. This is something to be investigated further as IrrC.3 was considered the most stringent selection pressure. As switching the peptide off PADI2 was required for recovery and enrichment of those sequences, this was most likely responsible for the disparities seen when compared to IrrC.1 or 2.

An additional takeaway of significant interest comes from cross-referencing the top 10 TES between each of the three IrrCs. No significantly enriched unique peptide motifs were identified, beyond 'WMD', 'IRL' or 'HHW'. More surprisingly, all the 'unique' sequences appeared to be the result of single amino acid substitutions around the three family motifs previously described. This suggests that the three irradiation conditions were all biasing the libraries towards the same family motifs, which can be rationalised as PADI2 only contains a limited number of accessible binding sites. More speculatively, from a methodology point of view, it may also be that the 'best' photoswitchable

binding peptides might not be the most enriched sequences, which is potentially problematic when using an mRNA display technology that biases for enhanced recovery through increased binding affinity. This is as the photoswitchable peptides could require weaker 'active isomer' binding such that the sufficient abolishment of binding when in the 'inactive' isomer can be achieved. If this were the case however, this should have been more dramatically seen through IrrC.3. Perhaps this is partially seen by preferential enrichment for the 'HHW' motif compared to IrrC.1 or IrrC2. Defining what the most optimal differentially binding peptide 'should look like', is perhaps the key defining question of this PhD work.

Following synthesis and characterisation, the PhotoRaPID v.10 methodology successfully generated peptide hits that display low nanomolar *Z*-isomer binding affinities to PADI2 with P001, P003 and P006 having $_{Z}K_{D}$s of 11.1 ± 3.6, 3.7 ± 2.9 and 41.9 ± 37 nM, respectively. Very excitingly, up to a 29-fold differential binding between *Z* and *E* isomers was also seen (P006). The tightest binding peptide (P003) and greatest differentially binding peptide (P006), each contain either the 'WMD' and 'HHW' motifs respectively. These motifs were two of the top three motifs ('MWD', 'HHW', 'IRL') identified from the sequencing data in the top 100 most enriched sequences (TES). From a methodology point of view, these peptides were identified from conditions IrrC.1 and IrrC.3, however, it is also interesting to note that the 'WMD' motif of P003 also dominated the TES seen in IrrC.2 (Figure 52). As the 'WMD' motif from P003 produced the tightest binding peptide identified so far, it could be beneficial to return to the sequencing data of the irradiation conditions to synthesise the top hit from IrrC.3 (cyclo-Photo-^{L}Y-QN**WMD**RAHDVSC(S-)G) which also contained this motif. This sequence was not initially synthesised to prioritise sequences diversity in the peptides selected for characterisation.

UV-Vis spectrophotometry experiments performed on the synthesised macrocyclic peptides demonstrated that the broad photochemical properties of the parent photoswitch (4pMe (7)) are maintained within the peptide, likely due to deconjugation of the photoswitch from the surrounding amino acids. This confirms that an active photoswitch can be successfully incorporated into macrocyclic peptides using the RaPID mRNA-peptide display system, validating use of RaPID technology in this way. Additionally, it appears that despite constraining the photoswitch within the cyclic peptides, the isomerisation and thermal relaxation were still typical of azopyrazoles. However, accelerated thermal reversion from the *Z* to *E* isomers of the photoswitchable cyclic peptides was seen (3.5 hrs to 31.2 hrs) when compared to the parent 4pzMe (7) photoswitch at 10-days. This may be due

to solvent effects of the aqueous buffer system or due to side-chain interactions within the cyclic peptides but requires further investigation to confirm the cause of this accelerated reversion.

Peptides P001, P003 and P006 were able to inhibit PADI2 citrullination activity, with up to a 6-fold difference in IC_{50}s between the two light-responsive *E* and *Z* macrocyclic isomers. With IC_{50} values in the low μM range, these peptides are not optimal for use as light-responsive inhibitors but could be converted to photoswitchable PROTACs (Proteolysis targeting chimeras) if desired. This would be a valuable application as PROTACs enable target degradation via the proteasomal degradation machinery, requiring only selective protein binding not inhibition.[157–159] Photoswitchable PROTACs have also been successfully demonstrated in the literature, further validating this potential approach.[160] This, lack of inhibitory potency, however, is an intrinsic drawback to the standard RaPID system, not specifically PhotoRaPID v1.0, as it is designed to preferentially enrich for peptide binders across the whole protein surface, not only the active site.

A possible future direction would be to perform parallel selections using PhotoRaPID v1.0 methodology with the active site of the protein of interest being blocked by a covalent inhibitor or suicide substrate as well as the unliganded protein. Deconvolution of the sequencing data to identify families that only appeared in the unliganded protein selection might enable the identification of potent photoswitchable cyclic peptide inhibitors that target the active site, thereby providing additional control over hit function.

In terms of method development, the discovery of peptides P003 ('WMD') and P006 ('HHW') that have tight nanomolar potency, display up to 29-fold differential light-responsive binding and modestly inhibit PADI2 in their *E* and *Z* forms, highlight the ability of PhotoRaPID v1.0 to identify potent photoswitchable hit peptides. This is particularly highlighted when considering the current published literature on photoswitchable macrocyclic peptide discovery platforms (Chapter 1.8).

Despite this initial success, it is also critical to understand areas where the PhotoRaPID v1.0 methodology can be improved. Most prominently, oversights were identified in the IrrC.2 methodology following SPR analysis of the synthesised peptide hits. The most enriched individual peptide from IrrC.2 was P002 at 60% enrichment, but SPR showed that this peptide binds streptavidin rather than PADI2. After returning to the workflow of IrrC.2, the source of these streptavidin binders

became clear. IrrC.2 (Figure 50), was the only irradiation condition not to contain a negative counter selection against streptavidin-coated beads. This was because the first light-responsive *E*-counter selections were discarded, and it was hypothesised that this step would remove non-specific streptavidin binders as well as undesirable *E*-binders. This does not appear to be the case. Despite this, the key 'WMD' peptide sequence motif was also enriched in IrrC.2, indicating that IrrC.2 could be a suitable approach if additional negative counter selections were to be employed. This would be to prevent 'swamping' of the viable sequences by unwanted streptavidin binders.

Further optimisation of PhotoRaPID could also include changes to the size of the macrocycles included in the starting library. From the peptide sequences recovered from PhotoRaPID v1.0, which was carried out using a mixed library of NNK6-12 macrocycles, NNK8-10 peptides dominated the recovered sequences. Interestingly, the largest macrocycles were absent from the TES of all three irradiation conditions. The absence of enriched NNK6-7 peptides is often observed during for RaPID selections as larger macrocycles typically contribute a greater number of productive target binding interactions and outcompete their smaller counterparts. However, the absence of NNK11-12 hits in PhotoRaPID v1.0 was surprising and is something to consider for future selections. One hypothesis is that the photoswitch may impart greater conformational changes between the two isomers when the macrocycle is smaller in size, and so this could translate to greater light-responsive differential binding for smaller peptide hits. As such, smaller starting NNK libraries could be beneficial for identifying larger differences in the binding of *E* and *Z* isomers, although this may be accompanied by a reduction in peptide binding affinity.

Overall, the PhotoRaPID v1.0 selection methodology was able to be successfully generated to discover photoswitchable cyclic peptides with light-responsive differential binding to PADI2. With the peptides displaying enhanced binding and photochemical properties than those reported in the literature (although different targets used). This work therefore demonstrated the applicability of PhotoRaPID v1.0 to identify light-responsive peptides for a target of interest. Further optimisation of the methodology was possible aiming to maximise the enrichment of peptides with largest *E*/*Z* binding differentials, and this is undertaken in the following chapters.

4 PHOTORAPID V2.0: METHOD OPTIMISATION

Chapter Aims:

This chapter describes work aiming to further improve the methodology of PhotoRaPID v1.0 to find cyclic peptides with the largest *Z/E* differential binding possible and in the most streamlined manner.

Key aspects of the chapter:

- Identify which irradiation condition from PhotoRaPID v1.0 is most suitable for optimisation.
- Repeat the PhotoRaPID selection with the optimised workflow – PhotoRaPID v2.0.
- Develop a tailored NGS data analysis workflow so suitably enriched peptides with large differential binding can be identified post-selection in a quick, predictive, and reproducible manner.
- Synthesise, then biophysically and photochemically characterise the peptide hits chosen.
- Define the preliminary success of the method optimisation and data analysis approaches.
- Assess whether this new methodology and tailored NGS analysis enabled the identification of improved differentially binding photoswitchable peptides compared to PhotoRaPID v1.0.

A visual representation of the chapter workflow is shown in (Figure 59).

Chapter Workflow:

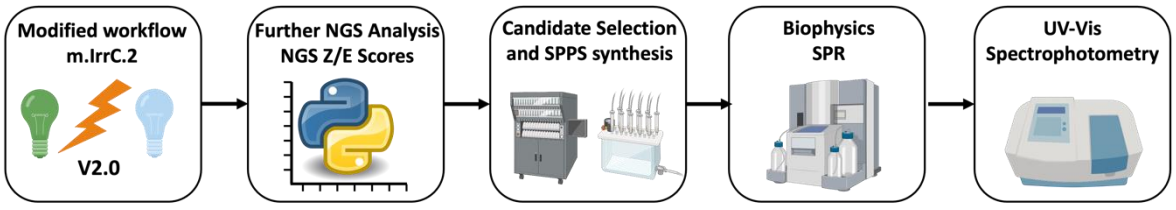


Figure 59: Workflow of the Chapter.

4.1 Choosing which PhotoRaPID method should be further developed:

To further build on the successes of the first PhotoRaPID v1.0 methodology, and due to the time constraints of the PhD, only one IrrC could be selected for further methodology development. This was to further enhance the enrichment of light-dependent differentially binding peptides during the selection. To decide which IrrC was best to pursue, two main criteria were considered:

(1) A selection pressure to enrich peptides with large differential binding: This was the primary factor for determining which IrrC to further develop. With the outright successes of IrrC.1 and IrrC.3 producing P003 and P006 in PhotoRaPID v1.0, these initially seemed the obvious choice. IrrC.1 however, as the closest workflow to a traditional RaPID selection, does not have a method to remove *E*-binding peptides and thus it was difficult to see how improved differential binding peptides could be obtained through this method. IrrC.1 was also designed as a benchmark against traditional RaPID if the more ambitious IrrC.2 or Irr.3 selection methods failed. But with enrichment seen in all three conditions, continuing this selection method felt unnecessary. From this, IrrC.3 appeared a logical candidate as it produced the greatest differentially binding peptide seen to date (P006).

(2) Applicability of this methodology to any protein target: When looking at the library quality of IrrC.3 after the PhotoRaPID v1.0 selection, however, it suggested that this IrrC may be too stringent (Figure 60). When taking the raw NGS sequencing data, further Python scripts were developed to see if truncated sequences from amber, ochre, opal, or frameshifts in the final round of each IrrC were being enriched. Enrichment of these truncated sequences suggests that the DNA recovered round-per-round was of poor quality as too few peptides were able to bind to PADI2 for substantial DNA recovery. From the library quality analysis, nested pie charts were generated (Figure 60). IrrC.3 had the largest proportion of truncated sequence reads at 25.66% when looking at the unfiltered sequence reads from the top 100 enriched peptides in the NGS *Z*-positive dataset for the final round of selection (R6). More crucially, 88.28% of these truncated peptide sequence reads arose from within the top 10 most enriched sequences. When looking at the raw sequences in IrrC.3, the most enriched peptide contained an internal amber stop codon (MSVVEamberIRLRCG - 20.21% of all IrrC.3 peptide reads). It also contained the 'IRL' motif which was found to be contributing to streptavidin binding in P002, despite a strep-negative counter selection being present in the workflow. As this methodology should be designed to be robust and applicable to any protein of interest, this raised concern over whether this IrrC would be able to reproducibly enrich light-dependent binding peptides robustly.

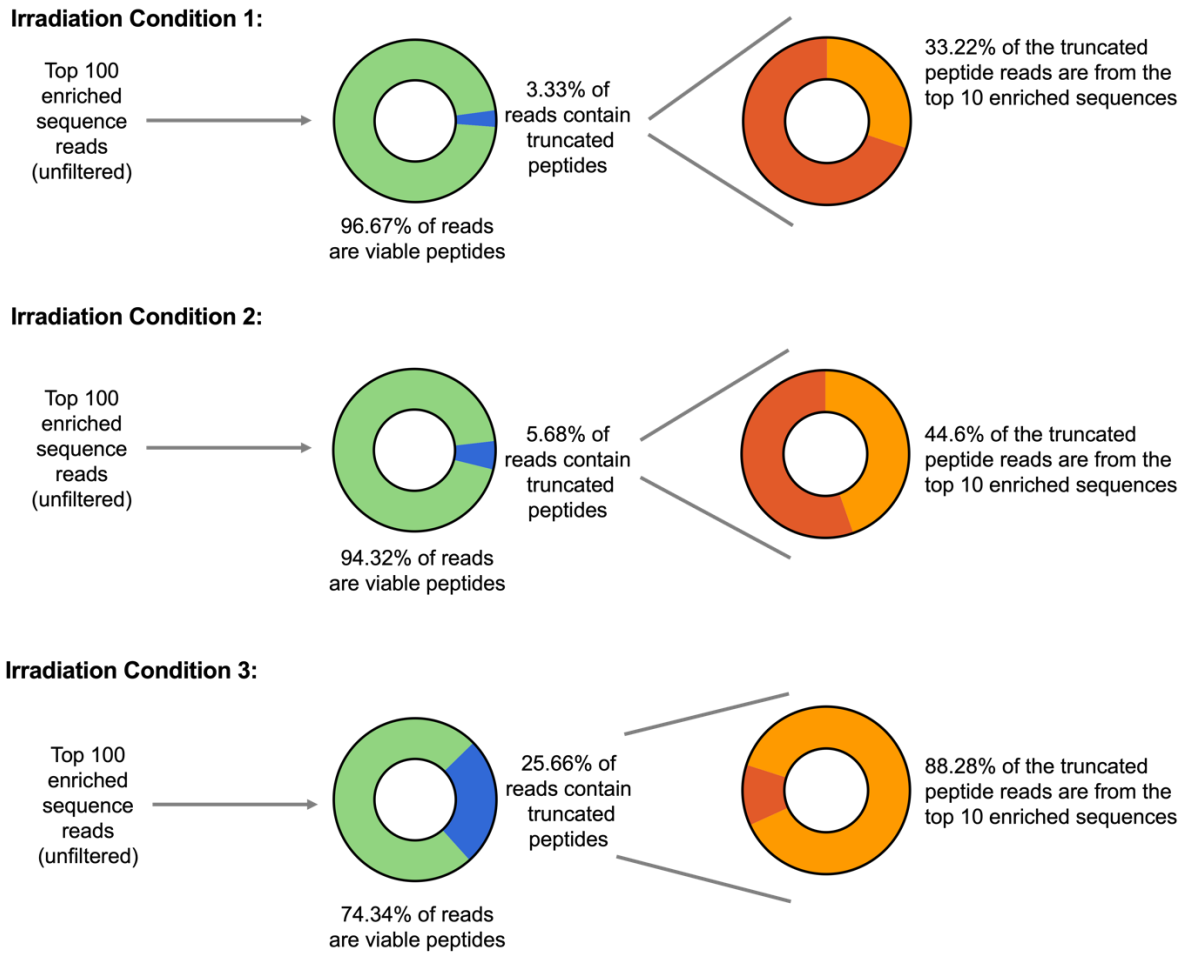


Figure 60: Library Quality Analysis post-selection – Raw NGS data (top 100 sequences) for each IrrC at the final round of selection (R5). Pie charts show the proportion of reads for sequences in the datasets which contain mutations resulting in internal STOP codons (amber, opal, ochre) or frameshifts. These are indicators of poor library quality post-selection.

This left IrrC.2 as the only condition that could be easily developed further. At first, this may not appear as the best option, as streptavidin binders contaminated this IrrC during PhotoRaPID v1.0 (Figure 52, Figure 54). However, the PADI2 binding motif 'WMD' still dominated this IrrC, despite a lack of a streptavidin negative counter selection (Figure 50). This suggested that IrrC.2 still had untapped potential as a method to find light-dependent macrocycles with larger binder differentials. Particularly, as it already contained the *E*-negative counter screen to remove any strong *E*-isomer binding peptides. Additionally, it also had comparable library quality to IrrC.1 (5.68% of the top 100 enriched peptide reads account for truncated peptides vs. 3.3% for IrrC.1). As such, it was the IrrC chosen as the workflow to further optimise.

Beyond improving the IrrC workflow, two additional features (i and ii) were trialled to see if they aided the discovery of peptides with large *Z/E* differential binding:

(i) **Using mixed starting libraries of peptides with smaller ring sizes**: This was chosen, as it was hypothesised that the smaller the total size of the cyclic peptide, the larger the degree of structural conformational change that could be imparted by the photoswitch between the two light-dependent isomers.

(ii) Using NGS sequencing results to predict large *Z/E* differentially binding peptides: In PhotoRaPID v1.0, peptides were chosen by their relative enrichment, but this did not consider their *Z/E* differential binding potential, a crucial factor when finding light-dependent cyclic peptide binders. By retaining the peptides recovered from the *E*-negative counter selection and sequencing these peptides, differential enrichment scores could be generated (*Z*-isomer enrichment / *E*-isomer enrichment). It was hoped that the calculated NGS *Z/E* enrichment differentials would correlate to the *Z/E* binding or IC_{50} differentials generated experimentally. If this were the case, excellent differentially binding peptides could be quickly chosen from the sequencing data in future selections. As a result, large numbers of SPPS peptides would not need to be synthesised and characterised after every selection.

4.2 Modified irradiation condition 2:

With IrrC.2 chosen for further optimisation, a modified workflow (m.IrrC.2) was developed (Figure 61). The main difference between IrrC.2 and m.IrrC.2 was a negative streptavidin (strep-negative) counter selection, which was undertaken before the *E*-isomer light-dependent counter selection. First, strep-negative, and *E*-isomer counter selections were undertaken. Then, the streptavidin and *E*-isomer binder depleted library was irradiated to the *Z*-isomer, such that the *Z*-positive selection could be undertaken. The *Z*-binding peptide sequences were recovered for the next round of selection or for NGS to deconvolute the enriched peptide hits. The strep-negative and E-negative were also kept for NGS sequencing to generate the NGS *Z/E* scores post-selection but critically, not enriched round-by-round unlike the *Z*-positive DNA recovered.

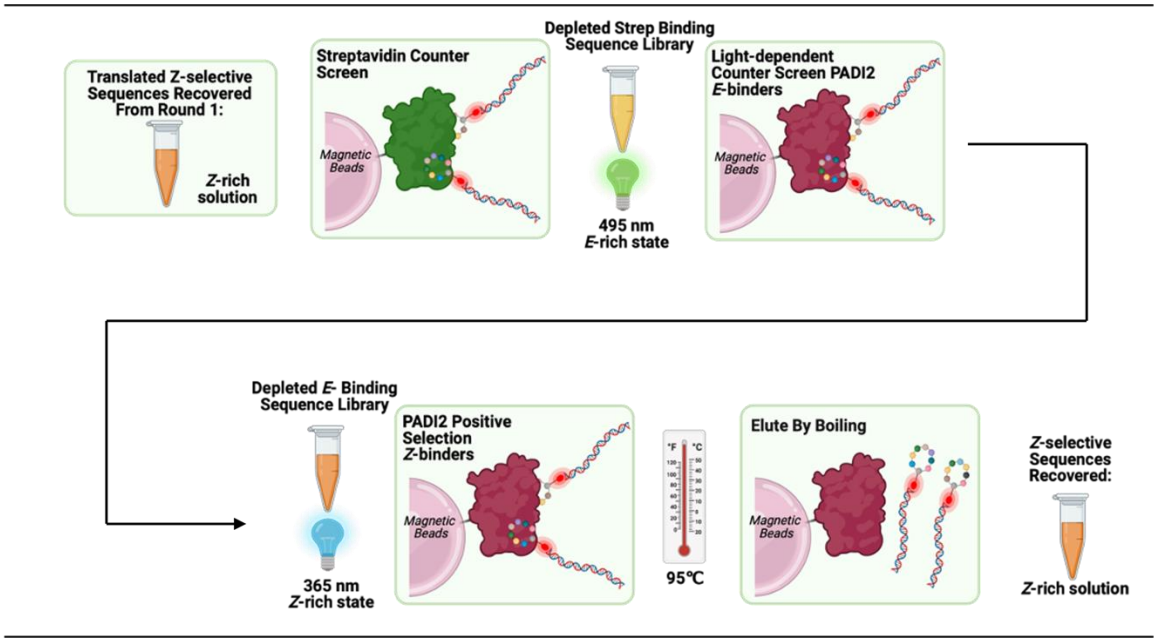


Figure 61: Optimised workflow for irradiation Condition 2 in PhotoRaPID v2.0 (m.IrrC.2). Note: The biotin-streptavidin interaction between the between the magnetic streptavidin beads and the biotinylated PADI2 is not visually depicted.

4.3 DNA recovery during the selection – PhotoRaPID v2.0:

Five rounds of the PhotoRaPID v2.0 methodology were undertaken using the m.IrrC.2 protocol. Two selections of mixed NNK4-8 and NNK6-12 libraries were undertaken in parallel (Figure 62). NNK4-8 was chosen to evaluate if the smaller ring sizes produced enriched pools of greater light-dependent differentially binding peptides. In contrast, the NNK6-12 was used as a reference against PhotoRaPID v1.0 which also used a NNK6-12 mixed library to show the effectiveness of the m.IrrC.2 approach. As shown in Figure 62, both libraries did not appear to enrich as well as the original PhotoRaPID v1.0 selection (Figure 51). However, a consistently increased percentage recovery of the *Z*-positive DNA over the negative DNA, suggested that PADI2 binding peptides were still enriching. Notably, there appeared to be very little difference between the enrichment profiles of the two mixed libraries, despite concern that the larger mixed library (NNK6-12) could contain more tight-binding peptides due to increased peptide-protein binding interactions possible with the larger macrocycles. This suggests that the influence of the DNA recovered round-by-round was largely the result of the new m.IrrC.2 methodology, not ring size (as NNK6-12 was the control relative to PhotoRaPID v1.0).

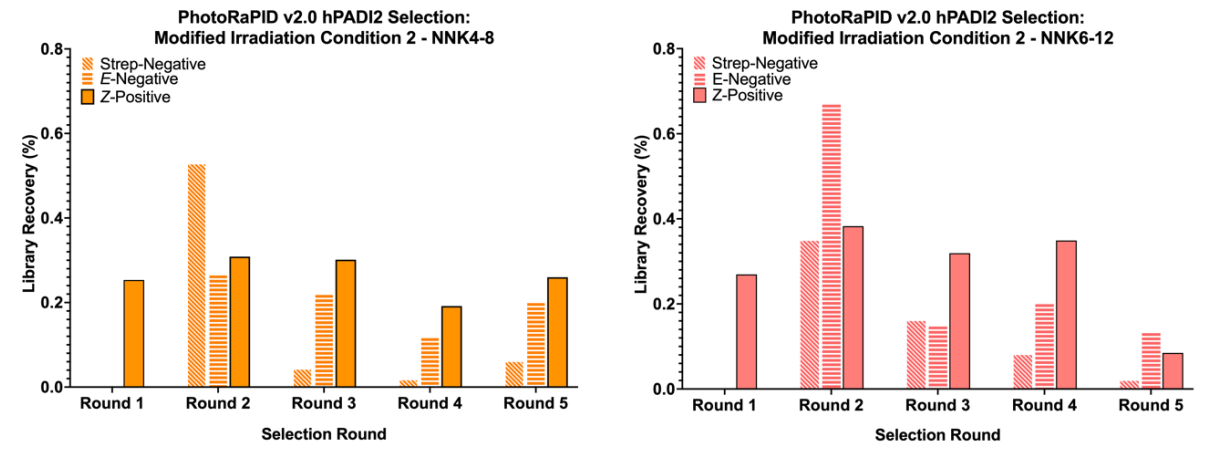


Figure 62: DNA recovery round-by-round during the PhotoRaPID v2.0 selection against PADI2, using the m.IrrC.2 methodology. Both mixed-library selections, NNK4-8 and NNK6-12, were undertaken in parallel.

Most interestingly, after R2, the combined recovery of the Strep-negative and *E*-negative DNA recovered was ~1% of the library for both selections, both greater than the *Z*-positive DNA recovered (when combined). From this, it could be argued that a greater number of Strep-binding and *E*-binding sequences were being depleted using this new irradiation protocol (m.IrrC.2).

The drop-off of enrichment for all DNA recovered in R5 of the NNK6-12 selection was attributed to experimental variance/ error and as such was not considered meaningful. The overall recoveries across both parallel selections (NNK4-8 and NNK6-12) were also much higher than traditionally seen during RaPID selections from the outgo. This may indicate that something went wrong with the selection or that some increased non-specific DNA binding event was occurring throughout both the NNK4-8 and NNK6-12 selections. This was a little concerning but regardless, following R5, the DNA recovered for all rounds (Strep-negative, *E*-negative, *Z*-positive), for both the NNK4-8 and NNK6-12 selections was prepared for NGS sequencing.

4.4 Identification of enriched peptide hits using NGS data and Python:

4.4.1 NGS data analysis of PhotoRaPID v2.0:

Following the same NGS data analysis workflow as for PhotoRaPID v1.0 (Figure 52), data processing generated 'top 10 enriched peptide hits' for the streptavidin-counter selection, *E*-counter selection and *Z*-positive selection from the DNA recovered in R5 (Figure 63(a)). The sequence homology and pie chart analysis (Figure 63(a and b)) revealed that the top two key PADI2 binding motifs 'WMD' and/ or 'HHW' were re-enriched in the TES for the *Z*-positive selections in both the NNK4-8 and NNK6-12 libraries (R5 data). This confirmed that despite the lower DNA recovery seen in the PhotoRaPID v2.0 selection the recovered hits appeared to be genuine PADI2 binders. Most importantly, following the addition of the streptavidin negative counter selection into the V2.0 workflow, enrichment of the 'IRL' motif was found to be drastically reduced within the *Z*-positive enriched peptide datasets for both the NNK4-8 and NNK6-12 selections. Accounting for 0% or 3% of reads for the top 100 *Z*-positive enriched peptides, respectively. This confirmed that the additional strep-negative counter selection successfully depleted the streptavidin binding associated motif peptides, as shown by the absence of the 'IRL' family.

Interestingly, the key difference seen between the top *Z*-positive sequences of NNK4-8 and the NNK6-12 libraries, was the relative enrichment of different motifs. In NNK4-8, there was a slightly biased distribution for 'HHW' over the 'WMD' motif, whilst the 'WMD' motif strongly dominated the NNK6-12 library (Figure 63(b)). This indicated that the library ring size directly influences the enrichment of certain core motifs. It could be rationalised that smaller macrocycles may favour specific clefts in the protein, that are not accessible to the larger macrocycles, accounting for the differences in motif enrichment seen.

It is also important to comment on the presence of the 'WMD' and 'HHW' motifs in the negative counter selections of both the NNK4-8 and NNK6-12 selections. Critically, despite the presence of the general core motifs, enrichment of specific sequences containing these motifs within the top negative sequences was not seen. This is relative to the enrichment seen of specific peptides containing these motifs in the top peptides of the *E*-negative or *Z*-positive selections. The most enriched negative sequence was at 0.02% of all strep-negative peptide reads, relative to 0.31% or 0.86% for the most enriched *E*-negative or *Z*-positive sequence respectively. This indicates that although the family motifs are present in the negative selection, they appear to be the background noise of an enriched library

where these motifs dominate. As such, perhaps it is unsurprising that some non-enriched peptides containing these motifs still have a slight tendency to bind to the negative streptavidin beads.

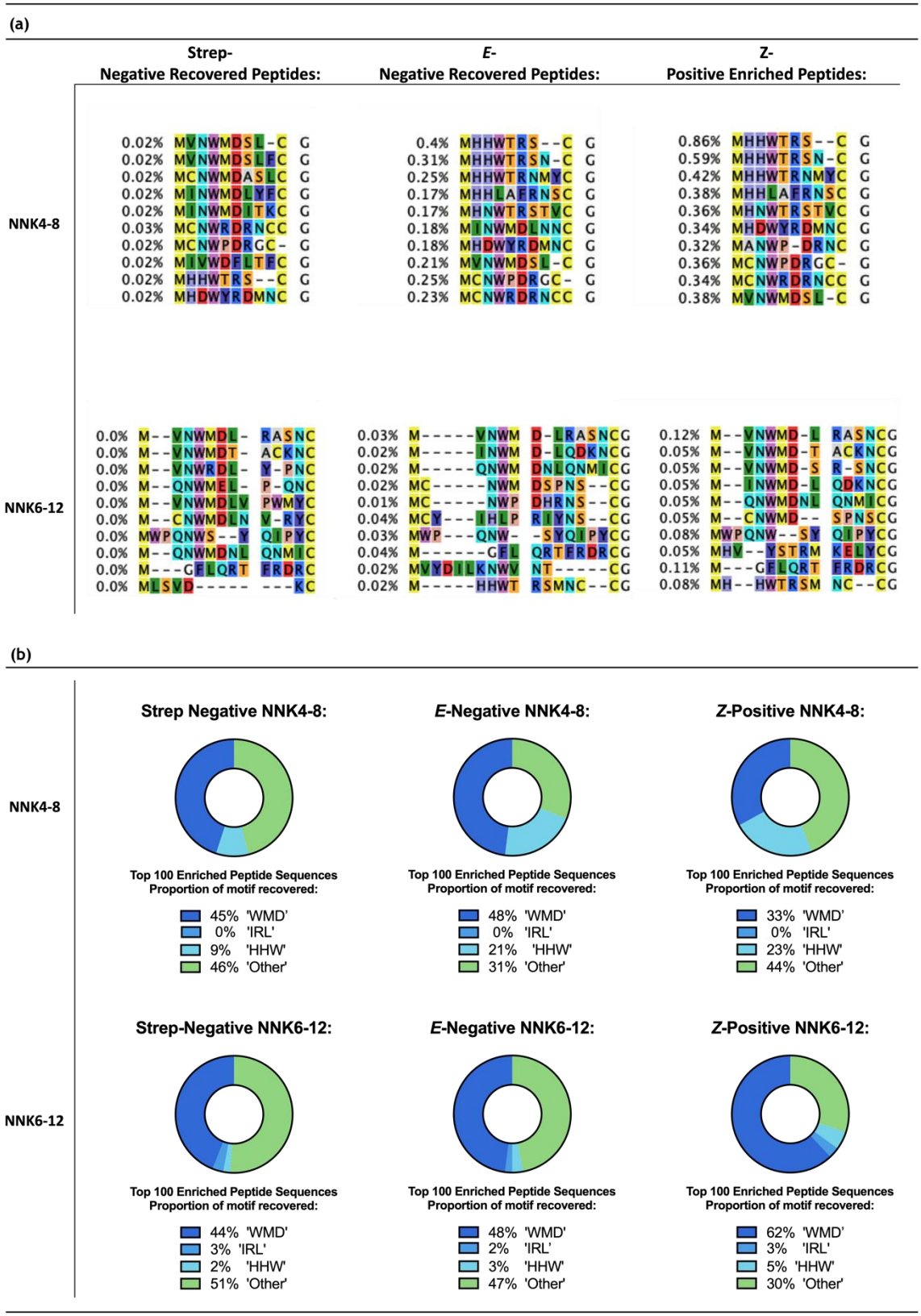


Figure 63: (a) Analysis of the enriched peptide hits identified from m.IrrC.2 of the PhotoRaPID v2.0 selection. The top enriched hits (TES) are shown aligned by sequence homology for the negative streptavidin, E-counter data

and Z-positive data. (B) Pie charts identifying the proportion that each enriched core motif identified from PhotoRaPID v1.0 was seen within the 100 TES of PhotoRaPID v2.0.

To reiterate, the addition of a streptavidin negative counter selection into the V2.0 workflow successfully prevented the enrichment of a known streptavidin binding motif ('IRL') and as such validated the changes implemented in the m.IrrC.2 workflow in PhotoRaPID v2.0.

4.4.2 A more tailored NGS analysis approach to find photoswitchable peptides:

As mentioned at the beginning of the chapter, in PhotoRaPID v1.0 the peptide candidates for further analysis were chosen by their relative enrichment, not considering the *Z/E* differential binding potential. Synthesis of the most enriched hits is traditionally the most common method of choosing which peptides to validate. But, for this work, it meant that the most enriched hits were chosen, only 'hoping' that they displayed good *Z/E* differential binding. *Z/E* differential binding is the crucial factor for developing light-dependent cyclic peptide ligands for photopharmacology applications. As such, for PhotoRaPID v2.0, greater emphasis was placed on seeing if large *Z/E* differentially binding peptides could be predicted from the NGS sequencing results. Accurate predictions would also save time and resources, as even for this field generally, false hit validation remains the bottleneck of all projects. This is further emphasised for this project as the photoswitch moiety (Fmoc-Photo-Acetic Acid (49)) required pre-synthesis on bulk scales (1-5 g) before SPPS synthesis started, adding further time and resource constraints. As such, ensuring that the hits chosen not only were excellent *Z*-binders, but also excellent *Z/E* differential binders was paramount.

To choose which peptides should be synthesised for further characterisation, for this tailored analysis, two factors were considered: **(1) the enrichment of the peptide sequence** and **(2) its potential to have large *Z/E* differential binding**.

(1) The enrichment of the peptide sequence: Peptide enrichment is the standard metric for choosing which peptides to validate and is readily available from the NGS sequencing data from the final round (R5), as used in PhotoRaPID v1.0.

(2) A peptide's potential to have large *Z/E* differential binding: To determine which peptides might have large *Z/E* differential binding from the sequencing data (R5), NGS *Z/E* enrichment scores were

generated. NGS *Z/E* enrichment was chosen as the predictive measure for *Z/E* differential binding under the following premise (Eq.4.4.2-1):

$$For\ Peptide\ x\ in\ NGS\ sequencing\ data:$$

$$\text{Let} \downarrow$$

$$Low\ enrichment\ of\ \text{peptide}\ x\ \text{in}\ E\ isomer\ (E\ negative\ selection\ NGS\ data) = weak\ binding\ to\ hPADI2$$

$$\text{Let} \downarrow$$

$$High\ enrichment\ of\ peptide\ x\ in\ Z\ isomer\ (Z\ positive\ selection\ NGS\ data) = tight\ binding\ to\ hPADI2$$

$$\text{Therefore} \downarrow$$

$$For\ peptide\ x\ where\ \text{Z enrichment} \gg E\ \text{enrichment}$$

$$\text{a large}\ \frac{Z}{E} score\ is\ seen\ \text{which is assumed to} = \frac{Z}{E}\ differential\ binding \qquad \text{(Eq.4.4.2-1)}$$

If a given peptide has low enrichment in the NGS *E*-negative counter-selection data, it 'must' bind poorly to PADI2. In contrast, if the same peptide binds tightly when in the *Z*-conformation, it 'must' have a high enrichment recovery in the *Z*-positive data. Therefore, if the same peptide displays high enrichment in the *Z*-positive dataset but low enrichment in the *E*-negative dataset then it will display a large *Z/E* differential score. This was assumed to be proportional to *Z/E* differential binding.

4.4.3 Calculating the *Z/E* enrichment scores:

To implement this new tailored NGS analysis, further Python scripts were developed. Following the workflow of Figure 64(a), the recovered R5 NGS sequences of the strep-negative, *E*-negative, and *Z*-positive datasets of the NNK4-8 selection were used. With limited time remaining in the PhD, only the enriched hits from NNK4-8, R5 were subjected to this further analysis. This was because peptides from the NNK4-8 library could begin to assess whether reduced ring sizes induce greater ring conformation change, hopefully producing peptides with greater light-dependent differential binding affinities.

The workflow to generate the *Z/E* enrichment scores using Python is as follows:

1. Standard filtering to remove unsuitable sequences (e.g. truncations) was first implemented, as done in PhotoRaPID v1.0 (Figure 52) and above (Figure 63).
2. The enrichment of a given sequence from the recovered *Z*-positive dataset (R5) was divided by its corresponding enrichment in the *E*-negative selection dataset (R5) to generate a fold difference termed a '*Z/E* enrichment score'.

3. The *Z*-isomer peptide enrichments recovered were also cross-referenced against recovery in the negative selection data generating a '*Z*/*N* enrichment score'.
4. With these scores in hand, the *Z/E* enrichment scores were then further categorised into:
   (1) The top 500 most enriched sequences (TES-500) recovered,
   or
   (2) The top 500 highest *Z/E* fold-change sequences (T*Z/E*-500) recovered.
5. With these two separate datasets in hand ((1) or (2)), the sequences from TES-500 and T*Z/E*-500 were then plotted.

4.4.4 Analysing the data – *Z*-positive peptide enrichment vs. predictive NGS *Z/E* enrichment scores:

When plotting the TES-500 and T*Z/E*-500 together (Figure 64(b)), most noticeably these peptides display two distinct populations. The T*Z/E*-500 peptides (orange) displayed much higher *Z/E* scores at the cost of much lower overall enrichment. In contrast, the peptides from the TES-500 sequence list (blue) had much lower fold changes but had dramatically increased overall enrichment (~10-fold higher enrichment for the most enriched peptide between TES-500 and T*Z/E*-500).

From this data, an inversely proportional relationship between the relative NGS *Z/E* score vs. its absolute enrichment was seen. This suggested that the 'best' photoswitchable peptides may not the be most enriched sequences, agreeing with results from previous chapters.

Replotting and scaling each graph individually, the inversely proportional relationship between peptide enrichment vs. *Z/E* score for the TES-500 peptides became more apparent (Figure 64(c)). In contrast, linear scattering was seen in the T*Z/E*-500 graph (Figure 64(d)), masking slightly, the same underlying inversely proportional trend of peptide enrichment vs. *Z/E* score. This was initially thought to be a rounding error of the peptide enrichment scores in the Python code, however, this was not the case. One possible explanation is that the sequences were so poorly enriched and as such the linear trends are artefacts. Sequences this poorly enriched are not typically chosen from RaPID selections, but these peptides appeared to have the highest *Z/E* scores. If these *Z/E* scores do translate well to *Z/E* binding and or inhibitory differentials, this would be significant. Particularly as these sequences would not have been typically chosen for SPPS under the standard NGS data analysis used in the field. All peptides plotted had high *Z*/*N* scores which was encouraging and as such not a metric that needed to be explored further during this analysis.

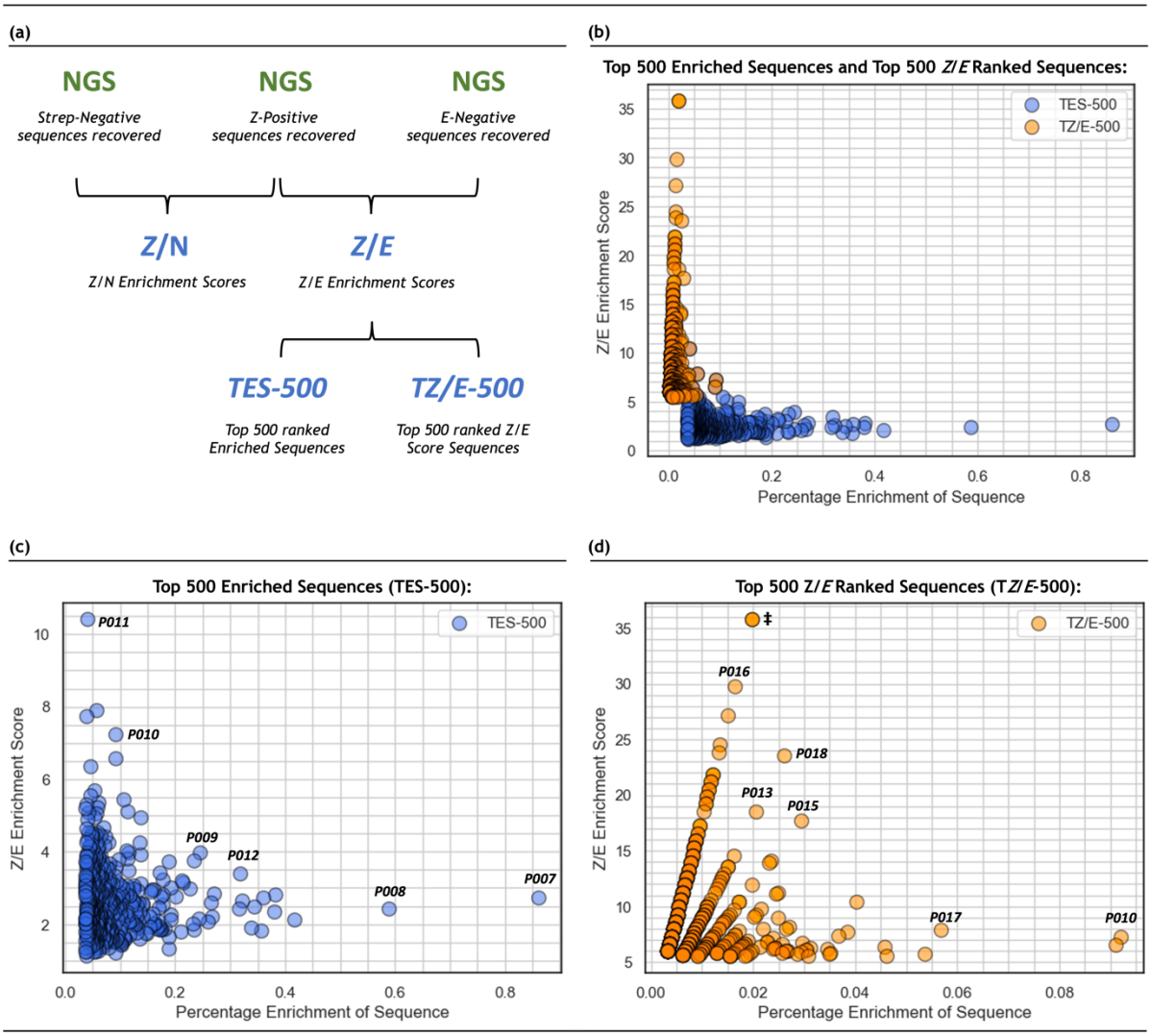


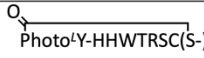
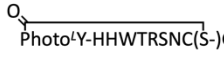
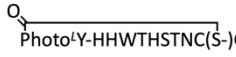
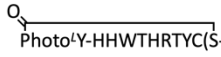
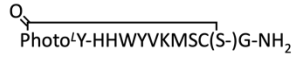
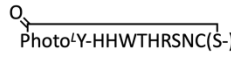
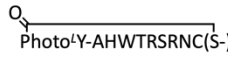
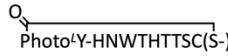
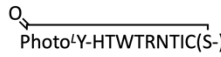
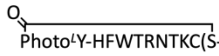
Figure 64: (a) The Python workflow to generate the NGS Z/E and Z/N enrichment scores, (b) Scatter graph of both the Top 500 enriched sequences (TES-500) and the Top 500 Z/E ranked peptides combined, (c) TES-500 and (d) TZ/E-500 Sequences plotted individually by Z/E scores. Peptides annotated are those which were selected for SPPS synthesis, ‡ indicates the most enriched peptide (P014) which could not be synthesised.

To assess whether these NGS *Z/E* scores are representative of the *Z/E* differentials observed during biophysical testing, twelve peptides (annotated in Figure 64(c), Figure 64(d) were chosen for synthesis on SPPS. A range of peptides displaying different enrichment and *Z/E* Scores were purposefully chosen to verify if the binding differentials observed in the sequence analysis could be recapitulated experimentally using SPR experiments.

4.5 SPPS of chosen peptide hits:

Of the twelve peptides chosen from across the NGS *Z/E* enrichment score graphs, six peptides were chosen from the TES-500 graph (Figure 64(c)). This was because these peptides are the most *Z*-enriched and so most likely to be successful candidates. From traditional analysis, they would also be the most likely candidates chosen and so was also a good reference point to the standard methods of choosing peptides. Additionally, six more peptides were chosen from the T*Z/E*-500 graph (Figure 64(d)). These peptides had much lower Z-enrichments and so would typically not be followed up. These peptides, however, displayed the highest predictive *Z/E* scores. As such, these peptides were the most interesting from the perspective of this project. With these peptides chosen, only eight out of the twelve could be fully synthesised. P010, P013, P014 and P018 proved challenging to synthesise on SPPS due to several truncations occurring during the automated synthesis and as such due to the limited time remaining in the PhD, these peptides were not brought forward at this time. These failed peptides included the peptide with the largest NGS *Z/E* score (‡ symbol in Figure 64(d), *Z/E* score = 35, P014), which would have been interesting to study. The successful peptides were synthesised via SPPS following resynthesis of Fmoc-Photo-Acetic Acid (49) on a 5 g scale, and the photoswitch was manually coupled as previously described. The peptides were then cyclised and purified without issue (Table 4).

Table 4: Peptide hits enriched against PADI2 synthesised, via SPPS from the NGS Z/E Scatterplots TES-500 or TZ/E-500 from PhotoRaPID v2.0.

Peptide ID:	Peptide Sequence:	Total Ring Size:	TES-500 or TZ/*E*-500:	NNK:	*Z/E* Score:	Enrichment in Library (%):	Expected Mass: $[M+H]^+$ / $[M/3]^+$	Observed Mass: $[M/3]^+$
P007		9	TES-500	6	2.74	0.86	1467.73 490.25	490.6
P008		10	TES-500	7	2.42	0.59	1581.77 528.27	528.5
P009		11	TES-500	8	3.98	0.25	1663.77 555.60	555.8
P010		11	TES-500	8	7.23	0.092	1781.86 594.96	Not observed
P011		11	TES-500	8	10.4	0.04	1668.85 557.29	557.5
P012		11	TES-500	8	3.40	0.32	1600.78 534.60	534.4
P013		11	TZ/E-500	8	18.6	0.021	1732.83 578.62	Not observed
P014		11	TZ/E-500	8	35.8	0.020	1719.81 574.28	Not observed
P015		11	TZ/E-500	8	17.7	0.03	1671.85 558.29	558.6
P016		11	TZ/E-500	8	29.9	0.02	1627.76 543.6	543.9
P017		11	TZ/E-500	8	7.9	0.06	1672.86 558.63	558.9
P018		11	TZ/E-500	8	23.5	0.026	1733.89 578.97	Not observed

4.6 Experimentally validating the photoswitchable cyclic peptides:

4.6.1 SPR – Obtaining SPR-*Z/E* binding differentials:

Preliminary binding affinities of the peptides for each isomer (*E*/*Z*) against PADI2 were then measured by SPR (Table 5). SPR was chosen as before, to obtain K_D values for each peptide as each conformational isomer (*E* and *Z*). This technique was chosen to enable easy comparison with the binding affinity of the peptides from PhotoRaPID v1.0. Either 300 nM or 75 nM biotinylated PADI2 was immobilised on a Biacore CAP chip, for 120 seconds and the peptides were flowed using a single cycle protocol of increasing peptide concentrations, except for one replicate of P011 where a multicycle protocol (full association and dissociation of each increasing concentration) was trialled (Table 5).

P008 was also not tested as it was not concentrated enough to ensure 0.1% DMSO f.c. whilst retaining a sufficiently high-top concentration.

Table 5: SPR table of KD values and Z/E binding differentials for the SPPS peptides synthesised from the NGS scatter plots. Asterisks (*) indicate that binding to PADI2 was observed but K_Ds could not be determined. Double daggers (‡) for P012 indicate that only affinity plots were used to generate the K_D shown, as such no kinetic parameters are presented. Kinetics data and associated sensograms shown are representative of a minimum of 2 independent experiments for each isomer of each peptide tested (replicates at Appendix 38).

Peptide ID:	*Z*-k_a (Ms^{-1})	*Z*-k_d (s^{-1})	*Z*-K_D (nM):	*E*-k_a (Ms^{-1})	*E*-k_d (s^{-1})	*E*-K_D (nM):	*Z*/*E* Binding differential:
P007	N/A	N/A	*	N/A	N/A	*	*
P009	N/A	N/A	*	N/A	N/A	*	*
P011	$6.7 \pm 4.5 \times 10^4$	$1.4 \pm 0.1 \times 10^{-2}$	271 ± 162	N/A	N/A	>5000 (n=2)	>18-fold
P012	‡	‡	710 ± 12	‡	‡	1020 ± 14	1.4-fold
P015	$7.7 \pm 7 \times 10^4$	$1.2 \pm 1.2 \times 10^{-2}$	202 ± 111	N/A	N/A	>5000	>25-fold
P016	N/A	N/A	>5000	N/A	N/A	>5000	N/A
P017	N/A	N/A	>3000	N/A	N/A	>3000	N/A

Peptides P011, P012 and P015 displayed nanomolar *Z*-isomer binding, $_ZK_D$ at 271 ± 162 nM, 710 ± 12 nM and 202 ± 111 nM respectively, while the remaining peptides were weakly binding (>5 µM), at top concentration tested (see table). P015 demonstrated the greatest *E*/*Z* binding differential from these peptides at >25-fold, however, accurate K_Ds in the *E*-isomer (*E*-P015) were challenging to determine and as such the top concentration tested (5µM) was used to generate the fold difference, this was the same for the *E*-isomer of P011. In general, it was challenging to obtain K_D values for the *E*-isomer of each peptide as they were largely weakly binding. This resulted in more pronounced bulk drift contributions and limit of detection liabilities which disrupted the SPR analysis software. Poorer *E*-isomer binding was an excellent indicator that the peptides are preferentially enriched for *Z*-isomer specifically but made the quantification of the E/Z binding differentials very challenging.

P007 and P009 (in table assigned *) displayed comparable binding affinities in both isomers of the peptide however, due to variability between the runs in the RU generated, the degree of contributing bulk drift and very limited time remaining in the PhD, accurate K_Ds could not be shown at this time.

P011 and P015 both displayed very good *Z* affinity and apparent weaker *E*-isomer binding affinities. From the data, predicted *Z/E* binding differentials at >18-fold and >25-fold respectively, could be

tentatively calculated. As mentioned, the exact $_EK_D$ for P011 and P015 could not be fully determined and as such only a preliminary *Z/E* differential is shown. This is something to investigate further, as the *Z/E* differential may change or be larger upon further *E*-isomer data collection.

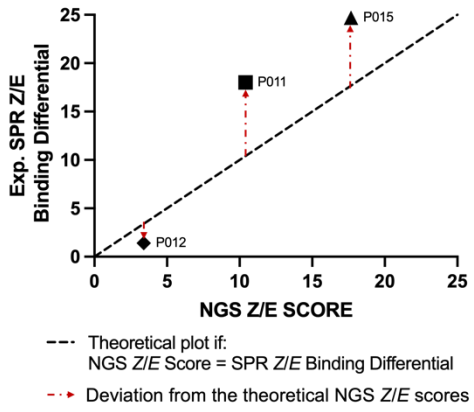
P012 displayed moderate nanomolar binding in the *Z*-isomer and weaker binding in the *E*-isomer as desired. The *Z/E* binding differential was very modest at 1.4-fold. However, as P012 contained a novel sequence motif (not containing 'WMD' or 'HHW'), it may be valuable to return to the sequencing data to identify any other peptides that have high enrichment and contain similar sequence homology to this peptide. Although with its modest *Z*-isomer binding and minimal *Z/E* differential binding, other peptide candidates should take priority.

Peptides P016 and P017 had very poor binding in both the *Z*-isomer and *E*-isomer. This could be attributed to the very low enrichment of each peptide sequence in the Z-isomer NGS data, at 0.02% and 0.06% respectively (Figure 64 and Table 4). From these results, it was more surprising that P015 was able to bind so potently ($_ZK_D$= 202 ± 111 nM) in comparison. This is despite a relatively poor sequence enrichment of 0.03% at a similar percentage to P016 and P017. This suggests that peptide enrichment is not the only key factor when choosing peptides for SPPS characterisation. This has been a consensus in the Walport group for some time and is reflected in many publications using the RaPID platform.[161–164] As P015 also had a large *Z/E* binding differential (~>25-fold), this may also imply that good photoswitchable peptides require weaker 'active isomer' binding, which could be related to weaker enrichment in the *Z*-positive dataset. This is to enable a sufficient abolishment of binding when in the 'inactive' isomer. This is also the case for P011 which demonstrates a large *Z/E* binding differential (>18-fold) despite a poor enrichment of 0.04%. As this is a very limited sample size, the overall consensus however, remains unclear.

4.6.2 Do the predicted NGS *Z/E* Scores correlate with the SPR *Z/E* binding differentials?

From the peptides whose differentials could be calculated, tentative preliminary correlations could be drawn. General agreement between the NGS *Z/E* scores vs. the experimental SPR K_D *Z/E* binding differentials could be seen, (Figure 65). But it is critical to state the preliminary nature of these conclusions. Additional data points and precise *E*-isomer K_D values are greatly needed. From this limited data set though, it does provide early speculation that this method of post-selection data

analysis could be valuable to identifying peptides with differential binding. With the caveat that this correlation does not account for peptides that do not bind (i.e., P016 and P017). More robust analysis in the future would be strongly required to fully validate these claims but, unfortunately, due to the time constraints of the PhD, that was not possible at this time.



Peptide:	Symbol:	NGS *Z/E* SCORE:	Experimental SPR *Z/E* Binding Differential:
P012	◆	3.40	1.4-fold
P011	■	10.41	>18-fold
P015	▲	17.67	>25-fold

Figure 65: Preliminary correlations between the NGS Z/E score vs. the Experimental Z/E binding differentials generated from SPR binding affinities.

4.6.3 UV-Vis spectrophotometry and thermal Half-life:

In parallel, UV-Vis spectrophotometry was also undertaken for peptides P011, P015, P016 and P017 to determine the photochemical properties of these representative macrocycles (Figure 66(a)). As this was undertaken in parallel with SPR, it was not known at the time which peptides were the most interesting, in relation to binding and *Z/E* differentials. For future work, the photochemical properties of the remaining peptides should also be characterised, particularly for the peptides that display the largest SPR-*Z/E* differential binding.

Similar to the analysis of PhotoRaPID v1.0, peptides were irradiated at each wavelength (365 nm or 495 nm) for varying lengths of time (1-5 mins). The absorption profile of the photoswitch within the macrocycles was then assessed across 200-800 nm to identify the proportion of *Z* or *E*-isomer at each wavelength upon reaching the PSS. Peaks around 200-300 nm result from residual DMSO absorption in the buffer and absorption from tyrosine and or tryptophan residues in the peptides if present (280 nm).[151]

All peptides evaluated had near identical photochemical properties as those discovered from PhotoRaPID v1.0, with a PSS_{365nm} = 85-97% ± 3% (*E*→ *Z*) and a PSS_{495nm} = 94-97% ± 3% (*Z*→ *E*) demonstrating that they could be near quantitatively switched back. This meant that these macrocycles had largely quantitative isomerisation as desired. P011 displayed the lowest *Z*-isomer and *E*-isomer PSS populations following irradiation at PSS_{365nm} 85% ± 3%, and PSS_{495nm} 94% ± 3%. The exact reason for this reduction remains unknown and may be due to intrinsic AA interactions with the azopyrazole motif or the conformation of peptide.

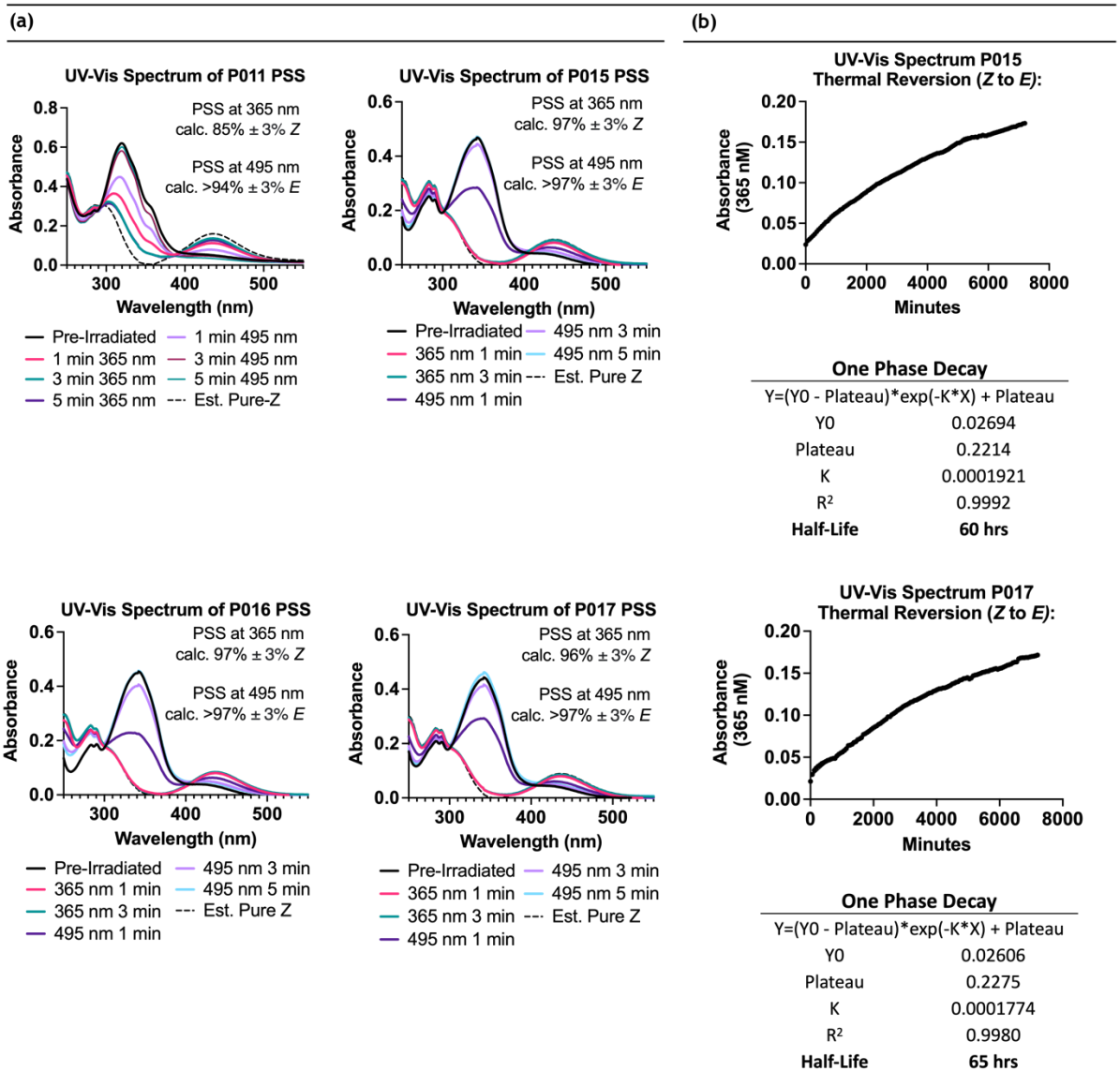


Figure 66: UV-Vis Spectroscopy of P011, P015, P016 and P017 peptides to assess their photochemical properties. Peptides were diluted into H_2O (at ~20 µM f.c. ~2% DMSO). PSS was achieved via 365nm (LED 3x 800 mW at 100% output) or 495 nm (LED 3x 750 mW at 100% output) irradiation <5 mins. Pre-irradiated is defined as the pure E-isomer.

With the PSS photoconversion efficiencies identified and with very limited time remaining, the thermal half-lives ($\tau_{1/2}$) of peptides P015 and P017 were then evaluated (Figure 66(b)). To identify the $\tau_{1/2}$, as before, the peptides were irradiated at 365 nm for 5 minutes and then left within the spectrophotometer (in darkness). Readings at 365 nm were taken every 60 mins tracking the increasing absorbance of the π-π* peak.

Both P015 and P017 displayed extended thermal half-lives of 60 and 65 hrs respectively, in comparison to the peptides identified from PhotoRaPID v1.0 (greatest 31 hrs (P001)). This is closer to the half-life of the photoswitch core alone which is 10 days. This suggests that the $\tau_{1/2}$ is highly dependent on the exact amino acid sequence and macrocycle ring size. The aqueous buffer with minimal DMSO was chosen specifically to mirror the thermal half-lives that could be expected within biological environments. The half-lives seen by these two representative peptides are in the optimal thermal reversion window (hours-to-days) required for use in cellular experiments, providing further support for the use of PhotoRaPID-derived photoswitchable macrocycles as chemical biology and photopharmacology tools.

4.7 Conclusions:

In conclusion, a modified irradiation condition, m.IrrC.2 was developed, incorporating an additional streptavidin-negative counter selection into the workflow. The aim of this was to remove streptavidin binding peptides which were found to be enriched in peptide hits from IrrC.2. This proved successful as the enriched peptides recovered from the *Z*-positive NGS data were completely absent of the 'IRL' motif associated with streptavidin binding (NNK4-8 library). Additionally, the m.IrrC.2 workflow re-enriched the 'HHW' motif, the motif from PhotoRaPID v1.0. This motif produced the peptide with the greatest *Z/E* differential binding (P006, $K_{D\ Z/E}$ = 29-fold) from PhotoRaPID v1.0. As such it was excellent to see that they could not only be re-enriched but also that m.IrrC.2 was able to achieve this.

Following on from the standard NGS data analysis, a tailored NGS workflow was also implemented where NGS *Z/E* enrichment scores were generated. This allowed for the relative enrichment of a given sequence in the *Z*-positive NGS data relative to the *E*-negative NGS data to be assessed. Scatterplots were generated showing the distribution of *Z*-peptide enrichment vs. *Z/E* enrichment against PADI2 from the R5 datasets. From this, peptide candidates across a distribution of *Z/E* scores and *Z*-peptide enrichments were chosen from the sequencing data for synthesis via SPPS. This was particularly

important as the current workflow from PhotoRaPID v1.0 relied solely on the enrichment of the peptides and did not consider whether the *Z/E* binding differential could be predicted. This is an extremely important metric for developing large *Z/E* differentially binding photoswitchable peptides and had been understudied thus far in the project.

From the eight peptides characterised via SPR, three *Z*-nanomolar binders were found (P011, P012 and P015), displaying *Z/E* differential binding ≥25-fold. From the *Z/E* TES-500 plot specifically (Table 5), P015 was identified, demonstrating the second largest *Z/E* differential binding seen to date in the project (~25-fold). This peptide was also a unique sequence motif identified from the m.IrrC.2 condition, not seen previously. This provided further evidence that the m.IrrC.2 was superior to the original IrrC.2 workflow. Furthermore, the location of P015 within the TES-500 scatterplot (Figure 64), demonstrated the utility of the graphs. This is because it would not have been a peptide traditionally chosen using a standard sequence homology and enrichment approach (enrichment of 0.03%). Its presence in the mid-range area of the graph between Z-positive enrichment and *Z/E* enrichment score also provides further evidence that the best photoswitchable binders may not be the most enriched sequences. This is because the photoswitchable peptides could require weaker 'active isomer' binding such that the sufficient abolishment of binding when in the 'inactive' isomer can be achieved.

With the data obtained to date it is not fully clear if the NGS *Z/E* scores can accurately predict the *Z/E* differential binding via SPR, however, of the peptides where preliminary *Z/E* binding differentials could be broadly calculated, a general correlation between the NGS *Z/E* score and *Z/E* SPR binding differential could be seen (Figure 65). From the limited data acquired thus far, it appears that the *Z/E* SPR binding differential is generally close to or much larger than what is predicted, which is promising, but some unexplained variances seen muddies the predictive aspect of this approach. Overall, it is important to stress that further validation of the SPR binding affinities underpinning the *Z/E* differentials is critical to ensure the validity of these conclusions and as such only rudimentary conclusions can be drawn at this time. One method to provide further validation would be to use a (competitive) fluorescence anisotropy assay, as an orthogonal method to SPR. Unfortunately, with the time remaining in the PhD, development of this orthogonal assay was not possible. One additional caveat is that this correlation only holds if artefact peptides are omitted (e.g. P016 and P017), which despite enrichment in the TES-500 and T*Z/E*-500 graphs did not appear to bind via SPR.

It may also be worthwhile re-evaluating the data collection used to feed the Python predictive scores. Perhaps, post-selection, performing one additional round where the enriched library (with multiple copies of each sequence) is equally split between two sets of immobilised PAD2 beads, each irradiated to the opposite isomer. This would ensure that there is no bias in the sequential panning step during m.IrrC2 (E-negative then Z-Positive during the selection) which is currently used to generate the predictive scores. Overall, these are still very exciting preliminary findings and would potentially enable the faster and more robust identification of light-dependent cyclic peptides immediately post-selection with further tuning. This would be vital for peptide candidate selection as the SPPS and validation of the hits remain the bottleneck of the RaPID workflow in general.

The role of ring size on *Z/E* differentials also remains largely inconclusive. This unfortunately could not be studied in detail during PhotoRaPID v2.0 as the macrocycles enriched within the TES-500 and T*Z/E*-500 datasets were largely NNK8 (the largest AA library size used in the selection – NNK4-8). This is perhaps unsurprising as larger macrocycles enable a greater number of binding interactions with the protein and therefore tend to bind with greater affinity. As a consensus, it appears that there is a trade-off between enrichment and ring size. Therefore, the largest ring-sized library possible should be used to still achieve sufficient potency ($_{Z}K_{D}$ in the nM range) for the target, whilst keeping the ring sizes small enough to ensure large conformationally induced *Z/E* differential binding.

Additional work to assess the inhibitory activity of these peptides against PADI2 using the *in vitro* COLDER assay also needs to be assessed. Speculatively, from the poorer inhibitory activities of the peptide hits from PhotoRaPID v1.0 also bearing the 'HHW' motif, it could be imagined that the IC_{50} differentials would not correlate well. But as P011, P012 and P015 all contain new motifs, assessing the IC_{50} inhibitory activity these new hits would be an important piece of future work.

In summary, this chapter demonstrated the successful optimisation of PhotoRaPID using an improved m.IrrC.2 workflow. Additionally, the generation of NGS *Z/E* enrichment scores as a predictive measure for SPR-*Z/E* binding differentials could hopefully begin to enable greater confidence when choosing which peptides to synthesise and characterise in the future. It is without a doubt that additional data is needed to fully validate if a stronger correlation is seen between this predictive measure and the experimental *Z/E* binding differentials currently identified.

Overall, the data generated thus far, despite its very preliminary nature, has laid an exciting foundation for any further work on PhotoRaPID v2.0 methodology in the future.

5 Conclusions and Future Directions

5.1 Conclusions:

In conclusion, this PhD led to the successful development of a next-generation *in vitro* light-responsive cyclic peptide display platform, termed PhotoRaPID. Azopyrazole-based molecular photoswitch analogues were first generated, with their loading efficiency tested on a model tRNA for integration within RaPID using Flexizyme technology. Two analogues, ClAc-Photo-^{L}Y-CME (36) and ClAc-Photo-Propyl-CME (41) could be successfully aminoacylated onto an $^{ini}Met_{CAU}$ tRNA and translated by the ribosome, within a known peptide sequence, using the FIT system under genetic code reprogramming. This represented the first reported example of a genetically encoded photoswitch compatible with mRNA display.

Development and testing of the first PhotoRaPID system (PhotoRaPID v1.0) then began by ensuring that light irradiation, required for photoswitching, did not adversely affect the RaPID library DNA. Preparation of the first protein target PADI2, followed. PADI2 was an important proof-of-concept target, as it is dysregulated in several disease profiles and the development of further tools to explore its substrate scope is still highly desired by the field. Three distinct PhotoRaPID workflows, IrrC.1, IrrC.2 and IrrC.3 were also subsequently designed. This was to promote the evolution of photoswitchable macrocyclic peptides with light-responsive differential binding during the selection process. All three IrrCs effectively enriched for light-responsive photoswitchable macrocyclic peptide motifs against PADI2, identified as 'WMD' and 'HHW'. More specifically, IrrC.1 and IrrC.3 produced P003 and P006, which have low nanomolar $_{Z}K_{D}$s of 3.7 ± 2.9 and 41.9 ± 37 nM, respectively. Most excitingly, P006 demonstrated 29-fold differential binding between the *Z*- and *E*-isomers. This constitutes the largest fold-change in differential binding seen within this project and within the literature to date.

Further work on these peptides identified that P003 and P006 were able to inhibit PADI2 citrullination activity, with up to a 6-fold difference in IC_{50}s between the two light-responsive *E* and *Z* macrocyclic isomers. But, with IC_{50} values in the low µM range, these peptides should be considered sub-optimal

for use as light-responsive inhibitors. An explanation for this weakened inhibitory activity may be that the peptides do not bind fully in the active site of PADI2 or bind somewhere completely different on the protein. They may cause partial displacement of the arginine mimetic substrate (BAEE) allosterically, promoting changes in the active site conformation. Further work, however, would be required to substantiate this definitively. Structural studies such as by CryoEM, X-ray crystallography or further biophysical studies in which the active site was blocked, and peptide binding affinities determined could be used to explore this further. However, as this was not a crucial aspect of the PhotoRaPID project it was not further examined. The weakened inhibitory activity seen however, highlights an intrinsic drawback to the standard RaPID system, not specifically PhotoRaPID v1.0, as the system is designed to preferentially enrich peptide binders across the whole protein surface, not only the active site. This could be something developed further if light-responsive photoswitchable cyclic peptide inhibitors are desired. Particularly, screening the active site of the protein of interest whilst blocked by a covalent inhibitor or suicide substrate in parallel to the unliganded protein, ignoring sequences that preferentially enriched in both selections. This would hopefully promote the identification of active-site binders.

The photochemical properties of these peptides were also investigated by assessing the photoconversion efficiency of the *E*→*Z* and *Z*→*E* isomerisation, as well as the thermal reversion profile ($\tau_{1/2}$) of these peptides. Gratifyingly, P003 and P006 demonstrated near quantitative isomerisation with $\geq$92-94 ± 3% to the *Z*-isomer under a PSS of 365 nm and $\geq$97± 3% *E*-isomer populations upon reaching the PSS at 495 nm. This isomerisation profile was similar to the parent azopyrazole (4pzMe (7)), even when directly integrated within the macrocyclic peptides. Importantly these integrated photoswitch motifs continue to surpass the photochemical properties of classical azobenzenes used within the literature as desired. It is also important to highlight that P003 and P006 did display a more accelerated $\tau_{1/2}$ of 3.5 and 10.8 hrs respectively. This was substantially faster than 4pzMe (7) (10 days), but critically as these peptides still had $\tau_{1/2}$ values on the order-of-magnitude of hours, still making them suitable for chemical biology applications.

With the success of PhotoRaPID v1.0, further optimisation of the IrrC.2 workflow to generate a more robust discovery platform was then undertaken (PhotoRaPID v2.0). This led to the re-enrichment of the previously identified photoswitchable binding motif 'HHW' for PADI2, with the peptides discovered displaying similar characteristics (nanomolar $_zK_D$s and similar *Z/E* binding differentials) as

PhotoRaPID v1.0. This demonstrates the reproducibly of the m.IrrC.2 workflow to enrich light-responsive cyclic peptides against PADI2 a second time.

In-depth data-driven analysis was also designed into PhotoRaPID v2.0, aiming to identify, in a more streamlined manner, enriched peptides with the maximal *Z/E* differential binding possible. This is especially important as current methods require trial-and-error selection from the sequencing data, something that remains a bottleneck of post-selection hit characterisation across the field. Using the post-selection DNA recovered, personally developed Python scripts were employed to see if a correlation between the NGS *Z/E* score vs the SPR K_D binding affinity could be identified. From the preliminary data acquired thus far, excitingly, a tentative general correlation could be seen. Further validation of the *Z/E* binding differentials calculated is essential, most probably with orthogonal assays such as a competitive FP anisotropy (providing there is indication that the peptides bind at the same region). One additional caveat is that this correlation only holds if artefact peptides are omitted (e.g., omission of enriched but non-binding sequences to the protein target). This is something where there appears to be no easy solution to identify such false positives, beyond the standard SPPS and characterisation. Overall, these are still exciting preliminary findings and would potentially enable the faster and more robust identification of light-dependent cyclic peptides immediately post-selection.

One additional research aim, the influence of smaller starting NNK libraries (smaller ring sizes) upon the maximal degree of *Z/E* differential binding, however, remains unclear. This is because during the photo-selections the smaller ring sizes were repeatedly outcompeted by the larger library members. Although not wholly unexpected, this was also seen even despite a smaller NNK4-8 library size being employed during PhotoRaPID v2.0. It may prove that an interplay between affinity enrichment vs differential binding, limits the enrichment of very large differentially binding peptides. As such, this would be an intrinsic limitation of affinity enrichment-based display platforms generally. Further work in this area would be perhaps very beneficial to further promote the development of novel light-responsive tool compounds with even larger differential binding.

5.2 Future directions:

Building on the successful development of PhotoRaPID applied to a single proof-of-concept target, future directions of the project may include expansion of the PhotoRaPID discovery platform to other protein targets of interest. This should also be coupled with the relevant *in vitro* cell-based

applications, employing the powerful photoswitchable nature of these new tool compounds to answer novel questions in chemical biology.

From a further methodology point-of-view, using more focused small library ring sizes (e.g., NNK4-5), could be valuable to fully interrogate if a small ring size is important to promote large, switchable, differential binding. These focused libraries could include a pre-enriched motif fixed within the randomised AA encoding region, possibly circumventing any issues in ensuring suitable affinity enrichment.

Other future directions could include optimisation of the translation efficiency of ClAc-Photo-Propyl-CME (41) by modifying the affinity of this loaded tRNA to the ribosome P-site, optimisation of the $^{Ini}Met_{CAU}$ tRNA stem would be required, in a similar fashion to those developed by Suga for elongation reprogramming. This remains unprecedented for direct ribosome P-site affinity optimisation however.[97] Finally, an exciting blue-skies future avenue could be the development of a single light-responsive cyclic peptide, targeting two independent protein targets, with its binding dependent upon the given isomerisation state. This could perhaps be achieved with a bicyclic photoswitchable cyclic peptide library and would be an extraordinary achievement.

Overall, these suggestions aim to ensure the robustness of this next-generation discovery platform for use against any recombinantly expressed protein of interest. To further expand its ability to identify *de novo*, potent, light-responsive cyclic peptides with large differential binding and, to begin exploring photoswitchable cyclic peptide chemical biology beyond the foundation set by this PhD work.

6 References

(1) Osorio, D.; Vorobyev, M. A Review of the Evolution of Animal Colour Vision and Visual Communication Signals. *Vision Res* **2008**, *48* (20), 2042–2051. https://doi.org/10.1016/J.VISRES.2008.06.018.

(2) Hohmann-Marriott, M. F.; Blankenship, R. E. Evolution of Photosynthesis. *Annu. Rev. Plant Biol.* **2011**, *62*, 515–548. https://doi.org/10.1146/ANNUREV-ARPLANT-042110-103811.

(3) Photochromism. *The IUPAC Compendium of Chemical Terminology* **2014**. https://doi.org/10.1351/GOLDBOOK.P04589.

(4) Emiliani, V.; Entcheva, E.; Hedrich, R.; Hegemann, P.; Konrad, K. R.; Lüscher, C.; Mahn, M.; Pan, Z. H.; Sims, R. R.; Vierock, J.; Yizhar, O. Optogenetics for Light Control of Biological Systems. *Nature Reviews Methods Primers 2022 2:1* **2022**, *2* (1), 1–25. https://doi.org/10.1038/s43586-022-00136-4.

(5) Velema, W. A.; Szymanski, W.; Feringa, B. L. Photopharmacology: Beyond Proof of Principle. *J Am Chem Soc* **2014**, *136* (6), 2178–2191. https://doi.org/10.1021/JA413063E/ASSET/IMAGES/LARGE/JA-2013-13063E_0015.JPEG.

(6) Volarić, J.; Szymanski, W.; Simeth, N. A.; Feringa, B. L. Molecular Photoswitches in Aqueous Environments. *Chem Soc Rev* **2021**, *50* (22), 12377–12449. https://doi.org/10.1039/D0CS00547A.

(7) Zbigniew L. Pianowski. Molecular Photoswitches. *Molecular Photoswitches* **2022**. https://doi.org/10.1002/9783527827626.

(8) Olesińska-Mönch, M.; Deo, C. Small-Molecule Photoswitches for Fluorescence Bioimaging: Engineering and Applications. *ChemComm.* **2023**, *59* (6), 660–669. https://doi.org/10.1039/D2CC05870G.

(9) Mangubat-Medina, A. E.; Ball, Z. T. Triggering Biological Processes: Methods and Applications of Photocaged Peptides and Proteins. *Chem Soc Rev* **2021**, *50* (18), 10403–10421. https://doi.org/10.1039/D0CS01434F.

(10) Fleming, C. L.; Li, S.; Grotli, M.; Andréasson, J. Shining New Light on the Spiropyran Photoswitch: A Photocage Decides between Cis-Trans or Spiro-Merocyanine Isomerization. *J Am Chem Soc* **2018**, *140* (43), 14069–14072. https://doi.org/10.1021/JACS.8B09523/ASSET/IMAGES/LARGE/JA-2018-09523X_0002.JPEG.

(11) Szymański, W.; Beierle, J. M.; Kistemaker, H. A. V.; Velema, W. A.; Feringa, B. L. Reversible Photocontrol of Biological Systems by the Incorporation of Molecular Photoswitches. *Chem*

Rev **2013**, *113* (8), 6114–6178. https://doi.org/10.1021/CR300179F/ASSET/IMAGES/LARGE/CR-2012-00179F_0053.JPEG.

(12) Bovoloni, M.; Filo, J.; Sigmundová, I.; Magdolen, P.; Budzák, Š.; Procházková, E.; Tommasini, M.; Cigáň, M.; Bianco, A. Unsymmetrical Benzothiazole-Based Dithienylethene Photoswitches. *PCCP* **2022**, *24* (38), 23758–23768. https://doi.org/10.1039/D2CP02325C.

(13) Ozhogin, I. V.; Zolotukhin, P. V.; Mukhanov, E. L.; Rostovtseva, I. A.; Makarova, N. I.; Tkachev, V. V.; Beseda, D. K.; Metelitsa, A. V.; Lukyanov, B. S. Novel Molecular Hybrids of Indoline Spiropyrans and α-Lipoic Acid as Potential Photopharmacological Agents: Synthesis, Structure, Photochromic and Biological Properties. *Bioorg Med Chem Lett* **2021**, *31*, 127709. https://doi.org/10.1016/J.BMCL.2020.127709.

(14) Rovira Algans, X.; Llebaría, A.; Gomez-Santacana, X.; Kirchner, S.; Pianowski, Z. Photopharmacology of Antimitotic Agents. *Int. J. Mol. Sci. 2022, Vol. 23, Page 5657* **2022**, *23* (10), 5657. https://doi.org/10.3390/IJMS23105657.

(15) Tavassoli, A.; Hamilton, A. D.; Spring, D. R.; Beharry, A. A.; Woolley, G. A. Azobenzene Photoswitches for Biomolecules. *Chem. Soc. Rev* **2011**, *40*, 4422–4437. https://doi.org/10.1039/c1cs15023e.

(16) Ash, C.; Dubec, M.; Donne, K.; Bashford, T. Effect of Wavelength and Beam Width on Penetration in Light-Tissue Interaction Using Computational Methods. *Lasers Med Sci* **2017**, *32* (8), 1909. https://doi.org/10.1007/S10103-017-2317-4.

(17) Anderson, R. R.; Parrish, J. A. The Optics of Human Skin. *JID* **1981**, *77* (1), 13–19. https://doi.org/10.1111/1523-1747.EP12479191.

(18) Kwon, K.; Son, T.; Lee, K. J.; Jung, B. Enhancement of Light Propagation Depth in Skin: Cross-Validation of Mathematical Modeling Methods. *Lasers Med Sci* **2009**, *24* (4), 605–615. https://doi.org/10.1007/S10103-008-0625-4/FIGURES/7.

(19) Young, A. R. Chromophores in Human Skin. *Phys Med Biol* **1997**, *42* (5), 789. https://doi.org/10.1088/0031-9155/42/5/004.

(20) Rastogi, R. P.; Richa; Kumar, A.; Tyagi, M. B.; Sinha, R. P. Molecular Mechanisms of Ultraviolet Radiation-Induced DNA Damage and Repair. *J Nucleic Acids* **2010**, *2010*. https://doi.org/10.4061/2010/592980.

(21) Klán, P.; Wirz, J. *Photochemistry of Organic Compounds: From Concepts to Practice*; Chichester, U.K. : Wiley, 2009.

(22) Jerca, F. A.; Jerca, V. V.; Hoogenboom, R. Advances and Opportunities in the Exciting World of Azobenzenes. https://doi.org/10.1038/s41570-021-00334-w.

(23) Merino, E.; Ribagorda, M. Control over Molecular Motion Using the Cis–Trans Photoisomerization of the Azo Group. *Beilstein J. Org. Chem.* **2012**, *8*, 1071. https://doi.org/10.3762/BJOC.8.119.

(24) Wagner-Wysiecka, E.; Łukasik, N.; Biernat, J. F.; Luboch, E. Azo Group(s) in Selected Macrocyclic Compounds. *J. Incl. Phenom. Macrocycl. Chem.* **2018**, *90* (3), 189–257. https://doi.org/10.1007/S10847-017-0779-4.

(25) Chang, C. W.; Lu, Y. C.; Wang, T. Te; Diau, E. W. G. Photoisomerization Dynamics of Azobenzene in Solution with S1 Excitation: A Femtosecond Fluorescence Anisotropy Study. *J Am Chem Soc* **2004**, *126* (32), 10109–10118. https://doi.org/10.1021/JA049215P.

(26) Bandara, H. M. D.; Burdette, S. C. Photoisomerization in Different Classes of Azobenzene. *Chem Soc Rev* **2012**, *41* (5), 1809–1825. https://doi.org/10.1039/C1CS15179G.

(27) Sun, S.; Liang, S.; Xu, W. C.; Xu, G.; Wu, S. Photoresponsive Polymers with Multi-Azobenzene Groups. *Polym Chem* **2019**, *10* (32), 4389–4401. https://doi.org/10.1039/C9PY00793H.

(28) *UV-Visible Spectroscopy*. https://www2.chemistry.msu.edu/faculty/reusch/virttxtjml/spectrpy/uv-vis/spectrum.htm (accessed 2023-09-27).

(29) Bortolus, P.; Monti, S. Cis-Trans Photoisomerization of Azobenzene. Solvent and Triplet Donor Effects. *Phys. Chem.* **1979**, *83* (6), 648–652. https://doi.org/10.1021/J100469A002/ASSET/J100469A002.FP.PNG_V03.

(30) Burdette, S. C.; Bandara, S.; Dhammika Bandara, H. M. Photoisomerization in Different Classes of Azobenzene. *Chem. Soc. Rev* **2012**, *41*, 1809–1825. https://doi.org/10.1039/c1cs15179g.

(31) Bléger, D.; Schwarz, J.; Brouwer, A. M.; Hecht, S. O-Fluoroazobenzenes as Readily Synthesized Photoswitches Offering Nearly Quantitative Two-Way Isomerization with Visible Light. *J Am Chem Soc* **2012**, *134* (51), 20597–20600. https://doi.org/10.1021/ja310323y.

(32) Hartrampf, N.; Leitao, S. M.; Winter, N.; Toombs-Ruane, H.; Frank, J. A.; Schwille, P.; Trauner, D.; Franquelim, H. G. Structural Diversity of Photoswitchable Sphingolipids for Optodynamic Control of Lipid Microdomains. *Biophys J* **2023**, *122* (11), 2325–2341. https://doi.org/10.1016/J.BPJ.2023.02.029.

(33) Hetzler, B. E.; Donthamsetti, P.; Peitsinis, Z.; Stanley, C.; Trauner, D.; Isacoff, E. Y. Optical Control of Dopamine D2-like Receptors with Cell-Specific Fast-Relaxing Photoswitches. *J Am Chem Soc* **2023**. https://doi.org/10.1021/JACS.3C02735.

(34) Xiong, H.; Alberto, K. A.; Youn, J.; Taura, J.; Morstein, J.; Li, X.; Wang, Y.; Trauner, D.; Slesinger, P. A.; Nielsen, S. O.; Qin, Z. Optical Control of Neuronal Activities with Photoswitchable Nanovesicles. *Nano Res* **2023**, *16* (1), 1033–1041. https://doi.org/10.1007/S12274-022-4853-X.

(35) Pritzl, S. D.; Morstein, J.; Kahler, S.; Konrad, D. B.; Trauner, D.; Lohmüller, T. Postsynthetic Photocontrol of Giant Liposomes via Fusion-Based Photolipid Doping. *Langmuir* **2022**, *38* (39), 11941–11949. https://doi.org/10.1021/ACS.LANGMUIR.2C01685.

(36) Ko, T.; Jou, C.; Grau-Perales, A. B.; Reynders, M.; Fenton, A. A.; Trauner, D. Photoactivated Protein Degrader for Optical Control of Synaptic Function. *ACS Chem Neurosci* **2023**. https://doi.org/10.1021/ACSCHEMNEURO.3C00390.

(37) Ko, T.; Oliveira, M. M.; Alapin, J. M.; Morstein, J.; Klann, E.; Trauner, D. Optical Control of Translation with a Puromycin Photoswitch. *J Am Chem Soc* **2022**, *144* (47), 21494–21501. https://doi.org/10.1021/JACS.2C07374/SUPPL_FILE/JA2C07374_SI_001.PDB.

(38) Kiora Pharmaceuticals. *Kiora Pharmaceuticals Completes Last Patient, Last Visit in Phase 1B Trial of KIO-301 in Patients with Retinitis Pigmentosa: Kiora Pharmaceuticals, Inc. (KPRX)*. https://ir.kiorapharma.com/news-events/press-releases/detail/199/kiora-pharmaceuticals-completes-last-patient-last-visit-in (accessed 2023-10-11).

(39) Daniels, E.; Barras, C.; Dwyer, A.; Strem, B.; Wykoff, C. C.; Gelder, R. N. Van; Casson, R. J. An Intravitreal "Photoswitch" Molecule (KIO-301) for Reanimation in Retinitis Pigmentosa: A First-in-Human Trial. *Invest Ophthalmol Vis Sci* **2023**, *64* (8), 5444–5444.

(40) Kiora Pharmaceuticals. *Corporate Overview 2022-3*. https://kiorapharma.com/wp-content/uploads/2022/01/Kiora-Pharma-Corp-Presentation-Jan2022.pdf (accessed 2023-10-11).

(41) Weston, C. E.; Richardson, R. D.; Fuchter, M. J. Photoswitchable Basicity through the Use of Azoheteroarenes. *ChemComm.* **2016**, *52* (24), 4521–4524. https://doi.org/10.1039/C5CC10380K.

(42) Gibson, R. S. L.; Calbo, J.; Fuchter, M. J. Chemical Z − E Isomer Switching of Arylazopyrazoles Using Acid. *ChemPhotoChem* **2019**, *3* (6), 372–377. https://doi.org/10.1002/cptc.201900065.

(43) Calbo, J.; Thawani, A. R.; Gibson, R. S. L.; White, A. J. P.; Fuchter, M. J. A Combinatorial Approach to Improving the Performance of Azoarene Photoswitches. *Beilstein J. Org. Chem.* **2019**, *15*, 2753–2764. https://doi.org/10.3762/bjoc.15.266.

(44) Weston, C. E.; Richardson, R. D.; Haycock, P. R.; White, A. J. P.; Fuchter, M. J. Arylazopyrazoles: Azoheteroarene Photoswitches Offering Quantitative Isomerization and Long Thermal Half-Lives. *J Am Chem Soc* **2014**, *136* (34), 11878–11881. https://doi.org/10.1021/ja505444d.

(45) Gerkman, M. A.; Gibson, R. S. L.; Calbo, J.; Shi, Y.; Fuchter, M. J.; Han, G. G. D. Arylazopyrazoles for Long-Term Thermal Energy Storage and Optically Triggered Heat Release below 0 °C. *J Am Chem Soc* **2020**, *142* (19), 8688–8695. https://doi.org/10.1021/jacs.0c00374.

(46) Calbo, J.; Weston, C. E.; White, A. J. P.; Rzepa, H. S.; Contreras-García, J.; Fuchter, M. J. Tuning Azoheteroarene Photoswitch Performance through Heteroaryl Design. *J Am Chem Soc* **2017**, *139* (3), 1261–1274. https://doi.org/10.1021/jacs.6b11626.

(47) Dong, M.; Babalhavaeji, A.; Samanta, S.; Beharry, A. A.; Woolley, G. A. Red-Shifting Azobenzene Photoswitches for in Vivo Use. *Acc Chem Res* **2015**, *48* (10), 2662–2670. https://doi.org/10.1021/ACS.ACCOUNTS.5B00270/ASSET/IMAGES/LARGE/AR-2015-00270S_0011.JPEG.

(48) Lam, P. Y.; Thawani, A. R.; Balderas, E.; White, A. J. P.; Chaudhuri, D.; Fuchter, M. J.; Peterson, R. T. TRPswitch - A Step-Function Chemo-Optogenetic Ligand for the Vertebrate TRPA1 Channel. *J Am Chem Soc* **2020**, *142* (41), 17457–17468. https://doi.org/10.1021/JACS.0C06811/SUPPL_FILE/JA0C06811_SI_006.CIF.

(49) Huang, Y.; Wiedmann, M. M.; Suga, H. RNA Display Methods for the Discovery of Bioactive Macrocycles. *Chem Rev* **2018**, *119* (17), 10360–10391. https://doi.org/10.1021/acs.chemrev.8b00430.

(50) *Macrocyclic peptides: the next wave of drug discovery - MSD*. https://www.msd.com/stories/macrocyclic-peptides-the-next-wave-of-drug-discovery/# (accessed 2023-09-27).

(51) Zhang, H.; Chen, S. Cyclic Peptide Drugs Approved in the Last Two Decades (2001–2021). *RSC Chem Biol* **2022**, *3* (1), 18–31. https://doi.org/10.1039/D1CB00154J.

(52) Walport, L. J.; Obexer, R.; Suga, H. Strategies for Transitioning Macrocyclic Peptides to Cell-Permeable Drug Leads. *Curr Opin Biotechnol* **2017**, *48*, 242–250. https://doi.org/10.1016/J.COPBIO.2017.07.007.

(53) Zhang, Y.; Zhang, Q.; Wong, C. T. T.; Li, X. Chemoselective Peptide Cyclization and Bicyclization Directly on Unprotected Peptides. *J Am Chem Soc* **2019**, *141* (31), 12274–12279. https://doi.org/10.1021/JACS.9B03623/ASSET/IMAGES/LARGE/JA-2019-036238_0005.JPEG.

(54) Choi, J. S.; Joo, S. H. Recent Trends in Cyclic Peptides as Therapeutic Agents and Biochemical Tools. *Biomol Ther (Seoul)* **2020**, *28* (1), 18. https://doi.org/10.4062/BIOMOLTHER.2019.082.

(55) Abdalla, M. A.; McGaw, L. J. Natural Cyclic Peptides as an Attractive Modality for Therapeutics: A Mini Review. *Molecules 2018, Vol. 23, Page 2080* **2018**, *23* (8), 2080. https://doi.org/10.3390/MOLECULES23082080.

(56) Ngo, K. H.; Yang, R.; Das, P.; Nguyen, G. K. T.; Lim, K. W.; Tam, J. P.; Wu, B.; Phan, A. T. Cyclization of a G4-Specific Peptide Enhances Its Stability and G-Quadruplex Binding Affinity. *ChemComm.* **2020**, *56* (7), 1082–1084. https://doi.org/10.1039/C9CC06748E.

(57) Joo, S. H. Cyclic Peptides as Therapeutic Agents and Biochemical Tools. *Biomol Ther (Seoul)* **2012**, *20* (1), 19–26. https://doi.org/10.4062/BIOMOLTHER.2012.20.1.019.

(58) Van Neer, R. H. P.; Dranchak, P. K.; Liu, L.; Aitha, M.; Queme, B.; Kimura, H.; Katoh, T.; Battaile, K. P.; Lovell, S.; Inglese, J.; Suga, H. Serum-Stable and Selective Backbone-N-Methylated Cyclic Peptides That Inhibit Prokaryotic Glycolytic Mutases. *ACS Chem Biol* **2022**, *17* (8), 2284–2295. https://doi.org/10.1021/ACSCHEMBIO.2C00403/ASSET/IMAGES/LARGE/CB2C00403_0006.JPEG.

(59) Wang, W.; Khojasteh, S. C.; Su, D. Biosynthetic Strategies for Macrocyclic Peptides. *Molecules 2021, Vol. 26, Page 3338* **2021**, *26* (11), 3338. https://doi.org/10.3390/MOLECULES26113338.

(60) Costa, L.; Sousa, E.; Fernandes, C. Cyclic Peptides in Pipeline: What Future for These Great Molecules? *Pharmaceuticals 2023, Vol. 16, Page 996* **2023**, *16* (7), 996. https://doi.org/10.3390/PH16070996.

(61) Singh, N.; Abraham, J. Ribosomally Synthesized Peptides from Natural Sources. *J. Antibiot.* **2014**, *67* (4), 277–289. https://doi.org/10.1038/ja.2013.138.

(62) Geurink, P. P.; Van Der Linden, W. A.; Mirabella, A. C.; Gallastegui, N.; De Bruin, G.; Blom, A. E. M.; Voges, M. J.; Mock, E. D.; Florea, B. I.; Van Der Marel, G. A.; Driessen, C.; Van Der Stelt, M.; Groll, M.; Overkleeft, H. S.; Kisselev, A. F. Incorporation of Non-Natural Amino Acids Improves Cell Permeability and Potency of Specific Inhibitors of Proteasome Trypsin-like Sites. *J Med Chem* **2013**, *56* (3), 1262–1275. https://doi.org/10.1021/JM3016987/SUPPL_FILE/JM3016987_SI_001.PDF.

(63) Cherf, G. M.; Cochran, J. R. Applications of Yeast Surface Display for Protein Engineering. *Methods Mol Biol* **2015**, *1319*, 155. https://doi.org/10.1007/978-1-4939-2748-7_8.

(64) Tavassoli, A. SICLOPPS Cyclic Peptide Libraries in Drug Discovery. *Curr Opin Chem Biol* **2017**, *38*, 30–35. https://doi.org/10.1016/J.CBPA.2017.02.016.

(65) Parwin, S.; Kalan, S.; Srivastava, P. Bacterial Cell Surface Display. *ACS Symposium Series* **2019**, *1329*, 81–108. https://doi.org/10.1021/BK-2019-1329.CH005.

(66) Jaroszewicz, W.; Morcinek-Orłowska, J.; Pierzynowska, K.; Gaffke, L.; Węgrzyn, G. Phage Display and Other Peptide Display Technologies. *FEMS Microbiol Rev* **2022**, *46* (2), 1–25. https://doi.org/10.1093/FEMSRE/FUAB052.

(67) Oppewal, T. R.; Jansen, I. D.; Hekelaar, J.; Mayer, C. A Strategy to Select Macrocyclic Peptides Featuring Asymmetric Molecular Scaffolds as Cyclization Units by Phage Display. *J Am Chem Soc* **2022**, *144* (8), 3644–3652. https://doi.org/10.1021/JACS.1C12822/ASSET/IMAGES/LARGE/JA1C12822_0007.JPEG.

(68) PELHAM, H. R. B.; JACKSON, R. J. An Efficient MRNA-Dependent Translation System from Reticulocyte Lysates. *Eur J Biochem* **1976**, *67* (1), 247–256. https://doi.org/10.1111/J.1432-1033.1976.TB10656.X.

(69) Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. Cell-Free Translation Reconstituted with Purified Components. *Nat. Biotechnol.* **2001**, *19* (8), 751–755. https://doi.org/10.1038/90802.

(70) Kuruma, Y.; Ueda, T. The PURE System for the Cell-Free Synthesis of Membrane Proteins. *Nat. Protoc.* **2015**, *10* (9), 1328–1344. https://doi.org/10.1038/nprot.2015.082.

(71) Shimizu, Y.; Kanamori, T.; Ueda, T. Protein Synthesis by Pure Translation Systems. *Methods* **2005**, *36* (3), 299–304. https://doi.org/10.1016/J.YMETH.2005.04.006.

(72) *Codon*. https://www.genome.gov/genetics-glossary/Codon# (accessed 2023-10-02).

(73) Roberts, R. W.; Szostak, J. W. RNA-Peptide Fusions for the in Vitro Selection of Peptides and Proteins. *Proc Natl Acad Sci U S A* **1997**, *94* (23), 12297–12302. https://doi.org/10.1073/PNAS.94.23.12297/ASSET/C01FD648-0190-40DB-B741-5888DB55E87E/ASSETS/GRAPHIC/PQ2272383005.JPEG.

(74) Nemoto, N.; Miyamoto-Sato, E.; Husimi, Y.; Yanagawa, H. In Vitro Virus: Bonding of MRNA Bearing Puromycin at the 3'-Terminal End to the C-Terminal End of Its Encoded Protein on the

Ribosome in Vitro. *FEBS Lett* **1997**, *414* (2), 405–408. https://doi.org/10.1016/S0014-5793(97)01026-0.

(75) Ishizawa, T.; Kawakami, T.; Reid, P. C.; Murakami, H. TRAP Display: A High-Speed Selection Method for the Generation of Functional Polypeptides. *J Am Chem Soc* **2013**, *135* (14), 5433–5440. https://doi.org/10.1021/JA312579U.

(76) Aviner, R. The Science of Puromycin: From Studies of Ribosome Function to Applications in Biotechnology. *Comput Struct Biotechnol J* **2020**, *18*, 1074. https://doi.org/10.1016/J.CSBJ.2020.04.014.

(77) Alberts, B.; Heald, R.; Johnson, A. (Alexander D.); Morgan, D. O.; Raff, M. C.; Roberts, K. (Keith); Walter, P. (Professor). Molecular Biology of the Cell. *W.W. Norton & Company* **2022**, 661.

(78) Kolitz, S. E.; Lorsch, J. R. Eukaryotic Initiator TRNA: Finely Tuned and Ready for Action. *FEBS Lett* **2010**, *584* (2), 396. https://doi.org/10.1016/J.FEBSLET.2009.11.047.

(79) Wingfield, P. T. N-Terminal Methionine Processing. *Curr Protoc Protein Sci* **2017**, *88*, 6.14.1. https://doi.org/10.1002/CPPS.29.

(80) Chauvin, C.; Salhi, S.; Le Goff, C.; Viranaicken, W.; Diop, D.; Jean-Jean, O. Involvement of Human Release Factors ERF3a and ERF3b in Translation Termination and Regulation of the Termination Complex Formation. *Mol Cell Biol* **2005**, *25* (14), 5801–5811. https://doi.org/10.1128/mcb.25.14.5801-5811.2005.

(81) Enam, S. U.; Zinshteyn, B.; Goldman, D. H.; Cassani, M.; Livingston, N. M.; Seydoux, G.; Green, R. Puromycin Reactivity Does Not Accurately Localize Translation at the Subcellular Level. *Elife* **2020**, *9*, 1–34. https://doi.org/10.7554/ELIFE.60303.

(82) Plais, L.; Scheuermann, J. Macrocyclic DNA-Encoded Chemical Libraries: A Historical Perspective. *RSC Chem Biol* **2022**, *3* (1), 7. https://doi.org/10.1039/D1CB00161B.

(83) Suga, H. Max-Bergmann Award Lecture:A RaPID Way to Discover Bioactive Nonstandard Peptides Assisted by the Flexizyme and FIT Systems. *J. Pept. Sci.* **2018**, *24* (1). https://doi.org/10.1002/psc.3055.

(84) Ohta, A.; Murakami, H.; Higashimura, E.; Suga, H. Synthesis of Polyester by Means of Genetic Code Reprogramming. *Chem Biol* **2007**, *14* (12), 1315–1322. https://doi.org/10.1016/J.CHEMBIOL.2007.10.015.

(85) Goto, Y.; Murakami, H.; Suga, H. Initiating Translation with D-Amino Acids. *RNA* **2008**, *14* (7), 1390–1398. https://doi.org/10.1261/RNA.1020708.

(86) Goto, Y.; Ohta, A.; Sako, Y.; Yamagishi, Y.; Murakami, H.; Suga, H. Reprogramming the Translation Initiation for the Synthesis of Physiologically Stable Cyclic Peptides. *ACS Chem Biol* **2008**, *3* (2), 120–129. https://doi.org/10.1021/CB700233T/SUPPL_FILE/CB700233T-FILE009.PDF.

(87) Kawakami, T.; Murakami, H.; Suga, H. Messenger RNA-Programmed Incorporation of Multiple N-Methyl-Amino Acids into Linear and Cyclic Peptides. *Chem Biol* **2008**, *15* (1), 32–42. https://doi.org/10.1016/J.CHEMBIOL.2007.12.008.

(88) Kawakami, T.; Murakami, H.; Suga, H. Ribosomal Synthesis of Polypeptoids and Peptoid-Peptide Hybrids. *J Am Chem Soc* **2008**, *130* (50), 16861–16863. https://doi.org/10.1021/JA806998V/SUPPL_FILE/JA806998V_SI_001.PDF.

(89) Goto, Y.; Suga, H. Translation Initiation with Initiator TRNA Charged with Exotic Peptides. *J Am Chem Soc* **2009**, *131* (14), 5040–5041. https://doi.org/10.1021/JA900597D/SUPPL_FILE/JA900597D_SI_001.PDF.

(90) Murakami, H.; Ohta, A.; Ashigai, H.; Suga, H. A Highly Flexible TRNA Acylation Method for Non-Natural Polypeptide Synthesis. *Nat Methods* **2006**, *3* (5), 357–359. https://doi.org/10.1038/nmeth877.

(91) Goto, Y.; Katoh, T.; Suga, H. Flexizymes for Genetic Code Reprogramming. *Nat Protoc* **2011**, *6* (6), 779–790. https://doi.org/10.1038/nprot.2011.331.

(92) Kawakami, T.; Ohta, A.; Ohuchi, M.; Ashigai, H.; Murakami, H.; Suga, H. Diverse Backbone-Cyclized Peptides via Codon Reprogramming. *Nat. Chem. Biol.* **2009**, *5* (12), 888–890. https://doi.org/10.1038/nchembio.259.

(93) Saito, H.; Watanabe, K.; Suga, H. Concurrent Molecular Recognition of the Amino Acid and TRNA by a Ribozyme. **2001**.

(94) Murakami, H.; Saito, H.; Suga, H. A Versatile TRNA Aminoacylation Catalyst Based on RNA U32-U35 and U40-U41 Consist of the Phe Binding Site. *Chem Biol* **2003**, *10*, 655–662. https://doi.org/10.1016/S.

(95) Niwa, N.; Yamagishi, Y.; Murakami, H.; Suga, H. A Flexizyme That Selectively Charges Amino Acids Activated by a Water-Friendly Leaving Group. *Bioorg Med Chem Lett* **2009**, *19* (14), 3892–3894. https://doi.org/10.1016/j.bmcl.2009.03.114.

(96) Lee, J.; Schwieter, K. E.; Watkins, A. M.; Kim, D. S.; Yu, H.; Schwarz, K. J.; Lim, J.; Coronado, J.; Byrom, M.; Anslyn, E. V; Ellington, A. D.; Moore, J. S.; Jewett, M. C. Expanding the Limits of the Second Genetic Code with Ribozymes. *Nat comm.* **2019**, *10* (1). https://doi.org/10.1038/s41467-019-12916-w.

(97) Iwane, Y.; Kimura, H.; Katoh, T.; Suga, H. Uniform Affinity-Tuning of N-Methyl-Aminoacyl-TRNAs to EF-Tu Enhances Their Multiple Incorporation. *Nucleic Acids Res* **2021**, *49* (19), 10807–10817. https://doi.org/10.1093/NAR/GKAB288.

(98) Kolitz, S. E.; Lorsch, J. R. Eukaryotic Initiator TRNA: Finely Tuned and Ready for Action. *FEBS Lett* **2010**, *584* (2), 396. https://doi.org/10.1016/J.FEBSLET.2009.11.047.

(99) Passioura, T.; Suga, H. A RaPID Way to Discover Nonstandard Macrocyclic Peptide Modulators of Drug Targets. *ChemComm.* **2017**, *53* (12), 1931–1940. https://doi.org/10.1039/C6CC06951G.

(100) Mukai, T.; Lajoie, M. J.; Englert, M.; Oll, D. S. ¨. Rewriting the Genetic Code. *Annual Review of Microbiology Annu. Rev. Microbiol* **2017**, *12* (4), 557–577. https://doi.org/10.1146/annurev-micro-090816.

(101) Wang aY, L.; Schultz bY, P. G. A General Approach for the Generation of Orthogonal TRNAs. *Chem Biol* **2001**, *8* (9), 883–890. https://doi.org/https://doi.org/10.1016/s1074-5521(01)00063-1.

(102) Iskandar, S. E.; Pelton, J. M.; Wick, E. T.; Bolhuis, D. L.; Baldwin, A. S.; Emanuele, M. J.; Brown, N. G.; Bowers, A. A. Enabling Genetic Code Expansion and Peptide Macrocyclization in MRNA Display via a Promiscuous Orthogonal Aminoacyl-TRNA Synthetase. *J Am Chem Soc* **2023**, *145* (3), 1512–1517. https://doi.org/10.1021/JACS.2C11294/ASSET/IMAGES/LARGE/JA2C11294_0003.JPEG.

(103) Yamagishi, Y.; Shoji, I.; Miyagawa, S.; Kawakami, T.; Katoh, T.; Goto, Y.; Suga, H. Natural Product-Like Macrocyclic N-Methyl-Peptide Inhibitors against a Ubiquitin Ligase Uncovered from a Ribosome-Expressed De Novo Library. *Chem Biol* **2011**, *18* (12), 1562–1570. https://doi.org/10.1016/J.CHEMBIOL.2011.09.013.

(104) Sako, Y.; Goto, Y.; Murakami, H.; Suga, H. Ribosomal Synthesis of Peptidase-Resistant Peptides Closed by a Nonreducible Inter-Side-Chain Bond. *ACS Chem Biol* **2008**, *3* (4), 241–249. https://doi.org/10.1021/CB800010P/SUPPL_FILE/CB800010P-FILE007.PDF.

(105) Torikai, K.; Suga, H. Ribosomal Synthesis of an Amphotericin-B Inspired Macrocycle. *J Am Chem Soc* **2014**, *136* (50), 17359–17361. https://doi.org/10.1021/JA508648S/SUPPL_FILE/JA508648S_SI_001.PDF.

(106) Ohshiro, Y.; Nakajima, E.; Goto, Y.; Fuse, S.; Takahashi, T.; Doi, T.; Suga, H. Ribosomal Synthesis of Backbone-Macrocyclic Peptides Containing γ-Amino Acids. *Chembiochem* **2011**, *12* (8), 1183–1187. https://doi.org/10.1002/CBIC.201100104.

(107) Rogers, J. M.; Kwon, S.; Dawson, S. J.; Mandal, P. K.; Suga, H.; Huc, I. Ribosomal Synthesis and Folding of Peptide-Helical Aromatic Foldamer Hybrids. *Nat Chem* **2018**, *10* (4), 405–412. https://doi.org/10.1038/s41557-018-0007-x.

(108) Wu, Y.; Bertran, M. T.; Joshi, D.; Maslen, S. L.; Hurd, C.; Walport, L. J. Identification of Photocrosslinking Peptide Ligands by MRNA Display. *Commun. Chem.* **2023**, *6* (1), 1–9. https://doi.org/10.1038/s42004-023-00898-2.

(109) Fujino, T.; Goto, Y.; Suga, H.; Murakami, H. Reevaluation of the D-Amino Acid Compatibility with the Elongation Event in Translation. *J Am Chem Soc* **2013**, *135* (5), 1830–1837. https://doi.org/10.1021/JA309570X/SUPPL_FILE/JA309570X_SI_001.PDF.

(110) Fujino, T.; Goto, Y.; Suga, H.; Murakami, H. Ribosomal Synthesis of Peptides with Multiple β-Amino Acids. *J Am Chem Soc* **2016**, *138* (6), 1962–1969. https://doi.org/10.1021/JACS.5B12482/ASSET/IMAGES/LARGE/JA-2015-12482E_0005.JPEG.

(111) Maini, R.; Kimura, H.; Takatsuji, R.; Katoh, T.; Goto, Y.; Suga, H. Ribosomal Formation of Thioamide Bonds in Polypeptide Synthesis. *J Am Chem Soc* **2019**, *141* (51), 20004–20008. https://doi.org/10.1021/JACS.9B11097/ASSET/IMAGES/LARGE/JA9B11097_0005.JPEG.

(112) Goto, Y.; Suga, H. The RaPID Platform for the Discovery of Pseudo-Natural Macrocyclic Peptides. *Acc Chem Res* **2021**, *54* (18), 3604–3617. https://doi.org/10.1021/ACS.ACCOUNTS.1C00391.

(113) Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z. G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth Biol* **2013**, *2* (2), 83–92. https://doi.org/10.1021/SB300037W/SUPPL_FILE/SB300037W_SI_001.PDF.

(114) Velema, W. A.; Szymanski, W.; Feringa, B. L. Photopharmacology: Beyond Proof of Principle. *J Am Chem Soc* **2014**, *136* (6), 2178–2191. https://doi.org/10.1021/JA413063E.

(115) Schoenberger, M.; Damijonaitis, A.; Zhang, Z.; Nagel, D.; Trauner, D. Development of a New Photochromic Ion Channel Blocker via Azologization of Fomocaine. *ACS Chem Neurosci* **2014**, *5* (7), 514–518. https://doi.org/10.1021/CN500070W/SUPPL_FILE/CN500070W_SI_001.PDF.

(116) Broichhagen, J.; Frank, J. A.; Trauner, D. A Roadmap to Success in Photopharmacology. *Acc Chem Res* **2015**, *48* (7), 1947–1960. https://doi.org/10.1021/ACS.ACCOUNTS.5B00129.

(117) Velema, W. A.; Van Der Berg, J. P.; Hansen, M. J.; Szymanski, W.; Driessen, A. J. M.; Feringa, B. L. Optical Control of Antibacterial Activity. *Nat. Chem.* **2013**, *5* (11), 924–928. https://doi.org/10.1038/nchem.1750.

(118) Stein, M.; Middendorp, S. J.; Carta, V.; Pejo, E.; Raines, D. E.; Forman, S. A.; Sigel, E.; Trauner, D. Azo-Propofols: Photochromic Potentiators of GABAA Receptors. *Angew. Chem., Int. Ed. Engl.* **2012**, *51* (42), 10500–10504. https://doi.org/10.1002/ANIE.201205475.

(119) Kobauri, P.; Galenkamp, N. S.; Schulte, A. M.; De Vries, J.; Simeth, N. A.; Maglia, G.; Thallmair, S.; Kolarski, D.; Szymanski, W.; Feringa, B. L. Hypothesis-Driven, Structure-Based Design in Photopharmacology: The Case of EDHFR Inhibitors. *J Med Chem* **2022**, *65* (6), 4798–4817. https://doi.org/10.1021/ACS.JMEDCHEM.1C01962/SUPPL_FILE/JM1C01962_SI_003.CSV.

(120) Liu, M.; Tada, S.; Ito, M.; Abe, H.; Ito, Y. In Vitro Selection of a Photo-Responsive Peptide Aptamer Using Ribosome Display. *ChemComm.* **2012**, *48* (97), 11871. https://doi.org/10.1039/c2cc36618e.

(121) Jafari, M. R.; Deng, L.; Kitov, P. I.; Ng, S.; Matochko, W. L.; Tjhung, K. F.; Zeberoff, A.; Elias, A.; Klassen, J. S.; Derda, R. Discovery of Light-Responsive Ligands through Screening of a Light-Responsive Genetically Encoded Library. *ACS Chem Biol* **2014**, *9* (2), 443–450. https://doi.org/10.1021/cb4006722.

(122) Bellotto, S.; Chen, S.; Rentero Rebollo, I.; Wegner, H. A.; Heinis, C. Phage Selection of Photoswitchable Peptide Ligands. *J Am Chem Soc* **2014**, *136* (16), 5880–5883. https://doi.org/10.1021/ja501861m.

(123) Jafari, M. R.; Yu, H.; Wickware, J. M.; Lin, Y. S.; Derda, R. Light-Responsive Bicyclic Peptides. *Org Biomol Chem* **2018**, *16* (41), 7588–7594. https://doi.org/10.1039/c7ob03178e.

(124) Janosko, C. P.; Shade, O.; Courtney, T. M.; Horst, T. J.; Liu, M.; Khare, S. D.; Deiters, A. Genetic Encoding of Arylazopyrazole Phenylalanine for Optical Control of Translation. *ACS Omega* **2023**, *8* (29), 26590–26596. https://doi.org/10.1021/ACSOMEGA.3C03512/ASSET/IMAGES/LARGE/AO3C03512_0007.JPEG.

(125) Garcia-Amorós, J.; Díaz-Lobo, M.; Nonell, S.; Velasco, D. Fastest Thermal Isomerization of an Azobenzene for Nanosecond Photoswitching Applications under Physiological Conditions. *Angew. Chem., Int. Ed. Engl.* **2012**, *51* (51), 12820–12823. https://doi.org/10.1002/ANIE.201207602.

(126) Katoh, T.; Suga, H. Flexizyme-Catalyzed Synthesis of 3-Aminoacyl-NH-TRNAs. *Nucleic Acids Res* **2019**, *47* (9). https://doi.org/10.1093/nar/gkz143.

(127) Murakami, H.; Ohta, A.; Ashigai, H.; Suga, H. A Highly Flexible TRNA Acylation Method for Non-Natural Polypeptide Synthesis. *Nature Methods 2006 3:5* **2006**, *3* (5), 357–359. https://doi.org/10.1038/nmeth877.

(128) Ohuchi, M.; Murakami, H.; Suga, H. The Flexizyme System: A Highly Flexible TRNA Aminoacylation Tool for the Translation Apparatus. *Curr Opin Chem Biol* **2007**, *11*, 537–542. https://doi.org/10.1016/j.cbpa.2007.08.011.

(129) Katoh, T.; Suga, H. Drop-off-Reinitiation at the Amino Termini of Nascent Peptides and Its Regulation by IF3, EF-G, and RRF. *RNA* **2023**, *29* (5), 663–674. https://doi.org/10.1261/RNA.079447.122.

(130) Tajima, K.; Katoh, T.; Suga, H. Drop-off-Reinitiation Triggered by EF-G-Driven Mistranslocation and Its Alleviation by EF-P. *Nucleic Acids Res* **2022**, *50* (5), 2736–2753. https://doi.org/10.1093/NAR/GKAC068.

(131) Siewertsen, R.; Neumann, H.; Buchheim-Stehn, B.; Herges, R.; Näther, C.; Renth, F.; Temps, F. Highly Efficient Reversible Z-E Photoisomerization of a Bridged Azobenzene with Visible Light through Resolved S1(Nπ*) Absorption Bands. *J Am Chem Soc* **2009**, *131* (43), 15594–15595. https://doi.org/10.1021/JA906547D/SUPPL_FILE/JA906547D_SI_002.PDF.

(132) *PADI2 - Protein-arginine deiminase type-2 - Homo sapiens (Human) | UniProtKB | UniProt*. https://www.uniprot.org/uniprotkb/Q9Y2J8/entry (accessed 2023-10-10).

(133) Slade, D. J.; Fang, P.; Dreyton, C. J.; Zhang, Y.; Fuhrmann, J.; Rempel, D.; Bax, B. D.; Coonrod, S. A.; Lewis, H. D.; Guo, M.; Gross, M. L.; Thompson, P. R. Protein Arginine Deiminase 2 Binds Calcium in an Ordered Fashion: Implications for Inhibitor Design. *ACS Chem Biol* **2015**, *10* (4), 1043–1053. https://doi.org/10.1021/CB500933J/SUPPL_FILE/CB500933J_SI_001.PDF.

(134) Cicek, E.; Monard, G.; Sungur, F. A. Molecular Mechanism of Protein Arginine Deiminase 2: A Study Involving Multiple Microsecond Long Molecular Dynamics Simulations. *Biochemistry* **2022**, *61* (13), 1286–1297. https://doi.org/10.1021/ACS.BIOCHEM.2C00158/SUPPL_FILE/BI2C00158_SI_002.ZIP.

(135) *PADI2 peptidyl arginine deiminase 2 [Homo sapiens (human)] - Gene - NCBI*. https://www.ncbi.nlm.nih.gov/gene/11240#gene-expression (accessed 2023-10-11).

(136) Barasa, L.; Thompson, P. R. Protein Citrullination: Inhibition, Identification and Insertion. *Philos Trans R Soc Lond B Biol Sci* **2023**, *378* (1890), 20220240. https://doi.org/10.1098/RSTB.2022.0240.

(137) Ishigami, A.; Ohsawa, T.; Hiratsuka, M.; Taguchi, H.; Kobayashi, S.; Saito, Y.; Murayama, S.; Asaga, H.; Toda, T.; Kimura, N.; Maruyama, N. Abnormal Accumulation of Citrullinated Proteins Catalyzed by Peptidylarginine Deiminase in Hippocampal Extracts from Patients with Alzheimer's Disease. *J Neurosci Res* **2005**, *80* (1), 120–128. https://doi.org/10.1002/JNR.20431.

(138) Chang, X.; Xia, Y.; Pan, J.; Meng, Q.; Zhao, Y.; Yan, X. PADI2 Is Significantly Associated with Rheumatoid Arthritis. *PLoS One* **2013**, *8* (12), e81259. https://doi.org/10.1371/JOURNAL.PONE.0081259.

(139) McElwee, J. L.; Mohanan, S.; Griffith, O. L.; Breuer, H. C.; Anguish, L. J.; Cherrington, B. D.; Palmer, A. M.; Howe, L. R.; Subramanian, V.; Causey, C. P.; Thompson, P. R.; Gray, J. W.; Coonrod, S. A. Identification of PADI2 as a Potential Breast Cancer Biomarker and Therapeutic Target. *BMC Cancer* **2012**, *12* (1), 1–17. https://doi.org/10.1186/1471-2407-12-500/FIGURES/7.

(140) Beato, M.; Sharma, P. Peptidyl Arginine Deiminase 2 (PADI2)-Mediated Arginine Citrullination Modulates Transcription in Cancer. *Int. J. Mol. Sci.* **2020**, *21* (4), 1351. https://doi.org/10.3390/IJMS21041351.

(141) Musse, A. A.; Zhen, L.; Ackerley, C. A.; Bienzle, D.; Lei, H.; Poma, R.; Harauz, G.; Moscarello, M. A.; Mastronardi, F. G. Peptidylarginine Deiminase 2 (PAD2) Overexpression in Transgenic Mice Leads to Myelin Loss in the Central Nervous System. *DMM* **2008**, *1* (4–5), 229–240. https://doi.org/10.1242/DMM.000729/265960/AM/PEPTIDYLARGININE-DEIMINASE-2-PAD2-OVEREXPRESSION.

(142) Muth, A.; Subramanian, V.; Beaumont, E.; Nagar, M.; Kerry, P.; McEwan, P.; Srinath, H.; Clancy, K.; Parelkar, S.; Thompson, P. R. Development of a Selective Inhibitor of Protein Arginine Deiminase 2. *J Med Chem* **2017**, *60* (7), 3198–3211. https://doi.org/10.1021/ACS.JMEDCHEM.7B00274/SUPPL_FILE/JM7B00274_SI_002.CSV.

(143) Zhang, X.; Shen, M.; Zhu, H.; Zhang, J.; Yang, M.; Su, K.; Zhang, Y.; Fu, W.; Ke, X.; Qu, Y. Small Molecule Activates Citrullination through Targeting PAD2. *Philos. Trans. R. Soc. B* **2023**, *378* (1890), 20220248. https://doi.org/10.1098/RSTB.2022.0248.

(144) Mondal, S.; Parelkar, S. S.; Nagar, M.; Thompson, P. R. Photochemical Control of Protein Arginine Deiminase (PAD) Activity. *ACS Chem Biol* **2018**, *13* (4), 1057–1065. https://doi.org/10.1021/ACSCHEMBIO.8B00053/ASSET/IMAGES/MEDIUM/CB-2018-000533_M003.GIF.

(145) Knipp, M.; Vašá, M. A Colorimetric 96-Well Microtiter Plate Assay for the Determination of Enzymatically Formed Citrulline 1. **2000**. https://doi.org/10.1006/abio.2000.4805.

(146) Clancy, K. W.; Weerapana, E.; Thompson, P. R. Detection and Identification of Protein Citrullination in Complex Biological Systems. *Curr Opin Chem Biol* **2016**, *30*, 1–6. https://doi.org/10.1016/J.CBPA.2015.10.014.

(147) Knight, J. S.; Subramanian, V.; O'dell, A. A.; Yalavarthi, S.; Zhao, W.; Smith, C. K.; Hodgin, J. B.; Thompson, P. R.; Kaplan, M. J.; Branch, S. A. Peptidylarginine Deiminase Inhibition Disrupts NET Formation and Protects against Kidney, Skin and Vascular Disease in Lupus-Prone MRL/Lpr Mice. *ARJ* **2015**, *74* (12), 2199–2206. https://doi.org/https://doi.org/10.1136/annrheumdis-2014-205365.

(148) Fairhead, M.; Howarth, M. Site-Specific Biotinylation of Purified Proteins Using BirA. *Methods Mol. Biol.* **2015**, *1266*, 171–184. https://doi.org/10.1007/978-1-4939-2272-7_12/COVER.

(149) Olson, C. A.; Liao, H. I.; Sun, R.; Roberts, R. W. MRNA Display Selection of a High-Affinity, Modification-Specific Phospho-Iκbα-Binding Fibronectin. *ACS Chem Biol* **2008**, *3* (8), 480–485. https://doi.org/10.1021/CB800069C/SUPPL_FILE/CB800069C-FILE006.PDF.

(150) Wranik, M.; Weinert, T.; Slavov, C.; Masini, T.; Furrer, A.; Gaillard, N.; Gioia, D.; Ferrarotti, M.; James, D.; Glover, H.; Carrillo, M.; Kekilli, D.; Stipp, R.; Skopintsev, P.; Brünle, S.; Mühlethaler, T.; Beale, J.; Gashi, D.; Nass, K.; Ozerov, D.; Johnson, P. J. M.; Cirelli, C.; Bacellar, C.; Braun, M.; Wang, M.; Dworkowski, F.; Milne, C.; Cavalli, A.; Wachtveitl, J.; Steinmetz, M. O.; Standfuss, J. Watching the Release of a Photopharmacological Drug from Tubulin Using Time-Resolved Serial Crystallography. *Nat. Commun.* **2023**, *14* (1), 1–12. https://doi.org/10.1038/s41467-023-36481-5.

(151) Anthis, N. J.; Clore, G. M. Sequence-Specific Determination of Protein and Peptide Concentrations by Absorbance at 205 Nm. *Protein Science* **2013**, *22* (6), 851–858. https://doi.org/10.1002/PRO.2253.

(152) Mourot, A.; Kienzler, M. A.; Banghart, M. R.; Fehrentz, T.; Huber, F. M. E.; Stein, M.; Kramer, R. H.; Trauner, D. Tuning Photochromic Ion Channel Blockers. *ACS Chem Neurosci* **2011**, *2* (9), 536–543. https://doi.org/10.1021/CN200037P/SUPPL_FILE/CN200037P_SI_001.PDF.

(153) Beharry, A. A.; Sadovski, O.; Woolley, G. A. Azobenzene Photoswitching without Ultraviolet Light. *J Am Chem Soc* **2011**, *133* (49), 19684–19687. https://doi.org/10.1021/JA209239M/SUPPL_FILE/JA209239M_SI_001.PDF.

(154) Verhoeven, J. W. Glossary of Terms Used in Photochemistry (IUPAC Reconunendations 1996). *Pure and Applied Chemistry* **1996**, *68* (12), 2223–2286. https://doi.org/10.1351/PAC199668122223.

(155) Berthet, J.; Agouridas, L.; Chen, S.; Allouchi, H.; Melnyk, P.; Champagne, B.; Delbaere, S. Synthesis and Switching Properties of New Derivatives of Azoresveratrol. *Dyes Prigm.* **2019**, *171*, 10766. https://doi.org/10.1016/j.dyepig.2019.107666.

(156) Ludwanowski, S.; Ari, M.; Parison, K.; Kalthoum, S.; Straub, P.; Pompe, N.; Weber, S.; Walter, M.; Walther, A. PH Tuning of Water-Soluble Arylazopyrazole Photoswitches. *Eur. J. Res.* **2020**, *26* (58), 13203–13212. https://doi.org/10.1002/CHEM.202000659.

(157) Sakamoto, K. M.; Kim, K. B.; Kumagai, A.; Mercurio, F.; Crews, C. M.; Deshaies, R. J. Protacs: Chimeric Molecules That Target Proteins to the Skp1-Cullin-F Box Complex for Ubiquitination and Degradation. *Proc Natl Acad Sci U S A* **2001**, *98* (15), 8554–8559. https://doi.org/10.1073/PNAS.141230798.

(158) Sakamoto, K. M.; Kim, K. B.; Verma, R.; Ransick, A.; Stein, B.; Crews, C. M.; Deshaies, R. J. Development of Protacs to Target Cancer-Promoting Proteins for Ubiquitination and Degradation. *Mol Cell Proteomics* **2003**, *2* (12), 1350–1358. https://doi.org/10.1074/MCP.T300009-MCP200.

(159) Sun, X.; Gao, H.; Yang, Y.; He, M.; Wu, Y.; Song, Y.; Tong, Y.; Rao, Y. PROTACs: Great Opportunities for Academia and Industry. *Signal Transduct. Target Ther.* **2019**, *4* (1), 1–33. https://doi.org/10.1038/s41392-019-0101-6.

(160) Zhang, Q.; Kounde, C. S.; Mondal, M.; Greenfield, J. L.; Baker, J. R.; Kotelnikov, S.; Ignatov, M.; Tinworth, C. P.; Zhang, L.; Conole, D.; De Vita, E.; Kozakov, D.; McCluskey, A.; Harling, J. D.; Fuchter, M. J.; Tate, E. W. Light-Mediated Multi-Target Protein Degradation Using Arylazopyrazole Photoswitchable PROTACs (AP-PROTACs). *ChemComm.* **2022**, *58* (78), 10933–10936. https://doi.org/10.1039/D2CC03092F.

(161) Mcallister, T. E.; Yeh, T.-L.; Abboud, M. I.; Leung, I. K. H.; Hookway, E. S.; King, O. N. F.; Bhushan, B.; Williams, S. T.; Hopkinson, R. J.; Loik, N. D.; Chowdhury, R.; Oppermann, U.; Claridge, T. D. W.; Goto, Y.; Suga, H.; Schofield, C. J.; Kawamura, A. Showcasing Research from Dr Akane Kawamura's Laboratory Non-Competitive Cyclic Peptides for Targeting Enzyme-Substrate Complexes †. *Chem. Sci* **2018**, 4569. https://doi.org/10.1039/c8sc00286j.

(162) Chowdhury, R.; Abboud, M. I.; Mcallister, T. E.; Banerji, B.; Bhushan, B.; Sorensen, J. L.; Kawamura, A.; Schofield, C. J. Use of Cyclic Peptides to Induce Crystallization: Case Study with Prolyl Hydroxylase Domain 2. *Sci. Rep.* **2020**, *10* (2020). https://doi.org/10.1038/s41598-020-76307-8.

(163) Johansen-Leete, J.; Passioura, T.; Passioura, T.; Passioura, T.; Passioura, T.; Foster, S. R.; Bhusal, R. P.; Ford, D. J.; Liu, M.; Jongkees, S. A. K.; Suga, H.; Stone, M. J.; Payne, R. J. Discovery of Potent Cyclic Sulfopeptide Chemokine Inhibitors via Reprogrammed Genetic Code MRNA Display. *J Am Chem Soc* **2020**, *142* (20), 9141–9146. https://doi.org/10.1021/JACS.0C03152/ASSET/IMAGES/LARGE/JA0C03152_0003.JPEG.

(164) Jongkees, S. A. K.; Caner, S.; Tysoe, C.; Brayer, G. D.; Withers, S. G.; Suga, H. Rapid Discovery of Potent and Selective Glycosidase-Inhibiting De Novo Peptides. *Cell Chem Biol* **2017**, *24* (3), 381–390. https://doi.org/10.1016/j.chembiol.2017.02.001.

(165) Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; Tinevez, J. Y.; White, D. J.; Hartenstein, V.; Eliceiri, K.; Tomancak, P.; Cardona, A. Fiji: An Open-Source Platform for Biological-Image Analysis. *Nat. Methods* **2012**, *9* (7), 676–682. https://doi.org/10.1038/nmeth.2019.

7 EXPERIMENTAL

7.1 Buffer compositions:

PADI2 selection buffer (pH 7.5 NaOH)	50 mM HEPES (pH 7.5) (Merck)
	2 mM DTT (Fluorochem)
	10 mM $CaCl_2$ (Merck)
	150 mM NaCl (Merck)
	0.1% Tween-20 (Thermo Bioreagents)

2x Blocking solution (pH 7.5 NaOH)	PADI selection buffer
	0.2% (w/v) BSA, acetylated (Invitrogen)
	2 mg/mL UltraPure™ Salmon Sperm DNA (Invitrogen)

50x TEA buffer (pH 8.3)	50 mM $Na_2EDTA.2H_2O$ (Merck)
	2 M TRIS base (Merck)
	1 M (Glacial) acetic acid (Merck)

5x TBE buffer (pH. 8.7)	5.4 % Tris BASE (w/v)
	2.8 % Boric acid (w/v)
	10 mM EDTA

DNase buffer (pH 7.6) (Store at 20 °C)	100 mM Tris-HCl (pH 7.5)
	25 mM $MgCl_2$
	5 mM $CaCl_2.6H_2O$ (only use hexahydrate (w/v))

10x T7 buffer (pH 8.0 HCl)	400 mM Tris-base (Merck)
	10 mM Spermidine (Merck)
	0.1% Triton X-100 (v/v) (Thermo Bioreagents)

1x KOD buffer (pH 8.0 HCl)	120 mM Tris-Base (Merck)
	10 mM KCl (Merck)
	6 mM $(NH_4)_2SO_4$ (Merck)
	0.1% Triton X-100 (Thermo Bioreagents)
	0.001% BSA (w/v) pH 8.0 (Merck)

2x DNA loading buffer (pH 7.5)	2X TAE
	30% Glycerol (Merck)
	0.05% Bromophenol Blue (w/v) (Merck)

2x RNA loading buffer (pH 7.5)	8 M Urea (Merck)
	2 mM $Na_2EDTA.2H_2O$ (Merck)
	2 mM TRIS base (Merck)
	2% BPB (w/v) (Merck)

1x Aminoacylation Acid PAGE loading buffer	150 mM NaOAc (Ph 5.2)
	10 mM EDTA
	93% Formamide
	0.02% BRB

PADI2 SPR buffer (pH 7.5 NaOH)	50 mM HEPES (pH. 7.5 using NaOH) (Merck)
	150 mM NaCl (Merck)
	10 mM $CaCl_2$ (Merck)
	2 mM DTT (Merck)
	0.02% Tween-200 (Thermo Bioreagents)
	0.1 % DMSO (Merck)

PADI Purification Buffers	
PADI2 His lysis/ binding buffer	50 mM HEPES pH 7.5
	500 mM NaCl
	20 mM Imidazole
	1 mM TCEP
	5% Glycerol

PADI2 Elution buffer	50 mM HEPES pH 7.5
	500 mM NaCl
	250 mM Imidazole
	1 mM TCEP
	5% Glycerol

PADI Gel filtration buffer	50 mM HEPES pH 7.5
	200 mM NaCl
	1 mM TCEP
	5% Glycerol

7.2 Reaction Mix compositions:

	Reagent:	Volume (µL):
Transcription Mix	10x T7 buffer	4
	100 mM DTT (Merck)	4
	250 mM $MgCl_2$ (Merck)	4.8
	25 mM NTPs (Jena Bioscience)	8
	T7 RNA polymerase (NEB)	1.2
	Water	10
	DNA template	8
	Final volume	40

	Reagent:	Volume (µL):
T4 RNA ligase Mix	10x T4 RNA ligase buffer (NEB)	6
	10 mM ATP	6
	Water	27
	7.5 µM DNA-PEG-Puro Linker	12
	T4 RNA Ligase (NEB)	3
	10 µM mRNA library	6
	Final volume	60

Reverse Transcription Mix (For making DNA standards) (cDNA)	Reagent:	Volume (µL):
	H_2O	25.2
	5×RT buffer (M-MLV RT 5× (Promega)	8

	10 mM dNTPs (NEB)	4
	100 µM Reverse Primer (f.c. 2 µM) (IDT)	0.8
	10 µM mRNA library/seq. (f.c. 0.2 µM)	0.8
	MLV RTase (H+) (Promega)	1
	RNAse inhibitor (Promega)	0.1
	Final volume	40

Reverse Transcription Mix (mRNA-DNA hybrids)	Reagent:	Volume (µL):
	RT Premix:	
	100 µM CGS13.r39 primer	6.3
	10 mM dNTPs (NEB)	9.75
	H_2O	9.75
	250 mMTris-HCl pH 8.3	37.8
	Final volume	62.55
	Reaction:	
	Translation Product	165
	RT Premix	46.2
	$Mg(OAc)_2$ (Merck)	27.5
	KOH (Merck)	13.75
	RTase H(-) (Promega)	6.88
	Final volume	260

1x Taq PCR Mix	Reagent:	Volume (µL):
	10x Taq PCR buffer B (KAPA Bioscience)	100
	25 mM $MgCl_2$ (Merck)	40
	10 mM dNTPs (NEB)	25
	100 µM T7g10M.F46 (IDT)	2.5
	100 µM CGS3an13.R22 (IDT)	2.5
	Water	830

1x Taq Real-time PCR Mix	Reagent:	Volume (µL):
	1x Taq PCR Mix	1000
	SYBR green I (Thermo Fisher Scientific)	10
	Taq DNA polymerase (KAPA Bioscience)	13.33

Next Generation Sequencing (NGS) Mixes		
NGS PCR 1 Mix	Reagent:	Volume (µL):
	Water	30
	10× KOD buffer	5
	10 mM each dNTPs (NEB)	1
	1× KOD DNA polymerase	0.4
	25 mM $MgSO_4$ (Merck)	3
	100 µM Rd1T7g10M.F67 (IDT)	0.2
	100 µM an13Rd2.R49 (IDT)	0.25
	DNA library sequence/ round	10
	Final volume	50

NGS PCR 2 Mix	Reagent:	Volume (µL):
	Water	70.2
	10× KOD buffer	10
	10 mM each dNTPs (NEB)	2

	1× KOD DNA polymerase	0.8
	25 mM $MgSO_4$ (Merck)	6
	100 µM P5S5**Rd1.F57 (IDT)	0.5
	100 µM Rd2N7**P7.R52 (IDT)	0.5
	DNA library sequence/ round	10
	Final volume	100

Flexizymes:		
Flexizyme PCR master Mix	Reagent:	Volume (µL):
	Water	957
	10x KOD buffer	115
	25 mM $MgCl_2$ (Merck)	46
	10 mM dNTPs (NEB)	23
	KOD DNA polymerase	9.2
	Final volume	1150.2

Flexizyme PCR extension mix	Reagent:	Volume (µL):
	Flexizyme PCR Master mix	99
	200 µM Fx-5'.F36	0.5
	200 µM eFx.R45 or dFx.R46	0.5
	Final volume	100

Flexizyme PCR reaction mix	Reagent:	Volume (µL):
	Flexizyme PCR Master mix	1000
	Extension reaction mix	5
	200 µM T7ex5.F22	2.5
	200 µM eFx.R18 or dFx.R19	2.5
	Final volume	1010

Flexizyme transcription mix	Reagent:	Volume (µL):
	H_2O	Up to 1000
	10x T7 buffer	100
	100 mM DTT (Merck)	100
	250 mM $MgCl_2$ (Merck)	120
	25 mM NTPs (Jena Bioscience)	200
	DNA template (PCR product)	100
	T7 RNA polymerase (Thermo Fisher Scientific)	30 (~ 0.1x the DNA conc.)
	Final volume	1000

tRNA		
tRNA PCR master mix	Reagent:	Volume (µL):
	MQ	940
	10x KOD buffer	125
	25 mM $MgCl_2$ (Merck)	50
	10 mM each dNTPs (NEB)	125
	KOD polymerase	10
	Total	1250

	Reagent:	Volume (µL):

tRNA PCR extension mix	tRNA PCR master mix	10
	20 µM $^{ini}tRNA_{CAU}$ Ini1-1G-5'.F49.1018	0.5
	20 µM $^{ini}tRNA_{CAU}$ Ini-cat.R44	0.5

tRNA PCR 1 mix	Reagent:	Volume (µL):
	tRNA PCR master mix	190
	tRNA PCR extension product	10
	200 µM T7ex5.F22	0.5
	200 µM $^{ni}tRNA_{CAU}$ Ini-3'.R38	0.5

tRNA PCR 2 mix	Reagent:	Volume (µL):
	tRNA PCR master mix	1000
	tRNA PCR 1 product	5
	200 µM T7ex5.F22	2.5
	200 µM $^{ni}tRNA_{CAU}$ Ini-3'. 20-Me	2.5

tRNA transcription mix	Reagent:	Volume (µL):
	H_2O	390
	10x T7 buffer	100
	100 mM DTT (Merk)	100
	250 mM $MgCl_2$ (Merk)	90
	25 mM NTPs (Jena Bioscience)	150
	100 mM GMP (Jena Bioscience)	50
	tRNA PCR 2 product	100
	T7 RNA polymerase (Thermo Fisher Scientific)	20
	Total	10000

7.3 Gel compositions:

Gels		
8% Denaturing PAGE	Reagent:	Volume:
	Water	12 mL
	13.3% acrylamide (19:1), 6.7 M Urea (Merck)	24 mL
	5x TBE	4 mL
	10% APS (w/v) (Merck)	400 µL
	TEMED (Merck)	30 µL
	Final Volume	40.43 mL
12% Denaturing PAGE	Reagent:	Volume:
	13.3% acrylamide (19:1), 6.7 M urea (Merck)	36 mL
	5x TBE	4 mL
	10% Amonium Persulfate (APS) (Merck)	400 uL
	Tetramethylethylenediamine (TEMED) (Merck)	30 uL

MiHx Gel	Reagent:	Volume:
	30% acrylamide	3.3 mL
	Urea	1.8 g
	3M NaOAc (pH 5.2)	83 µL
	Water	< 5 mL

	10% (w/v) APS	50 µL
	TEMED	4 µL
	Final volume	5 mL

7.4 FIT Translation under genetic code reprogramming:

Sol A (Homemade) Mix	50 mM HEPES-KOH (pH 7.6)
	2 mM ATP (Jena Bioscience)
	2 mM GTP (Jena Bioscience)
	1 mM CTP (Jena Bioscience)
	1 mM UTP (Jena Bioscience)
	20 mM Creatine phosphate (Roche)
	100 mM Potassium acetate (Merck)
	2 mM Spermidine (Merck)
	6mM Magnesium acetate (Merck)
	1.5 mg/ml *E.coli* tRNA mix (Roche)
	14 mM DTT (Fluorochem)

Sol B Mix – FIT	PURExpress™ Δaa, tRNA Kit (NEB Bioscience)
Sol B Mix – Selection FIT	PURExpress™ ΔRF123 Kit (NEB Bioscience)

FIT Reaction (5 µL)	Reagent:	Volume (µL):
	Homemade Sol A Mix FD(-)	0.78
	Sol B (PURExpress™ Δaa, tRNA)	1.5
	5 µM mRNA template	1
	2.5 mM 19 aa mix (Met-)	0.5
	250 µM ClAc-Photo-^{L}Y-CME (36)-$^{ini}Met_{CAU}$ tRNA (in 1 mM NaOAc)	1
	H_2O	0.22
	Total	5

Selection FIT Reaction (5 µL)	Reagent:	Volume (µL):
	Homemade Sol A Mix FD(-)	0.78
	Sol B (PURExpress™ ΔRF123)	1.5
	2.5 mM 19 aa mix (Met-) (Merck)	0.5
	5 µM mRNA-Puro linker (IDT)	1
	250 µM ClAc-Photo-^{L}Y-CME (36)-$^{ini}Met_{CAU}$ tRNA (in 1 mM NaOAc)	1
	H_2O	0.22
	Total	5

7.5 PCR cycling programmes:

	Temperature (°C):	**Time (Sec):**	**Cycles:**
Post-round rt.q.PCR	94	10	35x
	61	10	

	72	30	
	61 → 72	0.5℃/ sec	

Selection Post-round PCR	Temperature (℃):	Time (Sec):	Cycles:
	94	10	Rt.q.PCR C_T + 5
	61	10	
	72	30	
	61 → 72	0.5℃/ sec	

Next Generation Sequencing			
NGS PCR 1	Temperature (℃):	Time (Sec):	Cycles:
	95 (Initial denaturation)	60	
	95	20	10x
	63	10	
	72	40	
	72	120	

NGS PCR 2	Temperature (℃):	Time (Sec):	Cycles:
	95 (Initial denaturation)	60	
	95	20	12x
	52	10	
	72	40	
	72	120	

Flexizyme			
Flexizyme PCR extension	Temperature (℃):	Time (Sec):	Cycles:
	95 °C	60	
	50 °C	30	12x
	72 °C	30	

Flexizyme PCR extension	Temperature (℃):	Time (Sec):	Cycles:
	95 °C	20	15-30x
	50 °C	20	
	72 °C	20	

tRNA			
tRNA PCR extension	Temperature (℃):	Time (Sec):	Cycles:
	95 °C	60	
	50 °C	60	5x
	72 °C	60	

tRNA PCR 1	Temperature (℃):	Time (Sec):	Cycles:
	95 °C	40	7x
	50 °C	40	
	72 °C	40	

tRNA PCR 2	Temperature (℃):	Time (Sec):	Cycles:

	95 °C	40	20x
	50 °C	40	
	72 °C	40	

7.6 Primer/ oligo table:

Name:	Sequence:
General Primers:	
T7g10M.F48	5-TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATG-3
T7g10M.F46	5-TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATA-3
CGS3an13.R39	5-TTTCCGCCCCCCGTCCTAGCTGCCGCTGCCGCTGCCGCA -3
Rd1T7g10M (100bp).F67	5- CACTCTTTCCCTACACGACGCTCTTCCGATCTCACTATAGGGTTAACTTTAAGAAGGAGATATACAT-3
an13Rd2.R49	5- GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTTCCGCCCCCCGTCCT-3

Barcode Primers:	
P5S502Rd1.F57	5-AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGAC-3
P5S503Rd1.F57	5-AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGAC-3
P5S505Rd1.F7	5-AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGAC-3
P5S506Rd1.F57	5-AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGAC-3
P5S507Rd1.F57	5-AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGAC-3
P5S508Rd1.F57	5-AATGATACGGCGACCACCGAGATCTACACCTAAGCCTACACTCTTTCCCTACACGAC-3

Rd2N701P7.R52	5-CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTTCAGACGTG-3
Rd2N702P7.R52	5-CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGACGTG-3
Rd2N703P7.R52	5-CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGACGTG-3
Rd2N704P7.R52	5-CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTTCAGACGTG-3
Rd2N705P7.R52	5-CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTTCAGACGTG-3
Rd2N706P7.R52	5-CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTG-3

Flexizyme Primers:	
Fx-5.F36.1018	5-GTAATACGACTCACTATAGGATCGAAAGATTTCCGC-3
T7ex5.F22	5-GGCGTAATACGACTCACTATAG-3
eFx.R45.1018	5-ACCTAACGCTAATCCCCTTTCGGGGCCGCGGAAATCTTTCGATCC-3
eFx.R18.1018	5-ACCTAACGCTAATCCCCT-3

dFx.R46.1018	5-ACCTAACGCCATGTACCCTTTCGGGGATGCGGAAATCTTTCGATCC-3
dFx.R19.1018	5-ACCTAACGCCATGTACCCT-3

aFx.R47	5-ACCTAACGCCACTTACCCCTTTCGGGGGTGCGGAAATCTTTCGATCC-3

aFx.R19	5-ACCTAACGCCACTTACCCC-3

tRNA Primers:	
Ini1-1G-5'.F49.1018	5-GTAATACGACTCACTATAGGCGGGGTGGAGCAGCCTGGTAGCTCGTCGG-3
Ini-cat.R44	5-GAACCGACGATCTTCGGGTTATGAGCCCGACGAGCTACCAGGCT-3
T7ex5.F22	5-GGCGTAATACGACTCACTATAG-3
Ini-3'.R38	5-TGGTTGCGGGGGCCGGATTTGAACCGACGATCTTCGGG-3
Ini-3'. 20-Me	5-TGmGTTGCGGGGGCCGGATTT-3 (Gm=2'-Methoxylated G)

MiHx Oligos:	
MiHx FAM-23b oligo	5-56-FAM/rArGrGrCrUrCrUrGrUrUrCrGrCrArGrArGrCrCrGrCrCrA-3
MiHx FAM-5b oligo	5-56-FAM/rCrGrCrCrA-3

Puro-PEG-DNA linker	5-([5'Phos]CTCCCGCCCCCCGTCC[SP18][SP18][SP18][SP18][SP18]CC [Puromycin]-3

NNK-Library Primers: (where N1 = A, T,C, G, N2 = G,T)	
stdlib_NNK4_R57	5-GCTGCCGCTGCCGCTGCCGCA(N1:50500000)(N2:25252525) (N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)CATATGTATATCTCCTTCTTAAAG-3
stdlib_NNK5_R60	5-GCTGCCGCTGCCGCTGCCGCA(N1:50500000)(N2:25252525) (N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)CATATGTATATCTCCTTCTTAAAG-3
stdlib_NNK6_R63	5-GCTGCCGCTGCCGCTGCCGCA(N1:50500000)(N2:25252525) (N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)CATATGTATATCTCC TTCTTAAAG-3
stdlib_NNK7_R66	5-GCTGCCGCTGCCGCTGCCGCA(N1:50500000)(N2:25252525) (N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)CATAT GTATATCTCCTTCTTAAAG-3
stdlib_NNK8_R69	5-GCTGCCGCTGCCGCTGCCGCA(N1:50500000)(N2:25252525) (N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1) (N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)CATATGTATATCTCCTTCTTAAAG-3
stdlib_NNK9_R72	5-GCTGCCGCTGCCGCTGCCGCA(N1:50500000)(N2:25252525) (N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1) (N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)CATATGTATATCTCCTTCTTAAAG-3
stdlib_NNK10_R75	5-GCTGCCGCTGCCGCTGCCGCA(N1:50500000)(N2:25252525) (N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1) (N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)CATATGTATATCTCCTTCTTAAAG-3
stdlib_NNK11_R78	5-GCTGCCGCTGCCGCTGCCGCA(N1:50500000)(N2:25252525) (N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1) (N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2) (N1)(N2)(N2)CATATGTATATCTCCTTCTTAAAG-3
stdlib_NNK12_R81	5-GCTGCCGCTGCCGCTGCCGCA(N1:50500000)(N2:25252525) (N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1) (N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2) (N1)(N2)(N2)(N1)(N2)(N2)CATATGTATATCTCCTTCTTAAAG-3
stdlib_NNK13_R84	5-GCTGCCGCTGCCGCTGCCGCA(N1:50500000)(N2:25252525) (N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1) (N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2) (N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)CATATGTATATCTCCTTCTTAAAG-3
stdlib_NNK14_R87	GCTGCCGCTGCCGCTGCCGCA(N1:50500000)(N2:25252525)

	(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1) (N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2) (N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)(N1)(N2)(N2)CATATGTATATCTCCTTCTTAAAG

7.7 Preparation of DNA standards for quantitative real-time PCR:

A DNA standard was made for use in quantitative PCR (q.PCR) post-round within the photo-selections. The standards were based on a template with an average length of the library used for selection (NNK9). Using the Reverse Transcription Mix (Reaction mix composition table – Chapter 7.2) and the NNK9 mixed library mRNA, the reverse transcription reaction proceeded for 1 hour at 42°C. Assuming that there was 100% reverse transcription, 0.2 µM mRNA equated to 2 $x10^{-7}$ M, which was 1.204 x 10^{17} molecules/L or 1.204 x 10^{11}/µL. Diluting 1 µL of the reverse transcription product in 60.2 µL of 1x Taq PCR Mix (minus Taq) (Reaction mix composition table – Chapter 7.2), yielded a 2e9 molecule/µL sample as the top concentration DNA standards. Serial dilutions to generate 2 x 10^8, 2 x 10^7, 2 x 10^6 and 2 x 10^5 molecules/µL were also made with 1x Taq PCR Mix (minus Taq) and stored at -20°C.

7.8 3% Agarose DNA gel and electrophoresis protocol to assess sufficient DNA amplification:

3.60 g of Agarose was mixed with 120 mL of 1x TAE, and it was heated in the microwave in short bursts of 5-30 seconds with swirling in between until it dissolved (Gel composition Table 7.3). Then, 12 µL of 10,000x SYBR Safe DNA stain was added, and it was poured into a cast and left to set. To check for sufficient DNA amplification, 2 µL was taken from a PCR amplified solution and mixed with 2 µL of 2x DNA loading buffer (Buffer table Chapter 7.1). The resulting sample was analysed using the 3% agarose gel loading in parallel with 2 µL of GeneRuler Low Range DNA Ladder at 135 V for 15 minutes.

7.9 Next generation sequencing:

10 µL of PCR amplified DNA is prepared for NGS by adding it within 'NGS PCR 1 Mix' with the required primers (Primer table 7.6), forward – 'Rd1T7g10M(100bp).F67' and reverse – 'an13Rd2.R49'. PCR cycling then followed for 10 cycles using the 'NGS PCR 1' protocol (PCR programme table 7.5). After checking sufficient amplification using the '3% Agarose DNA Gel protocol' (Chapter 7.8). 10 µL of the PCR product was added within 'NGS PCR 2 Mix' (Reaction mix table 7.2) with unique barcodes provided to each sample. This was achieved using the primers, forward – P5S5**Rd1.F57 and reverse – Rd2N7**P7.R52, were ** = a unique barcoding sequence e.g., P5S5.502.Rd1.F57 (Barcode primer

table 7.6). PCR cycling then followed for 12 cycles using the 'NGS PCR 2' protocol (PCR programme table 7.5). After checking sufficient DNA amplification again, the samples were combined to make up an equally constituted multiplexed solution of 100 µL. Purification of this multiplexed barcoded DNA then occurred using a commercial EasyPure PCR Purification Kit (TransGen Biotech, Code#EP101-01). The samples submitted were double indexed libraries (Nextera XT indices) and sequenced on a MiSeq platform (Illumina) using a v3 chip as single 151 cycle reads (1.8 million reads of the first 100 bps in both the forward and reverse direction). Sequences were ranked by total read numbers and converted into their corresponding peptides sequences for subsequent in-house data analysis.

7.10 Preparation of Flexizymes (eFx, aFx, dFx):

eFx, afx and dFx DNA were prepared through sequential PCR reactions using KOD polymerase and the corresponding primers (Primer table 7.6). Extension of the Flexizyme primers proceeded using the Flexizyme PCR extension mix and Flexizyme PCR extension programme (Reaction mix table 7.2, PCR programme table 7.5). From this 5 µL was carried forward for PCR amplification using the Flexizyme PCR reaction mix (Reaction mix table 7.2) and Flexizyme PCR reaction programme (PCR programme table 7.5). After checking sufficient DNA amplification via 3% Agarose gel electrophoresis (Chapter 7.8), the PCR product was purified using phenol-chloroform extraction and ethanol precipitation. The DNA was transcribed with T7 RNA polymerase (Thermo Scientific) for 18 hours using the Flexizyme transcription mix (Reaction mix table 7.2). DNA was then degraded by the addition of 115 µL of 10x DNase buffer and 30 µL of DNase the followed at 37℃ for 60 mins. The RNA was precipitated using a mix of isopropanol (1190 µL), 0.5M EDTA pH 8 (200 µL), and 3M NaCl (135 µL), centrifuging at 13,000 revolution(s)-per-minute (rpm) (Thermo Scientific Heraeus Fresco 17 centrifuge) for 10 mins. The precipitate was then redissolved with 100 µL of H_2O and an equal volume of 2x RNA loading buffer (buffer table 7.1). The RNA was purified via 12% denaturing PAGE purification in 1x TBE running buffer, running the sample at 230 V for 60 mins (gel table 7.3). After cutting the desired band, visualised by UV, the gel was covered in 0.3M NaCl, flash frozen and the supernatant mixed at rt for 18 h. This was repeated two further times, but with 1 h incubation, retaining and combining the supernatant each time. The solution was filtered (0.45µm syringe filter), and RNA extraction proceeded via addition of 2x volume of EtOH with centrifugation of 13,000 rpm (Thermo Scientific Heraeus Fresco 17 centrifuge) for 15 mins. The RNA pellet was redissolved in ~50 µL of water and the Flexizyme concentration assessed via nanodrop and kept at either -20℃ or -80 ℃.

7.11 tRNA preparation:

DNA encoding $^{Ini}Met_{CAU}$ tRNA was made through sequential 3 step PCR reactions using KOD polymerase and the corresponding primers. tRNA PCR master Mix was first prepared (Reaction mix table 7.2). Extension of the first tRNA primers was done using the tRNA PCR extension mix and tRNA PCR extension programme (Reaction mix table 7.2, PCR programme table 7.5). All the extension product (10 µL) was then used for the amplification of the tRNA using the tRNA PCR 1 mix and amplified by tRNA PCR 1 programme (Reaction mix table 7.2, PCR programme table 7.5). Further elongation of the PCR1 product was done by using 5 µL of the PCR 1 product within the tRNA PCR 2 mix which was amplified by tRNA PCR 2 programme (Reaction mix table 7.2, PCR programme table 7.5). After checking sufficient amplification via the 3% Agarose gel electrophoresis (chapter 7.8), the PCR product was purified using phenol-chloroform extraction and ethanol precipitation. The DNA pellet was redissolved in 20 µL of H_2O and transcribed for 18 h at 37℃ using the tRNA transcription mix (reaction mix table 7.2). After this, remaining DNA was degraded using 115 µL of 10x DNase buffer and 30 µL of DNase, incubating again at 37℃ for an additional 1 hour.

The RNA was precipitated using a mix of isopropanol (1190 µL), 0.5M EDTA pH 8 (200 µL), and 3M NaCl (135 µL), centrifuging at 13,000 rpm for 10 mins (Thermo Scientific Heraeus Fresco 17 centrifuge). the precipitate was then redissolved with 100 µL of H_2O and an equal volume of 2x RNA loading buffer (buffer table 7.1). The RNA was purified via 8% denaturing PAGE purification in 1x TBE running buffer, running the sample at 230 V for 60 mins (gel table 7.3). After cutting the desired band, visualised by UV, the extracted gel was covered in 0.3M NaCl, flash frozen and the supernatant mixed for 18 h overnight. This was repeated three times retaining the supernatant. The solution was filtered (0.45µm syringe filter), and RNA extraction proceeded via addition of 2x volume of EtOH with centrifugation of 13,000 rpm (Thermo Scientific Heraeus Fresco 17 centrifuge) for 15 mins. The RNA pellet was redissolved in 10 µL of water and its concentration via nanodrop. The RNA solution containing $^{Ini}Met_{CAU}$ tRNA was adjusted to 250 µM and kept at either -20℃ or -80 ℃.

7.12 Assessing the effect of UV-irradiation on the mRNA/ DNA hybrid RaPID library:

A known mRNA sequence was reverse transcribed at a final concentration of 0.2.µM using the Reverse Transcription Mix using a H(-) reverse transcriptase (reaction mix table 7.2). Following this, 9.44 µL was taken, with 0.5 µL and set aside for the input of the real-time PCR. The input was further diluted by 1/500 in H_2O and set aside. The samples were diluted 1:1 with 2x Blocking solution to mimic the

RaPID buffer conditions (buffer table 7.1). The remaining 17.88 µL was irradiated for varying time points 0, 0.5, 1, 3, 5 and 15 minutes, using LEDs of a single wavelength (365 nm, 3x 800 mW at 100% output), taking aliquots of 2.98 µL at each time point into individual low-binding Eppendorf's. The samples were then diluted 1/ 100 in 1x Taq PCR Mix to mimic a 1% recovery post RaPID selection. These diluted samples were then diluted again 1 /100 in 1x Taq PCR Mix, again to mimic a general RaPID procedure (reaction mix table 7.2). Rt.QPCR: 1 µl of the input and final diluted samples and DNA standards (10^9-10^4 DNA molecules) (chapter 7.7), were placed into separate PCR tubes with 19 µL of 1x Taq Real-time PCR Mix (reaction mix table 7.2). Amplification by rt.q.PCR then proceeded using the 'post-round rt.q.PCR' cycle conditions (PCR cycling table 7.5). Bulk PCR: Upon identifying the cycle threshold (C_T) for each condition via rt.q.PCR, 1 µL of 1x Taq DNA polymerase was added to the remaining 99 µL of final sample-PCR mix PCR tubes. The DNA was then amplified to the C_T + 5 cycles using the 'post-round PCR' conditions (PCR cycling table 7.5), and sufficient amplification and checked using the '3% Agarose gel' protocol (chapter 7.8). 10 µL of the amplified irradiated DNA was then prepared using the Next Generation Sequencing Protocol to assess the proportion of sequence reads containing sequence mutation (chapter 7.9).

7.13 Aminoacylation of tRNA by Flexizymes and Determination of Aminoacylation Efficiency:

7.13.1 MiHx Assay - Aminoacylation of the FAM23b microhelix (10 µL Scale):

Aminoacylation of an activated ester/thioester non-standard amino acid/ photoswitch analogue was performed on a truncated (23 base-pair) tRNA N-terminally conjugated to 4-carboxyfluorescein (FAM23b or FAM5b) (IDT) (Primer table 7.6). 3 µL of nuclease-free water, 1 µL of 0.5 M HEPES (pH 7.5/ pH 8.0/ pH 9.0) and 1 µL of 250 µM FAM23b microhelix were mixed in a 0.5 mL tube with 1 µL of 10 µM eFx, dFx, or aFx. The mixture was heated for 2 min at 95°C and cooled to room temperature for 5 mins. 2 µL of 300 mM $MgCl_2$ was added and incubated for 5 mins at room temperature. The mixture was then transferred to ice for 2 mins and 2 µL of 25 mM activated substrate (non-standard amino acid) in dimethyl sulfoxide (DMSO) was then added to the reaction mixture. In cases of poor substrate aq. solubility, a final concentration (f.c.) of DMSO was adjusted <40% via decreasing the volume of nuclease-free water accordingly. The reaction mixture was incubated for a time-course ranging 6–120 h on ice in a cold room. For all the time course experiments a negative control (2 µl of DMSO) and a positive control (ClAc-^{D}Y-CME – optimal modifiers: efx, HEPES pH 7.5, 20% DMSO, 2 h) was added.

7.13.2 Precipitation of the aminoacylated FAM-23b (10 µL Scale):

Precipitation of the aminoacylated FAM-23b can be achieved via the addition of 40 µL of 0.3 M NaOAc and 100 µL EtOH, followed by mixing and centrifugation at 25°C at 13,000 rpm (Thermo Scientific Heraeus Fresco 17 centrifuge) for 15 minutes and subsequent discard of the supernatant. The precipitate was then washed with 50 µL of 70% EtOH containing 0.1 M NaOAc (pH 5.2) and centrifuged at 25°C at 13,000 rpm for 5 mins, followed by discarding of the supernatant. The pellet was then air dried for 5-10 mins to ensure complete EtOH evaporation. The RNA pellet was stored on ice as Acyl-RNA's are prone to hydrolysis.

7.13.3 Acid polyacrylamide-gel-electrophoresis (PAGE) (10 µL Scale):

Aminoacylated FAM-23b was then analysed by acid PAGE gel (gel table 7.2). FAM-23b pellets were resuspended in 1.5 µL of 10 mM NaOAc (5.2 pH) and added to 14.5 µL loading buffer (buffer table 7.1). 2 µL of the sample was loaded in the gel and electrophoresis was carried out using 50 mM NaOAc (pH 5.2) as a running buffer for 120 mins at 120 V. The gel was visualized on a Typhoon 9500 imager using the FAM filter. The aminoacylation yield was determined by quantifying the intensity of the microhelix bands using Fiji/ImageJ,[165] using the following equation, (1):

$$\%\ aminoacylation = \frac{Intensity\ of\ aminoacylated\ FAM23b}{(Intensity\ of\ FAM23b + Intensity\ of\ aminoacylated\ FAM23b)}\ x\ 100\ (1)$$

7.14 Non-standard amino acid loading - Aminoacylation of tRNA (10 µL scale):

Aminoacylation of an $^{ini}Met_{CAU}$ tRNA with a photoswitch analogue occurred via the exact procedure as above, replacing the FAM23b for $^{ini}Met_{CAU}$ tRNA and using the optimised conditions pre-determined via the MiHx assay.

7.15 *in vitro* Translation (IVT) of a known peptide sequence and MALDI to confirm non-standard amino acid incorporation:

7.15.1 Translation:

Translation was undertaken using the in-house protocol based upon a commercial PURExpress IVT translation system and the homemade Sol-A mix (reaction mix table 7.2), incubating at 37 ℃ for 1 h.

7.15.2 MALDI:

The matrix solution for MALDI-MS was also prepared by taking a small amount of α-CHCA (Bio Vision) into a tube, add 10-20 µL of C-tip solution Elute (80% MeCN, 0.5% AcOH), and vortex well to make a saturated solution. Transfer the supernatant to another tube and add equal amount of C-tip solution Elute (i.e. prepare "half-saturated" solution of α-CHCA). With the solutions prepared, desalting the resulting translation mixture by means of C-tip can occur. 15 µL of C-tip solution Elute (80% MeCN, 0.5% AcOH) was first added to the c-tip pipette tip followed by benchtop centrifugation to expel all liquid. 15 µL of C-tip solution Wash (4% MeCN, 0.5% AcOH) was then added to the c-tip pipette tip followed by benchtop centrifugation. The translation mixture was then added to the tip followed by benchtop centrifugation.15 µL of C-tip solution Wash (4% MeCN, 0.5% AcOH) was then added to the c-tip pipette tip followed by benchtop centrifugation, this then was repeated for a second time. Finally, 1.2 µL Finally, 1.2 µL of the half-saturated matrix solution was used to elute the sample directly into the MALDI plate and the sample was analysed using MALDI-TOFF spectrometry.

7.16 PADI2 protein construct:

The *Human Padi2* (residues 1-663) construct was pre-established in the lab. It was previously amplified and cloned into a p28BIOH-LIC vector by ligation independent cloning with addition of a His6 tag between the vector encoded Avi-Tag and the *Padi2* gene, (Full encoded N-terminal tag sequence: MSGLNDIFEAQKIEWHEGSAGGSGHHHHHHGSGG). The DNA sequence of the protein construct was confirmed by Sanger sequencing.

7.17 PADI2 protein expression and purification:

Human Padi2 (residues 1-663), was previously amplified and cloned into a p28BIOH-LIC vector containing a His6 tag and and Avi-Tag. The plasmid was transformed into BL21-DE3 cells, which co-express biotin ligase (BirA) (BPS Bioscience), to produce N-terminally biotinylated PADI2 ((biotinylated)Avi-His-PADI2). An overnight starter culture was used to inoculate Terrific broth at a 1:100 ratio. When the OD 600 reached 0.6, the cells were induced with 0.2 mM IPTG and 20 µM D-biotin, and the temperature was reduced to 18°C. After 16 hours, the cells were harvested by centrifugation. The pellets were resuspended in a binding buffer (buffer table 7.1), with the addition of included DNase1, lysozyme and 1xEDTA-free protease inhibitors. The cells were then lysed by sonication, and after centrifugation, the clarified lysate was purified on a Ni-NTA 5 mL column (Cytiva)

using an ÄKTA Pure (Cytiva). After washing with binding buffer and an additional 15 mM imidazole, the protein was step eluted with binding buffer containing 250 mM imidazole (buffer table - 7.1).

Fractions containing Bio-His-PADI2 were concentrated and further purified using an S200 size exclusion column (HiLoad 16/600 Superdex 200 (Cytiva)) using PADI gel filtration buffer (buffer table - 7.1). The desired fractions containing (biotinylated)Avi-His-PADI2 were aliquoted, flash frozen, and stored at -80°C until needed. The biotinylation was confirmed through binding with streptavidin beads, High-Res Mass spectrometry and western blotting for biotin using a Streptavidin-Horseradish Peroxidase (HRP) Conjugate antibody (Invitogen).

7.18 Bead binding assay

Bio-His-PADI2 (10 µL at 1 µM in PADI2 buffer) was mixed with varying amounts of magnetic streptavidin beads (16, 8, 4, 2 or 1 µL) (dynabeads M-280, Invitrogen), in separate low binding tubes, at 4°C for 15 minutes. As a negative control 16 µL of 10 µM buffer (no PADI2) was also incubated. Biotin was added then to a final concentration of 25 µM and incubated at 4°C for 15 minutes. The supernatant for each condition was removed into separate new tubes and the remaining beads were individually washed with 3x 50 µL of PADI2 selection buffer (buffer table - 7.1). The individual beads were then resuspended in 10 µL of buffer, with 3.3 µL of 4x SDS loading buffer added to all samples, whilst the supernatant samples only 3.3 µL of 4x SDS loading buffer was added. As a positive control 10 µL at 1µM of Bio-His-PADI2 was mixed with 3.3 µL of 4x SDS loading buffer as well. Following heating to 95°C of all 12 samples, they were run on a SDS PAGE gel at 100 V for 1 hr, (precast mPAGE™ 4-20% Bis-Tris 10x8 12-wells (Merck), 1x MOPS buffer (Invitrogen)). The gel was then incubated in quick Coomassie stain (Generon) overnight and washed off with water prior to in gel imaging (white light) using a Chemidoc Gel Imaging system (Bio-Rad). Bead binding efficiency was determined by assessing the proportion of PADI2 on the bead relative to its supernatant using the following equation:

$$\frac{F.I.\,on\ bead}{(F.I.\,on\ bead + F.I.\,in\ supernatant)}\ x\ 100$$

Where F.I. = fluorescence intensity.

7.19 Intrinsic PADI2-nucleic acid binding:

A library of reverse transcribed mRNA/ DNA library (e.g. NNK9) (f.c. 0.2 µM) was preincubated 1:1 with either bovine serum albumin (BSA), or a combination of BSA and sheared salmon sperm DNA (s-ssDNA) (buffer table - 7.1). This nucleic acid solution (f.c. 0.1 µM) with blocking buffer, was then mixed with immobilised PADI2 magnetic beads at a f.c. of PADI2 of 200 nM. After incubation at 4 ℃ for 30 mins, the library was washed 3x 1 bead volume of PADI selection buffer (buffer table -7.1). Any remaining library interacting with PADI2 was recovered by boiling into 100 µL of PCR mix (reaction mix table - 7.2). The proportion of the library that was recovered compared to the initial input library (1 µL of initial reverse transcribed library diluted 1/500 with H_2O) was then evaluated using real-time quantitative polymerase chain reaction (qRT-PCR) to assess the proportion of DNA recovered using each blocking solution.

7.20 PhotoRaPID:

PhotoRaPID is based of the previously described protocols established by Suga and co-workers.[103] DNA sequences containing a T7 polymerase binding site, ribosome binding site, variable length randomised peptide coding region (6-12 randomised amino acids), $(Gly\text{-}Ser)_3$, linker, 'TAG' stop codon and sequence for puromycin ligation were constructed by extensions and PCR from oligos purchased from Integrated DNA technologies (IDT) to produce NNK6-12 mixed libraries (primer table - 7.6).

Initial DNA libraries (including 6-12 degenerate NNK codons in a ratio 0.001 $NNK_{n=6}$:0.032 $NNK_{n=7}$:1 $NNK_{n=8}$:32 $NNK_{n=9}$:33 $NNK_{n=10}$:33 $NNKn_{=11}$:11 $NNK_{n=12}$) (primer table - 7.6) were transcribed to mRNA using T7 RNA polymerase (37 ℃, 18 hr) (reaction mix - table 7.2) and were ligated to a puromycin linker primer ([5'Phos]CTCCCGCCCCCCGTCC[SP18][SP18][SP18][SP18][SP18]ACC [Puromycin]) using T4 ligase in presence of ATP at 25 ℃ for 0.5 h (NEB) (reaction mix - table 7.2). First round translations were performed on a 150 µL scale, with subsequent rounds performed on a 5 µL scale. Translations were carried out (30 min, 37 ℃ then 12 min, 25 ℃) using a custom methionine(-) Flexible In vitro Translation system combining the PURExpress™ (DRF123) Kit (New England Biolabs) solution B, an in-house solution A Mix, and the synthesised ClAc-Photo-^{L}Y-CME (36) (Reaction mix table - 7.2). Ribosomes were then dissociated by addition of EDTA (18 mM final concentration, pH 8) and library mRNA were reverse transcribed using MMLV RTase, Rnase H Minus (Promega). Reaction mixtures were buffer exchanged into selection buffer using 1 mL homemade columns containing pre-equilibrated Sephadex resin (Cytiva) or via the addition equal addition of $CaCl_2$ (18 mM final

concentration, pH 8). Blocking buffer was added and libraries were incubated in accordance with the irradiation workflows (below).

Irradiation Condition 1:

The selection library was first irradiated to the *Z*-isomer (365 nm, 3x 800 mW at 100% output, 3 mins) and three consecutive negative counter selections were performed against streptavidin magnetic beads (3x 30 min, 4 °C). Following this, the remaining library was recovered and re-irradiated to the *Z*-isomer, ensuring Z-isomer PSS. The *Z*-library was panned against immobilised PADI2 (200 nM, 4 °C, 30 min), before washing (3x1 bead volume selection buffer (reaction mix table - 7.2), 4 °C) and elution of retained mRNA/ DNA/peptide hybrids in PCR buffer (95 °C, 5 min) (reaction mix table - 7.2).

Irradiation Condition 2:

A light-responsive counter selection was first undertaken by irradiating the library to the *E*-isomer (495 nm, 3x 750 mW at 100% output, 3 mins) and the *E*-library was panned against immobilised PADI2 (200 nM, 4 °C, 30 min). The library solution (library members not binding to PADI2) was then recovered from the beads and irradiated to the *Z*-isomer (365 nm, 3x 800 mW at 100% output, 3 mins). The *Z*-active light-responsive selection was then undertaken by incubating the *Z*-rich library with a fresh set of PADI2-immobilised magnetic beads (200 nM, 4 °C, 30 min). Beads were washed (3x1 bead volume selection buffer, 4 °C) (reaction mix table - 7.2). The two sets of beads (the *E*-counter selection (*E*-negative), and the *Z*-positive selection) were all diluted with PCR mix and incubating at 95 °C for 5 minutes to elute the retained mRNA/ DNA/peptide hybrids in PCR buffer (95 °C, 5 min) (reaction mix table - 7.2).

Irradiation Condition 3:

The selection library was first irradiated to the *Z*-isomer (365 nm, 3x 800 mW at 100% output, 3 mins) and three consecutive negative counter selections were undertaken (3x 30 min, 4 °C). The remaining library was recovered and re-irradiated to the *Z*-isomer (365 nm, 3x 800 mW at 100% output, 3 mins) to ensure that the *Z*-isomer PSS was maintained. The *Z*-library was then panned against immobilised PADI2 (200 nM, 4 °C, 30 min). Following this, the remaining library solution was discarded and to remove residual library, the beads were then washed from the PADI2-immobilised beads (3x1 bead

volume selection buffer, 4 °C) (reaction mix table - 7.2), and the beads were resuspended in PCR mix (reaction mix table - 7.2). The beads in the PCR mix were then irradiated to the *E*-isomer (495 nm, 3x 750 mW at 100% output, 3 mins), mixed around and were removed, leaving only the PCR mix containing any peptides that had eluted from the protein following 495 nm irradiation. The 3x combined negative counter selection DNA was recovered via the standard boiling method (95 °C, 5 min).

Following the workflows, library recovery for each workflow (negative, *Z*-positive, *E*-positive where applicable) was then assessed by quantitative real-time PCR relative to a library standard, and the input DNA library. Recovered library DNA were used as the input library for the subsequent round. Following completion of the selections, double indexed libraries (Nextera XT indices) were prepared (Chapter 7.9) and sequenced on a MiSeq platform (Illumina) using a v3 chip as single 151 cycle reads. Sequences were ranked by total read numbers and converted into their corresponding peptides sequences for subsequent in-house data analysis (for Python analysis scripts see Appendix 41-43).

7.21 Solid phase peptide synthesis (SPPS) chemistry:

All peptides were synthesised as the C-terminal amide on low-loading rink amide resin (100-200 mesh) (Sigma-Aldrich), using the standard fluorenylmethyloxycarbonyl (Fmoc)-strategy solid-phase chemistry on a 0.05 mmol scale. Synthesis was undertaken on a Gyros Protein Technologies PreludeX Peptide Synthesiser (Gyros Protein Technologies AB, Sweden). Couplings were performed via one of two methods: (1) in DMF using HATU (5 eq.) in DMF and 4-methylmorpholine (5 eq.) in DMF at room for 1 h at room temperature or (2) via Oxyma/DIC (5 eq.) for 5 mins at 90°C for all amino acids excluding: Cysteine at 60°C and Histidine at room temperature to reduce racemisation. When using SPPS method (2), amino acid capping using acetic anhydride (20 eq.) and DIPEA (20 eq.) was performed after each subsequent coupling step. Fmoc groups were removed with 20% piperidine in DMF for 3x 10 minutes.

Following PreludeX SPPS, manual coupling of the photoswitch was undertaking using a manual SPPS tank (Sigma-Aldrich) with the peptide resin (0.05 mmol, 1.eq) in DMF (2 mL) with the resin pre-swelled. The Fmoc-compatible photoswitch, (E)-2-(4-((1-(2-((tert-butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)acetic acid (49) (31.4 mg, 0.06 mmol, 1.2 eq.) was added to the resin vessel, with: HATU (22.8 mg, 0.06 mmol, 1.2 eq.), and DIPEA (20 µL, .0.12 mmol, 2.4 eq.). The reaction

was undertaken in darkness for 1 h at room temperature, followed by resin washing using DMF (3x 10 mL). Double coupling of the photoswitch was undertaken where the availability of (49) allowed. Following removal of the photoswitch Fmoc group, the resin was incubated with N-(chloroacetoxy)succinimide (47.9 mg, 0.25 mmol, 5 eq.) for 1 h at room temperature. The resin was then washed with DMF (3x 10 mL) and dried with DCM (3x 5 mL) with draining *in vacuo*. The final chloroacetylated, photoswitch containing linear peptides were bulk cleaved from the resin by incubation for 1 hr/ Arg, at room temperature with 5 mL of a trifluoroacetic acid (TFA) cleavage cocktail (TFA/2,2'-(ethylenedioxy)diethanethiol/triisopropyl silane/H2O (92.5:2.5:2.5:2.5)) before filtration, precipitation with ice-cold diethyl ether (10x cocktail volume). Crude peptides were washed with diethyl ether (1x 5mL), dried and resuspended in DMSO (1-2 mL). The pH was raised to ~8 using triethylamine (~100-200 µL) via pre-wetted (H_2O) pH paper, to allow for cyclisation for 1hr at room temperature.

The crude peptides were purified by reverse phase HPLC with a preparative C18 column (XBridge prep BEH C18, 19x250 mm, 10 µm). A flow rate of 8 mL/min was used with a linear gradient of 10-45 % for 32 mins (solvent A: 99.9 % (v/v) water, 0.1 % (v/v) trifluoroacetic acid; solvent B: 99.9 % (v/v) acetonitrile and 0.1 % (v/v) trifluoroacetic acid). The resultant peptides were analysed using a Waters Acquity UPLC H-Class PLUS system with a C18 column (Aquity UPLC BEH C18, 2.1x50 mm, 1.7 µm). A linear gradient of 97-5% of either 4 or 8 mins was used [solvent A: 99.9 % (v/v) water, 0.1 % (v/v) formic acid; solvent B: 99.9 % (v/v) acetonitrile and 0.1 % (v/v) formic acid to analyse the final peptide purity.

7.22 Surface plasmon resonance (SPR):

Measurements were conducted on a Biacore S200 (Cytiva) and the data was analysed using the Biacore Evaluation Software (Cytiva). Experiments were performed at 25°C using either the single or multicycle kinetics mode. 75 or 300 nM Biotinylated PADI2 (in SPR PADI buffer (buffer table - 7.1)) was immobilised on a CAP chip (Biotin CAPture kit, series S Cytiva) for 120 seconds with a target density of ~1000-1500 RU. PADI2 SPR Buffer was used as the running buffer (buffer table - 7.1). Between cycles, the chip was regenerated following the manufacturer's protocol. Samples were prepared as per manufacturers specifications in Biacore tubes, with serial peptide dilutions being undertaken for each isomer (*Z/E*) individually, and each pool of samples were independently irradiated prior to the initiation of the SPR run using either 365 nm (for *Z*-rich population) or 495 nm (for *E*-rich population) for 3 minutes at max LED output and kept in complete darkness during the run.

7.23 COLDER assay:

The PADI2 activity was assessed using the Colour Development Reagent (COLDER) assay,[145] which measured the formation of urea-containing compounds (e.g., citrulline). The reaction was carried out in a 96 well plate with a final volume of 50 µl. His-PADI2 (100 nM) was incubated with different concentrations of peptides (50 µM to 3 nM) in the presence of $CaCl_2$ (10 mM) for 10 min at 37 °C. Then, the reaction was started by the addition of 10 mM BAEE (Nα-Benzoyl-L-arginine ethyl ester hydrochloride, Merck), a model arginine substrate, for 30 min at 37 °C after which the reaction was quenched with EDTA (50 mM final concentration). For colour development, 200 µl of COLDER solution, consisting of 1 volume of solution A (80 mM Diacetyl monoxime/2,3-butanedione monoxime (Merck) and 2 mM thiosemicarbazide) and 3 volumes of solution B (3 M H_3PO_4, 6 M H_2SO_4 and 2 mM $NH_4Fe(SO_4)_2$), was added to each well and the mixture was heated at 95 °C for 15 minutes. After cooling to RT, absorbance at 540 nm was measured and citrulline concentration was determined using a standard curve of citrulline standards. The citrullination activity of each peptide for each wavelength was independently assessed. This was achieved by irradiating prior to the initial incubation using either 365 nm (for Z-rich population) or 495 nm (for E-rich population) for 3 minutes at max LED output. For the duration of the reaction up to and including quenching, the plate was kept in complete darkness. All studies were performed within the linear range of PADI2 activity and data analysis was performed using Prism GraphPad (Dotmatics). Data were presented as the average ± standard error of the mean from one independent experiment with three technical replicates for each isomer of each peptide.

7.24 Photochemistry:

Irradiation was performed using a custom irradiation setup developed by Sahlman Photochemical solutions (Germany). Two LED modules were used:

- LED Module 3x 365 nm: LED type 3x Nichia NCSU276A, Rank: U365, typical optical power: 3x 800 mW at 100% output.
- LED Module 3x 495 nm: LED type 3x Nichia NCSE219B-V1, Rank: k.A., Typical optical power: 3x 750 mW at 100% output.

UV spectra were recorded on a Jasco V-630 spectrophotometer, with samples held in a 1 window 1 cm x 1 cm quartz cuvette. The sample was irradiated just prior to entry into the UV-Vis spectrophotometer meter sample chamber in complete darkness. The Z-rich PSS was established when further irradiation time did not generate a difference in the UV-Vis spectrum obtained.

Photochemical Conversion at a given PSS – estimating Z%:

The spectrum of the *Z*-isomer was estimated based on the irradiation time upon which the PSS at 365 was reached (*Z*-rich). Initially, the fraction of *E*-isomer present in the maximal Z-isomer PSS was estimated from the absorbance at the 'pure' (pre-irradiated) *E*-isomer at 365 nm. Using the fractional estimate, the spectrum of the pure Z-isomer was calculated using the method of Weston and co-workers,[46] via the equation:

$$A_Z = \frac{A_{PSS} - A_E \times f}{1 - f}$$

Where:

A_{PSS} = The absorbance of the PSS spectrum with the highest concentration of Z-isomer identified (*Z*-rich).

A_Z = Absorbance of the pure Z-spectrum.

A_E = Absorbance of the pure *E* spectrum.

F = Fraction of *E* at the PSS producing max *Z* isomerisation.

This estimate for the pure Z-isomer spectrum was then modulated to ensure that absorbance remained positive at wavelengths (~300-500 nm) and that no obvious π-π* absorbance from the *E*-isomer remained, this was varied until a range was obtained that predicted a sensible UV/vis spectrum for the pure Z-isomer. From this, the percentage *E*-isomer remaining in the *Z*-rich PSS (365 nm) was calculated by subtracting the experimental *Z* absorbance at that wavelength from the estimated pure *Z* absorbance giving the isolated *E*-absorbance in the *Z*-rich PSS. This was then divided by the pure *E* absorbance at that wavelength and expressed as a percentage (% *E* remaining in Z-rich state). From this percentage *Z*-isomer in the *Z*-rich PSS was assigned with a 3% uncertainty in absorbance as this was convention from Weston and co-workers.[46]

Estimation of the back-switching efficiency was determined by assessing the absorbance intensity pre-irradiated (pure-*E*) and upon reaching the PSS at 495 nm.

Calculating thermal half-life (τ1/2):

The rate of thermal reversion was calculated by irradiating at 365nm for the duration of time required to reach the PSS, as determined by the UV-Vis spectrometry (described above). From this point the

increase in absorbance at 365 nm was tracked, indicating the reversion from *Z*→*E* until a plateau or near plateau in the absorbance reading was seen. From this, using Prism GraphPad (Dotmatics), a one-phase exponential decay (Y = (Y0 - Plateau)*exp(-K*X) + Plateau) was fitted to give the thermal half-life.

7.25 Small Molecule chemical synthesis:

7.25.1 General chemistry reagents and equipment:

All reagents and solvents were supplied from commercial sources and used as received unless otherwise indicated. Reactions requiring anhydrous conditions were conducted in oven-dried glassware under an inert atmosphere (N_2 or Ar) and using anhydrous solvents that were obtained commercially. All water used was distilled using a Milli-Q® Ultrapure Lab Water System. Reactions were monitored by thin-layer chromatography (TLC) using Merck silica gel 60 F254 plates (0.25 mm) and/or using a Waters UHPLC-MS system equipped ACQUITY UPLC BEH C18 Column, 130Å, 1.7 µm, 2.1 mm X 50 mm. TLC plates were visualized using UV light (254 nm) (handheld lamp) and or with appropriate staining in potassium permanganate, ninhydrin, or vanillin if required.

UHPLC Methods:

Method Name:	**Gradient:**
Method 1	97% - 5% over 4 minutes, 0.25 mL/min, H_2O (0.1% Formic acid): MeCN, 1 µL injection.
Method 2	97% - 5% over 8 minutes, 0.25 mL/min, H_2O (0.1% Formic acid): MeCN, 1 µL injection.

Flash column chromatography was performed using both free silica gel (Sigma Aldrich (Merck)) 40-63 µm 60 Å treated with a solvent system or through commercial pre-loaded silica columns of varying sizes (Biotage®). The eluents used are specified in individual procedures. Solvents were removed by rotary evaporation (BüCHI Rotavapor R-300) at 40-45 °C (BüCHI heating bath B-300 base) and the compounds were further dried using high vacuum pumps. In the case of reverse-phase column chromatography: commercial pre-loaded columns (Biotage®) were used of varying scaled sizes, using a Water: Acetonitrile (specific gradients specified in individual procedures). Lyophilization of the corresponding purified aq. compounds was undertaken using the BenchTop Pro with Omnitronics™.

^{1}H and ^{13}C NMR were recorded on a Bruker Advance 400 spectrometer at 400 MHz and 100 MHz respectively or a Bruker Advance 600 spectrometer at 500 MHz and 125 MHz respectively. Chemical shifts (δ) are quoted in ppm (parts per million) and referenced to residual solvent signals: ^{1}H δ = 7.26 ($CDCl_3$), 2.50 (d^6-DMSO); ^{13}C δ = 77.16 ($CDCl_3$), 39.52 (d^6-DMSO). Coupling constants (*J*) are given in Hz. The following abbreviations are used to designate multiplicity within 1H NMR analysis; s = singlet, d = double, t = triplet, q = quartet, quin = quintet, m = multiplet, br. = broad signal, *app* = apparent. Any signals corresponding to more than one carbon within ^{13}C NMR analysis are quantified in brackets. Peptides were purified using a Waters LC-MS system equipped with both an XBridge prep C18 10 µm,19 × 250 mm BEH column and an XBridge C18 5 µm (focused gradient 60% - 35%, over 36 minutes, H_2O (0.1% TFA acid): MeCN (0.1% TFA acid), <2 mL injection) , 4.6 × 250 mm BEH column (97% - 5% over 20 minutes, H_2O (0.1% TFA acid): MeCN (0.1% TFA acid), 2 µL injection - Method 3 for small molecule analytical runs). MALDI-TOF mass spectra were measured on a Shimadzu MALDI-8030 (Imperial College London, UK). Spectrophotometry was undertaken using a Jasco V-630 spectrophotometer between 200-800 nm using a fast scan.

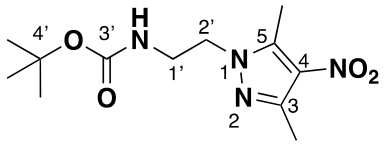
7.25.2 General Procedure – Zn Activation:

Commercial Zn dust (<9 µm), is added with stirring to a conical flask containing 2% hydrochloric acid (6.3 mL/g). Vigorous stirring is continued until the surface of the zinc becomes bright (ca. 5-10 minutes). The aqueous solution is decanted, and the solids in the flask are washed with four portions of distilled water (12 mL/g). The activated Zn is then transferred to Büchner filtration apparatus with distilled water (12 mL/g) and washed successively with ethanol (3.1 mL/g), acetone (6.3 mL/g), and dry ether (3.1 mL/g). Filtration and washing should be done as rapidly as possible to minimize exposure of the activated Zn to air. The Zn is finally dried in a vacuum desiccator for 1hr-overnight until a dry bright silver/grey fine powder is seen.

7.25.3 General Procedure – Boc deprotection using 4M HCl:Dioxane:

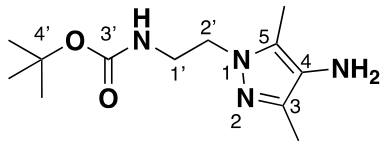
4 M HCl:Dioxane was added 1:1 to a solution of the Boc-protected photoswitch in DCM (f.c. of compound ~0.05 M) and stirred for 1 h at room temperature. During the reaction, the desired product as the HCl salt typically precipitates. After removal of the stirrer bar, the solution was directly concentrated *in vacuo* to afford an orange coloured solid and was then immediately used in the subsequent reactions as specified in text.

7.25.4 *tert*-Butyl (2-(3,5-dimethyl-4-nitro-1*H*-pyrazol-1-yl)ethyl)carbamate (13):



3,5-Dimethyl-4-nitro-1H-pyrazole (3.1 g, 22 mmol, 1eq.), caesium carbonate (7.8 g, 24 mmol, 1.1 eq.) and tert-butyl (2-bromoethyl) carbamate (4.9 g, 22 mmol, 1.eq.) were suspended in dry DMF (45 mL) under nitrogen and the mixture was heated at 60 °C for 16 hr. After completion, monitoring by LCMS, the reaction was cooled to room temperature, diluted with aq. 1 M NaOH (1x 50 mL) and extracted with ethyl acetate (3x 50 mL). The combined organic phases were then washed LiCl (5% w/v) solution (3x 50 mL), and the organics recovered. The organics were then concentrated until only residual DMF remained. This concentrated solution was then washed once more with a LiCl solution (3x 10 mL) and the final organics recovered. The solution was then dried over magnesium sulphate, filtered, and concentrated *in vacuo* to afford (13), as an off-white solid (5.5 g, 86%). The product can be carried through without purification. ^{1}H NMR (400 MHz, DMSO) δ 6.95 (t, *J* = 6.0 Hz, 1H, NH), 4.10 (t, *J* = 5.9 Hz, 2H, $C^{2'}H_2$), 3.25 (dt, *J* = 6.0, 5.9 Hz, 2H, $C^{1'}H_2$), 2.54 (s, 3H, C^{3}-CH_3), 2.38 (s, 3H, C^{5}-CH_3), 1.33 (s, 9H, $C^{4'}$-$(CH_3)_3$). ^{13}C NMR (101 MHz, DMSO) δ 156.0; $C^{3'}$, 145.4; C^{5}, 141.8; C^{3}, 130.7; C^{4}, 78.4; $C^{4'}$, 49.3; $C^{2'}$, 43.8; $C^{1'}$, 28.5; $C^{4'}$-$(\underline{C}H_3)_3$, 14.4; $C^{5}\underline{C}H_3$, 11.7; $C^{3}\underline{C}H_3$. LCMS ESI$^+$: Gradient: Method 1, Calc. Mass $C_{12}H_{20}N_4O_4$ $[MH]^+$ 284.15, Found at 185.1 $[M\text{-}Boc+H]^+$, rt. 2.41 min.

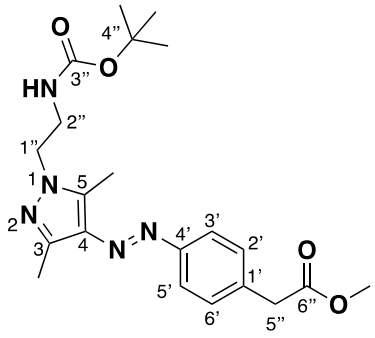
7.25.5 *tert*-Butyl (2-(4-amino-3,5-dimethyl-1*H*-pyrazol-1-yl)ethyl)carbamate (14):



To a solution of tert-butyl (2-(3,5-dimethyl-4-nitro-1H-pyrazol-1-yl)ethyl)carbamate (2.7 g, 9.5 mmol, 1 eq.) in ethanol (150 mL), water (75 mL) and acetic acid (15 mL), activated Zn dust (3 g, 46 mmol, 4.8 eq.); activated via General Procedure 1.21.2, was added. The reaction was stirred for 2 h at 60 °C. After cooling to room temperature, the suspension was filtered through Celite®, washed with ethyl acetate (3x 50 mL), and the filtrate was concentrated *in vacuo*. An aq. saturated $NaHCO_3$ solution was added (50 mL) to the residue and the aqueous phase extracted with ethyl acetate (3x 50 mL). The combined organic phases were then dried over magnesium sulphate, filtered, and evaporated *in vacuo*. Purification by reverse-flash column chromatography (H_2O 0.1% TFA: MeCN 0.1% TFA, 2%-10%)

afforded (14) as an off-white solid (2.3 g, 46%). The product can be carried through without purification. ^{1}H NMR (400 MHz, DMSO) δ 6.84 (t, *J* = 6.3 Hz, 1H, NH), 3.83 (t, *J* = 6.6 Hz, 2H, $C^{2'}$), 3.11 (dt, *J* = 6.6, 5.9 Hz, 2H, $C^{1'}$), 2.04 (s, 3H, C^{3}-CH_3), 1.96 (s, 3H, C^{5}-CH_3), 1.37 (s, 9H, $C^{4'}$-$(CH_3)_3$). ^{13}C NMR (101 MHz, DMSO) δ 155.6; $C^{3'}$, 136.7; C^{5}, 125.0; C^{3}, 124.1; C^{4}, 77.8; $C^{4'}$, 47.3; $C^{2'}$, 40.3; $C^{1'}$, 28.2; $C^{4'}$-($\underline{C}H_3)_3$, 10.9; $C^{5}CH_3$, 8.2; $C^{3}CH_3$. LCMS ESI^{+}: Gradient: Method 1, Calc. Mass $C_{12}H_{22}N_4O_2$ $[MH]^{+}$ 254.71, Found at 255.1 $[MH]^{+}$, rt. 1.25 min.

7.25.6 (*E*)-2-(4-((1-(2-((*tert*-Butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)acetate (17):

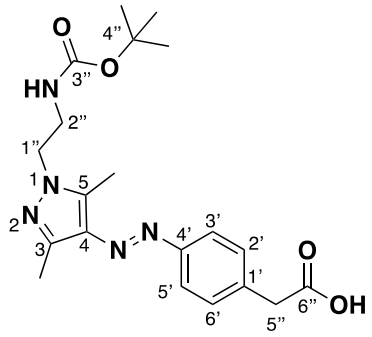


A solution of Oxone® (5.5 g, 36 mmol, 4 eq.) in H_2O (100 mL) was added to a solution of methyl 2-(4-aminophenyl)acetate (3.0 g, 18 mmol, 2 eq.), dissolved in DCM (50 mL). This was stirred under nitrogen, at rt and in darkness for 0.5 h, resulting in a bright green coloured solution. The solution was then transferred to a separatory funnel. Working promptly, the organics were recovered, in darkness, as the aqueous layer was extracted with DCM (2x 20 mL). The combined organic phases were kept in the dark, under nitrogen and then immediately added a pre-stirred solution at rt of tert-butyl (2-(4-amino-3,5-dimethyl-1H-pyrazol-1-yl)ethyl)carbamate (14) (2.3 g, 8.9 mmol, 1 eq.) in a mixture of chloroform (60 mL) and glacial acetic acid (150 mL), such that the final volume of glacial acetic acid to chloroform/DCM was 1:1. The nitroso solution was added slowly, via syringe and the combined solution was heated to 50 ℃ for 16 h in complete darkness. Upon reaching completion, monitored by LCMS, the reaction was cooled to room temperature and diluted with aq. 1 M HCl (200 mL), then added to a separatory funnel. The diluted reaction mixture was extracted with DCM (3x 50 mL). The combined organic phases were then dried over magnesium sulphate, filtered, and concentrated *in vacuo* to afford an orange coloured solid. Purification by flash column chromatography (cyclohexane: ethyl acetate, from 2% to 50%) afforded the purified product (17) as a bright yellow solid (3.7 g, 77%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.74 (m, 2H, 3′-CH, 5′-CH), 7.37 (m, 2H, 2′-CH, 6′-CH), 4.90 (t, 1H, NH), 4.19 (t, *J* = 5.6 Hz, 2H, $C^{1''}H_2$), 3.71 (s, 3H, OCH_3), 3.68 (s, 2H, $C^{5''}H_2$), 3.58 (dt, *J* = 6.0, 5.6 Hz, 2H, $C^{2''}H_2$), 2.57

(s, 3H, C^3CH_3), 2.50 (s, 3H, , C^5CH_3), 1.43 (s, 9H, $C^{4''}$-($C\underline{H}_3$)$_3$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 171.6; $C^{6''}$, 155.8, $C^{3''}$, 152.5; C^5, 142.; C^3, 139.7; C^4, 135.2; $C^{4'}$, 134.8; $C^{1'}$, 129.7; $C^{2'}C^{6'}$, 121.8; $C^{3'}C^{5'}$, 79.6; $C^{4''}$, 52.0; $C^{5''}$, 47.8; $C^{1''}$, 40.8; OCH_3, 40.1; $C^{2''}$, 28.2; O-($\underline{C}H_3$)$_3$, 13.7; $C^5\underline{C}H_3$, 9.5, $C^3\underline{C}H_3$.

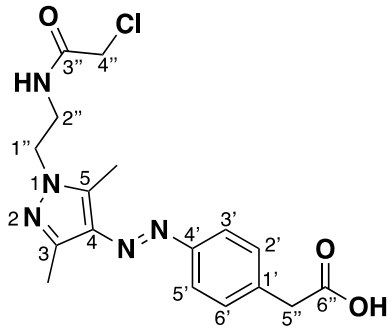
LCMS ESI$^+$: Gradient: Method 1, Calc. Mass $C_{21}H_{29}N_5O_4$ $[MH]^+$ 415.22, Found at 416.2 $[MH]^+$, rt. 2.38 min.

7.25.7 (*E*)-2-(4-((1-(2-((*tert*-Butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)acetic acid (18):



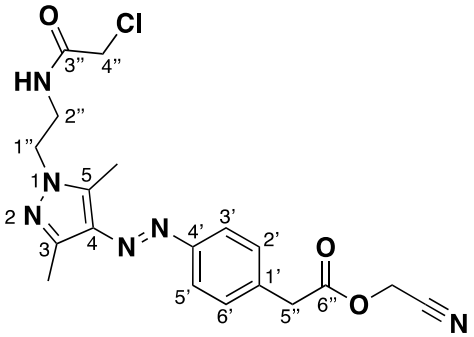
(*E*)-2-(4-((1-(2-((tert-Butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)acetate, (17) (2.9 g, 7.0 mmol, 1 eq.) in (150 mL) THF/ methanol/ H_2O (3:1:1.5) was mixed with LiOH.H_2O (1.8 g, 42 mmol, 6.0 eq.) at 0 ℃ for 2 h. The reaction mixture was acidified with 1 M HCl and extracted with ethyl acetate (3x 50 mL). The organic phases were then dried over magnesium sulphate, filtered, and concentrated *in vacuo* to afford an orange coloured solid (18). (1.5 g). The product was used in the next step without further purification. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.76 (m, 2H, 3′-CH, 5′-CH), 7.40 (m, 2H, 2′-CH, 6′-CH), 4.95 (t, 1H, NH), 4.27 (t, *J* = 5.6 Hz, 2H, $C^{1''}$), 3.72 (s, 2H, $C^{5''}$), 3.61 (br.(dt), 2H, $C^{2''}$), 2.59 (s, 3H, C^3CH_3), 2.54 (s, 3H, C^5CH_3), 1.43 (s, 9H, C4′′-(CH_3)$_3$). ^{13}C NMR (101 MHz, DMSO) δ 173.0; $C^{6''}$, 156.1; $C^{3''}$, 152.3, C^5, 141.2; C^3, 137.0; C^4, 134.9; $C^{1'}C^{4'}$, 130.7; $C^{2'}C^{6'}$, 121.7; $C^{3'}C^{5'}$, 78.4; $C^{4''}$, 48.4; $C^{1''}$, 40.9; $C^{2''}$, 32.5; $C^{5''}$, 28.7; $C^{4''}$-($\underline{C}H_3$)$_3$, 14.5; C^5, 9.7; C^3. LCMS ESI$^+$: Gradient: Method 1, Calc. Mass $C_{20}H_{27}N_5O_4$ $[MH]^+$ 401.21, Found at 402.3 $[MH]^+$, rt. 2.47 min.

7.25.8 (*E*)-2-(4-((1-(2-(2-Chloroacetamido)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)acetic acid (20):



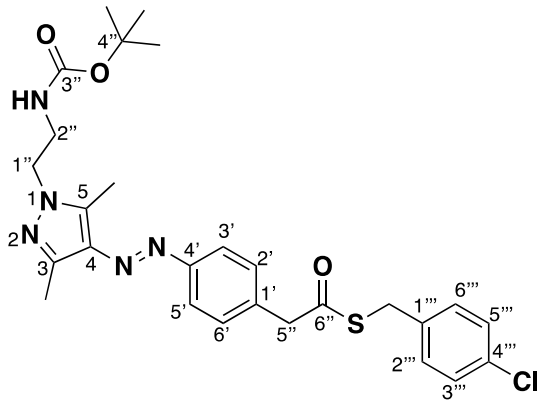
Boc deprotection of (18) was undertaken as according to the general procedure (Chapter 7.25.3) to generate (48). To a solution of (*E*)-2-(4-((1-(2-aminoethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)acetic acid.HCl (48) (84 mg, 0.25 mmol, 1 eq.) and $NaHCO_3$ (110 mg, 1.3 mmol, 5 eq.) in 1:1 Acetone: H_2O (5 mL), *N*-chloroacetoxy succinimide (250 mg, 1.3 mmol, 5 eq.) was added. The mixture was stirred at 0 °C for 2 h. Following the reaction, the solution was diluted with 1 M HCl then transferred to a separatory funnel and was extracted with ethyl acetate (3x 10 mL). The combined organic phases were then dried over magnesium sulphate, filtered, and concentrated *in vacuo* to afford an orange coloured solid (20). (50 mg). The product was used in the next step without further purification. ^{1}H NMR (400 MHz, DMSO) δ 8.42 (t, *J* = 5.8 Hz, 1H, NH), 7.67 (m, 2H, 3′-CH, 5′-CH), 7.39 (m, 2H, 2′-CH, 6′-CH), 4.13 (t, *J* = 6.1 Hz, 2H, $C^{1''}H_2$), 4.06 (s, 2H, $C^{4''}H_2$), 3.65 (s, 2H, $C^{5''}H_2$), 3.49 (dt, *J* = 6.1, 5.7 Hz, 2H, $C^{2''}H_2$), 2.54 (s, 3H, C^3CH_3), 2.39 (s, 3H, C^5CH_3). ^{13}C NMR (101 MHz, DMSO) δ 172.5; $C^{6''}$, 166.4; $C^{3''}$, 151.8; $C^{4'}$, 140.9; C^5, 140.1; C^3, 136.6; C^4, 134.3; $C^{1'}$, 130.3; $C^{2'6'}$, 121.3; $C^{3'5'}$, 47.4; $C^{1''}$, 42.5; $C^{4''}$, 40.4; $C^{5''}$, 39.0; $C^{2''}$, 14.0; C^5CH_3, 9.4; C^3CH_3. LCMS ESI$^+$: Gradient: Method 2, Calc. Mass $C_{17}H_{20}Cl^{35}N_5O_3$ $[MH]^+$ 377.13, Found at 378.25 $[MH]^+$, rt. 9.13 min.

7.25.9 Cyanomethyl (*E*)-2-(4-((1-(2-(2-chloroacetamido)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)acetate (11):



(*E*)-2-(4-((1-(2-(2-Chloroacetamido)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)acetic acid (20) (56 mg, 0.15 mmol, 1 eq.) was added to 1 mL of dry DMF, prior to the addition of bromoacetonitrile (13 µL, 0.18 mmol, 1.2 eq.) and potassium carbonate (9.3 mg, 0.22 mmol, 1.5 eq.). The reaction was stirred at rt for 1.5 h. Following the reaction, the solution was diluted with saturated aq. sodium bicarbonate and transferred to a separatory funnel before being extracted with ethyl acetate (3x 10 mL). The organic phases were then dried over magnesium sulphate, filtered, and concentrated *in vacuo*. Purification by reverse-flash column chromatography (H_2O 0.1% TFA: MeCN 0.1% TFA, 5% - 50%) afforded to afford the orange coloured solid ClAc-Photo-CME (11). (7.6 mg, 12%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.76 (m, 2H, 3′-CH, 5′-CH), 7.41 (t, NH), 7.37 (m, 2H, 2′-CH, 6′-CH), 4.75 (s, 2H, CH_2CN), 4.2 (t, *J* = 5.6 2H, $C^{1''}$), 4.05 (s, 2H, $C^{4''}$), 3.78 (s, 2H, $C^{5''}$), 3.77 (dt, *J* = 6.0, 5.6 Hz, 2H, $C^{2''}$), 2.58 (s, 3H, C^3CH_3), 2.50 (s, 3H, C^5CH_3). ^{13}C NMR (101 MHz, $CDCl_3$) δ 169.8; $C^{6''}$, 166.6; $C^{3''}$, 153.1. $C^{4'}$, 133.7; C^5, 139.6; C^3, 143.4; C^4, 135.3; $C^{1'}$, 130.0; $C^{2'}C^{6'}$, 122.4; $C^{3'}C^{5'}$, 114.3; $CH_2\underline{C}N$, 48.8; $\underline{C}H_2CN$, 47.3; $C^{1''}$, 42.6; $C^{4''}$, 40.3; $C^{5''}$, 39.7; $C^{2''}$, 14.2; C^5CH_3, 9.9; C^3CH_3. LCMS ESI$^+$: Gradient: Method 2, Calc. Mass $C_{19}H_{21}Cl^{35}N_6O_3$ $[MH]^+$ 416.14, Found at 417.25 $[MH]^+$, rt. 10.62 min.

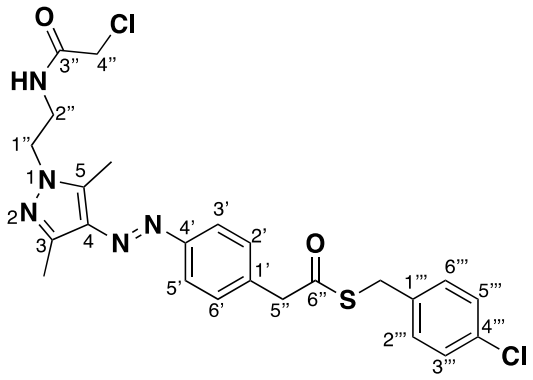
7.25.10 *S*-(4-Chlorobenzyl) (*E*)-2-(4-((1-(2-((*tert*-butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)ethanethioate (22):



(*E*)-2-(4-((1-(2-((*tert*-Butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)acetic acid (18) (50 mg, 0.12 mmol 1 eq.) was dissolved in anhydrous DMF (2.5 mL), followed by the addition PyBOP (73 mg, 0.14 mmol, 1.2 eq.) and stirred at rt for 15 mins. Triethylamine (50 µL, 0.36 mmol, 3 eq.) and 4-chlorobenzyl mercaptan (32 µL, 0.24 mmol, 2 eq.) were then subsequently added and the reaction was stirred at rt for 18 h. After completion, reaction was diluted with aq. $NaHCO_3$ in a separatory funnel (10 mL). The diluted reaction mixture was extracted with ethyl acetate (3x 10 mL). The combined organic phases were then washed with an aq. LiCl (5% w/v) solution (3x 10 mL) and extracted with ethyl acetate (3x 10 mL). The organics were subsequently dried over

magnesium sulphate, filtered and concentrated *in vacuo* to afford an orange coloured solid. Purification was achieved via flash column chromatography (ethyl acetate: DCM, 10% - 50%) afforded (22) as a yellow solid (14 mg, 21%). 1H NMR (400 MHz, CDCl3) δ 7.77 (m, 2H, 3′-CH, 5′-CH), 7.35 (m, 2H, 2′-CH, 6′-CH), 7.24 (m, 2H, 3‴-CH, 5‴-CH), 7.18 (m, 2H, 2‴-CH, 6‴-CH), 5.29 (s, 1H, NH), 4.16 (t, *J* = 5.8 Hz, 2H, $C^{1''}H_2$), 4.05 (s, 2H, S-CH_2), 3.88 (s, 2H, $C^{5''}H_2$), 3.54 (dt, *J* = 6.0, 5.8 Hz, 2H, $C^{2''}H_2$), 2.58 (s, 3H, C^3CH_3), 2.49 (s, 3H, C^5CH_3), 1.43 (s, 9H, $C^{4''}(CH_3)_3$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 196.4; $C^{6''}$, 156.0; $C^{3''}$, 153.0; $C^{4'}$, 146.6; C^5, 143.0; $C^{1'}$, 135.9; $C^{1'''}$, 134.4; C^3, 133.2; C^4, 130.3; $C^{4'''}$, 130.2; $C^{2'}C^{6'}$, 128.8; $C^{2'''}C^{5'''}$, 122.1; $C^{3'''}C^{5'''}$, 79.8; $C^{3'}C^{5'}$, 50.0; $C^{4''}$, 48.0; $C^{5''}$, 40.4; $C^{1''}$, 33.0; $C^{2''}$, 28.4; S$\underline{C}H_2$, 28.3; $C^{4''}(\underline{C}H_3)_3$, 11.4; $C^5\underline{C}H_3$, 9.7; $C^3\underline{C}H_3$. LCMS ESI$^+$: Gradient: Method 1, Calc. Mass $C_{27}H_{32}$ $Cl^{35}N_5O_3S$ $[MH]^+$ 541.19, Found at 542.3 $[MH]^+$, rt. 3.52 min.

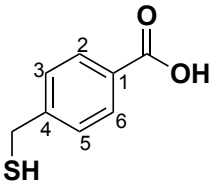
7.25.11 *S*-(4-Chlorobenzyl) (*E*)-2-(4-((1-(2-(2-chloroacetamido)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)ethanethioate (21):



Boc deprotection of (22) was undertaken as according to the general procedure (Chapter 7.25.3) to generate (23). *S*-(4-chlorobenzyl) (*E*)-2-(4-((1-(2-aminoethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl) phenyl)ethanethioate.HCl (23) (11 mg, 0.023 mmol, 1 eq.), was dissolved in anhydrous DCM (1 mL) followed by the addition of N-chloroacetoxy succinimide (21 mg, 0.11 mmol, 5 eq.) and potassium carbonate (11 mg, 0.11 mmol, 5 eq.) and the reaction was stirred at rt for 2 h. After completion, reaction was diluted with 1 M HCl (10 mL) and transferred to a separatory funnel. The diluted reaction mixture was extracted with ethyl acetate (3x 10 mL). The combined organic phases were then washed with a LiCl (5% w/v) solution (3x 10 mL) and extracted with ethyl acetate several times. The organics were subsequently dried over magnesium sulphate, filtered and concentrated *in vacuo* to afford an orange coloured solid. Purification by reverse-flash column chromatography (H_2O 0.1% TFA: MeCN 0.1% TFA, 5%-95%) afforded (21) as a yellow solid (11 mg, 93%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.75 (m, 2H, 3′-CH, 5′-CH), 7.42 (s, 1H, NH), 7.38 (m, 2H, 2′-CH, 6′-CH), 7.24 (d, *J* = 8.4 Hz, 2H, 3‴-CH, 5‴-CH), 7.19 (m, 2H, 2‴-CH, 6‴-CH), 4.20 (t, *J* = 5.2, 4.8 Hz, 2H, C1''H2), 4.06 (s, 2H, SCH_3), 4.05 (s, 2H,$C^{4''}H_2$), 3.88 (s,

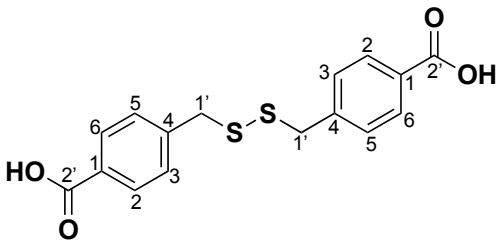
2H, $C^{5''}H_2$), 3.77 (dt, *J* = 6.0, 5.2 Hz, 2H, $C^{2''}H_2$), 2.57 (s, 3H, C^3CH_3), 2.50 (s, 3H, C^5CH_3). ^{13}C NMR (101 MHz, $CDCl_3$) δ 196.5; $C^{6''}$, 166.6; $C^{3''}$, 153.0; C^5, 143.4; C^3, 139.5; $C^{1'}$, 136.1; $C^{1'''}$, 135.3; $C^{4'''}$, 134.7; $C^{4'}$, 133.3; C^4, 130.4; $C^{2'}C^{6'}$, 130.4; $C^{2'''}C^{6'''}$, 128.9; $C^{3'''}C^{5'''}$, 122.3; $C^{3'}C^{5'}$, 50.2; $C^{1''}$, 47.3; $C^{4''}$, 42.6; $C^{2''}$, 39.7; $C^{5''}$, 33.1; SCH_3, 14.2; C^3CH_3, 9.9; C^5CH_3. LCMS ESI$^+$: Gradient: Method 1, Calc. Mass $C_{24}H_{25}Cl_2{}^{35}N_5O_2S$ $[MH]^+$ 517.11, Found at 518.2 $[MH]^+$, rt. 3.25 min.

7.25.12 4-(Mercaptomethyl)benzoic acid (29):



Methyl 3-(bromomethyl) benzoate (476 µL, 3.1 mmol, 1 eq.) in dry acetone (5 mL) was added dropwise to a solution of thiourea (240 mg, 3.1 mmol, 1 eq.) in dry acetone (20 mL) at room temperature. After 15 minutes, the suspension was concentrated *in vacuo*. The obtained solids were then then resuspended in H_2O (32.5 mL), and the pH was adjusted to pH 10.5 by the addition of aq. 10% (w/v) NaOH. The slurry mixture was refluxed at 110 °C for 2 h. After cooling to room temperature, the solution was extracted with ethyl acetate (3x 20 mL), and the organic phases were discarded. The aqueous solution was acidified to pH 1.5 and extracted with ethyl acetate (3x 20 mL). The combined organics were dried over magnesium sulphate, filtered, and concentrated *in vacuo* to afford a colourless solid (29) (310 mg, 60%). ^{1}H NMR (400 MHz, DMSO) δ 12.85 (s, 1H, OH), 7.88 (m, 2H, 1-CH, 2-CH), 7.46 (m, 2H, 3-CH, 5-CH), 3.79 (d, *J* = 8.0 Hz, 2H, S-C$\underline{H}_2$), 2.98 (t, *J* = 8.0 Hz, 1H, SH). ^{13}C NMR (101 MHz, DMSO) δ 167.6; COOH, 147.3; C^1, 130.0; C^3C^5, 129.6; C^4, 128.8; C^1C^2, 27.9; SCH_2). LCMS ESI$^-$: Gradient: Method 1, Calc. Mass $C_8H_8O_2S$ $[MH]^-$ 167.01, Found at 167 $[MH]^-$, rt. 2.05 min.

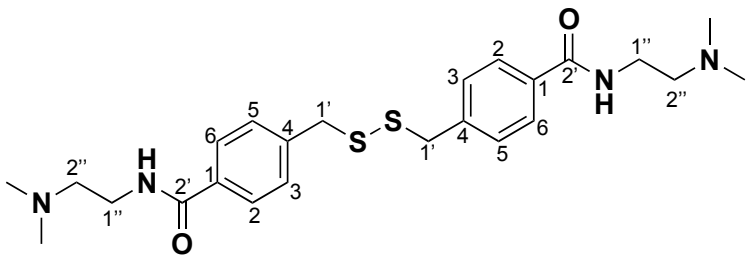
7.25.13 4,4′-(Disulfanediylbis(methylene))dibenzoic acid (30):



A solution of 4-(Mercaptomethyl)benzoic acid (29) (290 mg, 1.7 mmol, 1 eq.) in ethanol (2 mL) was cooled at 0°C by an ice bath. To the mixture was added a solution of 0.5 M I_2 in ethanol until the colour

of the solution turned red and the mixture was stirred at 0 °C for 1 hr. A solution of 0.5 M sodium thiosulfate in H_2O was then added to the mixture until the colour of the solution changed from red to white. Ethanol in the resulting mixture was evaporated by concentrating the solution *in vacuo*, and the precipitant was filtrated and washed with H_2O using vacuum filtration to afford a colourless solid (29) (200 mg, 71%). ^{1}H NMR (400 MHz, DMSO) δ 12.93 (s, 2H, $(COOH)_2$), 7.91 (m, 2.0 Hz, 4H, (2-CH, 6-CH)$_2$), 7.39 (m, 4H, (3-CH, 5-CH)$_2$), 3.84 (s, 4H, $(C^{1'})_2$). ^{13}C NMR (101 MHz, DMSO) δ 167.1; COOH, 142.5; C^4, 129.7; C^1, 129.6; C^3C^5, 129.4; C^2C^6, 40.9; $C^{1'}$. LCMS ESI$^-$: Gradient: Method 1, Calc. Mass $C_{16}H_{14}O_4S_2$ $[MH]^-$ 334.03, Found at 333.0 $[MH]^-$, rt. 2.42 min.

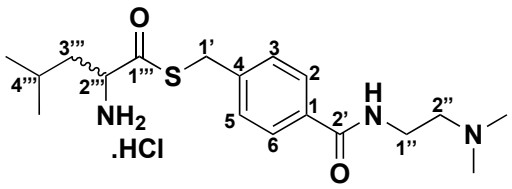
7.25.14 4,4′-(Disulfanediylbis(methylene))bis(*N*-(2-(dimethylamino)ethyl)benzamide) (31):



To generate the initial activated succinimide ester, N-hydroxysuccinimide (1.5 g, 13 mmol, 2.5 eq.) was added to a stirring solution of 4,4′-(Disulfanediylbis(methylene))dibenzoic acid (30) (1.7 g, 5.0 mmol, 1 eq.) in 1,4-dioxane (20 mL). 1-Ethyl-3-(3- dimethylaminopropyl)carbodimide hydrochloride (CDI) (2.1 g, 13 mmol, 2.5 eq.) in DCM (30 mL) was then added in small portions and the reaction mixture was stirred at rt for 1 hr. Following completion, monitored by LCMS, the organic solvent was concentrated *in vacuo* and ethyl acetate (30 mL) was added to the residue. Any precipitant was removed by filtration, the filtrate was washed with aq. 1 M HCl (3x 30 mL) and aq. saturated $NaHCO_3$ (3x 30 mL). The organics were subsequently dried over magnesium sulphate, filtered and concentrated *in vacuo* to afford a colourless solid as the succinimide ester intermediate. The intermediate (2.6 g, 5.0 mmol, 1 eq.) was then immediately dissolved 1, 4-dioxane (30 mL) and a solution of N,N-Dimethylethylenediamine (1.4 mL, 13 mmol, 2.5 eq.) in H_2O (15 mL) was added. After being stirred at 0 °C for 2 h, the reaction was concentrated *in vacuo* to afford an off-white solid. Purification by reverse-flash column chromatography (MeCN (0.1% FA): H_2O (0.1% FA), 20%-98%) afforded (31) as a yellow solid (950 mg, 40%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.47 (t, *J* = 5.6 Hz, 2H, $(NH)_2$), 6.63 (m, 4H, (2-CH, 6-CH)$_2$), 6.02 (m, 4H, (3-CH, 5-CH)$_2$), 2.58 (dt, *J* = 6.0, 5.6 Hz, 4H, $(C^{1''}H_2)_2$), 2.40 (s, 4H, $(C^{1'}H_2)_2$), 1.92 (t, *J* = 5.6 Hz, 4H, $(C^{2''}H_2)_2$), 1.50 (s, 12H, $((NCH_3)_2)_2$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 168.7; $(C^{2'})_2$, 141.3;

$(C^4)_2$, 132.8; $(C^1)_2$, 129.5; $(C^3C^5)_2$, 127.7; $(C^2C^6)_2$, 57.7; $(C^{1''})_2$, 43.8; $((NCH_3)_2)_2$, 43.0; $C^{1'}$, 35.7; $C^{2''}$. LCMS ESI$^+$: Gradient: Method 1, Calc. Mass $C_{24}H_{34}N_4O_2S_2$ $[MH]^+$ 474.21, Found at 475.3 $[MH]^+$, rt. 1.44 min.

7.25.15 S-(4-((2-(Dimethylamino)ethyl)carbamoyl)benzyl) 2-amino-4-methylpentanethioate.HCl (34):



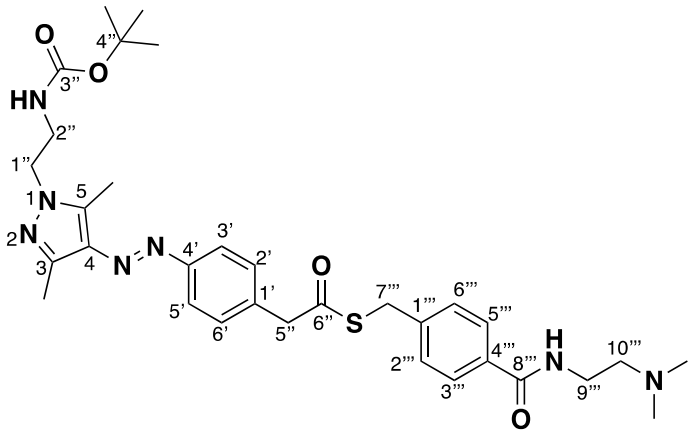
4,4′-(Disulfanediylbis(methylene))bis(*N*-(2-(dimethylamino)ethyl)benzamide) (31) (100 mg, 0.21 mmol, 1 eq.) and 10-molar equivalents of TCEP (600 mg, 2.1 mmol, 10 eq.))were added to 2 mL of H_2O and the solution was stirred for 10 minutes at room temperature until the reaction proceeded to completion. Following completion, monitored by LCMS, the reaction was concentrated *in vacuo*. Ethyl acetate (10 mL) was added to the residue. The organics were then washed with aq. 2 M NaOH (3x 30 mL). The organics were subsequently dried over magnesium sulphate, filtered and concentrated *in vacuo* to afford a colourless solid *N*-(2-(dimethylamino)ethyl)-4-(mercaptomethyl) benzamide (26).

(26) (62 mg, 0.26 mmol, 1.2 eq.) was then immediately added to N-^{t}Boc-L-leucine (50 mg, 0.22 mmol, 1 eq.), 1-ethyl-3-(3-dimethylaminopropyl)carbodimide hydrochloride (EDC.HCl) (63 mg, 0.33 mmol, 1.5 eq.) and dimethylaminopyridine (DMAP) (40 mg, 0.33 mmol, 1.5 eq.) in DCM (1 mL). The reaction was stirred at rt for 1 h and upon completion, monitoring by LCMS, the mixture was concentrated *in vacuo*. The solids were then redissolved in ethyl acetate (10 mL and the resulting solution was washed with aq. 1 M HCl (3x 15 mL), followed by aq. saturated $NaHCO_3$ (3x 3 mL). The organic phases were subsequently dried over magnesium sulphate, filtered and concentrated *in vacuo* to afford (33). Purification by reverse-flash column chromatography (MeCN (0.1% FA): H_2O (0.1% FA), 20%-98%) afforded *S*-(4-((2-(dimethylamino)ethyl)carbamoyl)benzyl)2-((tert-butoxycarbonyl)amino)-4-methylpentanethioate (33).

(33) (20 mg, 0.04 mmol, 1 eq.) was then immediately dissolved in DCM and added to a solution of 4 M HCl:Dioxane (f.c. of compound ~0.05 M) which was stirred at rt for 1 hr and monitored by LCMS until completion. the solution was directly concentrated *in vacuo* to afford a chalky amber coloured

solid as the di hydrochloride salt (34) (19 mg, 20% over 2-steps). ^{1}H NMR (400 MHz, DMSO) δ 8.89 (t, *J* = 5.6 Hz, 1H, NH), 8.56 (s, 3H, NH_2.HCl), 7.89 (m, 2H, 2-CH, 6-CH), 7.49 (m, 2H, 3-CH, 5-CH), 4.31 (s, 2H, $C^{1'}H_2$), 4.21 (t, *J* = 6.9 Hz, 1H, $C^{2'''}H$), 3.63 (dt, *J* = 6.0, 5.9 Hz, 2H, $C^{1''}H_2$), 3.25 (t, *J* = 5.9 Hz, 2H, $C^{2''}H_2$), 2.81 (s, 6H, $N(CH_3)_2$), 1.61 (t, 2H, $C^{3'''}H_2$), 1.23 (s, 1H, $C^{4'''}H$), 0.89 (S, 6H, $C^{4'''}(CH_3)_2$). ^{13}C NMR (101 MHz, DMSO) δ 197.1; $C^{1'''}$, 166.6; $C^{2'}$, 141.0; C^{4}, 133.4; C^{1}, 129.2; $C^{3}C^{5}$, 128.2; $C^{2}C^{6}$, 57.3; $C^{2'''}$, 56.5; $C^{2''}$, 42.8; $N(\underline{C}H_3)_2$, 40.9; $C^{3'''}$, 35.0; $C^{1''}$, 24.3; $C^{1'}$, 22.8; $C^{4'''}$, 22.3; $C^{4'''}(CH_3)_2$. LCMS ESI^{+}: Gradient: Method 1, Calc. Mass $C_{18}H_{29}N_3O_2S$ $[MH]^+$ 351.20, Found at 352.3 $[MH]^+$, rt. 1.26 min.

7.25.16 *S*-(4-Chlorobenzyl) (*E*)-2-(4-((1-(2-((*tert*-butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)ethanethioate (34):

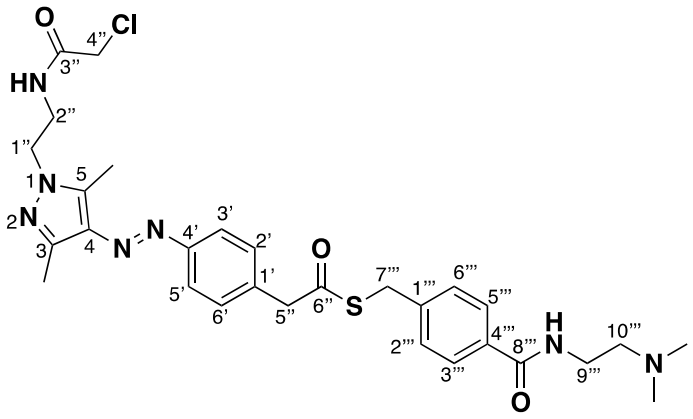


4,4′-(Disulfanediylbis(methylene))bis(*N*-(2-(dimethylamino)ethyl)benzamide) (31) (100 mg, 0.21 mmol, 1 eq.) and 10-molar equivalents of TCEP (600 mg, 2.1 mmol, 10 eq.))were added to 2 mL of H_2O and the solution was stirred for 10 minutes at room temperature until the reaction proceeded to completion. Following completion, monitored by LCMS, the reaction was concentrated *in vacuo*. Ethyl acetate (10 mL) was added to the residue. The organics were then washed with aq. 2 M NaOH (3x 30 mL). The organics were subsequently dried over magnesium sulphate, filtered and concentrated *in vacuo* to afford a colourless solid *N*-(2-(dimethylamino)ethyl)-4-(mercaptomethyl) benzamide (26).

(*E*)-2-(4-((1-(2-((*tert*-Butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl) acetic acid (18) (90 mg, 0.22 mmol, 1 eq.) was mixed *N*-(2-(dimethylamino)ethyl)-4-(mercaptomethyl)benzamide (26) (62 mg, 0.26 mmol, 1.2 eq.), 1-ethyl-3-(3-dimethylaminopropyl)carbodimide hydrochloride (EDC.HCl) (63 mg, 0.33 mmol, 1.5 eq.) and dimethylaminopyridine (DMAP) (40.3 mg, 0.33 mmol, 1.5 eq.) in DCM (1 mL). After being stirred at rt for 1 h, the mixture was concentrated *in vacuo* and was redissolved with ethyl acetate (10 mL). The

resulting solution was washed aq. sat. $NaHCO_3$ (3x 10 mL). The organics were subsequently dried over magnesium sulphate, filtered and concentrated *in vacuo* to afforded (34) as an orange solid (92 mg, 67%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.74 (m, 2H, 3‴-CH, 5‴-CH), 7.71 (m, 2H, 2‴-CH, 6‴-CH), 7.36 (m, 2H, 3′-CH, 5′-CH), 7.31 (m, 2H, 2′-CH, 6′-CH), 6.89 (s, 1H, NH-DiMe), 4.92 (s, 1H, NH-Boc), 4.15 (t, J = 5.7 Hz, 2H, C1′′H_2), 4.12 (s, 2H, $C^{7'''}H_2$), 3.88 (s, 2H, $C^{5''}H_2$), 3.56 (dt, J = 6.0, 5.7 Hz, 2H, $C^{2''}H_2$), 3.51 (dt, J = 6.0, 5.4 Hz, 2H,), 2.56 (s, 3H, C^5CH_3), 2.52 (t, J = 5.9 Hz, 2H, $C^{10'''}H_2$), 2.48 (s, 3H, C^3CH_3), 2.27 (s, 6H, $N(CH_3)_2$), 1.43 (s, 6H, 4′′-$(CH_3)_3$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 196.4; $C^{6''}$, 176.2, $C^{8''}$,167.2; $C^{3''}$, 154.5; C^5, 153.6; C^3, 153.1; C^4, 150*; $C^{4'}$, 141.0; C^4, 134.5; $C^{1'''}$, 133.8; $C^{1'}$, 130.4; $C^{4'''}$, 129.1; $C^{2'}C^{6'}$, 127.5; $C^{3'''}C^{5'''}$, 122.2; $C^{2'''}C^{6'''}$, 79.9; $C^{3'}C^{5'}$, 57.9; $C^{4''}$, 50.1; $C^{10'''}$, 48.1; $C^{1''}$, 45.2; $C^{2''}$, 37.2; $N(CH_2)_2$, 33.4; $C^{7'''}$, 28.5; $C^{5''}$, 14.2; C^5, 9.8; C^3.*$C^{4'}$ not observed due to DMAP impurity. LCMS ESI$^+$: Gradient: Method 1, Calc. Mass $C_{32}H_{43}N_7O_4S$ $[MH]^+$ 621.31, Found at 622.4 $[MH]^+$, rt. 2.24 min.

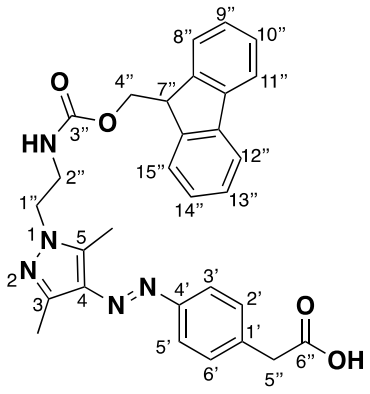
7.25.17 *S*-(4-Chlorobenzyl) (*E*)-2-(4-((1-(2-(2-chloroacetamido)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)ethanethioate (25):



Boc deprotection of (34) was undertaken as according to the general procedure (Chapter 7.25.3) to generate (35). To a solution of S-(4-((2-(dimethylamino)ethyl)carbamoyl)benzyl) (E)-2-(4-((1-(2-aminoethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)ethanethioate.HCl (35) (120 mg, 0.22 mmol, 1 eq.) and *N*,*N*′-diisopropylethylamine (DIPEA) (140 µL, 0.55 mmol, 2.5 eq.) in dry DMF (10 mL) was added a solution of N-chloroacetoxy succinimide (110 mg, 0.55 mmol, 2.5 eq.) in dry DMF 5.0 mL. The reaction was stirred for 2 h at room temperature. After removal of the stirrer bar, the solution was directly concentrated *in vacuo*. Purification by reverse-flash column chromatography (MeCN (0.1% FA): H_2O (0.1% FA), 5%-50%) afforded (33) as a yellow solid (6.6 mg, 5%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.61 (s, 1H,ClAc-NH), 8.14 (s, 1H, DiMe-NH), 7.97 (m, 2H, 3′-CH, 5′-CH), 7.75 (m, 2H, 3‴-CH, 5‴-CH), 7.37 (m, 2H, 2′-CH, 6′-CH), 7.34 (m, 2H, 2‴-CH, 6‴-CH), 4.20 (t, 3H, $C^{1''}$), 4.12 (s, 2H,$C^{7'''}$), 4.05 (s, 2H, $C^{4''}$), 3.89 (s, 2H, $C^{5''}$), 3.87 (dt, J = 6.0, 5.6 Hz, 2H, $C^{2''}$), 3.77 (dt, J = 6.0, 5.6 Hz, 2H, $C^{9'''}$), 3.21 (t,

2H, $C^{10'''}$), 2.83 (s, 6H, $(NCH_3)_2$), 2.57 (s, 3H, C^5), 2.50 (s, 3H, C^3). ^{13}C NMR (101 MHz, $CDCl_3$) δ 196.4; $C^{6''}$, 153.0; C^4, 141.3; $C^{1'''}$, 130.5; $C^{2'}C^{6'}$, 129.2; $C^{2'''}C^{6'''}$, 128.2; $C^{3'''}C^{5'''}$, 122.3; $C^{3'}C^{5'}$, 59.2; $C^{10'''}$, 50.0; $C^{1''}$, 47.5; $C^{2''}$, 44.2; $N(CH_3)_2$, 42.6; $C^{4''}$, 39.7; $C^{9'''}$, 35.3; $C^{7'''}$, 29.8; $C^{5''}$, 14.1; C^5, 9.9, C^3. Quaternary carbon C^8 weak signal and quaternary carbons $C^{3''}$, C^5, C^3, $C^{4'}$, $C^{1'}$, $C^{4'''}$, $C^{8'''}$ are not observed. LCMS ESI^+: Gradient: Method 1, Calc. Mass $C_{29}H_{36}Cl^{35}N_7O_3S$ $[MH]^+$ 597.23, Found at 598.3 $[MH]^+$, rt. 2.04 min.

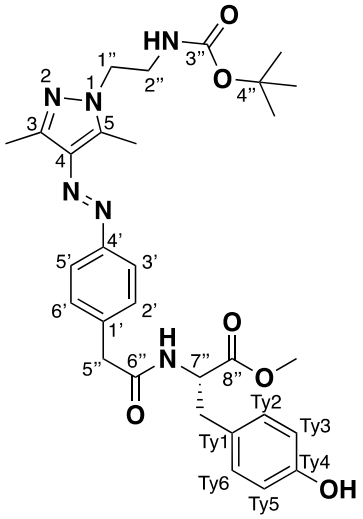
7.25.18 (E)-2-(4-((1-(2-((tert-Butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)acetic acid (49):



Boc deprotection of (18) was undertaken as according to the general procedure (Chapter 7.25.3) to generate (48). To a solution of (*E*)-2-(4-((1-(2-aminoethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)acetic acid.HCl (48) (2.9 g, 8.5 mmol, 1 eq.), N-(9-Fluorenylmethoxycarbonyloxy)succinimide (FmocOSu) (3.4 g, 10.2 mmol, 1.2 eq.) was dissolved in dry DMF 52 mL, with the addition of DIPEA (4.5 mL, 26 mmol 3 eq.). The reaction was left stirring at rt for 2 h. After completion, monitored by LCMS, the reaction was diluted with aq. 1 M HCl (100 mL) and transferred to a separatory funnel. The diluted reaction mixture was extracted with ethyl acetate (3x 40 mL). The combined organic phases were then washed with an aq. LiCl (5% w/v) solution (3x 50 mL) and extracted with ethyl acetate again (3x 40 mL). The organics were subsequently dried over magnesium sulphate, filtered and concentrated *in vacuo* to afford an orange coloured solid. Purification was achieved via washing of the solids with diethyl ether several times (5x 50 mL) to remove Fmoc-related impurities. The final product achieved was a yellow solid (49), (2.5 g, 56%/ over 3-steps). ^1H NMR (400 MHz, DMSO) δ 12.40 (s, 1H, OH), 7.87 (m, 2H, $3'$-CH, $5'$-CH), 7.66 (m, 4H, $11''$-CH, $12''$-CH, $2'$-CH, $6'$-CH), 7.48 (t, *J* = 5.9 Hz, 1H, NH), 7.40 (m, 4H, $9''$-CH, $10''$-CH, $13''$-CH, $14''$-CH), 7.30 (m, 2H, $8''$-CH, $15''$-CH), 4.32 (d, *J* = 6.8 Hz, 2H, $C^{4''}H_2$), 4.19 (t, *J* = 6.7 Hz, 1H, $C^{7''}H$), 4.09 (t, *J* = 5.9 Hz, 2H, $C^{1''}H_2$), 3.65 (s, 2H, $C^{5''}H_2$), 3.37 (dt, *J* = 6.0, 5.9 Hz, 2H, $C^{2''}H_2$), 2.47 (s, 3H, C^3CH_3), 2.36 (s, 3H, C^5CH_3). ^{13}C NMR (101 MHz, DMSO) δ 172.5; $C^{6''}$, 162.3; $C^{3''}$, 156.2; C^5, 151.8; C^3, 143.8; C^4, 140.8; $C^{7''}\underline{C}C^{8''}$, 140.7; $C^{11''}\underline{C}C^{12''}$, 136.6; $C^{1'}$, 130.2; $C^{10''}C^{13''}$, 127.6; $C^{9''}C^{14''}$, 127.1; $C^{8''}C^{15''}$, 126.6; $C^{4'}$, 125.1; $C^{11''}C^{12''}$, 121.2;

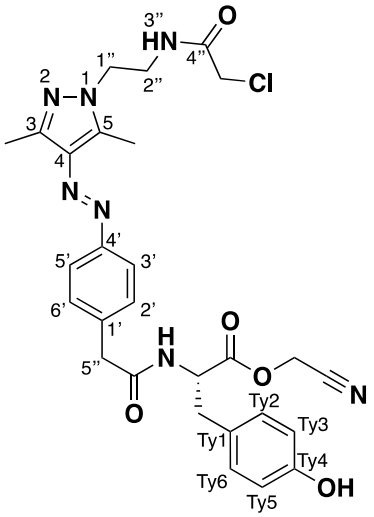
$C^{2'}C^{6'}$, 120.1; $C^{3'}C^{5'}$, 65.4; $C^{4''}$, 47.9; $C^{1''}$, 46.7; $C^{7''}$, 40.9; $C^{5''}$, 40.4; $C^{2''}$, 14.0; C^{5}, 9.3; C^{3}. LCMS ESI$^+$: Gradient: Method 1, Calc. Mass $C_{30}H_{29}N_5O_4$ $[MH]^+$ 523.22, Found at 524.2 $[MH]^+$, rt. 2.94 min.

7.25.19 Methyl (*E*)-(2-(4-((1-(2-((*tert*-Butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)acetyl)-*L*-tyrosinate (37):



(E)-2-(4-((1-(2-((tert-butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)acetic acid (18) (67 mg, 0.17 mmol, 1 eq.) was dissolved in dry DMF (2 mL). Methyl *L*-tyrosinate.HCl (60 mg, 0.25 mmol, 1.5 eq.), PyBOP (140 mg, 0.26 mmol, 1.5 eq.) and DIPEA (66 μL, 0.38 mmol, 2.2 eq.) were added and the reaction was stirred at rt for 1 h. The diluted reaction mixture was extracted with ethyl acetate (3x 5 mL). The combined organic phases were then washed with a LiCl (5% w/v) solution (3x 5 mL) and extracted with ethyl acetate (3x 5 mL). The organics were subsequently dried over magnesium sulphate, filtered and concentrated *in vacuo* to afford the crude product as an orange coloured solid. Purification by reverse-flash column chromatography (H_2O 0.1% TFA: MeCN 0.1% TFA, 10%-40%) afforded (37) as a yellow solid (68 mg, 69%). ^{1}H NMR (400 MHz, DMSO) δ 9.23 (s, 1H, OH), 8.54 (d, *J* = 8.4 Hz, 1H, TyrNH), 7.6 (m, 2H, Tyr^4HTyr^5H), 7.25 (m, 2H, Tyr^2HTyr^6H), 6.98 (m, 2H, $3'$-CH, $5'$-CH), 6.65 (m, 2H, $2'$-CH, $6'$-CH), 4.41 (dt, *J* = 8.3, 5.5 Hz, 1H, $C^{7''}H$), 4.08 (t, *J* = 5.9 Hz, 2H, $C^{1''}H_2$), 3.60 (s, 3H, OCH_3), 3.49 (d, *J* = 2.7 Hz, 2H, $C^{5''}H_2$), 3.28 (dt, *J* = 6.0, 5.7 Hz, 2H, $C^{2''}H_2$), 3.01 (d, *J* = 2.7 Hz, 2H, $TyrCH_2$), 2.53 (s, 3H, C^3CH_3), 2.37 (s, 3H, C^5CH_3), 1.35 (s, 9H, $C^{4''}(CH_3)_3$). ^{13}C NMR (101 MHz, DMSO) δ 172.3; $C^{8''}$, 156.0; $C^{3''}$, 147.6; $C^{4'}$, 141.0; C^3, 137.7; C^4, 130.2; $C^{2'}C^{6'}$, 129.9; Tyr^2,Tyr^6, 121.3; $C^{3'}C^{5'}$, 115.2; Tyr^4, Tyr^5, 78.0; $C^{3'}C^{5'}$, 54.1; $C^{7''}$, 54.0; $C^{1''}$, 51.9; OCH_3, 46.0; $C^{2''}$, 45.9; $TyrCH_2$, 41.7; $C^{5''}$, 28.3; $C^{4''}(CH_3)_3$, 14.1, C^5, 9.3; C^3. Quaternary carbons, C^5, $C^{1'}$ $C^{6''}$, Tyr^1, Tyr^4 are not observed. LCMS ESI$^+$: Gradient: Method 1, Calc. Mass $C_{30}H_{38}N_6O_6$ $[MH]^+$ 578.29, Found at 579.4 $[MH]^+$, rt. 2.45 min.

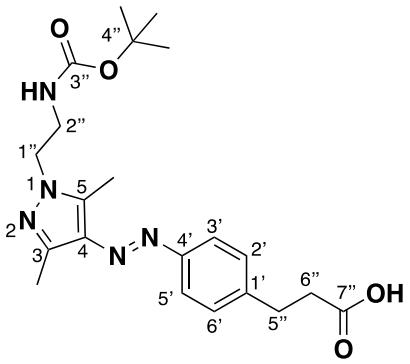
7.25.20 Cyanomethyl (*E*)-(2-(4-((1-(2-(2-chloroacetamido)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)acetyl)-*L*-tyrosinate (36):



Methyl (*E*)-(2-(4-((1-(2-((tert-butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl) phenyl)acetyl)-L-tyrosinate (37) (75 mg, 0.13 mmol, 1 eq.) in THF/ methanol/ H_2O (3:1:1.5) (2 mL) was mixed with LiOH.H_2O (33 mg, 0.78 mmol, 6.0 eq.) for 2 h at 0 ℃. Following completion, monitored by LCMS, the reaction was acidified with aq. 1 M HCl (10 mL), transferred to a separatory funnel and extracted with ethyl acetate (3x 5 mL). The organic phases were then dried over magnesium sulphate, filtered, and concentrated *in vacuo* to afford (38). Boc deprotection of (38) was then undertaken immediately as according to the general procedure (Chapter 7.25.3) to generate (39) as the hydrochloride salt. (39) was then immediately dissolved in anhydrous DCM (1 mL) followed by the addition of N-chloroacetoxy succinimide (63 mg, 0.33 mmol, 2.5 eq.) and DIPEA (57 µL, 0.33 mmol, 2.5 eq.) and the reaction was stirred at 0 ℃ for 1 h. After completion, monitored by LCMS, acidified to pH 7 with aq. 1 M HCl and concentrated *in vacuo* to generate (40). (40) was then dissolved into 1.5 mL of acetonitrile. Potassium carbonate (8.4 mg, 0.2 mmol, 1.5 eq.) and chloroacetonitrile (13 µL, 0.2 mmol, 1.2 eq.) were added, and reaction was left at rt for 1.5 h. Following completion, monitored by LCMS, the solution was directly concentrated *in vacuo* to afford an orange coloured solid. Purification by reverse-flash column chromatography (H_2O 0.1% TFA: MeCN 0.1% TFA, 10%-90%) afforded (36) as a yellow solid (17 mg, 23% over 4-steps). ^{1}H NMR (400 MHz, DMSO) δ 9.28 (s, 1H, TyrOH), 8.68 (d, *J* = 7.5 Hz, 1H, TyrNH), 8.42 (t, *J* = 5.8 Hz, 1H, $N^{3''}H$), 7.63 (m, 2H, TyrC^3HTyrC^5H), 7.26 (m, 2H, 3′-CH, 5′-CH), 7.00 (m, 2H, 2′-CH, 6′-CH), 6.65 (m, 2H, TyrC^2HTyrC^6H), 4.98 (s, 2H, CH_2CME), 4.47 (m, 1H, TyrCH), 4.13 (t, *J* = 6.1 Hz, 2H, $C^{1''}H_2$),4.06 (s, 2H. CH_2Cl), 3.50 (s, 2H, $C^{5''}H_2$), 3.49 (dt, *J* = 6.1, 5.9 Hz, 2H, $C^{2''}H_2$), 2.89 (m, 2H, TyrCH_2), 2.54 (s, 3H, C^3CH_3), 2.39 (s, 3H, C^5CH_3). ^{13}C NMR (101 MHz, DMSO) δ 171.3; Tyr<u>C</u>ONH,

170.5; COO, 166.9; $C^{4''}$, 156.6; $Tyr^{4'}$, 152.1; $C^{4'}$, 141.3; C^5, 140.5; C^5, 138.0; C^3, 134.8; C^4, 130.6; $C^{1'}$, 130.2; C^2C^6, 127.1; $C^{5'}C^{5'}$, 121.7; $TyrC^1$, 116.2; $TyrC^3TyrC^5$, 115.6; C≡N, 54.2; $TyrC^2TyrC^6$, 49.9; TyrCH, 47.9; CH_2CN, 42.9; $C^{5''}$, 41.9; CH_2Cl, 39.4; $C^{1''}$, 36.1; $TyrCH_2$, 14.5; C^5CH_3, 9.8; C^3CH_3. LCMS ESI^+: Gradient: Method 1, Calc. Mass $C_{28}H_{30}Cl^{35}N_7O_5$ $[MH]^+$ 579.20, Found at 580.2 $[MH]^+$, rt. 2.23 min.

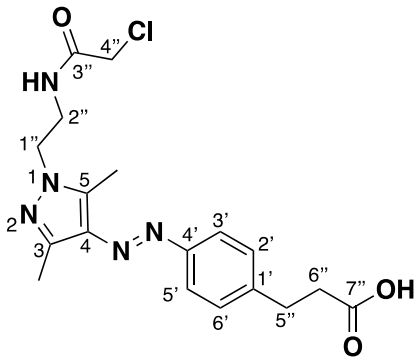
7.25.21 (*E*)-3-(4-((1-(2-((*tert*-Butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)propanoic acid (45):



A solution of Oxone® (1.2 g, 7.6 mmol, 4 eq.) in H_2O (16 mL) was added to a solution of methyl 3-(4-aminophenyl)propanoic acid (630 mg, 3.8 mmol, 2 eq.), dissolved in DCM (4 mL). This was stirred under nitrogen, at rt and in darkness for 0.5 h, resulting in a bright green coloured solution. The solution was then transferred to a separatory funnel. Working promptly, the organics were recovered, in darkness, as the aqueous layer was extracted with DCM (2x 10 mL). The combined organic phases were kept in the dark, under nitrogen and then immediately added to a rt pre-stirred solution of tert-butyl (2-(4-amino-3,5-dimethyl-1H-pyrazol-1-yl)ethyl)carbamate (14) (490 mg, 1.9 mmol, 1 eq.) in a mixture of chloroform (12 mL) and glacial acetic acid (36 mL), such that the final volume of glacial acetic acid to chloroform/DCM was 1:1. The nitroso solution was added slowly, via syringe and the combined solution was heated to 50 ℃ for 16 h in complete darkness. Upon reaching completion, monitored by LCMS, the reaction was cooled to room temperature and diluted with aq. 1 M HCl (100 mL), then added to a separatory funnel. The diluted reaction mixture was extracted with DCM (3x 25 mL). The combined organic phases were then dried over magnesium sulphate, filtered, and concentrated *in vacuo* to afford an orange coloured solid. Purification by reverse-flash column chromatography (H_2O 0.1% TFA: MeCN 0.1% TFA, 10%-90%) afforded the purified product (45) as a bright yellow solid (109 mg, 14%). ^{1}H NMR (400 MHz, DMSO) δ 12.18 (s, 1H, OH), 7.63 (m, 2H, 3′-CH, 5′-CH), 7.35 (m, 2H, 2′-CH, 6′-CH), 6.99 (t, *J* = 5.8 Hz, 1H, NH), 4.07 (t, *J* = 6.0 Hz, 2H, $C^{1''}H_2$), 3.27 (dt, *J* = 6.0, 5.7 Hz, 2H, $C^{2''}H_2$), 2.88 (t, *J* = 7.6 Hz, 2H, $C^{6''}H_2$), 2.57 (t, *J* = 7.6 Hz, 2H, $C^{5''}H_2$), 2.52 (s, 3H, C^3CH_3), 2.37 (s, 3H, C^5CH_3), 1.35 (s, 9H, $C^{4''}(CH_3)_3$). ^{13}C NMR (101 MHz, DMSO) δ 173.7; $C^{7''}$, 155.6; $C^{3''}$, 151.5;

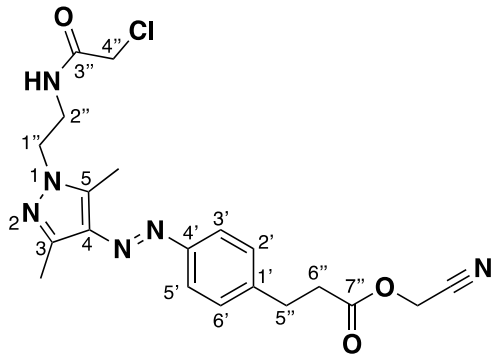
$C^{4'}$, 142.6; C^{5}, 140.7; C^{1}, 139.8; C^{3}, 134.2; C^{4}, 129.0; $C^{2'}C^{6'}$, 121.3; $C^{3'}C^{5'}$, 77.9; $C^{4''}$, 47.9; $C^{1''}$, 38.9; $C^{2''}$, 35.1; $C^{5''}$, 30.1; $C^{6''}$, 28.2; $C^{4''}(CH_3)_3$, 14.0; C^{5}, 9.2; C^{3}. LCMS ESI$^+$: Gradient: Method 1, Calc. Mass $C_{21}H_{29}N_5O_4$ $[MH]^+$ 415.22, Found at 416.3 $[MH]^+$, rt. 2.51 min.

7.25.22 (*E*)-3-(4-((1-(2-(2-Chloroacetamido)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)propanoic acid (47):



Boc deprotection of (45) was undertaken as according to the general procedure (Chapter 7.25.3) to generate (46). To a solution of (*E*)-3-(4-((1-(2-aminoethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)propanoic acid.HCl (46) (76 mg, 0.24 mmol, 1 eq.) and DIPEA (110 µL, 0.6 mmol, 2.5 eq.) in anhydrous DMF (1.5mL), N-chloroacetoxy succinimide (110 mg, 0.6 mmol, 2.5 eq.) was added. The mixture was stirred for 1 h at room temperature. Following the reaction, monitored by LCMS, the solution was diluted with 1 M HCl (15 mL) and transferred to a separatory funnel and was extracted with ethyl acetate (3x 10 mL). The organic phases were then dried over magnesium sulphate, filtered, and concentrated *in vacuo* to afford the crude product (47) as a yellow solid. ^{1}H NMR (400 MHz, DMSO) δ 12.16 (s, 1H, OH), 8.41 (t, *J* = 6.0 Hz, 1H, NH), 7.64 (m, 2H, 3′-CH, 5′-CH), 7.36 (m, 2H, 2′-CH, 6′-CH), 4.13 (t, *J* = 6.0 Hz, 2H, $C^{1''}H_2$), 4.06 (s, 2H, $C^{4''}H_2$), 3.49 (dt, *J* = 6.0, 5.8 Hz, 2H, $C^{2''}H_2$), 2.88 (t, *J* = 6.0 Hz, 2H, $C^{6''}H_2$), 2.57 (t, *J* = 7.4 Hz, 2H, $C^{5''}H_2$), 2.53 (s, 3H, $C^{3}CH_3$), 2.38 (s, 3H, $C^{5}CH_3$). ^{13}C NMR (101 MHz, DMSO) δ 174.14; $C^{7''}$, 166.89; $C^{3''}$, 164.25; $C^{4'}$, 151.92; C^{3}, 140.37; $C^{1'}$, 134.75; C^{3}, 134.43; C^{4}, 129.52; $C^{2'}C^{6'}$, 121.85; $C^{3'}C^{5'}$, 47.9; $C^{1''}$, 42.93; $C^{4''}$, 39.45; $C^{2''}$, 35.50; $C^{6''}$, 30.57; $C^{5''}$, 14.50; C^{5}, 9.80; C^{3}. LCMS ESI$^+$: Gradient: Method 1, Calc. Mass $C_{18}H_{22}Cl^{35}N_5O_3$ $[MH]^+$ 391.14, Found at 392.2 $[MH]^+$, rt. 2.15 min.

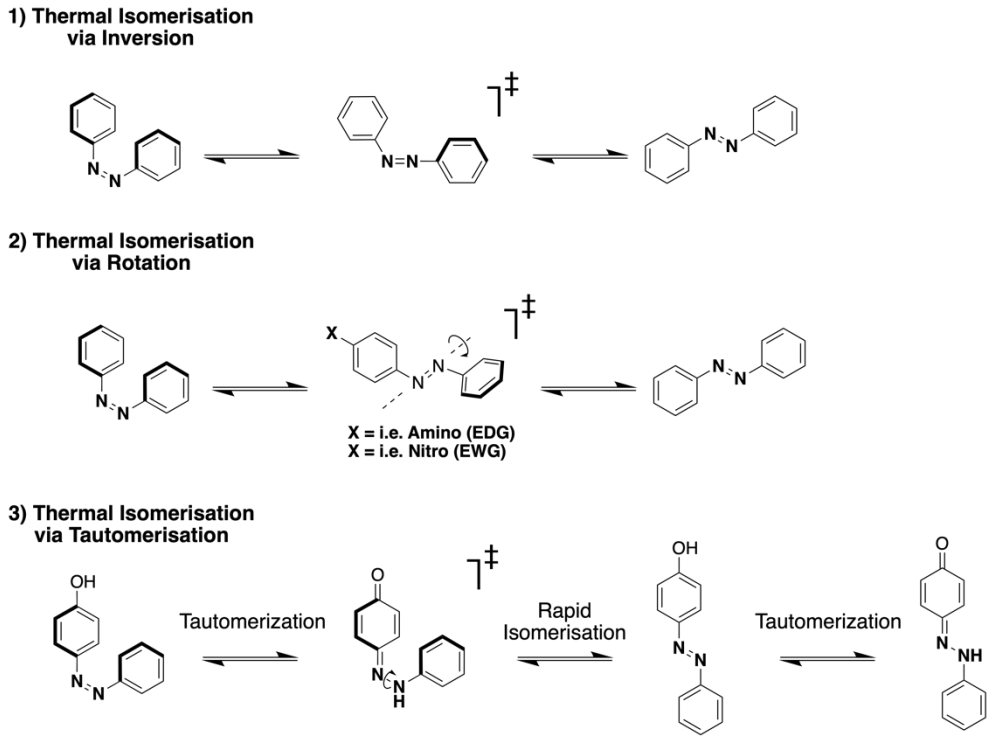
7.25.23 Cyanomethyl (*E*)-3-(4-((1-(2-(2-chloroacetamido)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)propanoate (41):



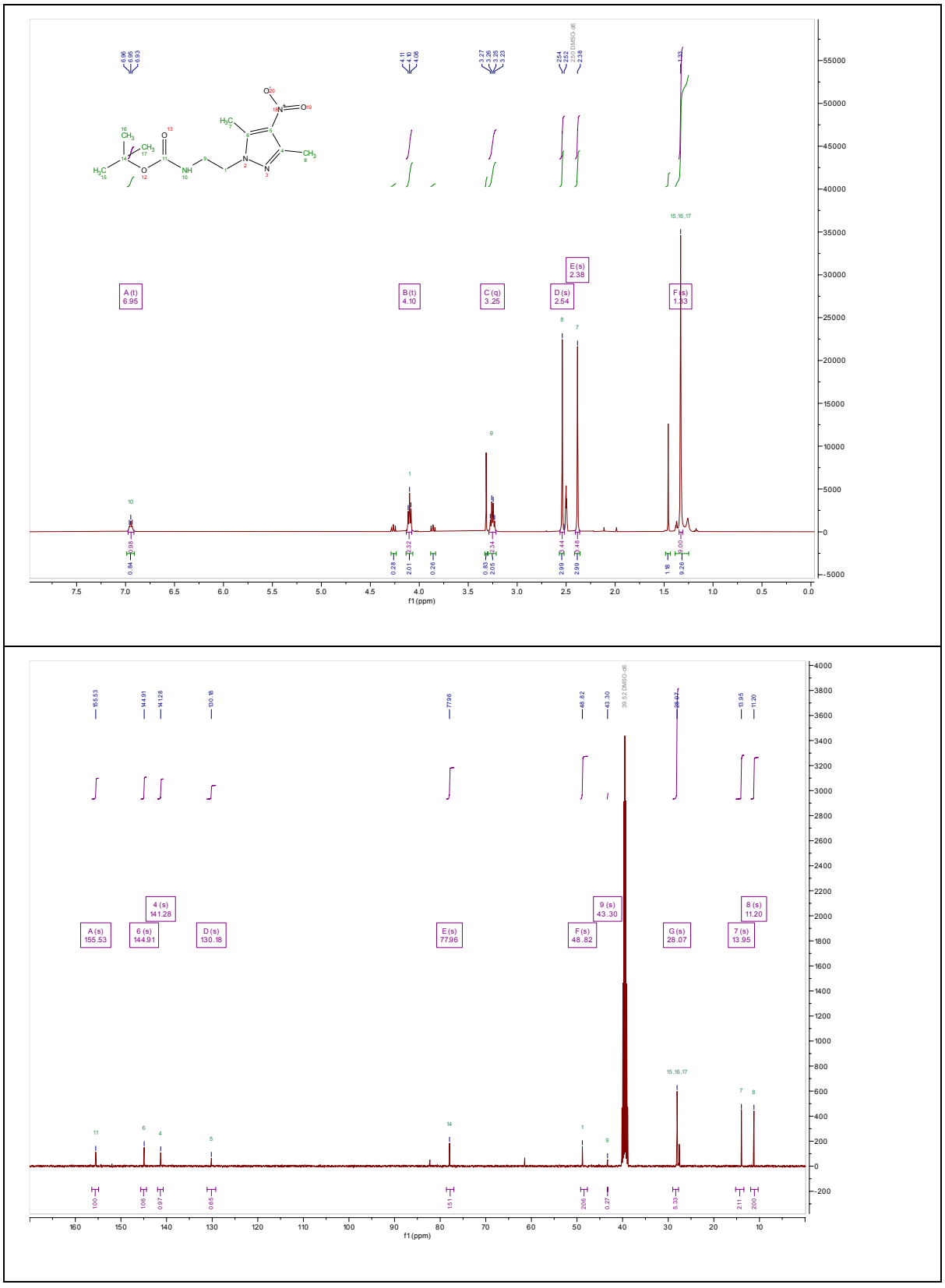
(*E*)-3-(4-((1-(2-(2-Chloroacetamido)ethyl)-3,5-dimethyl-1*H*-pyrazol-4-yl)diazenyl)phenyl)propanoic acid (47) (49 mg, 0.13 mmol, 1 eq.) was dissolved in anhydrous DMF (1.5 mL), followed by addition of potassium carbonate (8.4 mg, 0.2 mmol, 1.5 eq.) and chloroacetonitrile (10 μL, 0.6 mmol, 1.2 eq.). The reaction mixture was stirred at rt for 1 h. Following the reaction, monitored by LCMS, the solution was diluted with 1 M HCl (10 mL), transferred to a separatory funnel, and was extracted with ethyl acetate (3x 5 mL). The combined organic phases were then washed with a LiCl (5% w/v) solution (3x 10 mL) and extracted again with ethyl acetate (3x 5 mL). The combined organic phases were subsequently dried over magnesium sulphate, filtered, and concentrated *in vacuo* to afford the crude as an orange coloured solid. Purification by reverse-flash column chromatography (H_2O 0.1% TFA: MeCN 0.1% TFA, 10%-90%) afforded (41) as a yellow solid (13 mg, 23%). ^{1}H NMR (400 MHz, DMSO) δ 8.42 (t, *J* = 6.0 Hz, 1H, NH), 7.64 (m, 2H, 3′-CH, 5′-CH), 7.38 (m, 2H, 2′-CH, 6′-CH), 4.96 (s, 2H, CH_2CN), 4.13 (t, *J* = 6.0 Hz, 2H, $C^{1''}H_2$), 4.06 (s, 2H, CH_2Cl), 3.49 (dt, *J* = 6.2, 5.9 Hz, 2H, $C^{2''}H_2$), 2.94 (t, *J* = 7.2 Hz, 2H, $C^{5''}H_2$), 2.80 (t, *J* = 7.2, 1.0 Hz, 2H,$C^{6''}H_2$), 2.53 (s, 3H, C^3CH_3), 2.38 (s, 3H, C^5CH_3). ^{13}C NMR (101 MHz, DMSO) δ ^{13}C NMR (101 MHz, DMSO) δ 171.3; $C^{7''}$, 166.4; $C^{3''}$, 151.6; $C^{4'}$, 141.8; C^5, 140.8; $C^{1'}$, 139.9; C^3, 134.3; C^4, 129.1; $C^{2'}C^{6'}$, 121.4; $C^{3'}C^{5'}$, 116.0; CN, 48.9; CH_2CN, 47.3; $C^{1''}$, 42.5; CH_2CL, 39.0; $C^{2''}$, 34.1; $C^{6''}$, 29.7; $C^{5''}$, 14.0; C^5CH_3, 9.3; C^3CH_3. LCMS ESI$^+$: Gradient: Method 1, Calc. Mass $C_{20}H_{23}Cl^{35}N_6O_3$ $[MH]^+$ 430.15, Found at 431.2 $[MH]^+$, rt. 2.15 min.

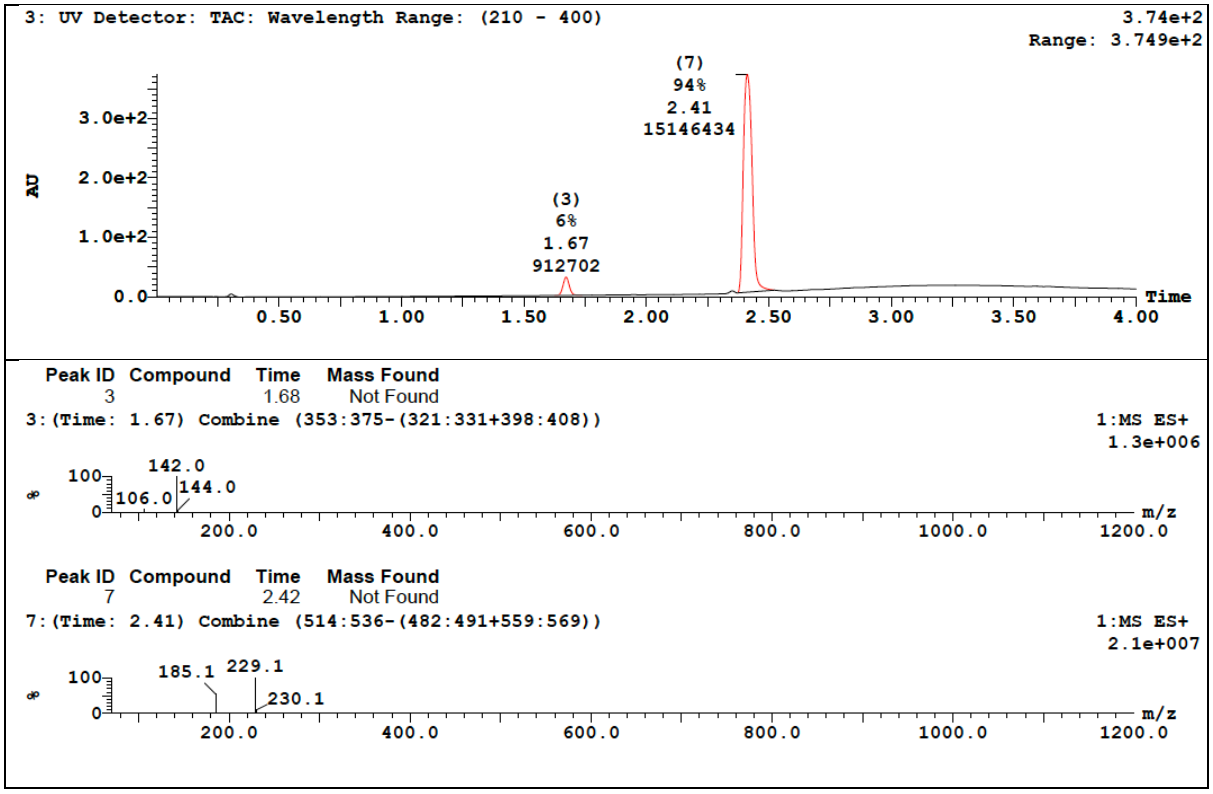
Appendix

Appendix 1: The three dominant mechanisms of thermal reversion displayed for the azo based class using azobenzene as the representative model.

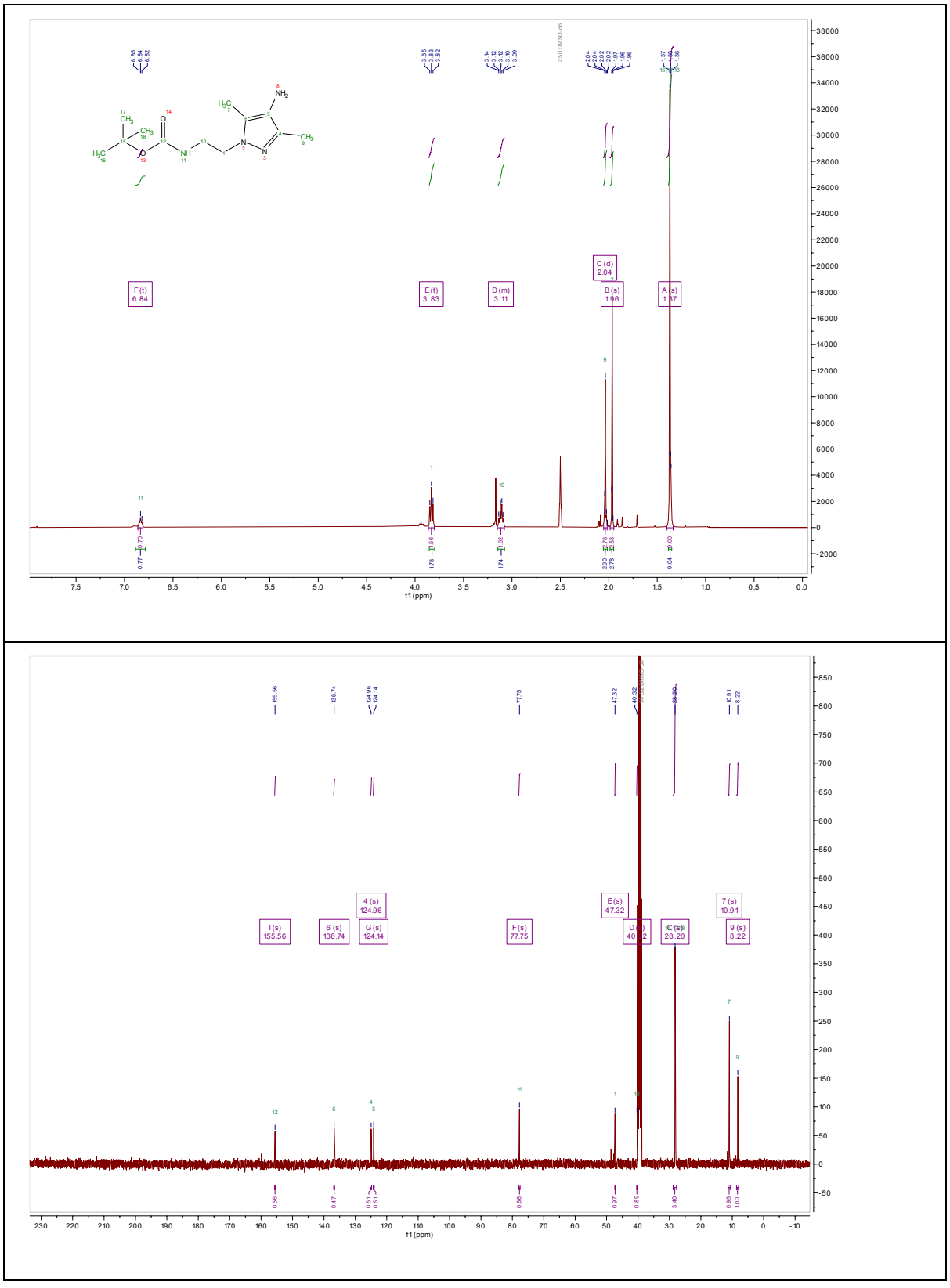


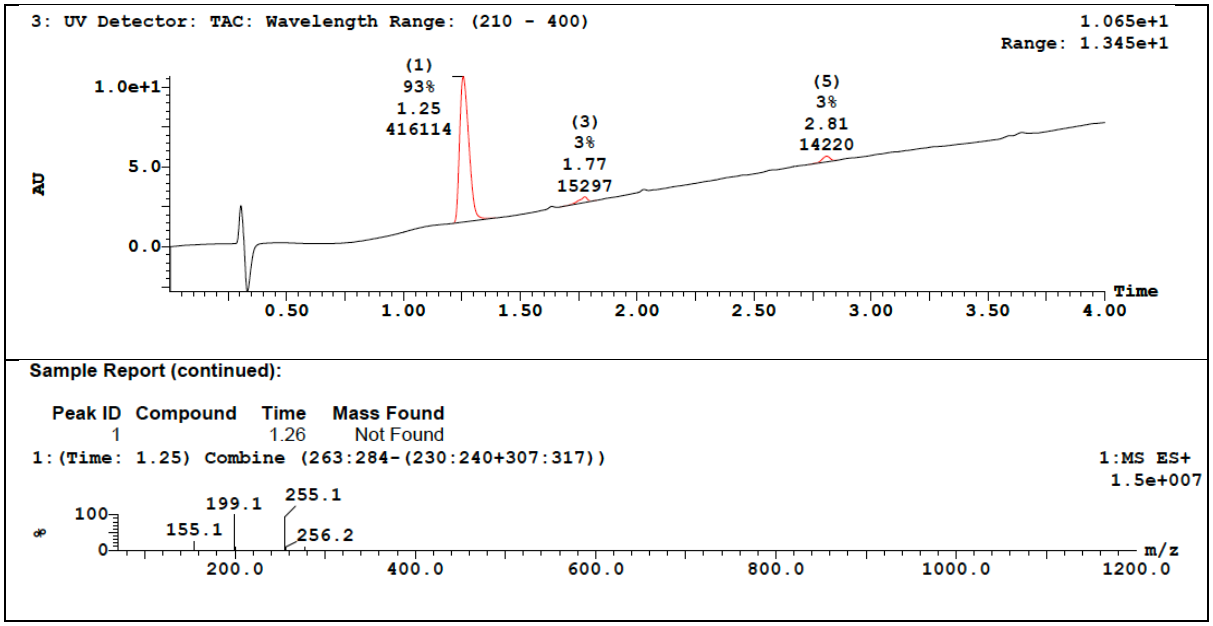
Appendix 2: tert-Butyl (2-(3,5-dimethyl-4-nitro-1H-pyrazol-1-yl)ethyl)carbamate (13)




3: UV Detector: TAC: Wavelength Range: (210 - 400)
3.74e+2
Range: 3.749e+2
(7)
94%
2.41
15146434
(3)
6%
1.67
912702
AU
3.0e+2
2.0e+2
1.0e+2
0.0
0.50
1.00
1.50
2.00
2.50
3.00
3.50
4.00
Time
Peak ID Compound Time Mass Found
3 1.68 Not Found
3:(Time: 1.67) Combine (353:375-(321:331+398:408))
1:MS ES+
1.3e+006
142.0
144.0
106.0
100
%
0
200.0
400.0
600.0
800.0
1000.0
1200.0
m/z
Peak ID Compound Time Mass Found
7 2.42 Not Found
7:(Time: 2.41) Combine (514:536-(482:491+559:569))
1:MS ES+
2.1e+007
185.1
229.1
230.1
100
%
0
200.0
400.0
600.0
800.0
1000.0
1200.0
m/z

Appendix 3: 6.28 tert-Butyl (2-(4-amino-3,5-dimethyl-1H-pyrazol-1-yl)ethyl)carbamate (14)

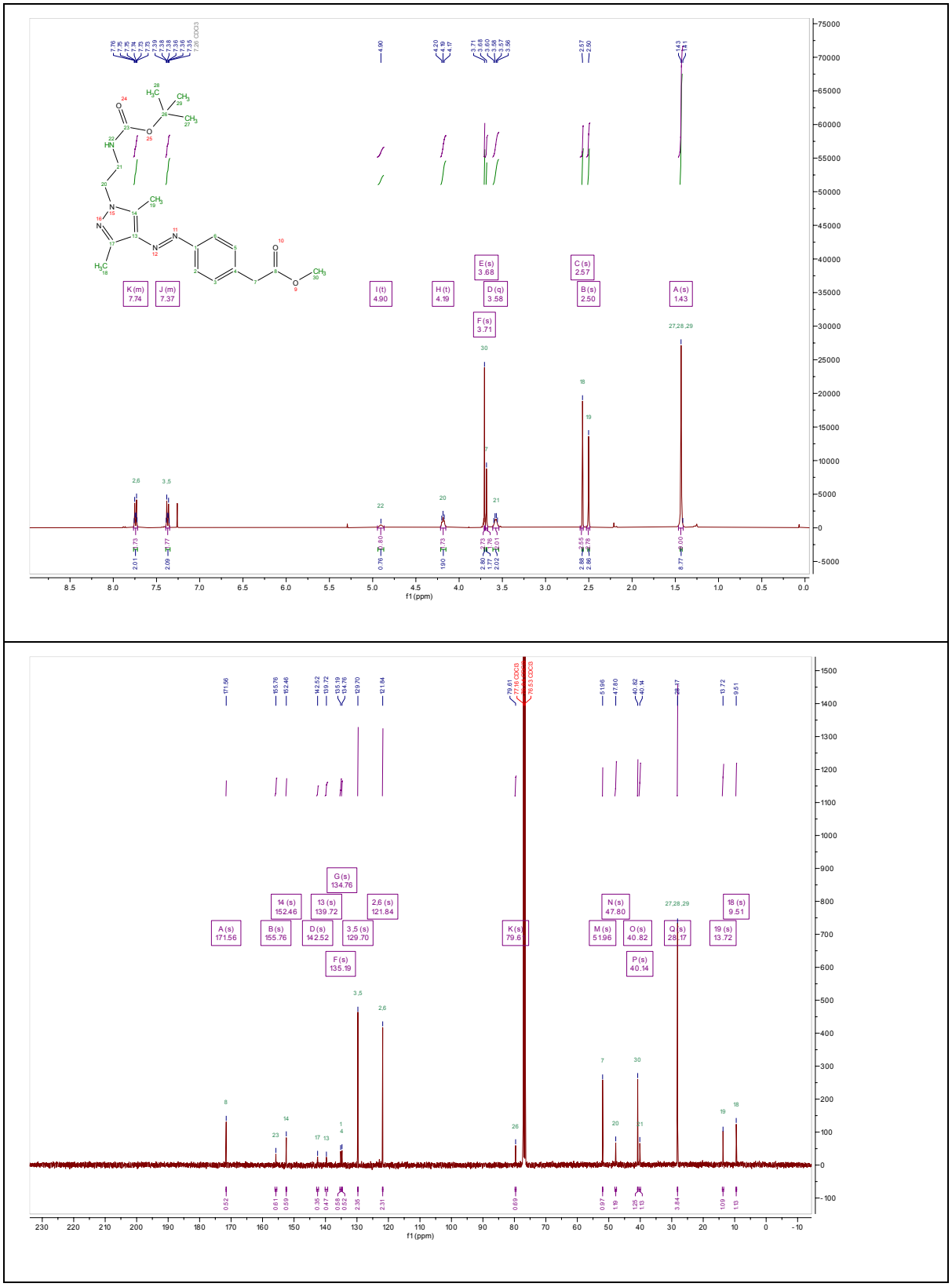


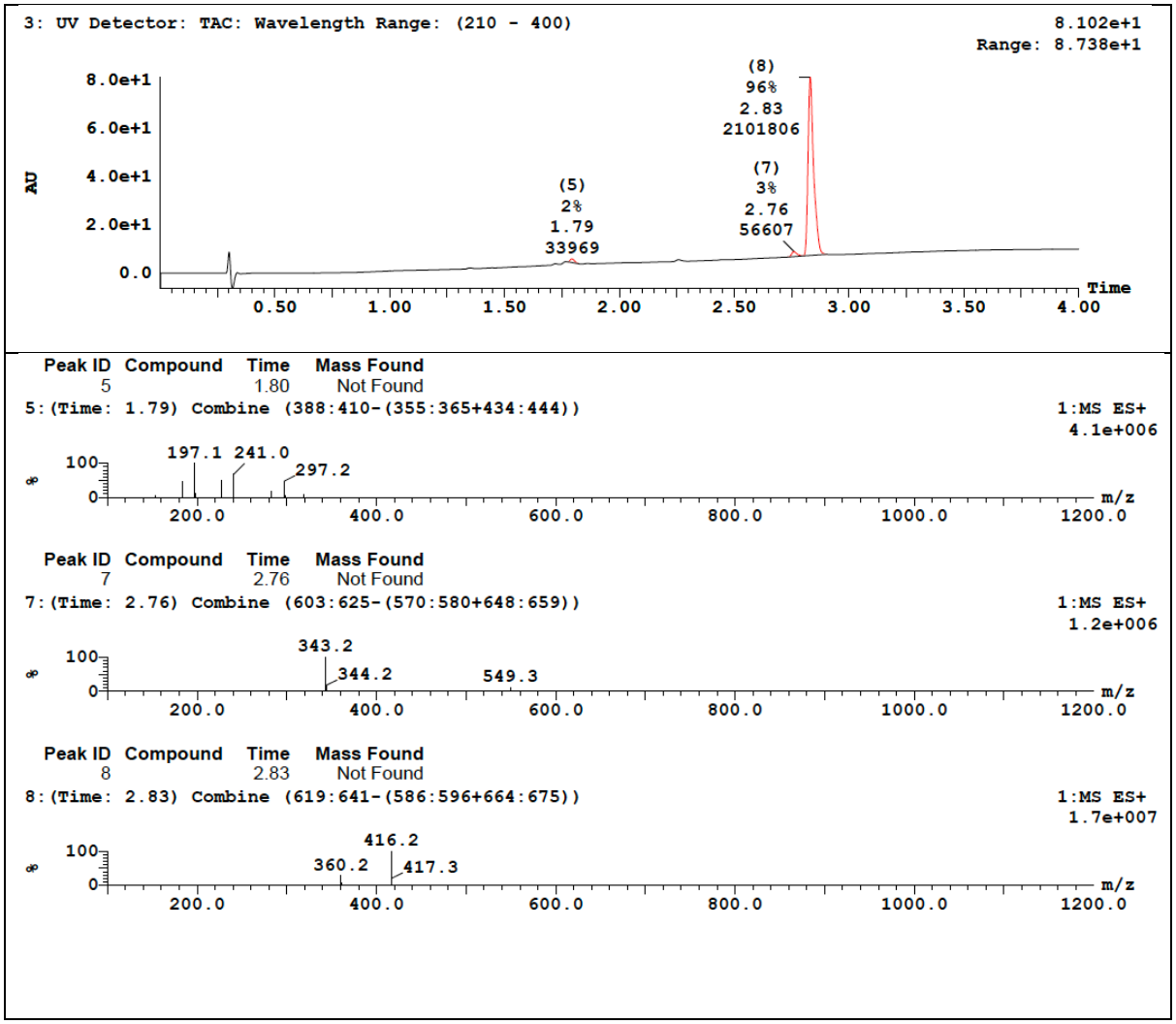


Sample Report (continued):

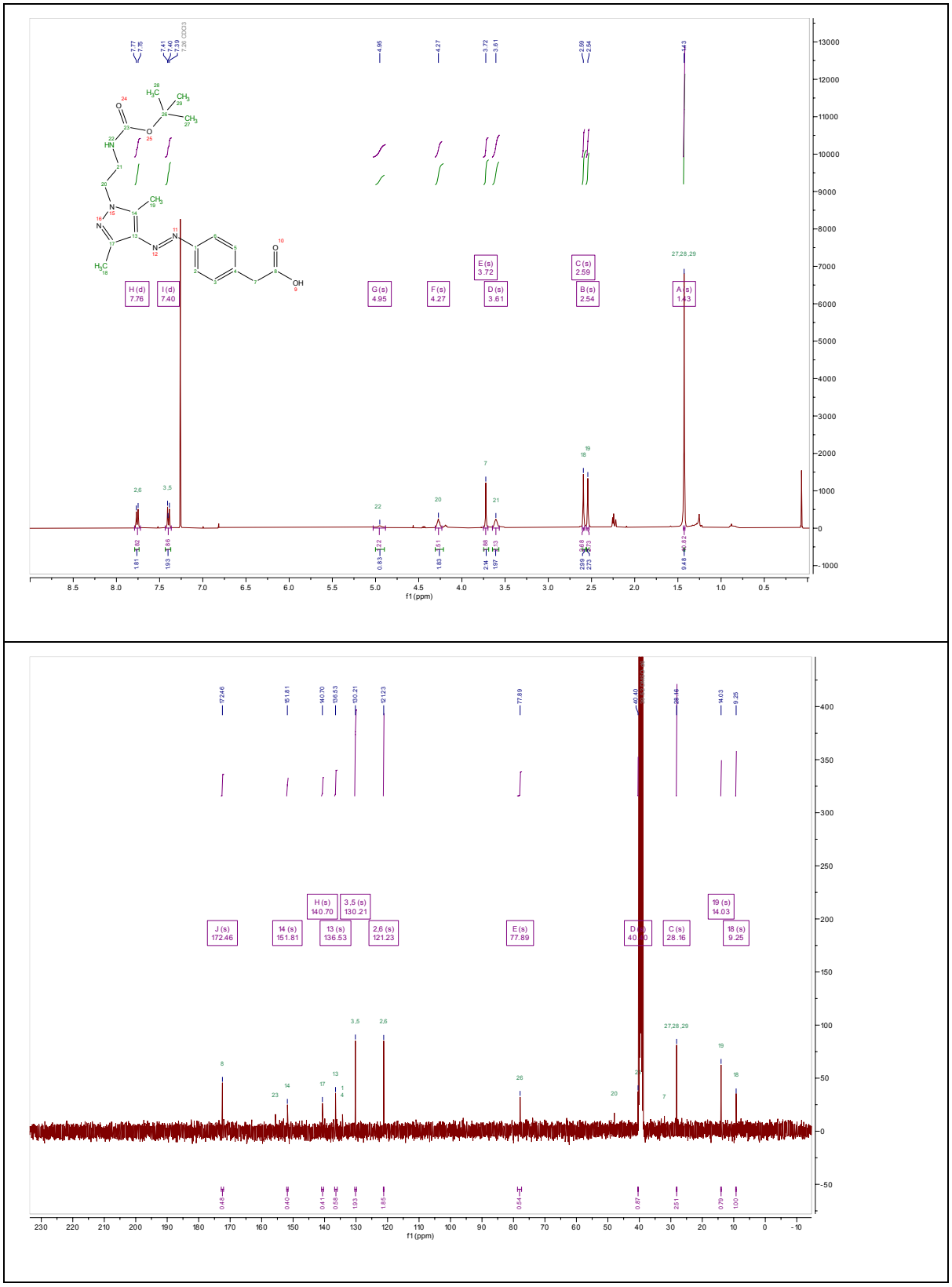
Peak ID	Compound	Time	Mass Found
1		1.26	Not Found

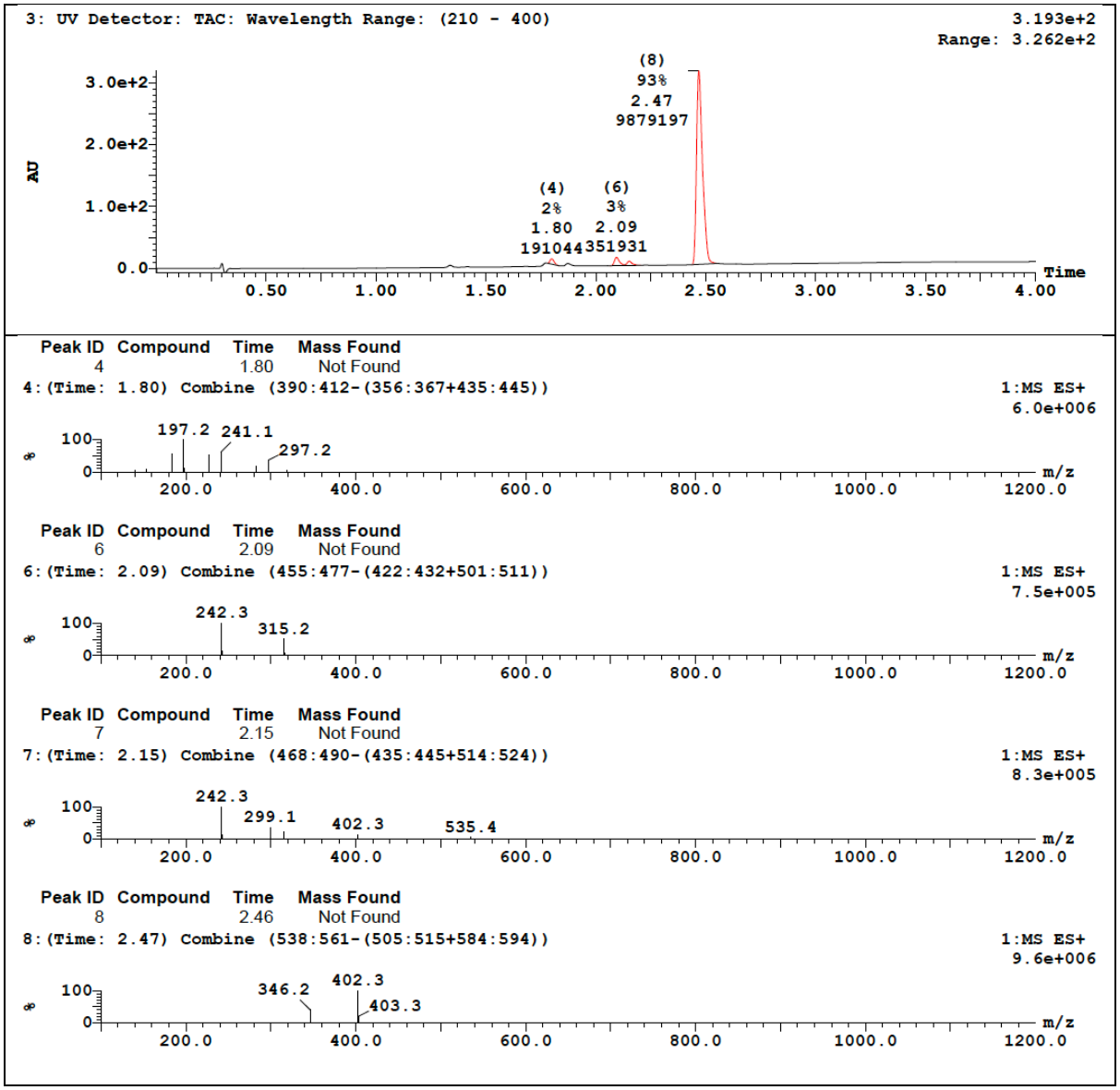
Appendix 4: Methyl I-2-(4-((1-(2-((tert-butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl) iazinyl)phenyl)acetate (17)



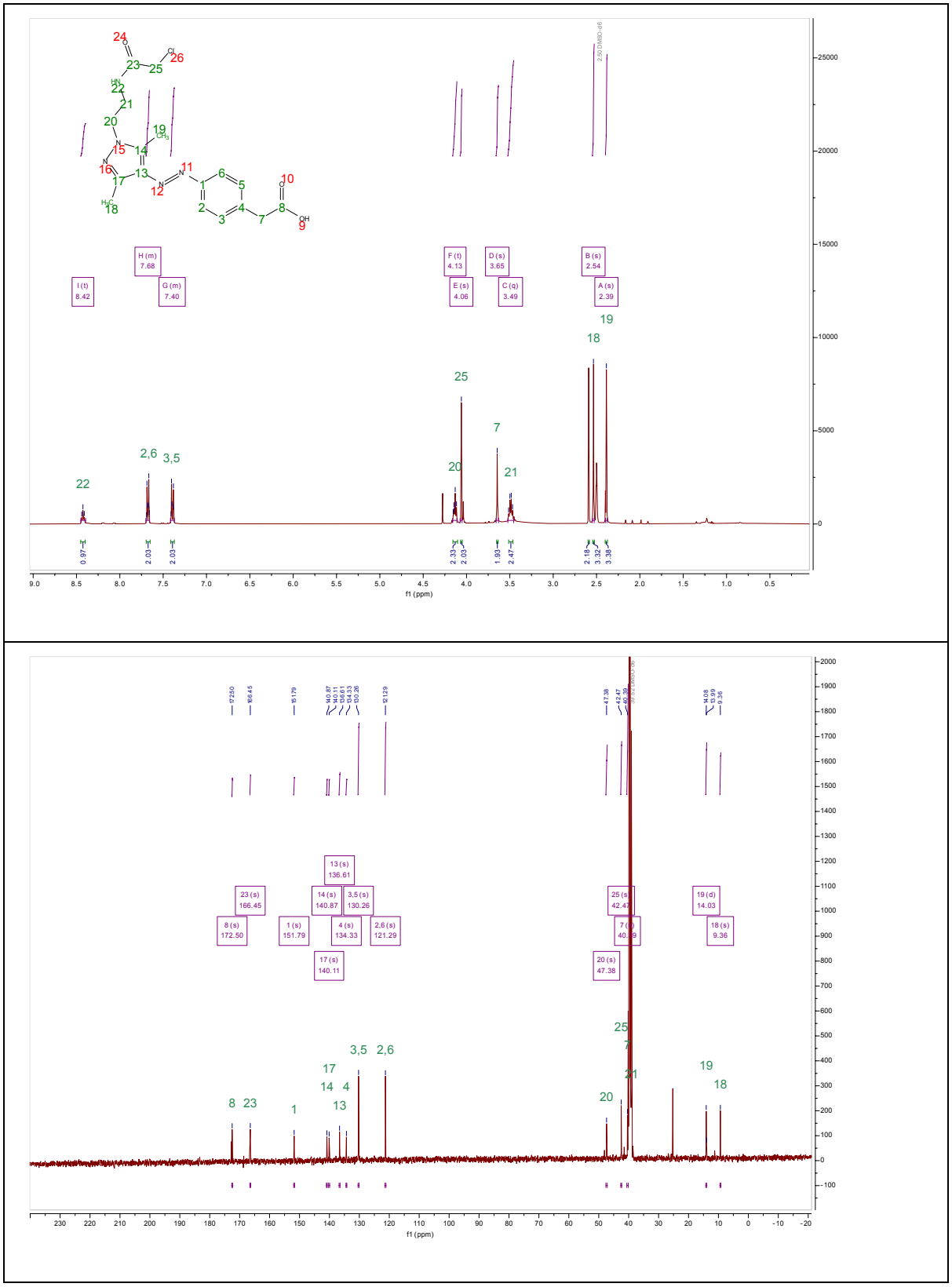

3: UV Detector: TAC: Wavelength Range: (210 - 400)
8.102e+1
Range: 8.738e+1
AU
8.0e+1
6.0e+1
4.0e+1
2.0e+1
0.0
(5)
2%
1.79
33969
(8)
96%
2.83
2101806
(7)
3%
2.76
56607
0.50
1.00
1.50
2.00
2.50
3.00
3.50
4.00
Time
Peak ID Compound Time Mass Found
5 1.80 Not Found
5:(Time: 1.79) Combine (388:410-(355:365+434:444))
1:MS ES+
4.1e+006
197.1
241.0
297.2
100
%
0
200.0
400.0
600.0
800.0
1000.0
1200.0
m/z
Peak ID Compound Time Mass Found
7 2.76 Not Found
7:(Time: 2.76) Combine (603:625-(570:580+648:659))
1:MS ES+
1.2e+006
343.2
344.2
549.3
100
%
0
200.0
400.0
600.0
800.0
1000.0
1200.0
m/z
Peak ID Compound Time Mass Found
8 2.83 Not Found
8:(Time: 2.83) Combine (619:641-(586:596+664:675))
1:MS ES+
1.7e+007
416.2
360.2
417.3
100
%
0
200.0
400.0
600.0
800.0
1000.0
1200.0
m/z

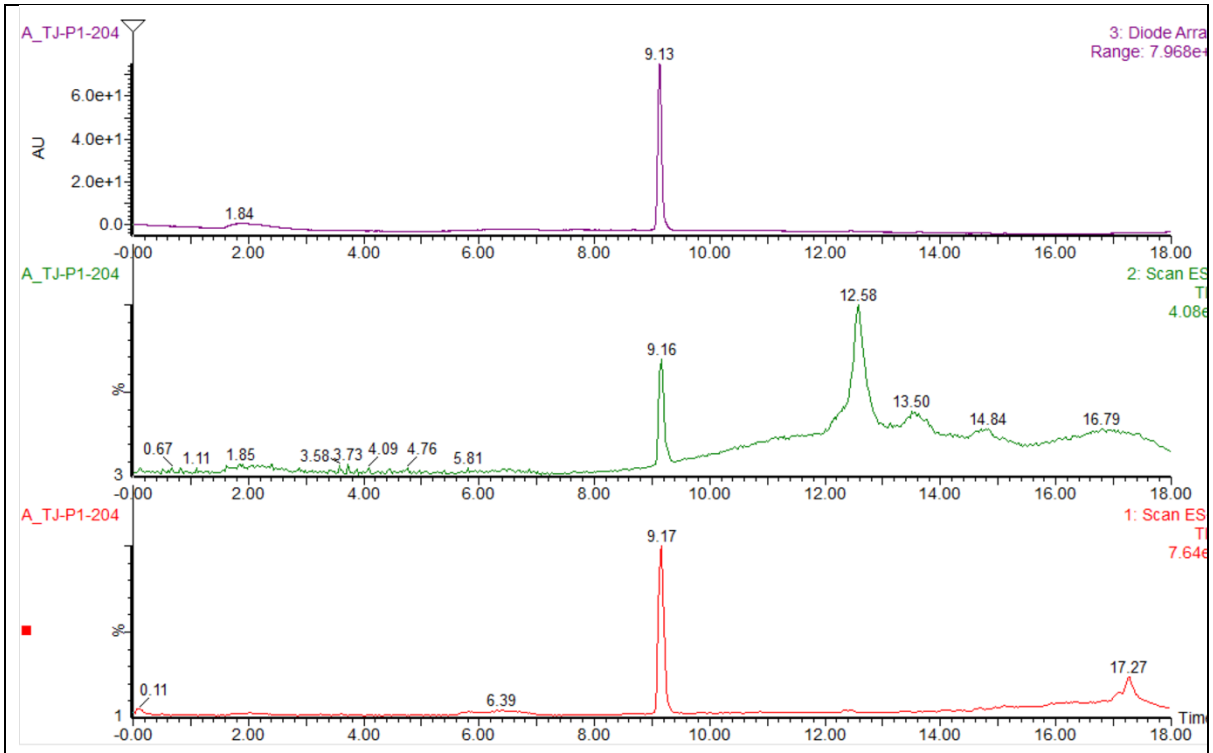
Appendix 5: (E)-2-(4-((1-(2-((tert-Butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl) acetic acid (18)

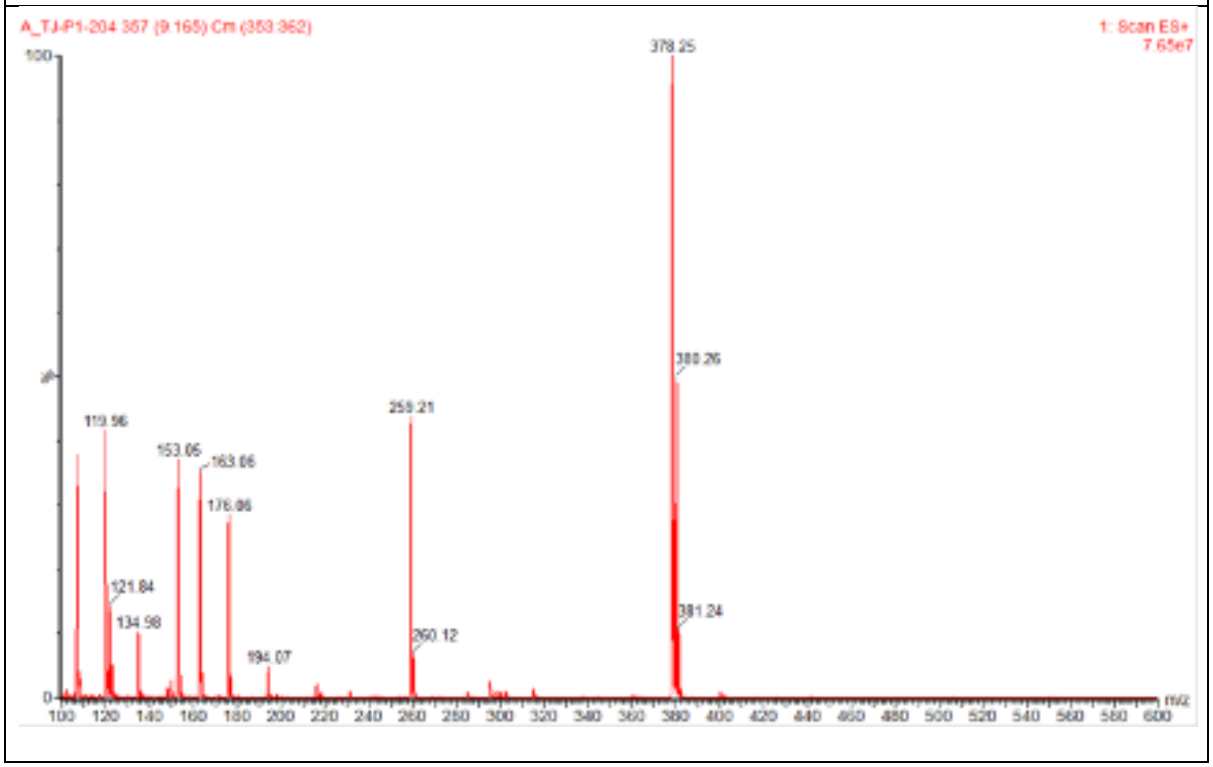



3: UV Detector: TAC: Wavelength Range: (210 - 400)
3.193e+2
Range: 3.262e+2
(8)
93%
2.47
9879197
(4)
2%
1.80
191044
(6)
3%
2.09
351931
AU
Time
Peak ID Compound Time Mass Found
4 1.80 Not Found
4:(Time: 1.80) Combine (390:412-(356:367+435:445))
1:MS ES+
6.0e+006
197.2
241.1
297.2
Peak ID Compound Time Mass Found
6 2.09 Not Found
6:(Time: 2.09) Combine (455:477-(422:432+501:511))
1:MS ES+
7.5e+005
242.3
315.2
Peak ID Compound Time Mass Found
7 2.15 Not Found
7:(Time: 2.15) Combine (468:490-(435:445+514:524))
1:MS ES+
8.3e+005
242.3
299.1
402.3
535.4
Peak ID Compound Time Mass Found
8 2.46 Not Found
8:(Time: 2.47) Combine (538:561-(505:515+584:594))
1:MS ES+
9.6e+006
346.2
402.3
403.3
m/z

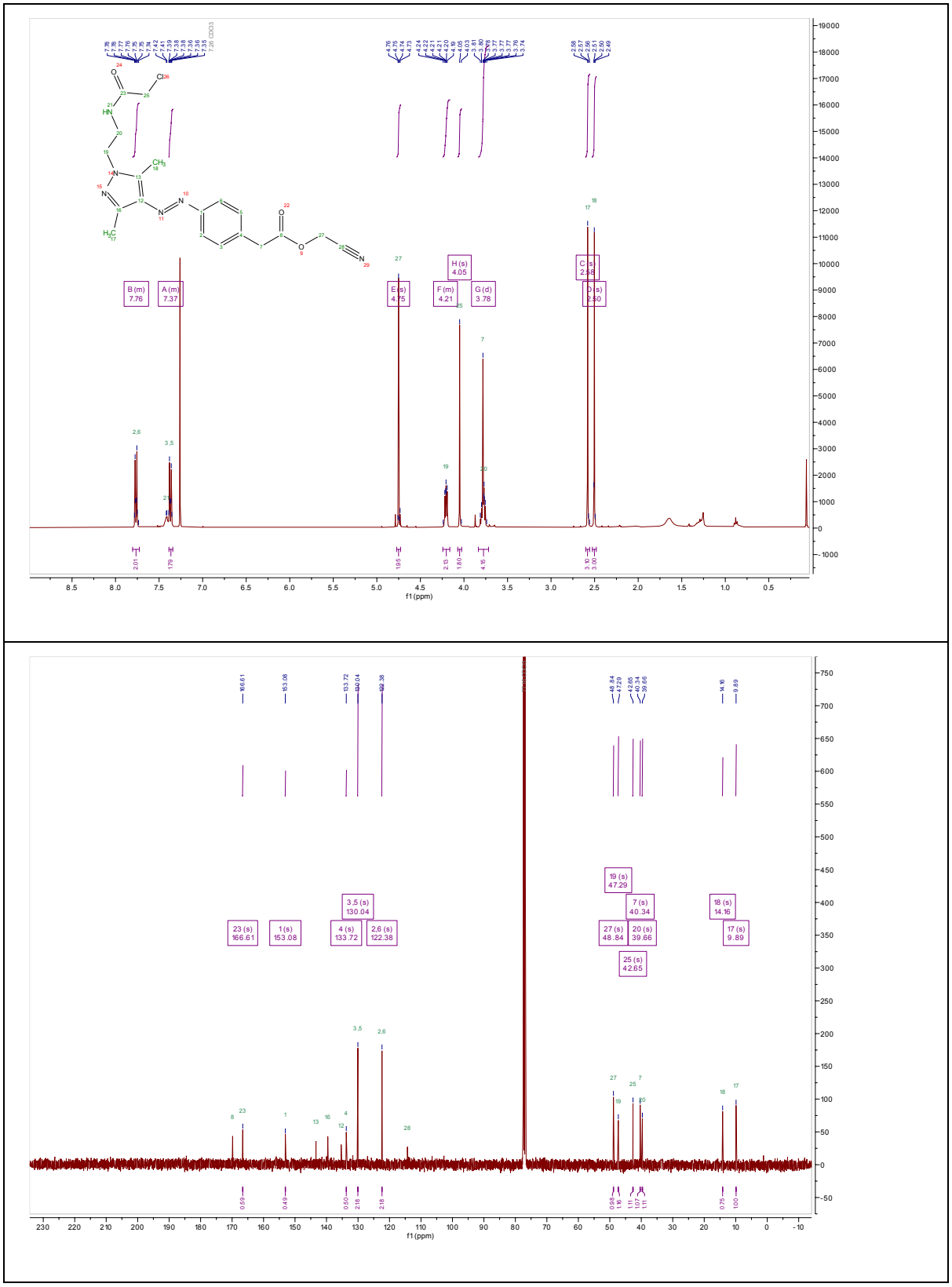
Appendix 6: (E)-2-(4-((1-(2-(2-Chloroacetamido)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl) phenyl)acetic acid (20)

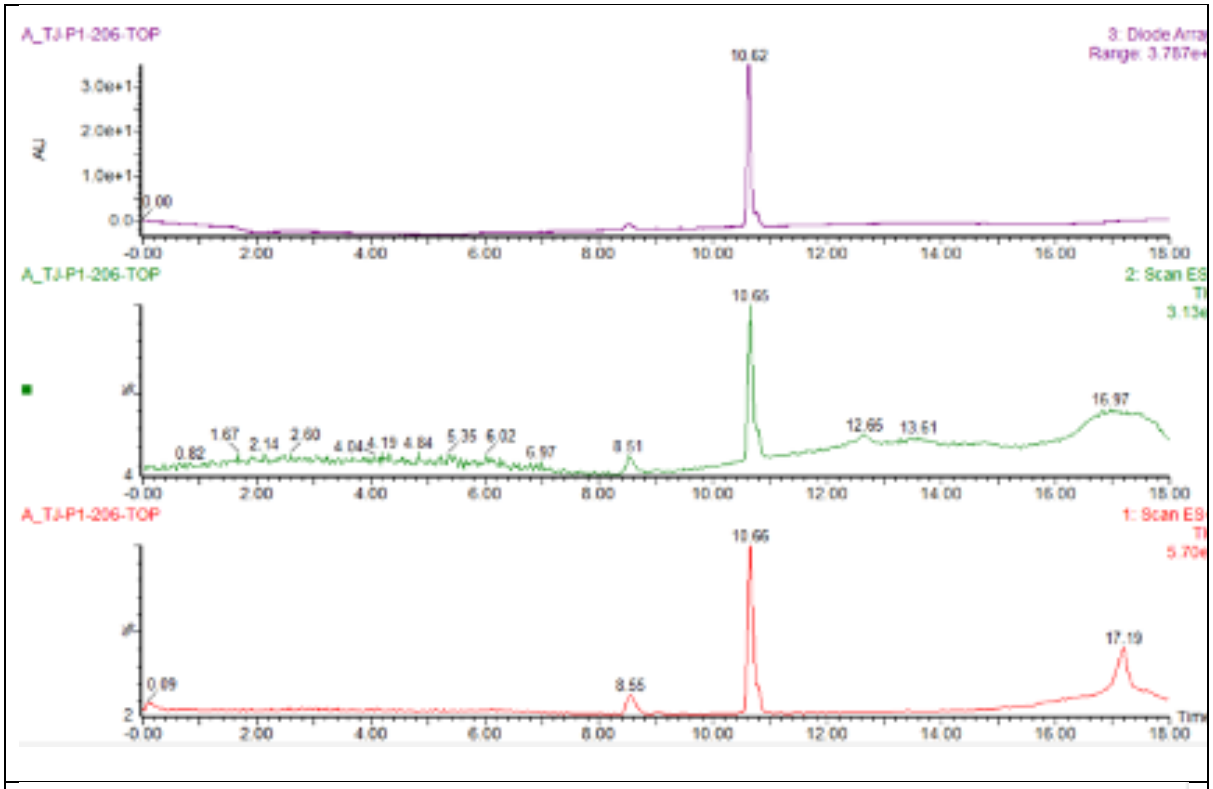


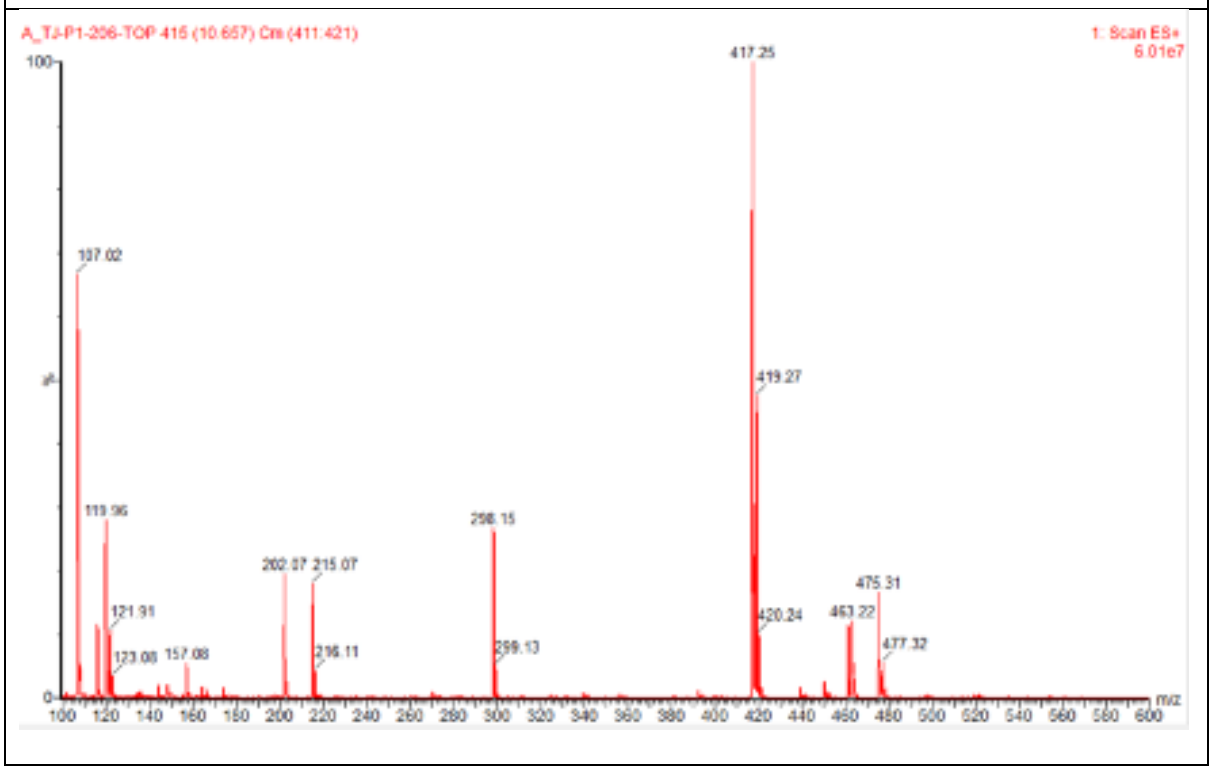

A_TJ-P1-204
3: Diode Arra
Range: 7.968e+
AU
6.0e+1
4.0e+1
2.0e+1
0.0
9.13
1.84
A_TJ-P1-204
2: Scan ES
TI
4.08e
%
12.58
9.16
13.50
14.84
16.79
0.67
1.11
1.85
3.58
3.73
4.09
4.76
5.81
A_TJ-P1-204
1: Scan ES
TI
7.64e
%
9.17
17.27
0.11
6.39
-0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00


A_TJ-P1-204 357 (9.165) Cm (353:362)
1: Scan ES+
7.65e7
100
%
0
378.25
380.26
259.21
119.96
153.05
163.06
176.06
121.84
134.98
381.24
260.12
194.07
100
120
140
160
180
200
220
240
260
280
300
320
340
360
380
400
420
440
460
480
500
520
540
560
580
600
m/z

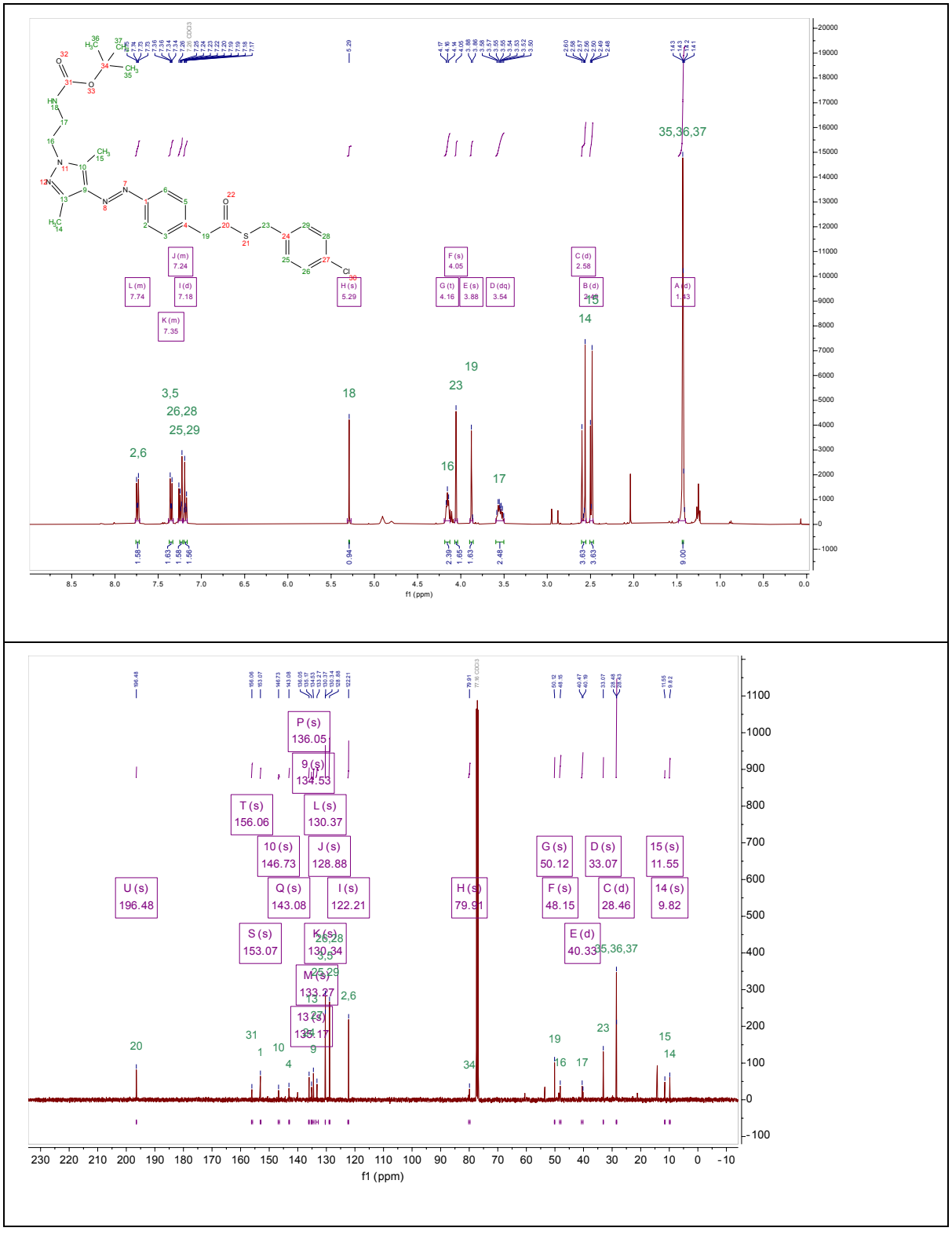
Appendix 7: Cyanomethyl (E)-2-(4-((1-(2-(2-chloroacetamido)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)acetate (11)

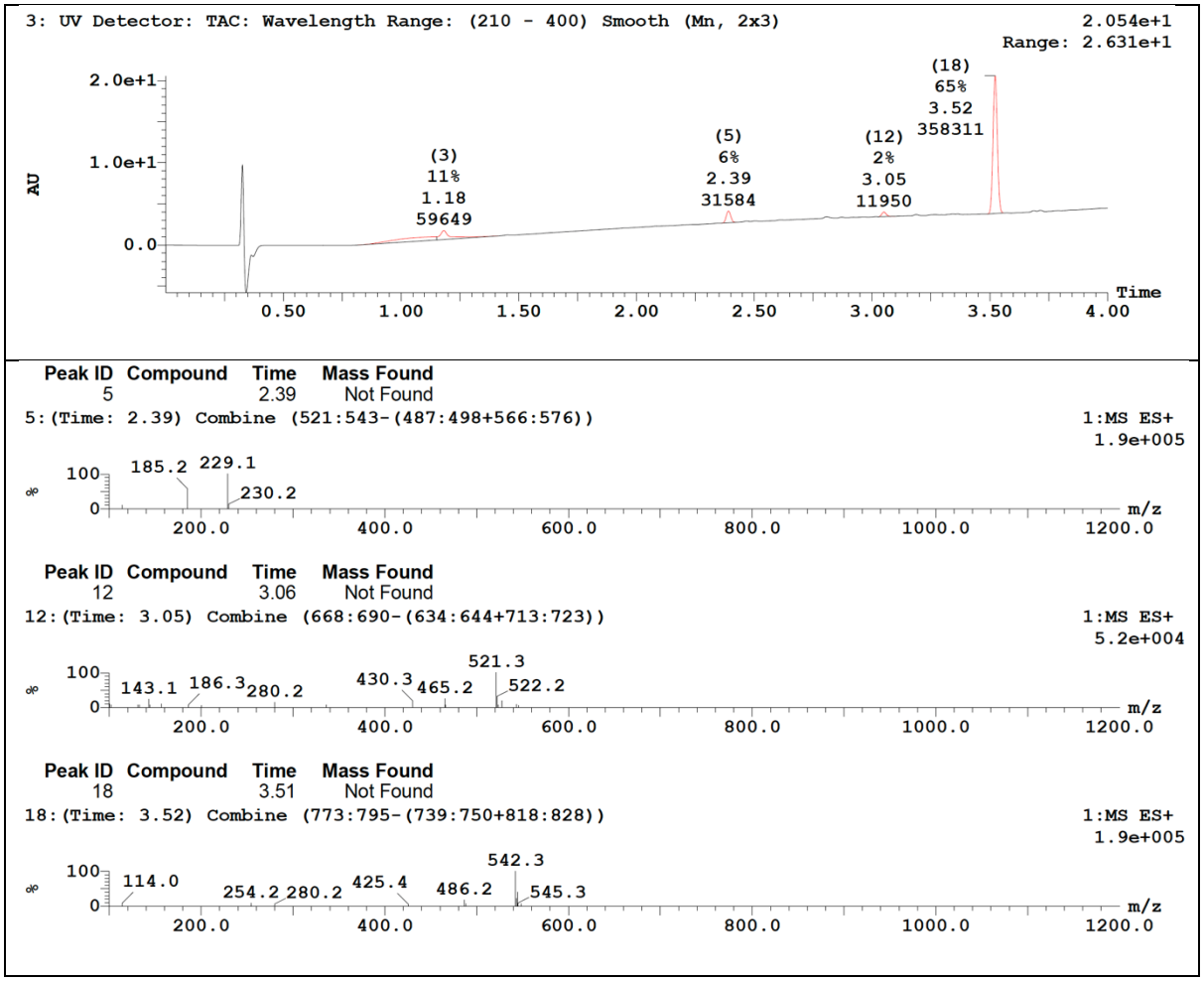



A_TJ-P1-206-TOP
3: Diode Array
Range: 3.757e+
10.62
AU
3.0e+1
2.0e+1
1.0e+1
0.0
0.00
-0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
A_TJ-P1-206-TOP
2: Scan ES
TI
3.13e
10.65
16.97
12.65
13.61
1.67
2.14
2.60
4.04
4.19
4.84
5.35
6.02
6.97
8.51
0.82
A_TJ-P1-206-TOP
1: Scan ES
TI
5.70e
10.66
17.19
8.55
0.89
Time

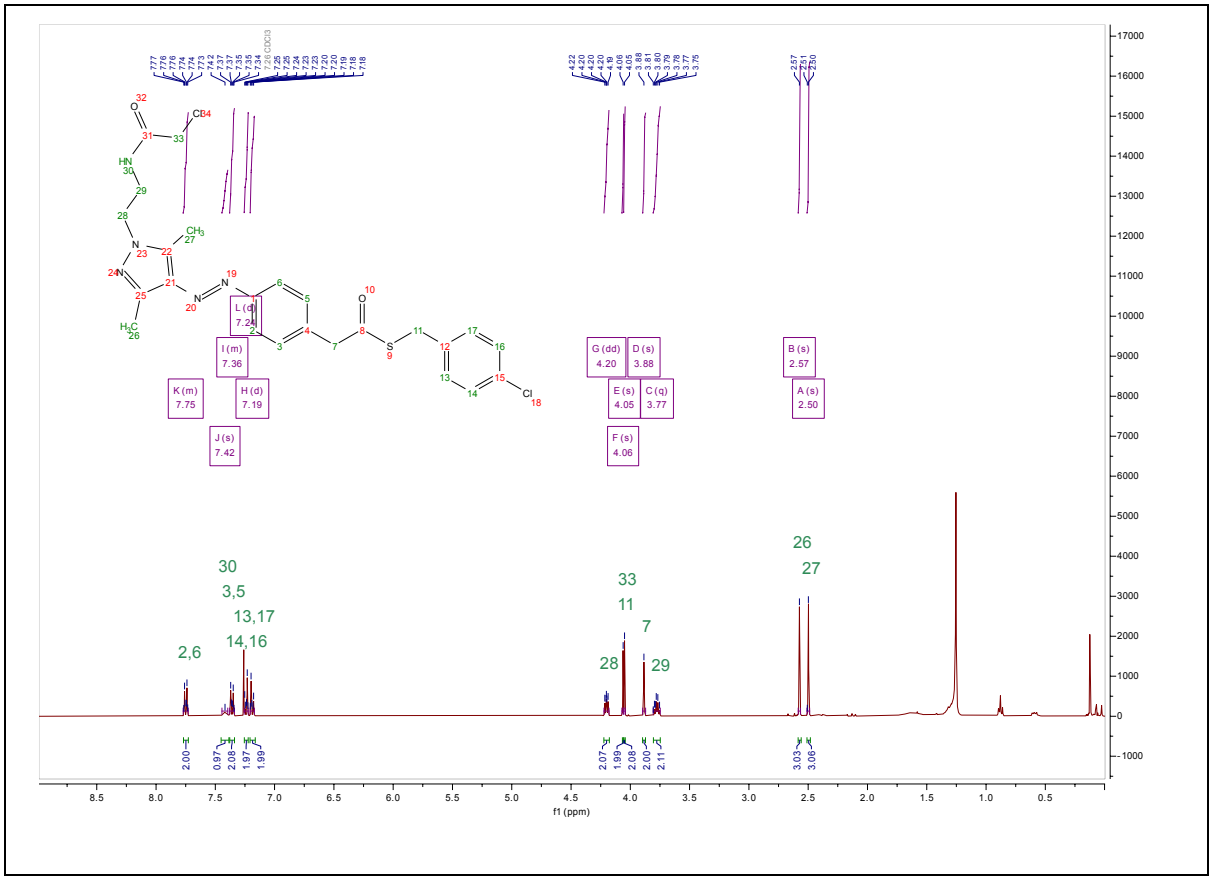

A_TJ-P1-206-TOP 416 (10.657) Cm (411:421)
1: Scan ES+
6.01e7
100
%
0
417.25
419.27
137.02
119.96
121.91
123.08
157.08
202.07
215.07
216.11
258.15
259.13
420.24
463.22
475.31
477.32
m/z
100
120
140
160
180
200
220
240
260
280
300
320
340
360
380
400
420
440
460
480
500
520
540
560
580
600

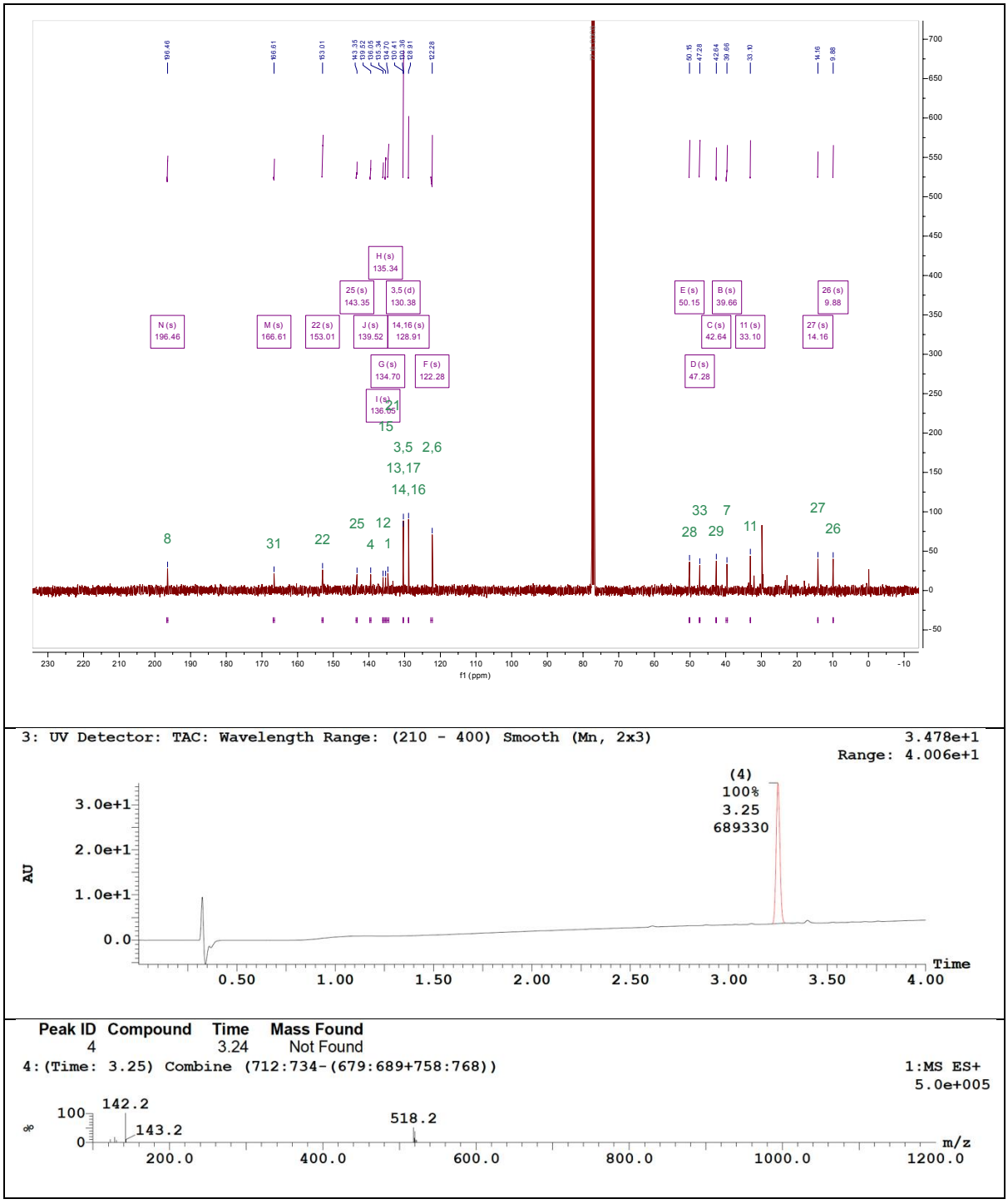
Appendix 8: S-(4-Chlorobenzyl) (E)-2-(4-((1-(2-((tert-butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)ethanethioate (22)



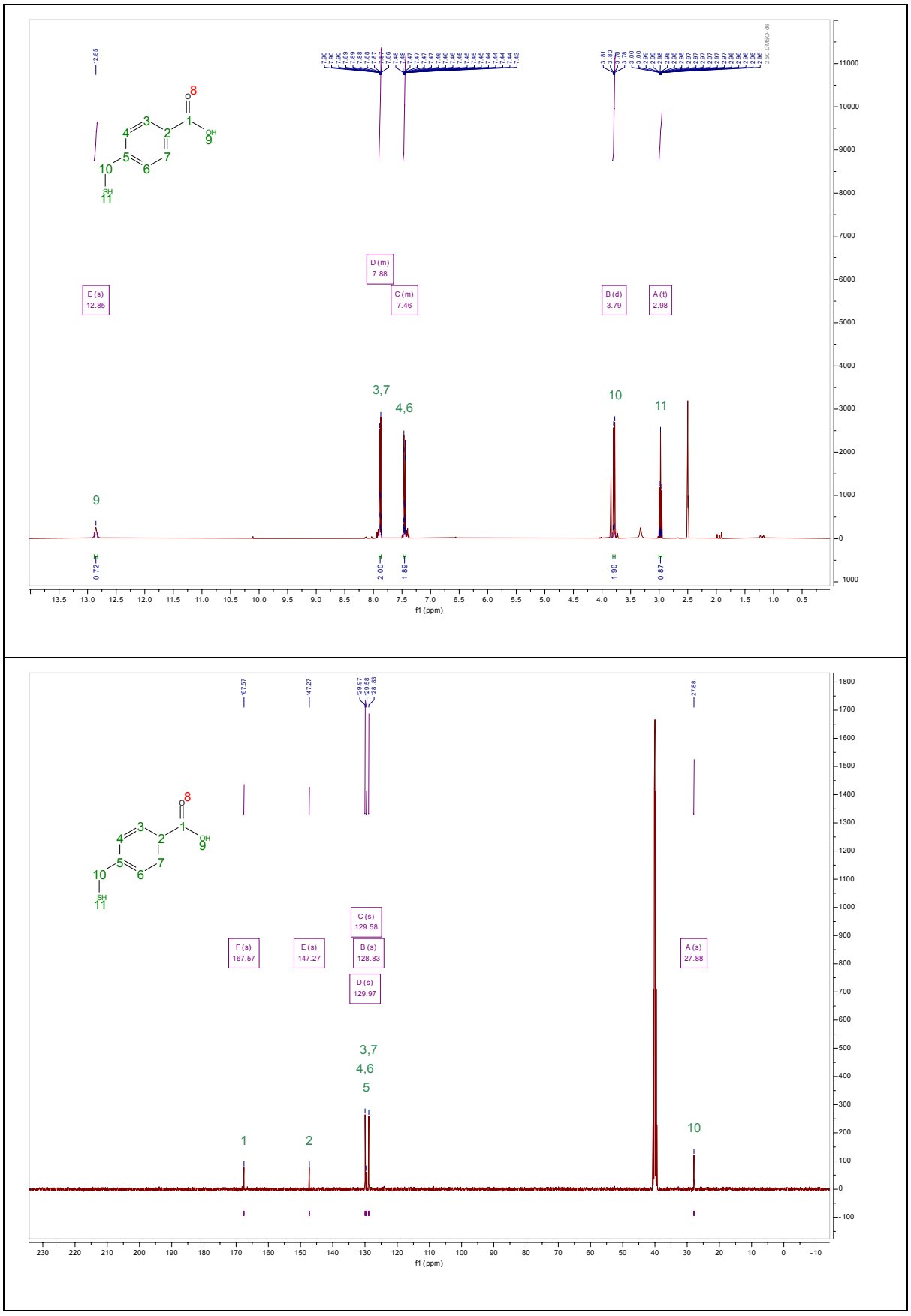

3: UV Detector: TAC: Wavelength Range: (210 - 400) Smooth (Mn, 2x3)
2.054e+1
Range: 2.631e+1
AU
2.0e+1
1.0e+1
0.0
(3)
11%
1.18
59649
(5)
6%
2.39
31584
(12)
2%
3.05
11950
(18)
65%
3.52
358311
0.50
1.00
1.50
2.00
2.50
3.00
3.50
4.00
Time
Peak ID Compound Time Mass Found
5 2.39 Not Found
5:(Time: 2.39) Combine (521:543-(487:498+566:576))
1:MS ES+
1.9e+005
185.2
229.1
230.2
100
%
0
200.0
400.0
600.0
800.0
1000.0
1200.0
m/z
Peak ID Compound Time Mass Found
12 3.06 Not Found
12:(Time: 3.05) Combine (668:690-(634:644+713:723))
1:MS ES+
5.2e+004
143.1
186.3
280.2
430.3
465.2
521.3
522.2
Peak ID Compound Time Mass Found
18 3.51 Not Found
18:(Time: 3.52) Combine (773:795-(739:750+818:828))
1:MS ES+
1.9e+005
114.0
254.2
280.2
425.4
486.2
542.3
545.3

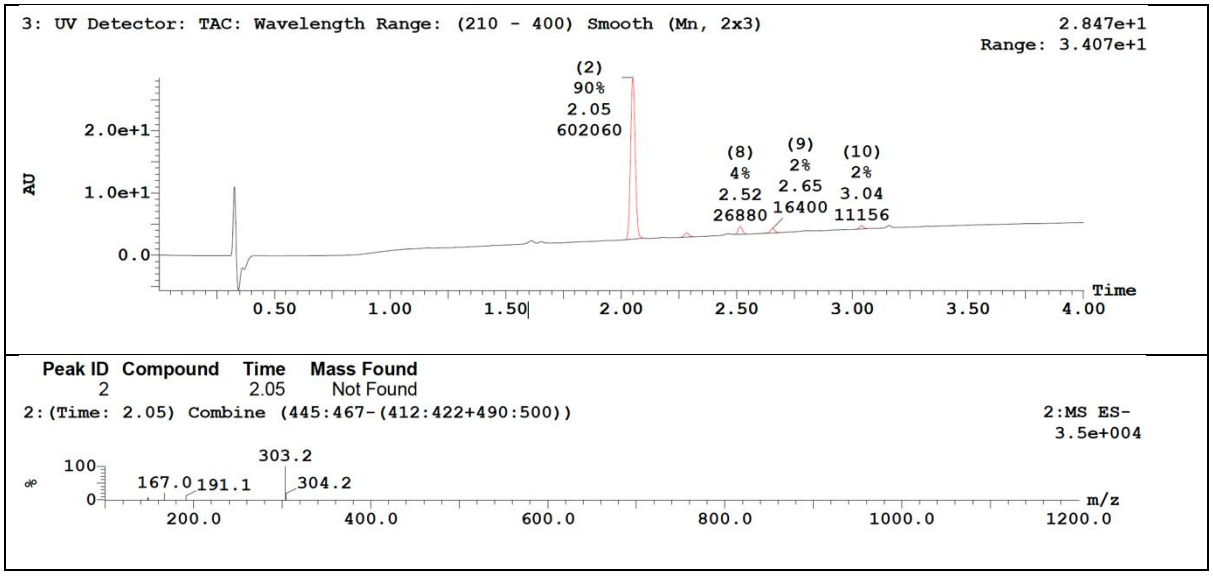
Appendix 9: S-(4-Chlorobenzyl) (E)-2-(4-((1-(2-(2-chloroacetamido)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)ethanethioate (21)




N (s) 196.46
M (s) 166.61
22 (s) 153.01
J (s) 139.52
14,16 (s) 128.91
H (s) 135.34
25 (s) 143.35
3,5 (d) 130.38
G (s) 134.70
F (s) 122.28
I (s) 136.05
E (s) 50.15
B (s) 39.66
C (s) 42.64
11 (s) 33.10
D (s) 47.28
26 (s) 9.88
27 (s) 14.16
f1 (ppm)
3: UV Detector: TAC: Wavelength Range: (210 - 400) Smooth (Mn, 2x3)
3.478e+1
Range: 4.006e+1
(4)
100%
3.25
689330
AU
Time
Peak ID Compound Time Mass Found
4 3.24 Not Found
4:(Time: 3.25) Combine (712:734-(679:689+758:768))
1:MS ES+
5.0e+005
142.2
143.2
518.2
m/z

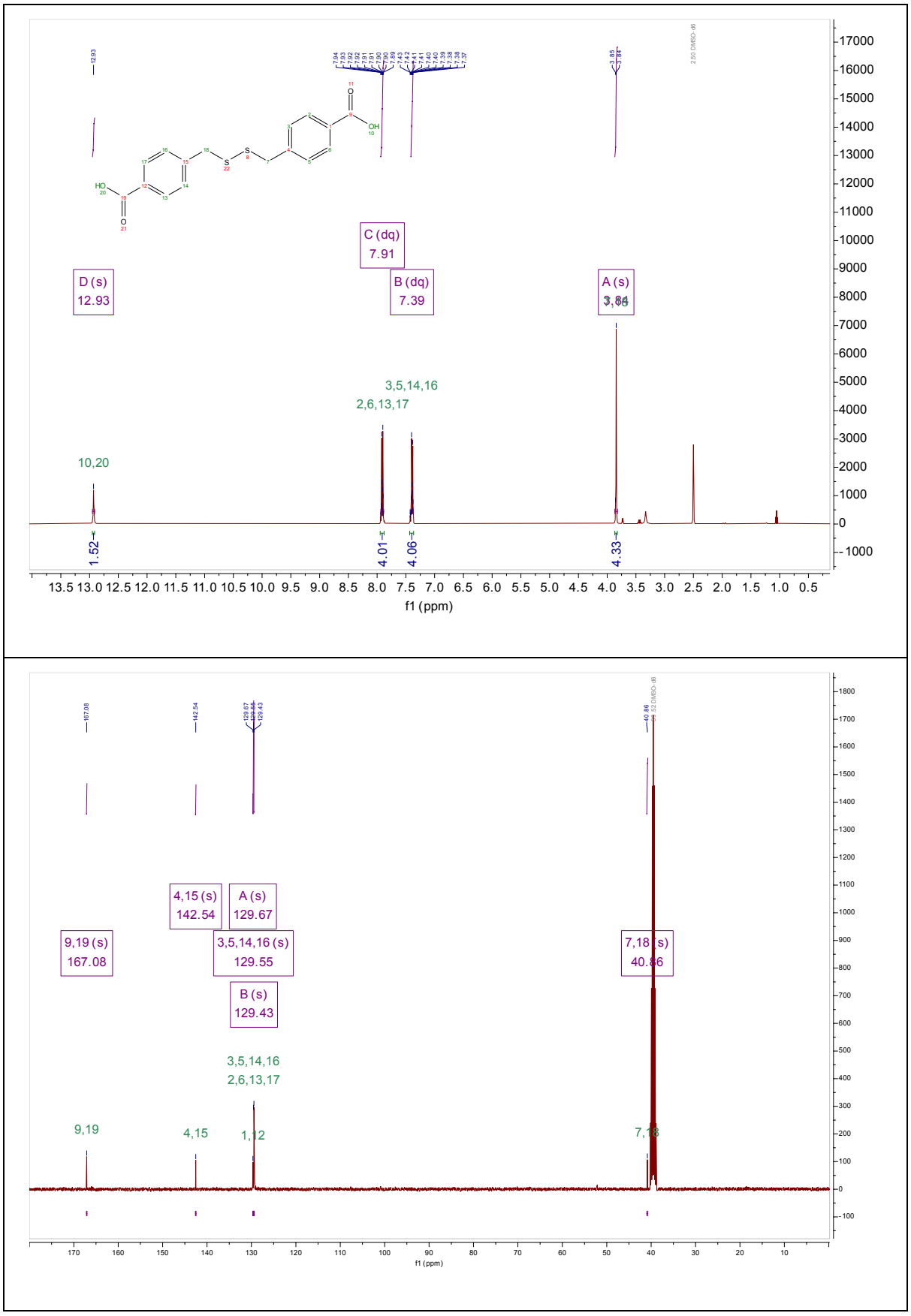
Appendix 10: 6.28 4-(Mercaptomethyl)benzoic acid (29)

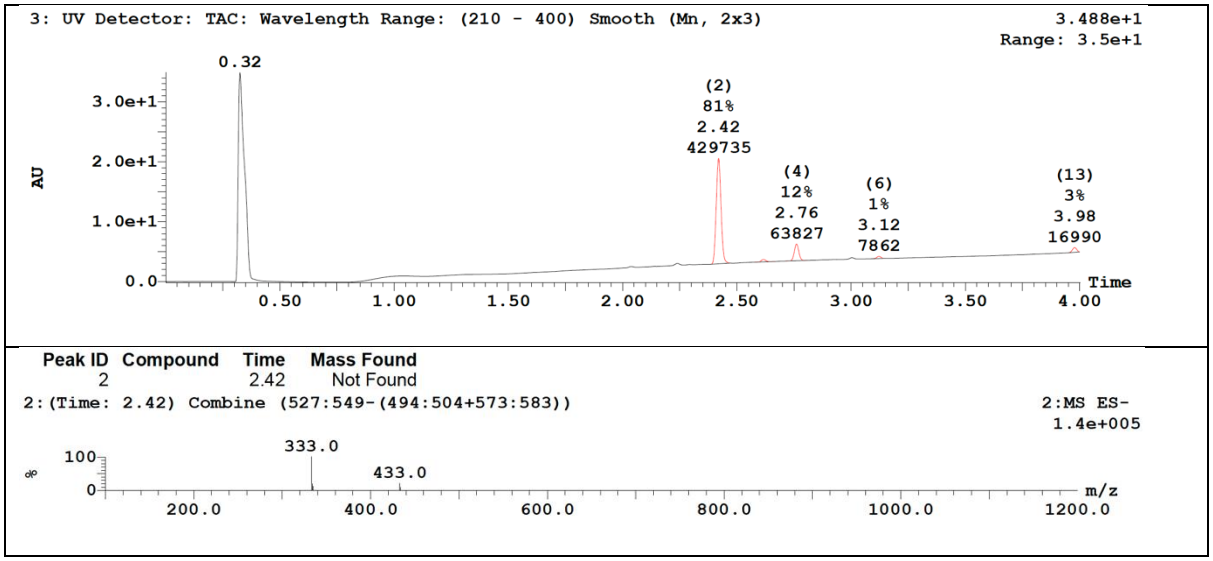




Peak ID	Compound	Time	Mass Found
2		2.05	Not Found

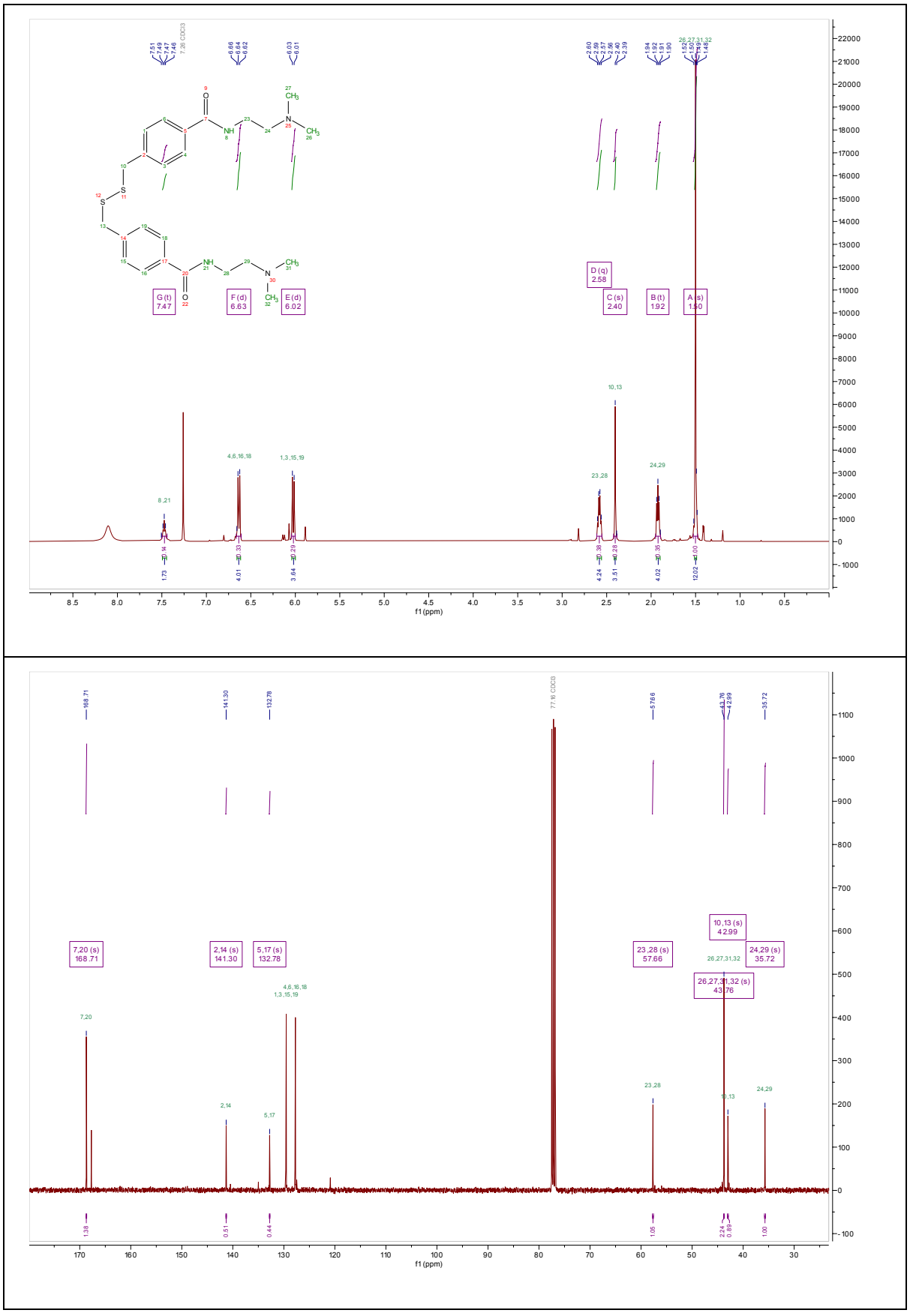
Appendix 11: 4,4'-(Disulfanediylbis(methylene))dibenzoic acid (30)

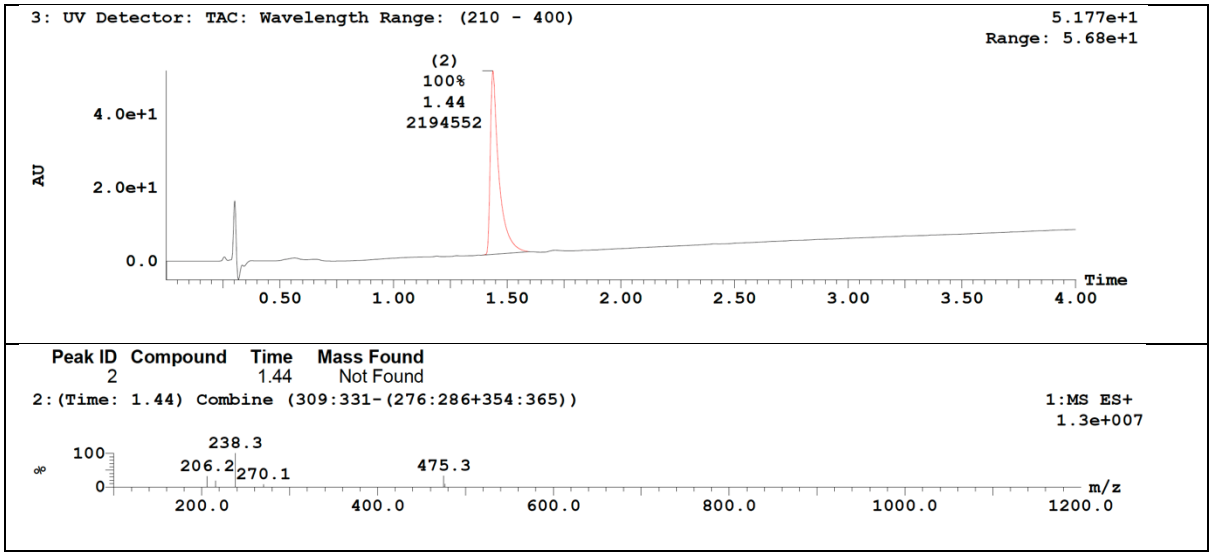




Peak ID	Compound	Time	Mass Found
2		2.42	Not Found

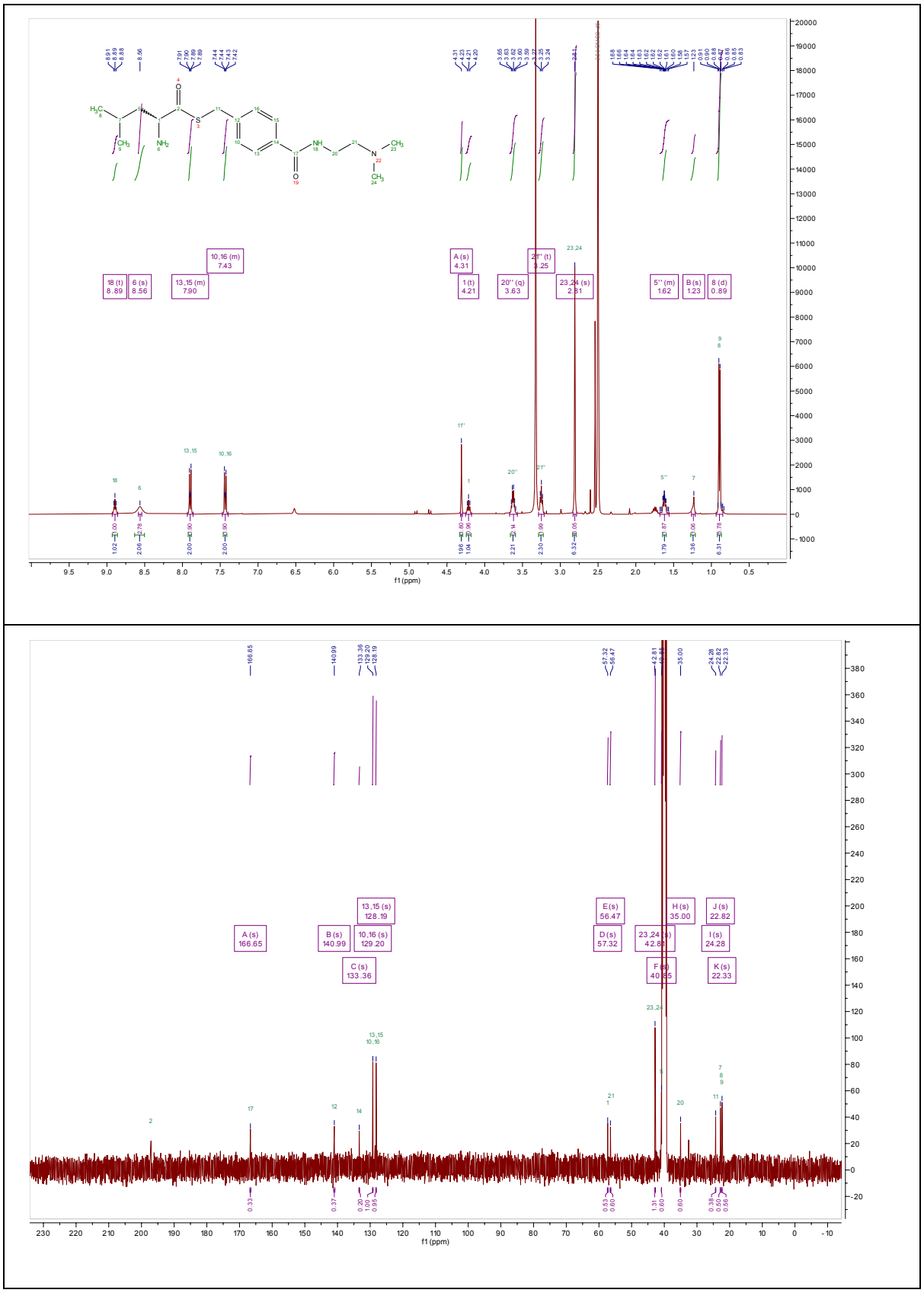
Appendix 12: 4,4'-(Disulfanediylbis(methylene))bis(N-(2-(dimethylamino)ethyl)benzamide (31)

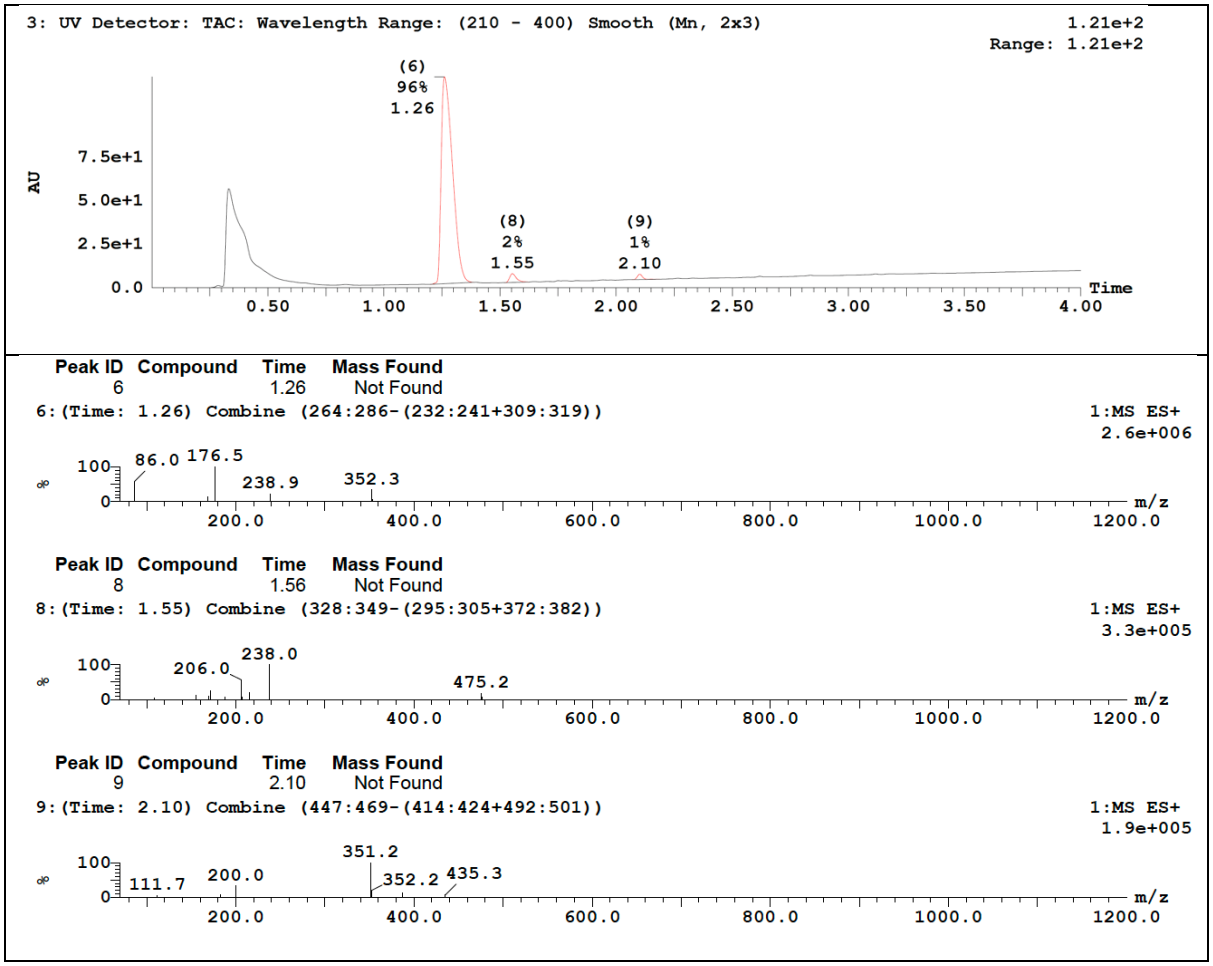




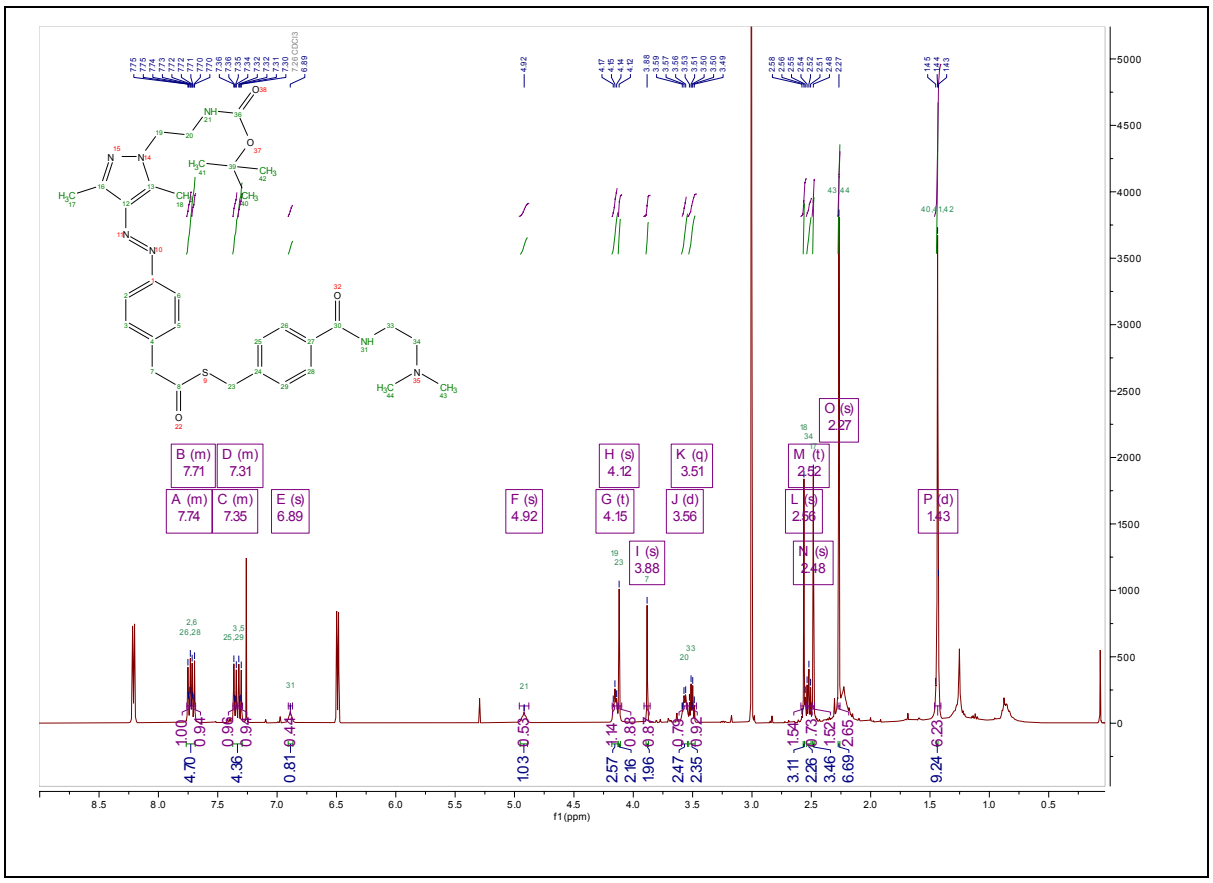
Peak ID	Compound	Time	Mass Found
2		1.44	Not Found

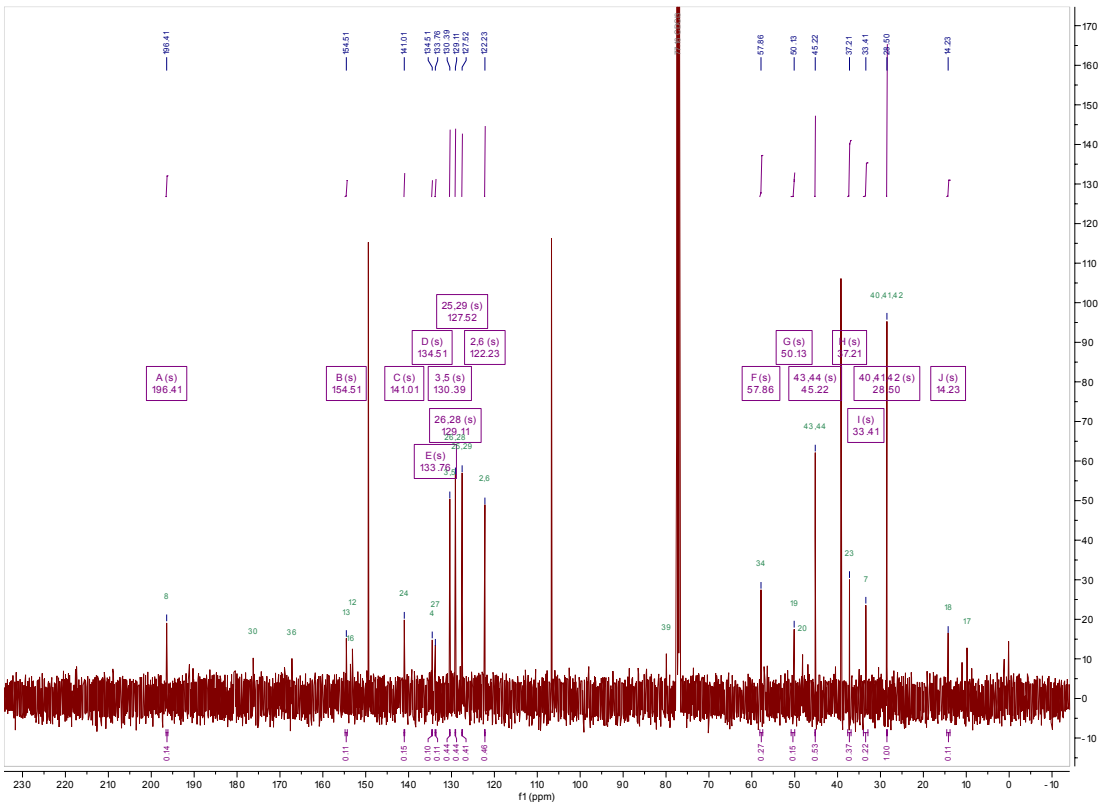
Appendix 13: S-(4-((2-(Dimethylamino)ethyl)carbamoyl)benzyl) 2-amino-4-methyl pentanethioate. HCl (33)

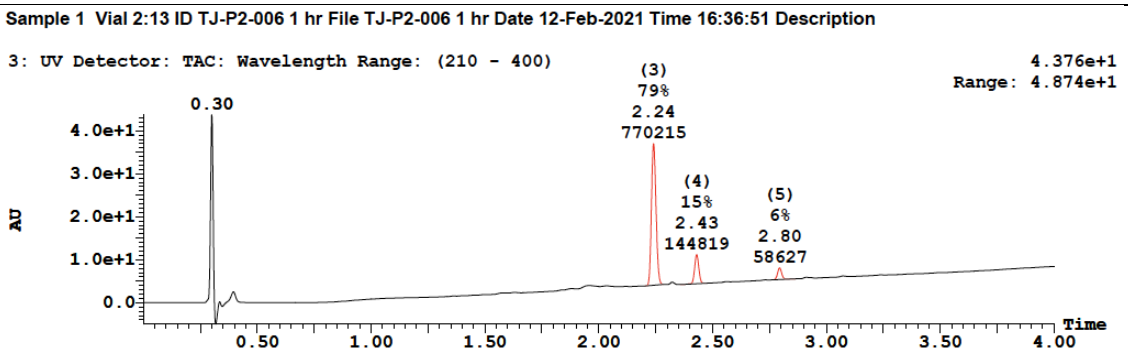


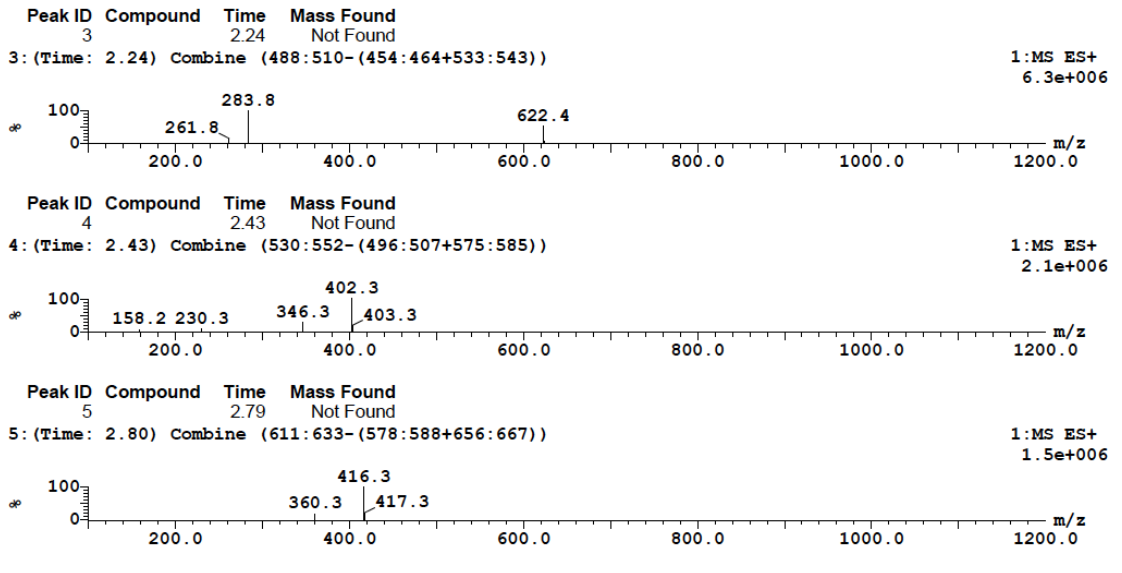

3: UV Detector: TAC: Wavelength Range: (210 - 400) Smooth (Mn, 2x3)
1.21e+2
Range: 1.21e+2
(6)
96%
1.26
(8)
2%
1.55
(9)
1%
2.10
AU
7.5e+1
5.0e+1
2.5e+1
0.0
0.50
1.00
1.50
2.00
2.50
3.00
3.50
4.00
Time
Peak ID Compound Time Mass Found
6 1.26 Not Found
6:(Time: 1.26) Combine (264:286-(232:241+309:319))
1:MS ES+
2.6e+006
100
%
0
86.0
176.5
238.9
352.3
200.0
400.0
600.0
800.0
1000.0
1200.0
m/z
Peak ID Compound Time Mass Found
8 1.56 Not Found
8:(Time: 1.55) Combine (328:349-(295:305+372:382))
1:MS ES+
3.3e+005
100
%
0
206.0
238.0
475.2
200.0
400.0
600.0
800.0
1000.0
1200.0
m/z
Peak ID Compound Time Mass Found
9 2.10 Not Found
9:(Time: 2.10) Combine (447:469-(414:424+492:501))
1:MS ES+
1.9e+005
100
%
0
111.7
200.0
351.2
352.2
435.3
200.0
400.0
600.0
800.0
1000.0
1200.0
m/z

Appendix 14: S-(4-((2-(Dimethylamino)ethyl)carbamoyl)benzyl) (E)-2-(4-((1-(2-((tert-butoxy carbonyl)amino)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)ethanethioate (34)

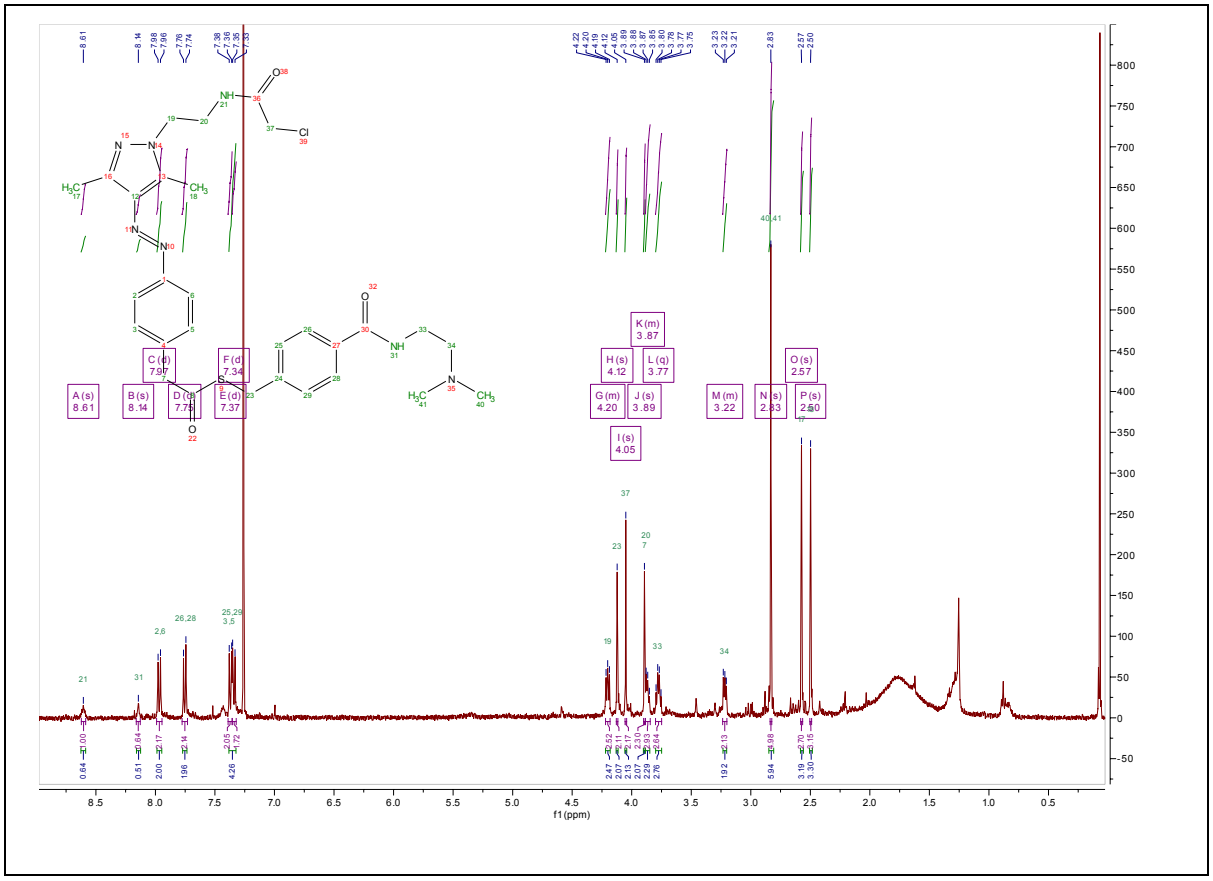


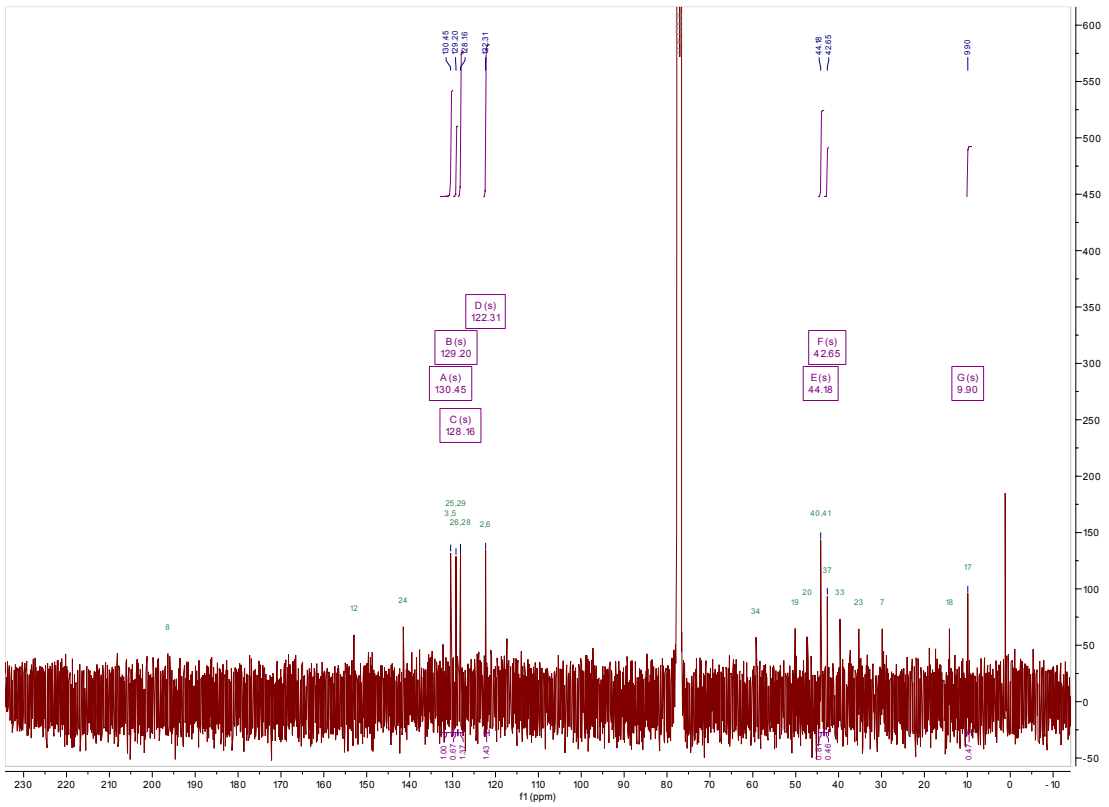

A (s) 196.41
B (s) 154.51
C (s) 141.01
D (s) 134.51
E (s) 133.76
3,5 (s) 130.39
26,28 (s) 129.11
25,29 (s) 127.52
2,6 (s) 122.23
F (s) 57.86
G (s) 50.13
43,44 (s) 45.22
H (s) 37.21
I (s) 33.41
40,41,42 (s) 28.50
J (s) 14.23
f1 (ppm)

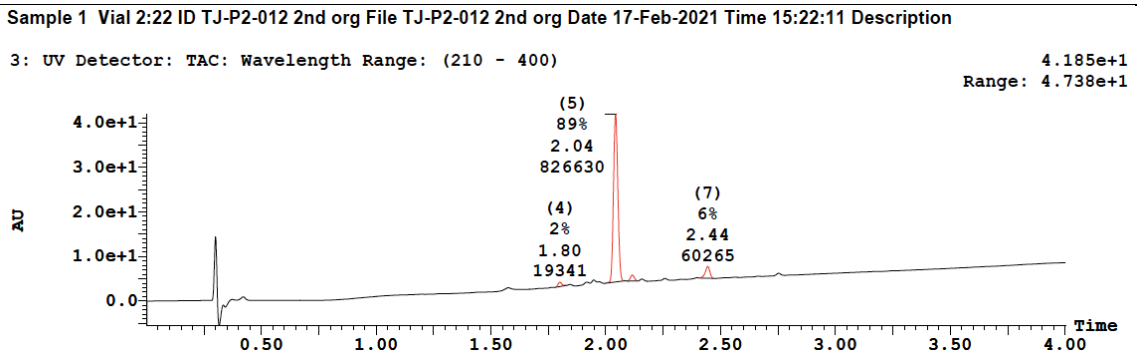

Sample 1 Vial 2:13 ID TJ-P2-006 1 hr File TJ-P2-006 1 hr Date 12-Feb-2021 Time 16:36:51 Description
3: UV Detector: TAC: Wavelength Range: (210 - 400)
4.376e+1
Range: 4.874e+1
0.30
(3)
79%
2.24
770215
(4)
15%
2.43
144819
(5)
6%
2.80
58627
AU
Time

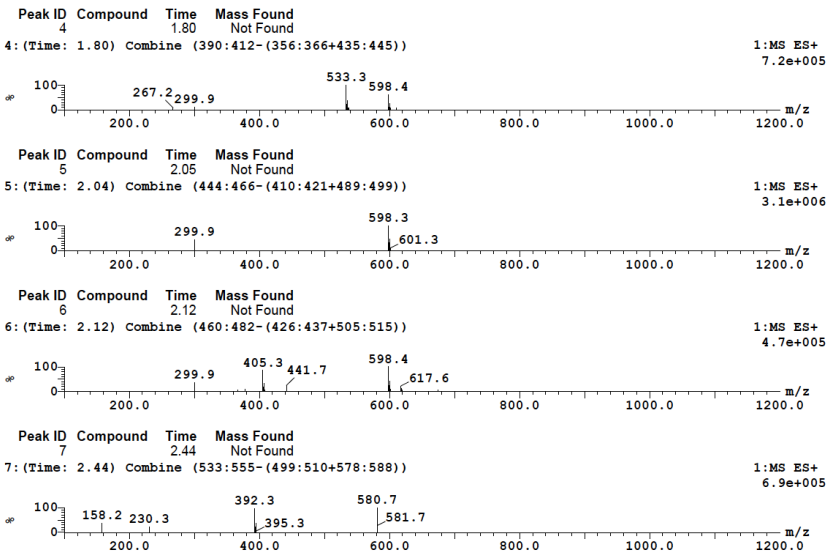

Peak ID Compound Time Mass Found
3 2.24 Not Found
3:(Time: 2.24) Combine (488:510-(454:464+533:543))
1:MS ES+
6.3e+006
283.8
261.8
622.4
m/z
Peak ID Compound Time Mass Found
4 2.43 Not Found
4:(Time: 2.43) Combine (530:552-(496:507+575:585))
1:MS ES+
2.1e+006
402.3
158.2 230.3
346.3
403.3
m/z
Peak ID Compound Time Mass Found
5 2.79 Not Found
5:(Time: 2.80) Combine (611:633-(578:588+656:667))
1:MS ES+
1.5e+006
416.3
360.3
417.3
m/z

Appendix 15: S-(4-((2-(Dimethylamino)ethyl)carbamoyl)benzyl) (E)-2-(4-((1-(2-(2-chloroacetamido)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)ethanethioate (25)

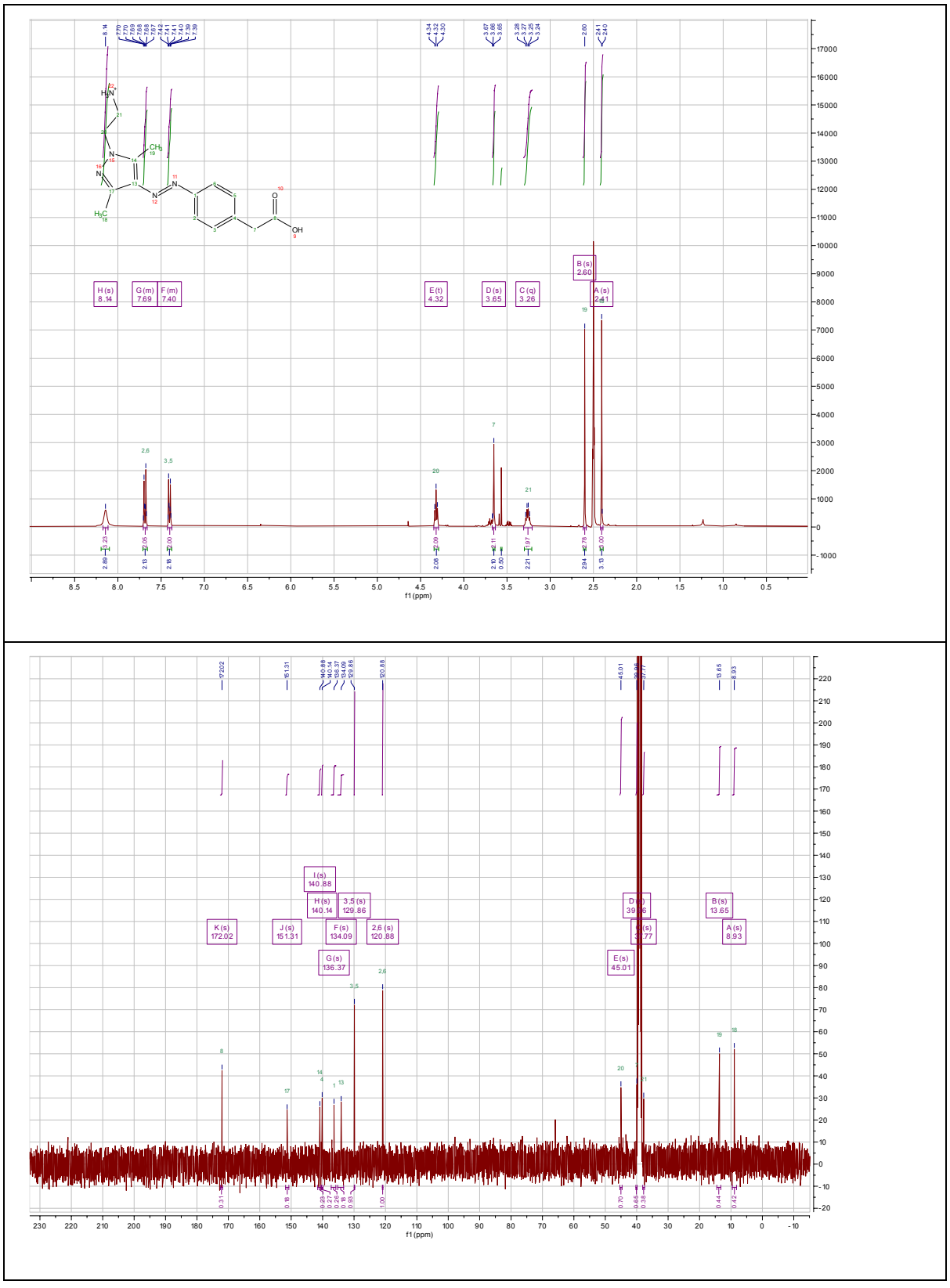


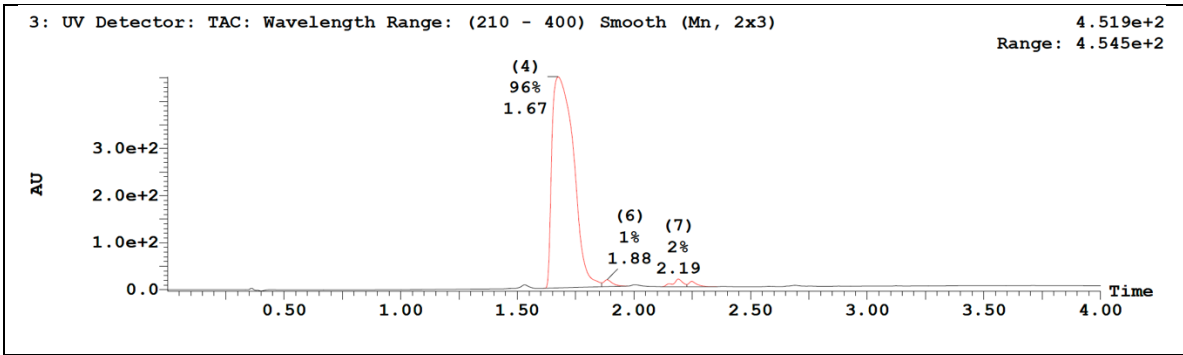

D (s)
122.31
B (s)
129.20
A (s)
130.45
C (s)
128.16
F (s)
42.65
E (s)
44.18
G (s)
9.90
f1 (ppm)


Sample 1 Vial 2:22 ID TJ-P2-012 2nd org File TJ-P2-012 2nd org Date 17-Feb-2021 Time 15:22:11 Description
3: UV Detector: TAC: Wavelength Range: (210 - 400)
4.185e+1
Range: 4.738e+1
(5)
89%
2.04
826630
(4)
2%
1.80
19341
(7)
6%
2.44
60265
AU
Time


Peak ID Compound Time Mass Found
4 1.80 Not Found
4:(Time: 1.80) Combine (390:412-(356:366+435:445))
1:MS ES+
7.2e+005
Peak ID Compound Time Mass Found
5 2.05 Not Found
5:(Time: 2.04) Combine (444:466-(410:421+489:499))
1:MS ES+
3.1e+006
Peak ID Compound Time Mass Found
6 2.12 Not Found
6:(Time: 2.12) Combine (460:482-(426:437+505:515))
1:MS ES+
4.7e+005
Peak ID Compound Time Mass Found
7 2.44 Not Found
7:(Time: 2.44) Combine (533:555-(499:510+578:588))
1:MS ES+
6.9e+005
m/z

Appendix 16: (E)-2-(4-((1-(2-Aminoethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)acetic acid (48)

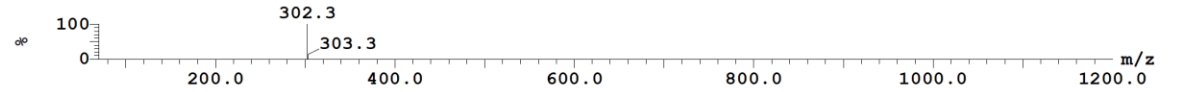




Sample Report (continued):

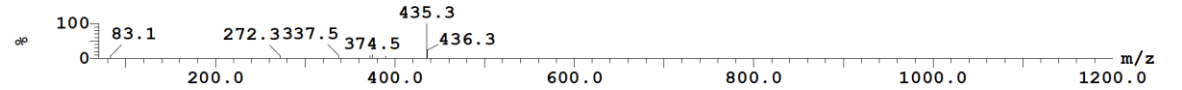
Peak ID	Compound	Time	Mass Found
4		1.67	Not Found

4:(Time: 1.67) Combine (354:376-(321:331+398:408))



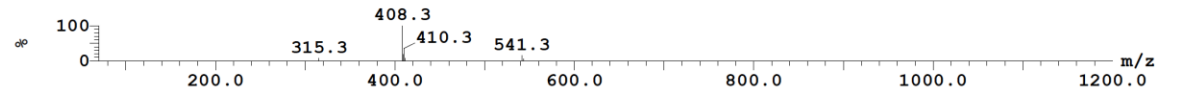
Peak ID	Compound	Time	Mass Found
6		1.88	Not Found

6:(Time: 1.88) Combine (400:421-(367:377+444:454))

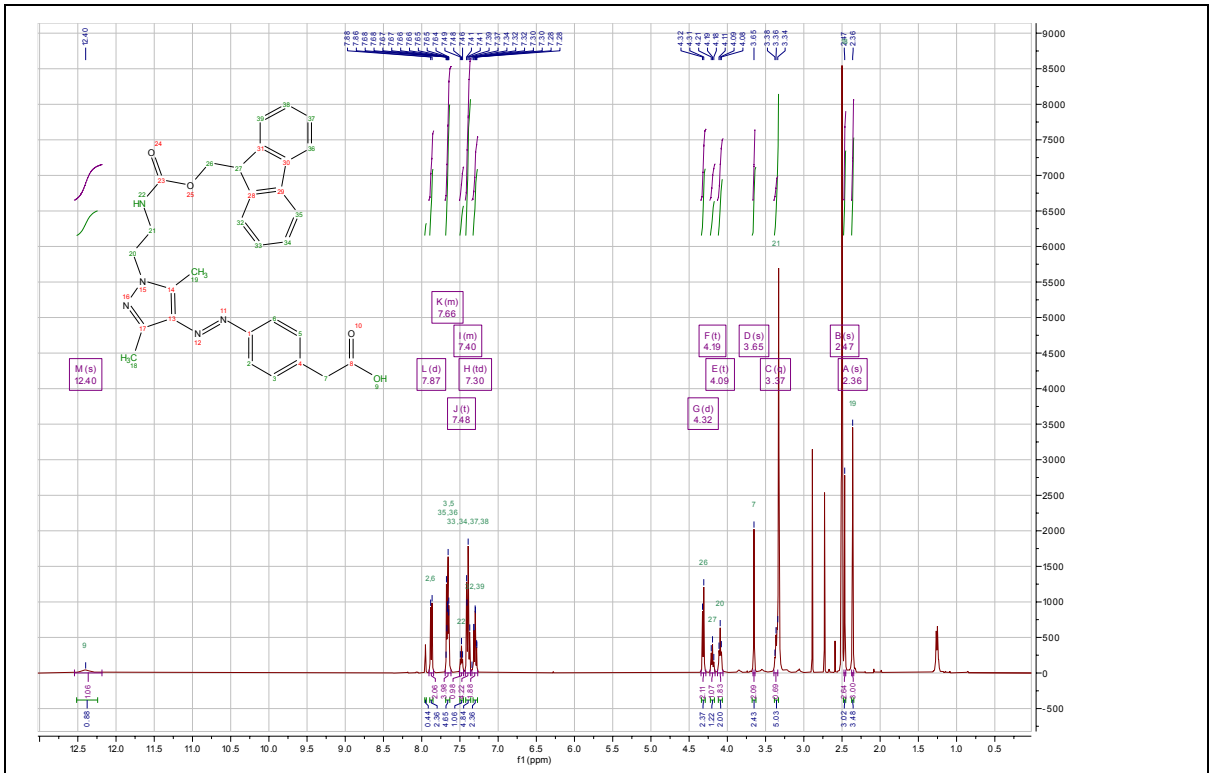


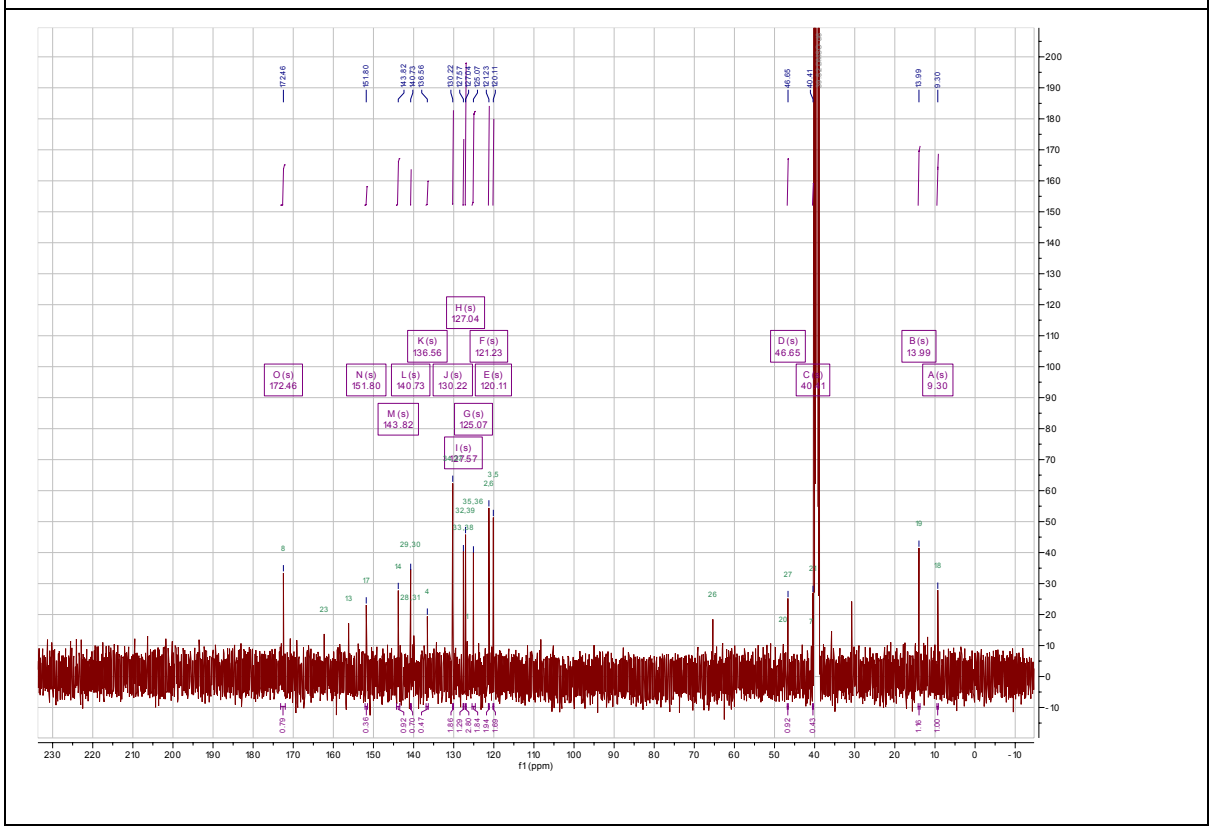
Peak ID	Compound	Time	Mass Found
7		2.15	Not Found

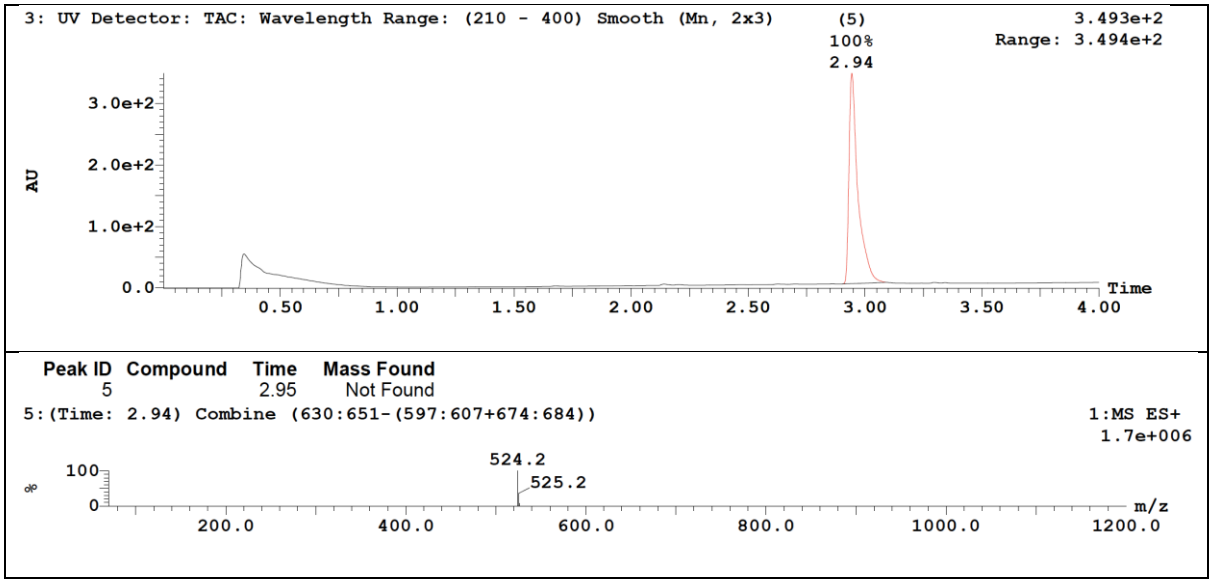
7:(Time: 2.19) Combine (466:488-(433:443+510:520))



Appendix 17: (E)-2-(4-((1-(2-((((9H-Fluoren-9-yl)methoxy)carbonyl)amino)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)acetic acid (49)

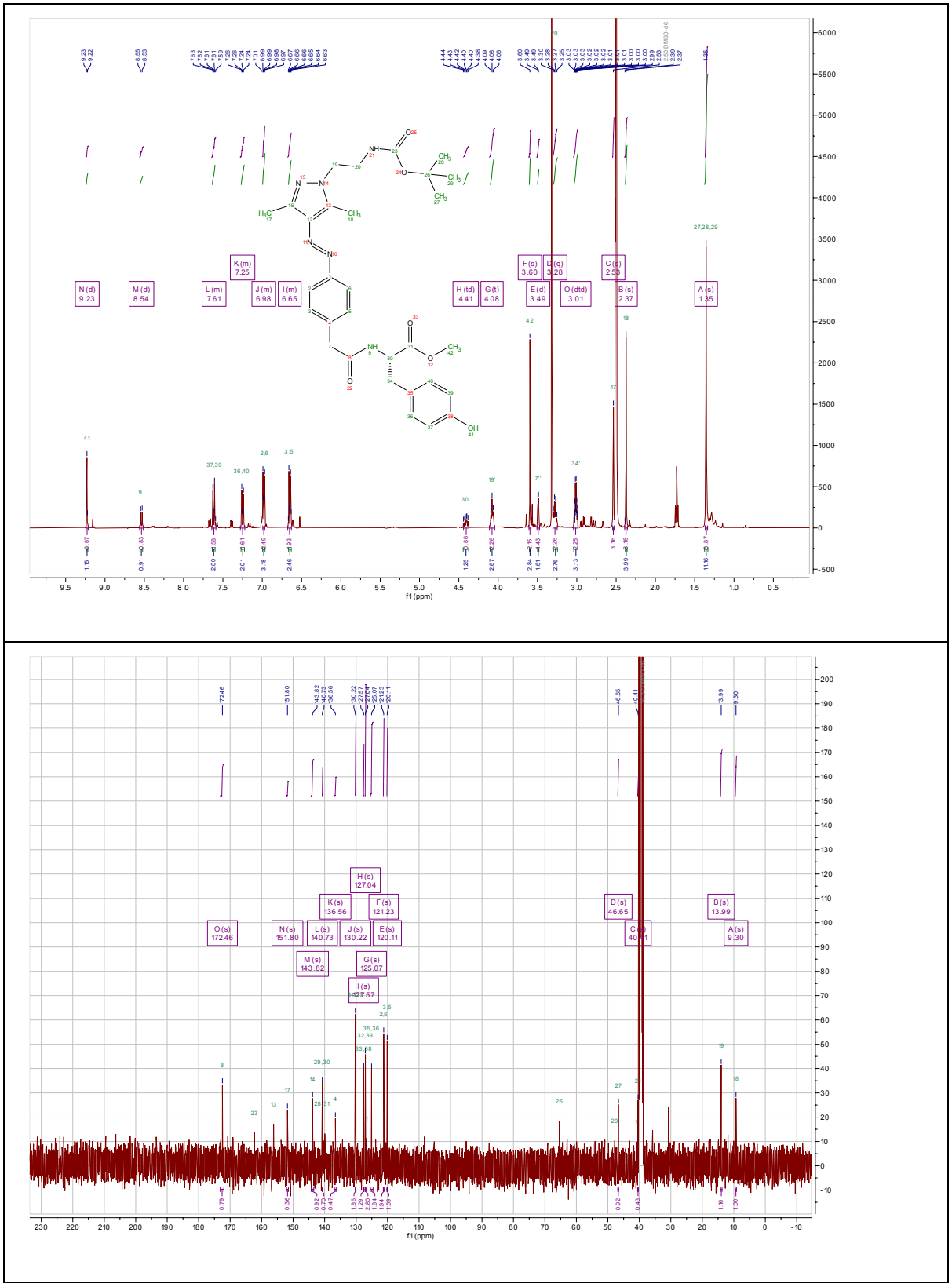


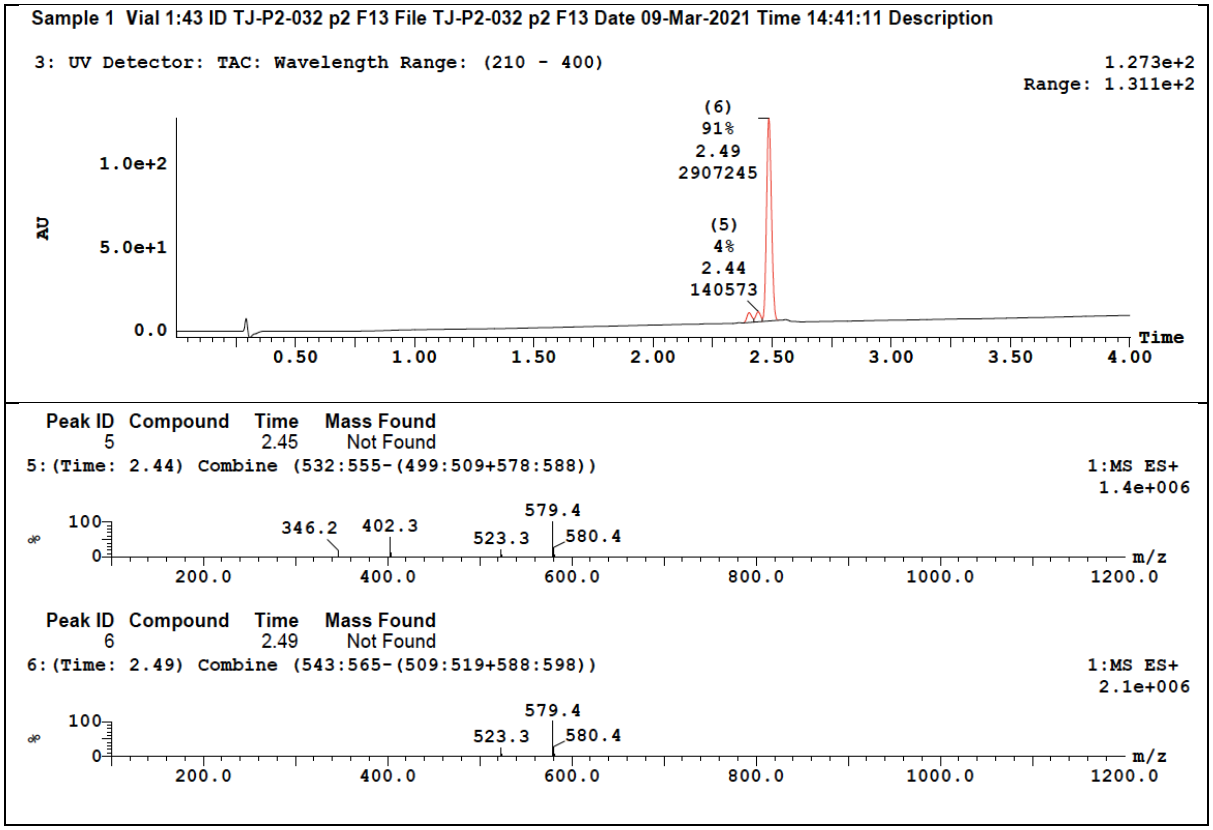




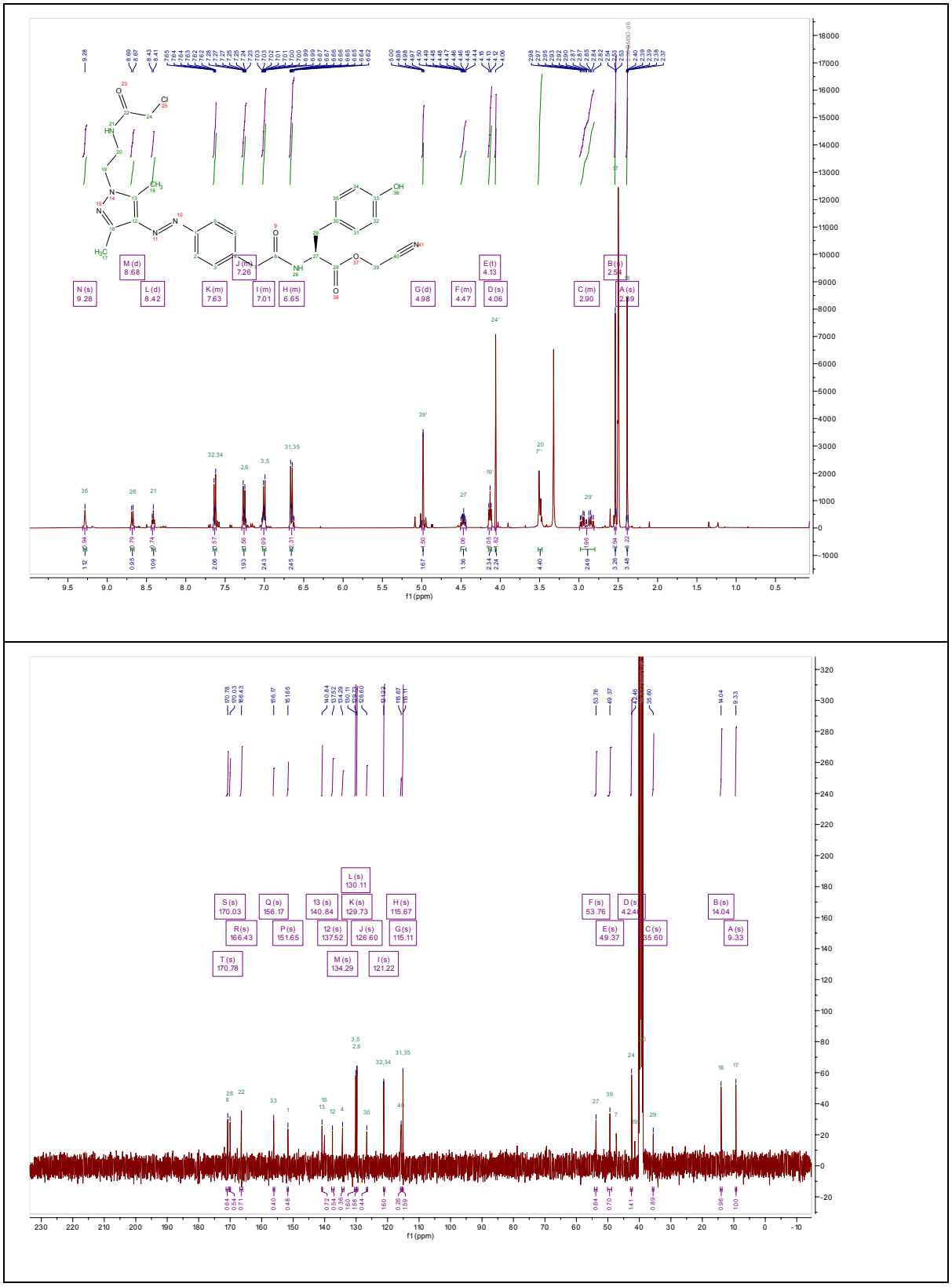
Peak ID	Compound	Time	Mass Found
5		2.95	Not Found

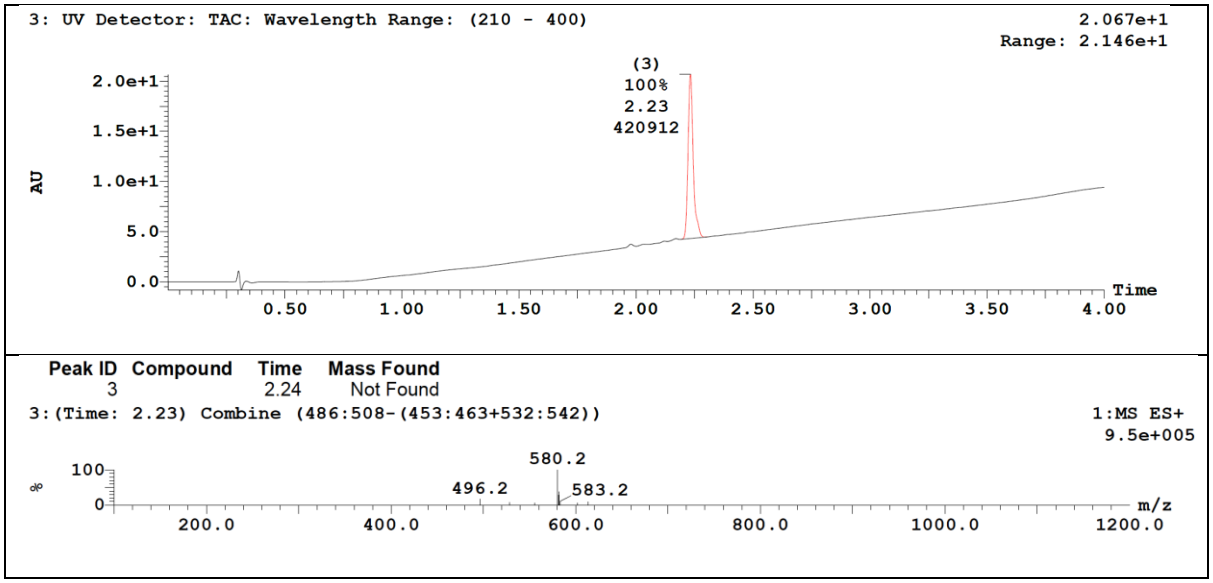
Appendix 18: Methyl (E)-(2-(4-((1-(2-((tert-butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)acetyl)-L-tyrosinate (37)




Sample 1 Vial 1:43 ID TJ-P2-032 p2 F13 File TJ-P2-032 p2 F13 Date 09-Mar-2021 Time 14:41:11 Description
3: UV Detector: TAC: Wavelength Range: (210 - 400)
1.273e+2
Range: 1.311e+2
(6)
91%
2.49
2907245
(5)
4%
2.44
140573
AU
1.0e+2
5.0e+1
0.0
0.50
1.00
1.50
2.00
2.50
3.00
3.50
4.00
Time
Peak ID Compound Time Mass Found
5 2.45 Not Found
5:(Time: 2.44) Combine (532:555-(499:509+578:588))
1:MS ES+
1.4e+006
100
%
0
346.2
402.3
523.3
579.4
580.4
200.0
400.0
600.0
800.0
1000.0
1200.0
m/z
Peak ID Compound Time Mass Found
6 2.49 Not Found
6:(Time: 2.49) Combine (543:565-(509:519+588:598))
1:MS ES+
2.1e+006
100
%
0
523.3
579.4
580.4
200.0
400.0
600.0
800.0
1000.0
1200.0
m/z

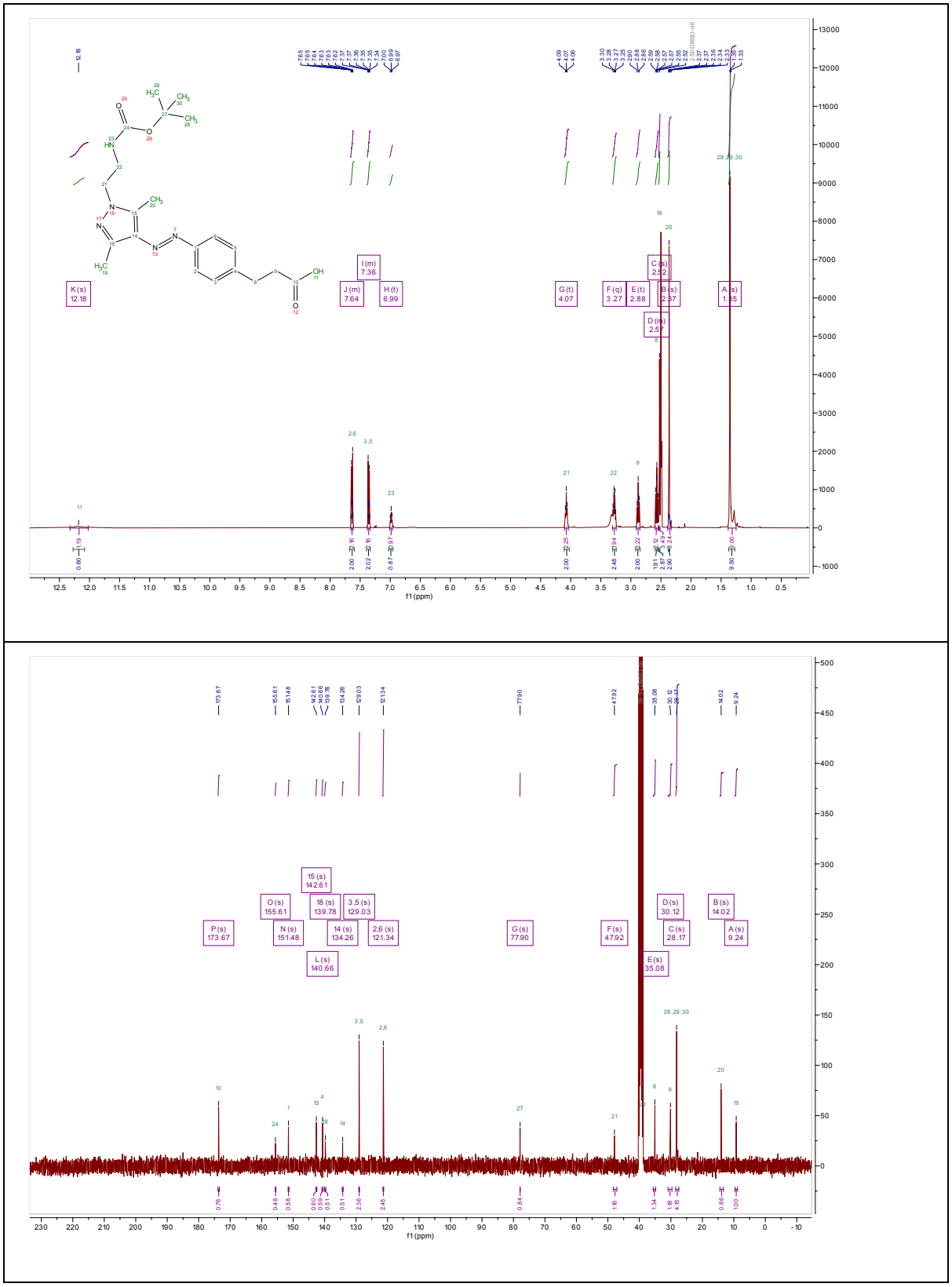
Appendix 19: Cyanomethyl (E)-(2-(4-((1-(2-(2-chloroacetamido)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)acetyl)-L-tyrosinate (36)

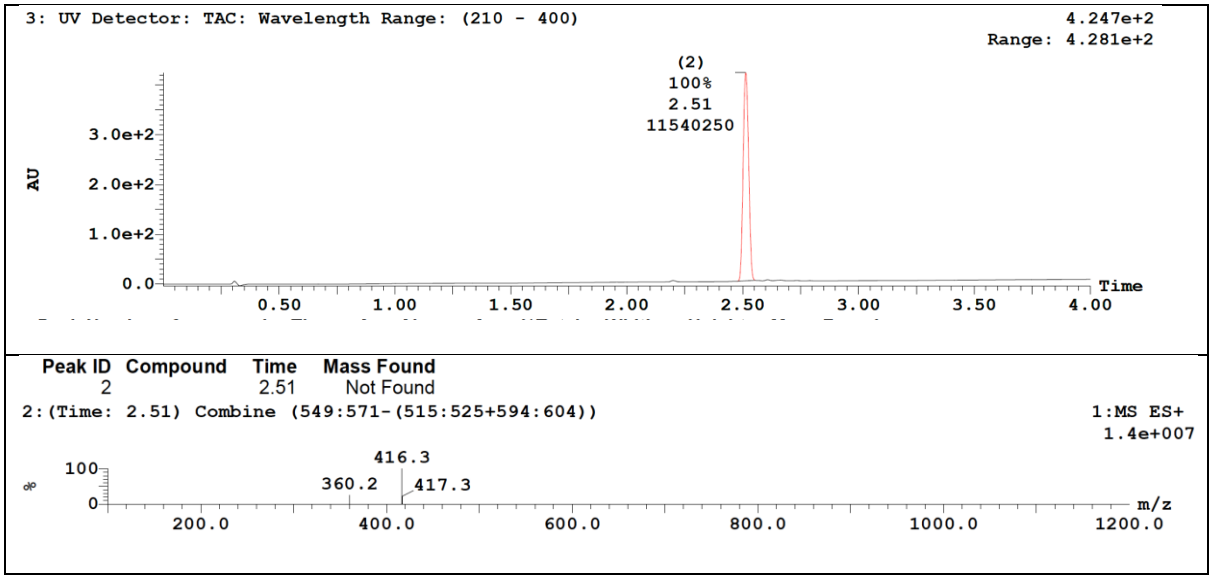




Peak ID	Compound	Time	Mass Found
3		2.24	Not Found

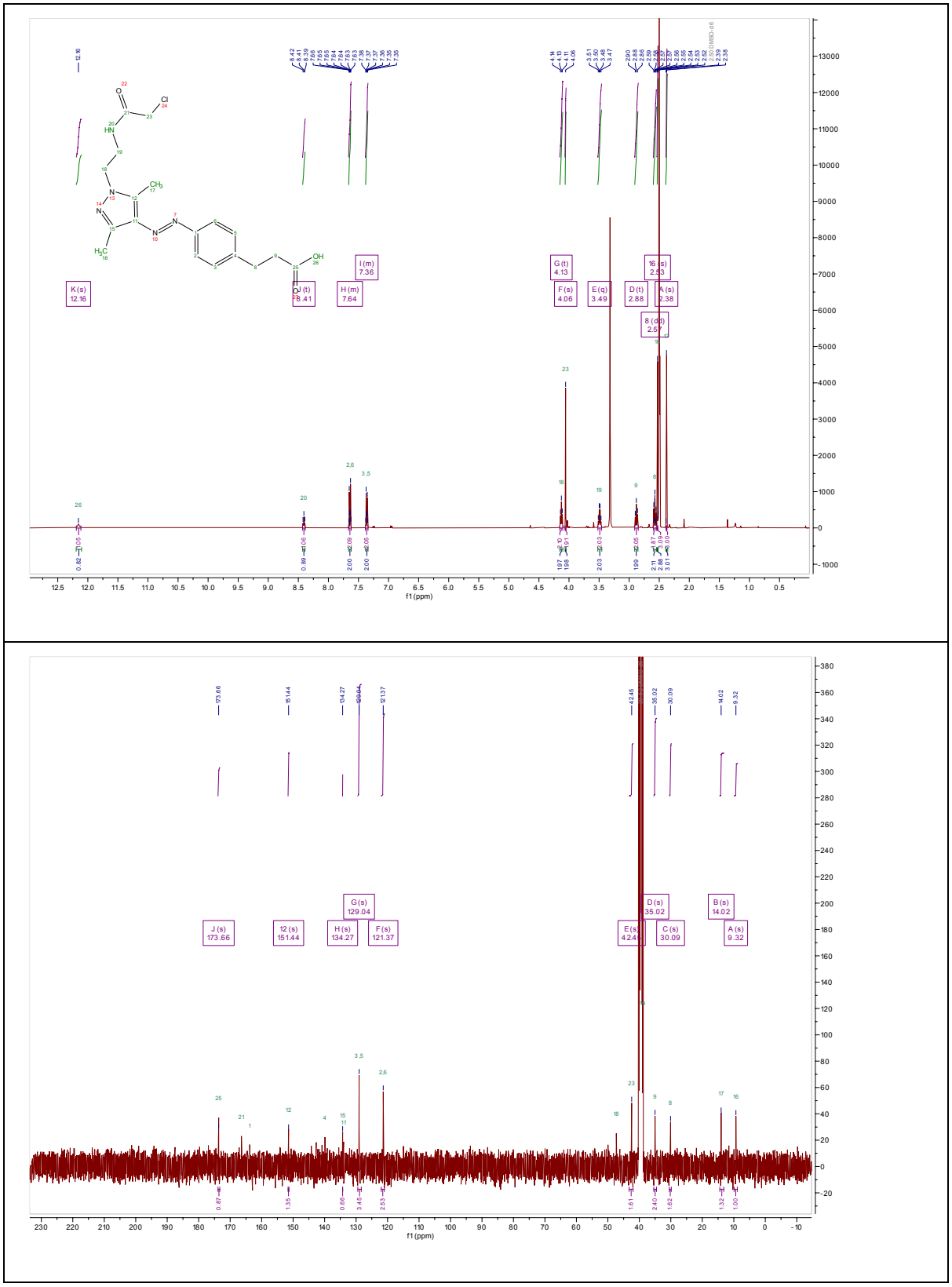
Appendix 20: (E)-3-(4-((1-(2-((tert-Butoxycarbonyl)amino)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)propanoic acid (45)

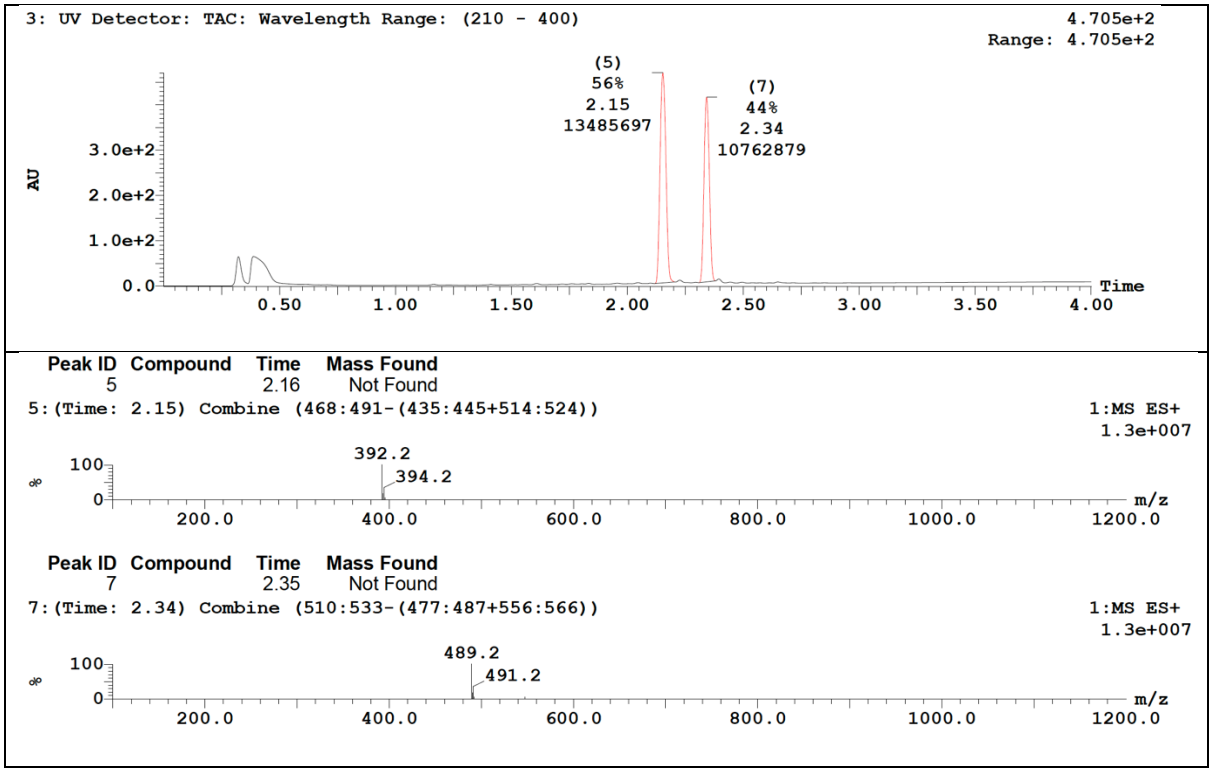




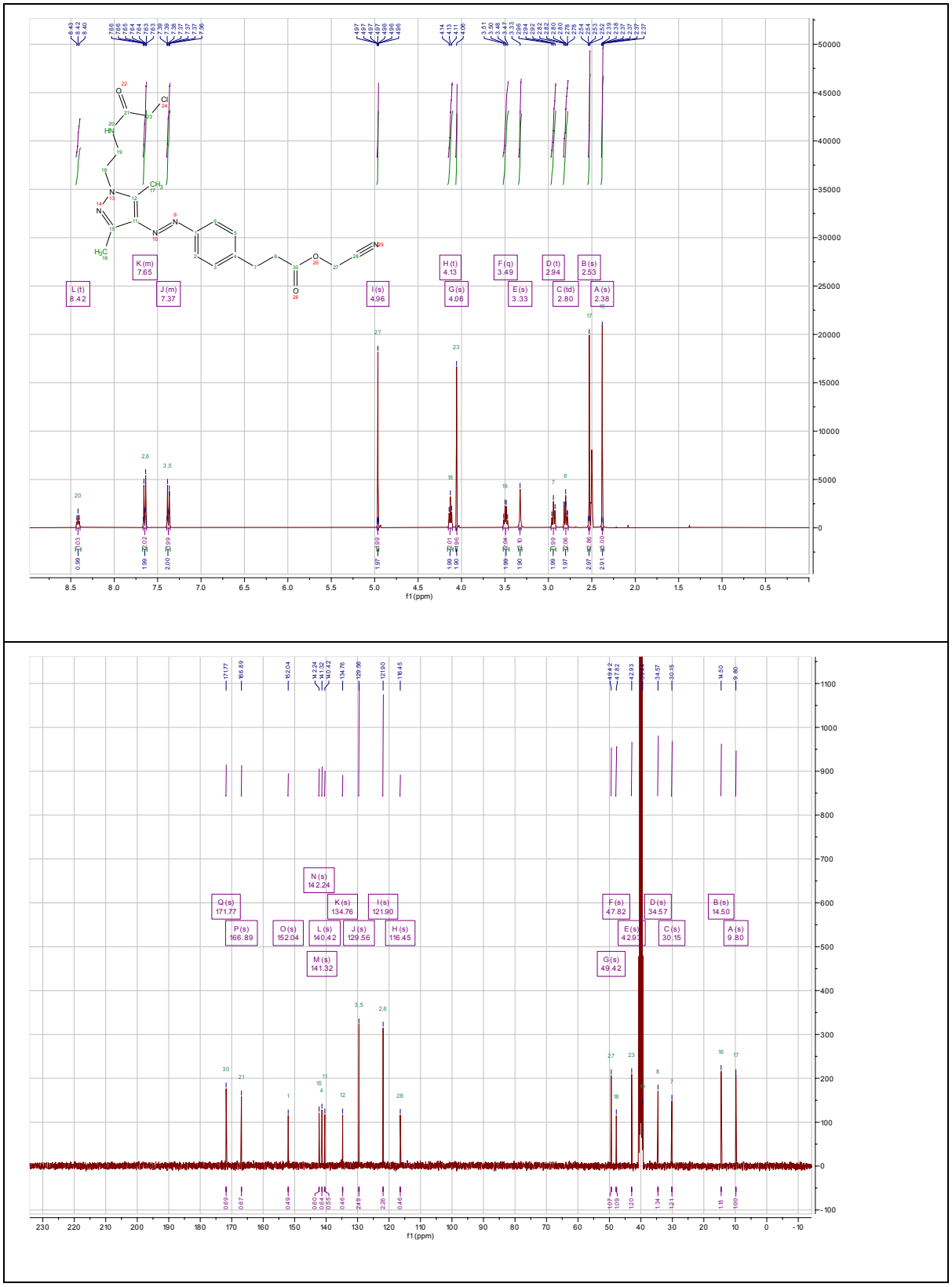
Peak ID	Compound	Time	Mass Found
2		2.51	Not Found

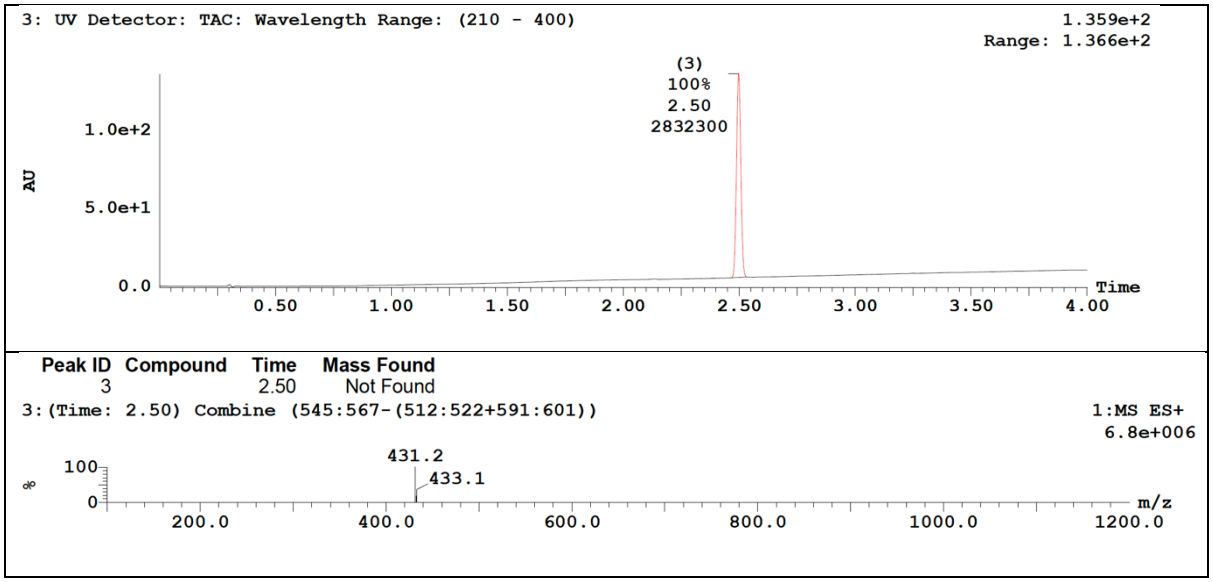
Appendix 21: (E)-3-(4-((1-(2-(2-Chloroacetamido)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl) diazenyl)phenyl)propanoic acid (47)




3: UV Detector: TAC: Wavelength Range: (210 - 400)
4.705e+2
Range: 4.705e+2
(5)
56%
2.15
13485697
(7)
44%
2.34
10762879
AU
3.0e+2
2.0e+2
1.0e+2
0.0
0.50
1.00
1.50
2.00
2.50
3.00
3.50
4.00
Time
Peak ID Compound Time Mass Found
5 2.16 Not Found
5:(Time: 2.15) Combine (468:491-(435:445+514:524))
1:MS ES+
1.3e+007
392.2
394.2
%
100
0
200.0
400.0
600.0
800.0
1000.0
1200.0
m/z
Peak ID Compound Time Mass Found
7 2.35 Not Found
7:(Time: 2.34) Combine (510:533-(477:487+556:566))
1:MS ES+
1.3e+007
489.2
491.2
%
100
0
200.0
400.0
600.0
800.0
1000.0
1200.0
m/z

Appendix 22: Cyanomethyl (E)-3-(4-((1-(2-(2-chloroacetamido)ethyl)-3,5-dimethyl-1H-pyrazol-4-yl)diazenyl)phenyl)propanoate (41)



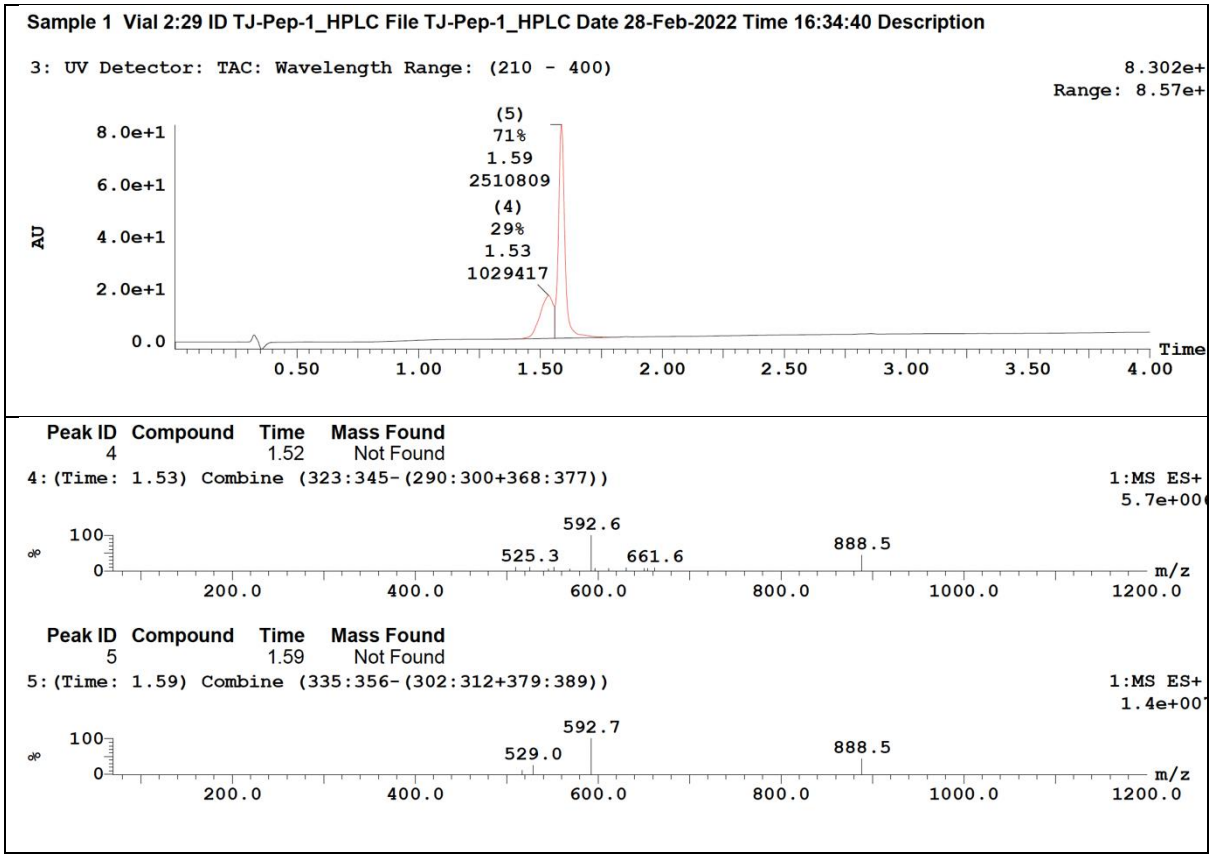


Peak ID	Compound	Time	Mass Found
3		2.50	Not Found

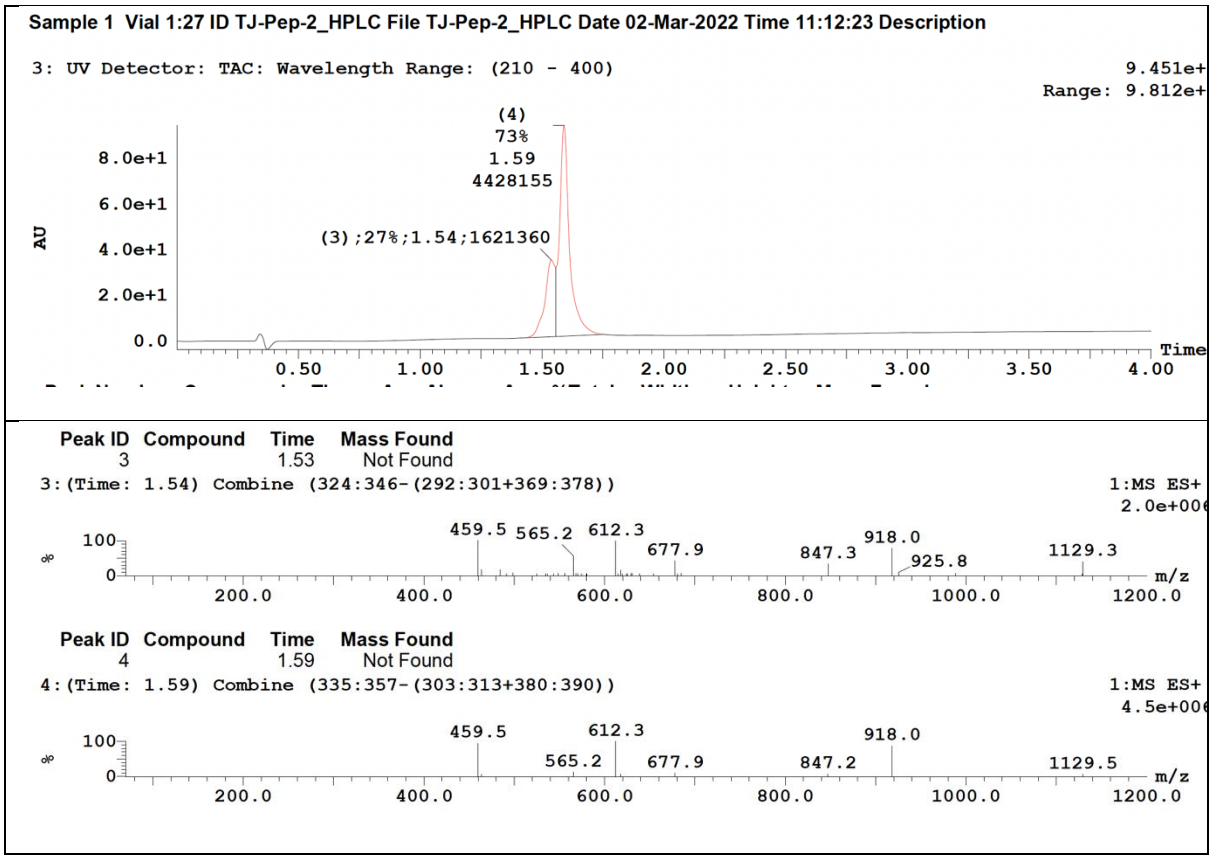
SPPS Peptide Characterisation:

For SPPS peptides with adjacent small shoulder peak – This relates to the other isomer (*E/Z*) confirmed by switching sample and is the result of a light within the LCMS sampler. Ion peak traces are the same for each isomer peak within the single run as shown below to also indicate the mixed isomer population. Some impurity peaks are also seen for some peptides.

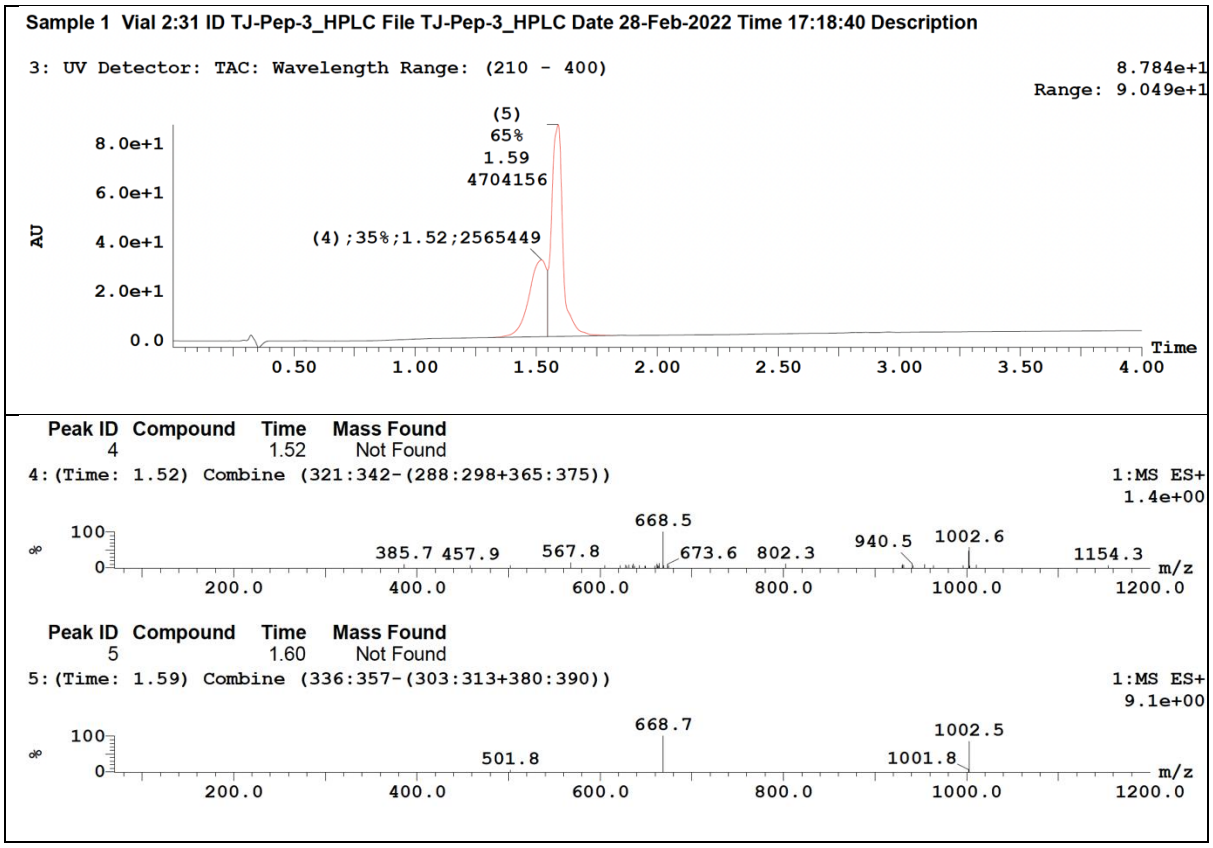
Appendix 23: SPPS Peptide Characterisation: P001



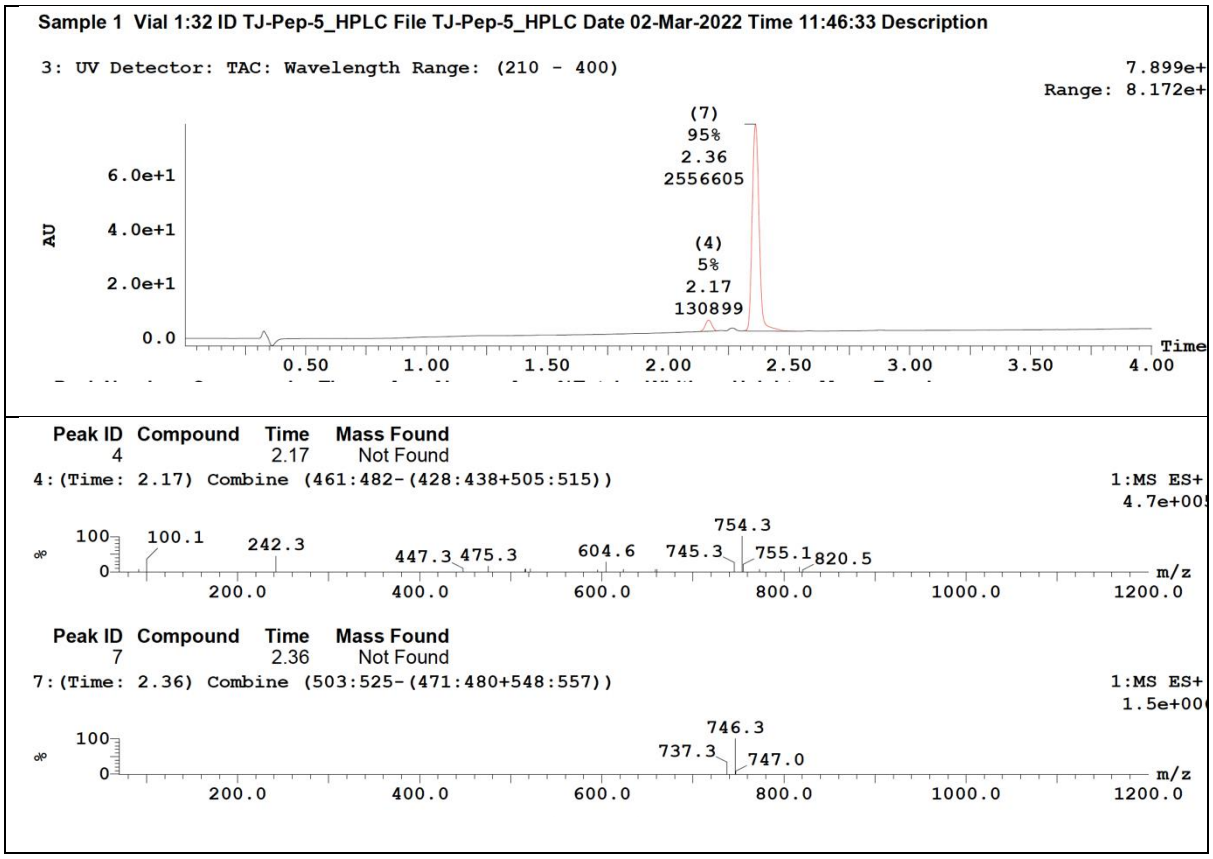
Appendix 24: SPPS Peptide Characterisation: P002



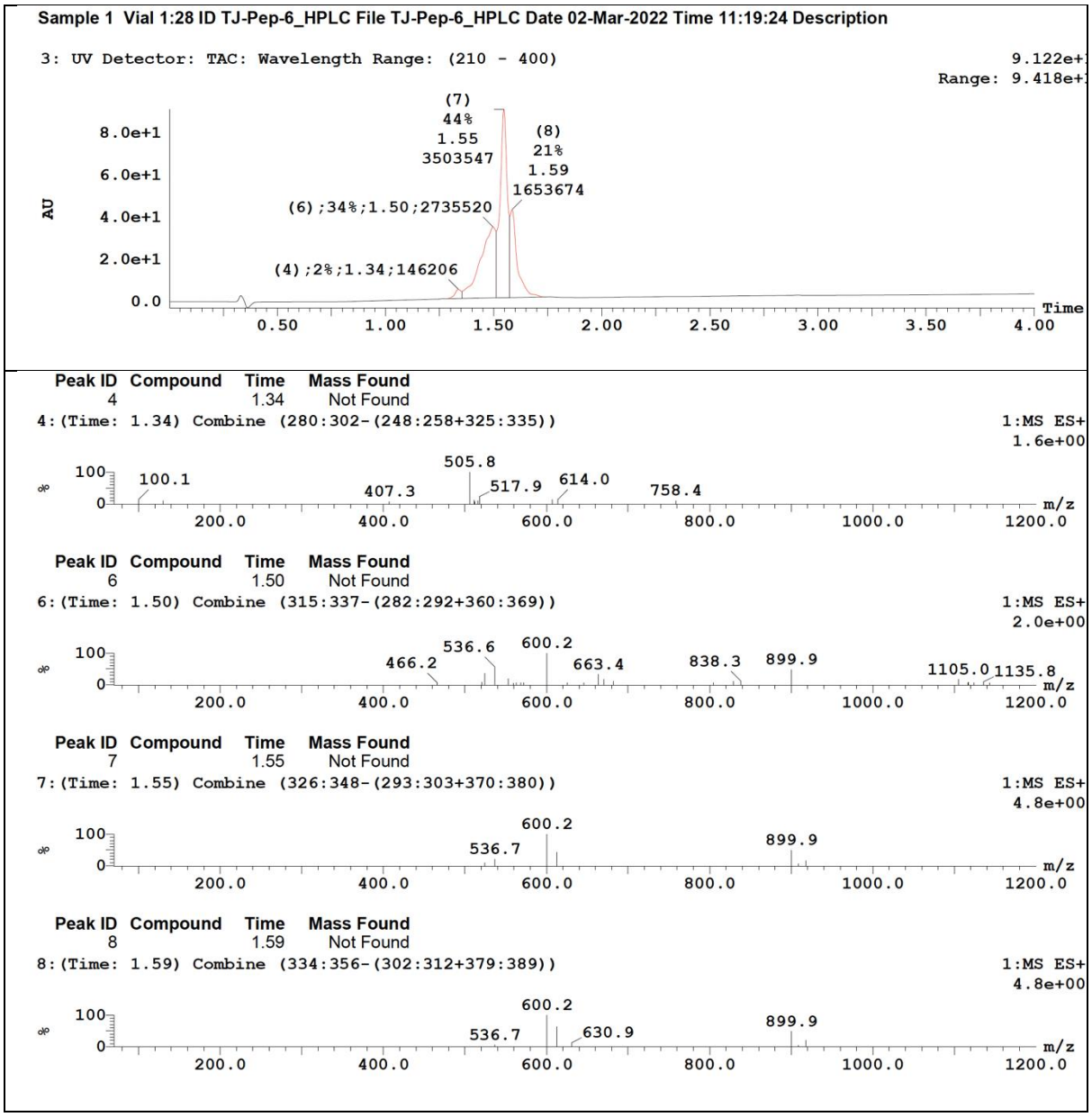
Appendix 25: SPPS Peptide Characterisation: P003



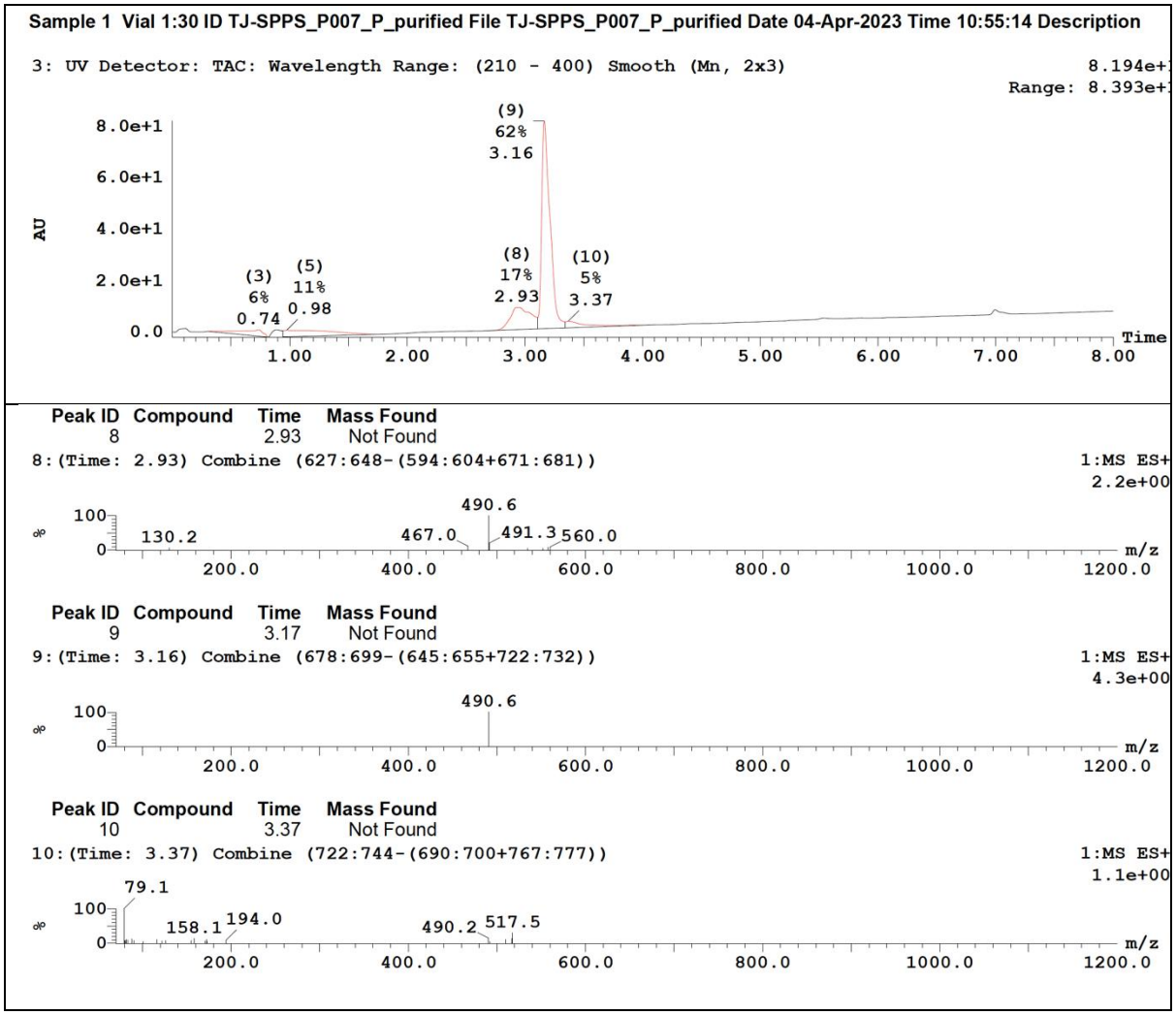
Appendix 26: SPPS Peptide Characterisation: P005



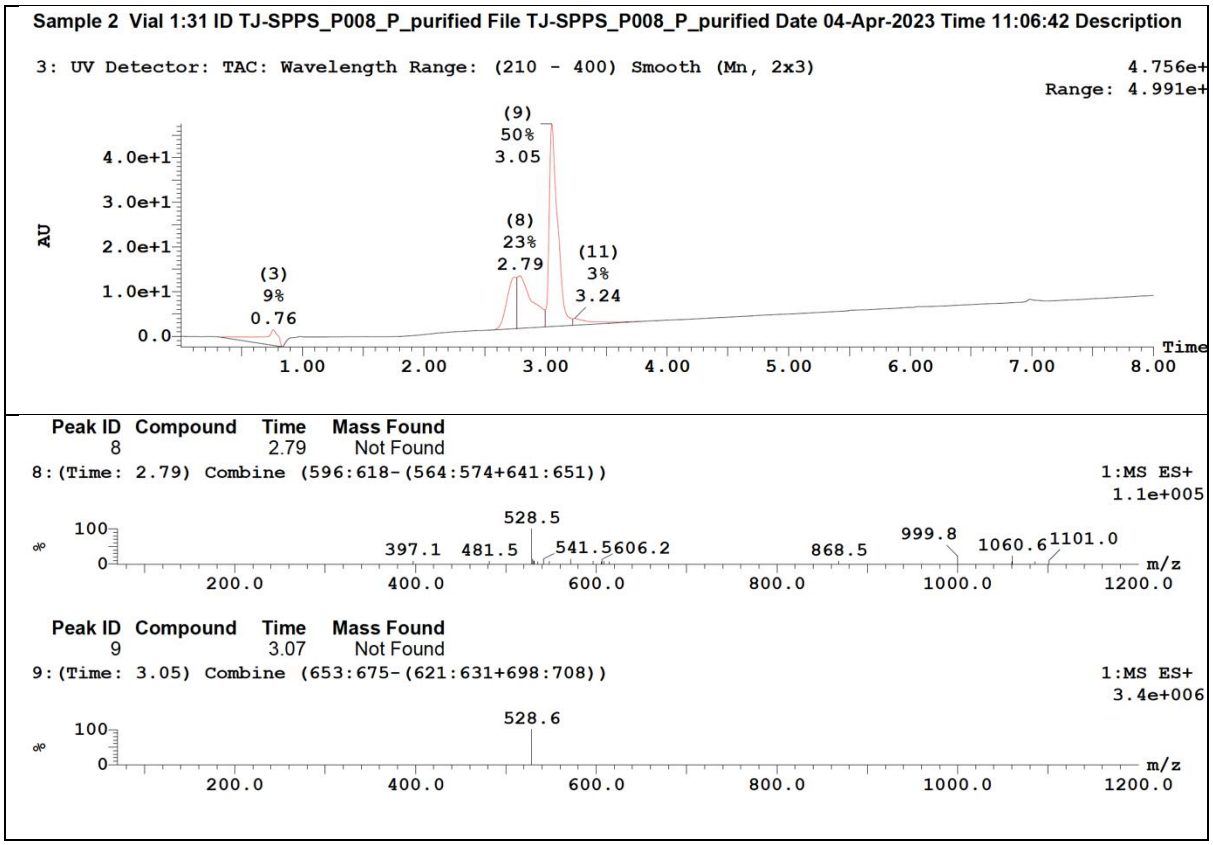
Appendix 27: SPPS Peptide Characterisation: P006



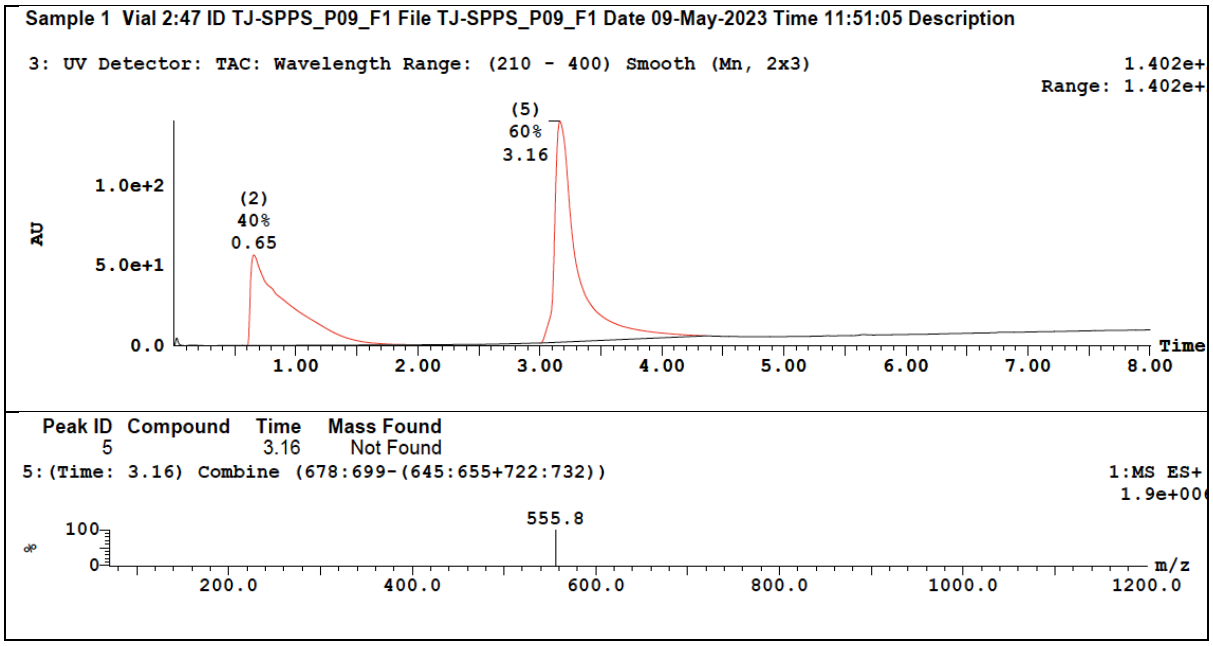
Appendix 28: SPPS Peptide Characterisation: P007



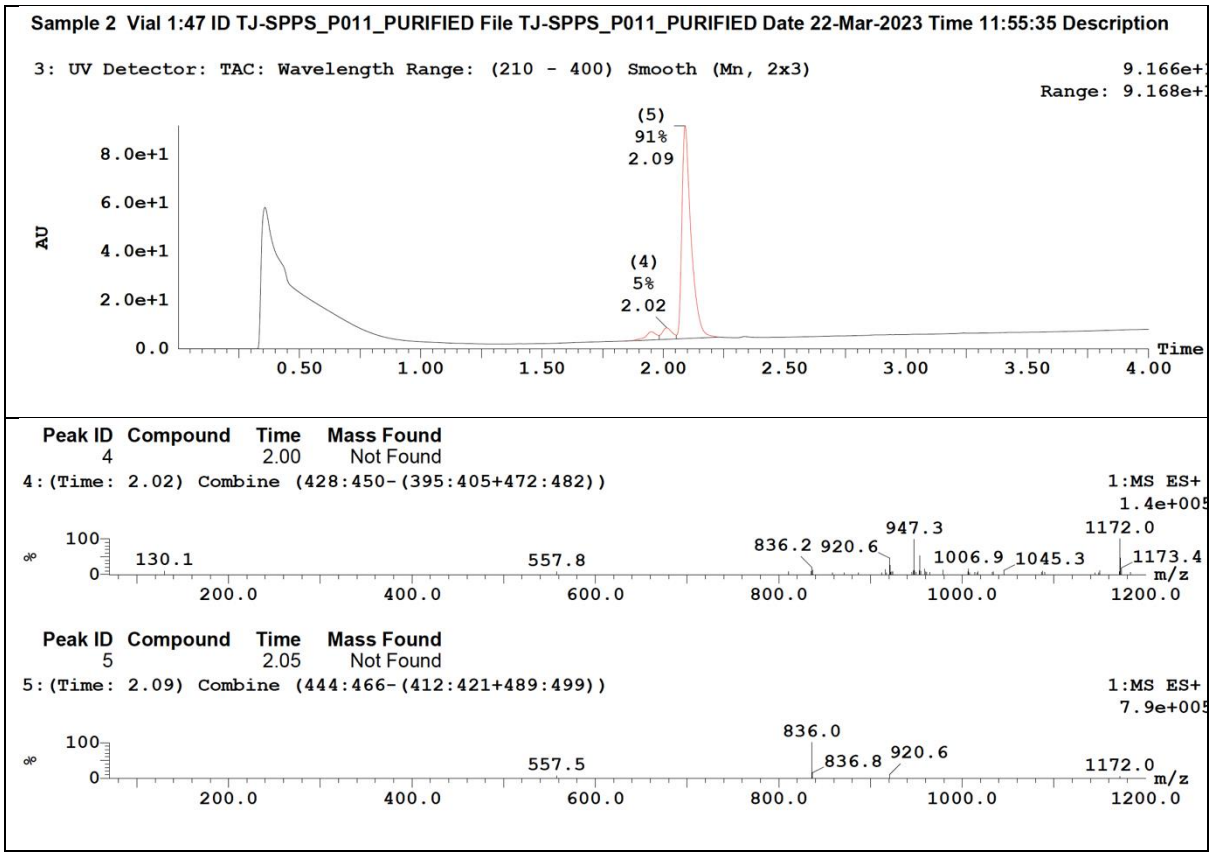
Appendix 29: SPPS Peptide Characterisation: P008



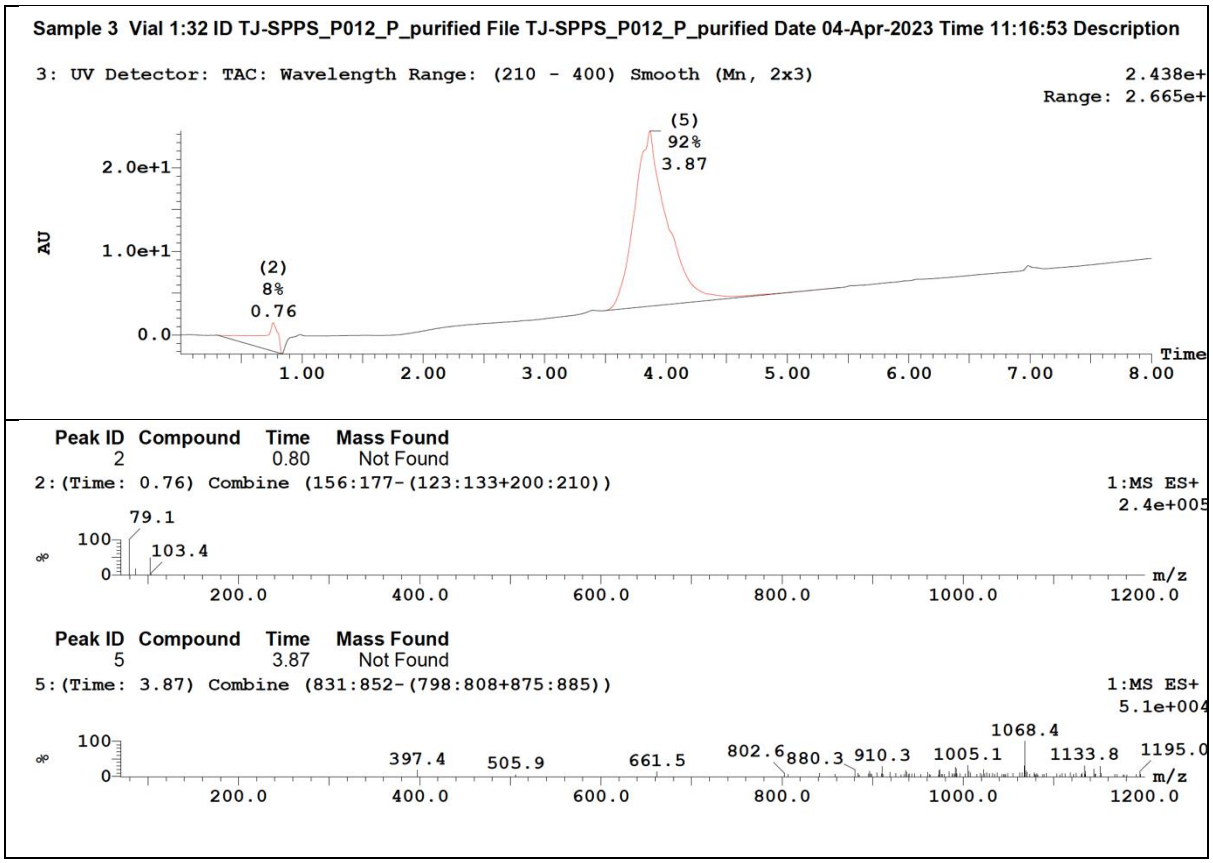
Appendix 30: SPPS Peptide Characterisation: P009



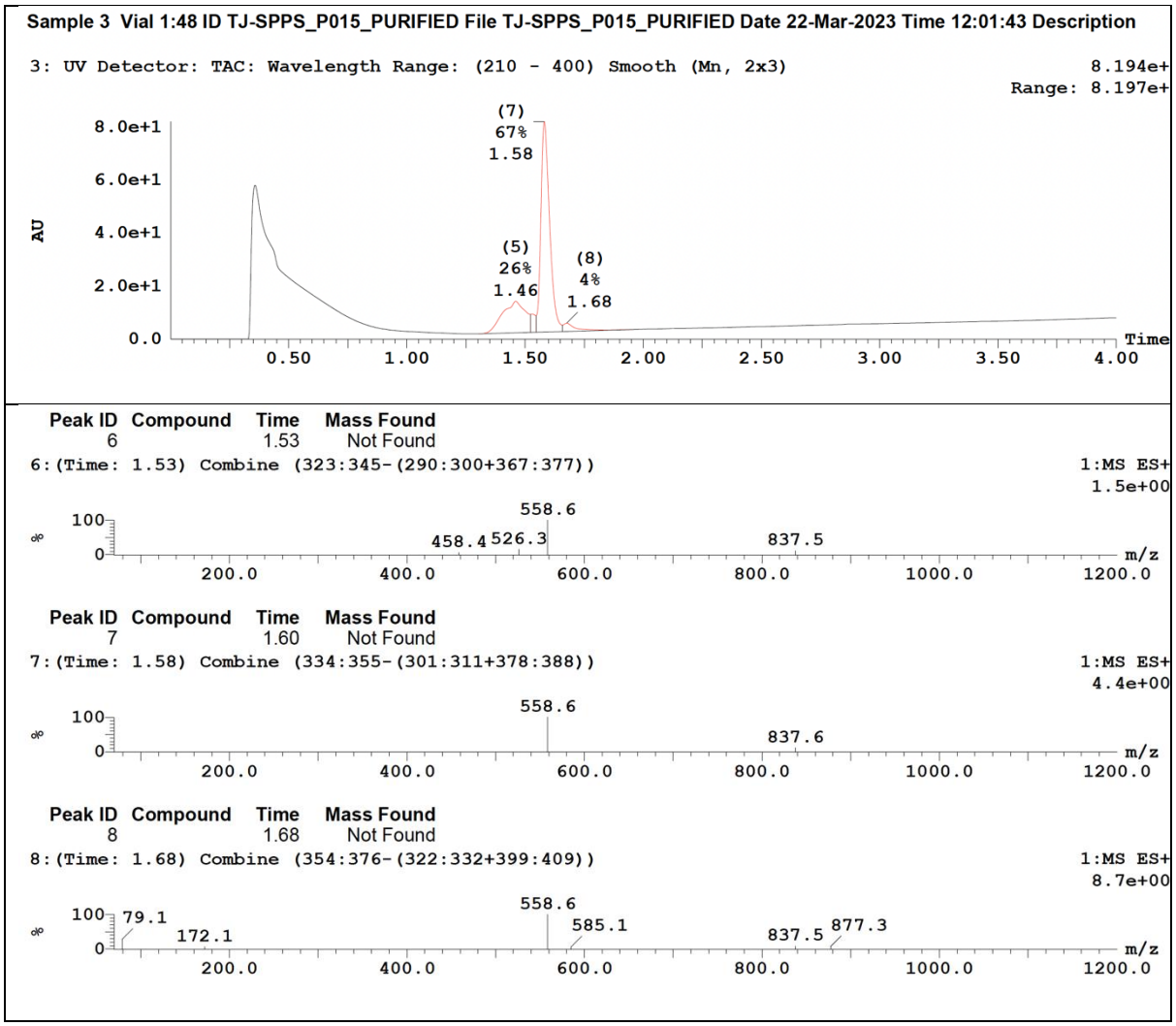
Appendix 31: SPPS Peptide Characterisation: P011



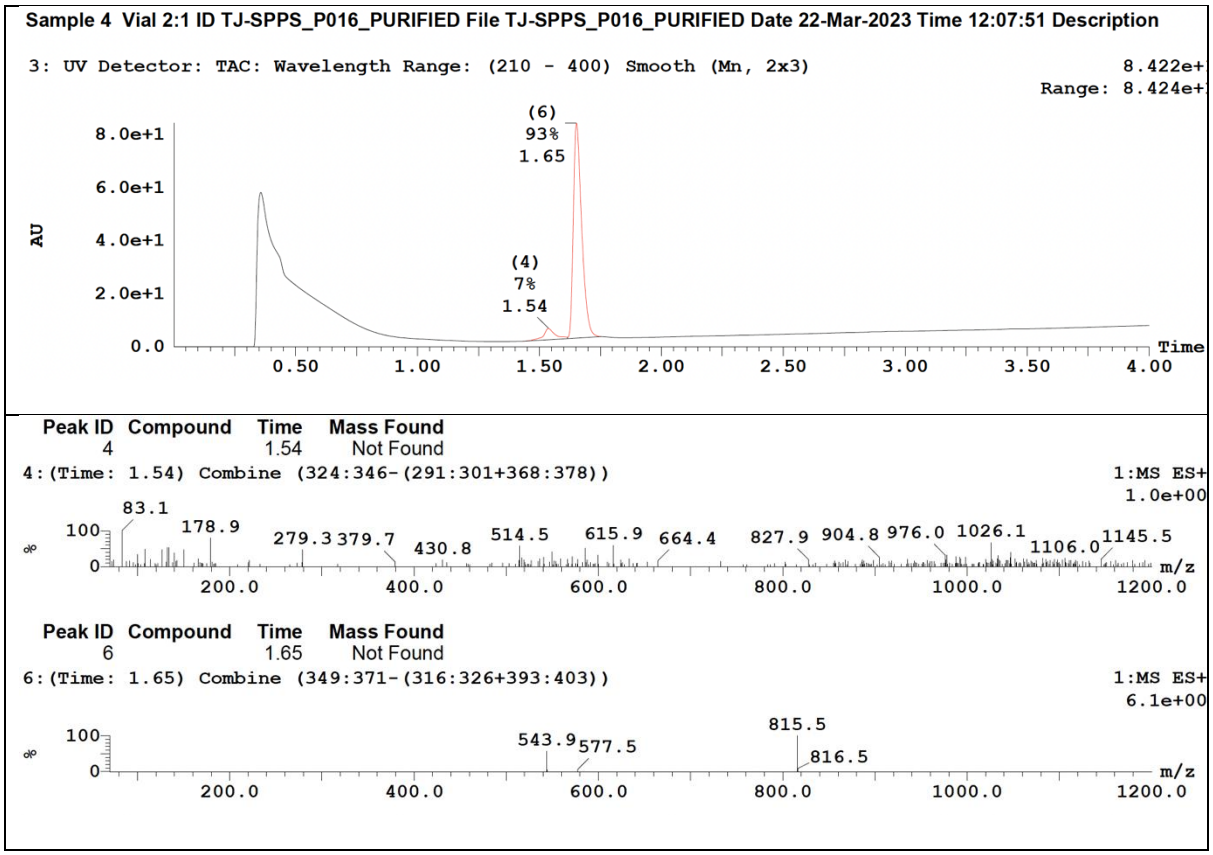
Appendix 32: SPPS Peptide Characterisation: P012



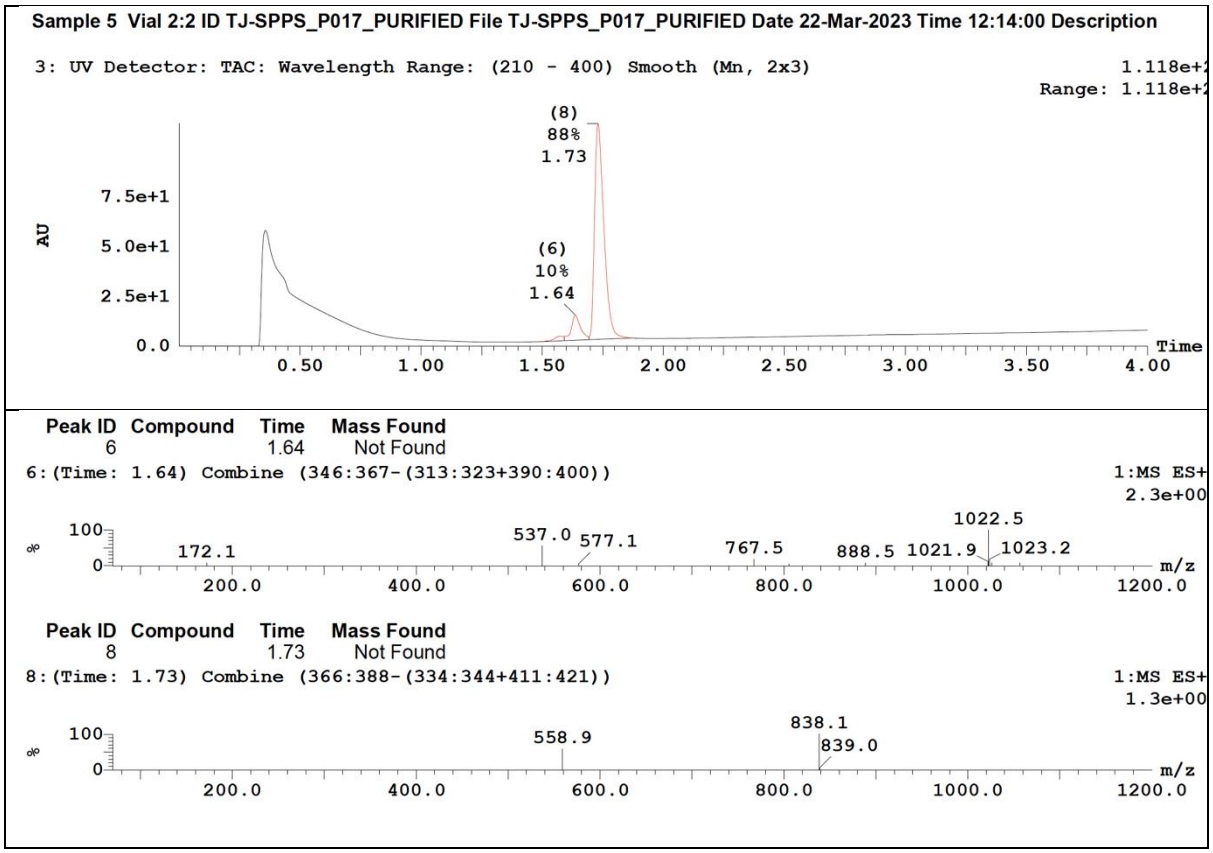
Appendix 33: SPPS Peptide Characterisation: P015



Appendix 34: SPPS Peptide Characterisation: P016



Appendix 35: SPPS Peptide Characterisation: P017



Appendix 36: RaPID DNA recovery tabulated data for PhotoRaPID v1.0

Proportion of Library Recovered Per Round Per Condition Panning Against hPADI2 (%):

	Irradiation Condition 1			Irradiation Condition 2			Irradiation Condition 3		
Round	***DNA input:***	***Strep-Negative:***	***Z-Positive:***	***DNA input:***	***E-Negative:***	***Z-Positive:***	***DNA input:***	***Strep-Negative:***	***Z-Positive:***
Round 1	1.10×10^{14}		0.00296						
Round 2	5.52×10^{10}	0.01502	0.00023		0.00012	0.00068		0.00003	0.00012
Round 3	2.17×10^{12}	0.00005	0.00006	1.78×10^{11}	0.00033	0.00004	3.18×10^{11}	0.00018	0.00041
Round 4	3.00×10^{11}	0.00003	0.01282	2.33×10^{11}	0.00027	0.01075	1.97×10^{11}	0.00010	0.00144
Round 5	5.49×10^{10}	0.00001	1.92906	9.88×10^{10}	0.00003	0.04463	5.59×10^{11}	0.00002	0.00796
Round 6	1.20×10^{11}	0.00029	33.25962	2.14×10^{10}	3.21184	96.22554	1.32×10^{11}	0.00139	3.42958

Appendix 37: RaPID DNA recovery tabulated data for PhotoRaPID v2.0

hPADI2 Recovery For Modified Irradiation Condition 2 (%):

	NNK4-8 Library:				*NNK6-12 Library:*			
Round	***DNA input:***	***Strep-Negative:***	***E-Negative:***	***Z-Positive:***	***DNA input:***	***Strep-Negative:***	***E-Negative:***	***Z-Positive:***
Round 1	3.96×10^{11}			0.25	1.83×10^{14}			0.27
Round 2	1.48×10^{12}	0.53	0.27	0.30	7.99×10^{11}	0.35	0.67	0.38
Round 3	1.25×10^{12}	0.05	0.22	0.30	2.81×10^{12}	0.16	0.15	0.32
Round 4	5.19×10^{12}	0.02	0.12	0.19	9.57×10^{12}	0.08	0.20	0.35
Round 5	3.23×10^{12}	0.06	0.20	0.26	4.35×10^{12}	0.02	0.14	0.09

Appendix 38: Supplementary SPR replicates tabulated data

			Ka			kd			KD			FINAL in nM	
Peptide ID	Isomer	ka	mean	sd	kd	mean	sd	KD	mean	sd		KD nM	Z/E
P001	Z	2.07E+06			1.78E-02			8.57E-09					
P001	Z	1.46E+06			1.98E-02			1.36E-08					
			1.77E+06	4.31E+05		1.88E-02	1.41E-03		1.11E-08	3.56E-09		11 +/ 3.6 nM	
P001	E	NOT POSSIBLE DUE TO STRANGE REF CHANNEL BINDING.				Top con = 1000 nM							

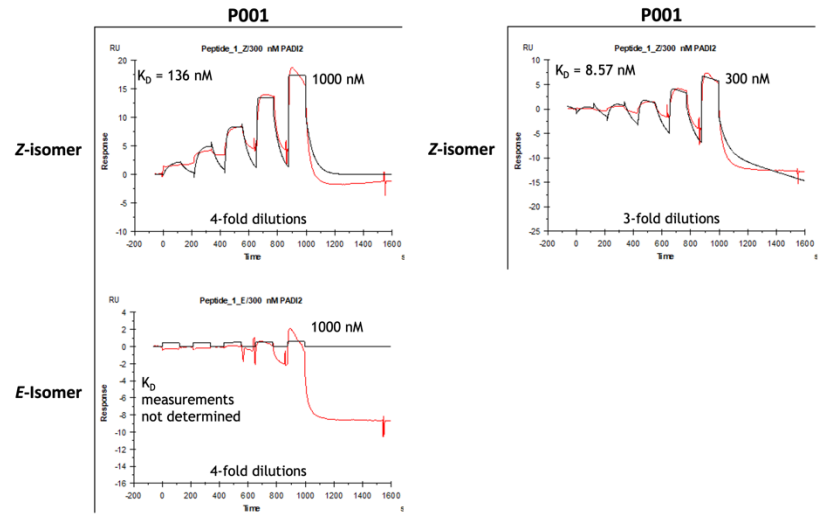
			Ka			kd			KD			FINAL in nM	
Peptide ID	Isomer	ka	mean	sd	kd	mean	sd	KD	mean	sd		KD nM	Z/E
P003	Z	9.30E+05			1.61E-03			1.73E-09					
P003	Z	3.06E+05			6.88E-04			2.25E-09					
P003	Z	1.57E+05			1.10E-03			7.00E-09					
			4.64E+05	4.10E+05		1.13E-03	4.62E-04		3.66E-09	2.90E-09		3.7 +/ 2.9 nM	5.83783784
P003	E	1.10E+05			4.11E-03			3.75E-08					
P003	E	2.02E+05			2.58E-03			1.27E-08					
P003	E	2.66E+05			3.85E-03			1.45E-08					
			1.93E+05	7.84E+04		3.51E-03	8.19E-04		2.16E-08	1.38E-08		21.6 +/ 13.8 nM	

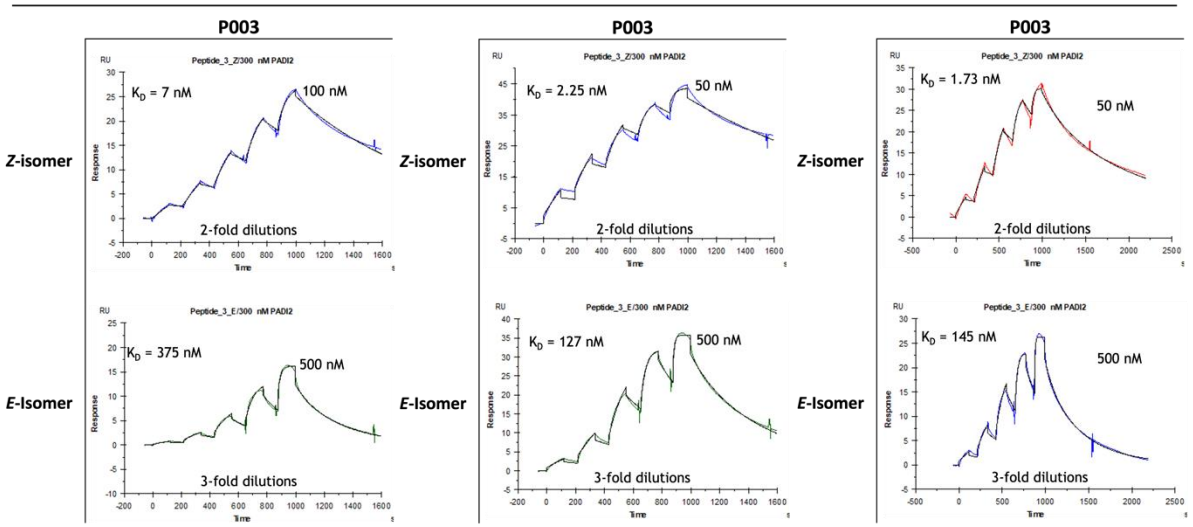
			Ka			kd			KD			FINAL in nM	
Peptide ID	Isomer	ka	mean	sd	kd	mean	sd	KD	mean	sd		KD nM	Z/E
P006	Z	4.89E+04			4.15E-03			8.48E-08					
P006	Z	5.90E+05			1.10E-02			1.87E-08					
P006	Z	9.73E+05			2.15E-02			2.21E-08					
			5.37E+05	4.64E+05		1.22E-02	8.74E-03		4.19E-08	3.72E-08		41.9 +/ 37 nM	28.8782816
P006	E	3.16E+04			4.30E-02			1.36E-06					
P006	E	4.25E+04			4.49E-02			1.06E-06					
			3.71E+04	7.71E+03		4.40E-02	1.34E-03		1.21E-06	2.12E-07		1210 +/- 212 nM	

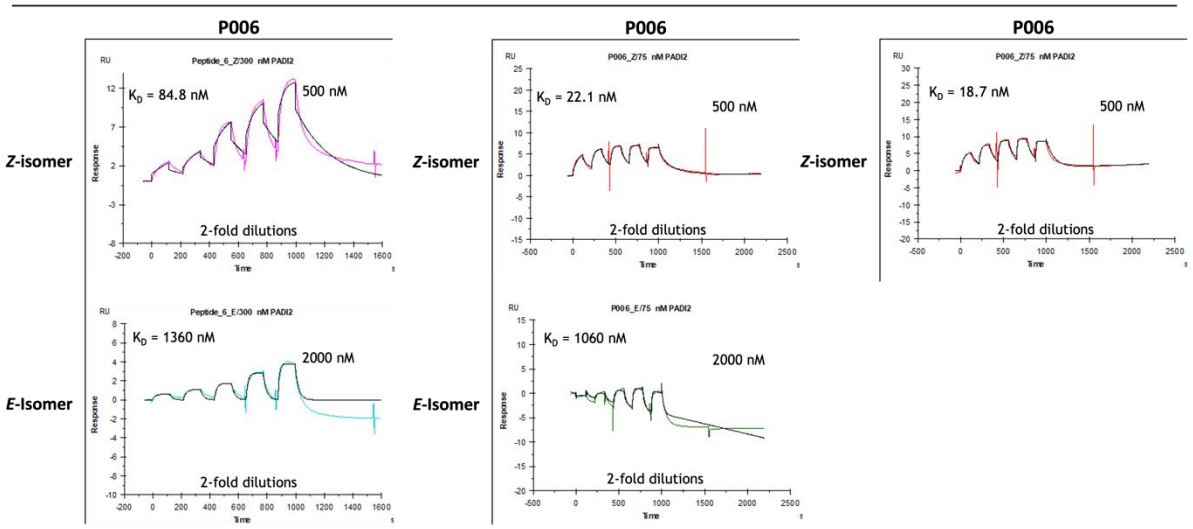
			Ka			kd			KD			FINAL in nM	
Peptide ID	Isomer	ka	mean	sd	kd	mean	sd	KD	mean	sd		KD nM	Z/E
P011	Z	9.87E+04			1.54E-02			1.56E-07					
P011	Z	3.47E+04			1.34E-02			3.85E-07					
			6.67E+04	4.53E+04		1.44E-02	1.41E-03		2.71E-07	1.62E-07		271 +/- 162 nM	18.4501845
P011	E	N/A			N/A			>2000				>2000	
P011	E	N/A			N/A			>5000				>5000	

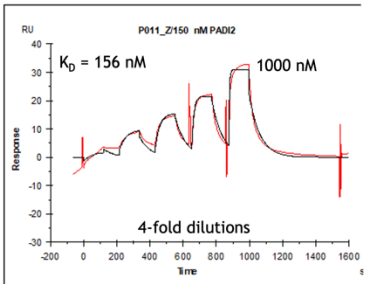
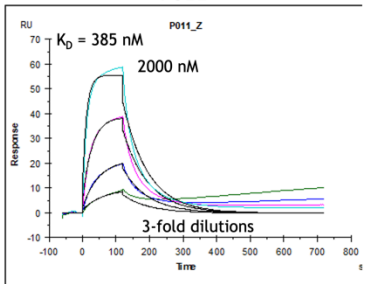
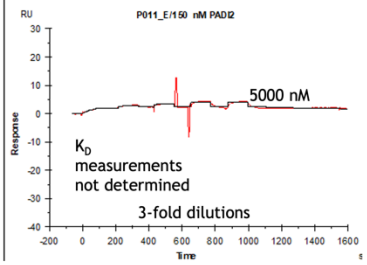
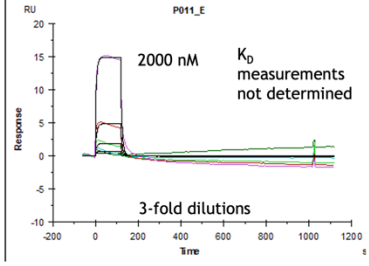
			Ka			kd			KD			FINAL in nM	
Peptide ID	Isomer	ka	mean	sd	kd	mean	sd	KD	mean	sd		KD nM	Z/E
P012	Z	AFFINITY			AFFINITY			7.18E-07				710 +/- 12 nM	1.43661972
	Z	AFFINITY			AFFINITY			7.01E-07					
									7.10E-07	1.20E-08		ONLY AFFINITY	
	E	AFFINITY			AFFINITY			1.03E-06					
P012	E	AFFINITY			AFFINITY			1.01E-06				1020 +/- 14 nM	
									1.02E-06	1.41E-08			
									1.02E+00	0.014142136			

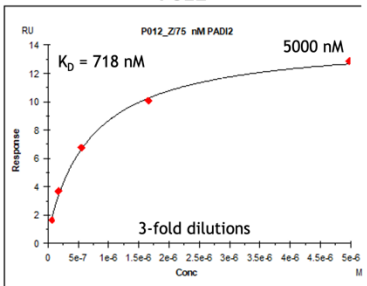
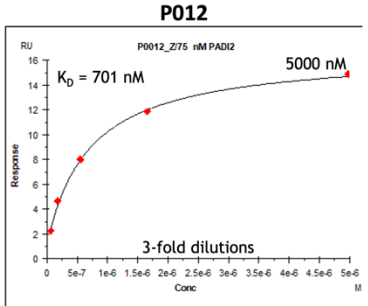
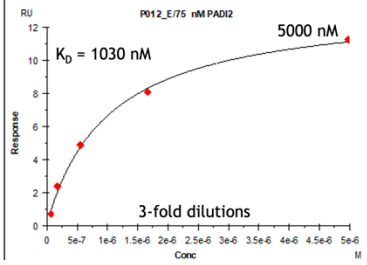
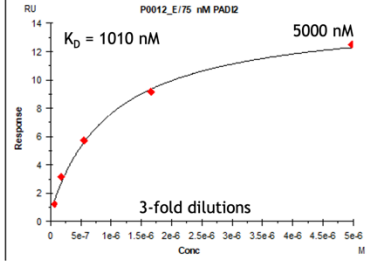
			Ka			kd			KD			FINAL in nM	
Peptide ID	Isomer	ka	mean	sd	kd	mean	sd	KD	mean	sd		KD nM	Z/E
P015	Z	2.74E+04			3.28E-03			1.20E-07	2.02E-07	1.10761E-07		202 nM +/- 111	24.7524752
P015	Z	AFFINITY			AFFINITY			3.28E-07					
	Z	1.26E+05	7.67E+04	6.97E+04	1.99E-02	1.16E-02	1.18E-02	1.58E-07					
P015	E	2.38E+04			1.46E-01			6.12E-06			Top con 5000 nM	outside range	
	E	2.03E+04			1.66E-01			8.15E-06			Top con 2000 nM	>5000	
			2.21E+04	2.47E+03		1.56E-01	1.41E-02		7.14E-06	1.44E-06			

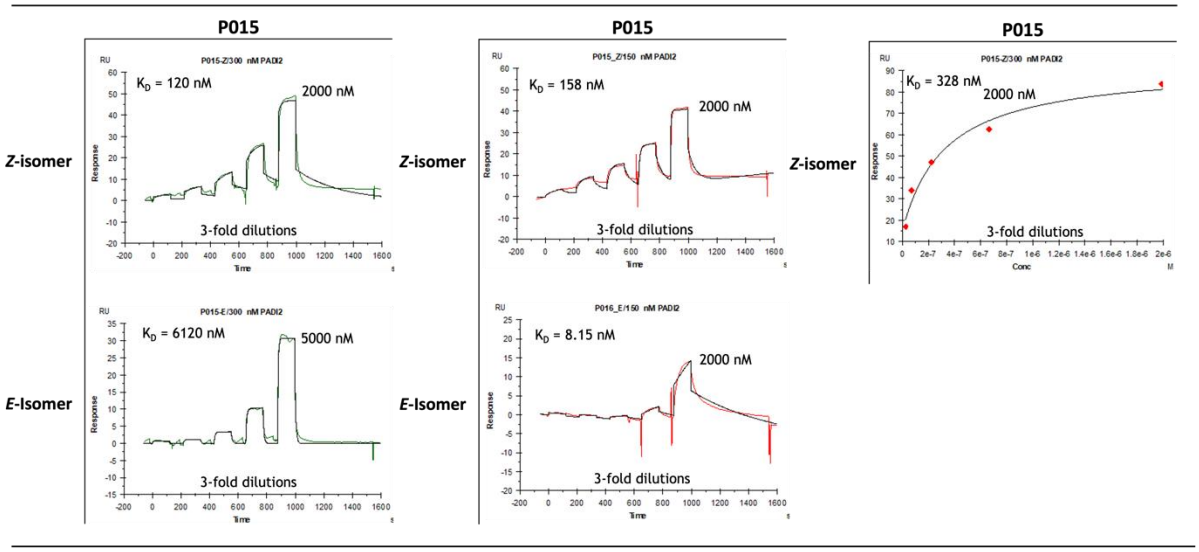

P001
Z-isomer
Peptide_1_Z/300 nM PADI2
K_D = 136 nM
1000 nM
4-fold dilutions
P001
Z-isomer
Peptide_1_Z/300 nM PADI2
K_D = 8.57 nM
300 nM
3-fold dilutions
E-Isomer
Peptide_1_E/300 nM PADI2
1000 nM
K_D measurements not determined
4-fold dilutions

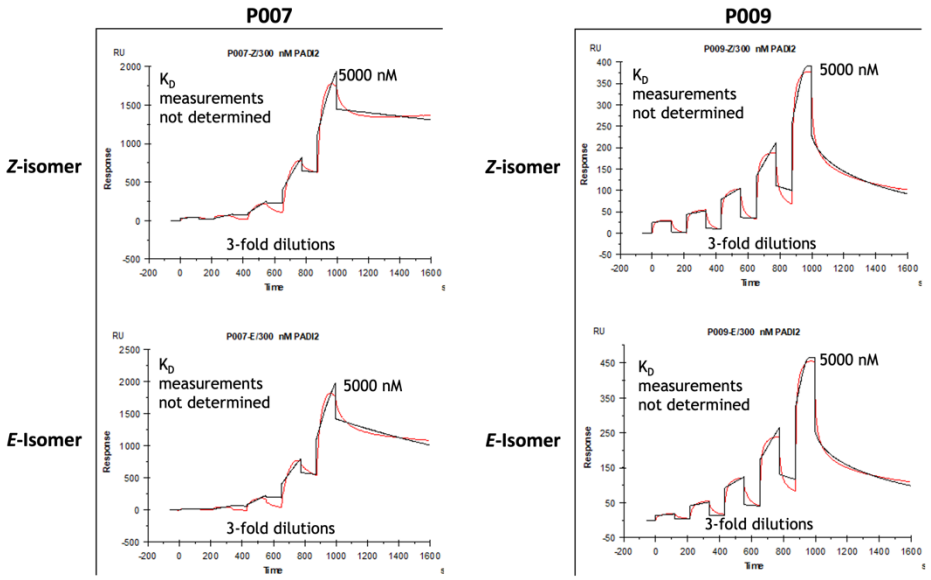

P003
Z-isomer
Peptide_3_Z/300 nM PADI2
K_D = 7 nM
100 nM
2-fold dilutions
P003
Z-isomer
Peptide_3_Z/300 nM PADI2
K_D = 2.25 nM
50 nM
2-fold dilutions
P003
Z-isomer
Peptide_3_Z/300 nM PADI2
K_D = 1.73 nM
50 nM
2-fold dilutions
E-Isomer
Peptide_3_E/300 nM PADI2
K_D = 375 nM
500 nM
3-fold dilutions
E-Isomer
Peptide_3_E/300 nM PADI2
K_D = 127 nM
500 nM
3-fold dilutions
E-Isomer
Peptide_3_E/300 nM PADI2
K_D = 145 nM
500 nM
3-fold dilutions


P006
Z-isomer
Peptide_6_Z/300 nM PADI2
K_D = 84.8 nM
500 nM
2-fold dilutions
P006
Z-isomer
P006_Z/75 nM PADI2
K_D = 22.1 nM
500 nM
2-fold dilutions
P006
Z-isomer
P006_Z/75 nM PADI2
K_D = 18.7 nM
500 nM
2-fold dilutions
E-Isomer
Peptide_6_E/300 nM PADI2
K_D = 1360 nM
2000 nM
2-fold dilutions
E-Isomer
P006_E/75 nM PADI2
K_D = 1060 nM
2000 nM
2-fold dilutions

	P011	P011
***Z*-isomer**		
***E*-Isomer**		

	P012	P012
***Z*-isomer**		
***E*-Isomer**		


P015
P015
P015
Z-isomer
Z-isomer
Z-isomer
E-Isomer
E-Isomer
P015-Z/300 nM PADI2
K_D = 120 nM
2000 nM
3-fold dilutions
P015_Z/150 nM PADI2
K_D = 158 nM
2000 nM
3-fold dilutions
P015-Z/300 nM PADI2
K_D = 328 nM
2000 nM
3-fold dilutions
P015-E/300 nM PADI2
K_D = 6120 nM
5000 nM
3-fold dilutions
P016_E/150 nM PADI2
K_D = 8.15 nM
2000 nM
3-fold dilutions
RU
Response
Time
Conc


P007
P009
Z-isomer
Z-isomer
E-Isomer
E-Isomer
P007-Z/300 nM PADI2
K_D measurements not determined
5000 nM
3-fold dilutions
P009-Z/300 nM PADI2
K_D measurements not determined
5000 nM
3-fold dilutions
P007-E/300 nM PADI2
K_D measurements not determined
5000 nM
3-fold dilutions
P009-E/300 nM PADI2
K_D measurements not determined
5000 nM
3-fold dilutions
RU
Response
Time

Appendix 39: Kingfisher protocol for automated PhotoRaPID Selections

Kingfisher Plate layout for the automated Irradiation protocol:

- Three columns are needed per one round of the selection. Three kingfisher protocol condition files have been generated to accommodate the irradiation steps.
- First the solution is switched to the Z-isomer via 3 mins irradiation (365 nm) using the LEDs and placed in C1-D position on the selection plate and 1.C2pnt2 kingfisher programme is run.
- Then the solution is removed from the plate and is placed into an epi - the ~20 µL is irradiated then using the LEDs for 3 minutes (495nm) to produce all E-library before the solution is transferred back into the C2-D position and 2.C2pnt2 programme is run.
- Finally, the solution is removed as above and irradiated to Z-again as above before 3.C2pnt2 is finally run - the elution strips with PCR mix are only needed in the final programme so this can be prepared whilst 1. and 2. programmes are running.

1.C2pnt2 Kingfisher Programme:

Steps data

NegTip Check		KingFisher Duo 12 tip comb	
	Pick-Up	Selection Tip Plate	(A) - Neg Tip
	Leave	Selection Tip Plate	(A) - Neg Tip
Neg Tip		KingFisher Duo 12 tip comb	
	Pick-Up	Selection Tip Plate	(A) - Neg Tip
	Neg 1 Collect	Selection Plate	(A) - Neg Bead 1
		Collect count	3
		Collect time [s]	2
	Neg 1 Mix	Selection Plate	(D) - RAPID
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Mixing time, speed	00:30:00, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	0
		Post-temperature	No
	Neg 1 Release	Selection Plate	(A) - Neg Bead 1
		Release time, speed	00:00:05, Fast
	Neg 2 Collect	Selection Plate	(B) - Neg Bead 2
		Collect count	3
		Collect time [s]	2
	Neg 2 Mix	Selection Plate	(D) - RAPID
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Mixing time, speed	00:30:00, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	2
		Post-temperature	No
	Neg 2 Release	Selection Plate	(B) - Neg Bead 2
		Release time, speed	00:00:05, Fast
	Neg 3 Collect	Selection Plate	(C) - Neg Bead 3
		Collect count	3
		Collect time [s]	2
	Neg 3 Mix	Selection Plate	(D) - RAPID
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Mixing time, speed	00:30:00, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	2
		Post-temperature	No
	Neg 3 Release	Selection Plate	(C) - Neg Bead 3
		Release time, speed	00:00:05, Fast
	Leave	Selection Tip Plate	(A) - Neg Tip

2.C2pnt2 Kingfisher Programme:

Steps data

Pos Tip Check		KingFisher Duo 12 tip comb	
	Pick-Up	Selection Tip Plate	(H) - Pos Tip
	Leave	Selection Tip Plate	(H) - Pos Tip
Pos Tip		KingFisher Duo 12 tip comb	
	Pick-Up	Selection Tip Plate	(H) - Pos Tip
	Pos Collection	Selection Plate	(E) - Pos Beads
		Collect count	3
		Collect time [s]	2
	Pos Mix	Selection Plate	(D) - RAPID
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Mixing time, speed	00:15:00, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	2
		Post-temperature	No
	ReleaseBeads1	Selection Plate	(E) - Pos Beads
		Release time, speed	00:00:05, Fast
	Leave	Selection Tip Plate	(H) - Pos Tip

3.C2pnt2 Kingfisher Programme:

Steps data

NegTip Check		KingFisher Duo 12 tip comb	
	Pick-Up	Selection Tip Plate	(A) - Neg Tip
	Leave	Selection Tip Plate	(A) - Neg Tip
Pos Tip Check		KingFisher Duo 12 tip comb	
	Pick-Up	Selection Tip Plate	(H) - Pos Tip
	Leave	Selection Tip Plate	(H) - Pos Tip
Pos Tip		KingFisher Duo 12 tip comb	
	Pick-Up	Selection Tip Plate	(H) - Pos Tip
	Pos Collection	Selection Plate	(E) - Pos Beads
		Collect count	3
		Collect time [s]	2
	Pos Mix	Selection Plate	(D) - RAPID
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Mixing time, speed	00:15:00, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	2
		Post-temperature	No
	Pos Wash 1	Selection Plate	(F) - Wash 1
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Mixing time, speed	00:00:15, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	2
		Post-temperature	No

	Pos Wash 2	Selection Plate	(G) - Wash 2
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Mixing time, speed	00:00:15, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	2
		Post-temperature	No
	Pos Wash 3	Selection Plate	(H) - Wash 3
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Mixing time, speed	00:00:15, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	3
		Collect time [s]	2
		Post-temperature	No
	Positive Bead PCR Elution	Pos PCR Mix	(A) - PCR Buffer
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Shake 1 time, speed	00:00:05, Bottom mix
		Shake 2 time, speed	00:01:00, Slow
		Shake 3 time, speed	00:00:30, Medium
		Heating during mixing	No
	End of step	Postmix	No
		Collect beads	No
		Post-temperature	No
	Leave	Selection Tip Plate	(H) - Pos Tip
Neg Tip Wash		KingFisher Duo 12 tip comb	
	Pick-Up	Selection Tip Plate	(A) - Neg Tip
	Neg Bead 3 Collection	Selection Plate	(C) - Neg Bead 3
	Beginning of step	Precollect	No
		Release beads	No
	Mixing / heating:	Mixing time, speed	00:00:10, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	4
		Collect time [s]	2
		Post-temperature	No

	Neg Bead 2 Collection	Selection Plate	(B) - Neg Bead 2
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Mixing time, speed	00:00:10, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	4
		Collect time [s]	4
		Post-temperature	No
	Neg Bead 1 Collection	Selection Plate	(A) - Neg Bead 1
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Mixing time, speed	00:00:10, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	5
		Collect time [s]	5
		Post-temperature	No
	Neg Bead Wash 1	Selection Tip Plate	(C) - Neg Wash 1
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Mixing time, speed	00:00:15, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	4
		Collect time [s]	5
		Post-temperature	No
	Neg Bead Wash 2	Selection Tip Plate	(D) - Neg Wash 2
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Mixing time, speed	00:00:15, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	4
		Collect time [s]	5
		Post-temperature	No
	Neg Bead Wash 3	Selection Tip Plate	(E) - Neg Wash 3
	Beginning of step	Precollect	No
		Release beads	Yes
	Mixing / heating:	Mixing time, speed	00:00:15, Slow
		Heating during mixing	No
	End of step	Postmix	No
		Collect count	4
		Collect time [s]	5
		Post-temperature	No

Appendix 40: RaPID post-selection script 1: Initial Data processing and generation of Z/N, E/Z and Z/E enrichment scores:

```
#Created: 20221101 - Jacksot

#Modified for hPADI2_C2pnt2_NNK6-12_NNK4-8_2nd_Attempt_Selection - 20230201 - Jacksot


import pandas as pd

import numpy as np


#***-----***-----***-----***


#*^^*^^*^^*^^*^^*^^*^^*^^*

#Loading the dataframes

dfNeg = pd.read_csv('hPADI2_C2pnt2_4_8_R5_Neg.csv')

dfE = pd.read_csv('hPADI2_C2pnt2_4_8_R5_E.csv')

dfZ = pd.read_csv('hPADI2_C2pnt2_4_8_R5_Z.csv')


#><><><><><><><><><><><><><><><><><><><><><><>

# Factoring Enrichment Score Ratio to be more representative to actual Negative Recovery From PCR

# - This is as we amp Pos + Neg to approx. equal skewing the P/N Ratio to be smaller amount of each peptide recovered in the neg.


# Input cDNA rt.Q.PCR Percentage Recoveries of the last round here WITHOUT the % sign.

#Positive Recovery Percentage:

ZqPCR_Recovery = 0.260

#Negative Recovery Percentage:

EqPCR_Recovery = 0.203
```

```
qPCR_Scaling_Factor = ZqPCR_Recovery / EqPCR_Recovery


#***-----***-----***-----***

#Data processing begins from here:



#MAKE SURE YOU ALSO MANUALLY ADD THE NO. OF READS TO RECALCULATE THE MORE ACCURATE ENRICH **^^**


print('Dataframe imports successful...')

#***-----***-----***-----***

dfNeg.dropna(inplace = True)

dfE.dropna(inplace = True)

dfZ.dropna(inplace = True)


#Truncating end of string to remove end of string SGSGSamber

dfNeg['Seq'] = dfNeg['Seq'].str.replace('SGSGSamber','')

dfE['Seq'] = dfE['Seq'].str.replace('SGSGSamber','')

dfZ['Seq'] = dfZ['Seq'].str.replace('SGSGSamber','')


#Filtering internal stop codons in string

dfNeg = dfNeg[~dfNeg['Seq'].str.contains('amber')]

dfNeg = dfNeg[~dfNeg['Seq'].str.contains('o')]

dfNeg = dfNeg[~dfNeg['Seq'].str.contains('fuck')]


dfE = dfE[~dfE['Seq'].str.contains('amber')]
```

```
dfE = dfE[~dfE['Seq'].str.contains('o')]

dfE = dfE[~dfE['Seq'].str.contains('fuck')]


dfZ = dfZ[~dfZ['Seq'].str.contains('amber')]

dfZ = dfZ[~dfZ['Seq'].str.contains('o')]

dfZ = dfZ[~dfZ['Seq'].str.contains('fuck')]


#% to decimal

#**^^**

#Getting a more accurate Percentage Enrichment manually formatted as a decimal replacing the original Enrich values:


Neg_total_reads = 258547

Z_total_reads = 272646

E_total_reads = 141040


dfNeg['Enrich'] = dfNeg[['Reads']].astype('float') / Neg_total_reads

dfE['Enrich'] = dfE[['Reads']].astype('float') / E_total_reads

dfZ['Enrich'] = dfZ[['Reads']] / Z_total_reads


#———————————————————————————————————————————————————————————————————————————
——

#Removing sequences less than 0.001% enriched

#dfNeg = dfNeg.loc[~((dfNeg['Enrich'] < 0.001))]

#dfE = dfE.loc[~((dfE['Enrich'] < 0.00001))]

#dfZ = dfZ.loc[~((dfZ['Enrich'] < 0.00001))]
```

```
#without cut-off rows = 43,3519 with 0.001% = 7,672

#————————————————————————————————————————
—

dfNeg.to_csv('1-Neg_processed.csv')

dfE.to_csv('1-E_processed.csv')

dfZ.to_csv('1-Z_processed.csv')



#VLOOKUP = pandas merge to put Enrich value of pos_Z and pos_E seq together where x = pos_Z and y = pos_E

dfPZE = pd.merge(dfZ,dfE, on ='Seq',how ='left')

dfPZE.dropna(inplace = True)


#divide Enrich_Int_x / Enrich_Int_y to give absolute enrichment score (Z/E)

dfPZE['Enrichment_Score'] = dfPZE[['Enrich_x']].div(dfPZE['Enrich_y'].values,axis=0)


#divide Enrich_Int_y / Enrich_Int_x to give absolute enrichment score (E/Z) - should not give scores as experiment was not designed to give E enriched peptides

dfPZE['E/Z_Enrichment_Score'] = dfPZE[['Enrich_y']].div(dfPZE['Enrich_x'].values,axis=0)


#Applying Scaling Factor from qPCR Recoveries to ensure Enrichment score is a more representative value of the pos/neg selectivity

#><><><><><><><><><><><><><><><><><><><><>

dfPZE['Enrichment_Score'] = dfPZE['Enrichment_Score'].astype(float)

dfPZE['Enrichment_Score'] = dfPZE['Enrichment_Score'] * qPCR_Scaling_Factor


#IF YOU WANT THE ENRICHMENT SCORE ROUNDED TO A WHOLE NUMBER

#dfPZE['Enrichment_Score'] = round(dfPZE['Enrichment_Score'] * qPCR_Scaling_Factor)
```

```
print('Enrichment Scores successfully scaled to the recoveries of the qPCR cDNA')

#Sort Enrichment by ascending
dfPZE = dfPZE.sort_values('Enrichment_Score', ascending = False)

dfPZE.to_csv('3-Z_over_E_Sequence_Enrichment_Scores.csv')

print('Sequence_Enrichment_Scores.csv processing successful...')

#-------------------------------------------------------------------------------------
#Seq that are only in Positive_Z and not in Positive_E AND visa versa
#CONFIRMED BY SPIKING IN SEQ MTESTRSKIYFCGSGSGSamber INTO THE ORF1 POSITION TO CHECK, IN NORMAL SELECTION  - IF RETURNS NOTHING THEN NO SEQ ARE FOUND IN POS ONLY - TRUE OUTPUT.
dfZPos_Only = dfZ.loc[~dfZ['Seq'].isin(dfE['Seq'])].copy()
dfZPos_Only.to_csv('2-Z_Seqs_Exclusive_to_Pos.csv')

dfEPos_Only = dfE.loc[~dfE['Seq'].isin(dfZ['Seq'])].copy()
dfEPos_Only.to_csv('2-E_Seqs_Exclusive_to_Pos.csv')

print('Seqs_Exclusive_to_Pos.csv processing successful...')

#-------------------------------------------------------------------------------------
#Reformatting decimal back to percentage

dfNeg['Enrich'] = dfNeg['Enrich']* 100.0
```

```
#dfNeg['Enrich'] = np.round(dfNeg['Enrich'], decimals = 5)

dfNeg['Enrich'] = dfNeg['Enrich'].apply(str) + '%'

dfNeg.to_csv('1-Neg_processed.csv')


dfE['Enrich'] = dfE['Enrich']* 100.0

#dfE['Enrich'] = np.round(dfE['Enrich'], decimals = 5)

dfE['Enrich'] = dfE['Enrich'].apply(str) + '%'

dfE.to_csv('1-E_processed.csv')


dfZ['Enrich'] = dfZ['Enrich']* 100.0

#dfZ['Enrich'] = np.round(dfZ['Enrich'], decimals = 5)

dfZ['Enrich'] = dfZ['Enrich'].apply(str) + '%'

dfZ.to_csv('1-Z_processed.csv')


dfPZE['Enrich_x'] = dfPZE['Enrich_x']* 100.0

#dfPZE['Enrich_x'] = np.round(dfPZE['Enrich_x'], decimals = 5)

dfPZE['Enrich_x'] = dfPZE['Enrich_x'].apply(str) + '%'


dfPZE['Enrich_y'] = dfPZE['Enrich_y']* 100.0

#dfPZE['Enrich_y'] = np.round(dfPZE['Enrich_y'], decimals = 5)

dfPZE['Enrich_y'] = dfPZE['Enrich_y'].apply(str) + '%'

dfPZE.to_csv('3-Z_over_E_Sequence_Enrichment_Scores.csv')


dfZPos_Only['Enrich'] = dfZPos_Only['Enrich']* 100.0

#dfZPos_Only['Enrich'] = np.round(dfZPos_Only['Enrich'], decimals = 5)

dfZPos_Only['Enrich'] = dfZPos_Only['Enrich'].apply(str) + '%'

dfZPos_Only.to_csv('2-Z_Seqs_Exclusive_to_Pos.csv')
```

```
dfEPos_Only['Enrich'] = dfEPos_Only['Enrich']* 100.0

#dfEPos_Only['Enrich'] = np.round(dfEPos_Only['Enrich'], decimals = 5)

dfEPos_Only['Enrich'] = dfEPos_Only['Enrich'].apply(str) + '%'

dfEPos_Only.to_csv('2-E_Seqs_Exclusive_to_Pos.csv')


#***-----***-----***-----***#***-----***-----***-----***#***-----***-----***-----***#***-----***-----
***-----***

print('--------------------------------------------------')

print("ALL OPERATIONS SUCCSSFULLY PARSED - END OF PROGRAMME")

print('--------------------------------------------------')

#***-----***-----***-----***#***-----***-----***-----***#***-----***-----***-----***#***-----***-----
***-----***
```

Appendix 41: RaPID post-selection script 2: Round-By-Round Enrichment Searching:

```
#Created: 20230203 - Jacksot

#for hPADI2_C2pnt2_NNK6-12_NNK4-8_2nd_Attempt_Selection - 20230203 - Jacksot


import pandas as pd

import numpy as np


#*^^*^^*^^*^^*^^*^^*^^*^^*

#Loading the dataframes

dfZN = pd.read_csv('3-Sequence Enrichment per Round of Candidate Enrichment Score Sequences.csv')

dfZE = pd.read_csv('3-Z_over_E_Sequence_Enrichment_Scores.csv')


#-----------------------------------------------------------------------------------------------

#Processing dfZN


# Removing percentages for dfZN

dfZN['Enrich_x'] = dfZN['Enrich_x'].str.replace('%','')

dfZN['Enrich_x'] = dfZN['Enrich_x'].astype(float)

#dfZN['Enrich_x'] = np.round(dfZN['Enrich_x'], decimals = 3)


#Populating all NaN values with 0

dfZN = dfZN.fillna(0)


#Renaming the ZN Enrichment Score Column for Clarity

dfZN = dfZN.rename(columns={"Enrichment_Score": "ZN_Enrichment_Score"})
```

```
#-------------------------------------------------------------------------------------------

#Processing dfZE

# Removing percentages for dfZE

dfZE['Enrich_x'] = dfZE['Enrich_x'].str.replace('%','')

dfZE['Enrich_x'] = dfZE['Enrich_x'].astype(float)

#dfZE['Enrich_x'] = np.round(dfZE['Enrich_x'], decimals = 3)


#Merging the Dataframe, filtering for desired columns and renaming columns for clarity

dfmerge = pd.merge(dfZE,dfZN, on ='Seq',how ='left')

dfmerge = dfmerge.fillna(0)


dfmerge =
dfmerge.filter(['ORF_x_x','Seq','Enrich_x_x','ZN_Enrichment_Score','E/Z_Enrichment_Score','Enrichme
nt_Score'], axis=1)


dfmerge = dfmerge.rename(columns={"ORF_x_x": "ORF","Enrich_x_x":
"Enrich","E/Z_Enrichment_Score": "EZ_Enrichment_Score","Enrichment_Score":
"ZE_Enrichment_Score"})


#Creating three df one with top 500 Enrich seqs and one with top 500 Z/E score seqs and one with top
500 E/Z socre seqs

df_Top_500_Enrich = dfmerge.sort_values('Enrich', ascending = False)

df_Top_500_ZE  = dfmerge.sort_values('ZE_Enrichment_Score', ascending = False)

df_Top_500_EZ  = dfmerge.sort_values('EZ_Enrichment_Score', ascending = False)


df_Top_500_Enrich = df_Top_500_Enrich.head(500)

df_Top_500_ZE  = df_Top_500_ZE.head(500)
```

```
df_Top_500_EZ     = df_Top_500_EZ.head(500)


#-------------------------------------------------------------------------------------------------

#Saving Table to Excel CSV

df_Top_500_Enrich.to_csv('1-- Top 500 Enriched sequences.csv')

df_Top_500_ZE.to_csv('2-- Top 500 ZE sequences.csv')

df_Top_500_EZ.to_csv('2--- Top 500 EZ sequences')
```

Appendix 42: RaPID post-selection script 3: Generating the Scatterplots, Z/E and Z/N

```
#Created: 20230203 - Jacksot

#for hPADI2_C2pnt2_NNK6-12_NNK4-8_2nd_Attempt_Selection - 20230203 - Jacksot


import pandas as pd

import numpy as np


#*^^*^^*^^*^^*^^*^^*^^*

#Loading the dataframes

dfZN = pd.read_csv('3-Sequence Enrichment per Round of Candidate Enrichment Score Sequences.csv')

dfZE = pd.read_csv('3-Z_over_E_Sequence_Enrichment_Scores.csv')


#-------------------------------------------------------------------------------------------------

#Processing dfZN


# Removing percentages for dfZN

dfZN['Enrich_x'] = dfZN['Enrich_x'].str.replace('%','')
```

```
dfZN['Enrich_x'] = dfZN['Enrich_x'].astype(float)

#dfZN['Enrich_x'] = np.round(dfZN['Enrich_x'], decimals = 3)


#Populating all NaN values with 0

dfZN = dfZN.fillna(0)


#Renaming the ZN Enrichment Score Column for Clarity

dfZN = dfZN.rename(columns={"Enrichment_Score": "ZN_Enrichment_Score"})


#-------------------------------------------------------------------------------------------

#Processing dfZE

# Removing percentages for dfZE

dfZE['Enrich_x'] = dfZE['Enrich_x'].str.replace('%','')

dfZE['Enrich_x'] = dfZE['Enrich_x'].astype(float)

#dfZE['Enrich_x'] = np.round(dfZE['Enrich_x'], decimals = 3)


#Merging the Dataframe, filtering for desired columns and renaming columns for clarity

dfmerge = pd.merge(dfZE,dfZN, on ='Seq',how ='left')

dfmerge = dfmerge.fillna(0)


dfmerge =
dfmerge.filter(['ORF_x_x','Seq','Enrich_x_x','ZN_Enrichment_Score','E/Z_Enrichment_Score','Enrichme
nt_Score'], axis=1)


dfmerge = dfmerge.rename(columns={"ORF_x_x": "ORF","Enrich_x_x":
"Enrich","E/Z_Enrichment_Score": "EZ_Enrichment_Score","Enrichment_Score":
"ZE_Enrichment_Score"})
```

```
#Creating three df one with top 500 Enrich seqs and one with top 500 Z/E score seqs and one with top
500 E/Z socre seqs

df_Top_500_Enrich = dfmerge.sort_values('Enrich', ascending = False)

df_Top_500_ZE     = dfmerge.sort_values('ZE_Enrichment_Score', ascending = False)

df_Top_500_EZ     = dfmerge.sort_values('EZ_Enrichment_Score', ascending = False)


df_Top_500_Enrich = df_Top_500_Enrich.head(500)

df_Top_500_ZE     = df_Top_500_ZE.head(500)

df_Top_500_EZ     = df_Top_500_EZ.head(500)


#-----------------------------------------------------------------------------------------------

#Saving Table to Excel CSV

df_Top_500_Enrich.to_csv('1-- Top 500 Enriched sequences.csv')

df_Top_500_ZE.to_csv('2-- Top 500 ZE sequences.csv')

df_Top_500_EZ.to_csv('2--- Top 500 EZ sequences')


#PLOT DATASET = TOP 500 MOST ENRICHED SEQUENCES - Z/E vs Enrichment


#NOTE: Data does not plot if one column variable contains NaN's

import seaborn as sns

import matplotlib.pyplot as plt


#White background of plot:

sns.set_theme(style='white')


#Dataframe to plot:

data = df_Top_500_Enrich
```

```
#Plot Parameters:

plt.scatter( x = data.Enrich, y= data.ZE_Enrichment_Score, alpha=.5, s=100, color="royalblue", edgecolor="black")


#Axes Limits:

#plt.ylim(0, 35)

#plt.xlim(0, 0.8)


#Axis and Title Name:

plt.title(label="Dataset - Top 500 Enriched Sequences \n Sequence Enrichment vs Z/E Enrichment Score", weight='bold')

plt.xlabel("Percentage Enrichment of Sequence")

plt.ylabel("Z/E Enrichment Score")


#Toggle if you want square outer grid and gridlines

#sns.despine(fig=None, ax=None, top=True, right=True, left=False, bottom=False, offset=None, trim=False)


plt.grid(True, which='both') #If you want the gridlines on

plt.minorticks_on()


#Legend

plt.legend(["TES-500"], loc ="upper right")


#Saving the plot to png - 600 dpi is publication quality

#plt.savefig("3-- Top 500 Enriched sequences - Sequence Enrichment vs ZE Enrichment Score.png", dpi=600, bbox_inches="tight")
```

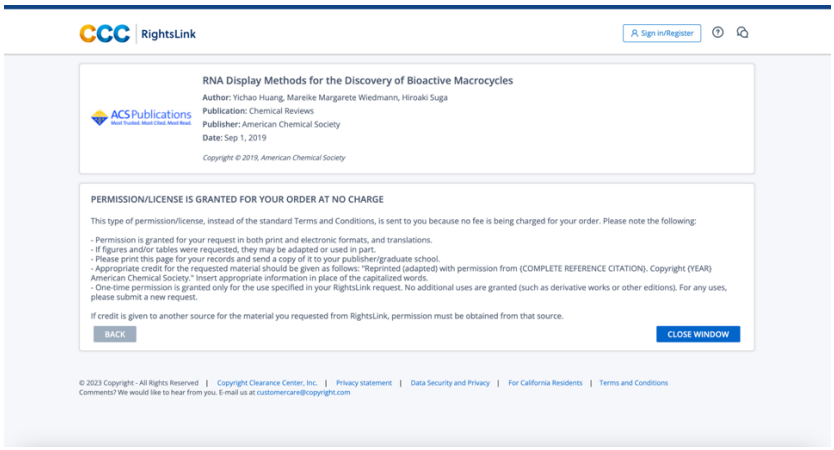
```
#PLOT DATASET = TOP 500 Z/E SEQUENCES - Z/E vs Enrichment

#NOTE: Data does not plot if one column variable contains NaN's

import seaborn as sns

import matplotlib.pyplot as plt


#White background of plot:

sns.set_theme(style='white')


#Dataframe to plot:

data = df_Top_500_ZE


#Plot Parameters:

plt.scatter( x = data.Enrich, y= data.ZE_Enrichment_Score, alpha=.5, s=100, color='darkorange', edgecolor="black")


#Axes Limits:

#plt.ylim(0, 35)

#plt.xlim(0, 0.8)


#Axis and Title Name:

plt.title(label="Dataset - Top 500 ZE Ranked Sequences \n Sequence Enrichment vs Z/E Enrichment Score", weight='bold')

plt.xlabel("Percentage Enrichment of Sequence")

plt.ylabel("Z/E Enrichment Score")


#Toggle if you want square outer grid and gridlines
```

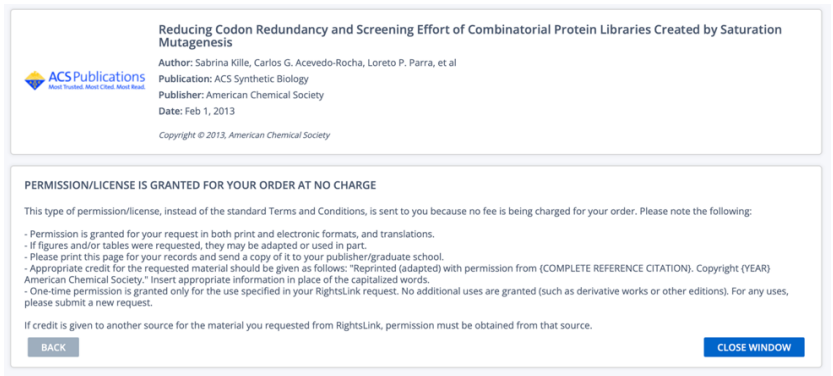
```
#sns.despine(fig=None, ax=None, top=True, right=True, left=False, bottom=False, offset=None, trim=False)


#Legend

plt.legend(["TZ/E-500"], loc ="upper right")


plt.grid(True, which='both') #If you want the gridlines on

plt.minorticks_on()


#Saving the plot to png - 600 dpi is publication quality

#plt.savefig("5-- Top 500 ZE Ranked sequences - Sequence Enrichment vs ZE Enrichment Score.png", dpi=600, bbox_inches="tight")


#Plotting all data together

import os

import seaborn as sns

import matplotlib.pyplot as plt


#White background of plot:

sns.set_theme(style='white')


#Dataframe to plot:

data  = df_Top_500_ZE

data1 = df_Top_500_Enrich


#Plot Parameters:
```

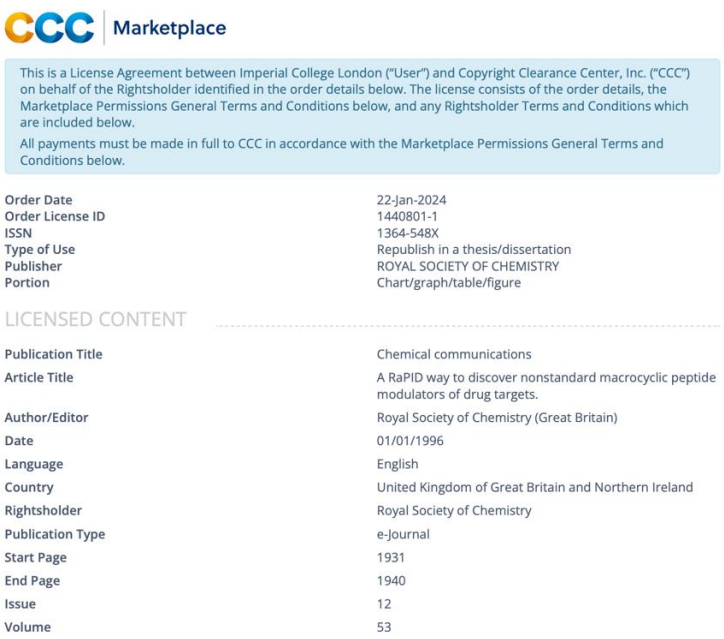
```
plt.scatter( x = data1.Enrich, y= data1.ZE_Enrichment_Score, alpha=.5, s=100, color='royalblue', edgecolor="black")

plt.scatter( x = data.Enrich, y= data.ZE_Enrichment_Score, alpha=.5, s=100, color='darkorange', edgecolor="black")


plt.title(label="Combined Datasets\n Top 500 Enriched Sequences and Top 500 ZE Ranked Sequences \n Sequence Enrichment vs Z/E Enichment Score:", weight='bold')

plt.xlabel("Percentage Enrichment of Sequence")

plt.ylabel("Z/E Enrichment Score")


#Toggle if you want square outer grid and gridlines

#sns.despine(fig=None, ax=None, top=True, right=True, left=False, bottom=False, offset=None, trim=False)


#Legend

plt.legend(["TES-500", "TZ/E-500"], loc ="upper right")


#If you want the gridlines on

plt.grid(True, which='both')

plt.minorticks_on()


#Saving the plot to png - 600 dpi is publication quality

#***-----***-----***-----***#***-----***-----***-----***#***-----***-----***-----***#***-----***-----***-----***

print('-------------------------------------------------')

print("ALL OPERATIONS SUCCSSFULLY PARSED - END OF PROGRAMME")

print('-------------------------------------------------')
```

#***-----***-----***-----***#***-----***-----***-----***#***-----***-----***-----***#***-----***-----***----***

Appendix 43: Image permissions:







Appendix 44: Software to generate some figure images.

Some figures were created with www.BioRender.com software.